基于克里金插值的湖底地形拟合

摘要

在对湖泊资源的勘探与利用中，运用有限的测量点来尽可能全面地测得湖泊的地理信息是一项重要的技术。本文围绕湖泊地形问题展开讨论。

问题一要求给出一种建立湖底曲面的模型的函数表达式，为此，首先确定湖面的边界信息，由于确定湖面边界为圆形，故采取确定的半径来固定湖面边界，确定半径为**28.91979km**。接着运用**克里金插值**的方法，对数据点进行丰富，尽可能减小拟合误差。接着运用**多项式拟合**的方法来求出湖底深度表达式，，采取最高次为 5 次，所得结果均方误差为 3.26565。

问题二要求对第一问结果进行误差分析，根据第一问所用的数学方法，猜测其误差可能的来源，并进行分析，发现初始点分布不均匀可能是这里误差产生的重要因素。

问题三要求给出新增测量点的分布建议，首先，根据问题一运用克里金插值所求得的各个点集的**变异系数**，按照位置加变异系数进行**加权 k-means 聚类**，取变异系数最大的 10 个聚类区域，取他们的聚类中心为需要增加的测量点的位置，最后再次进行变异系数的计算，发现增加测量点后，整体变异系数略有下降。

问题四要求计算湖水总量，已知湖地曲面表达式，通过**二重积分**的方法，即可求得湖水体积。湖水总量为 31726379459.8173m^3

关键字： 克里金插值；　多项式拟合；　变异系数；　加权 k-means 聚类；　二重积分

一、问题重述

1.1 问题背景

我国湖泊众多，湖内资源丰富，但由于内陆湖泊等测量区存在较多的水产养殖区(具有较多的水草、网簖、地笼等阻碍船只航行的固定装置)，使得此类测量区受到外界影响而不能直接获得水深值。因此，需要考虑在有限数据采集点的基础上，采用内插的方法来填补数据采集过程中的空白，以建立较为精确的湖底模型并用以估算湖水总量。

1.2 问题要求

本文围绕根据有限数据点建立湖底曲面模型算法，对数据点的有限增添与计算湖水总量方法这一系列问题的解决来展开。

1.2.1 问题一

以湖中心为原点建立坐标系，(x,y) 为点的坐标，单位是公里，f(x,y) 是 (x,y) 处湖底的深度，单位是米，根据给定的有限数据点，给出一种建立函数 f(x,y) 的表达式的方法。

1.2.2 问题二

假设湖底曲面足够光滑，试分析给出的这种方法的误差；

1.2.3 问题三

为提高精度，还可以增加十个测量点，试给出测量点分布的建议；

1.2.4 问题四

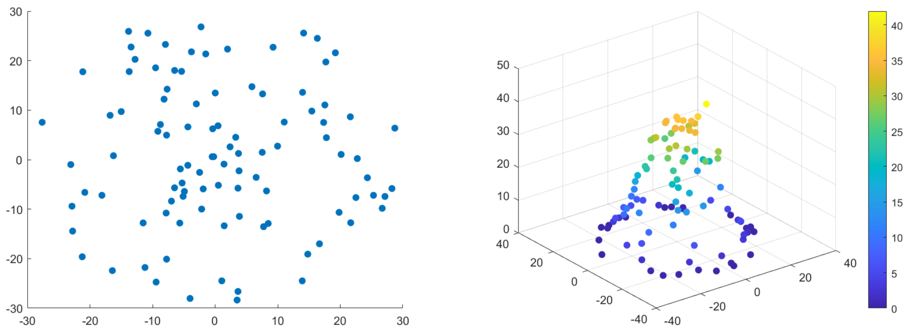
给出求湖水总量的一种近似计算方法，并对误差做出相应的理论分析。

二、问题分析

2.1 问题一分析

根据问题一题干及所给数据，我们不难得出要通过多项式曲面拟合来获得 f(x,y) 的表达式，但通过绘制原始数据点分布图，发现原始数据点有测量误差、分布过稀疏等问

题，直接进行曲面拟合效果不佳，故我们先要确定湖泊的正确边界，并采用先插值后拟合的方法获得更精确的表达式。



(a) 二维分布图　　(b) 三维分布图

图 **1**　原始数据分布图

2.2 问题二分析

问题一运用了克里金插值和多项式拟合的方法，这些方法各自存在优缺点，为此，应该综合两种方法可能出现的误差来进行分析。

2.3 问题三分析

问题三要求给出新增测量点的分布建议，在第一问得到的克里金插值的变异系数的基础上，运用加权 K—means 聚类的方法，以变异系数为主要指标，以位置坐标为次要指标进行聚类，按变异系数排列聚类编号，取前十个 cluster 的聚类中心为新增测量点坐标。

2.4 问题四分析

针对问题四，在已知湖底曲面表达式的情况下求湖水总量，即可以对表达式做二重积分求解得湖的体积，来近似求得湖水总量。

三、 模型假设

(1) 湖面边界假设：假设湖泊的边界为圆形，并且湖面边界的半径是固定的，即为 28.91979 公里。

(2) 湖底地形光滑性假设：假设湖底地形是足够光滑的，可以通过多项式拟合来有效地描述其深度变化。

(3) 数据点误差假设：假设在给定的数据点中，测量误差是均匀分布的，对整体模型的影响可以通过拟合和插值方法有效减小。

(3) 空间平稳性假设：假设在克里金插值过程中，数据的空间结构满足平稳性假设，即在较小范围内数据的空间依赖性相似。

(3) 体积计算的精确性假设：假设通过二重积分计算得到的湖水总量是一个近似精确的值，即误差可以忽略不计。

四、 符号说明

符号	说明
r	极坐标中的半径，单位为 km。
θ	极坐标中的角度，单位为度（°）
z	点的深度。
z'	克里金插值法预测的深度值。
ω_i	克里金插值中的权重。
γ_{ij}	已知点 i 和点 j 之间的半方差。
K_{ij}	克里金矩阵中，已知点 i 和点 j 之间的半方差矩阵。
λ	克里金插值的权重向量。
γ	已知点与插值点之间的半方差向量。
CV	变异系数，表示插值的可靠性。
$Z^*(x)$	位置 x 处的克里金估计值。
$Z(x)$	位置 x 处的真实值（若已知）。
$E[(Z^*(x) - Z(x))^2]$	克里金估计误差的均方误差（MSE）。
$E[Z(x)]$	真实值 $Z(x)$ 的期望值。
$D_w(x, c)$	加权距离，数据点 x 到质心 c 的加权距离。
c_j	第 j 个簇的加权质心。
C_j	第 j 个簇的数据点集合。

五、 模型的建立与求解

5.1 问题一模型的建立与求解

针对问题一，通过问题分析，我们确定了确定半径、插值、曲面拟合的三步走策略。

- **确定半径**

提取出原始数据中深度为 0 的点，通过转换极坐标，我们发现其半径并不相同，故我们先采用取平均值的策略确定较为准确的半径为 28.91979km。

表 **1**　深度为 **0** 的点的极坐标

极坐标半径 r/km	极坐标角度 θ
28.90214	48.32227
28.91169	42.74008
28.32391	81.87927
28.85179	-11.6718
29.21654	61.05315
28.6497	-15.2741
29.44775	56.31082
28.57606	-82.9286
29.36978	-61.9104
29.41526	12.50396
28.45305	-20.208

再根据假设 (1) 湖面为圆形，我们通过增添均匀分布在圆周的 30 个深度为 0 的数据点，使得数据更加完整，利于后续的插值与曲面拟合的准确性。

而对于插值预测数据点，我们采用国际主流的地理学算法————克里金插值法

- **克里金插值法的介绍**

克里金插值方法作为地统计学代表算法, 是一种利用协方差函数对随机场进行空间

建模和插值的回归算法, 它能够实现空间自协方差最佳插值, 故在众多领域都有着广泛的应用。

克里金插值法基于的假设是：一点的属性值与其周围点的属性值有关，且可以由其周围点的属性值导出。

克里金插值基于地理学第一定律产生其核心思想，即两点属性值差异性与二者间距离在一定距离范围内成正相关。

由此推导出克里金插值的基本公式为

$$z'_0 = \sum_{i=1}^{n} z_i \omega_i \tag{1}$$

- **采用克里金插值法的模型构建**

1 数据准备：收集和整理已知点的数据，包括二维位置和深度，并进行区域网格划分。

2 变差函数（Variogram）分析：通过计算不同距离下的数据对之间的半方差，构建变差函数模型。并通过高斯模型拟合。

已知点与插值点之间的半方差矩阵构建：

$$\gamma_{ij} = \gamma(d_{ij}) \tag{2}$$

3 克里金矩阵构建：构建已知点之间的克里金矩阵

$$K_{ij} = \gamma(d_{ij}) + \text{nugget} \cdot \delta_{ij} \tag{3}$$

其中，δ_{ij} 是 Kronecker delta 函数，当 $i = j$ 时，$\delta_{ij} = 1$，否则 $\delta_{ij} = 0$。nugget 为距离趋近于零时的变差。

4 计算权重：通过求解线性系统 $K\lambda = \gamma$ 得到权重 λ：

$$\lambda = K^{-1}\gamma \tag{4}$$

其中，γ 是已知点与插值点之间的半方差向量。

5 计算插值：利用公式（1）计算克里金插值，获得插值点深度。

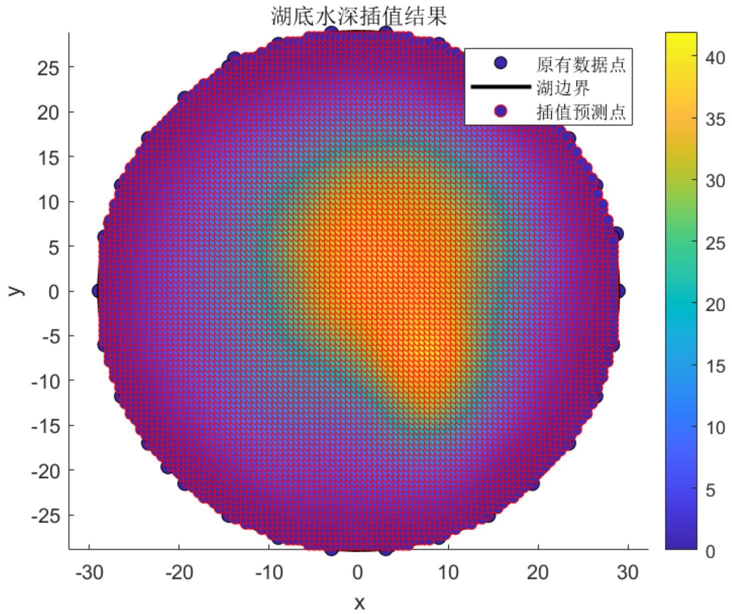


图 **2**　湖水水深插值结果

- **曲面拟合**

我们采用多项式拟合来得出表达式，先设出表达式基本形式：

$$z = f(x, y) = \sum_{i=0}^{d} \sum_{j=0}^{d-i} a_{ij} x^i y^j \tag{5}$$

为了防止拟合次数不足或过拟合，采取遍历拟合最高次数多次拟合的形式，通过比较预测值与实际值的均方误差，获得合适的多项式表达式。均方误差计算如下：

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (z_i - \hat{z}_i)^2 \tag{6}$$

表 **2**　多项式拟合不同次数的均方误差

多项式最高次数	均方误差
2	20.14712
3	14.87679
4	6.606152
5	3.265649

- **求解结果**

得到曲面拟合图像如图 3

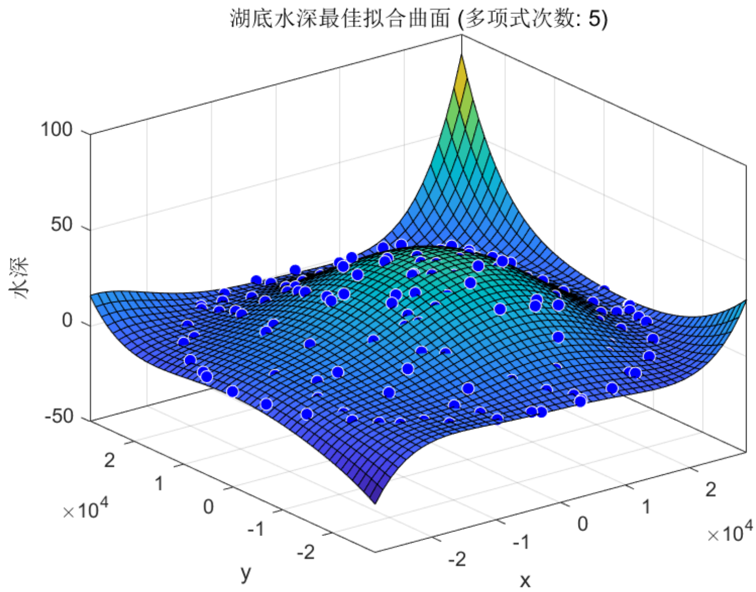


图 **3**　曲面拟合图像

求解得表达式：

$$
\begin{aligned}
\text{best\_fitresult}(x,y) = {} & 34.52 + 0.001009x + 0.0004063y - 7.845\times10^{-8}x^2 - 2.451\times10^{-9}xy - 7.743\times10^{-} \\
& - 3.315\times10^{-12}x^3 - 1.299\times10^{-12}x^2y - 1.706\times10^{-12}xy^2 - 1.106\times10^{-12}y^3 \\
& + 4.372\times10^{-17}x^4 + 7.27\times10^{-18}x^3y + 9.212\times10^{-17}x^2y^2 + 2.584\times10^{-18}xy^3 \\
& + 4.163\times10^{-17}y^4 + 2.619\times10^{-21}x^5 + 1.155\times10^{-21}x^4y + 3.127\times10^{-21}x^3y^2 \\
& + 1.667\times10^{-21}x^2y^3 + 4.21\times10^{-22}xy^4 + 7.413\times10^{-22}y^5
\end{aligned}
$$

5.2 问题二

克里金插值的误差来源

- 模型误差（Model Error）
  克里金插值依赖于半变异函数模型（如指数模型、高斯模型、球状模型等）的选择。若所选的模型不符合实际数据的空间结构，则会引入模型误差。
- 数据不充分或分布不均
  若数据点数量不足或分布不均匀，可能会导致插值结果不准确。克里金方法假设空间平稳性，如果该假设不成立，也会导致误差。本题测量点数据分布较为离散。

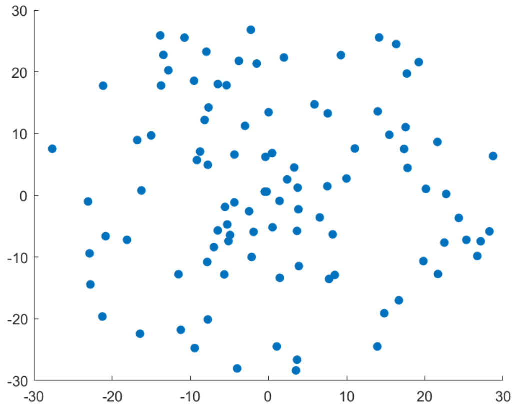


图 **4**　测量点分布图

- 参数估计误差（Parameter Estimation Error）

  半变异函数模型的参数估计不准确，会导致预测值的误差。这些参数包括基台（nugget）、基台（sill）、范围（range）等。
- 空间异质性

  当数据点间的空间依赖关系不一致时，克里金插值的假设（例如同质性）会失效，从而引入误差。

多项式曲面拟合的误差来源

- 模型误差（Model Error）

  多项式曲面拟合中选择的多项式阶数可能不足以捕捉数据的真实变化趋势，或者过度拟合（多项式阶数过高）导致的误差。
- 数据不充分或分布不均

  数据点不足或分布不均会影响拟合的准确性，特别是在边缘或数据稀疏区域。
- 测量误差（Parameter Estimation Error）

  与克里金插值类似，数据本身的测量误差也会影响拟合结果。。
- 外推误差

  多项式拟合的结果在数据范围之外（外推区域）通常不可靠，因为模型在这些区域没有数据支持，误差会显著增加。
- 多重共线性

  当自变量之间存在较强的相关性时，拟合的多项式模型参数估计会不稳定，导致拟合结果误差较大。

5.3 问题三模型建立

5.3.1 加权 k-means 聚类介绍

加权 K-means 聚类是一种在标准 K-means 聚类算法基础上进行扩展的方法，通过为每个数据点分配一个权重，从而对聚类结果施加更多的控制。这种方法特别适用于数据集中包含不同重要性或频率的数据点的情况。

标准 K-means 聚类算法的基本步骤如下：

- 初始化：随机选择 K 个初始聚类中心（质心）。
- 分配簇：根据每个数据点与质心的距离，将数据点分配给最近的质心所在的簇。
- 更新质心：计算每个簇的所有数据点的均值，并将该均值作为新的质心。
- 迭代：重复步骤 2 和 3，直到质心不再发生显著变化（即收敛），或达到预定的迭代次数。

加权 K-means 聚类在标准 K-means 的基础上引入了权重，使得每个数据点对质心更新的贡献由其权重决定。主要步骤如下：

- 初始化：随机选择 K 个初始聚类中心（质心）。
- 分配簇：根据每个数据点与质心的加权距离，将数据点分配给最近的质心所在的簇。加权距离通常为：

$$D_w(x,c) = w_i \times d(x,c) \tag{7}$$

  其中，w_i 是数据点的权重，$d(x,c)$ 是数据点 x 和质心 c 之间的距离。
- 更新质心：计算每个簇的加权质心。新的质心为簇内所有数据点的加权平均值：

$$c_j = \frac{\sum_{x_i \in C_j} w_i \times x_i}{\sum_{x_i \in C_j} w_i} \tag{8}$$

  其中，w_i 是数据点 x_i 的权重，C_j 是第 j 个簇的数据点集合。
- 迭代：重复步骤 2 和 3，直到质心不再发生显著变化或达到预定的迭代次数。

总体而言，处理具有不同重要性的特征数据时，加权 K-means 可以根据特征的重要性分配权重，从而提高聚类的准确性。加权 K-means 聚类提供了一种灵活的机制，可以根据数据点的重要性调整聚类结果，从而更好地捕捉数据的结构和特征。

5.3.2 模型求解

由问题一的克里金插值得到点的变异系数，变异系数代表插值的可靠性，变异系数越高，插值越不可靠，代表越需要新增测量点。

变异系数 CV 的计算公式为：

$$CV = \frac{\sqrt{E[(Z^*(x) - Z(x))^2]}}{E[Z(x)]} \tag{9}$$

其中：$Z^*(x)$ 是位置 x 处的克里金估计值。

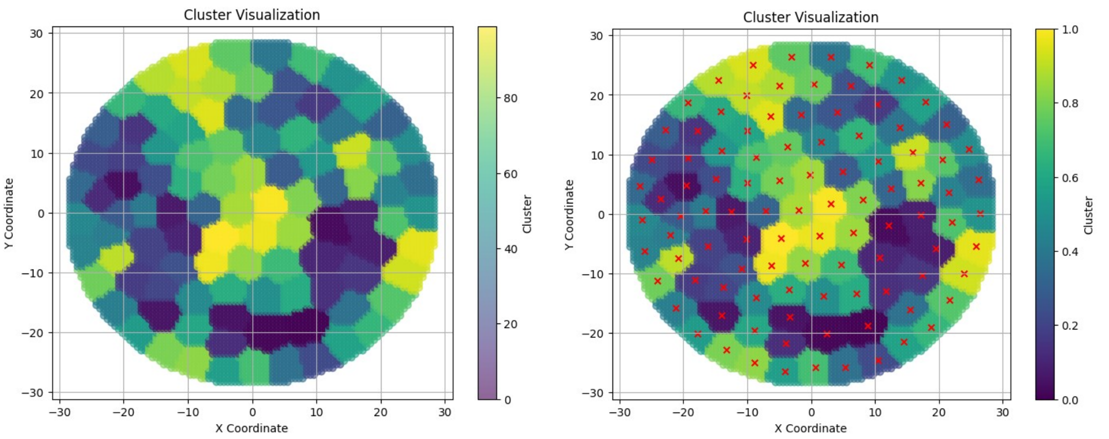
- $Z(x)$ 是位置 x 处的真实值（若已知）。

- $E[(Z^*(x) - Z(x))^2]$ 是克里金估计误差的期望值，也就是均方误差（MSE）。

- $E[Z(x)]$ 是真实值 $Z(x)$ 的期望值。

简化公式中，分子是估计误差的均方根（RMS），分母是真实值的期望值。该公式表示相对误差的大小，通常用于衡量克里金插值的精度或不确定性。如果真实值 $Z(x)$ 无法获得，可以使用克里金插值的估计值代替，分母可以用克里金估计值的期望值来近似。

将聚类结果 cluster 控制为 100 个，按聚类的变异系数高低做热图，其中，黄色区域表示，该区域平均变异系数高；绿色区域表示，该区域插值较为可靠。去变异系数高的聚类结果的聚类中心为新增测量点，如表3。



(a) 聚类结果　　(b) 聚类结果（含聚类中心）

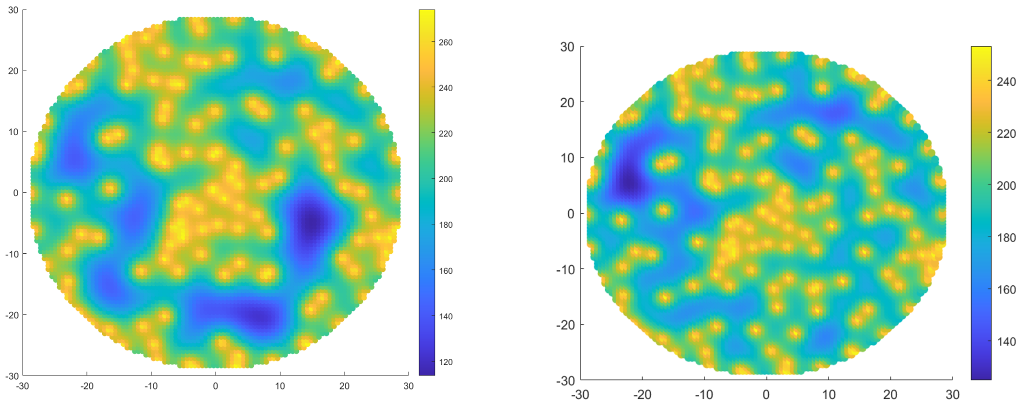
图 **5**　聚类结果图

5.3.3 模型效果分析

对比新增测量点前后的的变异系数热图，发现改进测量点前的的变异系数峰值超过260，而改进后的变异系数出现了可见的下调。

表 3　新增测量点坐标

x	y
15.86830921	-6.647035622
4.735425712	-21.34873735
-15.99952135	-15.01334755
-1.568785524	-20.61214852
16.37868403	-0.273102399
8.744931086	-17.22269266
12.71622968	-12.14971978
-11.04068199	-4.850651898
11.08176295	-1.999170606
1.944004806	-16.5788927



(a) 更新测量点之前　　(b) 更新测量点点之后

图 6　增加测量点前后变异值对比

5.4 问题四模型的建立与求解

5.4.1 湖水总量计算

已知表达式：

$$
\begin{aligned}
f(x,y) = {} & 34.52 + 0.001009x + 0.0004063y - 7.845\times10^{-8}x^2 - 2.451\times10^{-9}xy - 7.743\times10^{-8}y^2 \\
& - 3.315\times10^{-12}x^3 - 1.299\times10^{-12}x^2y - 1.706\times10^{-12}xy^2 - 1.106\times10^{-12}y^3 \\
& + 4.372\times10^{-17}x^4 + 7.27\times10^{-18}x^3y + 9.212\times10^{-17}x^2y^2 + 2.584\times10^{-18}xy^3 \\
& + 4.163\times10^{-17}y^4 + 2.619\times10^{-21}x^5 + 1.155\times10^{-21}x^4y + 3.127\times10^{-21}x^3y^2 \\
& + 1.667\times10^{-21}x^2y^3 + 4.21\times10^{-22}xy^4 + 7.413\times10^{-22}y^5
\end{aligned}
$$

对此表达式做 $\sqrt{x^2+y^2} \leq 28919.79$ 的二重积分,求解得近似湖水总量:31726379459.8173m^3

5.4.2 误差理论分析

1 测量点测量误差：测量点的地理位置可能存在偏差，导致测量点的位置不准确。在每个测量点测得的深度数据可能会受到湖底淤泥的影响存在误差，导致湖底高度数据不准确。

2 多项式拟合模型误差：多项式拟合的曲面可能无法准确反映湖底真实曲面，存在偏差，也会导致积分结果与真实湖水总量有偏差

3 测量点数量和分布导致的误差：在偌大的湖面只有 100 个数据点，无法反映湖底的真实情况，会存在偏差

4 在计算湖水总量时，默认湖中空间被水填满，但实际情况是湖中有水生生物、湖底废物残骸等的体积影响，会使估算值比真实值来得大。

六、 模型优缺点分析

6.1 克里金插值法（Kriging Interpolation）

优点：

- **空间自相关性利用**：克里金插值法利用半方差函数有效捕捉空间数据的自相关性，实现了高精度的空间预测。它在文章中通过选择适当的半变异函数模型，增强了插值结果的可靠性。
- **不确定性评估**：该方法提供了插值结果的标准误差（如变异系数），使得可以对插值的不确定性进行量化。这在文章中用于确定测量点的可靠性，帮助评估插值的准确性。
- **处理不均匀数据**：克里金插值法在处理数据分布不均匀的情况下表现良好，通过加权处理不均匀的数据点，从而减少数据稀疏对结果的影响。
- **灵活性**：该方法可以适应多种空间变异模型（如高斯模型、指数模型等），使其具有较强的适应性，可以应对不同的数据特性。

缺点：

- **计算复杂度高**：在处理大量数据时，克里金插值法的计算复杂度较高，特别是在计算克里金矩阵的逆矩阵时，可能导致计算时间长。
- **模型选择依赖**：插值结果依赖于半变异函数模型的选择，如果模型选择不当，可能会引入较大的预测误差。这在文章中表现为对模型选择的敏感性。
- **数据要求**：克里金插值法需要足够的样本点来确保结果的准确性。数据点不足时，插值结果可能不够可靠，这在处理数据稀疏区域时尤为明显。
- **对假设敏感**：方法假设数据具有空间平稳性，若实际数据不符合这一假设，则可能导致插值结果的不准确。

6.2 多项式曲面拟合（Polynomial Surface Fitting）

优点：

- **模型简单**：多项式曲面拟合的数学模型简单，易于实现和理解。对于数据点分布均匀且变化趋势明显的情况，拟合效果较好，能够提供明确的模型表达式。
- **计算效率高**：相较于克里金插值法，多项式曲面拟合的计算复杂度较低，尤其是在处理大量数据时，计算速度较快，适合快速拟合。
- **可解释性强**：多项式拟合的结果可以通过多项式系数直接解释，模型参数的物理意义较为清晰，便于理解和应用。

缺点：

- **过拟合风险**：多项式的阶数选择不当可能导致过拟合，尤其是高阶多项式容易对训练数据中的噪声过于敏感，影响预测的稳定性和准确性。
- **外推误差**：多项式拟合在数据范围之外的预测通常不可靠，因为模型在这些区域没有数据支持，误差会显著增加。
- **多重共线性**：高阶多项式可能导致自变量之间存在强相关性，影响模型参数的稳定性，从而导致拟合结果的误差增加。
- **无法处理复杂非线性关系**：对于复杂的非线性数据分布，多项式拟合可能无法捕捉真实的数据变化趋势，导致拟合效果不佳。

6.3 加权 K-means 聚类（Weighted K-means Clustering）

优点：

- **处理数据重要性**：加权 K-means 聚类通过为数据点分配不同的权重，能够考虑数据点的重要性或频率，从而改进聚类结果的准确性。这在文章中用于确定需要新增测量点的位置，以提高数据的可靠性。

- **提高聚类质量**：权重的引入使得模型能够处理数据点分布不均的问题，特别是对重要数据点的加权，使得聚类结果更为准确。
- **适应性强**：加权 K-means 允许调整权重以适应不同的应用需求，提供了较强的灵活性，能够根据实际数据的特点进行调整。
- **计算复杂度适中**：与一些其他复杂聚类算法相比，加权 K-means 的计算复杂度适中，较易实现和理解，同时能够处理较大规模的数据集。

缺点：

- **选择权重困难**：确定数据点的权重可能比较困难，特别是缺乏明确依据的情况下，权重选择不当可能会影响聚类效果的准确性。
- **对初始值敏感**：加权 K-means 聚类对初始质心的选择较为敏感。如果初始质心选择不当，可能导致聚类结果的不稳定和不准确。
- **局部最优解**：类似于传统 K-means，加权 K-means 聚类可能会陷入局部最优解，难以找到全局最优的聚类结果。
- **权重设置复杂性**：在实际应用中，权重的设置可能需要较多的领域知识和经验。如果权重设置不合理，可能会导致聚类结果的失真。

参考文献

[1] 傅珺. 邻域约束的克里金插值方法 [D]. 浙江大学,2022.DOI:10.27461/d.cnki.gzjdx.2022.002852.

[2] 靳国栋, 刘衍聪, 牛文杰. 距离加权反比插值法和克里金插值法的比较 [J]. 长春工业大学学报 (自然科学版),2003,(03):53-57.

[3] 张丹丹, 游子毅, 郑建, 等. 基于改进的局部异常因子检测的优化聚类算法 [J]. 微电子学与计算机,2019,36(11):43-48.DOI:10.19304/j.cnki.issn1000-7180.2019.11.009.

[4] 陈光, 任志良, 孙海柱. 最小二乘曲线拟合及 Matlab 实现 [J]. 兵工自动化,2005,(03):107-108.

[5] 白世彪, 陈晔, 王建. 等值线绘图软件 SURFER7.0 中九种插值法介绍 [J]. 物探化探计算技术,2002,(02):157-162.

[6] 李海涛, 邵泽东. 空间插值分析算法综述 [J]. 计算机系统应用,2019,28(07):1-8.DOI:10.15888/j.cnki.csa.006988.

附录 A　克里金插值

```
function [zi, var_zi] = kriging(x, y, z, xi, yi, range, sill, nugget)
% 数据点的数量
n = length(x);
% 插值点的数量
m = length(xi);

% 计算距离矩阵
d = pdist2([x, y], [xi, yi]);

% 计算半方差矩阵
gamma = spherical_variogram(d(:), range, sill, nugget);
gamma = reshape(gamma, [n, m]);

% 计算克里金矩阵
D = pdist2([x, y], [x, y]);
K = spherical_variogram(D, range, sill, nugget);
K = K + nugget * eye(n);

% 计算权重
weights = K \ gamma; % K_inv * gamma

% 计算预测值
zi = weights' * z; % n x 1 -> m x 1

% 计算变异数
var_zi = sill - sum(weights .* gamma, 1)';
end
```

附录 B　问题三变异系数的计算

```
    % 读取Excel文件中的数据
    data = xlsread('666.xlsx');
    % 提取前两列
    matrix_10x2 = data(:, 1:2);
    % 初始化一个新的10x4矩阵
    result = zeros(10, 4);
    % 对于每一行，在bianyizhi中找到最接近的行
    for i = 1:size(matrix_10x2, 1)
    % 计算欧氏距离
    distances = sqrt(sum((bianyizhi(:, 1:2) - matrix_10x2(i, :)).^2, 2));
```

```
% 找到最小距离对应的行索引
[~, row_index] = min(distances);

% 提取对应的行到结果矩阵中
result(i, :) = bianyizhi(row_index, :);
end
% 加载数据
data = readmatrix('t1_data.xlsx', 'sheet', 'add30');
% 提取x, y坐标和水深
x = data(:, 1);
y = data(:, 2);
z = data(:, 3);
x=[x;result(:,1)];
y=[y;result(:,2)];
z=[z;result(:,3)];
% 定义湖的边界
radius = 28.9197874968846; % 已知半径
theta = linspace(0, 2*pi, 100);
lake_boundary_x = radius * cos(theta);
lake_boundary_y = radius * sin(theta);

% 创建插值网格
grid_size = 100;
[xi, yi] = meshgrid(linspace(-radius, radius, grid_size), linspace(-radius, radius,
    grid_size));

% 排除网格中不在湖内的点
mask = sqrt(xi.^2 + yi.^2) <= radius;
xi = xi(mask);
yi = yi(mask);

% 计算半方差
vstruct = variogram([x, y], z, 'nrbins', 15, 'plotit', false); % 15个间隔

% 提取距离和半方差
distance_bins = vstruct.distance;
gamma_bins = vstruct.val;

% 初始参数猜测
initial_params = [max(distance_bins)/2, var(z), 0];

% 使用最小化函数来拟合参数
options = optimoptions('fminunc', 'Display', 'off');
fit_params = fminunc(@(p) variogram_error(p, distance_bins, gamma_bins), initial_params,
    options);
```

```
range = fit_params(1);
sill = fit_params(2);
nugget = fit_params(3);

% 输出拟合的半方差函数参数
fprintf('拟合的半方差函数参数:\n');
fprintf('范围 (range): %.4f\n', range);
fprintf('台阶 (sill): %.4f\n', sill);
fprintf('块金 (nugget): %.4f\n', nugget);

% 克里金插值
[zi, var_zi] = kriging(x, y, z, xi, yi, range, sill, nugget);
bianyizhi=[xi,yi,zi,var_zi];
x=bianyizhi(:,1);
y=bianyizhi(:,2);
z=bianyizhi(:,4);
figure;
scatter(x,y,36,z,"filled");
colorbar;
% 定义目标函数，用于最小化
function err = variogram_error(params, distance_bins, gamma_bins)
range = params(1);
sill = params(2);
nugget = params(3);
gamma_model = gaussian_variogram(distance_bins, range, sill, nugget);
err = sum((gamma_bins - gamma_model).^2); % 最小化平方误差
end

function gamma = gaussian_variogram(h, range, sill, nugget)
% 高斯模型的半方差函数
gamma = nugget + sill * (1 - exp(- (h / range).^2));
end
```

附录 C　问题四代码

```
% 定义半径
radius = 28919.7874968846;

% 定义五次多项式函数的系数
p00 =      34.52 ;
p10 =   0.001009 ;
p01 =  0.0004063 ;
p20 = -7.845e-08 ;
p11 = -2.451e-09 ;
```

```
p02 = -7.743e-08 ;
p30 = -3.315e-12 ;
p21 = -1.299e-12 ;
p12 = -1.706e-12 ;
p03 = -1.106e-12 ;
p40 =  4.372e-17 ;
p31 =   7.27e-18 ;
p22 =  9.212e-17 ;
p13 =  2.584e-18 ;
p04 =  4.163e-17 ;
p50 =  2.619e-21 ;
p41 =  1.155e-21 ;
p32 =  3.127e-21 ;
p23 =  1.667e-21 ;
p14 =   4.21e-22 ;
p05 =  7.413e-22 ;

% 定义五次多项式函数
f = @(x, y) p00 + p10*x + p01*y + p20*x.^2 + p11*x.*y + p02*y.^2 ...
+ p30*x.^3 + p21*x.^2.*y + p12*x.*y.^2 + p03*y.^3 ...
+ p40*x.^4 + p31*x.^3.*y + p22*x.^2.*y.^2 + p13*x.*y.^3 ...
+ p04*y.^4 + p50*x.^5 + p41*x.^4.*y + p32*x.^3.*y.^2 ...
+ p23*x.^2.*y.^3 + p14*x.*y.^4 + p05*y.^5;

% 转换为极坐标积分
integrand = @(r, theta) f(r.*cos(theta), r.*sin(theta)) .* r;

% 使用integral2函数计算积分
%total_volume = integral2(@(r, theta) integrand(r, theta), 0, radius, 0, 2*pi);
total_volume = integral2(@(r, theta) integrand(r, theta), 0, radius, 0, 2*pi, 'RelTol',
    1e-6, 'AbsTol', 1e-8);


% 显示结果
disp(['湖水总量: ', num2str(total_volume)]);
```