

经验正交函数与遗传算法结合的副热带 高压位势场非线性模型反演*

张 韧^{1,2,3}, 洪 梅¹, 孙照渤², 牛生杰²,
朱伟军², 闵锦忠², 万齐林³

1. 解放军理工大学 气象学院 海洋与空间环境系, 南京 211101;
2. 南京信息工程大学, 江苏省气象灾害重点实验室, 南京 210044;
3. 中国气象局 热带海洋气象研究所, 广州 510080)

(林建忠推荐)

摘要: 针对副热带高压的动力预报模型难以准确构建的困难, 基于 T106 数值预报产品 500 hPa 位势高度场序列, 用经验正交函数(EOF)分解方法对位势场序列进行了时、空分解, 引入了动力系统重构思想, 以 EOF 分解的空间模态的时间系数序列作为动力模型变量, 用遗传算法全局搜索和并行计算优势, 进行了动力模型参数的优化反演, 建立了客观合理的非线性动力模型。通过对动力模型积分和 EOF 的时、空重构, 实现了副热带高压的中、长期预报。试验结果表明, 本文反演的动力模型的副热带高压预报效果优于常规的数值预报产品, 该研究工作为副热带高压等复杂天气系统和要素场预报提供了新的方法思路和技术途径

关键词: 遗传算法; 经验正交函数; 非线性模型反演; 副热带高压

中图分类号: O175.14; P433 **文献标识码:** A

引 言

西太平洋副热带高压(简称副高)是影响夏季东亚天气气候的重要系统, 副高的活动异常是导致我国江淮流域出现洪涝和干旱等灾害性天气的主要原因。副高研究的主要目的和最大的难点是准确预报副高活动。由于副高不仅有规则的渐变, 更有异常的突变, 表现出明显的非周期性和非确定性, 使副高预报非常复杂和困难。

由于副高系统的非线性和复杂性, “精确”建立副高活动的动力预报模型非常困难, 为此我们拟用反问题思路进行副高的预报研究, 即从观测资料序列中反演重构描述副高活动的位势高度场动力预报模型。Takens 在其相空间重构理论^[1]中, 对从观测资料时间序列中重构动力系统的基本思想予以了严格的阐述和证明, 认为系统中任一分量的演化是由与之相互作用的

* 收稿日期: 2005_08_16; 修订日期: 2006-07-15

基金项目: 国家自然科学基金资助项目(40375019); 热带海洋气象科学研究基金资助项目(200609); 江苏省气象灾害重点实验室开放课题资助项目(KLME0507)

作者简介: 张韧(1963—), 男, 四川峨眉人, 博士, 教授(联系人. Tel: + 86_25_83720832; E_mail: zren63@126.com).

其它分量所决定,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,从有限的观测数据时间序列中能够重构出系统发展演变的动力模型。传统的时延相空间重构方法简单易行^[2,3],但由于仅考虑了单一要素与其时延序列,因此独立信息较少,在反演重构复杂非线性系统时存在较大的局限性。为此,我们拟从数据序列中通过模型参数反演,重构动力系统模型^[4]。首先用经验正交函数分解(EOF)方法对1995~1997年夏季月份的T106数值预报产品500 hPa位势高度场序列进行时、空分解,以EOF分解模态的时间系数序列为动力模型变量,用遗传算法进行动力模型参数的优化反演,建立起EOF时间系数的非线性动力模型,再通过EOF各分解模的时、空重构,以实现副高位势场的数值积分预报。

1 经验正交函数分解

经验正交函数(empirical orthogonal function, EOF)是地球科学中广泛应用的场分析方法^[5,6]。它对实际数据场序列作时空正交分解,将时空要素场转化为若干空间的基本模态和相应的时间系数序列的线性组合,进而得以客观定量地分析要素场的空间结构和时变特征。

设某一要素场有 n 个测点,进行了 m 次观测($m > n$),为消除季节变化的影响,一般需把要素场转化成距平场进行分析。将 n 个测点 m 次距平观测值排列成矩阵 X

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{pmatrix} \cdot$$

经验正交函数展开,即是把时空要素场序列分解成彼此正交的时间函数和空间函数的乘积之和:

$$\hat{x}_{ij} = \sum_{n=1}^N t_{ni} l_{nj}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n,$$

式中 l_{nj} 表示序号为 h 的空间典型场在第 j 个点的值,它只依赖于空间点变化,不随时间变化,称为空间函数; t_{ni} 表示序号为 h 的空间典型场在第 i 个时刻的权重系数,只随时间变化,称为时间函数(或时间权重系数)。上述资料阵可写成 $X_{m \times n} = T_{m \times m} L_{m \times n}$, 其中

$$T_{m \times m} = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{m1} & \cdots & t_{mm} \end{pmatrix}, \quad L_{m \times n} = \begin{pmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \vdots & \vdots \\ l_{m1} & \cdots & l_{mn} \end{pmatrix} \cdot$$

通常把空间函数 l_{nj} 视为典型场,时间函数 t_{ni} 视为典型场的权重系数。因此观测要素场时间序列可转化为空间典型场与时间权重系数的线性叠加,各场之间的差别主要表现在时间权重系数的不同。

2 动力模型反演原理

设任一非线性系统随时间演变的物理规律可表示为

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

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

$$\frac{q_i^{(j+1)\Delta t} - q_i^{(j-1)\Delta t}}{2\Delta t} = f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, \dots, q_i^{j\Delta t}, \dots, q_N^{j\Delta t}), \quad j = 2, 3, \dots, M-1, \quad (2)$$

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 $f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, \dots, q_i^{j\Delta t}, \dots, q_N^{j\Delta t})$ 有 G_{jk} 个包含变量 q_i 的函数展开项和对应的 P_{ik} 个参数(其中 $i = 1, 2, \dots, N; j = 1, 2, \dots, M; k = 1, 2, \dots, K$), 可设为

$$f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, \dots, q_i^{j\Delta t}, \dots, q_N^{j\Delta t}) = \sum_{k=1}^K G_{jk} P_{ik},$$

式(2)的矩阵形式为 $D = GP$, 其中,

$$D = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{pmatrix} = \begin{pmatrix} \frac{q_i^{3\Delta t} - q_i^{\Delta t}}{2\Delta t} \\ \frac{q_i^{4\Delta t} - q_i^{2\Delta t}}{2\Delta t} \\ \vdots \\ \frac{q_i^{M\Delta t} - q_i^{(M-2)\Delta t}}{2\Delta t} \end{pmatrix}, \quad G = \begin{pmatrix} G_{11}, G_{12}, \dots, G_{1K} \\ G_{21}, G_{22}, \dots, G_{2K} \\ \vdots \\ G_{M1}, G_{M2}, \dots, G_{MK} \end{pmatrix}, \quad P = \begin{pmatrix} P_{i1} \\ P_{i2} \\ \vdots \\ P_{iK} \end{pmatrix}.$$

上述未知方程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 D , 要求一个向量 P , 使上式满足. 对于 q 而言, 这是一个非线性系统, 但对 P 而言, 上式是一个线性系统, 可用经典的最小二乘估计, 使残差平方和 $S = (D - GP)^T(D - GP)$ 最小, 进而得到正则方程 $G^T GP = G^T D$.

由于 $G^T G$ 多为奇异矩阵, 所以可将其特征值与特征向量求出, 剔除其中的 0 值, 剩下 K 个 $\lambda_1, \lambda_2, \dots, \lambda_K$ 组成对角矩阵 Λ 与相应的 K 个特征向量组成的特征矩阵 U_L .

$$V_L = \frac{GU_i}{\lambda} \quad (i = 1, 2, \dots, K), \quad H = U_L \Lambda^{-1} V_L^T,$$

再求 $P = HD$, 即可求出方程组参数 P .

3 遗传算法与模型参数搜索

常规参数估计方法在参数空间中大多按顺序单向搜索, 需要遍历整个参数空间, 由于误差梯度收敛的局限性和对初始解的依赖性, 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^[7]. 为此, 我们采用遗传算法进行动力模型参数的优化反演, 以 $S = (D - GP)^T(D - GP)$ 极小为约束, 用遗传算法在参数空间中进行最优参数搜索. 设参数 P 为种群, 以残差平方和 $S = (D - GP)^T(D - GP)$ 为适应度函数, 通过种群(参数)编码、初始群体生成、适应值计算、父本选择、交叉操作、变异操作等遗传算法操作运算^[7], 即可实现非线性动力模型的参数优化反演(具体操作步骤略).

4 研究资料的位势场时空分解

选择 T106 数值预报产品为研究资料, 以 1995~1997 年夏季月份(每年 5 月 1 日至 8 月 31 日)共计 369 天的 500 hPa 位势高度初值场(代表实际位势场)序列用以描述副高活动(范围: $10^\circ N \sim 60^\circ N; 70^\circ E \sim 146^\circ E$).

采用 EOF 方法对上述位势场序列进行时间(系数)、空间(结构)分解, 各特征模的方差贡

献如表 1。表中可以看出,前 6 个特征模向量的方差贡献收敛较快,其累积方差贡献已达到原始场的 86.33%,基本上能够表现副高等大尺度系统的基本特征(各分解模态的空间场和时间系数序列图略)。因此,我们取前 6 个特征模向量为建模对象,并以其时、空重构逼近实际位势场。

表 1 前 10 个 EOF 分解模的方差和累积方差贡献

EOF 特征模	1	2	3	4	5	6	7	8	9	10
方差贡献 $S_i(\%)$	53.55	13.14	7.36	5.66	3.82	2.8	2.11	1.73	1.48	1.07
累计方差 $S_c(\%)$	53.55	66.69	74.05	79.71	85.53	86.33	88.44	90.17	91.65	92.72

上述 EOF 分解的空间模态除表示样本期间位势场基本空间结构特征外,也可表现未来相邻时段位势场的空间结构(除非出现异常突变)。为此 EOF 分解的空间模态可被近似为定常,位势场变化可认为是各空间模态时间系数的变化所致。这样对位势场的预报就转换为对 EOF 分解的空间模时间系数序列的预报建模,再用时间系数的预报结果进行 EOF 时空重构,进而实现位势场和副高的预测。

5 EOF 时间系数的动力模型重构

以 $T_1, T_2, T_3, T_4, T_5, T_6$ 表示 EOF 分解的前 6 个空间模的时间系数变量,并基于其“观测”数据,重构动力模型。设广义二阶非线性常微方程组为拟反演重构的非线性动力模型,取研究资料(共 369 天) EOF 分解的前 6 个空间模态时间系数序列的前 349 天作为“观测资料”,进行动力模型参数的优化反演(后 20 天样本数据留作动力模型预测检验)。

计算中取迭代步长为 1 天,采用遗传算法进行动力模型参数的优化反演。经 15 次遗传操作即可迅速收敛于目标适值,搜索出动力模型的优化参数。剔除量级很小的无关项后,得到如下 EOF 分解的前 6 项空间模态的时间系数的动力预报模型:

$$\frac{dT_1}{dt} = 6.9342 T_1 + 18.46 T_4 + 26.809 T_2^2 + 8.1656 T_6^2 + 15.755 T_1 T_6 + 21.284 T_2 T_3 + 17.042 T_4 T_5,$$

$$\frac{dT_2}{dt} = 12.4347 T_1 + 9.7325 T_6 + 5.9974 T_1^2 + 20.4432 T_3^2 + 17.3214 T_2 T_4,$$

$$\frac{dT_3}{dt} = 2.4735 T_2 + 14.2945 T_5 + 12.3552 T_2^2 + 25.7892 T_3^2 + 0.9872 T_5^2 + 1.7468 T_1 T_5 + 9.3275 T_5 T_6,$$

$$\frac{dT_4}{dt} = 5.7622 T_1 + 9.1524 T_3 + 35.1788 T_6 + 24.7215 T_1^2 + 1.7324 T_6^2 + 23.155 T_2 T_6 + 48.12 T_4 T_5,$$

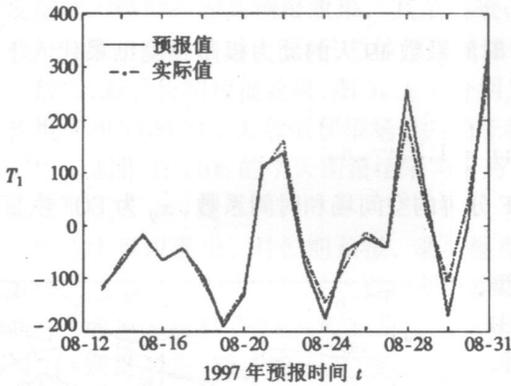
$$\frac{dT_5}{dt} = 41.1322 T_2 + 13.1725 T_5 + 9.1344 T_3^2 + 41.13 T_1 T_3 + 23.5213 T_2 T_4 + 7.23 T_4 T_6 + 1.334 T_5 T_6,$$

$$\frac{dT_6}{dt} = 9.1722 T_3 + 8.1542 T_6 + 12.344 T_1^2 + 4.2267 T_5^2 + 1.3435 T_1 T_6 + 6.7894 T_2 T_4.$$

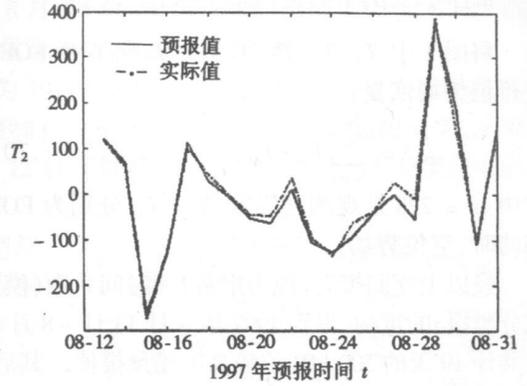
6 动力模型预报效果检验

为检验上述动力预报模型是否客观、准确,我们用研究资料中未参与模型反演的后 20 天

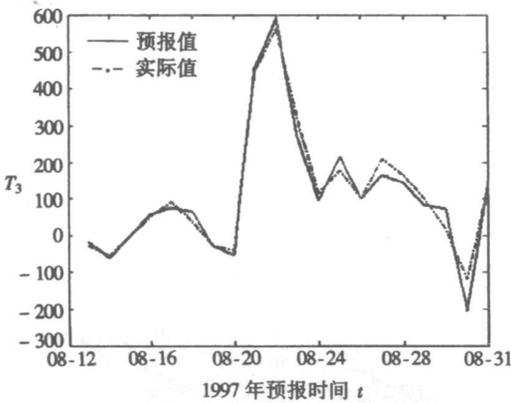
500 hPa 位势高度场 EOF 分解系数(1997 年 8 月 12 日到 8 月 31 日 20 天)来检验反演重构模型的积分预报效果。取 8 月 12 日的 6 个 EOF 分解系数作为初值,代入以上反演所得的非线性动力方程组,进行模型的数值积分运算,得到 1997 年 8 月 13 日至 8 月 31 日共 19 天的数值积分预报结果,如图 1。



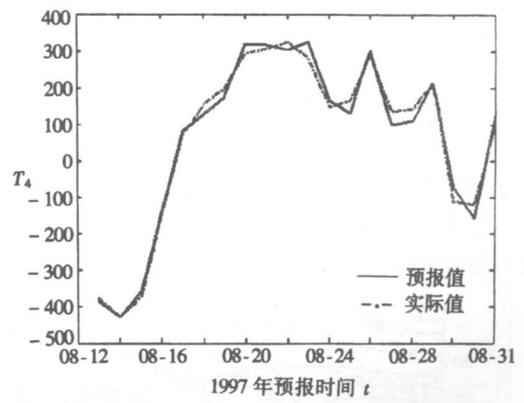
(a) 第 1 模态的时间系数积分预报 (T_1)



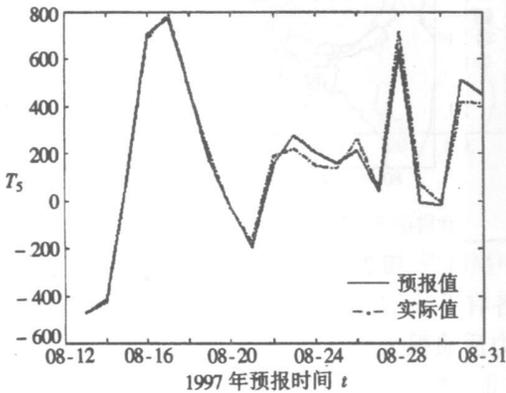
(b) 第 2 模态的时间系数积分预报 (T_2)



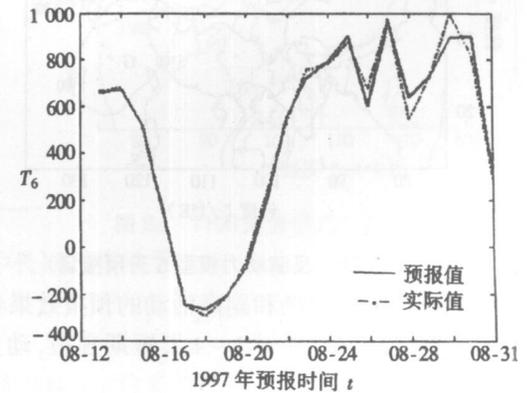
(c) 第 3 模态的时间系数积分预报 (T_3)



(d) 第 4 模态的时间系数积分预报 (T_4)



(e) 第 5 模态的时间系数积分预报 (T_5)



(f) 第 6 模态的时间系数积分预报 (T_6)

图 1 6 个空间模态的时间系数积分预报效果(1997 年, 预报时效 19 天)

图中可以看出, 前 10 天的积分预报结果(约 8 月 13 日~ 8 月 24 日)与真实值很吻合, 表现

出理想的逼近效果和预报准确率; 12 天以后的积分结果与真实值之间出现了略微的预报偏差, 但积分结果与真实值之间的变化趋势和峰谷位置仍保持一致, 表现出稳定、良好的中、长期预报效果。

7 副高位势高度场的预报试验

将图 1 中 $T_1, T_2, T_3, T_4, T_5, T_6$ 等 6 个 EOF 时间系数 19 天的动力模型预测结果代入下式进行位势场恢复:

$$\hat{x}_j = \sum_{n=1}^6 F_{nj} \cdot T_{nt}, \quad t = 1, 2, \dots, 20; j = 1, 2, \dots, n,$$

式中 $n = 228$ 是观测网络数, F_{nj}, T_{nt} 分别为 EOF 分解的空间场和时间系数, \hat{x}_j 为 EOF 恢复重构的时、空位势场。

经以上空间模态(视为定常)和时间系数(模型积分预报)的重构, 得到 1997 年 8 月 13 日~ 8 月 31 日共计 19 天的 500 hPa 位势高度场预报值。其后, 以实况场(同期初始场)、T106 数值预报场为参照, 对两种预报方法得到的副高强度、范围和位置等预报结果进行了对比分析和效果评估。

1997 年 8 月 13 日的 500 hPa 位势场实况与第 1 天的动力模型预报场和 T106 数值模式预报场的对比分析结果表明, 动力反演模型和 T106 模式两种预报结果与实况场的基本形式和高、低压中心均非常接近, 两种预报结果的副高范围和位置均与实况相符。其后 2~ 4 天(8 月 14 日~ 8 月 16

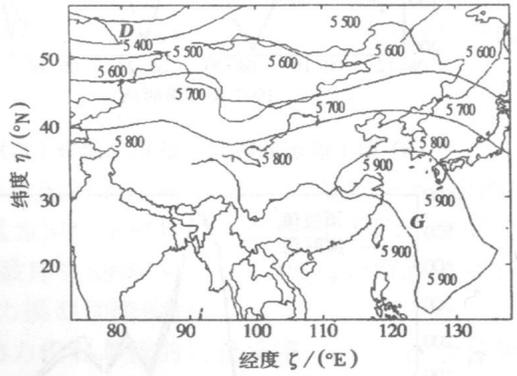


图 2a 500 hPa 位势高度实况场

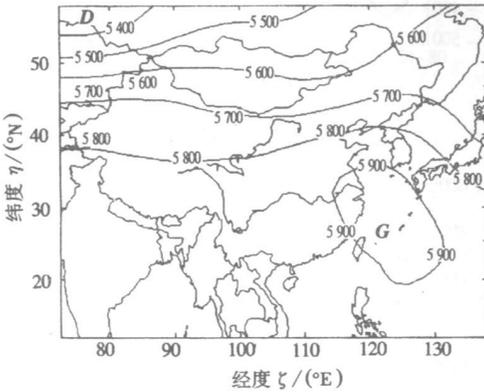


图 2b 反演动力模型 5 天预报场

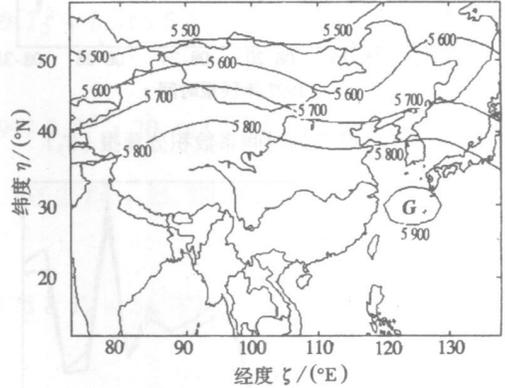


图 2c T106 数值模式 5 天预报场

日), 两种方法的位势场和副高活动的预报效果各有千秋、大体接近, 均能较好描述和逼近实况高度场和副高状况。即对 1~ 4 天短期预报, 动力反演模型与 T106 数值模式的副高预报效果相当(图略)。

图 2 a、b、c 分别为 1997 年 8 月 17 日的实况场、第 5 天的动力反演模型预报场和 T106 数值预报场。

图中可以看出, 7 月 18 日位势场实况图上, 5 900 线表示的副高呈东南—西北向, 位于菲律

宾以东、日本以南的西太平洋和东海海域,并伸入覆盖了长江中下游地区(图 2a);动力反演模型预报结果中的副高范围略小于了实况,副高西伸范围略小,仅覆盖了东南沿海地区,但副高的基本形式和强度、位置均十分接近副高实况(图 2b),预报结果基本正确;相比而言,T106 数值预报结果的副高范围和位置则明显偏小,与副高实际情况相差甚远(图 2c),预报效果远不及反演动力模型的副高预报效果。其余时效(8月18日~8月22日)反演动力模型方法的副高预报结果也基本接近实况场,总体预报效果优于 T106 数值预报(图略)。

最后,看看长期预报效果,图 3a、b、c 分别为 1997 年 8 月 25 日的实况场、动力反演模型 13 天预报场和 T106 的 7 天数值预报场(注:由于我们 T106 资料中仅有 1~7 天预报结果,故采用 8 月 18 日起报的 T106 的 7 天预报结果与 8 月 12 日起报的动力模型方法的 13 天预报结果进行比较)。

图 3 中可以看出,对长期预报,动力模型反演方法的位势场形势预报除对台湾和菲律宾以东洋面一个台风系统(图 3a)未能刻画外,基本环流结构,特别是对东(西太副高)、西(印度半岛以北)两个 5900 线副高范围均表现正确,对西太副高 5900 线与外围 5850 线形态位置的预报与实况相符(图 3b)。相比而言,T106 的 7 天数值预报结果对副高的描述存在较大偏差,除对西面的一个 5900 副高有所表现外(范围还有所扩大),对西太平洋副高未能描述,预报基本失败(图 3c)。

上述结果表明,本文反演建立的动力模型在副高的中长期预报方面较 T106 数值预报模式有

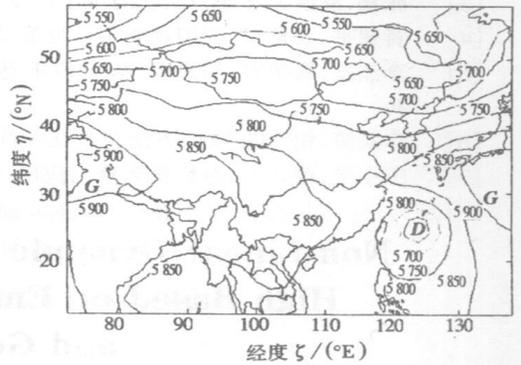


图 3a 500 hPa 位势高度实况场

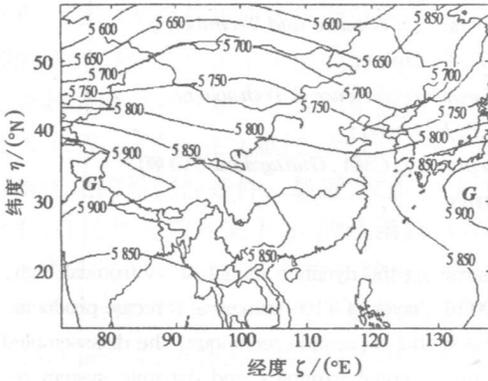


图 3b 反演动力模型 13 天预报场

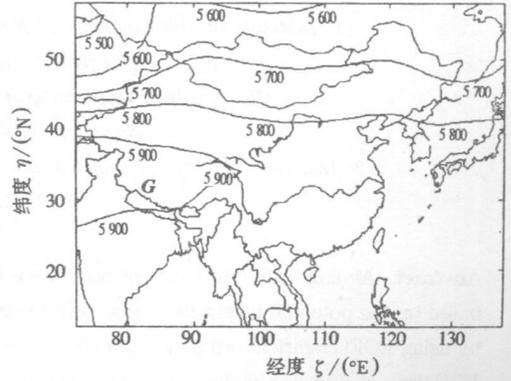


图 3c T106 数值模式 7 天预报场

一定的优势,能够得到较为准确、稳定的副高中长期预报结果。

8 结果与讨论

本文用经验正交函数(EOF)方法分解、重构 500 hPa 位势高度场序列,在此基础上,引入遗传算法对 EOF 的时间系数序列进行了动力模型重构和模型参数反演,建立了位势场 EOF 时间系数的非线性动力学模型,再通过时、空函数合成,构造了位势场的动力预报模型。对比试验结果表明,本文反演的位势场动力模型能够较好描述副高的变化特征,预报效果,尤其是中长

期预报效果优于 T106 数值预报结果, 为副高等复杂天气系统的预报提供了有益的方法、思路和参考借鉴。

本文提出的动力模型反演方法充分利用了实况“观测资料”信息, 所建动力模型能够较为客观、真实地反映位势场变化特征, 建模针对性较强, 兼备了数值预报和统计预报方法的优点, 因而具有较好的推广应用前景。

[参 考 文 献]

- [1] Takens F. Detecting strange attractors in fluid turbulence[J]. Lecture Notes in Mathematics, 1981, 898(2): 361—381.
- [2] 田纪伟, 孙孚. 相空间反演方法及其在海洋资料分析中的应用[J]. 海洋学报, 1996, 18(4): 1—10.
- [3] 张韧. 反演大气动力系统与中小尺度天气预报[J]. 海洋科学, 1997, 21(1): 17—20.
- [4] 黄建平, 衣育红. 利用观测资料反演非线性动力模型[J]. 中国科学, 1991, 3(3): 331—336.
- [5] 李杰友, 熊学农, 刘秀玉. 基于 EOF 迭代的月径流长期预报[J]. 河海大学学报, 2001, 29(2): 43—46.
- [6] 李跃清. 相空间 EOF 方法及其在气候诊断中的应用[J]. 高原气象, 2001, 20(1): 2—6.
- [7] 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, 2003.

Non Linear Dynamic Model Retrieval of Subtropical High Based on Empirical Orthogonal Function and Genetic Algorithm

ZHANG Ren^{1, 2, 3}, HONG Mei¹, SUN Zhao_bo², NIU Sheng_jie²,
ZHU Wei_jun², MIN Jin_zhong², WAN Qi_lin³

(1. Institute of Meteorology, PLA University of Science and Technology,
Nanjing 211101, P. R. China;

2. KLME, Nanjing University of Information Science & Technology,
Nanjing 210044, P. R. China;

3. Institute of Tropical and Marine Meteorology, CMA, Guangzhou 510080,
P. R. China)

Abstract: Aiming at the difficulty of accurately constructing the dynamic model of subtropical high, based on the potential height field time series over 500 hPa layer of T106 numerical forecast products, by using EOF(empirical orthogonal function) temporal_spatial separation technique, the disassembled EOF time coefficients series were regarded as dynamical model variables, and dynamic system retrieval idea as well as genetic algorithm were introduced to make dynamical model parameters optimization search, then, a reasonable non_linear dynamic model of EOF time_coefficients was established. By dynamic model integral and EOF temporal_spatial components assembly, a mid/ long term forecast of subtropical high was carried out. The experimental results show that the forecast results of dynamic model are superior to that of general numerical model forecast results. A new modeling idea and forecast technique is presented for diagnosing and forecasting such complicated weathers as subtropical high.

Key words: genetic algorithm; empirical orthogonal function; non_linear model retrieval; subtropical high