**空间插值（Spatial Interpolation）**

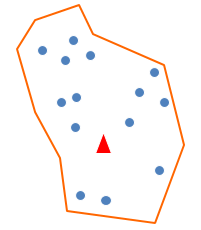
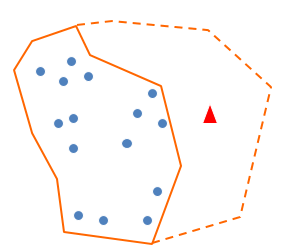
# 1. 概念

插值是用已知点来估算其他未知点的过程。需要插值的原因主要有：

* 现有数据不能完全覆盖所要求的区域；
* 现有离散曲面的分辨率、像元大小、方向与要求不符；
* 现有连续曲面的数据模型与要求不一致。

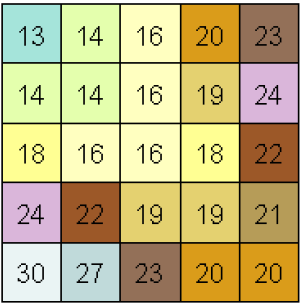
空间插值常用于将离散点的测量数据转换为连续的数据曲面，以便与其他空间现象的分布模式进行比较。它包括了空间内插和外推两种算法：

* 空间内插算法：通过已知点的数据推求同一区域未知点数据。
* 空间外推算法：通过已知区域的数据，推求其他区域数据。

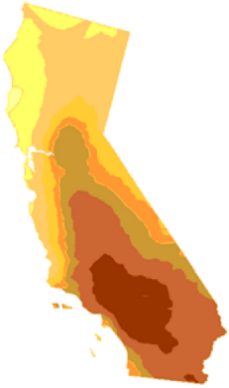
第一张图为内插，第二张图为外推。其中蓝色圆圈表示样本点，红色三角形表示预测点。

空间插值应用示例1——插值为降雨面：

第一张图表示输入降雨量点数据，第二张图表示插值后降雨量面。

空间插值应用示例2——插值为浓度面：

第一张图表示臭氧监测站的点位置，第二张图表示插值后臭氧浓度预测面。

# 2. 原理

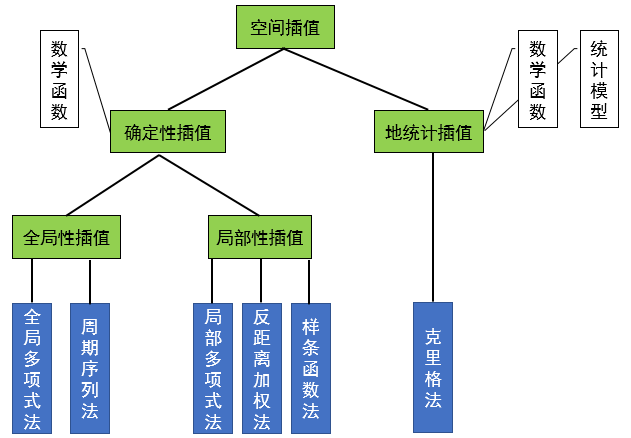
任何一种空间数据插值法都是基于空间相关性的基础上进行的。即空间位置上越靠近，则事物或现象就越相似；空间位置越远，则越相异或者越不相关。体现了事物/现象对空间位置的依赖关系。不同的插值方法对于如何获得最佳估计值都给予了一定的假设，但无论你选哪种插值方法，样本点越多，样本点分布越均匀，插值结果越接近实际值。

# 3. 分类

（1）根据数学原理：

**确定性插值法**是使用数学函数进行插值，以研究区域内部的相似性，或者以平滑度为基础，由已知样点来创建预测表面的插值方法。地统计插值依赖于**数学模型和统计模型**，基于自相关性（测量点的统计关系），根据测量数据的统计特征产生曲面。**地统计与确定性插值的最大区别在于，地统计插值引入了概率模型，即地统计插值认为从一个统计模型不可能完全精确地得出预测值，所以在进行预测时，应该给出预测值的误差，即预测值在一定概率内合理。**

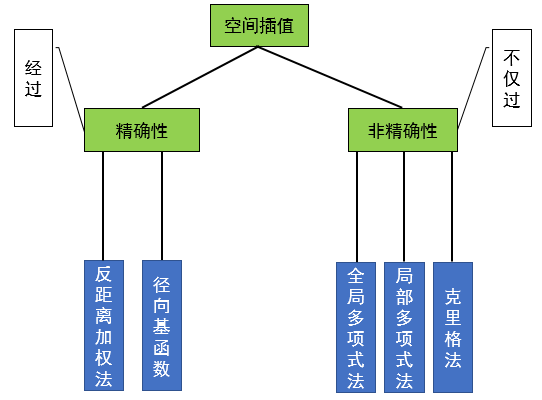
* 全局性插值（整体插值）：用研究区所有采样点数据进行全区特征拟合。整个区域的数据都会影响单个插值点，单个数据点变量值的增加、减少或者删除，都会对整个区域有影响。
* 局部性插值：只使用邻近的数据点来估计未知点的值。单个数据点的改变只影响其周围有限的数据点。



（2）根据是否能保证创建的表面经过搜有的采样点：

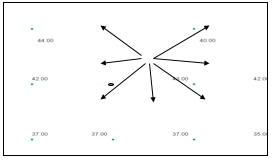
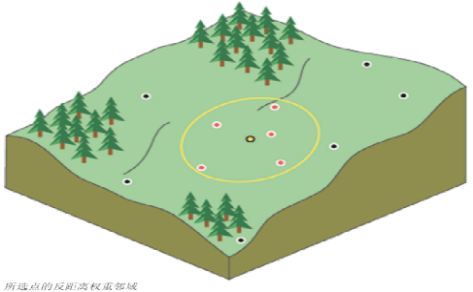
精确性插值：产生通过所有观测点的曲面。插值点落在观测点上，内插值等于估计值。（不是很清楚）

非精确（近似）插值：插**值产生的曲面不**通过所有观测点。当数据存在不确定性时，应该使用近似性插值，由于估计值替代了已知变量值，近似差值可以平滑采样误差。（插值时观察点值被改变了，用估计值替代）



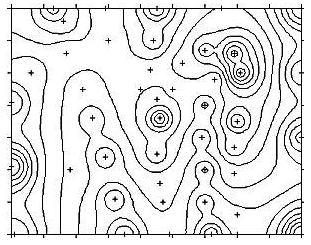
## 3.1 反距离加权法

反距离加权法(Inverse Distance to a Power)，也称为距离倒数乘方法。它认为某未知采样点的数据值与其周围一定范围内的已知采样点的数据值有关，是这些临近已知点的数据值综合贡献的结果，其贡献程度与距离成反比。该法综合了泰森多边形的邻近点法和多元回归法的长处，通过权重调整空间插值结构。

计算一个格网结点时给予一个特定数据点的权值与指定方次的从结点到观测点的该结点被赋予距离倒数成比例。当计算一个格网结点时，配给的权重是一个分数，所有权重的总和等于1.0。当一个观测点与一个格网结点重合时，该观测点被给予一个实际为 1.0 的权重，所有其它观测点被给予一个几乎为 0.0 的权重。换言之，该结点被赋给与观测点一致的值。

反距离加权法一般在数据点比较集中，且均匀分布的时候使用，是一种经常用来定位异常值的方法。但是如果数据点太少或者太稀疏，或分布不均匀，那这种方法就不太好用。如果你看到一幅等高线地图上过于频繁地出现下面的这种『牛眼』形状，那么这幅图很可能就是用反距离权重法画的。由于该方法在格网区域内会产生围绕观测点的“牛眼”，给电法与磁法数据解释带来不便，因此实际应用较少。



## 3.2 克里格法

克里格法(Kriging，也有翻译为克里金法)是地统计学的主要内容之一，从统计意义上说，是从变量相关性和变异性出发，在有限区域内对区域化变量的取值进行无偏、最优估计的一种方法；从插值角度讲是对空间分布的数据求线性最优、无偏内插估计一种方法。

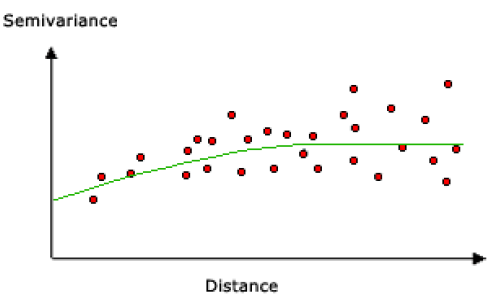
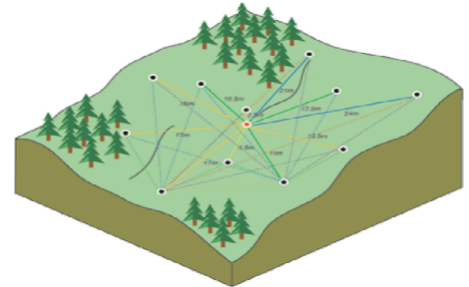
Kriging又称空间**自协方差最佳插值法**，是一种特定的滑动加权平均法，广泛地应用于地下水模拟、土壤制图、矿床中金属品位估计等领域。

根据样点数据统计特征的不同可将克里格分成多种不同的插值法：

（什么时候用什么方法，这里是理论的关键）

* 当样点数据是二进制值时，用指示克里格插值法进行概率预测；
* 对样点数据进行了未知函数变换后，可用该变换函数进行析取克里格插值；
* 当样点数据的趋势值是一个未知常量时，用普通克里格；
* 当样点数据的趋势可用一个多项式进行拟合，但回归系数未知时，用泛克里格插值法；
* 当样点数据的趋势已知时，用简单克里格插值法；

其中最常用的是普通克里格与泛克里格插值法；当加入了协变量进行插值时，则叫作协同普通克里格插值法和协同泛克里格插值法。

克里格插值与距离倒数加权插值的相似之处在于，二者都通过给已知样本点赋权重来求其他点的预测值。两种内插方法的通用公式如下：

其中，是已测得的第i个位置的属性值，是在第i个位置上测得值的权重，是待插值的位置，n是已知样点的数目。

反距离加权插值中，权重仅取决于样点到待插值点的距离。在克里格插值中，权重不仅考虑了已知点与插值点间的距离，而且考虑了己知点的位置和属性值整体的空间分布和格局。克里格插值中的权重来自**半方差函数模型（关键就是权重的选取）**（生成的表示地理现象连续表面的函数），在半方差函数模型和邻近已知点的空间分布的基础上，对研究区内的各个位置进行预测，权重取决于已知点的拟合模型、到插值点的距离和插值点周围的已知样点的空间关系。

## 3.3 最小曲率法

最小曲率插值法是一个非精确插值法，广泛应用于地球科学。其插值基准是生成一个具有最小曲率（即弯曲度最小），且到各样点的Z值的距离最小的曲面。最小曲率法试图在尽可能严格地尊重数据的同时，生成尽可能圆滑的曲面。

使用最小曲率法时要涉及到两个参数：最大残差参数和最大循环次数参数，来控制最小曲率的收敛标准。而且最小曲率法要求至少有四个点。实际应用中该法用于平滑估值，绘出的等值线主要用于定性研究。

## 3.4 径向基函数法

径向基函数(Radial Basis Function)即基函数是由单个变量的函数构成的，是一系列精确插值法的统称。所有径向基函数插值法都是准确的插值器，它们都能尽量适应的数据；若要生成一个更圆滑的曲面，对所有这些方法都可以引入一个圆滑系数。

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。

RBF是对最小曲率插值的改进，即属于精确的最小曲率插值法。RBF可以看作是一个高维空间中的曲面拟合（逼近）问题，学习是为了在多维空间中寻找一个能够最佳匹配训练数据的曲面，然后来一批新的数据，用刚才训练的那个曲面来处理（比如分类、回归）。RBF的本质思想是反向传播学习算法应用递归技术，这种技术在统计学中被称为随机逼近。RBF里的basis function（径向基函数里的基函数）就是在神经网络的隐单元里提供了提供了一个函数集，该函数集在输入模式（向量）扩展至隐空间时，为其构建了一个任意的“基”。这个函数集中的函数就被称为径向基函数。

## 3.5 趋势面分析法

作为一个非精确的插值方法，趋势面插值用多项式表示的线或面按最小二乘法原理对数据点进行拟合，并用于估算其他值的点，线和面多项式的选择取决于数据是一维还是二维。

在趋势面拟合中，空间位置以平面坐标为佳，即将经纬度坐标转换为以米为单位的平面大地坐标。通常趋势面分析用于分析趋势和异常，而不追求高的拟合精度，一般达到60-80%，阶数在1-4之间即可。拟合精度按R^2系数和F值检验。

趋势面分析是经典统计学在点数据进行空间展面上的应用，属于全局多项式插值，即对整个研究区域用一个多项式进行拟合。趋势面分析的一个基本要求就是，所选择的趋势面模型应该是剩余值最小，而趋势值最大，这样拟合度精确度才能达到足够的准确性。

## 3.6 样条插值法

样条插值(Spline Surface)的目标就是寻找一表面，使它满足最优平滑原则，也就是说，利用样本点拟合光滑曲线，使其表面曲率最小。样条函数是灵活曲线规的数学等式，为分段函数，一次拟合只有少数数据点配准，同时保证曲线段的连接处为平滑连续曲线。这就意味着样条函数可以修改曲线的某一段而不必重新计算整条曲线，插值速度快；保留了微地物特征，视觉上的满意效果。

样条插值法不适用于在短距离内属性有较大变化的地区，否则估计结果偏大。样条内插的误差不能直接估算。

## 3.7 谢别德法

谢别德法(Shepard's Method)使用距离倒数加权的最小二乘方的方法。因此，它与距离倒数乘方插值器相似，但它利用了局部最小二乘方来消除或减少所生成等值线的"牛眼"外观。谢别德法可以是一个准确或圆滑插值器。在用谢别德法作为格网化方法时要涉及到圆滑参数的设置，圆滑参数是使谢别德法能够象一个圆滑插值器那样工作。当你增加圆滑参数的值时，圆滑的效果越好。

## 3.8 线形三角网法

线性三角网法是最佳的徳洛内(Delaunay)三角形，连续样点数据间的连线形成三角形，覆盖整个研究区域，所有三角形的边都不相交。线性三角网法将在整个研究区域内均匀分配数据，地图上的稀疏区域会形成截然不同的三角面。

三角网插值器是一种严密的插值器，它的工作路线与手工绘制等值线相近。这种方法是通过在数据点之间连线以建立起若干个三角形来工作的。原始数据点的连结方法是这样：所有三角形的边都不能与另外的三角形相交。其结果构成了一张覆盖格网范围的，由三角形拼接起来的网。每一个三角形定义了一个覆盖该三角形内格网结点的面。三角形的倾斜和标高由定义这个三角形的三个原始数据点确定，给定三角形内的全部结点都要受到该三角形的表面的限制。因为原始数据点被用来定义各个三角形，所以样本数据是很受到尊重的。该方法适合于地层模型和断层的表示，也适合于大比例尺的磁法数据处理。

## 3.9 自然邻点法

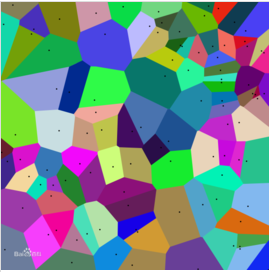
自然邻点(Natural Neighbor)插值法的基本原理是对于一组泰森(Thiessen)多边形，当在数据集中加入一个新的数据点（目标）时，就会修改这些泰森多边形，而使用邻点的权重平均值将决定待插点的权重，待插点的权重和目标泰森多边形形成比例。

该方法是对泰森多边形插值法的改进。它对研究区域内各点都赋予一个权重系数，插值时使用邻点的权重平均值决定待估点的权重。每完成一次估值就将新值纳入原样点数据集重新计算泰森多边形并重新赋权值，再对下一待估点进行估值运算。

## 3.10 最近邻点法

最近邻点法(Nearest Neighbor)又叫泰森多边形(Thiessen多边形)方法。在每个样点数据周边生成一个邻近区域，即Thiessen多边形，使得每个多边形内的任意一点离其内部的样点最近，在多边形内插值时只有其中心样点参与运算。

补充：泰森多边形是对空间平面的一种剖分，其特点是多边形内的任何位置离该多边形的样点（如居民点）的距离最近，离相邻多边形内样点的距离远，且每个多边形内含且仅包含一个样点。泰森多边形的特征：每个泰森多边形内仅含有一个离散点数据；泰森多边形内的点到相应离散点的距离最近；位于泰森多边形边上的点到其两边的离散点的距离相等。如下图：



# 4. 比较

|  |  |  |  |
| --- | --- | --- | --- |
| **常用插值方法方法** | **优点** | **缺点** | **应用** |
| **反距离加权插值法** | 算法简便易行，是一种精确插值算法。 | 对权重函数中幂次选择十分敏感，易受样本集群影响，造成“牛眼”效应。 | 样本点的分布均匀、密集且样本点含有区域特征信息。 |
| **克里格插值法** | 给出统计误差，充分考虑空间变量相关性，有效弥补数据集存在的聚类影响，插值精度高。 | 作为一种统计学方法，其计算步骤繁琐且插值速度慢。 | 适用于采样区域的样本数据存在随机性和结构性两特征。 |
| **径向基函数插值法** | 引入平滑因子，插值结果表面比较光滑。 | 当研究区域内采样点局部变异性大，采样数据具有较大的不确定时。难以保证数据的准确性。 | 采样数据局部变异不大时，像进行气象特征估计时可采用此法。 |
| **趋势面分析插值法** | 它是一种极易理解的技术，至少在计算方法上易于理解。另外，大多数数据特征可以用低次多项式来模拟。 | 当研究区域范围较大，地形很复杂时，需要用高阶多项式拟合以提高精度，但高阶将增加其计算成本，因而需要进行改进。 | 感兴趣区域的表面在各位置间出现渐变时，可以将该表面与采样点拟合，例如，工业区的污染情况；检查或排除长期趋势或全局趋势的影响。 |
| **最近邻点插值法** | 方法简单，计算率高，不受前提条件的约束。 | 插值结果受样本点的影响大，缺乏对其他空间因素和变量自身所具有的固有规律的考虑。 | 样本点分布均匀、完整，空间变异性不很明显，只存在少量数据的缺失。 |

# 5. 一般插值过程

①内插方法（模型）的选择；

②空间数据的探索性分析，包括对数据的均值、方差、协方差、独立性和变异函数的估计等；

③进行内插；

④内插结果评价；

⑤重新选择内插方法，直到合理；

⑥内插生成最后结果。

# 6. 插值方法选择的原则

①精确性。

②参数的敏感性：许多的插值方法都涉及到一个或多个参数，如距离反比法中距离的阶数等。有些方法对参数的选择相当敏感，而有些方法对变量值敏感。后者对不同的数据集会有截然不同的插值结果。希望找到对参数的波动相对稳定，其值不过多地依赖变量值的插值方法。

③耗时：一般情况下，计算时间不是很重要，除非特别费时。

④存储要求：同耗时一样，存储要求不是决定性的。特别是在计算机的主频日益提高，内存和硬盘越来越大的情况下，二者都不需特别看重。

⑤可视化、可操作性（插值软件选择）：三维的透视图等。

# 参考来源：

空间插值——插值方法的适用范围：

<http://www.cnblogs.com/zhuyuchen/p/3810811.html>

百度百科——空间插值：

<http://baike.baidu.com/link?url=R9r16MgZCp8iJKvPplziCABHHmcarY1uPWL567aBcVRa-TIOMjQeCV-Zd4FjwVxR19JccpgQHKPJjKIP6HWT8JexcKNp4EkxOm4FGG6xAJdt0-_fM8Zh1c9rNPbj1mM8>

泰森多边形(Voronoi图)生成算法：

<http://blog.csdn.net/gdut2015go/article/details/48208983>

地理科学学院地理信息系统专业网络课程教程第7章——空间数据分析模型：

<http://kc.njnu.edu.cn/dky/nb/page/2000-3-3/2000332117262480.htm>

空间插值方法对比整理版：

<https://wenku.baidu.com/view/d34f957827284b73f2425069.html>

空间插值方法简介：

<https://wenku.baidu.com/view/82c52c21192e45361066f5e3.html>

空间插值算法汇总：

<https://wenku.baidu.com/view/7941bbc0aa00b52acfc7ca90.html>

空间数据插值方法的分析与比较：

<https://wenku.baidu.com/view/c2476b75a417866fb84a8e6f.html>

10种插值方法在物探数据处理中的对比：

<https://sanwen8.cn/p/400ebZM.html>

克里格法（Kriging）：

<http://blog.sina.com.cn/s/blog_49babfed010007wd.html>

Kriging插值法：

<http://blog.sina.com.cn/s/blog_4aa4593d01014gwc.html>

克里金插值(kriging)：

<https://wenku.baidu.com/view/b497d80979563c1ec5da712a.html>

V-代约束的结构保持空间插值算法：

<http://www.fxyqpx.org/WHDXXBXXKXB/html/2015-6-823.htm>

插值算法（二）：反距离加权法IDW：

<http://blog.sina.com.cn/s/blog_816800900101f2lk.html>

径向基函数工作原理：

<http://blog.sina.com.cn/s/blog_63578f140100wvll.html>

插值算法（六）：趋势面插值（Trend Surface）：

<http://blog.sina.com.cn/s/blog_6316e2af0101l56y.html>

西南大学图书馆：

<http://202.202.96.94/>

地图上的各种等高线是怎么画出来的：

<http://daily.zhihu.com/story/9228981>