ArcGIS 地统计学习指南

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。

地统计学的功能在地统计分析模块的都能实现，包括：

（1）ESDA：探索性空间数据分析，即数据检查；

（2）表面预测（模拟）和误差建模；

（3）模型检验与对比。

地统计学起源于克里格。当时他用此法预测矿产分布，后来经过别人改进修改发展成为现在所用的克里格方法。虽然空间数据分析还有其他方法，如 IDW（反距离加权插值法）等，但克里格方法是最主要、最常用的空间分析方法，下面也以此法为主进行。

1.2 表面预测主要过程

ArcGIS 地统计扩展模块的菜单非常简单，如下所示，但由此却可以完成完整的空间数据分析过程。



一个完整的空间数据分析过程，或者说表面预测模型，一般为。拿到数据，首先要检查数据，发现数据的特点，比如是否为正态分布、有没有趋势效应、各向异性等等（此功能主要由 Explore Data 菜单及其下级菜单完成）；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；（后两种功能主要由 Geostatistical Wizard…菜单完成）。Create Subsets…菜单的作用是为把采样点数据分成两部分，一部分作为训练样本，一部分作为检验样本。

下面将按上述表面预测过程进行叙述。

（注：[1]文章示例中所使用的数据为 ArcGIS 扩展模块中所带的学习数据（某地测得的臭氧含量样本），整个过程均使用此数据；[2]文章以操作方法介绍为主，所涉及到的地统计方法和基本理论一般未进行解释，可查阅相关地统计理论资料；操作中所用到的某些参数为地统计中的标准名称的也未进行解释。）

我们下面的任务是根据测量所得到的某地臭氧浓度数据进行全区的臭氧浓度预测。首先检查数据的特点，然后根据数据特点用不同参数进行表面模型预测，随后比较不同模型的精确程序，选择最佳模型，最后制作成果图。

我们下面的任务是根据测量所得到的某地臭氧浓度数据进行全区的臭氧浓度预测。首先检查数据的特点，然后根据数据特点用不同参数进行表面模型预测，随后比较不同模型的精确程序，选择最佳模型，最后制作成果图。

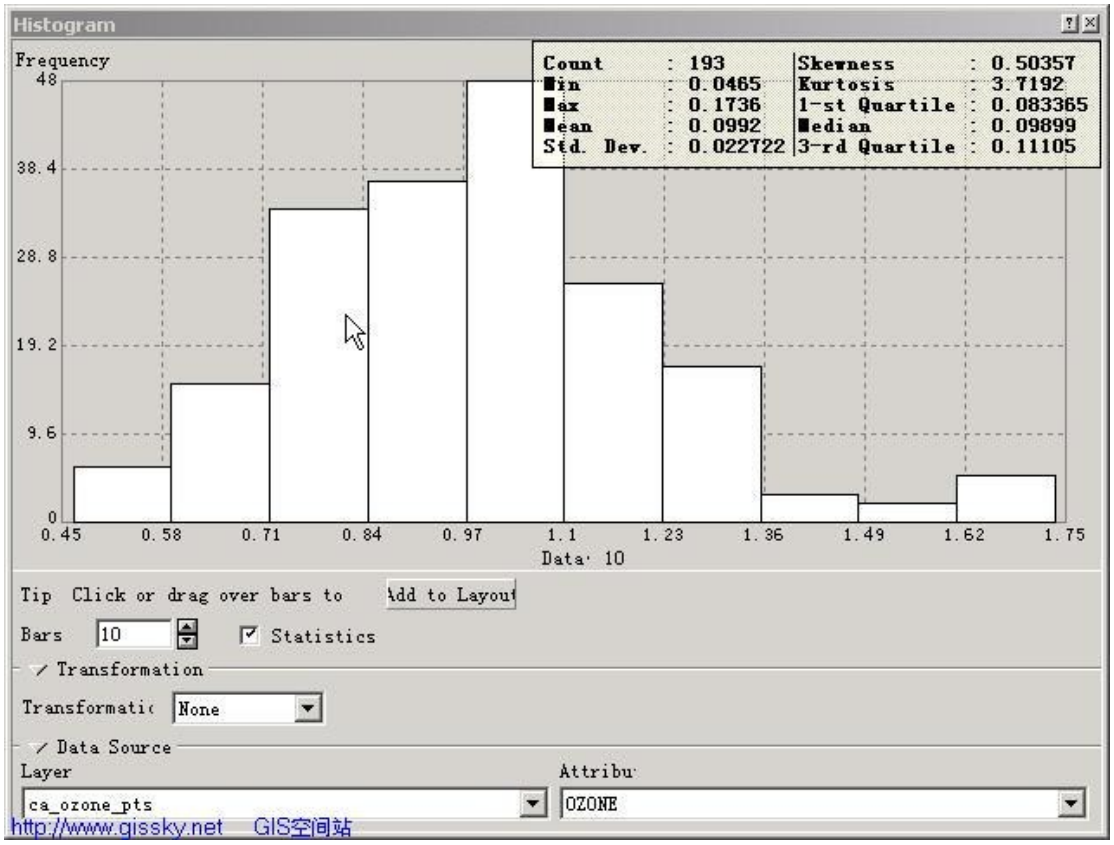
1.3 数据检查，即空间数据探索分析（ESDA）

此功能主要通过 Explore Data 菜单中实现。

扩展模块提供了多种分析工具，这些工具主要是通过生成各种视图，进行交互性分析。如直方图、QQ plot 图、半变异函数/协方差图等。

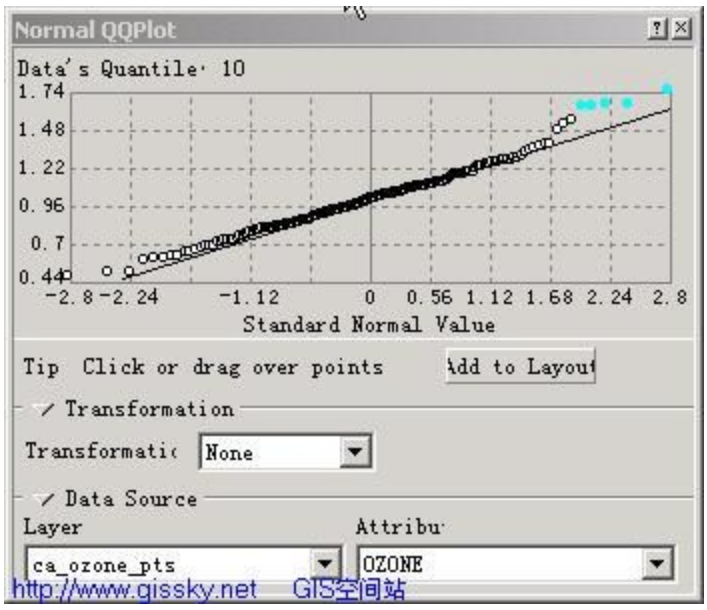
（1）直方图显示数据的概率分布特征以及概括性的统计指标。

下图中所展示的数据，中值接近均值、峰值指数接近 3。从图中观察可认为近似于正态分布。克里格方法对正态数据的预测精度最高，而且有些空间分析方法特别要求数据为正态分布



（2）正态 QQ Plot 图：

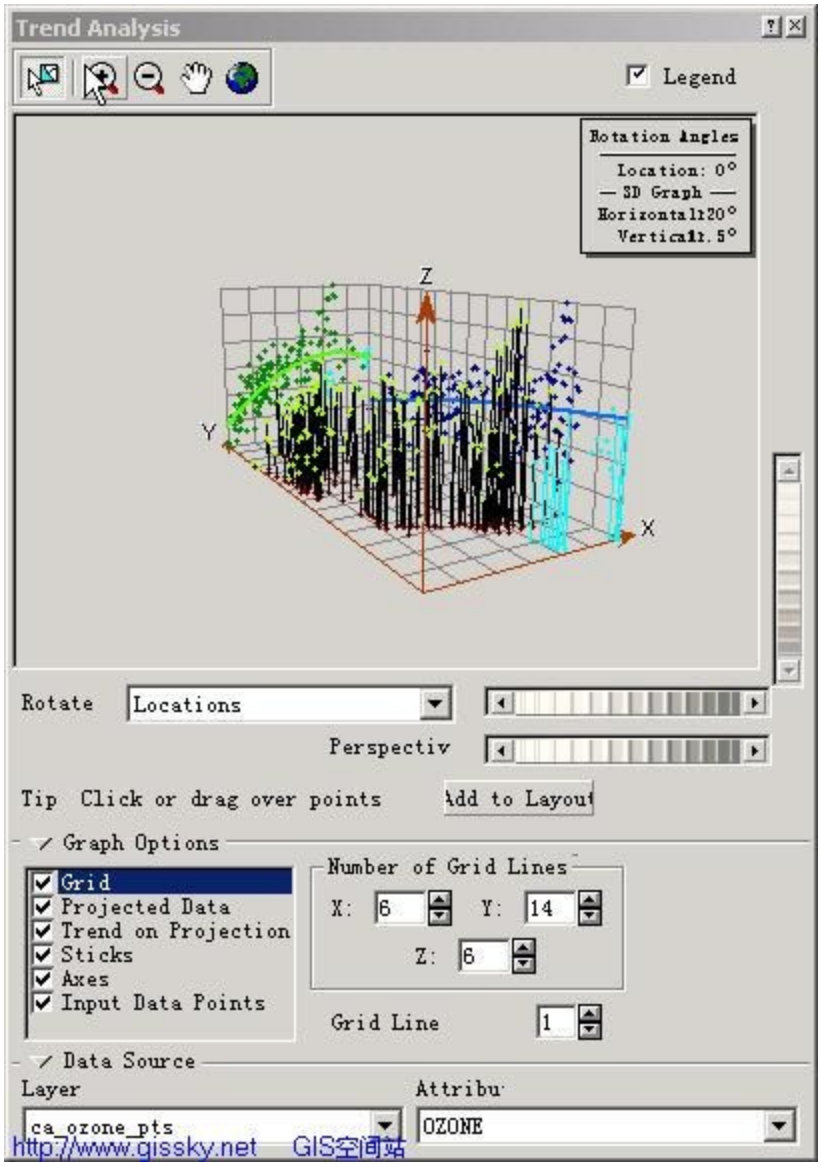
检查数据的正态分布情况。作图原理是用分位图思想。直线表示正态分布，从图中可以看出数据很接近正态分布（左上角几个偏离的点被选中）。



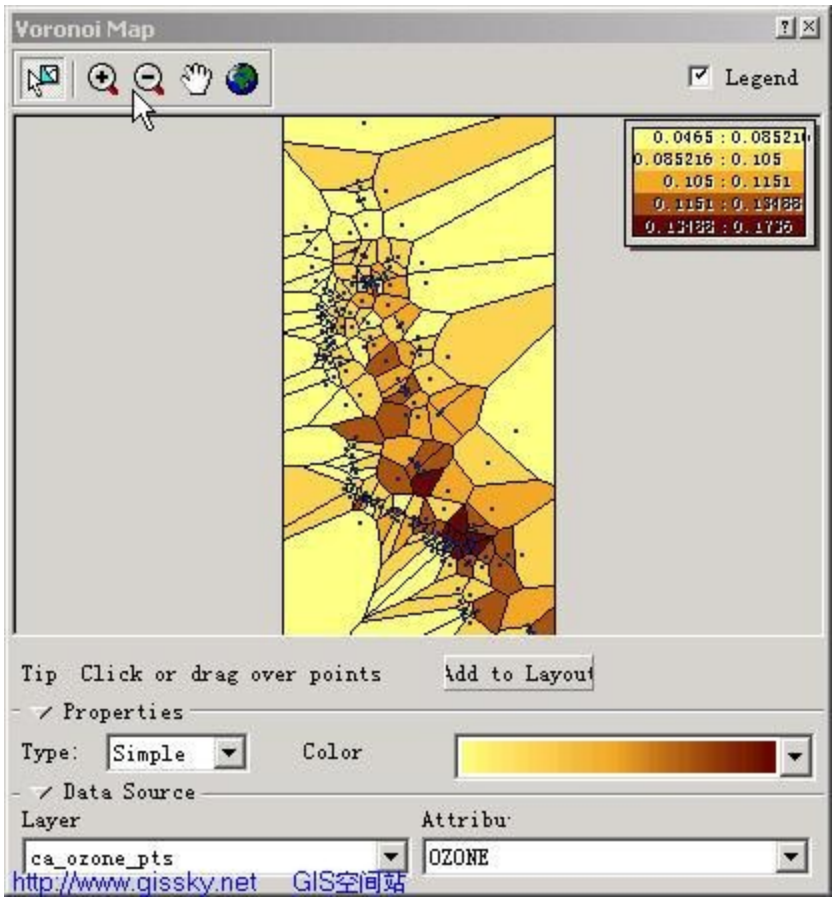
（3）趋势分析图。

蓝线表示南北方向，呈水平，可见南北方向无趋势。绿线表示东西方向，呈倒"U"形，

可用二阶曲线拟合，在后面进行表面预测时将会去除。
点击 Rotete 右边的方向旋转箭头（横向箭头），可旋转趋势图，更明显地显示某一个方向的趋势。



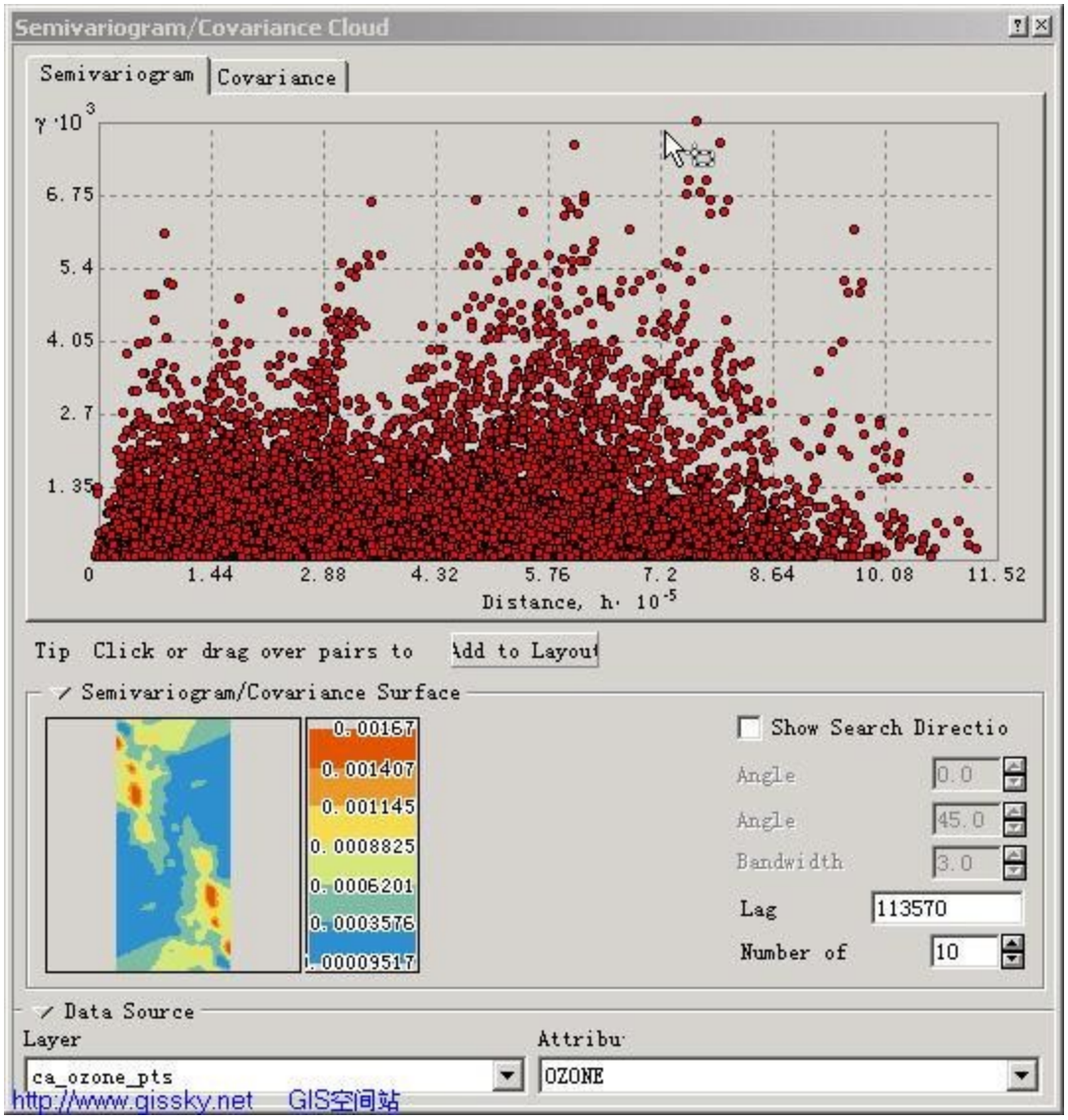
（4）Voronoi 图
用来发现离群值。Voronoi 图的生成方法：每个多边形内有一个样点，多变形内任一点到该点的距离都小于其他多边形到该点的距离，生成多边形后。某个样点的相邻样点便会与该样点的多边形有相邻边。至于多边形值的计算有多种方法，可以用生成多边形的样点值作为多边形的值（Simple 方法），也可以以相邻样点的平均值为多边形的值（Mean 方法），具体计算方法可以在 Type 下拉菜单中选择。



（5）半变异函数/协方差函数。

该图可以反应数据的空间相关程度，只有数据空间相关，才有必要进行空间插值法。图表的横坐标表示任两点的空间距离，纵标表示该两点的半变异函数值。根据距离越近越相似的原理，因而 x 值越小，y 值应该越小。

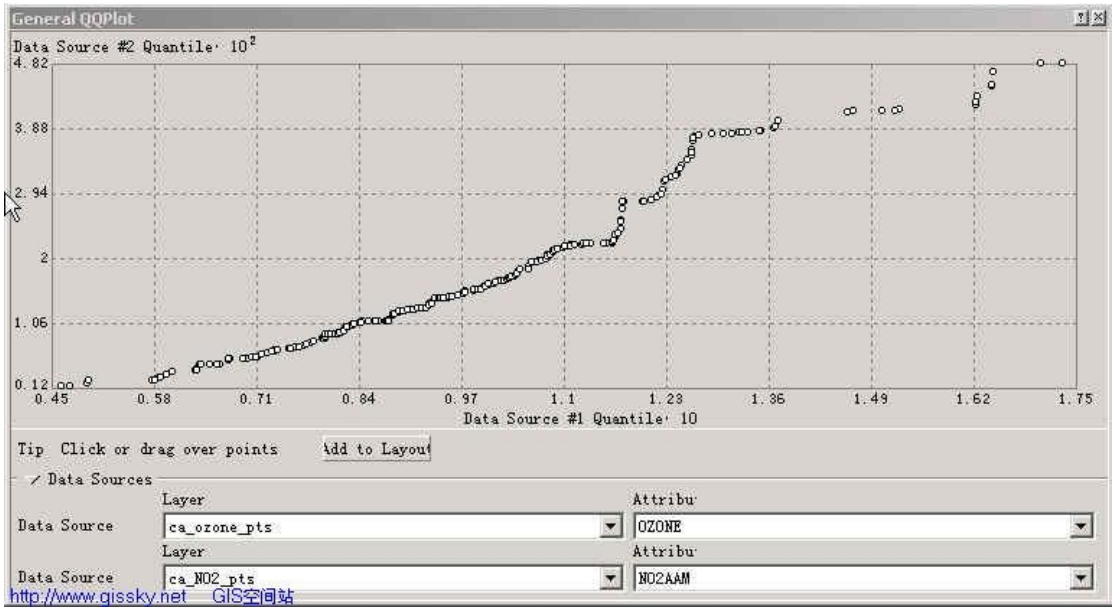
如果任意两点的值都要计算，当采样点很多时，数据量便很大，因而根据距离和方向对样点距离进行了分组。下列参数便是为此要求而设置：Lag，步长值；Number of，步长组数。步长值和步长组数之乘积应小于采样点区域的坐标范围的一半。如下图。



最后的两个图表是针对两个数据集而言的。

（6）普通 Qqplot 分布图

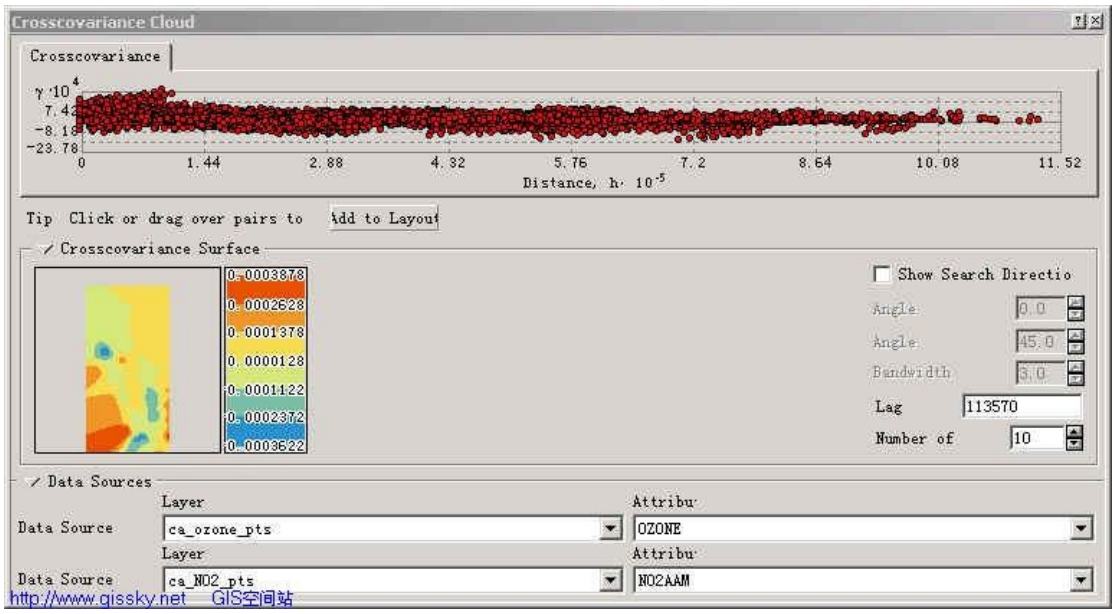
评估两个数据集分布的相似程度。利用两个数据集中具有相同累积分布值的数据值来作图。



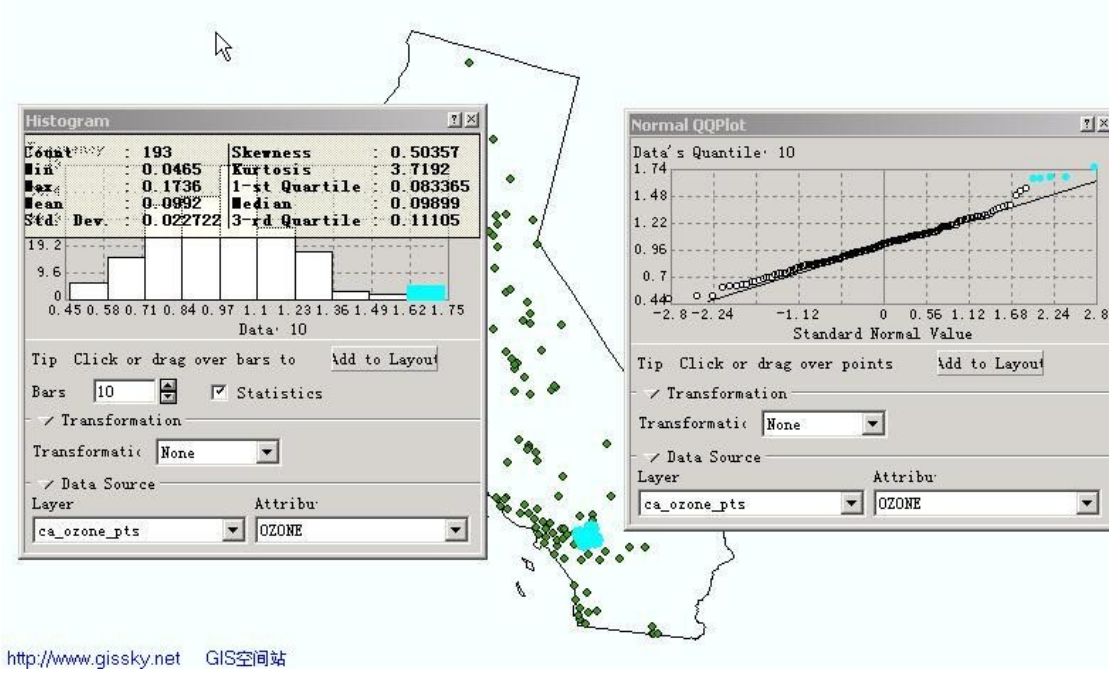
（7）正交协方差函数云。

横坐标：两点间的距离；

纵坐标：两点间的距离所对应的样点对的理论正交协方差。



这些图彼此相关联，并与 Arc Map 中的图层相关联。即，在某个分析图中选择某些样点，在 Arc Map 图层及其他分析图中同样会选中这些点。如下图。



后面将在数据检查的基础上进行表面预测。

1.4 制作表面预测图：

通过上面的数据检查，发现数据接近正态分布、有空间相关、无离群值、东西方向有倒"U"形趋势。决定使用普通克里格方法进行表面预测。下面的步骤是针对此数据进行的。

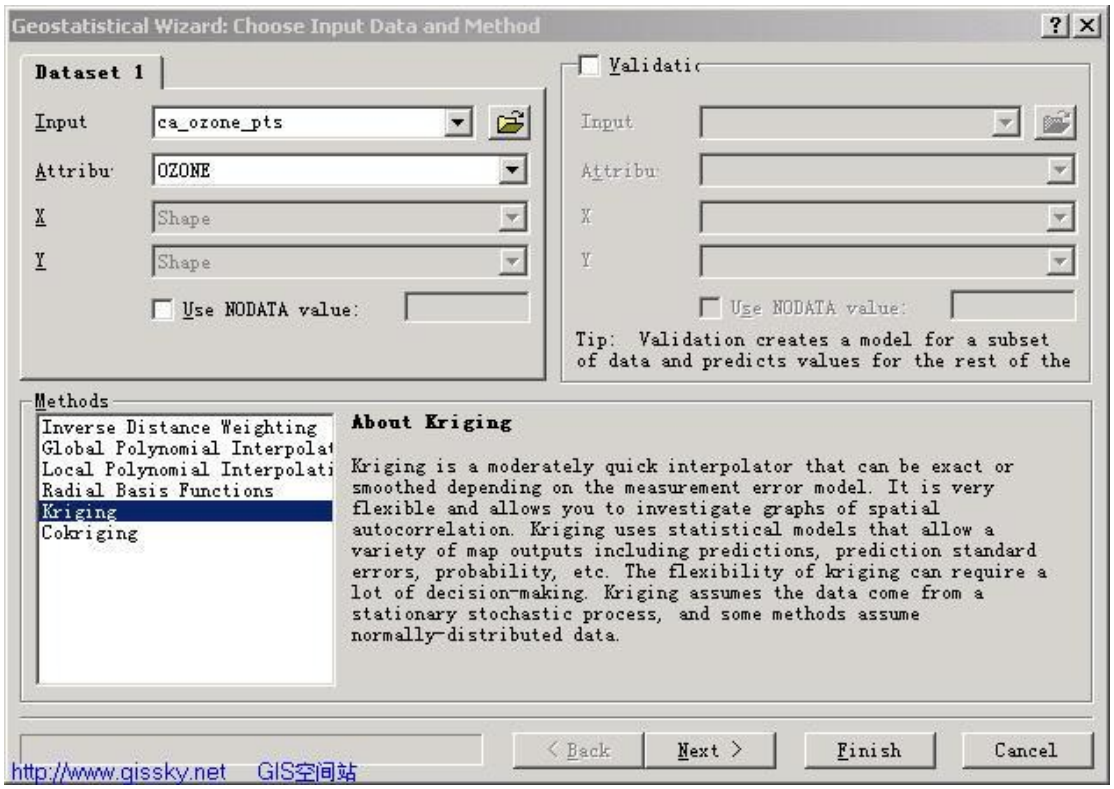
将使用地统计模块的第二个菜单 Geostatistical Analyst……。

第一步：选择输入数据和方法面板（Choose Input Data and Method）

选择使用的数据及其属性：分别在 Input 和 Attribute 中选择

选择预测方法：在 Methods 中选择。预测方法的选择要根据数据分析的结果而定。现在假如选择 Kriging 方法（其实所谓地统计方法，最主要并且用的最多的就是 Kriging 方法的几种变化形式）。

Validate 是个可选项，选择使用何种方法对生成的预测图进行检验，如果想用检验方法，则选中此项并设置检验数据集和属性；如果对结果进行交叉检验，则不要选择此项。

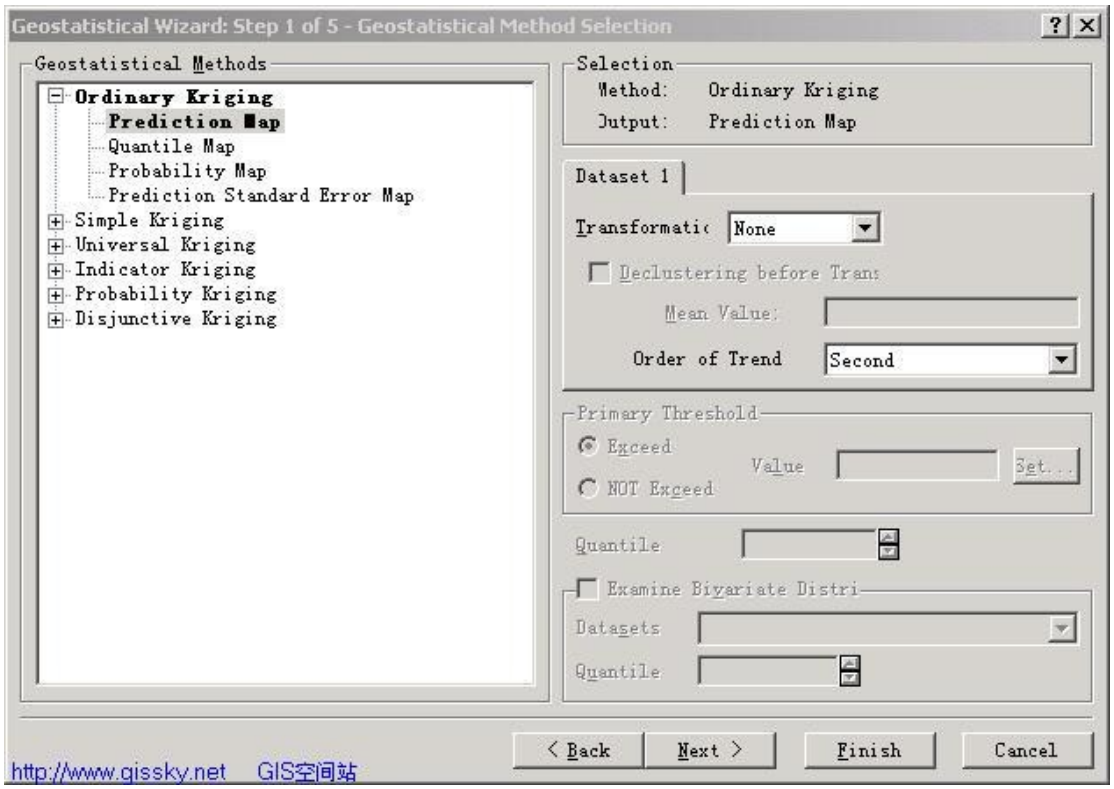


第二步：地统计方法选择面板（Geostatistical Method Selection）

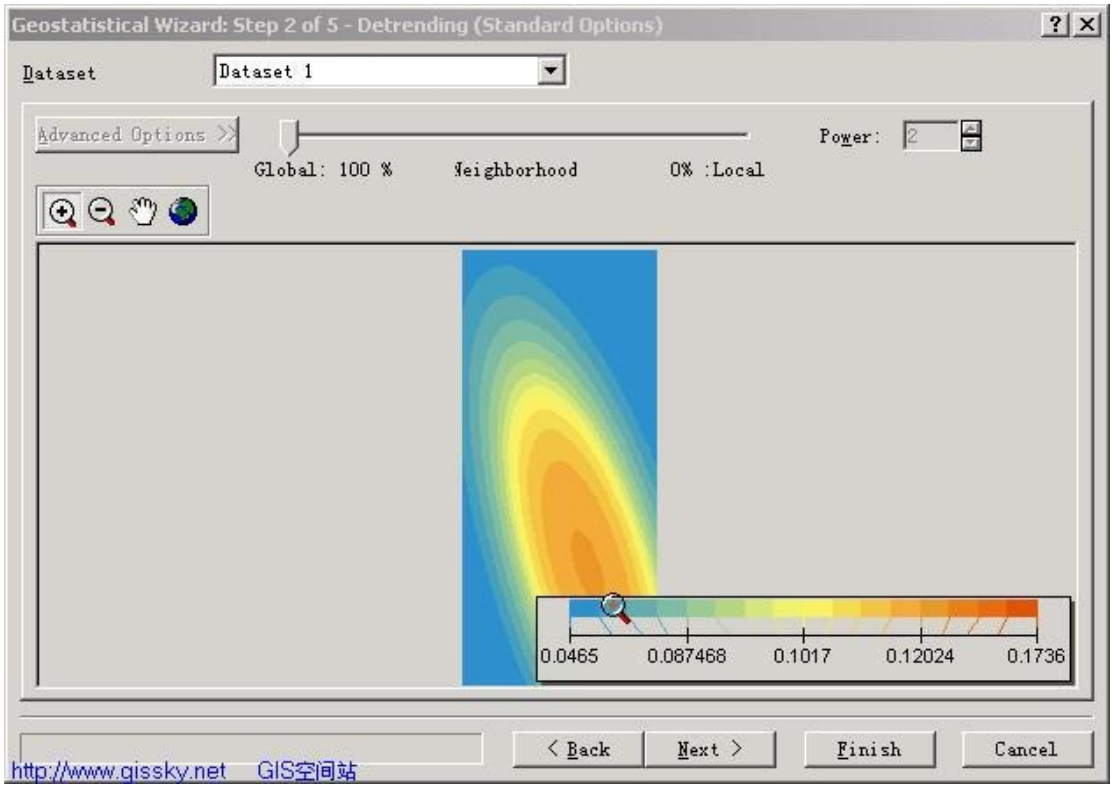
选择 Ordinary Kriging 中的 Prediction Map，即使用普通克里格方法生成一个表面预测图。普通克里格方法是最常用的地统计分析方法。其他几种依次为简单克里格、泛克里格、指示克里格、概率克里格、析取克里格。这集中克里格的区别是由于克里格的形式及其数据特点的不同。

Transmition 选项：对数据集进行转换，由于某些方法要求数据正态分布，因此如果数据与正态分布差距很大，可以在此选择一种方法对数据进行转换。

Order of trend：如果数据在某方向上存在趋势，则为了提高预测的准确性，一般要剔除趋势。在此处选择趋势方程的阶数：线性、一阶、或无趋势等。数据的趋势有无以及阶数在数据检查时得到，即用 Explore Data 菜单下的 Trend analysis 来分析得到。

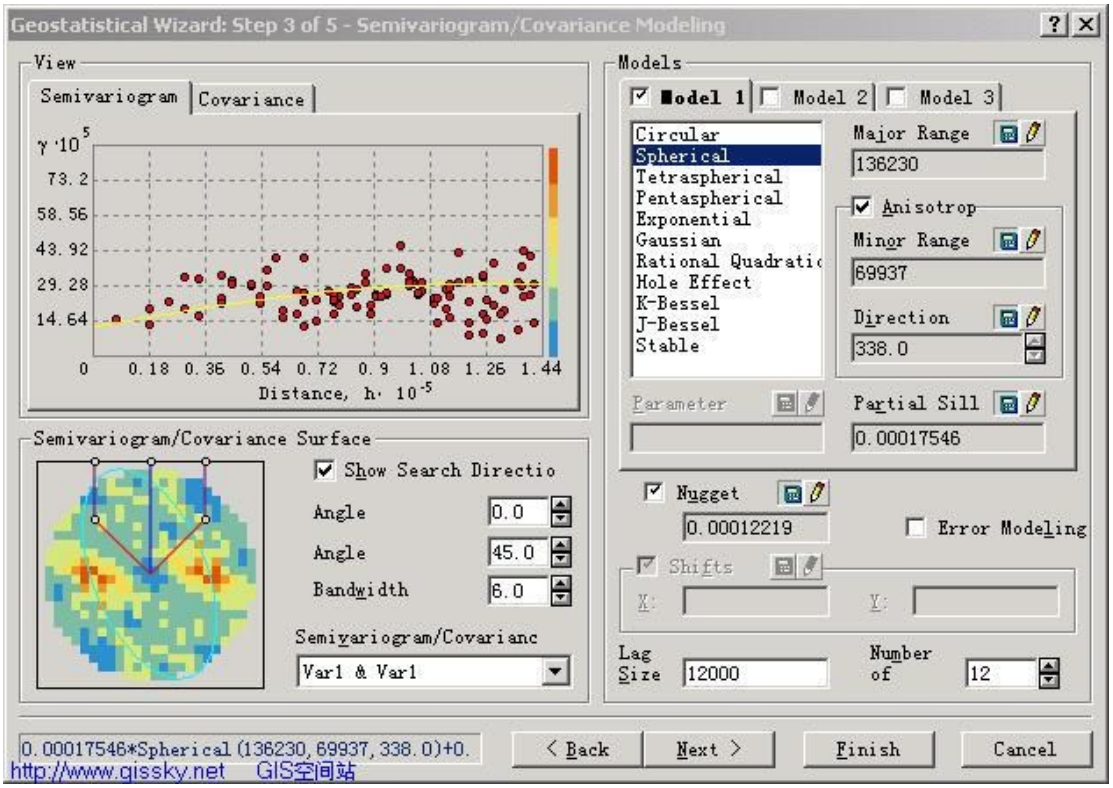


第三步：趋势剔除面板（Detrending） 此面板只有在第二步中选择了 Order of trend 选项是才会出现，一般为缺省即可。

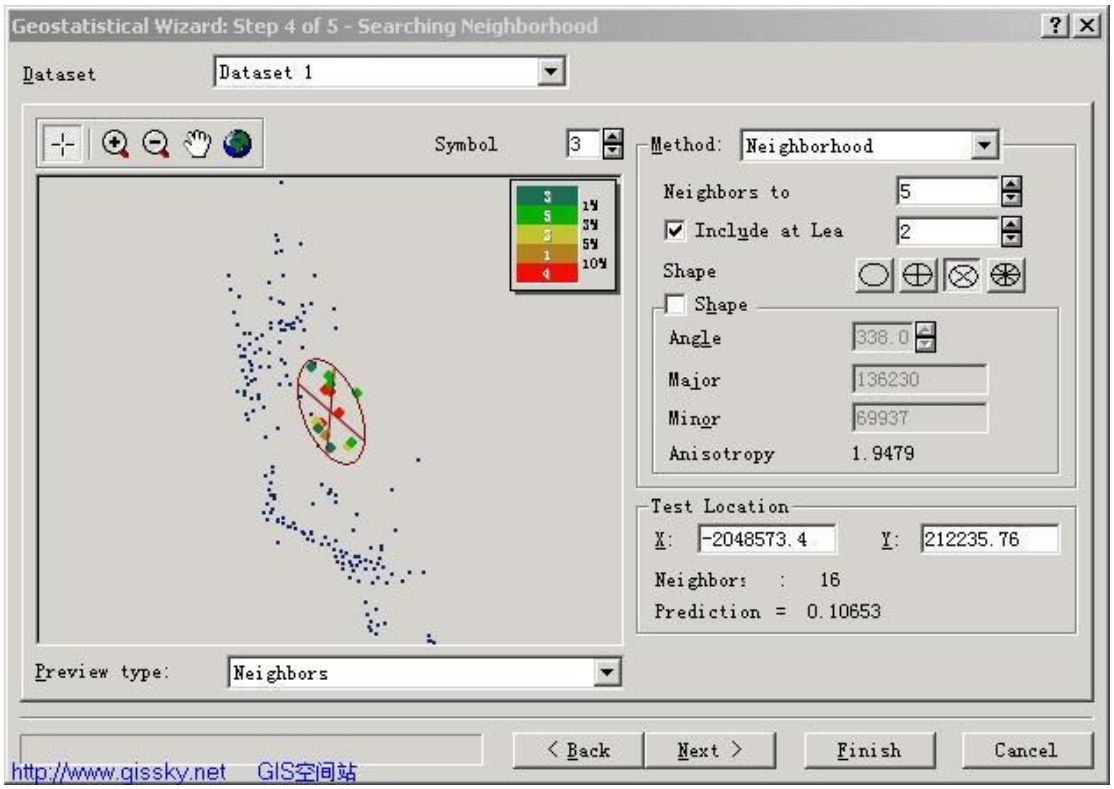


第四步：半变异函数/协方差模型面板（Semivariogram/covariance Modeling） 此步的主要功能为半变异函数建模，是预测过程中的实质性阶段。在此面板中需要社定许多与拟合半变异函数相关的选项以及半变异函数的参数。是克里格预测中十分关键的 Semivariogram/covariance 部分显示的是拟和的模型，黄线即半变异函数曲线。

Models 部分：model1,model2,model3 表示可以用多个通用函数来拟和半变异函数模型。
如果数据为各向异性，则需要选中 Anisotropy（其实大多数空间数据是各向异性的，各向同性只是相对的），当选中此选项时，黄线变为多条，表示多个方向的拟合函数。
Show Search Direction 选项选中后，表示只搜索某个方向的半变异函数。
Nugget：块金值，函数参数之一，即函数与 y 轴相交的 y 值。
Error Modeling：如果数据中有测量误差（比如一起原因等）的话，则选中此项，预测表面将光滑许多。

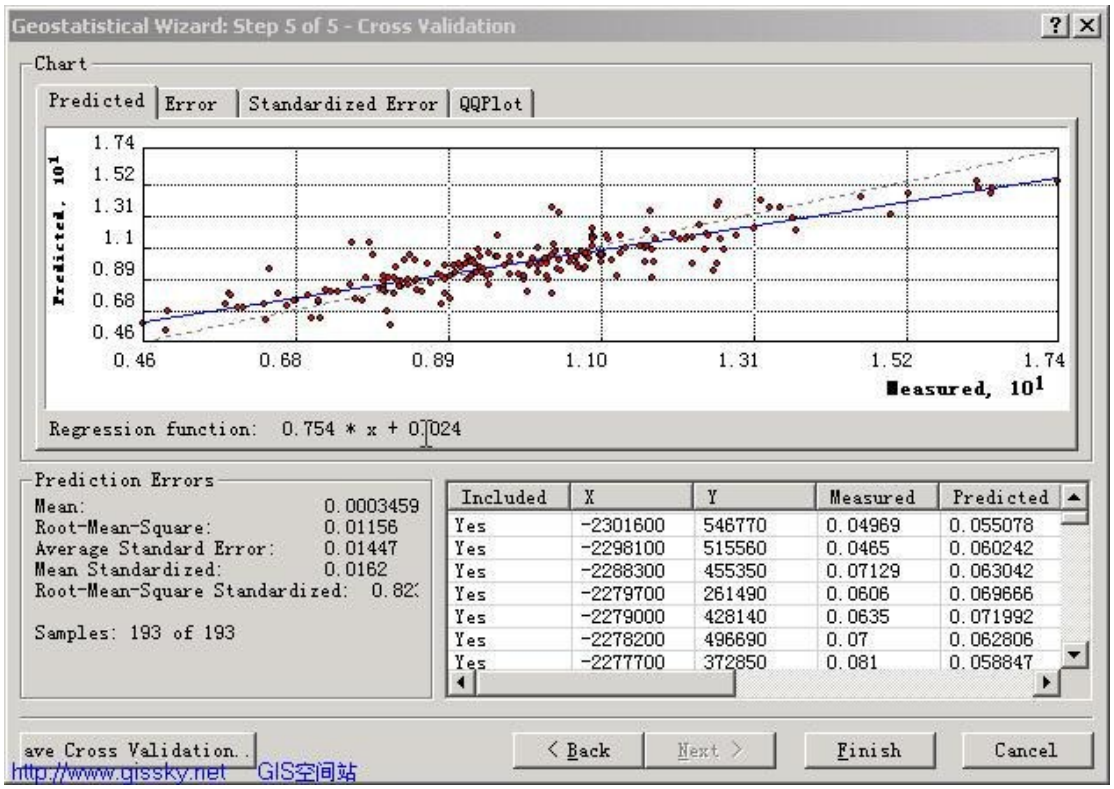


第五步：搜索区域面板（Searching Neighbourhood）
此面板的主要功能是设定预测某点数值时如何搜索邻近的已测量点。
主要有样点数（neighbours to）和搜索形状（shape）两个选项。
Neighbours to：最大搜索数目，离预测点太远的样点对预测无意义。
Include at least：最小样点数目。
Shape：设置如何搜索样点，有图解。



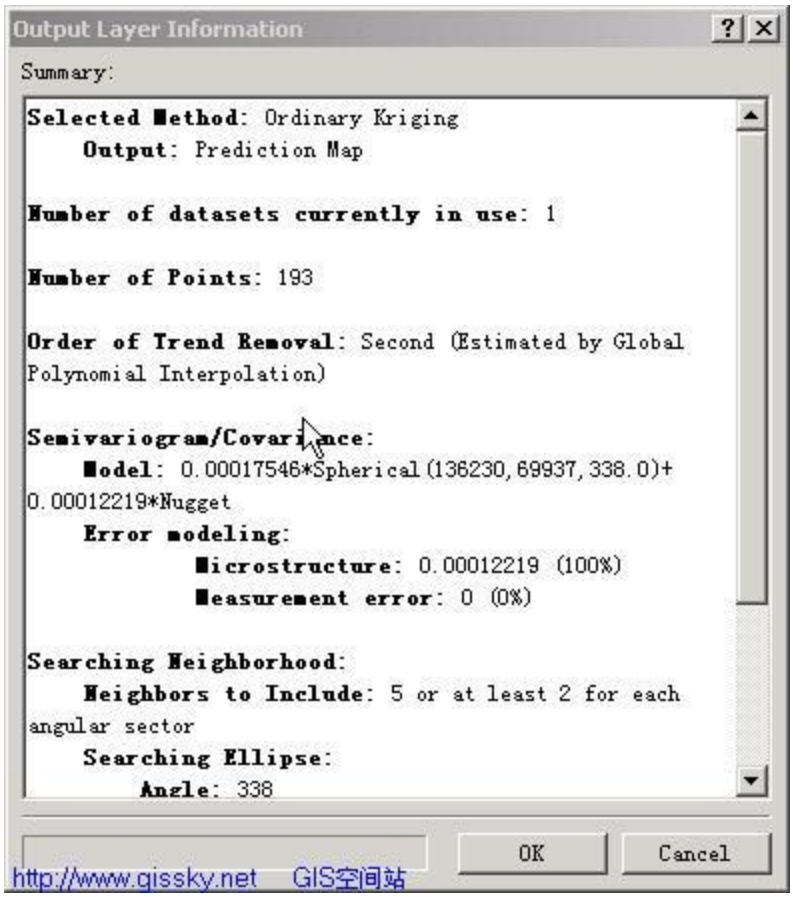
第六步：交叉验证面板（Cross Validation）
在此面板中查看预测的精度，有四个图表，现以最左边的"预测"图表进行说明。
图表的横坐标为测量制值，纵坐标为预测值，最理想的情况是数据呈 1:!线，即图中的破折线。

左下方的预测误差（precited error）项是预测误差的一些统计值，可很好的体现预测的好坏。其中，Mean:0.0005718（预测误差的均值）；Root-Mean-Square:0.01154（预测误差的均方根）；Average Standard Error:0.01456（平均预测标准差）、Mean Standardized:0.02688（平均标准差）；Root-Mean-Square Standardized:0.8463（标准均方根预测误差）。其中前四项越小越好，最后一项越接近 1 越好。
右下方的项含有每个点的误差、标准差等数据，

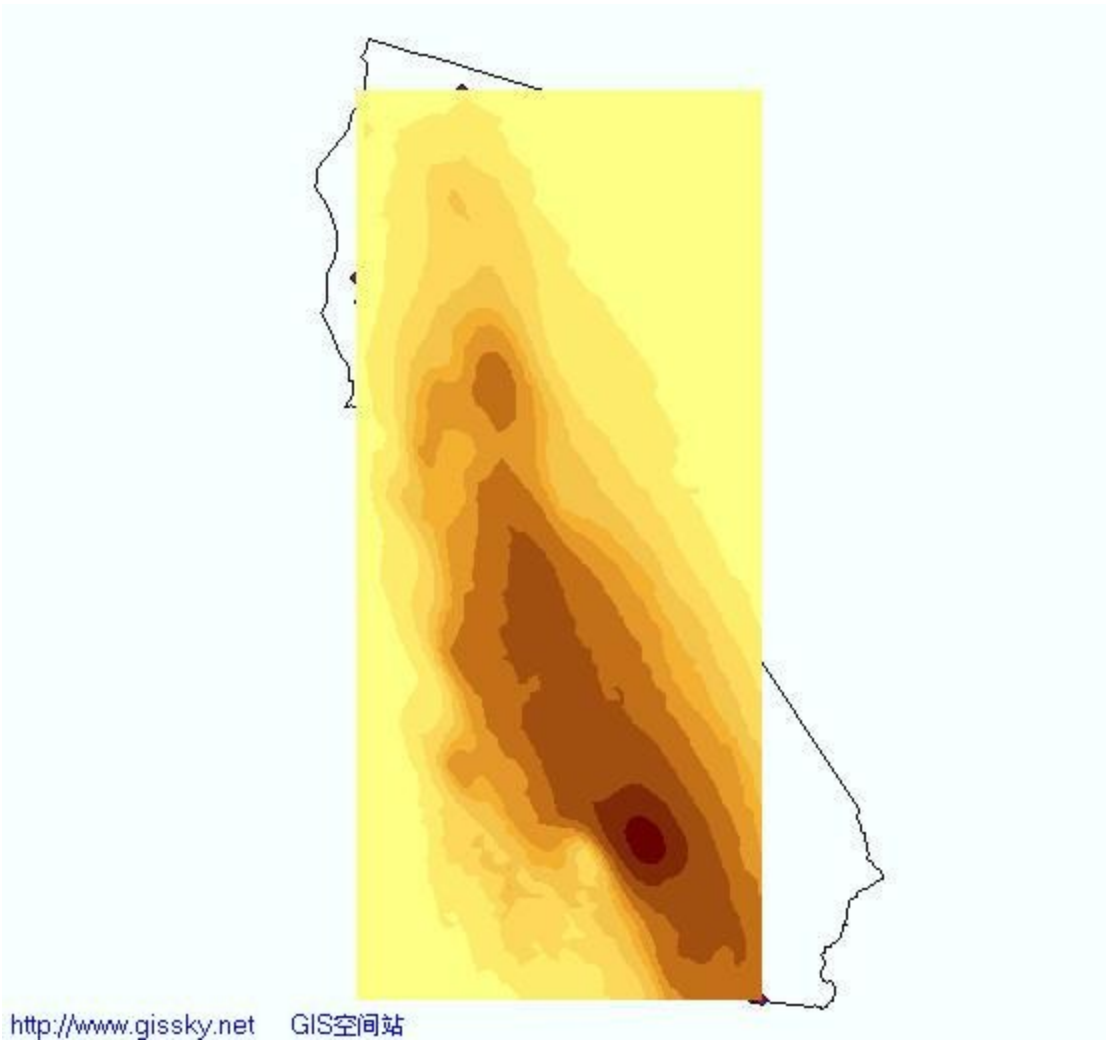


第七步：数据图层信息面板（Output Layer Information）

该面板中显示了在数据预测过程中设置的参数，可以查看。

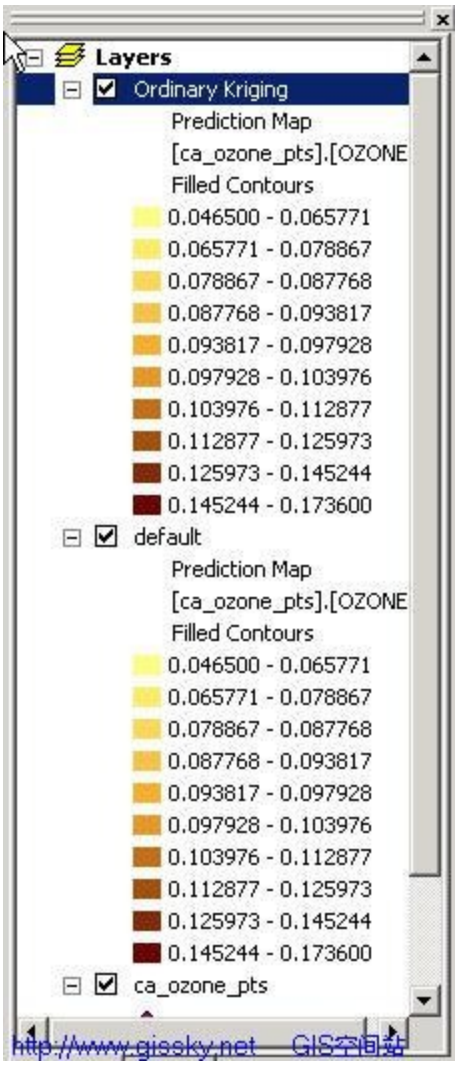


点击 OK，即可生成预测图。

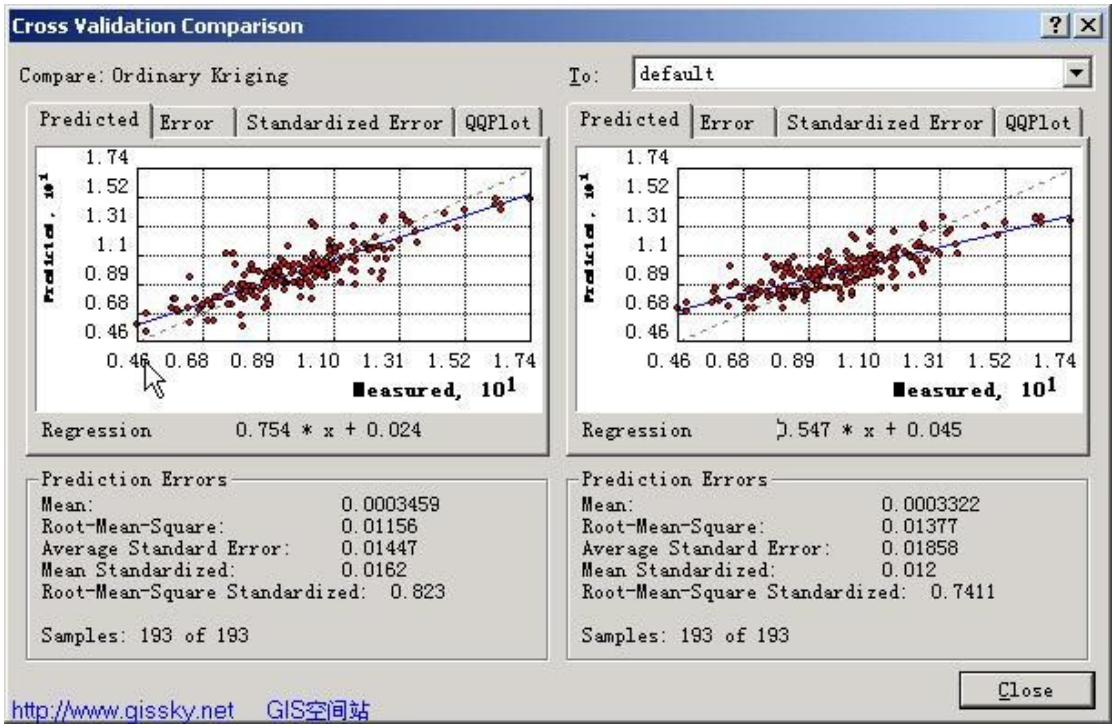


1.5 模型比较

一般情况下，有时候某些参数难以判断，因而会生成几个预测表面，然后比较不同表面的精度，选择精度最高的作为结果。（Ordinary Kriging 表面是用上述过程中的方法生成的预测表面，default 是用缺省的参数得到的预测表面）



右键点击 Ordinary Kriging 并选择 Compare…，即会出现下面的检验面板。To 后面即为要对比的预测表面。通过下面的预测参数，很容易便可看出，Ordinary Kriging 的精度明显高于 Default。



1.6 最终成果展示

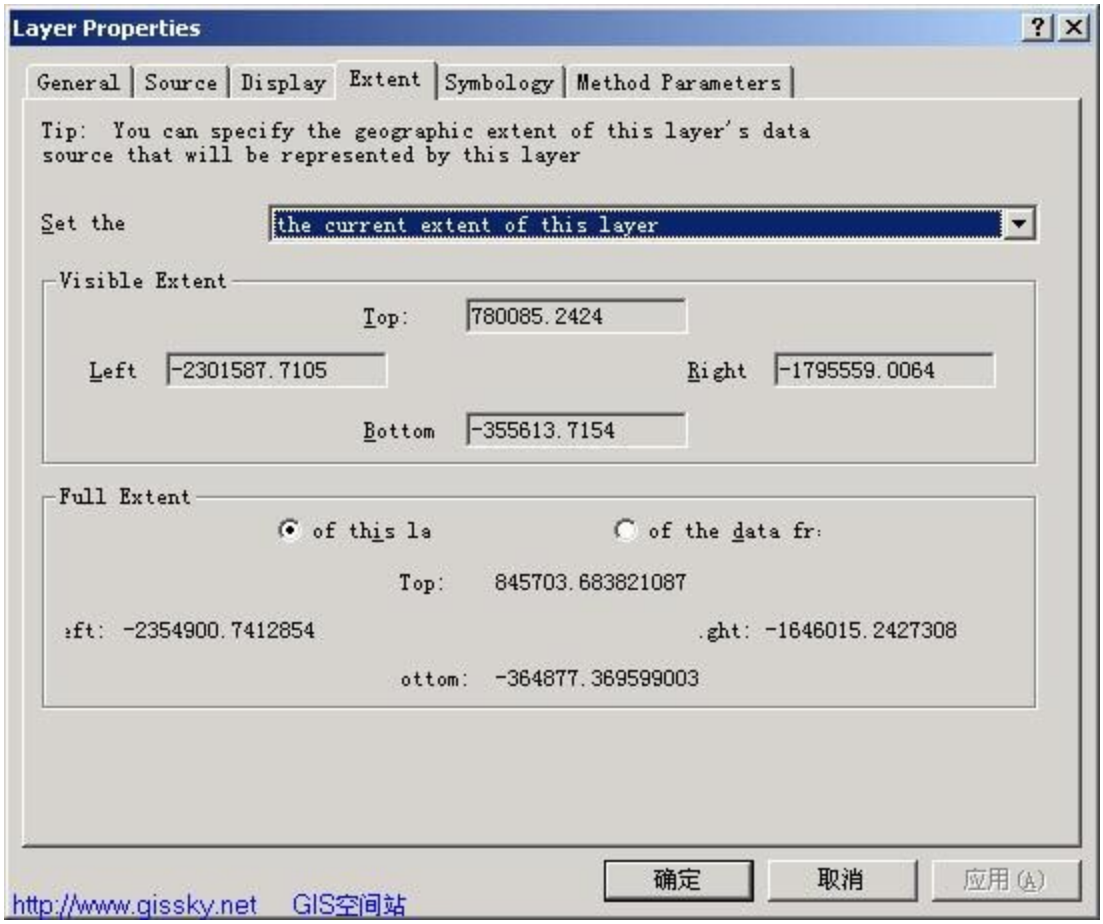
缺省情况下，生成的预测图按照采样数据的坐标范围显示成一个矩形。（如前面所示）现在要把它的范围显示到州界的范围。思路为先把预测表面外推，覆盖整个州界，然后再用州界进行限定，把表面限制在州界的范围。

第一步：外推。

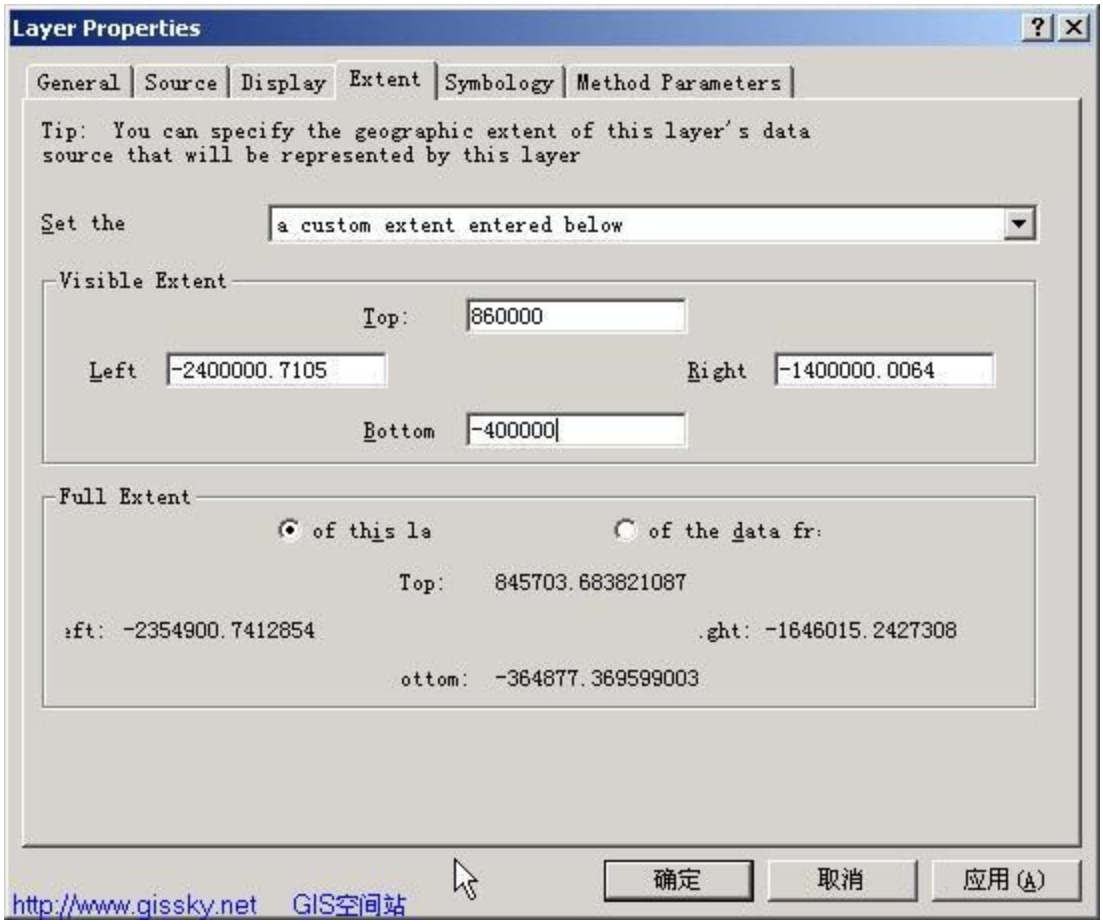
在 ArcMap 目录表中右键单击预测表面名，

在快捷菜单中选择 Properties，

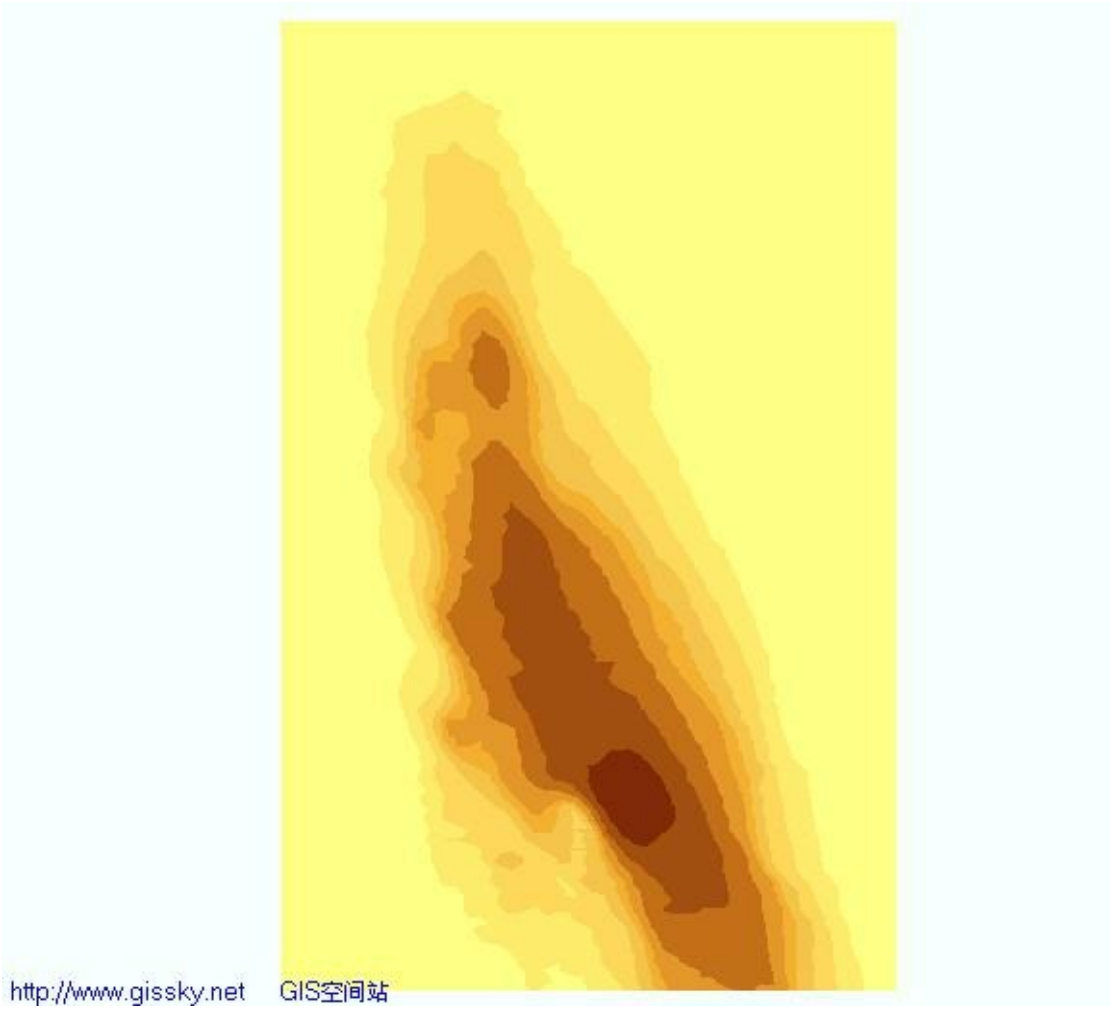
在 Layer Properties 面板中点击 Extent 页；



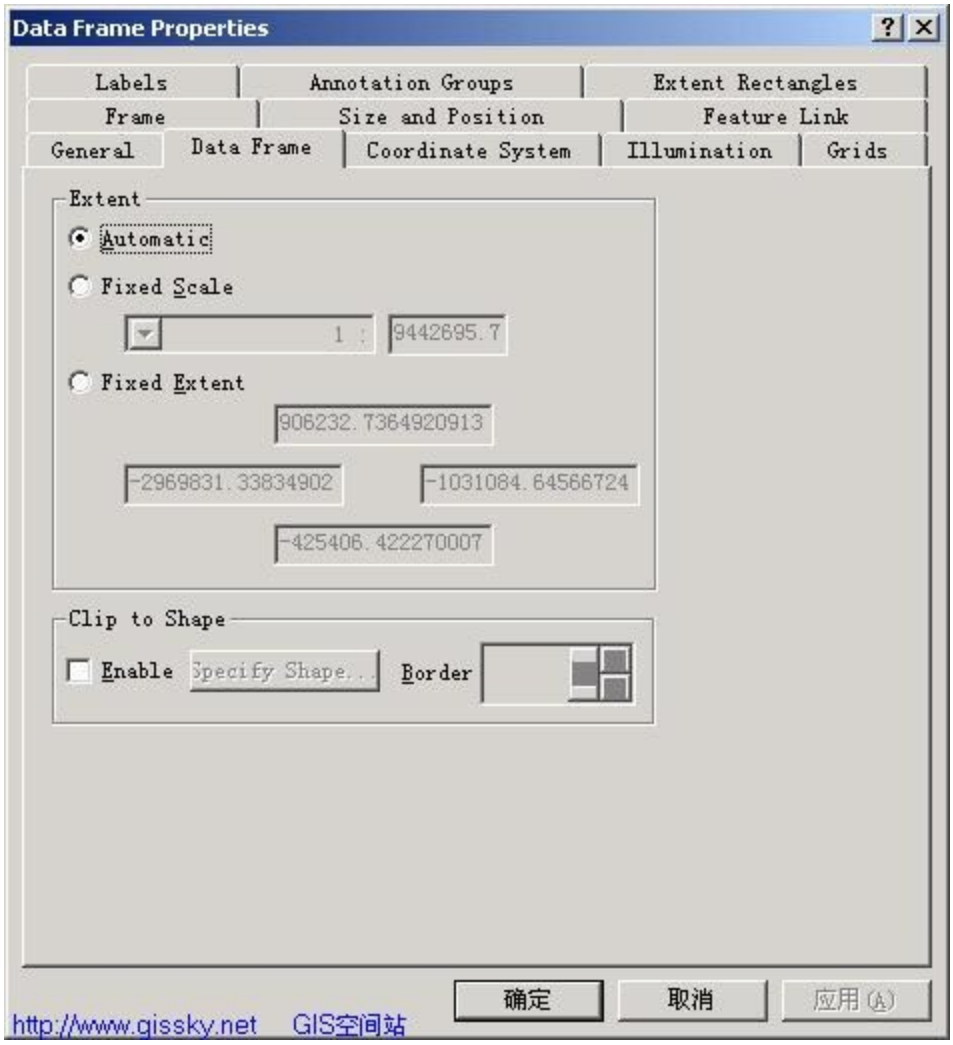
在 Set the extent to 下拉菜单中选择 a custom extent entered below，
然后在下面的 Visible Extent 项中设置坐标范围。（此图中分别设置为左：-240000，右：-1600000，上：860000，下-400000）。
设置后结果如图。


Layer Properties
General
Source
Display
Extent
Symbology
Method Parameters
Tip: You can specify the geographic extent of this layer's data source that will be represented by this layer
Set the
a custom extent entered below
Visible Extent
Top:
860000
Left
-2400000.7105
Right
-1400000.0064
Bottom
-400000
Full Extent
of this la
of the data fr:
Top:
845703.683821087
eft: -2354900.7412854
ght: -1646015.2427308
ottom:
-364877.369599003
http://www.gissky.net
GIS空间站
确定
取消
应用(A)

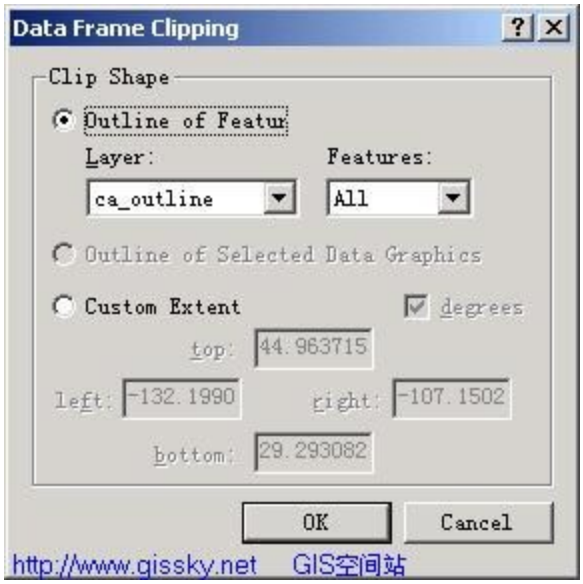
点击确定，得到结果：



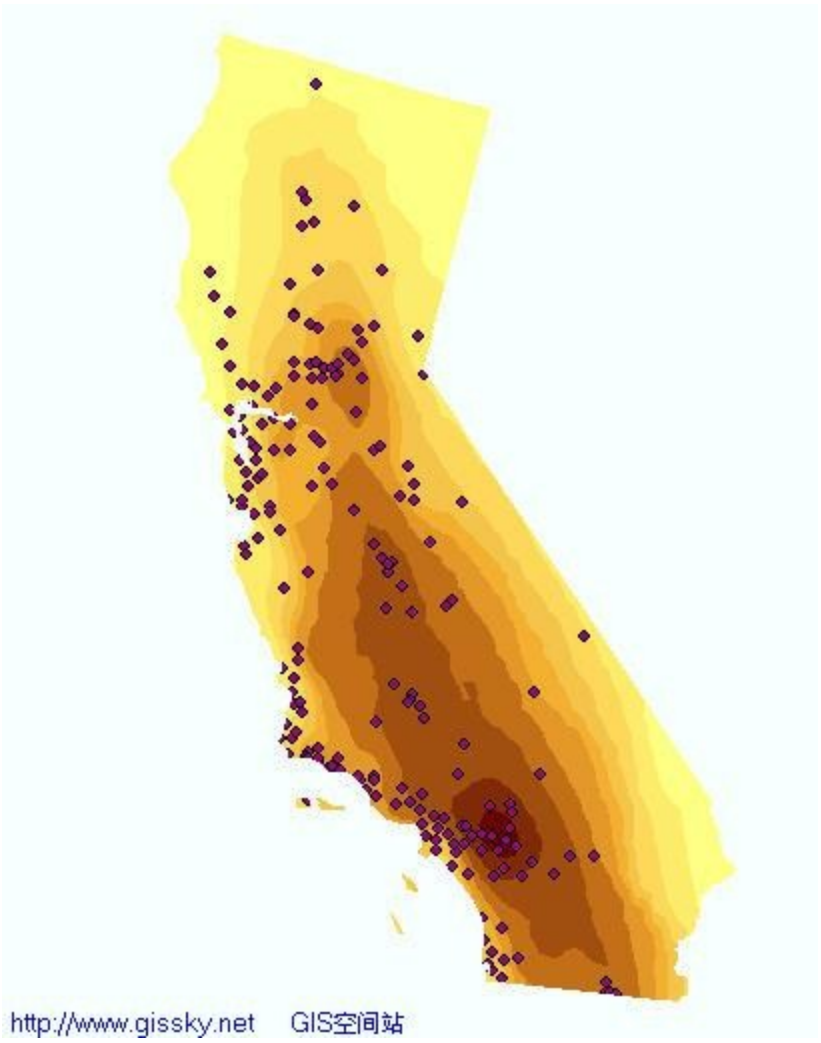
第二步：范围限制。
在 ArcMap 目录表中右键单击 Layers，
选择 Properties，
点击 Data Frame，



在 Clip to Shape 项中选中 Enable 前的复选框，然后点击 Specify shape…，在 Data Frame Clipping 面板中指定限制图形为 ca_outline，点击 OK，点击确定。



结果如下图。



后面可继续进行比例尺、图例等的设置，在此不一一赘述。（完）