

不确定性理论集对分析在预报模型 建立中的应用研究^{*}

薛根元

(南京气象学院, 南京, 210044; 浙江省绍兴市气象局, 绍兴, 312000)

王国强

(浙江省绍兴市气象局, 绍兴, 312000)

摘 要

文中应用集对分析原理, 在模型每次作出预报结论之前先对各预报因子的预报能力进行分析和判断, 对一些预报能力不强, 甚至可能干扰预报模型作出正确预报结论的所谓弱势因子, 进行抑制、甚至消除其影响, 而让其他预报能力较强的因子来决定预报结论。在每次预报中, 模型中哪个因子为弱势是不固定的, 而是随着预报环境的变化而变化, 也就是说, 随着环境的变化, 会有不同因子的作用被抑制或取消, 从而实现预报模型的因子结构动态优化, 增强了模型预报机制的合理性, 提高了模型的预报能力。大量的试验、多年的业务试用以及在数值预报产品释用中的应用例子表明, 这一方法具有较好的效果。

关键词: 集对分析, 不确定性, 因子势态, 数值预报产品释用。

1 引 言

天气预报模型一般由多个气象因子组成, 这组因子即为该预报模型的因子集合。可以进入上述模型的因子数量往往巨大, 但通过一定的技术方法, 可以在众多的气象要素和其他相关要素中为预报模型找到“最佳”因子集合。一般而言, 这一“最佳”因子集合中的每个因子都具有一定的预报能力, 当与其他因子在一起时具有良好的合作预报能力。在预报模型中所谓因子的优良性, 是相对样本整体或者次数众多的预报过程而言, 因而是一个统计性的概念。事实上, 对于样本中的某一个例或者应用中的某一次预报过程则并非完全如此, 有时因“最佳”因子集合中一个或某几个因子的表现不佳而导致该次预报的失败。这就涉及天气预报模型中预报因子结构的动态问题。

在多元分析与多元预报模型建立过程中, 上述现象普遍存在。它可描述为: 用 m 个元素构成某一数学模型 F 对命题 W 进行预测, 在各次预测中, 该

模型中各元素的相对重要性不断变化, 一些元素在某次预测中表现相对重要, 而在其他预测中可能变得次要, 甚至起负作用或干扰作用。多元线性模型和非线性模型通常难以对元素在各次预测中重要性程度的变化进行合理判断和动态处理。

可以将上述有关预报模型中预报因子的变异现象概括为这样一个问题: 假设某一“最佳”气象因子集合由 m 个因子组成, 在某次预报中有 p 个因子表现不佳, 那么能否设法使这 p 个因子在这次预报中失去作用, 而由其余 $(m-p)$ 个表现优良的因子来决定该模型的预报结论呢? 在下次预报中如果又有另外 q 个因子表现不佳, 同样能否使 q 个因子失去作用, 而由其余 $(m-q)$ 个表现优良的因子发挥作用呢? 可见这一思路就是要求能实现预报模型中因子结构的动态变化, 以实现该模型整体预报能力的提高。显然, 要达到这一目的, 用传统的统计学方法所建立的预报模型是无能为力的, 而集对分析 (Set Pair Analysis, 简记为 SPA) 恰好能提供实现上述思路的一种可能的数学手段。

* 初稿时间: 2003 年 3 月 12 日; 修改稿时间: 2003 年 7 月 11 日。

2 集对分析的基本原理

2.1 不确定性问题与 SPA 理论的提出

不确定性是自然界和经济社会中普遍存在的一种客观现象^[1]。为描述和分析这种不确定性,人们提出了研究模糊、随机、中介和信息不完全所致的不确定性的系统理论和方法,即 SPA 方法^[2]。该方法的基本思想就是对客观存在的种种不确定性予以承认,并把不确定性放入一个既确定又不确定的同异反系统进行辩证分析和数学处理。目前,SPA 理论和方法已广泛地应用于现代科学管理、系统控制、决策分析、人工智能、工业设计、军事指挥、农业技术、体育竞赛和社会经济学等诸多领域。

2.2 集对分析的一般原理

所谓“集对”是指具有一定联系的两个集合所组成的对子^[3]。SPA 理论认为,对于集对中两个集合的若干特征可以作同异反系统分析,其联系度 $\mu(w)$ 可表示为

$$\mu(w) = \frac{s}{N} + \frac{f}{N}I + \frac{p}{N}J \quad (1)$$

式中 $\frac{s}{N}$ 为在命题 w 下的同一度, $\frac{f}{N}$ 为差异度, $\frac{p}{N}$ 为对立度,式(1)也可记为

$$\mu(w) = a + bI + cJ \quad (2)$$

其中 $a = \frac{s}{N}$, $b = \frac{f}{N}$, $c = \frac{p}{N}$ 。式(1)或(2)表示在命题 w 下对某集对作分析,它们共有 N 个特性,即 $N = s + p + f$,其中 s 个特性为两个集合所共有, p 个特性为两个集合相对立,而 f 个特性则表现为既不对立又不同一。

联系度 $\mu(w)$ 的表达式(1)或(2)从同、异、反 3 个方面比较两个集合的特性,以描述事物的确定性和不确定性以及它们之间的转换关系。从表达式的形式看,它与复数的表达式相似,也有四则运算等运算法则。式中 I 和 J 既是差异度和对立度的标记,同时又可赋值,以对联系度进行计算。计算时取 $J = -1$,而 I 的取值则根据命题的专业特点而定^[4],其变化区间为 $[-1, 1]$ 。

2.3 天气预报的不确定性

大气运动同时具有确定性和随机性的双重特性^[5,6],这是人们在长期的天气预报实践中得到的辩证认识^[7]。大气运动的随机性以及信息和信息处理的不完全性导致了天气预报的不确定性。概率天气预报和集合天气预报^[8]等预报方法都是多年

来人们为解决大气运动的随机性和天气预报的不确定性问题而进行的有益探索。

有人曾经针对预报模型中自变量和因变量为(0/1)型变量的情况,提出了 SPA 的应用方法^[9]。但(0/1)型变量是特殊性质的离散变量,因此该方法不能用于其他类型的变量。文中试图探索在一般意义下,自变量和因变量为离散型或连续型变量时 SPA 的应用方法,当然作为特例也适用于自变量和因变量为(0/1)变量的情况。

3 预报因子的不确定性分析

集合预报^[10]的观点认为,从稍有差别的一组初始场出发,经过预报模式的积分,其预报轨迹可能发散到较大的区域。近年来集合预报在考虑初始场引起的预报发散的同时,又进一步注意到不同模式引起的发散问题。对于同样的一组初始场应用不同的模式去预报,如果这种发散越小,就认为预报模式的质量越好。

在回归分析中,近邻估计^[11]要在样本中为估计点 x 找到最为相近的 k 个个例,可记为 (x_i, y_i) ,其中 $i = 1, 2, \dots, k$ 。此时 k 个近邻的 k 个 x_i 是一组初始场, k 个 y_i 是一组预测值。如果 k 个近邻的 k 个 y_i 分布越集中,则表示用自变量 x 去预测 y 的效果越好,也就是自变量 x 的预报能力越强,自变量 x 关于 y 量的不确定性越小,此时称 x 为强势因子;反之 k 个 y_i 分布越分散,则表示用 x 去预测 y 的效果越差,也就是 x 的预报能力越弱,自变量 x 关于 y 量的不确定性越大,此时称 x 为弱势因子。如果仅从思路方面看,近邻估计就是一种集合预报^[12]。 k 个近邻的 y_i 分布集中或分散的程度称为离散度,可以用标准差或方差来定量描述。考虑不同数量级别的两组数字的离散度比较,则可采用标准差除以平均值所得到的变异系数^[13]来表示。如用 s 表示标准差, \bar{x} 表示平均值,则变异系数 c_v 表示为

$$c_v = \frac{s}{\bar{x}} \quad (3)$$

有时为了计算方便把变异系数的平方当变异系数使用,此时变异系数 c_v 可表示为

$$c_v = \frac{s^2}{\bar{x}^2} \quad (4)$$

即表示为方差平方与平均值平方之比。因此可以用 k 个近邻 y_i 的离散度来分析推断预报因子在本次预报中的不确定性程度。

4 基于 SPA 的动态因子结构的回归模型

4.1 因子势态的判断方法和实例

如前所述,“最佳”因子集合中的因子在大部分情况表现优良,有时个别因子表现不良。我们把前者称为因子处于强势,把后者称为因子处于弱势。用强势和弱势描述因子的势态,在多元回归分析中势态表示因子在多次预测中重要性的变化。

在非参数回归中,设有一维随机因变量 Y , p 维自变量 X , n 次观测 $X_i, Y_i (i = 1, 2 \dots n)$, 在自变量定义域中有任意点 $X = x$, 则最优估计函数

$$G(x) = E(Y | X = x) \quad (5)$$

即最优估计函数等于 $X = x$ 处的 Y 期望值。在预报业务中,最优估计函数可近似用近邻估计来求取。在 p 维自变量空间中,各次观测 X_i 与估计点 x_0 的距离为 $\rho(x)$, 按距离的近远排列,其中有最近的 k 个观测值,记第 k 个最近观测与 x 的距离为 $\rho_0(x)$ 。由于各自变分量的单位和数量的差异,此时不宜使用几何距离。本文采用统计距离,估计点 x_0 和 X_i 在 p 维自变量空间的位置写成列向量

$$x_0 = (x_{01}, x_{02} \dots x_{0p})' \quad (6)$$

和

$$X_i = (X_{i1}, X_{i2} \dots X_{ip})' \quad (7)$$

则它们之间的统计距离定义为

$$\rho_i(x) = \left[\frac{(x_{01} - X_{i1})^2}{s_{11}} + \frac{(x_{02} - X_{i2})^2}{s_{22}} + \dots + \frac{(x_{0p} - X_{ip})^2}{s_{pp}} \right]^{1/2} \quad (8)$$

其中 $s_{ii} (i = 1, 2 \dots p)$ 为自变量均方差。如果自变量为一维,即当 $p = 1$ 时,那么

$$\rho_i(x) = \frac{|x_0 - X_i|}{\sqrt{s_{11}}} \quad (9)$$

从式(8)和(9)可知,统计距离以样本标准差为基本单位。利用式(9)可以对回归模型中每一因子的各个例分别做近邻分析,找出 k 个近邻 $(X_{(1)}, Y_{(1)}), (X_{(2)}, Y_{(2)}) \dots (X_{(k)}, Y_{(k)})$, 对 k 个 Y 量 $Y_{(1)}, Y_{(2)} \dots Y_{(k)}$ 计算变异系数。在每一个例中,对 m 个因子计算出 m 个变异系数。

这里以浙江省绍兴市气象台(58453站)梅汛期雨量预报模型为例来说明分析与判断过程。模型中涉及样本容量为317,预报因子为6个。整个分析与判断过程分两步进行:(1)对均由数值预报产品格点资料组合而成的因子进行天气学和统计学方面的分析加工,这与本文关系不大,在此不再赘述;(2)根据各个例中每个因子的 c_v 值来判断因子的强势或弱势状态。

表1列出了各因子各个例的变异系数 c_v , 计算时取近邻数为 $k = 7^{[14]}$ 。根据对业务预报中34次天气预报失败的个例分析,由于模型的因子集合中有一个因子为弱势因子而造成预报失败的个例占2/3以上。如果对这一弱势因子作适当处理,则可较好地提高模型的预报能力。对于模型的因子集合中有2个或2个以上因子为弱势因子而造成预报失败的个例,则较难找到提高模型的预报能力的办法。因此这里仅对第一种情况进行分析处理,也就是在每一预报个例中,对因子集合中的 m 个因子只找出一个弱势因子。同时统计表明,在本例中当因子的离散度值大于45.0时,该因子干扰模型作出正确预报结论的可能性较大。因此当某因子的 c_v 值为6个因子 c_v 值中的最大值,且它的 $c_v \geq 45.0$ 时,该因

表1 各预报因子的变异系数

因子号 个例序号	X_1	X_2	X_3	X_4	X_5	X_6
001	17.700	155.800	40.600	11.900	10.700	999.0
002	45.000	37.300	20.500	18.200	66.600	999.000
003	9.700	22.400	55.300	4.500	20.900	3.800
004	60.500	321.000	81.800	51.200	97.800	66.000
005	0.033	0.049	0.049	0.009	0.009	0.002
006	22.600	68.700	18.700	4.400	15.500	3.400
007	26.900	16.600	4.800	17.600	19.600	0.000
008	42.400	30.400	51.300	18.200	5.000	999.000
009	48.300	24.000	93.200	42.200	13.400	17.700
010	46.100	69.900	17.900	9.500	7.900	3.700
...
317	28.700	59.200	21.900	21.400	10.700	9.000

子的预报能力较差, 可判定该因子为弱势因子。如在个例 1 中因子 X_6 的变异系数 $c_v = 999.0$, 为 6 个因子中的最大值, 表明该因子 X_6 处于弱势; 而在个例 2 和 3 中, 可以看到因子 X_6 和 X_3 分别处于弱势; 其余个例可类推。

4.2 适用于天气预报问题的联系度表达式及其应用

在预报模型的因子集合中, 因子之间相互联系、相互制约, 有机地组成一个整体, 缺一不可。在预报时如果发现一个或几个因子处于弱势, 说明它们在模型中的重要性已下降, 甚至可能干扰模型作出正确预报结论, 因而希望让这些因子在这次预报中失去作用。要达到此目的, 显然不能简单地剔除这些因子, 本节试图从 SPA 的同异反分析出发, 用联系度式(1)或(2)导出解决这一问题的方法。

式(2)中, 同一度 a 和对立度 c 是对系统确定性部分的量度, 差异度 b 是对系统不确定性的量度。对于普遍情况下的随机变量 $Y_i (i = 1, 2 \dots n)$ 和相应的某自变量 $X_i (i = 1, 2, 3 \dots n)$, 记 X_i 中的最大值为 X_{\max} 。把第 i 个例 X_i 与 X_{\max} 组成集对, 把第 i 个例 X_i 与 X_{\max} 进行对比分析可知, 当因子处于强势时, 表示此时因子的预报能力较强, 它对于预报量的不确定性较小, 此时它对于预报量的确定性占主导地位, 对预报量的不确定性占从属地位, 因此同一度 $a = s/N = X_i/X_{\max}$, 对立度 $c = p/N = (X_{\max} - X_i)/X_{\max}$, 差异度 $b = f/N = 0$ 。当因子处于弱势时, 自变量 X_i 对因变量 Y_i 的预测“态度”处于含糊状态, 表示此时因子的预报能力较弱, 它对于预报量的不确定性较大, 此时它对于预报量的不确定性占主导地位, 对预报量的确定性占从属地位, 因此同一度 $a = s/N = 0/X_{\max} = 0$, 对立度 $c = p/N = 0/X_{\max} = 0$, 差异度 $b = f/N = X_{\max}/X_{\max} = 1^{[4]}$ 。即有

$$a = \begin{cases} X_i/X_{\max} & x \text{ 处于强势} \\ 0 & x \text{ 处于弱势} \end{cases} \quad (10)$$

$$b = \begin{cases} 0 & x \text{ 处于强势} \\ 1 & x \text{ 处于弱势} \end{cases} \quad (11)$$

$$c = \begin{cases} (X_{\max} - X_i)/X_{\max} & x \text{ 处于强势} \\ 0 & x \text{ 处于弱势} \end{cases} \quad (12)$$

表 2 中所列是第 1 个因子的各个例的联系度, 同样可以统计出其他因子的各个例的联系度(表略)。在此基础上, 可以逐次计算所有个例 6 因子平均联系度(表 3)。

表 2 各个例中预报因子 X_1 的联系度统计

个例序号	同一度 a	差异度 b	对立度 c
1	0.154	0.000	0.846
2	0.060	0.000	0.940
3	0.112	0.000	0.888
4	0.060	0.000	0.940
5	0.133	0.000	0.867
6	0.133	0.000	0.867
7	0.000	1.000	0.000
8	0.042	0.000	0.958
9	0.071	0.000	0.929
10	0.029	0.000	0.971
...
317	0.018	0.000	0.982

在联系度表达式(1)中差异度 $b = f/N$, 它表示在 N 个特征中有 f 个特征表现为既不同一又不对立, 即有 f 个特征对天气事件持“含糊”态度。当因子处于弱势时, 该因子与其勉强参与表态, 还不如放弃“投票权”, 而把预报结论的决定权让给其他强势因子。为此, 这里令

$$I = \frac{\sum s}{\sum s + \sum p} + \frac{\sum p}{\sum s + \sum p} J \quad (13)$$

或表示为

$$I = \frac{\sum a}{\sum a + \sum c} + \frac{\sum c}{\sum a + \sum c} J \quad (14)$$

式中 Σ 表示对 m 个因子中所有的强势子因子求和。在上式中求和与求平均等价。

式(13)和(14)的含义是, 当因子处于弱势时, 它的差异度的 f 个特征按一定比分配给它的同一度和对立度, 这个比值就是其他几个强势因子的平均同一度与平均对立度之比。把式(13)和(14)分别代入式(1)和(2), 可得

$$\mu = \frac{s}{N} + \frac{f}{N} \frac{\sum s}{\sum s + \sum p} + \left(\frac{p}{N} + \frac{f}{N} \frac{\sum p}{\sum s + \sum p} \right) J \quad (15)$$

或

$$\mu = a + b \frac{\sum a}{\sum a + \sum b} + \left(c + b \frac{\sum c}{\sum a + \sum c} \right) J \quad (16)$$

表 3 6 个预报因子的平均联系度

个例序号	平均同一度 \bar{a}	平均差异度 \bar{b}	平均对立度 \bar{c}
1	0.106	0.167	0.727
2	0.087	0.167	0.746
3	0.157	0.000	0.848
4	0.099	0.167	0.734
5	0.062	0.333	0.605
6	0.209	0.167	0.625
7	0.135	0.000	0.865
8	0.077	0.167	0.757
9	0.074	0.167	0.759
10	0.109	0.167	0.724
...
317	0.057	0.167	0.776

按 SPA 的规定, 取 $J = -1$, 故有

$$\mu = \frac{s-p}{N} + \frac{f}{N} \frac{\sum s - \sum p}{\sum s + \sum p} \quad (17)$$

或

$$\mu = (a-c) + \frac{b(\sum a - \sum c)}{\sum a + \sum b} \quad (18)$$

式(17)和(18)是适用于气象问题的多元回归预报模型的联系度表达式。应用式(17)和(18)对表 2 和表 3 的数据进行计算, 可得到各因子的联系度值, 详见下表 4。用表 4 中各预报因子的值作为新的因子,

可以重新建立新的预报数学模型, 该模型即可用来对 y 值进行估计。

5 效果分析和讨论

在为预报模型选择一组性能优良的因子集合问题上^[15, 16], 回归分析为我们提供了许多内容丰富的理论和方法。如果不是滥用这些方法, 而是根据命题的特点有针对性地择用这些方法, 一般可以受到比较满意的效果。然而在选择模型的因子集合后, 又如何在预报中合理地使用这些因子? 换言之当因子处于强势时要发挥它的作用, 当因子处于弱势时则要抑制它的作用。其实合理使用因子与合理选择因子同样十分重要, 本文通过应用 SPA 理论和方法, 对此作了探索。

从 1997 年开始, 绍兴市气象台在降水概率预报、雨量预报和大一暴雨预报方面的大量试验和业务试运行表明, 这种方法有效地提高了天气预报的准确率。如在大一暴雨预报的应用中, 拟合的成功临界指数 CSI(Critical Success Index) 达到 0.56; 2002 年度业务运行的 CSI 达到 0.53。对全年共出现的 13 个大一暴雨报对了 10 个, 漏报的 3 个都是雨量相对较小的降水, 雨量较大的降水则一个未漏。下面从试验的情况来分析这一工作的效果。

表 4 预报因子的联系度

个例	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
1	-0.692	-0.772	-0.684	-0.750	-0.836	-0.745
2	-0.880	-0.870	-0.791	-0.864	-0.734	-0.866
3	-0.776	-0.674	-0.588	-0.762	-0.792	-0.526
4	-0.880	-0.358	-0.762	-0.898	-0.806	-0.868
5	-0.734	-0.814	-0.814	-0.788	-0.804	-0.928
6	-0.734	-0.499	-0.548	-0.658	-0.844	-0.022
7	-0.754	-0.626	-0.510	-0.854	-0.832	-0.800
8	-0.000	-0.796	-0.644	-0.864	-0.774	-0.816
9	-0.585	-0.820	-0.822	-0.776	-0.794	-0.866
10	-0.942	-0.738	-0.548	-0.792	-0.562	-0.908
...
317	-0.964	-0.863	-0.894	-0.894	-0.980	-0.982

5.1 基于 SPA 的预报模型的效果分析

绍兴市气象台梅汛期雨量预报中有 6 个预报模型因子, 分别为 X_1, X_2, X_3, X_4, X_5 和 X_6 。通过一系列的处理, 可得到它们相应的映射为 $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ 和 μ_6 , 映射主要应用式(18)以及因子的势态判断式(4)来进行的。式(18)是针对天气预报的特点从集对分析的联系度表达式推导而来。如果注意

一下前面的推导过程, 不难发现当因子处于强势时, 因子的映射只不过是线性变换; 而当因子处于劣势时, 该因子在模型中的作用已消失, 它的作用已由其他因子取代, 自然这是非线性变换, 而且是一种特殊的非线性变换。如果所有因子都是线性变换, 映射并不能使模型的质量有所变化。映射前后因子与预报量的相关系数见表 5, 可见大部分因子的相关系

数有了明显提高。对略有下降的因子 X_6 , 在映射时不再进行因子势态的判断和非线性变换, 使它的相关系数维持在原来水平上。因子映射前后的两种预

报模型的复相关系数列于表 5 末列, 复相关系数也有了提高。从表 5 可见, 映射后无论是因子独立预报能力还是整个模型的预报能力都有了提高。

表 5 预报因子映射前后的相关系数和复相关系数的比较

建立预报模型方法	X_1	X_2	X_3	X_4	X_5	X_6	复相关系数
一般回归的预报模型	0.556	0.331	0.297	0.229	0.293	0.461	0.593
基于 SPA 的预报模型	0.557	0.458	0.366	0.276	0.348	0.437	0.648
效果评定	略提高	提高	提高	提高	提高	略下降	提高

5.2 从一个预报实例看 SAP 的作用过程

SPA 有 3 个基本观点: 不确定性普遍存在; 不确定性和确定性可以放在同一个系统中进行分析和处理; 不确定性与确定性在一定条件下可以相互转换。其中后两个观点已充分反映在式(1)中。至于命题 w , 如何寻找不确定性部分和确定性部分, 如何让不确定性转化为确定性, SPA 没有也不可能给出具体的方法。在把 SPA 理论和方法应用于多元

预报模型的过程中, 一是定义了预报因子势态的意义, 设计了弱势因子和强势因子的判别方法, 也就是建立了辨认确定性部分和不确定性部分的方法; 二是推导了适用天气预报模型的联系度表达式, 也就是建立了天气预报问题中不确定性转化为确定性的方法。下面从一个具体的预报过程来剖析如何改进预报, 以分析这种改进的机制是否合理。

表 6 SPA 对因子的订正

因子类型	因子分析	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6
原因子 X (映射前)	变异系数 c_v	57.5	49.9	18.0	0.1	3.9	1.8
	一元回归值	6.05	6.59	22.61	14.26	12.13	15.63
	势态评定	弱势	强势	强势	强势	强势	强势
新因子 μ (映射后)	因子值 μ	-0.715	-0.870	-0.548	-0.716	-0.758	-0.688
	一元回归值	14.24	6.59	22.61	14.26	12.13	15.63

表 6 是以第 41 号个例为例所作的分析统计。根据变异系数判断, 第 1 因子处于弱势, X_1 对预报量的估计为 6.05, 以后出现的预报量实况为 25.6, 预报量比实况量明显偏小。由于 X_1 被判别为弱势, 用式(18)作非线性映射, 得到新因子 μ_1 (μ_1 是 X_2, X_3, X_4, X_5 和 X_6 的函数, 而不是 X_1 的函数), 用 μ_1 计算得到它对预报量的估计为 14.24, 预报值比原因子预报值有了明显提高。其他因子由于并不处于弱势, 它的一元回归值没有变化。原预报模型的预报值为 $\hat{y} = 15.2$ mm, 基于 SPA 的预报模型的预报值为 20.9, 后者比前者减少了误差 22.3%。可见通过因子的势态判断和因子的同异反分析和线性、非线性变换等一系列处理, 预报误差有了明显减少, 同时也可看到模型对因子的处理机制是一般统计预报方法难以实现的。

中的预报因子进行离散度和 SAP 分析, 即在因子参与作出预报结论之前先对各因子做势态判别和同异反分析, 根据判断结果, 让一些可能表现不良的弱势因子不再参加对预报结论的贡献, 而让其他强势因子去决定预报结论, 因而实现预报模型的因子结构动态优化, 增强了模型预报机制的合理性。本文所得到的主要结论有:

(1) 天气预报实践证明, 天气预报的不确定性是客观存在的, 这取决于大气运动的随机性以及初始场和预报模式的不完全性。集对分析(SPA)方法的基本思想就是对客观存在的种种不确定性予以承认, 并把不确定性放入一个既确定又不确定的同异反系统进行辩证分析和数学处理, 为我们对天气预报模型中实现动态因子结构提供了思路。借助于 SPA 方法, 可较好地解决一般多元回归等统计模型中所遇到的预报因子性能不稳定的问题。

(2) 本文探索在一般意义下, 自变量和因变量为离散型或连续型变量时 SPA 在天气预报中的应

6 结 语

文中应用集对分析原理, 对所建立的预报模型

用方法,当然也适用于自变量和因变量为(0/1)变量的情况。

(3) 预报模型中,“最佳”因子集中的因子在大部分情况表现优良,有时个别因子表现不良。前者称为因子处于强势,把后者称为因子处于弱势。通过对这些预报因子势态强弱程度的分析与判断,可以决定它们是否参与预报结论的确定。因子的势态主要由变异系数来判断。

(4) 发挥强势因子的作用,抑制弱势因子的作用,甚至取消弱势因子的作用,是通过同时用线性和

非线性变换,从而映射成一个新的因子集,进而进入多元回归预报模型进行运算来完成的。新的因子集由若干联系度数值所组成。

(5) 在绍兴市气象台一个数值预报产品释用预报模型中应用的例子表明,通过对预报模型中因子的上述分析、判断与变换,预报误差有了明显的减少,从而提高了模型的预报能力,说明这一方法对于数值预报产品释用预报模型的建立具有较好的效果。同时大量的试验和数年的业务试运行也说明了方法的稳定效果。

参考文献

- 1 赵克勤. 集对分析及其初步应用. 杭州: 浙江科学技术出版社, 2000. 1~ 15
- 2 蒋云良, 张裔智, 潘云鹤. 基于集对分析的同异反定量推理初探. 计算机工程, 1996, 3: 32~ 38
- 3 赵克勤. 集对论——一种新的不确定性理论方法和应用. 系统工程, 1996, 14(1): 18~ 23
- 4 赵克勤. 集对分析对不确定性的描述和处理. 信息与控制, 1995, 24(3): 165~ 168
- 5 薛根元. 天气预报哲学观若干问题思考. 新疆气象, 2001, 24(6): 28~ 30
- 6 张永山, 李维京, 丁一汇. 模糊优选法及其在月集合预报中的应用. 气象学报, 2000, 58(4): 428~ 434
- 7 史国宁. 概率天气预报的兴起及其社会意义. 气象, 1996, 22(5): 3~ 8
- 8 李小泉. 美国国家气象中心中期预报时段内的集合预报. 气象科技, 1994, 22(2): 7~ 11
- 9 王国强. 不确定理论——集对分析在 MOS 概率预报中的应用. 见: 天气预报技术论文集. 北京: 气象出版社, 1999. 6~ 13
- 10 范新岗. 集合预报方法的全局研究. 气象学报, 1999, 57(1): 75~ 83
- 11 陈孝孺等. 非参数统计教程. 上海: 华东师范大学出版社, 1993. 272~ 286
- 12 曹鸿兴, 封国林, 魏凤英. 一种区域气候预测模式的构建. 气象学报, 2000, 58(2): 159~ 166
- 13 中国科学院数学研究所. 常用数学统计方法. 北京: 科学出版社, 1973. 3~ 14
- 14 王国强, 陈红梅, 蒋延龙. 近邻估计——线性回归预报模型及其台风雨预报试验. 气象科技, 1999, 27(4): 25~ 29
- 15 金龙, 苗春生, 陈宁. 定性和定量长期预报模型的综合分析. 气象学报, 2000, 58(3): 479~ 483
- 16 杨昕, 张仁建. 均值生成函数的同期性延拓在回归分析中存在问题及其改进方案. 气象学报, 1998, 56(4): 493~ 499

THE APPLICATION RESEARCH OF THE THEORY OF UNCERTAINTY-SET PAIR ANALYSIS IN ESTABLISHMENT OF WEATHER FORECAST MODELS

Xue Genyuan

(Nanjing Institute of Meteorology, Nanjing 210044;

Shaoxing Meteorological Bureau of Zhejiang Province, Shaoxing 312000)

Wang Guoqiang

(Shaoxing Meteorological Bureau of Zhejiang Province, Shaoxing 312000)

Abstract

The forecast factors in weather forecast models have been selected strictly by means of the characteristics of forecasted objects, the physical meaning of forecasting factors and the experience in forecasting weather, and some technical ways. In general, these selected forecast factors have all better forecasting abilities, but the better abilities of these factors aren't always unchanged and sometimes, the badness ability of someone factor may produce error in forecast results. Set Pair Analysis (SPA) is a systemic theory and method used in researching and analyzing non-authenticity. Using the theory and method, this paper make judgment of station and analysis of same difference reverse about the factors before being used to forecast weather. In analyzing, the method weakens effect of those factors with badness abilities in forecast models, while those with better abilities in the models would play more important action. As a result the processes above have made the dynamic changes in the structure of the factors in the forecast models, which intensifies the rationality of forecasting mechanisms and the ability. The applications in interpret and use from the numerical forecast products indicate that the method has better effect.

Key words: Set Pair Analysis (SPA), Non-authenticity, Ability and state of factor, Interpret and use from the numerical forecast products.