

# 逐步引进因子场作相似预报

张延亭 单九生

(江西省气象台,南昌 330046)

## 提 要

首先通过数学模型计算出不同样本之间的气象因子场的相似系数和表征气象因子场与天气现象之间关系的相关特征量,然后用相似系数和相关特征量作参数,以预报拟合率为判据,逐个引进因子场,组建出最优相似预报方程,最终作出准确率较高的客观要素预报。

关键词: 因子场 相关特征量 相似系数

## 引 言

在我们的实际预报研究中,无论是对天气形势作纯外推,还是对数值预报产品进行线性回归作统计解释,都面临着大气的非线性干扰问题。作为一个替代工具,使用相似法有其直感性。

相似预报方法的基本思路是:假如两个样本的天气形势和气象要素场是相似的,那么,它们的天气发展过程也相似。所以,从众多历史样本中选取与预报日的形势及要素场最相似的样本,把其对应天气发展过程作为预报日的天气预报结论。目前,有不少气象台站利用相似技术作预报<sup>[1~4]</sup>,虽然基本思路雷同,但由于采用的技术路线不同,其预报效果也不一样。大多数相似预报方法都考虑了不同因子场在预报中的重要性的差异,选取与预报对象关系最为紧密的因子场参与相似计算。但是,这些方法都未涉及如何定量描述因子场重要性的问题,也没有把整个因子场作为一个因子引入预报方程的做法。本文提出了量化地表征气象因子场与天气现象之间关系的相关特征量处理技术和计算方法,同时还提出了用相似系数和相关特征量来组

建最优预报方程的办法。试验表明,这样作出的相似预报效果较好。

## 1 相似系数

假定一气象因子场  $H$ , 由  $n$  个气象要素组成, 样本长度为  $m$ , 用一矩阵表示为:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ h_{m1} & h_{m2} & h_{m3} & \cdots & h_{mn} \end{bmatrix} \quad (1)$$

上式用  $i, j$  作为要素  $h$  的下标,  $h_{ij}$  表示  $H$  场的第  $i$  个样本在第  $j$  个点的要素值。

用下式计算相似系数:

$$A_{ij}(H) = R_{ij}(H) \cdot (1 - D_{ij}(H)/D_{\max}) \quad (2)$$

其中,  $R_{ij}(H)$  是相关系数,  $D_{ij}(H)$  是平均距离, 它们的表达式分别为:

$$R_{ij}(H) = \frac{\sum_{k=1}^n (h_{ik} - \bar{h}_i)(h_{jk} - \bar{h}_j)}{\sqrt{\sum_{k=1}^n (h_{ik} - \bar{h}_i)^2 \cdot \sum_{k=1}^n (h_{jk} - \bar{h}_j)^2}} \quad (3)$$

$$D_{ij}(H) = \sqrt{\frac{1}{n} \sum_{k=1}^n (h_{ik} - h_{jk})^2} \quad (4)$$

$\bar{h}_i, \bar{h}_j$  分别是第  $i$  个样本与第  $j$  个样本的

$n$  个要素的平均值。 $R_{ij}(H)$  为  $H$  场的第  $i$  个样本与第  $j$  个样本之间的相关系数,它表征了两个样本  $H$  场的空间波形相似程度,其值越大,则两个样本  $H$  场的波形越相似。

$D_{ij}(H)$  为第  $i$  个样本与第  $j$  个样本  $H$  场之间的平均距离,它表征了两个样本  $H$  场之间空间位置的接近程度,其值越小,则两个样本的  $H$  场越接近。

$D_{\max}$  是一个可以自定义的适当大小的值。这里定义  $D_{\max}$  是  $H$  场的各样本之间平均距离的最大值,这样定义可以保证式(2)中的  $D_{ij}(H)/D_{\max} \leq 1$ ,从而保证相似系数 ( $A_{ij}$ ) 与相关系数 ( $R_{ij}$ ) 的符号相同。 $A_{ij}(H)$  是第  $i$  个样本与第  $j$  个样本  $H$  场之间的相似系数,它包括相关系数和平均距离两个量,所以它既表征了两个样本  $H$  场之间的波形相似程度,也表征了它们之间空间位置的间距大小,从公式(2)可知, $A_{ij}(H)$  随  $R_{ij}(H)$  的增大和  $D_{ij}(H)$  的减小而增大。也就是说,如果  $R_{ij}(H)$  越大、 $D_{ij}(H)$  越小,则  $A_{ij}(H)$  越大,即  $H$  场之间波形越相似、间距越小,则两个样本的  $H$  场越相似。可见,用  $A_{ij}(H)$  来刻画各样本因子场的相似程度是合适的。

## 2 相关特征量的计算

如何描述因子场与天气现象之间的关系,如何表达因子场在天气预报中的重要性,目前尚无现成的方法可以借鉴。通过以下步骤可计算出表征气象因子场与天气现象之间关系的相关特征量值,用其大小来权衡因子场在天气预报中的重要性。

### 2.1 求最佳相似样本

就  $H$  场而言,用公式(2)可以计算出各样本之间的相似系数,每个样本与其它样本之间的相似系数有  $m - 1$  个,从  $m - 1$  个相似系数中挑选出最大的相似系数 ( $A_{ij}(H)$ ) $_{\max}$  (以下简称为  $A_{id}(H)$ ),  $A_{id}(H)$  对应的样本就是第  $i$  个样本  $H$  场的最佳相似样本,最佳相似样本对应的天气现象用  $W_{id}(H)$  表示。

### 2.2 求样本天气与最佳相似样本天气之间的相关系数——相关特征量

假定各样本对应的天气现象依次是  $W_1, W_2, \dots, W_m$ , 而各样本  $H$  场的最佳相似样本对应的天气现象依次是  $W_{1d}(H), W_{2d}(H), \dots, W_{md}(H)$ ; 可以求出  $H$  场的样本天气序列与最佳相似样本天气序列之间的相关系数:

$$R_w(H) = \frac{\sum_{i=1}^m (W_i - \bar{W})(W_{id}(H) - \overline{W_d(H)})}{\sqrt{\sum_{i=1}^m (W_i - \bar{W})^2 \cdot \sum_{i=1}^m (W_{id}(H) - \overline{W_d(H)})^2}} \quad (5)$$

其中,  $\overline{W_d(H)}$  是  $H$  场的最佳相似样本天气序列的平均值,  $\bar{W}$  是样本天气序列的平均值。

从天气学原理和实践经验都可以这样认为:假如  $H$  场与天气现象  $W$  的关系密切的话,那么,  $H$  场很相似的两个样本所对应的天气现象也一定类似;反过来讲,如果  $H$  场很相似的任意两个样本所对应的天气大都较类似,那么,  $H$  场与天气现象的关系一定很

密切。这就是说,虽然式(5)计算的是天气现象之间的相关系数  $R_w(H)$ , 但  $R_w(H)$  却表征了因子场  $H$  与天气现象  $W$  之间关系的密切程度。所以我们将  $R_w(H)$  称为因子场  $H$  与天气现象  $W$  的相关特征量。

## 3 组建相似预报方程

### 3.1 综合相似预报方程

因为相关特征量表征了因子场与天气现象关系的密切程度,相似系数表征了两样本

间的相似程度,所以,可用各种因子场( $H_1, H_2, \dots, H_n$ )的相关特征量作为各因子场在方程中的权重系数,用相似系数作因子组成线性方程:

$$Y_{ij}(w) = A_{ij}(H_1) \cdot R_w(H_1) + A_{ij}(H_2) \cdot R_w(H_2) + \dots + A_{ij}(H_n) \cdot R_w(H_n) \quad (6)$$

就天气现象  $W$  而言,该方程表征了预报日( $i$ )与相似日( $j$ )间的各种因子场综合关系。这里把式(6)称为综合相似预报方程, $Y_{ij}(w)$ 称为第  $i$  个样本与第  $j$  个样本关于天气现象  $W$  的综合相似系数。

用式(6)可以计算出任一样本与其它样本间的综合相似系数,从中选出最大的综合相似系数,其对应的样本就是要求的最佳相似样本,最佳相似样本对应的天气现象就是预报的天气现象,这也就是各种因子场综合的相似预报结果。

### 3.2 逐步引进因子场,建立最优相似预报方程

预报效果的好坏,取决于相似预报方程的合理性。式(6)作为相似预报方程是否合理呢?众所周知,在建立多元回归方程时,并非因子越多预报效果就越好;另外,由于因子之间还存在相关的问题。因此不一定是单相关最好的几个因子组成的预报方程最优。在利用相关特征量和相似系数组建相似预报方程时也会遇到同样的问题,即:并非引进因子场越多,相似方程预报效果就越好;由于因子场相互间相关性,也不一定是单相关最好的几个因子场构成的相似预报方程最好。如果只引入少量因子场,怎样才能保证预报效果达到最优呢?为此,我们设计了一个因子场逐步引入法,这一方法的基本原理与逐步回归方法类似:从众多因子场中逐步挑选因子场进入预报方程,使预报拟合率(用样本天气与最

佳相似样本对应天气之间的相关系数表示)随着因子场的增加而提高,直至预报拟合率不再由于因子场的引入而提高为止。具体作法如下:

①用相关特征量最大的因子场作为首选因子场引入预报方程:

$$Y_{ij}(w) = A_{ij}(H_{d1}) \cdot R_w(H_{d1}) \quad (7)$$

用下标  $d1$  表示第一个入选的因子场, $R_w(H_{d1})$ 是各种因子场中与天气现象  $W$  相关特征量最大的一个。从前边的推导过程可知,如果只用一个因子场来组建方程,那么,方程(7)一定是预报效果最好的方程,预报拟合率(用  $ny_1$  来表示)就等于相关特征量。

②在式(7)首选因子场的基础上,将  $H_{d1}$  场与其它因子场逐一结合,建立由两个因子场组成的预报方程,分别计算各方程预报的天气与样本天气的拟合率,找出拟合率最大且超过  $ny_1$  的方程:

$$Y_{ij}(w) = A_{ij}(H_{d1}) \cdot R_w(H_{d1}) + A_{ij}(H_{d2}) \cdot R_w(H_{d2}) \quad (8)$$

这样,引入了第二个因子场  $H_{d2}$ ,预报拟合率为  $ny_2$ ,且  $ny_2 > ny_1$ 。

③仿照引入第二个因子场的办法,依次引入第三个、第四个……等因子场,直到预报拟合率不再提高为止,得到预报方程:

$$Y_{ij}(w) = A_{ij}(H_{d1}) \cdot R_w(H_{d1}) + A_{ij}(H_{d2}) \cdot R_w(H_{d2}) + \dots + A_{ij}(H_{dp}) \cdot R_w(H_{dp}) \quad (9)$$

该方程称作最优相似预报方程。试验表明,对短期降水预报来说,一般引入 4~6 个因子场后,拟合率就很难再提高。

### 3.3 举例

这里以南昌 5 月份 20~20 时的降水( $R_{22}$ )预报为例,说明用其前 12 小时(08 时)的高空各要素场来组建最优相似预报方程的具体过程。表 1 的第一列是因子场名,第二列

是各种不同因子场与  $R_{22}$  的相关特征量(即预报拟合率)。

按照前面所述,从表 1 第二列中可以找出,单因子场预报拟合率(即相关特征量)最

大的是 850hPa 风的  $V$  分量场,预报拟合率  $n_{y_1} = 0.3141$ 。因此, $V8$  就是首选因子场,其预报方程可写成:

$$Y_1 = 0.3141 \times A(V8) \quad (10)$$

表 1 5 月份 08 时高空要素场制作南昌 20~20 时的降水预报拟合率

因子场	单个场拟合率	+ V8 场拟合率	+ V8 场 + H5 场拟合率	+ V8 场 + H5 场 + A8 场拟合率	+ V8 场 + H5 场 + A8 场 + U7 场拟合率
D5	0.0503	0.1705	0.3537	0.3006	0.3094
D7	0.2091	0.1768	0.2261	0.2894	0.2941
D8	0.2256	0.2150	<u>0.4205</u>		
H5	0.0610	<u>0.3752</u>			
H7	0.1519	0.1484	0.2020	0.2890	0.2924
H8	0.2504	0.1758	0.1638	0.1766	0.1766
T5	0.1153	0.2085	0.3143	0.2502	0.2540
T7	0.1767	0.2679	0.2217	0.2385	0.2401
T8	$\approx 0$				
L5	0.0320	0.1735	0.3399	0.2981	0.2981
L7	0.0226	0.2456	0.3551	0.3152	0.3180
L8	0.1417	0.1993	0.1736	0.2298	0.2299
F5	$\approx 0$				
F7	0.1445	0.2495	0.2548	0.2257	0.2173
F8	0.2371	0.1034	0.1085	0.0564	0.0564
U5	$\approx 0$				
U7	0.0048	0.2701	0.3937	<u>0.4269</u>	0.2216
U8	0.2228	0.2148	0.2432	0.2179	0.3305
V5	0.0791	0.2560	0.3299	0.3305	0.3970
V7	0.1778	0.2724	0.3201	0.3933	
V8	<u>0.3141</u>				

注:场名的第一位 D:风向、H:高度、T:温度、;L:温度露点差、F:风速、U:东西风分量、V:南北风分量;  
场名第二位表示层次(500hPa、700hPa、850hPa)

用  $V8$  场分别与其它场逐个组合建立预报方程,预报拟合率列入表 1 的第三列。为了与单因子场的预报拟合率保持一致,在多因子场组合作预报时,仍以预报量与实况的相关系数表示预报拟合率。预报拟合率最高的

$V8$  和  $H5$  的组合,预报拟合率  $n_{y_2} = 0.3752$ ; 于是引进第二个因子场  $H5$ ;

$$Y_2 = 0.3141 \times A(V8) + 0.061 \times A(H5) \quad (11)$$

用  $V8$ 、 $H5$  场与其它场逐个组合建立预

报方程,预报拟合率列入表 1 的第四列。预报拟合率最高的是 V8、H5 和 D8 的组合,预报

拟合率  $ny_3 = 0.4205$ ; 于是引进第三个因子场 D8;

$$Y_3 = 0.3141 \times A(V8) + 0.061 \times A(H5) + 0.2256 \times A(D8) \quad (12)$$

同样可引进第四个因子场 U7;

$$Y_4 = 0.3141 \times A(V8) + 0.061 \times A(H5) + 0.2256 \times A(D8) + 0.0048 \times A(U7) \quad (13)$$

预报拟合率  $ny_4 = 0.4269$ 。

当引进第五个因子场时,最高预报拟合率只有 0.3970,拟合率没有提高,反而下降。所以可以认为式(13)是最优预报方程。

为了直观地表示最优预报方程的预报效果,现将部分计算结果(1983~1985 年)以曲

线形式绘成图 1。从图 1 可以看出,尽管预报曲线与实况曲线之间尚有差距,暴雨有空报现象,但该方程的晴雨预报能力较强,晴雨 TS 评分达 86%,方程对大雨以上降水也有一定预报能力,其 TS 评分为 33%。

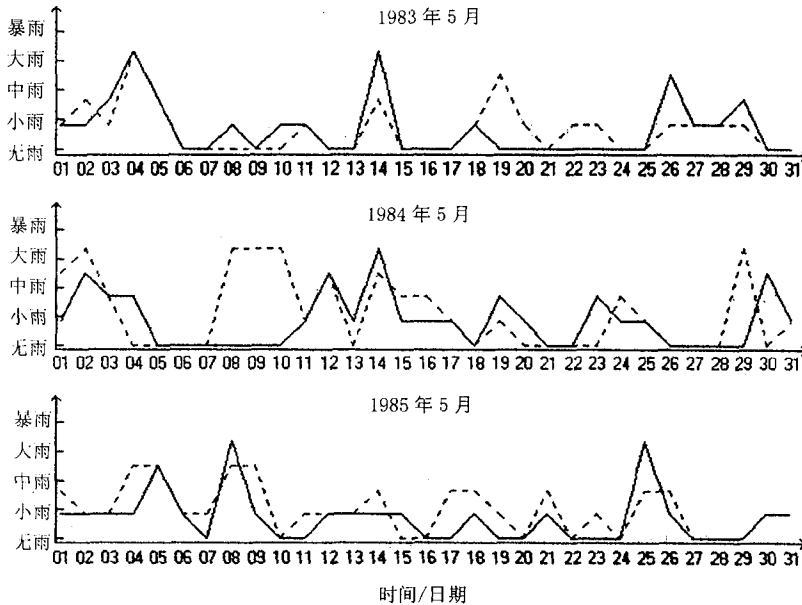


图 1 南昌 5 月预报(虚线)与实况(实线)降水曲线图

#### 4 对比试验分析

为了检验最优预报方程的效果,我们用不同的几种预报方程对 150 个样本进行试报,其结果见表 2。

从表 2 中可以看出,最优预报方程的试报拟合率较其它方程都高。而且试报拟合率还略高于挑选因子场时的历史拟合率,这表

明最优预报方程的性能比较稳定。

表 2 几种方程试报结果

方程名	方程中因子场	试报拟合率
最优预报方程	V8、H5、D8、U7	0.4456
方程 1	V8、H8、F8、D8	0.3850
方程 2	V8、D7、T5	0.1576
方程 3	D7、V7、T7、H7	0.4149

注:拟合率用预报值与实况值的相关系数表示。

## 5 小结

①相关特征量可以客观定量地表示因子场与预报对象的关系,它把复杂的场量关系浓缩转换成一个标量,为建立预报方程创造了条件。

②用逐步引进因子场法组建的最优预报方程较用其它方法建立的预报方程预报准确率高。

③从方法的制作过程可以看出,对各种预报对象都可以求解出其与因子场之间的相关特征量,并组建最优相似预报方程。所以,可用该方法制作多要素客观预报。

④相似预报可用于对数值预报产品的释用。条件是有足够长样本的数值预报产品历史资料,或是对数值预报产品的格点资料进行插值,与其它历史样本资料空间位置一一

对应。

⑤相似预报的准确率与相似范围密切相关<sup>[5]</sup>,试验表明,对短期预报来说,比较适当的相似范围约  $3.0 \times 10^6 \sim 4.0 \times 10^6 \text{ km}^2$ 。

## 参考文献

- 1 陈炎. 相似过滤作中期暴雨预报的探讨. 安徽气象, 1990, (2): 8~11.
- 2 贺懿华, 赵勤炳等. 用富里叶-车贝雪夫多项式制作中期客观相似预报. 湖北气象, 1996, (1): 27~29.
- 3 闫丽凤, 蔡则鹏, 王建国. 场相似在汛期暴雨中的应用. 气象, 1999, 25(2): 33~39.
- 4 Neil D Gorbon[新西兰]. 用相似统计法做甚短期预报. 中尺度天气分析和预报译文集. 北京: 气象出版社, 1998: 168.
- 5 张延亭, 单九生. 影响相似预报效果的几个因素. 江西气象科技, 1997, 20(3): 12~14.

## Similar Forecast with Step by Step Recommanding Factor Field

Zhang Yanting Shan Jiusheng

(Jiangxi Meteorological Observatory, Nanchang 330046)

## Abstract

The similar coefficient of meteorological factors between respective samples was calculated, and the correlated characteristic variables representing the relationship between the meteorological factors field and weather phenomena were given through a mathematical model. Using these results as the coefficient parameters, the optimal equation for similar forecast by introducing factors one by one according to fitting-rate of forecast was established. An accurate and objective forecast of various elements for all counties was made.

**Key Words:** factor field correlated characteristic variables similar coefficient