

气象降水分布图制作中的插值算法研究

张红杰¹ 马清云² 吴焕萍² 罗 兵² 唐 卫²

(1. 中国矿业大学(北京校区), 北京 100083; 2. 国家气象中心)

提 要: 针对气象服务中气象要素色斑图制作要求准确而美观的特点, 以及现有插值算法不能满足该需求的现状, 文章在分析了各类常用插值算法的基础上, 借鉴人工天气图的思路提出了一种基于 Cressman 算法的优化算法, 并对优化插值算法的关键技术问题进行了讨论, 最后通过实际应用展示了这种算法的应用效果。实际结果表明应用该算法所得制图效果美观, 增强了降水分布图的准确性和客观性, 满足业务应用的要求与系统建设的需要。

关键词: 降水分布图 插值算法 Cressman 算法

Research on the Interpolation Algorithm for Meteorological Precipitation Choroplethic Map

Zhang Hongjie¹ Ma Qingyun² Wu Huanping² Luo Bing² Tang Wei²

(1. China University of Mining & Technology, Beijing 100083;
2. National Meteorological Center, CMA)

Abstract: After comparing various interpolation methods and analysis of the practical requirements in meteorological services, an interpolation method is proposed to meet the requirements of both more accurate and higher quality visualization of meteorological elements choroplethic map. The method is based on the basic Cressman algorithm, and optimized by some key issues. The applications demonstrate that it works well in the meteorological service system, and improves the accuracy and objectivity of precipitation choroplethic map.

Key Words: precipitation choroplethic map interpolation methods Cressman algorithm

资助项目: 科技部项目十一五支撑(2006BAK01A29; 2006BAK01A18; 2006BAD04B10)、气象局新技术推广项目(CMATG2009MS33)资助。

收稿日期: 2008年1月9日; 修定稿日期: 2009年8月20日

引 言

气象要素的空间分布图,一般也称为色斑图^[1],要求精确性(如:避免丢值)、视觉美观性(如避免常规算法所带来的大量的碎部产生等问题),插值算法的选择直接影响到色斑图的表现质量。目前各类插值算法较多,如:克里格(Kriging)插值法、加权反距离法(IDW)、Delaunay 三角剖分线性内插、样条函数插值和 Cressman 客观分析等。

然而,一般的插值算法通常很难同时满足精确性和美观性要求,笔者在对常用的插值算法进行充分分析的基础上,借鉴人工天气图分析思路,采用逐级插值的分析方法,对逐步订正方法(Cressman 算法)进行了改进与优化,并提出了一种新的称之为 MCressman 的插值方法,从而解决了一些现有插值算法中的关键技术问题。

1 算法分析

1.1 常用插值算法分析

数据的空间插值方法很多,有多种分类方法。黄杏元等^[2]依据已知点和已知分区数据的不同,可将空间数据插值分为点的内插

和区域的内插,邬伦等^[3]则对应分为空间内插和外推。影响区域空间插值精度的主要因素就是空间插值方法。目前不同领域内较为常用的插值方法有克里格(Kriging)插值法、加权反距离法(IDW)、Delaunay 三角剖分线性内插、样条函数插值和 Cressman 客观分析等,这些方法在 IDL、MATLAB、Surfer、Arc/info 和 ArcMap 等数据分析处理软件以及气象数据客观分析研究中经常被采用。文献^[4-5]对克里格(Kriging)插值法、加权反距离法、Delaunay 三角剖分线性内插、样条函数插值和 Cressman 客观分析 5 种不同的空间内插方法进行比较研究,选用了中国区域内基本台站 10 年(1988 年 12 月至 1998 年 11 月)的月降水量作为 5 种插值算法的试验数据。不同方法的插值结果反插到原始台站后的对比分析结果如表 1 所示^[5]。结果表明,Cressman 客观分析法与原始台站最为接近,反距离加权和双谐样条方法次之,而克里格与三角化线性方法则与原始台站偏离最大。Cressman 客观分析方法是一种将离散点内插到规则格点引起误差较小的一种逐步订正的内插方法,因此被广泛应用于气象领域各种诊断分析和数值预报方案的客观分析中,因而 Cressman 方法第一次使客观分析成为了一门独立的科学^[6]。

表 1 插值方法比较

单位:mm

统计量	原始台站	克里格	反距离加权	三角化线性内插	双谐样条	Cressman
平均值	869.0	869.3	869.9	869.3	869.1	868.9
标准差	530.7	523.6	527.6	521.2	528.1	530.7
最大值	2576.9	2448.6	2555.7	2402.3	2529.2	2576.9
最小值	16.9	30.3	18.2	30.4	22.9	16.9

1.2 Cressman 客观分析方法

Cressman 客观分析方法是由 Cressman^[7]等在 1959 年提出的,其基本原理是先给定第一猜测场,然后用实际观测场逐步修

正第一猜测场,直到订正后的场逼近观测记录为止。

$$\alpha' = \alpha_0 + \Delta\alpha_{ij} \quad (1)$$

其中

$$\Delta\alpha_{ij} = \frac{\sum_{k=1}^K (W_{ijk}^2 \Delta\alpha_k)}{\sum_{k=1}^K W_{ijk}} \quad (2)$$

式中, α_k 为任意一气象要素(如:降水、风、温度等); α_0 是变量 α 在格点 (i, j) 上的第一猜测; α' 是变量 α_k 在格点 (i, j) 上的订正值; $\Delta\alpha_k$ 是观测点 k 上的观测值与第一猜测值之差; W_{ijk} 是权重因子, 分别由香蕉形、椭圆形和圆形权重函数决定, 在 $0.0 \sim 1.0$ 之间; K 是影响半径 R 内的台站数。Cressman 客观分析方法最重要的是权重函数 W_{ijk} 的确定, 它的一般形式为

$$W_{ijk} = \begin{cases} \frac{R^2 - d_{ijk}^2}{R^2 + d_{ijk}^2}, & \text{当 } d_{ijk} < R \\ 0, & \text{当 } d_{ijk} \geq R \end{cases} \quad (3)$$

其一般取为 1、2、4、7 和 10 几个常数^[5]。 d_{ijk} 是格点 (i, j) 到观测点 k 的距离。

2 基于关键值的 Cressman 方法

在气象降水分布图制作过程中, 通常按 24 小时降水量(mm)的大小将其分为: 小雨(1~9.9)、中雨(10~24.9)、大雨(25~49.9)、暴雨(50~99.9)、大暴雨(100~250)、特大暴雨(>250)这六个级别。尽管 Cressman 在气象方面应用最为广泛, 最为贴近实际情况, 但在降水插值中存在的最明显的问题是: 如果在实况观测站数据(即离散点数据)中只有一个站点而其降水量恰好处于降水分级值附近, 那么常规 Cressman 算法则会在其逐步订正过程中

将该重要的站点值因内插订正而丢失。显然, 这一至关重要的降水量信息在决策服务产品的制作过程中是不能丢失的。

另外, 当台站资料分布极为不均时, 如在一大片区域内如果只有一个较小集中区域有降水时(本文称为孤点), 采用 Cressman 插值后该点会出现圆形插值区域, 且范围过大, 这样从视觉上会显得有些单调, 不够自然。而且插值的边界会确定不合理——即有时会过大或应该有的地方没填上, 有不知名碎块, 区域性表现不强。

针对以上问题, 为了满足气象决策服务的需求, 本文在保留了传统 Cressman 优势的基础上对其进行了优化, 提出了基于关键值的 Cressman 插值算法(以下称为 MCressman 算法), 该方法既解决了传统 Cressman 算法在插值中存在的问题, 又满足了精确性和美观性的要求。

MCressman 算法的基本思想如下: 借鉴人工天气图分析思路, 采用逐级插值的分析方法, 在分析某一级时, 只选择该级以上的观测点资料参与分析, 以保证该级别不被平滑掉; 并在 Cressman 算法中加入对不同的分级分配不同的权重, 即在各自级别权重的基础上根据级别的不同加重或减轻权重随影响距离的增加而衰减的程度; 以表现同一值在不同级别中的层次感, 并对分析得到的格点值进行光滑是图形色块边界平滑、美观。同时为了保证插值边界的准确性, 提出了离散点边界的提取、数据插补技术、格点光滑等技术。其主要原理如图 1 所示。



图 1 改进算法的处理流程示意

3 关键技术

MCressman 需要解决的几个关键技术

包括: 边界搜索算法、数据插补、格点光滑。

3.1 边界搜索算法

为有效保证插值的边界在一定的合理范

围内,本文对离散点的边界作了如下处理:

(1) 对所有离散点进行缓冲区(buffer)分析,形成一系列圆形多边形区域,即找出每一离散点与其相邻的其他离散点,距离的选取可以参照数据的分布情况,如在全国范围内距离的取值为 10km。

(2) 对每个离散点形成的缓冲区逐一求合并(union)操作,最终将形成若干个新的不规则多边形。对每一个参与合并过的多边形给出标记,记 flag 为 1。

(3) 在(2)中如果多边形的 flag 为 0 即表示在一定范围内,它不与别的多边形区域相交,也表明了生成该圆形多边形的原离散

点在一定的距离内周围没有其他离散点。因此,为了色斑图的美观,以该点为中心做一个较小的椭圆,以避免插值后形成规则的圆形区域。

(4) 对所有合并处理后的多边形求其凸包,并对该凸包的边界进行光滑处理,在本文中采用了贝塞尔曲线算法进行处理。

(5) 对光滑后的所有多边形再进行一遍循环求交,以确保由于求凸包过程中两个距离很近的多边形有相交的情况。

经过以上处理,就完成了离散点所形成的边界提取。其流程图如图 2。

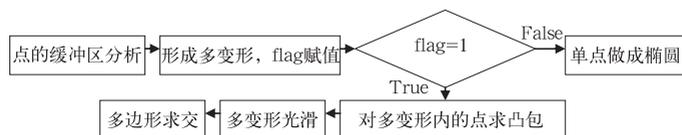


图 2 边界搜索流程示意

3.2 数据插补

经过边界搜索,此时边界已经确定。为了确定无效值是无降水资料还是遗漏降水资料。本文站点数据做了如下处理。因为边界内的站点的值并非都是有效值。非有效值(或无效值)并不意味着这个站点的降水量为零,而是表示该站点的降水量未被记录下来或其他原因不能被应用。所谓数据插补是指对一定区域内的气象站点的空缺的值和无效值的处理。数据补差的基本原理如下:

首先在一个多边形(polygon)中查找无效值并添加到无效值(novalue)点集中,同时把有效值点放到有效值(value)点集中;

其次确定缓冲区的距离 d ,这里根据最大最小距离的方法来设定。

(1) 计算出有效值点集(value)中的第 i 个点到无效值点集(novalue)中其他测站之间距离中的最小值 d_i 。

(2) 取所有的 d_i 中最大的距离作为缓冲区距离 d 。

最后,根据缓冲区距离 d ,对多边形(polygon)中的无效值点集(novalue)的所有站点做缓冲区分析,此缓冲区分析是为了确定插值的边界问题,确保插值范围的准确性。计算落入该缓冲区内的有效值点集(value)的测站数目。如果数目大于或等于 1,就认为该测站降水量是漏记,用落入该缓冲区中有效值点集(value)的平均值代替,反之,则认为该站点无降水。

3.3 核心算法

MCressman 的核心算法的主要步骤如下:

(1) 获得分级的个数(n)。对于降水来说就是分为小雨(1~9.9mm)、中雨(10~24.9mm)、大雨(25~49.9mm)、暴雨(50~99.9mm)大暴雨(100~250mm)、特大暴雨

(>250mm)这六个级别,各级值分别为 1, 10, 25, 50, 100, 250。

(2) 计算权重 W_{ijk} 。用 Cressman 方法计算出权重 W_{ijk} ,这是 MCressman 对 Cressman 进行优化关键步骤。MCressman 要对不同的级别按级别的大小给以限制,即级别越高其权重随着站点与分析格点距离的增加衰减越快。

$$W_{ijk} := W_{ijk}^{(2+a)(6+2a)} \quad (4)$$

其中 a 表示的是分级的序号,且 $a=1, 2, \dots, n$ 。“:=”表示替代。这种变化呈渐进趋势,能够更接近于真实的状态。实际意义就是同一降水量在图中可以表现出不同级别的范围和层次。

(3) “孤点”权重的处理。分析格点(i, j)时,在最大步长范围内找出参与插值的“孤点”,以便单调处理。即在 5 个步长的资料扫描中,若最小步长中已有资料,随着步长的增大,资料数没变即认为是“孤点”,如果是孤点该点的权重在(2)的基础上进一步降低。

$$W_{ijk} := W_{ijk}^2 \quad (5)$$

这样就会使孤点的范围进一步缩小。使插值出现良好的视觉效果。

(4) 得到分级的格点分析场。把上面的权重参与到 Cressman 的插值计算中,计算出第 m 级别的相应格点的值 G_m 。

(5) 获得综合分析场。把所有级别的分析场合成一个综合分析场。由于 MCressman 是分级的,而且是分析低级别时就会包含有高级别的数据,所以在进行格点值的合并时一定要做比较,并遵循格点值大覆盖格点值小的原则。

如果($G[j] < G_m[j]$)则 $G[j] = G_m[j]$ 。其中 $G[j]$ 表示总的格点的值, $G_m[j]$ 则表示分级的格点的值。

其流程见图 3。

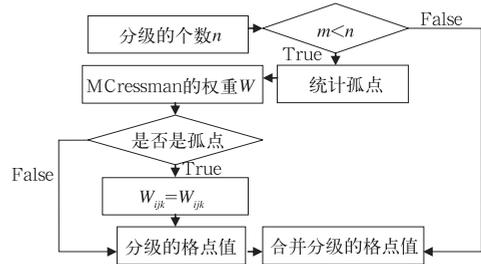


图 3 插值过程流程示意

3.4 格点光滑

格点的光滑主要采用的是九点光滑,其基本原理是对任意一格点周围的八个点和该格点本身的值求均值,再将其值赋给该格点,起到一定的滤波作用,达到图形边界光滑、美观的效果。同时,为了保证数据的临界值不会丢失,对格点值进行了分级光滑,其原理是对任意一级别的任一格点周围的八个点进行判断,如果判断结果存在无效值,就将该格点的值赋予无效值点,然后再将九个点的均值赋予该格点。

4 应用实验

实践中,利用 VC6.0 实现了 MCressman 算法,并与传统 Cressman 算法进行了对比实验分析。实验选用了 2007 年 10 月 12 日 08 时至 2007 年 10 月 13 日 08 时的 24 小时降水量实况观测数据,观测站(离散点)个数为 430 个。其中在实况观测数据中,重庆出现了一个观测值为大雨量级(值为 25mm)的站点,而在云南省与海南省分别出现了一个观测值为暴雨量级(50mm)的站点。Cressman 算法和 MCressman 算法采用的步长均为 1、2、4、7 和 10。图 4 为 Cressman 算法插值生成的色斑图,图 5 则为 MCressman 插值算法生成的色斑图。从图 4、5 对比分析,没有改进的 Cressman 插值过程损失了云南省与海南省的暴雨级别的站点

以及重庆的大雨量级的站点,而且内蒙古的大雨量级的站点在插值后的分布图呈圆形,效果不理想。青海省西南部、西藏的东南部还有碎块出现,观测文件中这些碎块所在区域的站点值都是小于1的。而 MCressman 插值分析结果形成的分布图,不仅保证了数据的准确性,也达到了良好的图形质量与视觉效果,很好地弥补了 Cressman 算法的不足。

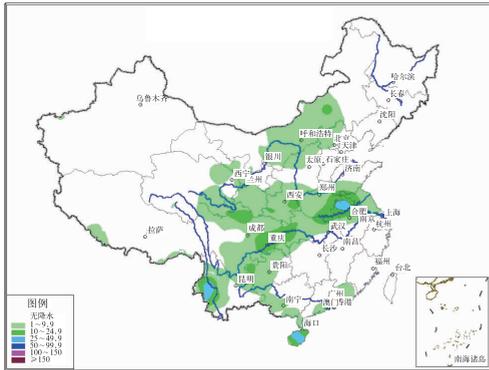


图 4 Cressman 的插值结果

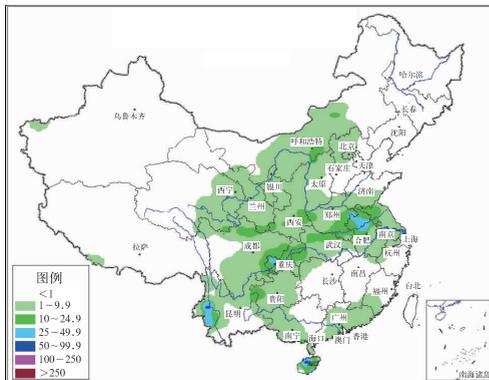


图 5 MCressman 的插值结果

5 结论

本文在分析各类常用插值算法的基础上

提出了一种基于 Cressman 方法的优化算法,并对改进插值算法的关键技术问题进行了详细讨论。实验表明,MCressman 算法有效地保证了气象降水色斑图的精确性和美观性的要求。在业务系统中得到了很好的应用,从而使每天的降水量实况图的生成体现了完全的自动化业务运行能力,大大地节省了人力,比常规人工分析的雨量图所含信息更多,减少了人为随意性,增强了插值的精确性和客观性。该算法虽然是针对降水描述的,但也同样适用于其他气象要素的分布图的分析制作。

参考文献

- [1] 吴焕萍,罗兵,王维国,等. GIS 支持下的决策气象服务系统建设[J]. 应用气象学报, 2008, 19(3): 380-383.
- [2] 黄杏元,马劲松,汤勤. 地理信息系统概论[M]. 北京:高等教育出版社, 2001, 93-103.
- [3] 郭伦,刘瑜,张晶,等. 地理信息系统原理方法和应用[M]. 北京:科学出版社, 2001, 178-192.
- [4] 赵天保,艾丽坤,冯锦明. NCEP 再分析资料和中国站点观测资料的分析与比较[J]. 气候与环境研究, 2004, 6: 278-294.
- [5] 冯锦明,赵天保,张英娟. 基于台站降水资料对不同空间内插方法的比较[J]. 气候与环境研究, 2004, 6: 261-277.
- [6] 王跃山. 客观分析和四维同化——(II)客观分析的主要方法(1)[J]. 气象科技, 2001, 1: 1-9.
- [7] GEORGE P. CRESSMAN. AN OPERATIONAL OBJECTIVE ANALYSIS SYSTEM [J]. Monthly Weather Review: 1959. 87(1): 367-374.
- [8] 朱海燕. GIS 空间分析方法在热带气旋研究中的应用[D]. 华东师大硕士学位论文, 2005, 1-54.
- [9] 张立祥,陈力强,周小珊,等. 中尺度数值预报中不同初值方案的检验对比[J]. 气象, 2001, 27(7): 8-12.
- [10] 周青,赵凤生,高文华. 中国实测地表温度和地面气温对比分析[J]. 气象, 2008, 34(2): 83-91.