

田武文. 中国各省月预报与无技巧预报的评估分析[J]. 气象, 2010, 36(9): 100-105.

中国各省月预报与无技巧预报的评估分析^{*}

田武文

陕西省气候中心, 西安 710015

提 要: 在 Ps 评分中, 预报技巧是由一段时间内实际业务的 Ps 评分与随机预报评分之差确定。通过对中国各省无技巧预报评分的评估, 定量分析了在现行评分办法中影响无技巧评分存在差异的两个因素。统计结果及分析表明: (1) 由于各省测站要素各等级气候概率分布的差异, 决定了随机预报准确率存在差异。以省份为单位, 随机预报评分降水最大差异为 14 分, 气温最大差异为 7 分; (2) 目前预报正确性判定标准的设计, 使得在只做二级预报或持续预报时, 可以大幅度提高无技巧预报准确率, 给无技巧预报评分的定量评估带来不确定因素; (3) 对比持续预报和随机预报的评估结果, 全国平均降水提高了 9 分, 温度提高了 13 分。研究结果表明, 由于各省之间的无技巧预报评分存在明显差异, 因此直接利用 Ps 评分比较预报技巧, 不能客观反映出预报能力的差异。

关键词: Ps 评分, 各等级气候概率分布, 随机预报, 持续预报, 预报准确率

The Quantitative Assessment of Random Forecast Ps Score Among the Provinces in China

TIAN Wuwen

Shaanxi Climate Center, Xi'an 710015

Abstract: In the Ps score assessment, the forecast skills are determined by the difference between Ps score of operational forecast and random forecast accuracy over a period of time. The paper attempts to make quantitative evaluation of the random forecast accuracy among all the provinces and regions in China, and the statistical results are as follows. (1) Owing to the obvious difference of the random forecast accuracy about station elements determined by climate probability in each grade in every province, the maximum difference of precipitation score is 14, and the maximum difference of temperature score is 7. (2) Since the judging standard of forecast accuracy has been designed vulnerably, the accuracy of two grade forecast and continuous forecast may be higher than random forecast accuracy, it brings uncertainty to the quantitative evaluation of random forecast accuracy; (3) The score of random forecast might be higher than that of the actual forecast if these factors affecting the score are applied to quantitative assessment. The result shows that the current scoring methods could not access the forecast skills in various provinces properly.

Key words: Ps score, distribution of climate probability in each grade, random forecast, continuous forecast, forecast accuracy

引 言

自 2007 年中国气象局发文并规定各省按规定的测站每月上报月预测信息, 并按年公布各省的 Ps 排名之后, Ps 评分中按区域评定的方式便自动终

止, 只能采用站点评分一种方式。Ps 评分排名的目的在于, 通过对一段时期内各省份之间的 Ps 评分进行对比, 展示预报技巧的高低, 以此激励短期气候预测研究, 提高预报准确率。

目前气候预测业务上使用的检验评分体系, 是基于 Ps、SS、ACC 和 Ts 4 个评分指标^[1-12], 其中

* 陕西省气象局重点科研项目《省级滚动预测业务系统及国家动力气候模式统计降尺度解释应用研究》; 陕西省气象局科技创新基金项目“气候变暖背景下低温冷害规律及对策研究”共同资助

2009 年 3 月 6 日收稿; 2009 年 12 月 21 日收修定稿

作者: 田武文, 主要从事短期气候预测业务及研究. Email: t-w-w58@sohu.com

SS、ACC 和 Ts 是技巧评分。要利用 Ps 评分判断预报技巧的高低,就需要和某种无技巧预报进行比较,比如随机预报或持续预报^[1-2]。即在一段时间之内,Ps 评分高于无技巧预报评分,预报才能被认为具有技巧。一般说来,Ps 评分与无技巧预报评分的差越大,预报技巧就越高,预报的应用价值就越大;反之,预报技巧就越低,相应的就没有多少应用价值。利用 Ps 评分结果进行预报能力排名,就是假定所对比的省份之间,无技巧预报准确率相同而做的简化处理。在上述 4 种评分方法中,Ps 评分使用得最广。

随机预报或持续预报评分,是由预报要素各等级的概率分布所决定的^[1-2]。我国的 Ps 评分方法中,测站要素等级是按照要素实况值与气候均值偏差的大小,使用统一的标准划分的。由于我国疆域辽阔,南北、东西气候差异大,测站的位置位于不同的气候带,预报要素之间的变差差异显著,就使得无技巧预报准确率随着测站所在地理位置的变化而变化。

本文以测站的月预报为目标,根据评分办法规定的等级划分标准,利用历史资料对全国各省的无技巧预报的准确率进行了估算。目的在于在比较预报技巧时,能够采用 Ps 评分与随机预报准确率的差来衡量,使得预报技巧的比较结果更加客观。文中同时分析了在目前评分办法中,影响无技巧预报评分的其他因素,发现持续预报可以提高无技巧预报评分。

1 标准分布下的随机预报准确率

随机预报准确率的计算是基于如下假设:对要预报某测站某月的要素,是在已知其各等级气候概率分布的条件下,保持随机预报各等级的概率与其气候概率相同,其预报评分就是随机预报准确率。很显然,本文所计算的测站之间随机预报准确率的差别就是由测站间各等级气候概率的分布所决定的。

以 6 级 Ps 评分为例,假设各等级分布概率为表 1 中所列的分布,作为标准分布以供参考。传统 Ps 评分是在预报级别与实况级别完全相同时,才判定为正确。6 级随机预报准确率为表 1 中对角线上的数据之和,为 0.18。中国现行的 Ps 评分办法对预报正确性的判定标准进行了修改,即表 1 中所列的概率均判定为正确。

通过表 1,就可以计算出 6 级预报的随机预报准确率。把表 1 中所列的各级别的概率相加,可以得出随机预报准确率为 0.58。表 1 中的最后一列,为按等级分布预报时,各等级的预报准确率。与之相对应的等概率 4 级 Ps 评分,随机预报的准确率为

0.625(表略)。

对于任何一个测站,当气候要素各等级的气候分布概率确定之后,代入表 1,就可以计算出该站的随机预报准确率。

为了方便讨论样本等级气候概率分布的变化情况,在此约定:在 6 级评分中,异常等级是指 1 级(降水特少,气温异常偏低)和 6 级(降水特多,气温异常偏高)两个等级;正常等级是指 3 级(降水略偏少,气温略偏低)和 4 级(降水略偏多,气温略偏高)两个等级;而 2 级(降水偏少,气温偏低)和 5 级(降水偏多,气温偏高)为正常、异常之间的过渡级别。

对于任何测站,如果预测的月要素各等级的气候概率分布与表 1 相同,则随机预报的准确率就相同。而如果使表 1 中的异常等级的概率分布减少,正常等级的概率分布增加,把计算出的随机预报准确率与表 1 的结果相比,可以发现其随机预报准确率是上升的,反之则是下降的。这说明各等级的概率分布向正常等级的方向集中时,计算出的随机预报准确率高于标准分布;反之,向异常等级的方向发散时,计算出的随机预报准确率低于标准分布。由于现行评分办法规定:在全国范围内,降水按统一的距平百分率标准划分等级,气温按统一的距平值标准划分等级,这就必然使得不同测站之间各等级的概率分布出现差异,进而导致随机预报准确率出现差异。

表 1 6 级预报随机预报准确率

Table 1 The random forecast accuracy in 6-grade forecasts

	观测						各等级预报准确率
	1 级	2 级	3 级	4 级	5 级	6 级	
	0.1	0.2	0.2	0.2	0.2	0.1	
1 级	0.1	0.01	0.02	0.02			0.05
2 级	0.2	0.02	0.04	0.04			0.10
预 报	3 级	0.2	0.02	0.04	0.04		0.14
	4 级	0.2		0.04	0.04	0.02	0.14
	5 级	0.2			0.04	0.04	0.10
	6 级	0.1			0.02	0.02	0.05

用概率分布求 Ps 评分中的随机预报准确率是最基本的方法。它根据 Ps 评分规定的等级划分标准,把气候年限内的预报要素划分成等级,统计出某测站某月各等级出现的气候概率,同时假定在随机预报结果中,各等级的概率分布与其相应等级的气候分布概率相同,把结果列入表 1,计算出预报各等级时的随机预报准确的概率;把各等级的概率相加,即可得出该站该月的随机预报准确率。

2 由实际等级概率分布决定的随机预报准确率

下面利用全国 160 个观测站,1971—2000 年共

30 年的气候资料,对 12 个月的月平均气温和月降水总量两个气候要素,按评分办法规定的等级划分标准,计算出各站各月各个等级的气候分布频率(其中气温全部按 6 级分布。降水量 $< 10 \text{ mm}$ 时,不分级,不预报; $50 \text{ mm} >$ 降水量 $\geq 10 \text{ mm}$ 时分 4 级;降水量 $\geq 50 \text{ mm}$ 时分 6 级)。根据表 1,即可计算出其随机预报准确率。12 个月随机预报准确率的平均,

就是年随机预报准确率。

2.1 测站间、月份间随机预报评分的差异

经过计算和统计,把同一测站不同月份之间,不同测站年平均之间,不同测站不同月份之间,全国 160 站不同月份之间的随机预报评分的最大差异列于表 2。

表 2 测站间、月份间随机预报评分高低的差异
Table 2 The difference of random forecast scores in each month among different stations

		高评分			低评分			差
		测站	时间	评分	测站	时间	评分	
降水	同一测站月差	毕节	2 月	76.2%	毕节	10 月	52.0%	24.2%
	测站年平均差	且末	年平均	71.3%	厦门	年平均	55.0%	16.3%
	不同测站月差	岷县	3 月	76.7%	南阳	10 月	50.0%	26.6%
	160 测站月差		12 月	61.5%		8 月	55.6%	5.9%
气温	同一测站月差	蒙自	8 月	83.8%	蒙自	4 月	52.0%	31.8%
	测站年平均差	景洪	年平均	64.9%	乌鲁木齐	年平均	51.9%	13.0%
	不同测站月差	景洪	8 月	85.1%	*	*	50.0%	35.1%
	160 测站月差		8 月	58.8%		2 月	52.5%	6.3%

*: 月气温随机预报评分等于 50% 的测站分别是: 呼玛 11 月、海拉尔 2 月、哈尔滨 11 月、乌兰浩特 2 月、延吉 7 月、营口 2 月、北京 2 月、阜阳 7 月、福州 2 月、广州 12 月、海口 3 月、武威 12 月、阿勒泰 3 月和 11 月、乌鲁木齐 3 月共 14 站次。

从表 2 中可以看出,对于降水和气温各自所列出的 4 个项目,均有较大的差异。

2.2 以省为单位的随机预报评分

对各省内所有测站年随机预报准确率进行平均,就得出全国各省(自治区、直辖市)月气候要素的年平均随机预报准确率。按评估结果由高到低的顺序排列,其结果列于表 3。

从表 3 中可以看出:(1) 由于各省所在地区的气候差异远较全国范围的气候差异小,测站的这种差异能使得各省随机预报准确率出现明显的差异。如果设定预报完全报对为 100 分,按省份对比,降水最大相差 14 分,气温最大相差 7 分。(2) 按省份等权重平均,降水的随机预报准确率为 62 分,气温为 55 分,与表 1 所列的标准模型相比,降水准确率高于 6 级标准模型评分,略低于 4 级等概率模型的 62.5 分;气温准确率则低于 6 级标准模型评分。

3 影响随机预报准确率的其他因素

按照目前的评分办法,无技巧预报评分仍有提升的空间。提升的途径有两种:一种是只做 2 级预报,另一种是持续预报。下面通过对这两种情景下预报评分的评估,讨论影响随机预报准确率的其他因素。

3.1 只做 2 级预报时的随机预报准确率

相对于 6 级分级评分,如果采用只报 3 级和 4 级两个级别,对其他级别不予预报,即预报这两个等级的概率均为 0.5,利用表 1 可以算出,这时 6 级评分下的随机预报准确率就变为 0.7;与之相应的 4 级(各级概率分布相等)随机预报准确率为 0.75。

通过以上分析,上述问题是由预报正确性判定的规定引发出来的,属于评分模型设计的缺陷(或称漏洞)。那么,这种设计缺陷在实际等级概率分布(根据评分办法规定的标准划分要素等级)的情况下,对单站的随机预报准确率是否有影响呢?表 4 给出统计说明。

表 4 中,降水和气温随机预报准确率大于 0.80 的站次数分别为 11 站和 39 站,虽然站数尚达不到参加评分总站数的 1% 和 2%,但对于与其相比较的站,显然是不公平的。在 2 级预报下,全国 160 个测站降水和气温的平均随机预报准确率均为 0.65,高于 6 级概率分布的随机预报准确率。

在表 4 的气温随机预报准确率统计中,发现 7—9 月的景洪站、8 月的临沧站共 4 站次随机预报评分竟然达到 0.90 以上。

对于所有测站、所有月份的统计发现,采用 2 级预报的随机预报评分,月降水量有 93.1% 站次大于由实际等级概率分布决定的随机预报评分,月平均

表 3 由概率分布计算的全国各省随机预报准确率

Table 3 The random forecast accuracy calculated by probability distribution in every province

降水				气温			
排序	省份	全年评分月数	随机预报	排序	站数	省份	随机预报
1	新疆	12	0.70	1	8	云南	0.60
2	宁夏	7	0.66	2	1	海南	0.59
2	河北	9	0.66	3	7	广东	0.57
4	青海	7	0.65	3	12	四川	0.57
4	甘肃	9	0.65	3	6	广西	0.57
4	重庆	12	0.65	6	4	浙江	0.56
7	四川	12	0.64	6	3	青海	0.56
7	西藏	7	0.64	6	4	河南	0.56
9	黑龙江	9	0.63	6	6	江西	0.56
9	辽宁	11	0.63	6	5	贵州	0.56
9	内蒙古	8	0.63	6	2	西藏	0.56
9	海南	12	0.63	6	1	天津	0.56
9	天津	8	0.63	6	4	福建	0.56
14	吉林	10	0.62	6	2	宁夏	0.56
14	山西	9	0.62	15	5	湖北	0.55
14	贵州	12	0.62	15	7	湖南	0.55
17	山东	12	0.61	15	5	陕西	0.55
17	云南	12	0.61	15	5	辽宁	0.55
17	北京	7	0.61	15	1	重庆	0.55
20	江苏	12	0.60	15	5	江苏	0.55
20	陕西	9	0.60	15	5	安徽	0.55
22	湖北	12	0.59	15	7	山东	0.55
22	河南	12	0.59	15	9	甘肃	0.55
22	广西	12	0.59	15	3	山西	0.55
25	安徽	12	0.58	25	1	北京	0.54
25	广东	12	0.58	25	4	河北	0.54
25	上海	12	0.58	25	12	新疆	0.54
28	湖南	12	0.57	25	13	内蒙古	0.54
28	浙江	12	0.57	29	1	上海	0.53
28	福建	12	0.57	29	3	吉林	0.53
31	江西	12	0.56	29	9	黑龙江	0.53
	全国平均		0.62		160	全国平均	0.55

注:(1) 全年评分月数只是指降水;气温全年 12 个月均评分;(2) 气温栏中各省参与评分的站数,对于降水应为单月最多评分站数,即不一定该省某月所有的站降水量都在 10 mm 以上,当某站某月降水量的气候均值在 10 mm 以下时,该站便不评分;(3) 表中最后一栏为全国 31 个省区等权重平均。

表 4 2 级预报下随机预报评分的参数统计

Table 4 The parameter statistics of random forecast score in two grade forecasts

	随机预报评分 >0.80 的站次数	平均 准确率	参加评分 的站次数
降水	11	0.65	1526
气温	39	0.65	1920

气温则有 99.8% 站次大于由实际等级概率分布决定的随机预报评分。

以上分析说明,由于不同测站之间、不同月份之间,各等级样本气候概率分布的不均等,评分模型设计的缺陷,只做 2 级预报时,对测站之间随机预报评分不等的问题,起到了一种放大作用。

3.2 持续预报的随机预报准确率

持续预报是指在预报要素均值不变的时期内,某站某月某要素在只预报指定的一个等级。在所预报的要素各等级出现的概率相同的情况下,无论预报那个有利的等级,对于不同的测站,持续预报的评分总是相同的;如果各等级出现的概率不同,但参与评分的测站,各等级都按相同的概率分布,这时所有测站的 Ps 评分仍然具有可比性;目前我国气候预测评分,对于不同测站、不同月份,都是按实际样本距离气候均值的大小,实行统一的等级划分标准,所以等级划分中,不同测站、不同月份,各等级的分布就

不相同,这时的持续预报评分,对不同的测站就不具有可比性。

对于 6 级预报,假设预报员只预报 3 级或者只预报 4 级,(4 级预报下只报 2 级或 3 级),可以根据测站各等级的气候概率分布计算出持续预报评分。持续预报准确率,是在对每一测站分别计算其持续预报降水略偏少/气温略偏低、降水略偏多/气温略偏高两项气候评分后选其较高(有利)的评分所得的年平均评分。利用 30 年各等级的气候分布概率,可以计算出持续预报的准确率。

在表 5 中,第 2 列和第 3 列的数据为持续预报下随机预报准确率大于 0.80 的站次数;从表 5 可以看出:降水和气温在持续预报时,评分 >0.80 的站次数及与参加评分的站次数之间的比例基本持平;但降水多集中在预报偏少,气温则两者基本相等;第 4 列为全国 60 个测站的年平均持续预报准确率。

通过对比,不难发现,持续预报准确率高于 6 级分级 6 级预报、6 级分级 2 级预报下的随机预报评分,说明持续预报对评分模型缺陷放大的最明显。

表 5 持续预报下随机预报评分的参数统计

Table 5 The parameter statistics of forecast score in continuous forecast

	持续预报降水 偏少/气温偏低	持续预报降水 偏多/气温偏高	平均随机预 报准确率	参加评分 的站次数
降水	117	9	0.71	1526
气温	66	69	0.68	1920

4 中国各省持续预报评分的评估

通过以上讨论可知,在上述三种随机预报策略下,持续预报的评分是最高的。以持续预报为例,计算出中国各省年平均持续预报评分的评估结果,列于表 6。

表 6 全国各省持续预报准确率评估结果

Table 6 The continuous forecast accuracy calculated in every province

降水				气温			
排序	省份	全年评分月数	持续预报	排序	站数	省份	持续预报
1	青海	7	0.75	1	8	云南	0.73
2	海南	12	0.74	2	1	海南	0.71
2	四川	12	0.74	3	2	西藏	0.70
4	河北	9	0.73	3	12	四川	0.70
4	重庆	12	0.73	3	3	青海	0.70
6	新疆	12	0.72	3	6	广西	0.70
6	辽宁	11	0.72	3	4	浙江	0.70
6	贵州	12	0.72	3	5	贵州	0.70
6	西藏	7	0.72	3	4	河南	0.70
6	内蒙古	8	0.72	10	6	江西	0.69
6	吉林	10	0.72	10	7	广东	0.69
6	云南	12	0.72	10	5	江苏	0.69
13	宁夏	7	0.71	10	7	湖南	0.69
13	黑龙江	9	0.71	10	1	天津	0.69
13	甘肃	9	0.71	10	5	安徽	0.69
13	江西	12	0.71	10	4	福建	0.69
13	浙江	12	0.71	17	2	宁夏	0.68
13	广西	12	0.71	17	1	重庆	0.68
13	湖北	12	0.71	17	7	山东	0.68
13	陕西	9	0.71	17	5	湖北	0.68
13	广东	12	0.71	17	5	辽宁	0.68
13	山西	9	0.71	17	9	甘肃	0.68
23	安徽	12	0.70	17	3	山西	0.68
23	上海	12	0.70	17	5	陕西	0.68
23	江苏	12	0.70	25	1	上海	0.67
23	湖南	12	0.70	25	4	河北	0.67
23	山东	12	0.70	27	1	北京	0.66
23	北京	7	0.70	27	12	新疆	0.66
23	天津	8	0.70	29	13	内蒙古	0.65
23	福建	12	0.70	29	3	吉林	0.65
31	河南	12	0.69	29	9	黑龙江	0.65
	全国平均	12	0.71		160	全国平均	0.68

对比表3(随机预报)和表6(持续预报),可以看出:(1)对于所有省份的降水和温度评分,持续预报都高于实际概率分布下的随机预报评分,全国平均降水提高了9分,温度提高了13分。(2)在评分高低的排名中,虽然排名的顺序有所不同,但是表3中排名较高的测站,在表6的排名中仍然较高。如降水预报,在表3中评分的前16位,在表6中,有15位仍在并列13位之前;温度预报中,表3的前14位,在表6有13位仍在并列10位之前。(3)各省降水无技巧预报评分,随机预报最大相差14分,持续预报最大相差16分;温度随机预报最大相差7分,持续预报最大相差8分。因此,表3和表6均可以说明,各省的无技巧预报评分均存在着较大的差异。(4)随机预报评分的差异是由于测站预报对象等级分布不均造成的;而持续预报评分的差异则是由测站等级分布不均和预报评分表的漏洞所共同造成的。(5)从本文的分析中,可以给出3种不同的无技巧评估结果,这反而给无技巧预报评分的定量评估带来不确定因素。

5 小 结

在Ps评分中,预报技巧是由一段时间内实际业务的Ps评分与无技巧预报评分的差来衡量的。本文通过对中国各省无技巧预报评分的评估,讨论了在现行评分办法中,影响无技巧评分存在差异的两个因素。

在Ps评分中,由于全国使用了统一的距平等级划分标准,形成了影响无技巧预报评分差异的第一个因素。结果造成了不同测站之间、同一测站不同月份之间,各等级的气候概率分布出现差异,进而导致随机预报评分存在差异。

利用实际等级气候概率分布,评估了各省的随机预报评分。结果表明:对于不同的省份,降水随机预报准确率最大可相差14分,气温最大相差7分。这个结果说明,各省之间无技巧预报评分差异明显。

预报正确性判定的规定,形成了影响无技巧预报评分存在差异的第二个因素。其结果使得在只做2级预报或持续预报的情况下,可以大幅提高无技巧预报准确率,给无技巧预报评分的定量评估带来不确定因素。

对随机预报和持续预报的两种评估结果的比较说明:对于所有省份的降水和温度评分,持续预报都高于实际概率分布下的随机预报评分。持续预报降水全国平均提高了9分,温度平均提高了13分。

研究结果说明:由于各省之间的无技巧预报评分存在明显差异,因此直接利用Ps评分比较预报技巧,不能客观反映出预报能力的差异。

致谢:本文获赵振国先生的批评指正,特此鸣谢。

参考文献

- [1] 王绍武,朱锦红.短期气候预测的评估问题[J].应用气象学报,2000,11(增刊):1-10.
- [2] 陈桂英,赵振国.短期气候预测评估方法和业务初估[J].应用气象学报,1998,9(2):178-185.
- [3] 王雨,闫之辉.降水检验方案变化对降水检验评估效果的影响分析[J].气象,2007,33(12):53-61.
- [4] 马振锋,杨佑洪.成都区域气象中心短期气候预测业务评估[J].气象,2001,27(12):29-32.
- [5] 李辑,金巍,赵连伟.辽宁省近10年短期气候预测质量评估检验[J].气象,2007,33(4):82-87.
- [6] 吉庭艳.“八五”期间我省几种长期预报方法的评估检验[J].贵州气象,1999,23(3):36-42.
- [7] 郑峰,苗长明,张小伟.NRSM短期气候模式对温州台汛期预测能力检验(I):月和季降水检验[J].科技导报,2008,26(2):67-75.
- [8] 陈丽娟,李维京.月动力延伸预报产品的评估和解释应用[J].应用气象学报,1999,10(4):486-490.
- [9] 王会军,陈丽娟,李维京.中国区域月平均温度和降水的模式可预报性分析[J].2007,气象学报,65(5):725-732.
- [10] 姚筠,马晓群,许莹,等.2008年安徽省雪灾农业影响定量评估研究[J].气象,2010,36(1):86-90.
- [11] 朱艳峰,肖子牛,温娜.广义平衡反馈分析方法[J].气象,2010,36(4):1-6.
- [12] 王遂缠,吉慧敏.重要灾害性天气评分系统[J].气象,2004,30(9):42-45.