

离散数据网格化参数的确定和数学模型的选择 ——以 Sufer 7.0、Mapgis 6.0 为例

高艳芳

(中国地质科学院地球物理地球化学勘查研究所, 廊坊 065000)

[摘要] 使用计算机绘制等值线首先要对离散数据进行网格化处理。针对物化探数据的特点, 详细地讨论了利用 DEM 模型进行网格化时如何设定所需的系列参数, 同时对采用的数学模型进行了应用比较。

[关键词] 等值线 网格化参数 数学模型

[中图分类号] P628 [文献标识码] A [文章编号] 0495-5331(2002)S0-0139-04

0 前言

等值线图(contour map)是物化探工作中经常使用的图件。它利用区域上离散点的信息, 来反映区域上客观存在的连续变化的面特征。计算机绘制等值线图常用的方法有三角网法和格网法。

最常用的制图软件—Sufer 7.0、Mapgis 6.0, 都有很强的等值线绘制功能(数字高程模型), 而且皆采用格网法绘制等值线图(Mapgis 6.0 也可以采用三角网法绘制)。根据格网法等值线的绘制原理, 必须有一个将离散数据转化为网格数据的过程, 也就是采用一定的网格化方法(数学模型)对不规则分布的原始数据点进行插值, 生成所需范围内的规则间距分布的数据点。在这个过程中就要涉及到网格化参数的设定、选择最优的网格化数学模型等几方面的问题。

1 网格化参数的设定

1.1 网格化的范围选择

网格化的范围也就是要获得的网格节点的最大最小坐标值, 也就是等值线的分布范围。程序默认的值是离散点的分布空间。因此要获得比离散点范围大的等值线分布, 就需要进行扩边。选择扩边后, 程序网格化时将采用外推的方法获取空区处的值。外推获得的值, 准确度肯定有所下降。因此网格化的范围即等值线的范围不应偏离采样点的分布区域

太多。当然, 网格化的范围可以小于离散点分布的空间, 但这肯定给野外工作带来负担。因此工作中要兼顾采样点分布与成图区的关系。

1.2 网格间距和网格数

网格间距和网格数是网格化时最重要的参数。它关系到所派生数据的密度, 并直接影响等值线模型的精度, 最终影响异常的空间结构特征。若网格过大, 会丢失一些特征信息, 也可能造成等值线扭曲, 将使野外工作达不到一定的效果。理想状态下, 网格数越多, 所绘的结果越详细。但是, 实际上当网格达到一定的间隔后, 无论再怎样增加网格数, 等值线图的轮廓几乎不再发生任何变化。过细的网格, 会产生更多冗余的“游离”数据, 即相邻网格点值的差别微小而不包含有效的等值线特征信息。这样的网格化过程不仅增加了计算和存储资源的开销, 同时也使异常空间分布特征复杂化, 导致一些假的、没有意义小线圈的出现。

图 1 中的原始数据是 $3\text{ km} \times 3\text{ km}$ 分布的。从图中可以发现, 网格化后的数据为原来的 2 倍以上, 等值线既可反映其原始的情况, 又能满足要求。所以, 在网格化时, 网格间距要小于原始采样点的分布间隔, 但也不能过细。通常采用的间距以网格化后的数据点数为原数据点数的 2 至 4 倍左右为宜。实际应用中, 对地球化学调查获取的 1 km^2 一个点的数据进行网格化, 网格间距采用 0.5 km 、 2 km^2 一个点的采用 1.5 km 或 1 km 的间距即可, 依此类推。

[收稿日期] 2002-07-29 [责任编辑] 徐大良。

[作者简介] 高艳芳(1965年-)女, 1984年毕业于长春地质学校, 高级工程师, 主要从事物化探处理、制图及软件开发等工作。

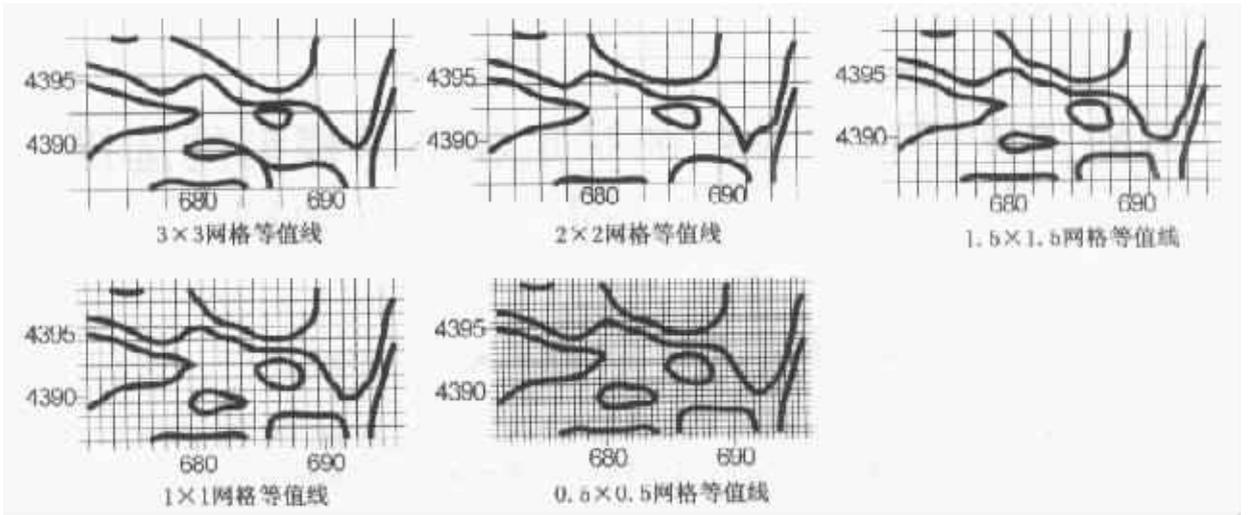


图 1 设定网格间距示意图

1.3 搜索方式和搜索半径的确定

1.3.1 搜索方式

搜索方式决定了网格化形成新的网格节点所使用的数据。可分为所有点搜索、简单搜索(单方向)、象限搜索(4 方向)、卦限搜索(8 方向)等方式。理论上数据点小于 250 个时,而且数据分布比较均匀,可以使用所有的数据进行网格化,即不选择其搜索方式。当数据量增大,而且数据分布疏密不均,须要采用象限搜索、卦限搜索这两种方式。距离倒数、克里格和半径函数等数学方法允许在数据量小于 200 个点时,采用所有点搜索。使用所有的数据点这一搜索方式,在计算过程中也应用权因子来插值计算网格节点的值,离节点近的数据点给予的权较远距离的点权大。

1.3.2 搜索半径和搜索规则

搜索半径决定了所用的数据范围。它和其它 3 项搜索规则一起来决定网格化时所使用的点。这 3 项规则是:每一个扇区内所用的最多点数、必须使用的最少有效点数和允许的最多空方向数。网格化时推荐使用的最佳的有效点数为 5。实际设定时,给定这 3 项的值最好宽松一些(取其默认值即可)。

采用 1 km^2 一个点的化探数据,在搜索规则不变的条件下,任意改变搜索半径的大小,从图 2 可以发现,搜索半径的变化对数值点区域的等值线不产生影响,但是要影响到空区的填充。当设定搜索半径为 1 或 2 时,等值线圈定了实际的点分布范围。随着半径的增大,添满了所有的空区。所以搜索半径是一个重要的参,它直接影响网格化后的结果。若半径太小则找不到形成网格节点的

就不可能形成企望的等值线,太大则将空值点变为了数据点。

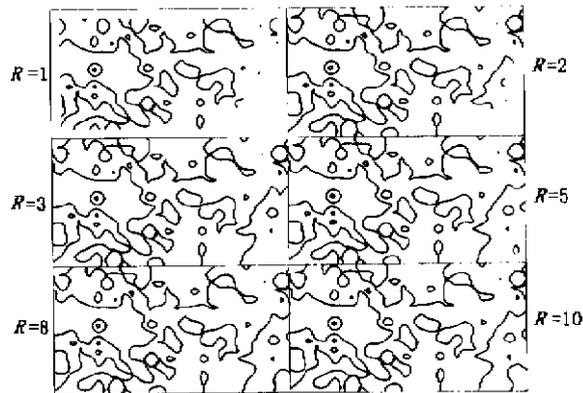


图 2 不同的搜索半径网格化后的等值线图对比
1—原始数据为 $1 \text{ km} \times 1 \text{ km}$ 2—其它的网格化参数一致

鉴于此,搜索半径的确定以客观反映实际的地质情况为条件。例如 1 km^2 一个点的原始数据要获取间隔为 0.5×0.5 的网格数据,考虑到扩边和特殊点的缺乏,搜索半径设为 2 km 即可。具体应用中,可视需要适当的增大或减少。

2 网格化模型的选择

网格化使用的数学方法是网格化的核心问题。所有的插值方法,都是希望通过插值加密数据点,使一个粗糙的初始地质模型变成一个新的精确模型,这仅是理想状态。相反,经过插值计算得到的新模型,相对于原始模型来讲总是有一定的变形。也就是说,相对于实测的、精确的原始来说,插值点往往会有一定的误差。因此最大限度地减少这个误

差,使原始数据中包含的客观特征能无明显损失地传递到内插计算的数字模型中去,是对所有的数学模型的最基本的要求。

可用来对离散数据进行网格化的数学模型很多。在 Sufer7.0 中,给定了 9 种数学模型,分别是:距离倒数(Inverse distance to a power)、克里格(Kriging)、最小曲率(Minimum Curvature)、最近临近点(Natural Neighbor)、自然临近点(Natural Neighbor)、多项式回归(Polynomial Regression)、径向基本函数(Radial Basis Function)、改进谢别德(Modified Shepard's Method)、三角网/线性插值(Triangulation with Linear Interpolation)。Mapgis6.0 也给出了 4 种之多,分别是 距离倒数、(泛)克里格法、稠密数据中值选取法、稠密数据高斯距离权法。不同的数学模型有各自的特点,不同类型的数据可以选择相应的数学模型。

距离倒数、克里格、径向基本函数、最近点这 4 种数学模型可以采用搜索方式设定参数进行网格化。其他的几种进行网格化插值计算时,采用的是全部数据进行网格插值计算。尽管所有的 DEM 数字高程模型类的应用程序提供给用户的网格化的方法很多,但并不是所有都能满足物化探数据处理的要求。

距离倒数是一种加权平均插值法,可以进行确切的、圆滑的插值。多年来,在物化探数据处理过程中被广泛采用。它的特征之一是在网格区域内产生围绕观测点位置的“牛眼”。“牛眼”可以通过改变圆滑参数进行消除。

普通克里格法或泛克里格方法是一种适合于许多领域的网格化插值方法,目前应用也比较广。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,同时也兼顾了数据空间分布特征,其内插的结果可信度高。它能够比较精确地表示出隐含在原始数据中的趋势,避免产生距离倒数中过多“牛眼”点。

径向基本函数是多个数据插值方法的组合。其中的复二次函数被许多人认为是最好的方法,既能反映原始数据的特征,又能光滑等值线。网格化结果相似于克里格插值结果。使用它来制作地球化学图,效果最好。

最近临近点法,严格地说,它不是一种插值法,它只是将距离网格最近的点的安插到网格的节点上去。这个法是针对点均匀分布而且空值点又少的原始来说,这样的不需要插

值就可以绘制等值线。所以它是将一个数据点分布均匀的文件转为网格文件的过程,只是添加了一个数据头、填充了空格。

最小曲率法不是一个很精确的插值方法。它过多地追逐了等值线的圆滑,而降低了数据的准确度。最小曲率法在插值一个节点时,其值要经过多次的反复计算,满足它的最小残留值,达到光滑的目的。

自然临近点是在 Sufer7.0 中出现的一种新的网格化方法。在许多领域被广泛应用。因为要严格地构造一种多边形,同时只能在其所构造的多边形内部进行插值,所以会造成许多不应该出现的空值点。表现在等值线图上是不应有的空白区域的出现。

多项式回归法实际上不是插值,只是被用来确定数据的空间趋势。又称为趋势面分析。

谢别德法是使用距离倒数加权的最小二乘方的方法。因此,它与距离倒数乘方插值原理相似,但它利用了局部最小二次乘方来消除或减少所生成的等值线的“牛眼”外观。

三角网/线性插值方法工作路线和手工绘制等值线相近。这种方法是通过在数据点之间按一定的规则连接以建立起若干个三角形,进行插值计算。

在图 3 和图 4 中,我们选用了所有的网格化模型进行了网格化处理,并绘制了等值线图。图 3 采用的是均匀分布的原始数据点,图 4 采用的是不均匀分布的原始数据点。它们进一步说明了以上各个数据模型的特征。其中多项式回归是我们在做网格化时首先排除在外的。若数据量小于 250 个点,可

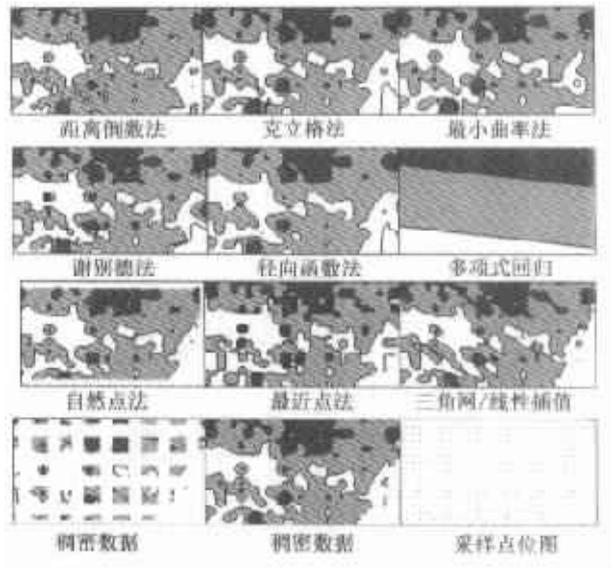


图3 不同网格化 学模型所形成的等值线图对比 (分布均匀的离散点)

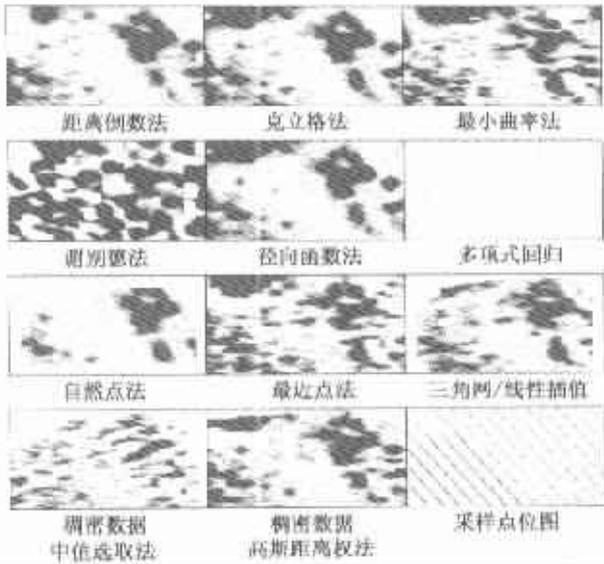


图4 不同网格化数学模型所形成的等值线结果对比
(分布不均匀的离散点数据)

大而且数据分布不均匀,我们可以利用的方法只有距离倒数、克里格、径向基本函数这3种数学模型,更适合化探数据的是克里格、径向基本函数中的复二次函数。

3 结论和建议

当为绘制等值线准备网格化数据时,在了解原始数据特征的前提下,合情合理地选择网格化参数和数据模型是必要的。也许由于软件本身的原因,比如 Sufer7.0 在追踪等值线时会出现等值线的交叉。这种情况下,你可以将原始数据网格化时的间距不要设得太小,在网格化后再对网格数据进行样条插值处理,利用这样的数据结果来做等值线将非常圆滑。这是另外一种做等值线图的途径。

[参考文献]

[1] 黄健全,罗明高,胡雪涛.实用计算机地质制图[M].北京:北京地质出版社,1998.

以考虑使用谢别德、最小曲率方法。但是若数据量

SPECIFYING THE PARAMETERS OF GRIDDING AND CHOOSING GRIDDING ALGORITHM

GAO Yan - Fang

(Institute of Geophysical and Geochemical Exploration CAGS Langfang 065000)

Abstract Produce a contour map by computer must need a gridding process of scattered data. Be dead against the specialty of Geophysical & Geochemical data, how to specify the parameters of gridding and choose gridding algorithm are given in this paper.

Key words contour map, parameters of gridding, gridding algorithm