

客观预报中多时刻因子的应用及其效果

张建海^{1,2} 王国强²

(1. 南京信息工程大学大气科学系, 210044; 2. 浙江省绍兴市气象局)

提 要

分析了在客观预报中单时刻因子与预报量之间存在着时间尺度上的不匹配问题, 提出使用多时刻因子的必要性和可能性。对比试验表明, 多时刻因子可改善大多数预报方程的质量, 有效率达 90.0%; 多时刻因子使预报误差明显减小, 预报方程的残差平方和平均减少 23.8%。残差平方和的减少在各种场合有所不同, 但呈现出一定的规律性。同时又指出, 在多时刻因子中并不是包含的时刻越多越好, 以防出现“维数灾”现象。

关键词: 多时刻因子 时间尺度 对比试验 有效率 残差平方和

引 言

在动力-统计释用方法中, 以往受客观条件的限制, 往往用某一时效的数值天气预报(以下简记为 NWP)产品制作相应时效的预报。如初始时刻为 $t_0 = 0$, 要制作 0~24 小时的降水和温度预报, 则使用国家气象中心全球模式 T213 和 ECMWF 全球模式预报时效为 $t = 24$ 小时的气象要素场和物理量场的预报资料。同样要制作 24~48 小时时段预报则使用 NWP 产品时效为 $t = 48$ 小时的 NWP 产品资料。显然这里的预报量与预报因子有时间尺度不匹配的问题^[1]。

在实际天气过程中, t_0 时刻的大气状态可以有若干不同的发展趋势。要预测 $t_1 \sim t_2$ 时刻的大气状态, 除了掌握 t_2 时刻的环境场, 还需要了解 t_1 时刻的环境场, 甚至需要了解 $t_1 \sim t_2$ 中间时刻的环境场。某一时段内的降水量是在特定条件下多个中尺度雨团降水量累积的结果。某日的平均气温更与该日的多时刻气温有关, 最高气温和最低气温是一定大气范围内空气热能不断积累和逐步释放的结果。

有人在用 NWP 产品制作热带气旋路径客观预报的研究^[2]中指出:在统计释用过程中,从初始时刻直至预报时刻之间每一时刻的环境场因子都可成为热带气旋路径的预报信息。可见多时刻预报因子的应用有着合理性,因此在构造预报模式时不必局限于同时刻因子,可应用多时刻的 NWP 产品因子,为预报找到更多的有关信息。随着 NWP 的发展,提供的产品日益丰富,在客观天气预报中使用多时刻因子已经有了可能。

1 试验方案设计

数值预报的简单模式可以表示为^[3]:

$$\frac{dA}{dt} = \mathcal{L}(A_1 \cdots) \quad (1)$$

对式(1)积分:

$$A = A_0 + \int \mathcal{L}(A_1 B_1 \cdots) dt \quad (2)$$

其中 \mathcal{L} 是一个算子,也可以看作是函数。对不同时刻,我们可以把式(2)改写为

$$A_k = A_{k-1} + \int \mathcal{L}(A_1 B_1 \cdots)_{k-1} dt \quad (3)$$

其中下标 $k, k-1$ 表示资料时刻, k 为 $k-1$ 时刻的下一步预报时刻。按照式(3)可以写出一系列递推公式:

$$\begin{aligned} A_{k-1} &= A_{k-2} + \int \mathcal{L}(A_1 B_1 \cdots)_{k-2} dt \\ &\quad \dots\dots \\ A_{k-i} &= A_{k-i-1} + \int \mathcal{L}(A_1 B_1 \cdots)_{k-i-1} dt \end{aligned} \quad (4)$$

MOS 方法是用数值预报得到的预报场(要素场和形势场)选取因子,与所要预报的要素进行统计、建立预报方程。现在设要从 A_k 场中读取 n 个因子,和所要预报的某要素 Y 进行筛选、拟合,所选取的因子有 $m_1 (< n)$ 个,其他 $n - m_1$ 个因子不能选上。所得的预报方程为:

$$Y = a_k^0 + a_k^1 A_k^1 + \cdots + a_k^{m_1} A_k^{m_1} \quad (5)$$

我们对余下的因子按式(3)进行分解,从而得到 $n - m_1$ 个 A_{k-1} 场的因子,并从中选取 m_2 个因子,它们 m_1 和 A_k 个场因子一起对 Y 组成预报因子:

$$Y = a_{k-1}^0 + a_{k-1}^1 A_{k-1}^1 + \cdots + a_{k-1}^{m_1} A_{k-1}^{m_1} + a_{k-1}^1 A_{k-1}^1 + \cdots + a_{k-2}^{m_2} A_{k-2}^{m_2} + \cdots + a_{k-2}^{m_1} A_{k-2}^{m_1} \quad (6)$$

其中 a_{k-1}^0 包含了式(5)中的 a_k^0 。对于没有选上的因子再按式(4)进行分解,从而得到 $n - m_1 - m_2$ 个 A_{k-2} 场的因子,再在其中选取因子与前面已经选取的 A_k, A_{k-1} 场因子一起组成 Y 的预报方程。以此类推,最后可以得到一个预报方程:

$$Y = a_0 + a_k^1 A_k^1 + \cdots + a_k^{m_1} A_k^{m_1} + a_{k-1}^1 A_{k-1}^1 + \cdots + a_{k-2}^{m_2} A_{k-2}^{m_2} + \cdots + a_{k-i-1}^1 A_{k-i-1}^1 + \cdots + a_{k-i}^{m_i} A_{k-i}^{m_i} \quad (7)$$

绍兴市气象台短期和中期 MOS 预报系统是已投入业务运行的科研项目。它用国家气象中心全球模式 T213 和 ECMWF 全球模式 2001~2003 年资料建立了绍兴市所属 5 个气象站的预报模型,预报时段有 24~48、48~72、72~96、96~120、120~144 小时共 5 个时段,预报项目有降水概率、强降水概率、平均气温、最高气温、最低气温、相对湿度和降温幅度共 7 项。试验选用 1 月份和 7 月份的资料进行,1 月和 7 月分别代表冬季和夏季,因此共有 $5 \times 5 \times 7 \times 2 = 350$ 个预报方程参加试验。试验时应用预报建模系统分别用多时刻因子资料和单时刻因子资料各自建立 350 个预报方程。本文把使用多时刻因子构建的预报方程简称为多时刻预报方程,把使用单时刻因子构建的预报方程简称为单时刻预报方程,对比多时刻预报方程和单时刻预报方程的质量,以分析使用不同资料的不同效果。

2 效果分析

2.1 多时刻因子的有效率

令多时刻预报方程的复相关系数为 R_2 , 单时刻预报方程的复相关系数为 R_1 , 则 $\Delta R = R_2 - R_1$ 表示同一预报量的两个方程的复相关系数之差。 ΔR 越大表示多时刻预报方程越优于单时刻预报方程,反之亦然。表 1 给出复相关系数的差值等级。表 1 中的“优”表示多时刻预报方程质量优于单时刻预报方程,其余类推。把“优”和“略优”等级的方程数占有所有参加试验的方程总数的百分率称为多时刻因子的有效率。表 2 是 1 月份 175 个方程的质量对比情况,由表 2 可见,多时刻预

报方程的质量普遍比单时刻预报方程好,“优”占45.1%，“略优”占46.3%，而“差”的没有出现,“略差”的仅占0.6%，多时刻因子的有效率为91.4%，可见采用多时刻资料构建预报方程的效果十分明显。7月份的情况也类似(表略)，多时刻因子的有效率为88.5%，1月和7月平均有效率为90.0%。统计事实说明多时刻因子能普遍改善预报方

程的质量。

表1 复相关系数差值等级表

等级	$\Delta R = R_2 - R_1$
优	[0.05, 1]
略优	[0.01, 0.05]
相当	(-0.01, 0.01)
略差	(-0.05, -0.01]
差	[-1, -0.05]

表2 1月份复相关系数之差的等级统计(单位:次)

	绍兴站					诸暨站					上虞站					嵊县站					新昌站					5站合计				
	优	略优	相当	略差	差	优	略优	相当	略差	差	优	略优	相当	略差	差	优	略优	相当	略差	差	优	略优	相当	略差	差	优	略优	相当	略差	差
48小时	5	2	0	0	0	4	3	0	0	0	5	1	1	0	0	3	4	0	0	0	4	3	0	0	0	21	13	1	0	0
72小时	3	4	0	0	0	3	4	0	0	0	3	3	1	0	0	5	1	1	0	0	4	2	1	0	0	18	14	3	0	0
96小时	3	4	0	0	0	5	1	1	0	0	4	3	0	0	0	1	5	1	0	0	4	3	0	0	0	17	16	2	0	0
120小时	2	5	0	0	0	5	1	1	0	0	4	2	1	0	0	3	3	1	0	0	3	4	0	0	0	17	15	3	0	0
144小时	2	5	0	0	0	1	5	1	0	0	3	2	2	0	0	0	5	2	0	0	6	0	1	0	0	6	23	5	1	0
5时次合计	15	20	0	0	0	18	14	3	0	0	19	11	5	0	0	12	18	5	0	0	15	18	1	1	0	79	81	14	1	0
百分比/%	42.9	57.1	0.0	0.0	0.0	51.4	40.0	8.6	0.0	0.0	54.3	31.4	14.3	0	0	34.3	51.4	14.3	0.0	0.0	42.9	51.4	2.9	2.9	0.0	45.1	46.3	8.0	0.6	0.0

2.2 残差平方和对比方法

总离差平方和分解式为:

$$S_{yy} = U + Q \quad (8)$$

其中, $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$, $U = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, 分别为总离差平方和、回归平方和及残差平方和。由式(7)可知,当总离差平方和不变时,回归平方和愈大,残差平方和愈小,可以用残差平方和来衡量多时刻预报方程和单时刻预报方程的效果好坏。已知相关系数的平方等于回归方差和预报量的方差之比^[4]为:

$$R_{xy}^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{U}{S_{yy}} \quad (9)$$

得 $Q = \frac{S_{yy} - U}{S_{yy}} = (1 - R_{xy}^2) S_{yy} \quad (10)$

设 R_1, R_2, Q_1, Q_2 分别为同一预报量用单时刻因子和多时刻因子建立的两个模型的相关系数和残差平方和,令 $E = \frac{Q_1 - Q_2}{Q_1}$ 为多时刻预报方程比单时刻预报方程减小残差平方和的百分率,由式(10)可得

$$E = \frac{R_2^2 - R_1^2}{1 - R_1^2} \quad (11)$$

2.3 多时刻因子作用的残差分析

从前面的分析知道,多时刻因子的有效率为90.0%,即多时刻因子的使用使90.0%的预报方程得到改善。然而这些方程的改善使方程的误差又有多少减小呢?表3给出了1月和7月绍兴市5个气象站平均残差平方和减小的百分率。

表3 1月和7月5站平均残差平方和减小的百分率 E(单位:%)

时效/小时	降水概率		强降水概率		最高温度		最低温度		平均温度		相对湿度		降温幅度		平均	
	1月	7月	1月	7月	1月	7月	1月	7月	1月	7月	1月	7月	1月	7月	1月	7月
48	21.7	31.3	40.4	22.0	40.0	36.6	50.2	32.1	48.4	37.4	20.9	30.3	37.4	17.7	38.4	28.3
72	16.6	24.9	2.9	30.1	32.9	17.6	39.8	27.3	34.7	23.2	19.7	17.0	31.4	22.8	26.6	22.1
96	17.4	19.7	25.8	17.5	38.2	11.7	28.8	10.6	31.3	13.3	16.3	9.1	30.3	18.0	27.2	13.9
120	23.3	26.3	24.9	12.4	26.5	20.4	41.3	19.7	31.8	17.4	17.0	16.5	22.0	23.8	27.1	19.1
144	7.2	15.7	20.0	19.7	22.0	21.7	16.8	23.8	20.6	18.1	3.9	11.1	20.7	20.4	17.1	17.4
平均	17.2	23.6	22.8	20.3	31.9	21.6	35.4	22.7	33.4	21.9	15.6	16.8	28.4	20.5	27.3	20.2

统计结果表明,采用多时刻因子构建预

报方程可明显减少残差平方和,1月平均减

少 27.3%, 7 月平均减少 20.2%, 2 个月平均减少 23.8%, 减少十分明显。但是 E 值的分布并不均匀, 不同预报时效, 不同的季节和不同的预报项目, E 值各不相同。从表 3 末列可见从 48 小时到 144 小时 1 月份的 E 值从 38.4% 下降到 17.1%, 基本上是递减的, 7 月份的情况也类似。比较 2 个月的相应数据, 发现 1 月份的 E 值要大于 7 月份的 E 值。分析不同预报项目的 E 值分布则可看到, 最高温度、最低温度和平均温度的 E 值要大于其它预报项目的 E 值。我们知道, 预报时效短的比预报时效长的预报难度小, 冬季的预报比夏季的预报难度小, 温度类预报项目比降水类预报项目的预报难度小。因此从统计数据反映, 使用多时刻因子减小残差平方和的幅度与预报难度有关, 难度大的幅度小, 难度小的幅度大。这仅是初步统计结果, 还有待于进一步的分析研究。

3 结论和讨论

(1) 客观预报中的单时刻因子与预报量之间存在着时间尺度不匹配的问题。NWP 的发展使得在客观预报中使用多时刻因子有了可能。

(2) 对冬季和夏季共 350 个预报方程的对比试验表明, 多时刻因子对改善预报方程的有效率平均为 90.0%, 多时刻因子减少预报方程的残差平方和平均为 23.8%。从这

两方面可见多时刻因子对预报方程质量的改善作用十分明显。

(3) 多时刻因子减少预报方程残差平方和的幅度在各种情况下有所不同, 减少幅度与预报时效、预报季节和预报项目有关。统计数据反映, 减少幅度与预报难度有关, 难度大的幅度小, 难度小的幅度大。

(4) 多时刻因子的使用, 实际上增加了预报方程的因子数。在样本长度不变的条件下, 过多的因子会产生“拟合好, 预报差”的现象^[5,6], 有人称之为“维数灾”。从我们的试验情况看, 多时刻因子一般不超过 3 个时刻, 样本长度不足的情况下, 以 2 个时刻为宜。如何处理好使用多时刻因子和控制自变量维数的关系, 有待进一步研究。

参考文献

- 1 陈炎, 王兴荣, 刘忠平等. 用相应时段累积因子预报累积降水的初步探讨. 气象科学, 2002, 22(3): 357~360.
- 2 钟元. 数值预报产品统计释用技术在热带气旋路径客观预报中的应用. 热带气象学报, 1998, 14(4): 314~321.
- 3 钮学新等. 关于在动力统计预报中使用多时刻资料的可能性. 浙江气象科技, 1985, (1): 1~3.
- 4 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 2000: 33~34.
- 5 张承福. 人工神经网络在天气预报中的应用研究. 气象, 1994, 20(6): 43~47.
- 6 曹鸿兴, 谷湘潜. 一种集成预报技术—最优调合法. 气象, 1999, 25(5): 3~7.

Application of Multi-time Factors in Objective Forecast

Zhang Jianhai^{1,2} Wang Guoqiang²

(1. Department of Atmosphere Sciences, Nanjing University of Information Science & Technology, 210044;

2. Shaoxing Meteorological Office, Zhejiang Province)

Abstract

The problem of no match on time scale between single-time factors and predictor in objective forecast is analyzed, and the necessity and the possibility of using multi-time factors are put forward. Contrast experiment shows that multi-time factors can improve the quality of prediction equations and make effective rate reach 90.0 percents and reduce sum of residual squares 23.8 percents meanly. The reduction appears some regularity in spite of it's variation in various situation. It is also indicated that it is not benefit to contain more time number in multi-time factors in order to avoid "dimension disaster".

Key Words: multi-time factor time scale contrast experiment effective rate sum of residual squares