研究论文

自忆性方程与自忆模式[1)]

曹鸿兴

（中国气象科学研究院，北京 100081）

提 要

从不可逆过程的记忆概念出发，讨论了大气运动的自忆性方程。随着记忆函数的不同求取方式，自忆性方程可构成数值、统计-动力和多时刻模式。在仅考虑单点序列，把其他空间点的影响视为随机干扰时，自忆预测模式蜕变为均生函数模型。

关键词 记忆 多时刻模式 天气预报 气候预测

引 言

在数值天气预报中只使用一个初始场，显然其信息量甚为有限，因此，在天气预报中使用除初始场外多时次资料，一直是人们关心的问题，本文通过介绍和讨论大气运动的自忆性方程阐明了通过引进记忆函数和建立自忆模式，为使用多时次观测资料开辟了一个新途径，且能从自忆性方程出发展开一系列饶有兴趣的学术问题。

1 记忆性

客观世界的演变是不可逆的，可逆过程只是一种科学抽象。例如，一滴墨水撒入水杯中，慢慢通过布朗运动使其分子散布到整个水杯中，一个炸弹爆炸后变成碎片，这些都是不可逆过程。在许多运动现象中，不可逆过程起着建设性作用，例如在化学反应过程中，未来和过去是有区别的[1]。

要研究不可逆过程，就要引入记忆概念，也就是未来不仅与现在有关，还与过去有关。据此，可把方程写为

$$x(t) = F(x(t-1), x(t-2) \quad , x(t-p), \lambda) \tag{1}$$

式中 λ 为参数，方程表示系统在 t 时刻的状态 $x(t)$ 与过去时刻 $x(t-1)$，$x(t-2)$，乃至 $x(t-p)$ 都有关，反映系统的记忆特征。若 $x(t)$ 仅与上时刻 $x(t-1)$ 有关，则称系统为马尔柯夫型的。描写大气运动的微分方程只与一个时刻（即现在）有关，即

$$\partial x(t)/\partial t = F(x(t), \lambda, t) \tag{2}$$

在概率论中讨论一种随机游动，系统每次以概率 p 左移一步，以概率 $q=1-p$ 右移一步，未来时刻的状态与现在及任何时刻的状态无关，则是无记忆的。象醉汉一样，下一步怎么走是任意的。白噪声也是一种无记忆过程，因为它由一个前后无关的随机序列构成。

2 自忆性

所谓自忆性就是强调系统状态本身前后的联系，把着眼点放在系统自身演化的规律上，而把环境因素视为一种外强迫。外强迫也必须通过系统自身的演化来起作用。

在自然现象和社会现象中可以举出许多自忆性例子。例如，地球系统的演化，一个生命的成长过程，疾病的家族遗传史等等，未来的演变都和现在、过去状态紧密相关。

研究自忆性的两个关键问题是 第一，如何建立系统的自忆性模型，它的数学表达式，第二，自忆性在系统演化中起多大作用，相对

1) 自然科学基金资助

于外强迫占多少份量，记忆长度有多少，也就是经多长时间后，记忆丧失，不再对未来起作用。

大家知道，一个预报员作预报时，要看最后一个观测实况图以前的好几张图，报的时间越长，向前追踪查看的图越多，也就是这是一种考虑天气演变的多时次预报。如果把预报员预报过程视为一种计算机运算过程，那么预报过程是一种记忆过程。我们下面推导的大气运动自忆性方程在物理上源于不可逆过程动力学，在现象上则是对预报过程的一种数学模拟。

3 自忆性方程

由方程(2)出发，定义希尔伯特空间中的内积，引进记忆函数 $\beta(t)$，可以导得大气运动在空间点 r 的自忆性方程[2]

$$\begin{aligned} \beta(t)x(t) &= \beta(t_{-p})x(t_{-p}) \\ &+ \sum_{i=-p}^{0} x_i^m(\beta(t_i+1)-\beta(t_i)) + \\ &\int_{t_{-p}}^{t} \beta(\tau)F(x,\lambda,\tau)d\tau = \mathrm{SM}+\mathrm{OE} \quad (3) \end{aligned}$$

方程各量中省略了书写 r，式中 x_i^m 为时间间隔 $t_i<t_m<t_{i+1}$ 中的 x 中值。t_0 为数值预报中的初始时刻，t_{-p} 表示 t_0 前的 p 个时刻。由于式(3)强调了局地点上状态变量本身前后的联系，因此称为大气的自忆性方程。式(3)中右端第一和第二项只与局地点的 x 值有关，称为自忆项(SM)，而第三项为其他空间点的影响称为他效项(EO)。由式(3)可见，若已知 t_0, t_{-1}　t_{-p} 时次的记忆函数和观测值，根据式(3)就可以作 t 时刻的预报，所以式(3)构成了一个预测方程，称为自忆模式。若用式(3)作预报时，模式始终要用到 t_0 以前 p 个值。因此式(3)中 $\beta(t)$ 起了记忆 $p+1$ 个时次 x 的效应的功能，所以称 $\beta(t)$ 为记忆函数，这是引进记忆函数的数学理由。引进记忆函数的物理基础是，大气运动方程组中包含热力学方程，由于大气是一个开放系统，不时接受太阳能和发射红外辐射，因此大气运动是一种不可逆过程。不可逆过程研究对物理学的杰出贡献是将记忆概念引进物理中，也就是说，大气的未来发展不仅与上一时刻的、而且与它过去的状态有关，即大气并不遗忘它的过去。对这种记忆性，在这里引进记忆函数来表征。

在导出自忆性方程后，如何求取记忆函数？这是运用自忆性方程进行分析和预测的关键。有 4 种途径可以求取记忆函数

(1)运用观测资料通过最小二乘法来估计记忆函数。

(2)用系统辨识中若干参数辨识法求取，如极大似然递推算法，鲁棒松弛算法，随机迫近法等。

(3)基于模糊集理论对记忆函数进行模糊指派。

(4)基于物理和经验考虑，给记忆函数以特定值，构成一种差分计算格式。

4 自忆性方程与其他预报模型的关系

当用不同途径来求取记忆函数时，自忆性方程就变成不同的预报模型，就此而言，自忆性方程是一个中介媒体，或者比喻性地称它为预报方法的多媒体。

4 1　当取记忆函数 $\beta(t)$ 为简单特定值时，自忆性方程变成一种差分格式。差分格式是数值预报计算的基本手段。可以证明[2]，当 $\beta_0=1$，$\beta_1=1$ 时导得隐式差分格式，当 $\beta_0=1$，$\beta_1=0$ 时导得向前差分格式。可以推断，许多已有的差分格式可以从给定记忆函数来导得，而且给记忆函数以特定的数值后还可以导出新的差分格式，这种格式在时间、空间上还具有内在统一性。

4 2　可以证明[2]，当记忆函数取为变高(压)时，自忆性方程与郑庆林-杜行远的多时刻模式等价。取 $\beta(t)$、$\beta(t-1)$、$\beta(t-2)$ 为某些特定值时，可导得丑纪范的微分方程广义解为原理的多时刻模式。

4 3　可以证明，作者于 1981 年提出的动力-随机差分模式是不考虑中值项加上随机噪声

后得到的自忆性方程的一个特例。

4 4　如下一节所述，当自忆性方程离散化后，不考虑他忆项，就在形式上变为一个自回归模型[3]。从这一模型又可导出均生函数。这就是说，均生函数模型是自忆性方程一种统计求解的结果[4]。

4 5　自忆性方程可由系统的动态方程导得，反过来，当取 $\beta(t_0)=\beta(t)=1$ 时，其他 $\beta=0$ 时，自忆性方程蜕化为对动态方程的通常的积分求解。

4 6　对自忆性方程中的记忆系数进行模糊指派时就得到模糊动态方程，这类模糊动态方程至今尚很少涉足过。

将上述 6 点综合成图 1。

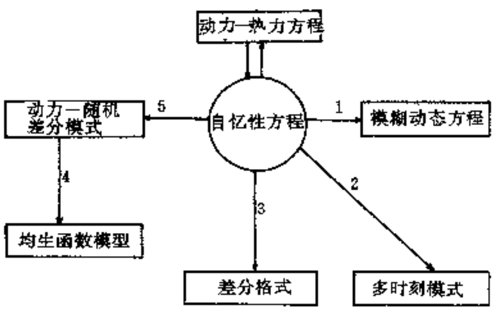

图 1　自忆性方程与其他预报模型的关系
1　模糊指派记忆函数，2　对多时刻就记忆函数取特定值，3　记忆函数取简单特定值，4　单点序列统计求记忆函数，5　时间上统计离散化

5　自忆时序模型

若考虑 3 个时刻 t_0,t_1,t，令 $x(t_m)=x(t_1)$，积分用求和代替，则式(3)可化为

$$x(t) = \varphi_0 x(t_0) + \varphi_1 x(t_1) + \sum_{i=0}^{t}\theta(r_1 t_i)F(x,\lambda,t_i)\Delta t_i \qquad (4)$$

式中 $\varphi_0,\varphi_1,\theta(t_i)$ 为与记忆函数有关的系数[3]。

令 B_t 替代式(4)的最后一项，考虑初始时刻 t_0 前有 p 个时刻，则不难将式(4)推广为

$$x(t) = \sum_{i=0}^{p-1}\varphi_i x(t_i) + B_t \qquad (5)$$

尽管式(5)在形式上与 p 阶自回归模型相同，但其含意是不同的，因为 φ_i 是由 p 个时刻的记忆函数组合而成，具有不可逆过程中记忆特征方面的物理意义。为了能用最小二乘法来求式(5)中的系数 φ_i，令 $B_t\equiv a t$，即 B_t 被假定为白噪声。

5 1　生成函数

设样本量为 N 的时间序列

$$x(t) = \{x(1), x(2),\quad x(N)\} \qquad (6)$$

对时刻 t 用 N 个取样建立自回归

$$x(t) = \varphi_1 x(1) + \varphi_2 x(2) + \quad + \varphi_N x(N) + a_1(t) \qquad (7)$$

从式(7)出发可形式上推导得[4]

$$x(t) = \sum_{i=1}^{l}\alpha_i^l \bar{x}_l(i) + a_l(t),$$
$$l = 1,2,\ \cdot M, M = \mathrm{INT}(N/2) \qquad (8)$$

式中 INT 表示取整。

$$\bar{x}_l(i) = \sum_{j=1}^{n_l} p_i^l(i+jl)x(i+jl)$$
$$n_l = \mathrm{INT}(N/l) \qquad (9)$$

称 $\bar{x}_l(i)$ 为一般生成函数，称 p_i^l 为贡献率 $0\leqslant p_i^l\leqslant 1$，若取 $p_i^l(j)=1/n_l, j=1,2,\quad ,N, i=1,2,\quad ,l, l=1,2,\quad ,M$，则式(11)变为

$$\bar{x}_l(i) = \frac{1}{n_l}\sum_{j=1}^{n_l}x(i+jl)$$
$$i = 1,2,\quad ,l, l = 1,2,\quad ,M \qquad (10)$$

这就是已研究过的均值生成函数[5]，简称均生函数，它是一般生成函数的特例。式(8)启示我们用生成函数 $\bar{x}_l(i)$ 来建立 $x(t)$ 的预测模型，而不是象通常时域建模中做的那样用时间截口上的 x 值来建模。由于 $\bar{x}_l(i)$ 是经取均值得到的，平缓了极值的影响，因此有可能建立稳健的回归。

5 2　均生函数模型

象在傅氏分析中所做的那样，将函数定义域延拓到整个数轴上，即作周期性延拓

$$f_l(t) = \bar{x}_l\left(t - l\cdot\mathrm{INT}\left(\frac{t-1}{l}\right)\right)$$
$$l = 1,2\quad ,M, t = 1,2,\quad ,N,N+1, \qquad (11)$$

我们称 $f_l(t)$ 为均生函数延拓序列或经验周期函数(Empirical periodic function)，其含意是，这一函数是对观测数据通过象式(10)这

样的公式计算随后作周期延拓得到的。

建立原序列 $x(t)$ 与 $f_l(t)$ 间的回归

$$x(t)=\sum_{l=1}^{q}\varphi_l f_l(t)+a(t) \quad q<M \tag{12}$$

它是时间序列 $x(t)$ 与经验周期函数 $f_l(t)$ 间的同时性相关。与自回归模型的延时性相关是不同的。式中 q 为自回归的阶数。由于 $f_l(t)$ 是周期函数，所以当作长时期多步预测时，$x(t)$ 不会趋向平均值。

为了拟合原序列中的高频成分，对原序列 $x(t)$ 进行差分运算，这一运算实际上起着高通滤波的作用，即有

$$\Delta x(t)=x(t+1)-x(t)$$
$$t=1,2,\quad,N-1$$
$$\Delta^2 x(t)=\Delta x(t+1)-\Delta x(t)$$
$$t=1,2,\quad,N-2$$

随之建立相应的序列 $x^{(1)}(t)=\{\Delta x(t)\}$，$x^{(2)}(t)=\{\Delta^2 x(t)\}$，同理计算 $x^{(1)}(t)$ 与 $x^{(2)}(t)$ 的生成函数及其延拓序列 $\overline{x}_l^{(1)}(t)$，$\overline{x}_l^{(2)}(t)$ 和 $f_l^{(1)}(t)$，$f_l^{(2)}(t)$。

为了拟合时序中向上递增或向下递减的趋势，进一步建立累加延拓序列[5]

$$f_l^{(3)}(t)=x(1)+\sum_{i=1}^{t-1}f_l^{(1)}(i+1)$$
$$t=2,\quad,N,l=1,2,\quad,M \tag{13}$$

其中 $f_l^{(3)}(1)=x(1)$。累加延拓实际上是用一阶差分生成函数代替不同时刻差分值。

最后共求得 $4M$ 个经验周期序列 $f_l(t)$，$f_l^{(1)}(t)$，$f_l^{(2)}(t)$，$f_l^{(3)}(t)$，$l=1,2,\quad M$。其目的是为了通过筛选获得高精度拟合的模型，即有更多的自变量可供筛选，达到精选的目的。

通过双评分准则粗选[5]，再用所有可能子集回归精选2)，筛选得 $q(<m)$ 个均生函数序列，随后建立如式(12)的预报方程。

均生函数模型有两个长处，一个是它可以制作长程多步预测，如制作10年气候预测、制作下一年12个月的逐月预报，能报得这样长是目前其他统计模型和数值模式难以做到的。另一个长处是它能很好地拟合和预报极值，也是其他预报方法难以做到的。均生函数模型存在问题是，当样本量不大，如小于30时，建模效果不稳定，而样本量很大，如大于300时，所占计算机内存很大，小的微机上无法计算。

6 应用

从正压涡度方程出发，已导出一个动力-随机差分模式，这一模式是自忆模式在不考虑中值项时的特例。试验研究表明，在制作96小时的中期形势预报时，距平相关系数可达0 75[6]。

从 phillips 二层大气环流模式出发，导出了一个谱展开的动力随机差分模式，用1981年欧洲中期天气预报中心的逐日资料，试作1981年5月9日—6月8日的30天平均场，700hPa 高度场的距平符号率为56%3)。

一个从水汽平衡方程、大气和土壤热量平衡出发的自忆模式正在研制中，旨在制作我国的月降水场和气温场的预报。

用均生函数模型制作长期天气预报和气候预测已在国内广泛使用，将降水场或气温场按逐年-月排列，组成一个连续序列，资料经标准化处理后用经验正交函数 EOF 展开，再对主分量序列用均生函数模型外推，就可“滚动式”地制作多个月的场预报。

7 实例

取徐州、淮阴、清江、信阳等淮河流域9站的6—8月降水量平均值作为预报对象，资料取自中央气象台长期预报科，资料长度为1951—1992年，$N=42$，于1993年3月作当年6—8月汛期降水量预报。$M=42/2=21$。本例为业务预报之成果。

先将数据作标准差标准化，随后用式(10)对原序列 $x(t)$、一阶差分序列 $x^{(1)}(t)$、二阶差分序列 $x^{(2)}(t)$ 计算均生函数，再用式(11)作延拓得序列 $f_l(t)$，$f_l^{(1)}(t)$，$f_l^{(2)}(t)$，用式(13)求得累加延拓序列 $f_l^{(3)}(t)$，这样总共

2) 曹鸿兴，自忆分析与预测，中国气象科学研究院，广东省气象局印，1994

3) 蒋维东，二层斜压随机差分模式及其在长期预报中的应用，成都气象学院硕士论文，1990

得到84个序列，即备选因子。随后通过CSC粗选和精选，求得预报方程

$$
\begin{aligned}
x(t) = & -189\ 692 + 1\ 26067 f_7^{(3)}(t) \\
& - 0\ 234029 f_7^{(1)}(t) - 0\ 634339 f_7(t) \\
& - 0\ 710836 f_9^{(3)}(t) + 1\ 73568 f_9(t) \\
& + 1\ 04220 f_6^{(3)}(t) - 0\ 983295 f_3^{(3)}(t)
\end{aligned}
$$

图2给出了观测与拟合曲线。由图可见，大涝年1954、1956、1963、1965和1991年都报得不错。大旱年除1966年报得较差(距平趋势仍然正确)外，1978、1985和1988年报得也相当好。预报1993年6—8月降水为441mm，实况为463mm，预报正确。

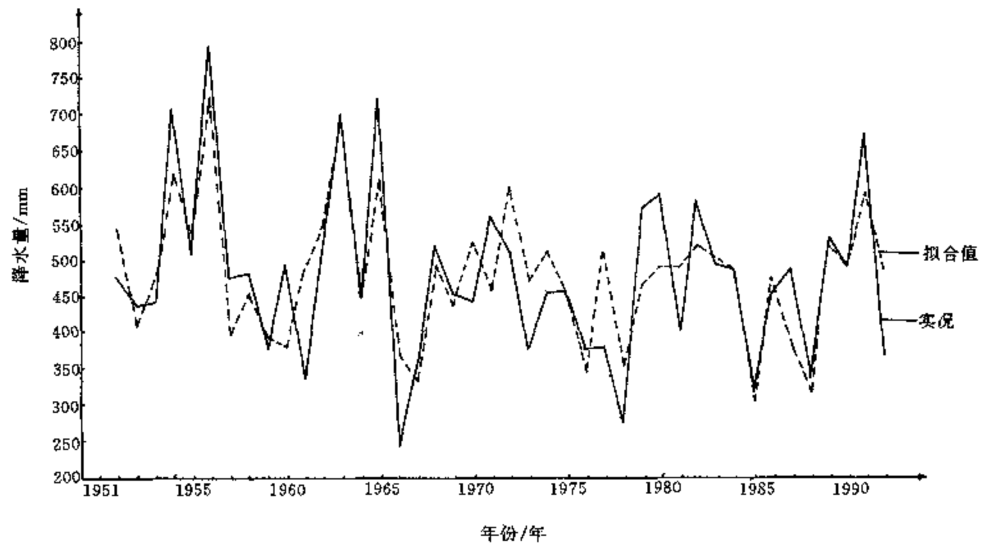

图2 淮河6—8月降水量实测与拟合曲线

实测 实线，拟合 虚线，无虚线处为与实测重合

参考文献

1 尼科里斯，普利高津 探索复杂性 罗久里，陈奎宁译 四川教育出版社，1986

2 曹鸿兴 大气运动的自忆性方程 中国科学B辑，1993，23 104—112

3 曹鸿兴，孙强，刘生长 月海平面气压场的自相关性与可预报性 大气科学增刊，1993，20—28

4 Cao Hongxing，Wei Fengying Time series modelling with empirical period function，in proceedings of Asian Conference on statistical Computing，27 — 31 October 1993，Beijing，1—4

5 Hongxing Cao，Fengying Wei，Multi-step prediction model in time series analysis with mean generating function，in proceedings on 5th International Meeting on Statistical Climatology，22--26 June 1992，Toronto，287—290

6 Zhu Shenming，Cao Hongxing Preliminary study of barotropic difference model applied to weather prediction，Acta Meteor Sinica，1991，5 90—100

Self-memorization Equation and Self-memory Model

Cao Hongxing

(Chinese Academy of Meteorological Sciences，Beijing 100081)

Abstract

Based on memory concept in irreversible process，the self-memorization equation of atmosphere is discussed With several approaches getting the memory function，the equation can become numerical，statistic-dynamic and multi-time models Under conditions of taking account of time series only at one hand and considering the effects at the other as a random process，the self-memory prediction model degenerates into mean generating function model the computational case is given

Key Words Memory multi-time model weather forecasting climate prediction