《地理信息系统原理》实验报告

实验名称: 地统计分析

姓 名： 马骁

班 级： 21级地信1班

学 号： 07212393

中国矿业大学环境与测绘学院

2023年6月23日

# 实验八 地统计分析

## 一、实验目的与主要原理

1、实验目的

地统计分析方法被广泛应用于许多领域，已成为空间统计学的一个重要分支。利用地统计分析模块，可以根据一个点要素层中已测定的采样点、栅格层或者利用多边形质心，轻而易举地生成一个连续表面。这些采样点的值可以是海拔高度、地下水位的深度或者污染浓度值等。通过本实验，掌握地统计分析的基本知识、基本技能，具有应用地统计分析解决实际问题的能力。熟悉ArcGIS地统计分析模块提供的一整套创建表面的工具，以及通过这些表面来可视化、分析及理解各种空间现象。

2、实验内容

1. 探索性空间数据分析；
2. 表面预测（五个插值模型）和误差建模；
3. 模型检验与对比。

## 二、过程和结果

1、功能模块介绍

ArcGIS地统计分析模块由三个主要功能模块和两个帮助文档组成，探索性数据分析、地统计分析向导以及生成数据子集。

1) 探索性数据分析工具

探索性数据分析是为了全面了解所使用的数据，以便于选取更合适的参数和方法。如数据是否服从正态分布，是否存在某种趋势等。在ArcGIS地统计分析模块中，内嵌了多种探索性空间数据分析工具，包括直方图、正态QQPlot分布图、趋势分析等。

2) 地统计分析向导工具

地统计分析模块提供了一系列利用已知样点评估内插精度，完成表面预测和误差建模。主要的操作界面包括：

(1) 选择内插方法和数据集界面。通过此界面可选择内插方法、内插数据及辅助数据等。

(2) 参数设置界面。通过此界面设置模型所用的各项参数，如所使用的模型、块金值、基台值、步长等，参数设置应根据实例而定。对输出表面质量评价的最严格方法就是将观测值与预测值进行比较。

(3) 精度评定界面。主要包括误差分布图、误差标准化值分布图等。通过这些界面可以了解模型的精确度，也为模型的改进提供了必要的信息。

(4) Create Subsets对话框。可以生成测试和训练数据集。

2、启动ArcMap并激活地统计分析模块

按图操作加载地统计分析模块。

3、探索性空间数据分析

1) 数据分布

数据服从正态分布时，插值方法生成表面的效果最佳。如果数据是偏态分布的，即向一边倾斜，则需要通过数据变换使之服从正态分布。因此在插值表面之前分析数据分布十分必要。频率直方图描绘了数据属性的频率直方图分布，能够针对数据集的每一种属性检测其单变量分布。

(1) 加载实验八的实验数据，并进行直方图统计，选择属性OZONE，结果如图3-1所示。

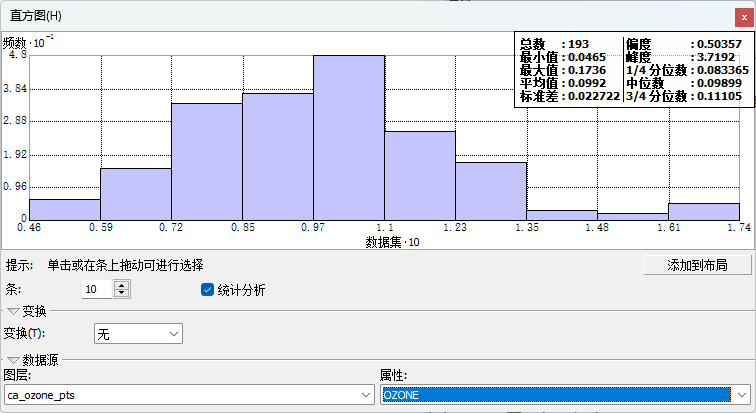


图3-1 数据源设置

臭氧属性的分布情况用一个直方图描述，该直方图将浓度值分为10级，每一级别中数量的相对比例通过每一个直方条柱子的高度来表示。上面显示的直方图表面数据是单峰分布的，而且具有较好的对称性，接近于正态分布。

(2) 点击直方图中任意一列，图层中对应的要素将被高量显示。

2) 正态QQ图

正态QQ图提供了另外一种度量数据正态分布的方法，利用QQ图可以将现有数据的分布与标准正态分布对比，数据点越接近一条直线，则它们越接近于服从正态分布如图3-2所示。

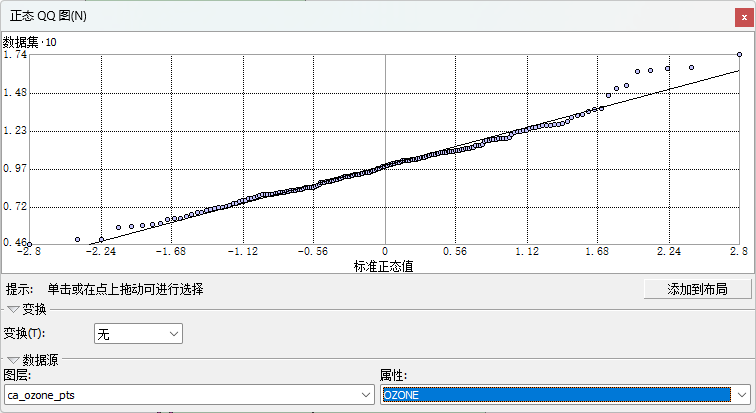


图3-2 正态QQ图

在一个普通的正态QQ图上，两种分布的对应点一一对应。对于两种相同类型的分布，正态QQ图应该是一条直线。因此，通过绘制相对应的臭氧数据的分布点与标准正态分布的分布点，能够检查臭氧数据的正态分布情况。从上述正态QQ图可以看出，该图形非常接近于一条直线。而偏离直线的情况主要发生在臭氧浓度值较高时。

3) 识别数据的全局趋势

当数据中存在某种趋势时，可以利用某些数学公式对表面的非随机(确定性的)成分进行表达。例如，一个缓倾斜的山坡可以用一个平面来表达，而山谷则可以利用一个能够生成“U”字形的更加复杂的一个二次多项式来表示。数学公式虽然能够生成表面，但大多数时候，数学公式太过于平滑而不能精确地描述表面。因此，趋势面不能精确地描绘实际需要的表面，需要将其移去，通过建立趋势剔除后的残差模型来继续分析，有助于提高插值精度。

趋势分析图中的每一根竖棒代表了一个数据点的值和位置。这些点被投影到一个东西向的和一个南北向的正交平面上如图3-3。通过投影点可以作出一条最佳拟合线，并用它来模拟特定方向上存在的趋势。上图中的东西方向的线从较低的值开始，向东移动时逐渐增加直到变平稳。这表明该数据在东西向上显示出一个很强的趋势，而在南北向的趋势则较弱。

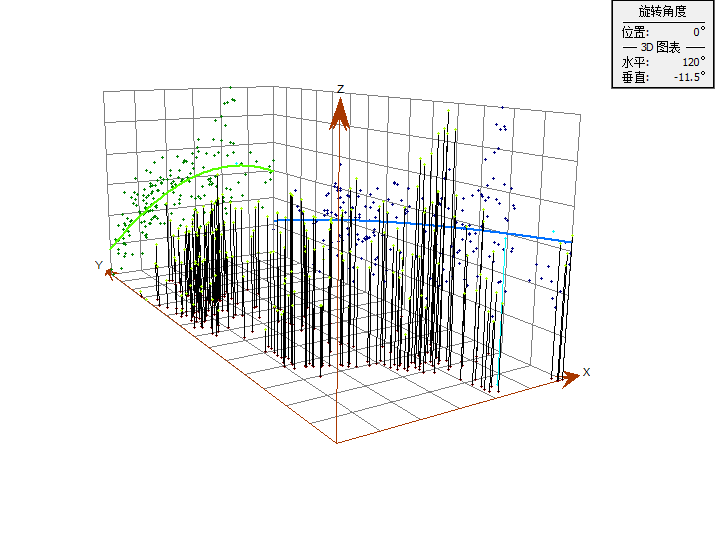


图3-3 趋势分析图

通过旋转旋转角为30°，东西向趋势的形状可以看得更清楚，显示为一个倒置的“U”型。既然该趋势呈“U”型，所以可以选择一个二阶多项式对其全局趋势进行较好地模拟，如图3-4。尽管这个趋势显示在了东西向的投影平面上，但因为把数据点旋转了30°，所以实际的趋势是北东-南西向。造成该趋势的一个可能的事实是，在沿海地区污染较轻，而在向内陆推进时，人口增多，污染增大。到了山区则人口又减少，污染也随之减少。

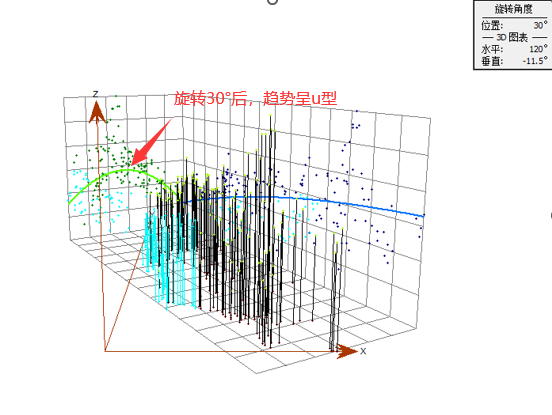


图3-4 旋转30°后的趋势分析图

4) 数据自相关和空间分析效应

半变异函数/协方差函数云图能够检测已测样点间的空间自相关。空间自相关理论认为彼此之间距离越近的事物越相似。半变异函数/协方差函数云图能够对这种关系进行检测。Y轴表示半变异函数值，即每一样点对间测量值之差的平方，而相应地用X轴表示每对样点之间的距离。

在半变异函数/协方差函数云图中，每个红点表示一对采样点。既然越近的点越相似，在半变异函数云图中邻近的点，应该有较小的半变异函数值。随着样点对间距离的增加，变异函数值也要相应增加。当到达一定的距离后，云图变平，这表明超出这个距离时，样点对之间不再具有相关关系了。

除了全局趋势外，影响数据的还有方向效应。这些方向效应的原因可能并不明了，但它们可以在统计上给予量化。

点击并将方向指针移动到任意角度。指针指向的方向决定了哪些样点对将会出现在变异函数图中，这就能够去除不感兴趣的那些点对，从而来检查数据中的方向效应。

单击并拖动选择工具，选中那些具有最大半变异函数值的点，使之在半变异函数图及地图中高亮显示。可以看到，无论距离大小，大多数相连的样点对都会对应到某一采样点上，如图3-5。



图3-5 具有大变异函数值的点

4、空间确定性插值

确定性空间插值是使用数学函数进行插值，以研究区域内部的相似性（如反距离加权插值法）、或者以平滑度为基础（如径向基函数插值法）由已知样点来创建预测表面的插值方法。

1) 反距离权重插值法

加载实验数据，选择地统计分析下的地统计向导，选择反距离权重法，源数据集中选择ca\_ozone\_pts，数据字段中选择OZONE，进行主要参数设置，如幂的设置，最大搜索半径等，然后获得交叉验证结果如图4-1。X轴代表样点的真实值，Y轴代表内插出来的样点值，细线（灰色）代表理论上的点值得拟合曲线，粗线（蓝色）代表内插点值得拟合曲线。二者吻合的越好，说明内插的效果越好。

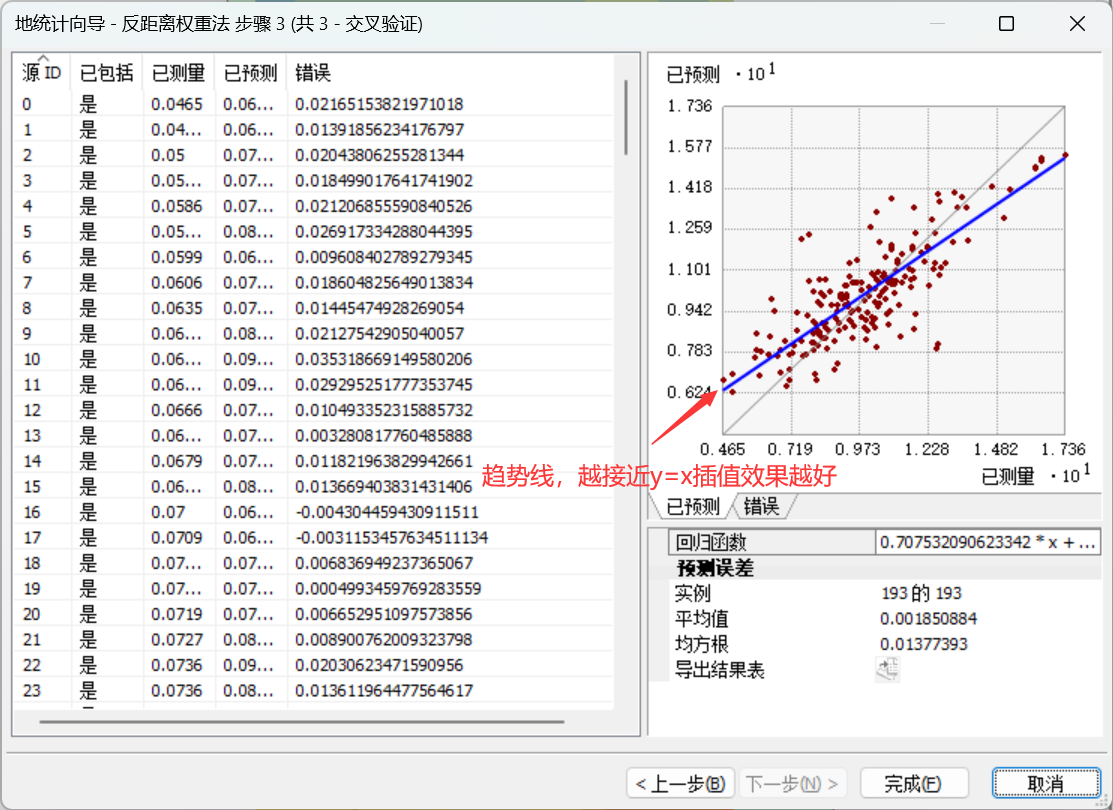


图4-1 交叉验证

对得到的结果进行裁剪，使插值结果在图层范围内。

2) 径向基函数插值

径向基函数插值法如同将一个软膜插入并经过各个已知样点，同时又使表面的总曲率最小，属于精确插值方法。所谓精确插值方法就是指表面必须经过每一个已知样点。

径向基函数包括五种不同的基本函数：平面样条函数，张力样条函数，规则样条函数，高次曲面函数和反高次曲面样条函数。径向基函数插值法适用于对大量点数据进行插值计算，同时要求获得平滑表面的情况。将径向基函数应用于表面变化平缓的表面，如表面上平缓的点高程插值，效果较好。

点击地统计分析下的地统计向导选择径向基函数插值法，不同于反距离权重法，径向基插值法通过选择核函数来进行插值。

核函数下拉菜单中提供了几种进行插值的基本函数：平面样条函数，张力样条函数，规则样条函数，高次曲面函数，反高次曲面样条函数。

其他参数的含义同反距离权重法。

核参数是用于设置控制表面光滑度的参数。对于反高次曲面样条函数来说，控制表面光滑度的参数越大，所得到的表面越不平滑；除此之外的其它所有径向基函数，控制表面光滑度的参数越大，所得到的表面越平滑。

3) 全局多项式函数插值

全局性插值方法以整个研究区的样点数据集为基础，用一个多项式来计算预测值，即用一个平面或曲面进行全区特征拟合。全局多项式插值所得的表面很少能与实际的已知样点完全重合，所以全局插值法是非精确的插值法。利用全局

插值法生成的表面容易受极高和极低样点值的影响。

全局多项式插值法适用的情况有：当一个研究区域的表面变化缓慢，即这个表面上的样点值由一个区域向另一个区域的变化平缓时，可以采用全局多项式插值法利用该研究区域内的样点对该研究区进行表面插值；检验长期变化的、全局性趋势的影响时一般采用全局多项式插值法，被称为趋势面分析。

选择多项式的阶进行全局多项式插值。

4) 局部多项式插值

局部多项式插值采用多个多项式，每个多项式都处在特定重叠的邻近区域内。通过使用搜索邻近区域对话框可以定义搜索的邻近区域。局部多项式插值法不是一个精确的插值方法，但它能得到一个平滑的表面。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，数据集中含有短程变异时，局部多项式插值法能描述这种短程变异。

在局部多项式插值法中，邻近区域相关属性需要进行设定如图4-2。局部多项式插值法产生的表面更多地用来解释局部变异。

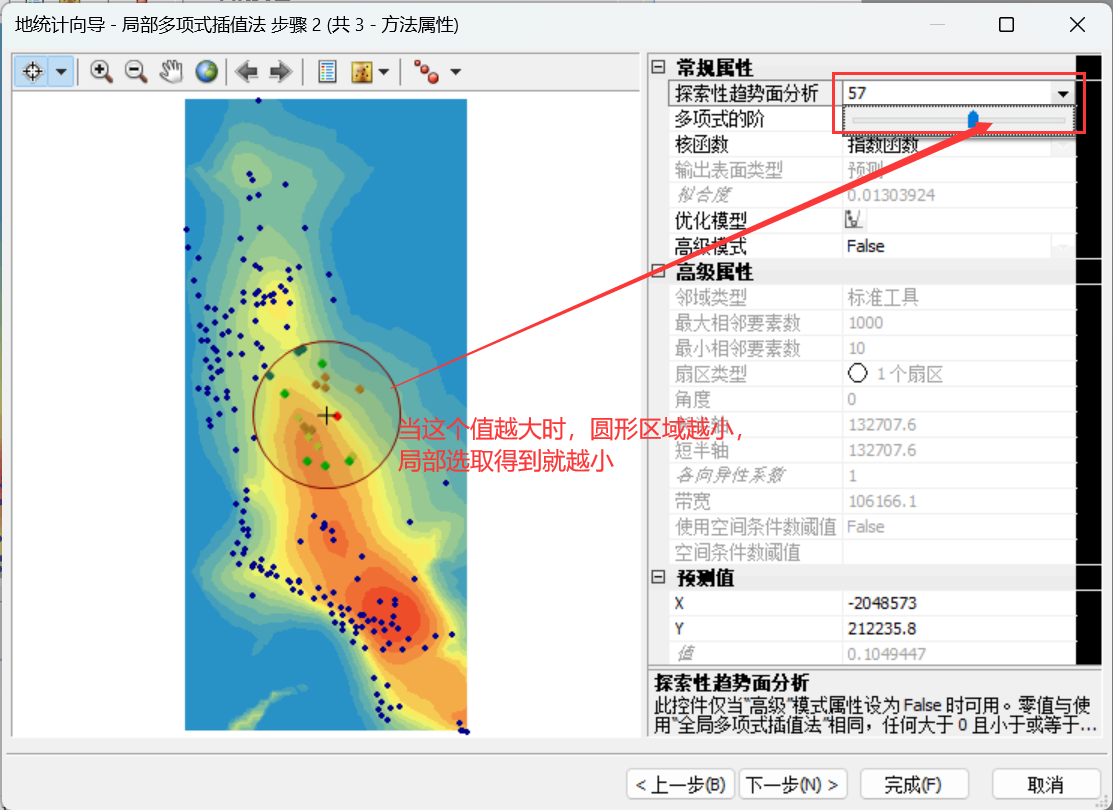


图4-2 局部多项式插值参数设置

5) 克里金插值

进入地统计向导，选择克里金/协同克里金法，选择普通克里金，变化类型无，趋势的移除阶数选择为2次。

在本实验的探索性空间数据分析中，通过趋势分析已经检测到一条南西-北东方向的“U”型曲线，选择二阶多项式拟合效果较好。

南东-北西方向的趋势可以从数据中剔除。剔除全局趋势后，就可以对表面残差或表面的短程变异成分进行统计分析。在创建最终表面之前，该趋势还将自动添加回来以产生正确的结果。全局趋势剔除后所进行的分析将不再受其影响，而一旦将全局趋势再添加进来，就能够生成一个更加精确的表面。

可以看出数据在南西-北东方向变化最快，下一步进行半变异函数参数设置如图4-3。

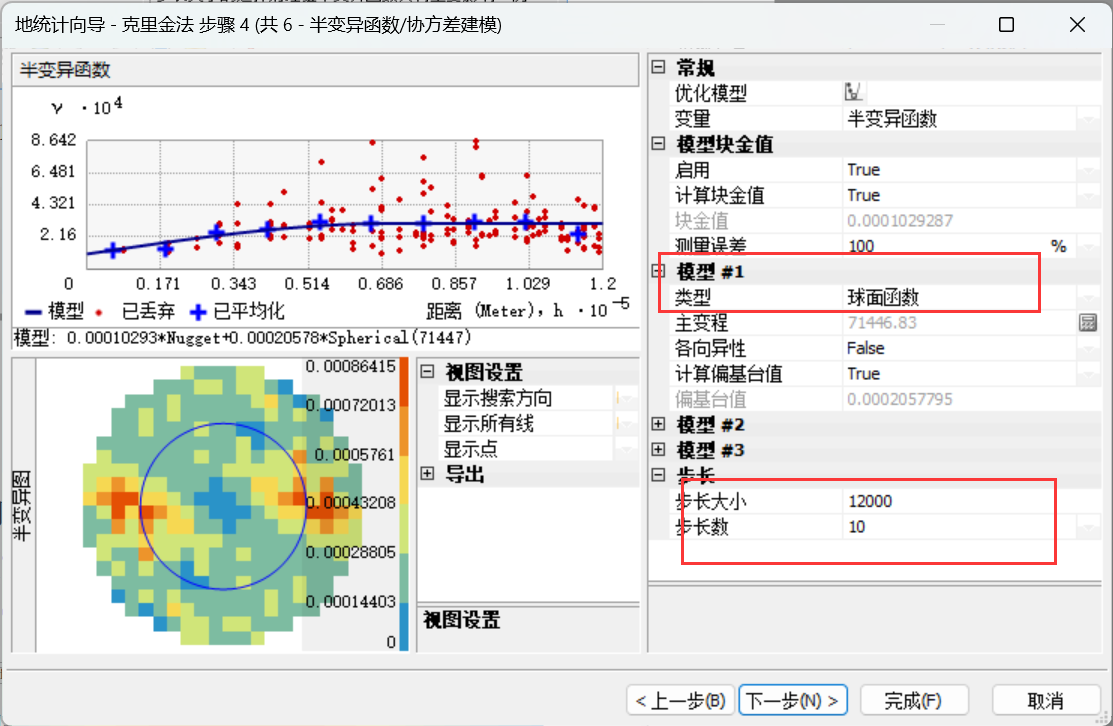


图4-3 半变异函数参数设置

半变异函数显示了具有不同距离的样点对的方差。用半变异函数/协方差函数进行模拟的目的在于为其确定一个最佳拟合模型，该模型将穿过半变异函数图中的那些点。

半变异函数是一个关于数据点的半变异值与数据点间距离的函数。半变异函数值的划分确定一个合适的步长大小，可以减少大量可能的合并而将数据点对分成不同的距离级。这种方法称为步长分组。在步长分组后可拟合一个球面半变异函数模型。

颜色比例尺代表了计算后的半变异函数值。蓝色和绿色代表的值较低，而橙色和红色代表的值较高。从半变异函数表面图可以看山，在南西-北东向的变异性要比南东-北西方向增加得要快。

进行方向半变异函数参数设置。方向效应会对半变异函数中的点以及将要拟合的模型产生影响。邻近的事物在某些方向上的相似性比其他方向的相似性更强。方向效应又被称为各向异性。

通过调整方向、容差、步长使方向指针分别于各向异性椭圆的长轴和短轴重合，观察半变异函数曲线。

在上述两步操作中，半变异函数达到稳定水平时的值相同，这个值称之为基台值。半变异函数达到其限值(基台值)时经过的距离，即为模型的自相关阈值（变程）。超出这个自相关阈值后，各点之间的变异性将随着步长距离的增加而变为常数。步长是通过点对间的距离来界定的。步长距离大于自相关阈值的样点对之间空间无关。块金效应代表了测量误差、微观尺度的变异。

对模型进行交叉验证结果如图4-4，该模型标准均方根误差值为0.86，效果较好。

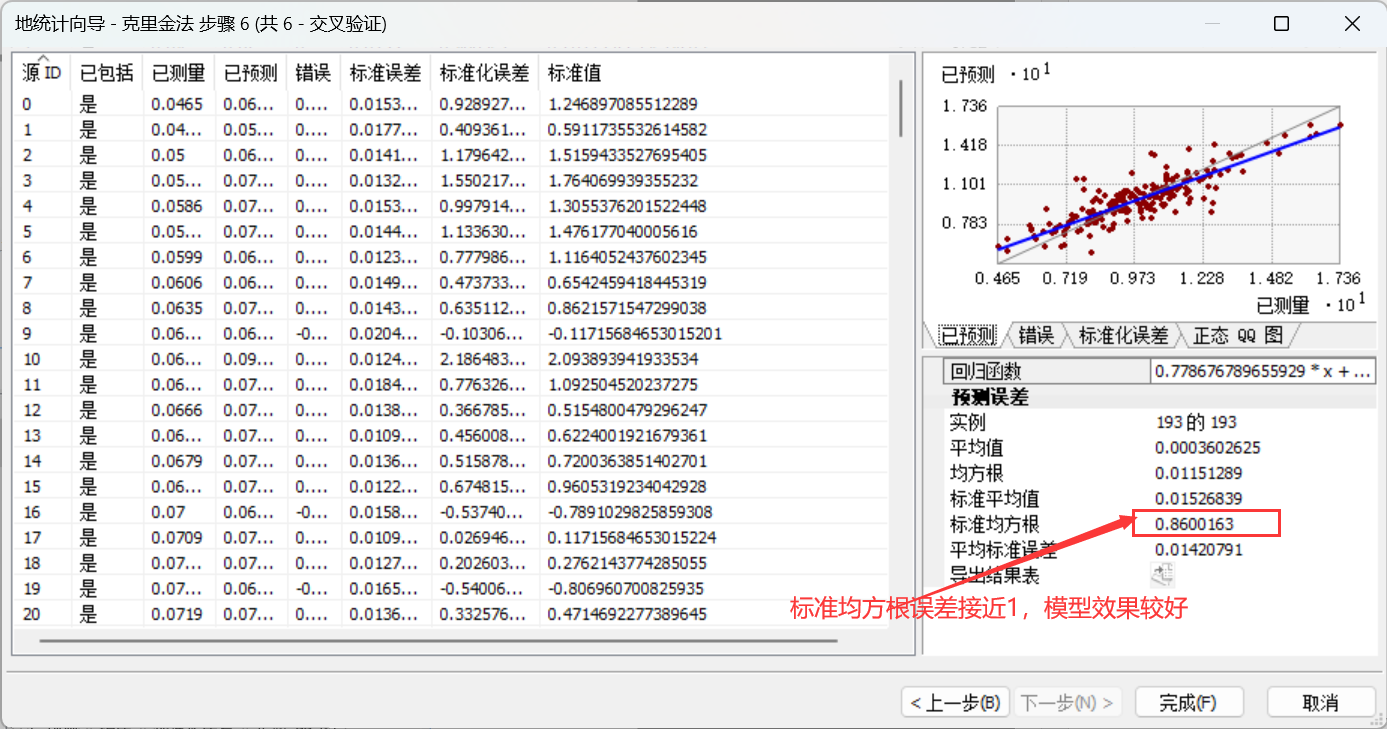


图4-4 克里金插值交叉验证结果

## 三、讨论

1、常用的半变异函数有哪些，如何得到半变异函数图？

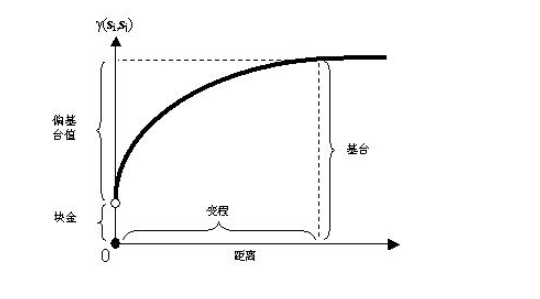
指数函数、高斯函数、幂函数、球状函数等。

收集样本数据，计算样点之间的空间距离。计算半变差：对于每对样点之间的距离，计算它们之间的半变差，即两点之间属性值的差异值。分组计算平均半变差：将不同距离范围内的半变差进行分组，并计算每个距离范围内的平均半变差。绘制半变异函数图：将距离范围作为横轴，平均半变差作为纵轴，绘制半变异函数图。根据具体的半变异函数类型，选择合适的函数公式进行拟合，并将拟合曲线绘制在图上。

2、何为基台值、块金值？

查看半变异函数的模型时，模型会在特定距离处呈现水平状态。模型首次呈现水平状态的距离称为变程，半变异函数模型在变程处所获得的值（y 轴上的值）称为基台。

理论上讲，在零间距处，半变异函数的值是 0。但是，在极小的间距处，半变异函数通常显示块金效应，即值大于 0。如果半变异函数模型在 y 轴上的截距为 2，则块金为 2。它也表示区域化变量的随机性大小。



3、常用的插值模型有哪些，各有什么优缺点？

1) 线性插值：

优点：简单、易于实现，计算效率高。

缺点：对于地形变化剧烈的区域，线性插值可能无法准确地捕捉到地形的细节变化，导致插值结果不够平滑。

2) 反距离加权插值：

优点：简单、易于实现，计算效率高。

缺点：对于数据点密集区域和离散值较大的情况，IDW插值可能会导致插值结果过于依赖局部数据点，对整体趋势的反映不够准确。

3) 克里金插值：

优点：克里金插值考虑了空间相关性，能够根据样本点之间的空间距离和方向来估计未知位置的值，因此可以更好地反映地形的变化趋势。

缺点：对于大规模数据集，计算复杂度较高，需要进行半方差函数的拟合和参数的估计。

4、如何验证插值模型的精度？

1) 交叉验证：

将数据分为训练集和验证集。使用训练集进行插值模型的参数估计和插值操作，然后使用验证集中的观测数据与插值结果进行对比。可以计算验证集中观测数据与插值结果之间的误差，如均方根误差RMSE或平均绝对误差MAE，来评估插值模型的精度。

2) 误差分析：

对于已知的观测数据点，计算插值结果与实际观测值之间的误差，并分析误差的分布特征。可以绘制误差分布图、误差直方图或误差散点图，以了解插值模型在不同地点和不同高程变化情况下的误差情况

5、影响插值精度的主要因素有哪些？

1. 数据密度和分布
2. 数据误差和异常值
3. 插值方法和参数
4. 空间相关性

《地理信息系统原理》实验报告

实验名称: 地形分析

姓 名： 马骁

班 级： 21级地信1班

学 号： 07212393

中国矿业大学环境与测绘学院

2023年6月23日

# 实验九 地形分析

## 一、实验目的和主要原理

1、实验目的

1. 加深对TIN建立过程的原理、方法的认识；
2. 掌握ArcGIS中建立DEM、TIN的技术方法；
3. 结合实际，掌握基于DEM地形分析的原理、与算法。

2、实验内容

DEM是对地形地貌的一种离散的数字表达，是对地面特性进行空间描述的一种数字方法、途径。实验内容主要包括TIN的构建以及DEM生成，以及DEM的应用，主要包括坡度、坡向、等高线提取、计算地表阴影以及可视性分析以及地形剖面等。

## 二、过程和结果

1、TIN及DEM生成

1) 创建TIN

由高程点、等高线矢量数据生成TIN并转为DEM。

选择3D Analyst模块下的数据管理工具进行创建TIN，选额要素并设置高度字段和类型如图1-1。

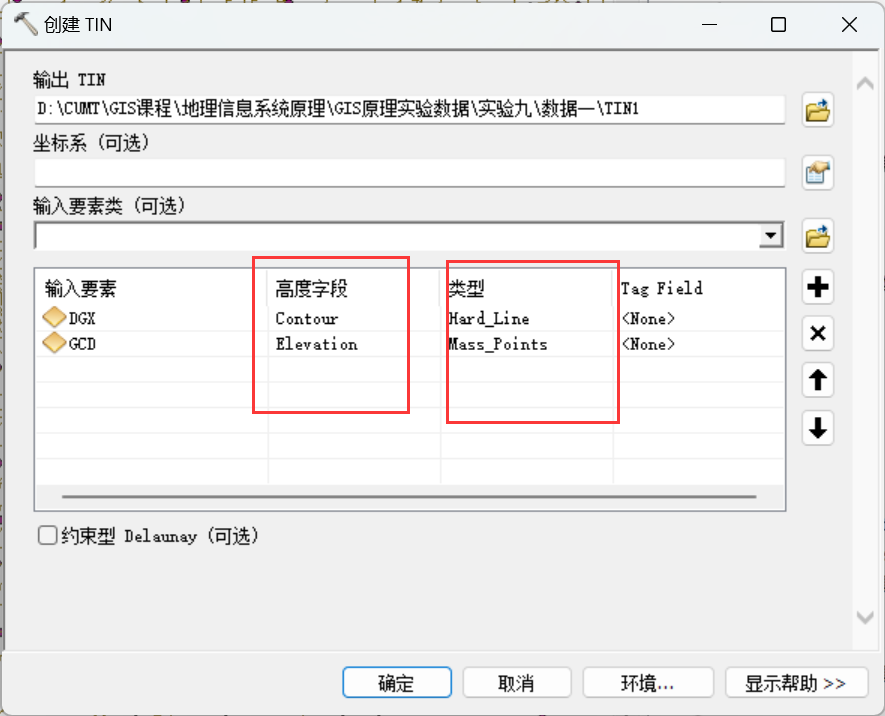


图1-1 创建TIN及设置

创建TIN成功后，进行TIN转为栅格。

2) TIN的显示及应用

为TIN进行渲染，打开TIN图层的符号系统，取消显示中的边类型和高程，添加渲染选择相同符号的边和相同符号的点这两项进行显示。

将具有分级色带的表面坡度和具有分级色带的表面坡向这两项添加到TIN的显示中；添加完成后，在对话框中选中坡度进行分类，手动设置中断值，图层渲染效果如图1-2。

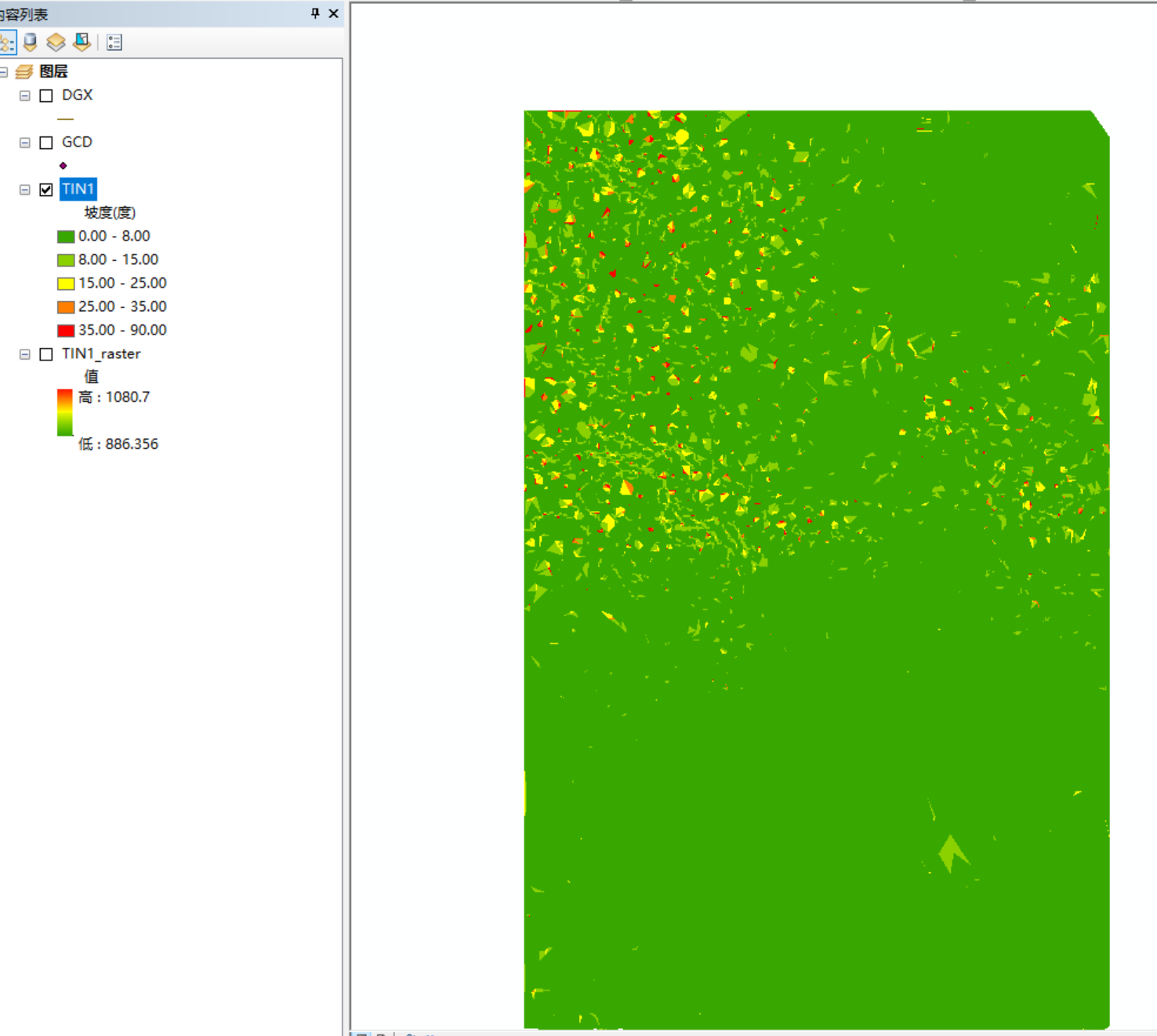


图1-2 坡度分类显示效果

2、DEM的应用

1) 求取坡度

坡度是指经过地表面任意一点的切平面与水平地面之间的夹角。坡度值较低表明地势较为平坦，坡度值高则地势陡峭。求取坡度的操作步骤如下：

加载TIN转栅格DEM数据，打开栅格表面工具集选择坡度，点击环境设置栅格分析的大小为30，结果如图2-1。

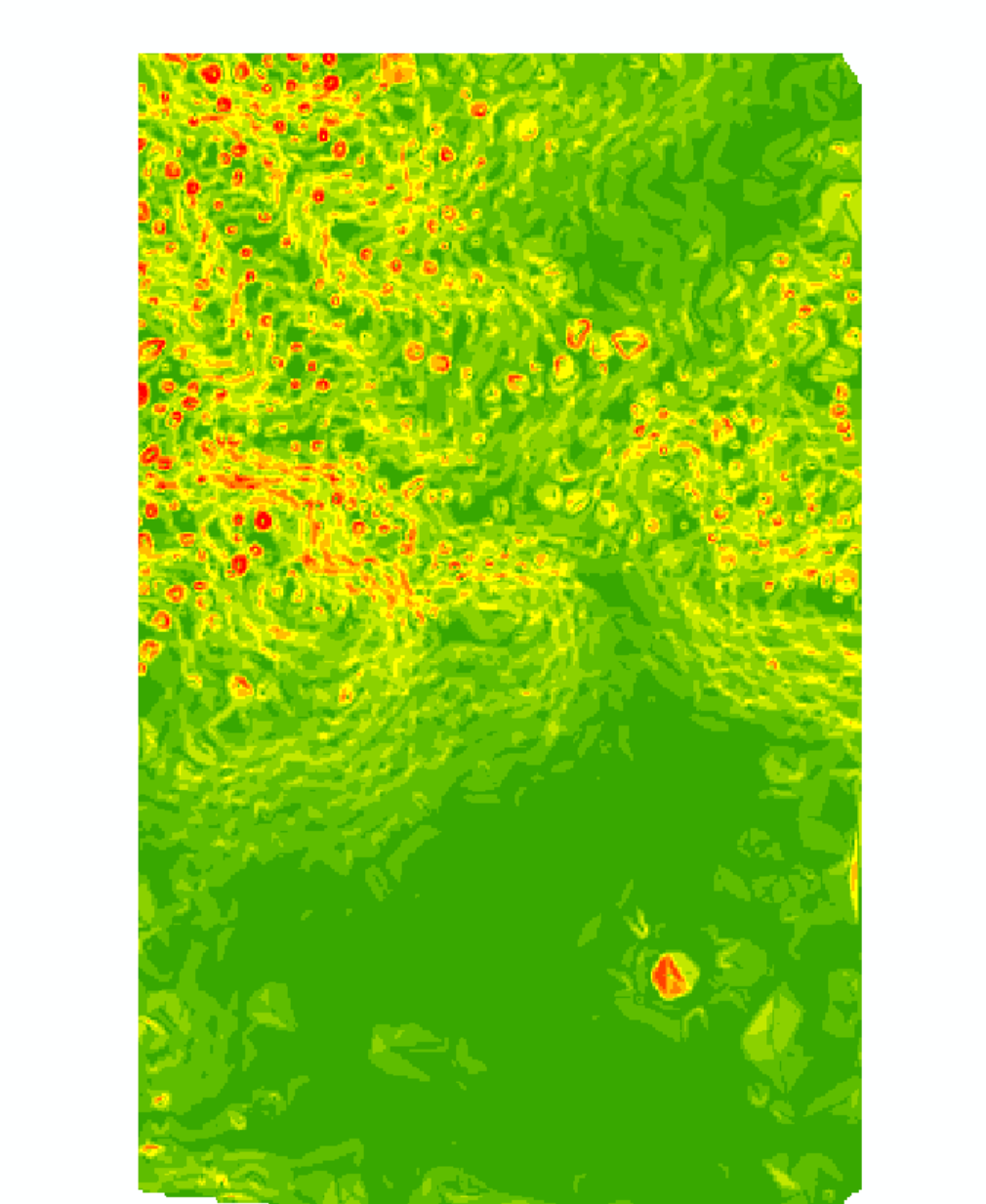


图2-1 坡度计算结果

并对坡度进行重分类。

2) 计算剖面曲率

曲率是对地形表面上一点扭曲变化程度的定量化度量因子。曲率为正，说明该像元表面向上凸；曲率为负，说明表面开口朝上凹入；值为0说明表面是平的。计算曲率的操作步骤如下：

选择栅格表面工具集下的曲率，选择刚刚计算的坡度图层，结果如图2-2。



图2-2 剖面曲率计算结果

3) 计算坡向

坡向是指地表面上任意一点的切平面的法线矢量在水平面的投影与过该点的正北方向的夹角。坡向表明了该点高程值改变量的最大变化方向。坡向值正北方向为0°，按顺时针方向计算，取值范围为0°~360°。坡向的提取过程如下：

打开栅格表面工具集选择坡向，输入图层选择由TIN创建的栅格数据，结果如图2-3。

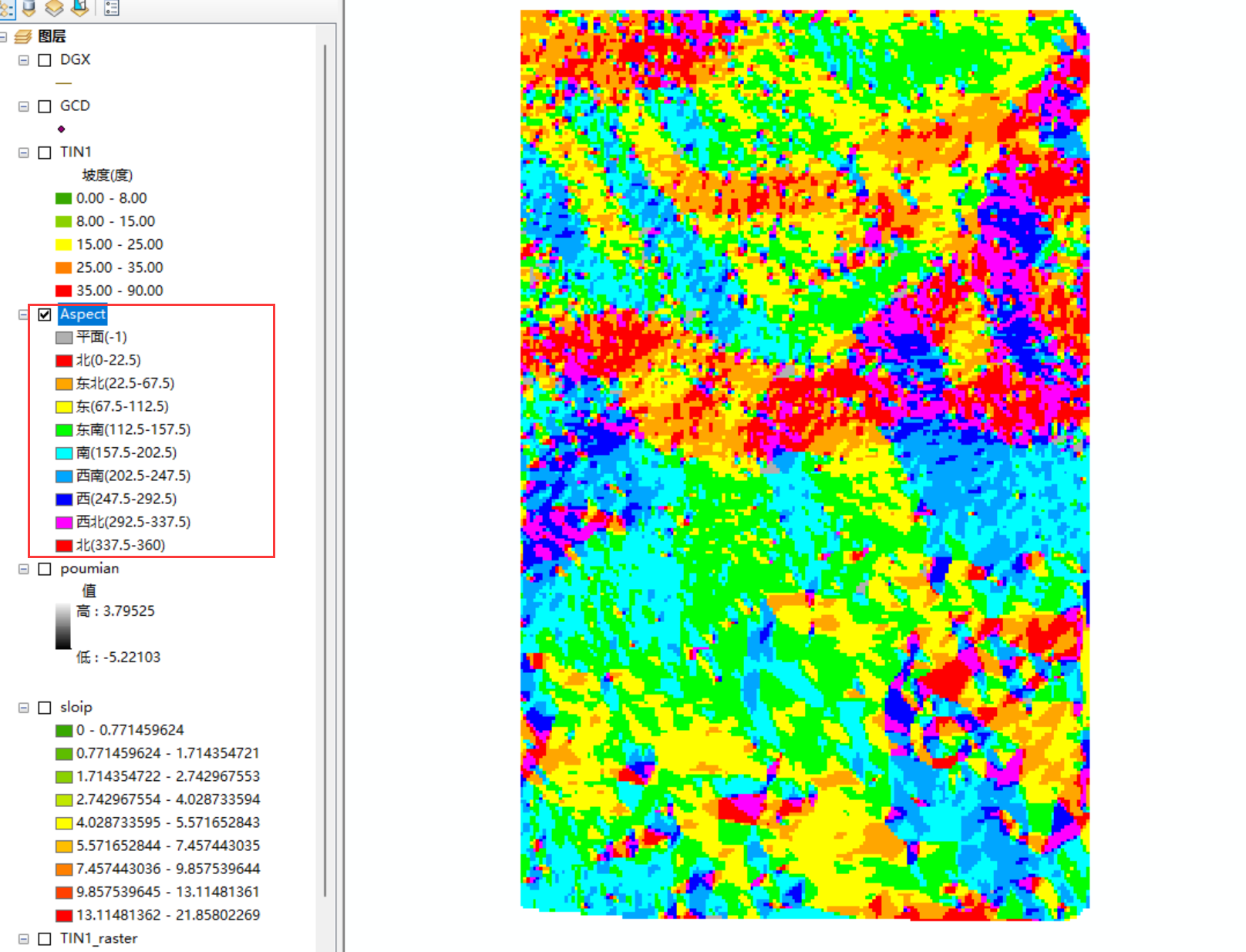


图2-3 坡向栅格

4) 计算平面曲率

平面曲率又叫地面坡向变率，是指在提取坡向的基础上提取坡向的变化率，也叫坡向之坡度（Slope of Aspect，SOA）。它可以很好地反映等高线的弯曲程度。地面坡向变率在所提取的地表坡向矩阵的基础上沿袭坡度的求算原理，提取地表局部微小范围内坡向的最大变化情况，操作步骤如下：

打开栅格表面工具集下的坡向，输入图层为刚刚计算的坡向，得到结果如图2-4。

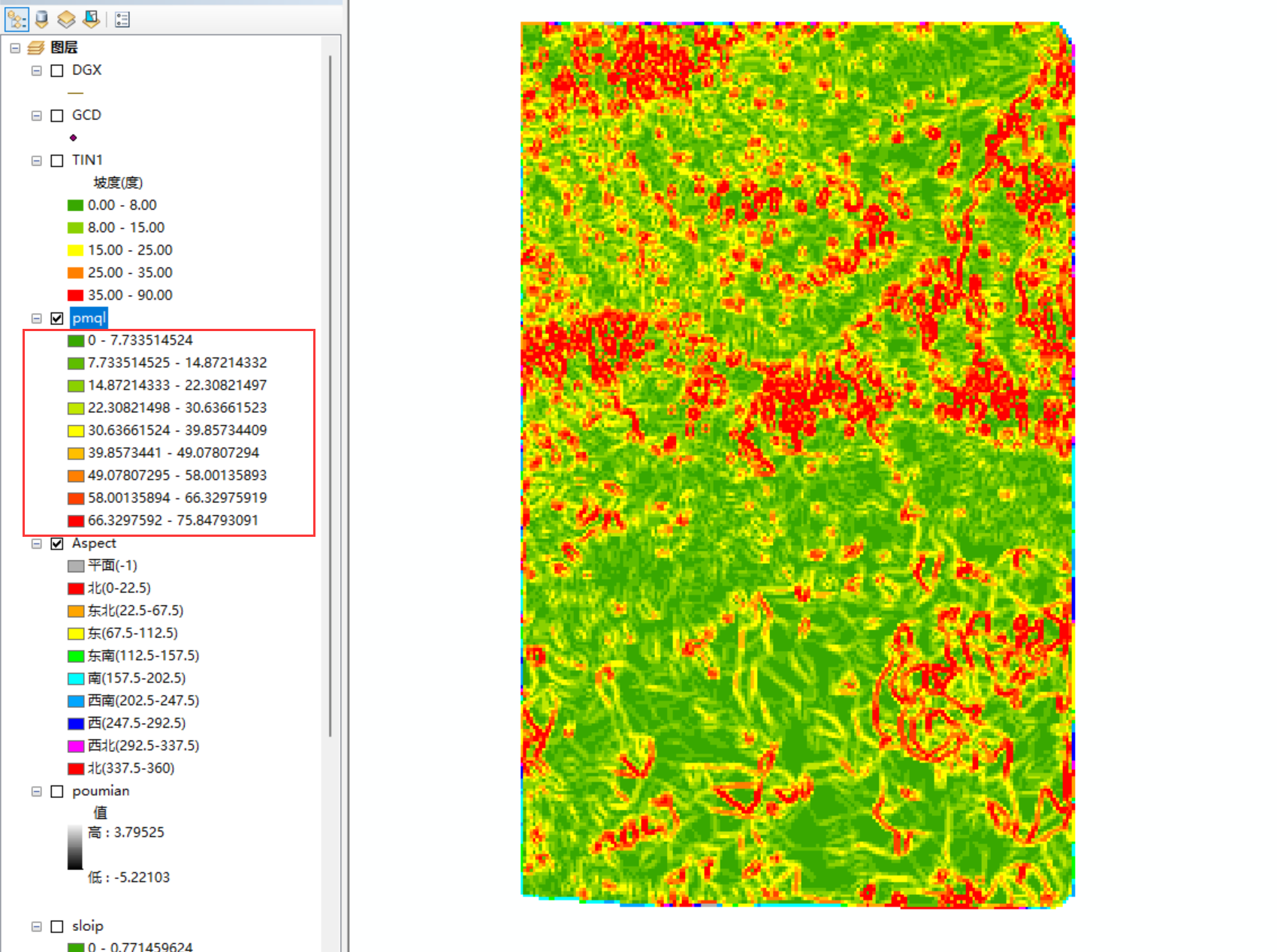


图2-4 平面曲率栅格

5) 根据DEM数据提取等值线

等值线是将表面上相邻的具有相同值的点连接起来的线（如等高线、等温线等）。等值线分布的疏密一定程度上表明了表面值的变化情况。通过研究等值线可以获得对表面值变化的基本趋势。由DEM数据提取等值线的步骤如下：

打开栅格表面工具集下的等值线，输入图层为TIN创建的栅格数据，结果如图2-5。

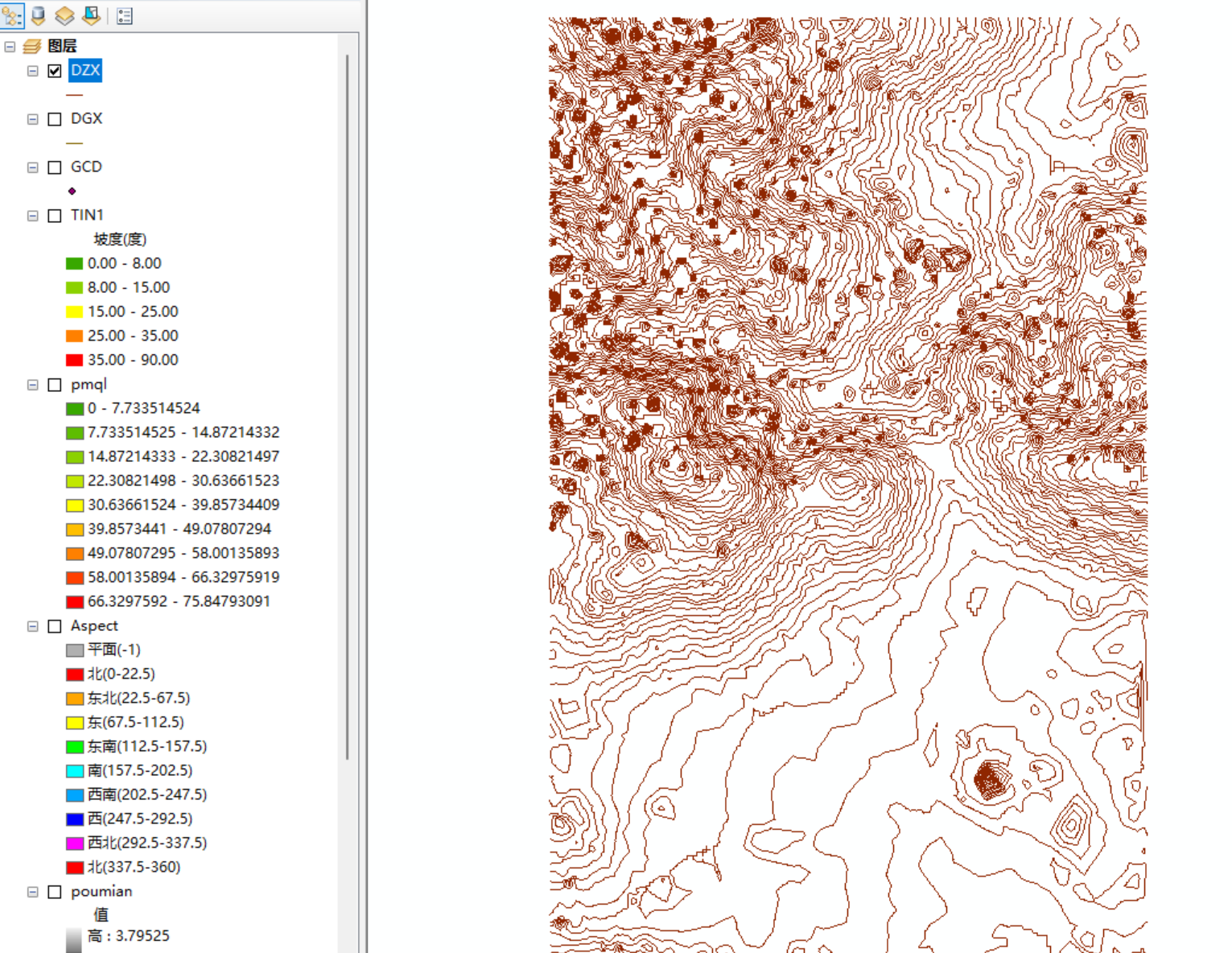


图2-5 等高线矢量图层

6) 计算地形表面的阴影值

山体阴影是根据假想的照明光源对高程栅格图的每个栅格单元计算照明值，其表达了地形的立体形态。计算阴影的过程中包括三个重要参数：太阳方位角、太阳高度角以及表面灰度值。计算山体阴影的操作过程如下：

打开栅格表面下的山体阴影工具，输入图层为由TIN转换后的栅格，进行设置，结果如图2-6。

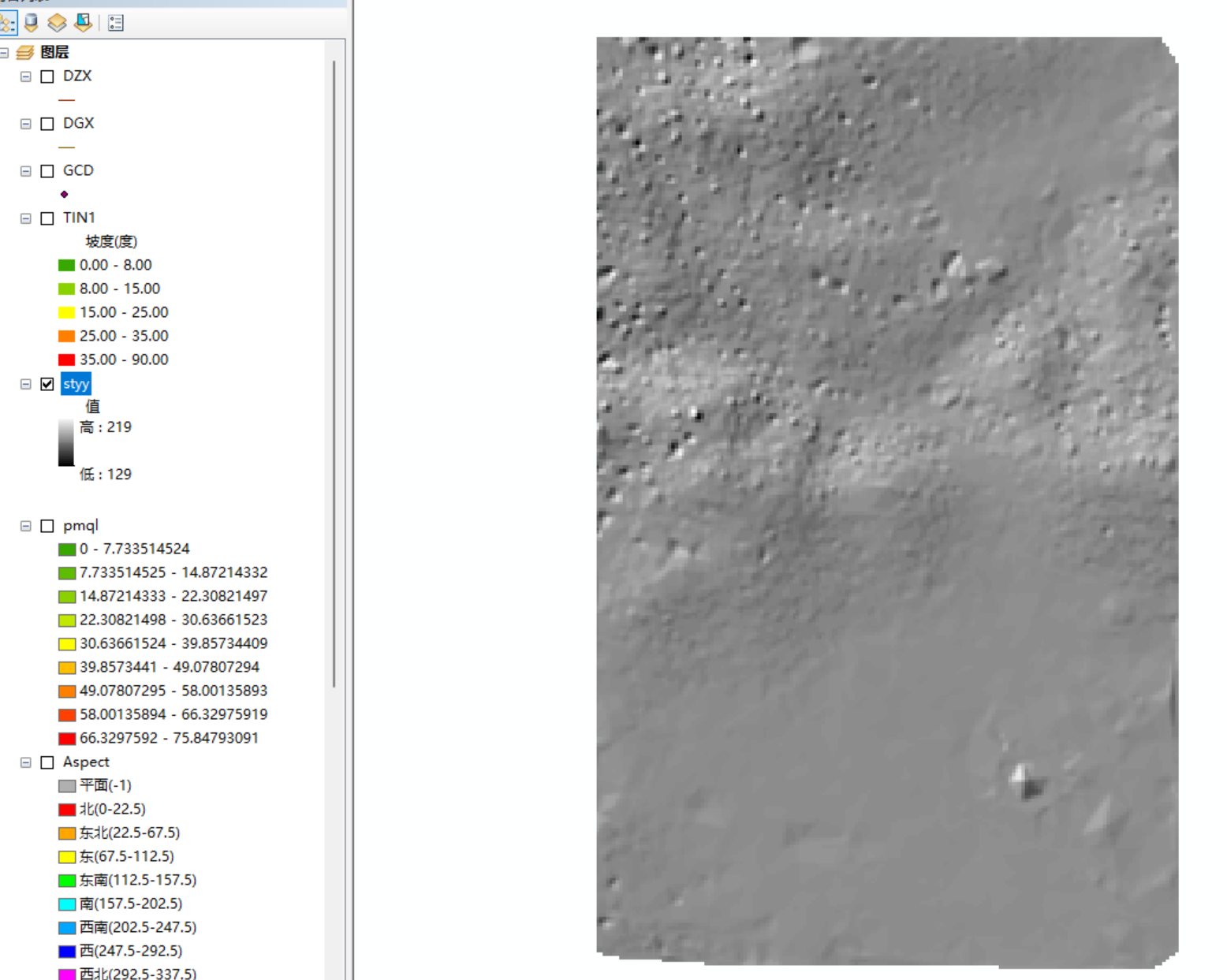


图2-6 山体阴影

将山体阴影和由TIN转换后的栅格数据叠加显示，设置透明度便于显示，结果如图2-7。

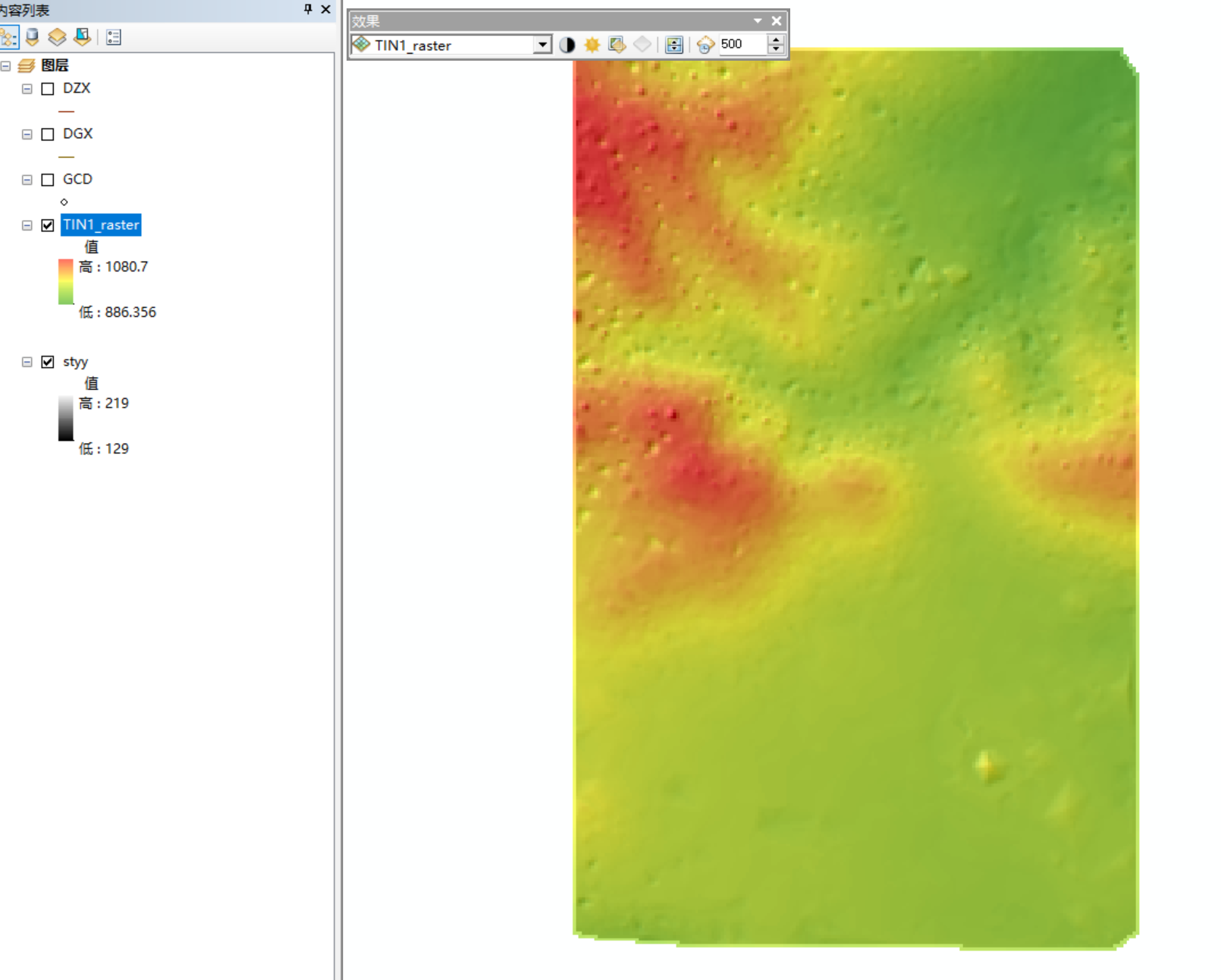


图2-7 DEM渲染设置

7) 可视性分析

在3D Analyst工具条下选择创建通视线工具，在通视分析对话框中输入观察者偏移量和目标偏移量，可以认为是到地表的距离。随后在图上点击输入观察者位置并绘制到目标的直线，可以得到观察点到目标点的通视性。其中绿色部分代表的是可视的部分，红色线段表示不可见的部分，结果如图2-8。

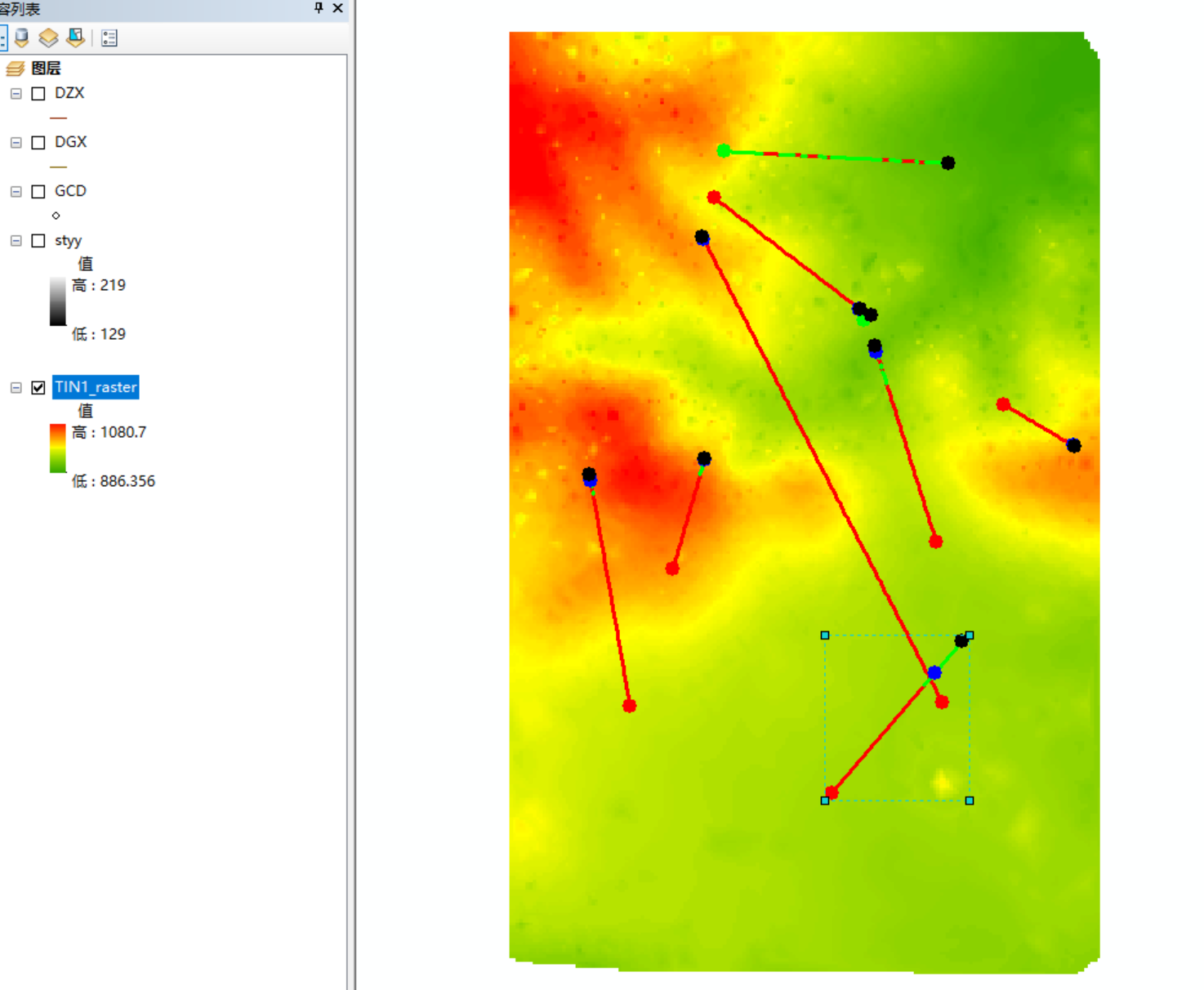


图2-8 通视分析结果

8) 地形剖面

在3D Analyst工具条中点击线插值工具，跟踪一条线段，该线段从由TIN转后后的栅格数据中获得高程值。随后点击创建剖面图按钮创建地形剖面图，如图2-9。

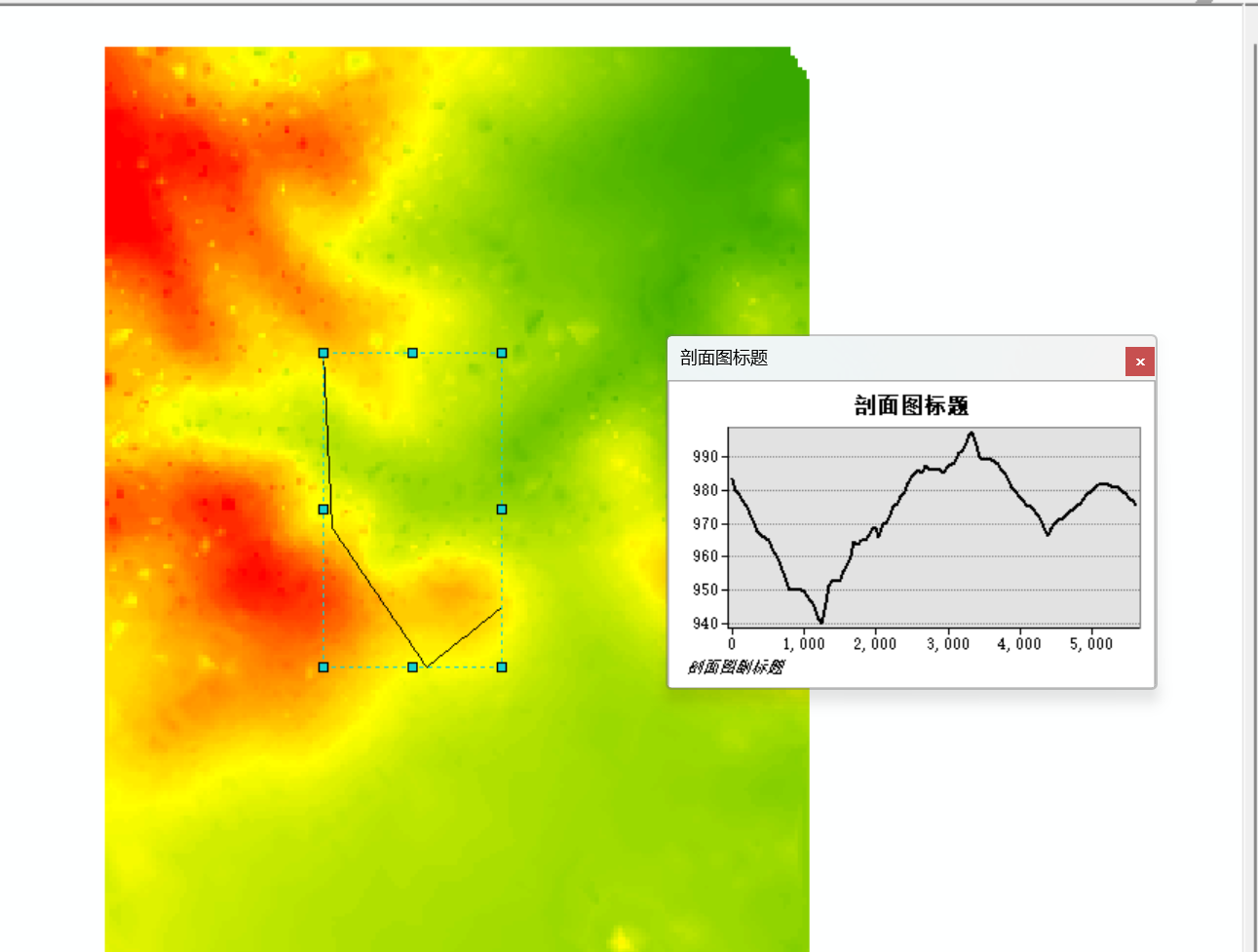


图2-9 创建地形剖面图

## 三、讨论

1. 地形坡度计算有哪几种方法？
2. 简单斜坡算法，通过计算两个像元之间的高程差，并将其转化为坡度值。
3. 多方向斜坡算法，简单斜坡算法只考虑了水平和垂直方向上的高程变化，而多方向斜坡算法考虑了更多方向上的变化。它使用一个环形邻域来计算每个像元的坡度，考虑了像元周围的高程变化情况。
4. 累积斜坡算法，累积斜坡算法考虑了从起点到当前像元的累积坡度。它通过将像元与其邻域像元的坡度值相加，计算从起点到当前像元的累积坡度。
5. 能否从TIN中直接计算坡度？考虑其算法。

计算TIN中某个点的坡度需要考虑其周围的三角形，首先确定目标点P及其周围的三角形：遍历TIN中的所有三角形，找到包含目标点P的三角形。计算每个三角形的法向量，再计算累积法向量，得到一个平均法向量。

计算坡度：使用平均法向量来表示地形表面的法线方向。地形表面的坡度可以通过计算该法线向量在水平方向上的分量来确定。

这种计算适合连续的点，若为离散的数据点还应插值为栅格，再计算坡度。

1. 在构建TIN时，加入等高线有什么作用？
2. 数据辅助和验证：构建TIN之前，可以使用等高线数据来辅助确定地形特征，识别地形凹凸点和沟谷线等。等高线数据可以用于验证TIN的准确性和一致性。
3. 控制TIN的拓扑关系：等高线数据可以作为TIN构建的辅助约束。在构建TIN时，可以使用等高线作为约束条件来确保TIN中的三角形与等高线相交，从而更好地捕捉地形的细节和特征。
4. 填补数据空白区域：在一些地区，可能缺乏高程数据或高程数据的密度不足。等高线数据可以用于填补这些数据空白区域，从而提供更完整的地形模型。
5. 可视化和分析：等高线数据具有直观的特点，可以用于地形的可视化和分析。

加入等高线可以提供对地形的可视化表示，并在TIN构建过程中起到辅助、验证和约束的作用，从而改善TIN的质量和精度。

1. 比较根据DEM生成的等值线与原等值线数据。
2. 平滑度：DEM生成的等值线可能相对于原等值线数据具有更平滑的特征。这是因为DEM是基于离散的高程点插值得到的，生成的等值线可能会在一定程度上平滑地连接高程点。
3. 细节损失：由于DEM的分辨率和插值方法等因素的限制，DEM生成的等值线可能会丢失一些细节。原等值线数据可能包含了更多的地形特征和细微的地形变化。
4. 数据缺失：如果原等值线数据在某些区域缺失或不完整，DEM生成的等值线可能会填补这些数据缺失区域，从而提供更完整的地形表示。
5. 地形剖面是如何生成的？

地形剖面是地形表面在特定方向上的高程变化曲线。它可以用来可视化地形的垂直变化。生成地形剖面可以通过以下步骤实现：

首先，确定生成剖面的路径或剖面线，使用DEM或其他高程数据源，提取剖面线上的高程数据。在剖面线上选择一系列离散的点，并获取每个点的高程值。然后绘制剖面线上的高程变化曲线。横轴表示剖面线上的位置或距离，纵轴表示对应位置上的高程值。通过将这些点连接起来，即可绘制出地形剖面。