第16卷 第5期
1992 年 9 月

大 气 科 学
SCIENTIA ATMOSPHERICA SINICA

Vol.16,No.5
Sept. 1992

长期预报的相空间近邻等距法 *

林 振 山

(北京大学地球物理系，北京，100871)

提 要

考虑到确定性系统的外在随机因素和内在随机性(浑沌)所造成的长期预报的不准确性，本文将有关浑沌的理论和数理统计理论结合起来，提出了 d 维相空间距平符号传播的"正、反规则"假说和近邻等距模式．实例检验表明，经调试后距平符号的报准率一般为66%—80%，而相对误差一般不大于 8%.

关键词：浑沌系统；相空间；模式；长期预报.

一、引 言

如何从资料或时间序列对未来做出预测，如长期天气预报、地震预报、病虫害预报等，是当今科学领域中关系到国计民生的一个重要问题． 而长期天气预报则是气象科学的一个重要分支．目前，许多国家的气象部门对长期预报，越来越重视，并积极开展这方面的研究．由于，采用动力模式(如 GCM 模式)不仅计算昂贵，且只能做 1 个月左右的预报．其预报准确率还不能达到业务预报的最低要求．所以，人们在努力发展动力模式的同时，主要还是采用统计预报．遗憾的是，无论是国内，还是国外，目前，统计预报的结果尚不理想．这就促使我们认真思考导致长期预报不准确的因素．

系统的外在随机因素，如随机初始条件、随机参数、随机噪声、物理场的随机起伏等，是我们早已熟悉并公认的导致长期预报不准确的一个重要原因．对此，人们已采用了许多提高模式系统真实性的方法．然而，长期以来，人们一直忽略了导致长期预报不准确的另一重要因素，即系统的内在随机性．所谓系统的内在随机性，指的是某些完全确定性的方程(系统)，不需附加任何随机因素，亦可表现出随机行为，即浑沌．混沌系统的最大特点是系统的演化，对初始条件十分敏感，从而导致长期行为的随机性，即不可预报性．大量的研究表明许多天气系统，气候系统都是浑沌系统．而传统的统计方法，是无法消除或滤去内在随机性的影响．幸运的是，浑沌理论早已指出，在 d 维(d 是大于关联维 D_2 的某一整数)相空间里，系统的短期行为是可预测的．可预测的时间尺度为 $1/h$(h 为系统所有正的 Lyapunov 特征指数之和)．有必要指出：我们常说的长期预报的长期概念，是相对天气过程而言，而它们在 d 维相空间里，则可以对应于一个短期(甚至是最短期)过程，这取决于我们是用什么间隔的时间序列来重建这个 d 维相空间.

1990 年 6 月 3 日收到，1991 年 1 月 6 日收到再改稿.

* 本文系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开放实验室资助课题.

如用 $\Delta t=1$ 个月的时间序列来重建 d 维相空间，则在 d 维相空间里做提前一个月的预报，这对应于该相空间里的一个最短期(只演化第一步)过程．所以，如果在 d 维相空间里(而不是在传统的时、空空间里)，用数理统计方法来研究长期预报问题(浑沌系统是各态历经的，服从统计原理)，将有一定的实际意义．

本文的第二部分阐述了重建相空间的方法；第三、四部分阐述了我们所提出的相空间近邻等距预报法和相空间距平符号传播的正、反规则假说；第五部分则是一些实例和检验结果．

二、相空间

本世纪80年代以前，人们习惯于用时间序列找其平均值、方差及能谱等．而在近几年，人们还利用时间序列来计算系统吸引子的维数和 Lyapunov 指数．从而，揭示系统的演化特性和可预报时间尺度．如果，从时间序列 $\{x_t\}(t=1,2,\cdots,n)$ 里求得关联维 D_2 为非整数，则 $\{x_t\}$ 为一浑沌时间序列，其所表征的系统的行为是浑沌的．

对于浑沌系统，我们可以嵌入一个 $d=\mathrm{INT}(l+D_2)$ 维的相空间．这里的 INT 为取整函数，$l=1,2,\cdots$．

设时间序列的时间间隔 $\Delta t=1$ 单位(如：$\Delta t=1$ 个月)，将该时间序列 $\{x_t\}(t=1,2,\cdots,n)$ 分为两段：x_1 至 x_m 和 x_{m+1} 至 x_n．后者很短，仅供检验用(在正式预报时，则不再分段)，而前者按如下排列，构成相型，

$$\begin{array}{lllll}
x(t_m) & x(t_{m-1}) & \cdots\ x(x_i) & \cdots\ x(t_1+(d-1)\tau) \\
x(t_m-\tau) & x(t_{m-1}-\tau) & \cdots\ x(t_i-\tau) & \cdots\ x(t_1+(d-2)\tau) \\
\vdots & \vdots & \vdots & \vdots \\
x(t_m-(d-1)\tau) & x(t_{m-1}-(d-1)\tau) & \cdots\ x(t_i-(d-1)\tau) & \cdots\ x(t_1) \\
Y(t_m) & Y(t_{m-1}) & \cdots\ Y(t_i) & \cdots\ Y[t_1+(d-1)\tau]
\end{array}$$

这里的 τ 为延滞时间，$Y(t_i)$ 为 d 维相空间里的一个相点(态)，$Y(t_i)$ 在 d 个基矢上有 d 个分量：$x(t_i)$、$x(t_i-\tau)\cdots x(t_i-(d-1)\tau)$．为研究方便，我们分别把它们称为 $Y(t_i)$ 的第1分量、第2分量……第 d 分量．而第1分量的时间标志和相点 $Y(t_i)$ 的时间标志是一致的．$\{Y(t_i)\}$ 构成了 d 维相空间里的一个相型，表示该系统在某一瞬时的状态．而这 $(m-d)$ 个相点的连线，构成了点在 d 维相空间的轨道，该轨道表征了系统状态随时间的演化．

设参考态为 $Y(t_m)$(其第1分量为 $x(t_m)$)，经提前预报时间 T 后，态 $Y(t_m)$ 演化为 $Y(t_m+T)$．只要预报时间 $T\leqslant\tau$，则态 $Y(t_m+T)$ 中，只有其第1分量 $x(t_m+T)$ 是未知的，其余的 $(d-1)$ 个分量 $x(t_m-\tau+T)$、$x(t_m-2\tau+T)$、$\cdots$ $x(t_m-(d-1)\tau+T)$ 都是已知的．故 $x(t_m+T)$ 即为我们的预报对象．当 $T=\tau$ 时，$Y(t_m+T)$ 是由 $Y(t_m)$ 经过 τ 演化而成．若我们在重建相空间时取 $\tau=\Delta t=1$(月)，则我们提前1个月、2个月、3个月的长期预报问题，都对应于 d 维相空间里的短期演化问题．因为 $Y(t_m+T)$ 是由 $Y(t_m)$ 至多演化了3步而成的．是属于浑沌系统里的可预报时间尺度范畴．

三、相空间近邻等距法

1. 最近邻点

用$\|\ \|$表示d维Euclid模，则$\|Y(t_i)-Y(t_j)\|$表示在d维相空间里，态$Y(t_i)$到态$Y(t_j)$的距离.

设参考态$Y(t_m)$的最近邻态为$Y_{\rm nbt}(t_i)$，则$Y_{\rm nbt}(t_i)$与$Y(t_m)$的关系为，

$$\|Y(t_m)-Y_{\rm nbt}(t_i)\|=\min_{j=1,\cdots,(m-1)}[\|Y(t_m)-Y(t_j)\|]\ ,\tag{1}$$

(1) 式里的min为取最小函数.

2. 基本假设

由于我们的预报时间尺度与延滞时间τ同量级，它们相对于时间序列长度而言，都很小. 故而，我们假设：给定时间序列的结构，在预报区域内不变.

3. 相空间近邻等距离假设

设参考态$Y(t_m)$及其最近邻$Y_{\rm nbt}(t_i)$经时间T(即提前预报时间)后，分别发展成为$Y(t_m+T)$和$Y_{\rm nbt}(t_i+T)$. 当$T\leqslant\tau$时，$Y(t_m+T)$的d个分量中，只有第1分量$x(t_m+T)$是未知的. 分别用L_0和L_T表示态$Y(t_m)$和$Y_{\rm nbt}(t_i)$、$Y(t_m+T)$和$Y_{\rm nbt}(t_i+T)$之间的距离，即

$$L_0=\|Y(t_m)-Y_{\rm nbt}(t_i)\|\ ,\tag{2}$$

$$L_T=\|Y(t_m+T)-Y_{\rm nbt}(t_i+T)\|\ .\tag{3}$$

在传统的长期统计预报中，相似预报是人们常用的一种较有效的近似方法. 即利用前期资料选相似，利用相似年的演变来作预报. 将这一原理应用到我们的相空间里，并假设

$$L_0=L_T\tag{4}$$

或$\|Y(t_m)-Y_{\rm nbt}(t_i)\|=\|Y(t_m+T)-Y_{\rm nbt}(t_i+T)\|$.

上式就是我们所提出的相空间近邻等距假设. 显然，利用(4)式，我们可以求出(4)式里唯一的一个未知量$x(t_m+T)$.

四、距平符号预报的相空间正、反规则假说

显然，直接应用(4)式，只能报出$x(t_m+T)$与$x_{\rm nbt}(t_i+T)$的偏差的绝对值. 因为，(4)式运算到最后的形式为

$$[x(t_m+T)-x_{\rm nbt}(t_i+T)]^2=\text{某一常数}\tag{5}$$

或　$x(t_m+T)=x_{\rm nbt}(t_i+T)\pm$ 某一数，

在业务预报时，对(5)式里的正、负号的取舍是至关重要的. 而实际上，目前长期预

报的最大困难则是距平符号的预报．尽管许多国家的业务人员、科研人员采用了许多方法，当前距平符号的报准确率若能保证为60％ — 65％ 就已不错了．把数理统计的相似原理，运用到相空间来，我们提出了以下的距平符号预报的相空间正、负规则假说，试图用以预报距平符号．

当 $T=\tau$ 时，从态 $Y(t_m)$ 演化到 $Y(t_m+T)$ 以及态 $Y_{nbt}(t_i)$ 演化到 $Y_{nbt}(t_i+T)$ 的过程中，具有代表意义的距平符号是 $Y(t_m+T)$ 和 $Y_{nbt}(t_i+T)$ 的第 1 分量的距平符号和第 2 分量的距平符号，以及 $Y(t_m)$ 和 $Y_{nbt}(t_i)$ 的第 d 分量 $x(t_m-(d-1)\tau)$ 和 $x_{nbt}(t_i-(d-1)\tau)$ 的距平符号．这是因为：

(1) $T=\tau$ 时，$Y(t_m+T)$、$Y_{nbt}(t_i+T)$ 的第 2 分量，分别也就是 $Y(t_m)$ 和 $Y_{nbt}(t_i)$ 的第 1 分量．

(2) $T=\tau$ 时，$Y(t_m)$ 和 $Y_{nbt}(t_i)$ 的前 $(d-1)$ 个分量，分别就是 $Y(t_m+T)$ 和 $Y_{nbt}(t_i+T)$ 的第 d 分量．

$x(t_m)$、$x(t_m-(d-1)\tau)$ 和 $x(t_m+T)$ 以及 $x_{nbt}(t_i)$、$x_{nbt}(t_i-(n-1)\tau)$ 和 $x_{nbt}(t_i+T)$ 在相点演化过程中的重要性和特殊性，可以从以下的"暗箱"图解得到直观的说明：

$$\begin{array}{ccc} & \begin{array}{|ll|}\hline x(t_m) & x(t_m-\tau)\ \cdots \\ x(t_m)\ \cdots & x(t_m-(d-2)\tau) \\ \hline\end{array} & x(t_m-(d-1)\tau) \\ x(t_m+T) & & \end{array}$$

图 1　标以记号 $x(t_m)$ 的"暗箱"图

$$\begin{array}{ccc} & \begin{array}{|ll|}\hline x_{nbt}(t_i)\ \cdots & x_{nbt}(t_i-(d-2)\tau) \\ x_{nbt}(t_i)\ \cdots & x_{nbt}(t_i-(d-2)\tau) \\ \hline\end{array} & x_{nbt}(t_i-(d-1)\tau \\ x_{nbt}(t_i+T) & & \end{array}$$

图 2　标以记号 $x_{nbt}(t_i)$ 的"暗箱"图

图 1 说明了第 1 分量为 $x(t_m)$ 的参考态 $Y(t_m)$ 的演化．图 2 则说明了 $Y(t_m)$ 的最近邻态 $Y_{nbt}(t_i)$(其第 1 分量为 $x_{nbt}(t_i)$)的演化．

设 $Y(t_m)$ 和 $Y_{nbt}(t_i)$ 的第 d 分量的距平符号分别为：M、N；$Y(t_m+T)$ 和 $Y_{nbt}(t_i+T)$ 的第 1、2 分量的距平符号分别为 M_T、M_t 和 N_T、N_t．将 N、N_T、N_t 和 M、M_T、M_t 按下列方式排列：

$$N=N_T\cdot N_t, \tag{6}$$

$$M=M_T\cdot M_t. \tag{7}$$

显然，(6)、(7)式分别表征了 $Y(t_m)$ 及其最近邻态 $Y_{nbt}(t_i)$ 在演化过程中的某些特征(可参见图 1 和图2)．为研究上的方便，我们提出了以下的距平符号传播规则：

正规则：正正得正，负负得正，正负得负．(与乘法规则相同)．

反规则：正正得负，负负得负，正负得正．(与乘法规则相反)．

根据统计预报的相似原理，我们提出如下假说：

M 与 N 应该遵守相同的符号传播规则．即若 $N=N_T\cdot N_t$ 遵守正（反）规则，则 $M=M_T\cdot M_t$ 亦遵守正(反)规则．其中若有一个符号为零距平，则该假说自动失效．

我们提出上述假说的根据是，由于 $Y(t_m)$ 和 $Y_{nbt}(t_i)$ 是最近邻关系，它们应该具有某些相似特点．换言之，它们具有相似性的几率要比不具有相似性的几率大．

例 1：设 $d=5$，取广州 1988 年 3 月的平均温度为参考态 $Y(t_m)$，则求得其 $Y_{nbt}(t_i)$ 的第 1 分量为 1985 年 3 月的平均温度．当 $T=\tau=1$ 时，有关的符号图为：

$$Y_{nbt}(t_i):\ (-----+);\quad Y_{nbt}(t_i+1):\ (------).$$

$$Y(t_m):\ (--+-+);\quad Y(t_m+1):\ (M_T--+-).$$

所以：　$N=+$；　　$N_T=-$；　　$N_t=-$

　　　$M=+$；　　$M_t=-$．

由于 $N=N_T\cdot N_t$ 遵守正规则，故假设 $M=M_T\cdot M_t$ 亦遵守正规则；所以：$M_T=$ "$-$"．

例 2：设 $Y_{nbt}(t_i)$、$Y_{nbt}(t_i+T)$、$Y(t_m)$ 和 $Y(t_m+T)$ 的符号图为，

$Y(t_m)$：$(-++0+)$；$Y(t_m+T)$：(M_T-++0)．

$Y_{nbt}(t_i)$：$(-++-+)$；$Y_{nbt}(t_i+T)$：$(+-++-)$．

则：　$N=+$；$N_T=+$；$N_t=-$．

　　　$M=+$；$M_t=-$．

由于 $N=N_T\cdot N_t$ 遵守反规则，故假设 $M=M_T\cdot M_t$ 亦遵守反规则．所以，$M_T=+$．

严绍瑾、彭永清等 [1] 利用广州1873 — 1980 年的月平均气温序列，求得关联维 $D_2=2.3$，表 1、2 是分别以不同的 d 和 τ 抽样计算的结果．

表 1　不同的 d 和 τ 的抽样计算结果

d	τ	T	$x(t_m)$ (℃)	实际 $x(t_m+T)$ 的距平符号	$N=N_T\cdot N_t$ $M=M_T\cdot M_t$	正或反规则	预测结果	报对与否
3	1	1	14.4 (1985.12.)	+ (1986.1.)	$+=-\cdot+$ $+=M_T\cdot-$	反	+	对
3	1	1	28.0 (1985.7.)	+ (1985.8.)	$-=+\cdot+$ $+=M_T\cdot-$	反	+	对
3	2	2	28.0 (1985.7.)	− (1985.9.)	$-=+\cdot-$ $-=M_T\cdot-$	正	+	错
4	1	1	28.0 (1985.7.)	+ (1985.8.)	$-=-\cdot+$ $-=M_T\cdot-$	正	+	对
4	2	2	28.0 (1985.7.)	− (1985.9.)	$+=+\cdot-$ $-=M_T\cdot-$	反	−	对
4	3	3	28.0 (1985.7.)	+ (1985.10.)	$-=-\cdot-$ $+=M_T\cdot-$	反	+	对

平均距平符号报准率 =5/6

表 2　不同的 d 和 τ 的抽样计算结果

d	τ	T	$x(t_m)$ (℃)	实际 $x(t_m+T)$ 的距平符号	$N=N_T\cdot N_t$ $M=M_T\cdot M_t$	正或反规则	预测结果	报对与否
5	1	1	27.7 (1985.6.)	− (1985.7.)	$+=+\cdot+$ $-=M_T\cdot+$	正	−	对
5	2	2	27.7 (1985.6.)	+ (1985.8.)	$-=+\cdot-$ $+=M_T\cdot+$	正	+	对
5	1	1	28.0 (1985.7.)	+ (1985.8.)	$-=+\cdot+$ $-=M_T\cdot-$	反	−	错
5	2	2	28.0 (1985.7.)	− (1985.9.)	$+=-\cdot-$ $+=M_T\cdot-$	正	−	对
5	1	1	13.7 (1985.2.)	− (1985.3.)	$-=+\cdot-$ $+=M_T\cdot-$	正	−	对
5	2	2	13.7 (1985.2.)	− (1985.4.)	$+=-\cdot+$ $+=M_T\cdot-$	反	+	错

平均距平符号报准率 =4/6

表 3　经调试后的预报和检验结果

d	τ	T	$x(t_m)$ (℃)	实际 $x(t_m+T)$ 的距平符号	$N=N_T\cdot N_t$ $M=M_T\cdot M_t$	正或反规则	预测结果	报对与否
5	1	1	24.8 (1987.10.)	+ (1987.11.)	$+=+\cdot+$ $+=M_T\cdot+$	正	+	对
5	1	1	19.7 (1987.11.)	− (1987.12.)	$+=-\cdot+$ $+=M_T\cdot+$	反	−	对
5	1	1	14.7 (1987.12.)	+ (1988.1.)	$+=+\cdot-$ $+=M_T\cdot-$	反	+	对
5	1	1	15.3 (1988.1.)	− (1988.2.)	$+=+\cdot+$ $0=M_T\cdot+$	失效	/	/
5	1	1	13.6 (1988.2.)	− (1988.3.)	$+=-\cdot-$ $+=M_T\cdot-$	正	−	对
5	1	1	14.5 (1988.3.)	− (1988.4.)	$+=-\cdot-$ $+=M_T\cdot-$	正	−	对
5	1	1	20.4 (1988.4.)	+ (1988.5.)	$+=+\cdot-$ $-=M_T\cdot-$	反	−	错
5	1	1	27.0 (1988.5.)	+ (1988.6.)	$+=+\cdot+$ $+=M_T\cdot+$	正	+	对
5	1	1	28.7 (1988.6.)	+ (1988.7.)	$-=+\cdot+$ $-=M_T\cdot+$	反	+	对
5	1	1	27.5 (1988.7.)	− (1988.8.)	$-=+\cdot-$ $-=M_T\cdot+$	正	−	对
5	1	1	28.0 (1988.8.)	+ (1988.9.)	$-=-\cdot-$ $-=M_T\cdot-$	反	−	错
5	1	1	27.5 (1988.9.)	+ (1988.10.)	$+=+\cdot0$ $+=M_T\cdot+$	失效	/	/

平均距平符号报准率 ≃9/12

表3是经调试后，利用广州1951年1月— 1988年12月的月平均温度序列，做了从1987年11月至1988年10月共1年(12个月)的预报($T=\tau=1$个月)和检验结果.

五、检 验

利用相空间距平符号传播正、反规则假说和相空间近邻等距假设，我们可以方便地做出有物理量的长期预报. 表4是在 D=2.3 的基础上[1]，用广州1951年1月至1988年12月的月平均气温时间序列，做出的一些预报和检验结果.

表4 用月平均气温时间序列做出的预报和检验结果

d	τ	T	$x(t_m)$ (℃)	实测 $x(t_m+T)$ (℃)	实际距平符号	预测距平符号	预测 $x(t_m+T)$	相对误差 %	距平符号报对与否
3	1	1	14.4 (1985.12.)	14.4 (1986.1.)	+	+	注1	/	对
3	1	1	28.0 (1985.7.)	28.8 (1985.8.)	+	+	29.0	0.7	对
3	2	2	28.0 (1985.7.)	26.4 (1985.9.)	−	+	29.3	11	错
4	1	1	28.0 (1985.7)	28.8 (1985.8.)	+	+	29.0	0.7	对
4	2	2	28.0 (1985.7.)	26.4 (1985.9.)	−	−	26.9	1.9	对
5	1	1	27.7 (1985.6.)	28.0 (1985.7.)	−	−	27.8	0.7	对
5	2	2	27.2 (1985.6.)	28.8 (1985.8.)	+	+	30	4.1	对
5	1	1	28.0 (1985.7.)	28.8 (1985.8.)	+	−	28.2	2	错
5	2	2	28.0 (1985.7.)	26.4 (1985.9.)	−	−	26.2	0.8	对
5	1	1	13.7 (1985.2.)	14.8 (1985.3.)	−	−	17.5	18	对
5	2	2	13.7 (1985.2.)	20.6 (1985.4.)	−	+	注2	/	错
平均距平符号报准率 =8/11=72.7%									
平均绝对误差 =4.4%									

注1：预报结果 $x_{pred}(t_m+T)$ 的距平符号与距平符号正、反规则预报法的结论相悖，故不做出具体预报，而考虑用其它方法.

注2：预报结果 $x_{pred}(t_m+T)$ 为虚数. 在该点上，相空间近邻等距预报法失效. 但这种情况的出现几率是非常小的.

六、结 语

从上面的实例检验里，可以看出：本文所提出的相空间近邻等距法和相空间距平符号传播的正、反规则假说，不仅具有一定的实用意义，而且是对传统统计预报理论的补充.

由于在重建相空间时，我们可以取 $INT(D_2+1)$、$INT(D_2+2)$、… 维；而对每一个 d 维相空间，我们又可取不同的 τ 值进行调试. 故相空间预报模式有较大的可调性. 此外，取提前预报时间 T 等于 τ，则可十分方便地做出超前 1 个月、2 个月、3 个月的定量长期预报.

致谢：北京大学王绍武先生、刘式达先生，对本文的撰写提出了有益意见，谨此致谢.

参 考 文 献

[1] 严绍瑾等，一维气候时间序列的拓展及其相空间中浑沌吸引子维数的确定，热带海洋学报，5，第 2 期，97 — 105.

The Model of Equal Distance of Near Neighborhood in the Phase Space for Long-term Forecast

Lin Zhenshan

(Geophysics Department, Peking University, Beijing 100871)

Abstract

Taking into accouat of long-term forecast caused by external random factors and internal randomness (chaos)of the deterministic system, the chaos theory is combined with the statistics theory in this paper. Then, a model of equal distance of near neighborhood in the phase space and a hypothesis of positive and anti-positive law about anomaly sign propagation are set up. With these models, the forecast accuracy will reach 66 — 80 per cent, and the relative error is usually less than 5 per cent.

Key words: Chaos system Phase space; Model; Long-term forecast.