**基于马尔科夫网格模型的地质属性建模**

【摘要】

【关键词】

# 第1章 绪论

## 1.1研究背景及意义

地质属性建模属于“三维地质建模”这一大的范畴，目前主要被用于矿藏、油藏开采中的储量估计和品位评估。“三维地质建模”是一项综合运用了地质学、信息科学、计算机三维可视化等邻域知识的交叉学科技术，其概念由加拿大学者Simon W. Houlding于1993年首次提出，二十多年来相关的技术和理论发展迅速。其技术进步得益于现代信息技术的高速发展，特别是计算机硬件性能的增强提高了运算速率，扩大了储存容量，让“三维地质建模”得以投入实际应用；再者，科研工作者们辛勤钻研出了一套较为完整的“三维地质建模”方法体系，以此为基础，各个方向的研究得以深入发展，创造出了更多高效可靠、分类细致的地质模型。地质属性建模近年来成为三维地质建模研究领域中的热点，一方面由于并行计算的广泛应用为其技术发展创造了空间，另一方面也与采矿产业步入攻坚阶段的实际情况紧密相连。

一般而言，三维地质模型作为一种可视化工具，能够提供直观的地质情况解读，从而辅助决策者进行分析判断，帮助对地质情况不甚了解的用户高效获取有用信息。对于采矿业应用来说，三维地质模型还用于后续的数值模拟和储量分析。目前，矿产开发的难度日益增大，开采精细化程度加深。其中，属性模型运用得最为广泛的石油开采业正面临非常规油气开采的挑战。面对更高的要求，“三维地质建模”尤其是其中的属性建模方法研究迎来了新一轮发展的契机。

地质属性建模现在仍然处在理论完善的阶段，实际运用比较局限，其中随机建模的方法以地质统计学为核心，近年来相关研究备受关注，产生了很多改进的算法和思路。两点统计的方法已经能够投入生产，但其能够模拟出的属性特征质量较差，模拟效果较好的多点统计法由于较高的内存占用率和CPU需求难以投入实际运用，实用而高效的模拟方法成为人们的研究目标。与之形成对比，“三维地质建模”的另一模块——结构建模在工业生产中的运用已经比较成熟：GOCAD、Petrel等成为科研和生产工作中的主流软件，它们可处理的数据类型广泛，算法效率高，能在适量人工干预的前提下快速建出矢量结构的三维地质模型，保证拓扑关系正确。结构建模的应用现状为属性建模的未来勾画了商业化的发展蓝图。

本文从多点统计学的基本统计学模型——马尔科夫随机场模型出发，实现了一个基于马尔科夫网格模型的有参的多点统计学算法，参数计算方法为属性建模过程赋予了明确的数学意义，是对多点统计学理论和方法的进一步探索。

## 1.2研究现状

地质统计学自从被引入地质研究就成为地质属性建模的核心，多点统计学作为地质统计学中的新方法于1993年被Guadiano F.等人提出。传统的两点统计方法在进行属性估测时，仅考虑了两个属性值对待测值的影响，不能捕捉到复杂的属性分布结构和纹理特征；多点统计学的方法能取得更好的模拟效果，但目前尚未成熟，众多算法都有各自的缺点和局限，运行起来对计算机性能的要求也比较高。

根据多点统计学的统计学原理，多点统计学的算法分为无参算法和有参算法：无参算法通过直接粘贴属性分布模式进行模拟；有参算法需要建立明确的随机模型以估测模拟参数，再根据这些模拟参数计算出得出新的属性分布概率，又称基于概率的模拟。

多点统计学开始被用于实际研究当中是从有参方法Snesim开始的，但Snesim是基于概率估计的算法，存在不能再现非平稳地质现象的问题。后来在数字图像形态学和计算机可视化等研究的基础之上又产生了诸如Simpat，Filtersim等无参模拟算法。这些方法虽在计算效率等方面不断进步，但许多本质问题如训练图像处理、平稳性问题等仍然没有突破性进展。

马尔科夫随机场模型是多点统计学算法所依据的基本统计学模型，其理论体系在计算机图形学中被集中讨论，blahblahblah。。。（栗子）但在地质统计学中的相关研究较少，。。。提出。。。在二维地质属性剖面的模拟中取得了良好效果，但该方法由于计算量过大，难以用在“三维地质建模”当中。2011年，最新的一项研究提出了一种简化的马尔科夫随机场模型——马尔科夫网格模型，用于计算马尔科夫随机模拟的参数，被用在三维地质属性模拟当中。由于计算量大幅度减少，该方法模拟三维地质体所需的内存空间减小，模拟时间缩短，具备了一定的实用价值。

## 1.3研究内容

前人对马尔科夫网格模型的研究取得了良好的三维模拟效果，但建模所需的三维训练图像在实际情况中不易获得，且求取参数的过程采用三维训练图像将产生很大的运算量，对计算机性能要求很高。本文以该研究为基础，首先实现了二维隧道图像的有参模拟，然后探索了如何利用垂直相交的二维隧道剖面分别作为不同维度的训练图像，接着从序贯模拟的角度构建三维地质体模型，最后分别分析和评价了二维和三维的模拟结果。

## 1.4组织结构

# 第2章 地质属性建模概述

## 2.1地质属性建模的一般方法

地质属性建模旨在将地质体内物理、化学参数的分布情况进行定量的可视化表述，其技术手段多采用基于体元的建模方法，又分为基于规则体元的建模和基于不规则体元的建模。规则体元建模中体元由规则分布的节点构成，

数据来源；数据预处理；建模方法（确定性建模和不确定性建模，分别有哪些方法）；栅格建模中不确定性的产生和处理。。。

## 2.2地质属性建模面临的难题

属性建模旨在确定地质属性的空间数据分布，假设我们获得了所有空间几何要素对应的属性值，要将属性值与几何对象对应起来就变得简单了。但在实际情况中，由于资金和工具的限制，可用于建模的硬性数据如钻孔数据等数量十分有限，我们只能获得少量在地质空间中位置确定的属性值，因此需要借助数学工具进行插值或模拟，从而达到从已知推未知的目的。插值属于确定性建模，在已知数据一定的情况下，得到的计算结果是固定的。这虽然满足了人们对确定性的要求，但实际上并不符合真实情况。随机模拟属于不确定性建模，得出的结果反映了数据源的不确定性带来的很多可能性，更加符合客观事实，但却不能满足对精度要求较高的工业应用需求。

不确定性一直是制约地质属性建模发展的关键问题，要降低不确定性有两种思路，一是提升建模资料的数量和质量，二是改进建模方法。在可以获得的地质资料有限的情况下，我们只能将主要精力放在算法的改进上。而且事实发展也证明，算法改进虽然不能从根本上解决不确定性问题，却可以大大提高人们对地质数据的利用效率，这也是我们进行算法研究的主要目标。

## 2.3地质统计学的主要算法介绍

地质统计学是地质属性建模中用于随机模拟的主要理论和方法依据。

# 第3章 马尔科夫网格模型介绍

## 3.1马尔科夫随机场(MRF)

多点地质统计学算法本质上都是随机模拟算法，它们的基本模型被称为“Markov随机场”(Markov Random Field，简称MRF)模型。如果某一空间内任意一个点的值只与其邻域内的点的值有关，而与其他点的值无关，那么这个空间内所有点的值的集合就是一个马尔科夫随机场，用公式表达为：

MRF的原始模型——马尔科夫链描述的是一类随机过程，这类过程在已知目前状态的情况下，未来的演变不依赖于过往的演变。类比到二维空间中，即空间中某种属性值的变化可以被看作该属性值数据事件的随机分布，每个属性值的出现概率可以从其周围属性值的分布情况推断。该属性值与其周围所有对它施加影响的属性值构成了一个集团，我们可以通过集团内的相互关系推测未知点的属性值。

马尔科夫随机场的形状不是一成不变的，具体构成取决于模拟需求。如图所示为最简单的相互影响集团，每个方格所代表的值由其余三个方格共同决定。简单的两两相互影响不能反映目标值复杂的变化行为，于是计算机图形学中又提出了多值相互影响的马尔科夫随机场模型。如图所示，六边形面元扩大了集团包含的值数量，每个目标值至少受到周围六个已经存在的值的影响。

马尔科夫随机场(MRF)模型运用于地质空间模拟或时空模拟时有一个问题，它描述了广义的模型，涉及到很多模糊的参数，难以提供具体的计算方法，且需要估计的参数太多，计算量非常庞大，故而无法投入实际应用。

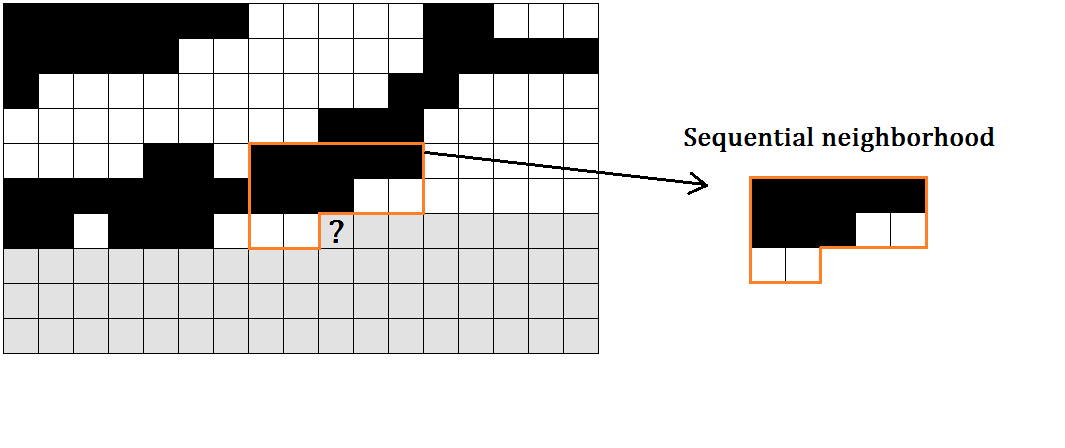
## 3.2马尔科夫网格模型(MMM)

### 3.2.1简化的马尔科夫随机场模型

马尔科夫网格模型(Markov Mesh Model，简称MMM)可以被看作部分有序的马尔科夫随机场模型，它没有使用一般的邻域模型（见公式），而是将二维空间进行网格划分后对所有网格进行排序，然后按照这个顺序进行邻域构建。这样一来，MRF模型中待求参数过多的问题得到了解决，求取参数的取样过程也变得高效了很多。

与马尔科夫随机场模型不同，马尔科夫网格模型的模型设计是不对称的。但正是这样的不对称结构便利了参数估计和样本模拟的过程。马尔科夫网格模型用一系列严格排序的邻域控制模型参数，由规则网格结点构成的栅格路径定义了这些邻域。与MRF模型相同，这样的顺序模拟基于以下假设，即假设待模拟的随机过程具有马尔科夫特性：

可以把马尔科夫网格模型看作一个一维时间系列向二维或三维空间系列的拓展。就像时间系列所表现的那样，一段已经存在的序列，就像“历史”，“预示”待模拟的值。于是，上述模型可以被简化为：



本文中的模型被定义为沿单边路径进行模拟，每个当前单元格的值由其顺序邻域内各单元格的值决定。该顺序邻域是一个已经模拟好的单元格的子集，如图(a)所示。设想一个二维或多维的有限规则网格，用一维下标 标注网格中的所有单元格。这些单元格的集合记作，其中每个单元格的值可以取到个不同的属性值，即。假设下标为 的单元格取到某个属性的条件概率仅仅决定于单元格子集 ，满足，可以写作：

马尔科夫网格模型中所有未知单元格取到某种属性值的概率全部由上式得出，则整个模型满足某种属性分布的联合分布概率为：

我们可以通过广义线性模型创建目标函数并对模型参数进行求解。

### 3.2.2单边路径

马尔科夫网格模型采用单边路径进行模拟，保证了模型的连续性，从而能够直接进行参数估计。然而一旦划定了顺序领域网格的范围，单边路径可能会导致模型的偏斜。不过这种偏斜在参数估计过程中会被部分抵消，偏斜的程度也会随着顺序邻域网格的增大而减小。

单边路径还带来一个问题，就是导致整个模型难以进行条件模拟，因为位于路径末端的数据不能被考虑到在它之前的顺序模拟中。为此，作者K某某提出了引入指示克里金插值的方法，将约束数据的影响考虑在内，更新马尔科夫网格模型模拟得出的概率值。具体方法将在后文给出。

### 3.2.3广义线性模型(GLM)

引入广义线性模型是为了定义马尔科夫网格模型中的条件概率。广义线性模型指的是一串独立变量 的线性估计值与因变量 的预测值 通过一个非线性的连接函数相联，即。连接函数的种类取决于具体的回归模型，本文的解决方案采用的是逻辑回归模型，连接函数为：

与经典线性回归中连续分布的因变量不同，逻辑回归中因变量 为二值分布，即。在本文中，我们引入独立变量，

另外，定义 为由编号为*i*的网格的顺序邻域内所有领域方程组成的向量，。有关邻域方程的构建细节将在第三章继续讨论。对于每个属性值*k* = 0，1，…，*K*-1来说，预测值 可被写作：

都是大小为的向量，分别代表每种属性对应于 中不同领域方程的系数。将以上结果与式联系起来，就得到以下计算条件概率的表达式：

.

据此，式可被表述为：

.

于是，式中的联合概率被表述为：

广义线性模型产生二值变量，要用最大可能性方程求解，应用在本文模型中的具体解法将在第三章模型算法一节中详述。

# 第4章 基于马尔科夫网格模型的三维地质建模方法

## 4.1二维网格的无约束数据模拟

### 4.1.1邻域方程

求取合适的邻域方程是成功用马尔科夫网格模型进行模拟的关键，因为我们用不同的邻域方程组合描述每个参考单元格邻域内的属性分布情况，再通过求解GLM获得相应属性分布模式的模拟系数。对于地质结构来说，属性的连续性和属性值的改变是最重要的特征，我们的模型在求解模拟系数时主要考虑了这两个特征，因此邻域方程也将重点描述这两个特征。在遍历训练图像之前，我们先将参考单元格与其邻域网格的属性关联情况分解为两点内联和多点内联，再为每个参考单元格（当前单元格）的每个内联关系建立其相应的邻域方程。

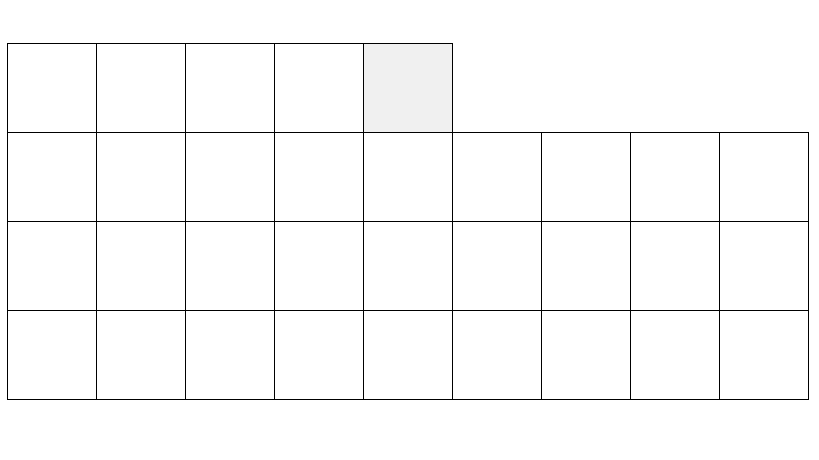
本文的目标是利用二维训练图像构建三维属性模型，模拟系数全部由二维训练图像获得，故仅为二维模型定义邻域方程，*j* = 1, …, *P*。反映了第*j*个参考单元格的邻域网格 内属性值的变化情况：上文提到的两点内联指的是参考单元格与任意一个它的顺序邻域内的单元格的属性值间的关系；多点内联根据给定的参数值分为 个等级， 等级的多点内联指的是参考单元格与其顺序邻域内  *-*1个单元格的值构成的邻域方程间的关系。

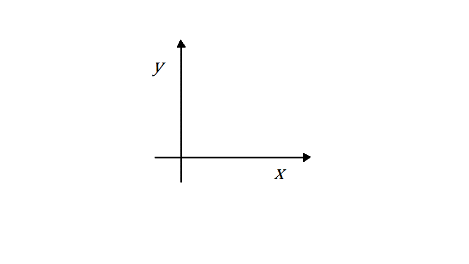
用多点统计学的方法进行有参模拟时，如果要提取出训练图像中属性分布的所有特征，即考虑到所有点与点间的内联关系，将得到很多可能性，导致无法模拟出合理有序的属性分布模式，如Snesim方法模拟出的结果就比较散乱。多点统计学中的无参模拟方法（如Simpat）通过粘贴模式的途径隐式地模拟出了训练图像中的属性分布特征，与之类似，马尔科夫网格模型的方法提取出了一个反映重要属性分布特征的邻域方程的集合。本文中的参数提取方案并非适用于所有训练图像，增减邻域方程可以获得不同的属性分布特征。

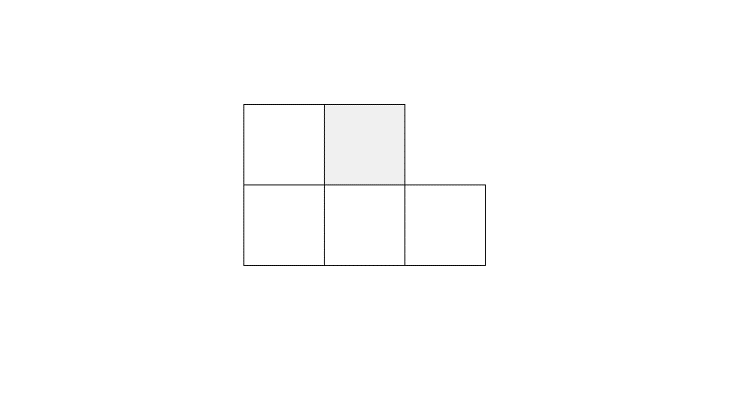
两点内联反映最简单的点与点间的关系，我们的模型包括了顺序邻域内一个子集中所有的两点内联。如图所示， 和 是描述该子集规格大小的参数，模型总共统计了个网格与参考网格间的两点内联关系，我们为其中每个网格针对每个属性值*K*建立一个指示方程，如果下标为*j*的单元格属性值为*K*，，否则。这样就为顺序邻域内每个网格的每种可能的属性值生成了一个参数，一共生成了个参数。

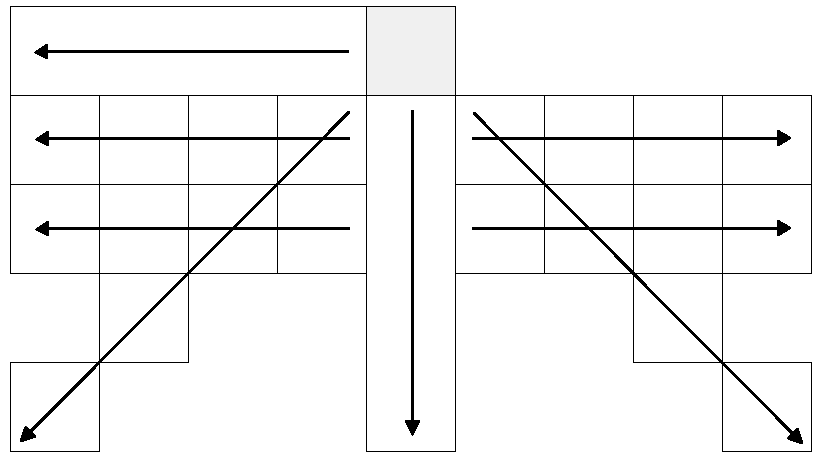
与参考单元格相邻的四个单元格构成网格，如图所示，其属性分布与参考单元格属性值关联性最大，因此所有其内部的属性组合都将体现在邻域方程中，于是又产生了个参数。对于网格形状延伸较长的多点内联，我们关注的是每种属性的连续性和属性值的改变，要将并且仅将属性值组合相同的多点模式考虑在内。在这些模式中，我们选择能够捕获到每个属性对象的形状和延伸状况的单元格集合。同时，我们将选择控制在如图所示的几个方向内。沿着图示的每个方向，我们首先包括了参考单元格与最近两个单元格间的内联关系，即所谓的“三点内联”（Three point interaction）。接着每次增加一个网格，直到达到设定好的最大值 ，如图5所示流程。设 为 点内联项中 个邻域单元格的集合，其指示方程为：

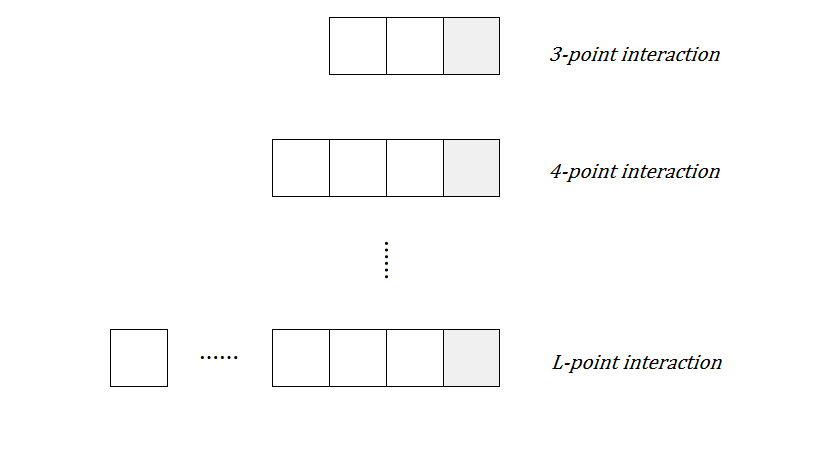
对每个属性的每个方向来说， 每增加1就有一个上述邻域方程被包括在我们的模型中，直到 等于最大值 为止，共产生了 个新的参数。











以上参数化过程的含义十分明确而且容易被理解：两点统计步骤直指点与点间属性值的依赖关系；临近单元格集合内的属性分布指示细微尺度下的特定结构；如果指示某个属性连续性的领域方程对应的参数值偏大，则该领域方程促进连续性的产生，否则其促进属性值改变。上述邻域方程中的一些是冗余的，例如由于考虑了参考单元格临近四个单元格的属性分布情况，没有必要再考虑这些单元格与参考单元格间的两点内联关系。另外还会遇到一些特殊情况，如模拟过程中取到了训练图像中没有出现过的解释变量（邻域方程的值）组合，或者某个参考单元格没有四个相邻的单元格，甚至所有属性在所有多点内联关系中都不连续。为了解决这些问题，模拟之前需要进行参数简化。

### 4.1.2最大主成分分析(PCA)

我们通过提取解释变量主成分的方法简化模拟参数，相当于抽取了属性分布模式的主要特征，这样不仅解决了3.1.1中提到的问题，也避免了求参过程中需要对奇异矩阵求逆的情况。简单来说，就是要用PCA对解释变量矩阵 进行降维处理。

PCA的实现一般有两种：特征值分解和奇异值分解，特征值分解只适用于方阵，而奇异值分解能用于任意矩阵，本文用它来提取解释变量矩阵的特征。本文用SVD方法将解释变量矩阵分解如下：

，

设每个属性的待测模拟系数总个数为 ，简化后的解释变量矩阵为 ，的大小是，，并且，

，

式中 是大小为的矩阵，由 个奇异向量构成，对应矩阵 最大的 个奇异值。 中的每个行向量都是 中行向量的线性组合。

集合是与 相对应的简化后的模拟系数集，它满足以下关系：

.

因此，系数集可以用 进行估测，真实的系数可通过下面式子计算求得：

.

### 4.1.3算法流程

标准的广义线性模型假设一个地质空间数据分布中有N个独立的模型观察值，则该空间分布的总体可能性方程为：

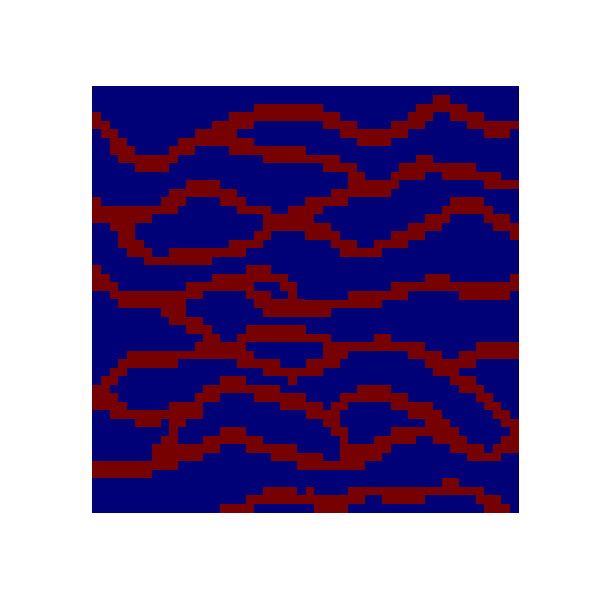
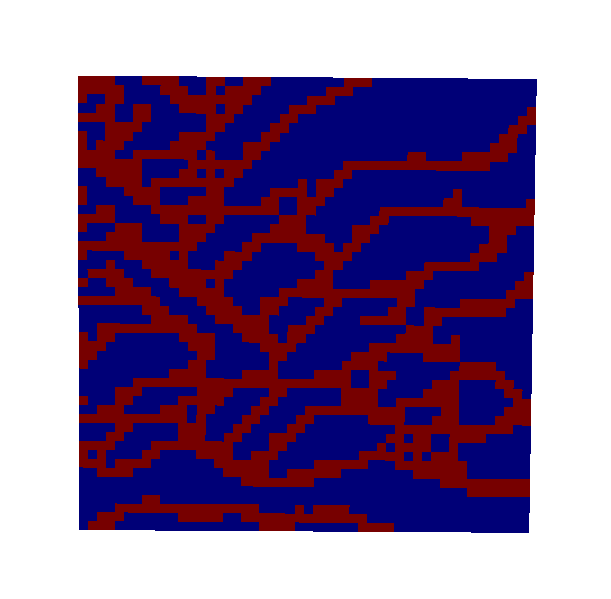
通过求解最大可能性方程得到以下相等关系：

各个变量用矩阵表示如下：

式通过梯度查询的方法求解，得到迭代加权最小二乘方中的每一次迭代公式如下：

上式中是一个大小为的权值的矩阵，其表示如下：

### 4.1.4模拟结果



训练图像： 模拟结果：

分析：从隧道连续性、模式重复性和volume fraction等方面作分析。（图像规格：50\*50，参数设置：Lx=Ly=5，L=8）

要还做的工作：换用更高精度的训练图像；用隧道形状不同的训练图像做模拟。

### 4.1.5优化模拟

（如何运用Volume fraction steering的方法进行优化，优化结果分析）

## 4.2二维网格的有约束数据模拟

### 4.2.1指示克里金插值

### 4.2.2算法流程

### 4.2.3模拟结果

## 4.3基于二维属性剖面的三维属性建模

### 4.3.1训练图像

### 4.3.2序贯模拟

### 4.3.3算法流程

### 4.3.4模拟结果

# 第5章 结论与展望

### 5.1论文总结

### 5. 2研究展望

# 参考文献:

[1]

[2]

# 致谢