

杨松¹ 杞明辉² 姚德宽¹

(1. 云南省保山市气象台, 678000; 2. 云南省气象科学研究所)

提 要

根据误差的特性,即误差由系统误差和随机误差组成,对保山5月雨量的几种预报方法的预报结果进行分析,寻找系统误差的订正方法,降低系统误差的影响。将几种预报方法订正后的结果平均从而消除随机误差,其预报结果明显改善,平均误差和均方根误差明显降低,预报准确率明显提高,可以在业务预测中广泛应用。

关键词: 预报集成 回归分析 5月雨量 系统误差

引 言

由于气候系统本身具有高度的非线性,影响它的因子非常多,物理过程非常复杂,而多种气候预测的模式和模型都是对实际系统的简化和抽象,其包含的因子必定是有所选择并十分有限,因此需对它们的预测结果进行客观集成。近年来,预报集成方法的研究越来越受到国内外的研究,解明恩等提出了一些可用的集成方法^[1],本文从预报误差产生的机制及其性质入手,分析保山5月雨量的几种预报方法的系统性误差,确定系统误差并进行订正,对几种预报方法的预测结果进行简单平均消除随机误差从而做到预报集成。

1 误差的订正原理^[2]

目前国内外统计预报大都以线性回归作为其数学模型,由于预报因子和预报对象间的关系并非线性回归中的线性关系,而多为非线性关系,不可避免地造成预报误差,如果我们将预报方程计算出的预报量记为 Y_f ,预报方程产生的系统误差表示为 ζ ,随机误差部分表示为 ϵ ,则有:

$$Y = Y_f + \zeta + \epsilon \quad (1)$$

其中 $Y = \{Y_1, Y_2, Y_3, \dots, Y_n\}$ 为预报

对象向量;

$Y_f = \{Y_{f1}, Y_{f2}, Y_{f3}, \dots, Y_{fn}\}$ 为预报方程计算出的预报向量;

$\zeta = \{\zeta_1, \zeta_2, \zeta_3, \dots, \zeta_n\}$ 为预报值中含有的随机误差向量,它具备随机系列的特征和统计特征;

$\epsilon = \{\epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_n\}$ 为预报向量中系统误差。

在通常情况下,含有随机误差 ϵ 的时间序列样本 Y_f 总是满足正态分布,可通过平均的办法消除随机误差:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^N \epsilon_i = 0 \quad (2)$$

在实际业务中 $N \rightarrow \infty$ 是不可能的,但当 N 达到一定的临界值时,随机正负相互抵消后可以接近0值或某个可以忽略不计的小值。本文主要针对保山日常业务上对保山5月雨量预报比较好的4种方法的预报集成, N 取3,值比较小,不能完全消除随机误差。

由于预报方法是线性回归,用一条直线去拟合一条复杂的曲线,必然存在系统误差,我们对误差值与预报因子预报对象进行分析,发现系统误差值与预报对象存在正相关,相关系数较高,我们对预报值近似代替预报

① 云南省“十五”科技攻关项目 2001NG4304 资助

对象并按值的大小等分成 $i = 3$ 级,对历史样本分级确定系统误差值。

$$ER_i = \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - Y_{fi}) \quad (3)$$

其中 i 值为分别表示分成 i 等级订正, ER_i 第 i 级的系统误差, N_i 为 i 级的样本数, Y_{ij} 、 Y_{fi} 分别为第 i 级的第 j 个样本的 5 月雨量实况值和预报值。

2 预报模式系统误差的确定

4 种预报 5 月雨量的方法分别为: 500hPa 高度场回归模型(A)、海温资料回归预测模型(B)、大气特征资料回归预测模型(C)、均生函数预测模型(D)。其中 A 至 C 模型的回归方程为:

$$A \text{ 模型: } Y_A = -249.743 + 6.613x_1 - 3.107x_2 - 4.297x_3 + 2.263x_4$$

$$B \text{ 模型: } Y_B = -1092.438 + 0.98x_5 + 1.384x_6 + 1.898x_7 + 1.145x_8$$

$$C \text{ 模型: } Y_C = -18.781 + 2.638x_9 + 1.133x_{10} + 3.664x_{11} - 4.378x_{12}$$

式中 Y_A 、 Y_B 、 Y_C 均为当年保山 5 月雨量的预报值, x_1 为上年 10 月 500hPa 高空 (35°N 、 120°W) 格点高度值, x_2 当年 1 月 500hPa 高空 (85°N 、 75°W) 格点资料, x_3 上年 500hPa 高空 9 月 (70°N 、 125°W) 格点资料, x_4 上年 11 月 500hPa 高空 (65°N 、 40°W) 格点资料; x_5 为上年 3 月 (40°N 、 150°W) 格点的海温资料, x_6 为上年 3 月 (15°N 、 125°W) 格点的海温资料, x_7 为上年 10 月 (25°N 、 170°W) 格点的海温资料, x_8 为上年 11 月 (40°N 、 165°W) 格点的海温资料, x_9 为上年 9 月大西洋欧洲环流型 E, x_{10} 为当年 3 月亚洲经向环流指数 ($60 \sim 150^\circ\text{E}$), x_{11} 为上年 10 月印度副高脊线 ($65 \sim 95^\circ\text{E}$); x_{12} 为上年 11 月东太平洋副高脊线。

建立上述预报方程样本 1952~1993 年共 42 年,以 A 种模式为例,计算预报误差与预报因子、预报对象和预报结果的相关系数,

从表 1 中可得到预报误差与实况有明显的正相关,同时绘制 B 模型预报误差与实况的点聚图(如图 1),均说明了预报误差与实况值存在线性相关,随着实况值的增加,误差由负值向正值过渡,并逐步增加。而预报误差与预报因子和预报值相关较低,是因为建立预报方程时预报因子对预报对象已进行线性拟合。

表 1 A 模型预报误差与预报因子和对象的相关分析

| 因子 | X_1 | X_2 | X_3 | X_4 | 实况 | 预报结果 |
|------|-------|-------|--------|-------|-------|-------|
| 相关系数 | 0.00 | 0.00 | -0.003 | 0.001 | 0.668 | 0.001 |

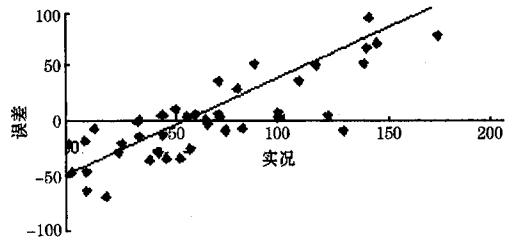


图 1 B 模型预报误差与预报对象点聚图

我们计算了 4 种模型的雨量实况与预报结果的相关系数如表 3 所示,实况与预报结果也存在较高的相关性,因此用 5 月雨量预报值代替实况值,对预报结果进行系统误差订正值的确,由于样本数太少,要消除随机误差,我们将 42 年样本预报值从小到大进行排列等分成 3 等分,每份 14 个样本,将其误差值平均得到各个级别的系统误差,即预报订正值。4 种模型的订正值如表 2 所示。

表 2 4 种模型预报订正值

| 项目 | 1 级 | 2 级 | 3 级 |
|-----------|-----------------|-------------------|-----------------|
| A 模型预报值范围 | $Y_A \leq 49.5$ | $49.5 < Y_A < 75$ | $Y_A \geq 75$ |
| A 模型订正值 | 5.9264 | -17.5065 | 9.6628 |
| B 模型预报值范围 | $Y_B \leq 53$ | $53 < Y_B < 74$ | $Y_B \geq 74$ |
| B 模型订正值 | 5.9561 | -10.2221 | 4.8891 |
| C 模型预报值范围 | $Y_C \leq 52$ | $52 < Y_C < 71.9$ | $Y_C \geq 71.9$ |
| C 模型订正值 | -1.4464 | -11.9791 | 13.3913 |
| D 模型预报值范围 | $Y_D \leq 47$ | $47 < Y_D < 72$ | $Y_D \geq 72$ |
| D 模型订正值 | 3.8714 | -2.5864 | -1.027 |

3 预报集成

上述误差订正工作消除了部分系统误差,假设我们已消除全部系统误差,则订正后

的预报结果仅存在随机误差,根据误差订正原理,随机误差的时间序列样本总是满足正态分布,可通过平均的办法消除随机误差,这里我们通过对4种预报结果订正后平均来消除随机误差,从而完成预报集成。

4 误差订正的统计特性

我们利用4种预报模型历史(1952~1993年)拟合值,计算预报与实况的相关系数、预报均方根误差、预报平均误差,如表3所示,每种预报模型通过误差订正后,相关系数增加,误差降低,最终预报集成后,相关系数最高,而平均误差和均方根误差达到最低。

表3 4种预报模型误差修正结果对比分析(mm)

| 项目 | 预报与实况 相关系数 | 平均误差 | 均方根误差 |
|--------|---------------|-------|-------|
| A模型 | 0.753 | 23.98 | 28.76 |
| A模型订正后 | 0.795 | 21.55 | 26.53 |
| B模型 | 0.566 | 26.67 | 36.02 |
| B模型订正后 | 0.587 | 27.50 | 35.37 |
| C模型 | 0.648 | 25.73 | 33.48 |
| C模型订正后 | 0.698 | 24.56 | 31.30 |
| D模型 | 0.84 | 19.51 | 23.69 |
| D模型订正后 | 0.846 | 19.33 | 23.45 |
| 预报集成 | 0.899 | 18.00 | 23.10 |

5 预报集成试验

我们利用上述集成方法,对保山市1994~2002年5月雨量进行预报集成试验,结果如表4和表5所示,预报集成后,误差明显降低,均方根误差和平均误差均低于任何一种预报模型,如表6所示。

表4 1994~2002年5月雨量预报结果(mm)

| 年份 | 实况 | A模型 | B模型 | C模型 | D模型 | 集成 |
|------|-------|--------|--------|--------|--------|--------|
| 1994 | 22.8 | 68.92 | 55.01 | 54.20 | 13.04 | 38.83 |
| 1995 | 95.7 | 76.59 | 98.02 | 146.36 | 79.76 | 106.91 |
| 1996 | 110.6 | 162.72 | 48.49 | 112.88 | 99.15 | 112.81 |
| 1997 | 45.1 | 45.57 | 67.24 | 70.39 | 67.77 | 57.86 |
| 1998 | 73.6 | 64.90 | 63.12 | 76.11 | 88.92 | 69.42 |
| 1999 | 90.0 | 136.91 | 105.16 | 95.67 | 24.18 | 98.43 |
| 2000 | 156.4 | 141.97 | 149.41 | 130.87 | 110.64 | 139.95 |
| 2001 | 219.5 | 183.93 | 146.56 | 146.00 | 117.64 | 155.26 |
| 2002 | 185.7 | 141.52 | 79.02 | 176.53 | 51.69 | 118.53 |

根据中国气象局制定的短期气候预测6级评分标准对预报结果进行评定如表7所

示,集成后的预报准确率达100%,这与预报时间短有一定关系,但是通过误差订正集成后预报准确率明显提高了。

表5 1994~2002年5月雨量预报
结果距平百分率(%)

| 年份 | 实况 | A模型 | B模型 | C模型 | D模型 | 集成 |
|------|-----|-----|-----|-----|-----|-----|
| 1994 | -64 | 8 | -14 | -15 | -80 | -39 |
| 1995 | 50 | 20 | 53 | 129 | 25 | 67 |
| 1996 | 73 | 154 | -24 | 76 | 55 | 76 |
| 1997 | -30 | -29 | 5 | 10 | 6 | -10 |
| 1998 | 15 | 1 | -1 | 19 | 39 | 8 |
| 1999 | 41 | 114 | 64 | 49 | -62 | 54 |
| 2000 | 144 | 122 | 133 | 104 | 73 | 119 |
| 2001 | 243 | 187 | 129 | 128 | 84 | 143 |
| 2002 | 190 | 121 | 23 | 176 | -19 | 85 |

表6 1994~2002年预报检验统计值

| 项目 | 预报与实况 相关系数 | 平均误差 | 均方根误差 |
|------|---------------|-------|-------|
| A模型 | 0.825 | 29.73 | 36.82 |
| B模型 | 0.675 | 36.78 | 48.19 |
| C模型 | 0.863 | 25.11 | 35.99 |
| D模型 | 0.571 | 46.95 | 53.19 |
| 预报集成 | 0.931 | 22.52 | 32.41 |

表7 1994~2002年保山5月雨量预测评定

| 年份 | A模型 | B模型 | C模型 | D模型 | 集成 |
|------|-----|-----|-----|-----|------|
| 1994 | - | + | + | + | + |
| 1995 | + | + | + | + | + |
| 1996 | + | - | + | + | + |
| 1997 | + | - | - | - | + |
| 1998 | + | + | + | + | + |
| 1999 | + | + | + | - | + |
| 2000 | + | + | + | + | + |
| 2001 | + | + | + | + | + |
| 2002 | + | + | + | - | + |
| 准确率 | 89% | 78% | 89% | 67% | 100% |

6 结语

误差订正预报集成试验表明:集成后的预报准确率明显提高,预报平均误差和均方根误差明显降低,集成效果较好,对日常预报有指导作用。

参考文献

- 1 解明恩,张万诚,沙文生.云南短期气候预测集成方法的试验研究.云南气象,2002,(1):27~32.
- 2 杞明辉,刘建宇,杨松.一种数值预报产品的误差修订方法及效果检验.云南气象,2002,(2):17~20.

Application of Error Correction in Forecast Ensemble

Yang Song¹ Qi Minghui² Yao Dekuan¹

(1. Baoshan Meteorological Office, Yunnan Province 678000; 2. Yunnan Province Meteorological Institute)

Abstract

In consideration of the error comprised of systematic and random error, the error analysis of the May rainfall forecast of several methods in Baoshan, Yunnan Province is made. And first, the systemic error correction method of each forecast technique is found out, to reduce its effect. Then, the corrected forecast is averaged, to eliminate the random error. The results show that mean error and mean square root error decrease obviously, and the forecast accuracy is improved obviously.

Key Words: forecast ensemble rainfall systematic error random error