多维空间点聚方法预报月降水量

苏州地区气象台　江苏师范学院数学系应用组

图解预报是气象台站常用的统计预报方法，具有简单直观的优点。比如预报量 y 与两个前期要素（预报因子）x_1，x_2有关，（x_1，x_2）的一列历史资料在平面上表现为一个点组，在每一点处对应有一个 y 值。若因子挑选适当，这些 y 值在平面中的分布应有较明显的规律，这时，或者划出 y 的等值线，或者按 y的不同等级，把平面划分为互斥的指示区域，就可应用来作预报，这就是平面点聚图。由于作图只能在平面中进行，因此上述方法对多因子的应用受到了限制。有一种逐次图解（或称点聚图过滤）的方案，它是把因子两两分组，在双因子点聚图上获取初步的信息，然后以这些信息作为新的预报因子，再作点聚图分析，这种方法普遍认为还存在一定的问题。下面我们采用了一种直接处理多因子的方法。因为它同样具有简单、直观的特点，所以称之谓多维空间点聚方法。实质上它也是一种相似法，只不过在找相似的思路上有所不同。

一、原　理

在多因子的情况，因子组的一列历史数据表现为多维空间中的一个点组，在每一点处对应有一个 y 值，这些y值在空间中的分散情况也应有一定的规律。如果在挑选因子时进行合理的筛选，虽然点组不能用图表示出来，但它们所具有的分辨力（或者说它们提供的信息）都包含在历史资料的数据之中，问题在于如何实现预报。

如果想类似于平面的情况，按照 y 的一定等级，把空间划分为互斥的指示区域，这时由于不能借助几何直观，在确定分界曲面时，只能使用数学方法来搜索改进，不仅工作量大，又由于因子选择不当，使效果的提高不甚明显。我们可以避开这种以历史资料为中心，以预报量的等级划分为基础的方法，采用一个更自然的概念，以实况资料为中心，考虑历史资料对它的“围绕”、“聚集”情况，这时，事先把空间划分为互斥的固定不变的区域（连带它们的分界曲面）就成为不必要的了。

二、方　法

首先要适当地挑选因子。最合理的标准是：把因子与预报量的历史资料作成点聚图，以能较清楚地看出点子集聚在一条线附近为最好。

其次是找相似。设用上法选得M个预报因子，而数据个数为 N。

记 X_i：$(x_1^{(i)}, x_2^{(i)}, \cdots x_M^{(i)})$　$i=1,2,\cdots,N$

X：$(x_1, x_2, \cdots x_M)$ 是实况数据。

这里数据都已作标准化处理。

计算M维空间中两点：X与X_i的距离：

$$\rho_i = \rho(X, X_i) = \sqrt{\sum_{j=1}^{M}(x_j^{(i)} - x_j)^2}$$

$$i = 1, 2, \cdots, N$$

找 $X_{k_1}, X_{k_2}, \cdots, X_{k_t}$　$(t \geqslant 3)$

使得 $\rho_{k_1} \leqslant \rho_{k_2} \leqslant \cdots \leqslant \rho_{k_t} \ll \rho_i$

$i \neq k_1,\ k_2, \cdots, k_t$

这表明 $X_{k_1}, X_{k_2}, \cdots X_{k_t}$ 围绕X而聚集。

作预报时，设想X_{k_i} 偏离X是由于随机的原因，从而 y_{k_i} 偏离 y 也是随机因素的影响，于是

$$\bar{y} = \frac{\sum_{i=1}^{t} y_{k_i}}{t}$$

可作为 y 的点估计。

$$(\bar{y} - S,\ y + S)，其中\ S = \sqrt{\frac{\sum_{i=1}^{t}(y_{k_i} - \bar{y})^2}{t-1}}$$

可作为 y 的区间估计。

三、操作实例

以预报苏州 5 月分月降水量（1974年）为例。

预报量 y（R_5）与预报因子x_1（上一年 1 月分平均相对湿度），x_2（上一年 3 月分降水量），x_3（上一年 1 月分平均温度），x_4（上一年 3 月分平均气压），x_5（上一年 9 月分平均最低温度）的历史资料为1952—1973 年共 22 年。连同用来预报 1974 年的实况资料均列入表 1 。（X 的资料均经标准化）

表 1

年	序号	x_1	x_2	x_3	x_4	x_5	y (毫米)
1952	1	0.98	0.04	0.24	0.64	−1.23	94.4
1953	2	0.30	0.70	1.58	−0.64	0.25	76.5
1954	3	−0.11	0.72	0.47	−0.42	0.25	190.5
1955	4	1.65	−1.02	1.50	1.38	0.65	64.7
1956	5	0.03	2.00	−2.13	−0.21	1.63	210.9
⋮	⋮		⋮		⋮		⋮
1970	19	0.71	0.03	−0.24	0.64	1.23	146.3
1971	20	−0.24	0.36	−1.26	1.64	1.06	96.1
1972	21	−0.65	−1.00	−0.47	1.43	−0.33	60.7
1973	22	0.43	−0.51	0.87	−0.11	−1.23	146.9
1974	23	0.84	0.11	0.79	0.32	−0.98	

距离平方 ρ_1^2，$\rho_2^2 \cdots\cdots \rho_{22}^2$ 的计算见表 2。

表 2

年	序号	$(x_1^{(1)}-x_1)^2$	$(x_2^{(1)}-x_2)^2$	$(x_3^{(1)}-x_3)^2$	$(x_4^{(1)}-x_4)^2$	$(x_5^{(1)}-x_5)^2$	ρ_i^2
1952	1	0.0196	0	0.3047	0.1011	0.0600	0.4854
1953	2	0.2916	0.3481	0.6209	0.9063	1.5000	3.6669
1954	3	0.9025	0.3721	0.0999	0.5476	1.5000	3.4221
1955	4	0.6561	1.2770	0.5027	1.1220	2.6670	6.2248
1956	5	0.6561	3.5720	8.5150	0.2798	6.8280	19.8509
1958	7	0.0169	0.2209	0.0562	0.0697	0.3260	0.6897
1970	19	0.0169	0	1.0530	0.1011	4.8620	6.0330
1971	20	1.1660	0.0625	4.2070	1.7500	4.1660	11.3515
1972	21	2.2200	1.2320	1.5604	1.2370	0.4264	6.6758
1973	22	0.1681	0.3844	0.0064	0.1789	0.0600	0.7978

因为我们的着眼点在最小的那几个 ρ^2值，故实际计算中若发现某些 $(x_i^{(1)}-x_i)^2$ 已很大，这时相应的 ρ_i^2值就不必具体算出，这样可减少计算量。

由表 2 知，$\rho_1 < \rho_7 < \rho_{22} << \rho_i \quad i \neq 1,7,22$

故预报 1974 年

$$y=\frac{(94.4+88.0+146.9)\text{毫米}}{3}=109.8\text{毫米}$$

区间估计 S＝32，〔109.8－32，109.8＋32〕毫米

实况是 122.8 毫米。

四、讨　论

1. 以上假定存在与 X 特别接近的 Xk_1，$Xk_2 \cdots\cdots Xkt$，若不存在这样的点，预报的精确性就难以保证，这意味着历史资料不足，只要历史资料积累得足够丰富，这个缺点可以解决。

2. 在确定预报值y时，也可求加权平均，如对 y_{k_1} 加上$1/\rho_{k_1}^2$ 的权重。本例如按此法，可得预报值为 106.5毫米。