第22卷 第7期
2005年12月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 22 No. 7
Dec. 2005

文章编号:1005-3085(2005)07-0133-04

雨量预报方法的评价

陈 赟，宋 杰，卞 超
指导教师：数模教练组
(南京信息工程大学 南京，210044)

编者按：该文在讨论两种预报方法的准确度时，采用了相对误差的概念，并对相对误差分母为零（雨量的实测值为零）的情形作了相应的处理。

摘 要：本文建立了科学评价雨量预报方法的数学模型，对所给网格点上的数据进行插值计算，得到两种方法的预报值，再结合题目提供的实测数据，并考虑公众的满意程度，通过建立相对误差模型，比较两种方法误差的大小来评价两种方法的准确度。

关键词：插值；相对误差；满意度.

分类号：AMS(2000) 00A71　　**中图分类号**：O241.1　　**文献标识码**：A

1 参数说明

A_{ijk} 表示第 i 天第 j 时段第 k 个站点实际测量的降雨量；
B_{ijk} 表示第 i 天第 j 时段第 k 个站点预报的降雨量；
u 实际观测到的降雨量的等级；
v 预报方法得到的预报值的等级；
$\mathrm{H}=(h_{uv})$ 由预测降雨量的等级与实测降雨量等级之间的差构造的满意度矩阵；
$\mathrm{H}=(f_{uv})$ 考虑公众感受，用权重修正之后的满意度矩阵。

2 模型的建立

2.1 问题一模型的建立

2.1.1 简单模型的建立

我们通过分析数据和查阅资料，这些预报的数据是位于特定点（经度，纬度等步长）上的值，即落在网格点上的。而我们需要的是每个观测站的预报值，这些站点坐标并不一定是正好落在网格点上，所以第一步工作就是通过插值（具体的插值算法在后文描述）计算出每个观测站的预报的降雨量，然后计算预报值和观测值的绝对误差：

$$S=\sum_{i=1}^{41}\sum_{j=1}^{4}\sum_{k=1}^{91}|A_{ijk}-B_{ijk}|. \tag{1}$$

2.1.2 改进模型

对题目深入分析后，发现上述误差并不能很客观地比较两种方法的好坏。比如，某一站点某天某一时段实际降雨量为10毫米，若方法一预报值为0毫米，两者相差10毫米。另一站点某天，某时段实际降雨110毫米，方法二预报值为130毫米，两者相差20毫米，从绝对误差看前者好于后者，可实际上方法二虽然差了20毫米，但这20毫米对于降雨110毫米来说，差的不多，而方法一一共才下了10毫米雨就误差了10毫米的雨量，从常识上判断显然方法一不如方法二好。可见绝对误差模型并不能很客观地评价两种预报方法的好坏，因此在计算逐点误差时引入

相对误差，建立改进模型。但考虑到分母上实测值可能为0，故将所有的点91 * 4 * 41个实测数据分成两类：为0与不为0。在不为0的点考虑相对误差，为0的点就考察绝对误差。这样考虑是合理的，因为对两种预报方法的实际观测值都分无雨（值为0）和有雨（值不为0）的情况分类讨论，对两种方法的衡量标准一样，不影响评价的准确性。故建立相对误差的模型

$$\begin{cases} \delta 1_j = \dfrac{1}{91}\displaystyle\sum_{k=1}^{91}\left(\dfrac{1}{41}\sum_{i=1}^{41}\dfrac{|A_{ijk}-B_{ijk}|}{A_{ijk}}\right) & A_{ijk}\neq 0 \\ \delta 0_j = \dfrac{1}{91}\displaystyle\sum_{k=1}^{91}\left(\dfrac{1}{41}\sum_{i=1}^{41}|A_{ijk}-B_{ijk}|\right) & A_{ijk}=0. \end{cases} \tag{2}$$

2.2 问题二模型的建立

为了将公众的感受作为一个评价因素引入评价体系，必须要将其量化，然后作为权重加入问题一建立的误差模型上。

2.2.1 权重值的选取

首先对降雨量的大小划分成 $0\sim6$ 个等级见表1：$D(x)$

表1: 降雨量的大小划分

降雨量 x	<0.1	0.1~2.5	2.6~6	6.1~12	12.1~25	25.1~60	>60.1
等级 D	0	1	2	3	4	5	6

例如，降雨量 $A_{ijk}=10$，则 $D(A_{ijk})=3$。

为了量化公众的满意程度，我们先定义满意度值，当预报雨量等级与实测雨量等级相一致时，公众的满意度最好，其值为1，当预报值雨量等级与实测雨量等级相差一个等级时，公众的满意度次之，其值为2，依次类推，得出公众对降雨预报的满意度矩阵 H

$$H=(h_{uv})=\begin{pmatrix} 1&2&3&4&5&6&7\\ 2&1&2&3&4&5&6\\ 3&2&1&2&3&4&5\\ 4&3&2&1&2&3&4\\ 5&4&3&2&1&2&3\\ 6&5&4&3&2&1&2\\ 7&6&5&4&3&2&1 \end{pmatrix} \tag{3}$$

其中 H 中的元素用 h_{uv} 表示，下标 $u=D(A_{ijk})$，$v=D(B_{ijk})$。即 u，v 分别对应实际观测到的降雨量的等级和预报方法得到的预报值的等级。

但考虑到预报值和实测值的降雨量等级差虽然一样，但不同情况公众的满意度不一定一样。例如预报雨量属于无雨，但实际降雨量为六等，两个量相差六等；而如果预报雨量为六等，实际无雨，两个量同样差六等。而前者对公众造成的困扰肯定比后者大，就好比有人听了天气预报下雨带了伞，而实际没下雨；另一种情况是预报不下雨，没带伞，结果下大雨。

为解决上述问题，我们考虑将对称的满意度矩阵的元素的值做改进。因为公众的感受可以认为是人的主观因素，我们采用层次分析法来寻求适当的值。满意度矩阵的列向量是按预报雨量的递增等级来排列的，用层次分析法来将预报量等级的满意度（作为准则层）对于实测值满意度影响（作为目标层）的权向量计算出来。具体采用特征根法[1]可求得权向量（H 矩阵的最大特征根对应的特征向量）：$(\alpha_1, \alpha_2, \alpha_3\ldots,\alpha_7)$，将所求的权重对应乘到满意度矩阵的下三角区域，就是将下三角区域中的 $(1, 2, 3\ldots,7)$ 对应换成 $(\alpha_1, 2\alpha_2, 3\alpha_3\ldots,7\alpha_7)=(0.45,0.75,0.99,1.27,1.66,2.24,3.13)$。得到修正后的满意度矩

阵 $H=(f_{uv})$，其中 f_{uv} 表示公众感受因素的影响，视为权重，进而建立以下模型。

2.2.2 模型的建立

我们根据每个站点每天每时段的实际值 A_{ijk}，和预报值 B_{ijk}，的雨量等级和得到每个站点每天每时段的误差的权重 f_{uv}，加入到问题一的误差模型中，就是问题二考虑公众满意度对两种模型预报评价影响的模型

$$\begin{cases} \delta 1_j = \dfrac{1}{91}\sum\limits_{k=1}^{91}\left(\dfrac{1}{41}\sum\limits_{i=1}^{41}\dfrac{|A_{ijk}-B_{ijk}|f_{uv}}{A_{ijk}}\right) & A_{ijk}\neq 0 \\ \delta 0_j = \dfrac{1}{91}\sum\limits_{k=1}^{91}\left(\dfrac{1}{41}\sum\limits_{i=1}^{41}|A_{ijk}-B_{ijk}|f_{uv}\right) & A_{ijk}=0. \end{cases} \tag{4}$$

此公式算出了对于41天的同一时刻中91个观测点的误差的平均值。

3 模型的求解及结果分析

3.1 预报值的计算

计算每个观测站的预报雨量值的插值算法如下：

设题目所给的预报值数据是位于坐标 (X,Y) 处的数据。X 表示经度（题目数据包文件 lon 中的值），Y 表示纬度（文件 lat 中的值）。观测站坐标设为 (u,v)。

Step1. 固定 X 的值为 X_0，按不同的 Y 寻找 (u,v) 所在的区间或者位置，并将所在区间的上限或者是正好的位置所在的点 (X_0,Y_k)，向前推三个点即 $(X_{0-i}，Y_k)$ $(i=1,2,3)$ 和向后增加四个点 $(X_{0+i}，Y_k)$ $(i=1,2,3,4)$，这样就形成了一共八个点，但是由于经、纬度的增量不等，所以往往每行（不同的 Y_k）中 u 的位置并不完全一样可能选到的点的个数少于八个（这对插值结果影响不大），我们记下所有行的共同区间。

Step2. 在 Y 中每一列上寻找 v 所在的区间或者位置，同样的方法，我们记下所有列 (X_l) 的共同的区间。

Step3. 先对每个 X 所选的八个点（可能更少的点）进行拉格朗日插值，将数据插值到相同的 Y（纬度）上。这样就形成了同一纬度的八个点（或者更少的点）的值。

Step4. 在上述所得的八个值（或者更少）进行同样的拉格朗日插值。因为不能确定 v 所在纬度对应的经度情况，这里假定了经度的分布和所选纬度上八个点（或者更少）的中间位置的经度分布情况相对应。

Step5. 经过两次拉格朗日插值就可以得出所插的一个站点的数据。

step6. 重复上面的1-5过程，就可以得出所有站点的值。

3.2 问题一的误差值的计算及结果分析

通过计算机程序实现公式(2)，得到的结果如表2、表3所示：

表2: 第一种预报误差

Mssall					Mssall0				
1_dis1	2_dis1	3_dis1	4_dis1	avg	1_dis1	2_dis1	3_dis1	4_dis1	avg
0.8095	0.8166	0.6518	0.4653	0.6858	0.05	0.0610	0.0837	0.0762	0.067

表3: 第二种预报误差

Mssall					Mssall0				
1_dis1	2_dis1	3_dis1	4_dis1	avg	1_dis1	2_dis1	3_dis1	4_dis1	avg
0.8112	0.8296	0.6576	0.4956	0.6986	0.0518	0.0634	0.0841	0.0797	0.06975

其中表2表3中的 i_dis1，i_dis2 分别为第一种预报方法的第 i 个时段和第二种方法的第 i 个时段 ($i = 1,2,3,4$) 的误差；Msssall 所含四个值为4个时段所有实测值不等于0的点总体的相对误差值，Msssall0 所含四个值为四个时段所有实测值等于0的点总体的绝对误差值。avg 对应4个时段误差的平均值。由表2可以看出在用第一种预报方法预报的降雨量在第一时段时的误差为0.80954，相应的在表3里可以得到用第二种预报方法预报的在同一时段的降雨的误差为0.8112，因为误差越小越好，所以方法一是比较准的预报方法。同理，在其它时段方法一都比方法二好。与此同时，在表2和表3里，误差对时段的平均值也是方法一小于方法二的，即 $0.68581 < 0.6986$。同样可分析各观测站观测数据为零时的两种预报方法的预报值的准确程度见表2表3 Msssall0 所对应的数据。所以对问题一不管是哪个时段或实际降雨量是否为0，第一种预报方法都是最好的雨量预测方法。

3.3 问题二的误差值的计算及结果分析

通过简单的计算机程序我们可以按公式(4)计算出问题二考虑了公众的满意度以后的修正评价模型，得到结果见表4：

表4: 问题2的误差值

	方法一的误差	方法二的误差
第一时间段	0.2840	0.2765
第二时间段	0.2312	0.2623
第三时间段	0.2604	0.2674
第四时间段	0.2698	0.2641
平均值	0.2614	0.2676

分析结果我们发现不管使用方法一还是方法二，两者都所产生的误差相差不大，比较两者的误差值，第二种预报方法的在第一个时段和第四个时段产生的误差比方法一小，说明在这两个时段方法二预报的效果更另公众满意，但是在二，三两时段比方法一的误差大（见表第二行和第三行），说明方法一在这两个时段预报效果令公众比较满意，因此仅从表面看不出是哪个方法好，于是列出了平均值，以此来衡量最终是哪个方法误差小，从总体上令公众感到满意的次数多。由表4最后一行可见方法一的平均误差小，说明总体上方法一的预报效果能更令公众满意。

参考文献:

[1] 姜启源. 数学模型[M]. 北京：高等教育出版社，1992
[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,l9(2):178-185

Evaluation of Rainfall Forecast Methods

CHEN Yun, SONG Jie, BIAN Chao
(Nanjing University of Information Science and Technology, Nanjing 210044)

Abstract: In this paper we has made mathematical model to evaluate the rainfall forecast methods scientifically. According to given points of grid we take interpolation and get the prediction value of two kinds of methods. Combining the data of surveying that the question offers and considering public satisfying degree, we set up relative error model to compare the value of two kinds of method errors, and estimate the accuracy of two kinds of methods.

Keywords: interpolation; relative error; degree of satisfaction