

几种建立 DEM 模型插值方法精度的交叉验证

顾春雷^{1,2}, 杨 漾^{1,3}, 朱志春²

(1. 中国科学技术大学 地球和空间科学院,安徽 合肥 230026; 2. 安徽省地震局,安徽 合肥 230031;
3. 海南省地震局,海南 海口 570203)

摘 要:建立 DEM 模型时需要离散的高程点进行空间插值,实现这一目的的插值方法有很多种。文章主要选择了 6 种常用的插值方法,分别在平原、丘陵和高山几种不同复杂程度的环境下对其插值的精度使用交叉验证方法进行评估。结果表明,同样的插值方法对不同复杂程度的地形效果是不一致的,多种插值方法中以克里金插值方法的插值适用性最强,精度最高。

关键词:DEM 模型; 克里金插值; 交叉验证; 均方根预测误差

中图分类号:P208 文献标识码:B 文章编号:1672 - 5867(2011)05 - 0099 - 04

Accuracy Cross - validation of Several Interpolation Methods of DEM

GU Chun - lei^{1,2}, YANG Yang^{1,3}, ZHU Zhi - chun²

(1. College of Earth and Space Science, University of Science and Technology of China, Hefei 230026, China;
2. Earthquake Administration of Anhui Province, Hefei 230031, China;
3. Earthquake Administration of Hainan Province, Haikou 570203, China)

Abstract: There are many methods can be used for spatial interpolating of discrete elevation points when build DEM model. This paper chose six kinds of interpolating methods applied in plain, hilly and mountain region, and applied cross - validation method to evaluate the interpolating accuracy. The result shows, the effectiveness of same interpolating methods are different at different complex degree of terrain, the precision interpolation of Kriging interpolation method has strongest applicability and highest accuracy of the several interpolation methods.

Key words: DEM model; Kriging interpolation method; cross - validation; Root - Mean - Square error

0 引 言

数字高程模型(DEM) 是针对地形地貌的一种数字模拟,其精度一直是使用和研究人员的特别关注和重要的研究议题。其主要误差分为离散点本身的观测误差和离散点数学模型逼近误差,离散点自身误差来源包含了采样点的分布、密度和采样点的观测误差,对于数学模型逼近方面与选择的函数、模型方法有关。多年来众多国内外学者在该方面开展了许多研究,先后使用了试验方法、傅里叶方法、区域变异理论等统计学分析研究 DEM 模型精度^[1-3]。在数学模型逼近方面朱长青等使用重构等高线的方法、任志峰等使用 Strahler 积分方法对 DEM 模型精度进行评判^[4,5]。DEM 模型反映的是一定区域地表变化理论模型,自然地表是没有确定性规律的,而理论模型则有,因此同样的数学逼近模型不能完全适用于不同区域特点^[6]。

离散点还原趋势面的方法很多,DEM 模型常用的插值算法有加权平均法、移动拟合法、多面叠加法等几类,具体的常用方法如反距离加权插值法、自然邻点插值、局部多项式插值和克里金插值等。这些插值方法除在 DEM 模型建立方面发挥重要作用,在诸多领域有着广泛的应用,如地震、气象、煤矿、水文、地质、大地测量等^[7-9]。本文主要使用交叉验证的方法对常用的几种插值方法进行精度评定,讨论对于不同区域特征该如何选择适合的插值模型。

1 几种常用插值方法

从数学过程看,DEM 模型插值是对待插未知点高程的一种估算,从物性角度分析,DEM 模型插值是具备空间地理意义的,每种插值算法都直接或者间接地表达了地理目标之间的空间相关关系。

收稿日期:2011 - 05 - 10

作者简介:顾春雷(1982 -)男,江苏南通人,工程师,固体地球物理专业硕士研究生,主要从事测绘工程、地震监测、分析等工作。

1.1 反距离加权插值法 (Inverse Distance to a Power)

反距离加权插值法是假设平面上分布一系列离散点 $P(x, y, z)$, 已知其位置坐标 $P(x_i, y_i)$ 和属性值 $z_i (i=1, 2, \dots, n)$ 。根据周围离散点的属性值, 通过距离加权插值求 P 点属性值。反距离加权插值法综合了泰森多边形的邻近点法和多元回归法的渐变方法的长处, 它假设 P 点的属性值是在局部邻域内中所有数据点的距离加权平均值, 可以进行确切的或者圆滑的方式插值。其插值原理是待插值点邻域内已知散乱点属性值的加权平均, 权的大小与待插点的邻域内散乱点之间的距离有关, 是距离 $k (0 \leq k \leq 2, k$ 一般取 2) 次方的倒数。即:

$$P(Z) = \frac{\sum_{i=1}^n \frac{Z_i}{[d_i(x, y)]^k}}{\sum_{i=1}^n \frac{1}{[d_i(x, y)]^k}}$$

其中, $d_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, 表示由离散点 (x_i, y_i) 至 $P(x, y)$ 点的距离。这种方法是一种精确性插值法, 其优点是可以通过权重调整空间插值等值线的结构。

1.2 克里金插值法 (Kriging)

克里金 (Kriging) 插值法是一种空间自协方差最佳内插法。该方法是以空间结构分析为基础进行估值, 它首先考虑的是空间属性在空间位置上的变异分布, 确定对一个待插点值有影响的距离范围, 然后用此范围的采样点来估计待插点的属性值。该方法在数学上可对所研究的对象提供一种最佳线性无偏估计 (某点处的确定值) 的方法。克里金插值法依赖与数学模型和统计模型。用克里金插值法预测的结果将和概率联系在一起, 即从一个统计模型中不可能完全精确地得出预测值, 但它仍是一种光滑的内插方法, 在数据点多时, 其内插的结果可信度较高。

1.3 最小曲率法 (Minimum Curvature)

最小曲率法广泛应用于地球科学, 是构造出具有最小曲率的曲面, 使其穿过空间场的每一点, 并尽可能使曲面变得光滑。在插值过程中要尽可能严格地适合于采样数据空间分布特点, 同时生产尽可能平滑的曲面。最小曲率法不是一个精确的插值方法, 也就是说在插值过程中不可能总是完全适合于采样数据空间分布的特点。

1.4 线性插值三角网法 (Triangulation with Linear Interpolation)

线性插值三角网法是使用的最佳 Delaunay 三角形, 将连接数据点间的连线形成三角形, 而且所有三角形的边都不能与另外的三角形相交, 其结果构成了一张由三角形拼接起来的覆盖网格范围的网作为原始数据点的连接方法。

每一个三角形定义了一个覆盖该三角形内格网节点的面, 三角形的倾斜和标高由定义这个三角形的三个原始数据点确定。给定三角形内的全部节点都要受到该三角形的表面的限制。因为各个三角形都是用原始数据点定义的, 这样就把三角形和数据紧密联系起来。线性

插值三角网法将在网格范围内均匀分配数据, 地图上稀疏的区域将会形成截然不同的三角面。

1.5 移动平均法 (Moving Average)

移动平均法是指在格网点搜索椭圆内, 通过指定格网节点值的平均值作为待定点上的属性值。

设在一个区域内采位置坐标为 X_i , 变量观测值为 $Z(X_i) (i=1, 2, \dots, n)$, 则预测点 X_0 的估计值 $\hat{Z}(X_0)$ 为搜索邻域内 n 个样本点的平均值: $\hat{z}(x_0) = \frac{1}{n} \sum_{i=1}^n z(x_i)$ 。

1.6 局部多项式法 (Local Polynomial)

局部多项式插值法是指在格网节点搜索椭圆内通过对数据进行最小平方拟合来指定格网节点的值。该方法用特定的一阶、二阶和三阶多项式方程仅仅对指定的搜索邻域内所有点进行插值。所搜邻域之间相互重叠, 预测值是拟合的多项式在区域内中心点的值。

局部多项式插值法不是一个精确的插值方法, 但它能够得到一个平滑的表面。对于建立平滑表面和确定变量的小范围的变异可以使用局部多项式插值法, 特别是数据集中含有短程变异时, 局部多项式插值法插值生成的表面就能描述这种短程变异。

2 交叉验证及评定指标

交叉验证是在建模时评价某一插值方法的质量, 基本思路是在给定的建模样本中, 拿出大部分样本建立模型, 并使用建立的模型对预留的小部分样本进行预测, 求出这小部分样本的预测误差。通过交叉验证可以确定设置的插值方法相关参数是否合理, 或者用来比较不同插值方法的相对质量, 以寻求最适合的插值方法。

交叉验证的评定指标有平均估计误差百分比 (PAEE)、相对均方差 (RMSE)、均方根预测误差 (RMSPE) 以及计算残差分析等。其中均方根预测误差可用于评判插值方法之间的优劣比较, 均方根预测误差越小, 待插点的预测值就越接近于真实值。

2.1 均方根预测误差 RMSPE

$$RMSPE = \sqrt{\frac{1}{n} \sum_{k=1}^n [Z_i^*(X_k) - Z_i(X_k)]^2}$$

其中 $Z_i^*(X_k)$ 为未知 X_k 随机变量 Z_i 的估计值; $Z_i(X_k)$ 为未知 X_k 上的样点真实值; n 为用于交叉验证的预测样点个数。

2.2 残差分析 (Residuals Statistics)

残差值可用于定量估计源数据的 Z_{real} 值与数学逼近方法内插 $Z_{calculate}$ 值之间的一致性。根据残差计算结果可以计算平均残差和残差标准差等。

$$Z_{res} = Z_{real} - Z_{calculate} \quad \sigma_{res} = \sqrt{\frac{\sum_{i=1}^n (Z_{real,i} - Z_{calculate,i})^2}{n - 1}}$$

其中 Z_{res} 为残差值; σ_{res} 为残差标准差。

3 实例计算与结论

选择多种插值方法对不同区域特征的 DEM 模型进

行插值计算,并使用交叉验证评定各种插值方法对于不同复杂程度的高程模型插值的有效性。本文共计选择了 3 个特征区域,分别为长江三角洲平原区、大别山东段丘陵区和秦岭高山区,各区域的 DEM 模型原始采样点均为

10 201 个,区域范围为 100 km²,用于交叉验证的点数为 1 020 个,占总样本点数的 10%。图 1 为本文用于实验的 3 个不同区域特征的 DEM 模型;表 1、表 2 和表 3 分别为其对应区域的插值方法有效性交叉验证结果。

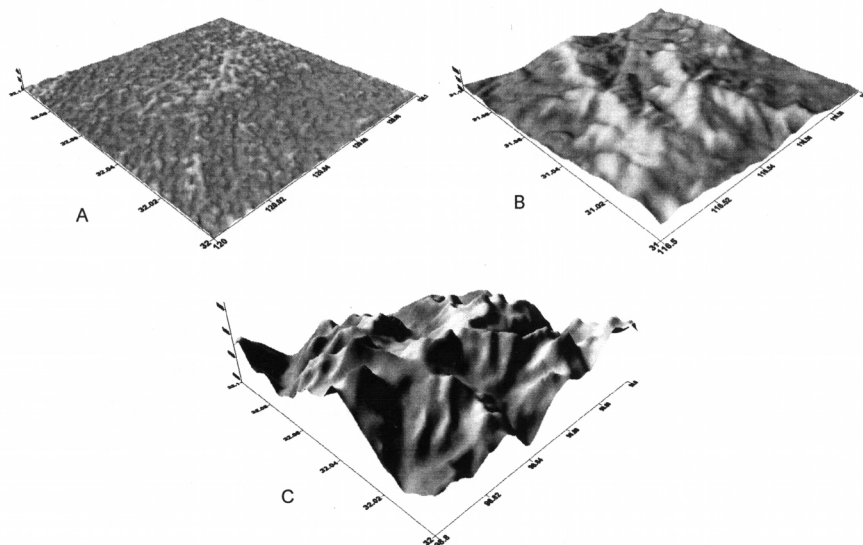


图 1 3 种不同区域特征的 DEM 模型(克里金插值方法建立)

A. 平原 B. 丘陵 C. 高山

Fig. 1 Digital elevation model of three difference region by Kriging interpolation method

A. Plain region B. Hilly region C. Mountain region

表 1 6 种插值方法在平原地区插值有效性交叉验证结果(单位:m)

Tab. 1 Cross-validation results of 6 interpolation methods at plain region(Unit: m)

插值方法	最大偏离	最小偏离	平均偏离	均方根预测误差	残差标准差
反距离加权	3.941 0	0.003 0	-0.070 0	1.027 1	1.025 2
克里金	1.698 6	0.000 1	0.005 4	0.792 5	0.792 9
最小曲率	2.820 2	0.000 0	-0.079 0	0.925 3	0.922 3
线性插值三角网	2.474 4	0.000 1	-0.033 2	0.817 5	0.817 2
移动平均	4.697 0	0.000 3	0.078 5	1.441 0	1.439 7
局部多项式	3.011 3	0.000 9	-0.009 1	1.165 5	1.166 0

表 2 6 种插值方法在丘陵地区插值有效性交叉验证结果(单位:m)

Tab. 2 Cross-validation results of 6 interpolation methods at hilly region(Unit: m)

插值方法	最大偏离	最小偏离	平均偏离	均方根预测误差	残差标准差
反距离加权	57.303 8	0.082 8	0.926 0	15.444 5	15.440 2
克里金	16.217 2	0.000 5	-0.061 3	5.568 3	5.571 8
最小曲率	44.665 3	0.011 2	-1.648 2	11.449 6	11.374 8
线性插值三角网	44.844 8	0.000 0	-0.329 0	9.8048	9.806 0
移动平均	269.308 6	0.086 5	24.237 4	135.0044	135.865 9
局部多项式	98.181 8	0.057 5	-0.660 6	20.827 5	20.842 2

表3 6种插值方法在高山地区插值有效性交叉验证结果(单位:m)

Tab.3 Cross-validation results of 6 interpolation methods at mountain region(Unit:m)

插值方法	最大偏离	最小偏离	平均偏离	均方根预测误差	残差标准差
反距离加权	141.611 1	0.043 2	-1.888 6	28.386 8	28.337 8
克里金	39.085 4	0.000 3	-0.345 2	8.340 3	8.337 2
最小曲率	122.254 4	0.016 7	-4.452 0	20.501 6	20.022 2
线性插值三角网	96.299 4	0.000 0	-0.369 9	12.110 3	12.110 6
移动平均	806.345 6	0.508 8	33.740 5	315.164 7	313.507 2
局部多项式	192.450 2	0.034 6	0.049 6	45.599 1	45.621 4

交叉验证结果的均方根预测误差说明 6 种插值方法的有效性由好到差依次为克里金、线性插值三角网、最小曲率、反距离加权、局部多项式和移动平均方法。

对于地势低平,起伏和缓,相对高度一般不超过 50 m,坡度在 5°以下的平原地区多种插值方法都是有效的,其中克里金、最小曲率和线性插值三角网三种方法效果较好,克里金插值方法的最大偏离误差为总高差的 3.40%;对于起伏不大,坡度较缓,由连绵不断的低矮山丘组成的地形,坡度在 25°左右的丘陵地区,克里金插值方法效果最好,其最大偏离误差为总高差的 4.28%,最小曲率和现行插值三角网次之,最大偏离误差约为总高差的 11.80%;对于高度较大,坡度较陡高差超过 1 000 m 的高山地区,克里金插值方法是有效的,其最大偏离误差为总高差的 3.74%,而其他插值方法最大偏离误差相对于总高差在 9.21%~77.16%之间。

4 结束语

综合分析得出以下几点结论:

1) 交叉验证方法可以对多种 DEM 插值方法进行科学评价,择选出最适合的建模方法。

2) 在众多插值方法中,克里金插值方法有很好的适应性,对于不同复杂程度的地形均能实现有效的空间插值,精度较高。对于地形起伏不大的平原地区,最小曲率和线性插值方法也是可以采用的。

参考文献:

- [1] Lay R. Accuracy assessment of digital terrain models [J].

Auto-Carto London, 1986(1): 455-464.

- [2] Kidner D B. Higher-Order interpolation of regular grid digital elevation models [J]. International Journal of Remote Sensing, 2003(14): 2981-2987.
- [3] Kumler M P. An intensive comparison of triangulated irregular networks and digital elevation models [J]. Cartographica, 1994(1): 1-9.
- [4] 朱长青,王志伟,刘海砚. 基于重构等高线的 DEM 精度评估模型 [J]. 武汉大学学报(信息科学版), 2008, 33(2): 153-156.
- [5] 任志峰,刘学军,卢华兴,等. 基于 Strahler 积分的 DEM 精度评价模型 [J]. 南京师大学报(自然科学版), 2008, 31(2): 119-123.
- [6] 汤国安,龚健雅,陈正江,等. 数字高程模型地形描述精数量化模拟研究 [J]. 测绘学报, 2001, 30(4): 361-365.
- [7] 余卫东,李湘阁,王靖. 气象场等值线自动绘制 [J]. 气象科技, 2002, 30(4): 222-225.
- [8] 蔡占川,梁延研,李坚,等. 基于嫦娥一号卫星激光测高数据的月球数字高程模型 [J]. 地球物理学进展, 2010, 25(4): 1153-1160.
- [9] 李振海,田辉. 离散重力数据等值线生成算法 [J]. 地理空间信息, 2010, 8(2): 6-9.
- [10] 王建,白世彪,陈晔. Surfer 8 地理信息制图 [M]. 北京: 中国地图出版社, 2004.

[编辑:宋丽茹]

(上接第 98 页)

参考文献:

- [1] Ziems M, Gerke M, Heipke C. 2007. Automatic road extraction from remote sensing imagery incorporating prior information and colour Segmentation [J]. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. ISPRS 2007, Vol36(3/W49A): 141-147.
- [2] Luo J, Ming D, Liu W. Extraction of bridges over water from IKONOS panchromatic data [J]. International Journal of Remote Sensing, 2007, 28(16): 3633-3648.

- [3] 杨朝辉. 基于数学形态学的 SPOT 影像桥梁提取 [J]. 苏州科技学院学报, 2009, 26(1): 53-57.
- [4] 许玮,卢凌,聂明新. 基于“3S”的大型水上桥梁识别系统 [J]. 武汉理工大学学报, 2007, 31(1): 35-38.
- [5] 骆剑承,明冬萍,沈占锋,等. 高分辨率遥感影像桥梁特征提取方法研究 [J]. 计算机应用研究, 2006(10): 151-153.

[责任编辑:王丽欣]