

西北 4 省(区)2009 年气候预测业务 质量新旧评估办法对比

林 纾, 郭俊琴, 张东方, 成青燕

(中国气象局西北区域气候中心, 甘肃 兰州 730020)

摘 要:通过对甘肃、青海、陕西、宁夏 4 省(区)2009 年预测业务新旧评估办法的对比发现, 无论是降水还是温度, 年平均和最高值旧办法均比新办法系统性偏高 15% 左右, 而最低值旧办法多数情况比新办法系统性偏低。旧办法最高值达满分的概率相对较高, 尤其是温度预测, 但新办法得满分的概率很小; 但一旦报反旧办法得零分的概率很高, 而新办法一般至少得 20 分。与短期预报和预测业务旧评估办法相比, 气候预测新评估办法更严格。

关键词:气候预测; 质量评估; 办法; 对比

中图分类号: P459.9

文献标识码: A

引 言

短期气候预测业务主要面对的是政府决策部门的公益服务, 在农业、交通、军事、气候资源等各方面的防灾减灾工作中起到重要的参谋作用, 故短期气候预测业务质量的高低是衡量服务效果的重要依据之一。因此如何科学地评估短期气候预测业务质量一直是不断探讨的问题, 气候预测质量评估办法也随着业务的发展不断更新。2009 年以前, 沿用的是 1999 年中国气象局减灾司出台的《短期气候预测质量评定暂行办法》^①(以下简称旧办法), 但从 2010 年 1 月起, 中国气象局预报与网络司开始执行《短期气候预测质量分级检验办法》^②(以下简称新办法), 新办法有较大幅度的改动, 为了更好地领会并在预测业务中灵活应用新办法, 本文就甘肃、青海、陕西、宁夏 4 省(区)2009 年预测业务质量应用新旧评估办法进行了对比分析和评述。

1 资料与方法

1.1 资 料

甘肃、青海、陕西、宁夏 4 省(区)2009 年预测与

实况资料分别来自各省(区)气候中心, 各有 21、20、14、4 个代表站参加评估, 与国家信息中心评估的站点一致。最高(低)分是指参加评估的所有月份中成绩最好(差)月份的分数。本文均使用预报定性评分 P_s , 单位: %。

1.2 旧、新办法评估规定及对比

为了新旧办法具有可比性, 均按照 6 级评分制来评定(表 1)。

在表 1 中看到, 新旧办法的降水等级划分没有变, 只在略少(多、高、低)级的预报用语上略有差别, 新办法中增加了“正常”2 字; 温度的等级划分差异较大, 新办法将各等级间的跨度加大到 1 °C。

旧办法中规定, 预测与实况符号相符为正确, 预测与实况符号不符时只允许跨一个量级算正确, 即预测与实况均为略多(少、高、低)量级算正确, 预测与实况跨 2 个量级为不正确。如果预测偏多(少、高、低), 实况也为偏多(少、高、低), 则奖励 0.5 的权重系数; 如果预测特多(少、高、低), 实况也为特多(少、高、低), 则奖励 1.0 的权重系数。

预报定性评分 P_s 可表示为: $P_s = [N_0 + 0.5 N_1$

收稿日期: 2010-04-22; 改回日期: 2010-04-30

基金项目: 国家自然科学基金面上项目“中国西部秋季降水机制研究”(40675066)及财政部和国家发改委“甘肃省风能资源详查和评价”项目共同资助

作者简介: 林纾(1964-), 女, 福建福州人, 正研级高工, 主要从事短期气候预测及相关领域研究工作。E-mail: treewest@163.com

① 关于印发执行《短期气候预测质量评定暂行办法》的通知. 中气预发[1999]15 号.

② 关于下发《短期气候预测质量分级检验办法》的通知. 气预函[2009]141 号.

$$+ 1.0 N_2] / [N + 0.5 N_1 + 1.0 N_2]$$

上式中, P_s 为预报定性评分, N 为参加评分的总站数, N_0 为预报与实况的距平符号相同的站数以及预报与实况虽然距平符号不同, 但都属略多、略少级的站数; N_1 为偏多(少)异常预报正确的站数, N_2 为特多(少)异常预报正确的站数。各级评分情况见表2。

新方法规定当预测与实况的距平符号和量

级均一致时, 评分为 100%; 当预测与实况的量级相差 1 个级别时, 减 20%; 量级相差 2 个级别时, 减 40%; 依次类推, 减至 0 为止; 当预测与实况的距平符号不一致时, 在量级减分的基础上再减 20%; 减至 0 为止; 鼓励预报异常, 当预报为异常级且实况与预报相差 1 个量级时, 可以在上述得分的基础上再加 10%。各级评分情况见表 3。

表 1 新旧办法气温、降水趋势预测 6 级评分制用语及各等级划分标准

Tab. 1 Terms for six levels evaluation of temperature and precipitation prediction by the old and new methods and criteria for the six levels

	预测用语	特少(低)	偏少(低)	略少(低)	略多(高)	偏多(高)	特多(高)
旧办法	ΔR	$\Delta R \leq -50$	$-50 < \Delta R \leq -20$	$-20 < \Delta R < 0$	$0 \leq \Delta R < 20$	$20 \leq \Delta R < 50$	$50 \leq \Delta R$
	ΔT	$\Delta T \leq -1.0$	$-1.0 < \Delta T \leq -0.5$	$-0.5 < \Delta T < 0.5$	$0 \leq \Delta T < 0.5$	$0.5 \leq \Delta T < 1.0$	$1.0 \leq \Delta T$
	预测用语	特少(低)	偏少(低)	正常略少(低)	正常略多(高)	偏多(高)	特多(高)
新办法	ΔR	$\Delta R \leq -50$	$-50 < \Delta R \leq -20$	$-20 < \Delta R < 0$	$0 \leq \Delta R < 20$	$20 \leq \Delta R < 50$	$50 \leq \Delta R$
	ΔT	$\Delta T \leq -2.0$	$-2.0 < \Delta T \leq -1.0$	$-1.0 < \Delta T < 0$	$0 \leq \Delta T < 1.0$	$1.0 \leq \Delta T < 2.0$	$2.0 \leq \Delta T$

ΔR :降水距平百分率/%; ΔT :气温距平/°C

表 2 旧办法气温、降水趋势预测 6 级检验评分制单站评分表(单位:%)

Tab. 2 Table for single station evaluation of temperature and precipitation prediction by the old method in the six levels(Unit:%)

实况	预测	特少(低)	偏少(低)	略少(低)	略多(高)	偏多(高)	特多(高)
特少(低)	100 + 100 × 1.0	100	100	100	0	0	0
偏少(低)	100	100 + 100 × 0.5	100	100	0	0	0
略少(低)	100	100	100	100	100	0	0
略多(低)	0	0	100	100	100	100	100
偏多(低)	0	0	0	100	100 + 100 × 0.5	100	100
特多(低)	0	0	0	100	100	100 + 100 × 1.0	100

表 3 新办法气温、降水趋势预测 6 级检验评分制单站评分表(单位:%)

Tab. 3 Table for single station evaluation of temperature and precipitation prediction by the new method in the six levels(Unit:%)

实况	预测	特少(低)	偏少(低)	正常略少(低)	正常略多(高)	偏多(高)	特多(高)
100	80 + 10	60	20	0	0	0	0
偏少(低)	80 + 10	100	80	40	20	0	0
正常略少(低)	60	80 + 10	100	60	40	20	20
正常略多(高)	20	40	60	100	80 + 10	60	60
偏多(高)	0	20	40	80	100	80 + 10	80 + 10
特多(高)	0	0	20	60	80 + 10	100	100

2 2009 年新旧办法评估结果对比

2.1 降水

表 4 是西北 4 省(区)2009 年降水新旧办法的对比结果。可以清楚地看到,新办法比旧办法成绩偏低较多,且旧办法成绩愈高新办法下降幅度愈大。如 2009 年青海旧办法评估为 74.4%,新办法为 60.9%,下降 13.5%;甘肃旧办法 69.4%,新办法为 58.6%,下降 10.8%,但陕西和宁夏的成绩用旧办法评估原本较低,新办法仅下降约 5%。

最高分新办法比旧办法成绩高得多,4 省平均相差 23.1%,宁夏 2 种办法的差异最大,达 35%。但最低分新办法多数情况比旧办法高,最大能够偏高 17.2%。

举 2 个特别的例子更能凸显新旧办法的差异。

以宁夏 4 月降水为例:预测 4 月降水均偏少 2 成以内,实况是全区偏少 5 成以上。旧办法评分是各站趋势正确,得满分,而新办法预测与实况相差 2 个量级,各站均得 60%。同样的预测与实况但 2 种不同评估办法的结果能差 40%。同样宁夏 6 月降水:预测 6 月降水为偏多 1 成,属略多的范畴,实况是全区特少,偏少 9 成以上。旧办法是各站趋势相反,得零分,而新办法各站能得 20%,这时新旧办法的评估结果相差 20%。

表 4 2009 年降水新旧办法评估对比结果(单位:%)

Tab.4 Results of comparison of precipitation prediction evaluation by the old and new methods in 2009

	方法	平均 Ps	平均时段	最高分	最低分
甘肃	旧	69.4	3~10 月	92.3	45.0
	新	58.6		75.4	47.0
陕西	旧	58.9	3~10 月	85.7	21.4
	新	53.9		71.4	38.6
宁夏	旧	50.0	1~10 月	100.0	/
	新	45.6		65.0	20.0
青海	旧	74.4	3~10 月	90.0	65.0
	新	60.9		64.0	55.5

2.2 温度对比

温度与降水相似,新办法比旧办法成绩平均偏低 17.4%,同样也是旧办法成绩愈高新办法下降幅

度愈大。如 2009 年宁夏旧办法评分为 100%,新办法仅为 71%,下降 29%;青海旧办法评分 94.6%,新办法为 75.0%,下降 19.6%。

从表 5 中可看到,4 省(区)使用旧评估办法的最高分都是满分,且一年中有 80% 以上的月份能够达到这个成绩,但新办法很难达到 100%,就 2009 年看最高分是 90.5%。最高分新办法比旧办法成绩偏高 10%~15%;除青海外,其它省份最低分新办法比旧办法平均偏高 30% 左右。

表 5 2009 年温度新旧办法对比结果(单位:%)

Tab.5 Results of comparison of temperature prediction evaluation by the old and new methods in 2009

	方法	平均 Ps	平均时段	最高分	最低分
甘肃	旧	87.7	1~12 月	100.0	14.0
	新	75.2		90.5	41.0
陕西	旧	82.2	3~10 月	100.0	14.3
	新	73.8		85.7	48.6
宁夏	旧	100.0	1~10 月	100.0	/
	新	71.0		85.0	60.0
青海	旧	94.6	1~11 月	100.0	70.0
	新	75.0		87.0	64.0

3 新旧评估办法评述

自从发布短期气候预测产品以来,评估预测质量一直以趋势是否正确为依据。旧办法基本遵循这个原则,报对趋势即可得满分,并能够充分鼓励报异常,对预测不同量级的异常给不同的奖励;但新办法的评估规则比旧办法严格得多,必须报对量级才能得满分,这点就目前短期气候预测的技术发展水平而言很难做到,且仅从 2009 年的成绩看,新办法无论降水还是温度系统性下降的幅度较大。

4 与短期预报评估办法的比较

同样是预报评估,现在与短期预报评估进行简单的比较。

以短期预报的晴雨预报评估为例,评估办法很简单^③,对单站晴雨预报而言,如果报有降水,只要实况出现微量以上的降水就得满分,否则不得分,预报量级的跨度有无限大;但对月平均降水 < 50 mm

③ 关于下发《中短期天气预报质量检验办法(试行)》的通知. 气发[2005]109 号.

的站点,气候预测的跨度只有5 mm,月平均降水 > 50 mm 同时 < 100 mm 的站点,气候预测的跨度为10 mm,即使月平均降水 > 100 mm,预测跨度也仅20 mm,这本身对气候预测已经具有很高的要求。

以兰州站4月降水预测为例。兰州站4月降水多年平均降水为14.7 mm,假如预测4月降水量为13~18 mm,其下限13 mm 比多年平均偏少12%,上限18 mm 比多年平均偏多22%,应该是偏少12%到偏多22%之间(理论上气候预测也应该有一个预报值的范围,类似短期预报中报小雨、中雨等,其实已经界定了一个雨量范围),但目前都是以13~18 mm 的中间值15.5 mm 来评估的,即比多年平均偏多5%,实质上就认为预测兰州4月降水是15.5 mm,已经是一个定量的概念。新办法偏多5%是略多的范畴,只有实况值落在14.7~17.6 mm 之间才能得满分,得满分的跨度只有3 mm,这显然对短期气候预测预报员而言过于苛刻。而旧评分办法只要实况降水 > 10.2 mm (-29.9%) 就可以得满分,不过降水若 < 10.3 mm 不得分,由此看到2种评估办法的差异甚大。笔者认为,还是要从短期气候预测业务的实际情况出发,比较完整地考虑气候业务评估技术,既要与旧办法有衔接和延续,又要横向比较,与短期预报评估有一定的可比性,这样才能更好地促

进短期气候预测业务评估技术向着科学化方向发展。

5 小结与讨论

(1) 新办法与旧办法相比,无论是降水还是温度,年平均和最高值新办法均比旧办法系统性偏低15%左右,而最低值新办法多数情况比旧办法系统性偏高。

(2) 旧办法最高值达满分的概率相对较高,尤其是温度预测,但新办法得满分的概率很小;但一旦报反旧办法得零分的概率很高,而新办法一般至少得20分。

(3) 通过新旧办法的对比,新办法更是让预报员从得分的角度预测“正常略多(少)”量级,不利于鼓励预测异常量级。

(4) 通过与短期预报评估办法的对比,气候预测评估办法更严格且有一定的不合理性,所以建立一套科学的短期气候预测业务评估技术和办法是迫在眉睫的工作。

致谢:感谢青海、陕西、宁夏3省(区)气候中心提供2009年预测与实况资料。

Comparison of the Quality Evaluation of Climate Prediction Operation for Northwest China in 2009 by Old and New Evaluation Methods

LIN Shu, GUO Junqin, ZHANG Dongfang, CHENG Qingyan

(Northwest Regional Climate Center, Lanzhou 730020, China)

Abstract: Through comparison of the evaluation of climate prediction operation for Gansu, Qinghai, Shaanxi and Ningxia in 2009 by the old and new evaluation methods respectively, for precipitation and temperature, it has been found that both the average annual score and the maximum score by the old method were systematic higher about 15% than that by the new method, but the minimum score by the old method was systematic lower than that by the new method in most cases. It was relatively higher probability to reach full mark for the maximum score by the old method especially to temperature prediction, and less probability by the new method. A zero score was high probability by the old method when the prediction was opposite to the fact, but 20 points by the new method at least. The new evaluation method for climate prediction was more rigorous than not only the old method, but also the evaluation method for weather forecast.

Key words: climate prediction; quality evaluation; method; comparison