

Variogram Tutorial

Randal Barnes

Golden Software, Inc.

变异函数教程

作者：Randal Barnes(兰德尔·巴恩斯)

Golden Software 软件公司

翻译：程贤辅 2012.12.24

目录

- 1 - 引言
- 2 - 什么是一个变异函数，它代表什么呢？
- 3 - 什么是变异函数？
- 4 - 变异函数网格
- 5 - 建模全方位的变异函数
- 6 - 变异函数模型各向异性
- 7 - 经验法则
- 8 - 常见问题
- 9 - 一些地质统计学的参考文献

1 引言

变异函数刻画了数据集在空间上的连续性或粗糙程度。普通一维的两个数据集的统计数据几乎相同，但在空间上的连续性可能是完全不同的。其理由请参阅第2节变异函数的部分内容。

变异函数的分析包括从数据计算出来的实验变异函数和变异函数模型拟合出来的数据。实验变异函数的计算方法是有关Z值的均方差的一半，该Z超过指定的分隔距离和方向观测到的所有数据对。它被绘制为一个二维图。用于计算实验变异函数的数学公式的详细信息，请参阅第3节。

该变异函数模型是选自一组描述空间关系的数学函数。相应的模型是选自与实验变差函数的曲线形状相匹配的数学函数的形状曲线。

请参阅网上Surfer用户指南和Surfer帮助中有关每个函数的曲线形状图中的变异函数模型图形主题。要交代的几何异向性（在不同方向上的可变空间连续性），单独的实验和模型变异函数可以计算数据集中的不同的方向。

	A	B
Count	15251	15251
Average	100.00	100.00
Standard Deviation	20.00	20.00
Median	100.35	100.92
10 Percentile	73.89	73.95
90 Percentile	125.61	124.72

表1.1

2 什么是一个变异函数所代表的？

考虑两个人工数据集，表1.1给出了A和B这两个数据集的一些常见的描述性统计，等一下我们将利用这些数据。

图1.1和1.2中给出这两个数据集的直方图。根据这方面的证据看来两组数据几乎是相同的。然而，这两个数据集在不包含常见的描述性统计信息和直方图方面有着显著的不同方式。

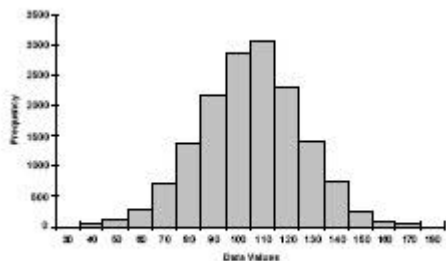


图1.1 数据集A的直方图

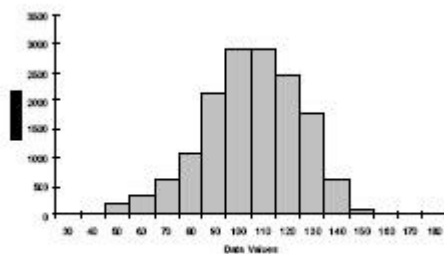


图1.2 数据集B的直方图

通过比较相关的等高线图（见图1.3和1.4）中可以看出，数据集A要比数据集B来得粗糙。请注意，我们不能说该数据集A是数据集B的“变异”，因为两个数据集的标准偏差是一样的，高低幅度都是一样的。这两个数据集之间的明显的视觉差异是在纹理质感上和没有可变性上。

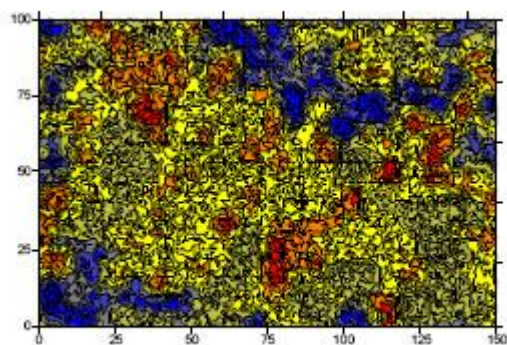


图1.3 数据集A的等值线图

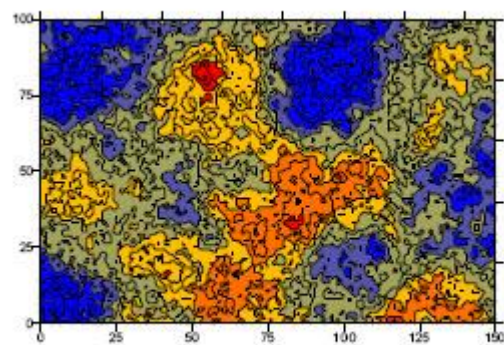


图1.4 数据集B的等值线图

特别地，数据集A比数据集B在空间上的变化更迅速。连续的高区（红色色块）和连续的低区（蓝色色块）平均而言，数据集A比数据集B要更小。这种差异在一般的样本设计、场地特征和空间预测上有显著的影响。

这并不奇怪，常见的描述性统计和直方图无法识别，更不用说在量化上和质感上区别这个例子中的两个数据集。常见的描述性统计和直方图无法用定义的计算方法去包括数据的空间位置。

变异函数是一个定量的描述性统计，可以以图形方式显

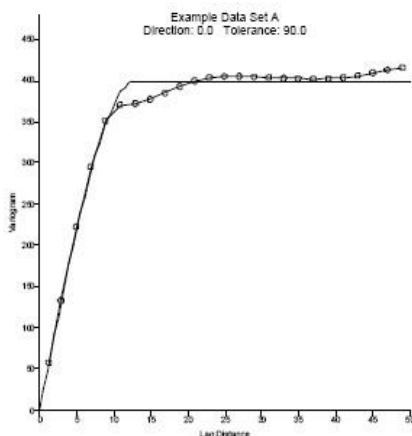


图1.5 数据集A的变异函数和模型

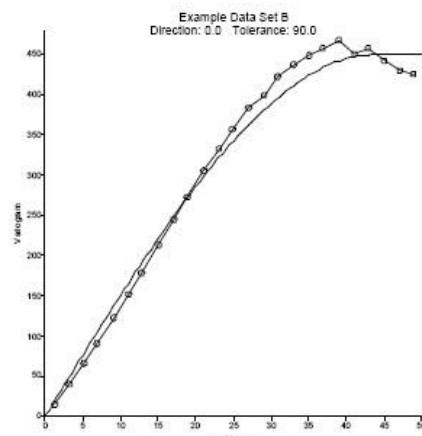


图1.6 数据集B的变异函数和模型

示在数据集空间特征上的连续性(即粗糙度)的方式。图1.5和1.6所示为这两个数据集的变异函数图。两组曲线在初始斜率上的差异是明显的。

3 变异函数是什么？

变差函数的数学定义是这样的：

$$\gamma(\Delta x, \Delta y) = \frac{1}{2} \varepsilon \left[\{Z(x + \Delta x, y + \Delta y) - Z(x, y)\}^2 \right] \quad (3.1)$$

其中， $Z(x, y)$ 是相关变量在点 (x, y) 的有效值，而 $\varepsilon[\]$ 为统计期望因子。请注意，变异函数， $\gamma(\)$ 函数是其中的点与点之间的距离 $(\Delta x, \Delta y)$ 的函数，而不是特定位置 (x, y) 的函数。这个数学定义是一个有用的抽象概括，但又不是那么容易应用到观测值的分析上。

我们现在考虑观测到的 n 组数据： $\{(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \dots, (X_n, Y_n, Z_n)\}$ ，其中点 (X_i, Y_i) 为观测点 i 的位置，而 Z_i 则是与该观测点相关的观测值。对于 $n(n-1)/2$ 的位置，我们有(前后)唯一的观测值对，在每一个这样的数据对中，我们可以计算出相关联的间隔矢量：

$$(\Delta x_{ij}, \Delta y_{ij}) = (x_i - x_j, y_i - y_j) \quad (3.2)$$

当我们要推导出某特定间隔矢量的变异函数 $(\Delta x, \Delta y)$ 时，我们将使用所有的间隔矢量的数据对，它是近似等于这种间隔的有效值：

$$(\Delta x_{ij}, \Delta y_{ij}) \approx (\Delta x, \Delta y) \quad (3.3)$$

让 $S(\Delta x, \Delta y)$ 代表所有数据对的集合：

$$S(\Delta x, \Delta y) = \{ (i, j) \mid (\Delta x_{ij}, \Delta y_{ij}) \approx (\Delta x, \Delta y) \} \quad (3.4)$$

此外，我们让 $N(\Delta x, \Delta y)$ 等于 $S(\Delta x, \Delta y)$ 的数据对的数目。从观测数据来推断变异函数，然后，我们将采用实验变异函数的公式。

$$\gamma(\Delta x, \Delta y) = \frac{1}{2N(\Delta x, \Delta y)} \sum_{(i, j) \in S(\Delta x, \Delta y)} (Z_i - Z_j)^2 \quad (3.5)$$

即，有效值的特定间隔矢量的实验变异函数，是该矢量的增量的观测间隔的所有数据对的 Z 值的差的平方的平均值的一半。(很拗口的句子，参看上面的公式才好理解——译者)。

4 变异函数网格

如果有 n 个观测数据，那么就有 $n(n-1)/2$ 个唯一的观测对。因此，即使是一个中等尺度的数据集也会产生大量的数据对。例如，如果 $n=500$ ，那么就有 $n(n-1)/2 = 124745$ 个数据对(应该是124750个数据对——译者)。即使对于一个快速的计算机，对这样大量的数据对进行操作也是非常耗时的。Surfer将预



图4.1

计算所有的数据对，并且将必要的和与差在变异函数网格中储存起来。(注：该变异函数网格和用于绘制图形的网格是不一样的。)

要创建一个新的变异图，选择“Grid|Variogram|New Variogram(网格|变异图|新建变异图)”菜单命令，在“打开”对话框中，指定数据文件的名称，然后单击“打开”按钮。指定X、Y、Z列，“重复数据”的设置，“数据排除滤波器”的设置(如果有的话)，并审查“数据统计”。

点击“General(常规)”选项卡，查看“Variogram Grid(变异图网格)”和“Detrending(反趋势)”选项。“Max Lag Distance(最大滞后距离)”是变异函数建模过程中要考虑的最大间隔距离。在默认情况下，这是大约三分之一的对角线范围内的观测数据。角度分割值180和径向分割值100对于几乎任何设置都是足够的。

“Detrending(反趋势)”提供了普遍应用的克里金算法的高级选项。在通常情况下，“Do not detrend the data (不做数据反趋势)”是个适当的选择。但是，如果你知道数据中有一个强大的趋势存在，你可能要考虑使用“Linear(线性)”反趋势。选择“Generate Report(生成报告)”，就会创建一个“Data Filter Settings(数据过滤设置)”和“Data Statistics(数据统计)”的列表。

如果不改变任何设置，点击“OK(确认)”。显示图形如图4.3。

黑色的线和圆点就是全方位的我们的实验变异函数，而蓝色线是在拟合变异函数模型中第一个通过的(尽管是非常勉强的)。

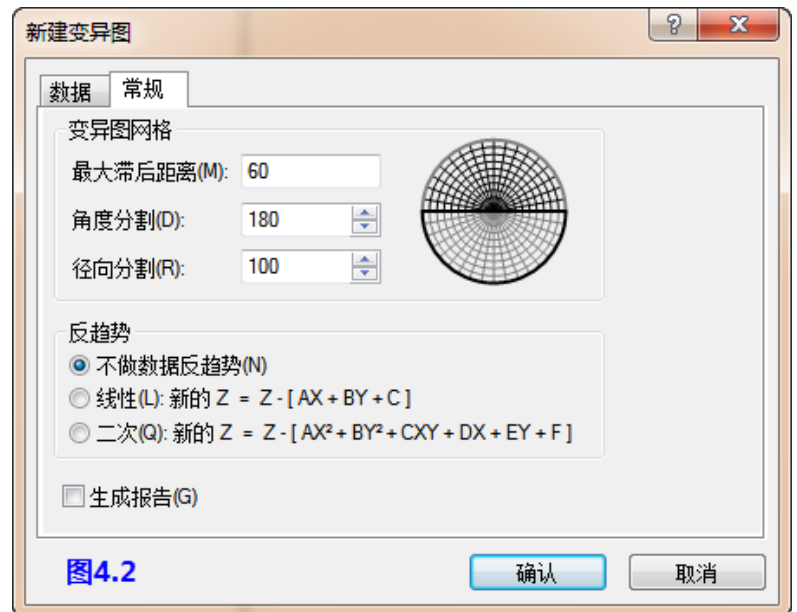


图4.2

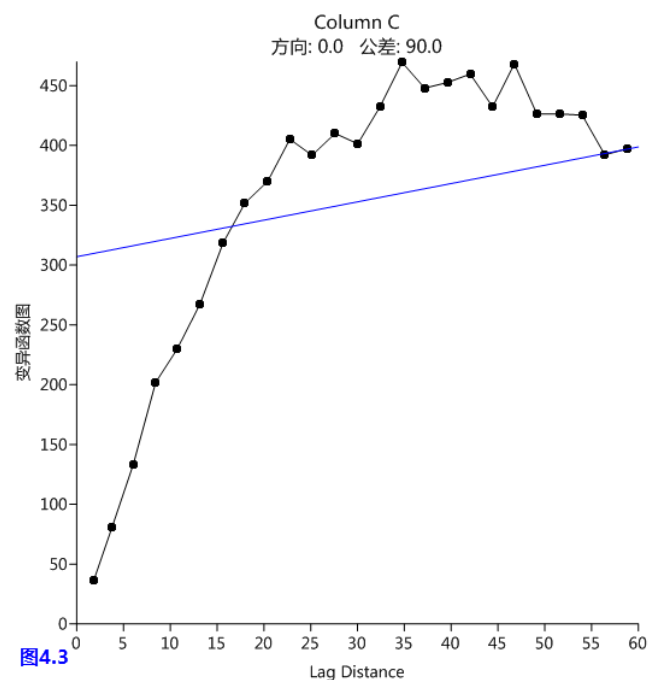


图4.3

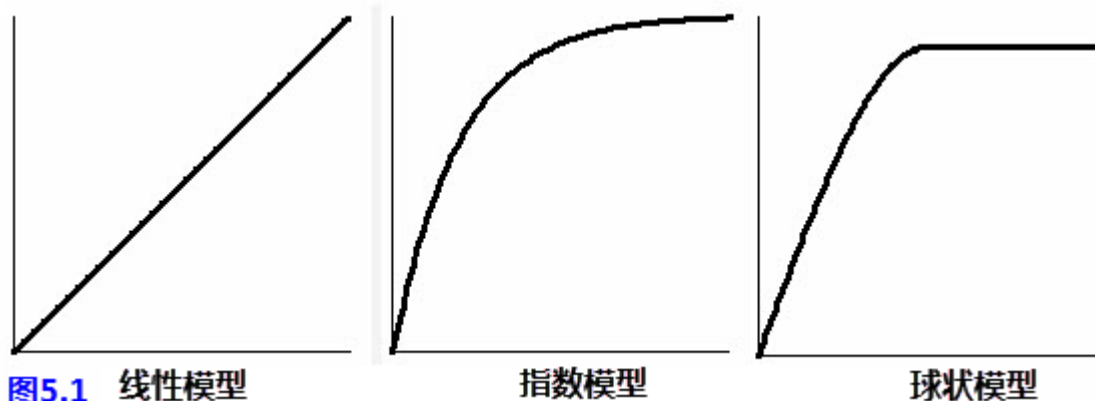
5 全方位变异函数的建模

默认情况下，第一个图是全方位的变异函数(定向公差是90度)。根据全方位的变异函数的阈值、以及金砖效应来选择模型类型。

5.1 选择变异函数的模型类型

有无限多的可能的变异函数模型。Surfer允许成千上万不同的基于十种(Surfer11版是12种)选定的可用组件类型组合起的变异函数模型的创建。

含有金砖效应组合的三种模型之一，它们足以应付大多数的数据集：就是线性模型、指数模型、还有球状模型。这三种模型的例子，如图5.1所示。



如果实验变异函数从来没有突破回落值，则线性模型通常是合适的。如果实验变异函数超出回落值，而且是“弯曲地”一路上涨，则指数模型应予以考虑。如果实验变异函数开始时就急剧地超过回落值，笔直地向上发展，然后再弯曲，则球状模型是很好的首选。

对于文件ExampleDataSetC.xls中的数据，会出现一个球状模型(也可以尝试指数模型)。双击变异函数曲线，然后选择“Model(模型)”选项卡。(在Surfer11中，单击图形，然后在属性管理器中选择“Model(模型)”选项卡。并没有对话框弹出——译者)

按两次“删除”按钮，删除不适当的默认模型。然后按下“添加”按钮，选择球状模型，然后按“OK(确认)”。

5.2 选择变异函数模型的高值(Scale)和长度(Length)参数

现在，我们必须采用迭代的方法(就是猜测和检验)来设置高值和长度(A)。这里“高值”是变异函数回落值在Y轴的高度。通过简单的图形看来，400到450之间的一个数值似乎是合理的，所以输入425。球状模型的长度(A)，是变异函数的回落值的滞后距离。同样，从图中看30和40之间的一个值似乎是合理的，所以输入35。按应用按钮，新的供选择的变异函数模型就绘制出来。

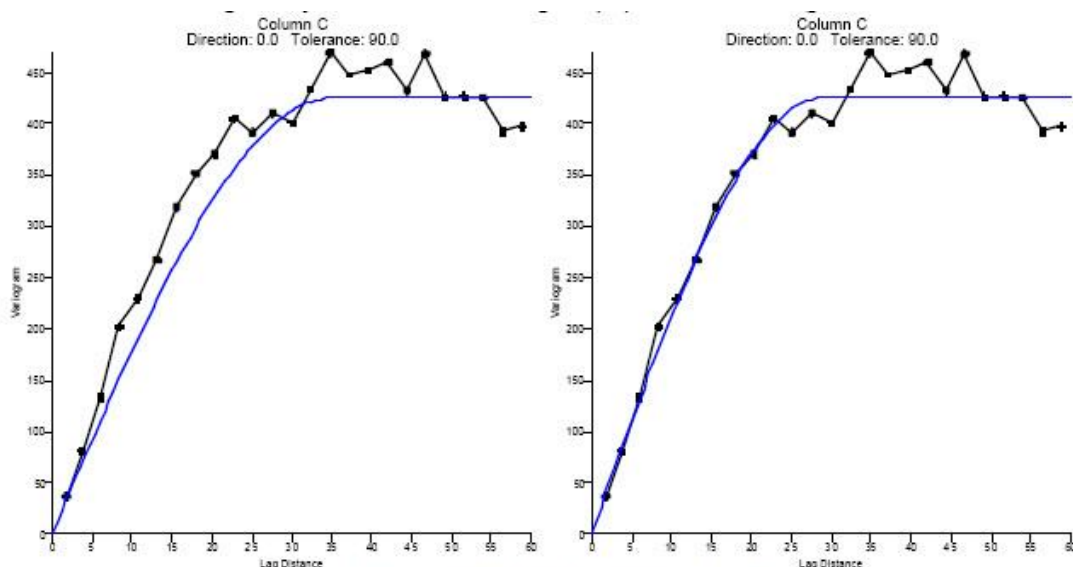


图5.3 初始假设的变异函数模型(这是教程原图)

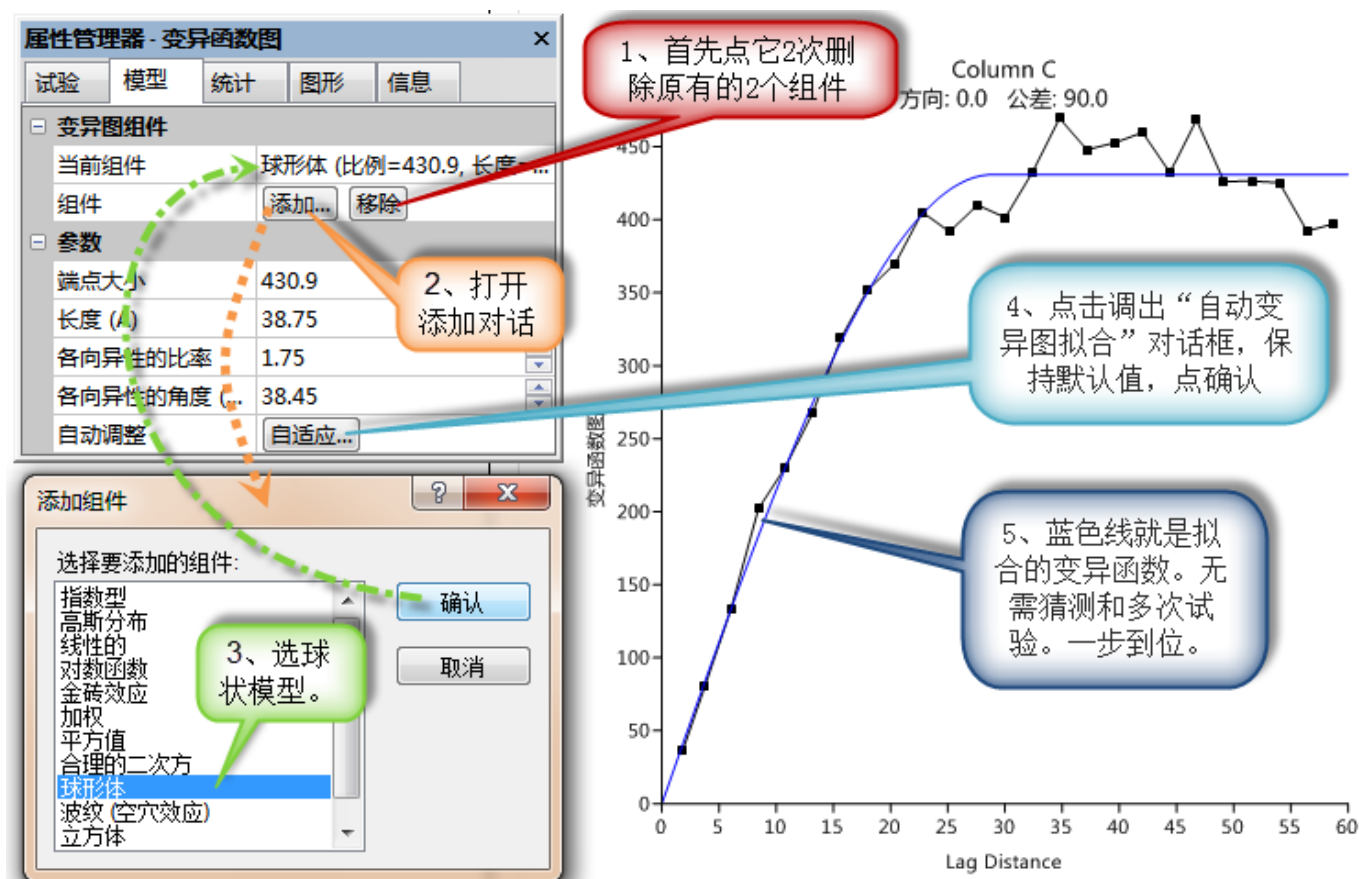
左边：高值=425，长度(A)=35。

右边：高值=425，长度(A)=29。

这是一个不坏的猜测，但是经审查绘制的曲线长度(A)，似乎有点太长，因为现在模型(蓝线)位于

实验变异函数(黑线和圆点)的右侧。重新设置长度为30, 然后按“应用”按钮, 结果似乎还有点长。尝试长度(A)为29。结果看来是变异函数一个不错的选择。参见图5.3。

译注: 这篇教程发表时可能尚未发行Surfer11版, 变异函数的拟合过程采用猜测和多次迭代试验的做法, 显然是非常麻烦和不准确的。时代进步了, 软件也在进步。Surfer11已经有自动拟合的功能了。看下图, 按步骤, 很快就绘制出变异函数图。



5.3 选择变异函数的金砖效应

如果实验变异函数出现在垂直轴上有一个非零的截距, 则该模型可能需要有一个具有金砖效应的组件。增量Z的方差在变异函数网格报告里的“最近的相邻的数据的统计信息”部分, 金砖效应在大多数情况下都提供了一个定量的上限。

在Surfer中, 金砖效应被划分为两个子部件: 误差方差和微方差。这两个子部件都是非负的, 而且这两个子部件的总和应该等于显著的非零截距。

误差方差测量一些可重复性的观测值。这些同时包含了取样和化验(分析)的误差。误差方差是在计算重复样品之间差异的方差时最好的选择。

微方差是在间隔距离小于典型样品间距的未知变异函数的替代品。这是根据在实验变异函数的明

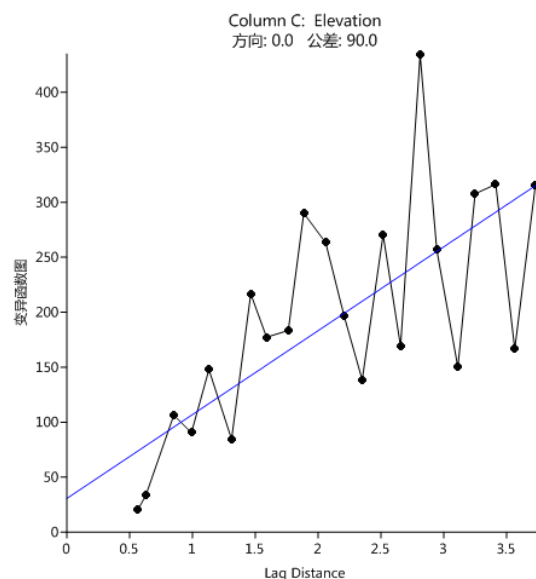


图5.4 Demogrid.dat数据集的带金砖效应的线性变异函数模型

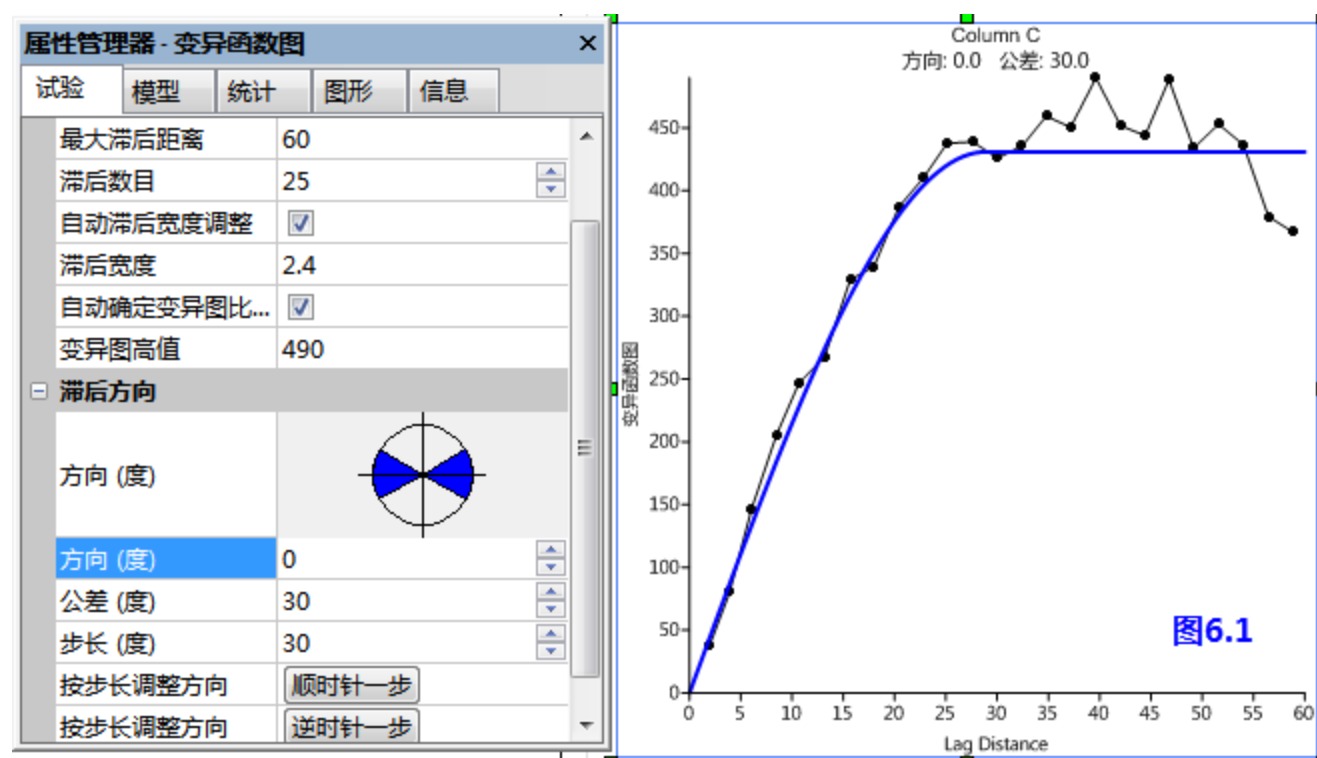
显的非零截距和误差方差之间采用的差异的最好选择。

该模型在我们的例子中函数与垂直轴相交于0(截距等于0)，因此我们将不应用金砖效应。

6 变异函数模型的各向异性

通常,实验变异函数显示出在不同方向上的不同尺度(长度和高值)。这就是所谓的几何各向异性。对于线性变异函数模型,这将显示为不同方向上的不同斜率,而球状模型则体现了在不同方向上明显的不同长度(A)的几何各向异性。

了解变异函数的各向异性,选择“Variogram Properties | Experimental(变异图属性|试验)” (在Surfer11中,在属性管理器中选择“试验”)选项卡。全方位的实验变异函数的平均值在所有方向上都有取值。可以查看在某个特定方向上改变方向和公差参数。现在我们开始设置公差为30度。注意,属性表中的Lag(滞后)方向部分现在显示了一个集中在0度(模型中的东部)的角度为(正负)30度公差的蓝色扇区。



拖动对话框到一边,让变异函数可见。按动“Step CCW(逆时针一步)”按钮,可以迅速转遍各个方向。按一次,就按照当前设定的方向旋转一个步长30度,再按一次就如你选择了60度。再次按下该键,再而三,注意圆形的图形可以帮助你跟踪当前的方向。

扫过各个方向几次后,请留意垂直轴上高值的变化。要使图形垂直刻度适应此变化,设置垂直刻度为515,并且按“Apply(应用)”按钮。(在Surfer11中,将Y轴属性的“缩放比例”中的“轴刻度的极大值”改为515,“最后的主刻度”改为500——译者)

更改步长为15度,同样扫描所有方向一次。你可以按住“Step CCW(逆时针一步)”按钮以产生一个连续的扫描(在Surfer11中做不到)。似乎在90度到180度之间是一致的,全向的模型位于实验变

异函数上升部分的右侧。相反，在0度到90度时，全向的模型位于实验变异函数上升部分的左侧。这个明显的迹象就是几何的各向异性。

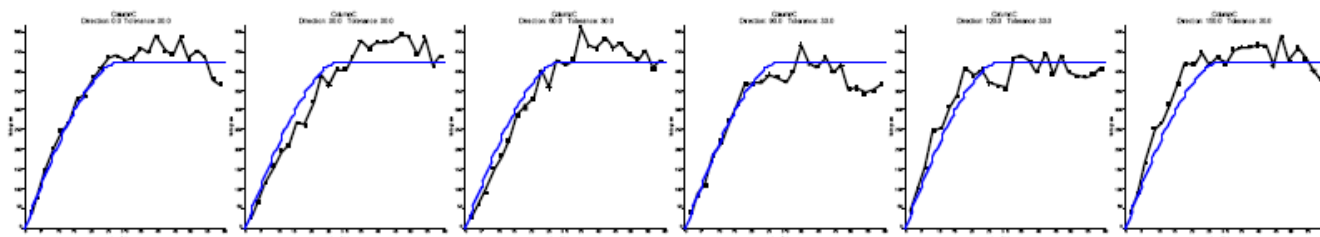


图6.2 定向变异函数曲线容差设置为30°
从左到右：方向为0°，30°，60°，90°，120°，150°。

按“Step CCW(逆时针一步)”按钮，使实验变异函数的长度和高值达到明显的最大化那个角度：30度。转到属性中“Model(模型)”选项卡，并更改“长度(A)”的参数，以更好地适应这个方向的变异函数。40太长，30又太短，看来37吻合得十分好(Surfer11中软件设定该长度为38.74)。更改各向异性角度为30度(Surfer11中软件设定该角度为38.47)，然后按“应用”按钮(图形在这一点上应该没有改变)。(在Surfer11中由于是自动拟合，所以上述的实际操作数据稍有出入。——译者)

点击“Experimental(试验)”选项卡，并改变滞后方向为120度(这是垂直方向上的最大长度， $120=30+90$)。再点击“Model(模型)”选项卡。仅仅改变各向异性比率使变异函数模型去适应当前方向的实验变异函数(当你改变设置，按“应用”按钮即可看到更新后的图形)(Surfer11中没有“应用”按钮，也不需要此按钮，修改参数后图形立即更新)。看来比率1.6似乎是个不错的选择(Surfer11中软件设定该比率为1.75)。



图6.3

为了得到一个最终的视觉以验证你的模型，把它“旋转”。点击“Experimental(试验)”选项卡，单击并按住“Step CCW(逆时针一步)”按钮(Surfer11中按住此按钮不放没有作用)。请注意变异函

数模型模拟实验变异函数的形状变化。你可以微调模型参数以更好地匹配实验变异函数，但也不要过分：一到两个重要的参数就是你所期望的。

单击“AutoFit...(自适应...)”按钮来微调模型，在弹出的对话框中选择适当的参数。单击“确认”，接受默认设置，“自动适应”就返回各向异性比率为1.75，角度为38.49(看来自动拟合功能是有的，前面只是讲课需要才一步一步手工调整设置——译者)。

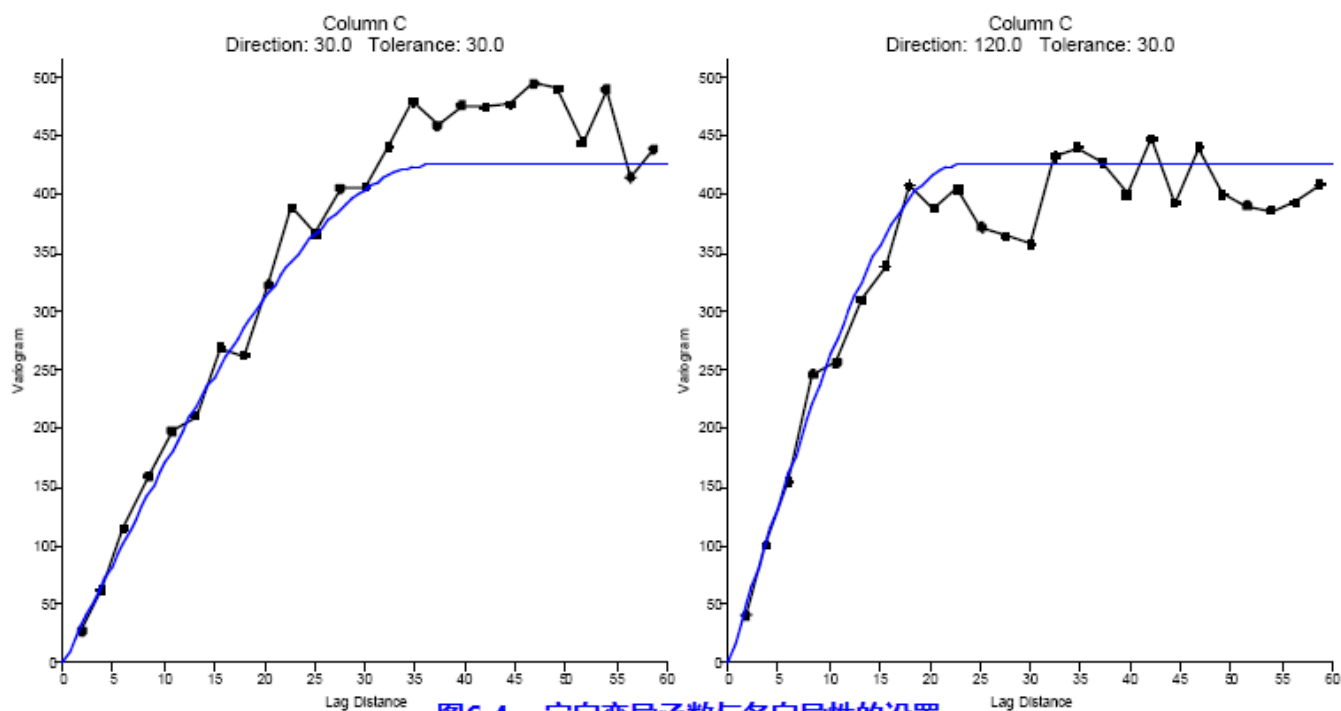


图6.4 定向变异函数与各向异性的设置。

左边：方向=30°，右边：方向=120°

各向异性比率=1.6，角度=30°

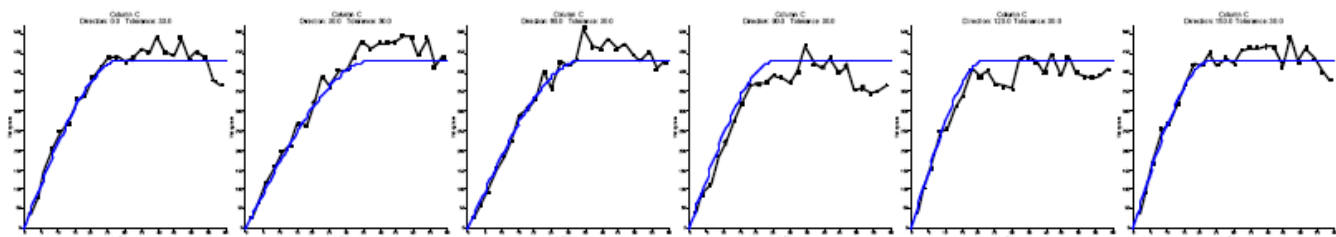


图6.5 各向异性比率为1.75和角度为38.49°时的定向变异函数曲线图

从左到右：方向=0°，30°，60°，90°，120°，150°。

7 经验法则

- 了解你的数据！在实验变异函数计算之前，首先进行常规的非空间统计。利用“Variogram Report(变异函数报告)”显示数据的最小值、最大值、中位数、平均数、标准差、方差和斜率。在Surfer中创建一个散点图或者归类散点图来显示散点分布。使用绘图器来创建直方图以及累积频率曲线图。

- 模型拟合不要过分。最简单的模型，能够再现重要的试验变异函数的功能就是最好的模型。
- 当有疑问时，请使用默认的变异函数模型进行网格化。一个简单的线性变异函数模型往往会产生一个可以接受的网格；在初始数据的分析中这是特别真实的。然而请记住，克里金标准偏差网格的生成使用默认的变异函数是毫无意义的。
- 除非有一个明确的、毫不含糊的、物理的理由，不要设定各向异性比率大于3比1。如果实验变异函数出现支持大于3比1的各向异性，并且有这样严重的各向异型却没有明确的物理理由，也有可能是数据中有这样的趋势。可以考虑将数据首先去趋势，然后再进行您的变异函数分析。
- 尝试一个模型，并在等值线图中观察它如何产生网格。如果你有多个变异函数模型，则从每一个来生成网格和等值线图。如果这些图形没有显著的差异，应该选择最简单的变异函数模型。
- 变异图的范围通常是Z值的空间波动中的物理异常的平均大小的靠近值。在一个可靠的实验变异函数的情况下，这条经验规则可以用来推测一个变异函数的范围。
- 在一个恒定值附近波动的实验变异函数，并非是一种病态的变异函数。这显示一个迹象，表示典型抽样间隔最大值和Z值不相关。在这样的情况下的等值线图，不管使用何种网格插值算法，所代表的数据都是不可靠的，并且更近距离的抽样间隔的更多的数据对于更为详细的局部特征是必须的。
- 如果以下三个条件得到满足，那么抽样方差是作为一个变异函数的合理的逼近值。
 - 1、数据被均匀地分布在我们所期望的区域中，显示在一个场景图中。
 - 2、在我们所期望的区域中的数据中没有显著的趋势。
 - 3、我们所期望的区域的尺寸是有效变异函数范围的3倍以上。

8 常见问题

问：什么是变异函数？

答：变异函数是一种基于统计的、定量的、说明表面的粗糙程度的函数表达。变异函数是一个函数的分离矢量：包括间隔距离和方向，或 Δx 和 Δy 。变异函数在指定向量的间隔的点与点之间得出的平均差异(该差异是用Z值的平方差来测量)。

问：一个变异函数和半方差函数之间的区别是什么？

答：区别仅仅是2的因数。最初定义的变异函数如下

$$2\gamma(\Delta x, \Delta y) = \varepsilon \left[\{Z(x + \Delta_x, y + \Delta_y) - Z(x, y)\}^2 \right] \quad (8.1)$$

这里被命名为变异函数。而 $\gamma()$ 是实际感兴趣的半方差函数的名称，因为它是变异函数的一半。因为只有函数 $\gamma()$ 被用在克里金算法中，这个前缀“半”是经常省略的，因此该 $\gamma()$ 函数在统计学文献中交替地称为变异函数和半方差函数。

问：变异函数曲线图中使用的单位是什么？

答：设(X,Y)表示观察的位置，并让Z代表相关联的观测值。横轴(滞后距离轴)和X、Y单位相同。纵轴(变异函数轴)的单位等于Z值的平方单位(类似Z的方差)。例如，让我们：

- X 代表向东方向的基准测量单位[米]，
- Y 代表向北方向的基准，测量单位也是[米]，
- Z 代表降雨强度，单位[厘米/小时]。

之后，变异函数的横轴的单位是[米]，以及纵轴的单位为[(厘米/小时)²]

问：为什么我在变异函数参数中不能指定阈值？如何决定相关的阈值大小呢？

答：阈值是指变异函数回落时的纵坐标的值。一个变异函数模型可以包括多个组件(例如，一个球形的组件加一个金砖效应组件)。虽然对于单个的组件，可以说有一个阈值，这种用法往往导致总体模型的各个组成部分的阈值和阈值之间产生混淆。混乱的另一个原因是，一些组件没有阈值，以及包含这样组件的模型(例如线性模型)也没有阈值。

为了减少混淆，阈值(sill)这个词是用来描述整个模型的回落值，而不是各个组件。当讨论一个单独的组件的回落值时，我们使用的术语是高值(scale)。

问：变异函数的方向是做什么的？为什么当我改变方向后变异函数跟着改变？

答：这种特性称为各向异性。就是变异函数测量中作为间隔距离和方向的函数的差异。在许多物理设置中，它是在某个方向比其他方向能更迅速改变的可能感兴趣的变量。例如，配送的晶粒尺寸的变化在垂直方向上就比水平方向上变化得更加迅速。同样，某些种类的植物在干旱气候条件下患病率的变化，在垂直河道方向移动时发生的变化要比沿着河道移动而发生的变化来的大；即在不同的方向有不同的特性。方向参数可以让你去调查在不同方向时的变异函数。因此，可以对各向异性识别和进行定量分析。

问：Surfer在网格化时使用各种变异函数的方向吗？

答：是的。明确地说，该变异函数的各向异性就用于当指定的变异函数采用克里金算法网格化时。

问：Surfer有一个多方向的网格化算法，仅仅是为了各向异性（在不同的方向上的不同阈值）可以被改变吗？

答：变异函数模型包含了必要的信息来描述在各个方向上的变异函数的特性。比如球状模型，是一个漏斗状的表面。变异函数图上绘制的一根线条是代表通过变异函数模型的垂直切片。

各向异性比率描述的是长度(A)在不同方向(几何各向异性)的比率。有时在不同的方向(地带性的各向异性)具有实验变异函数的明显的不同阈值。例如，一个实验变异函数很好地描述了一个球状模型的长度(A)是100米。它在东西方向(角度=0度)上的回落值(阈值)为2.0，而它在南北方向(角度=90

度)上的回落值为3.0。Surfer可以使用嵌套的模型再现这种特性。我们使具有各向同性的球状模型为第一个组件，它的长度为100，高值为2，而使第二个组件为各向异性的模型，取值为：长度100，比率0.001，方向为90度，以及高值为1。在一个复合的模型中，添加两个组件，它们之一，东西方向具有阈值为2，另一个，南北方向具有阈值为3。

问：为什么在导出变异函数数据时，有些单元格为空白？

答：当被指定的滞后距离间隔没有分离的数据对，以及该单元格表示的是角度公差时，单元格是空白的。

问：我要如何进行克里金算法的交叉验证，jack-knifing(留出一个模型)，还是其他？

答：在Surfer中“网格|数据(Grid | Data)”命令提供了一套广泛的交叉验证工具。交叉验证报表(使用克里金网格化算法)，提供了许多量化的指标，可以用来作为一个不错的变异函数的拟合统计。我们建议作为经常有用的三条统计数据：交叉验证残差的平均绝对偏差、交叉验证的标准残差，以及测量和估算之间的排名相关性。

问：如何绘制三维变异函数曲面图？

答：在Surfer中此选项不可用。

问：Surfer计算了 S^2 (数据的方差)，我们怎样在程序的输出中找到这个值？

答：Surfer使用的公式在实验变异函数计算范例方差的帮助和手册中。该数值在变异函数属性对话框的“统计”页里提供，还在“变异函数报告”里提供。

问：当使用不同的斜率值的线性变异函数时，为什么网格没有变化？

答：出现这种情况是因为没有采用金砖效应。改变线性变异函数的斜率相当于仅仅改变观察位置的高值，因此，变异函数中任意两个对点之间的值的比率通过改变斜率是不会改变的。然而，重要的是要注意，即使是敏感的斜率内插值的权重和插补值也不是克里金方差。

问：我怎样才能获得有关数据点之间间距的信息？

答：将你的数据绘制成散点图(参见“Map|Post Map|New Post Map”菜单)(Surfer 11中，菜单是“Map|New|Post Map[图形|新建|散点图]”)。Surfer的变异函数报告中，以及在变异函数属性对话框中的“统计”这一页的底部，都提供了一些有用的定量统计信息。例如，在文件ExampleDataSetC.xls中的相邻值(Nearest Neighbor values)是：

- Avg. distance to nearest neighbor(平均相邻距离): 2.90403
- Min. distance to nearest neighbor(最小相邻距离): 1
- Max. distance to nearest neighbor(最大相邻距离): 9.48683

- Gamma for nearest neighbors(相邻距离的伽玛值): 64.735

此外，在“网格|数据”菜单上的数据指标网格化的算法中提供了广泛的具有数据几何特性的工具套件。这包括了相邻数的数目、数据密度、距离最近的相邻的数，等等。

问：当使用克里金算法的指数变异函数模型时，参数长度(A) 是输入相关长度(即阈值达到63%的变异函数的长度)，还是实际长度(即阈值达到95%的变异函数的长度)？

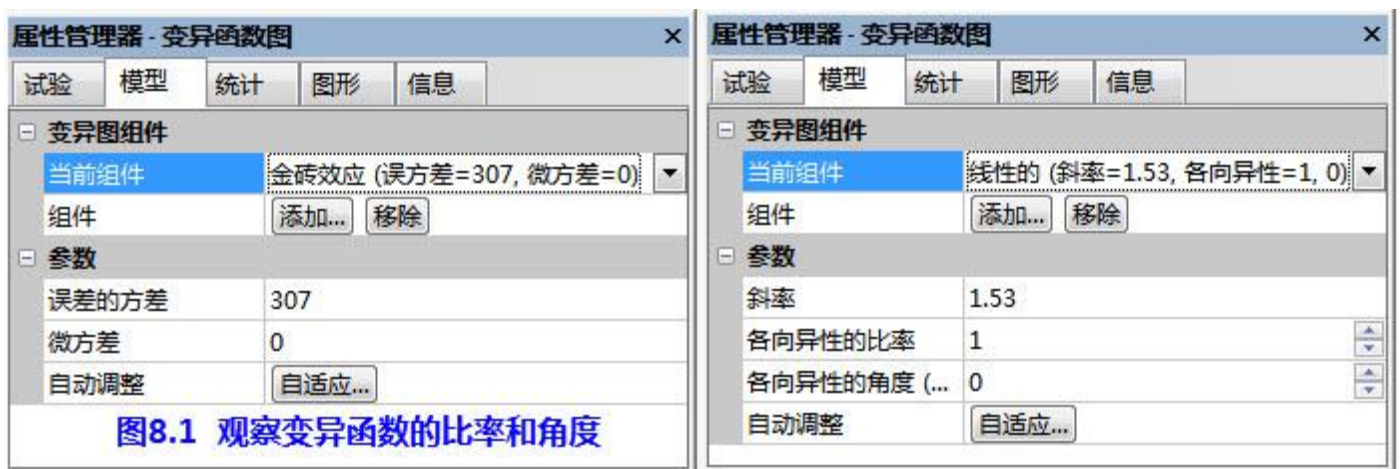
答：不同的地质统计软件包对于指数模型会使用不同定义的长度参数。Surfer定义的长度(A)参数是相关长度。图示来说，相关长度是指，当滞后距离为零时的曲线的切线与指数模型的阈值相交时对应的滞后距离。指数模型达到约63%的阈值在滞后距离(A)。而滞后距离为3A时指数模型的阈值大约达到95%。

问：我在应用高斯模型时，似乎得到一个在我所见到的变异函数的图形和由软件“自适应...”结束后报告的长度之间不相匹配的现象。阈值的长度要比属性中“模型”选项卡最下面的“自适应...”结束后报告的长度值要大很多。

答：对于指数模型，不同的地质统计软件包使用高斯模型不同定义的长度(A)参数。Surfer定义的长度(A)参数是相关长度。图形上看，相关长度就是对应曲线的拐点的滞后距离。高斯模型在滞后距离为(A)时达到大约阈值的63%。当滞后距离为3A时，高斯模型达到大约阈值的99.99%。

问：在新的Surfer软件中，在变异函数的过程中是否有一个函数可以向我提供确切的金砖、范围以及阈值？

答：要查看金砖、范围以及阈值，请选择“Grid | Variogram | New Variogram(网格|变异图|新建变异图)”菜单命令。指定一个数据文件，然后按“Open(打开)”按钮，采用默认值，并点击“确认”。在变异函数图上双击，单击“Variogram Properties | Model(变异函数|模型)”选项卡，查看“金砖效应”误差方差和微方差，以及默认的线性模型的斜率和各向异性参数(比率和角度)。



问：我使用“自适应...”和所有的标准设置，试图来适应球形变异函数。可是适应得很差。你能解释一下怎样调整Surfer来适应吗？我要怎样做才能够适应球面函数的使用？

答：“自适应...”调整功能需要一些合理的初始参数才能达到一个不错的结果。请检查实验变异函数图并估算长度(A)和高值。如果直接从你的数据集中是很难估计出高值的，可以参考“变异函数报告”中的 Z 方差列表(Z variance listed)以及变异函数属性的“统计”对话框中的内容。

问：我设置的滞后宽度是一个特定的值。为什么变异函数图上该点绘制的是数倍的该值呢？

答：滞后宽度定义了正或者负的滞后宽度的一半为中点。该点是绘制在该范围内所有点的距离平均距离值。

问：我怎样才能得到平均间隔距离的一个列表？

答：导出变异函数到一个文件中。此文件包含了平均滞后距离、二分之一的平均方差、还有所在范围内的数据对的数目。

问：是否可以从具有两个并列的不同变异函数模型应用领域上创建一个网格文件？该领域具有空间对比的形态。

答：如果该区域被定义为矩形，你可能可以为每个具有单独模型的区域建立网格，将这些网格转换成XYZ数据文件，在工作表中合并这些数据文件，并使用“Nearest Neighbor(最近邻点插入法)”防止进一步的插值。

问：偏移在克里金算法里面是如何使用的？是否应该让模型在转换后的数据(数据减去偏移)上运行完成，或者在非转换的数据上运行完成？事先去除趋势的克里金法与一般的克里金法是不同的。

答：我们建议用趋势数据计算变异函数模型时首先要套用偏移期间的克里金法，而后参阅Kitanidis所著，1997年，143页，第6.12节。

问：在Surfer中我还没有发现任何步骤可以帮助告诉我，理论上的变异函数模型被安装到了我的变异函数的实验数据中。我知道我可以“自适应...”按钮，但是我喜欢一个定量的数字这样子来达到完善的拟合。

答：“如何”确定理论模型的拟合数据，最好的措施是尝试，以及观察网格化后的数据与选定的变异函数模型。可以创建一个等值线图与该数据文件的散点图叠加，通过检查等直线图来判断模型的质量。交叉验证报告(使用克里金网格算法)提供了许多量化的指标，可以用来作为一个完善的变异函数的拟合统计信息。我们建议三条作为最经常用到的特别的统计数据：交叉验证残差的绝对平均偏差，交叉验证残差的标准偏差，和测量值和估计值之间的相关性的排名。

问：Surfer可以从Variowin、GeoEAS、GSLIB或者其他软件导入变异函数模型吗？

答：Surfer不具备此功能。可以提交请求到surferupport@goldensoftware.com，告知你想支持的软件名称。

问：我在其他地方读到的高值不应超过方差。这是真的吗？对于ExampleDataSetC.xls数据集，Z的方差为405.371，但是你推荐的高值却为425。

答：在一个合理的和普遍适用的假定之下，变异函数的阈值应该和基础总体的方差相等，但是，样本方差往往和总体方差的测量互相不适应(对于假设独立的数据进行方差估算的标准公式，在大多数地质环境中是无效的)。因此，象“高值不应该超过方差”这样的规则不是一个有用的概括。

关键的结果是

$$\varepsilon[S^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(x_i - x_j, y_i - y_j)$$

此结果指出的预期值的样本方差， S^2 ，等于所有 n^2 个数据对的样本值之间的变异函数的平均值。如果 n 个样本值均匀地分布在是变异函数范围的许多倍的区域范围内，然后将突出长距离的数据对，忽视短距离的数据对。因此，所有数据对之间的变异函数的平均值，将有许多值等于阈值而仅有少量值偏离阈值。在这样的情况下，样本方差是一个变异函数阈值的合理的估算值。

9 一些地质统计学的参考文献

Armstrong, M. [ed.], 1989, *Geostatistics*, volumes 1 and 2, Kluwer Academic Publishers, 1027 pp.

Armstrong, M., 1997, *Basic Linear Geostatistics*, Springer-Verlag, New York, ISBN: 3540618457, 220 pp.

Baafi, E. Y., 1997, *Geostatistics Wollongong '96 – Quantitative Geology and Geostatistics*, Kluwer Academic Publishers, ISBN: 0792302044, 1333 pp.

Bras, R., and Rodriguez-Iturbe, R., 1985, *Random Functions and Hydrology*, Addison-Wesley, 559 pp.

Chiles, J., and Delfiner, P., 1999, *Geostatistics: Modeling Spatial Uncertainty*, John Wiley and Sons, New York, ISBN: 0471083151, 695 pp.

Christakos, G., 2001, *Modern Spatiotemporal Geostatistics*, Oxford University Press; ISBN: 0195138953, 312 pp.

Clark, I., 1979, *Practical Geostatistics*, Applied Science Publishers, 129 pp.

Clark, I., and W. V. Harper, 2000, *Practical Geostatistics 2000*, Ecosse North America, LLC; ISBN: 0970331703, 442 pp.

Cressie, N., 1991, *Statistics for Spatial Data*, John Wiley and Sons, New York, 900 pp.

David, M., 1977, *Geostatistical Ore Reserve Estimation*, 2nd edition, Elsevier, 364 pp.

David, M., 1988, *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*, Elsevier, 216 pp.

Davis, J., 1973, *Statistics and Data Analysis in Geology*, Wiley, 550 pp.

deMarsily, G., 1986, *Quantitative Hydrogeology*, Academic Press, 440 pp.

Deutsch, Clayton, V., Journel, Andre G., 1992, *GSLIB - Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 338 pp.

Draper, N., and Smith, H., 1981, *Applied Regression Analysis*, second edition, Wiley-Interscience, 709 pp.

Dubrule, O. 1998, *Geostatistics in Petroleum Geology*, American Association of Petroleum Geologists; ISBN: 0891811877.

Gelhar, L. W., 1993, *Stochastic Subsurface Hydrology*, Prentice Hall, Englewood Cliffs, New Jersey, 390 pp.

Geostatistics, 1980, McGraw-Hill, 168 pp.

Gilbert, R.O., 1987, *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold Company, 320 pp.

Gomez-Hernandez, J., A. Soares, and R. Froidevaux [eds.], *Geoenv II -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, 562 pp.

Goovaerts, P. 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 483 pp.

Guarascio, M., David, M., and Huijbregts, C. [eds.], 1976, *Advanced Geostatistics in the Mineral Industry*, Reidel, 491 pp.

Gy, P., 1979, *Sampling of Particulate Materials*, Elsevier, 431 pp.

Houlding, S. W., 2000, *Practical Geostatistics - Modeling and Spatial Analysis*, Springer-Verlag, New York, 160 pp.

Hohn, M. E., 1998, *Geostatistics and Petroleum Geology*, 2nd edition, Kluwer Academic Publishers; ISBN: 041275780X, 238 pp.

Isaaks, E. H., and Srivastava, R. M., 1989, *An Introduction to Applied Geostatistics*, Oxford

University Press, New York, 561 pp.

Johnson, M. E., 1987, *Multivariate Statistical Simulation*, John Wiley and Sons, New York, 230 pp.

Journel, A.G., and Huijbregts, C., 1978, *Mining Geostatistics*, Academic Press, 600 pp.

Journel, A.G., 1988, *Geostatistics for the Environmental Sciences*, EPA Project Report, Project No. CR 811893, EPA/EMSL, Las Vegas, Nevada.

Journel, A.G., 1989, *Fundamentals of Geostatistics in Five Lessons*, American Geophysical Union, Washington D.C.

Kitanidis, P. K., 1997, *Introduction to Geostatistics: Applications in Hydrogeology*, ISBN 0-521-58747-6.

Koch, G., and Link, R., 1971, *Statistical Analysis of Geological Data*, volumes I and II, Wiley, 375 pp. now available through Dover reprints.

Krige, D., 1978, *Lognormal DeWijsian Geostatistics for Ore Evaluation*, South African I.M.M., 55 pp.

Knudsen, H.P., 1988, *A Short Course on Geostatistical Ore Reserve Estimation*, Division of Mining and Minerals Engineering, Montana Tech.

Manoukian, E.B., 1986, *Modern Concepts and Theorems of Mathematical Statistics*, Springer-Verlag, 156 pp.

Matheron, G., 1965, *Les Variables Regionalisees et leur Estimation*, Editions Masson, Paris.

Matheron, G., and Armstrong, M. [eds.], 1987, *Geostatistical Case Studies*, Reidel, 248 pp.

Monestiez, P., D. Allard, R. Froidevaux [eds.], 2001, *Geoenv III -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, ISBN: 0792371070, 555 pp.

Olea, R., 1999, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, Boston, MA, 303 pp.

Pannatier, Y., 1996, *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York, 91 pp.

Rendu, J., 1978, *An Introduction to Geostatistical Methods of Mineral Evaluation*, South African I.M.M., 100pp.

Ripley, B., 1981, *Spatial Statistics*, John Wiley and Sons, New York, 252 pp.

Rouhani, S., and M. Srivastava, 1997, *Geostatistics for Environmental and Geotechnical Applications*, ASTM Special Technical Publication 1283, American Society for Testing and Materials, ISBN: 0803124147.

Rivoirard, J., J. Simmonds, K. G. Foote, P. Fernandes, and N. Bez [eds.], *Geostatistics for Estimating Fish Abundance*, Blackwell Science Inc; ISBN: 0632054441, 206 pp.

Soares, A. [ed.], 1992, *Geostatistics Troia '92*, volumes 1 and 2, Reidel, 1088 pp.

Soares, A., Gomez-Hernandez, J., and R. Froidevaux [eds.], *Geoenv I -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, ISBN: 0792345908, 516 pp.

Thompson, S. K., 1992, *Sampling*, John Wiley and Sons, New York, 343 pp.

Verly, G., David, M., Journel, A.G., and Marachel, A. [eds.], 1984, *Geostatistics for Natural Resources Characterization*, volumes 1 and 2, Reidel, 1092 pp.

Webster, R. and M. A. Oliver, 2001, *Geostatistics for Environmental Scientists*, John Wiley and Sons, Ltd., Chichester, England, 271 pp.