

用遗传算法重构副热带高压特征指数的 非线性动力模型

洪梅^{1, 2} 张韧^{1, 2, 3} 吴国雄¹ 何金海³

1 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室, 北京 100029

2 解放军理工大学气象学院海洋与空间环境系, 南京 211101

3 南京信息工程大学江苏省气象灾害重点实验室, 南京 210044

摘 要 用遗传算法从离散时间序列资料中反演重构了非线性动力模型。首先, 用 Lorenz 系统的时间积分数据进行模型重构试验, 随后, 对十年平均的副热带高压形态指数时间序列进行动力模型参数反演和仿真预报试验。结果表明, 遗传算法具有的全局搜索和并行计算优势能够较为准确地描述和刻画副热带高压活动, 能对副高活动进行较为准确的描述与建模, 是诊断和预测副热带高压等复杂天气系统活动的一条有效途径。

关键词 遗传算法 Lorenz 系统 副热带高压

文章编号 1006-9895 (2007) 02-0346-07

中图分类号 P424

文献标识码 A

A Non-linear Dynamic System Reconstruction of the Subtropical High Characteristic Index Based on Genetic Algorithm

HONG Mei^{1, 2}, ZHANG Ren^{1, 2, 3}, WU Guo-Xiong¹, and HE Jin-Hai³

1 State Key Laboratory of Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

2 Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101

3 Nanjing University of Information Science and Technology, Jiangsu Key Laboratory of Meteorological Disaster, Nanjing 210044

Abstract By using Genetic Algorithm (GA), a non-linear dynamic model is retrieved and reconstructed from T106 numerical forecast fields time series data. From this non-linear dynamic model, the dynamic significance is found, and the model is used as the forecast model to diagnose and predict more accurately.

After introducing the basal idea of the dynamic system reconstruction and elementary flow of GA, the Lorenz chaos dynamic model is first retrieved for examining the reconstruction effect of GA. Using the known Lorenz chaos dynamic system to prove that the calculational precision of GA is higher and the retrieved model parameters are closer to reality, the authors objectively and accurately reconstruct the Lorenz chaos dynamic system model.

Then the authors retrieve and reconstruct an actual dynamic model of the Pacific subtropical high characteristic index from a time series of the 10-year averaged Pacific subtropical high characteristic index. Using the 10-year average (1958-1997) subtropical high ridge index of pentad, the subtropical high area index of pentad and the subtropical high western extension index of pentad as the time series of observational information, the authors reconstruct

收稿日期 2005-09-29, 2006-02-23 收修定稿

资助项目 国家自然科学基金项目资助 40375019, 江苏省气象灾害重点实验室开放课题 KLME0507, 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室开放课题

作者简介 洪梅, 女, 1982 年出生, 在读研究生, 主要研究方向为热带海洋天气与海洋卫星遥感应用。E-mail: flowerainhm@126.com

the non-linear dynamic model of the Pacific subtropical high characteristic index and get the retrieved parameters, then eliminate the illusive parameters which are take up less proportion. Finally the authors get the reconstructed non-linear dynamic model of the Pacific subtropical high characteristic index. On this basis an integral forecasting examination is processed. By initializing actual data which are got from the series of subtropical high characteristic index, the numerically integral results of model are gotten. Comparing with actual data, the model has the characters of longer forecasting cycle and better forecasting effect. In the paper, the authors also give the comparison between the integral results and the actual results.

The simulated forecast results of the retrieved dynamic model show that for GA has the connotative advantages of global optimization and parallel calculation, the dynamic model reconstruction method by GA can accurately describe and simulate the subtropical high, and can be used for diagnosing and predicting such complicated weather system as the subtropical high. The efficiency and veracity of the retrieved model parameter are better than the conventional methods, the forecasting temporal effect and maneuverability are better than the conventional statistical methods, such as Neural Net. A new method is found to diagnose and predict the dynamics of complicated weather system especially in which the dynamic model cannot be got.

Key words genetic algorithm, Lorenz system, subtropical high

1 引言

副热带高压是影响、制约夏季东亚天气气候的重要天气系统, 副高活动具有明显的非线性和非周期性, 由于描述副高活动的动力模型很难准确构建, 因此副高预测 (尤其是副高的中、长期预测) 一直是气象学研究中的一个难题^[1]。Takens^[2]提出的重构相空间理论, 其基本思想认为系统中任一分量的演化是由与之相互作用的其他分量所决定, 因此这些相关分量的信息隐含在任一分量的发展过程中, 这样就可从一些仅与时间相关的观测数据中提取和恢复系统原来的规律。上述思想为从观测资料中反演副高动力模型提供了一条客观合理的方法途径。

目前, 常用的动力模型重构方法主要有传统的时延相空间重构方法和动力模型参数反演方法。前者通过分维计算和寻找最优嵌入维与时延数, 得到要素时间序列演变的动力学模型, 如田纪伟等^[3]用延迟时间模型讨论了海表温度在嵌入相空间中的动力系统重构; 魏恩泊等^[4]利用混沌系统单物理参量时间序列的重构相空间点条件概率密度及其标准化分析, 提出了一种新的隐含变量标准化资料的反演方法。但是, 上述方法由于仅考虑了单个要素与其时延序列的信息, 包含的独立信息较少, 因此在反演重构复杂非线性动力系统时表现出较大的局限性。另一种途径则是基于广义的常微动力系统模型, 以模型输出与实际资料间的误差极小为目标约

束, 进行模型参数的优化搜索和动力模型重构^[5]。黄建平等^[6]曾用经典的最小二乘估计讨论了从观测资料中重构非线性动力模型的途径, 并从数值积分序列中较好地重构恢复了 Lorenz 系统, 但该方法未应用于实际天气系统的重构。不过, 目前常规的优化搜索方法 (如最小二乘估计) 大多存在着参数空间单向搜索的效率较低和误差收敛局部极小等缺陷^[7]。

遗传算法是近年发展起来, 并得到广泛应用的一种全局优化算法, 其特点在于全局搜索和并行计算, 因而具有很好的参数优化能力和误差收敛速度^[8]。基于此, 本文引入遗传算法对模型参数的搜索优选过程进行改进, 首先对 Lorenz 模型进行试验, 以验证方法的科学性和实用性, 其后基于副高活动指数的观测资料, 进行副高动力模型的参数反演和模型预测试验。

2 动力模型反演的基本思想

设任一非线性系统随时间演变的物理规律可表示为

$$\frac{dq_i}{dt} = f_i(q_1, q_2, \dots, q_i, \dots, q_N), \quad i = 1, 2, \dots, N, \quad (1)$$

函数 f_i 为 $q_1, q_2, \dots, q_i, \dots, q_N$ 的广义非线性函数, N 为状态变量个数, 一般可根据动力系统吸引子的复杂性 (可通过计算其分维数来衡量) 来确定^[9]。方程 (1) 的差分形式可写成:

$$\frac{q_i^{(j+1)\Delta t} - q_i^{(j-1)\Delta t}}{2\Delta t} = f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, \dots, q_i^{j\Delta t}, \dots, q_N^{j\Delta t}),$$

$$j = 2, 3, \dots, M-1, \quad (2)$$

其中, M 为观测资料的时间序列长度, 模型参数和系统结构可以通过反演计算从观测数据中获取。 $f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, \dots, q_i^{j\Delta t}, \dots, q_N^{j\Delta t})$ 为未知非线性函数, 设其有 G_{jk} 个包含变量 q_i 的函数展开项和对应的 p_k 个参数 (其中, $i = 1, 2, \dots, N, j = 1, 2, \dots, M, k = 1, 2, \dots, K$) 可设为:

$$f_i(q_1, q_2, \dots, q_n) = \sum_{k=1}^K G_{jk} p_k,$$

式 (2) 的矩阵形式为 $\mathbf{D} = \mathbf{GP}$, 其中:

$$\mathbf{D} = \begin{Bmatrix} d_1 \\ d_2 \\ \dots \\ d_M \end{Bmatrix} = \begin{Bmatrix} \frac{q_i^{3\Delta t} - q_i^{\Delta t}}{2\Delta t} \\ \frac{q_i^{4\Delta t} - q_i^{2\Delta t}}{2\Delta t} \\ \dots \\ \frac{q_i^{M\Delta t} - q_i^{(M-2)\Delta t}}{2\Delta t} \end{Bmatrix}, \quad (3-1)$$

$$\mathbf{G} = \begin{Bmatrix} G_{11}, G_{12}, \dots, G_{1K} \\ G_{21}, G_{22}, \dots, G_{2K} \\ \dots \\ G_{M1}, G_{M2}, \dots, G_{MK} \end{Bmatrix}, \quad (3-2)$$

$$\mathbf{P} = \begin{Bmatrix} p_{i1} \\ p_{i2} \\ \dots \\ p_{iK} \end{Bmatrix}. \quad (3-3)$$

上述广义的未知方程组的系数项可通过实际观测数据予以反演确定。给定一个向量 $\mathbf{D}$, 要求一个向量 $\mathbf{P}$, 使上式满足。对于 q_i 而言, 这是一个非线性系统, 但是我们换个角度, 对 $\mathbf{P}$ 而言 (即拿 $\mathbf{P}$ 当作未知数), 上式正好是一个线性系统, 可以用经典的最小二乘估计, 使残差平方和 $S = (\mathbf{D} - \mathbf{GP})^T (\mathbf{D} - \mathbf{GP})$ 最小, 进而得到正则方程 $\mathbf{G}^T \mathbf{GP} = \mathbf{G}^T \mathbf{D}$ 。由于 $\mathbf{G}^T \mathbf{G}$ 经常是奇异矩阵, 所以可将其特征值与特征向量求出, 剔除其中为 0 的那些, 剩下 K 个 $\lambda_1, \lambda_2, \dots, \lambda_i$ (λ_i 表示第 i 个特征向量) 组成对角矩阵 $\mathbf{\Lambda}_i$ 与相应的 K 个特征向量组成的特征矩阵 $\mathbf{U}_L$ 。

$$\mathbf{H} = \mathbf{U}_L \mathbf{\Lambda}_i^{-1} \mathbf{V}_L^T,$$

$$\mathbf{V}_L = \frac{\mathbf{G} \mathbf{U}_i}{\lambda_i},$$

其中 $\mathbf{U}_i$ 表示第 i 个特征矩阵, 再求 $\mathbf{P} = \mathbf{HD}$, 求出参数 $\mathbf{P}$ 。

基于上述方法途径, 即可反演确定出非线性动力系统诸系数, 进而得到对应观测数据序列的非线性动力学方程组。

3 遗传算法的基本流程

常规的参数估计方法 (如最小二乘估计) 在对模型参数进行优化时, 通常取参数空间中的模型计算值与实际值的误差最小二乘作为约束, 然后按顺序单向搜索, 以获取最佳参数 (对应最小二乘误差极小), 通常需要遍历整个参数空间, 且由于误差梯度收敛的局限性和对初始解 (通常为随机给出的参数初值) 和误差函数极小值域位置的依赖性, 其参数估计易陷入局部最优、而非全局最优。而遗传算法的全局搜索和并行计算优势正是对上述常规方法缺陷的一个很好的补充和完善。为此, 我们引入遗传算法进行模型参数的优化反演, 即以上述残差平方和 S 最小作为约束条件, 用遗传算法在参数空间中进行最优参数搜索。设参数组成的种群为 $\mathbf{P}$, 取残差平方和 S 为适应度函数, 具体操作步骤如下:

(1) 编码: 设 N 个样本被分为 c 类: $M_1, M_2, M_3, \dots, M_c$, 由上节可知, 对于第 i 个方程对应着 p_k ($k=1, 2, 3, \dots, K$) 个参数, 那么对于包含有多个方程的方程组, 其参数就组成了参数矩阵 $\mathbf{P} = (p_k)$, 其中共有 $N \times c$ 个元素。编码方案如下: 将每个 p_k 用三位二进制编码, 组成 $N \times c$ 个基因链, 最后将所有基因链拼接在一起形成一个染色体, 按照编码组合规则, 染色体中共有 $(3N \times c + c)$ 个基因。此时 p_k 对应的二进制串转换关系为

$$p_k = \frac{\text{decimal}(k)_2}{2^3 - 1},$$

其中, $(k)_2$ 是 p_k 对应的二进制串的值。

(2) 初始化群体生成: 初始化群体的生成以基因编码为基础, 基因是描述生物染色体的最基本单位, 染色体也称为个体。基因用一定的代码表示, 可以是数字串也可以是字符串, 若干代码组成基因链, 即所谓的染色体。基因代码是基因操作的基本单元, 当类数 c 给定时, 可随机生成 p_k 组成参数矩阵 $\mathbf{P} = (p_k)$, 其中 p_k 满足 (1) 编码方案。这样根据编码可得到多个个体。

(3) 适应值计算: 将每个个体的残差平方和 S_m 作为目标函数值, S_m 越小, 个体的适应值就越高。因此, 个体适应值可取 $f_i = -S_m$, 总的适应值为

$$F = \sum_{i=1}^n f_i。$$

(4) 父本选择: 计算每个个体的选择概率 $pp_i = f_i/F$ 及累积概率 $q_i = \sum_{j=1}^i pp_j$, 选择方法采用旋转花轮法, 旋转 m 次即可选出 m 个个体。在计算机上实现的步骤为: 产生 $[0, 1]$ 的随机数 r , 若 $r < q_1$, 则第一个个体入选, 否则, 第 i 个个体入选, 且 $q_{i-1} < r < q_i$ 。

(5) 交叉操作:

① 对每个个体产生 $[0, 1]$ 间的随机数 r , 若 $r < p_c$ (p_c 为选定的交叉概率), 则该个体参加交叉操作, 如此选出交叉操作后的一组后, 随机配对。

② 对每一配对, 产生 $[1, (3n+c)]$ 间的随机数以确定交叉的位置。

(6) 变异操作:

① 对基因编码中的前 c 位, 采用如下变异方法:

(a) 随机选择一个个体和一个随机整数 h ($1 \leq h \leq c$);

(b) 随机选取 h 个样本 $x_{i_1}, x_{i_2}, \dots, x_{i_j}, \dots, x_{i_h}$;

(c) 随机产生 h 个不同的整数 r_j ($1 \leq r_j \leq h$);

(d) 将该个体的第 r_j 个基因换成 x_{i_j} 。

② 对基因编码中第 c 位以后的基因, 采用如下变异方法:

(a) 对每一串中的每一位产生 $[0, 1]$ 间的随机数 r , 若 $r < p_m$ (p_m 是变异概率), 则该位变异;

(b) 实现变异操作, 即将原串中的 0 变为 1, 1 变为 0, 如果新的个体数达到 N 个, 则已形成一个新群体, 转向第 (3) 步; 否则转向第 (4) 步继续遗传操作。

(7) 终止操作: 在遗传算法中, 起止条件往往是人为给定的, 根据本问题的特点, 取终止条件为最优目标函数值 $S_m \leq \epsilon$ (本文取 $\epsilon = 0.2\%$)。

基于遗传算法的上述操作步骤, 可实现广义非线性动力模型参数的全局性优化反演, 遗传算法搜索得到的目标函数最小的进化种群即为非线性动力方程组的最优系数解, 即遗传算法反演的动力系统最优化参数。另外, 需要说明的是, 遗传算法是对一个种群进行操作, 种群中一般包含若干个体, 每个个体都是参数的一个解, 最后得到的系统参数从进化过程最后一代的若干优化解中产生, 进而保

证了系统参数反演的全局最优化特性。

4 Lorenz 混沌动力系统反演

为验证遗传算法的模型参数反演效果和精度, 我们首先用如下 Lorenz 混沌动力系统进行试验:

$$\frac{dX}{dt} = 10Y - 10X,$$

$$\frac{dY}{dt} = 28X - Y - XZ,$$

$$\frac{dZ}{dt} = XY - \frac{8}{3}Z.$$

首先, 用 Longkuta 方法对上述方程组进行数值积分, 将每一步的积分结果看成是一次的观测记录, 于是得到 X 、 Y 、 Z 时间序列^[10]。随后, 将拟反演的非线性模型设为如下广义二阶常微分程组形式:

$$\begin{aligned} \frac{dX}{dt} = & a_1X + a_2Y + a_3Z + a_4X^2 + a_5Y^2 + a_6Z^2 + \\ & a_7XY + a_8XZ + a_9YZ, \end{aligned}$$

$$\begin{aligned} \frac{dY}{dt} = & b_1X + b_2Y + b_3Z + b_4X^2 + b_5Y^2 + b_6Z^2 + \\ & b_7XY + b_8XZ + b_9YZ, \end{aligned}$$

$$\begin{aligned} \frac{dZ}{dt} = & c_1X + c_2Y + c_3Z + c_4X^2 + c_5Y^2 + c_6Z^2 + \\ & c_7XY + c_8XZ + c_9YZ. \end{aligned}$$

显然, 方程组中既包含了 Lorenz 系统中的真实项, 也包含了一些虚假项。我们需要做的工作即是利用 X 、 Y 、 Z “观测” 资料序列, 反演重构方程组中的真实参数值, 剔除虚假项。求解过程为:

(1) 首先画出残差平方和 S 的三维截面图 (图略), 以辅助观察最小适应值的参数区域。

(2) 随机产生方程组中 $a_1, a_2, \dots, a_9, b_1, b_2, \dots, b_9, c_1, c_2, \dots, c_9$ 参数系列的初始种群值, 每个种群 (参数) 包含 9 个个体, 每个个体代表一个解, 其中 a_1 参数初始种群个体分布如图 1 (其余参数略)。

(3) 经 15 次遗传迭代后, 得到的寻优结果如图 2。可以看出, 经 15 次迭代后, a_1 初始种群只剩下 5 个个体, 其中 4 个在遗传迭代过程中已被淘汰, 分别计算剩下的 5 个个体的适应度函数值, 并选择适应度函数值最小的个体作为最优解 (为 -10.01)。

(4) 绘出遗传算法的寻优动态图 (图略), 结果

表明, 约经 13 代遗传迭代操作后的解已逼近真实值。

同理, 其余 $a_1, a_2, \cdots, a_9, b_1, b_2, \cdots, b_9, c_1, c_2, \cdots, c_9$ 参数系列亦可按上述计算方案和操作步骤求得最优解参数。为检验遗传算法的计算效率和反演参数的准确率, 我们还采用最小二乘估计方法进行了 Lorenz 模型的参数反演, 并将两种方法的参数反演结果进行比较。

对比结果表明, 两种方法的参数反演值与模型真值均很接近, 最小二乘估计的反演结果与模型真值最大误差为 19%, 而遗传算法反演结果的最大误差仅 3%。在虚假项系数上, 最小二乘估计的反演结果比模型真值 (0) 小 3 个量级以上, 而遗传算法的反演结果则小 5 个量级以上。因此, 遗传算法的

反演结果较传统最小二乘估计结果更精确地接近模型真值。

为定量比较模型各项对系统演变的相对贡献大小, 我们计算了各项相对方差贡献, 计算公式为:

$$R_i = \frac{1}{m} \sum_{j=1}^m (T_i^j / \sum_{i=1}^9 T_i^j), \quad i = 1, 2, \cdots, 9,$$

其中, $m = 1000$ 为资料序列的长度, $T_i = a_1 X, a_2 Y, \cdots, a_9 YZ$ 为模型方程中的各项。用两种方法分别得出模型各项的 R_i 值 ($b、c$ 系数序列的方差贡献比同理计算给出)。

计算结果表明, 反演出的模型真实项的方差贡献占有较大比重, 而虚假项的方差贡献几乎为零, 对比发现, 遗传算法反演的各项方差贡献较最小二乘估计结果更合理, 虚假项的方差贡献更趋于零。基于上述计算分析结果, 根据反演参数的量级大小剔除模型虚假项, 进而得到如下 Lorenz 模型:

$$\begin{aligned} \frac{dX}{dt} &= a_1 Y - a_2 X, \\ \frac{dY}{dt} &= b_1 X - b_2 Y - b_8 XZ, \\ \frac{dZ}{dt} &= c_3 Z + c_7 XY. \end{aligned}$$

两种方法反演得到的方程组中各项的参数如表 1。

对比结果表明, 遗传算法的计算精度更高, 反演的模型系数更接近真实情况, 客观准确地重构出了 Lorenz 动力系统模型。Lorenz 混沌动力系统的反演结果表明, 遗传算法不仅计算快捷、操作方便, 而且计算精度和反演效果也优于常规的最小二乘估计。由于遗传算法能够有效地反演重构复杂的混沌动力系统, 因此, 下面将其应用于从观测资料中反演重构副高特征指数的动力系统模型。

表 1 两种反演方法得到的模型参数比较
Table 1 The comparison of models' parameters for the two retrieval methods

方法 Method	系数 Parameters						
	a_1	a_2	b_1	b_2	b_8	c_3	c_7
实际模型 Actual model	—10	10	28	—1	—1	—2.66	1
最小二乘法 The last square method	—10.20	10.07	28.16	—1.19	—1.01	—2.66	1.00
遗传算法 Genetic Algorithm	—10.01	10.01	28.01	—1.03	—1.002	—2.66	1.00

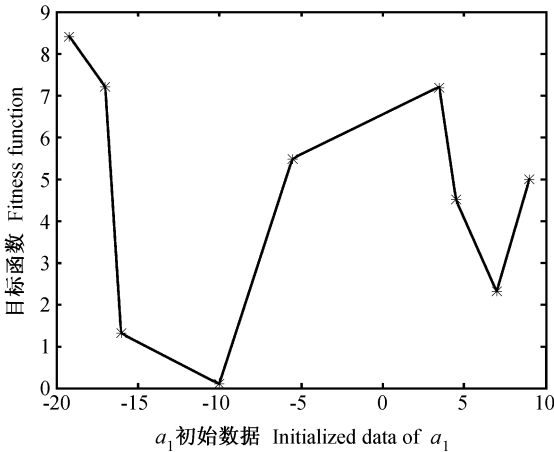


图 1 a_1 参数的初始种群个体 (9 个) 分布
Fig. 1 The distribution of initial population of a_1 (nine)

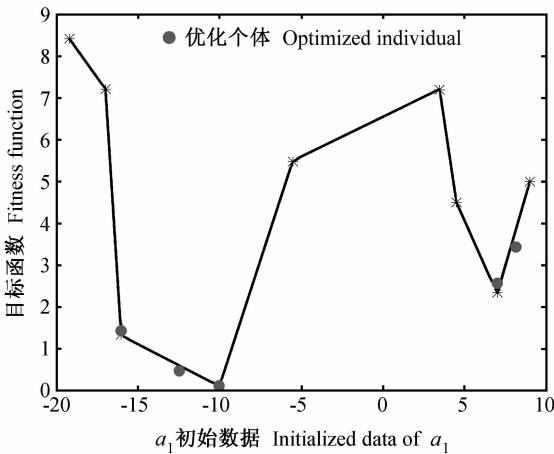


图 2 经 15 次遗传迭代后 a_1 参数的寻优结果
Fig. 2 The optimized result of a_1 of 15 generations

5 副高指数的非线性动力模型反演

利用 1958~1997 年 10 年平均的逐候副高脊线指数 (X)、副高面积指数 (Y) 和副高西脊点指数 (Z) 的观测资料时间序列, 进行副热带高压特征指数的非线性动力模型反演, 模型的基本形式为:

$$\frac{dX}{dt} = a_1X + a_2Y + a_3Z + a_4X^2 + a_5Y^2 +$$

$$a_6Z^2 + a_7XY + a_8XZ + a_9YZ,$$

$$\frac{dY}{dt} = b_1X + b_2Y + b_3Z + b_4X^2 + b_5Y^2 +$$

$$b_6Z^2 + b_7XY + b_8XZ + b_9YZ,$$

$$\frac{dZ}{dt} = c_1X + c_2Y + c_3Z + c_4X^2 + c_5Y^2 +$$

$$c_6Z^2 + c_7XY + c_8XZ + c_9YZ.$$

采用与前面相同的计算方案和操作步骤, 计算中迭代时间步长为 0.1 候, 最后得到副高特征指数非线性动力模型中各项优化反演参数。剔除量级很小的模型虚假项后, 反演得到如下副高形态指数的非线性动力系统模型:

$$\frac{dX}{dt} = -3.696 \times 10^{-4}X + 1.191 \times 10^{-5}Y + 5.178 \times$$

$$10^{-5}Z + 4.199 \times 10^{-6}X^2 + 1.582 \times 10^{-6}XZ,$$

$$\frac{dY}{dt} = -6.626 \times 10^{-2}X + 2.136 \times 10^{-3}Y + 9.282 \times$$

$$10^{-3}Z + 7.53 \times 10^{-4}X^2 + 2.836 \times 10^{-4}XZ,$$

$$\frac{dZ}{dt} = 1.877 \times 10^{-2}X - 6.1 \times 10^{-4}Y - 2.63 \times$$

$$10^{-3}Z - 2.1 \times 10^{-4}X^2.$$

上述反演所得的副高特征指数动力学模型是否准确合理, 需要进行实际检验。为此, 我们用该模型进行了积分预报试验, 通过设定真实的预报初值 (从副高指数序列中选取), 对模型进行了数值积分。该模型以第 30 候的副高指数为初值进行的数值积分结果如图 3, 从图中可见, 前期的积分预报结果 (从初始积分点第 30 候到第 55 候之间) 与实际值均很吻合, 表现出较理想的预报效果和准确率, 55 候以后模型积分结果与实际值之间的误差逐渐增大 (而面积指数的积分结果一直保持较好的准确率), 表现出相应的可预报性限制。

以该方程组作为预报模型, 只需提供模型的初值 (副高面积指数、脊线和西脊点指数) 即可, 无需统计回归和神经网络方法那样需要提供众多的预

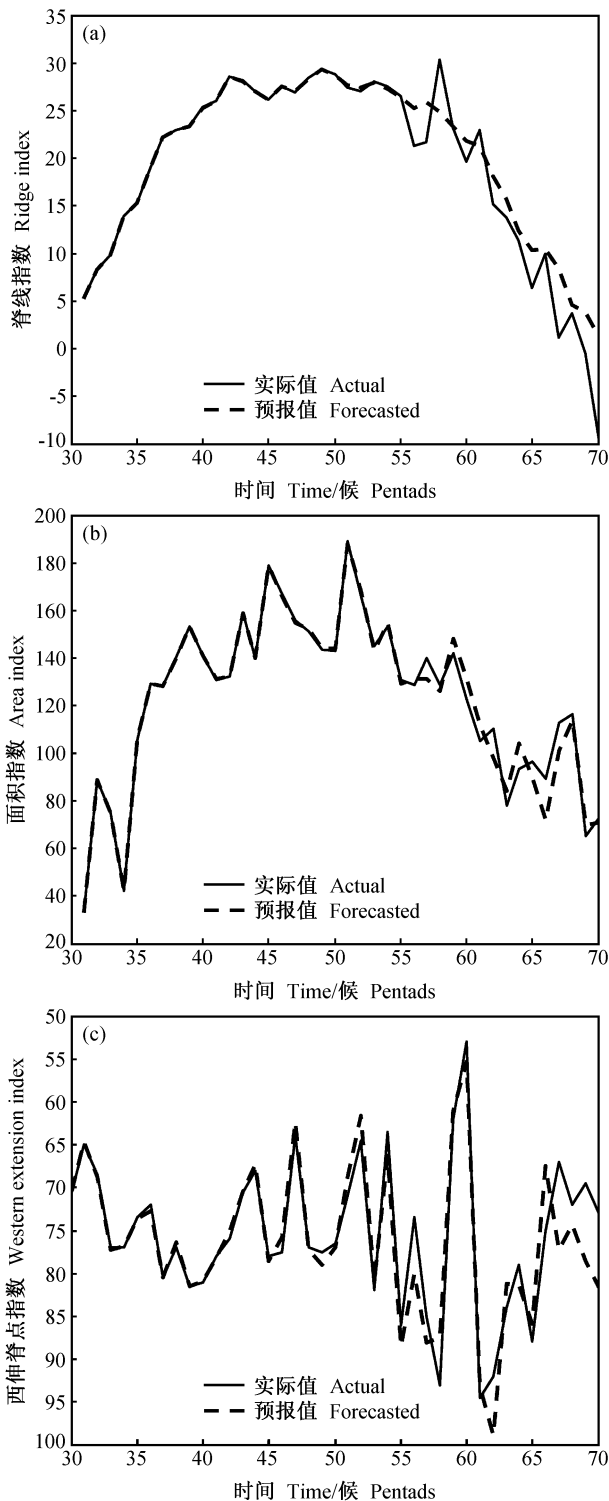


图 3 候平均副高特征指数反演模型的预报效果: (a) 副高脊线指数; (b) 副高面积指数; (c) 副高西脊点指数
Fig. 3 The forecasted results of the retrieval model for the pentad-averaged subtropical characteristic indexes: (a) the average subtropical high ridge index; (b) the average subtropical high area index; (c) the average subtropical high western extension index

报因子。模型可提供多时效的预报(时间积分即可),而无须像统计方法那样需要分别建立多个时效的预报模型,因而兼备了数值预报和统计预报方法的优点,具有较好的实用意义和应用前景。

6 结果与讨论

采用遗传算法,从数据时间序列中反演重构了非线性动力学模型, Lorenz 混沌动力系统的试验结果证明了该方法的可行性和有效性,随后从实际观测数据的时间序列中反演重构了副高特征指数的非线性动力学模型,模型的参数反演效率和准确率优于传统的最小二乘估计^[11],反演模型的预报时效和可操作性 优于常规统计预报方法(包括神经网络等)^[12],为复杂天气系统(特别是无法准确获取其动力模型)的动力学研究和诊断预测探索了新的途径。

在模型反演的具体应用^[13]时,需要注意以下问题:

(1) 资料滤波:由于实际观测资料包含了多种因素的共同影响,为了突出系统的主要特征,在进行模型反演之前必须对原始资料进行滤波处理,尽可能消除噪声的影响。

(2) 数据标准化:选择适当的特征时间,以便使观测采样步长与数值积分(差分)时间步长一致。如果观测数据的采样时间间隔过长,可适当进行插值加密处理。

(3) 对不同的建模对象,需要作针对性的诊断分析,以便客观合理地选取确定模型变量。

(4) 本文所做的仅是初步的应用探索,尚有许多不完善之处(如怎样客观确定反演模型的变量数和阶数,如何在模型反演中充分提取和吸收实际数据序列的动力学性质,如分维、李氏指数特性等)。

参考文献 (References)

- [1] 吴国雄,丑纪范,刘屹岷,等. 副热带高压研究进展及展望. 大气科学, 2003, **27** (4): 503~517
Wu Guoxiong, Chou Jifan, Liu Yimin, et al. Review and prospect of the study on the subtropical anticyclone. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2003, **27** (4): 503~517
- [2] Takens F. Detecting strange attractors in fluid turbulence. *Lecture Notes in Mathematics*, 1981, **898** (2): 361~381
- [3] 田纪伟,孙孚,楼顺里,等. 相空间反演方法及其在海洋资料分析中的应用. 海洋学报, 1996, **18** (4): 1~10
Tian Jiwei, Sun Fu, Lou Shunli, et al. A phase space EOF method and it's application to analysis oceanic data. *Acta Oceanologica Sinica* (in Chinese), 1996, **18** (4): 1~10
- [4] 魏恩泊,田纪伟,许金山. 隐含变量新反演方法及在海洋资料中的应用. 中国科学(D辑), 2001, **31** (2): 171~176
Wei Enbo, Tian Jiwei, Xu Jinshan. New reconstructing method with connotative variable and it's application to analysis oceanic data. *Science in China (Series D)* (in Chinese), 2001, **31** (2): 171~176
- [5] 马军海,陈予恕,刘曾荣. 动力系统实测数据的非线性混沌模型重构. 应用数学和力学, 1999, **20** (11): 1128~1134
Ma Junhai, Chen Yushu, Liu Zengrong. A non-linear chaotic model reconstruction for the exerimental data obtained from different dynamic system. *Applied Mathematics and Mechanics* (in Chinese), 1999, **20** (11): 1128~1134
- [6] 黄建平,衣育红. 利用观测资料反演非线性动力模型. 中国科学, 1991, **3** (3): 331~336
Huang Jianping, Yi Yuhong. A non-linear dynamic model retrieval of observational data. *Science in China* (in Chinese), 1991, **3** (3): 331~336
- [7] 王凌. 智能优化算法及其应用. 北京: 清华大学出版社, 2001. 2~10
Wang Ling. *Intelligent Optimization Algorithms and Its Application* (in Chinese). Beijing: Tsinghua University Press, 2001. 2~10
- [8] 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, 2003. 7~9
Wang Xiaoping, Cao Liming. *Algorithms—Theory, Applications and Soft Realization* (in Chinese). Xi'an: Xi'an Jiaotong University Press, 2003. 7~9
- [9] Gouesbet G. Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time series. *Phys. Rev. A*, 1991, **43**: 5321~5331
- [10] Gouesbet G. Reconstruction of vector fields: The case of the Lorenz system. *Phys. Rev. A*, 1992, **46**: 1784~1796
- [11] Le Sceller L, Letellier C, Gouesbet G. Global vector field reconstruction including a control parameter dependence. *Phys. Lett. A*, 1996, **211**: 211~216
- [12] Ortega G J. A new method to detect hidden frequencies in chaotic time series. *Phys. Lett. A*, 1995, **209**: 351~355
- [13] 巢纪平,周德刚. 大气边界层动力学和植被生态过程耦合的一个简单解析理论. 大气科学, 2005, **29** (1): 37~47
Chao Jiping, Zhou Degang. A simple analytical theory of coupling between the atmosphere boundary layer and plant. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2005, **29** (1): 37~47