

文章编号:1005-3085(2005)07-0139-04

## 基于插值的雨量预报评价模型

谭永基, 蔡志杰

(复旦大学数学科学学院, 上海 200433)

**摘 要:** 本文讨论了雨量预报方法的评价问题, 给出了散乱数据拟合的若干方法及误差确定方法, 同时在顾及公众反应的情形下考虑了评价准则, 最后针对评阅中发现的一些问题作了评述。

**关键词:** 散乱数据插值; 误差分析

**分类号:** AMS(2000) 41A05

**中图分类号:** O174.42

**文献标识码:** A

## 1 问题的提出

今年全国大学生数学建模竞赛 C 题来源于气象部门的一个实际问题。天气预报与人民群众的生活密切相关, 如何提高预报的准确性是气象部门非常关注的问题。本题考虑的是对预报方法的评价问题。气象部门研究6小时雨量的两种预报方法, 预报位置位于一个等距网格的  $53 \times 47$  个网格点上。同时在91个分布不均匀的观测站点实测各时段的实际雨量。如何评价两种预报方法的准确性?

## 2 插值方法

要评价预报方法的准确性, 必须对同一位置上的预报值和实测值进行比较, 计算它们之间的误差大小。然而, 由于条件的限制, 我们得到的预报数据和实测值并不处于同一位置, 这就需要根据已知信息推算出其它位置的信息。有两种方法: 由实测站点的实值推算出预报网格点上的实测值, 或者由预报网格点上的预报值推算出实测站点处的预报值。

若采用第一种方案, 由于观测站点的分布是散乱的, 应采用散乱数据插值方法。散乱数据插值方法很多, 主要有 Shepard 插值和径向基插值。

设  $P_k(x_k, y_k)$  ( $k = 1, 2, \dots, N$ ) 为观测站点的坐标,  $f_k$  为  $P_k$  处的实测值,  $Q(x, y)$  为某个预报网格点。为了由  $P_k$  ( $k = 1, 2, \dots, N$ ) 处的实测数据推算出  $Q$  点的实测数据, 容易想到的原则是, 距离越近的对  $Q$  的影响越大, 距离越远的点对  $Q$  的影响越小。一个典型的方法是“反距离加权平均”, 即 Shepard 方法。

记  $r_k = \sqrt{(x - x_k)^2 + (y - y_k)^2}$  为  $P_k$  到  $Q$  点的距离, 则定义插值函数

$$f(x, y) = \begin{cases} f_k, & \text{当 } r_k = 0 \text{ 时,} \\ \sum_{k=1}^N \frac{f_k}{r_k^2} / \sum_{k=1}^N \frac{1}{r_k^2}, & \text{否则.} \end{cases} \quad (1)$$

于是由散乱点  $P_k$  处的实测值可以得到任一点处的实测值。

另一种常用的方法是径向基插值。例如用多二次径向基插值, 设插值函数为

$$f(x, y) = \sum_{k=1}^N \alpha_k (r_k^2 + c)^{-\frac{1}{2}}, \quad (2)$$

其中  $\alpha_k$  为待定系数,  $c$  为常数, 可取为  $c = 1$ 。根据插值条件

$$f(x_k, y_k) = f_k \quad (k = 1, 2, \dots, N), \quad (3)$$

求解关于  $\alpha_k$  的线性代数方程组, 可得到系数  $\alpha_k$  的值。

若采用第二种方案, 上述两种方法仍然有效, 也有参赛队采用双线性插值等其他插值方法进行计算。具体情况我们在第5节中加以说明。

### 3 误差分析

记  $Q_{ij}$  ( $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ ) 为预报网格点,  $g_{ij}$  为  $Q_{ij}$  处的预报值。利用插值方法, 得到  $Q_{ij}$  上的实测值, 记为  $f_{ij}$ , 这样可定义两者之间的平方误差为

$$E_{ij} = (f_{ij} - g_{ij})^2. \quad (4)$$

考虑到不同的实测雨量的预报误差对公众的感受是不同的, 应采用相对平方误差的概念, 即定义

$$ER_{ij} = \frac{(f_{ij} - g_{ij})^2}{f_{ij}^2}. \quad (5)$$

当  $f_{ij} = 0$ , 即  $Q_{ij}$  处无雨时, (5)式的分母为零, 必须对这一情形作相应的处理。一种处理方式是定义

$$ER_{ij} = \begin{cases} \min \left\{ 1, \frac{(f_{ij} - g_{ij})^2}{f_{ij}^2} \right\}, & \text{当 } f_{ij} \neq 0 \text{ 时,} \\ 1, & \text{当 } f_{ij} = 0, g_{ij} \neq 0 \text{ 时,} \\ 0, & \text{当 } f_{ij} = g_{ij} = 0 \text{ 时.} \end{cases} \quad (6)$$

也可以有其他的处理方法, 例如

$$ER_{ij} = \frac{(f_{ij} - g_{ij})^2}{1 + f_{ij}^2}. \quad (7)$$

在此基础上定义相对均方误差

$$MER = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n ER_{ij}. \quad (8)$$

也可用连续区域中实测数据与预报数据之间的误差值。用(1)或(2)式等插值方法分别对实测数据和预报数据进行插值, 得到整个区域(记为  $\Omega$ ) 内实测值和预报雨量的分布函数, 分别记为  $f(x, y)$  和  $g(x, y)$ , 相对均方误差可定义为

$$MER = \frac{\iint_{\Omega} (f(x, y) - g(x, y))^2 dx dy}{\iint_{\Omega} f^2(x, y) dx dy}. \quad (9)$$

对问题提供的41天共164个时段的数据分别求出相对均方误差的平均值和标准差, 作为评价预报方法优劣的依据。

## 4 考虑公众感受的模型

在分级模型中, 将不同等级的雨量与数值建立一一对应关系, 如用 0 ~ 6 分别对应于无雨、小雨、中雨、大雨、暴雨、大暴雨和特大暴雨七个等级。然后建立公众不满意度函数。最简单的定义方法是

$$S = \sum_{i=1}^m \sum_{j=1}^n |d_{ij} - e_{ij}|, \quad (10)$$

其中  $d_{ij}$  表示预报网格点处预报雨量的等级,  $e_{ij}$  表示网格点处实测雨量的等级。(10)式反映了实测等级与预报等级相差越大, 公众的不满意程度也越大这一事实。但是这个定义较为粗糙, 没有考虑到不同等级的雨量的误报及不同时段的影响对公众的不同影响。

一种处理方法是, 对不同等级的实测雨量, 定义公众不满意度函数为

$$S = \sum_{i=1}^m \sum_{j=1}^n \alpha_{e_{ij}} |d_{ij} - e_{ij}|, \quad (11)$$

其中  $\alpha_k$  表示第  $k$  等级的雨量的权系数, 当然也可取权系数为  $\alpha_{kl}$ , 表示权系数与预报等级和实测等级均有关系。然后对一天中4个时段的不满意度再加权求和, 得到

$$S_d = \sum_{k=1}^4 \beta_k S_k, \quad (12)$$

其中  $\beta_k$  表示第  $k$  个时段的权系数,  $S_k$  为由(11)式计算得到的第  $k$  个时段的公众不满意度函数,  $S_d$  表示公众对一天中各时段预报的总体不满意度。这样, 根据41天各时段的数据分别求出两种预报方法41天的总体不满意度的平均值和标准差, 作为评价依据。

## 5 评阅中发现的问题

在今年的评阅过程中, 我们发现同学们对问题一般都能很好地理解, 查阅了很多有关的书籍, 对数学软件(如 MATLAB 等)也能很好地使用。但也发现一些问题。

1) 对数据的处理问题, 很多参赛队未能采用插值或拟合的方法, 而简单地采用了最邻近点法。预报网格点及实测点之间不可能离得很近, 一般都有很大的距离, 这是一个常识。因此, 最邻近点法会产生很大的误差。

有些参赛队采用了简单平均的方法, 即某点的预报雨量用围绕其周围的4个预报网格点上的预报值作平均, 由于没有考虑实测点与预报点之间距离的影响, 同样会产生较大的误差。

还有些参赛队采用了多项式插值, 如双线性插值, 理论上来说, 这是一种可以接受的方法, 但它过于局部化, 实测点上的预报值仅依赖于周围4个预报网格点上的值。一种解决方法是扩展区域, 用周围16个或更多点作多项式拟合。

2) 在数学软件的应用中, 绝大多数参赛队都采用了 MATLAB 软件, 并使用了其中的 griddata 函数来作数据插值。但只有少数参赛队说明了拟合的方法, 并对 griddata 函数中的插值方法加以说明。事实上, griddata 函数有四种插值方法可供选择, 缺省是线性插值。大多数参赛队在没有弄清方法的情况下, 盲目使用数学软件, 这是一种不好的倾向, 应加以克服。

3) 很多参赛队在处理数据误差时, 采用了绝对误差。对本题来说, 最终结果与采用相对误差是相同的, 但从问题出发, 这不是很好。在工程问题中, 涉及误差的问题, 通常都采用相

对误差。试想, 实际降雨量为0.01 mm, 而预报值为0.02 mm; 与实际降雨量为10.01 mm, 预报值为10.02 mm, 它们的误差带给公众的感受难道是一样的吗?

4) 在考虑等级问题中, 只有极少数参赛队考虑了不同等级的降雨量的误报对公众的不同影响, 而考虑不同时段对公众的影响的参赛队就更少了。有一个参赛队考虑用层次分析法进行处理, 这是很好的想法, 可惜未能深入下去, 给出相应的矩阵及计算结果, 给人以虎头蛇尾的感觉。

#### 参考文献:

- [1] 叶其孝主编. 大学生数学建模竞赛辅导教材(二)[M]. 长沙: 湖南教育出版社, 1997
- [2] 谭永基, 曾舍荣, 王健. 用MQ方法数值求解偏微分方程[J]. 数学的实践与认识, 1997,27(4):327-334
- [3] Hardy R L. Multiquadric equations of topography and other irregular surfaces[J]. Journal of Geophysical Research, 1971,76(8):1905-1915

## Rain Quantity Forecasting Evaluation Model Based on Interpolation

TAN Yong-ji, CAI Zhi-jie

( School of Mathematical Sciences, Fudan University, Shanghai 200433 )

**Abstract:** In this paper, we study the problem of evaluating different methods of rain quantity forecasting. We describe some interpolate methods for scattering data and the method to estimate the error. We set up an evaluating criterion to consider the public reaction on the forecasting. We also made some comments on the contest articles.

**Keywords:** interpolation for scattering data; error analysis