第3卷 第2期
1979 年 6 月

大 气 科 学
SCIENTIA ATMOSPHERICA SINICA

Vol. 3, No. 2
June, 1979

多因子综合的降维法
——非线性映象在天气预报中的应用

曹鸿兴 陈国范
（中央气象局气象科学研究所）

点聚图是一种简便有效的预报工具，但通常只能制作二因子的点聚图，多因子时如何综合为一张点聚图至今在理论上、实践上都没有很好解决．综合更多因子而又能发挥点聚图的长处，自然成为人们工作中提出的一个问题．非线性映象原理是解决这个问题的途径之一．

在统计天气预报中，将多因子压缩为少因子降维的方法，现常用的是将组合因子表为原因子的线性函数，使组合因子能最大限度地表示原因子，属于这种方法的有主成分分析（自然正交分解）、因子分析和典型相关分析．非线性映象是从另一角度一保持数据的内部结构来降维的．我们将在本文中论述它的一般原理，同时给出了在天气预报中应用的实例．

一、基 本 原 理

设由 L 个变量组成的向量 $X_i=(x_{1i}, x_{2i}, \cdots, x_{Li})^T$，$(i=1, 2, \cdots, N)$，$T$ 表示转置．这 N 个样本点组成高维空间 $\mathscr{L}$；现将 N 个点映象到低维空间 $\mathscr{D}$ 中，$\mathscr{D}$ 的维数为 D，即 $Y_i=(y_{1i}, y_{2i}, \cdots, y_{Di})^T$，要求 $D<L$，使得数据的固有构形近似地保持．这里构形一词是指样本点间的几何关系．构形近似保持的含意是，在 $\mathscr{D}$ 中的 N 个点的点间距离相应地近似于 $\mathscr{L}$ 空间中的点间距离．不言而喻，低维空间主要是指一维、二维、三维的，因为我们可以在这样的低维空间中看到直观的图象．

非线性映象的基本作法是，给定一个判据，用解非线性方程的算法，如最速下降法、牛顿-雷费逊法、共轭斜量法、变尺度法等，通过迭代求得判据达极小（或极大）时样本点在低维空间的坐标值，最后画出可供分析的图象．

连续性指数是常用的判据之一，其表达式为[1][2]

$$\kappa=\frac{1}{NF}\sum_{i<j}^{N}W_{ij}\left(\frac{d_{ij}^{*}}{d_{ij}}\right)^2$$

$$=\frac{1}{\left(\sum_{i<j}^{N}\frac{1}{d_{ij}^2}\right)^2}\sum_{i<j}^{N}\frac{1}{d_{ij}^2}\left(\frac{d_{ij}^{*}}{d_{ij}}\right)^2$$

1977 年 10 月 2 日收到．

式中，$\sum_{i<j}^{N} \equiv \sum_{j=2}^{N}\sum_{i=1}^{j-1}$，$d_{ij}^*$ 为 $\mathscr{L}$ 空间中点 i 与 j 之间的距离，d_{ij} 为 $\mathscr{D}$ 空间中点 i 与 j 之间的距离，$NF=\left(\sum_{i<j}^{N}\frac{1}{d_{ij}}\right)^2$ 为标准化因子，$W_{ij}=\frac{1}{d_{ij}^2}$ 为权重系数．κ 的含意是当 $\mathscr{L}$ 中的 X 连续地从一个值移向下一个值时 $\mathscr{D}$ 中的 Y 的变化比当 X 有跳跃时要小，也就是高维空间点间距离和相应低维空间点间距离之间的函数关系是连续的．连续性指数独立于 d_{ij} 的尺度，且当 d_{ij}^*/d_{ij} 对所有 i，j 是常数时达到极小．

另一个非线性映象的判据为[3][4]

$$K=\frac{1}{\sum_{i<j}^{N} d_{ij}^*}\sum_{i<j}^{N}\frac{(d_{ij}^*-d_{ij})^2}{d_{ij}^*} \tag{2}$$

其中 $NF=\sum_{i<j}^{N} d_{ij}^*$，$W_{ij}=\frac{1}{d_{ij}^*}$．$K$ 的含意是使原空间的距离与新空间的距离之差的平方和达最小来求得新空间的数据构形．由于采取了 $1/d_{ij}^*$ 的权重，映象的结果是，低维空间 $\mathscr{D}$ 中的构形，其点间距离较小的差不多等于原空间 $\mathscr{L}$ 中的距离，相反，$\mathscr{L}$ 中点间距离较大的，到了 $\mathscr{D}$ 中有较大的变化．这就是说，用 K 来作非线性映象只能使数据的局地结构得到保持，即邻近点的相对关系在两个空间中是一致的，而远离点的关系则可能有很大的改变．非线性映象的这一结果正是我们所需要的，因为通常在使用平面图示法时，或将图面划为几个区，或找靠近的点作为相似样本，一般考虑邻近点间的关系就可以了．

二、松弛解法

需要指出，在用通常的非线性方程的解法对(1)或(2)求解时，每次迭代都得计算距离 d_{ij}，由于样本点间距离个数等于 $\frac{N(N-1)}{2}$，因此即使对于中等数值的 N 来说，距离矩阵也大得惊人，势必占用大量的计算机内存单元和消耗计算时间．因此，我们在计算中采用了一个求 K 最小的松弛解法，它的计算量比较小，其原理如下：

对判据(2)来说，由于仅当对所有 i、j，$d_{ij}=d_{ij}^*$ 时，$K=0$，因此，我们的思路不是一下子使全体 d_{ij} 等于 d_{ij}^*，而是在迭代的一步中，仅使 K 中的一项，即两点 X_p 与 X_q 的 $(d_{pq}-d_{pq}^*)^2/d_{pq}^*$ 达极小，此等价于极小化 $(d_{pq}-d_{pq}^*)^2$．随后依次轮流使成对的原、新空间距离间的误差平方和极小，通过循环迭代最后达到使 K 最小，这样就能在一定程度上克服上述计算上的困难．

令 Y_p 和 Y_q 为相应于 $\mathscr{L}$ 中的 X_p 和 X_q 在 $\mathscr{D}$ 中的两个点，用梯度法求解，迭代后的新点为 Y_p' 和 Y_q'

$$\begin{aligned} Y_p' &= Y_p - C\nabla_{Y_p}(d_{pq}^*-d_{pq})^2 \\ Y_q' &= Y_q - C\nabla_{Y_q}(d_{pq}^*-d_{pq})^2 \end{aligned} \tag{3}$$

式中 C 为正的订正因子，∇_{Y_p} 和 ∇_{Y_q} 分别为对于 Y_p 和 Y_q 的梯度算子，如用欧氏距离，

令

$$d^*_{pq} = \|X_p - X_q\|^2,\ d_{pq} = \|Y_p - Y_q\|^2$$

则

$$\nabla_{Y_p}(d^*_{pq} - d_{pq})^2 = -4(d^*_{pq} - d_{pq})(Y_p - Y_q)$$
$$\nabla_{Y_q}(d^*_{pq} - d_{pq})^2 = 4(d^*_{pq} - d_{pq})(Y_p - Y_q)$$

因此,(3)变为

$$\begin{aligned} Y'_p &= Y_p + 4C(d^*_{pq} - d_{pq})(Y_p - Y_q) \\ Y'_q &= Y_q - 4C(d^*_{pq} - d_{pq})(Y_p - Y_q) \end{aligned} \tag{4}$$

我们使 $d^*_{pq} = d'_{pq}$ 来决定 C, d'_{pq} 表示 Y'_p 与 Y'_q 间的距离,即有

$$\|Y'_p - Y'_q\|^2 = d^*_{pq}$$

将(4)代入得

$$\|(Y_p - Y_q)[1 + 8C(d^*_{pq} - d_{pq})]\|^2 = d^*_{pq}$$
$$[1 + 8C(d^*_{pq} - d_{pq})]^2 = d^*_{pq}/d_{pq}$$

即解出 C

$$C = -\frac{1}{8}(1 - \sqrt{d^*_{pq}/d_{pq}})/(d^*_{pq} - d_{pq}) \tag{5}$$

将(5)代入(4)得

$$Y'_p = Y_p - \frac{1}{2}(1 - \sqrt{d^*_{pq}/d_{pq}})(Y_p - Y_q) \tag{6}$$

$$Y'_q = Y_q + \frac{1}{2}(1 - \sqrt{d^*_{pq}/d_{pq}})(Y_p - Y_q)$$

容易明白,如果我们调整 Y_p 和 Y_q 使有一个比较好的 d_{pq}, 这时 d_{pk} 和 d_{qk} ($k \neq p$, $k \neq q$) 跟着改变了. 由于不可能对所有 i、j, 使 $d^*_{ij} = d_{ij}$ 成立,所以,我们的策略是使局地结构保持,即当 d^*_{ij} 较小时,使 d_{ij} 非常接近于 d^*_{ij},若 d^*_{ij} 较大时,则不要求 d_{ij} 接近 d^*_{ij},为此,引入校正因子 $\frac{1}{(1 + d^*_{pq})}$,则(6)变为

$$\begin{aligned} Y'_p &= Y_p - \frac{1}{2}\frac{1 - \sqrt{d^*_{pq}/d_{pq}}}{(1 + d^*_{pq})}(Y_p - Y_q) \\ Y'_q &= Y_q + \frac{1}{2}\frac{1 - \sqrt{d^*_{pq}/d_{pq}}}{(1 + d^*_{pq})}(Y_p - Y_q) \end{aligned} \tag{7}$$

校正因子使得对大的 d^*_{pq} 其订正因子 $C = \frac{1}{2}\frac{\sqrt{d^*_{pq}/d_{pq}}}{(1 + d^*_{pq})}$ 的值变小.

具体的做法是,令 $D = 2$, 即固定为对二维平面映象,先任意给定

$$Y_1 = \begin{pmatrix} y_{11} \\ y_{12} \end{pmatrix} \quad Y_2 = \begin{pmatrix} y_{21} \\ y_{22} \end{pmatrix} \cdots\cdots Y_N = \begin{pmatrix} y_{N1} \\ y_{N2} \end{pmatrix}$$

然后按下列顺序系统地迭代

$$\begin{aligned} &(Y_1, Y_2), (Y_1, Y_3), \cdots, (Y_1, Y_N), \cdots, (Y_2, Y_1), (Y_2, Y_3), \\ &\cdots, (Y_2, Y_N), \cdots, (Y_N, Y_1), \cdots, (Y_N, Y_{N-1}) \end{aligned} \tag{8}$$

这样迭代循环一圈称一轮迭代,迭代轮次达到事先指定次数所得 $Y_1, Y_2, \cdots Y_N$ 值即为

在二维平面上的映象.

(7)的迭代公式中只涉及两点间距离，因而计算量较小. 此外，由于应用(8)的顺序，某点刚迭代得的 Y 值就可在紧跟着的迭代中使用，加速了迭代收敛.

三、 在天气预报中的应用

在气象中，应用点聚图(单变量或双变量相关图)作预报被广泛采用，因它直观且能发挥人的判断能力. 因此，如果能将多个预报因子组成的高维空间的图象映射为平面图象，就能起到多因子综合的作用. 我们用(2)式 K 作为判据，用松弛解法对一些预报问题进行了试验，今举下面三例来说明.

例一，暴雨集中期预报

用江西广昌县气象站 1955—1973 年资料，凡 6 月分出现日雨量 >100 毫米或 5 天雨量 >150 毫米的定为有暴雨集中期，预报问题就是在春季确定当年 6 月分是否有暴雨集中期.

预报因子为：

x_1：上年 5 月极端最高气温；x_2：上年 7 月下旬平均气温；x_3：上年 7 月雨量；x_4：当年 3 月日照；x_5：当年 3 月极端最高气温与极端最低气温之差.

经 61 轮迭代，K 值已稳定，$K = 0.28$，二维平面图如图 1，有暴雨集中期(A 类)和无

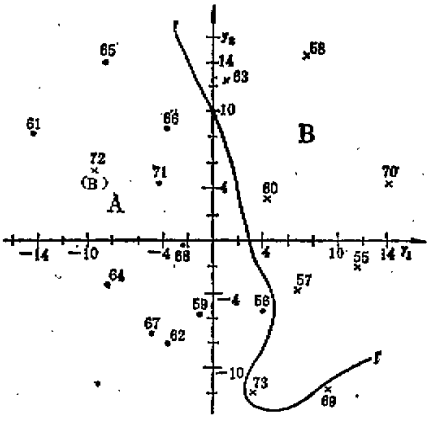

图 1　点旁数字为年分，省略 19

暴雨集中期(B 类)分得很清楚，分别在分界线 l—l' 的两边，只有 1972 年的点子错落在 A 区. 本资料广昌气象站建立过两类判别公式，拟合率为 18/19，这就佐证了非线性映象在天气预报上的可用性.

例二. 月雨量预报

用浙江嘉兴 1953—1975 年资料，预报 6 月分降水量，预报因子为：

x_1：1 月降水量；x_2：2 月降水量；x_3：1 月平均温度；x_4：2 月平均温度；x_5：1—2 月霜日数；x_6：1—2 月雪日数；x_7：月平均风速，1 月与 2 月之差.

用松弛解法迭代 20 轮得二维图象见图 2. $K = 0.23$,图上雨量呈有规律的分布，两个多雨中心,四周为少雨区. 若以小于 110 毫米为少雨,大于 230 毫米为多雨，110—230 毫米为正常,则图上三类区域基本分开,象多雨年 1954、1953、1957 年都集中在一起，少雨的 1958、1960、1968 年也汇聚在一起. 如试报 1975 年，按图中位置应报在 140—180 毫米间,与实况 174 毫米相吻合.

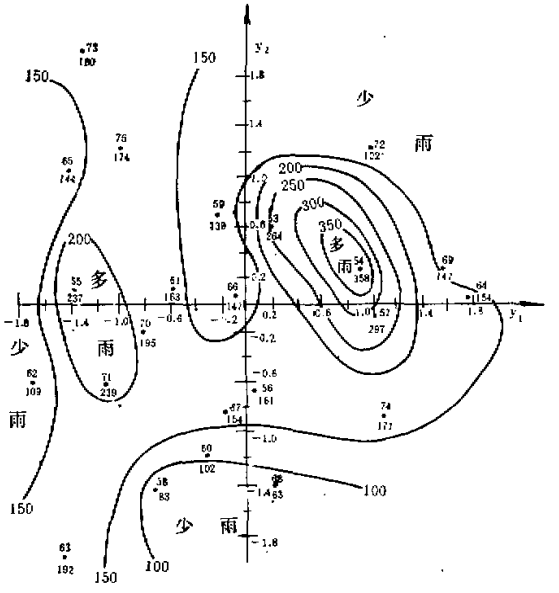

图2 点子上面数字为年分,下面数字为 6 月分雨量(单位：毫米)

例三. 北京 10 月 1 日天气中期预报

用北京单站资料于 9 月 25 日作国庆节天气展望,先用逐步逦归筛选因子得如下 5 个(只用了温度作备选因子):

x_1: 9 月 22 日日平均气温； x_2: 9 月 21 日最高气温； x_3: 9 月 23 日最高气温； x_4: 9 月 24 日最高气温； x_5: 9 月 24 日最低气温.

用 1952—1977 年共 26 个样本资料,迭代 20 轮时 $K = 0.24$，求得各样本点的 y_1, y_2 值,作图(图从略),找与 1977 年相近的点为 1967, 1968, 1972 年的样本,这三年10月 1 日北京均无雨,所以预报 1977 年 10 月 1 日无雨,实况是 10 月 1 日无雨，2 日下了雨. 同样用上述因子建立 y 只取双值的逦归,预报亦为无雨.

由以上三例可看出使用非线性映象作预报的一些特点. 先按天气和物理考虑选定若干因子,通过筛选淘汰一些因子,再把由这些因子 (一般在 10 个以下) 组成的样本资料映象为平面图. 在图上或按预报要求划分几个区域，或分析等值线,或找相近点样本,经分析判断最后把预报结论作出. 与现有的统计方法不同的是，预报样本必须与历史样本一起作映象,才能得到预报样本点在低维空间的坐标值. 此外，现有的统计方法象逦归,判别分析等作预报时只给定一个固定值，而映象结果如同使用点聚图一样，有一定的灵活

性，这时人的判断能力将起相应的作用．现有的统计方法大多受到线性或正态等数学模型的限制，非线性映象除受判据的约束外主要揭示样本资料的固有结构，亦即反映几个预报因子对预报对象的综合效果。

四、讨　论

非线性映象作为多元分析的一种方法，有它独特的思路，它是多元分析图法的一种．由于它能获得直观的图象，为使用带来了方便．在计算条件具备时本法不失为一个方便而又实用的预报方法．

在我们运用松弛解法中发现，给定初值不同，映象所得样本点的二维坐标值是不尽相同的，但点间相对几何关系是基本一致的，即其总的图象是一致的．这是由于算法只保持数据的局地结构，往往一种初值所得结果是另一种初值所得结果的线性变换，即图象的旋转、平移或反射，所以并不影响使用．如果要获得固定结果，可令与预报对象（应变量）相关最大和次大的因子（自变量）分别为第一座标 y_1 和第二座标 y_2 的初值．这样，一般只要迭代20轮或更少K值即稳定．这种作法的统计解释是，以相关大的因子为基底，通过迭代算法把其余因子的作用添加进去，这比 Sammon[4] 建议的以第一、第二主成分作为 y_1、y_2 的初值无疑要省事得多．

对大样本资料，采取压缩样本的办法使之变为小样本后再使用非线性映象法．如用因子分析来选取有代表性的典型样本，或用聚类分析分成若干类后在每一类中抽取一个"种子样本"组成小样本集．

对大样本数，也可采用所谓"架子法"，即在N个样本点中先选M个：$X_1, \cdots, X_M$，$M < N$，应用(7)式的算法得到M个点的二维映象，这M个点称为架子，然后让其余$(N-M)$个点与这M个架子点 $Y_p(1 \leqslant P \leqslant M)$调整距离 d_{pj}，$(M < j \leqslant N)$，而$(N-M)$个点之间不作调整，即 $Y_j(M < j \leqslant N)$不与任何 $Y_k(M < k \leqslant N)$调整．这样就可大大减少计算量和内存，系统迭代的顺序为

$$(Y_1, Y_{M+1}), (Y_2, Y_{M+1}), \cdots, (Y_M, Y_{M+1}), (Y_1, Y_{M+2}), \cdots (Y_1, Y_N), \cdots, (Y_M, Y_N)$$

显然，架子法并不能如实地保持数据的局地结构，且计算结果随取M的不同而不同，但对分类问题来说，只要在平面上的类点分得清楚就行．

非线性映象除能降维外，还能使其具有增加类间分离度的性能，这在图象识别中是研究得较多的，对一些类别判定、分类预报的问题这样做是很需要的。

参 考 资 料

[1] R. N. Shepard, J. D. Carrol, parametric representation of nonlinear data structures, 1966, Multivariate analysis I, Academic Press.

[2] T. W. Calvert, Nonorthogonal projections for feature extraction in pattern recognition IEEE Trans. Comput., Vol. C-19 447—452, May 1970.

[3] C. L. Chang, R. C. T. Lee, A heuristic relaxation method for nonlinear mapping, IEEE Trans. on Systems, Man and Cybernetics. Vol. SMC-3, No. 2, March, 1973.

[4] J. W. Sammon, a nonlinear mapping for data structure analysis IEEE Trans. Computers, Vol. C-18, 401—409, May, 1969.

THE DIMENSION DESCENDING METHOD OF MULTIPLE-PREDICTOR SYNTHESIS——APPLICATION OF NONLINEAR MAPPING TO WEATHER FORECASTING

Cao Hong-xing and Chen Guo-fan

(Meteorological Institute, Central Meteorological Bureau)

The nonlinear mapping is one of the multivariate analysis graphical methods. Using this method, a map higher space may be descend to one of lower dimension while maintaining the local geometric configuration. As the mapping is on a plane, it gives us much convenience for analysis and decision. In this paper a few weather forecasting cases were computed with the relaxation method and compared with the results obtained by using the discriminant analysis and the regression analysis. Since the computation is simple and the presentation is clear and intuitive, the method is a convenient and practical forecasting one when a computer is available.