

# 物理量场在客观定量预报中的使用方法

王双一

(总参气象局气象室)

## 提 要

本文以文献[1]中的论证为依据,提出了物理量场在客观定量预报中的使用方法,使用该方法,既可以避免对备选因子集选取的繁琐过程,又可使所建立的预报方程在各种物理量场中获取到尽可能多的预报信息,使预报方程的预报效果得到明显的改善和提高。使用该方法,有利于统计预报模型建立和实际预报自动化的实现。

## 一、引 言

随着数值预报业务的不断发展和完善,导致了“动力-统计”预报方法的提出和兴起,其中主要方法有完全预报方法和模式输出统计预报方法,在这些预报方法的使用中,更多地涉及到场因子的选取问题。对于场因子的选取方法,已有人作过一些研究<sup>(2, 3, 4)</sup>,现在较为常用的方法是:根据某物理量场的历史样本,求出表征与预报量之间相关性好坏的相关场,根据所得的相关场,选出关键区,对关键区各格点上的物理量值求平均,然后对各关键区内的平均值通过适当的组合,从而得到与预报量之间具有较好相关性的场因子。

从天气、动力学意义上看,某个预报量的形成过程,是各种物理过程在时间上的连续影响作用的过程。对于某个特定时间上的物理量场来说,该物理量场对预报量的形成过程的作用,是分布在不同空域中的物理过程对预报量的连续影响作用。在场因子的选取问题中,如何使所选的场因子体现出各种物理过程对预报量形成的作用特征及使场因子尽量好地反映出对预报量的预报指示性,

这是本文进行详尽研究的一个问题。

## 二、在单个物理量场中的 场因子选取方案

文献[1]中的论证表明,引言所说的目前常用的场因子选取方法,并不能获取对预报量具有最佳预报指示性的场因子。所以必须解决在某个物理量场中获取具有最佳组合信息的场因子的方法。

使用最小二乘法所建立的多元线性回归方程,使预报量残差的平方和达到最小值。当我们使用预报量的残差平方和的大小来衡量因子的好坏时,残差平方和越小,因子的预报性能越好。所以在场因子的选取过程中,以物理量场中各格点值作为备选因子值,使用逐步回归方法来获取各格点上的物理量值与预报量之间的线性回归方程,由该方程对预报量的拟合值作为各格点上物理量值之间的组合值,以该组合值作为所要选取的场因子。使用这一场因子的选取方法,既可以在物理量场中获取尽可能多的预报信息,又使得所获取的场因子是各格点上物理量值的最佳线性组合。

上面的论述指出,对于某一个物理量场

来说,使用逐步回归的方法可以得到能够表示出该物理量场中包含有的具有最大限度预报信息的最佳线性组合因子。在同时有几个物理量场的情况下,根据文献[1]的论证,不必使用逐步回归的方法分别求出各物理量场中的最佳线性组合因子。因为这些组合因子之间所具有的较好相关性,造成了对预报量在相互配合上的不好。所以当使用多个物理量场对某个预报量建立预报模型时,尽管可以在某一物理量场中获取到具有最佳线性组合的场因子,但是如何从所有物理量场中,获取到对预报量的预报具有最佳组合效果的预报信息(或者说是具有最佳因子组合的预报模型),这是需要继续探讨的。

### 三、物理量场的使用方法

在文献[1]中,根据多元线性回归方程形成过程的特性,提出了备选因子集中因子选取的一种可行的方法——配合主因子法。当关于预报量 $Y$ 的回归方程中入选主因子 $X_1$ 后,则因子 $X_2$ 关于 $X_1$ 的残差 $X'_2$ 和 $Y$ 关于 $X_1$ 的残差 $Y'$ 之间的相关系数表达式为:

$$R_{X'_2 Y'} = (R_{X_2 Y} - R_{X_1 Y} R_{X_1 X_2}) / [(1 - R_{X_1 Y}^2)(1 - R_{X_1 X_2}^2)]^{1/2} \quad (1)$$

(1)式中 $R_{X_1 Y}$ 、 $R_{X_2 Y}$ 及 $R_{X_1 X_2}$ 分别为 $X_1$ 和 $Y$ 、 $X_2$ 和 $Y$ 及 $X_1$ 和 $X_2$ 之间的相关系数。

当使用若干个物理量场来建立关于预报量 $Y$ 的预报方程时,为了使所建立的方程能够从所有物理量场中获取到具有最佳组合效果的总体预报信息,我们撇开备选因子集的选取过程,根据上面提出的从单个物理量场中获取场因子的方法,用类似于文献[1]中的配合主因子法来直接建立关于预报量 $Y$ 的预报方程,具体实施过程如下。

设共有 $G$ 个物理量场,分别记为 $_1F, _2F, \dots, _gF$ 。以每一个物理量场中每一个格点值作为一个因子,记该因子为 $_gF_{ij}$  ( $g = \overline{1, G}, i = \overline{1, I_g}, j = \overline{1, J_g}$ ),其中 $I_g, J_g$ 分别为第 $g$ 个物理量场 $_gF$ 所取网格区域的列数

和行数。建立方程时所需的控制参数为最小 $F$ 水平值 $F_a$ ,在某一个物理量场中每一步允许选入的最少因子数 $L_1$ 和允许选入的最多因子数 $L_2$ 。

首先求出主因子场。分别求出各物理量场与预报量 $Y$ 之间的相关场,记 $_gR$ 为物理量场 $_gF$ 与 $Y$ 之间的相关场,求出 $_gR$ 中相关系数绝对值最大的那个格点,即 $_gR_{i_0 j_0} = \max \{ |{}_gR_{ij}|, |i = \overline{1, I_g}, j = \overline{1, J_g}\} (g = \overline{1, G})$ ,设 $_{K_1}R_{i_0 j_0} = \max \{ {}_gR_{i_0 j_0} | g = \overline{1, G} \}$ ,则第 $K_1$ 个物理量场即为所要求的主因子场。求出 $_{K_1}R_{i_0 j_0}$ 的 $F$ 水平值为:

$$F_1 = ({}_{K_1}R_{i_0 j_0})^2 \times (n-1) \div [1 - ({}_{K_1}R_{i_0 j_0})^2] \quad (2)$$

当 $F_1 < F_a$ 时,将无法建立符合条件的回归方程,此时应调整 $F_a$ 的取值,以使 $F_1 \geq F_a$ 。

当 $F_1 \geq F_a$ 时,以第 $K_1$ 个物理量场中的 $I_{K_1} \times J_{K_1}$ 个格点值作为备选因子集中的因子,根据所给定的允许选入的最少因子数 $L_1$ 和最多因子数 $L_2$ ,使用逐步回归方法求出关于 $Y$ 的回归方程,记为 $\hat{Y}_1$ 。

第二步,以回归方程 $\hat{Y}_1$ 对预报量 $Y$ 的拟合值 $\hat{Y}_1$ 作为主因子,根据(1)式求出 $Y$ 关于 $\hat{Y}_1$ 残差和 $_gF (g = \overline{1, G})$ 场关于 $\hat{Y}_1$ 残差场之间的相关场,记为 $_{g\hat{Y}_1}R^{(1)}$ ,求出 $_{g\hat{Y}_1}R^{(1)}$ 场中相关系数绝对值最大的那个格点上的相关系数,即 $_{g\hat{Y}_1}R_{i_0 j_0}^{(1)} = \max \{ |{}_gR_{ij}^{(1)}|, |i = \overline{1, I_g}, j = \overline{1, J_g}\} (g = \overline{1, G})$ ,设 $_{K_2\hat{Y}_1}R_{i_0 j_0}^{(1)} = \max \{ {}_gR_{i_0 j_0}^{(1)} | g = \overline{1, G} \}$ ,则第 $K_2$ 个物理量场即为所要求的备选因子场。求出 $_{K_2\hat{Y}_1}R_{i_0 j_0}^{(1)}$ 的 $F$ 水平值为:

$$F_2 = ({}_{K_2\hat{Y}_1}R_{i_0 j_0}^{(1)})^2 \times (n-2) \div [1 - ({}_{K_2\hat{Y}_1}R_{i_0 j_0}^{(1)})^2] \quad (3)$$

当 $F_2 < F_a$ 时,建立方程过程结束。当 $F_2 \geq F_a$ 时,以第 $K_2$ 个物理量场中的 $I_{K_2} \times J_{K_2}$ 个格点值和 $\hat{Y}_1$ 作为备选因子集,根据所给定的参数 $F_a, L_1$ 和 $L_2$ ,用逐步回归的方法求出关于 $Y$ 的回归方程,记为 $\hat{Y}_2$ 。

同上面原理,依此类推,当到第 $l+1$ 步时,有 $F_{l+1} < F_0$ ,此时建立方程过程结束,由上面各步所建立的预报方程 $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_l$ 原则上均可作为实际使用的预报方程,当然这还需通过独立样本的检验结果或者实际预报结果来确定,一般取独立样本检验结果较好或者实际预报效果较好的那个方程作为实际预报方程。在建立预报方程的过程中,当所取的最小 $F$ 水平值 $F_0$ 较大和拟合样本长度较长时,一般情况下,取最后所建立的那个方程(即 $\hat{Y}_l$ )直接作为实际使用的预报方程是可行的。

(2)、(3)式中的 $n$ 为历史拟合样本数。

从以上建立预报方程的过程中可以看出,在多个物理量场中,通过采用分若干步建立预报方程的方法,其各步首先考虑了在某个物理量场中获取具有最佳组合效果的线性组合因子,构成该组合因子的各格点上的物理量与预报量之间的相关性并不一定都好,这就表明了该选取场因子的方法与以往选取场因子方法的本质区别;以前一步的预报方程对预报量的拟合值作为后一步建立预报方程时的主因子,则考虑了各物理量场对预报量的预报信息之间的相互配合问题。所以以上这种在多个物理量场下建立预报方程的方法,既将分布在不同物理量场中的对预报量具有较好预报指示性且相互之间具有较好配合的物理因子选入到预报方程中,又避免了以往只追求因子与预报量之间的相关性而不能考虑在不同物理量场中所选取的因子之间的相互配合的不利因素。这种建模方法,既避免了备选因子集中因子选取的繁琐而不客观的过程,又能选取到更佳的因子组合,从而使预报方程具有更好的预报效果。

#### 四、物理量场的分布区域及大小的选取

对于物理量场在预报量形成过程中的作用,从天气学和动力学意义上看,制作某一

定点上的预报量在一定预报时效上的预报时,没有必要对全球分布的物理量场进行处理,只要选取对预报量的形成起主要作用的那一有限区域上的物理量场,然后对该有限区域上的物理量场进行处理。这种有限区域上的物理量场的选取,符合天气过程的形成原理,不致于使预报信息大量丢失;另一方面,由于在建立预报模型时,受到计算条件的客观限制,要求包含在物理量场内的网格点数不能过多,这就要对物理量场的所取区域范围有一定的限制。

根据所取预报量的地理位置及预报时效,根据天气分析预报的实践经验和天气动力学的理论基础,我们可以得到对预报量的产生起作用的天气系统及各种环流情况的概念,这种概念也可以在诊断分析的基础上得到。根据这种经验,不难使物理量场的分布区域得以确定。在确定出物理量场的分布区域以后,对于区域大小的确定,应该根据客观计算条件的允许情况及预报量预报时效的长短,区域取得或大或小,一般来说,应尽可能取得大一些,特别是对预报时效较长的预报量,物理量场分布区域应取得尽可能地大。

对于物理量场分布区域的取定问题,以往通常是通过相关场的普查方法来确定,通过对相关场的分析,取相关系数较大的一个或几个区域来作为物理量场的分布区域,这势必会丢失大量对预报量具有较好预报指示性、但与预报量之间的相关性却不好的预报信息。

#### 五、实例

我们将预报量取为北京12月、1月及2月的日平均气温,样本取为1984—1987年的12月、1985—1988年的1、2月,共有12个月中的351个样本(不包括缺测)。利用当天的资料制作第二天(即24小时)的日平均温度的预报。

表 1

各物理量场的选取情况表

序号g	场的物理意义	场名	$I_g, J_g$	区域的选取
1	500hPa位势高度分析场	HA	8, 7	95-130°E, 30-60°N
2	地面气压分析场	PA	8, 7	95-130°E, 30-60°N
3	24小时500hPa位势高度预报场	HF	8, 7	95-130°E, 30-60°N
4	24小时地面气压预报场	PF	8, 7	95-130°E, 30-60°N
5	500hPa地转风u分量场	UGHA	6, 5	100-125°E, 35-55°N
6	地面地转风u分量场	UGPA	6, 5	100-125°E, 35-55°N
7	24小时500hPa预报地转风u分量场	UGHF	6, 5	100-125°E, 35-55°N
8	24小时地面预报地转风u分量场	UGPF	6, 5	100-125°E, 35-55°N
9	500hPa地转风v分量场	VGHA	6, 5	100-125°E, 35-55°N
10	地面地转风v分量场	VGPA	6, 5	100-125°E, 35-55°N
11	24小时500hPa预报地转风v分量场	VGHF	6, 5	100-125°E, 35-55°N
12	24小时地面预报地转风v分量场	VGPF	6, 5	100-125°E, 35-55°N
13	500hPa地转涡度场	VOHA	6, 5	100-125°E, 35(55°N)
14	地面地转涡度场	VOPA	6, 5	100-125°E, 35-55°N
15	24小时500hPa预报地转涡度场	VOHF	6, 5	100-125°E, 35-55°N
16	24小时地面预报地转涡度场	VOPF	6, 5	100-125°E, 35-55°N

物理量场使用欧洲中期天气预报中心的数值预报产品, 包括500hPa及地面的网格分析资料及24小时的预报资料, 计算的各层地转风u、v分量和地转风涡度资料, 共有16个物理量场。根据上面所述的物理量场的分布区域及大小的取定方法, 各物理量场的选取情况见表1。

### (一) 预报模型的建立

使用上面所给定的16个物理量场, 控制参数取为: 最低F水平值 $F_a = 5.0$ , 每一步允许在某个物理量场中选取最少和最多因子数 $L_1$ 和 $L_2$ 分别为 $L_1 = L_2 = 2$ 。根据物理量场的使用方法, 建立日平均气温的24小时预报方程。方程的建立过程如表2所示。

表 2

各方程的建立及其预报效果

参数		$F_a = 5.0, L_1 = L_2 = 2$			
步骤	选入场	所建方程	复相关系数	拟合平均误差	预报平均误差
1	HA	$\hat{Y}_1 = -132.633 + .05882HN_{26} + .18065HA_{38}$	0.5728	2.171°C	2.296°C
2	PA	$\hat{Y}_2 = 35.830 + .10500 \hat{Y}_1 - .28190PA_{38}$	0.7584	1.745°C	1.631°C
3	HA	$\hat{Y}_3 = -31.069 + .10839 \hat{Y}_2 + .05871HA_{39}$	0.7795	1.672°C	1.612°C
4	VGPA	$\hat{Y}_4 = -0.012 + .09801 \hat{Y}_3 - .09293VGPA_{11}$	0.7954	1.614°C	1.563°C
5	VGHF	$\hat{Y}_5 = -0.233 + .10588 \hat{Y}_4 - .06189VGHF_{24}$	0.8063	1.596°C	1.563°C
6	PF	$\hat{Y}_6 = 6.450 + .10221 \hat{Y}_5 - .05502PF_{42}$	0.8163	1.575°C	1.622°C
7	VOPF	$\hat{Y}_7 = 0.126 + .09613 \hat{Y}_6 + .04065VOPF_{20}$	0.8245	1.529°C	1.599°C
8	VOPA	$\hat{Y}_8 = -.019 + .10249 \hat{Y}_7 - .04202VOPA_{31}$	0.8328	1.490°C	1.610°C
9	VOPF	$\hat{Y}_9 = .015 + .10908 \hat{Y}_8 + .04964VOPF_{18}$	0.8399	1.475°C	1.624°C
10	VOPF	$\hat{Y}_{10} = .142 + .09967 \hat{Y}_9 + .04008VOPF_2$	0.8431	1.450°C	1.633°C
11	VOHA	$\hat{Y}_{11} = -.025 + .10236 \hat{Y}_{10} + .08855VOHA_{13}$	0.8469	1.419°C	1.658°C
12	VOPA	$\hat{Y}_{12} = .201 + .09878 \hat{Y}_{11} - .02521VOPA_{26}$	0.8509	1.398°C	1.650°C
13	UGPA	$\hat{Y}_{13} = .229 + .09779 \hat{Y}_{12} - .04448UGPA_{26}$	0.8556	1.373°C	1.654°C
14	VGHA	$\hat{Y}_{14} = .183 + .09993 \hat{Y}_{13} + .04260VGHA_{30}$	0.8591	1.367°C	1.627°C
15	HF	$\hat{Y}_{15} = -12.865 + .10205 \hat{Y}_{14} + .02529HF_8$	0.8627	1.361°C	1.623°C
16	UGHA	$\hat{Y}_{16} = -.994 + .10767 \hat{Y}_{15} - .04201UGHA_{30}$	0.8660	1.350°C	1.623°C

注: 表中所示的方程中的因子的脚标为相应物理量场区域内的网格点序号

## (二) 实际预报效果及比较

使用上述所得的预报方程,对1988—1989年的12月、1989—1990年的1月及1989年2月中的150天(不包括缺测)进行了实际预报,预报结果见表2。由表2可以看出,随着建立方程过程的步数增加和所选入的场因子的增多,方程对预报量的拟合效果不断提高,但由各步所建立的方程的实际预报效果看,由第六步以后的实际预报效果略差于由第四、五步的实际预报效果,这可能是由于所用物理量场的单一性,即所有16个物理量场都是由500hPa位势高度场和地面气压场的导出场,由于资料的限制,没有充分考虑到使用影响平均温度变化的其它各种物理量场,从而导致了方程中因子数的增加反而造成了方程不稳定性的增强,因而造成了随着方程中因子数的增加,并没有使实际的预报效果得以改善。根据前面给出的确定实际使用预报方程的准则,由于由第四、五步所建立的方程的实际预报效果最好,所以使用由第四步或者第五步所得出方程 $\hat{Y}_4$ 或 $\hat{Y}_5$ 作为实际使用的预报方程是恰当的。

在文献[1]中,使用以往通常使用的场因子选取方法,并利用配合主因子法来建立备选因子集,用以上同样的物理量场和历史样本,在最小 $F$ 水平值取为 $F_{\alpha} = 5.0$ 时,使用逐步回归的方法,得到24小时日平均温度的预报方程为[1]:

$$\hat{Y} = -1133.676147 + 2.477134X_1 - 1.616685X_7 + 1.521886X_{11} - 1.795453X_{16} - 0.614396X_{18} \quad (4)$$

(4)式对历史样本的拟合效果为:复相关系数为0.75298,拟合平均误差为1.727℃,对以上的150天进行了实际预报,其预报

的平均误差为1.80℃。

比较(4)式的预报效果和本文所提出的方法所得的预报方程的预报效果,我们可以看出,使用本文的方法,可以明显地提高预报方程的拟合效果和实际预报效果。如果以本文中由第四步所得方程的预报结果和(4)式的预报结果进行比较,可以看出,本文中的方法使实际预报的平均误差降低了0.237℃,即使实际预报误差平均降低了13%,明显提高了对预报量的实际预报效果。在(4)式和使用本文方法所建立的预报方程中的因子构成是不同的,这说明,使用本文的方法可以从不同的物理量场中获取到具有较好相互配合的预报信息,从而使实际预报水平得以提高。

## 六、讨论

根据以往使用物理量场的经验,对物理量场中各格点上的值进行适当的平滑处理,可以提高物理量场中某些单个格点值对预报量的预报效果。对物理量场的这种处理方法,对本文中所提出的物理量场的使用方法仍然适用,即在使用物理量场建立预报方程以前,对各物理量场中各格点上的值进行适当的平滑处理,然后用经过平滑处理后的物理量场来建立预报方程,这更有利于提高所建方程对预报量的预报效果。

本文中仅给出单独使用物理量场时,建立预报模型的处理方法。当使用物理量场和另外一些因子来建立预报模型时,可以将那些因子作为一个或者分成几个物理量场来处理。

## 参考文献

- [1] 王双一,论备选因子集中的因子构成及选取,气象,1990,2期,P16—22。

The application of the physicas fields to the objective forecast

Wang Shuang Yi

(Meteorological Bureau, Headquarters of the General Staff, PLA)

### Abstract

In this paper, the method to apply the physicas fields to the objective forecast is given based on the article[1]. The advantages of this method are given as well as.