第十章 地统计分析

地统计分析方法被广泛应用许多领域，已成为空间统计学的一个重要分支。很长时间以来，地统计分析一直没能很好的和 GIS 分析模型紧密结合在一起，这成为 GIS 软件一大遗憾。ArcGIS 地统计分析模块在地统计学与 GIS 之间架起了一座桥梁，使得复杂的地统计方法可以在软件中轻易实现，体现了以人为本、可视化发展的趋势。这种结合具有重要的开创性意义，通过测定预测表面的统计误差，GIS 应用人员首次能够对预测表面的模型质量进行量化。本章主要通过对地统计分析的概念介绍，逐步引导读者在 ARCGIS 中如何应用地统计分析解决实际问题。

10.1 地统计基础

10.1.1 基本原理

地统计（Geostatistics）又称地质统计，是在法国著名统计学家 G. Matheron 大量理论研究的基础上逐渐形成的一门新的统计学分支。它是以区域化变量为基础，借助变异函数，研究既具有随机性又具有结构性，或空间相关性和依赖性的自然现象的一门科学。凡是与空间数据的结构性和随机性，或空间相关性和依赖性，或空间格局与变异有关的研究，并对这些数据进行最优无偏内插估计，或模拟这些数据的离散性、波动性时，皆可应用地统计学的理论与方法。

地统计学与经典统计学的共同之处在于：它们都是在大量采样的基础上，通过对样本属性值的频率分布或均值、方差关系及其相应规则的分析，确定其空间分布格局与相关关系。但地统计学区别于经典统计学的最大特点即是：地统计学既考虑到样本值的大小，又重视样本空间位置及样本间的距离，弥补了经典统计学忽略空间方位的缺陷。

地统计分析理论基础包括前提假设、区域化变量、变异分析和空间估值。

1. 前提假设

（1）随机过程

与经典统计学相同的是，地统计学也是在大量样本的基础上，通过分析样本间的规律，探索其分布规律，并进行预测。地统计学认为研究区域中的所有样本值都是随机过程的结果，即所有样本值都不是相互独立的，它们是遵循一定的内在规律的。因此地统计学就是要揭示这种内在规律，并进行预测。

（2）正态分布

在统计学分析中，假设大量样本是服从正态分布的，地统计学也不例外。在获得数据后首先应对数据进行分析，若不符合正态分布的假设，应对数据进行变换，转为符合正态

分布的形式，并尽量选取可逆的变换形式。

（3）平稳性

对于统计学而言，重复的观点是其理论基础。统计学认为，从大量重复的观察中可以进行预测和估计，并可以了解估计的变化性和不确定性。对于大部分的空间数据而言，平稳性的假设是合理的。这其中包括两种平稳性：一是均值平稳，即假设均值是不变的并且与位置无关；另一类是与协方差函数有关的二阶平稳和与半变异函数有关的内蕴平稳。二阶平稳是假设具有相同的距离和方向的任意两点的协方差是相同的，协方差只与这两点的值相关而与它们的位置无关。内蕴平稳假设是指具有相同距离和方向的任意两点的方差（即变异函数）是相同的。二阶平稳和内蕴平稳都是为了获得基本重复规律而作的基本假设，通过协方差函数和变异函数可以进行预测和估计预测结果的不确定性。

2. 区域化变量

当一个变量呈现一定的空间分布时，称之为区域化变量，它反映了区域内的某种特征或现象。区域化变量与一般的随机变量不同之处在于，一般的随机变量取值符合一定的概率分布，而区域化变量根据区域内位置的不同而取不同的值。而当区域化变量在区域内确定位置取值时，表现为一般的随机变量，也就是说，它是与位置有关的随机变量。在实际分析中，常采用抽样的方式获得区域化变量在某个区域内的值，即此时区域化变量表现为空间点函数：

$$Z(x) = Z(x_u, x_v, x_w) \qquad (10.1)$$

根据其定义，区域化变量具有两个显著特征：即随机性和结构性。首先，区域化变量是一个随机变量，它具有局部的、随机的、异常的特征；其次，区域化变量具有一定的结构特点，即变量在点 x 与偏离空间距离为 h 的点 x+h 处的值 Z(x)和 Z(x+h)具有某种程度的相似性，即自相关性，这种自相关性的程度依赖于两点间的距离 h 及变量特征。除此之外，区域化变量还具有空间局限性（即这种结构性表现为一定范围内）、不同程度的连续性和不同程度的各向异性（即各个方向表现出的自相关性有所区别）等特征。

3. 变异分析

（1）协方差函数

协方差又称半方差，表示两随机变量之间的差异。在概率论中，随机变量 X 与 Y 的协方差定义为：

$$Cov(X,Y) = E[(X - E(\mathrm{X}))(Y - E(\mathrm{Y})] \qquad (10.2)$$

借鉴上式，地统计学中的协方差函数可表示为：

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - \overline{Z}(x_i)][Z(x_i + h) - \overline{Z}(x_i + h)] \qquad (10.3)$$

其中，Z(x)为区域化随机变量，并满足二阶平稳假设，即随机变量 Z(x)的空间分布规律不因位移而改变；h 为两样本点空间分隔距离；$Z(x_i)$为 Z(x)在空间点 x_i 处的样本值；

$Z(x_i+h)$是 Z(x)在x_i处距离偏离 h 的样本值[i=1,2,…,N(h)]；N(h)是分隔距离为 h 时的样本点对总数；$\overline{Z}(x_i)$和$\overline{Z}(x_i+h)$分别为$Z(x_i)$和$Z(x_i+h)$的样本平均数，即：

$$\overline{Z}(x_i)=\frac{1}{n}\sum_{i=1}^{n}Z(x_i) \tag{10.4}$$

$$\overline{Z}(x_i+h)=\frac{1}{n}\sum_{i=1}^{n}Z(x_i+h) \tag{10.5}$$

上式中，n 为样本单元数。一般情况下，$\overline{Z}(x_i)\neq\overline{Z}(x_i+h)$（特殊情况下可以认为近似相等）。

（2）半变异函数

半变异函数又称半变差函数、半变异矩，是地统计分析的特有函数。区域化变量 Z(x)在点 *x* 和 *x+h* 处的值 *Z(x)*与 Z(x+h)差的方差的一半称为区域化变量 Z(x)的半变异函数，记为 r(h)，2r(h)称为变异函数。

根据定义有：

$$r(x,h)=\frac{1}{2}Var[Z(x)-Z(x+h)] \tag{10.6}$$

即

$$r(x,h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2-\frac{1}{2}\{E[Z(x)]-E[Z(x+h)]\}^2 \tag{10.7}$$

区域化变量 Z(x)满足二阶平稳假设，因此对于任意的 h 有：

$$E[Z(x+h)]=E[Z(x)] \tag{10.8}$$

因此，半变异函数可改写为：

$$r(x,h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2 \tag{10.9}$$

由上式可知，半变异函数依赖于自变量 x 和 h，当半变异函数 *r(x, h)*仅仅依赖于距离 h 而与位置 x 无关时，*r(x, h)*可改写为 *r(x)*，即：

$$r(h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2 \tag{10.10}$$

具体表示为：

$$r(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i)-Z(x_i+h)]^2 \tag{10.11}$$

各变量的含义同前。也有将 r(h)称为变异函数，两者使用上不引起本质上的差别。

（3）变异分析

半变异函数和协方差函数把统计相关系数的大小作为一个距离的函数，是地理学相近相似定理定量量化。图 10.1 和图 10.2 为一典型的半变异函数图和其对应的协方差函数图。

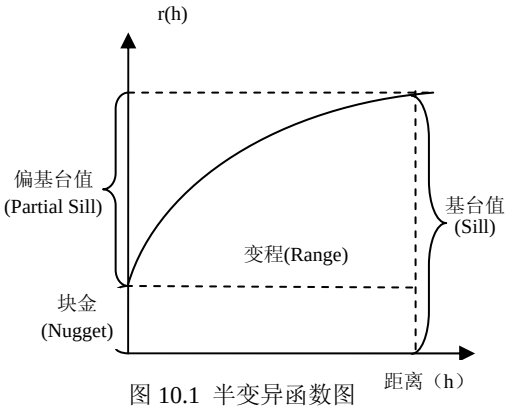


图 10.1 半变异函数图

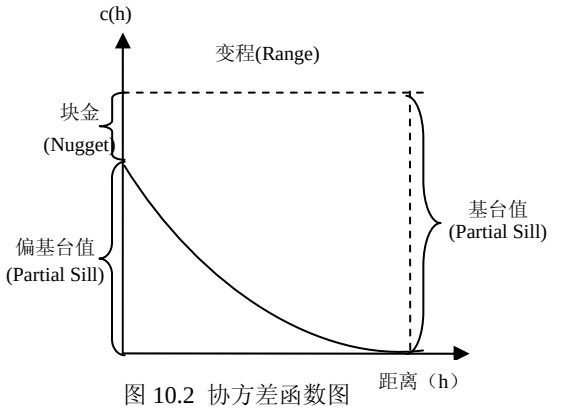


图 10.2 协方差函数图

图 10.1 和图 10.2 显示，半变异值的变化随着距离的加大而增加，协方差随着距离的加大而减小。这主要是由于半变异函数和协方差函数都是事物空间相关系数的表现，当两事物彼此距离较小时，它们是相似的，因此协方差值较大，而半变异值较小；反之，协方差值较小，而半变异值较大。此外，协方差函数和半变异函数随着距离的加大基本呈反向变化特征，它们之间的近似关系表达式为：

$$r(h) = sill - C(h) \qquad (10.12)$$

半变异函数曲线图和协方差函数曲线反映了一个采样点与其相邻采样点的空间关系。此外，它们对异常采样点具有很好的探测作用，在 ArcGIS 地统计分析模块中可以使用两者的任意一个，一般采用半变异函数。在半变异曲线图中有两个非常重要的点：间隔为 0 时的点和半变异函数趋近平稳时的拐点，由这两个点产生四个相应的参数：块金值（Nugget）、变程（Range）、基台值（Sill）、偏基台值（Partial Sill）它们的含义表示如下：

块金值（Nugget）：理论上，当采样点间的距离为 0 时，半变异函数值应为 0，但由于存在测量误差和空间变异，使得两采样点非常接近时，它们的半变异函数值不为 0，即存在块金值。测量误差是仪器内在误差引起的，空间变异是自然现象在一定空间范围内的变化。它们任意一方或两者共同作用产生了块金值。

基台值（Sill）：当采样点间的距离 h 增大时，半变异函数 r(h)从初始的块金值达到一个相对稳定的常数时，该常数值称为基台值。当半变异函数值超过基台值时，即函数值不随采样点间隔距离而改变时，空间相关性不存在。

偏基台值（Partial Sill）：基台值与块金值的差值。

变程（Range）：当半变异函数的取值由初始的块金值达到基台值时，采样点的间隔距离称为变程。变程表示了在某种观测尺度下，空间相关性的作用范围，其大小受观测尺度

的限定。在变程范围内，样点间的距离越小，其相似性，即空间相关性越大。当 h>R 时，区域化变量 Z(x)的空间相关性不存在，即当某点与已知点的距离大于变程时，该点数据不能用于内插或外推。

当限定的样本点间隔过小时，可能出现曲线图上所有 r(h)≈ Nugget，即曲线为一近似平行于横坐标的直线，此时半变异函数表现为纯块金效应。这是由于所限定的样本间隔内，点与点的变化很大，即各个样点是随机的，不具备空间相关性，区域内样点的平均值即是最佳估计值。此时只有增大样本间隔，才能反映出样本间的空间相关性。

空间相关性的强弱可由 Partial_Sill/Sill 来反映，该值越大，空间相关性越强。相应地，Nugget/Sill 称为基底效应，表示样本间的变异特征，该值越大，表示样本间的变异更多得是由随机因素引起的。

4. 空间估值

一个完整的地统计分析过程，或者说空间估值过程，一般为：首先是获取原始数据，检查、分析数据，找寻数据暗含的特点和规律，比如是否为正态分布、有没有趋势效应、各向异性等等；然后选择合适的模型进行表面预测，这其中包括半变异模型的选择和预测模型的选择；最后检验模型是否合理或几种模型进行对比。尽管在 ArcGIS 中利用地统计分析模块完成上述过程非常简单，但是遵循一个结构化处理过程仍很重要，如图 10.3 所示。

（1）数据显示

在 ArcMap 数据视图窗口中添加并显示待分析的数据图层。

（2）数据检查

分析数据集的统计属性，对数据进行深入了解。数据检查内容包括检验数据分布、寻找数据离群值、全局趋势分析、探测空间自相关及方向变异，以及多数据集协变分析。

（3）模型拟合

基于对数据的认识，初步选择一个认为合适的模型创建表面。全面的数据检查有助于选择出合适的模型。

（4）模型诊断

评估模型的输出（表面），了解所选模型对未知值的预测效果。诊断的主要内容包括：①预测的准确性。②模型的有效性。

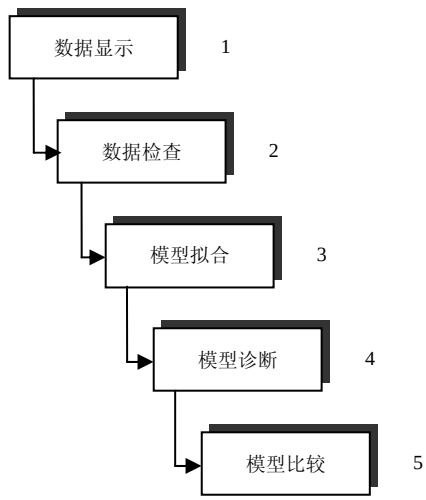


图 10.3 空间估值流程图

（5）模型比较

通过设置不同参数或者选择多个可选模型创建表面，通过对比分析可以确定哪个模型对未知值的预测更好。

10.1.2 克里格插值

克里格插值（Kriging）又称空间局部插值法，是以变异函数理论和结构分析为基础，在有限区域内对区域化变量进行无偏最优估计的一种方法，是地统计学的主要内容之一。南非矿产工程师 D.R.Krige（1951 年）在寻找金矿时首次运用这种方法，法国著名统计学家 G.Matheron 随后将该方法理论化、系统化，并命名为 Kriging，即克里格方法。

克里格方法的适用范围为区域化变量存在空间相关性，即如果变异函数和结构分析的结果表明区域化变量存在空间相关性，则可以利用克里格方法进行内插或外推；否则反之。其实质是利用区域化变量的原始数据和变异函数的结构特点，对未知样点进行线性无偏、最优估计。无偏是指偏差的数学期望为 0，最优是指估计值与实际值之差的平方和最小。也就是说，克里格方法是根据未知样点有限邻域内的若干已知样本点数据，在考虑了样本点的形状、大小和空间方位，与未知样点的相互空间位置关系，以及变异函数提供的结构信息之后，对未知样点进行的一种线性无偏最优估计。

地统计分析的核心就是通过对采样数据的分析、对采样区地理特征的认识选择合适的空间内插方法创建表面。插值方法按其实现的数学原理可以分为两类：一是确定性插值方法，另一类是地统计插值，也就是克里格插值，如图 10.4 所示。

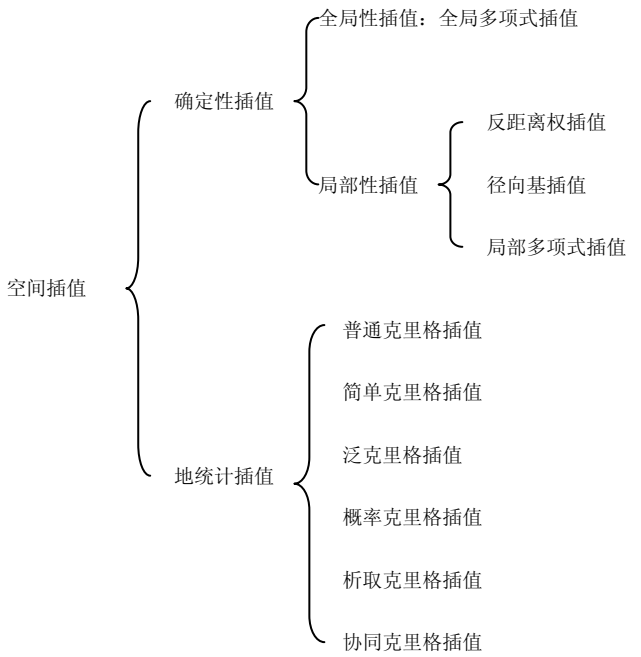


图 10.4 空间插值分类示意图一

确定性插值方法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）由已知样点来创建表面。地统计插值方法（例如克里格法）利用的则是已知样点的统计特性。地统计插值方法不但能够量化已知点之间的空间自相关

性，而且能够解释说明采样点在预测区域范围内的空间分布情况。

确定性插值方法有可以分为两种：即全局性插值方法和局部性插值方法，如图 10.4 所示。全局性插值方法以整个研究区的样点数据集为基础来计算预测值，局部性插值方法则使用一个大研究区域内较小的空间区域内的已知样点来计算预测值。

克里格方法与反距离权插值方法有些类似，两者都通过对已知样本点赋权重来求得未知样点的值，可统一表示为：

$$Z(x_0)=\sum_{i=1}^{n}\lambda_i Z(x_i) \qquad (10.13)$$

式中，$Z(x_0)$为未知样点的值，$Z(x_i)$为未知样点周围的已知样本点的值，λ_i为第i个已知样本点对未知样点的权重，n为已知样本点的个数。

不同的是，在赋权重时，反距离权插值方法只考虑已知样本点与未知样点的距离远近，而克里格方法不仅考虑距离，而且通过变异函数和结构分析，考虑了已知样本点的空间分布及与未知样点的空间方位关系。

空间插值方法根据是否能保证创建的表面经过所有的采样点，又可以分为精确性插值和非精确性插值（图 10.5）。精确性插值法预测值在样点处的值与实测值相等，非精确性插值法预测值在样点处的值与实测值一般不会相等。使用非精确性插值法可以避免在输出表面上出现明显的波峰或波谷。反距离权插值和径向基插值属于精确性插值方法，而全局多项式插值、局部多项式插值，以及克里格插值都属于非精确性插值方法。

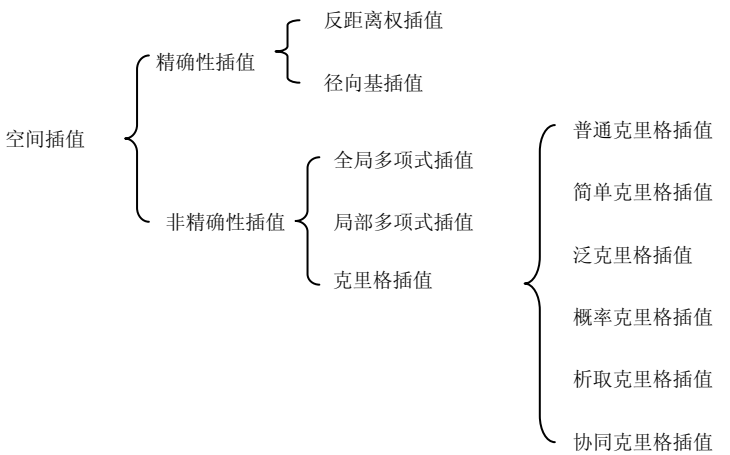


图 10.5 空间插值分类示意图二

10.1.3 ArcGIS 地统计分析

1. 功能模块介绍

ArcGIS 地统计分析模块主要由三个功能模块组成，探索性数据分析（Explore Data）、地统计分析向导（Geostatistical Wizard），以及生成数据子集（Create Subsets）。利用这些基本功能模块，可以方便的完成多种地统计分析，创建完善的专题地图（表面预测）。

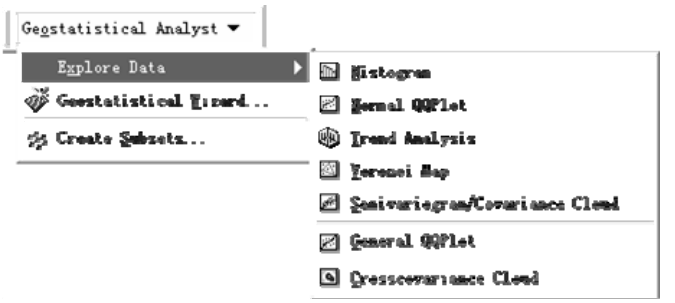


图 10.6 地统计模块菜单

（1）探索性数据分析（Explore Data）

数据分析工具可以让用户更全面地了解所使用的数据，以便于选取合适的参数及方法。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 Histogram(直方图)、Voronoi Map(Voronoi 地图)、Normal QQPlot(正态 QQPlot 分布图)、General QQPlot(普通 QQPlot 分布图)、Trend Analysis(趋势分析)、Semivariogram/Covariance Cloud(半变异/协方差函数云)、Crosscovariance Cloud(正交协方差函数云)。这些探索性空间数据分析工具对数据从不同的视图，提供用户用多种方式检测空间数据。

（2）地统计分析向导（Geostatistical Wizard）

地统计分析模块提供了一系列利用已知样点进行内插生成研究对象表面图的内插技术。地统计分析向导通过完善的图形用户界面，引导用户逐步了解数据、选择内插模型、评估内插精度，完成表面预测（模拟）和误差建模。地统计分析向导能提供用户的主要图形界面包括：

1） 选择内插方法与数据集界面（如图 10.7）

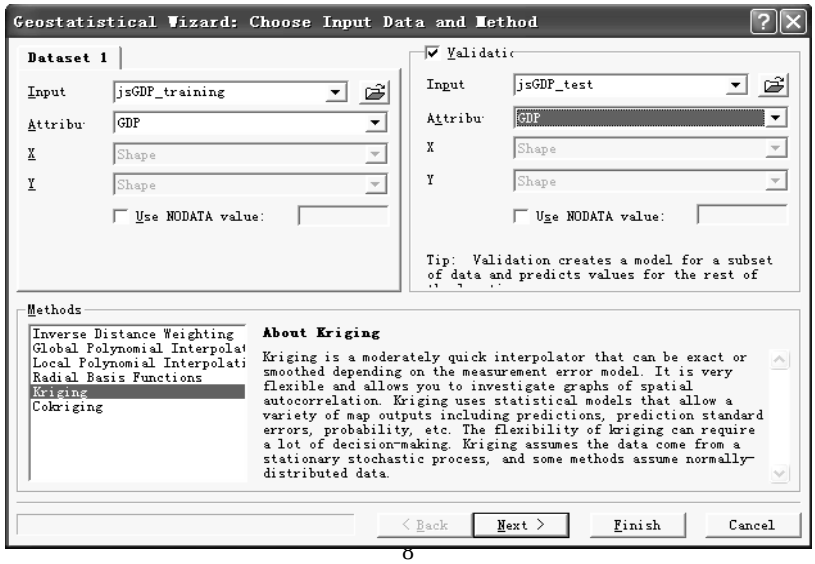


图 10.7 数据及方法选择界面

通过此界面可选择所用数据、实验方法以及检验数据等，在 DataSet1 中，选择输入数据及其属性；在 Validation 中，选择检验数据及其属性；在 Methods 中，选择相应的内插方法。

2） 参数设置界面（图 10.8）

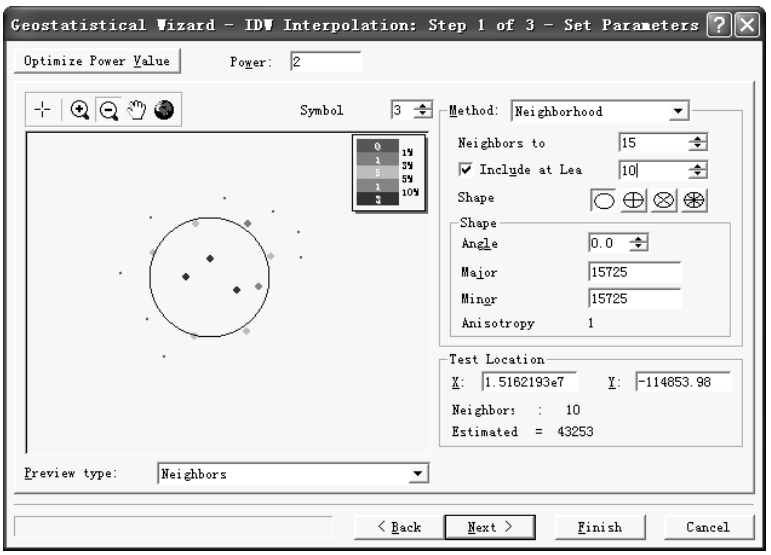


图 10.8 参数设置图

通过此界面设置模型所用的各项参数，如模型所用的幂指数（Power）、最少包含的样点数（Include at Least）等，参数设置应根据实例而定。

3） 精度评定界面（图 10.9）

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，为模型的改进也提供了必要的信息。对于不同的内插方法，上述界面提供的信息一般是不一样的，弹出的对话框的个数、参数设置界面或精度评定界面有时不止一个，应根据具体内容具体对待。

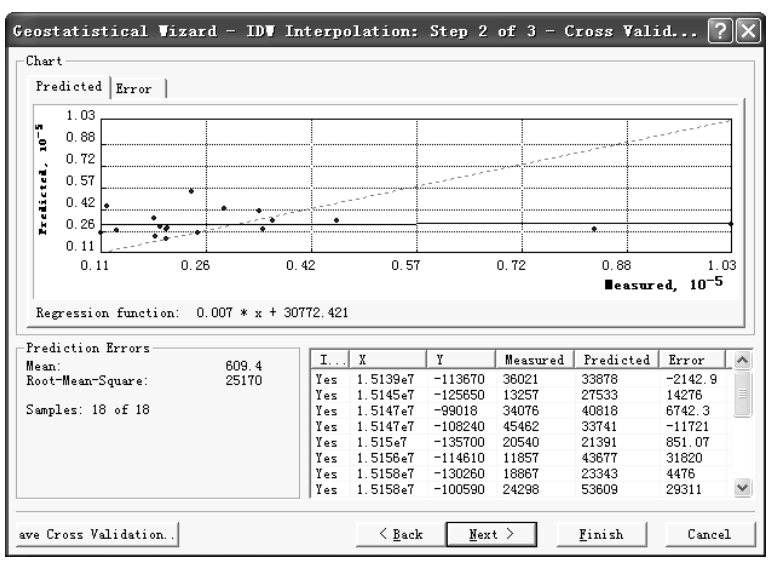


图 10.9 精度评定图

4） 生成数据子集（Create Subsets）

对输出表面质量评价的最严格方法就是将观测值的观测值与预测值进行比较。通常情况下，这种作法无需到研究区采集独立的验证数据集。一个办法就是将原始数据分割成两部分，一部分用来空间结构建模及生成表面，另一部分用来比较和验证预测的质量。Create Subsets 对话框（图 10.10）可以让用户生成测试和训练数据集。

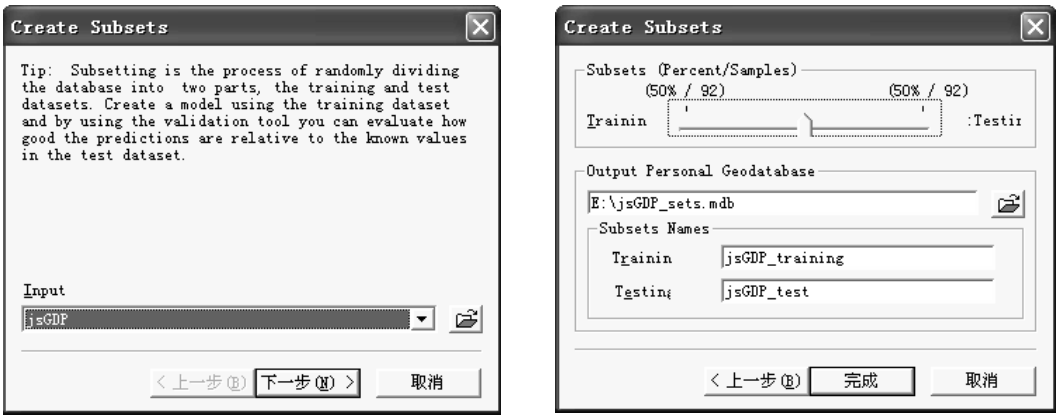


图 10.10 生成数据子集图

2. Geostatistical Analyst 的启动

（1）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（2）单击 Geostatistical Analyst 下的 Explore Data 菜单。

10.2 探索性数据分析

探索性数据分析是利用 ArcGIS 提供的一系列图形工具和适用于数据的插值方法。可以确定统计数据属性，探测数据分布、全局和局部异常值（过大值或过小值）、寻求全局的变化趋势、研究空间自相关、和理解多种数据集之间相关性。探索性空间数据分析能让用户更深入了解数据，认识研究对象，从而对与其数据相关的问题做出更好的决策。

10.2.1 数据分析工具

1. 刷光（Brushing）与链接（Linking）

刷光指在 ArcMap 数据视图或某个 ESDA 工具中选取对象，被选择的对象高亮度显示。链接指在 ArcMap 数据视图或某个 ESDA 视图中的进行选取对象操作，则在所有视图中被选取对象均会执行刷光操作。如在下面章节将要叙述的探索性数据分析过程中，当某些 ESDA 工具（如：直方图、Voronoi 地图、QQplot 图，以及趋势分析）中执行刷光时，ArcMap 数据视图中相应的样点均会被高亮度显示。当在半变异/协方差函数云中刷光时，ArcMap

数据视图中相应的样点对及每对之间的连线均被高亮度显示。反之，当样点对在 ArcMap 数据视图中被选中，在半变异/协方差函数云中相应的点也将高亮度显示。

2. 直方图

直方图指对采样数据按一定的分级方案（等间隔分级、标准差分，等等）进行分级，统计采样点落入各个级别中的个数或占总采样数的百分比，并通过条带图或柱状图表现出来。直方图可以直观的反映采样数据分布特征、总体规律，可以用来检验数据分布和寻找数据离群值。

在 ArcGIS 9 中，可以方便的提取采样点数据的直方图，基本步骤为：

（1）在 ArcMap 中加载地统计数据点图层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Histogram 命令。

（3）设置相关参数，生成直方图。

1） Bars：直方图条带个数，也就是分级级数。

2） Translation：数据变换方式。

3） None：对原始采样数据的值不作变换，直接生成直方图。

4） Log：首先对原始采样数据取对数，再生成直方图。

5） Box-cox：首先对原始采样数据进行博克斯-考克斯变换（也称幂变换），再生成直方图。

6） Layer：当前正在分析的数据图层。

7） Attribute：生成直方图的属性字段。

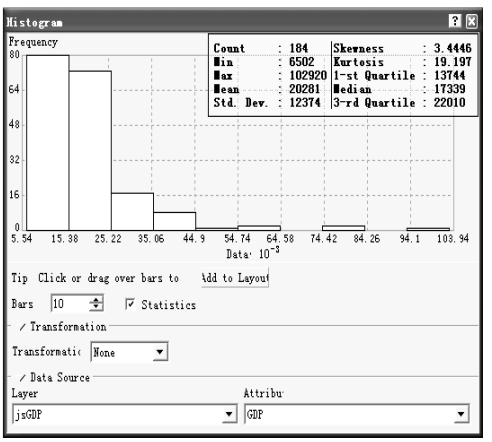


图 a 原始数据直方图

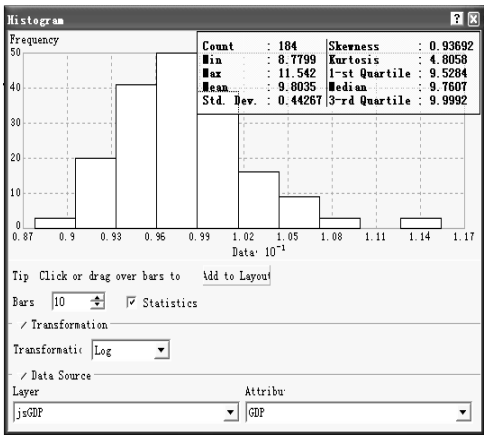


图 b 经变换直方图

图 10.11 直方图示意图

从图 a 和图 b 的对比分析可以看出，该地区 GDP 原始数据并不服从正态分布，经过对数变换处理，分布具有明显的对数分布特征。并且在最右侧有一个明显的离群值。

在直方图右上方的小视窗中，显示了一些基本统计信息，包括：个数（Count）、最小值（Min）、最大值（Max）、平均值（Mean）、标准差（Std. Dev.）、峰度（Kurtosis）、偏态（Skewness）、1/4 分位数（1-st Quartile）、中数（Median）、3/4 分位数（3-rd Quartile），通

过这些信息可以对数据有个初步的了解。

四分位数：如果将N个数值由小至大排列，第$\frac{1}{4}N$个数就是第一个四分位数，通常以Q_1来表示；第$\frac{1}{4}N$个数就是第二个四分位数（Q_2），即中位数；第$\frac{1}{4}N$个数就是第三个四分位数（Q_3）。四分位距即为：$Q= Q_3- Q_1$，它将极端的前 1/4 和后 1/4 去除，而利用第三个与第一个分位数的差距来表示分散情形，因此避免了极端值的影响。但它需要将数据由小到大排序，且没有利用全部数据。

峰度（Kurtosis）：用于描述数据分布高度的指标，正态分布的峰度等于 0。如果数据的峰度大于 0，那么该数据的分布就会比正态分布高耸且狭窄，此时数据比正态分布集中于平均数附近。反之，如果峰度小于 0，数据的分布就比正态分布平坦且宽阔，此时数据比正态分布分散。

偏态（Skewness）：用于描述数据分布左右对称性的指标，正态分布的偏态等于 0。如果数据的直方图向右延伸，即大部分的数据集中于左边，则偏态大于 0，称为正偏态或右偏态。如果数据的直方图向左延伸，即大部分的数据集中于右边，则偏态小于 0，称为负偏态或左偏态。

如图 10.11，由原始数据的直方图上可看出，原始数据的分布属于正偏态，经过变换后的数据比正态分布的数据更加集中于平均值附近。

3. Voronoi 地图

Voronoi 地图是由在样点周围形成的一系列多边形组成的。某一样点的 Voronoi 多边形按下述方法生成：多边形内任何位置距这一样点的距离都比该多边形到其它样点的距离要将要近。Voronoi 多边形生成之后，相邻的点就被定义为其 Voronoi 多边形与选择样点的 Voronoi 多边形具有公共边的其它样点。

在 ArcGIS9.0 中生成数据的 Voronoi 地图的基本步骤为：

（1）在 ArcMap 中加载地统计数据点图层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Voronoi Map 命令。

（3）设置参数，生成 Voronoi 图，如图 10.12。

1） Type：分配和计算多边形值的方法。

2） Layer：当前正在分析的数据图层。

3） Atrribute：生成直方图的属性字段。

Voronoi Map 对话框 Type 选项提供了多种分配和计算多边形值的方法：

（1）简化（Simple）：分配到某个多边形单元的值是该多边形单元的方法。

（2）平均（Mean）：分配到某个多边形单元的值是这个单元与其相邻单元的平均值。

（3）模式（Mode）：所有的多边形单元被分为五级区间，分配到某个多边形单元的值是这个单元与其相邻单元的（级发生频率的）模式。

（4）聚类（Cluster）：所有的多边形单元被分配到这五级区间中，如果某个多边形单元的级区间与它的相邻单元的级区间都有不同，这个单元用灰色表示，以区别于其它单

元。

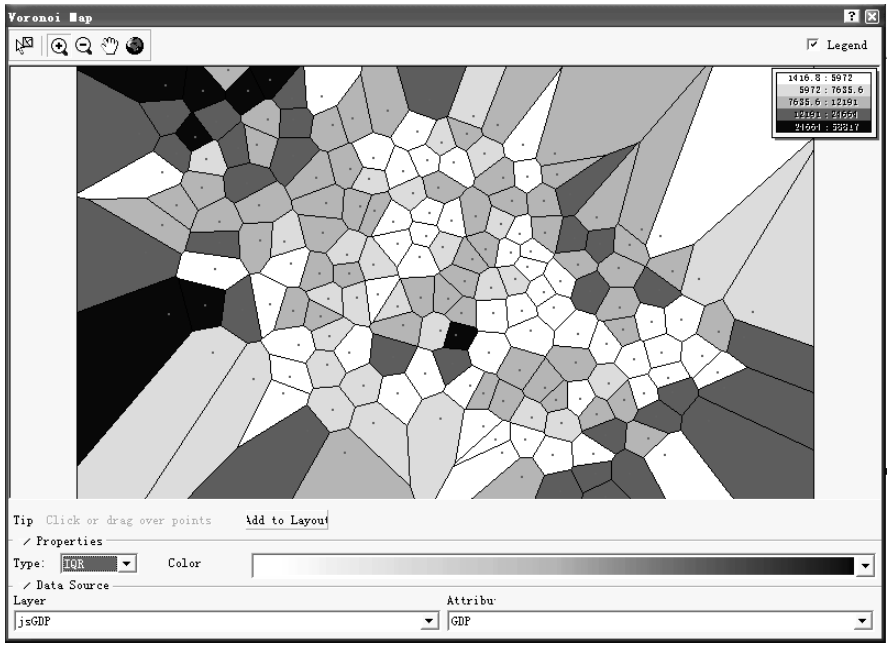


图 10.12　Voronoi 地图

（5）熵（Entropy）：所有单元都根据数据值的自然分组分配到这五级中。分配到某个多边形单元的值是根据该单元和其相邻单元计算出来的熵。

（6）中值（Median）：分配给某多边形的值是根据该单元和其相邻单元的频率分布计算的中值。

（7）标准差（StDev）：分配给某多边形的值是根据该单元和其相邻单元计算出的标准差。

（8）四分位数间间隔（IQR）：第一和第三四分位数是根据某单元和其相邻单元的频率分布计算出来的。分配给某多边形单元的值是用第三四分位数减去第一四分位数得到的差。

Voronoi 地图可以了解到每个采样点控制的区域的范围，也可以体现出每个采样点对区域内插的重要性。利用 Voronoi 地图中就可以找出一些对区域内插作用不大且可能影响内插精度的采样点值，可以将它剔除。用聚类和熵的方法生成的 Vonoroi 图可也可用来帮助识别可能的离群值。自然界中，距离相近的事物比距离远的事物具有更大的相似性。熵值是量度相邻单元相异性的一个指标。因此，局部离群值可以通过高熵值的区域识别出来。同样，一般认为某个特定单元的值至少应与它周围单元中的某一个的值相近。因此聚类方法也能将那些与它们周围单元不相同的单元识别出来。

4. QQPlot 分布图

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，如果数据越接近一条直线，则它越接近于服从正态分布。

（1）正态 QQPlot（Normal QQPlot）分布图

正态 QQPlot 分布图主要用来评估具有 n 个值的单变量样本数据是否服从正态分布。构建正态 QQPlot 分布图的通用过程为（图 10.13）：

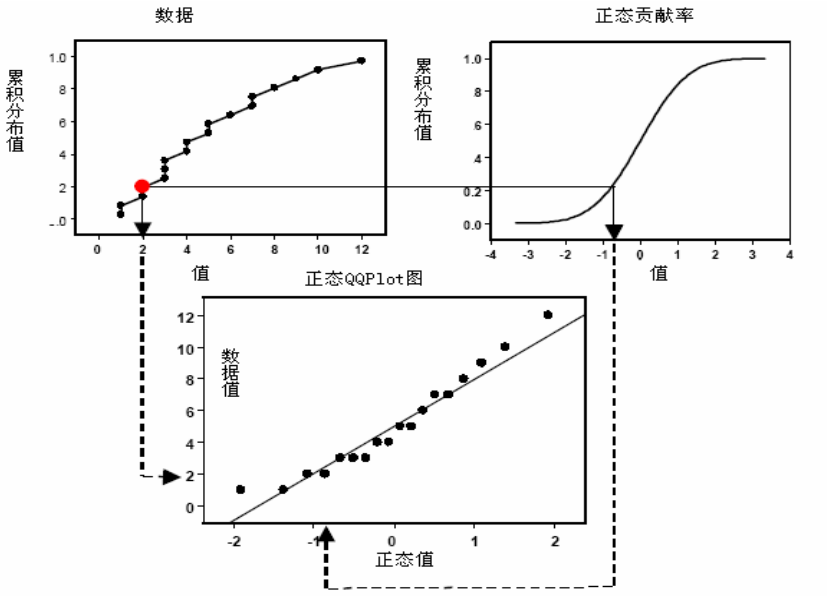


图 10.13 正态 QQPlot 示意图

1） 首先对采样值进行排序；

2） 计算出每个排序后的数据的累积值（低于该值的数据的百分比）；

3） 绘制累积值分布图；

4） 在累积值之间使用线性内插技术，构建一个与其具有相同累积分布的理论正态分布图，求出对应的正态分布值。

5） 以横轴为理论正态分布值，竖轴为采样点值，绘制样本数据相对于其标准正态分布值的散点图，此图为样本数据的正态 QQPlot 分布图。

如果采样数据服从正态分布，其正态 QQPlot 分布图中采样点分布应该是一条的直线。如果有个别采样点偏离直线太多，那么这些采样点可能是一些异常点，应该对其进行检验。此外，如果在正态 QQ 图中数据没有显示出正态分布，那么就有必要在应用某种克里格插值法之前将数据进行转换，使之服从正态分布。

在 ArcGIS9.0 中生成数据的正态 QQPlot 分布图的主要步骤为：

1） 在 ArcMap 中加载地统计数据点图层。

2） 单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Normal QQPlot 的命令。

3） 设置参数，生成 Normal QQPlot 图。

A Translation：数据变换方式。

a) None：对原始采样数据的值不作变换，直接生成直方图。

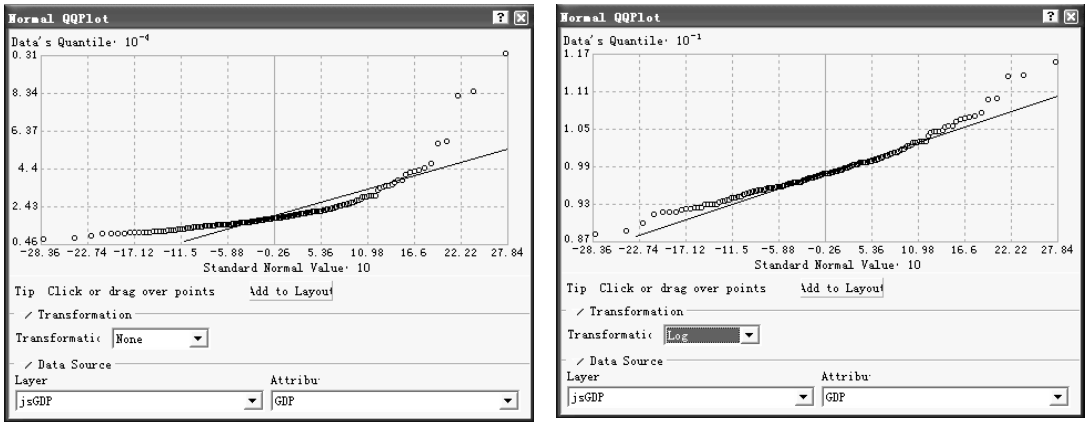
b) Log：首先对原始采样数据取对数，再生成直方图。

c) Box-cox：首先对原始采样数据进行博克斯-考克斯变换（也称幂变换），

再生成正态 QQPlot 图。

B　Layer：当前正在分析的数据图层。

C　Atrribute：生成 Normal QQPlot 分布图使用的属性字段。



a. 原始数据正态 QQPlot 图　　b. 经 log 变换后的正态 QQPlot 图

图 10.14 正态 QQPlot 分布图

从图 a 可以看出，该地区 GDP 的采样数据不符合正态分布，但对其进行对数变换处理后（图 b），对数变换数据近似符合正态分布。仅从采样点值的分布看，在小值区域和大值区域，存在个别离群点值。

（2）构建一个普通 QQPlot 分布图

普通 QQPlot（General QQPlot）分布图用来评估两个数据集的分布的相似性。普通 QQPlot 分布图通过两个数据集中具有相同累积分布值作图来生成，如图所示。累积分布值

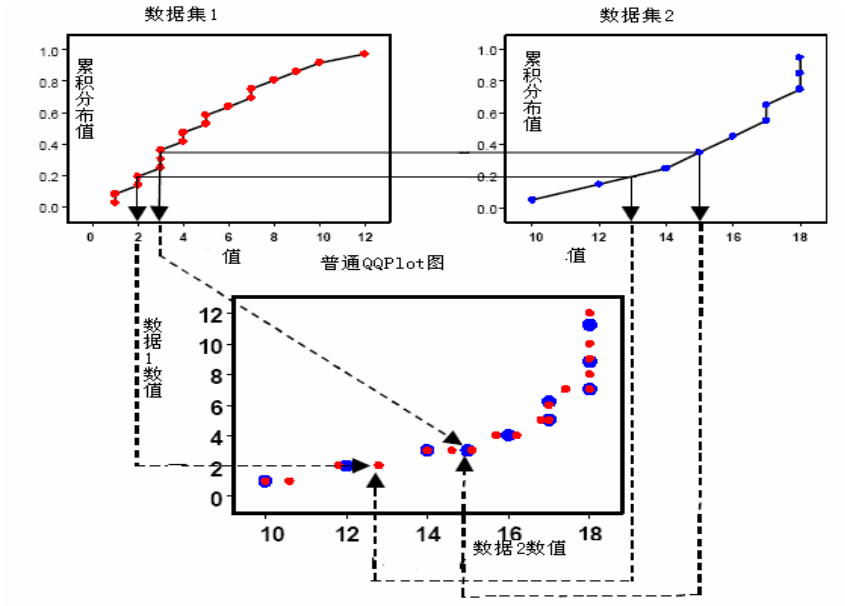


图 10.15　普通 QQPlot 示意图

的作法参阅正态 QQPlot 分布图内容。

在 ArcGIS9.0 中生成数据的普通 QQPlot 分布图的主要步骤为：

1） 在 ArcMap 中加载地统计数据点图层。

2） 单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 General QQPlot 命令。

3） 设置参数，生成 General QQPlot 图。

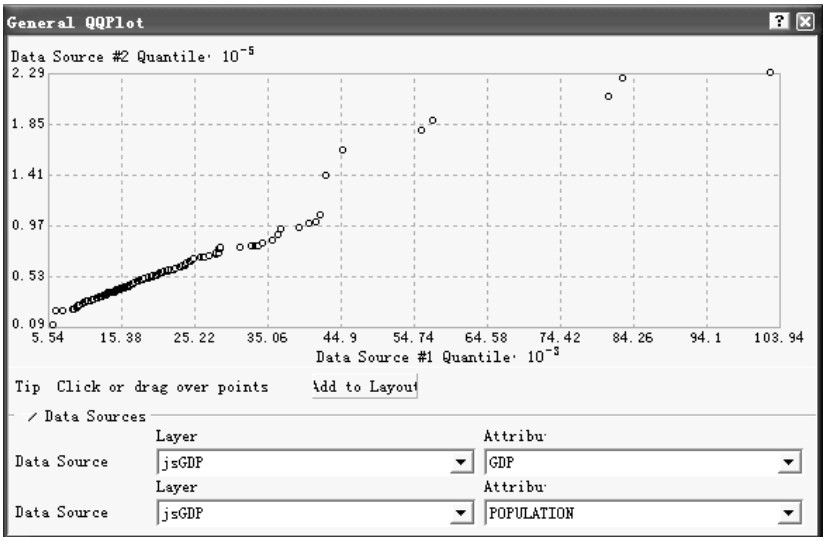


图 10.16 普通 QQPlot 分布图

普通 QQPlot 图揭示了两个物体（变量）之间的相关关系，如果在 QQPlot 图中曲线呈直线，说明两物体呈一种线性关系，可以用一个一元一次方程式来拟合。如果 QQPlot 图中曲线呈抛物线，说明两物体的关系可以用一个二元多项式来拟合。

5. 趋势分析工具

趋势分析工提供用户研究区采样点转换为以感兴趣的属性值为高度的三维透视图，允许用户从不同视角分析采样数据集的全局趋势。

在 ArcGIS9.0 中对使用趋势分析的主要步骤有：

（1）在 ArcMap 中加载地统计数据点图层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Trend Analysis 的命令。

（3）检查中数据层层名对话框(Layer)的设置是否正确，在字段属性对话框(Attribute)选择参见趋势分析的字段名称。

（4）Rotate 选择 Locations，通过其右侧旋转水平螺旋按钮可以任意改变投影视角；选择 Graph，通过其右侧旋转按钮可以任意改变整个投影视图的视角，即同时投影图已经固定，只是用户在不同水平方向观察它，此时试图右侧上下螺旋按钮可以动态改变透视图的仰角（观测者与视图的相对高低位置）。

（5）趋势分析工具 Graph Options 中提供了多种显示功能：

（6）Grid：调整 X、Y、Z 方向上的格网数，以及格网线条的粗细。

（7）Projected Data：选择落在 X、Y、Z 方向格网上投影点的颜色及点的大小。

（8）Trend on Projection：选择东西、南北方向趋势面投影线用多项式拟合的次数以及趋势线的粗细。

（9）Sticks：选择点到 X、Y 平面垂线的粗细和颜色。

（10） Axes：选择 X、Y、Z 坐标轴的粗细和颜色。

（11） Input Data Points：选择高程点的颜色和大小。

趋势分析图中的每一根竖棒代表了一个数据点的值（高度）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。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。如果该线是平直的，则表明没有趋势存在。在图 10.17，在图中，可以看到投影到东西向上的较细的趋势线，从东往西呈阶梯状平滑过渡，而南北方向上，趋势线（较粗的黑色线条）呈 U 形，从中我们可以大致可以得知，此区域的地势为从东往西逐渐下降，南北方向上两边高、中间低的地形。分析出的结果和此区域的真实地形基本吻合。可见趋势分析工具对观察一个物体的空间分布具有简单、直观的优势，还可以找出拟合最好的多项式对区域中的散点进行内插，得到趋势面。

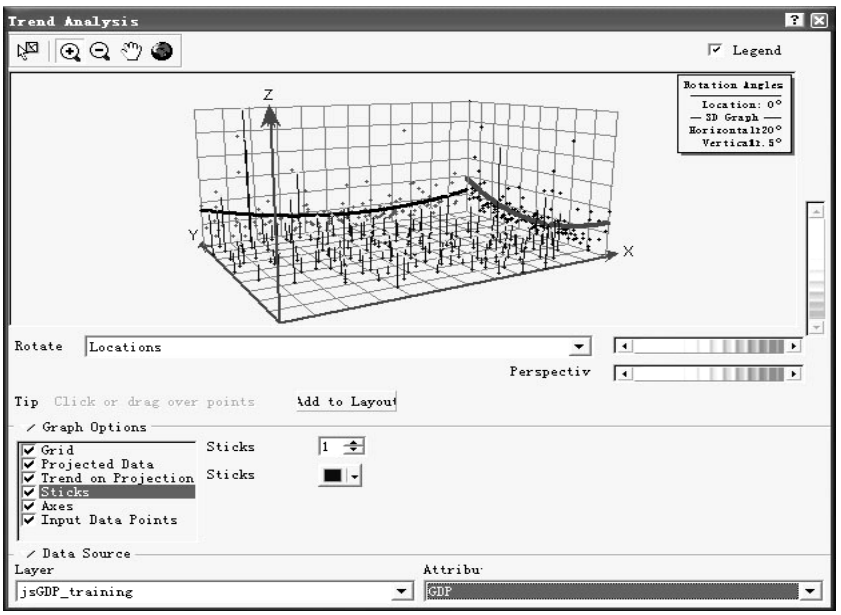


图 10.17 趋势分析图

6. 方差变异分析工具

（1）半变异 / 协方差函数云

半变异 / 协方差函数云表示的是数据集中所有样点对的理论半变异值和协方差，并把它们用两点间距离的函数来表示，用此函数作图来表示。

在 ArcGIS9.0 中生成数据的半变异 / 协方差函数云图主要步骤有：

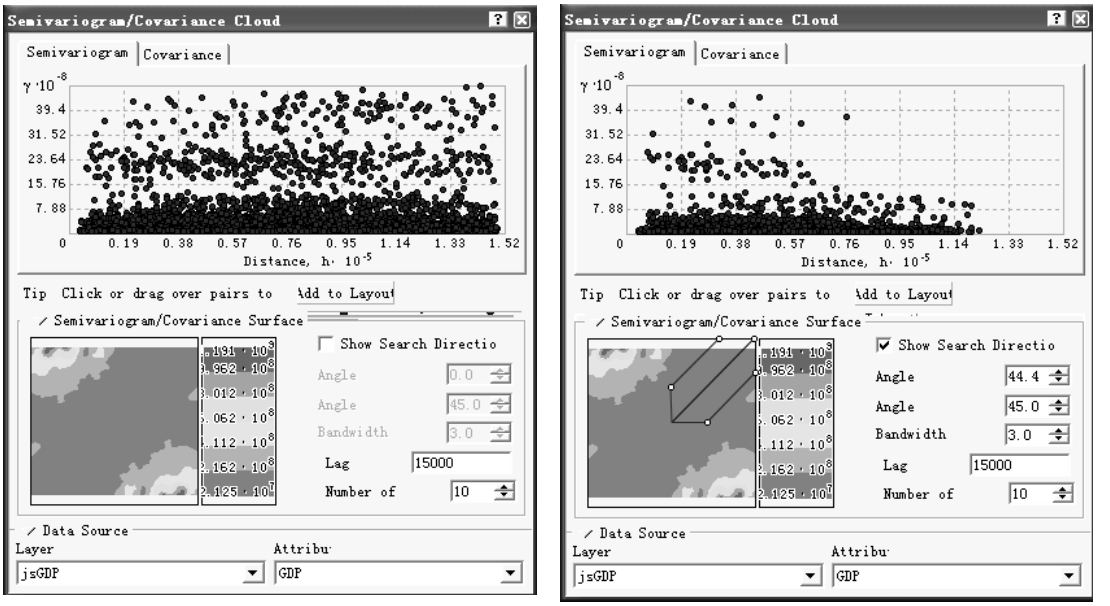
1） 在 ArcMap 中加载地统计数据点图层。

2） 单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Semivariogram/Covariance Cloud 的命令。

3） 检查中数据层层名对话框（Layer）的设置是否正确，在字段属性对话框（Attribute）选择参见趋势分析的字段名称。

4） Lag Size 为最大步长，Number of Lags 为步长分组个数，如图 10.18.a 所示。

5） 如果空间变异具有方向性，可以选择 Show Search Direction（方向搜索），然后点击方向控制条、重设它或改变它的方向来浏览半变异函数云的某个方向子集，如图 10.18.b 所示。



a. 无方向性的半变异 / 协方差函数云图　　b. 有方向性的半变异 / 协方差函数云图

图 10.18 半变异 / 协方差函数云图

（2）正交协方差函数云

正交协方差函数云表示的是两个数据集中所有样点对的理论正交协方差，并把它们用两点间距离的函数来表示。

在 ArcGIS9.0 中生成数据的正交协方差函数云图的主要步骤：

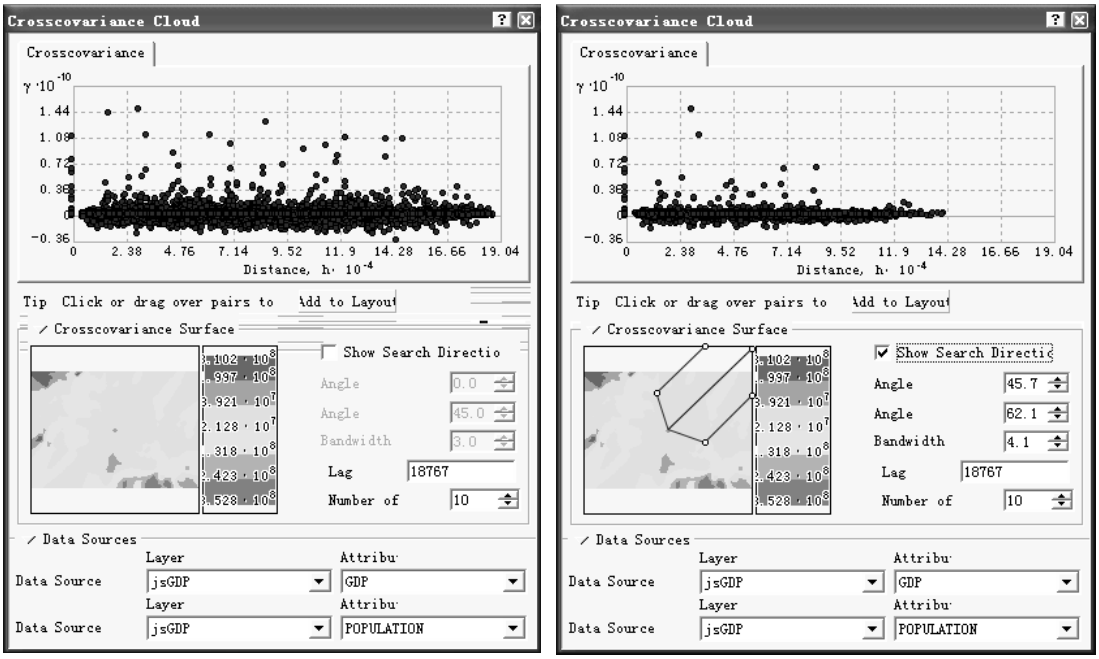
1） 在 ArcMap 中加载地统计数据点图层。

2） 单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Crosscovariance Cloud 的命令。

3） 检查中数据层层名对话框（Layer）的设置是否正确，在字段属性对话框（Attribute）选择参见趋势分析的字段名称。

4） Lag Size 为最大步长，Number of Lags 为步长分组个数。

如果空间变异具有方向性，可以选择 Show Search Direction（方向搜索），然后点击方向控制条、重设它或改变它的方向来浏览半变异函数云的某个方向子集，如图 10.19 所示。具体方向的确定应根据形成该现象的成因及各方向结果的比较。



a. 无方向性的正交协方差函数云图　　b. 有方向性的正交协方差函数云图

图 10.19 正交协方差函数云图

10.2.2 检验数据分布

在地统计分析中，克里格方法是建立在平稳假设的基础上，这种假设在一定程度上要求所有数据值具有相同的变异性。另外，一些克里格插值（如普通克里格法、简单克里格法和泛克里格法等）都假设数据服从正态分布。如果数据不服从正态分布，需要进行一定的数据变换，从而使其服从正态分布。因此，在进行地统计分析前，检验数据分布特征，了解和认识数据具有非常重要的意义。数据的检验可以通过直方图和正态 QQPlot 分布图完成。

1. 用直方图检测数据的分布

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Histogram 命令。

2. 用 Normal QQplot 图检测数据分布

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Normal QQPlot 命令。

如果数据服从正态分布，数据的直方图应该呈钟形曲线，在 QQPlot 图中，数据的分布近似成为一条直线。否则，需要通过直方图和 QQPlot 图工具提供的对数变换（Log）和幂变换（Cox-Box）功能，对数据进行变换处理，尽可能让变换后的数据服从正态分布。

10.2.3 寻找数据离群值

数据离群值分为全局离群值和局部离群值两大类。全局离群值是指对于数据集中所有点来讲，具有很高或很低的值的观测样点。局部离群值值对于整个数据集来讲，观测样点的值处于正常范围，但与其相邻测量点比较，它又偏高或偏低。

离群点的出现有可能就是真实异常值，也可能是由于不正确的测量或记录引起的。如果离群值是真实异常值，这个点可能就是研究和理解这个现象的最重要的点。反之，如果它是由于测量或数据输入的明显错误引起的，在生成表面之前，它们就需要改正或剔除。对于预测表面，离群值可能引起多方面的有害影响，包括影响半变异建模和邻域分析的取值。

离群值的寻找可以通过三种方式实现：

1. 利用直方图查找离群值

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Histogram 命令。

离群值在直方图上表现为孤立存在或被一群显著不同的值包围，如图 10.20 上，直方图上最右边被选中的一个柱状条即是该数据的离群值，相应地，数据点层面上对应的样点也被刷光。但需注意的是，在直方图中孤立存在或被一群显著不同的值包围的样点不一定是离群值。

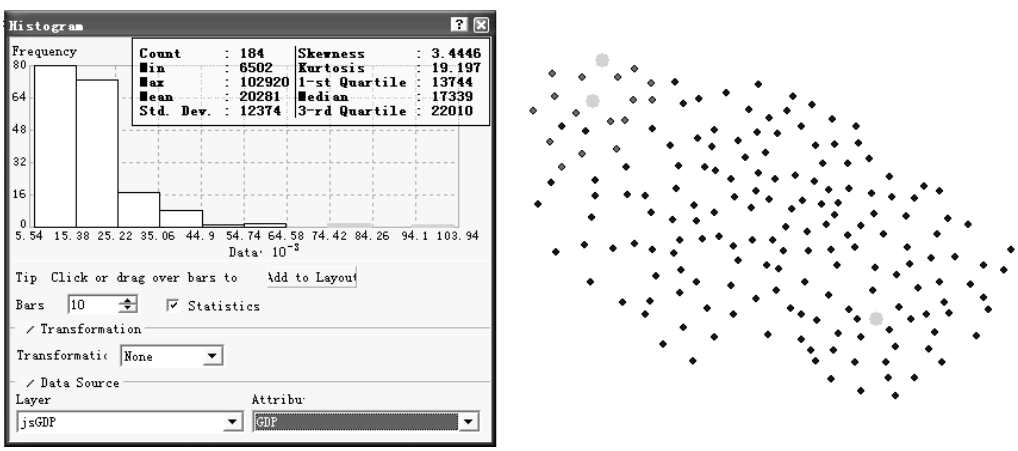


图 10.20 直方图查找离群值图

2. 用半变异/协方差函数云识别离群值

如果数据集中有一个异常高值的离群值，则与这个离群值形成的样点对，无论距离远近，在半变异/协方差函数云图中都具有很高的值。如下图所示，这些点可大致分为上下两层，对于上层的点，无论位于横坐标的左端或右端（即无论距离远近）都具有较高的值。刷光上层的一些点，右图是对应刷光的样点对。可以看到，这些高值都是由同一个离群值的样点对引起的，因此，需要对该点进行剔除或改正。

其具体操作步骤如下：

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Semivariogram/Covariance Cloud 命令。

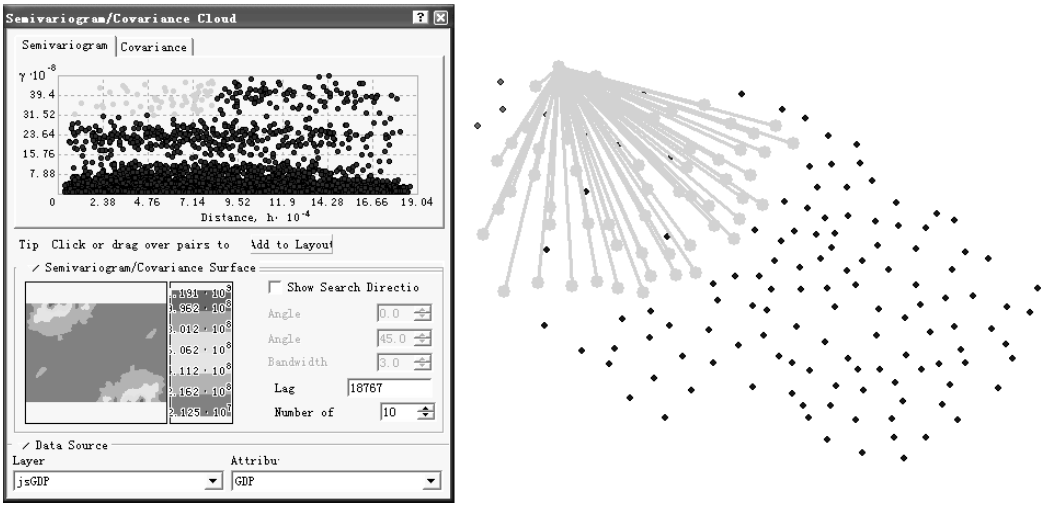


图 10.21 半变异/协方差函数云识别离群值图

3. 用 Voronoi 图查找局部离群值

用聚类和熵的方法生成的 Voronoi 图可用来帮助识别可能的离群值。熵值是量度相邻单元相异性的指标。通常，距离近的事物比距离远的事物具有更大的相似性。因此，局部离群值可以通过高熵值的区域识别出来。同样的原理，聚类方法也可将那些与它们周围单元不相同的单元识别出来。

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Voronoi Map 命令。

（3）选择 Type 类型为 TQR；

（4）选择 Voronoi Map 图中颜色和周围所有邻接面域颜色皆然不同的面域，如图 10.22 中的晕线所示区域，这些面域所代表的点可能就是离群值点。

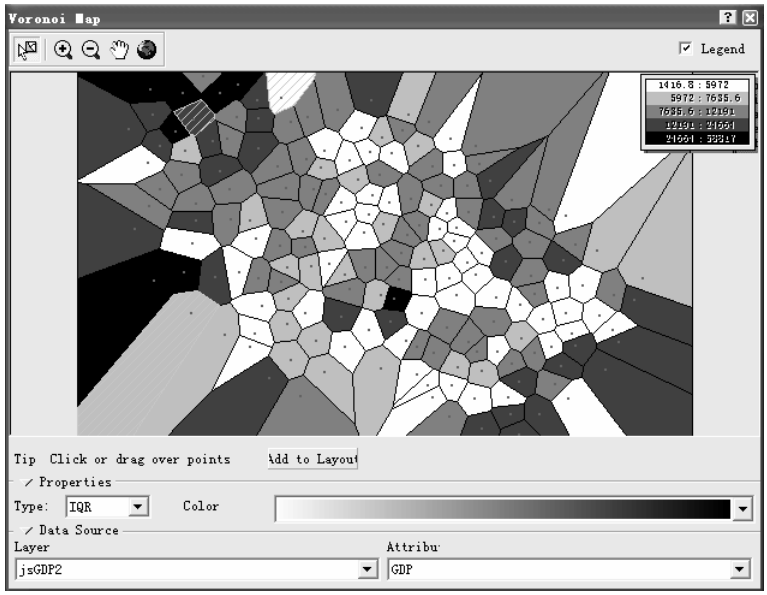


图 10.22　Voronoi 图查找局部离群值图

10.2.4 全局趋势分析

一个表面主要由两部分组成：确定的全局趋势和随机的短程变异。空间趋势反映了空间物体在空间区域上变化的主体特征，它主要揭示了空间物体的总体规律，而忽略局部的变异。趋势面分析是根据空间抽样数据，拟合一个数学曲面，用该数学曲面来反映空间分布的变化情况。它可分为趋势面和偏差两大部分，其中趋势面反映了空间数据总体的变化趋势，受全局性、大范围的因素影响。如果能够准确识别和量化全局趋势，在 ArcGIS 地统计建模中可以方便的剔除全局趋势，从而更能准确地模拟短程随机变异。

在 ArcGIS9.0 中使用趋势分析工具分析全局趋势的具体操作步骤如下：

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Trend Analysis。

在 ArcGIS9.0 中使用趋势分析工具探测全局趋势的关键在于选择合适的透视角度，准确的判定趋势特征。同样的采样数据，透视角度不同，反映的趋势信息也不相同。图 10.23（a）为显示东西方向和南北方向的趋势图。图 10.23（b）逆时针旋转 45 度后，显示东南-西北方向和西南-东北方向的趋势图。

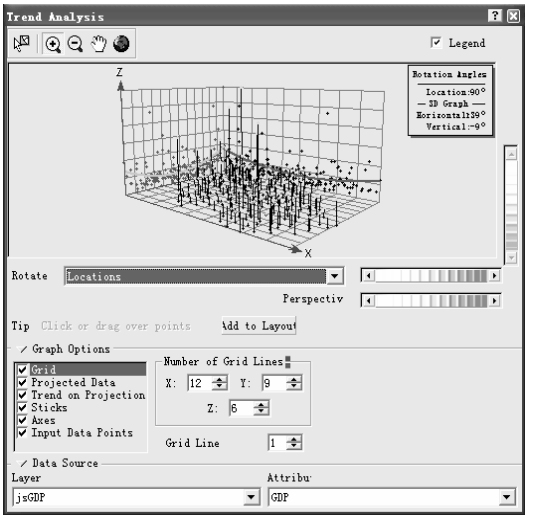


图 10.23（a）

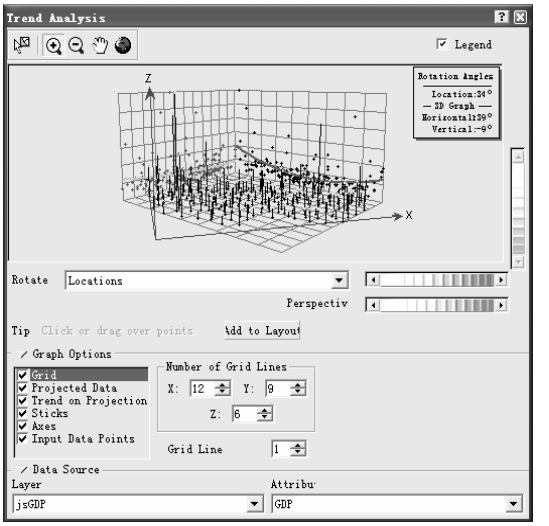


图 10.23（b）

图 10.23　　趋势面分析图

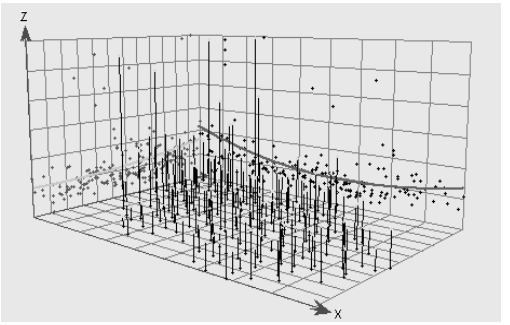


图 10.24（a）

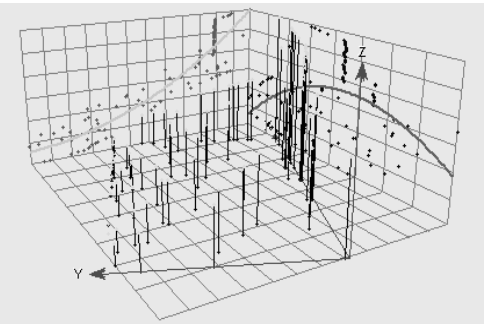


图 10.24（b）

图 10.24　　趋势面分析透视面图

对比分析可以发现，图 10.23（a）显示采样数据在东西方向和南北方向具有微弱的 U 型趋势；图 10.23（b）显示采样数据在东南-西北方向具有明显的 U 型趋势，而在南北方向基本不具有任何趋势。

趋势分析过程中，透视面的选择应尽可能采样数据在透视面上的投影点分布比较集中，通过投影点拟合的趋势方程才具有代表性，才能有效反映采样数据集全局趋势。图 10.24（a）反映的趋势显然要比图 10.24（b）要更为准确。

10.2.5 空间自相关及方向变异

大部分的地理现象都具有空间相关特性，即距离越近的两事物越相似。这一特性也是地统计分析的基础。半变异/协方差函数云图就是这种相似性的定量化表示。

在 ArcGIS9.0 中使用趋势分析工具分析全局趋势的具体操作步骤如下：

（1）在 ArcMap 目录表中点击需要进行数据检测分析的点要素层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Semivariogram/Covariance Cloud。

（3）设置相关参数。

半变异/协方差函数值越小，就越相似。如图 10.25 所示，jsJDP1 中 GDP 采样值具有很强的空间相关性，空间相关半径（变程值）约在 16 附近。图 10.26（a）所示，jsJDP2 中 GDP 采样值在空间基本不具有空间相关性，虽然在左侧有一个明显的突变局势，但它反映的采样点（图 10.26（b）中线段相连接的数据点）的连线距离过于短小，不具有实际意义。

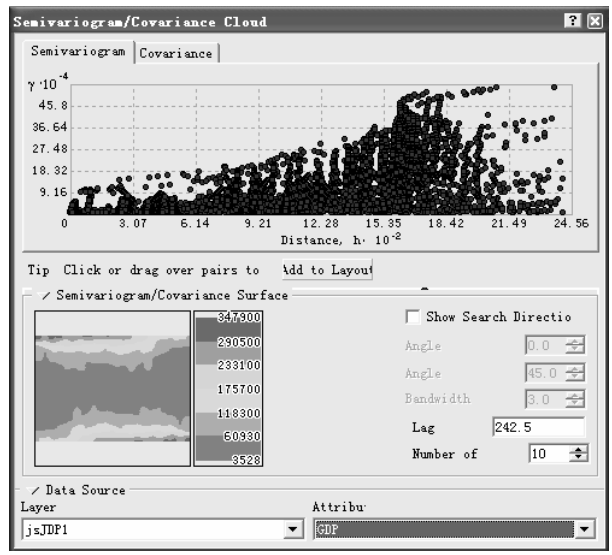


图 10.25　　半变异/协方差函数云

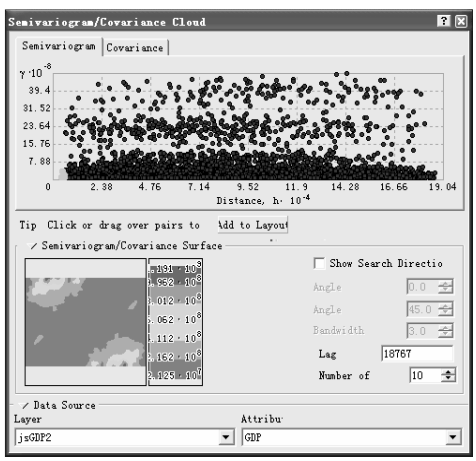


图 10.26（a）

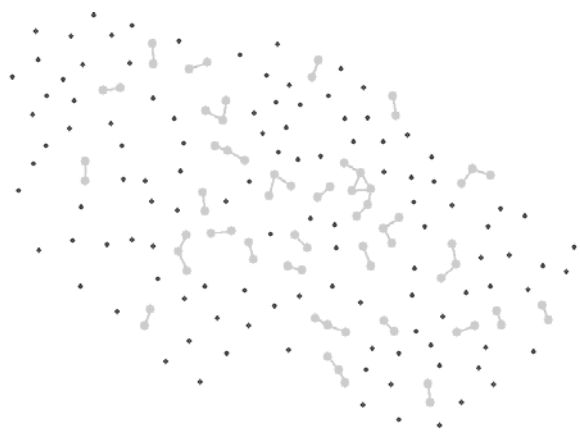


图 10.26（b）

图 10.26　　半变异/协方差函数云对比分析图

空间相关性也可能仅仅与两点间距离有关，这时称为各项同性。在实际应用中，各项异性现象更为普遍，也就是说，当考虑方向影响时，有可能在某个方向距离更远的事物具有更大的相似性，这种现象在半变异和协方差分析中成为方向效应。分析各项异性具有很重要的意义，如果能探测出自相关中方向效应，就可以在半变异或协方差拟合模型中考虑这个因素。

选择 Semivariogram/Covariance Cloud 视图中的 Show Search Direction，动态改变搜索方向和角度，观察半变异/协方差视图中半变异/协方差函数云的变化，确定方向效应。从图 10.27 中 a、b 可以看出，采样数据在西北-东南方向比东西方向更远距离的空间相关性。

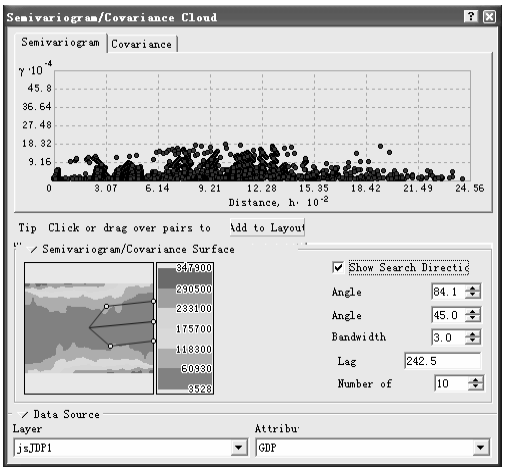


图 10.27（a）

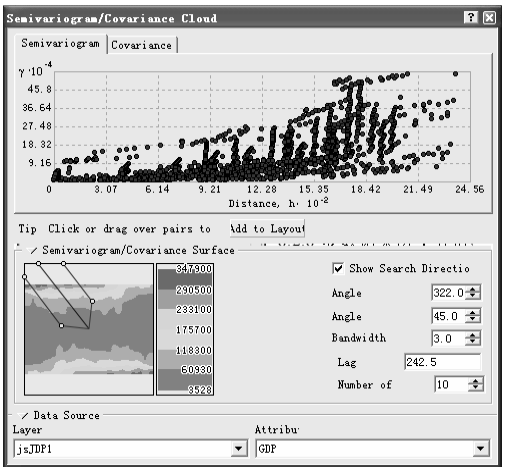


图 10.27（b）

图 10.27　　半变异/协方差函数云方向图

10.2.6 多数据集协变分析

事物都是处于广泛得联系之中，发展发生都会受相互的制约和影响，绝对独立的事物是不存在的。协变分析主要通过分析多因素（数据集）关联特征，在地统计空间分析中可以有效利用这种相关特征增强建模效果，如协同克里格插值分析。

正交协方差云用来研究两个数据集间的交叉相关。在 ArcGIS9.0 中生成数据的正交协方差函数云图的主要步骤：

（1）在 ArcMap 中加载地统计数据点图层。

（2）单击 Geostatistical Analyst 模块的下拉箭头选择 Explore Data 并点击 Crosscovariance Cloud 的命令。

（3）设置相关参数。

图 10.28 为某地区 GDP 与人口的正交协方差云图。从图中可以看出，该地区人口数量

和 GDP 的交叉相关性似乎并不对称，具有明显的西北-东南方向性。选择 Crosscovariance Cloud 视图中的 Show Search Direction，改变搜索方向为西北，观察半变异/协方差视图中半变异/协方差函数云的变化（图 10.29（a））；再次，改变搜索方向为西北，观察半变异/协方差视图中半变异/协方差函数云的变化（图 10.29（b））。可以发现，该地区人口数量和 GDP 的交叉相关性具有的西北-东南方向性，并且该两个方向的交叉相关性比较稳定。交叉相关性的分析可以为地统计分析中的协同克里格插值分析重要的建模依据。

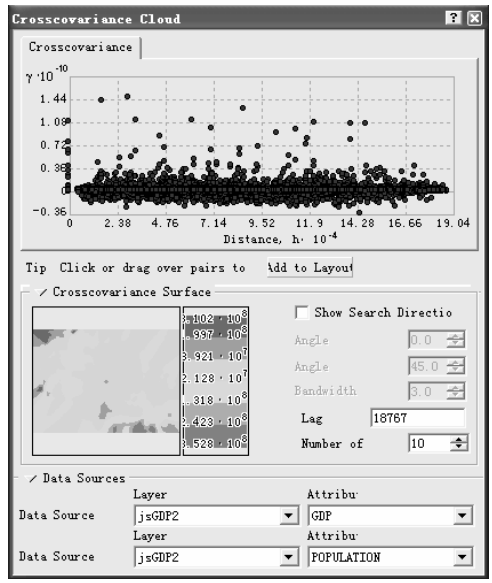


图 10.28　正交协方差云图

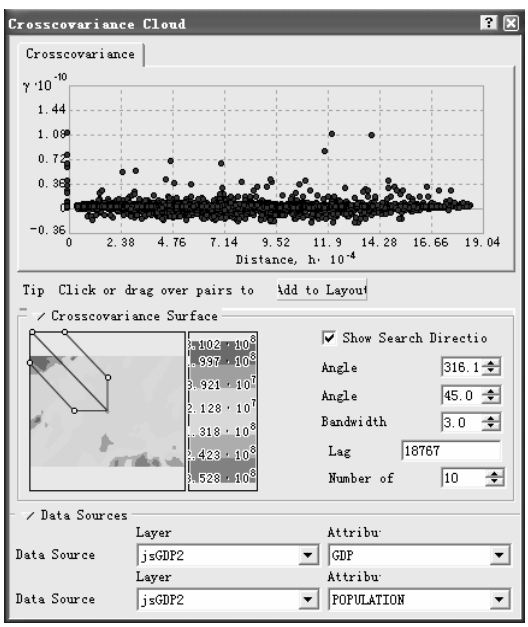


图 10.29（a）

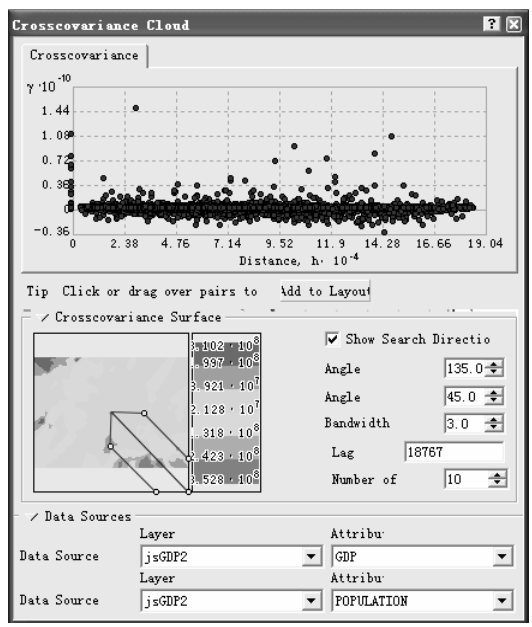


图 10.29（b）

图 10.29　正交协方差云方向图

10.3 空间确定性插值

确定性插值方法以研究区域内部的相似性、或者以平滑度为基础，由已知样点来创建表面。本节主要对确定性空间插值中的反距离权插值、全局多项式插值、局部多项式插值和径向基函数插值的概念、原理及在 ArcGIS 9 中的实现过程进行详细介绍。

10.3.1 反距离加权插值

反距离权（IDW　Inverse Distance Weighted）插值法是基于相近相似的原理：即两个物体离得近，它们的性质就越相似，反之，离得越远则相似性越小。它以插值点与样本点间的距离为权重进行加权平均,离插值点越近的样本点赋予的权重越大。

反距离加权插值法的一般公式如下：

$$\hat{Z}(s_0)=\sum_{i=1}^{N}\lambda_i Z(s_i)$$

其中，$Z(s_0)$为s_0处的预测值；

N 为预测计算过程中要使用的预测点周围样点的数量；

λ_i 为预测计算过程中使用的各样点的权重，该值随着样点与预测点之间距离的增加而减少；

$Z(s_i)$ 是在 s_i 处获得的测量值。

确定权重的计算公式为：

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^{N} d_{i0}^{-p} \qquad \sum_{i=1}^{N}\lambda_i = 1,$$

其中，P 为指数值；

d_{i0}是预测点s_0与各已知样点s_i之间的距离。

样点在预测点值的计算过程中所占权重的大小受参数 p 的影响；也就是说，随着采样点与预测值之间距离的增加，标准样点对预测点影响的权重按指数规律减少。在预测过程中，各样点值对预测点值作用的权重大小是成比例的，这些权重值的总和为 1。

在 ArcGIS9.0 中利用反距离加权工具进行空间插值的基本步骤为：

1. 在 ArcMap 中加载地统计数据点图层。

2. 单击 Geostatistical Analyst 模块的下拉箭头点击 Geostatistical Wizard 命令，弹出图 10.30 所示界面；

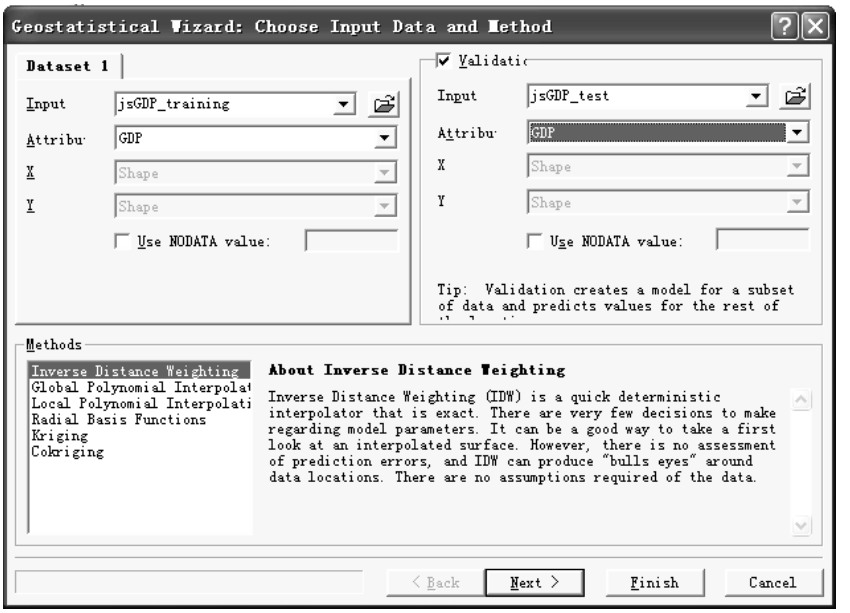


图 10.30　　输入数据和方法选择对话框

参数说明：

（1）Dataset1

1） Input:选择进行内插的实验数据。

2） Attribute: 选择进行内插的实验数据的属性。

（2）Validation

1） Input:选择进行内插的检验数据。

2） Attribute: 选择进行内插的检验数据的属性。

（3）Method:选择数据的内插方法。

为了保证内插的精度，在此选择检验数据。为了能够将表面精确地描绘出来，需保证训练数据集中有足够的样本。若训练数据集中数据太少或含有异常值，会造成模型参数错误及输出结果变形。

3. 选择 Inverse Distance Weighting ,然后单击 Next, 弹出图 10.31 所示界面

参数设置：

（1）Power:即公式中的 p 参数值。

（2）Symbol:设置上图中点符号的大小。

（3）Preview type

1） Neighbors:预览预测点的效果图。

2） Surface: 预览反距离内插表面生成的效果图。

（4）Method

1） Neighbors to:在搜索半径内使用预测点最大的个数。

2） Include at least:在搜索半径内使用预测点最小的个数。

3） Shape:区域扇区形状的选择。

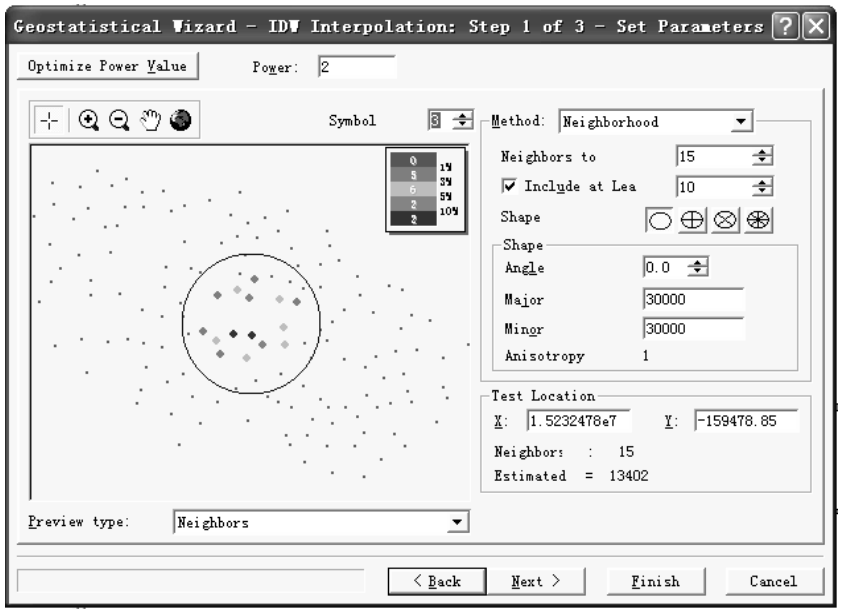


图 10.31　　反距离权内插参数设置对话框

4. 单击 Next， 弹出内插实验数据样点值预测图 10.32 所示界面；

上图 X 轴代表样点的真实值，Y 轴代表内插出来的样点值，虚线代表理论上的点值的拟和曲线，实线代表内插出点值的拟合曲线。实线的趋势越与虚线吻合，说明内插的效果越好。

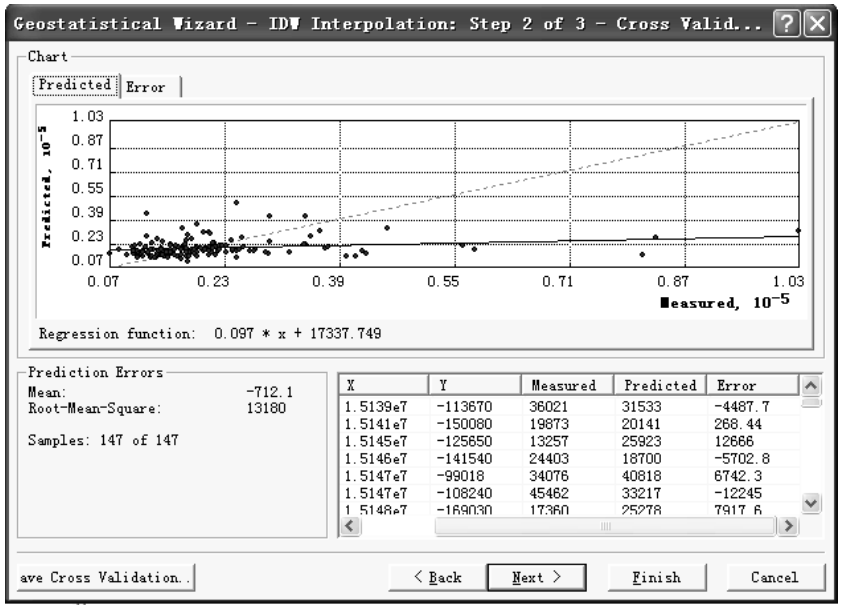


图 10.32　　反距离权内插正交验证对话框

5. 单击 Next， 弹出检验数据样点预测值误差图 10.33 所示界面，各参数含义同上：

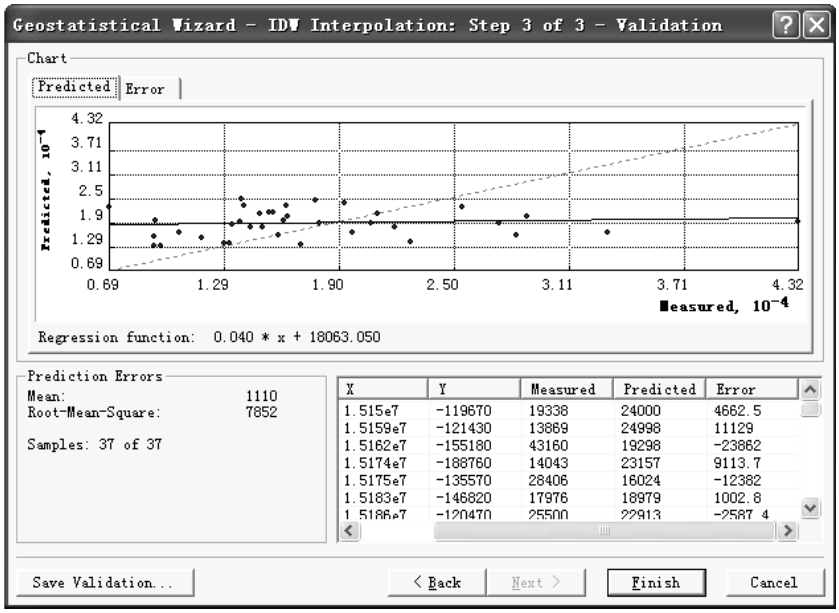


图 10.33　　反距离权内插验证对话框

6. 单击 Finish，弹出反距离内插图 10.34 所示界面。

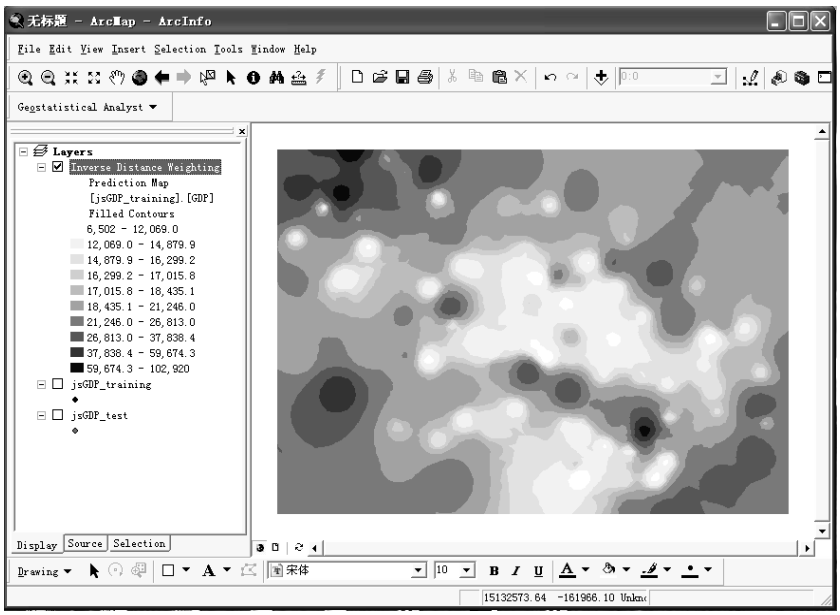


图 10.34　　反距离权内插结果图

利用该方法进行插值时，样点分布应尽可能均匀，且布满整个插值区域。对于不规则分布的样点，插值时利用的样点往往也不均匀的分布在周围的不同方向上，这样，每个方

向对插值结果的影响是不同的，插值结果的准确度也会降低。

10.3.2 全局多项式插值

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。利用全局性插值法生成的表面容易受极高和极低样点值的影响，尤其在研究区边沿地带，因此用于模拟的有关属性在研究区域内最好是变化平缓的。全局多项式插值法适用的情况有：1. 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。2. 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，在这种情况下应用的方法通常被称为趋势面分析。

在 ArcGIS9.0 中利用全局多项式插值方法进行空间插值的基本步骤为：

1. 在 ArcMap 中加载地统计数据点图层。
2. 单击 Geostatistical Analyst 模块的下拉箭头点击 Geostatistical Wizard 命令，弹出图 10.35 所示界面。

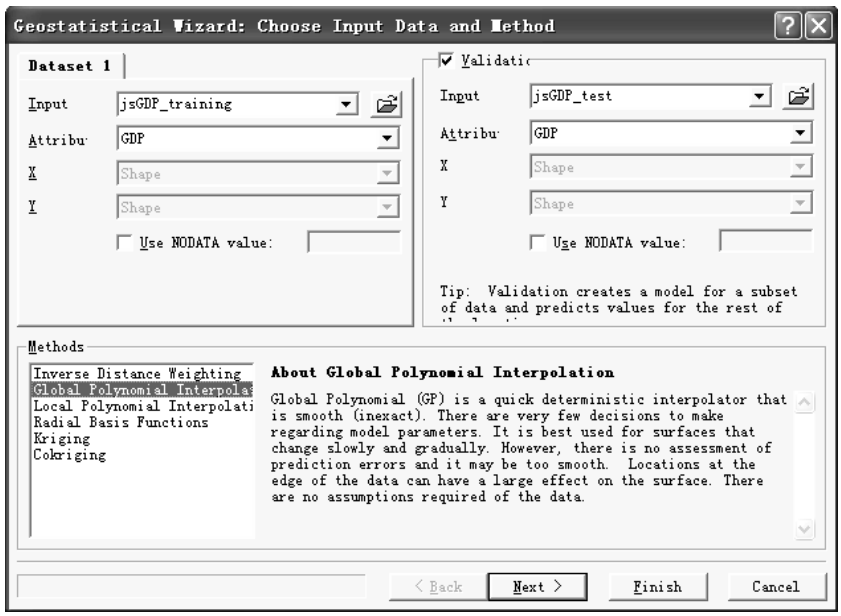
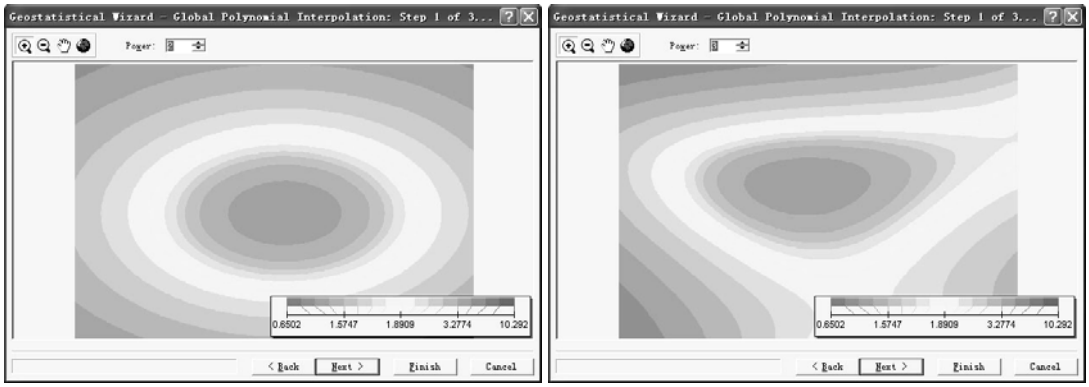


图 10.35　　输入数据与方法选择对话框

3. 选择 Global Polynomial Interpolation，然后单击 Next，　弹出图 10.36 所示界面。参数说明：

   Power:设置拟合表面多项式的次数。次数越低，拟合的表面越粗糙，实际表面拟合的效果越差，大致代表了此区域的宏观趋势。次数越高，拟合面越光滑，拟合的结果更

接近实际的表面。但需注意的是，并不是次数越高越好，次数过高使得计算量大大增加而精度提高不大。因此，一般选用到三次即可。下图分别为二次拟合及三次拟合的结果图，可看出，三次拟合的结果更接近真实情况。



a. 二次趋势拟合　　b. 三次趋势拟合

图 10.36　全局多项式趋势拟合对话

4. 单击 Next，弹出内插实验数据样点值预测图 10.37 所示界面，各参数含义同前。

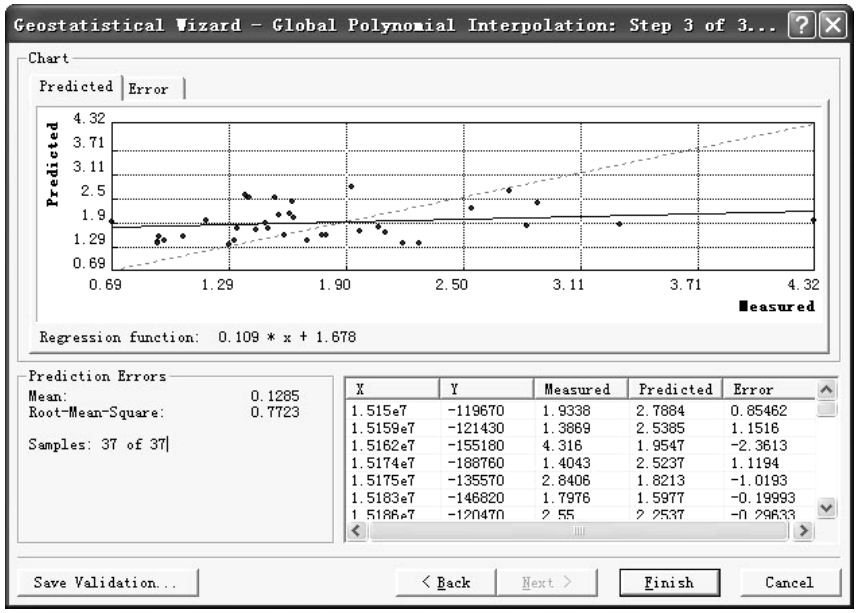


图 10.37　全局多项式拟合正交验证结果

5. 单击 Next，弹出内插检验数据样点值预测图 10.38 所示界面；

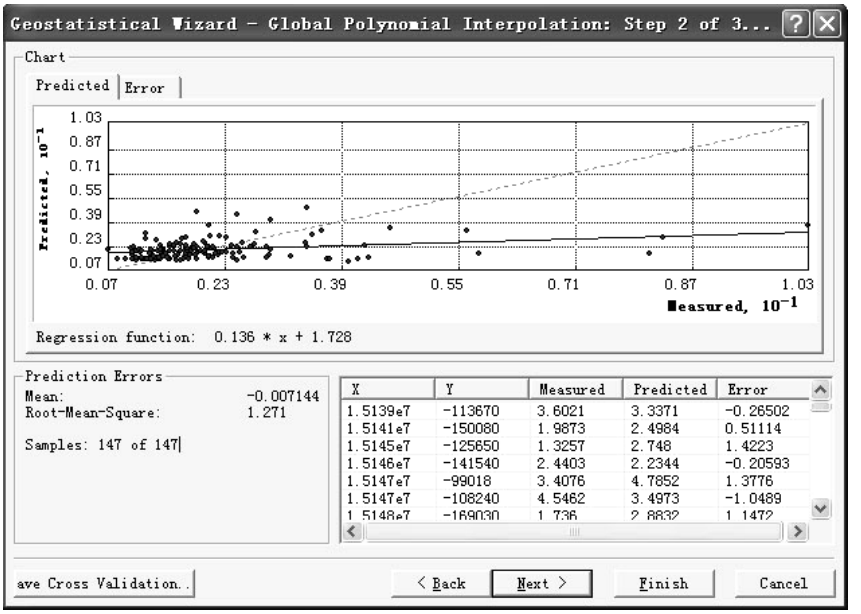


图 10.38 全局多项式拟合的交叉检验结果

6. 单击 Finish，弹出全局多项式内插图 10.39 所示界面。

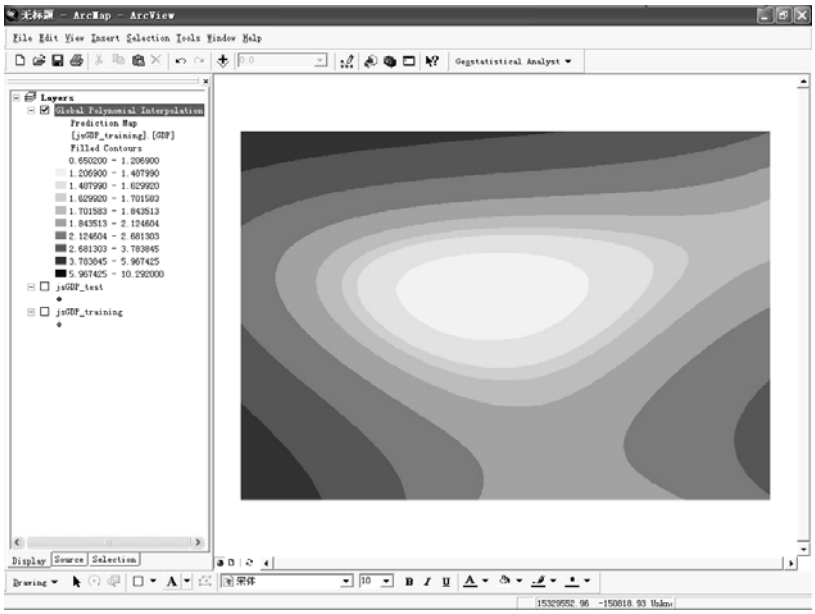


图 10.39 全局多项式内插结果图

10.3.3 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，特别是数据集中含有短程变异时，局部多项式插值法生成的表面就能描述这种短程变异。

在局部多项式插值法中，邻近区域的形状、要用到的样点数量的最大值和最小值以及扇区的构造都需要进行设定，还可以使用另外一种方法，就是通过拖动一个滑块改变参数值定义邻近区域的宽度，这个参数以预测点与已知样点之间的距离为基础，所用的邻近区域内的采样点的权重随着预测点与标准点之间距离的增加而减小。因此，局部多项式插值法产生的表面更多地用来解释局部变异。

在 ArcGIS9.0 中利用局部多项式插值方法进行空间插值的基本步骤为：

1. 在 ArcMap 中加载地统计数据点图层。
2. 单击 Geostatistical Analyst 模块的下拉箭头点击 Geostatistical Wizard 命令，弹出图 10.40 所示界面。

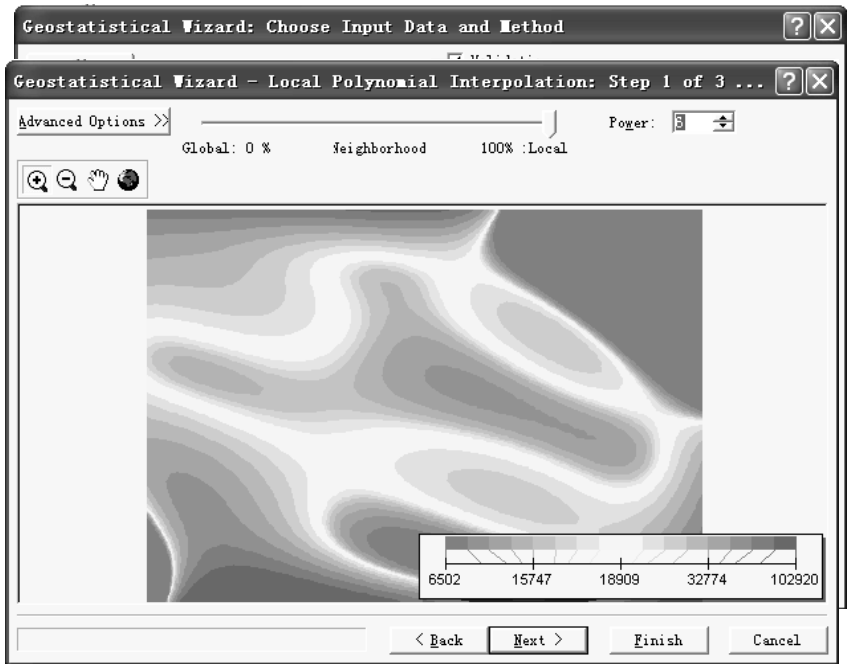


图 10.41　　局部多项式趋势拟合对话框

3. 选择 Local Polynomial Interpolation ，然后单击 Next，　弹出图 10.41 所示界面。

参数说明：

Power: 设置拟合表面多项式的次数。

滑块值用来设置局部和全局插值的程度。

4. 单击 Next，弹出内插实验数据样点值预测图 10.42 所示界面。

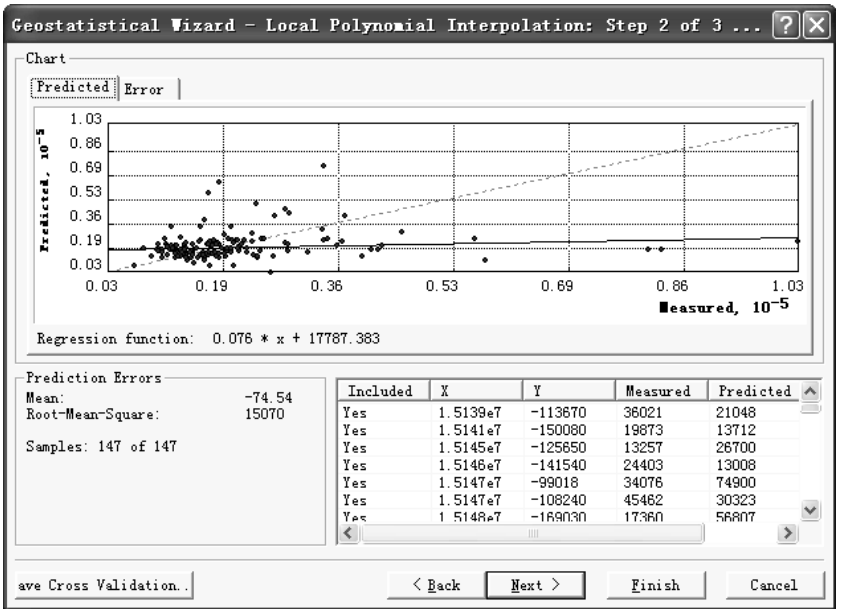


图 10.42　局部多项式插值方法的交叉验证

5. 单击 Next，弹出内插检验数据样点值预测图 10.43 所示界面。

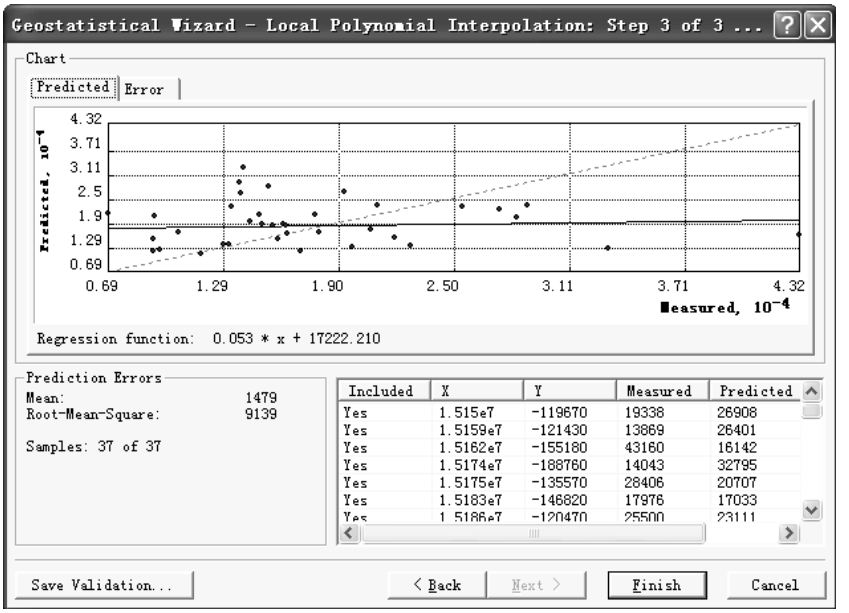


图 10.43　局部多项式拟合的交叉检验结果

6. 单击 Finish，弹出局部多项式内插图 10.44 所示界面。

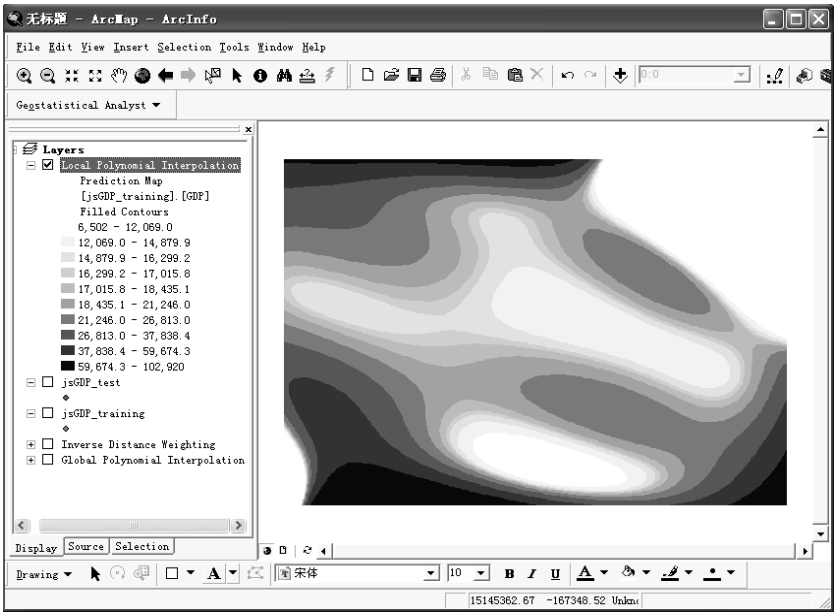


图 10.44　局部多项式内插的结果图

10.3.4 径向基函数插值

从概念上来说，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。它不同于全局多项式和局部多项式插值方法，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。选择何种基本函数意味着将以何种方式使径向基表面穿过一系列已知样点。

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，能得到令人满意的结果。而在一段较短的水平距离内，表面值发生较大的变化，或无法确定采样点数据的准确性，或采样点数据具有很大的不确定性时，径向基函数插值的方法并不适用。

在 ArcGIS9.0 中利用径向基函数插值方法进行空间插值的基本步骤为：

1. 在 ArcMap 中加载地统计数据点图层。
2. 单击 Geostatistical Analyst 模块的下拉箭头点击 Geostatistical Wizard 命令，弹出图 10.45 所示界面。

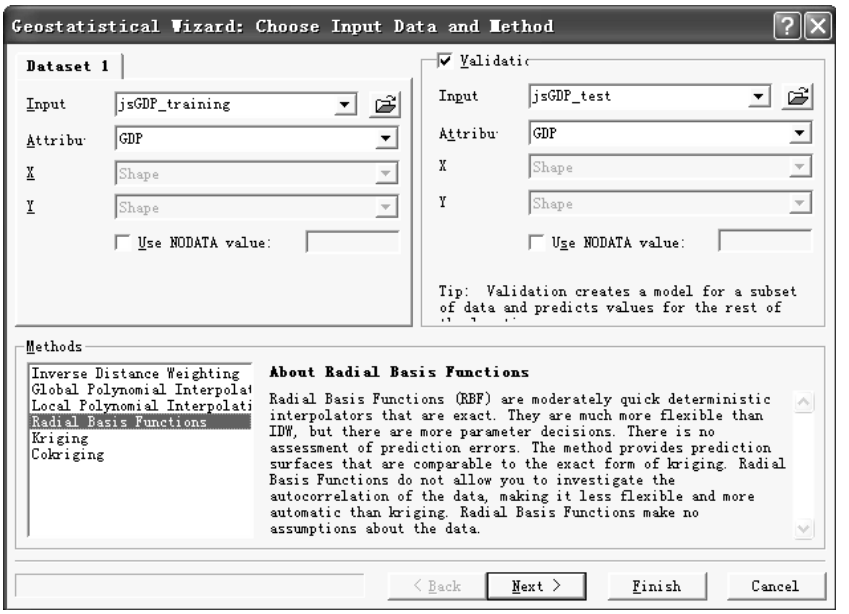


图 10.45　方法与数据选择对话框

3.　选择 Radial Basis Function ，单击 Next，　弹出图 10.46 所示界面。

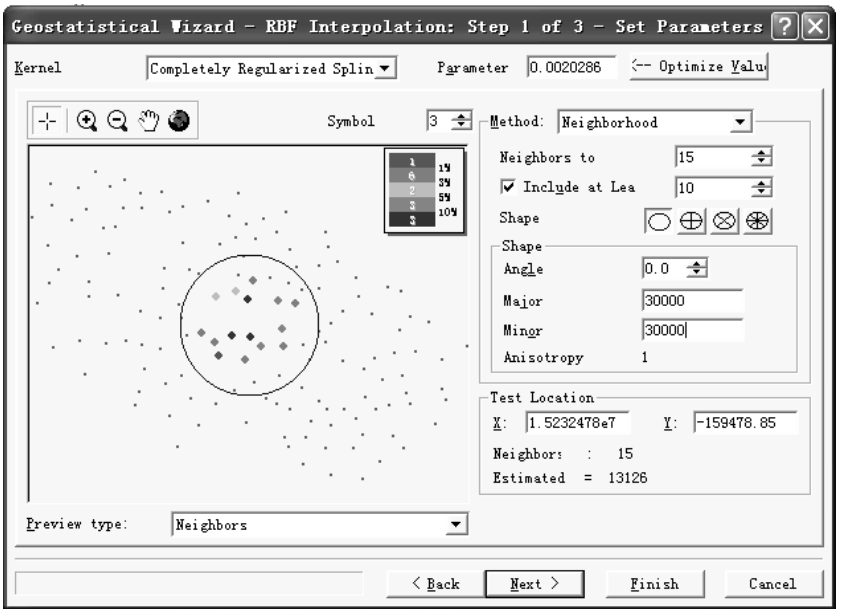


图 10.46　　径向基插值方法的参数设置

参数说明：

Kernel 下拉菜单中提供了几种进行插值的基本函数：

（1）Thin-plate spline:平面样条函数

（2）Spline with tension:张力样条函数

（3）Completely regularized spline:规则样条函数

（4）Multiquadric functions:高次曲面函数

（5）Inwerse multiquadric spline:反高次曲面样条函数

在 Parameter 对话框中可以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，所得到的表面却不平滑；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。

4. 单击 Next， 弹出内插训练数据样点值交叉验证对话框，如图 10.47 所示界面。

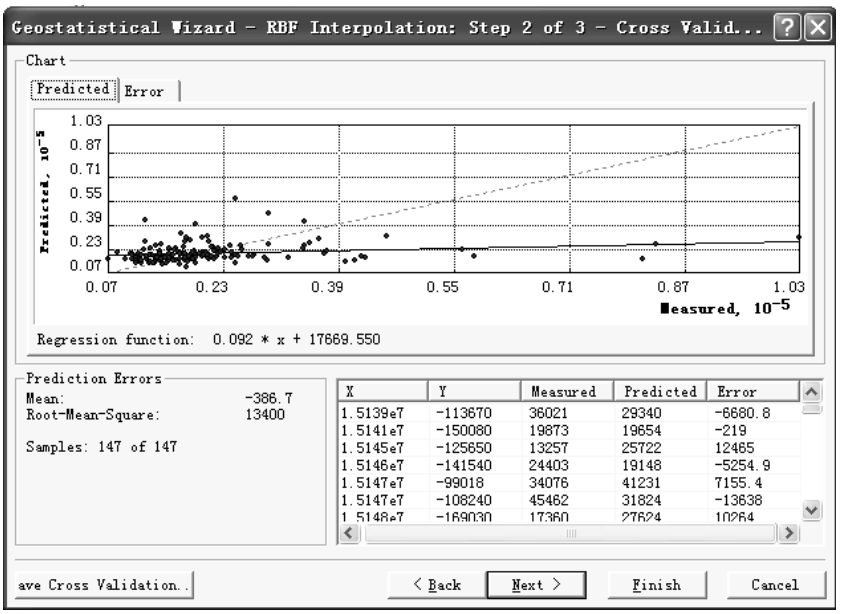


图 10.47 径向基插值方法的交叉验证

5. 单击 Next， 弹出内插检验数据样点值验证对话框，如图 10.48 所示界面。

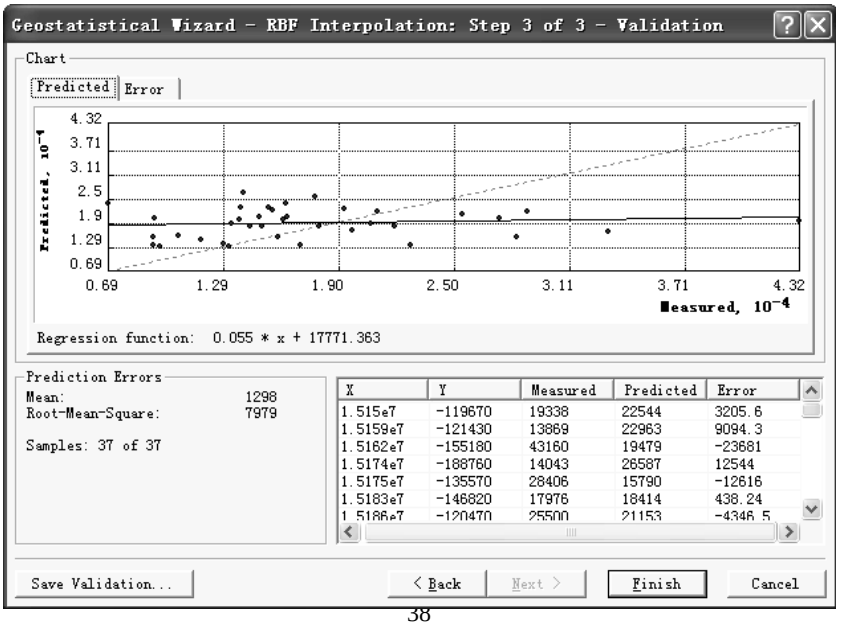


图 10.48 径向基插值方法的结果验证

6. 单击 Finish，弹出径向函数内插图 10.49 所示界面。

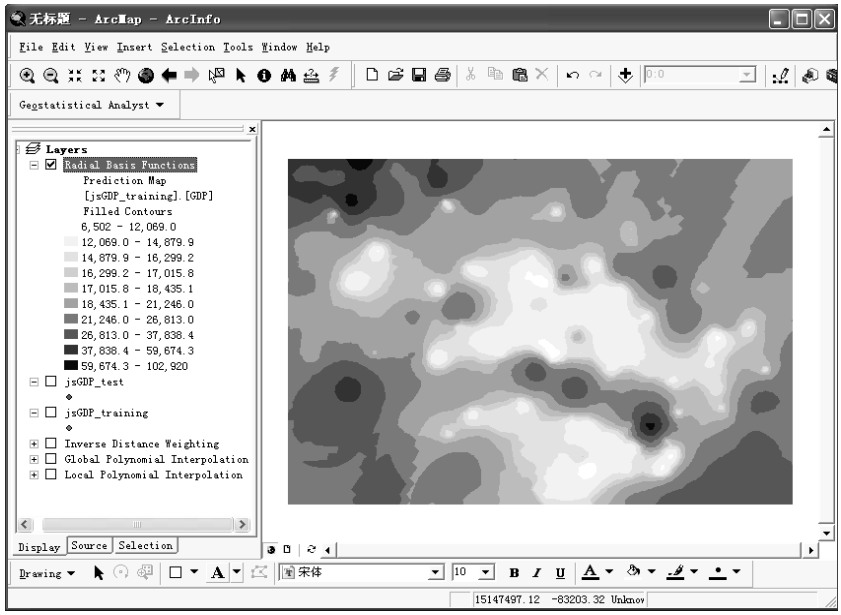


图 10.49　　径向基插值结果图

10.4 地统计插值

10.4.1 克里格插值基础

1. 克里格方法概述

克里格方法（Kriging）又称空间局部插值法，是以变异函数理论和结构分析为基础，在有限区域内对区域化变量进行无偏最优估计的一种方法，是地统计学的主要内容之一。南非矿产工程师 D.R.Krige（1951 年）在寻找金矿时首次运用这种方法，法国著名统计学家 G.Matheron 随后将该方法理论化、系统化，并命名为 Kriging，即克里格方法。

克里格方法的适用范围为区域化变量存在空间相关性，即如果变异函数和结构分析的结果表明区域化变量存在空间相关性，则可以利用克里格方法进行内插或外推；否则，是不可行的。其实质是利用区域化变量的原始数据和变异函数的结构特点，对未知样点进行线性无偏、最优估计。无偏是指偏差的数学期望为 0，最优是指估计值与实际值之差的平方和最小。也就是说，克里格方法是根据未知样点有限邻域内的若干已知样本点数据，在考虑了样本点的形状、大小和空间方位，与未知样点的相互空间位置关系，以及变异函数提供的结构信息之后，对未知样点进行的一种线性无偏最优估计。

克里格方法与反距离权插值方法类似的是，两者都通过对已知样本点赋权重来求得未

知样点的值，可统一表示为：

$$Z(x_0)=\sum_{i=1}^{n}\omega Z(x_i)$$

式中，$Z(x_0)$为未知样点的值，$Z(x_i)$为未知样点周围的已知样本点的值，w_i为第i个已知样本点对未知样点的权重，n为已知样本点的个数。

不同的是，在赋权重时，反距离权插值方法只考虑已知样本点与未知样点的距离远近，而克里格方法不仅考虑距离，而且通过变异函数和结构分析，考虑了已知样本点的空间分布及与未知样点的空间方位关系。

2. 克里格方法的具体步骤

用克里格方法进行插值的主要步骤如图 10.50 所示：

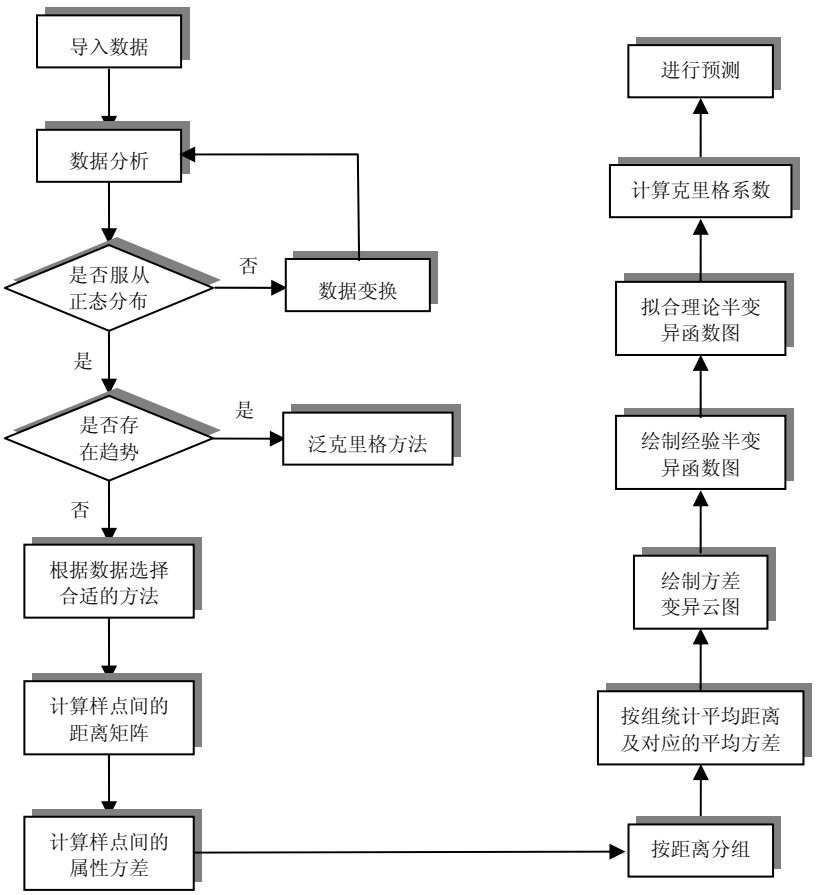


图 10.50 克里格方法的主要步骤

在克里格插值过程中，需注意以下几点：

（1）数据应符合前提假设

（2）数据应尽量充分，样本数尽量大于 80，每一种距离间隔分类中的样本对数尽量多于 10 对

（3）在具体建模过程中，很多参数是可调的，且每个参数对结果的影响不同。如：块金值：误差随块金值的增大而增大；基台值：对结果影响不大；变程：存在最佳变程值；拟合函数：存在最佳拟合函数

（4）当数据足够多时，各种插值方法的效果相差不大。

3. 克里格方法的分类

目前，克里格方法主要有以下几种类型：普通克里格（Ordinary Kriging）；简单克里格（Simple Kriging）；泛克里格（Universal Kriging）；协同克里格（Co-Kriging）；对数正态克里格（Logistic Normal Kriging）；指示克里格（Indicator Kriging）；概率克里格（Probability Kriging）；析取克里格（Disjunctive Kriging）等。下面简要介绍一下 ArcGIS 中常用的几种克里格方法的适用条件，其具体的算法、原理可查阅相关文献资料。

不同的方法有其适用的条件，按照以上流程图所示步骤，当数据不服从正态分布时，若服从对数正态分布，则选用对数正态克里格；若不服从简单分布时，选用析取克里格。当数据存在主导趋势时，选用泛克里格。当只需了解属性值是否超过某一阈值时，选用指示克里格。当同一事物的两种属性存在相关关系，且一种属性不易获取时，可选用协同克里格方法，借助另一属性实现该属性的空间内插。当假设属性值的期望值为某一已知常数时，选用简单克里格。当假设属性值的期望值是未知的，选用普通克里格。

10.4.2 普通克里格插值

普通克里格（Ordinary Kriging）是区域化变量的线性估计，它假设数据变化成正态分布，认为区域化变量 Z 的期望值是未知的。插值过程类似于加权滑动平均，权重值的确定来自于空间数据分析。

ArcGIS 中普通克里格插值包括 4 部分功能：创建预测图（Prediction Map）、创建分位数图（Quantile Map）、创建概率图（Probability Map）、创建标准误差预测图（Prediction Standard Error Map）。

1. 创建预测图（Prediction Map）

其在 ArcGIS 中的实现步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 模块的下拉箭头点击 Geostatistical Wizard 命令。

（4）在弹出的对话框（如图 10.51)中，在 Dataset 选择训练数据 jsGDP_test_training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP_test_test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

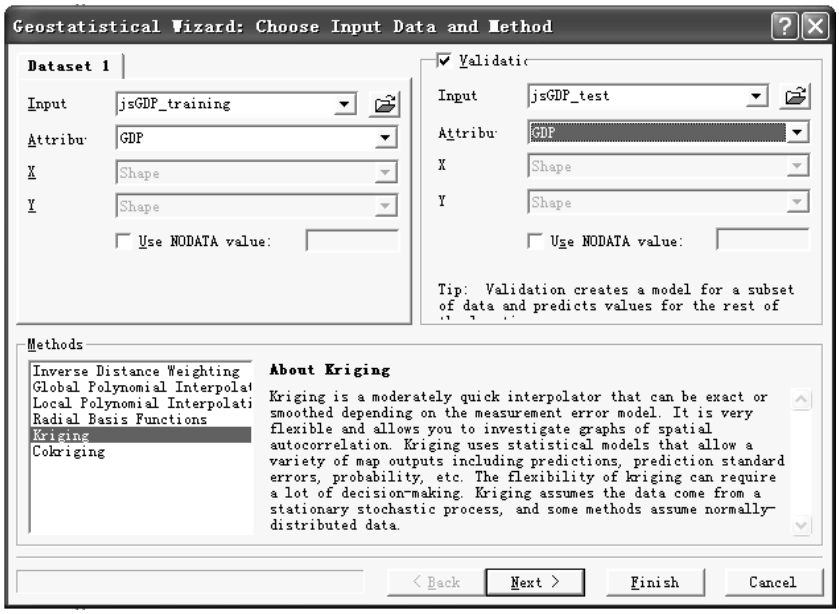


图 10.51　　输入数据和方法选择对话框

（5）在弹出的对话框（如图 10.52)中，展开普通克里格（Ordinary Kriging），在下面的选项中点击预测（Prediction），在 DataSet1 里的 Transformation 里选择 log 变换方式，点击 Next 按钮。

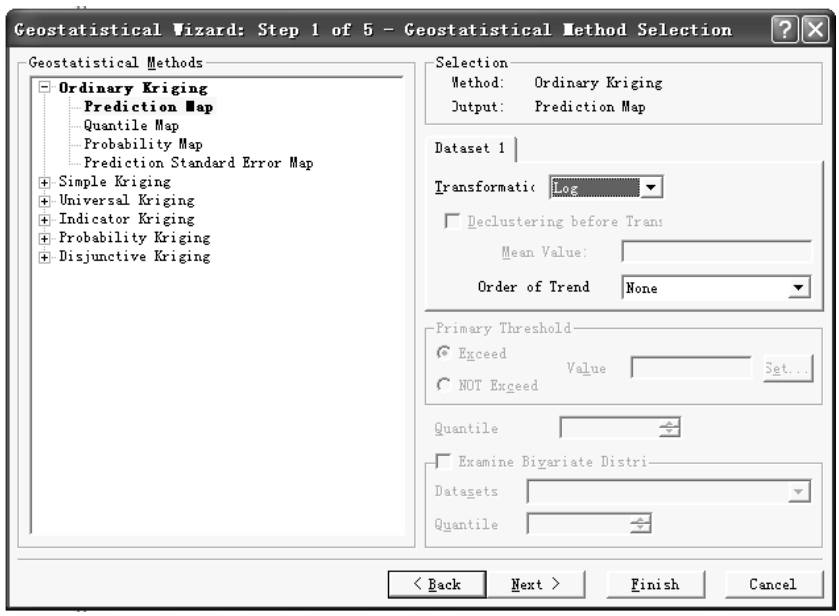


图 10.52　　统计内插方法选择对话框

（6）在弹出的 Semivariogram/Covariance Modeling 对话框（如图 10.53)中，选中 Show Search Direction 选项，移动左图中的搜索方向，然后点击 Next 按钮。

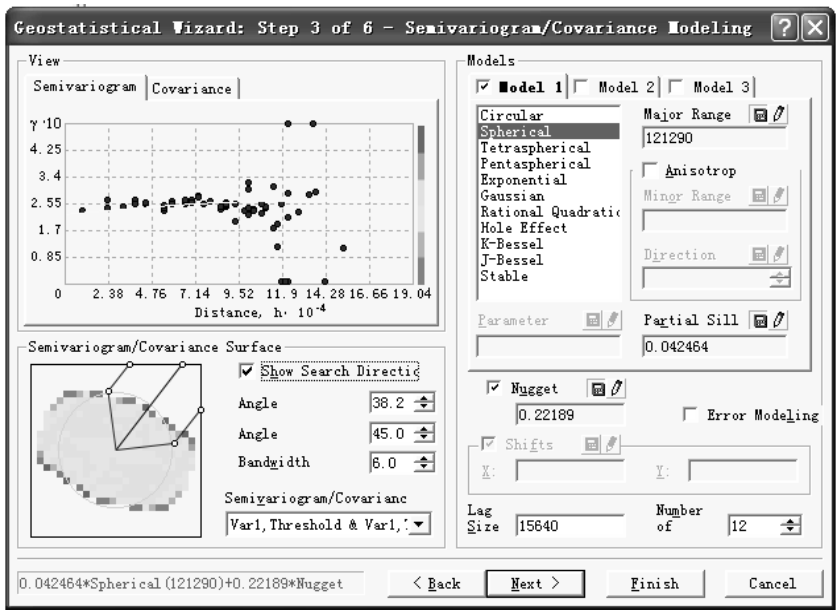


图 10.53　　半变异/协方差模型对话框

（7）在弹出的 Searching Neighborhood 对话框（如图 10.54)，点击 Next 按钮。

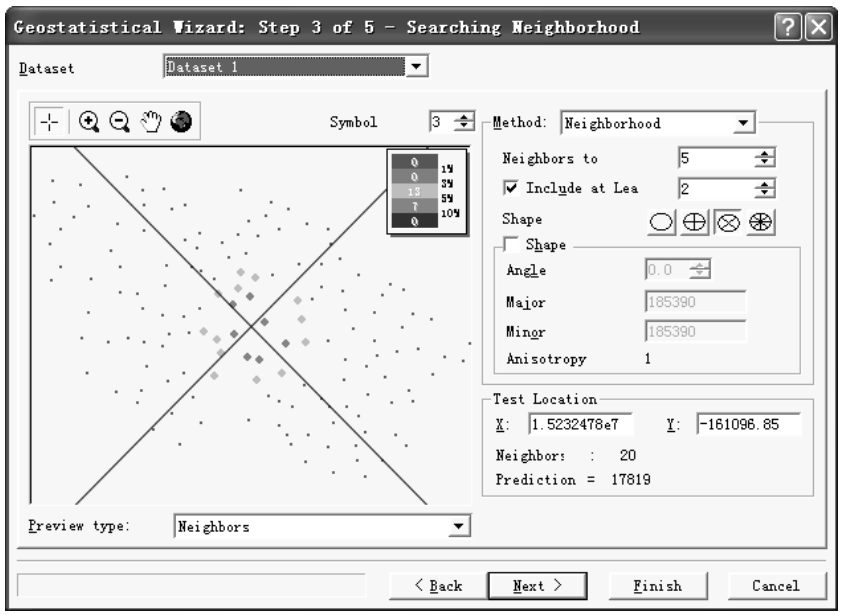


图 10.54　　邻域搜索对话框

（8）在弹出的 Cross Validation 对话框（如图 10.55)中，列出对上述参数的训练数据模型精度评价。在对不同参数得到模型的比较中，可参考 Prediction Error 中的几个指标。符合以下标准的模型是最优的：标准平均值（Mean Standardized）最接近于 0，均方根预测误差（Root-Mean-Square）最小，平均标准误差（Average Mean Error）

最接近于均方根预测误差（Root-Mean-Square），标准均方根预测误差（Root-Mean-Square Standardized）最接近于 1。点击 Next 按钮。

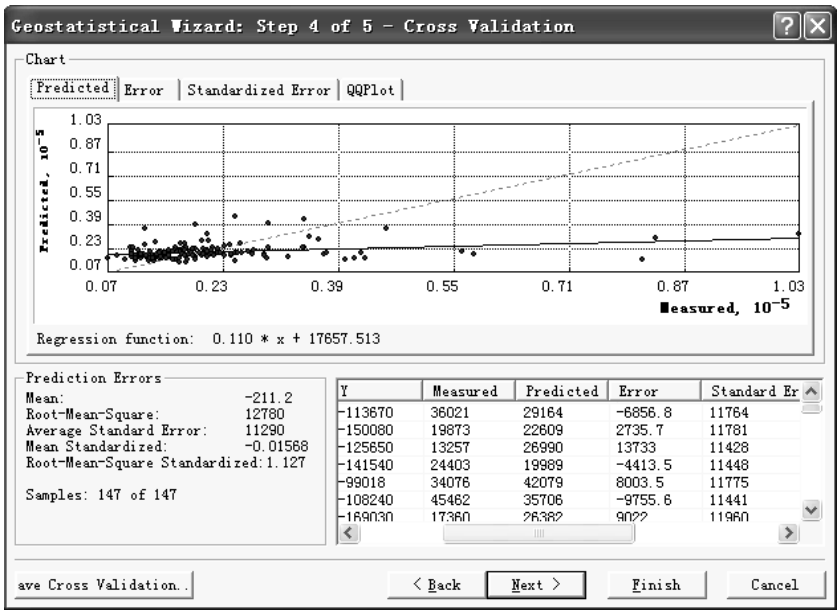


图 10.55　　交叉验证对话框

（9）在弹出的 Validation 对话框（如图 10.56)中，点击 Finish 按钮。普通克里格法内插结

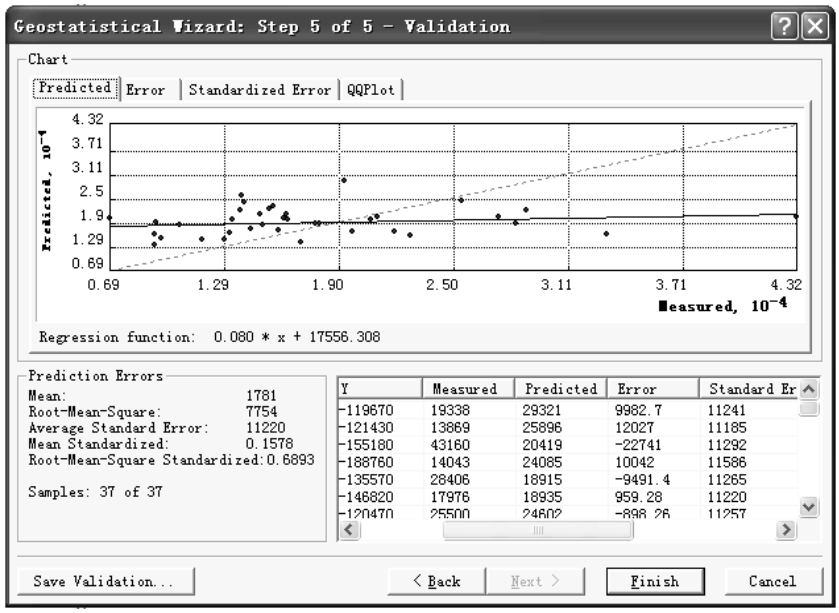


图 10.56　　验证对话框

果图如图 10.57。

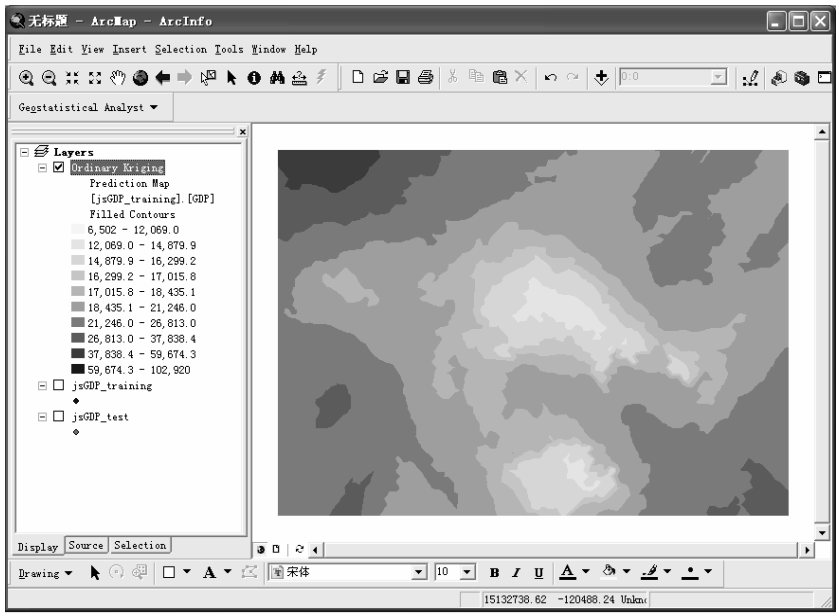


图 10.57　　普通克里格内插生成的预测图

2. 创建分位数图（Quantile Map）

以类似方法可创建普通克里格的分位数图，对 jsGDP_training 创建分位数图的结果如下图 10.58 所示：

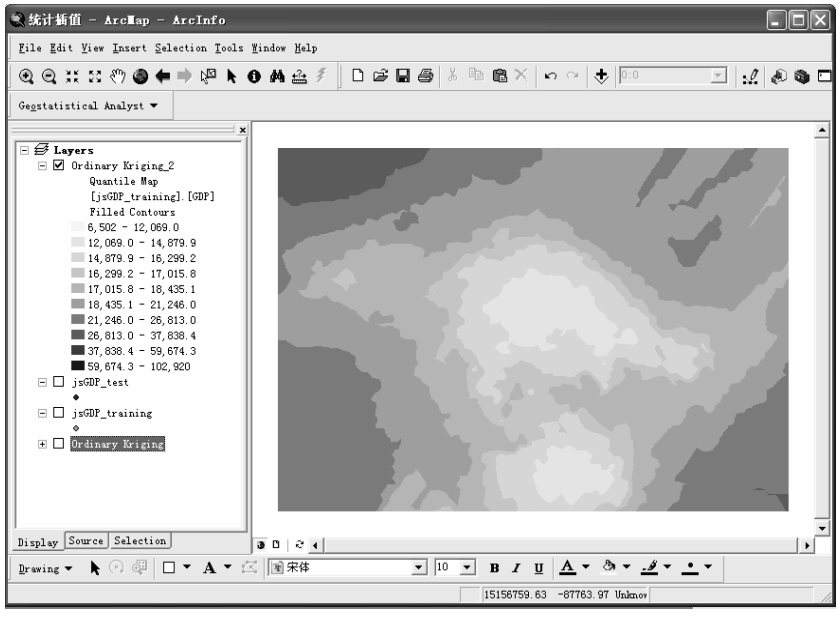


图 10.58　　普通克里格内插生成的四分位图

3. 创建概率图（Probability Map）

以类似方法可创建普通克里格的概率图，对 jsGDP_training 创建概率图的结果如下图 10.59 所示：

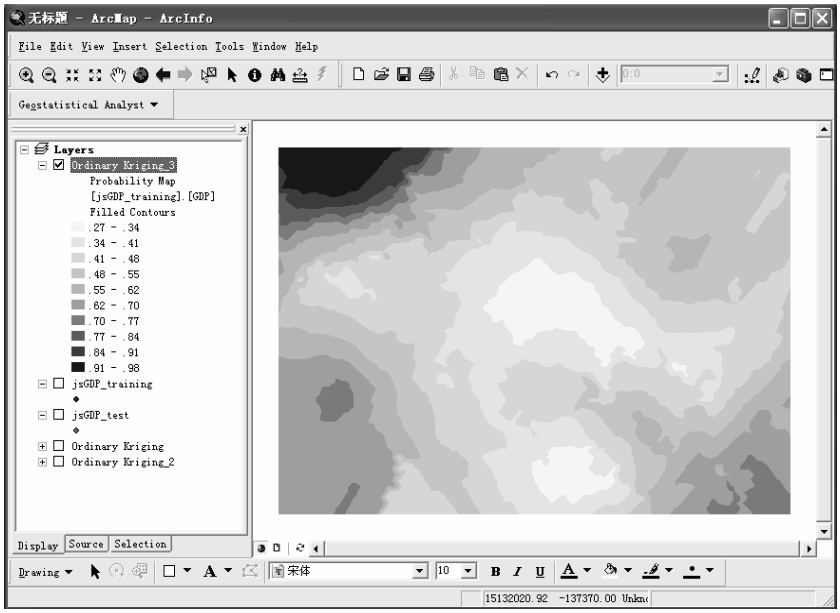


图 10.59　普通克里格内插生成的概率图

4. 创建标准误差预测图（Prediction Standard Error Map）

以类似方法可创建普通克里格的标准误差预测图，对 jsGDP_training 创建标准误差预测图的结果如下图 10.60 所示：

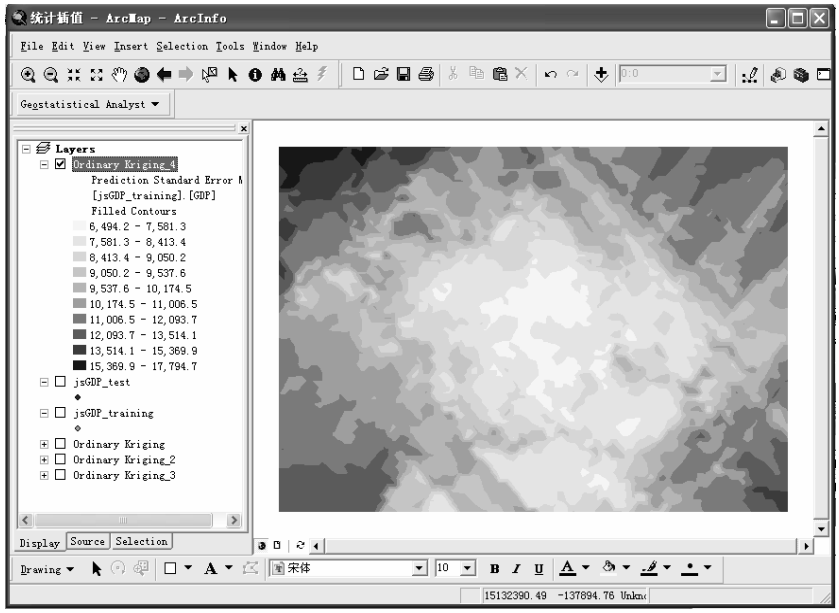


图 10.60 普通克里格内插生成的标准误差预测图

10.4.3 简单克里格插值

简单克里格是区域化变量的线性估计，它假设数据变化成正态分布，认为区域化变量 Z 的期望值为已知的某一常数。ArcGIS 中普通克里格插值包括 4 部分功能：创建预测图（Prediction Map）、创建分位数图（Quantile Map）、创建概率图（Probability Map）、创建标准误差预测图（Prediction Standard Error Map）。

1. 创建预测图（Prediction Map）

其在 ArcGIS 中的实现过程与普通克里格的方法雷同，具体步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框中，在 Dataset 选择训练数据 jsGDP _training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP _test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

（5）在弹出的对话框中，展开简单克里格（Simple Kriging），在下面的选项中点击预测图（Prediction Map），然后数据集选择 Log 变换，平均值（Mean）设为 20651，最后点击 Next 按钮。

（6）在弹出的 Semivariogram/Covariance Modeling 对话框中，钩选 Show Search Direction，移动左图中的搜索方向，然后点击 Next 按钮。

（7）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（8）在弹出的 Cross Validation 对话框中，点击 Next 按钮。

（9）在弹出的 Validation 对话框中，点击 Finish 按钮。简单克立格法内插结果如图 10.61 所示。

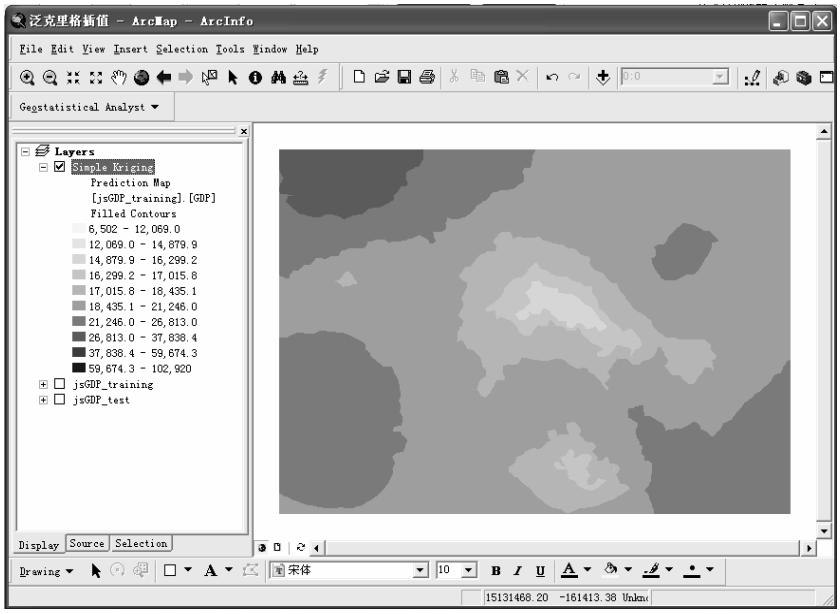


图 10.61　简单克里格内插生成的预测图

2. 创建分位数图（Quantile Map）

以类似方法可创建简单克里格的分位数图，对 jsGDP_training 创建分位数图的结果如下图 10.62 所示：

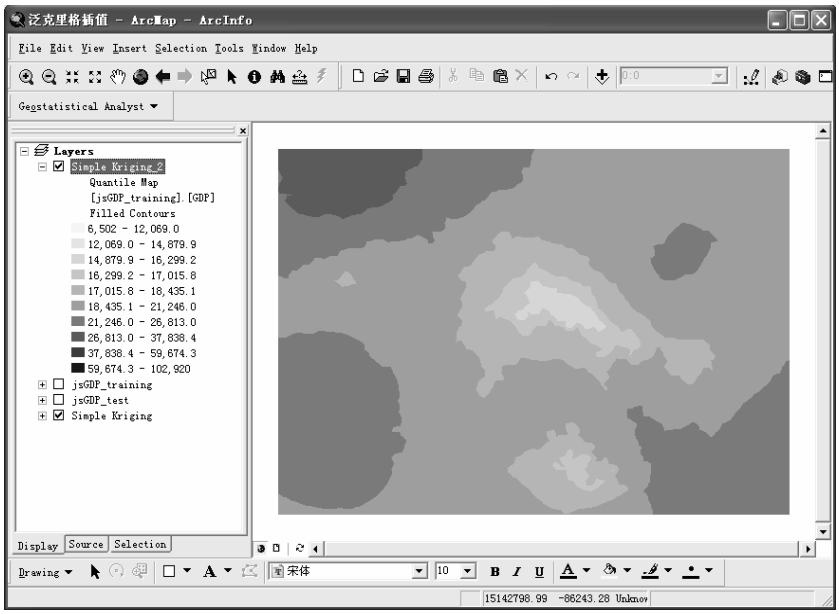


图 10.62　　简单克里格内插生成的分位数图

3. 创建概率图（Probability Map）

以类似方法可创建简单克里格的概率图，对 jsGDP_training 创建概率图的结果如下图 10.63 所示：

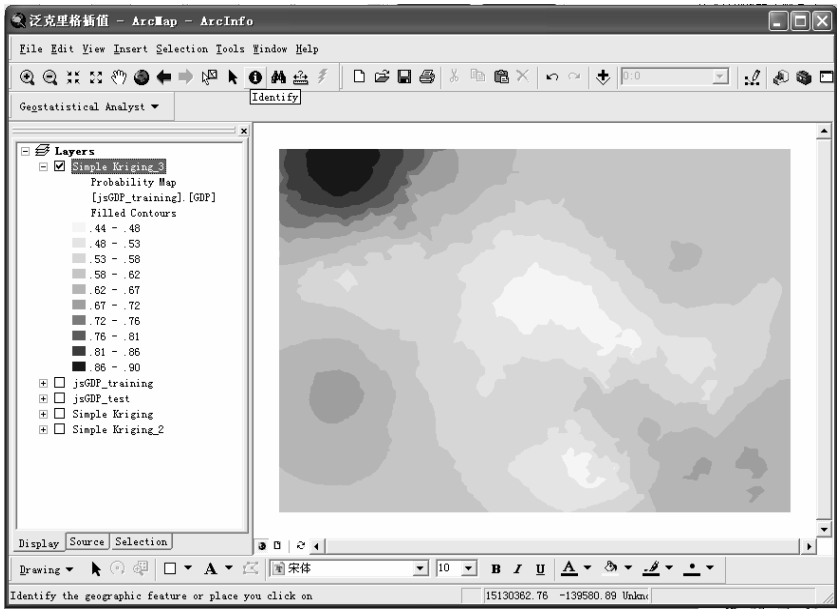


图 10.63　　简单克里格内插生成的概率图

4. 创建标准误差预测图（Prediction Standard Error Map）

以类似方法可创建简单克里格的标准误差预测图，对 jsGDP_training 创建标准误差预测图的结果如下图 10.64 所示：

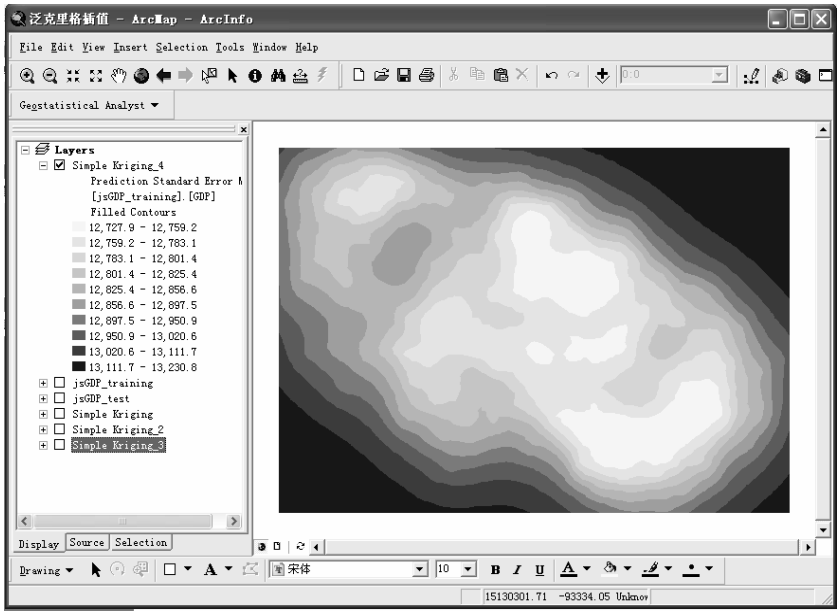


图 10.64 简单克里格内插生成的标准误差预测图

10.4.4 泛克里格插值

泛克里格假设数据中存在主导趋势，且该趋势可以用一个确定的函数或多项式来拟合。在进行泛克里格分析时，首先分析数据中存在的变化趋势，获得拟合模型；其次，对残差数据（即原始数据减去趋势数据）进行克里格分析；最后，将趋势面分析和残差分析的克里格结果加和，得到最终结果。由此可见，克里格方法明显优于趋势面分析，泛克里格的结果也要优于普通克里格的结果。

ArcGIS 中普通克里格插值包括 4 部分功能：创建预测图（Prediction Map）、创建分位数图(Quantile Map)、创建概率图(Probability Map)、创建标准误差预测图(Prediction Standard Error Map)。

1. 创建预测图（Prediction Map）

其在 ArcGIS 中的实现过程与普通克里格的方法雷同，具体步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框中，在 Dataset 选择训练数据 jsGDP _training 及其属性 GDP，在

Validation 中选择检验数据 jsGDP _test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

（5）在弹出的对话框(如图 10.65)中，展开泛克里格（Universal Kriging），在下面的选项中点击预测图(Prediction Map)，然后数据集选择 Log 变换，趋势选项中选择 Third（在前面的趋势分析中，jsGDP_training 存在三次趋势），最后点击 Next 按钮。

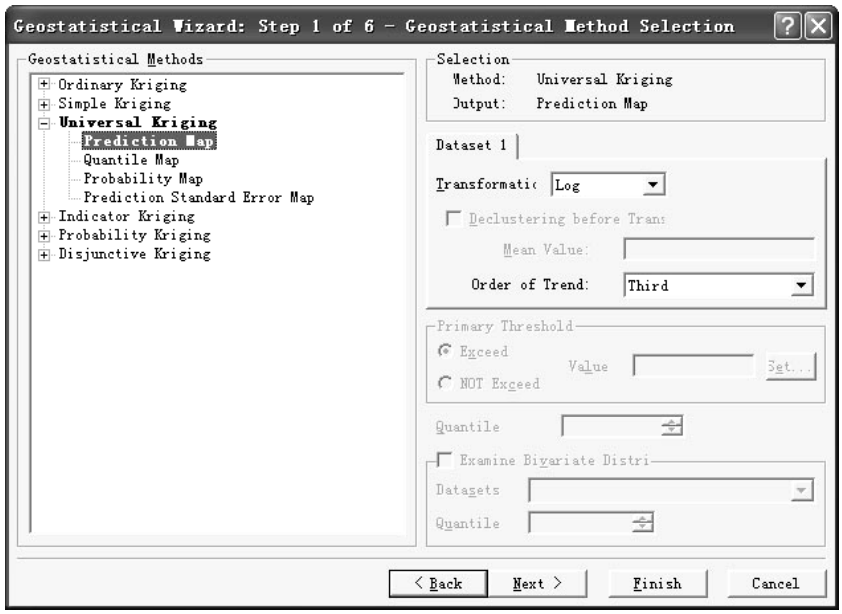


图 10.65　　地统计内插方法选择对话框

（6）在弹出的 Detrenting 对话框（如图 10.66），点击 Next 按钮。

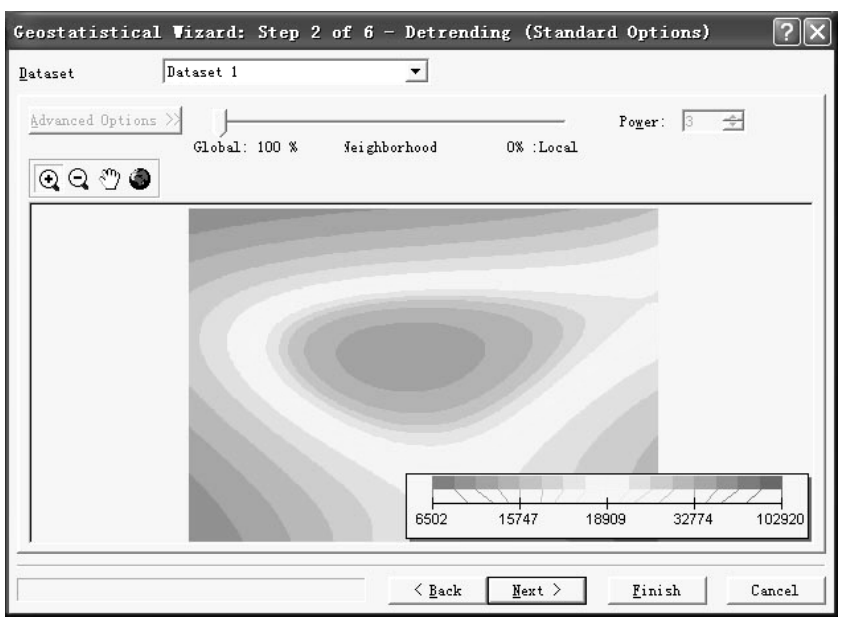


图 10.66　　剔出趋势对话框

（7）在弹出的 Semivariogram/Covariance Modeling 对话框中，钩选 Show Search Direction，移动左图中的搜索方向，然后点击 Next 按钮。

（8）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（9）在弹出的 Cross Validation 对话框中，点击 Next 按钮。

（10） 在弹出的 Validation 对话框中，点击 Finish 按钮。泛克立格法内插结果如图（10.67）。

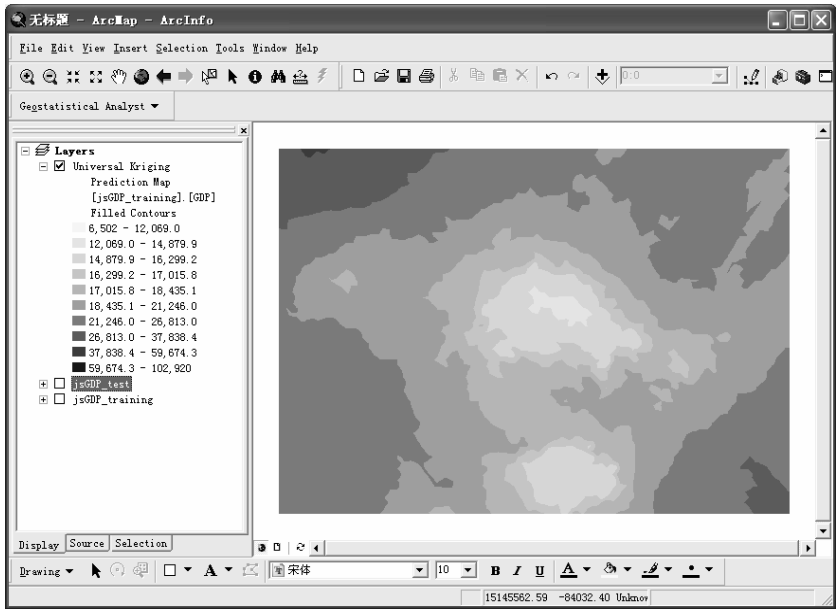


图 10.67　　泛克里格内插生成的预测图

2. 创建分位数图（Quantile Map）

以类似方法可创建泛克里格的分位数图，对 jsGDP_training 创建分位数图的结果如图 10.68 所示：

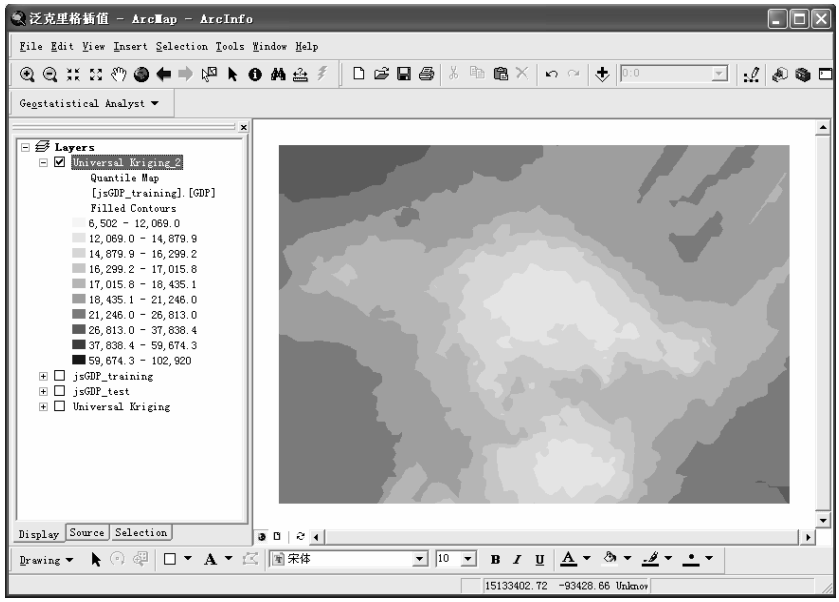


图 10.68　泛克里格内插生成的四分位图

3. 创建概率图（Probability Map）

以类似方法可创建泛克里格的概率图，对 jsGDP_training 创建概率图的结果如下图 10.69 所示：

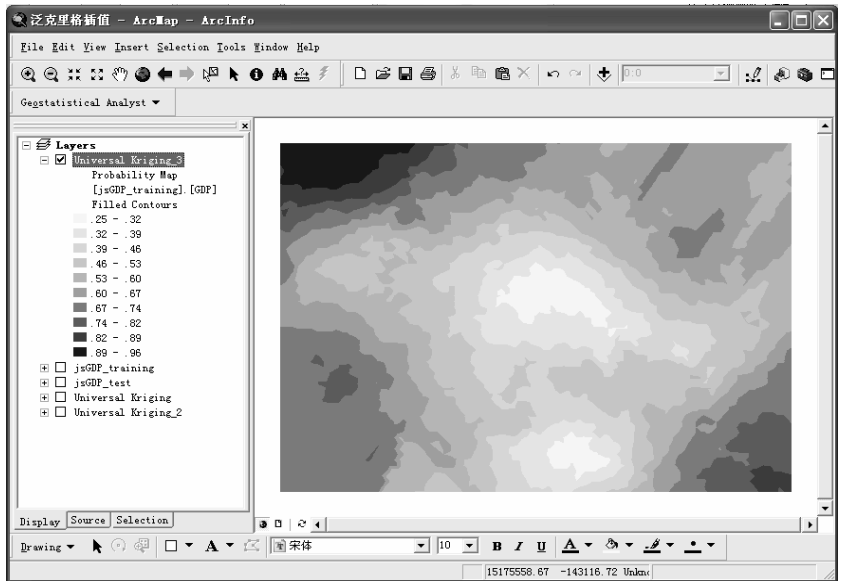


图 10.69　泛克里格内插生成的概率图

4. 创建标准误差预测图（Prediction Standard Error Map）

以类似方法可创建泛克里格的标准误差预测图，对 jsGDP_training 创建标准误差预测图的结果如下图 10.70 所示：

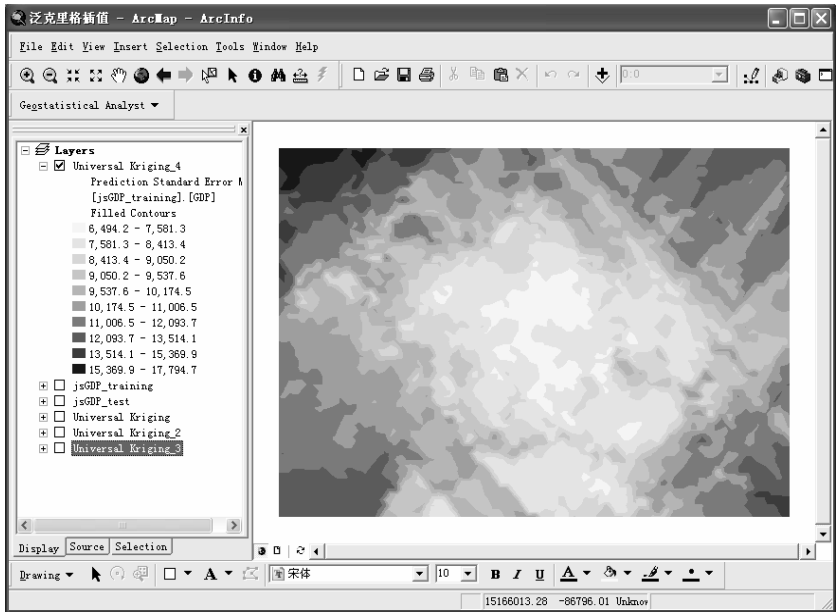


图 10.70　泛克里格内插生成的标准误差预测图

10.4.5 指示克里格插值

在很多情况下，并不需要了解区域内每一个点的属性值，而只需了解属性值是否超过某一阈值，则可将原始数据转换为（0，1）值，选用指示克里格法（Indicator Kriging）进行分析。ArcGIS 中普通克里格插值包括 2 部分功能：创建概率图（Probability Map）和创建标准误差指示图（Standard Error of Indicator Map）。

1. 创建概率图（Probability Map）

其在 ArcGIS 中的实现步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框中，在 Dataset 选择训练数据 jsGDP _training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP _test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

（5）在弹出的对话框中，展开指示克里格（Indicator Kriging），在下面的选项中点击概率图（Probability Map），Primary Threshold 对话框中的阈值按默认值，最后点击 Next 按钮。

（6）在弹出的 Additional Cutoffs Selection 对话框（如图 10.71 所示），点击 Next 按钮。

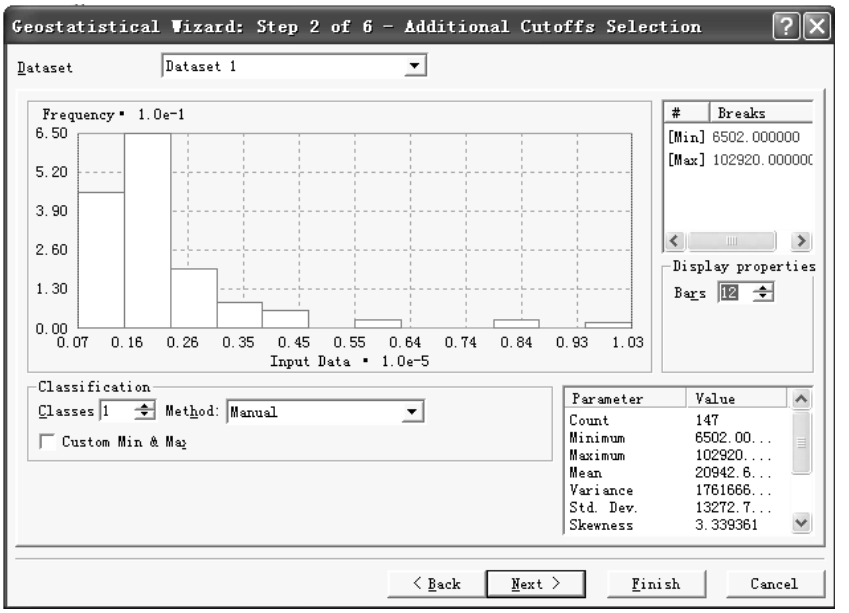


图 10.71　　去除离群值对话框

（7）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（8）在弹出的 Cross Validation 对话框中，点击 Next 按钮。

（9）在弹出的 Validation 对话框中，点击 Finish 按钮。指示克立格法内插结果（如图 10.72）。

2.　创建标准误差指示图（Standard Error of Indicators）

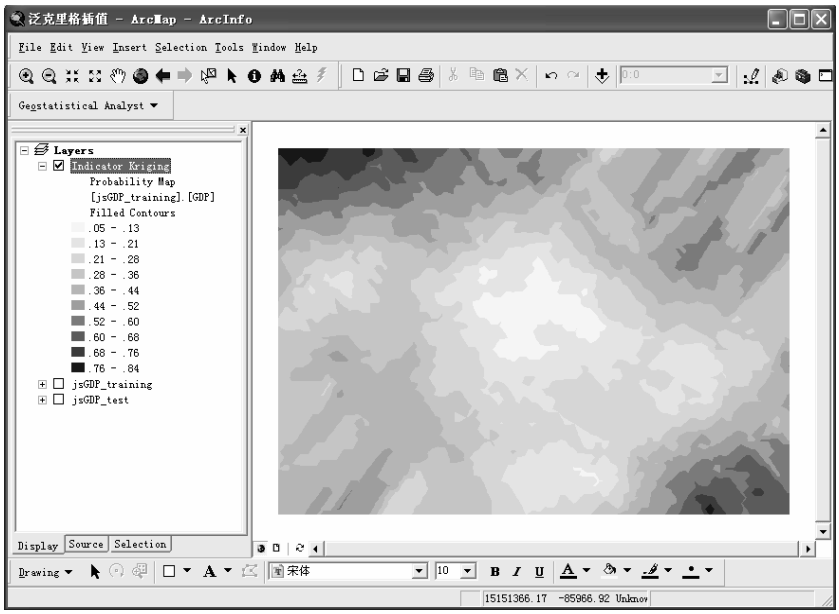


图 10.72　　指示克里格内插生成的概率图

以类似方法可创建指示克里格的标准误差指示图，对 jsGDP_training 创建标准误差指示图的结果如下图 10.73 所示：

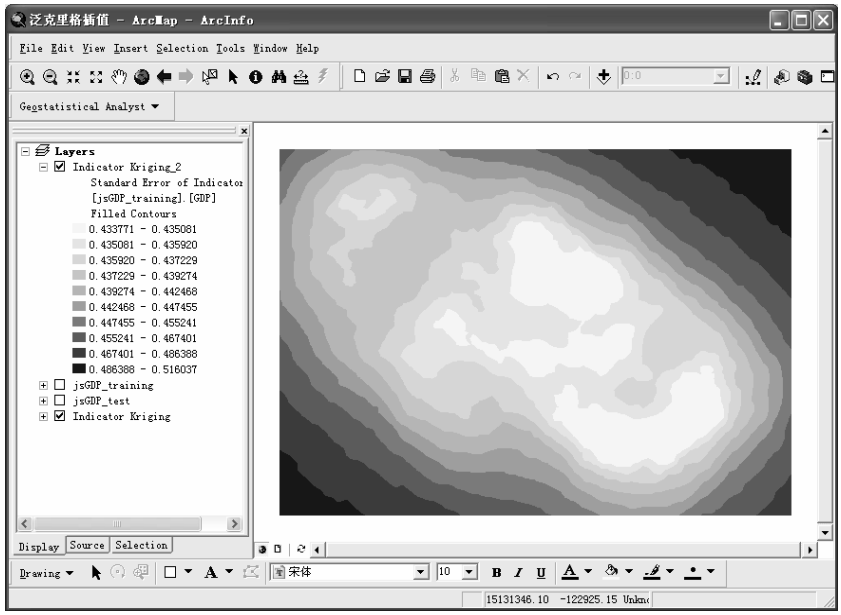


图 10.73 指示克里格内插生成的标准误差指示图

10.4.6 概率克里格插值

ArcGIS 中普通克里格插值包括 2 部分功能：创建概率图（Probability Map）和创建标准误差指示图（Standard Error of Indicator Map）。

1. 创建概率图（Probability Map）

其在 ArcGIS 中的实现过程与指示克里格的方法雷同，具体步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框中，在 Dataset 选择训练数据 jsGDP _training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP _test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

（5）在弹出的对话框中，展开概率克里格（Probability Kriging），在下面的选项中点击概率图（Probability Map），最后点击 Next 按钮。

（6）在弹出的 Semivariogram/Covariance Modeling 对话框中，钩选 Show Search Direction，移动左图中的搜索方向，然后点击 Next 按钮。

（7）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（8）在弹出的 Cross Validation 对话框中，点击 Next 按钮。

（9）在弹出的 Validation 对话框中，点击 Finish 按钮。概率克立格法内插结果（如图 10.74)

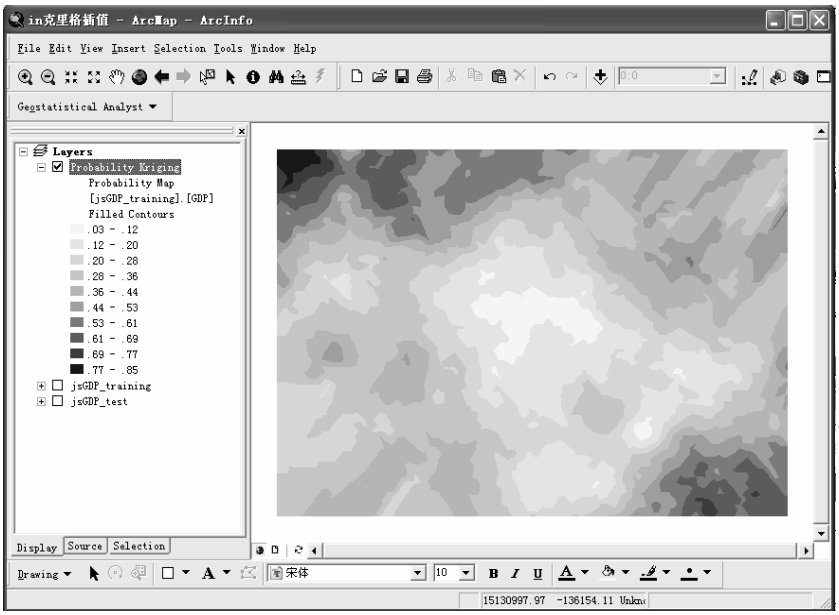


图 10.74　　概率克里格内插生成的概率图

2.　创建标准误差指示图（Standard Error of Indicators）

以类似方法可创建概率克里格的标准误差指示图，对 jsGDP_training 创建标准误差指示图的结果如下图 10.75 所示：

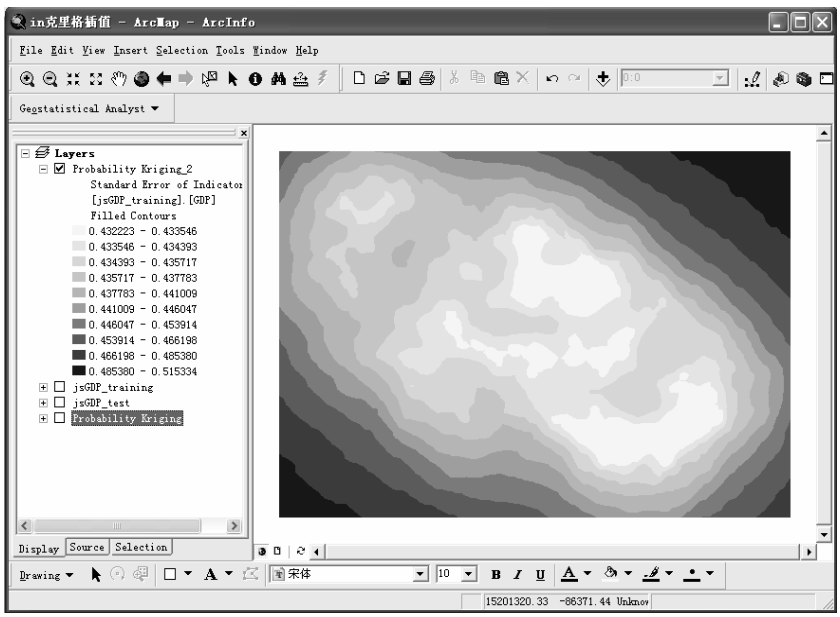


图 10.75 概率克里格内插生成的标准误差指示图

10.4.7 析取克里格插值

如果原始数据不服从简单的分布（高斯或对数正态等），则可选用析取克里格法（Disjunctive Kriging），它可以提供非线性估值方法。ArcGIS 中普通克里格插值包括 4 部分功能：创建预测图（Prediction Map）、创建概率图（Probability Map）、创建标准误差预测图（Prediction Standard Error Map）和创建标准误差指示图（Standard Error of Indicator Map）。

1. 创建预测图（Prediction Map）

其在 ArcGIS 中的实现过程与普通克里格的方法雷同，具体步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框中，在 Dataset 选择训练数据 jsGDP _training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP _test 及其属性 GDP，选择 Kriging 内插方法，最后点击 Next 按钮。

（5）在弹出的对话框中，展开析取克里格（Disjunctive Kriging），在下面的选项中点击预测图（Prediction Map），然后数据集选择 Log 变换，最后点击 Next 按钮。

（6）在弹出的 Normal Score Transformation 对话框(如图 10.76)，点击 Next 按钮。

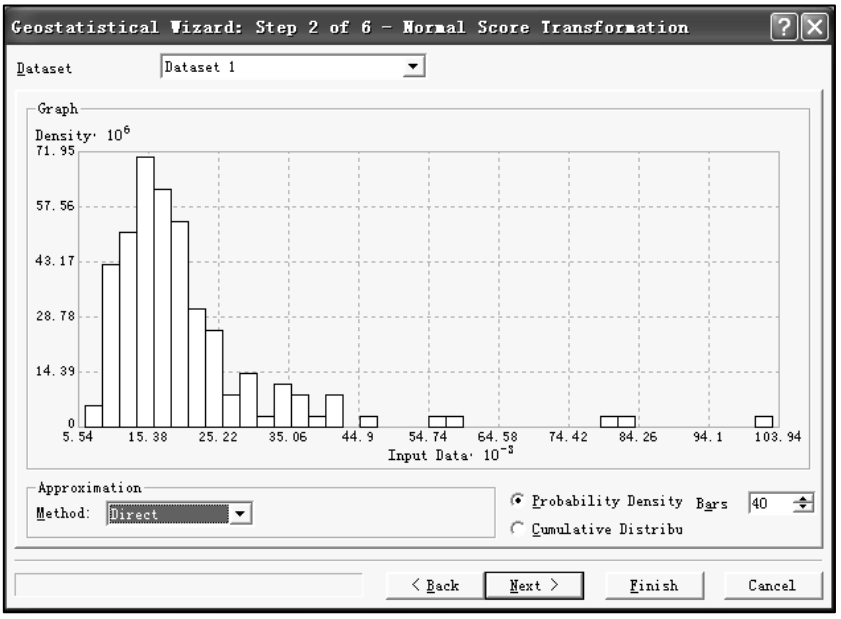


图 10.76　NST 变换对话框设置

（7）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（8）在弹出的 Cross Validation 对话框中，点击 Next 按钮。

（9）在弹出的 Validation 对话框中，点击 Finish 按钮。析取克立格法内插结果如图 10.77 所示。

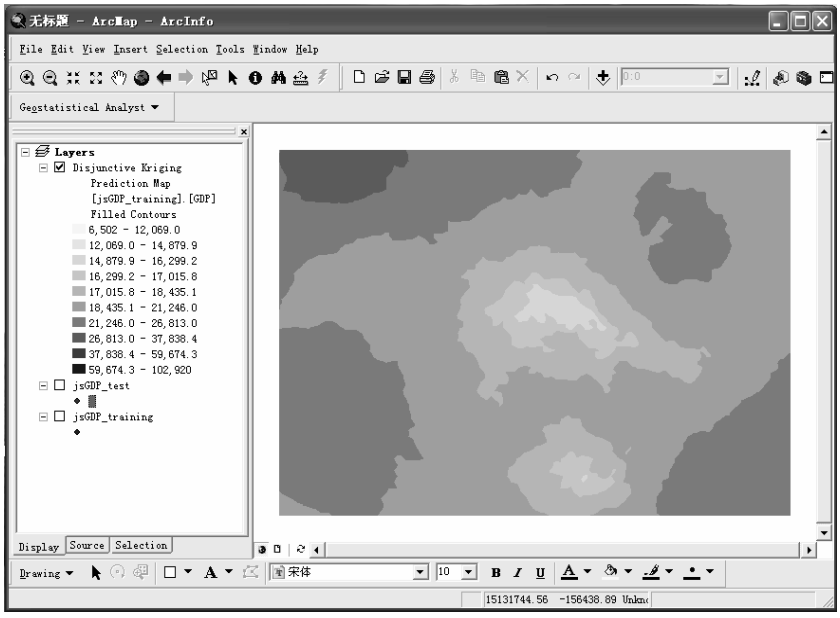


图 10.77　　析取克里格内插生成的预测图

2.　创建概率图（Probability Map）

其在 ArcGIS 中的实现过程与指示克里格的方法雷同，对 jsGDP_training 创建概率图的结果如下图 10.78 所示：

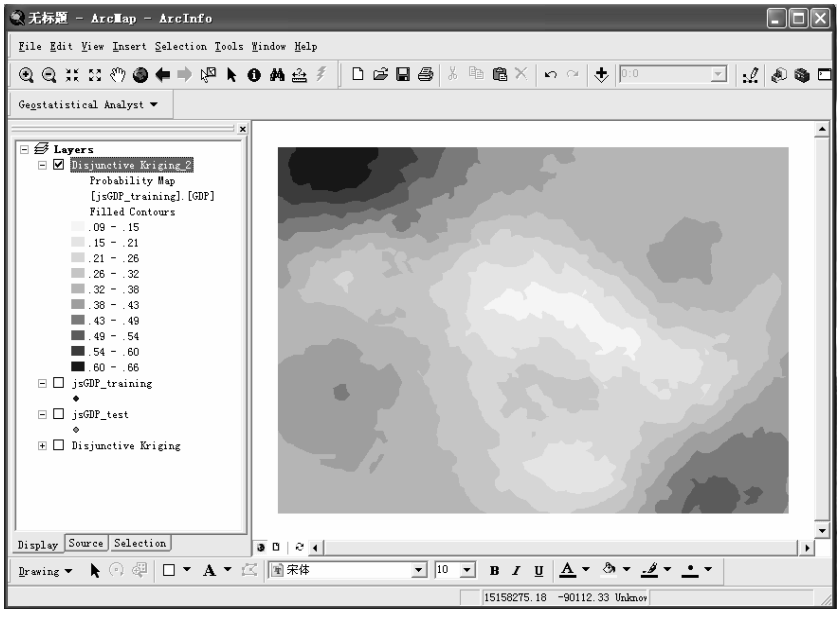


图 10.78　　析取克里格内插生成的概率图

3. 创建标准误差预测图（Prediction Standard Error Map）

其在 ArcGIS 中的实现过程与普通克里格的方法雷同，对 jsGDP_training 创建标准误差预测图的结果如下图 10.79 所示：

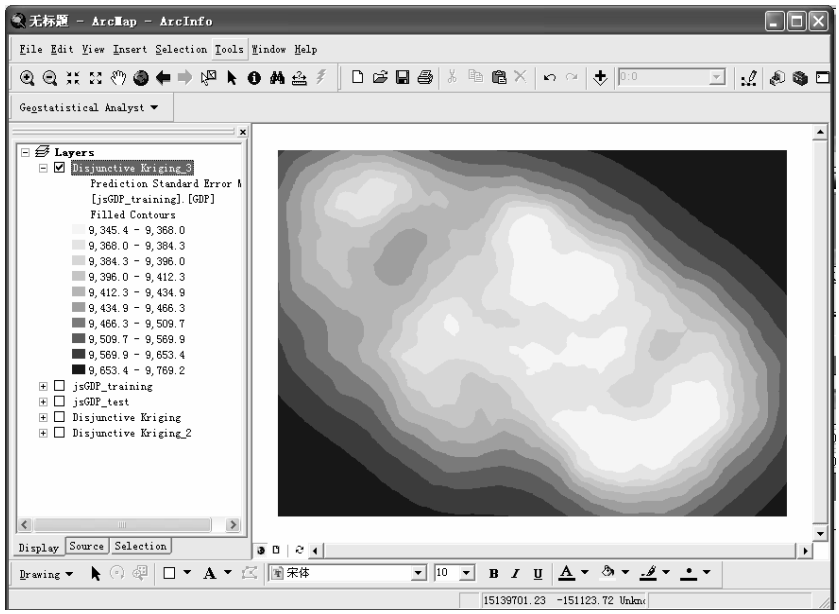


图 10.79 析取克里格内插生成的标准误差预测图

4. 创建标准误差指示图（Standard Error of Indicators）

其在 ArcGIS 中的实现过程与指示克里格的方法雷同，对 jsGDP_training 创建标准误差指示图的结果如下图 10.80 所示：

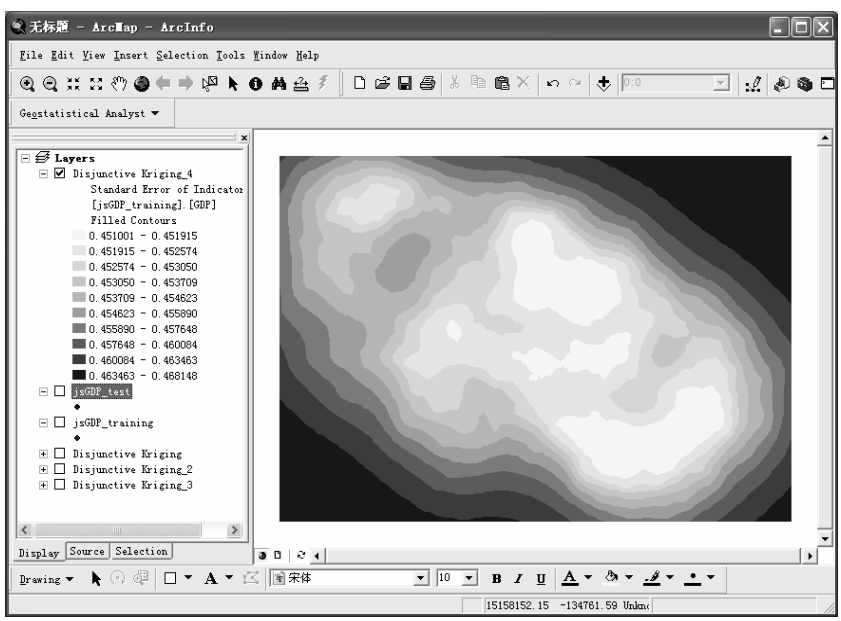


图 10.80 析取克里格内插生成的标准误差指示图

10.4.8 协同克里格插值

当同一空间位置样点的多个属性之间存在某个属性的空间分布与其它属性密切相关，且某些属性获得不易，而另一些属性则易于获取时，如果两种属性空间相关，可以考虑选用协同克里格法。协同克里格法把区域化变量的最佳估值方法从单一属性发展到二个以上的协同区域化属性。但它在计算中要用到两属性各自的半方差函数和交叉半方差函数，比较复杂。

其在 ArcGIS 中的实现步骤为：

（1）在 ArcMap 中加载 jsGDP _training 和 jsGDP _test。

（2）右击工具栏，启动地理统计模块 Geostatistical Analyst。

（3）单击 Geostatistical Analyst 下的 Geostatistical Wizard 命令。

（4）在弹出的对话框(如图 10.81)中，在 Dataset1 选择训练数据 jsGDP _training 及其属性 GDP，在 Validation 中选择检验数据 jsGDP _test 及其属性 GDP，然后点击 Dataset2，选择训练数据 jsGDP _training 及其属性 POPULATION,最后选择 CoKriging 内插方法，最后点击 Next 按钮。

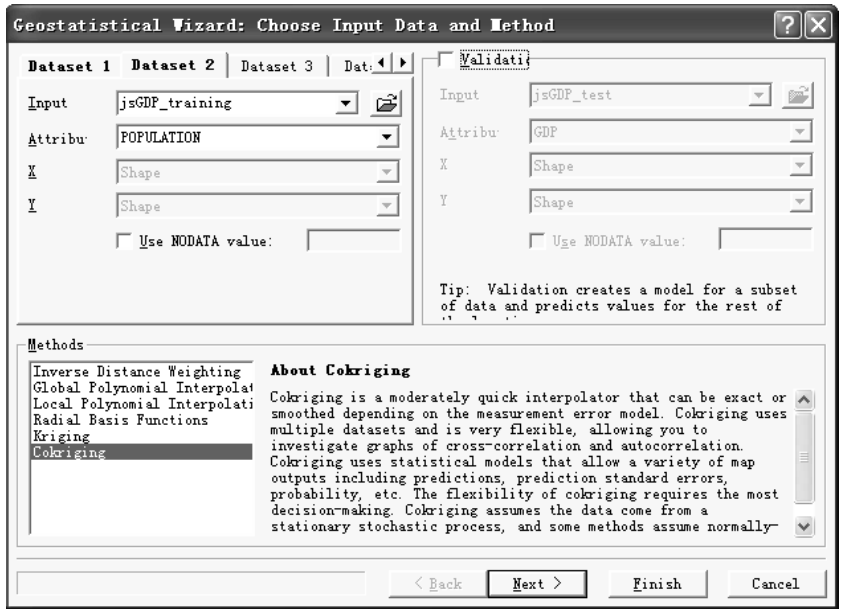


图 10.81　　输入数据和方法选择的对话框

（5）在弹出的对话框中，根据数据情况可选择上述六种克里格方法中的任意一种，此例中选择展开普通克里格（Ordinary Kriging），在下面的选项中点击预测图（Prediction Map），最后点击 Next 按钮。

（6）在弹出的 Searching Neighborhood 对话框，点击 Next 按钮。

（7）在弹出的 Cross Validation 对话框中，点击 Finish 按钮。协同克立格法内插结果（如图 10.82）。

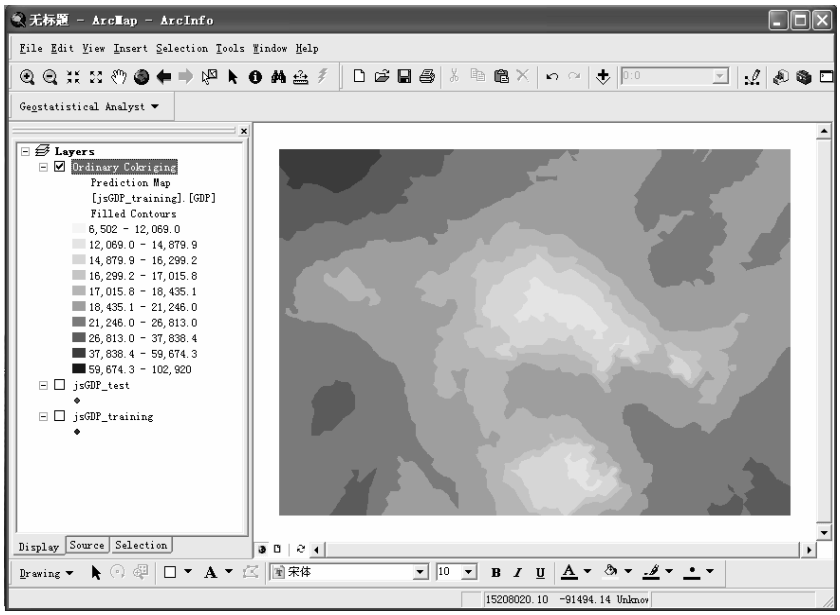


图 10.82　　协同克里格内插生成的预测图