



本科毕业论文（设计）

题 目 空间插值技术的开发与实现

学 院 计算机与信息科学学院软件学院

专 业 网 络 工 程

年 级 2015 级

学 号 222015321210063

姓 名 陈 典 超

指 导 教 师 伍 胜

成 绩 _____

2019 年 02 月 28 日

目录

| | |
|--|----|
| 摘要: | 1 |
| Abstract: | 1 |
| 第 1 章 导论..... | 2 |
| 第 2 章 正文..... | 3 |
| 2.1 空间插值的基本原理..... | 3 |
| 2.2 两种插值方式（反距离加权插值法、普通克里金插值法的比较） | 3 |
| 2.2.1 基本思想比较..... | 3 |
| 2.2.2 算法比较..... | 4 |
| 2.2.3 实验结果比较..... | 5 |
| 2.3 变差函数..... | 6 |
| 2.3.1 变差函数的理论模型..... | 6 |
| 2.3.2 有 sill 的理论模型..... | 6 |
| 2.3.3 无 sill 的理论模型..... | 7 |
| 2.4 克里金插值法..... | 7 |
| 2.4.1 假设条件..... | 8 |
| 2.4.2 变差函数 γ | 8 |
| 2.4.3 无偏约束条件之期望..... | 9 |
| 2.4.4 克里金插值之估计误差分析..... | 12 |
| 2.5 克里金插值的实现..... | 17 |
| 2.5.1 系统流程图..... | 17 |
| 2.5.2 克里金插值流程图..... | 18 |

| | |
|----------------------|----|
| 2.5.3 数据结构..... | 19 |
| 2.5.4 输出结果..... | 20 |
| 第 3 章 结论..... | 22 |
| 参考文献: | 23 |
| 致谢: | 23 |
| 附录..... | 24 |
| 插值算法的一些参考链接..... | 24 |
| 普通克里金插值的大致流程图: | 25 |

空间插值技术的开发与实现

陈典超

西南大学计算机与信息科学学院 软件学院，重庆 400715

摘要：气温对生态系统的结构和功能具有十分重要的影响。因此，掌握气温空间分布的不同形态特征，然后使得气温的空间分布精确的展现出来具有十分重要的意义。而在实际的生活，采集数据时，不可能按照规则网格布点采集。而空间插值也由此应运而生。

本文是在分析了现实应用中采用一些插值方法，在数字高程模型的基础上，采用了克里金法的进行插值。因为克里金法能够充分考虑采样点所在的空间数据变量相关性，达到有效的削弱数据集聚类影响，实现高精度的空间插值。本文中开发出的插值算法，实现了离散数据集的网格化并达到插值的目的。相比于反距离权重法，克里金法更适用于样本数据分布具有随机性和结构性的采样区域。

关键词：空间插值；克里金法；离散采集点数据集

Development and implementation of spatial interpolation technology

CHEN Dianchao

School of computer and information science, Southwest University, Chongqing 400715, PR China

Abstract: Temperature has a very important influence on the structure and function of ecosystem. Therefore, it is of great significance to grasp the different morphological characteristics of the spatial distribution of temperatures, and then make the spatial distribution of weather more accurate. However, when collecting data, it is impossible to collect data according to regular grid arrangement, and the space interpolation comes into being.

In this paper, with some interpolation methods analyzed in practical application, the interpolation method of kriging is adopted on the basis of digital elevation model, Because kriging method can fully consider the correlation of spatial data variables where sampling points are located, it can effectively weaken the influence of data clustering and achieve high-precision spatial interpolation. In this paper, the interpolation algorithm realize the meshing of discrete data sets and achieve the purpose of interpolation further. Compared with the IDW, kriging method is more suitable for sampling areas with random and structural distribution of sample data.

Key words: interpolation; Kriging method; discrete collection point data sets

第 1 章 导论

自 20 世纪以来，全球气温升高、温室效应加剧等问题一直受到人们的广泛关注。地球历史上就曾出现过因气温的重大变化导致生态系统遭受巨大破坏而造成动植物的灭绝，地球冰河期等重大生态、地质变化；化石记录就曾经显示过去的 5 亿年里的物种大灭绝都和主要的气温变化有联系。然而，当前的气候变暖却不同以往，它已经被人类的社会生产极大的加速了。因此，气温的变化很可能对物种及其所需的生活环境产生重大影响一简而言之，就是对生态系统产生巨大影响。不仅如此，气温的升高还可能会导致温室效应、酸雨、臭氧层变化等对人类的日常生产生活造成巨大影响的极端气候。所以，观测人员需要对气温形势有更确切的认知，能够更加精细的描述气温的变化。人们只有在了解气温的大致分布后，才能也才有需要更细致的了解到某个区域一段时间内气温变化形势。因此，掌握气温的分布状态，并通过科技手段展现出气温的变化趋势对于研究人员来说具有十分重大的意义。

在实际生活中，为了更加准确、形象、具体的展现出气温分布图，在生态系统的保护和人类生产生活中及时的应对气温的变化，气温图也因此应运而生。因此所采集的数据点是否充足、绘图方式是否先进、绘图技术是否精湛等都直接影响生态环境的保护力度。卫星通过对全球的云层、地形等重要因素以及热成像等科学手段来采集大量的气温分布数据。根据采集方式的不同，大致分为网格状数据即规则的数据集以及不规则采样（如根据地形采样得到的数据）数据即离散数据集。虽然依据空间插值的原理，对于规则的采集点，气温图的绘制更加的方便合理准确。但是，在实际生活中，规则的采样点数据是不现实的，即数据集并不是沿网格状分布，而是根据地形或者其他空间因素变化的。正是基于不规则采样点数据的绘图，各种空间插值技术随之出现，以期解决这种不规则采样点可能导致的数据误差。

第 2 章 正文

2.1 空间插值的基本原理

空间插值 (https://en.wikipedia.org/wiki/Multivariate_interpolation) 是在空间自相关 (https://en.wikipedia.org/wiki/Spatial_analysis#cite_note-19) 的基础上根据大量已知点对部分未知点进行估计插值。Tobler 曾指明：任何事物（现象）与其他事物（现象）之间都是相互关联的，但近处的事物（现象）比远处的事物（现象）相关性更强即更依赖。也就是说：在空间（一维、二维、三维）中，空间距离越短，未知变量也就与观测数据之间越相似即越依赖；而空间位置越远，未知变量与观测数据的差异也就越大即相关性越弱。空间自相关体现了数据之间在空间位置上的依赖程度。插值算法的不同都是基于这一定律在对空间数据之间的关系做出假设后得出该数据在不同位置的估计值。但是，无论采用的是哪种插值算法，都与采样得到的观测值有关。即采样点越多，采样点分布越均匀，那么插值算法得到的估计值也越接近现实的实际值。

2.2 两种插值方式（反距离加权插值法、普通克里金插值法的比较）

2.2.1 基本思想比较

反距离加权插值法（Inverse Distance Weighted 维基百科：http://en.wikipedia.org/wiki/Inverse_distance_weighting）是根据反距离权重法（又称距离倒数乘方法）的思想使用数学方法在计算机上实现的。IDW 根据空间的自相关性认为插值点即（未知坐标点）的数据与该点周围一定区域内的已知采样点数据有关，是这些已知数据综合作用的结果，其影响大小与采样点到插值点的距离成负相关的关系。反距离权重法的重要组成部分之一是反距离的幂参数。幂值是基于观测点与插值点之间的距离来控制离散数据对内插值的影响。这种方法综合泰森多边形邻近点法等多种方法的优点，通过改变权值的方式来调整空间插值的结果。

克里金插值法（又称克里格法、克立格法、kriging 法等）（维基百科 <https://en.wikipedia.org/wiki/Kriging>）是一种另类的反距离加权法。从地理数据的角度上来说：它受到空间的相互作用、空间的扩散作用，在有限的范围对区域化变量取值在无偏约束条件下进行的最优估计的一种方式。从计算公式上来看：它与反距离插值法仅仅

只是权重的确定方式不同。但是从整个权重的计算过程来看，可以发现：反距离加权插值仅仅是根据空间距离直接确定权重。而克里金插值确实站在整个空间的角度上，根据空间的相关性、变异性对所有离散点进行综合分析得到变差函数后再来确定权重。

2.2.2 算法比较

反距离权重插值和克里金插值都是基于范围内的已知离散数据，来计算每个离散数据到内插点的权重，然后通过权重与范围内采样点数据的乘积累加的方式得到估计值，所以称克里金插值为一种另类的反距离加权法。

反距离权重插值使用的是一组基于幂参数的权重来确定插值点的估计值，它的权重是一个与距离成负相关的函数。其权重具体表达式为：

$$w_i = \frac{h_i^{-p}}{\sum_{j=1}^n h_j^{-p}}$$

其中 h_i 表示的是采样点与插值点的距离， p 表示其幂参数（一般情况下 p 值取 2）， w_i 表示其权重系数。通过加权函数我们可以看出：在使用较高的幂值的情况下，邻近区域的采样点数据将对内插点造成更大影响。插值估计值也就越接近其临近采样点，这也将造成插值后的数据在经过可视化后得到的表面更加不平滑。因此，常将 30 认为超大幂，不建议使用。然而在定义较小幂值时，采样点所影响的范围也就越大，即插值点受到了更多的采样点影响。此外，使用反距离加权法时需要注意：如果插值确定的范围或者幂值较大时可能得出误差较大的估计值（即估计值错误）。

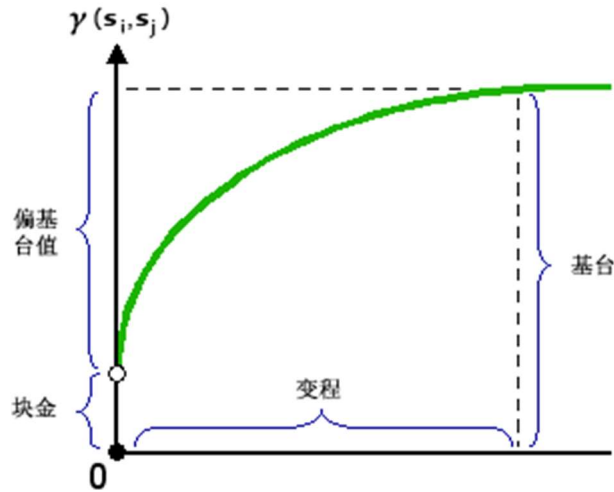
为了避免由于采样点过多造成的运算量过大等劣势，在进行反距离权重插值时，通常会确定可能影响插值点的范围，通过范围内的离散点来确定插值结果。反距离权重插值对于范围的确定有两种方式：一种是通过固定的搜索半径内的采样点数据来确定半径内的插值点；另一种是不断地改变半径，通过半径的变化确定一定数目的采样点所需要的半径大小。

克里金插值对于权重的确定方式与反距离加权法则完全不同。它是通过对于所有采样点两两之间距离与采样点数据半方差函数值构成的一组组数据，通过将这些数据通过画图拟合成一个函数即变差函数（又称半方差函数、协方差函数），通过变差函数，可以确定克里金插值采用的模型算法。根据所选择的模型得到系数矩阵、采样点与插值点之间的距离矩阵，最后根据两个矩阵得到权重系数矩阵。通过克里金插值的公式：

$$\widehat{Z(s_0)} = \sum_{i=1}^n \lambda_i Z(s_i)$$

得到插值点的估计值。

气象采样得到的函数拟合模型通常为球状模型。本文中普通克里金算法实现也是采用球状模型。画图拟合得到的变差函数如图：



可以确定基台值、块金常数和变程这三个重要参数。其中变程 a 表示的是范围，也就是说只有与插值点数据小于等于 a 的采样点数据才会对插值估计值产生影响。块金常数 C_0 （又称块金效应）指的是距离为零的方差值，表示的是相距很近的两点之间数据变化情况，基台值 $sill$ 表示的则是变量的变化幅度。

2.2.3 实验结果比较

反距离权重法综合泰森多边形等多种插值方法的长处，对插值点可以进行更圆滑的插值。而且，算法的实现更加简单。同样的，简单的算法也意味着其忽略了更多的细节。IDW 在插值时往往只根据距离插值，而忽略空间位置中变量的依赖关系，这也就导致插值时往往忽略空间中最大（最小）值，即插值如果是最大（最小）值时往往会被错误估值。最后造成插值的数据与真实值存在较大出入。IDW 在插值时还会在采样点分布不均时，因为插值数据往往不会存在最大最小值，造成插值数据点所产生的等值线图产生围绕一片区域内采样点最大值的“牛眼”现象

克里金插值法是一种在无偏约束条件下的代价函数最优解中所产生的插值点。它充分考虑了变量的空间的相关性、变异性，所以它在插值过程中常常会得到最佳的估计值。

但是，由于其充分利用了变量空间性质，所以在计算过程中存在计算量过大、插值时间过长等问题。

2.3 变差函数

空间插值的基本原理：空间中距离越短，事物也就越相似；距离越远，事物差异性也就越大。基于空间插值的基本原理，普通克里金插值认为：事物的相似度 γ 与空间距离 d 存在着一种函数关系，这种函数关系反应了空间的自相关性。它可能是线性函数，也可能是非线性函数。为了确定这种函数，克里金插值中将它分为理论变差函数和实验变差函数。

实验变差函数是在计算采样点空间位置距离 d 和采样点两两之间平方差构成的点的基础上，通过设定步长 k ，计算 $0\sim k$ 、 $k\sim 2k$ 、 $2k\sim 3k$ 等步长之间的平均距离和采样点数据在步长内的半方差得到这样的一组数据所构成的函数。但是由于空间具有变异性，所得到的函数往往不能用一些具体的表达式所呈现出来，而在插值过程中需要用到这个函数计算估计值。因此，需要通过实验变差函数画图拟合得到理论变差函数。

克里金法通过对平均距离和采样点数据对半方差所构成的数据对画出的散点图，使用一些经过总结、抽象然后加以概括出来的函数模型来对散点图进行函数拟合。拟合后的函数可以得到函数模型的重要参数：基台值 $sill$ 、块金常数 C_0 和变程 a 。

2.3.1 变差函数的理论模型

变差函数的理论模型是克里金插值中的一个重要板块，它往往包含三个重要参数基台值 $sill = C_0 + C$ 、块金常数、变程。克里金插值通过采用一些经典的理论模型来拟合实验变差函数，最符合散点图的函数模型即为最符合此次插值结果的理论模型。本文采用的便是对生态系统影响较大的气象因素所常使用的球状模型。常见的理论模型根据基台值 $sill$ 有无分为两大类。

2.3.2 有 $sill$ 的理论模型

1) 球状模型公式：

$$\gamma(d) = \begin{cases} 0 & d = 0 \\ C_0 + C \left(\frac{3}{2} * \frac{d}{a} - \frac{1}{2} * \frac{d^2}{a^2} \right) & 0 < d \leq a \\ C_0 + C & d > a \end{cases}$$

球状模型中的变程为 a ;

2) 指数模型公式:

$$\gamma(d) = \begin{cases} 0 & d = 0 \\ C_0 + C \left(1 - e^{-\frac{d}{a}} \right) & d > 0 \end{cases}$$

指数模型中的变程约为 $3a$;

3) 高斯模型公式:

$$\gamma(d) = \begin{cases} 0 & d = 0 \\ C_0 + C \left(1 - e^{-\frac{d^2}{a^2}} \right) & d > 0 \end{cases}$$

高斯模型中的变程为 $\sqrt{3} a$;

2.3.3 无 sill 的理论模型

1) 幂函数模型公式:

$$\gamma(d) = d^a \quad 0 < a < 2$$

2) 对数函数模型公式:

$$\gamma(d) = \log d$$

3) 纯块金效应模型公式:

$$\gamma(d) = \begin{cases} 0 & h = 0 \\ C_0(>0) & h > 0 \end{cases}$$

2.4 克里金插值法

与反距离权重法相似但更为抽象, 克里金插值的公式也为:

$$\widehat{Z(s_0)} = \sum_{i=1}^n \lambda_i Z(s_i)$$

它表示的是估计值 $\widehat{Z(s_0)}$ 在栅格点 s_0 处的插值等于权重系数 λ_i 与离散点值 $Z(s_i)$ (在 s_i 处离散点值的大小) 的累加。其中 $Z(s_i)$ 表示在范围内的离散点。此处的权重系数 λ_i , 并

非距离倒数的平方。它是基于无偏估计条件下：

$$E(\hat{z}_0 - Z_0) = 0$$

满足插值点估计值 \hat{z}_0 与真实值 Z_0 ：

$$\min_{\lambda_i} \text{Var}(\hat{z}_0 - Z_0)$$

最小的一组解。

2.4.1 假设条件

克里金插值有这许多假设条件，这些假设也对应着众多克里金插值方法。普通克里金插值方法的假设条件：对于整个空间来说，空间中的每个变量都有着相同的期望 E 和方差 σ^2 。即整个空间中不同范围内变量的期望、方差恒定（插值便是针对整个空间中不同范围内的某个变量估值）。

$$E[z(x, y)] = c$$

$$\text{Var}[z(x, y)] = \sigma^2$$

这也就意味着：空间内任意一点的数据 $z(x, y)$ 都等于常量 c 和随机偏差 $R(x, y)$ 之和：

$$z(x, y) = E[z(x, y) + R(x, y)] = c + R(x, y)$$

可以将它认为是克里金插值的变式。此处的随机偏差方差等于空间方差：

$$R(x, y) = \text{Var}[z(x, y)] = \sigma^2$$

2.4.2 变差函数 γ

在克里金插值中，为了更好的推导最优权重系数，我们定义了这么一个函数：变差函数（<https://en.wikipedia.org/wiki/Variogram>）

$$\gamma_{ij} = \frac{1}{2} E[(Z_i - Z_j)^2]$$

Z_i 、 Z_j 分别为空间中位置 i 、位置 j 对应的变量值。普通克里金插值认为空间中任一点都等于恒定的空间期望 c 与随机偏差 R 之和。由此，可以推出：

$$Z_i - Z_j = R_i - R_j$$

因此：

$$\begin{aligned} \gamma_{ij} &= \frac{1}{2} E[(R_i - R_j)^2] \\ &= \frac{1}{2} E[R_i^2] + \frac{1}{2} E[R_j^2] - E[R_i R_j] \end{aligned}$$

$$= \sigma^2 - cov(R_i, R_j)$$

由于克里金插值中假设空间中数据既有相近的属性，也在空间中有一定的相似度，而变差函数 γ_{ij} 表示了数据的相近属性，相似的空间则有空间距离 d_{ij} 来表示：

$$\begin{aligned} d_{ij} &= d(Z_i, Z_j) = d[(x_i, y_i), (x_j, y_j)] \\ &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \end{aligned}$$

为了得到这个数据模型的变差函数，我们将 (d, γ) 组成的散点图（常称为实验变差函数）对变差函数的理论模型进行拟合得到最优的变差函数：

$$\gamma = \gamma(d)$$

2.4.3 变差函数的实现

为了在设计中更好的实现变差函数的拟合，根据变差函数的基本原理，可以将变差函数实现分为以下几步：

- 1) 求采样点两两之间的二维数组 **data** 和高程差组成二维数组 **ZD**；
- 2) 将得到的距离数组按距离（步长 **h**）分组，分组可根据采样点多少来确定分组个数；
- 3) 根据分组计算每组的平均距离与平均高程差的平方；
- 4) 根据半方差公式：

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N [Z_i - Z_{i+h}]^2$$

计算出每组对应的变差函数值，得到点对 (h, γ) 。**N** 表示该步长内的点对值；

- 5) 将得到的点对画图得到散点图，求得三个重要参数基台值、块金常数、变程，用变差函数球状理论模型画图拟合；

sill 基台值表示的是变差函数随着距离的变化最终趋向的值，也就是变量在超过一定距离后，变差函数的值恒为基台值。可以认为是区域变量变异性的度量；

C₀ 块金常数表示的是距离为零，即距离无限接近时变差函数值。可认为是观测误差等引起的随机变化；

a 变程表示的是变量影响的范围。即在距离小于 **a** 的情况下，变量值与距离具有正相关性；距离大于 **a**，则变量与距离不具有相关性；

6) 得到变差函数理论模型;

7) 设计实现变差函数;

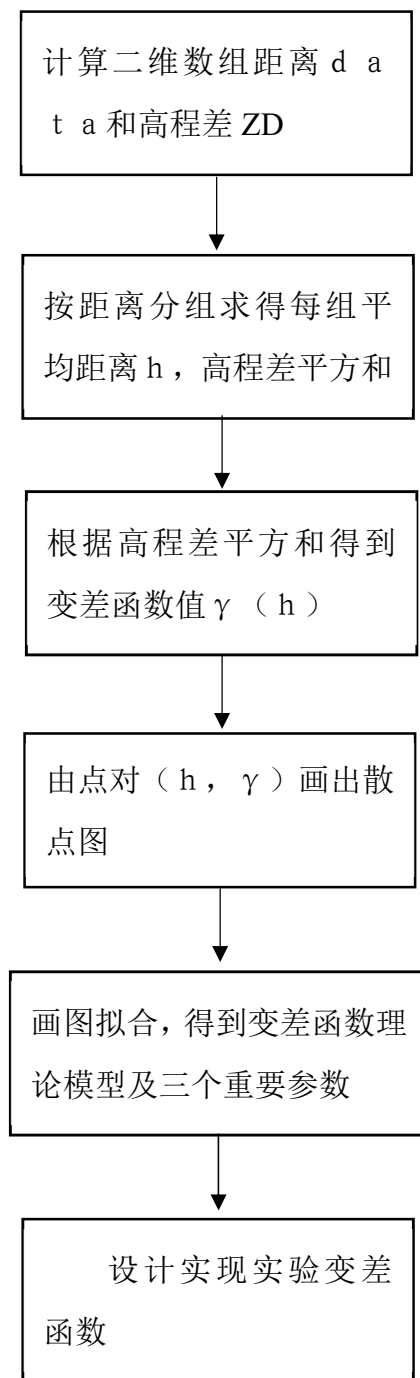
8) 实验变差函数的点对(h, V)实现:

```
for(k=h;k<=ZDmax+h;k+=h)//分组
{
    for(i = 0; i < data.length; i++)
    {
        for(j=0;j<data[0].length;j++)
        {
            if(k<=data[i][j]<k+h)
            {
                count++;
                distance+=data[i][j];
                Z+=ZD[i][j]2
            }
            r=Z/(2*count)
            h=distance/count
        }
    }
}
```

9) 理论变差函数的实现 (本设计采用球状模型)

```
function (double h)
{
    double r=0;//步长, 变差函数值
    if(h==0) r=C0;
    else if(h>=a) r=sill;
    else r=C0+sill*((3*h^4)/(4*a^4));
    return r;
}
```

2.4.4 变差函数流程图



2.4.5 无偏约束条件之期望

普通克里金法是基于无偏估计条件：

$$E(\hat{z}_0 - Z_0) = 0$$

下的，得出空间中所有的任何一个离散点对整个空间来说都有一个相同的期望 c 。因此在已知克里金插值公式的情况下，可得无偏估计条件为：

$$E\left(\sum_{i=1}^n \lambda_i Z_i - Z_0\right) = 0$$

普通克里金法插值假设条件中假定空间期望恒定且为 c ，则可化简为：

$$c \sum_{i=1}^n \lambda_i - c = 0$$

从而可以推断出：

$$\sum_{i=1}^n \lambda_i = 1$$

因此，可以认为当且仅当权重系数 λ_i 累加为一，插值估计值满足普通克里金假设条件中的无偏估计条件之一：空间中的每个变量对于整个空间中有一个相同的期望。

2.4.6 克里金插值之估计误差分析

克里金插值假设中无偏估计的另一条件便是空间内的任意一点对于整个空间都有唯一的方差，为了求得权重系数为一和方差最小时权重系数数组与变差函数值的关系，有了如下推导，克里金假设方差为常数：

$$\text{Var} [z(x,y)] = \sigma^2$$

同样的代入克里金插值公式可得：

$$\begin{aligned} f(x,y) &= \text{Var}(\hat{z}_0 - Z_0) = \text{Var}\left(\sum_{i=1}^n \lambda_i Z_i - Z_0\right) \\ &= \left[\text{var}\left(\sum_{i=1}^n \lambda_i Z_i\right) - 2\text{cov}\left(\sum_{i=1}^n \lambda_i Z_i, Z_0\right) + \text{cov}(Z_0, Z_0) \right] \\ &= \left[\sum_{i=1}^n \sum_{j=0}^n \lambda_i \lambda_j \text{cov}(Z_i, Z_j) - 2 \sum_{i=1}^n \lambda_i \text{cov}(Z_i, Z_0) + \text{cov}(Z_0, Z_0) \right] \end{aligned}$$

$\text{var}()$ 表示随机变量方差； $\text{cov}()$ 表示协方差

已知克里金插值公式的变式是空间期望与随机变量之和，由协方差的计算公式我们可以得到：

$$\text{cov}(Z_i, Z_j) = \text{cov}(R_i, R_j)$$

可以得到：

$$f(x, y) = \left[\sum_{i=1}^n \sum_{j=0}^n \lambda_i \lambda_j \text{cov}(R_i, R_j) - 2 \sum_{i=1}^n \lambda_i \text{cov}(R_i, R_0) + \text{cov}(R_0, R_0) \right]$$

由 2.4.2 可知变差函数变形公式：

$$\gamma_{ij} = \sigma^2 - \text{cov}(R_i, R_j)$$

整理可得：

$$\text{原式} = \sum_{i=1}^n \sum_{j=0}^n \lambda_i \lambda_j (\sigma^2) - \sum_{i=1}^n \sum_{j=0}^n \lambda_i \lambda_j (\gamma_{ij}) - 2 \sum_{i=1}^n \lambda_i (\sigma^2) + 2 \sum_{i=1}^n \lambda_i (\gamma_{i0}) + \sigma^2 - \gamma_{00}$$

已知只有当权重系数累加为一时满足无偏约束条件期望恒定，因此：

$$f(x, y) = 2 \sum_{i=1}^n \lambda_i (\gamma_{i0}) - \sum_{i=1}^n \sum_{j=0}^n \lambda_i \lambda_j (\gamma_{ij}) - \gamma_{00}$$

为了求得最小的一组 λ_i ，使得方差函数 $f(x, y)$ 最小，可以令函数 $f(x, y)$ 对 λ_i 求偏导，使得函数值为零，求得的函数值 λ_i 即为最小值。为了得到满足无偏约束条件下期望恒定（权重累加为一）的最优的权重系数 λ_i ，这是一个条件极值问题，就需要构造一个满足条件极值拉格朗日乘法方程组：

$$\psi = f(x, y) + 2\phi \left(\sum_{i=1}^n \lambda_i - 1 \right)$$

依次对该方程的每个权重系数 λ_i 和拉格朗日中构造的 ϕ 求偏导，得到一个方程组。这个方程组的解便是在无偏约束条件下的最优（最小）的权重系数

$$\begin{cases} \frac{\partial \psi}{\partial \lambda_j} = 0; & j = 1, 2, \dots, n \\ \frac{\partial \psi}{\partial \phi} = 0 \end{cases}$$

整理得：

$$\begin{cases} r_{j0} - \sum_{k=1}^n r_{jk} \lambda_k + \phi = 0; k = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

通过上述计算可以得到一个求解权重系数的方程组，化为矩阵形式即为：

$$[r_{ij}][\lambda] = [r_{i0}]$$

其中

$$r_{ij} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} & 1 \\ r_{21} & r_{22} & \cdots & r_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\phi \end{bmatrix}$$

$$\lambda_{i0} = \begin{bmatrix} \lambda_{10} \\ \lambda_{20} \\ \vdots \\ \lambda_{n0} \\ 1 \end{bmatrix}$$

最后得到的矩阵 λ 即为满足无偏约束条件的最优权重系数矩阵。

在对普通克里金插值的无偏约束条件中恒定期望与估计误差的最小值的推导后，可知只有在范围内使用的是基于无偏约束条件且方差（估计误差）最小情况下得到的最优权重系数，才会使得在权重系数与变量累加求和后得到的变量值为满足无偏估计条件的最优估计值。

2.4.7 无偏约束条件下权重系数的实现

- 1) 根据离散点距离数组求得离散点距离经变差函数得到的变差函数数组 **vdata1**（系数矩阵数组）；

变差函数数组 **vdata1** 的实现（注意：最后一列拉格朗日乘数系数为 1，最后一行为权重系数累加求和所以权重系数 λ 的系数为 1，最后一行的最后一个元素为对拉格朗日乘数求导后的拉格朗日乘数的系数，为 0）；

```

for(i = 0; i <= data.length; i++)
{
    for(j=0;j<=data[0].length;j++)
    {
        if(i=data.length||j=data.length) data[i][j]=1;
        else if (i=data.length&&=data.length) data[i][j]=0;
        else vdata[i][j]=function(data[i][j]);
    }
}

```

- 2) 根据离散点与插值点之间的距离经变差函数得到变差函数数组 `vdata2`。数组最后一个为函数对拉格朗日乘数求导后的系数，所以该系数为 1；

```

for(int i=0;i<=XLength;i++)
{
    if(i<XLength)
    {
        vdata2=function(math.sqrt((x-Xs[i])2+(y-Ys[i])2));
    }
    else vdata2 =1;
}

```

- 3) 将数组 `vdata1` 化为矩阵，根据线性代数矩阵公式 $a^{-1}a = e$ ；矩阵的逆乘以矩阵为单位矩阵.对矩阵 `vdata1` 进行求逆；

矩阵求逆

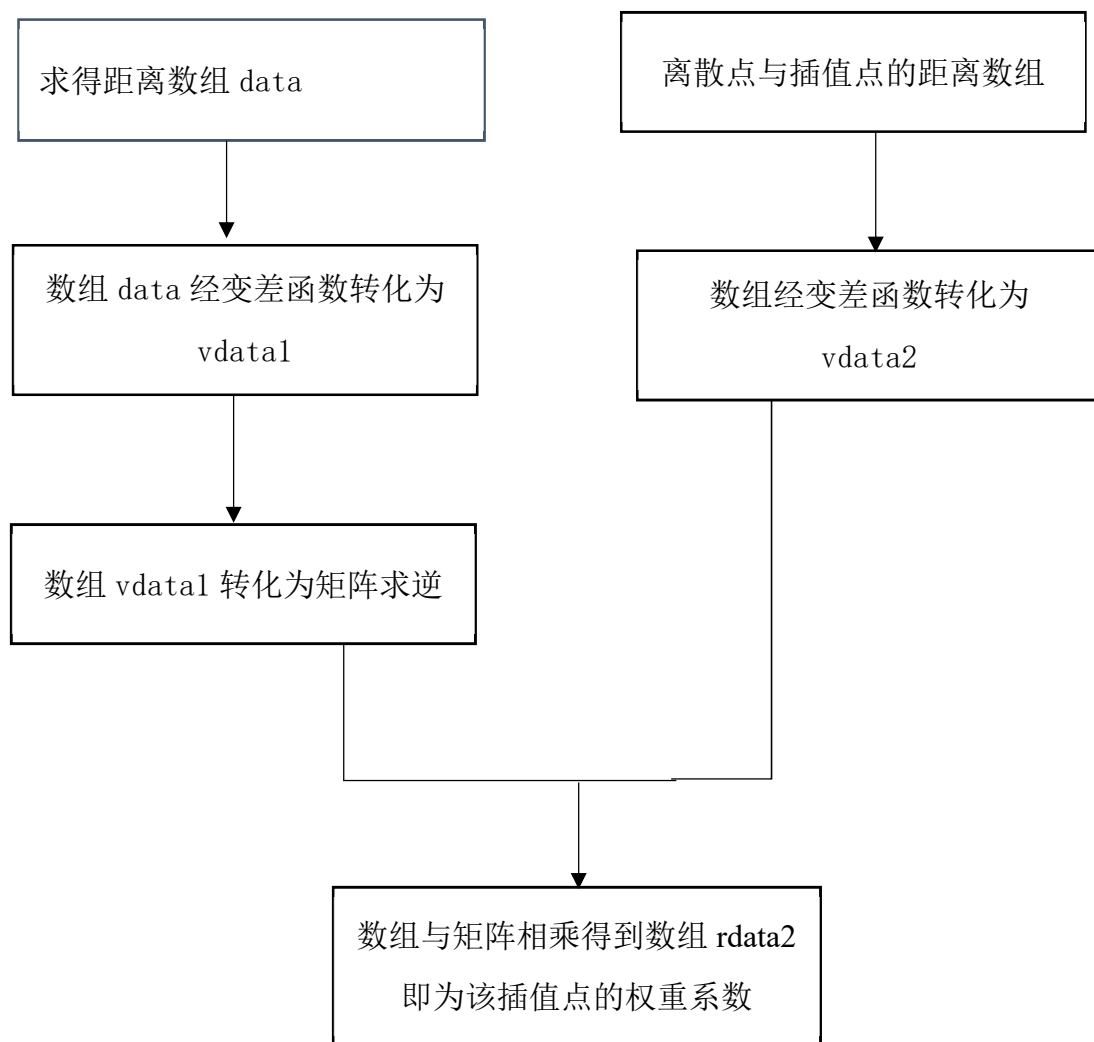
```

Matrix m = new Matrix( vdata );
Matrix M = m.inverse;

```

- 4) 将求逆后的矩阵 `vdata-1` 与 `vdata2` 相乘得到的数组 `rdata` 即为最优权重系数。利用矩阵的 `get()`函数即可得到矩阵某一行元素，通过某一行元素与 `vdata2` 相乘

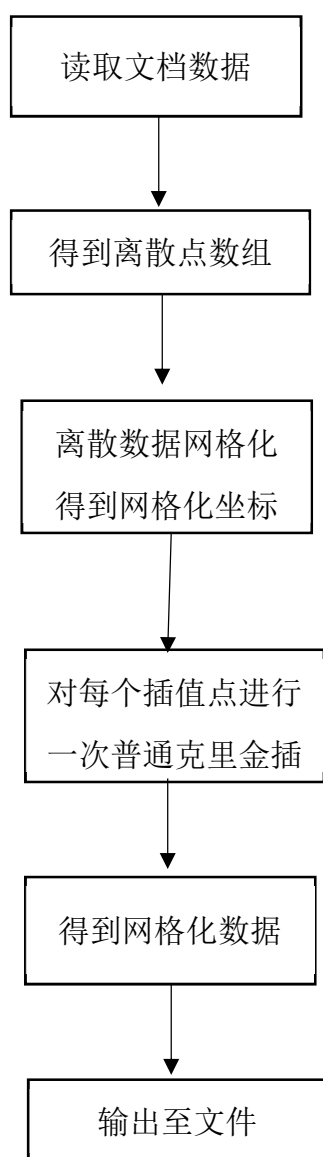
2.4.8 权重系数实现流程图



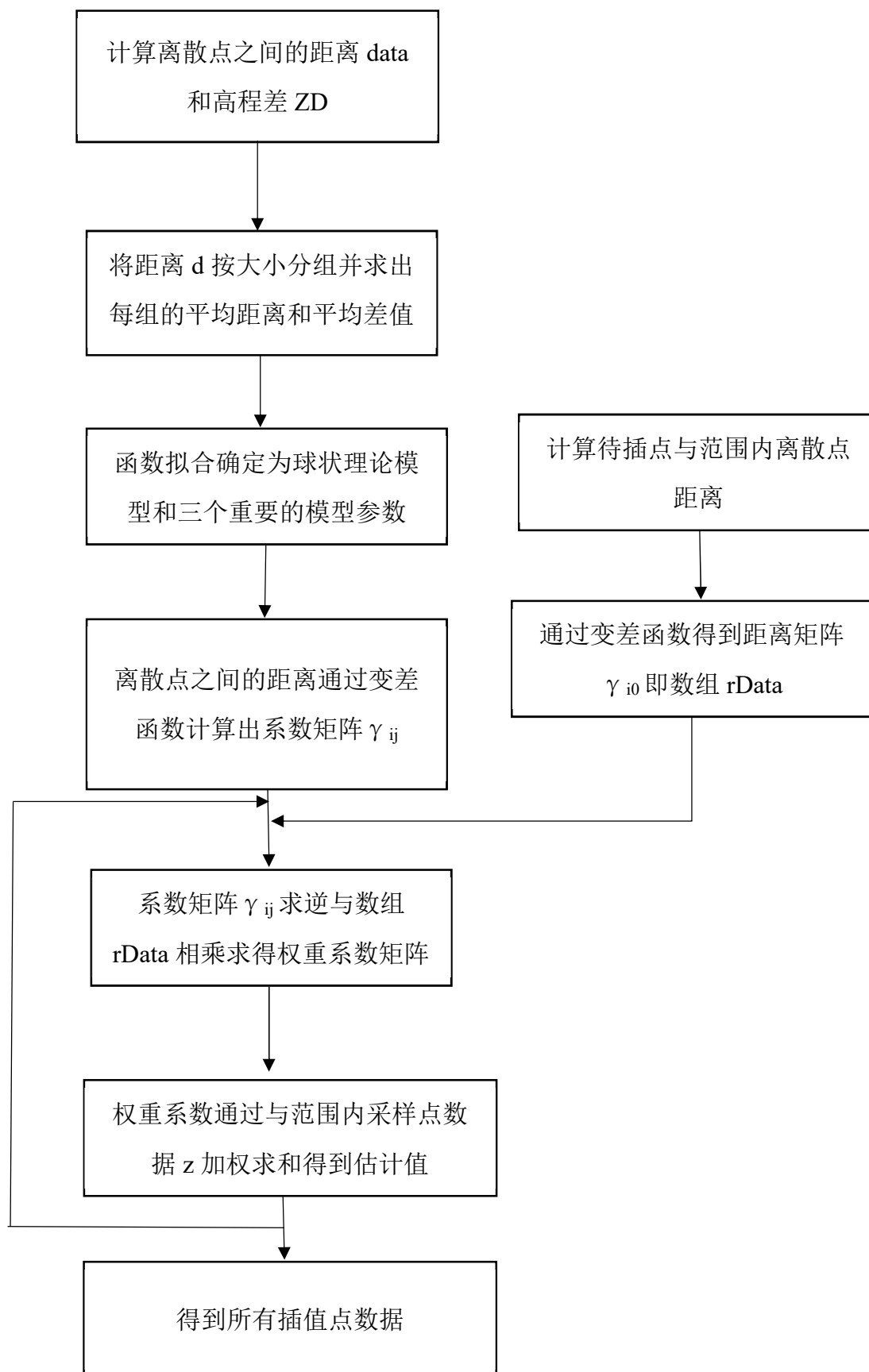
2.5 克里金插值的实现

本文基于上述的理论及方法实现克里金插值。代码是以 `eclipse` 为开发平台的。代码主要板块有文档的读取、数据的存储、数据分析、数据计算以及文档输出。

2.5.1 系统流程图



2.5.2 克里金插值流程图



2.5.3 数据结构

栅格点数据结构的定义：

```
public class DEM_Point {  
    private double x;  
    private double y;  
    private double z;  
  
    public DEM_Point( double x, double y, double z)  
    {  
        this.x=x;  
        this.y=y;  
        this.z=z;  
    }  
  
    public DEM_Point(double x,double y)  
    {  
        this.x=x;  
        this.y=y;  
    }  
  
    public double getX()  
    {  
        return x;  
    }  
  
    public void setX( double x)  
    {  
        this.x=x;  
    }  
}
```

```

public double getY()
{
    return y;
}

public void setY( double y)
{
    this.y=y;
}

public double getZ()
{
    return z;
}

public void setZ( double z)
{
    this.z=z;
}

public String toString() {
    return x + "    " + y + "    " + z;
}
}

```

2.5.4 输出结果

克里金插值得到的结果将以 txt 文件格式输出。Txt 文本内存储了栅格点的 x 坐标、y

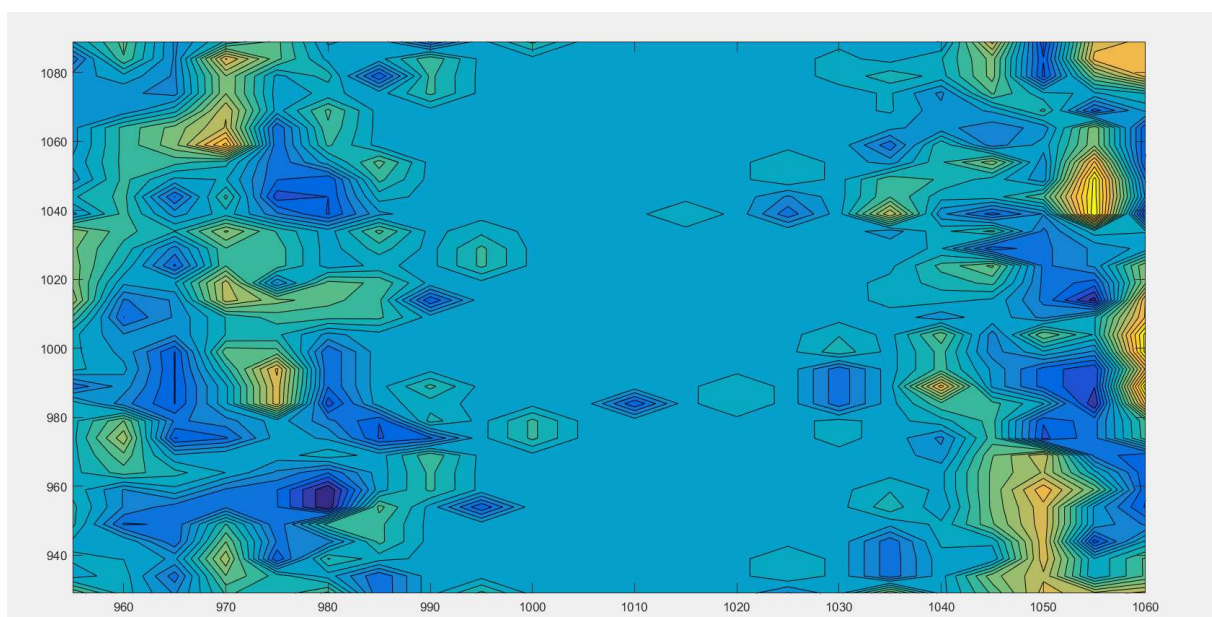
坐标以及 (x, y)坐标对应的属性值，根据文件可以通过软件 matlab 得到相应的等值线图。

文本文件输出结果：

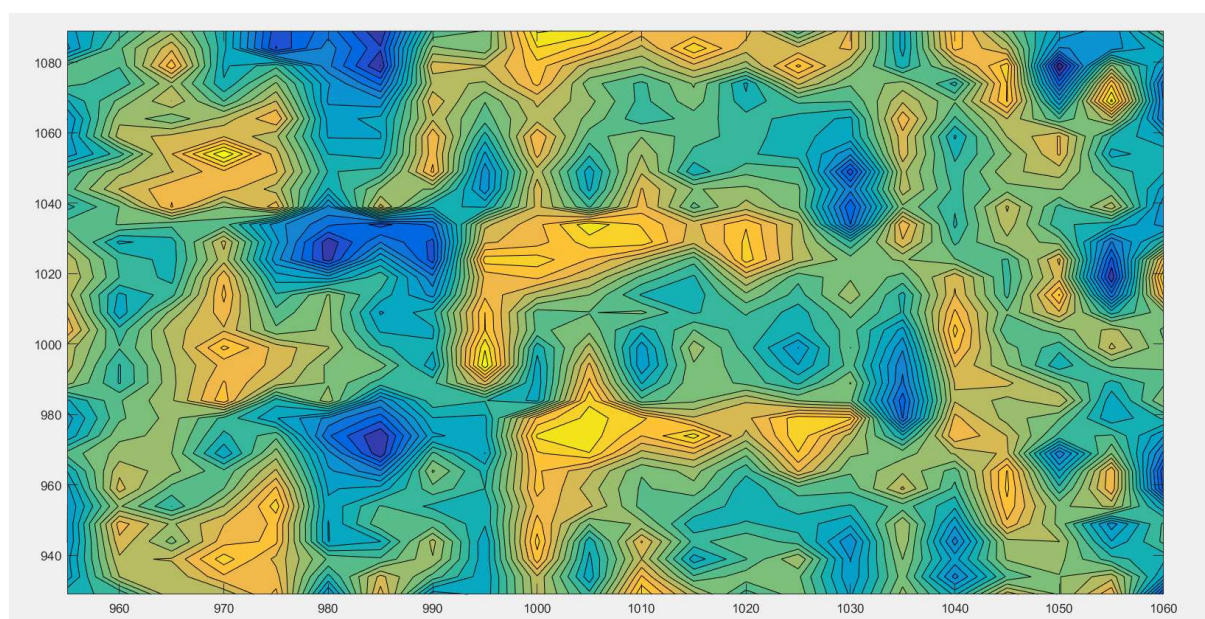
| X | Y | (x, y) |
|-------|--------|---------|
| 929.0 | 955.0 | 123.769 |
| 929.0 | 960.0 | 47.433 |
| 929.0 | 965.0 | 57.643 |
| 929.0 | 970.0 | 51.71 |
| 929.0 | 975.0 | 77.129 |
| 929.0 | 980.0 | 76.968 |
| 929.0 | 985.0 | 67.349 |
| 929.0 | 990.0 | 67.196 |
| 929.0 | 995.0 | 66.978 |
| 929.0 | 1000.0 | 66.68 |
| 929.0 | 1005.0 | 66.369 |
| 929.0 | 1010.0 | 66.12 |
| 929.0 | 1015.0 | -9.069 |
| 929.0 | 1020.0 | 60.049 |
| 929.0 | 1025.0 | 60.078 |
| 929.0 | 1030.0 | 60.182 |
| 929.0 | 1035.0 | 103.907 |
| 929.0 | 1040.0 | 154.392 |
| 929.0 | 1045.0 | 136.89 |
| 929.0 | 1050.0 | 134.167 |
| 929.0 | 1055.0 | 128.25 |
| 929.0 | 1060.0 | 137.12 |
| 934.0 | 955.0 | 33.187 |
| 934.0 | 960.0 | 70.119 |

克里金插值与 IDW 插值后得到的结果如图所示：

克里金插值画等值线后的结果：



IDW 插值画等值线后的结果：



根据结果可以发现：IDW 插值后得到的等值线图容易形成所谓的“牛眼”效应，即数据点容易围绕周围一个较大的变量来形成等值线。相对来说，普通克里金插值后得到的数据没有特别明显的牛眼现象；此外，可以看出 IDW 插值后，得到的数据没有特别明显的极大极小值，数据分布大体上较为均匀。而普通克里金插值后，数据分布差异极大，数据极大极小值十分显著。因此，可以看出克里金插值相比于 IDW 来说更加准确。

第 3 章 结论

本文基于空间气象研究中针对气象（如气温、云层等）的采样进行空间插值。基于空间插值的一些基本原理，对两种基本插值方法从不同的方面进行比较，最后采用插值更准确的普通克里金法。综合以上研究内容，可得出实验结论：

无论是从算法思想、算法实现、还是从插值结果的角度来看，普通克里金法在空间中的插值都远胜于 IDW。相对于反距离权重法，普通克里金插值对于空间自相关性具有更精确的定义，并且用数学方式更准确的将空间的相关性、差异性描述出来。因此，在空间中克里金插值法得到的估计值更接近真实值。

参考文献:

- [1] 李军龙, 张剑. 气象要素空间插值方法的比较分析[J]. 草业科学, 2006, 23(8): 6-10.
- [2] 林忠辉, 莫兴国, 李宏轩. 中国陆地区域气象要素的空间插值[J]. 地理学报, 2002, 57(1): 47-56.
- [3] Dubrule, Olivier. Two Methods with Different Objectives: Splines and Kriging[J]. Journal of the International Association for Mathematical Geology, 1983, 15(2): 245-257.
- [4] S. J. Yakowiz, F. Szidarovsky. A Comparison of Kriging with Nonparametric Regression Methods[J]. Journal of the multivariate analysis, 1985, 16(1): 21-23.
- [5] 卢耀华, 王胜明, 郭剑. Kriging方法在输电线路气象环境建模中的应用[J]. 计算机系统应用, 2019, 28(2): 264-269.
- [6] 周成虎, 裴涛. 地理信息系统空间分析原理[M]. 北京: 科学出版社, 2011: 70-72.
- [7] 著者. 文献题名[C]. 编者. 论文集名. 出版地: 出版者, 出版年: 起止页码.
- [8] 颜慧敏. 空间插值技术的开发与实现[D]. 四川: 西南石油学院, 2005.
- [9] 周体鹏. 基于克里金插值法的昆明市PM2.5预测[D]. 昆明: 云南大学, 2016.
- [10] 翟进乾. 克里金插值方法在煤层分布检测中的应用研究[D]. 太原: 太原理工大学, 2008.
- [11] 王子良. 空间数据预处理及插值方法对比研究 - 以铜陵矿区土壤元素为例[D]. 合肥: 合肥工业大学, 2010.
- [12] 彭思岭. 气象要素时空插值方法研究[D]. 长沙: 中南大学, 2015.

致谢:

本设计是在伍胜老师的细心指导下完成的。从设计选题、开题报告、毕业设计实现到论文撰写, 整个设计及论文的每一个细节与老师的悉心教导有着密不可分的关系。在论文撰写及设计实现的过程中, 遇到过许许多多的难题, 在老师的精心指导下, 这些问题都一一得到解决。所以在此, 向老师表示衷心的感谢

古人云: 不积跬步, 无以至千里。本设计能够成功完成, 也离不开各位老师渊博的知识和认真负责的教学态度。感恩之情难以表述, 谨以最诚恳的言语致以最崇高的敬意: 谢谢你们!

其次, 感谢大学四年朝夕相处的室友和孜孜求学的同学。从遥远的东南来到这陌生的城市, 是室友们的帮助, 让我渐渐熟悉这座城市, 使我逐渐融入这大西南; 是同学们孜孜不倦的求学态度感染了我, 让我慢慢的对计算机这个专业感兴趣。

最后, 感谢我的亲人。他们不仅支持我完成学业, 还在日常生活与学习中给予我莫大的支持。他们的默默支持, 是我前进的最大精神动力!

附录

插值算法的一些参考链接

Wikipedia 首页:

<https://www.wikipedia.org/>

反距离加权法的实现:

<https://blog.csdn.net/r1254/article/details/48159515>

反距离权重法:

https://blog.csdn.net/IT_bigstone/article/details/82753486

反距离加权插值的 C#实现:

<https://blog.csdn.net/dahongdahong/article/details/51959018>

空间插值的一些插值方法及对比:

<https://blog.csdn.net/kikitamoon/article/details/7857375/>

空间中散乱点的插值:

<https://wenku.baidu.com/view/64b723c54128915f804d2b160b4e767f5acf80c4.html>

K r i g i n g 插值的原理及公式推导:

<https://xg1990.com/blog/archives/222>

变差函数的概念及分析:

<https://wenku.baidu.com/view/438282296edb6f1aff001fee.html>

克里金插值法:

<https://www.cnblogs.com/einyboy/p/3196679.html>

克里金插值算法中的理论变差函数模型和算法实现的步骤:

http://blog.sina.com.cn/s/blog_695752850101lug3.html

整理空间插值学姐的 GitHub 网址:

<https://github.com/IronLavender/Spatial-Interpolation>

我的 GitHub 地址:

<https://github.com/ChenDianchao/Kriging>

普通克里金插值的大致流程图：

插值算法的理论步骤：

普通克里金方法的基本步骤如下：

- 1、计算采样点之间的距离，求解普通克里金方程组的关键在于先构建。

$$\begin{bmatrix} 0 & Y_{12} & Y_{10} & \dots & Y_{1n} & 1 \\ Y_{21} & 0 & \dots & \dots & Y_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & Y_{n2} & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} Y_{10} \\ Y_{20} \\ \vdots \\ Y_{n0} \\ 1 \end{bmatrix}$$

的系数矩阵以及右侧的列阵，即先求变差 Y_{ij} 与 Y_{i0} ，而变差函数只与采样点之间的距离有关，而与采样点的坐标无关，因此可以先计算出各测量点之间的距离。

- 2、距离分组，先将上一步计算所得的距离值按从小到大排成顺序，然后将距离值分成若干组，每组包含一定数量的距离值。

- 3、拟合变差函数 $\gamma(h) = h$ ，针对每组距离值，计算出平均距离 \bar{h}_i ，并按式

$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$ 即算出该组内变差函数的估计值；然后选择一定的变差函数理论模型，进行函数拟合，求出模型参数，从而得到变差函数 $\gamma(h)$ 的表达式。

- 4、计算式

$$\begin{bmatrix} 0 & Y_{12} & Y_{10} & \dots & Y_{1n} & 1 \\ Y_{21} & 0 & \dots & \dots & Y_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & Y_{n2} & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} Y_{10} \\ Y_{20} \\ \vdots \\ Y_{n0} \\ 1 \end{bmatrix}$$

的系数矩阵与右侧的列阵，将相应采样点、待插点之间的距离值代入变差函数中，计算出式中的所有变差。

- 5、计算待插点的估计值，针对每个待插点，求解

$$\begin{bmatrix} 0 & Y_{12} & Y_{10} & \dots & Y_{1n} & 1 \\ Y_{21} & 0 & \dots & \dots & Y_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & Y_{n2} & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} Y_{10} \\ Y_{20} \\ \vdots \\ Y_{n0} \\ 1 \end{bmatrix}$$

，即可得到所有对应的加权系数，然后按照 $Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$ ，即可求得区域变量在待插点 x_0 处的估计值。由于

$$\begin{bmatrix} 0 & Y_{12} & Y_{10} & \dots & Y_{1n} & 1 \\ Y_{21} & 0 & \dots & \dots & Y_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & Y_{n2} & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} Y_{10} \\ Y_{20} \\ \vdots \\ Y_{n0} \\ 1 \end{bmatrix}$$

的系数矩阵是不随待插点的变化而变化的，因此，对于所有待插点而言，系数矩阵可以一次形成，但是右侧的列阵因待插点不同而不同，因此，对于每个待插点需要单独计算。

插值算法的实现步骤：

