

文章编号:1005-3085(2005)07-0127-06

雨量预报方法的评价模型

伍利兵, 雷中博, 王翠

指导教师: 周国鹏

(咸宁学院, 湖北咸宁 437005)

编者按: 论文较完整的解决了雨量预测方法的评估问题。论文合理的选用了与距离有关的插值函数, 得到了观测站处的预报值, 给出了评价两种预报方法优劣的指标。特别是论文不仅考虑了预报值与测量值之间误差的大小, 还考虑了预报的稳定性, 使评价指标更加完善。不过, 该论文的第二问稍显单薄, 未能考虑不同等级雨量的误报对公众的不同影响。

摘 要: 本文建立了“最邻近点插值法”、“反距离加权平均法”等两个降雨量预报算法模型, 给出各观测站的雨量预报值, 并且用三项指标对两种雨量预报准确性进行了评价。对于问题二, 给出了满意度函数用来评价公众满意程度。结果表明两种预报方法公众的满意度都在95%以上。

关键词: 最邻近点插值法; 反距离加权平均法; 满意度函数

分类号: AMS(2000) 65D17

中图分类号: O29; P456

文献标识码: A

1 模型的基本假设

1) 假设所有预报数据和实测数据及预报点和观测站的经纬度坐标值均有效, 即不考虑人为因素造成的无效数据。

2) 假设当两地距离大于某给定 $\delta > 0$ 时, 两地之间的降雨量没有必然联系。

2 符号说明

$A_i = (a_i, b_i)$: 表示91个观测站的经纬度坐标 ($i = 1, 2, \dots, 91$)。

(x_{jk}, y_{jk}) : 表示 53×47 个等距网格点上第 j 行、第 k 列的经纬度坐标 ($j = 1, 2, \dots, 53; k = 1, 2, \dots, 47$)。

$z_1(j, k, d, t), z_2(j, k, d, t)$: 分别表示第一、二种预报方法对预报点 (x_{jk}, y_{jk}) 第 d 天第 t 时段的降雨量预报值 ($d = 1, 2, \dots, 41; t = 1, 2, 3, 4$)。

$z^*(i, d, t)$: 表示观测站 A_i 第 d 天第 t 时段的降雨量实测值。

$\hat{z}_1(i, d, t), \hat{z}_2(i, d, t)$: 分别表示两种预报方法对观测站 A_i 第 d 天第 t 时段的降雨量预报值。

r_{ijk} : 表示观测站 A_i 与预报点 (x_{jk}, y_{jk}) 之间的距离。

$\varepsilon_1(i, d, t), \varepsilon_2(i, d, t)$: 表示两种预报方法对观测站 A_i 预报值的绝对误差。

E_1, E_2 : 表示两种预报方法预报值绝对误差的数学期望。

D_1, D_2 : 表示两种预报方法预报值绝对误差的方差。

G_1, G_2 : 表示按降雨量等级预报时公众满意度函数。

Sa : 表示公众对预报方法的平均满意度。

3 分析与建模

本题给出了一段时间内若干预报点用两种不同方法进行降雨量预报的预报数据,同时给出了91个观测站降雨量实测值,要求我们评价两种6小时降雨量预报方法的准确性。要检验预报的准确性,只能通过对预报点的预报数据进行分析,建立预报模型,通过预报模型预报观测站的预报值,然后将预报值与实测值进行比较,得到预报准确性的评价方法。我们首先将利用两种插值算法,建立相关数学模型,对观测站的降雨量进行预报。然后再利用相关的评价算法,对前两种预报方法进行相应的评价。

3.1 预报算法设计

由于此问题数据量大,选择算法非常关键。模型结构简单,计算量小,预报准确性高,是人们追求的目标。然而,熊掌鱼翅不可兼得,我们将给出两种预测模型,它们各有所长。

3.1.1 预报算法1(最邻近点插值法)^[2]

人们最容易想到、最简单的一种算法是“最邻近点插值法”。这种算法的思想是将与 A_i 最邻近的点 (x_{jk}, y_{jk}) 的预报值 $z_1(j, k, d, t), z_2(j, k, d, t)$ 作为此观测站预报值,即

$$\hat{z}_1(i, d, t) = z_1(j, k, d, t), \quad \hat{z}_2(i, d, t) = z_2(j, k, d, t), \quad d = 1, 2, \dots, 41, \quad t = 1, 2, 3, 4. \quad (1)$$

我们用此算法对观测站的雨量进行了预报,得到如图1所示的各观测站的雨量预报值。

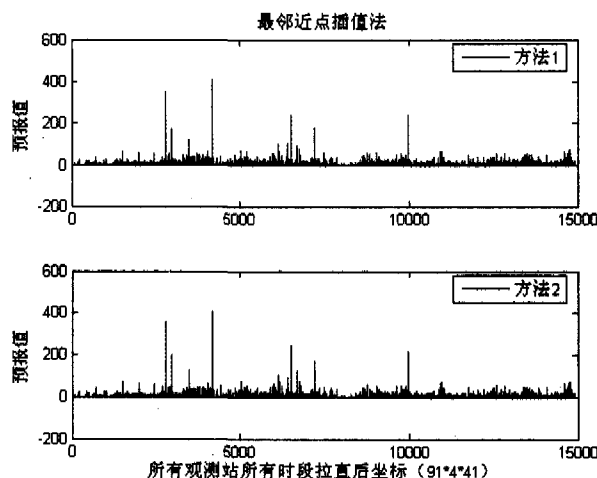


图 1: 最邻近点插值算法对所有观测站所有时段两种预报方法的预报值

3.1.2 预报算法2(反距离加权平均法)^[2]

由于观测站 A_i 处的降雨量不仅与它最邻近点的降雨量有关,而且与它周围的降雨量存在着密切的关系,对 A_i 处降雨量的预报,应考虑其邻近网格点预报值对它的影响。我们将寻找 A_i 的一个邻域 $U(A_i, \delta)$, 并假设此邻域外的网格点降雨量对 A_i 的影响可以忽略不计。其中 (x_{jk}, y_{jk}) 到 A_i 的距离为

$$r_{ijk} = \sqrt{(a_i - x_{jk})^2 + (b_i - y_{jk})^2}. \quad (2)$$

对于 $(x_{jk}, y_{jk}) \in U(A_i, \delta)$, 两种预报方法对 A_i 点在第 d 天、第 t 时间段的预报值定义为

$$\hat{z}_i(i, d, t) = \begin{cases} z_1(j, k, d, t), & \exists (x_{jk}, y_{jk}), \quad r_{ijk} < \frac{1}{10}\delta, \\ \sum_{(x_{jk}, y_{jk}) \in U(A_i, \delta)} W_{jk}(i) z_1(j, k, d, t), & \forall (x_{jk}, y_{jk}), \quad r_{ijk} \geq \frac{1}{10}\delta, \end{cases} \quad (3)$$

其中

$$W_{jk}(i) = \frac{1}{r_{ijk}^2} \bigg/ \sum_{(x_{jk}, y_{jk}) \in U(A_i, \delta)} \frac{1}{r_{ijk}^2}.$$

此算法在“反距离加权平均法”(又称 Shepard 方法)基础上进行了适当改进, 其基本思想是对于 $(x_{jk}, y_{jk}) \in U(A_i, \delta)$, (x_{jk}, y_{jk}) 对 A_i 的预报值的影响随它们之间距离的增大而减小。值得注意的是, 当存在预报点 (x_{jk}, y_{jk}) 离 A_i 足够近时(如: $r_{ijk} < \frac{1}{10}\delta$), 按“反距离加权平均法”, 预报点 (x_{jk}, y_{jk}) 对 A_i 的影响很大, 同时, 为了防止直接利用“反距离加权平均法”计算时出现奇异, 我们直接将预报点 (x_{jk}, y_{jk}) 的预报值当成观测站 A_i 预报值。

取 $\delta = 0.3$, 通过计算, 得到两种预报方法各观测站的全部预报值如图2所示。

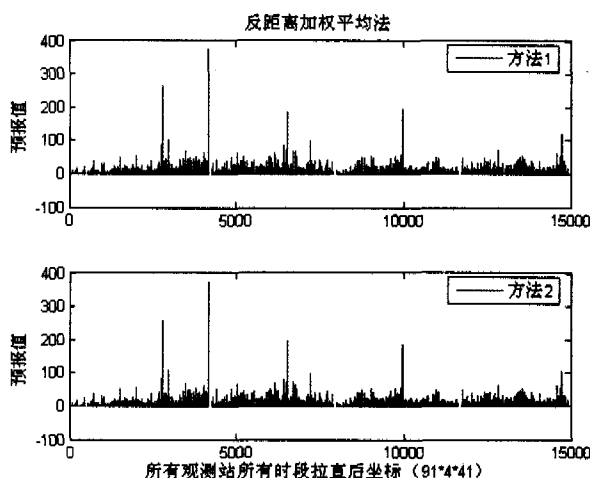


图 2: 反距离加权平均法对所有观测站所有时段两种预报方法的预报值

3.2 6小时雨量预报方法准确性评价指标算法

观测站 A_i 第 d 天 t 时段预报绝对误差为

$$\varepsilon_l(i, d, t) = |\hat{z}_l(i, d, t) - z^*(i, d, t)|, \quad l = 1, 2. \quad (4)$$

91个观测站41天 t 时段的预报绝对误差的均值为

$$E_l(t) = \frac{1}{41 \times 91} \sum_{d=1}^{41} \sum_{i=1}^{91} \varepsilon_l(i, d, t), \quad l = 1, 2. \quad (5)$$

按时段分, 第 t 时段预报绝对误差的方差为

$$D_l(t) = \sum_{d=1}^{41} \sum_{i=1}^{91} (\varepsilon_l(i, d, t) - E_l(t))^2, \quad l = 1, 2. \quad (6)$$

ε_l 为观测站降雨量的实测值和预报值之间的偏差, 反映了预报值偏离观测值的程度, ε_l 越小越好。本题, 我们关心的是预报值和观测值之间的差距是多少, 至于是偏大还是偏小都不影响对两种预报方法的评价, 而且运用统计方法对误差绝对值进行处理时, 不会出现误差相互抵消的情况, 所以误差只有定义为预报值偏离实测值的绝对值, 才有较好的实际意义。 $D_l(t)$ 用来评价第 l 种方法在41天对91个观测站第 t 时段预报效果的好坏。

3.3 问题二的求解算法

问题2中, 气象部门将降雨量分成7个等级, 我们用 0, 1, 2, 3, 4, 5, 6 等7个数字分别表示无雨、小雨、中雨、大雨、暴雨、大暴雨及特大暴雨等, 将降雨量定义为如下分段函数

$$f(x) = \begin{cases} 0, & \forall x \in [0, 0.1) \\ 1, & \forall x \in [0.1, 2.5) \\ 2, & \forall x \in [2.5, 6) \\ 3, & \forall x \in [6, 12) \\ 4, & \forall x \in [12, 25) \\ 5, & \forall x \in [25, 60) \\ 6, & \forall x \in [60, \infty). \end{cases} \quad (7)$$

为了考虑公众对天气预报的感受, 定义如下公众满意度函数

$$G_l(\hat{z}_l(i, d, t), z^*(i, d, t)) = \frac{6 - |f(\hat{z}_l(i, d, t)) - f(z^*(i, d, t))|}{6}. \quad (8)$$

这里, 我们对公众的满意度进行了量化, 取等级误差为零 ($|f(\hat{z}_l(i, d, t)) - f(z^*(i, d, t))| = 0$) 时满意度为1, 等级误差的差距最大 ($|f(\hat{z}_l(i, d, t)) - f(z^*(i, d, t))| = 6$) 时满意度为0. 其它的情况满意度介于0和1之间, 且与级差成反比. 利用此算法, 容易得到公众对预报的满意程度. 定义如下公众对两种预报方法的平均满意度函数

$$S_{al}(t) = \frac{1}{91 \times 41} \sum_{d=1}^{41} \sum_{i=1}^{91} G_l(\hat{z}_l(i, d, t), z^*(i, d, t)), \quad t = 1, 2, 3, 4, \quad l = 1, 2; \quad (9)$$

$$S_{al} = \frac{1}{91 \times 41 \times 4} \sum_{t=1}^4 \sum_{d=1}^{41} \sum_{i=1}^{91} G_l(\hat{z}_l(i, d, t), z^*(i, d, t)), \quad l = 1, 2. \quad (10)$$

$S_{al}(t)$ 表示41天内第 t 时段公众对第 l 种预报方法的满意度; S_{al} 表示公众对第 l 种预报方法的平均满意度. 用上述算法, 对题中给出的两种预报数据进行分析、计算, 得到了相关满意度结果 (见模型的结论)。

4 问题的结论

我们利用“最邻近点插值法”及“反距离加权平均法”所得到的预报数据, 对问题一中预报方法预报降雨量的准确性进行相关分析, 得到如下结果:

4.1 最邻近点插值法

1) 两种预报方法绝对误差平方和的平方根分别为

$$\begin{aligned} \varepsilon_1 &= \sqrt{\sum_{t=1}^4 \sum_{d=1}^{41} \sum_{i=1}^{91} \varepsilon_1(i, d, t)^2} = 449.4401; \\ \varepsilon_2 &= \sqrt{\sum_{t=1}^4 \sum_{d=1}^{41} \sum_{i=1}^{91} \varepsilon_2(i, d, t)^2} = 474.2242. \end{aligned}$$

- 2) 两种预报方法绝对误差的均值分别为

$$E_1 = 0.0022, E_2 = 0.0024.$$

- 3) 两种预报方法绝对误差的方差分别为

$$D_1(t): 25.9718, 14.3807, 7.8352, 5.9273;$$

$$D_2(t): 27.9725, 16.6639, 9.7564, 5.8832.$$

通过上述二组数据的比较,可以看出在“最邻近点插值法”建立起的算法模型下,预报方法一的预报降雨量效果比预报方法二的准确度要好一些。从这些数据还可看出,两种预报方法的误差的均值都非常小,说明两种方法都比较可靠,但是方法一的误差均值更小,说明它的预报值更接近于观测值。另外,方法一的方差比方法二小,说明它的预报更“稳定”,误差的波动范围小,不会出现预报错误。

- 4) 两种预报方法在41天中四个时段的公众满意度及整体满意度分别为

$$Sa_1(t): 0.9718, 0.9717, 0.9704, 0.9627$$

$$Sa_2(t): 0.9716, 0.9715, 0.9698, 0.9627$$

$$Sa_1: 0.9692, Sa_2: 0.9689.$$

从这组数据可以看出,公众对两种预报方法的满意度都在95%以上,满意度都比较高。

4.2 反距离加权平均法

- 1) 两种预报方法绝对误差平方和的平方根分别为

$$\varepsilon_1 = \sqrt{\sum_{t=1}^4 \sum_{d=1}^{41} \sum_{i=1}^{91} \varepsilon_1(i, d, t)^2} = 647.1429;$$

$$\varepsilon_2 = \sqrt{\sum_{t=1}^4 \sum_{d=1}^{41} \sum_{i=1}^{91} \varepsilon_2(i, d, t)^2} = 664.3092.$$

- 2) 两种预报方法绝对误差的均值分别为

$$E_1 = 0.0012, E_2 = 0.0016.$$

- 3) 两种预报方法绝对误差的方差分别为

$$D_1(t): 36.4778, 65.4187, 11.5453, 5.4461;$$

$$D_2(t): 36.9732, 65.9509, 11.5886, 5.5904.$$

和“最邻近点插值法”相似,在“反距离加权平均法”建立的算法模型下,得到预报方法一的降雨量预报效果比预报方法二的准确度略好一些。

- 4) 两种预报方法在41天中四个时段的公众满意度及整体满意度分别为

$$Sa_1(t): 0.9660, 0.9658, 0.9636, 0.9574;$$

$$Sa_2(t): 0.9660, 0.9661, 0.9635, 0.9568;$$

$$Sa_1: 0.9632, Sa_2: 0.9631.$$

从这组数据可以看出,公众对两种预报方法的满意度都在95%以上,得到公众满意度都比较高的结论。通过两种模型对上述两种预测方法的有效性的研究,我们得到,这两种预测方法都可以接受,而预测方法1对问题一的降雨量的预报准确度及问题二公众的满意度都要好一些。

5 模型的评价

通过用“最邻近点插值法”及“反距离加权平均法”建立两种预报算法模型,对观测站的降雨量进行预报,并且用“预报绝对误差平方和的方根”、“预报绝对误差的期望”及“预报绝对误差的方差”三项指标对两种雨量预报准确性进行了评价。对于问题二,我们建立了公众满意度函数,并求出了两种预报方法四个时间段的预报的公众满意度,得到较满意的结论。

我们所用的两种算法各有特点。其中“最邻近点插值法”算法简单直观,计算量小,由于进行数据处理时用到的数据量小,只用了2491个预报点中的91个点的的历史数据,可能产生较大的误差。对于数据量大,对预测精度要求不是太高,预报值随时间变化比较快时,用这种快速算法不失为一种有意义的算法。“反距离加权平均法”基本避免了算法1的缺点,它充分考虑了观测站邻近点对观测站预报值的影响。同时,认为预报点对观测站预报值的影响随距离增大而减小。由于数据量大,计算速度比较慢,需要较多的计算时间。

虽然本题第一种算法,即“最邻近点插值法”,得出的结果较好,但这是具体问题的特殊性所造成的(因为从地理知识可知,被预报位置在为江苏、上海、浙江的沿海城市,这些位置夏季降雨量极不均匀,在这种情况下,最邻近点插值法效果较好)。通常算法2会得出比算法1更合理的结果,对于更一般的情况,算法2更有效。

参考文献:

- [1] 姜启源. 数学模型(第二版)[M]. 北京: 高等教育出版社, 1993,8
- [2] 赵静, 但琦. 数学建模与数学实验[M]. 北京: 高等教育出版社, 施普林格出版社, 2000,11
- [3] 盛骤等. 概率论与数理统计[M]. 北京: 高等教育出版社, 2001,12
- [4] 李丽, 王振领. MATLAB工程计算及应用[M]. 北京: 人民邮电出版社, 2001,9

The Evaluation Model of the Rainfall Forecast Method

WU Li-bing, LEI Zhong-bo, WANG Cui

Advisor: ZHOU Guo-peng

(Xianning College, HuBei Xianning 437005)

Abstract: Two rainfall forecast models, “nearest point interpolation method” and “anti-distance weighted average method”, were proposed in this paper. And the rainfall forecasts of the observation stations were given. Three indices were used to evaluate the given rainfall forecast methods. The satisfiable degree function was presented to evaluate to personnel satisfied degree. As a result, the satisfiable degree values of the given methods were also greater than 95%.

Keywords: nearest point interpolation method; anti-distance weighted average method; satisfiable degree function