

一个包含外强迫因子的非平稳时间序列的预测方法

王革丽^①, 杨培才^①, 卞建春^①, 周秀骥^②

① 中国科学院大气物理研究所, 中层大气环境探测与遥感开放实验室, 北京 100029;

② 中国气象科学研究院, 灾害天气国家重点实验室, 北京 100081

E-mail: wgl@mail.iap.ac.cn

2011-03-02 收稿, 2011-06-15 接受

国家自然科学基金(40890052, 41075061, 40940023)资助项目

摘要 气候系统所具有的非平稳性质已经得到越来越多的气象学家的认同. 事实上, 产生非平稳行为的根本原因在于作用在系统上的外部强迫是随时间而变化的. 鉴于此, 本文提出了一个新的非线性和非平稳时间序列的预测方法. 与以往的某些方法最大的不同在于: 在建模过程中, 外强迫因子直接参与了预测方程的构建. 还利用时变控制参数条件下, Logistic 映射和 Lorenz 方程产生的非平稳时间序列, 对上述方法进行了预测试验. 初步的试验结果表明, 外强迫因子在预测中扮演着与状态变量同等重要的角色, 它们的参与可以有效地改善预测精度.

关键词

外强迫
非平稳系统
气候预测
时间序列预测

短期气候预测是气候变化研究的热点和难点问题之一. 由于短期气候过程内在的复杂性, 以及人们对控制它的规律缺乏全面的认识, 因而, 短期气候预测水平没有明显的提高^[1,2]. 对于 20 世纪的气候变化来说, 全球变暖是其重要特征, 目前对其成因尚有很多争议. 可能的原因包括自然强迫因子的变化(例如太阳活动、火山喷发), 以及人类活动的影响(包括人为排放的温室气体和硫酸盐气溶胶, 以及土地利用的变化等). 另外, 气候系统的内在变率的影响同样不容忽视. 然而, 无论其变化机制怎样, 全球变暖的事实已经证明了气候过程是非平稳的.

目前, 几乎所有的非线性时间序列预测理论, 都是建立在平稳性(状态分布与时间无关)假定的基础之上, 人们依靠这样的假定, 便可应用“遍历性”理论, 借助于一个观测现实恢复系统的动力学, 并在它上面建立系统的预测方程.

然而, 对于天气或气候系统来说, 这样的做法显然是不合理的, 人们不可能设想驱动气候系统的外部强迫会是一成不变的. 实际上, 除了全球变暖之外, 在一些天气和气候资料的分析中, 还发现了大气过

程的平稳性被破坏的事实. Tsonis^[3]通过对近百年的全球降水过程的年际变率的分析, 证明了降水的平均值没有变化, 但是其相对于平均值的涨落(二阶矩)却明显地增大(表明旱涝极端事件的频发), 也就是说, 全球降水资料描述了一个非平稳过程. 另外, 北太平洋地区冬季海平面气压等在 1977/1978 年前后的突变^[4]也说明大气过程的平稳性结构被改变了. 非平稳特征标志着, 控制系统的外部条件是随时间变化的^[5].

然而, 迄今为止, 还没有建立任何有关非平稳过程预测的一般性理论. 科学家们所能做到的, 仍然是通过某种假定或近似, 设法在非平稳过程中找出一些保持平稳的段落, 或者是通过某些变换, 把相应的过程平稳化.

近年来, 我国的大气科学家在大量的观测事实的支持下, 提出了“气候层次”理论, 指出气候是一个由处于不同“层次”上的物理系统形成的串级结构^[6-9]. 在这一理论的指导下, 他们提出了“复合重构”和“分解隔离”等方法, 并用它们来建立非平稳气候系统的预测方程. 从技术上来说, 这些方法是试图通过分解

信号的非平稳性, 来寻找近似的平稳的信号分量.

Hegger 等人^[10]还提出“过嵌入”(over-embedding)方法, 通过提高嵌入维数, 把系统中暗藏的时变参数当作状态变量来处理, 以此来化解产生非平稳性的物理上的原因. 王革丽等人^[11]还利用一个在理论上更为成熟的“过嵌入”或“升维”方法(“支持向量机”算法)预测了几个非平稳过程, 并且得到了较好的预测结果.

事实上, 非平稳行为起因于外部强迫随时间的变化, 而“升维”技术则是在缺少外部强迫信息的情况下, 在重建的动力系统中, 对它所做的某种修补. 因而从这个意义上讲, 将外强迫因子引入到重建的系统之中, 应该是一个直接和有效的方法. 然而, 这样一来, 获取外强迫信息就成为一个必须面对的十分重要的问题. 当然, 我们可以从观测到的不同的气候因子之间的关联性, 来确定它们之间联系. 但是, 这种联系必须建立在因果关系的基础之上^[12], 简单的相关分析有可能产生错误的结论^[13]. 因此, 在这样的思想指导下, 并且考虑到实际气候系统中存在着诸如噪音及资料长度等等不确定影响因素, 因而本文从“理想”时间序列入手, 尝试将外强迫因子引入到重建的系统之中, 并在此基础上建立非平稳时间序列的预测模式.

1 预测模型建立方法

非线性时间序列预测分析方法中, 最重要的理论基础是状态空间重构理论^[14]和嵌入定理^[15]. 根据 Takens^[15]嵌入定理, 对于一个给定的时间序列, 只要选择适当的嵌入维数和延滞时间, 就可以在其状态空间恢复系统的动力学, 它表现为一条描述系统状态演变的轨线, 通过对轨线段的研究, 了解系统的动力学行为, 并预测其状态未来的演变. 有关非线性时间序列预测建模方法有很多工作可参考^[16-19], 这里不再赘述. 值得注意的是, Takens 定理只适用于自治的动力系统, 对于非平稳时间序列来说也可以将外强迫因子嵌入到同一个已知的状态空间中^[20].

为方便描述, 考虑包含两个时间序列 $\{x_i\}_{i=1,2,\dots,n}$ 和 $\{\alpha_i\}_{i=1,2,\dots,m}$ 的一个非线性非平稳过程, 前者是系统的状态随时间的变化, 而后者表示其外强迫随时间的演变. 借助于一个适当的时滞参数 τ , 可以将它们嵌入到 $m_1 + m_2$ 维的状态空间中, 并得到如下的状态轨线:

$$\bar{y}(i) = \{x_i, x_{i-\tau}, \dots, x_{i-(m_1-1)\tau}; \alpha_i, \alpha_{i-\tau}, \dots, \alpha_{i-(m_2-1)\tau}\}_{i=1,2,\dots,N},$$

或者简写为

$$\bar{y}(i) = \{\bar{x}_i; \bar{\alpha}\}_{i=1,2,\dots,N}, \tag{1}$$

这里, m_1 和 m_2 分别为 $\{x_i\}$ 及 $\{\alpha_i\}$ 的嵌入维数, $N = n - (\max(m_1, m_2) - 1)\tau$ 为轨线上的相点数目. 在本文中, 根据这条轨线, 利用全局近似方法^[17]建立了如下的预测模型:

$$\bar{y}(i) = \{\bar{x}_i; \bar{\alpha}\}_{i=1,2,\dots,N}, \tag{2}$$

这里 \hat{f} 是一个待定的函数, ε_i 为拟合误差. 这样一来, 问题就转化为寻求 \hat{f} , 使其残差平方和

$$\eta = \sum_{k=1}^N [\hat{f}(\bar{x}_k, \bar{\alpha}_k) - x_{k+1}]^2, \tag{3}$$

达到最小. 在本文中 \hat{f} 被设定为二阶多项式函数.

2 预测试验分析

为了叙述方便, 在下面的预测试验中, 对于包含和不包含外强迫因子的预测模式, 将分别称其为“外强迫模式”和“平稳性模式”.

上述预测方法被应用到两组不同的时间序列. 第一组时间序列来自于 Logistic 映射:

$$x_{t+1} = \mu_t x_t (1 - x_t). \tag{4}$$

它由两个时间序列组成, 这两个时间序列分别对应于两个不同的参数 μ_t . 在这里, μ_t 是 Logistic 系统的外强迫因子, 它是一个随时间变化的参数. 对于 Logistic 映射来说, 当 μ_t 在 3.57~4.0 之间取值时, 系统具有混沌行为. 将两个不同的 μ_t 的变化限制在上述范围, 并使其随时间的变化服从下面两种函数形式(图 1):

$$\mu_t^{(1)} = 3.95 - 0.4e^{-2.5t}, \tag{5.1}$$

$$\mu_t^{(2)} = 3.45 + 0.5e^{-1.5t}. \tag{5.2}$$

在这样的条件下, 得到两个不同的具有混沌特征的非平稳时间序列. 为了描述方便, 分别将它们记为 $S^{(1)}$ 和 $S^{(2)}$, 其长度皆取为 2000.

在试验中, 用前 1900 个数据构造映射 \hat{f} , 后 100 个数据用于检验预测效果. 时滞参数 τ 取为 1, $\{x_t\}$ 的嵌入维数 m_1 取为 1 和 2, $\{\mu_t\}$ 的嵌入维数 m_2 取为 0 和 1. 这里, $m_2=0$ 表示预测方程中不考虑外强迫因子, 也就是说预测方程是基于平稳时间序列建立的. 表 1 给出了不同预测试验的预测误差, 可以看出, 当考虑外强迫因子, 并将其用入预测模型时, 预测精度都大

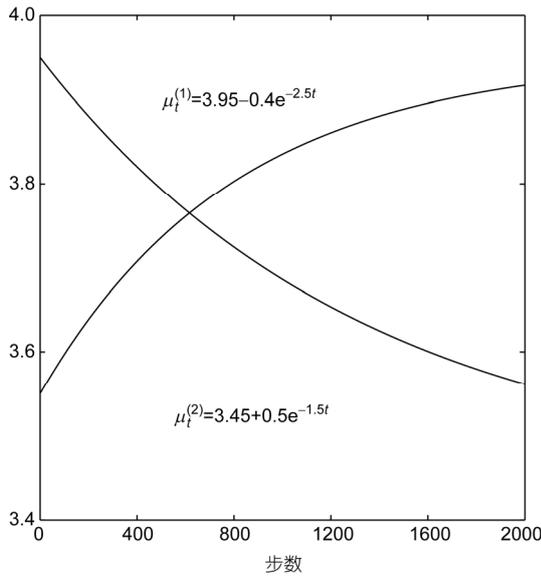


图1 两种不同的外强迫因子

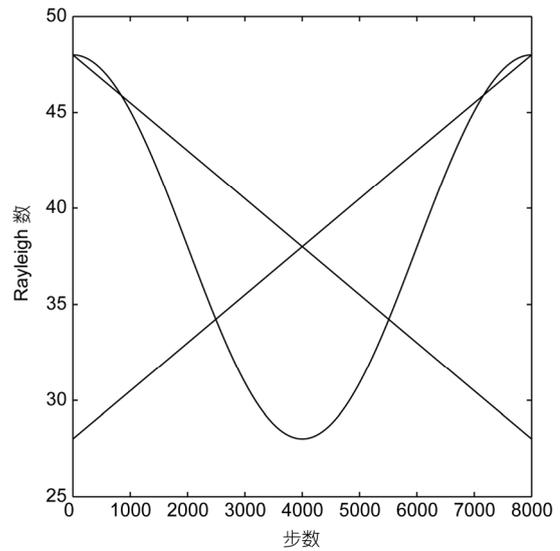


图2 Rayleigh 数的3种不同的时间变化

表1 第一组预测试验的均方根误差及比较

	$E_{S^{(1)}} (%)$	$E_{S^{(2)}} (%)$
平稳性模式, $m_1=1$ 和 $m_1=2$ 的平均	1.76	2.29
外强迫模式, $m_1=1$ 和 $m_1=2$ 的平均	0.62	0.63

大提高. 这表明, 对于非平稳时间序列的预测, 引入外强迫因子可以有效地改善预测结果, 更多的细节和例子见文献[21].

第二组预测试验序列来自于 Lorenz 系统:

$$\begin{cases} \dot{x} = -\sigma x + \sigma y, \\ \dot{y} = r(t)x - y - xz, \\ \dot{z} = xy - bz, \end{cases} \quad (6)$$

这里 Rayleigh 数 $r(t)$ 是系统的外强迫因子, 它是一个随时间变化的参数. 令 $r(t)$ 在区间[28,48]上取值, 且服从如图 2 所示的 3 种不同的函数变化形式. 另外两个参数 σ 和 b 则分别取为 10 和 8/3. 采用四阶 Runge-Kutta 方法积分方程(6), 并在它们的吸引子上截取一个长度为 8000 个相点的轨线段, 得到了 3 个关于变量 z 的非平稳时间序列(分别记为 $T^{(1)}$, $T^{(2)}$ 和 $T^{(3)}$). 在试验中, 用前 7200 个数据建立预测模型, 后 800 个数检验预测误差. 除了嵌入维数 m_1 取为 3~5 之外, 其他重构参数皆取与上面的试验相同的值. 预测试验的结果列在表 2 中. 由表 2 可以看到两个重要的事实: (1) 外强迫模式的误差, 除个别情况外, 都远低于平

稳性模式, 这说明, 外强迫信息的引入可有效地提高了预测精度; (2) 外强迫模式的误差(随预测步数)增长速度远低于平稳性模式, 这表明外强迫信息的引入可以大大地改善系统的可预报性.

3 讨论

考虑气候系统的非平稳性质, 对于揭示气候变化的机理, 发展新的气候预测理论和方法, 将产生重要的影响. 本文发展了一个建立在“外强迫”基础上的非平稳气候预测方法. 几个理想时间序列给出的试验结果表明, 该方法可以有效地提高预测精度.

表2 第二组预测试验的精度(均方根误差)及比较

预报步数	m_2	m_1 (平均)	$E_{T^{(1)}} (%)$	$E_{T^{(2)}} (%)$	$E_{T^{(3)}} (%)$
1	0	3~5	0.76	2.44	2.50
	1	3~5	0.75	1.18	0.88
2	0	3~5	1.14	5.48	1.66
	1	3~5	0.51	0.84	1.14
3	0	3~5	1.94	9.41	4.71
	1	3~5	0.82	1.46	1.51
4	0	3~5	3.10	16.51	7.12
	1	3~5	0.95	2.63	2.06
5	0	3~5	4.66	11.16	9.17
	1	3~5	1.25	2.35	2.88
6	0	3~5	6.44	16.78	9.89
	1	3~5	1.23	3.69	3.65

本质上说,如果系统的外强迫是平稳的,那么重构的动力系统也具有平稳特征,也就是说,方程(2)只不过描述了一个包含两个层次结构的平稳过程^[6].相反,如果其外强迫过程是非平稳的,那么,就必需找到与它有关的外强迫因子,以便在方程(2)中加入一个新的层次.

为了进一步完善上面提出的预测理论和方法,还应考虑以下3个工作:(1)如何获取外强迫信息:这是一个必须面对的十分重要的问题.一方面可以从气候因子之间的关联性,来确定它们之间联系.但是,这种联系必须建立在因果关系的基础之上,依据因果关系寻找外强迫因子.然而,目前还缺乏行之有效的方法.最近,Verdes等人^[22]利用“转移熵”概念,提出了一个分析变量间的因果关系的方法,并应用于气候问题研究,他的结果指出,太阳辐射和CO₂(做为温室气体)两者与全球平均温度之间存在着因果关系;另一个获取外强迫信息的方法是,直接从

时间信号中提取得到外强迫因子,这是一个有重要意义的科学问题.从本质上说,它是一个反问题(其数学提法常常是不适定的).要解决这样的问题,不仅要求人们对系统有更多的物理上的了解,甚至不得不依赖于一些事先的假定,例如,假定作用在气候系统上的外部强迫是随时间缓慢变化的等等;(2)把Takens嵌入定理延伸到非平稳时间序列.我们知道,Takens嵌入定理只适用于自治系统,然而大多数的系统并不满足这样的条件,因此修改Takens定理,使之应用于非平稳时间序列分析,就成为一个必须解决的问题.Stark^[20]很好地讨论了这个问题,他的工作给出了与驱动系统相关的Takens嵌入定理的两种形式,一种适用于驱动因子不清楚的情况,而另一种则适用于驱动系统能够独立确定的情况;(3)针对由外强迫因子的加入而引起的嵌入维数增加的问题,可能会对观测数据的数量和质量提出新的要求,因此,发展合适的建模技术也是十分必要的.

参考文献

- 1 王绍武. 短期气候预测的可预报性与不确定性. 地球科学进展, 1998, 13: 8-14
- 2 李维京, 张培群, 李清泉, 等. 动力气候模式预测系统业务化及其应用. 应用气象学报, 2005, 16(增刊): 1-11
- 3 Tsonis A A. Widespread increases in low-frequency variability of precipitation over the past century. Nature, 1996, 382: 700-702
- 4 Trenberth K E. Recent observed interdecadal climate changes in the northern hemisphere. Bull Amer Meteor Soc, 1990, 7: 988-993
- 5 Manuca R, Savit R. Stationarity and nonstationarity in time series analysis. Phys D, 1996, 99: 134-161
- 6 陈伯民, 纪立人, 杨培才, 等. 改善月动力延伸预报水平的一种新途径. 科学通报, 2003, 48: 513-520
- 7 杨培才, 卞建春, 王革丽, 等. 气候系统的层次结构和非平稳行为: 复杂系统预测问题探讨. 科学通报, 2003, 48: 1470-1476
- 8 Wang G L, Yang P C. A compound reconstructed prediction model for nonstationary climate process. Int J Clim, 2005, 25: 1265-1277
- 9 Yang P C, Wang G L, Bian J C, et al. The prediction of non-stationary climate series based on EMD. Adv Atmos Sci, 2010, 27: 845-854
- 10 Hegger R, Kantz H, Matassini L, et al. Coping with non-stationarity by over-embedding. Phys Rev Lett, 2000, 84: 4092-4101
- 11 王革丽, 杨培才, 毛宇清. 基于支持向量机方法对非平稳时间序列的预测. 物理学报, 2008, 57: 714-719
- 12 Granger C W. Testing for causality: A personal viewpoint. J Econ Dyn Control, 1980, 2: 329-352
- 13 耿直. 因果推断与Simpson悖论. 统计与信息论坛, 2000, 15: 9-12
- 14 Packard N H, Crutchfield J P, Farmer J D, et al. Geometry from a time series. Phys Rev Lett, 1980, 45: 712-715
- 15 Takens F. Detecting strange attractors in turbulence. In: David R, Lai-Sang Y, eds. Dynamical Systems and Turbulence. Berlin: Springer-Verlag, 1981. 366-381
- 16 Farmer J D, Sidorowich J. Predicting Chaotic time series. Phys Rev Lett, 1987, 59: 845-848
- 17 Casdagli M. Nonlinear prediction of chaotic time series. Phys D, 1989, 35: 335-356
- 18 Yang P C, Zhou X J, Bian J C. A nonlinear regional prediction experiment on a short-range climatic process of the atmospheric ozone. J Geophys Res, 2000, 105: 12253-12258
- 19 Elsner J B, Tsonis A A. Nonlinear prediction, chaos and noise. B Amer Meteorol Soc. 1992, 73: 49-60
- 20 Stark J. Delay embeddings for forced systems: Deterministic forcing. J Nonlinear Sci, 1999, 9: 255-332
- 21 Wang G L, Yang P C. A recent approach incorporating external forces to predict nonstationary processes. Atmos Oceanic Sci Lett, 2010, 3: 151-154
- 22 Verdes P F, Granitto P M, Ceccatto H A. Overembedding method for modeling nonstationary systems. Phys Rev Lett, 2006, 96, doi: 10.1103/PhysRevLett.96.118701