第51卷 第3期　　　　气　象　学　报　　　　Vol. 51, No. 3
1993年8月　　　　ACTA METEOROLOGICA SINICA　　　　Aug., 1993

长期预报的相空间分量组合法*

林　振　山

（南京大学大气科学系，210008）

在做长期预报时，我们常用的统计建模理论有这么一个缺点：当建模者事先对系统一无所知；或虽在建模前已经可以决定应变量与变量之间的关系，但赖以建模的数据又太少；我们则无法根据资料来建模，或不可避免地将自己的偏见带入模式中。为此，Ivakhnenko 提出了数据处理的组合法(GMDH)[1]，而其回归方程的阶数是根据某些判据而较客观地自动给出。此外，传统的统计预报理论一般是基于时间这一一维"空间"的。由于一维空间无法"容纳"分数维$D_2>1$的吸引子，故必将丢失许多关于系统吸引子演化的重要信息，从而导致某些方面的预报不准确。为此，我们把GMDH技术与相空间理论结合起来，提出了以下的相空间分量组合法(GMPSC)。

1 相空间

设时间序列$\{x(t_i)\}(i=1,2\cdots n)$所表征的吸引子维数为$D_2$，取相空间的维数为：

$$d=\mathrm{INT}(D_2+l),\ l=1,2,\cdots \tag{1}$$

这里的INT为取整函数。

将该时间序列延拓成d维相空间的一个相型分布：

$$\begin{array}{llll}
x(t_n) & x(t_{n-1}) & \cdots & x(t_1+(d-1)\tau) \\
x(t_n-\tau) & x(t_{n-1}-\tau) & \cdots & x(t_1+(d-2)\tau) \\
\vdots & \vdots & \vdots & \vdots \\
x(t_n-(d-1)\tau) & x(t_{n-1}-(d-1)\tau) & \cdots & x(t_1) \\
Y(t_n) & Y(t_{n-1}) & \cdots & Y(t_1+(d-1)\tau)
\end{array} \tag{2}$$

这里的τ为延滞时间，$Y(t_i)$为相点。我们把$Y(t_i)$的d个分量：$x(t_i),\cdots,x(t_i-(d-1)\tau)$分别叫做$Y(t_i)$的第一、第二、$\cdots$第$d$分量。相点的连线构成相轨道，该轨道表征了系统状态随时间的演化。

2 相空间分量组合法(GMPSC)

将GMDH方法应用于d维的相空间，就构成了所谓的相空间分量组合法(GMPSC)。

将矩阵(2)重新写成如下的形式：

	$y=x_1$	x_2	$\cdots$	x_{d-1}	x_d
(A)	$x(t_n)$	$x(t_n-\tau)$	$\cdots$	$x(t_n-(d-2)\tau)$	$x(t_n-(d-1)\tau)$
	$x(t_{n-1})$	$x(t_{n-1}-\tau)$	$\cdots$	$x(t_{n-1}-(d-2)\tau)$	$x(t_{n-1}-(d-1)\tau)$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$x(t_m)$	$x(t_m-\tau)$	$\cdots$	$x(t_m-(d-2)\tau)$	$x(t_m-(d-1)\tau)$
(B)	$x(l_{m-1})$	$x(t_{m-1}-\tau)$	$\cdots$	$\tau(t_{m-1}-(d-2)\tau)$	$x(t_{m-1}-(d-)\tau)$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$x(t_1+(d-1)\tau)$	$x(t_1+(d-2)\tau)$	$\cdots$	$x(t_1+\tau)$	$x(t_1)$
	(a)	(b)			

(3)

* 1991年6月14日收到原稿。1991年9月6日收到修改稿。

对矩阵(3)输入变量中的每一对x_i和x_j与输出变量y进行如下的多项式回归：

$$y=a^0+b^0x_i+c^0x_j+d^0x_i^2+e^0x_j^2+f^0x_ix_j \tag{4}$$

$$i\neq j,\quad 2\leqslant(i,j)\leqslant d$$

这将产生$(d-1)(d-2)/2$个较高阶变量z_t，即令：

$$z=a^0+b^0x_i+c^0x_j+d^0x^2+e^0x_j^2+f^0x_ix_j$$

则可从矩阵(3)的(b)部分里算出以下的新矩阵：

z_1	z_2	…	$z_{k'}$
$z_{11}(x_2,x_3)$	$z_{12}(x_2,x_4)$	…	$z_{1k'}(x_{d-1},x_d)$
$z_{21}(x_2,x_3)$	$z_{22}(x_2,x_4)$	…	$z_{2k'}(x_{d-1},x_d)$
⋮	⋮	⋮	⋮
$z_{n'1}(x_2,x_3)$	$z_{n'2}(x_2,x_4)$	…	$z_{n'k'}(x_{d-1},x_d)$

(5)

其中：

$$k'=(d-1)(d-2)/2;\quad n'=n-(d-1)\tau/\Delta t \tag{6}$$

这里的Δt为时间间隔，而τ的单位为Δt。

用矩阵(3)中的检验矩阵(B)里的应变量$y=x_1$的矩阵元值，与矩阵(5)里的对应元素值按列计算以下的均方根：

$$r_j=\left[\frac{\sum\limits_{i=m+1}^{n'}(x_{1i}-z_{ij})^2}{\sum\limits_{i=m+1}^{n'}x_{1i}^{\ 2}}\right]^{1/2},\quad j=1,2,\cdots,k' \tag{7}$$

这里已根据矩阵(3)取检验矩阵(B)的行序号从$(m+1)$至n'。

根据调试或经验取一L值，从矩阵(5)中淘汰掉那些$r_j\geqslant L$的列。设有k''列满足$r_j<L$条件，将这k''列所组成的矩阵记为$z_{k''}$。用$z_{k''}$取代矩阵(3)中的(b)部分，从而组成新矩阵：

	$y=x_1$	$z_1\ z_2\cdots\cdots z_{k''}$
(A)		
(B)		
	(a)	(b)

(8)

由于已应用了优选法(去掉$r_j\geqslant L$的列)，故可认为矩阵(8)里的z是比矩阵(3)里的$x_2,\cdots,x_d$具有更好预报能力的变量。我们把z称为系统自组织的第二代产物。记r_j的最小值为R_{imin}。

对矩阵(8)进行如下的运算：

(1) 回归：$y=a^1+b^1z_i+c^1z_j+d^1z_i^2+e^1z_j^2+f^1z_iz_j$ (9)

$$i\neq j,\quad 1\leqslant(i,j)\leqslant k''$$

(2) 计算：$u_{ij}=a^1+b^1z_i+c^1z_j+d^1z_j^2+e^1z_j^2+f^1z_iz_j$ (10)

(3) 计算：

$$r_j^{(2)}=\left[\frac{\sum\limits_{i=m+1}(y_i-u_{ij})^2}{\sum\limits_{i=m+1}^{n'}y_i^2}\right]^{1/2} \tag{11}$$

并求出其最小的$r_j^{(2)}$值，记为R_{2min}。

(4) 去掉矩阵$u_{ij}^{(2)}$中$r_j^{(2)} \geq L$的那些列，用所余下的列所组成的矩阵取代矩阵u_{ij}，从而产生第三代变量u。

(5) 重复(1)—(4)步骤，直至Ri_{min}值从下降变为上升，我们便认为R_{min}曲线降到了它的最低点。记最小的Ri_{min}所对应的迭代次数为W，与其对应的回归方程为：

$$y = A^W + B^W Q + C^W V + D^W Q^2 + E^W V^2 + F^W QV \tag{12}$$

综上所述，我们可知预报模式的形式为：

$$y_{pred} = A + \sum_{i=2}^{d-1} B_i x_i + \sum_{i=2}^{d-1} \sum_{j=2}^{d-1} C_{ij} x_i x_j + \cdots \tag{13}$$

由于GMPSC是在可以容纳系统吸引子的d维相空间里应用GMDH技术的，故GMPSC较GMDH具有以下的优点：

(1) 具有较好的动力学性质(GMDH不具有动力学性质)；(2) 由于l和τ均有一定的取值范围，故GMPSC较GMDH具有较好的可调性；(3) 可以提供有关系统吸引子演化的大量信息；(4) GMPSC与一般的统计建模方法相比，模式具有较好的客观性；所需的资料较少；模式系统不出现高价"病态"回归方程，切实可解。

GMPSC的缺点是：(1)计算量大。正是由于l，τ有一定的取值范围，故要调试出一个较好的模式需花费较多的机时；(2)计算机的贮存量要大，尤其是W数较大时，更是如此；(3)预报效果很大程度上依赖于L的取值，而后者是因人因经验而异的。

3 预测和检验

我们根据广州月平均气温时间序列，做了从1981年1月至1981年12月的试验预报。这里的D_2取2，3[1)]，而l取1，τ取$1\Delta t$。

表 1　Ri_{min}与迭代次数的试验结果

迭代次数	1	2	3	4
Ri_{min}	0.61	0.27	0.09	0.16

表 2　GMPSC 预测结果与实际资料对比

月份	1	2	3	4	5	6	7	8	9	10	11	12
实测℃	14.2	15.1	19.3	24.3	24.3	27.1	27.8	29.4	27.4	23.1	18.4	13.9
GMPSC	13.6	14.5	18.6	22.8	25.1	29.8	29.1	28.9	27.4	22.7	19.2	14.1

表 3　GMPSC与其它方法的距平预测对比

月份	1	2	3	4	5	6	7	8	9	10	11	12
实测	+1.1	+1.3	+1.8	+2.4	−1.5	−0.1	−0.5	+1.2	+0.4	−0.7	−1.4	−1.3
GMDH	+0.8	+0.4	+0.9	+1.9	+0.2	−0.6	+0.2	+0.9	+0.8	−0.3	−0.4	−0.4
统计相似	+2.6	+3.1	+1.1	−0.7	−0.6	+0.6	−0.6	+0.2	−0.2	−0.6	−0.6	−0.8
GMPSC	+0.5	+0.7	+1.1	+0.9	−0.7	+2.6	+0.8	+0.7	+0.4	−1.1	−0.7	−0.9

从表3可以看出，GMDH和GMPSC的距平符号报准率要比统计相似法的高，而GMPSC的平均相

1) 林振山，长期预报的相空间模式和气候层次理论，1991年北京大学博士论文，第30—80页。

对误差要比GMDH的小。

4 结　语

本文所提出的GMPSC技术是一种基于相空间分量的"自组织"的主动式高阶非线性回归建模方法。所得的试验模式对实测资料有较好的拟合能力，个别的试验结果是令人满意和鼓舞的。随着研究的深入和试验规模的扩大，我们认为GMPSC将为长期预报、非线性时间序列的内插和外延提供一个行之有效的工具。

致谢：北京大学刘式达教授和沈心悼博士对本文提供了一些有益的意见和资料，谨此致谢。

参考文献

[1] Farlow S J. Self-organizing Method in Modeling. Dekker, 1984, 80—210.

THE GROUP METHOD OF PHASE SPACE COMPONENTS FOR LONG-TERM FORECAST

Lin Zhenshan*

(*Department of Atmospheric Sciences, Nanjing University*, 210008)

Abstract

The Group Method of Phase Space Components(GMPSC) is presented in the paper by combining Group Method of Data Handling with phase space theory. Some trial tests show that the GMPSC provides an efficient tool for the long-term forecast, the interpolation and the extention of nonlinear time series.

* Present affiliation, Department of Atmospheric Sicences, Nanjing university.