

143-150

## 天气预报评分方法评述

丁金才

(上海中心气象台, 上海, 200030)

p 456.9

A

**摘要** 本文按定性预报、百分率预报和形态预报 3 类预报形式, 分别概述了常用的验证评分方法。提出了预报评分方法应遵循的 7 条准则, 并以此为依据, 针对强对流天气等小概率事件的预报, 讨论了各种评分方法的适用性。最后, 对各类预报形式的验证评分, 提出了最佳的方案。

**关键词** 验证评分, 定性预报, 百分率预报, 形态预报, 小概率事件预报

**分类号** P456

天气预报验证评分方法的研究已有 100 多年的历史, 近 40 年来, 它越来越受到人们的重视, 发展成天气预报研究领域的一个重要的分支。

不同类型的预报, 需要采用相应的评分方法进行评估。到目前为止, 天气预报大致可分为 4 种类型: (1) 定性预报: 对某一事件(指天气现象、要素等, 下同)出现的可能性作出有、无两级判别预报, 或多级判别预报。(2) 概率预报(亦称百分率预报): 以百分率形式对事件出现的可能性大小作出判断。一般以 10% 间隔来描述事件出现可能性的差别。现在, 概率预报已在许多国家成为主要的公众天气预报形式。(3) 形态预报(亦称场预报): 对某一气象要素的场分布或某一天气系统的形态变化作出预报, 如雨区预报、卫星云图的云团预报、雷达回波预报等等。近年来, 随着强对流天气短时预报的发展, 形态预报已成为业务预报的一个重要内容。(4) 定量预报: 对某一气象要素作出定量的预报。一般用于客观预报工具。虽然每种类型预报, 已有多种验证评分方法, 但每种方法都有一定的局限性。由于定量预报的评分方法可转化成定性预报和形态预报的评分方法, 本文将评述前三种预报类型的评分方法, 并针对强对流天气等小概率事件的预报, 着重分析各种评分方法的适用性, 以便能合理地使用它们。

## 1 评分方法简介

### 1.1 定性预报

美国的 Finley 在预报验证评分方法研究方面做了开创性的工作。1884 年, 他首先提出了叫正确率的评分法, 对他的龙卷风预报试验进行了评估<sup>[1]</sup>。此后, 不少学者提出了各种评分方法, 但都没有系统性。直至 1965 年, Brounlee 设计了两变量预报验证列联表  $\{n_{ij}\}$ <sup>[2]</sup>, 把预报和观测的关联系统地列出(见表 1)。这样, 利用表中任意两个元素, 可组成意义清楚的各种预报评分方法。如命中率  $POD = n_{11}/n_{.1}$ , 伪警率  $FAR = n_{12}/n_{.1}$ , 伪击率  $POFD = n_{12}/n_{.2}$ , 失误率  $FOM = n_{21}/n_{.1}$ , 无事件报对率  $FOCN = n_{22}/n_{.2}$  等。

表 1 预报和观测列联表  
Table 1 contingency table for the observation and prediction

观 测		有	无	
预 报	有	$n_{11}$	$n_{12}$	$n_{1\cdot}$
	无	$n_{21}$	$n_{22}$	$n_{2\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

1975 年, Donaldson 提出了成功临界指数<sup>(3)</sup>:

$$CSI = n_{11} / (n_{11} + n_{12} + n_{21}) \quad (1)$$

他利用列联表中 3 个元素的组合来评估预报, 显然比两要素评分法进了一步。但是, 直至 70 年代末, 所有的评分法都局限于简单的百分比表达式, 共同的不足之处是(1)没有利用列联表中全部元素, 因而不能全面反映出预报能力; (2)它们反映的不是预报的真实技巧水平; (3)不能用统计假设量对预报技巧的显著性进行检验。1982 年, Mason 提出了真实技巧统计法(True Skill Statistic), 简称 TSS 法<sup>(4)</sup>。

$$TSS = \frac{tr(R)}{tr(R_p)} = \frac{[n_{11} - \frac{n_{1\cdot}n_{\cdot 1}}{n_{\cdot\cdot}}] + [n_{22} - \frac{n_{2\cdot}n_{\cdot 2}}{n_{\cdot\cdot}}]}{[n_{\cdot 1} - \frac{n_{\cdot 1}n_{\cdot 1}}{n_{\cdot\cdot}}] + [n_{\cdot 2} - \frac{n_{\cdot 2}n_{\cdot 2}}{n_{\cdot\cdot}}]} \quad (2)$$

上式分子中的  $\frac{n_{1\cdot}n_{\cdot 1}}{n_{\cdot\cdot}}$  为用气候概率  $\frac{n_{\cdot 1}}{n_{\cdot\cdot}}$  预报事件出现的正确次数, 同理  $\frac{n_{2\cdot}n_{\cdot 2}}{n_{\cdot\cdot}}$  为用事件不出现的气候概率  $\frac{n_{\cdot 2}}{n_{\cdot\cdot}}$  预报正确的次数。所以,  $tr(R)$  反映的是实际预报技巧。现假定预报完全正确,

即  $n_{12} = n_{21} = 0$ , 则  $n_{\cdot 1} = n_{1\cdot} = n_{11}$ ,  $n_{\cdot 2} = n_{2\cdot} = n_{22}$ 。分母中  $\frac{n_{\cdot 1}n_{\cdot 1}}{n_{\cdot\cdot}} + \frac{n_{\cdot 2}n_{\cdot 2}}{n_{\cdot\cdot}}$  应为理想预报假设下的气候概率预报的正确次数。TSS 含义是实际预报技巧和理想预报技巧之比。1990 年, Doswell 发现在对小概率事件预报的评分时, TSS 常趋近于 POD, 会导向过分预报倾向。为此, 他对 TSS 法作了修改, 提出了 Heidke Skill Score 法, 简称 S 法<sup>(5)</sup>。

$$S = \frac{tr(R)}{tr(R_p^*)} = \frac{[n_{11} - \frac{n_{1\cdot}n_{\cdot 1}}{n_{\cdot\cdot}}] + [n_{22} - \frac{n_{2\cdot}n_{\cdot 2}}{n_{\cdot\cdot}}]}{[n_{\cdot 1} - \frac{n_{\cdot 1}n_{\cdot 1}}{n_{\cdot\cdot}}] + [n_{\cdot 2} - \frac{n_{\cdot 2}n_{\cdot 2}}{n_{\cdot\cdot}}]} \quad (3)$$

S 仅仅把分母中 TSS 的理想预报假设下的气候概率预报正确次数 ( $\frac{n_{\cdot 1}n_{\cdot 1}}{n_{\cdot\cdot}} + \frac{n_{\cdot 2}n_{\cdot 2}}{n_{\cdot\cdot}}$ ) 改成实际预报下的气候概率预报正确次数 ( $\frac{n_{1\cdot}n_{\cdot 1}}{n_{\cdot\cdot}} + \frac{n_{2\cdot}n_{\cdot 2}}{n_{\cdot\cdot}}$ ), 因此  $tr(R_p^*)$  可称为理想预报实际技巧。

## 1.2 概率预报

概率预报在我国尚未开展, 表 2 列出的是美国国家强风暴实验室(NSSL)在 1988~1989 年中尺度应用试验期(简称 MAP)的强对流天气中午展望预报的评分结果。

概率预报的评分法最早由美国的 Brier 在 1950 年提出的, 叫半贝里尔得分(half-brier score), 简称 B<sup>(6)</sup>

$$B = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2 \quad (4)$$

式中  $F_i$  为某次预报的百分率,  $O_i$  为该次实况百分率, 事件出现为 1, 不出现为 0,  $N$  为预报总次

数。 $B = 0$  表示预报完全正确。1963 年, Sander 又提出了技巧分法<sup>[7]</sup>, 简称 BS

$$BS = 100 \cdot (B_c - B_f) / B_c \quad (5)$$

式中  $B_f$  为由(4)式求得的实际预报的半贝里尔得分,  $B_c$  为以气候概率  $R_0$  为预报概率所得到的半贝里尔得分。 $BS > 0$ , 表示有预报技巧。1980 年, Hughs 又提出了偏离率法来衡量预报过分或偏少的倾向, 简称 Bias 法。

$$Bias = 100 \cdot (R_f - R_0) / R_0 \quad (6)$$

式中  $R_f$  为平均预报概率。 $Bias > 0$  表示有过分预报倾向, 反之表示有过少预报倾向。

表 2 美国 MAP 试验期强对流天气概率预报(中午展望预报)

Table 2 Severe convective weather probability outlook for noon hours during U. S. MAP Experiment

百分率(%)	预报次数	实况次数	FOH	POD	FAR	CSI	POFD	TSS	S
100	5	5	1.00	0.07	0.00	0.07	0.00	0.07	0.10
90~99	10	8	0.80	0.19	0.13	0.19	0.01	0.18	0.24
80~89	7	6	0.86	0.28	0.14	0.27	0.02	0.26	0.33
70~79	7	6	0.86	0.37	0.14	0.35	0.02	0.34	0.42
60~69	10	7	0.70	0.47	0.18	0.43	0.04	0.43	0.49
50~59	19	13	0.68	0.66	0.22	0.56	0.07	0.59	0.61
40~49	17	5	0.29	0.74	0.33	0.54	0.14	0.59	0.57
30~39	12	4	0.33	0.79	0.38	0.53	0.19	0.60	0.56
20~29	16	6	0.38	0.88	0.42	0.54	0.25	0.64	0.55
10~19	29	4	0.14	0.94	0.52	0.47	0.39	0.55	0.43
1~9	58	3	0.05	0.99	0.65	0.35	0.71	0.28	0.18
0	52	1	0.02	1.00	0.72	0.28	1.00	0.00	0.00

$$b_f = 0.12 \quad R_0 = 0.28 \quad R_f = 0.27 \quad B_c = 0.20 \quad B_s = 42.52 \quad Bias = -3.31$$

### 1.3 形态预报

目前,常用的形态预报的评分法有 3 种:

相关系数法,即

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7)$$

均方差法,即

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (8)$$

真实技巧法,即

$$TS = S_{fc} \sqrt{(S_0 + S_f - S_{fc})} \quad (9)$$

上式中  $S_0$  为事件出现的覆盖面积,  $S_f$  为预报事件所覆盖的面积,  $S_{fc}$  为预报正确的面积。

近年来,随着强对流天气短时预报的发展,预报员常需要对天气系统形态,如卫星云图的

云团, 雷达回波等作出预报。1990 年 Doswell 和丁金才又提出了形态评分法 (pattern verification), 简称 P 法。该方法把形态预报误差分解成 6 个基本要素, 如图 1 所示。

(1) 距离误差 (distance error), 简称 DIS。指预报的形态中心到实况的形态中心的距离。若图示的距离方向与系统移动方向一致, DIS 为负值, 反之则为正值。单位以格距或其他长度单位。(2) 方向误差 (direction error), 简称 DIR。指预报和实况的形态质量中心连线与垂直轴的夹角。以度为单位, 范围为  $0 \sim 180^\circ$ 。(3) 强度误差, 或称幅值误差 (amplitude error), 简称 AMP。指预报的最大值与实况的最大值之差。(4) 方位误差 (orientation error), 简称 ORI。指预报的形态长轴与实况的形态长轴之间的夹角。范围为  $0 \sim 180^\circ$ 。它在一定程度上反映了系统分布方向的预报误差。(5) 形态误差 (shape error), 简称 SHP。系统形状用形态的短、长轴之比 (偏心率) 来描述。范围为  $0 \sim 1$ , 它反映了在天气系统类型上的预报误差。偏心率小的可认为是锋面、飑线等带状系统, 偏心率大的是对流单体、MCC 等云团系统。(6) 面积误差 (coverage error), 简称 COV。指预报的形态面积与实况形态面积之差。单位为格点数或其他面积单位。对各次预报误差先进行极值标准化处理, 最后由下式得出形态评分值

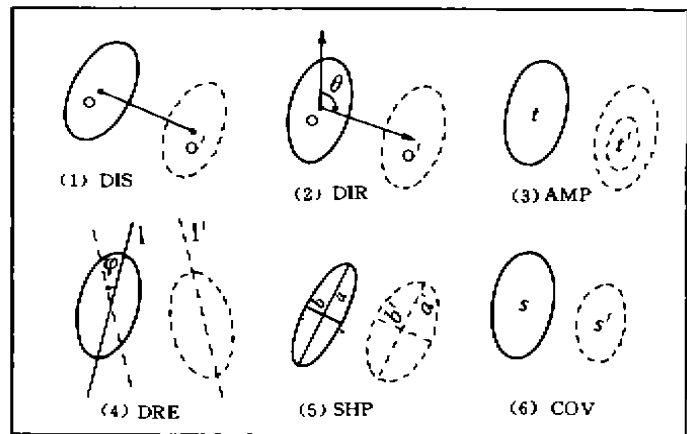


图 1 形态预报验证方法的 6 个要素

实线为预报, 虚线为实况

- (1) 距离误差:  $O \sim O'$
- (2) 方向误差: 用  $\theta$  角表示
- (3) 强度误差:  $t' - t$
- (4) 方位误差: 用长轴夹角  $\varphi$  表示
- (5) 形状误差: 用偏心率差  $\frac{b'}{a'} - \frac{b}{a}$  表示
- (6) 面积误差:  $S' - S$

Fig. 1 Elements from decomposition of the pattern forecast error with solid (dashed) line denoting condition forecasted (observed)

(1) distance error  $O - O'$ ; (2) direction error  $\theta$ ;

(3) amplitude error  $t' - t$ ; (4) orientation error  $\varphi$ ;

(5) shape error  $b'/a' - b/a$ ; (6) coverage error  $S' - S$

$$P = \frac{1}{6} \sum_{i=1}^6 \left( \frac{1}{N} \sum_{j=1}^N |P_{ij}| \right) \quad (10)$$

式中  $P_{ij}$  为第  $i$  次预报中第  $j$  类的经标准化处理后的误差,  $N$  为预报总次数。P 值范围为  $0 \sim 1$ 。台风路径预报的评分实质是简化的形态评分法, 只取其中的距离误差和方向误差。

## 2 评分方法的讨论

### 2.1 评分方法准则

Flueek 在 1987 年曾提出建立一个评分方法应遵循 5 条准则<sup>[9]</sup>。我们根据多年来的研究和使用的体会, 提出了如下 7 条准则: (1) 构造一个简单易懂的值来表示评分结果; (2) 构造评分值时应包括列联表中的所有元素; (3) 该值与事件本身概率是独立的, 使评分结果有可比性; (4) 避免导向预报偏于某种错误倾向; (5) 反映预报能力的变化有一定的灵敏度; (6) 能用统计假设值对预报技巧的显著性进行检验; (7) 对预报的改进方向有一定的指示作用。

根据上述准则, 针对强对流天气等小概率事件预报的特点, 讨论各种评分方法的适用性。

## 2.2 定性预报

POD、FAR 等用两要素组成的评分法,显然不符合准则(2)。若仅以 POD 来衡量预报能力,会导致宁空不漏,即过分预报倾向。而 FAR 会导致宁漏不空的倾向,即过少预报倾向。在小概率事件预报中,更易导致这两种错误倾向,不符合准则(4)。还须指出的是 FAR 等两要素评分法有时与事件本身概率有关,不符合准则(3)。假设每次都预报事件发生,则列联表中  $n_{1.} = N, n_{11} = n_{.1}, n_{12} = n_{.2}, POD = n_{.1}/n_{..}$ , 为事件发生气候概率,  $FAR = n_{.2}/n_{..}$  为事件不发生时气候概率。这种情况下,POD 和 FAR 值不能反映预报好坏。

CSI 由列联表中 3 个元素组成,很大程度上避免了过分预报和过少预报倾向。但是它没有包括  $n_{22}$  这一要素,即预报事件不发生的能力,没有完全满足准则(2)。由于强对流天气等小概率事件的预报往往存在过分预报倾向,CSI 在这种情况下与事件发生的气候概率有一定关联,不满足准则(3)。在上述完全过分预报的假设下,  $n_{21} = 0, n_{11} = n_{.1}, n_{12} = n_{.2}, CSI = n_{.1}/n_{..}$ , 即为事件发生的气候概率。

TSS 和 S 包含了列联表中的所有元素,能综合反映预报事件发生和不发生的能力,完全满足准则(2)。它们与事件的气候概率是独立的。若每次都预报事件发生,或预报事件不发生,则  $n_{11}$  和  $n_{22}$  中必有一个为零, TSS 和 S 均为 0, 显示无预报能力, 满足了准则(3)。

值得一提的是 TSS 和 S 在反映预报能力方面比 POD 和 CSI 更趋合理。表 3 列出了 NSSL 在 1990~1991 年的 QED 业务试验的两种预报评分结果。虽然 POD、FAR 和 CSI 3 种方法都显示强对流天气的预报能力优于中尺度气旋的预报能力,但 TSS 和 S 的评分结果表明后者优于前者。这主要是两个原因:一是 POD、FAR 和 CSI 没有包括预报事件不发生的能力,二是它们都是一种简单的百分比,不一定反映出预报技巧水平。

表 3 美国 QED 试验期中午展望预报

Table 3 Noon-time outlook of severe convective weather and mesoscale cyclone during U. S. QED Experiment

		强对流天气预报			中尺度气旋预报				
观 测		有	无		有		无		
预 报	有	42	17	59	预 报	有	13	6	19
	无	24	76	100		无	11	117	128
		66	93	159			24	123	147
		$f_1 = 0.41 \quad POD = 0.64 \quad CSI = 0.51$			$f_2 = 0.16 \quad POD = 0.54 \quad CSI = 0.43$				
		$FAR = 0.29 \quad TSS = 0.45 \quad S = 0.46 \quad FOM = 0.36$			$FAR = 0.32 \quad TSS = 0.49 \quad S = 0.54 \quad FOM = 0.46$				

由于 TSS 和 S 表达式的分子是表示实际预报频数和气候预报的理论频数之差,因此可用  $\chi^2$  统计假设量来检验预报技巧水平的显著性。表 3 中强对流天气预报的  $\chi^2 = 32.2$ , 中尺度气旋预报的  $\chi^2 = 39.1$ , 均大于  $\chi^2(\alpha = 0.01) = 10.8$ , 表明其预报能力的显著性有 99% 以上可信度。它们满足了准则(6)。

对小概率事件的预报,预报事件不出现的正确次数  $n_{22}$  一般远大于其他 3 个要素  $n_{11}, n_{12}$  和  $n_{21}$ 。可以证明,在这种情况下, TSS 趋于 POD, S 趋于  $2 \times (1 + CSI)^{-1}$  (证明略)。因此对强对流天气等小概率事件预报,用 TSS 法来评估容易导致过分预报倾向,而 S 法可避免这一不足。在下一节的百分率预报评分法的讨论中,将有例子进一步说明 S 法比 TSS 法优越。

1990 年 Schaefer 把技巧得分引入 CSI,提出了技巧成功临界指数 CSI,

$$CSI_s = (n_{11} - c) / ((n_{11} - c) + n_{12} + n_{21}) \quad (11)$$

式中  $c$  为气候概率事件发生的正确次数。这样使  $CSI_s$  满足了准则(6),但它仍只反映预报事件发生的技巧。事实上,在  $S$  表达式中去掉预报事件不发生的技巧,即  $(n_{22} - n_{21} \cdot n_{12} / n_{11})$ ,就可得  $CSI_s$ 。

值得一提的是  $CSI$ 、 $TSS$  和  $S$  都具有对称性。在它们的表达式,对换  $n_{12}$  和  $n_{21}$ ,表达式无变化,说明它们不能反映预报误差是由过分预报偏向造成的,还是由过少预报偏向造成的。这不符合准则(7)。可见至今尚未有一种评分方法能全部满足 7 条准则。但是,  $FAR$  和  $FOM$  能单独反映出过分预报和过少预报倾向。若某个预报的评分结果,  $FOM > FAR$ ,则可认为该预报主要倾向是过少预报,反之为过分预报。因此我们建议对小概率事件定性预报的最佳评分方案是  $S$  法和  $FAR$ 、 $FOM$  法一起用。

### 2.3 概率预报

概率预报的评分,除用  $B$ 、 $BS$  和  $Bias$  对总预报进行评估外,还可对每一百分率段的预报,经过有、无两级处理后,用定性预报的评分法进行分别评估(见表 2)。对小概率事件预报,低百分率的预报次数总是占了预报总次数的极大部分,如表 2 中低于预报概率 0.20 的次数为 139 次,占总次数的 57%,因此  $B$  值总是很小,不符合准则(5)。而且事件发生的气候概率越小,  $B$  值往往也越小,这又不符合准则(3)。如 NSSL 在 MAP 试验中,中尺度气旋概率预报的评分结果为  $B = 0.06$ ,  $R_s = 0.08$ ,  $R_f = 0.12$ ,  $B_s = 0.08$ ,  $BS = 25.72$ ,  $Bias = 43.75$ (表略)。其  $B$  值仅为表 2 中强对流天气概率预报的一半,似乎前者的预报优于后者。但  $BS$  值表明前者的预报技巧水平远低于后者。 $Bias$  值也表明前者存在着明显的过分预报倾向,而后者的  $Bias = -3.31$ ,几乎无过分预报和过少预报倾向,应该说强对流天气概率预报能力优于中尺度气旋预报能力。可见半贝里尔评分法不适用于小概率事件的预报评分。

值得注意的是,表 2 中  $TSS$  的最大值落在 20%~29% 的低百分率段内,证实了上一节提及的  $TSS$  法在小概率事件预报中会导致过分预报倾向。而  $S$  的最大值落在 50%~59% 的百分率段内,表明不会导致过分或过少的预报倾向。

### 2.4 形态预报

对于强对流天气等小概率事件预报,  $r$ 、 $RMS$  和  $TS$  法会出现诸多不合理之处。现以图 2 假设的 3 个形态预报例子进行说明,表 4 列出了 4 种评分法的评分结果。

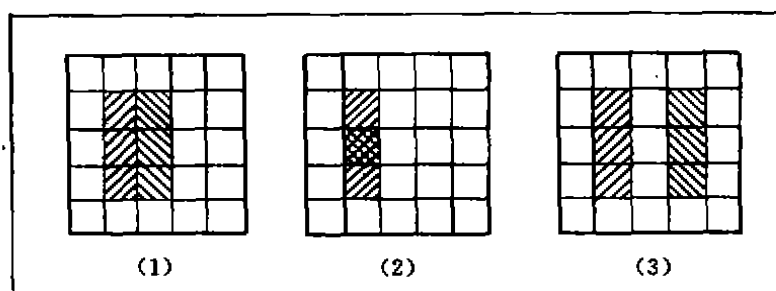


图 2 3 个形态预报的例子

// 为预报事件发生的区域, \\ 为观测到的事件发生区域

Fig. 2 Examples of the pattern forecast

Area of the event predicted (observed) is forward (backward) slashed

表 4 形态预报的 3 种评分方法结果  
Table 4 Results of the pattern forecast verification

方 法	r	RMS	TS	P
预报 1	-0.136	0.49	0.00	0.05
预报 2	0.55	0.28	0.33	0.13
预报 3	-0.136	0.49	0.00	0.08

由于小概率事件的发生区域往往只占整个预报区域的很小一部分,常会由于系统移向、移速上较小的预报误差,造成预报位置和实况位置完全错开,如预报 1 所示。结果  $r$  为负值,指示预报和实况相反; $TS$  为零,指示无预报能力, $RMS$  较大,指示误差最大。这样的结论显然对预报员欠公正。从形态预报评分法  $P$  的观点来看,预报 1 和 3 只是在系统的移速预报上差了 1~2 格距,其他 5 个要素都预报正确。但预报 2 虽然在系统移速上无预报误差,但天气系统形状由块状预报成条状,预报面积比实况大 2/3。 $P$  法评分结果表明预报 1 的预报能力最好,比较接近预报的实际情况。还应指出的是  $r$ 、 $RMS$  和  $TS$  法对预报 1 和 3 的评分结果完全相同。这说明一旦预报的位置与实况完全错开,这 3 种方法无法分辨出误差的大小,不符合准则(5)和(7)。而  $P$  法能区分出 3 种预报的误差大小,并能从 6 个方面反映预报误差,指出预报改进的方向,符合准则(5)和(7)。但由于  $P$  法使用时间不长,尚有许多不成熟之处,有待进一步改进。

### 3 结 语

对定性预报的验证评分,POD、FAR 等两要素评分方法简单、明了,但会导致过分预报和过少预报倾向。CSI 法避免了这一缺点,但当预报存在较明显的过分预报倾向时,CSI 值会接近于事件发生的气候概率,丧失对预报能力评估的可比性。POD、FAR 和 CSI 只反映了预报事件发生的能力,没有把小概率事件中大量的预报事件不出现的预报能力包括进去。TSS 和 S 避免了这些缺点,能全面地反映出预报技巧水平,并能用  $\chi^2$  检验预报技巧的显著性。但是对小概率事件预报而言,TSS 往往趋于 POD,易导致过分预报倾向,而 S 没有这一缺点。但是 CSI、TSS 和 S 法都具有对称性,不能分辨出预报误差是由过分预报还是过少预报倾向造成的。因此,目前对强对流天气等小概率事件的预报的验证评分,最佳方案是 S 法结合 FOM 和 FAR 一起用。

对小概率事件的概率预报,B 法失去了评分价值,比较理想的方案是 BS 和 Bias 一起用,既对综合预报能力作出评估,又能发现预报存在的何种偏向。

对小概率事件的形态预报,由于预报事件发生的位置常会与实况位置错开, $r$ 、 $RMS$  和  $TS$  评分会产生许多不合理之处,而且对预报误差大小也无法分辨。 $P$  法能避免这些缺点,具有较清楚的天气预报意义,能从 6 个方面反映出预报误差大小。但  $P$  法还有待进一步改进和完善。

**致谢:**上海市气象局总工程师朱永祺和南京气象学院张培昌教授对本工作给予了热情的帮助,并提出了宝贵的修改意见,在此谨表谢意。

## 参 考 文 献

- 1 Finley J P. Tornado predictions. *Amer Meteor J.* 1884, 1: 85~88
- 2 Brownlee K A. *Statistical theory and methodology in science and engineering*, 2nd ed. John Wiley and sons, 1965
- 3 Donaldson R J, Dyer R M, Krauss M J. An objective evaluator of techniques for predicting severe weather events. In: *Amer Meteor Soc eds. 9th conf severe local storms*. Norman, Oklahoma, 1975. 321~326
- 4 Mason I. Dependence of the critical success index on sample climate and threshold probability. *Aust Meteor Mag.* 1989, 37: 75~81
- 5 Doswell C A III, Davies-jones R, Keller D L. On summary measures of skill in rare events forecasting based on contingency tables. *Weather Forecasting*, 1991, 6: 576~585
- 6 Brier G W. Verification of forecasts expressed in terms of probability. *Mon Wea Rev.* 1950, 78: 1~3
- 7 Sanders F. On subjective probability forecasting. *J Appl Meteor.* 1963, 2: 191~201
- 8 Flueck J A. A study of some measures of forecast verification. In: *Amer Met Soc eds. 10th Conf Probability and Statistics in Atmospheric Science*. Edmonton, Alberta, 1987, 69~73

## REVIEW OF WEATHER PREDICTION VERIFYING TECHNIQUES

Ding Jincai

(Shanghai Meteorological Center, Shanghai, 200030, PRC)

**Abstract** Usual verification techniques are outlined for categorical, probability and pattern forecasts with seven criteria proposed for the verifications. Accordingly, the applicability of each of the three methods is addressed with reference to small-probability events like severe convective weather. Finally, an optimal scheme is developed for each of the techniques.

**Keywords** verification, categorical forecast, probability forecast, pattern forecast, small-probability event forecast