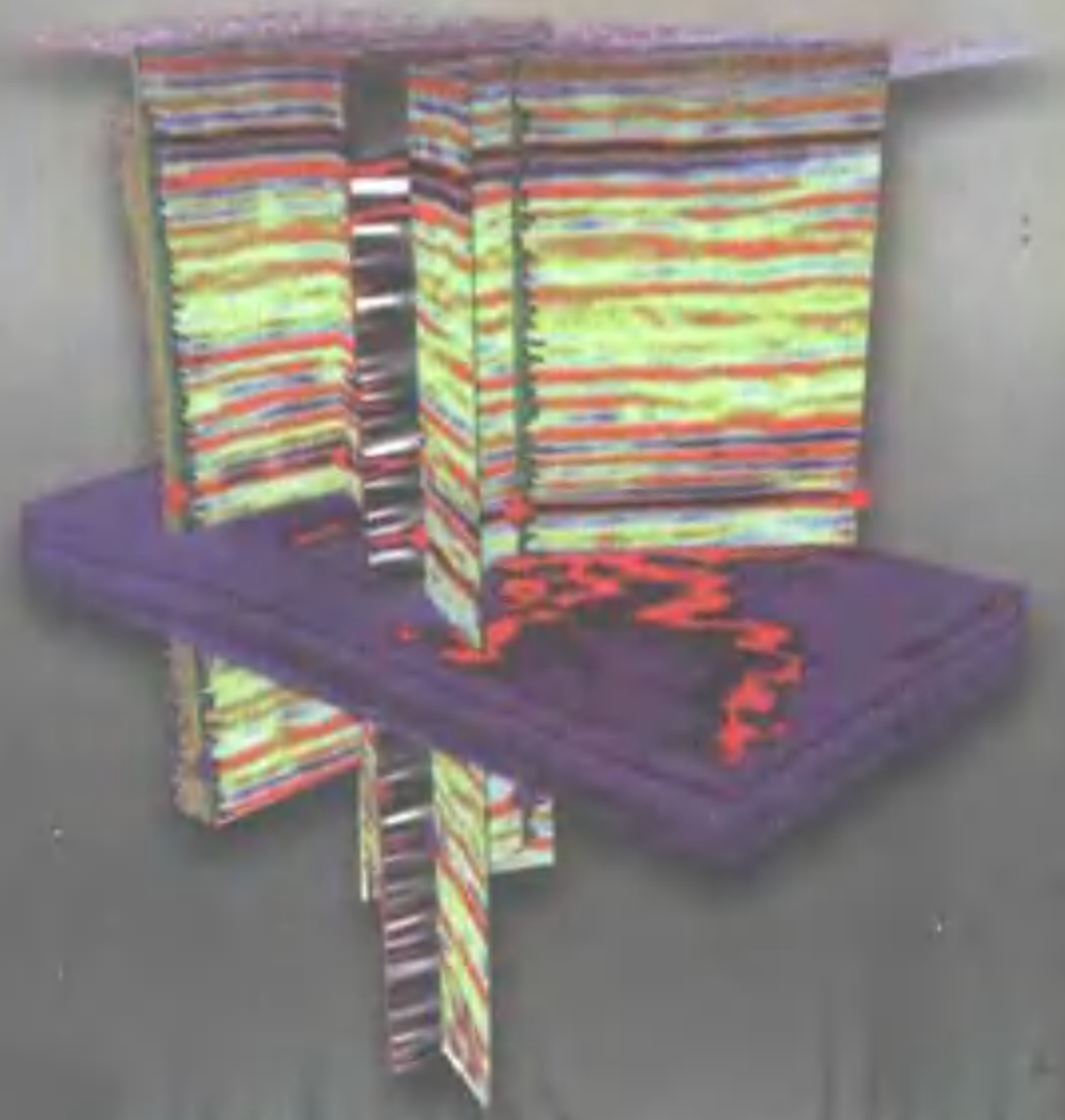
STOCHASTIC RESERVOIR MODELING



王家华　张团峰　著

油气储层随机建模

北京)
2,2

石油工业出版社
PETROLEUM INDUSTRY PRESS

油气储层随机建模

王家华　张团峰　著

石油工业出版社

图书在版编目（CIP）数据

油气储层随机建模/王家华，张团峰著．
北京：石油工业出版社，2001.4
ISBN 7-5021-3319-4

Ⅰ.油…
Ⅱ.①王…②张…
Ⅲ.油气聚集-储集层-建立模型
Ⅳ.P618.130.2

中国版本图书馆 CIP 数据核字（2001）第 14086 号

石油工业出版社出版
（100011 北京安定门外安华里二区一号楼）
石油工业出版社印刷厂排版印刷
新华书店北京发行所发行
*
787×1092 毫米 16 开本 10.75 印张 266 千字 印 1—1000
2001 年 4 月北京第 1 版　2001 年 4 月北京第 1 次印刷
ISBN 7-5021-3319-4/TE·2488
定价：35.00 元

序　　言

由国人自己编写的《油气储层随机建模》专著即将问世了。这是西安石油学院王家华教授和他所领导的研究小组，十多年来在油藏描述这一新型领域中辛勤耕耘的结晶，为我国储层地质学科填补了一个非常重要的空白。

油气储层随机建模是20世纪80年代后期刚刚萌芽兴起的一项油藏描述高新技术。它是为适应油气田开发的深入，应用先进的二次采油和三次采油技术进一步提高油气采收率的需求，应运而生的。这就为储层地质工作者提供了一个既是定量化、精细化的，又能实事求是地反映“不确定性”的储层地质模型。随机建模技术既尊重油气储层固有的地质规律，反映某些客观存在的随机性，又能定量地描述由于资料信息的不足和人们认识的局限给储层地质模型带来的不确定性。它为油气田开发战略决策中的风险分析提供了一个重要依据。这正是这一技术刚刚出现就受到石油工业界极大重视的原因所在。

油气储层随机建模技术，是储层地质工作本身高新技术化的体现。储层地质学作为石油地质学的一支，也受到地质传统的一些局限。定性和半定量表征是它常用的手段。如果说近代测井、地震、油藏数值模拟、工程测试等高新技术的发展，为储层建模提供了大量的、精确的定量化的数据，为储层地质定量化创造了很好的环境条件；那么随机建模技术的诞生，可以说是地质学与数学、计算机结合，本身向着高新技术发展的结果。正如王家华教授他们在这专著中阐明的，随机建模是以地质统计学为基础，模拟过程的两步建模程序又是以依据储层沉积学理论建立储层沉积相估价模型为第一步，以此来控制物性等其它参数的空间分布。应该说，这是地质学科发展的一个重要方向。

随机建模技术在国内的应用，应该承认最初是引进学习的。王家华教授他们在这一新技术刚刚出现之际，就敏锐地捕捉到它的潜在前景，成立了我国第一个专题研究小组，跟踪学习消化，请进来，派出去，自己动手研制软件，到各油田去宣传实践。他们除承担了中国石油天然气总公司的科研攻关任务外，还在国内一些处于高含水期的主力大油田，如大庆、胜利等，开展了多个实例研究。十多年的孜孜攻研，终于结出了丰硕的成果。他们不仅结合油田生产实际，为开展油藏精细描述作出了贡献，更重要的是，发展了随机建模新的算法，提出了适合我国广为分布的河道砂体储层的随机游走建模方法；编制了我国自己版权的第一个随机建模的软件系统“GASOR”，并不断完善升版。应该说，他们在国内推动和发展这一新技术上，功不可没。

《油气储层随机建模》这一著作的出版，不仅集中反映了他们在这一领域的成就，还将为进一步在国内推广这一新技术作出重要贡献。在该书中，他们介

绍了基本原理和基础理论，详细描述了目前通用流行的算法，以及如何应用于储层地质工作，实例丰富，深入浅出，具体详实。为我国从事油气储层地质的广大科研和现场实际工作的技术人员，为即将投身于这一事业的在校学子们，为从事油气田开发工程的其它专业的科技人员，提供了一个新的学习园地、一个学习应用参考随机建模的有力工具。我作为储层地质工作的一个老兵，非常欣慰，并坚定地相信，这本专著的传播，必将引发更多的有志者，象王家华教授他们那样，踊跃参加到这一科研领域中来，为共同发展适合我国陆相储层特色的储层建模技术，为壮大我国石油工业作出更大的贡献！

裘怿楠

于 2000 年 8 月盛夏

前　言

一个16岁的少年，还有两年的时间才到社会公认的18岁成年。自1984年Lake和Haldorsen（1984）发表的一篇论文开始，仅过去了16年，油气储层随机建模技术已经得到了世界各国众多石油地质学家、油藏工程师、地质统计学家和地球物理学家的广泛承认。在短短的十多年间，由于他们和计算机软件专家共同努力，把储层随机建模作为石油地质统计学的一个应用领域，有力地促进了油藏描述技术进一步向定量化方向发展。储层随机建模已经发展成为系统化的 -个方法、一门技术，成为油藏描述技术的有机组成部分，为促进建立三维定量地质模型，改善油藏管理，提高油气采收率和油气产量作出了重要贡献。

目前，从事储层随机建模的人员专业背景相当广泛，从从事数学地质的，应用数学的和概率统计的，到从事石油地质，油藏工程的和计算机软件的，也有从事地震，测井的。人们对储层随机建模技术寄予厚望，试图运用这门技术解决沉积相带和渗透率、孔隙度的空间分布，以及断层、裂缝等的空间分布。同时，人们还在试图运用地震资料进行储层随机建模，并已获得不同程度的成功。进而，人们希望在预测剩余油的分布和加密井的位置，进行非均质性分析，计算油气储量，改善油藏数值模拟的效果，进而提高油气田采收率方面取得经济效益。

在20世纪80年代的最后一年，随着我们从事地质统计学中的克里金估计在油藏描述绘图中应用不断取得成果，我们承担的“七五”国家重点科技攻关项目“山东牛庄油田油藏描述技术研究”的成果初见端倪。这时，我们开始了随机模拟方面的探索。.

最初，我们是从A.Journel教授的著作“矿业地质统计学”中了解到随机模拟（Stochatic Simulation）的基本概念和基本算法。在那本书中，矿体的随机模拟的算法是用“旋转带法”来完成的。随后，我们对沉积相带空间分布的随机模拟，以及沉积相带空间分布控制下的物性参数空间分布随机模拟进行了长达十年之久的深入研究。沉积相带空间的随机模拟是整个储层随机建模技术中的一个核心问题，也是最为困难的。我们的研究内容包括方法，算法，软件系统和油田实例研究，涉及的内容有地质统计学，随机过程理论，随机几何学，油藏描述技术，油藏工程，储层沉积学，软件工程学和计算机图形学等。

储层随机建模技术和其他技术相比，在信息的获取，加工和转换方面有着明显的优点。

1）以储层沉积学作为理论基础，用沉积相带的空间分布控制着物性参数的空间分布。

2）为把地质，地震，测井和油藏工程等各种不同信息的结合提供了切实可行的途径。

3）利用随机性的概念，较合理处理了测井数据和岩心数据比较稀少造成的不确定性。

4）较好地解决了各种数据分辨率不同给数据和信息的结合带来的矛盾。

在过去的十年，我们承担了储层随机建模方面的十多个科研项目。其中，有“八五”、“九五”国家重点攻关项目的有关任务，有中国石油天然气集团公司的项目，以及各油田的横向合同项目。由于各级领导的支持和关怀，各油田现场的密切配合，这些项目都取得了圆满的成功，先后获1996年中国石油天然气总公司（部级）科技进步二等奖一项（“胜利孤东油田馆陶组河流相三维定量储层地质模型技术研究”），局级科技进步奖四项。在这过程中，我们对储层随机建模技术的方法特点、算法设计、结果分析、软件编写诸方面的理解、掌握

逐步深化。这里，经历了一个以读文献，照搬国外现成的方法到根据国内油田的现场需要，发展具有中国特色的储层随机建模方法及其软件系统的过程。随机游走建模方法的提出和实施，和“储层地质统计分析系统 GASOR”实现，升级及推广应用就是最好的证据。通过这些项目的实施，我们加强了为各油田服务，以及和各油田的沟通，改善了自我的研究条件和手段。我们深深地扎根于国内油田开发这片沃土上，储层随机建模的一朵朵鲜花正在盛开。

计算机软件系统的设计、开发和使用是储层随机建模的一个重要部分，作者多年来在这方面做了大量的工作。把三维储层随机建模的结果数值地计算出来，计算量相当大，对当代功能最强大的工作站系统来说都是一个相当不轻的负担。其数据量巨大，对它们的管理又要求相当灵活、方便。此外，储层随机建模系统的各种各样的输入参数十分繁多。因此，系统的设计在软件工程方面的考虑，要求十分之高。将储层随机建模的结果直观灵活地显示绘制出来并非容易。按地质学家不同的要求，把一个三维的储层地质体的整体趋势和局部细节，表面和实体内部的特征同时清晰地表达出来，往往是比较困难的。人们正在利用计算机图形学和虚拟现实技术的最新发展成果，来解决储层随机建模结果用图件的显示问题。因此，储层随机建模系统是一个十分庞大而复杂的软件系统。限于篇幅，我们在本书中不作专门的表述。

我们在研究储层随机建模的过程中，化去了相当大的精力发展了一个储层随机建模软件系统，并将它定名为“储层地质统计分析系统 GASOR”。GASOR 经历了从最初的微机版本，到目前的 SUN 工作站版本。本书的实例研究的图幅都是利用 GASOR 得到的。只有开发出了我们自己版权的软件系统，并不断完善，我们才能为油田现场服务，才能不断根据现场需要提出问题发展新的适合国情的储层方法，才能在国内外发表理论和实践相结合的研究论文，总之，才能使整个研究走上良性循环的正确道路。通过了十余年研究，我们初步体会到所走出的是一条符合中国国情的，具有我们学术优势的发展油气储层随机建模技术的道路。

我们在为油田现场服务的过程中，逐步懂得了如何运用储层随机建模的先进方法为现场服务。我们千方百计地尽量满足油田各方面的要求，从随机建模方法的选取，到软件系统模块设置，一直到数据预处理的程序的编制，从研究人员的配置到具体的时间安排。

在过去的十年中，我们通过各渠道加强了国际交流，钻研国际上随机建模方面的各种成果和动向，始终使研究水平跟上国际潮流的发展。我们在这里要衷心感谢美国斯坦福大学（Stanford）的 A.Journel 教授，法国地质统计学中心（Centre de Geostatistique）的 M.Armstrong 和 A.Galli 两位博士，挪威国家石油公司（Statoil）的 A.Henriquez，A.MacDonald，A.Nass 等博士，挪威科技大学的 H.Omre 教授，挪威计算中心（Norweigian Computing Centre）L.Holden 博士，还有加拿大地质建模公司（GeoModeling）的中国旅加学者文仁君博士。他们之中，有的是应邀来华讲过学，有的是邀请我们作访问学者去访问。

作为一本专门叙述储层随机建模技术原理的著作，我们力求以比较严谨和比较正宗的语言把其中的数学原理叙述清楚，且用我们在国内油田完成科研项目的例子进行说明。同时，我们又注意运用通俗的观点，使得读者对本书的内容有一个全面的理解。

本书以指示模型、截断高斯模型、示性点过程模型（第 5、6、8 章）等国际上流行的储层随机建模方法为核心。这些方法主要是来解决确定沉积相带的空间分布。作为对储层随机建模基本方法的一个发展，我们在第 7 章中，首次提出了对于密井网的条件下，利用随机游走模型来确定水下辫状分流河道的流向和位置。这种河道狭窄，分叉会合频繁，给确定沉积

相带的空间分布带来了一定的困难。在国内外的文献未见过类似的报道。我们相信，随着应用实践的不断深入，随机游走模型在储层随机建模中的应用有着广阔的前途。本书也讨论了渗透率粗化的问题（第9章）。它是油藏描述和油藏数值模拟之间的桥梁。作为储层随机建模技术的基础，“地质变量的空间变异性分析”问题和“随机模拟与地质统计学的基本原理”问题，分别在第2章、第3章进行了详细地讨论。为了宣传储层随机建模，本书首先在第2章讨论了“储层随机建模的地质应用”。第1章“绪论”则讨论了油藏描述技术的产生和发展，随机模拟理论，石油地质统计学的形成和储层随机建模技术的特点等问题。

在第5、7、8、9等章中详细讨论了有关的实例应用研究。在这些章节中，展示了用我们开发的“储层地质统计分析系统 GASOR”制作的各种二维和三维的储层随机建模成果图。这些图件的大部分是第一次公开发表。

本书叙述的重点是油气储层随机建模在模拟储层沉积相的空间分布方面的应用。这是储层随机建模中一个极为重要的方向。它在储层随机建模发展过程中最早被提出的，并得到充分研究的，最能体现储层建模特点。

我们在从事储层随机建模技术工作的过程中，积累和阅读了大量的国内外文献。本书的有关章节引用了它们，以便读者可以进一步参考。这些文献大多是研究基本原理和基本方法的，和每年发表的大量储层随机建模方面的文献相比，仅是相当少的一部分。

本书为了提高储层随机建模结果图件的显示效果，将其中的一些图件用彩色图幅来表示，并且集中附在本书的最后。

我们都是数学专业的背景。我们在长期的实践中体会和理解了油田地质学的基本内容和基本过程，本书无疑为广大读者提供有关储层随机建模原理的一个概貌。然而从如何把更多的地质观点注入储层随机建模的原理方面来看，本书的内容选择和编排还存在着一定的不足。

王家华教授撰写了绪论、第8，9章，还撰写了第1，7章中的部分内容。张团峰副教授撰写了第1，2，3，4，5，6，7章的有关内容。最后，王家华教授完成了全书的统稿工作。

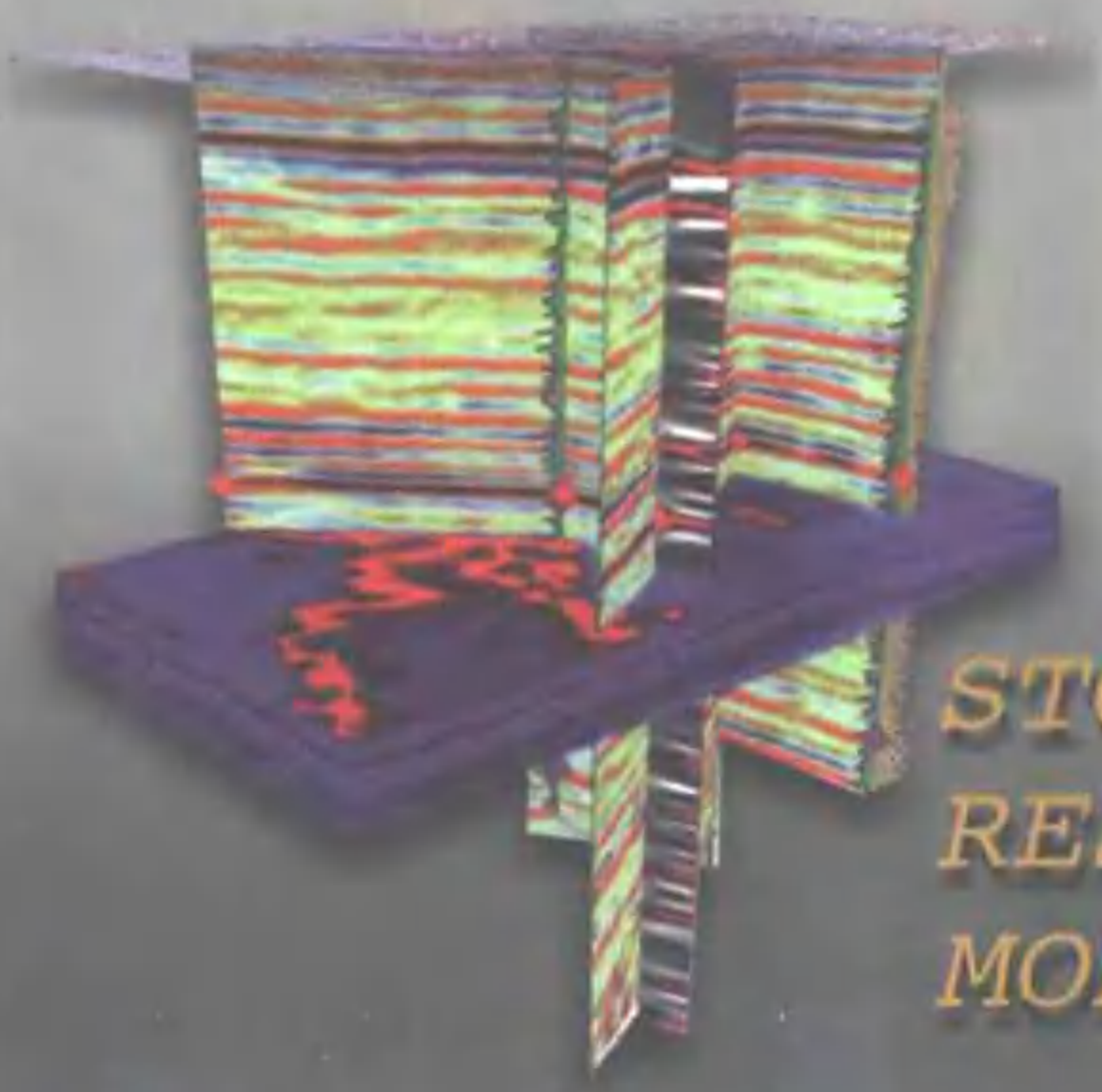
我们在这里要特别感谢韩家新、任长林、黄沧佃等年轻教师。没有他们的创造性工作，就没有“储层地质统计系统 GASOR”，也就没有储层随机建模的今天。我们十分感谢原野、何聚厚、王红霞、周崇礼、刘崇涛等硕士研究生。他们为计算机软件“储层地质统计系统 GASOR”的开发、应用和推广做了许多工作。

在我们多年从事储层随机建模的过程中，曾得到过许多领导、专家的帮助、支持和指导。他们中间有：赵鹏大、裘怿楠、周成勋、刘宝和、阎存章、刘振武、巢华庆、张一伟、熊琦华、潘兴国、岳登台、罗志斌、薛叔浩、陈丽华、贾承造、陈永武、刘昌玉、章术英、隋军、赵翰卿、赵永胜、赵世发、吕晓光、李伯虎、赵国忠、甄鹏、张明禄、张明礼、阎汉杰、刘昌玉、车翔飞、马宏飞、郑容植、李焕鹏、夏季庄、隋志强、景平、夏键、李宁、卢涛、李元觉、郑强等。我们谨向他们表示衷心的感谢。

我们特别要衷心感谢北京石油勘探开发科学研究院的裘怿楠教授为本书撰写了序言。裘怿楠教授是我国杰出的油田开发地质学专家。他重视和倡导储层随机建模在中国的应用，对我们所从事的工作给予一贯的支持和指导。北京科技大学的侯景儒教授、中国地质大学的王仁铎教授分别对本书的有关章节进行仔细的审阅，我们衷心地感谢他们。这两位教授做了许多开创性的工作，对地质统计学理论在中国的推广和应用作出了杰出的贡献。我们还要感谢我院的陈军斌博士对本书的第9章进行仔细的审阅。

本书的主要读者是从事油气田储层地质学、油藏描述和油藏工程的生产、科研人员以及研究生、本科生，同时也面向从事测井和地震的专业解释人员。本书假定读者已经对期望和方差等概率论的基本概念具有一定的了解，最好是对地质统计学和克里金估计的原理有一定的了解。

王家华　张团峰
2000 年 4 月
于西安石油学院　计算机系
710065 西安市电子二路东段 18 号
Tel.：(029) 8382714
E－Mail：jhw@xapi.edu.cn



STOCHASTIC RESERVOIR MODELING

责任编辑：郎 杰　封面设计：赛维玉　责任校对：王 颜

石油大
P618
2

ISBN 7-5021-3319-4

9 787502 133191 >

ISBN 7-5021-3319-4/TE · 2488

定价：35.00 元

目　　录

绪论……（1）
0.1 油藏描述技术的产生和发展……（1）
0.2 随机模拟理论……（5）
0.3 石油地质统计学的形成……（5）
0.4 储层随机建模技术的特点……（7）
第 1 章　储层随机建模的地质应用……（12）
1.1　非均质性的评价和定量化……（14）
1.2　储层静态行为的研究……（14）
1.3　运用地震数据进行建模……（18）
1.4　两阶段随机建模……（21）
1.5　用于油气产量预报……（21）
1.6　应用于早期的评价到后期开发的整个过程……（23）
第 2 章　地质变量的空间变异性分析……（26）
2.1　随机函数，协方差函数和变异函数……（26）
2.2　理论变异函数模型……（30）
2.3　各向异性的变异函数模型及其套合……（31）
2.4　实验变异函数的计算及拟合……（33）
2.5　实验变异函数的拟合……（35）
2.6　结构分析的一般步骤……（37）
第 3 章　随机模拟与地质统计学的基本原理……（38）
3.1　地质变量的分类……（38）
3.2　地质统计学基本工具……（40）
3.3　估计与模拟……（43）
3.4　抽样基本理论……（47）
3.5　随机模拟基本原理……（48）
第 4 章　高斯场模型……（52）
4.1　基本理论……（52）
4.2　高斯场模型的检验……（55）
4.3　模拟方法……（57）
4.4　随机分数维场及其模拟方法……（64）
第 5 章　指示模型……（68）
5.1　单变量指示模型……（68）
5.2　协同指示模型……（72）
5.3　软硬信息相结合的指示模型……（74）
5.4　后验概率的次序关系校正及内插和外推……（78）

5.5 DQ 油田的实例研究 …… (81)
第 6 章 截断高斯模型 …… (86)
6.1 比例曲线 …… (86)
6.2 相序规律与高斯函数的截断 …… (87)
6.3 截断高斯模型 …… (88)
第 7 章 随机游走模型 …… (93)
7.1 区域网格化 …… (94)
7.2 河道源头 …… (94)
7.3 河道主流线位置 …… (95)
7.4 分支河道位置 …… (95)
7.5 河道主流线位置的修正 …… (96)
7.6 河道加宽与光滑 …… (96)
7.7 实例研究 …… (97)
第 8 章 示性点过程模型 …… (98)
8.1 点过程和示性点过程 …… (98)
8.2 模拟退火算法 …… (102)
8.3 示性点过程建模的原理和算法 …… (104)
8.4 先验模型和后验模型 …… (110)
8.5 应用实例 …… (113)
第 9 章 渗透率粗化 …… (119)
9.1 引言 …… (119)
9.2 基本理论 …… (121)
9.3 粗化方法 …… (124)
9.4 实例研究和方法比较 …… (135)
参考文献 …… (144)
后记 …… (157)
彩色图幅

绪　论

本章研究了油气储层随机建模技术产生的背景。油气生产和油藏描述技术为油气储层随机建模技术的发展提供需求，随机模拟方法的发展为它奠定了理论基础，而储层沉积学、测井地震处理解释技术和计算机软硬件的发展则为它创造了有利的条件。接着，本章简单叙述了油气储层随机建模技术产生和发展的过程。最后，讨论油气储层随机建模技术的目的和特点。

0.1　油藏描述技术的产生和发展

我国著名的开发地质学家裘怿楠教授在油藏描述、储层三维定量地质模型的建立和储层沉积学等方面做过许多工作，有过精辟的论述（裘怿楠，1982，1990a，1990b，1991，1993，1994，1996a，1996b，1997a，1997b，1997c，1997d）。

0.1.1　油气生产的需要

在主要产油国，随着勘探领域转向了自然经济和地质条件比较复杂的边远地区，促使成本上升。而同时，已投入开发的含油气盆地的勘探开发成熟度很高。已开发的老油田在现有技术条件下预计采收率仅35%左右，大有潜力可挖。挖掘这部分潜力，引起人们更大的重视。且不说采用化学剂的三次采油，仅依靠现有的二次采油技术（注水），通过深入认识储层的非均质性，依靠各种改善了的采油技术，还有19%的储量潜力可供挖掘动用。

在我国石油工业的发展中，为了稳定东部油田油气产量，近一步挖掘已开发的主力油田的潜力，提高其采收率，具有重要的意义。由于采用了注水开发，这些主力油田目前已进入高含水期。搞清目前高含水条件下油藏内剩余油的分布状况，是正确采取挖潜措施的前提。油藏精细描述正是为适应这一要求发展起来的。

油田进入高含水期以后，油层内剩余油分布已高度分散。均匀地成批钻加密井一般已不具备条件，必须优选剩余油富集区有针对性地采取措施。我国一些主力油田都已进入这一阶段。

近年来我国油田开发正面临着从易开发区向难开发区和边远地区、从部分油藏的高含水、高采出程度，到储采基本平衡向严重不平衡过渡的严峻形势。再考虑到中国陆相储层的复杂性，这些都需要对储层进行更加全面深入的研究。目前地下油水分布情况发生了巨大的变化，呈现出高度分散、局部集中，剩余油多分布在差、薄、边部位，开采难度增大等特点。这些主要是由于储层的各种非均质性及复杂的构造断层切割所造成的。因此，为了搞清高含水期老油田地下剩余油分布规律，进一步提高滚动勘探开发工作水平，需要更加精细的储层描述，以预测井间砂体及各种油藏参数的分布规律，建立精细的三维定量预测地质模型（穆龙新等，2000）。

油藏精细描述的目的就是为了建立一个可供研究剩余油饱和度分布的油藏地质模型。这涉及如何估计井间未钻井地区的储层变化和各种参数值。这是进行油藏精细描述遇到的一个技术难点。油田开发的特殊性决定了这一问题的重要性。只能通过井孔了解油藏，而井间

（一般数百米井距）油藏体积比井孔体积大千百万倍。这部分体积的油藏参数要通过少量井孔资料来推测估计。为了估计值接近于合理和实际，目前出现了两大技术，即以储层地震技术和地质统计学为基础的随机建模技术。

0.1.2　储层沉积学的发展

储层沉积学（Reservoir Sedimentology）的任务是应用沉积学理论和相分析方法，描述各种环境下沉积的油气储层特征（Reservoir Description），是沉积学的一个应用分支。它兴起于60年代，70年代后期比较趋于成熟，在勘探和开发实践中得到比较成功的应用。促使其发展的生产动力是海上油田的大量勘探和开发和改善二次、三次采油提高采收率的需要。海上勘探和开发石油耗资大，发现油气田后希望以较少的探井资料，能对油气藏作出比较正确的估计，以正确设计开发井位和钻井平台，在开发中得到较好的经济技术效果。这就要求对油气藏、特别是储层特征作出比较正确的预测和描述。二次采油注水技术的发展，不断地揭露控制和影响水驱油效果的地质特征，使人们认识到假如开发早期能正确描述出这些储层特征，二次采油的效率还可以大为提高。三次采油方法的大量试验，更进一步说明储层的一些宏观和微观的特性，是筛选三次采油方法的关键；一些三次采油方案的失败，往往是由于没有正确认识某一方面储层特性。出于这些生产实际的需要，应用沉积学理论和方法解决石油开发中的储层描述问题，就应运而生了。

储层沉积学在石油生产实践中，所以有如此旺盛的生命力，仍离不开沉积学的近代进展。关键在于：一是现代沉积调查的大量积累，建立了各种沉积概念模式；二是以水槽试验为主的基础理论研究，加深认识了各种沉积现象的形成机理。

石油生产对储层描述的要求日益从宏观向微观方向发展；储层沉积学工作者力图由定性发展到定量地描述储层。储层沉积学的出现为油藏描述的诞生创造了条件，并为它的进一步发展提供了理论依据。

0.1.3　数字地震、数字测井的出现

地震资料的数字记录和处理，测井资料的数字记录和处理，使得这些资料处理和解释的质量提高了许多。计算机系统的不断提高，计算机地质绘图的功能逐渐完善。这些新技术，有利于对油藏的认识，促进油藏描述技术的诞生。

0.1.4　油藏描述和储层研究

油田开发工作包括两大部分：认识油藏和改造油藏。它们是贯穿于油田开发全过程的两个永恒的课题。它们的最终目的是取得最大的经济效应。认识油藏是改造油藏的基础。油藏地下情况认识清楚了，油田开发的成功则取得了一大半。在认清油藏地下情况基础上，才能确定开发措施，优化开发方法，才能以最少的人力财力投人，获得最大的油气采收率和经济效益。所以，油藏描述作为油田开发的一项基础工作，一直受到极大的重视。随着石油工业的发展，油田开发的深入，油藏描述一直在不断地发展提高。

80年代以来世界上油藏描述的动向，可以用“精细油藏描述”来形象化的概括。“精”就是要提高定量化和精确度；“细”是指描述内容和尺寸越来越细，也就是分辨率要求越来越高。西方在术语上也有所改变，近年来逐渐以“Reservoir Characterization”（油藏表征）来代替原来的“Reservoir Description”（油藏描述），其内涵就是反映了油藏描述这一动向。再具体地说，油藏描述的发展动向可归纳为，宏观向微观方向发展；定性向定量方向发展；单一学科向多学科综合方向发展。

以上诸因素，终于使人们提出了油藏描述的概念，使人们能综合地应用地质、地震和测

井等资料，以前所未有的程度定量地描述所关心的油藏。油藏描述的概念被广大的地质学家、地球物理学家和计算机工程师所接受，对于石油的勘探和开发起很大的推动作用。同时，油藏描述本身从80年代初期以来，在90年代有了很大的发展。这主要表现在以下三个方面。

1）对已有的数据稀少和存在着误差等原因引起的地下数据的不确定性有了进一步的认识，提出用随机的方法解决油藏描述的问题，产生了储层随机建模技术乃至石油地质统计学等一系列的方法。

2）对储层非均质性进行了全面的认识，提出了针对储层非均质性进行建模的概念。

3）提出了油藏描述的核心是建立储层三维定量地质模型。

从1986年开始，“油藏描述技术研究”作为“七五”国家重点科技攻关项目正式立项。当时的石油工业部组织了有关单位，在胜利、中原和江汉等三个油田开展了油藏描述的研究工作。随着研究项目成果的不断取得，油藏描述开始在油气田开发中受到越来越大的重视，在生产中发挥了越来越大的作用。国内，一大批有关的专著和论文相继出版发表，代表了我国油藏描述和储层研究的发展水平。

中国石油天然气总公司在“七五”、“八五”期间开展了“全国油气储层评价”研究。作为该项目的成果之一，“油气储层研究技术”（陈丽华等，2000）总结了储层定量评价和计算机模拟技术，包括储层随机建模方法、湖盆三角洲沉积过程的数值模拟、储层裂缝定量模拟、油气储层综合评价等，还总结了储层地球化学、包裹体测试、激光显微取样、核磁共振等新的实验技术方法。

“大庆油田河流—三角洲储层研究”（隋军等，2000）系统地总结了大庆油田从60年代直至90年代储层研究的成果。他们的研究经历了四个阶段：60年代至70年代早期，为开创小层对比阶段；70年代中期至80年代中期，为动静结合，从成因上认识储层阶段；80年代中、后期，为全面认识储层阶段；90年代以来，为密井网条件下的储层精细地质研究阶段。该书分为三篇。第一篇概述了大庆油田储层研究工作，重点介绍了大庆长垣大型河流—三角洲沉积物的沉积环境、三角洲模式和砂体成因类型。第二篇主要介绍了密井网条件下储层精细描述的方法和技术，包括应用沉积学手段对河道砂、河间砂、三角洲前缘砂精细描述方法和实例，也包括地质统计学等定量表征方法对储层井间参数预测的探讨。第三篇着重介绍储层研究在油田开发中的应用。

十余年来，以张一伟教授、熊琦华教授为代表的石油大学（北京）油藏描述与预测研究所，在生产实践中不断探索、创新，逐步发展了一套适用于中国地质特点的油藏描述的理论、技术和方法（张一伟，1992，1997；吴胜和等，1998）。他们撰写并出版的《中国典型油气藏描述与预测丛书》包括了“现代油藏描述技术”等十二个分册（王志章等，1999）。该套著作比较系统地阐述了我国陆相油田不同勘探开发阶段的、不同油藏类型的动、静态特征，及其地质模型．提出了关于陆相油藏、复杂油藏的储层岩石物理相、相控储层对比、相控储层预测、视标准层构成和视储集空间构成的多方面的新理论、新观点和新概念，发展完善了储层地震预测、多参数识别油气水层、剩余油形成机理及预测、裂缝性储层评价、碳酸盐岩储层表征的方面的配套技术和方法。他们在地质统计学在油藏描述中的应用、储层随机建模应用等方面做了不少工作（张一伟等，1992；吴胜和等，1998；王志章等，1999）。吴胜和等（1999）讨论了储层建模的目的和意义、储层模型类型、储层建模流程及建模策略、对各种建模方法进行了叙述和对比。

张金亮在早期的油藏描述研究方面，在已有的少量井的古生物、古地磁及孢粉研究的基础上，结合波形相似性研究及地震地层学研究，首次解决了研究区争论已久的地层划分对比问题（1993）。他们在剩余油分布的研究方面，将储层沉积学方法和油藏工程方法有机结合起来，颇具创新和特色（1995）。林壬子等在储层地球化学、湖泊沉积相及储集砂体、层序地层学、砂岩成岩演化模式、储层地质模型和区域储层评价等六个方面的研究，总结了我国目前在陆相储层沉积学研究的新进展（1995）。

目前，储层随机建模技术已逐步被国内广大从事储层地质的专家们所接受。穆龙新等（2000）总结了储层精细研究的特点和内容、理论基础和方法、储层非均质性表征及定量建模、应用露头和现代沉积进行类比性储层地质建模、露头研究成果的应用。他们总结了国内外研究河流、三角洲等沉积储层的各种定性和定量地质知识库，储层预测方法尤其是随机建模的基本理论和应用。吕晓光等（2000）研究了储层随机建模的意义和目前出现的主要随机建模方法。穆龙新等（1994）论述了建立系列定量储层地质概念模型的一套新的研究思路和方法，指出把地质学原理和现代随机建模方法、计算机技术及开发应用紧密结合在一起，代表了储层建模向定量化发展的方向。李元觉（1994）利用假三维条件模拟法储层建模原理，提出了控制层状定容气藏几何形体和储渗体空间连续性的随机变量分布函数以及总体结构的保持算法，以模拟沉积相在各小层的展布为基础，形成了陕甘宁盆地中部气田中区马五Ⅰ储层的数值模型。利用该数值模型所作的储量评价及气藏数值模拟结果表明：该数值模型对井眼之间的储层参数估值更为合理，该方法能自动划出较为准确的有效储层边界，对该气藏较强的非均质性作出了比较确切的定量描述，并在微机上实现了“Windows”环境下中文界面的“条件模拟法储层建模软件”，可作为层状定容气藏储层建模的有效工具。聂昌谋（1996）在胡状集油田的油藏描述中，应用随机建模技术中的顺序指示建模方法对胡 12 断块沙三中油藏 7 油组的渗透率进行了研究，研究结果表明顺序指示建模方法比常用的插值方法更能反映储层参数的变化，而且能够描述由于资料的缺乏而引起的模型的不确定性。杜启振等（1997）对储层随机建模方法的国内外研究现状及其在油藏描述中的应用进行讨论。张永贵等（1997）讨论了在储层随机精细建模中使用模拟退火组合优化算法的方法和步骤，包括目标函数的构造、选择初始状态、抽样和扰动机制、退火策略参数的选择以及怎样综合多种类型的资料等问题。他们指出在储层建模中使用模拟退火优化算法有两方面的优点，即易于综合多种类型的资料，可以客观地再现储层变量的空间相关结构。其主要缺点是计算量大，有必要进行并行处理。文健等（1998）研究了埕岛油田馆陶组上段储集层随机模型。钟宝荣等（1998）对储层随机建模与条件模拟的计算机实现进行了讨论。张永贵等（1998）综合论述了储层地质随机模拟的意义、方法和技术及其在油气勘探与开发中的应用。赵永胜（1998）指出了随机建模结果使用的局限性。陈霞等（1998）讨论了油藏地质属性的随机建模用于储层研究的两个方向。李龙滃等应用地质统计学提出了用于井间砂体预测的一个方法。徐宁余（1999）提出了随机模拟技术在石油地质研究中面临的挑战和存在的问题，最后提出了使用该技术对陆相储层地质和油藏描述研究的经济价值和事件意义。俞启泰（1999）提出了利用河流相、三角洲相各 5 个具有代表性的小层平面等渗透率图，合成河流相、三角洲相渗透率平面分布数值抽象地质模型的方法和结果。吴晓光等（1999）研究了储层地质模型的类型、研究内容、建模步骤和建模方法。

0.2　随机模拟理论

在英文字典中，模拟（to simulate）一词的解释有好几种：假装，借口，装作，象，类似于，伪装成，模仿，为了方便或训练，按一定的模型来仿制一些条件（或形势）等等。这里，每种意义都有贬义的成分。但在技术方面，模拟一词意味着产生某种结果的模型，而不是利用被研究的真实系统进行的实验。如果这个模型具有随机的成分，那么这种模拟就称为随机模拟。

在第二次世界大战期间，随机模拟就称为蒙特卡洛方法（Monte－Carlo），用以对原子分裂的建模。

至于模拟的目的，Tocher（1963）写了这方面的第一个教科书“模拟的艺术”。他认为模拟是一种艺术，尽管对模拟装置所用的工具也应有相当多的认识。

对于随机模拟，首先需要的是随机性的来源。在随机模拟的算法中，最主要的，也是最基本的随机数发生器。它可以随机地产生一个数的序列，可以做到序列中前一个数完全不能预测下一个数是什么。

然而，目前人们还没有设计出真正的随机数发生器。在计算机中提供的只是一种伪随机数发生器。数学家们还在继续努力，使得设计出的伪随机数发生器如何更加随机化。只有设计出了随机数发生器（实际上只是伪随机数发生器），人们才能对一个一定分布的随机函数进行等概率的抽样。

模拟的目的是和模型的建立紧紧地联系在一起的。在统计学和运筹学的范畴内，建模的目的通常有两个：

1）对数据进行概括。用于回归中的统计线性模型就是对数据进行的一种概括。

2）对观测值的预测。回归方程可以在一个新的条件下预测一个系统的响应。这是模型在回答“是什么”方面的功能。

通常存在着两种模型：机械的和简便的。线性模型是简便型的，而遗传的模型被认为是表达了一个真实的过程。工程师们对物理世界的所有的模型，往往是确定性的和机械的。然而，大部分统计模型都是简便的。这些模型，都可用于理解、预测和决策。

如何使用一个模型，人们可以有两种选择。

1）利用数学分析支持帮助的方法，理解模型的功能和特点。对于线性模型，这相当简单。然而，对岩石结构中的流体流动方程和复杂排队系统，这几乎是不可能的。

2）对这个模型去进行实验。一个随机模型的输出可以是变化的，于是人们希望对每组参数，建立若干个实现。

对模型的选择将依赖于使用该模型的目的。模拟善于回答“是什么”的问题，而数学分析往往是用于加深对该模型了解。模拟的应用是一种方法混合的过程：做一次模拟的实验，再对它进行分析产生一个简便的模型，然后再用这个模型去预测和决策。

0.3　石油地质统计学的形成

地质统计学作为随机过程理论在地质上的一个应用，已经存在了近 50 年。自然界有许多随机现象，如地震，台风的发生，在经济领域，如期货，保险业的发展，以及医学领域中

的传染病学等问题，都可以用随机过程理论来描述和分析。认识这些自然现象和客观事物的本质，确定决策方案，减少损失，具有重要意义。

地质统计学最初的应用仅是在采矿业中的储量计算。直至 80 年代油藏描述技术出现后，地质统计学才应用于石油工业，最初是用在地质绘图。在近十年来，地质统计学才以储层随机建模的面貌应用于油气田的开发。

石油地质统计学作为一个学科曾被学者们偶然提及。只有当储层随机建模的应用日益广泛，被越来越多的石油地质学家、石油工程师们所接受的时候，这个术语才会被人们接受。然而，被国际上权威学者正式系统地加以叙述，则是在 1999 年（Journel，1999）。斯坦福大学石油工程系的 A.Journel 教授在他的论文中，以已召开的五次国际地质统计学大会取得的成果为纲，系统地叙述了石油地质统计学的过去、现在和将来。

如果矿业地质统计学可以追溯到 20 世纪 50 年代早期的 H.Sichel（1949）和 D.Krige（1951）的论文，那么地质统计学的理论可以追溯到 50 年代的晚期的年轻的数学家 G.Matheron（1955，1957）的两篇论文。美国斯坦福大学石油工程系教授儒尔奈耳在 1974 年发表的专著“矿业地质统计学”全面地总结了地质统计学在当时的发展（Journel and Huijbregts，1978）。这本著作的影响很大，至今仍被广泛地引用。

儒尔奈耳教授在 1974 年发表的论文中，讨论了随机模拟在矿业中的应用。那里，他称随机模拟为条件模拟。随后，他在 1978 年出版的专著“矿业地质统计学”第七章“矿床的模拟”中作了详细的讨论。地质统计学中的随机模拟在油藏描述中的应用，已被世界各国的同行称为油气储层的随机建模。它从挪威 Hydro 石油公司的哈得逊博士（H.Haldorsen）和美国德州大学奥斯汀分校的莱克教授（L.Lake）在 1984 年共同发表的油气储层随机建模方面的第一篇论文开始，经历了 15 年的发展历程。石油地质统计学直到 80 年代才日趋成熟，取得了地质统计学中的主流地位。

第一届国际地质统计学大会于 1975 年作为北大西洋公约组织高级学院（NATO Advanced Institute），在罗马附近的 Frascari 举行。届时，仅有一篇石油方面的文章发表。作者是 A.Haas 和 C.Jousselin。前者目前是这方面的带头人之一。第二届国际地质统计学大会则是在加里福尼亚的 Tahoe 湖于 1983 年举行的。那时，仅有少数几篇石油方面的文章发表，其中包括来自埃克森石油公司的 T.Jones 。他指出；“至于地质统计学在石油工业方面应用较少的原因，我并不认为石油本身是一个原因。”其实，真正的原因是人们对石油勘探开发数据的特点没有足够的重视。很清楚，克里金估计对铁矿石的储量的估计是合理的，而对渗透率等参数的等值线图的绘制就缺乏合理性。

石油地质统计学在 1988 年法国的 Avignon 的第三届国际地质统计学大会上取得了较好的成果。15 篇论文分成了两个分会场，构成了论文集的第二集。O.Dubrule 和 S.Rijswick 在他们的论文中提出了油藏的随机模型。石油地质统计学的出现，是因为现有的克里金估计方法无法解决目前的石油问题。

从 20 世纪 90 年代起，石油地质统计学和随机建模的发展迎来了新的历史发展阶段，出现了比克里金算法更为复杂的模拟算法。变异函数，多点统计学，地质体目标形状等结构的重现，以及对石油生产预测的不确定性的认识超过了各种克里金估计变量所能提供的信息。在这过程中，斯坦福大学油藏预测中心（SCRF）在宣传石油地质统计学的基本概念方面功不可没。

在 1992 年于葡萄牙的 Troia 举行的第四届国际地质统计学大会上，石油行业方面的论

文的数量比以往大大增加。这些论文构成了大会论文集的第一卷，对地质统计学的理论作出了重要贡献。其中，出现的标准算法有：离散数据，指示值的编码处理，概率场的模拟。这次大会还出现了模拟退火和更加一般的迭代随机模拟算法，包括马尔可夫链和序贯算法（不一定是高斯）。Troia 的第四届国际地质统计学大会，播下了石油地质统计学方面的概念的种子，必将在以后发芽开花。它们是以重建特殊的几何模式为目的的面向对象模拟，和以训练图像的多点统计集合。在这次会上，挪威学派在石油应用方面初露头角。Troia 会议的一个特点是将地质统计学家从严格的随机函数的理论框架中解放了出来。在 1993 年的以“下一个世纪的地质统计学”为题的学术会议上，C.Deutsch 运用了“算法定义的随机函数”的术语，意为通过所有的实现所产生的一个随机函数，这里的实现有一个给定的算法产生，而每一个实现则由一个随机种子完全确定。目前，地质统计学中所用的模拟算法的绝大多数属于算法定义的随机函数，包括大多数面向对象的算法。

在 1992 年前后，由于缺少免费的软件，妨碍了地质统计学广泛使用。人们需要一个自己可以单独使用和修改的公开的源程序。斯坦福大学的 C.Deutsch 教授和儒尔奈耳教授开发的软件系统 GSLIB 满足了这个要求（Deutsch and Journel，1992）。

使得石油开发地质学家和油藏工程师都感到十分鼓舞的是，石油地质统计学不仅可以为开发地质学服务，而且已经显示出可以为油藏工程服务。Journel（1999）指出石油地质统计学具有如下四大特点：

1）在石油地质统计学的应用中，一个特别重要的变量是渗透率。它不仅和所在的多孔介质有关，而且还依赖于压力和含油饱和度的条件。由于这种依赖关系，渗透率的平均规律并不符合线性规律。渗透率粗化和细化则是石油地质统计学的一项重要任务。

2）油气从一个地方向另一个地方的流动而被采出，靠的是泵的抽动，而不是依靠开采出一块多孔介质。这种油气开采的动力学方向特点对石油地质统计学绝对是至关重要的。因为油气流动路径和遮挡的模式，要比克里金估计理论中的局部精度更为重要。随机模拟等的算法能再现这种空间模式和结构，而且比传统的滤波式的绘图算法更为合理。

3）从静态的油藏模型的不确定性向开采预报的不确定性的转换是复杂的，要求利用油藏数值模拟来解决。这种模拟十分复杂，不能用任何的解析函数来表达。它是油藏模型的全部物性参数的一个隐含的函数。这种经过多点函数的不确定性的转换要求能产生多孔介质的多点统计量的油藏模型。这和传统的变异函数，协方差函数的两点统计量是不同的。

4）岩心和测井等硬数据的极端缺乏，将地震，测井等软数据的引入使数据量大大增加。由于大量的数据集成算法的产生，有力地促进了地质统计学在石油方面的应用的发展。这里，最著名的要数指示算法和贝叶斯算法。它们利用软数据，更新了硬数据的先验概率分布。

0.4　储层随机建模技术的特点

储层随机建模是在地质统计学在石油工业中广泛应用的过程中，逐步发展起来的。它的特点也是逐步形成的。

0.4.1　随机建模相对克里金估计的特点

地质统计学一般包括三个基本的组成部分：空间函数的相关性分析，克里金估计和随机模拟。空间函数相关性分析是指对变异函数和协方差函数的分析，包括它们的定义和估计方

法，它是克里金估计和随机模拟的基础。克里金估计和随机模拟之所以具有生命力，正是依赖于变异函数和协方差函数所反映的空间函数的空间相关性。

根据估计理论，克里金估计实际上是一种线性的最佳无偏估计。设 $\{Z(x) \mid x \in D\}$ 是研究空间 D 内的油藏属性，可将其看成是一个二阶平稳随机场。若已获得了在 D 内的点 x_i，$(i = 1,2,\cdots,N)$ 处的观察数据 $\{z(x_i) \mid i = 1,2,\cdots,N\}$，则可利用已知数据估计 $x \in D$ 处的值 $z(x)$。人们可以用随机变量的线性加权平均

$$Z^*(x) = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

来预测随机变量。为了得到一个合理的预测结果，必须采取一种衡量标准。那就是：

$$E[Z^*(x) - Z(x)] = 0$$

和

$$E[Z^*(x) - Z(x)]^2 = \min$$

由第二式可知 $\sum_{i=1}^{N} \lambda_i = 1$。

这样，便可以使用拉格朗日乘数法得到克里金方程组：

$$\sum_{i=1}^{N} C(x_i - x_j)\lambda_i - \mu = C(x - x_j)$$

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

在此，$C(h)$ 为随机场的协方差模型，μ 为拉格朗日乘数因子。求解这个克里金方程组可得到权系数 $\lambda_i(i = 1,2,\cdots,N)$，从而可给出未知位置属性的克里金估计。通过分析，克里金估计原理具有如下特点。

1）克里金方法区别于其它传统方法关键之处，就在于它不仅考虑已知点对待估点的影响，而且也考虑已知点之间的相互影响，即强调数据构形的作用。不同位置相互影响的大小是用协方差函数或变异函数来定量描述的。

2）从克里金方程组可看出克里金估计通过已知数据点，故克里金估计方法是严格内插方法。

3）利用最小均方误差标准式知，克里金估计实际上是对条件数学期望 $E\{Z(x) \mid z(x_i), i = 1,2,\cdots,N\}$ 的估计，在线性无偏估计类中是最优的。若 $\{Z(x) \mid x \in D\}$ 是高斯模型，则易知克里金估计在所有无偏估计类中是最优的。

综上所述，克里金估计实际上是一种线性平滑的低通滤波器，是一种对条件数学期望的估计，从而具有平滑效应。对于变化平缓的地质参数，使用克里金估计，并以等值线图形式来表达，结果会合理美观，可以反映平均变化趋势。

克里金估计在石油工业中应用主要是用于地质绘图（Olea，1984；Journel，1990；王家华等，1999）。

随机模拟技术是地质统计学中继克里金估计技术之后，迅速发展的一个新工具。它的起源和发展主要是由于油藏工程领域越来越重视对油藏非均质性的认识和评价，认为油藏非均质性是影响提高采收率的重要因素。这种非均质性必须用新的方法来刻画和定量描述。克里金估计方法由于是一种光滑的内插方法而不能定量评价储层空间的非均质性，由于克里金估计给出的是一种单一的数值模型，是一种对平均值的估计，所以不能反映隐含在随机场概率

模型中的整体相关结构，不能反映储层空间的非均质性，不能反映储层内的沉积相和沉积微相的空间分布。然而，随机模拟通过建立油藏属性的概率模型，再施行抽样过程，抽取各种相互独立的，等概率的来自模型各个部分的联合实现，用网格化数据来表达的一系列实现

$$\{Z^{(l)}(x) \mid x \in D\}, l = 1,2,\cdots,L$$

表示区域 D 中属性值 $Z(x)$ 空间分布的 L 个可能图像。这些图像的差异就反映了储层属性空间分布的不确定性。

随机模拟结果强烈地依赖于建立的概率模型。随机建模过程必须充分利用各种类型的地质数据并强调以地质概念模型为基础。正由于此，随机模拟不象克里金方法那样，有统一的公式和算法。克里金方法本身不与模型发生关系，它仅仅利用协方差或变异函数这个矩信息。而对随机建模而言，对于不同地质问题，需要建立不同的概率模型。建模本身是一门艺术，寻找一种快速实用的模拟算法对于特定的问题是至关重要的。

随机模型通常分为条件模拟和非条件模拟。非条件模拟所产生的随机图像仅要求再现概率模型所包含的关于储层属性空间分布的相关结构，而条件模拟所产生的随机图像不仅要求再现储层属性空间分布的相关结构，还要求条件化到已知地质数据和其它数据，比如井位数据。因此，通常讲的随机模拟一般指条件模拟。

下面通过序贯高斯模拟阐述模拟思想：设 $\{Z(x) \mid x \in D\}$ 是储层的属性分布，该随机场可看成高斯随机场。否则，可对之进行统计检验和变换处理，得到一个高斯随机场。将区域离散化为 $\{Z_i \mid i \in D_L\}$，D_L 为网格，网格数目为 N_L。利用高斯分布的经典理论知，高斯分布的条件分布仍是高斯的。这样可采用序贯思想模拟网格结点处的值：

$$Z_1 = z_1 \sim N(0,\sigma^2);$$

$$(Z_2 \mid z_1) = z_2 \sim N(\mu_{2|1},\sigma^2_{2|1});$$

$$(Z_3 \mid z_2,z_1) = z_3 \sim N(\mu_{3|2,1},\sigma^2_{3|2,1});$$

……;

$$(Z_{N_L} \mid z_{N_L-1},\cdots,z_1) = z_{N_L} \sim N(\mu_{N_L|N_L-1,\cdots,1},\sigma_{N_L|N_L-1,\cdots,1})。$$

在此：$\mu_{i|i-1,\cdots,1} = E\{Z_i \mid z_{i-1},\cdots,z_1\}$；$\sigma^2_{i|i-1,\cdots,1} = \mathrm{Var}\{Z_i \mid z_{i-1},\cdots,z_1\}$。这样得到一个模拟实现 $\{z_i \mid i \in D_L\}$，可模拟许多次得到多种实现。

由以上模拟过程可知，相对于克里金估计，随机模拟有下述特点：

1）模拟算法是依赖于模型的，不同模型可能要采用不同的模拟算法。

2）随机模拟结果强调结果的整体相关性。它从整体上对储层属性空间分布提供了不确定性的度量。而克里金方法不能保证条件化到统计量，所得到的结果，其直方图和协方差与原始数据计算结果会有很大偏差。

3）克里金方法只给出一个数值结果，而随机模拟能给出多种数值结果。这些结果的异同正好反映了隐含在概率模型之中的不确定性。

鉴于在油藏描述和储层表征中，人们所关心的核心问题是建立储层三维定量地质模型。采用地质统计学中的随机模拟技术，来建立储层三维定量地质模型，就产生了储层随机建模技术。

0.4.2　沉积（微）相空间分布的建模

目前从事储层随机建模的地质学家和油藏工程师们大多把 Haldorsen 和 Lake 在 1984 年发表的论文，看成是储层随机建模技术发表最早的一篇文章。该文讨论了储层中分布不连

续、不相关的泥岩的分布规律。然而，两位作者在该文中并没有明确利用地质统计学中的观点，而是利用了一般统计学的方法。鉴于泥岩的分布在储层对流体流动有着重要的作用，该文开创了一个新领域，利用统计学来研究储层的非均质性。

储层非均质性一般认为由沉积环境、成岩作用和构造作用等因素所致。它包括如下几种：层间非均质性、单一储层内平面和纵向非均质性、孔间非均质性、断层和裂缝。一个油田的采收率和其非均质性关系密切，取决于非均质性的分布，人们认识非均质性的程度，以及人们解决非均质性的能力（陈永生，1993）。储层随机建模明显地可研究前两种非均质性和断层与裂缝空间分布的问题。至于利用储层随机建模来解决孔间微观的非均质性问题，目前还未见到报导。

储层随机建模技术具有生命力的重要原因之一，是其具有石油地质学和储层沉积学坚实的理论根据。石油地质学和储层沉积学认为，渗透率，孔隙度，泥质含量和含油饱和度等储层物性参数的分布，主要受储层的沉积相或沉积微相的空间分布的控制。主河道和河间泥岩的渗透率相差几倍、几十倍，甚至几百、几千倍。为了要把渗透率的空间分布估计准确，应该先把沉积相或沉积微相的空间分布搞准。

分析利用计算机绘制的渗透率等值线图时，往往会觉得比较困难，并不容易发现渗透率空间分布的地质规律。然而，根据储层沉积学的原理，渗透率的空间分布主要受到沉积（微）相带空间分布的控制。河道内的渗透率明显地要大于河间泥岩的渗透率。若不考虑到沉积相带分布，仅考虑各井处的渗透率数值来进行在各个网格节点处估值，绘出的渗透率等值线图就显得杂乱无章，无规律可循。

0.4.3　地质、地震、测井、试井等数据的结合

条件化是储层随机建模技术的一个非常重要的特点。正因为具有这个特点，储层随机建模才能够把不同的数据结合起来。

0.4.4　不同尺度的非均质性分析

非均质性是一个相对的概念，以不同的尺度存在于储层。成因的不同，测量的手段不同，和使用目的的不同，都会造成非均质性具有不同的尺度。

一般来说，由成岩作用形成的非均质性都是小尺度的，从小于 1mm 到几米。由沉积作用形成的非均质性的尺度较大一些，从几米到几十米，几百米。构造作用造成的非均质性的尺度则要更大一些。

在油田中得到孔渗饱资料主要包括：一是岩心分析的结果，另外是测井解释的结果。这两种资料所测量的空间比例尺都非常小。岩心资料大概是厘米级的左右，再精密一些所分析的范围是厘米或毫米级的。测井解释的结果大概是米级左右范围的非均质性，在纵向上的分辨率可达几厘米到几十厘米左右。这种测量结果非常精确；是直接在实验室所观测到的或用分辨率很高的测井仪器所测量到的。这种非均质性在厘米级和米级左右范围都会有变化，其尺度是比较小的。相比之下，地震数据的分辨率就比较低一些，在纵向和横向上都达几十米，于是反映的非均质性的尺度是比较大一些。

考虑非均质性的形成过程中，应考虑非均质性形成的时间顺序和尺度规模。如对河道砂体而言，首先是河道的侧向迁移、垂直叠加，然后到小比例尺上的一些波痕运动。这些都不是一次性瞬时间形成的。它们都有一定时间的形成过程。我们知道假如按照一般的确定性的模型，把确定性模型用偏微分方程来表示一套过程，相当复杂。因为涉及到的变量太多了，好多参数都是未知的。

164202

0.4.5　随机建模和决策科学化（Holden，1999）

储层随机建模的最终目的是决策科学化。本文讨论了随机建模是如何被利用来抽取科学化决策的。为了在科学决策过程中做到定量化，有必要形成一个目标函数。最优决策就是使这个目标函数达到最大。利用随机建模结果进行决策的例子包括：

1. 在河流相储层中确定加密井的位置

在一个河流相储层中，假若有着若干口井，和两条河道。为了从这些井中使油气产量达到最大，要布置若干口注水井，使得在生产井和注水井之间的河道的体积达到尽可能的大。这样，在评价中需要对这个河流相储层的随机建模，和计算油气体积。

2. 钻井时的建模

在海相储层钻井时，储层建模的一个重要的任务就是使得钻成的一口井，在一个给定的层位内使其井身尽可能长。对每米井身，对于同一个统计模型，产生了一组所在层所对应的统计实现。基于这一组实现，井身的方向可以这样来确定，使所在层的井身的长度为最长。这时，随机建模可使用截断高斯模型，来评价不同的井身轨迹。

3. 历史拟合油藏中的加密井

在一个油藏内，建立了孔隙度和渗透率空间分布的随机模型。这个模型利用了一些井中的数据和其中一定数量的井（例如 7 口井）在一定时间（例如 7 年）内的生产数据。这时，决策的目的是经过一段时间（例如 10 年）的生产后，是否需要再打另外一定数量的井（例如 5 口），在何处布井为最优。基于历史拟合的随机现实，从这些新打的井中增加的油气产量是可以预测的。把预测的收入和这些新钻井的费用相比较，以决定这 5 口井的位置是否为最优。

第 1 章　储层随机建模的地质应用

自从 20 世纪 60 年代初期，法国的马特隆教授提出区域化变量理论并创立地质统计学以来（Matheron，1963），地质统计学应用于油藏描述领域已取得了很大的进展。地质统计学的早期应用，主要应用于绘制一个或多个地层参数的空间分布图，强调使用变异函数模型和克里金估计方法绘图，并把克里金误差作为衡量精度的指标。克里金方法是一种线性平滑内插方法，它优于传统方法（如三角剖分法、距离反比加权法等）之处在于它不仅考虑到被估点位置与已知数据位置的相互关系，而且还考虑到已知点位置的相互关系。这相当于从整体上强调数据构形的作用，并使用变异函数对相互关系进行定量的刻画。但是，由于克里金估计方法仍是一种数据内插方法，把它用于储层评价常常会平滑掉储层特征在空间展布的变异性，从而对研究储层的非均质性和不确定性是不合适的（Journel and Isaaks，1984）。不难验证，克里金方差独立于数据值，因而用克里金方差作为估计精度的一种衡量指标是不妥当的。

储层描述的目的，就在于尽可能真实的模拟储层特征在三维空间的展布。因此，最近，地质统计学已不再强调克里金估计方法在绘图方面的应用。克里金估计方法现在被用来对储层空间的不确定性进行随机建模，或者把克里金估计方法作为应用于对储层进行随机建模（Stochastic Modeling）的一个基本工具（Haldorsen and Damsleth，1990）。不难发现，储层分布中的不确定性除了与数据构形有关外，还依赖于具体的观测数值。如果建立了储层的随机地质模型，那么可以采用相应的随机建模（Stochastic Simulation）方法产生储层属性空间分布各种变化等的实现。随机建模是一个抽样过程，该过程抽取各种可能的、等概率的随机模型各个部分的联合实现，一系列实现（通常用网格化数据来表达）$\{Z^{(l)}(\vec{u});\vec{u}\in A\}l=1,2,\cdots,L$，表示区域 A 中属性值 $Z(\vec{u})$ 空间分布的 L 个可能的图像。每个实现也称为随机图像（Stochastic Imaging），反映了强加在随机模型 $Z(\vec{u})$ 的某些性质，如果来自样本的与匹配的随机模型 $Z(\vec{u})$ 相适应的性质越多，那么随机模型就越好。

模拟在以下两个方面不同于克里金或任何内插方法：

1）在大多数内插方法中，主要目的在于每一次只取一个未抽取位置，然后对该抽样位置提供一种“最优”的估计 $z_k^*(\vec{u})$，而并不特别考虑关于估计 $\{Z_k^*(\vec{u});\vec{u}\in A\}$ 的最终空间统计量。然而，在随机建模中，主要目的就在于考察模拟值 $\{Z^{(l)}(\vec{u});\vec{u}\in A\}$ 的最终总体特征或结构以及统计量，而此时局部的精确性是次要的。

2）对指定的数据集和条件统计量，克里金估计作为一种内插算法，是一种单一的数值模型 $\{Z_k^*(\vec{u});\vec{u}\in A\}$。该模型在局部精确定义下是“最好的”。而模拟提供许多不同的数值模型 $\{Z^{(l)}(\vec{u});\vec{u}\in A\},l=1,2,\cdots,L$。它们的每一个在整体上可以说是对现实情况的“优良”描述。这 L 个变化图像之间的差异提供了对空间联合不确定性的一种度量（Journel and Gomez－Hernandez，1989）。

随机建模通常分为条件模拟和非条件模拟。非条件模拟所产生的随机图像仅要求再现随机模型所要求的关于储层属性空间分布的相关结构，而各种模拟所产生的随机图像不仅要再现储层属性空间分布的结构，且要求条件化到已知的井位数据，或者说这些随机图像在已知

井位观察位置必须与已知数据相一致。通常讲的随机建模一般是指条件模拟。值得指出的是条件模拟技术所产生的随机图像能够从整体上对未抽样位置提供一种不确定性的度量，而不是局部的，一个位置接一个位置的反映这种不确定性（Journel and Huijbreg，1978）。

使用统计概念和随机性概念来建立地质模型把油藏工程的研究提高到了一个新的水平。建立储层随机地质模型的过程简称为储层随机建模（Haldorsen and Damsleth，1990；Dubrule，1989）。早期研究储层行为主要是使用定性的或半定量的确定性地质模型，称之为储层描述（Reservoir Description）。这样建立的地质模型只能是对储层大范围特征的定性的认识，而且常把储层的属性分布看成是连续性和持续性都比较好的理想模式，对储层物性参数在大块段上常赋予一个平均值来刻画。但是，真正的储层行为要比想象的复杂得多，储层分布及结构的非有序性、各向异性、非连续性、非均质性是普遍存在的。为了探索这种复杂性，就有必要利用充分的地质资料结合现代新技术来对现代储层的复杂性进行定量化的评价。这样建立的模型不同于早期的储层描述，称之为储层表征（Reservoir Characterization）技术。由于获取的地质资料、信息、数据的分散性、稀疏性、跳跃性、间接性和随机性，因而利用这些资料来建立模型其结果肯定包含不确定性和多解性。所以储层随机建模技术提供了解决上述问题的一种有力工具。

已经发表的工作表明，随机建模技术可使用于对储层从静态行为到动态行为整个全过程的研究。使用于研究储层静态行为包括下列几个方面：

1）岩相的划分和空间分布。

2）沉积相带的划分和空间分布。

3）孔隙度、渗透率和含油饱和度等参数的空间分布。

4）砂体中悬挂着泥岩的空间分布和尺寸。

5）储层中遮挡带的空间分布。

6）不同砂体之间的连通性。

7）储层的空间位置和形状。

8）断层和裂缝的位置、方向、长度及其空间分布。

研究储层动态行为有如下几个方面：

1）储层可采油气的动态分布。

2）储层的产量预报。

3）油田开发中的风险评价。

4）提供油田开发方案。

常用的随机建模方法有：

1）序贯高斯模拟（Sequential Gaussian Simulation）。

2）序贯指示模拟（Sequential Indicator Simulation）。

3）布尔模拟（Boolen Simulation）。

4）截断高斯模拟（Trancated Gaussian Simulation）。

5）退火模拟（Simulated Annealing）。

6）分形模拟（Fractal Simulation）。

7）示性点过程模拟（Marked Point Process Simulation）。

8）镶嵌过程模拟（Mosaic Process Simulation）。

以下具体讨论随机模拟应用于储层评价的几个方面。

1.1　非均质性的评价和定量化

近几年来，储层非均质性研究已愈来愈受到地质人员的关注，而且已经认识到储层的非均质性是影响提高采油率的主要原因（Weber and Meatschappy，1986）。没有详细的储层评价就实施开发方案常会归于失败，这是由于未能识别出储层中一些遮挡带对流体流动的影响。例如在储层开发阶段，由于井跨度越来越小，大量的取心数据、测井数据、试井数据、产量数据提供了比储层早期评价阶段更为详尽的资料，这些都有助于对储层的非均质性进行更为详尽的描述。同时，大容量，高速度的电子计算机的出现以及各种模拟方法的形成，使得评价储层的非均质性成为可能。

提供收集露头数据和各种地质数据的结合，也使得可以对各种类型的非均质性进行定量的处理。不同类型的非均质性只能借助于一定的模型才能更好的揭示出来。非均质性伴随着不确定性，对储层非均质性评价的好坏依赖于是否利用已知数据正确的揭示了储层中所包含的不确定性。随机建模技术能够给出在已知数据条件控制下，相应于一定地质概念模型储层的各种可能的展布，各种结果的综合就反映了储层评价中所包含的非均质性和不确定性。基于特定的模型，可以研究在储层的各种非均质性影响下，所采用的开发方案的风险性（Damsleth，Tjølsen，Omre and Haldorsen，1990）。

储层的非均质性体现在储层的各种观察尺度上，即从宏观的非均质性到微观的非均质性。如来自于沉积地质特征的大范围储层构造特征；无断层的储层内部渗透体之间的边界，这相应于储层沉积相的边界，或称为生成元的边界；在每个特定生成元内的渗透带，与沉积环境和成岩作用有关的遮挡带；裂缝的出现也常常是难以定量化的非均质性。孔隙类型和多孔介质内部的非均质性属于微观的非均质性。这些非均质性的观察尺度从几十公里到几个微米。

非均质性可以看成一种弱连续性或不连续性，或者说成是地质特征在一定尺度上的剧烈变异性。因此，传统的内插线性平滑方法显然是不适合描述储层的非均质性，必须寻找能反映储层非均质性的模型和相应的处理方法（Journel and Alabert，1990；Journel and Gomez－Hernadez，1989）。众所周知，建立的储层模型都必须考虑到尺度效应，这一点是不容忽视的。模型的可靠性和精度是相应于一定的井间跨度或观察尺度的，也是与井的密度有关的。例如，利用大跨度井位数据推断出的结论只能反映储层宏观的非均质性，若要用该模型对小尺度范围内储层特征进行预测就必须收集更多的小跨度数据，例如，与研究区块具有类似地质特征的露头资料的收集。否则，模型对储层小尺度的非均质性预测的可靠性就差一些。正由于模型的可靠性依赖于数据密度和观察尺度，因而使用特定模型来预测储层的详细特征就具有不确定性，从而模型的结果具有多解性，这样才能反映由于掌握信息有限所引起的不确定性。储层随机建模技术的应用，可以有效地刻画和定量评价储层的非均质性，下面就具体的讨论储层随机建模技术应用于油气储层评价中的若干方面。

1.2　储层静态行为的研究

1.2.1　对河流三角洲砂体露头研究和随机建模（Matheron and others，1987）

根据露头资料建立相应的随机模型，并利用露头岩相资料和已知数据确定不同岩相在不

同参照水平面上相应的频率，用不同的频率值对高斯随机函数进行截断，然后确定不同岩相的变异函数与交互变异函数，最后可给出河流三角洲砂体岩相的模拟结果。这样处理的优点在于如下几个方面：

1）可以反映三角洲砂体岩相在空间剖面上的沉积次序关系。

2）可以反映不同岩相在空间累积频率分布情况。

3）利用露头资料可以使二维岩相的研究容易地过渡到岩相在三维空间的展布。

4）砂泥岩夹层在三角洲砂体中分布较广，直接影响着三角洲砂体物性在空间的非均质性分布，而露头研究对搞清砂泥岩夹层的分布和相互关系是至关重要的。

可用截断高斯模拟方法来解决上述问题。

1.2.2　河流沉积相分布的随机建模（Doyen and，Psaila，1984；Syversveen and Omre，1996）

沉积相的模拟也是随机建模技术应用的一大方面。地质沉积相的描述和刻画对于研究储层行为具有决定作用。对河流相沉积，如何有效的寻找河道砂体位置和延伸范围对搞清储层的动态行为至关重要。对曲流河道砂体的分布的描述可用以下两种基本的方法来处理：

1）示性点过程方法，又称为面向目标（地质体）的方法。

2）指示方法。

面向目标的方法主要是把描述的沉积对象看成是在三维空间中其位置、长度、宽度、厚度等是具有一定分布且有相互作用的地质体，建立相应的随机模型并进行相应的模拟。这种方法的缺点是难以条件化到已知的实验数据，优点是服从几何体形状的空间分布。指示方法的缺点是不能完全描述几何形体，其优点是易于操作并易于条件化到已知的实验数据。

对于辫状分流河道沉积相的模拟，可以建立相应的随机模型——随机游走模型（Wang and Zhang，1997）。效果表明，这种方法是解决这一难度较大问题的有用工具。

1.2.3　储层内部物性参数非均质性分布的随机建模（Gomez－Hernandez and Journel，1994）

以前指出储层物性分布的非均质性已愈来愈受到地质工程人员的重视。正是这种非均质性使得预测储层的动态行为增加了很大的难度。对储层物性非均质性认识不足常常导致乐观地估计了采收率，过于简单的布置了注水井和开发井，从而收到的效果并不令人满意。为此必须使用随机建模方法对储层物性空间分布的非均质性进行充分的描述，只有这样，才能了解利用已知资料和已知模型究竟在多大程度上控制着储层的行为，才能了解到底还有多少不确定性还有待去探索。

经常使用的模拟储层物性空间分布的地质统计学方法有：序贯高斯方法、序贯指示方法、LU 分解法等。

1.2.4　河流相储层砂体连通性分布的随机建模（MacDonald and others，1992）

河流相储层常具有难以描述的几何结构，这往往可归结为砂体伴随着较复杂的沉积环境。好的砂岩（高渗性）和不好的粉砂岩（低渗性）、泥质砂岩和泥岩经常混合在一起。油气在采收过程中，主要流动路径是通过砂岩进行的，穿过不好的岩石的流动速度大慢而变得没有经济价值。很清楚，在特定的井位构形下，砂岩的连通性对于确定油气的采收率来讲是至关重要的。若进一步能确定整个连通砂岩体的形状，那么对获得储层流动性态的信息是很有意义的。通过相应的随机建模技术来揭示河流相重叠砂岩的连通性和传导性的一般特征是可行的。可以考虑这样一个简单的砂体模型，那就是把砂体看成是随机放置的几何形体，它

嵌入在一种非渗透的岩石之中。虽然在地质上这种假设不具有一般性，但它确定能给出洞察这种复杂系统行为的一种强有力的工具。可通过随机点过程中的渗透理论（Percolation Theory）来解决这个问题。通过模拟结果表明：当有效厚度低于某个特定的阈值后，砂体不具有连通性。当有效厚度大于这个特定的阈值时，连通性将有很大程度的提高，然而有效渗透率的增长却比较缓慢。而且，将渗透理论用于不同观察尺度时，不难发现连通性和传导性随着井跨度的增加而减少。

1.2.5　断层带的随机建模和模拟（Omre and Salna，1990）

许多油藏的特征受到断层和裂缝的影响，一般来讲，落差大于 20m 的较大断层带可以从地震记录中观察到的。断层可以是一个单元的地层沿着某个连续平面的下滑，也可以是由许多的小裂缝组成。虽然，断层带中实际的断裂类型支配着流体的流动是一个众所周知的事实，但是对于流体穿过断层流动的准确机理，目前还未有满意的解释。然而，地质学家对断层类型进行了大量的研究，对其中的一些作用力也给出了一定的解释。因此，利用随机建模对断层带及其断裂类型的识别和研究的必要性和可能性是十分明显的。

小的断层在地震记录中是无法观察到的。它们的外观方向和大小依赖于它们所构成的主断带。建立随机模型要根据一般的构造地质方面的知识。利用随机建模，不仅会产生和断裂类型有关的各种实现，而且还可以用断层中心线，垂直落差和水平范围等可以观测到的信息对实现进行条件化控制。每一个实现有可作为油田产量模拟的输入，以便评价断裂类型对油气产量的影响。

有一种方法可以对一组断层的空间分析进行建模，还可以对油藏中小于地震分辨率的小断层进行模拟。考虑到地震观测到的断层的系统性的特性，一组断层可以通过一个示性点过程（MPP）的一个实现来建模。单条断层可以被视为空间中的一个示性点，示性的内容包括断层的长度、落差、倾角和倾斜方位角。断层的模式具有两个的特点：成对相互吸引和成对相互排斥。

以上问题可用示性点过程随机建模方法来解决。

1.2.6　储层裂缝分布的建模

碳酸盐岩是沉积岩三大类型之一，我国近十多年来油气勘探在碳酸盐岩领域中有重大突破，已愈来愈清楚地认识到碳酸盐岩是一个重要的找油气领域。因此，它已成为石油地质很重要的研究内容之一。认识和识别碳酸盐岩储层对于油气勘探具有重要的实践价值。由于构造作用而形成的裂缝型碳酸盐岩储层在全部碳酸盐岩储层中占有相当大的比重，如何有效的描述和预测其中裂缝的位置和分布是一个非常重要的问题。裂缝储层比较复杂而且定量描述有相当大的困难。在很多情形下，对这种储层的描述方法并不令人满意。但是，如果把裂缝模型定量地用一种迭代函数系统（简记为 IFS）来描述，并用分形模拟方法，可以给出具有类似特征的裂缝分布模拟结果。

具体应用时需使用的信息有：

1）形成储层的随机演化过程、地质特征、地质构造运动过程、应力定律等一系列的地质知识。这些对于我们形成储层裂缝分布的地质概念模型是至关重要的。

2）裂缝数据：包括已知的裂缝位置、长度、密度等（这些可利用地震与倾角测井资料给出）。

有一种方法可利用模拟退火的最优化技术来模拟一个裂缝系统。这个方法具有如下的两种特点：

1）每一条裂缝视为一个个别的对象。它具有空间中的各种状态：它能生长、压缩、转移、旋转、死亡。

2）一组裂缝的空间演变由下列因素控制：双直方图、裂缝的大小和方向，分形维数，交叉的数目，不同支撑的协方差。

对裂缝性油藏进行三维几何描述发表了一种方法。它遵循不同裂缝组之间的传播和相交的一般规则。这个方法要求裂缝密度和方向以及统计量。其目的在于综合使用关于地下裂缝系统的完全不同的数据。这些信息来自井中的岩心、测井、露头，地应力和地震资料。这些数据可以是定量的或定性的，可以有不同的可信度。

1.2.7　储层含油气体积的随机建模（Berteig and others，1988）

由于储层结构的复杂性和所占有资料的局限性，在利用传统方法预测含油气体积的过程中，存在着不同程度的不确定性。这种不确定性反映了预测储量和实际储量之间的差异。为了在油气储量预测过程中考虑到这种不确定性，并将它进行定量化，需要采用随机建模方法。通过对储层几何特征及其中物性参数的模拟，可以求取储量的各种可能值，从而得到储量预测值的概率分布曲线。该概率曲线既包含了人们对实际问题的认识，也反映了预测所包含的不确定性。

评价油藏的各个阶段的行为都要对含油气体积进行预测，为此需要以下信息：

1）地震资料，它们具有较好的空间覆盖性，但其精确度较差。

2）井中的观测数据，包括储层的埋深和物性参数，由于井的数量有限，这种数据的数量较少，但比较精确。

3）来自邻近地区的储层评价经验，它们难以公式化。

4）一般的地质知识。

如果对一个储层的各项参数完全了解，那么其含油气体积 V_H 可以利用下式求取

$$V_H = \int_{D_r} \varphi(u)[1 - S_w(u)]\mathrm{d}u$$

这里 D_r 表示该储层含油气的范围，$\varphi(u)$ 表示孔隙度，$S_w(u)$ 表示含水饱和度。如果对储层的认识不充分，那么含油气体积的预测将包含不确定性。为了评价这种不确定性，就必须采用随机模型。

含油气体积的随机表达式可写为如下形式：

$$V_H = \int_{D_r} \varphi(u)[1 - S_w(u)]\mathrm{d}u = \int_{D_r} I_r(u)\varphi(u)[1 - S_w(U)]\mathrm{d}u$$

这里，随机函数 $\{I_r(x); x \in D_0\}$ 表示储层的位置和范围：

$$I_r(x) = \begin{cases} 1, x \text{ 在储层内} \\ 0, \text{其他} \end{cases}$$

x 称之为几何变量。随机函数 $\{\varphi(x), S_w(x); x \in D_0\}$ 称为物性变量，它们分别代表孔隙度和含水饱和度。区域 D_0 是一个包含了储层的固定空间区域。根据实际的地质知识，$\varphi(x)$ 和 $S_w(x)$ 互相有一定关系的。

对以上的每一个随机函数，都必须给出相应的随机模型，且具有相应的模型参数。可使用传统的蒙特卡洛模拟方法生成含油气体积的值，从而得到它的概率分布：

$$\mathrm{Prob}\{V_H < v\} = F(v)$$

这就是要得到的储层含油气体积预测值的累积分布函数。

1.3　运用地震数据进行建模

近几年来，在利用地震数据来进行储层的随机建模方面已经发表了相当多的成果。这方面的研究成果得到了许多地球物理学家的承认。

储层的随机建模是利用井中数据，变量的变异函数和直方图来进行约束。它的不确定性可以利用产生一些可能的图像或实现，而所有这些图像和实现满足上述的统计量和井中数据。由于井中数据的缺少，特别在油田评价阶段，可以产生影响非均质性模型的很大的不确定性。因此，任何其它可用的数据，甚至受不确定性影响较严重的数据，必须用来对模型进行更好的限制。

有许多数据可以获得，用于更好地对井间的非均质性建模，例如测井数据，地震数据和动态数据（表 8.1）。为了把这些数据有机地结合起来，人们提出了两类方法：直接的方法和非直接的方法。直接的方法用于这样的场合，被进行约束的数据可以写成被模拟的变量的确定性函数。例如，地震响应可以以一个波阻抗模型的一维褶积来导出，试井获得的渗透率，可以通过围绕这口井的若干个网格块的渗透率的平均来获得。非直接方法则假定在一个位置上的被模拟的变量和限制的信息之间存在着一个统计关系。例如，波阻抗模型可以利用孔隙度与波阻抗剖面之间的关系，提供模拟孔隙度的直接方法。

表 8.1　影响油藏生产的主要非均质性及其早期确定的方法

储层类型	主要非均质性	早期确定的方法			
		地震	测井	生产	模拟
层状、饼状	高度连续的反射层； 渐灭； 渗透率剖面	层序地层学分析	岩心； 用于相分析和渗透率估计的测井数据； 重复式地层测试器的结果/多重动态图像测试器的结果（RFT/MDT）	垂直脉冲测试； 长时间测试； 生产测试	泥岩长度的统计量； 碳酸盐胶结的地层的同位素特征； 沉积模式
交错状	泥浆漏失层； 有方向性的流动单元； 渗透率的明显不同； 局面的反射层	3D—地震露头； 层序地层学分析	岩心； 井中图像测井（RFT/MDT）	长时间测试； 生产测井； 水平脉冲测试	砂体大小特性； 沉积模式； 泥岩长度统计量
迷宫状	有方向性的流动单元； 相互连通性（砂泥比小于 0.35）； 堵塞端； 重力圈闭	3D—地震露头； VSP； 井井之间的地形	岩心（有方向性）； 图像测井； RFT/MDT； 用于概率模型的测井； 水平测井	水平脉冲测试； 生产测试； 生产测井	砂体大小特性； 沉积模式

1.3.1　直接的方法（利用地震信息限制波阻抗模型）

如果波阻抗（密度乘地震速度）的一个三维分布用条件模拟已经被产生，那么就发生了一个问题：这个已被产生的波阻抗模型是否和已知的数据相一致。同样地可以说，如果在井间的一个渗透率条件模拟已经被产生，测井数据可以被用来限制在这些井的一个邻域内的渗透率变化。这就是下述的方法的目的。

在应用过程中，地球物理学家们常常利用著名的褶积模型，把一个合成地震振幅响应和波阻抗数值的一维垂直地震道联系在一起。一个称为“地质统计学反演”的方法由 Haas 和 Dubrule（1994）提出，可以描述如下：一个三维的波阻抗条件模拟可以从井中数据进行建造，这个经过一维褶积后产生的模型是和真正的三维地震振幅块相一致的。

利用条件模拟的序贯算法，三维波阻抗的一个实现可以用序贯方法来计算，一个波阻抗在垂直方向的值计算出后，然后再计算一个波阻抗的值。这个算法的输入数据是沿着每口井处的波阻抗值的垂直序列。对每一步，水平面上的一个点被随机地抽取。在这个位置上，若干个可能的波阻抗值的垂直序列被序贯高斯模拟方法所采样，这时已有的井中数值作为输入（Gómer - Hernándey and Journel，1993）。每个可能的垂直序列和地震子波相褶积，然后与在相同位置是真实地震振幅值相比较。这时的比较可以基于各种标准，诸如相关系数或最小平方差。一旦一个序列被发现满足拟合标准，那么在这个位置处的采样就停止，所发送出的序列被放入数据库，而且一个新的迭代又开始了。

具体方法如下：

1）以地震时间来建立储层的地质模型。储层的顶底是用定义了经过解释的地震数据的一组尖峰导出的。

2）网格间距在垂向上是 1～4ms。

3）井中数据需用地震数据进行校正。这意味着通过比较反映反射率的测井响应和在同一位置地震响应，地震子波必须匹配于井中数据。

4）地质统计学参数利用测井和地震数据计算而求取。侧向变异函数用地震数据求取。测井数据用来计算对每一层的先验模型（波阻抗的均值和标准离差）和垂向变异函数。

5）地质统计学反演过程这时就可以开始了。一个模拟过程跟随着一个随机路径。在每一个地震道的位置，由序贯高斯模拟产生一个波阻抗道。在每一个位置产生了大量的波阻抗道，且相应的反射系数被计算出来。经过和子波褶积后，波阻抗剖面将导致和真实地震道的最佳拟合。

这个称为“地质统计学反演”的方法至今为止已经给出了有说服力的结果，已经被若干商品化软件所采纳。地质统计学反演不同于标准的“层序反演技术”在于，它并不产生唯一的最佳解，而是产生一组三维的波阻抗实现。它们都和真实的三维地震相匹配。它们之中一个和另一个的差别是相当大的，在这个例子中，两个层展示的实现都通过井中的数据，直方图和变异函数等先验统计量也是同样的。

在当前例子中，各实现之间的变化说明了，尽管有由波阻抗和井中数据来限制，波阻抗的实现之间还是存在差别，也就是不确定性相当大的。正如所预料的那样，井的数目太少，地震分辨率还是不足以确定的方式来限制波阻抗值的分布。

尽管已有高分辨率的地震技术以及处理技术不断改进，但是不唯一性仍然是地震反演中的一个基本问题。引入不同的假设会导致不同的模型。最简单的假设是考虑波阻抗的均值和方差在地层学的水平上是独立的，没有任何随深度的变化。一张地震反演图像可能沿着垂直

方向的构造变化很少；而另一张地震反演图像可能受到较多的约束，且以随着深度变化均值和方差为特征，层理显得更为明显。然后，对应于这些不同实现的合成地震图几乎是相同的，尽管当地层约束成为较强时，合成地震图会有一点小的变化。影响地震反演的非唯一性以使人信服的方式被显示出来了。

直接方法必须用于条件模拟而不是克里金估计。这时因为条件模拟显示了所研究变量的变化和所期望的是一致的。但是，直接方法只能用于连续的变量，例如波阻抗、渗透率。对于这些变量，可以利用进行限制的那些测量值建立起确定的关系。

现在，假定已经具有在井间的波阻抗信息，如何来导出关于岩相分布的信息。这一点是非常重要的，人们可以期望这一点。

Doyen 等人（1994）的论文是关于从地震数据导出岩相研究的主要文献。这篇文章利用了这样的思想：有两种相互独立的信息可以减少确定岩相的不确定性：

1）在井中的任一种岩相，在井间任何一点发生的概率可以利用克里金进行估计。

2）这个连续参数（例如波阻抗）在这个位置的值，可以被表示为“可能”的值。

这个由克里金估计求得的概率被 Doyen 等人称为先验概率，和可能的值相乘产生一个后验概率。

1.3.2　非直接的方法

利用这些一类方法，例如 Markov－Bayes 方法，外部漂移方法，误差协同克里金估计等等，能解决一个变量的地质图件绘制的一类问题。这个变量在各井处是已知的。利用这些井间的信息，这类方法会产生和这些被内插的变量有一定关系的一个辅助变量。例如：

1）井中速度和地震速度相结合产生一个速度模型。

2）把岩心平均的物性参数和非取心的井中参数的结合。

3）把地震属性和井中的砂泥比相结合产生油藏各点的砂泥比。

4）把地震波阻抗和井中的孔隙度相结合，产生油藏各点的孔隙度。

1．外部漂移方法

这点也许从理论上不大十分满意，但是它却是应用十分广泛的一种方法。其原因是它是较早地应用于将井中数据和地震数据相结合的一种方法，并给了惊人的结果。这种方法将地震信息视为一种全局的趋势，而井中的信息解释为关于全局趋势的参数的局部变化。Chambers 等人和 Wolf 等人已经分别运用了外部漂移技术将波阻抗和井中孔隙度结合起来了，和将地震幅度信息和砂体厚度信息相结合了。

2．同位协同克里金方法

从理论上讲，Xu 等人（1992）的方法比外部漂移方法要好一些。Doyen 等人（1996）的论文显示了并列协同克里金估计可以简化成如外部漂移方法那样简单的方法。并列协同克里金估计的优点是地震变量和人们研究的变量之间的相关性可以被真实地进行考虑。这意味着这种相关性越好，那么越多的权系数就会自动地赋给地震数据（和仅利用井中数据的克里金估计相比）。

1.3.3　结论：数据结合的理论

目前，多种数据结合的地质统计学方法已被大量地采用。然而，目前的发展仍然满足不了实际需要。在地震、地质和生产数据很有效地结合方面，还有大量的研究需要进行。目前，大多数研究集中于将生产数据和地质数据相结合，以及将地震数据和地质数据相结合。数据相互结合的困难在于不同的油藏数据对于不同的参数是敏感的，所以这些参数并不能一

对一地相互联系起来。例如，叠后的三维地震数据对应于波阻抗的对比；而叠前数据则受到P波和S波速度的影响。另外一方面，试井数据受到渗透率变化的影响。作为例子，在河道—三角洲沉积中，往往很难用地震方法把砂体从泥岩中区分出来，其原因就在于煤和其它岩心界面的波的对比，相对砂体和泥岩界面的波的对比是相当强，然后，区分煤和泥岩对储层建模是毫无用处的。岩心物理学在多种约束数据的结合方面将发挥一个基础的作用。众多的参数之间的相互关系必须建立起来，并且这些参数将在油田范围内和各种比例上是相互依赖的。

1.4　两阶段随机建模

储层的两阶段随机建模方法综合了近年来出现的各种模型和方法，适用于描述储层在各种尺度上的非均质性。这种非均质性可以由于物性不同而引起的。该方法的主要步骤如下：

在第一阶段，不同地质体或不同沉积相所引起的大范围非均质性用离散的示性点过程来建模，其模型参数是从地震数据、测井数据和露头数据来获取。在第二阶段，基于第一阶段模拟结果的基础上，利用多元高斯场对不同沉积带和各种物性参数分别建模。对不同的沉积相带，相应的高斯随机场的参数是有差异的。露头数据和井中的岩心数据、测井数据可用来确定有关变量的变化，各变量的关系和空间结构，进而求取有关模型的参数。

两阶段建模和模拟的主要优点在于，第一阶段的模拟结果提供了变量的主要空间结构和形态，而第二阶段模拟的则是揭示储层小范围的变化。无论是哪个阶段，随机建模皆以井中的观察数据为条件。

事实上，分两阶段进行建模和模拟，体现了将地质变量按沉积相分类，然后分别处理的思想。在第一阶段，沉积相类型的标签是离散的，可用离散随机变量进行描述，因此建立模型可用离散的示性点过程，其中往往隐含着离散的马尔柯夫性。第二阶段处理每个沉积相带内部的物性参数。由于物性参数在空间是连续变化的，因而建模可用连续的高斯随机场。

两阶段随机建模是一种通用的方法。通过两个阶段的随机建模，将再现油藏在各个尺度上的变异性和不确定性，而且可以进一步将这些模拟结果作为预测的输入信息，评价出预测的可靠性（Damsleth and Haldorsen，1990；Wang and Zhang，1995）。

1.5　用于油气产量预报

1.5.1　建立描述储层非均质性的地质模型

储层工程技术研究基本上考虑如下问题：

1）估计储层在空间的位置以及可采油气的分布。

2）提出一种油田开发方案。

3）进行油田产量预报。

这些预测的质量依赖于对“系统成分”的认识，那就是：

1）容器（储层）。

2）内容（空间的流体）。

3）相互作用以及对产量的影响。

特别值得指出的是，必须充分了解下列诸多方面：

1）多孔介质的结构以及流体迁移的地质环境。

2）空间流体的热力学性质。

3）物理规律和物性参数，它们可用于对流体和外部作用力之间的关系进行建模。

4）边界条件，包括流体压力条件，输出井的情况等。

早在70年代的研究工作主要集中在如何增加油气采收率的各种基础上，例如如何注入混相或非混相气体，活性剂和聚合物溶液，蒸汽或热水等。其主要目的在于提高油气产量，或者可以认为研究主要集中在空间流体的热力学性质与物理规律和物性参数两个方面上。但自从80年代中期以来，储层表征和非均质性对流体性态的影响的研究工作加快速度发展。这方面的研究工作已成为许多研究项目的主要内容（Wolotta and other，1993）。

非均质性对产量的影响已愈来愈重要，它对流体流动的物理化学性态有重要的影响。事实上，所有储层无论产油还是产气，物性在空间的变化有广泛而强烈的变异性。这来自于地层的历史、沉积模式、物理压实和埋实，变形和胶结作用。这些都影响着流体的初始分布。即使对于渗透性好的储层，非均质性的存在使采用通常的井跨度和传统的采收手段也难以对流体进行完全的驱替。在非均质性严重以及分块型储层，利用一次和二次采油只能采出不到可采油气的50%。当储层非常薄、或流体粘性大，流动缺乏活力时，非均质性对采收率的影响也是非常大的。

因此，如果要全面考虑1）到4）诸方面时，非均质性应予以充分重视。

已经指出，非均质性在空间的各个尺度上都能表现出来。垂向上从厘米级到米级的非均质性，横向上从十米级到百米级的非均质性都是应该考虑的。所有尺度上的非均质性皆控制着驱替效率。在垂直尺度上，地震普查相对来讲是比较模糊的，而试井数据只能对非均质性提供少量的信息，这些都使得井间预测变得困难。

考虑到这种不确定性及各种尺度上的变异性，应该使用随机地质模型，并应用于产量预报。特别需要指出的是根据岩相所建立的随机地质模型是十分有用的，这由于：

1）它们可以再生来自露头数据而获得的空间相关结构，而这些露头结构类似于所研究的油藏。

2）如果模拟是条件化的，它们匹配已知的观察数据。

3）它们只需较少的评价参数（一般地，旨在建立岩相的直接和交互变异函数）。

根据模型的选择和可用的数据，用随机建模技术可产生一系列可能的储层三维“图像”。这些储层图像的随机特征来源于由于井位信息稀少而引起内部分布广泛的不确定性。

1.5.2　从随机地质模型到确定的流动模型

首先可以假定储层有若干个不同的几何形态，然后对它们逐个进行研究。这些几何形态可通过不同的实现得到，然后保留参数值不变或同时改变这些参数（如岩性变异函数、物性特征等）。为了评价不确定性对地质模型的影响，更愿意得到在井中各个可能结果的整体分布。

1. 图像的选择（实现的选择）

为了减少流动模拟的数目，一种可考虑的方法就是识别有代表性的储层图像，它们可以给出极端情形下的产量剖面。已有学者建议进行连通性分析，通过几何度量来选择三种图像（高、中、低连通性）。也有学者提出了与井跨度有关的连通性概念，即“砂岩被井穿透的面积”和“相互联结的砂岩面积”等概念。这些都有利于选择储层随机图像。一旦选择了这些具有代表性的储层图像之后，就可进行进一步的储层分析，给出产量预报的可靠性和风险性

评价。

2. 物性数据的合成和流体流动的模拟

当考察一个三维储层图像时，可将其看成是在储层尺度范围内的“连续几何体”。事实上，所考察的储层模拟的块段体积比用于数值模拟中的块段体积细得多。因此又常讲随机建模的图像是一种高分辨率的图像。

为了把描述储层几何体的细段块（又称为地质块）组装成用于数值模拟的大块段（又称为储层块），必须把信息进行平均以便给用于数值模拟的储层块指定参数，此时可能要损失一些信息。

非可加变量的合成问题（如绝对渗透率、相对渗透率、毛管压力等）并不是一件容易的事情。已有许多学者提出了一些用于渗透率数据合成的有效方法（Gomez － Hernandez and Journel，1994；Pickup and other，1995）。如果得到了储层块的参数，那么就可进行相应的数值模拟并可给出产量预报。

1.6　应用于早期的评价到后期开发的整个过程

随机建模方法特别适用于解决油田开发过程中所遇到的这样一些实际问题如储层的早期评价以及进入开发阶段后，开发方案的选择等。

1.6.1　储层的早期评价阶段

把随机建模技术应用于早期评价阶段已受到足够重视。在储层早期评价阶段，只有少量的钻井信息，只有个别主要的可使用的描述特征：如一般的储层构造、流体接触面、流体组成、井的产能以及利用试井手段获得的储层渗透率估计等等。此时利用井位数据的地质统计学分析由于井数量少而难以实施。但是，储层的类型常常是可以辨认的，可类比的生产储层或露头资料可以使用。因而，就有可能考虑这些类比的地质层系所确定的参数并合理的运用于所研究的储层。这样就能使地质人员可把它们有价值的分析转化成对储层的定量描述。

储层评价中使用随机建模技术的最终目的，就在于它能比传统的方法更精确的模拟真实的储层分布，而且使用少量的数据和参数就可进行随机建模技术给地质人员和油藏工程人员提供了解决下列问题更为容易的办法：

1）使用随机方法探讨储层的连续性和不连续性。

2）更好的考察物性参数在空间的变异性。它可以给出高分辨率的地质模拟结果，然后使用一定的物性参数的平均化方法来得到可用于数值模拟的参数值。

利用随机建模技术对储层进行早期评价可以分析流体的驱动机制的影响和井跨度的影响。

随机建模技术用于储层的早期评价可以研究由沉积模式所确定的流动路径遮挡带的形成。（Haldorsen and Lake，1982；Hϕiberg and others，1990）更重要的是断层系统的刻画。在评价阶段，只有主要的断层才能被描述，次一级的断层，只有当获得更多的钻井资料和地震资料后才能被发觉。这些断层应该在储层的早期评价阶段通过建立随机模型并使用随机建模技术来研究。

1.6.2　开发方案的选择

油田处于开发阶段后，常可获得大量数据，如观察井和生产井位数据。此时地质人员和油藏工程人员必须使用下面两个工具：

1）岩相、物性参数等井位数据的地质统计学分析。

2）产量数据分析。

基于井位数据的地质统计分析可以告诉下面事实：利用诸如变异函数、交互变异函数等一些空间相关函数给出一些结果，这些空间相关函数是依赖于由跨度所决定的步长的。远离原点（步长=0）时候，空间相关函数的第一个实验点相应于平均井跨度，它具有重要的统计意义。由原点到第一个可观察的点变化范围之内，要直接确定空间相关函数是不可能的，因而，必须面临和需要解决下面两个问题：

1）由于空间相关模型是由现存的实验点导出的，因而要把它应用于更短的距离，必须给出一些假设，这些假设可由露头研究来验证。

2）由于空间相关模型的变程要比井的平均跨度小得多，此时模型就不可能正确导出。但是，可使用类似于储层早期评价阶段的思路，用可类比的储层或露头数据提取信息加入到模型之中去或把随机模型中的参数看成是未知的，并进行敏感性分析。

这里，敏感性研究（Sensitivity Study）包括：

1）改变随机模型参数（如频率、变异函数等）。

2）产生相应的储层图像。

3）对物性特征进行某种平均之后，进行流体流动模拟。

为了进行敏感性研究，就必须采用随机建模技术来产生满足已知参数（如频率、变异函数等）控制下的各种等可能的储层图像，并进行相应的产量模拟并与一些关键井的产量历史数据进行比较，以便选择能给出最优拟合的一套参数。这也说明了储层表征中产量数据所扮演的角色。产量数据可以增强储层表征的能力，这可以从以下两个方面来看：

1）历史拟合步骤验证选取的参数是否能够代表所研究的储层特征；

2）产量数据可直接参与控制随机建模的结果。

因此，一旦得到了一套最优的历史拟合参数，那么可通过改变注水井、生产井的位置、密度、跨度等参数，利用随机建模技术给出储层产量剖面并进而确定有利的储层开发方案。

1.6.3　随机建模用于油气储层评价的一般流程

通常假定沉积地层在油藏中可以被划分成若干时间带，对于它们的每一个可以分别进行建模和模拟。这些时间带由不同的时间曲面所分隔，而且把同一时期沉积的地质体连结在一起。时间曲面是比较容易识别的地质特征，比如薄的而且大范围分布的基质层位。可以进一步假设储层在初始沉积时是水平展布而且有同样的厚度。由于沉降、剥蚀和压实作用使得储层变成弯曲的、厚度变化的当前模样。这样可以建立相应的储层模型，然后利用获得的沉积相资料、露头资料、渗透率数据，通过下面完整的四阶段储层建模过程对储层进行评价：

1）地质构造模型。

2）沉积模型。

3）物性模型。

4）均一化模型。

储层评价的一般流程可用下列框图示意（图 1.1）：

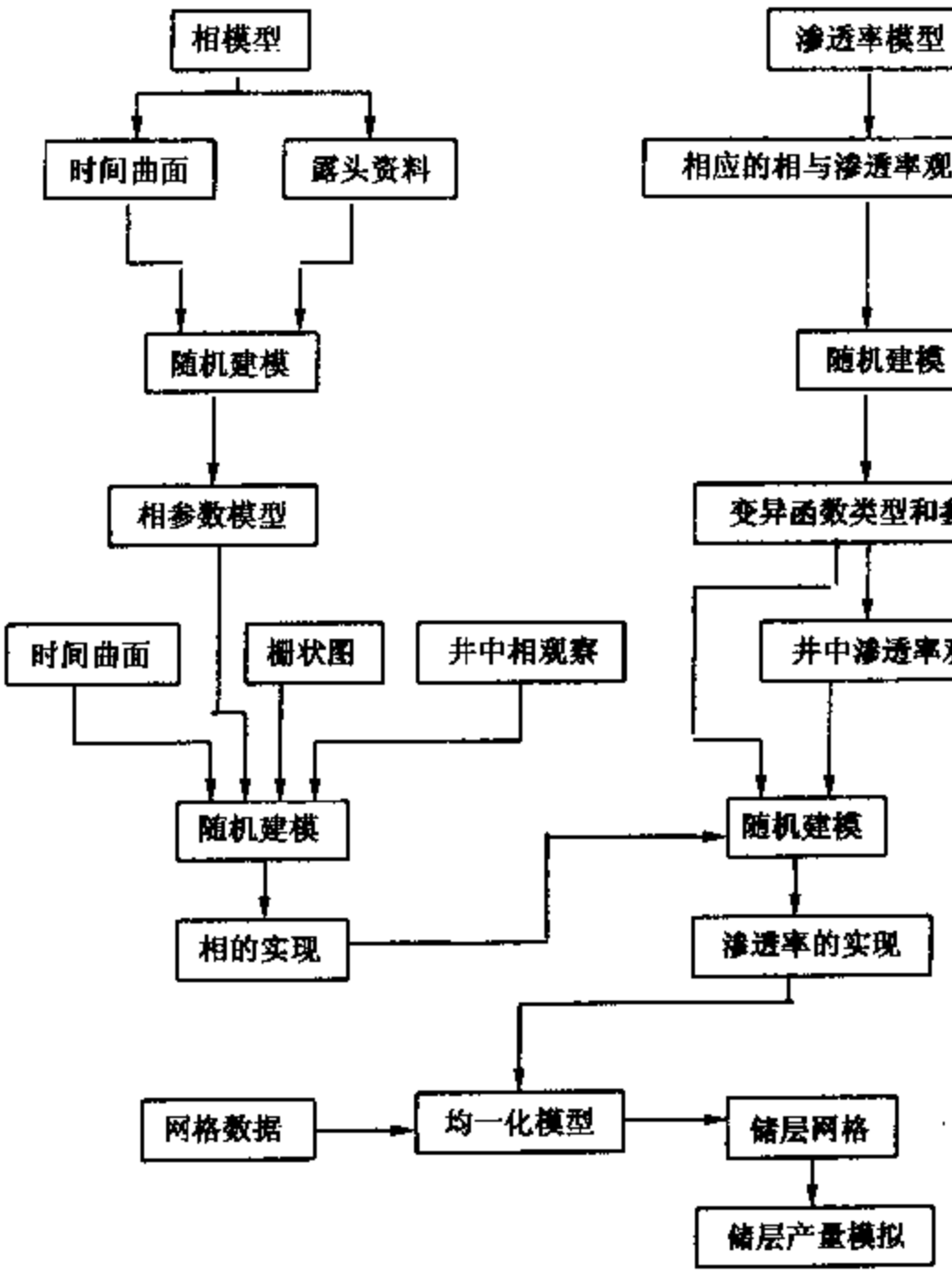


图 1.1　储层评价的一般流程

第 2 章　地质变量的空间变异性分析

地质变量是指在油藏范围内变化的地质参数或地质属性，如储层构造深度、厚度、储层物性参数孔隙度、渗透率、含油饱和度等在一个三维区域中的变化，储层沉积相带在空间的分布变化等。储层表征（Reservior Characterization）的最终目的，就在于尽可能充分地把地质学家的已知信息，地震数据，井的数据和其他类型的数据综合起来，对油藏中的地质变量的三维空间分布进行尽可能真实地刻画，从而为进一步油藏开发提供可靠依据。进行井间的预测是石油地质统计学的核心。它要求充分利用包含在已知数据中的信息进行建模和推断（Haldorsen and Damsleth，1990）。所有这些都与基于地质变量的空间分布是有一定联系的，信息之间的相关程度及其自身空间位置的联系使得统计推断成为可能。地质统计学中的协方差函数和变异函数是定量刻画地质变量空间相关性的重要工具，也是理解地质统计学和随机建模原理的基础。因此，本章将主要讨论协方差函数和变异函数的作用及意义。

2.1　随机函数，协方差函数和变异函数

2.1.1　随机函数

设 D 是研究的三维油藏区域，在 D 中任意位置 $\vec{u}$ 用点 (x,y,z) 表示。点 (x,y,z) 对应着一个地质变量 $z(\vec{u})$。值得注意的是，由于真实的油藏分布事先不清楚，所以认为地质变量 $z(\vec{u})$ 对于指定的 $\vec{u}$ 来讲对应于一个随机变量。这样，研究区域 D 中，相应地就拥有一个随机变量集合 $\{Z(\vec{u}):\vec{u}\in D\}$，称之为随机函数或随机场（Matheron，1963；Hijort and Omre，1990）。随机场 $Z(\vec{u})$ 和经典的随机过程是不同的。主要区别就在于随机过程的指标集 $t\in T$ 是一维的，而随机场 $Z(\vec{u})$ 中的指标集 $\vec{u}\in D$ 是二维或三维空间。这导致了地质统计学的研究方法和常规的随机过程研究方法有很大的不同，这些也恰好是地质统计学特色之所在（Hjort and Omre，1990）。

将地质变量的空间分布与随机函数联系起来是很重要的，这使得可以在概率和统计的框架中来研究储层的性质，从而使得可以对储层属性的井间分布不确定性进行定量研究，对勘探开发的风险性给予定量评价。值得指出的是，真实的油藏分布是唯一的，然而又是不完全清楚的。地质环境，地质沉积过程的复杂性导致了油藏分布的多样性，从而可以认为这种油藏演化过程是由某一种概率模型［指 $Z(\vec{u})$ 的概率模型］来支配的，真实的油藏是来自于该模型的一个特定的实现，记为 $Z(\vec{u})$。在一个简单的投硬币的例子中，如果投掷一枚两面均匀的硬币，它落下时迅速用双手盖住该硬币，此时也看不出国徽那一面朝上还是朝下。但是，可以推测国徽朝上的概率为 1/2。因此，硬币具体状态的确定性和唯一性并不妨碍对这一问题进行概率的描述。当然，真正的地质演化过程要比投掷硬币例子复杂得多。但是二者又有共同之处，那就是可将最终人们不能完全了解而又是确定的唯一状态置于概率模型之中，并将其看成是该模型的一个具体的实现。

在具体油藏描述中，常见信息往往是已知随机函数 $Z(\vec{u})$ 的一个实现 $z(\vec{u})$ 在某些位置的数据。这些已知的数据表示为 $\{z(\vec{u}_\alpha):\alpha=1,2,\cdots,n\}$，为已知数据集。油藏描述的最终

目的就在于如何利用这些数据再现油藏的真实分布或恢复整个实现 $\{z(\vec{u}):\vec{u}\in D\}$。因此，主要的任务就是如何利用已知的信息来建立 $Z(\vec{u})$ 的概率模型，进一步对井间分布进行预测。

如图（2.1）所示，假设研究的是储层的构造，那么知道在一些已知井位处的储层的构造深度 $\{Z(\vec{u}_\alpha):\alpha=1,2,\cdots,n\}$，要对整个储层构造深度进行预测。

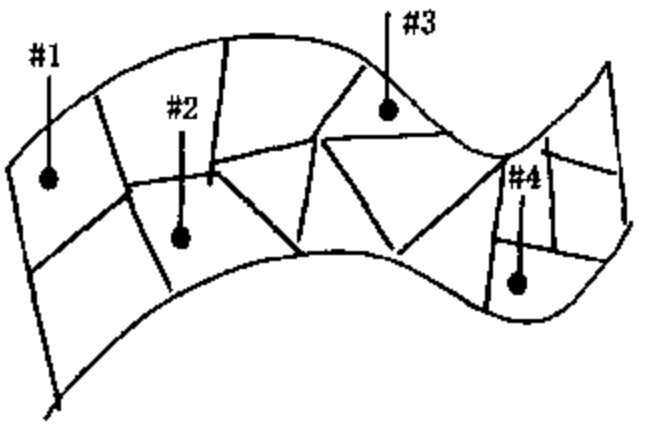


图 2.1　储层构造深度的预测

2.1.2　协方差函数和变异函数

对于给定的随机函数 $\{Z(\vec{u}):\vec{u}\in D\}$，将 $m(\vec{u})=E[Z(\vec{u})]$ 称之为该随机场的数学期望或均值。它反映了该随机场在空间的平均变化情况，有时又称之为趋势面。为了刻画该随机场空间相关性，引入协方差函数：

$$\begin{aligned}\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]&=E\{[Z(\vec{u})-EZ(\vec{u})][Z(\vec{u}+\vec{h})-EZ(\vec{u}+\vec{h})]\}\\&=E\{[Z(\vec{u})-m(\vec{u})][Z(\vec{u}+\vec{h})-m(\vec{u}+\vec{h})]\}\end{aligned}\tag{2.1}$$

该协方差函数反映了随机函数 $Z(\vec{u})$ 在滤掉趋势面后在空间两个位置 $\vec{u}$ 和 $\vec{u}+\vec{h}$ 值之间的相关程度或相似程度。具体来讲，若令

$$\varepsilon(\vec{u})=Z(\vec{u})-m(\vec{u})\tag{2.2}$$

表示随机函数 $Z(\vec{u})$ 滤掉趋势面 $m(\vec{u})$ 之后的误差。则协方差表达式（2.1）可改写成

$$\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]=E[\varepsilon(\vec{u})\cdot\varepsilon(\vec{u}+\vec{h})]\tag{2.3}$$

由此可见，当 $\varepsilon(\vec{u})>0$ 且增加时，若 $\varepsilon(\vec{u}+\vec{h})>0$ 且增大，即有 $\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]>0$ 且取较大的值。这说明 $Z(\vec{u})$ 在两个空间位置 $\vec{u}$ 和 $\vec{u}+\vec{h}$ 取值有正相关性。反之，若 $\varepsilon(\vec{u})$ 与 $\varepsilon(\vec{u}+\vec{h})$ 二者符号相反，且当一个绝对值增大时，另一个绝对值减小。若 $\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]<0$，且绝对值较大，从而说明 $Z(\vec{u})$ 在两个空间位置 $\vec{u}$ 和 $\vec{u}+\vec{h}$ 取值有负相关性。这说明了（2.1）式定义的协方差函数确实反映了随机函数在空间的相关程度。

如果假定随机函数 $Z(\vec{u})$ 在空间的变化是平稳的，即 $E[Z(\vec{u})]=m(\vec{u})=m$。此时均值为一常数 m，即 $Z(\vec{u})$ 在空间的变化没有明显的趋势，而是围绕着 m 值上下波动，此时（2.1）式可写成

$$\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]=E\{[Z(\vec{u})-m][Z(\vec{u}+\vec{h})-m]\}\tag{2.4}$$

即
$$\operatorname{cov}[Z(\vec{u}),Z(\vec{u}+\vec{h})]=E[Z(\vec{u})Z(\vec{u}+\vec{h})]-m^2$$

亦可写成
$$C(\vec{u},\vec{h})=E[Z(\vec{u})Z(\vec{u}+\vec{h})]-m^2\tag{2.5}$$

若进一步假定协方差式（2.5）是平稳的，那么 $C(\vec{u},\vec{h})=C(\vec{h})$，即协方差表达式不依赖于空间绝对位置 $\vec{u}$，$\vec{u}+\vec{h}$，而依赖于相对位置 $\vec{h}$。或者可以讲，协方差函数（2.5）具有空间的平稳不变性。这时，就有

$$C(\vec{h})=E[Z(\vec{u})Z(\vec{u}+\vec{h})]-m^2\tag{2.6}$$

（2.6）式就是在随机场 $Z(\vec{u})$ 的均值和协方差平稳假设下 $Z(\vec{u})$ 协方差表达式。这时，称为 $Z(\vec{u})$ 满足二阶平稳性假设。

在除了用（2.6）式刻画 $Z(\vec{u})$ 的空间相关性之外，地质统计学中另一个更常用的工具就是变异函数：

$$\gamma(\vec{h}) = \{E[Z(\vec{u})] - E[Z(\vec{u}+\vec{h})]\}^2/2 \tag{2.7}$$

分析（2.7）式，不难知道它揭示了 $Z(\vec{u})$ 在空间相距 $\vec{h}$ 的两个位置取值的相似性程度或变异性程度。当 $\gamma(\vec{h})$ 值较小时，说明二者取值较为接近，反之说明差异较大。因此，（2.7）式定量地刻画 $Z(\vec{u})$ 的空间变异性程度。易知在二阶平稳假定下，（2.6）和（2.7）有如下简单关系

$$C(\vec{h}) = C(0) - \gamma(\vec{h}) \tag{2.8}$$

该式说明了相关性程度度量和变异性程度度量之间的关系。

在地质统计学中，变异函数的角色更为重要。变异函数在评价解释方面比协方差函数来得更为容易。（2.6）和（2.7）显然是一种向量值函数，即 $C(\vec{h})$ 和 $\gamma(\vec{h})$ 依赖于向量 $\vec{h}$，这称之为各向异性模型，即在空间不同方向，相关性和变异性是不同的。在特殊情形下，即对应于各向同性的模型时，此时（2.6），（2.7）和（2.8）可写成：

$$C(h) = E[Z(\vec{u})Z(\vec{u}+\vec{h})] - m^2 \tag{2.9}$$

$$\gamma(h) = \frac{1}{2}E[Z(\vec{u}) - Z(\vec{u}+\vec{h})]^2 \tag{2.10}$$

$$C(h) = C(0) - \gamma(h) \tag{2.11}$$

这里，$h = |\vec{h}|$。

2.1.3　协方差函数和变异函数的解释

以下研究各向同性的变异函数模型：

$$\gamma(h) = \frac{1}{2}E[Z(\vec{u}) - Z(\vec{u}+\vec{h})]^2, h = |\vec{h}|, \vec{u} = (x,y,z) \in D$$

一般来讲，当 h 增大时，$Z(\vec{u})$ 和 $Z(\vec{u}+\vec{h})$ 差异增加，导致 $\gamma(h)$ 变大。但是二者差异不能无限增大。通常情况下，当 $h > a$（a 为某个常数）时，$\gamma(h)$ 将会基本保持在一个确定值。这揭示了当 $h > a$ 时，$z(\vec{u})$ 和 $Z(\vec{u}+h)$ 的相互独立性，即二者不具有影响关系。而当 $0 \leqslant h \leqslant a$ 时，$z(\vec{u})$ 和 $Z(\vec{u}+h)$ 二者有影响。若再结合下式 $C(h) = C(0) - \gamma(h)$，其中 $C(0) = \mathrm{Var}[Z(\vec{u})]$ 为随机函数的总方差。当 $h \leqslant a$ 时，可认为具有相关性；当 $h > a$ 时，此时可认为不具有相关性。此时，$C(h) = 0$，从而 $\gamma(h) = C(0)$。通常将 $C(0)$ 称为变异函数的基台值，a 称为变异函数的变程。$C(0)$ 刻画了随机函数 $Z(\vec{u})$ 在研究区域中的总的变差，而 a 刻画了 $Z(\vec{u})$ 的相关距离或影响范围。

值得指出的是，变异函数有时会在原点间断。对于有的随机函数，尽管定义了 $\gamma(0) = 0$，但当 h 趋于零时，$\gamma(h)$ 并不一定趋向于零。这样，变异函数在原点处就发生了间断。这样的随机函数不再具备均方意义下的连续性。从相邻的两点取得的值 $z(x)$ 和 $z(x+h)$，即使 h 相当小，其差别可能较大。$\gamma(h)$ 在原点的间断程度越大，这种差别也就会越大。

变异函数在原点处的间断性称为“块金效应”，相应的常数

$$C_0 = \lim_{h \to 0} \gamma(h) > 0$$

称为“块金常数”。“块金”这个术语是从金的颗粒分布不均匀而导致在很短距离内样品的品位有较大差异性的现象引申来的。块金效应可以由测量误差和随机函数的微观变异性引起。

图 2.2 和 2.3 反应了 $C(h)$ 和 $\gamma(h)$ 的变化和关系，分别称之为变异函数图和协方差函数图。

另一个反应相关程度关系的量是相关函数：

$$\rho(h) = C(h)/C(0) = 1 - \gamma(h)/C(0) \tag{2.12}$$

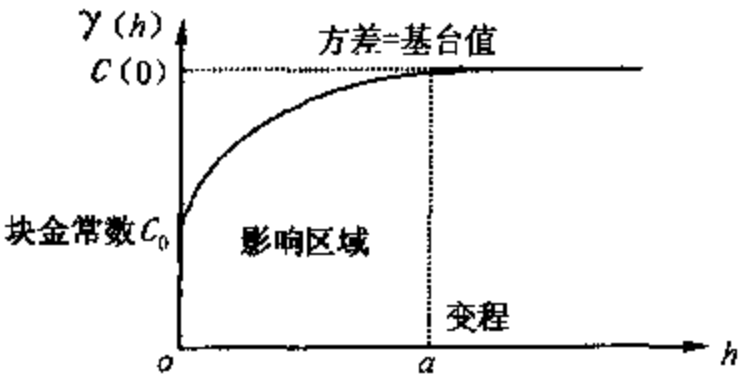


图 2.2　变异函数图

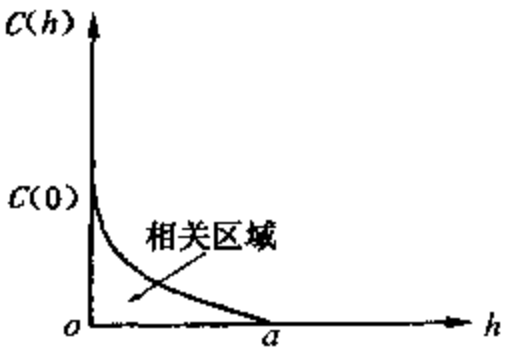


图 2.3　协方差函数图

它也反映了 $Z(\vec{u})$ 和 $Z(\vec{u}+\vec{h})$ 的相关性。它可以看成是对协方差函数 $C(h)$ 的正规化处理。显然 $|\rho(h)|\leqslant 1$，$\rho(h)$ 是一个无量纲的量。图 2.4 称之为相关函数图。

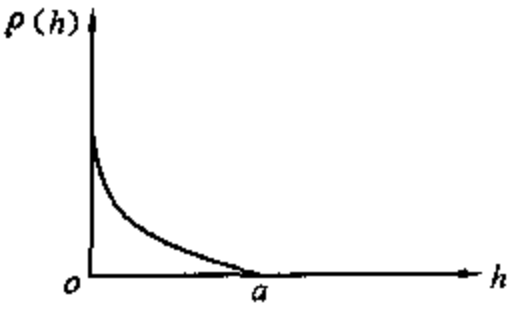


图 2.4　相关函数图

当进行多个变量协同估计或模拟时，此时需使用交互协方差函数或交互变异函数来刻画变量之间的相关性和变异性。如结合已知孔隙度和渗透率数据，对未知位置的渗透率数据进行估计或模拟时，此时未知位置的孔隙度数据可能是容易得到的。此时利用孔隙度和渗透率之间相关关系来预测未知位置渗透率值是很有用的。而此时就需要对孔隙度和渗透率相关或变异程度进行度量。因而前面的协方差函数和变异函数概念可进行推广。设 $Z(\vec{u})$ 和 $Y(\vec{u})$ 是定义在区域 D 上的两个随机函数。则称

$$C_{zy}(h)=\operatorname{cov}[Z(\vec{u}),Y(\vec{u}+\vec{h})]=E[Z(\vec{u})Y(\vec{u}+\vec{h})]-m_z m_y \tag{2.13}$$

为 $Z(\vec{u})$ 和 $Y(\vec{u})$ 的交互协方差函数。在此 $h=|\vec{h}|$，$\vec{u}$，$\vec{u}+\vec{h}\in D$。一般地，$C_{zy}(h)\neq C_{yz}(h)$，此时，可将某个随机函数所对应的协方差函数称为自协方差函数，以示与交互协方差函数的区别。

相应的交互变异函数如下：

$$\gamma_{zy}(h)=\frac{1}{2}E\{[Z(\vec{u})-Z(\vec{u}+\vec{h})][Y(\vec{u})-Y(\vec{u}+\vec{h})]\} \tag{2.14}$$

在此，$h=|\vec{h}|$，$\vec{u}$，$\vec{u}+\vec{h}\in D$。由此可见前面讲述的单个随机函数的变异函数（可称为自变异函数）是交互变异函数的特殊情形。类似地可以推导（2.13）和（2.14）之间的关系：

$$\gamma_{zy}(h)=C_{zy}(0)-\frac{1}{2}[C_{zy}(h)+C_{yz}(h)] \tag{2.15}$$

若定义交互相关函数 $\rho(h)$ 如下：

$$\rho_{zy}(h)=\frac{C_{zy}(h)}{\sqrt{C_y(0)}\sqrt{C_z(0)}} \tag{2.16}$$

则有

$$\gamma_{zy}(h)/(\sqrt{C_z(0)}\sqrt{C_y(0)})=\rho_{zy}(0)-\frac{1}{2}[\rho_{zy}(h)+\rho_{yz}(h)] \tag{2.17}$$

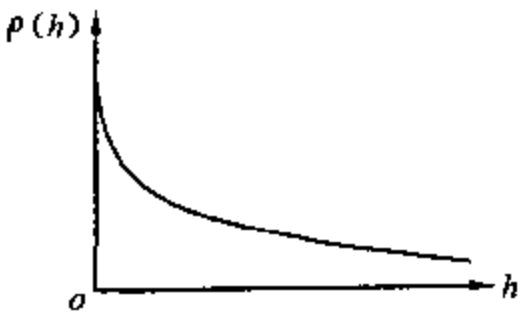


图 2.5　交互相关函数图

交互相关函数如图 2.5 所示，

2.2　理论变异函数模型

在具体问题中，若利用已知数据计算出实验变异函数或称之为理论变异函数模型的估计(参见下节)。然后根据实验变异函数，选择合适的理论变异函数模型。但是，并非任一条拟合曲线都可作为选择的变异函数模型，这是因为变异函数模型必须满足条件非负定性。对任一条曲线模型，要验证该条件非负定性并不是容易的事情。为此，从应用角度来考虑，经常选择四种理论变异函数模型的一个来作为需要的模型，这四种模型是满足条件非负定性的，且各自具有相应的参数，只需要根据实验变异函数和选定的模型来求取相应的模型参数就可以了。常用的四种理论变异函数模型为：球状模型，指数模型，幂函数模型和高斯模型。

2.2.1　球状模型

$$\gamma(h)=\begin{cases}C[(3h/2a)-\frac{1}{2}(h/a)^3]+C_0, & 0\leqslant h\leqslant a\\ C+C_0, & h>a\end{cases} \tag{2.18}$$

在此 C_0 为块金常数，C_0+C 为基台值，a 是变程。靠近坐标原点，该模型接近于直线形状。该模型称为可迁的，因为在有限的变程内，它可以达到基台值。该模型图示如图 2.6。

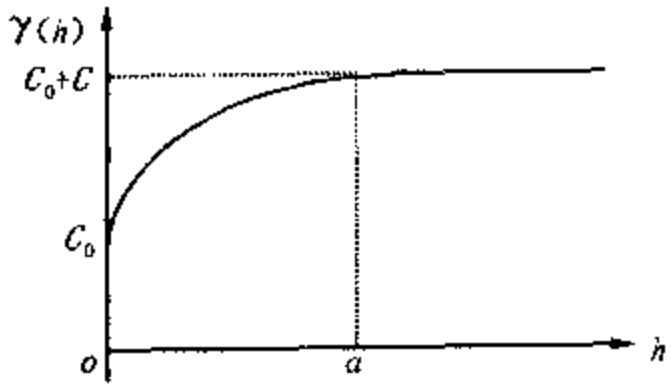


图 2.6　球状模型

球状模型是一种较常用的理论变异函数模型。该模型可做这样的解释，在三维空间中，相距为 h 的两个位置，他们的相互影响大小是以该两个位置为中心，做半径为 a 的两个球体，其相交部分的大小来度量的。很显然，当两个球体中心位置相距大于 a 时，说明两个位置之间不产生影响。该模型对于低于三维的空间也是适用的。

2.2.2　指数模型

$$\gamma(h)=C_0+C[1-\exp(-3h/a)]\qquad h\geqslant 0 \tag{2.19}$$

在此，C_0 为块金常数，C_0+C 为近似基台值。该模型的变程定义为 a，因为 $\gamma(a)=0.95C+C_0$ 很接近于基台值 C_0+C。该模型示意图如图 2.7。

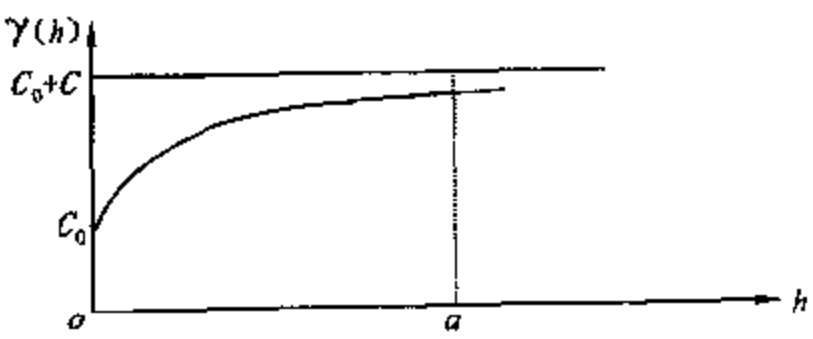


图 2.7　指数模型

2.2.3　幂函数模型

$$\gamma(h)=ch^{\alpha},0<\alpha<2,h>0 \tag{2.20}$$

该模型是不可迁的，因为 $\lim\limits_{h\to\infty}\gamma(h)=+\infty$。该模型所描述的随机函数 $\{Z(\vec{u}):\vec{u}\in D\}$ 常具有特殊性质，那就是具有随机分形特征，该随机场常具有高斯增量。在该模型中 α 与分形维数 D 有如下关系 $D=d+1-\alpha/2$，其中 d 为欧氏空间的维数。当 $\alpha=1$ 时，该幂函数模型成为线性模型，它对应于布朗运动函数。幂函数模型示意如图 2.8。

2.2.4　高斯模型

$$\gamma(h)=C_0+C\{1-\exp[-3(h/a)]^2\}\qquad h\geqslant 0 \tag{2.21}$$

在此，C_0 为块金常数，C_0+C 为近似的基台值，a 称为变程。这是因为 $\gamma(a)=0.95C+$

C_0。当 $C_0 = 0$ 时，该模型接近原点的形状类似于抛物线。这也是该模型与球状模型和指数模型的主要区别。这一特点保证了该模型刻画的随机函数具有较好的连续性。然而，在实际问题中，该模型的这一特点也常常导致了其不稳定的特点。高斯模型示意图如下（图 2.9）：

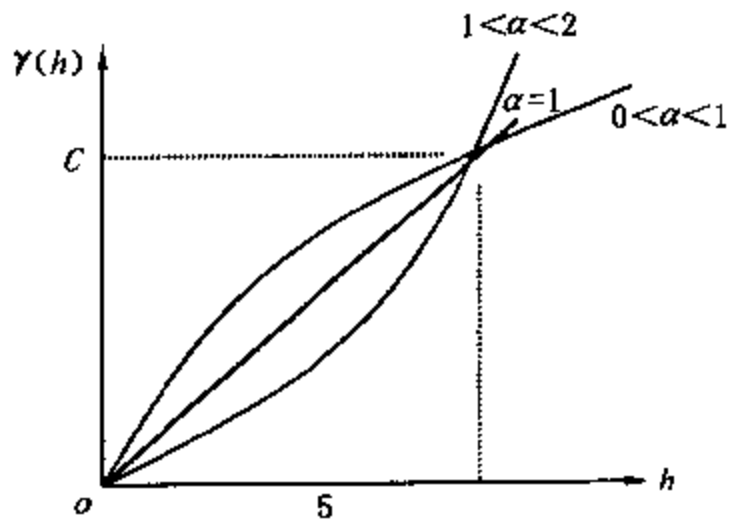


图 2.8　幂函数模型

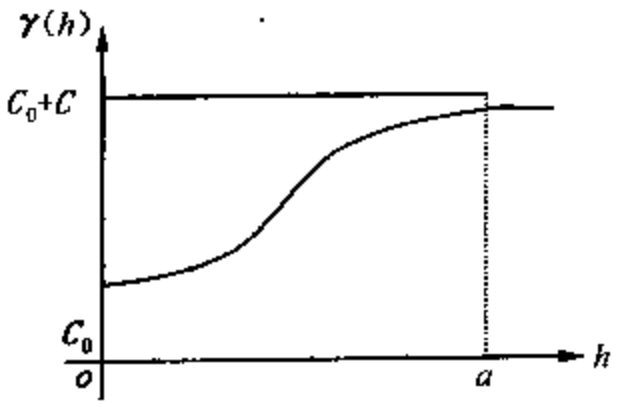


图 2.9　高斯模型

在实际问题中，计算出的实验变异函数常使得用一个单一的理论模型去拟合效果不能令人满意。此时根据实验变异函数特征并结合具体的地质背景，应用多个理论变异函数模型来拟合，这种方法称之为变异函数的套合。一般地，N 个各向同性理论变异函数模型 $\gamma_i(h)$，$i = 1,2,3,\cdots,N$。套合成一个各向同性的理论变异函数模型 $\gamma(h)$，可如下表示：

$$\gamma(h) = \sum_{i=1}^{N} \gamma_i(h)$$

值得指出的是，套合结构中不同的变异函数模型反映了不同的规模或尺度随机函数的相关结构，因而常具有相应的地质解释。如果没有明显的套合结构可用，则可选择一个理论模型来拟合。

2.3　各向异性的变异函数模型及其套合

在实际问题中，平面或空间中两个点的相互影响不仅依赖于两点的距离，更依赖于它们两点连线所处方向。因此，变异函数的计算经常依赖于两点连线所处方位，这就是各向异性。变异函数各向异性的物理背景在于地质体在沉积过程中常具有方向性或有利的延展范围。例如河流沉积的储层，物源方向或古水道方向常决定了河道的主要流向。因此，交切河道方向物性参数的连续性要差一些或变异性大一些。

在平面区域上，若在各个方向上的变异函数模型其基台值相同，而方向—变程图是一个椭圆状，则称之为平面几何各向异性。设 u 方向变程最大，令其为 a_u；v 方向变程最小，令其为 a_v。那么 u 方向称为变异函数的主方向，而 $a_v/a_u = k(k < 1)$ 称为各向异性比。如图 2.10 所示，坐标系 xoy 代表着指标集 $\vec{u} \in D$ 的所在的二维空间，D 为随机函数 $Z(\vec{u})$ 的定义域；而 θ 称为主方向的方位角。将坐标系 xoy 中 D 变换到坐标系 uov 中的一个区域，并将此新的区域视为随机函数 $Z(\vec{u})$ 的定义域，那么具有平面几何各向异性的随机函数 $Z(\vec{u})$ 就变成了各向同性的了。这就使二维变异函数的套合。这时，需要从坐标系 xoy 到坐标系 uov 的一个线性变换。

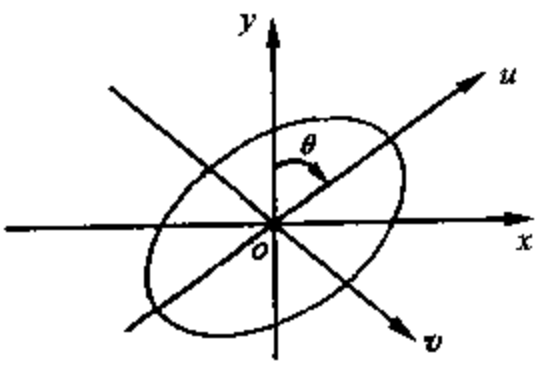


图 2.10　平面几何各向异性示意图

详情请参见“克里金地质绘图技术”（王家华等，1999）的5.3节“变异函数的套合结构”。

同理，现来讨论三维变异函数的套合。在一个三维区域中，若各个方向的变异函数模型，其基台值相同，而方向—变程图是一椭球状，则称之为空间几何各向异性。设 u 方向变程最大，与 u 方向垂直的 v 方向变程次之，而垂直于 u，v 轴的 w 轴变程最小（三个方向变程依次记为 a_u, a_v, a_w）。则称 $a_v/a_u = k_1(k_1 < 1)$ 为第一方向异性比，而 $a_w/a_u = k_2(k_2 < 1)$ 为第二方向异性比。在此引入三个角度 $\theta_1, \theta_2, \theta_3$；其中 θ_1 称为方位角，θ_2 称为倾角，θ_3 称为伏角。

三维空间 $oxyz$ 到三维空间 $ouvw$ 的一个线性变换，可使得具有空间几何各向异性的随机函数变为各向同性。具体来讲，这个线性变换可由如下的坐标系 $oxyz$ 三次旋转而成。

1）y 轴在 xoy 平面轴顺时针方向旋转 θ_1（主方向），此时保持 z 轴不动。所得的新坐标系为 $ox'y'z$（见图2.11（a））。

2）主方向 y' 轴在 $y'oz$ 平面内逆时针方向旋转 θ_3，此时保持 y'' 轴不动。所得的新坐标系为 $ox'y''z'$（见图2.11（b））。

3）z 轴在 $x'oz'$ 平面内顺时针方向旋转 θ_2，此时保持 y'' 轴不动。所得的新坐标系为 $ox''y''z''$（见图2.11（c））。

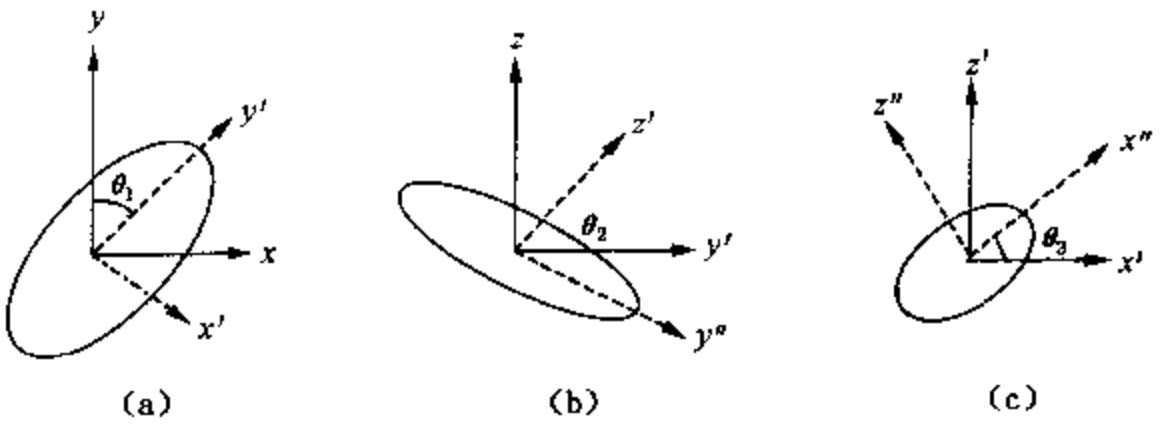


图2.11　空间几何各向异性

总之，旧坐标系 $oxyz$ 总可以分别根据三个角度 $\theta_1, \theta_2, \theta_3$，经过上述的三次旋转而到达任意一个新的坐标系 $ouvw$（Deutsch and Journel，1990）。

平面或空间上的非几何各向异性统称带状各向异性。在实际应用中必须将各向异性模型统一转化成一个各向同性模型。这种转化相当于坐标变换，使平面上的椭圆面转化成一个圆面，使空间的椭球转化成一个球体。

2.3.1　平面几何各向异性变异函数模型的转化

设 $P(x_1, y_1)$，$Q(x_2, y_2)$ 为平面上的两点，令

$\alpha = 90.0 - \theta_1$，$\mathrm{d}x = x_2 - x_1$，$\mathrm{d}y = y_2 - y_1$，$r_{11} = \cos\alpha$，$r_{12} = \sin\alpha$，$r_{21} = -\sin\alpha$，$r_{22} = \cos\alpha$，$\mathrm{d}x_1 = (r_{11} \cdot \mathrm{d}x + r_{12} \cdot \mathrm{d}y)$；$\mathrm{d}y_1 = (r_{21} \cdot \mathrm{d}x + r_{22} \cdot \mathrm{d}y)$

各向异性距离为：$h = \sqrt{(\mathrm{d}x_1)^2 + (\mathrm{d}y_1)^2}$。

此时，各向异性的变异函数模型就转化成了一个各向同性模型 $\gamma(h)$（具有与主方向理论变异函数模型相同的参数）

2.3.2　空间几何各向异性变异函数模型的转化

设 $P(x_1, y_1, z_1)$，$Q(x_2, y_2, z_2)$ 为空间两点，$\gamma(h)$ 为各向同性的理论变异函数模型，令 $\alpha = 90.0 - \theta_1$，$\beta = -1.0 \times \theta_2$，$\gamma = \theta_3$

$$r_{11} = \cos\beta \cdot \cos\alpha, r_{12} = \cos\beta \cdot \sin\alpha, r_{13} = -\sin\beta$$

$$r_{21} = (-\cos\gamma \cdot \sin\alpha + \sin\gamma \cdot \sin\beta \cdot \cos\alpha)/k_1$$

$$r_{22} = (\cos\gamma \cdot \cos\alpha + \sin\gamma \cdot \sin\beta \cdot \sin\alpha)/k_1$$

$$r_{23} = (\sin\gamma \cdot \cos\beta)/k_1$$

$$r_{31} = (\sin\gamma \cdot \sin\alpha + \cos\gamma \cdot \sin\beta \cdot \cos\alpha)/k_2$$

$$r_{32} = (-\sin\gamma \cdot \cos\alpha + \cos\gamma \cdot \sin\beta \cdot \sin\alpha)/k_2$$

$$r_{33} = (\cos\gamma \cdot \cos\beta)/k_2$$

$$\mathrm{d}x = x_1 - x_2, \mathrm{d}y = y_1 - y_2, \mathrm{d}z = z_1 - z_2;$$

$$\mathrm{d}x_1 = r_{11}\mathrm{d}x + r_{12}\mathrm{d}y + r_{13}\mathrm{d}z, \mathrm{d}y_1 = r_{21}\mathrm{d}x + r_{22}\mathrm{d}y + r_{23}\mathrm{d}z$$

$$\mathrm{d}z_1 = r_{31}\mathrm{d}x + r_{32}\mathrm{d}y + r_{33}\mathrm{d}z$$

$$h = \sqrt{(\mathrm{d}x_1)^2 + (\mathrm{d}y_1)^2 + (\mathrm{d}z_1)^2}$$

那么空间几何各向异性变异函数模型转化为各向同性变异函数模型 $\gamma(h)$。

2.3.3　几何各向异性变异函数模型的解释

对于平面几何各向异性函数模型来讲，表明对空间平面的任一位置 P，其对椭圆上任一点处的影响是一样的，或者讲，椭圆上任一点 Q 与 P 两点处变异是相同的，从而椭圆形状刻画了平面上研究属性值的影响范围。将各向异性变异函数模型化成各向同性的模型相当于

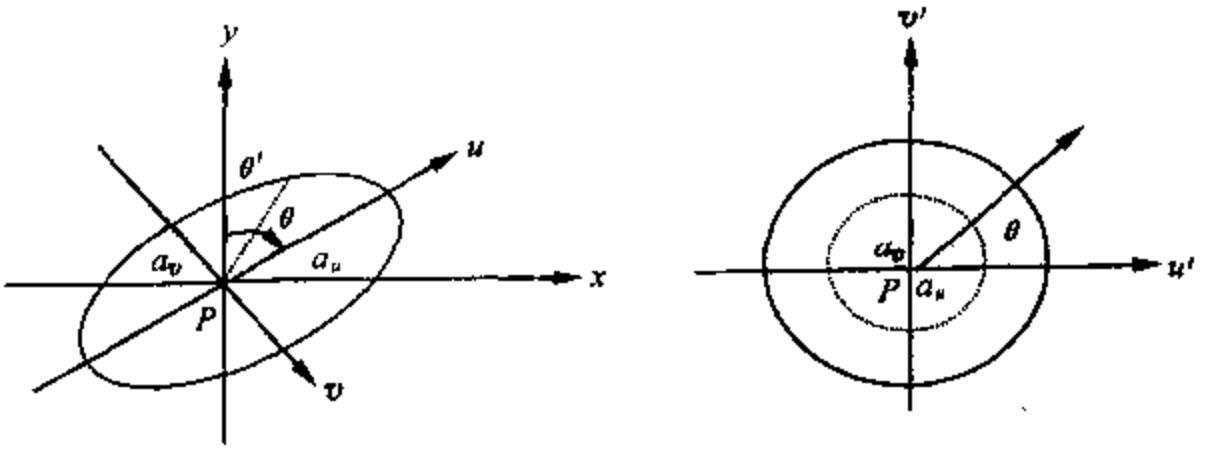


图 2.12　平面几何各向异性模型转化示意图

利用一个旋转及伸缩的变换矩阵

$$A = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$$

将椭圆变成圆，从而可以用一个各向同性的变异函数模型来替代各向异性的变异函数模型。

对于空间几何各向异性变异函数模型的解释是类似的，此时使用了旋转及伸缩矩阵

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

2.4　实验变异函数的计算及拟合

2.4.1　二维实验变异函数的计算

根据 (2.7)，理论变异函数为 $\gamma(\vec{h}) = \{E[Z(\vec{u}) - Z(\vec{u} + \vec{h})]^2\}/2$。在实际问题中，首

先要利用已知的信息（如地质，测井数据等）对理论变异函数进行估计。假设已知区域的井位图如下：

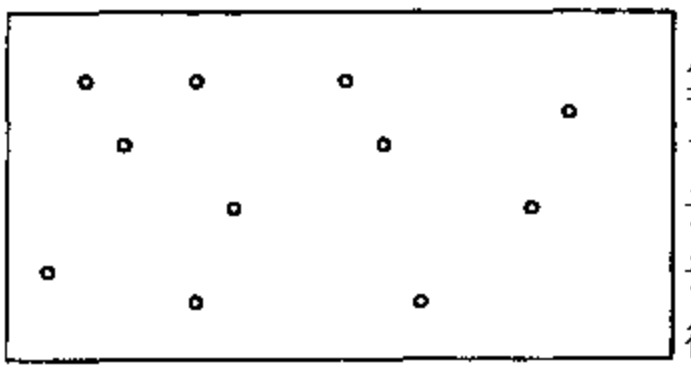


图 2.13　井位图

已知井位数据 $\{Z(x_i): i = 1,2,\cdots,N\}$。为了计算某个方向的实验变异函数，通常计算该方向上在若干不同距离的实验变异函数值。此时选取某个距离为计算实验变异函数的基本距离 L，称之为步长，分别计算 $L,2L,\cdots,mL$ 距离的实验变异函数值，此时就可得到 m 个实验变异函数点，m 称为步长个数。实际情况通常是不可能在精确的某个方向上，或精确的某个步长上能获得需要的点对数目来计算实验变异函数值，此时，一个变通的方法就是给方向一个容许的范围，称之为角度容限，只要点对方向不超出该容限就认为该点对可以参与到计算中来。同时，可以给每个步长一个容许范围，称之为步长容限，只要点对距离落入到该容限内，就认为该点对可以参与到计算中来。这样，可以给出计算某个方向相应于特定步长的实验变异函数。

$$\hat{\gamma}(h) = \sum_{i=1}^{N(\vec{h})}[Z(x_i) - Z(x_i + \vec{h})^2]/2N(\vec{h}) \tag{2.22}$$

在此 $h = |h|$，$N(h)$ 为落在特定方向容限及特定步长容限范围内的点对个数。

值得指出的是，随着步长的增加，虽然有时点对符合方向容限与步长容限，但是偏差将会增大，为此需要用偏离方向主线的一个固定宽度的带子来限制，使超出该范围的点对不参与运算，这样，有利于得到合理实验变异函数计算值。二维实验函数计算如图 2.14 所示。如此，可计算 $\gamma(L),\gamma(2L),\cdots,\gamma(mL)$ 的值，从而得到实验变异函数的散点图（图 2.15）。

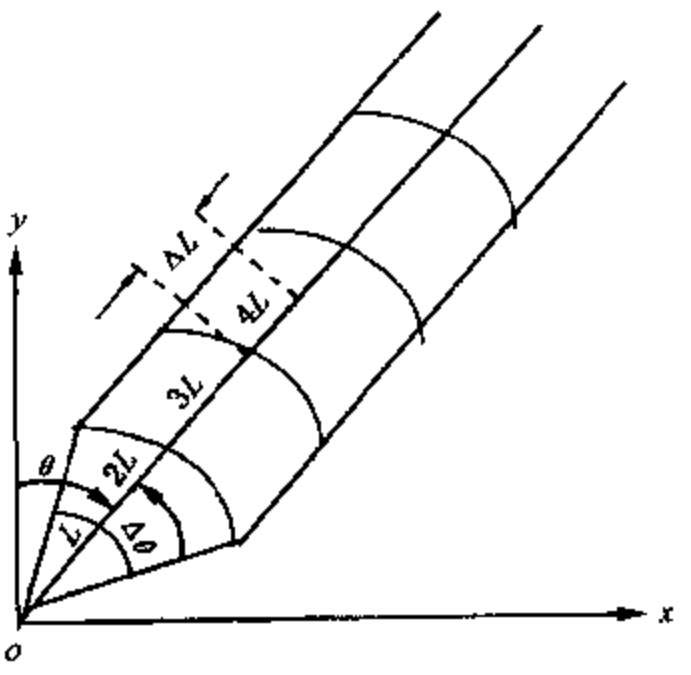


图 2.14　实验变异函数计算

对每个方向都可以按上述方法计算出实验变异函数。但是一般来讲，较为经济的方法就是计算两个方向（主方向，次方向）的实验变异函数即可，在此主方向的选择对应于区域的物源方向或物性持续性最好的方向，而次方向垂直由于主方向。

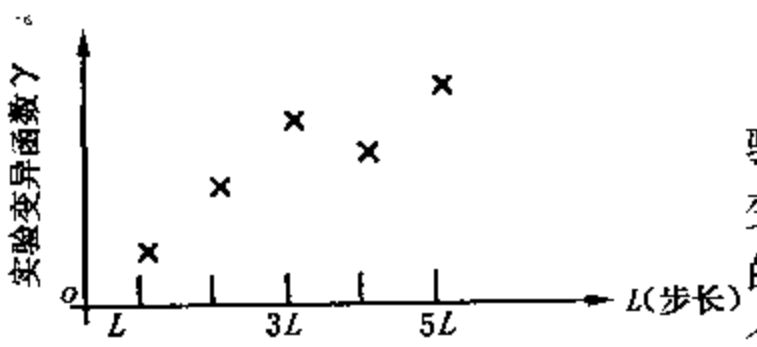


图 2.15　实验变异函数散点图

1．三维实验变异函数的计算

三维实验变异函数的计算其基本原理与二维实验变异函数基本相同。需要指出的是对于三维实验变异函数，在某个方向上的计算较为复杂的是方位的确定。在三维空间中，某个直线方向的确定由两个角度确定，如下图（图 2.16）。

因此，具体计算时由这两个角度来定方位，同时角度容限和步长容限及窗宽参数需相应指定。

2．实验交互变异函数及实验协方差函数的计算

类似于前面的实验变异函数计算，有时需计算实验交互变异函数及实验协方差函数，从

而给出交互变异函数及协方差函数理论模型。利用曾讲过的理论交互变异函数及协方差函数定义可知实验交互变异函数计算公式如下（见图 2.17）：

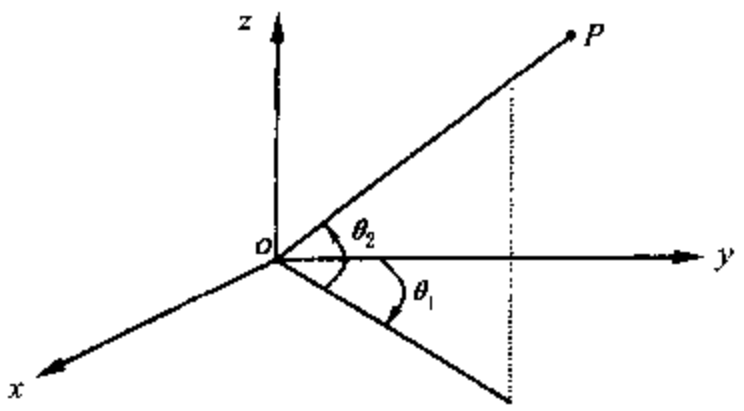


图 2.16　三维空间中一条直线方位的确定

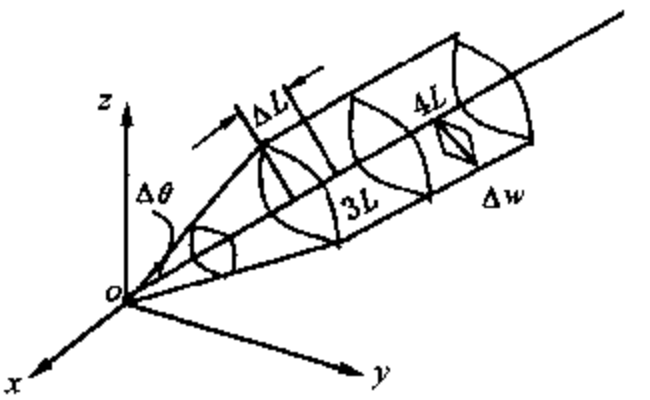


图 2.17　三维实验变异函数计算示意图

$$\hat{\gamma}_{zy}(h) = \frac{1}{2N(\vec{h})}\sum_{i=1}^{N(\vec{h})}[Z(x_i) - Z(x_i + \vec{h})][y(x_i) - y(x_i + \vec{h})] \tag{2.23}$$

在此 $\{z(x_i); i = 1,2,\cdots,N\}$ 和 $\{y(x_i); i = 1,2,\cdots,N\}$ 为已知的两个属性数据集。$h = |\vec{h}|$，$N(\vec{h})$ 为方向 $\vec{h}$ 上落入指定角度容限，步长容限的点对数目。

中心化实验协方差函数计算公式如下：

$$\hat{C}_{zy}(h) = \frac{1}{N(\vec{h})}\sum_{i=1}^{N(\vec{h})} z(x_i)y(x_i + \vec{h}) - m_z(-\vec{h}) \cdot m_y(\vec{h}) \tag{2.24}$$

在此 $N(\vec{h})$ 表示在方向 $\vec{h}$ 上落入指定角度容限，步长容限的点对数目。$z(x_i)$ 和 $y(x_i + \vec{h})$ 是相应点对位置处两个属性观察值。

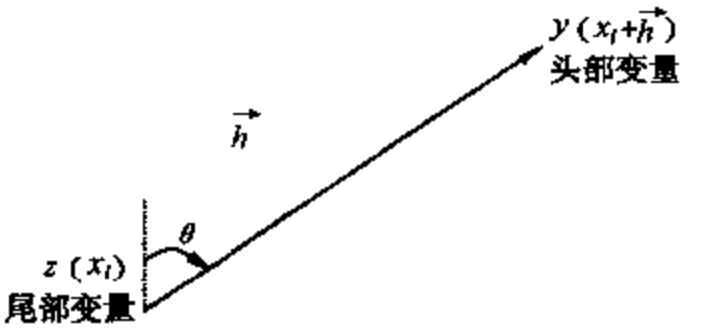


图 2.18　交互时的头尾变量示意

$m_z(h)$ 和 $m_y(h)$ 分别是尾部变量 z 和头部变量 y 均值的估计（见图 2.18）。

$$m_z(-\vec{h}) = \frac{1}{N(\vec{h})}\sum_{i=1}^{N(\vec{h})} z(x_i)$$

$$m_y(\vec{h}) = \frac{1}{N(\vec{h})}\sum_{i=1}^{N(\vec{h})} y(x_i + \vec{h})$$

类似地计算非中心化实验协方差公式如下：

$$\hat{C}_{zy}(h) = \{\sum_{i=1}^{N(\vec{h})}[z(x_i)y(x_i + \vec{h})]\}/N(\vec{h})$$

2.5　实验变异函数的拟合

一旦计算出了实验变异函数，那么就必须根据计算结果选取合适的理论变异函数模型。在几何各向异性情形，就必须选取几个特殊方向的实验变异函数进行拟合，并进行各向同性模型的转化。

2.5.1　二维实验变异函数的拟合

在二维情形，通常找出变程最大，最小的两个方向的实验变异函数并分别拟合。根据实验变异函数散点的分布情况，结合前面讲过的四种理论变异函数模型及合适参数，从而确定最终的理论变异函数模型。

在图 2.19 中，符号×表示 u 方向（主方向）的实验变异函数，符号△表示 v 方向（次方向）的实验变异函数，a_u、a_v 分别表示两个方向的变程，那么 $a_v/a_u = k(k<1)$ 表示各向异性比，C_0、C 分别为块金常数和拱高。

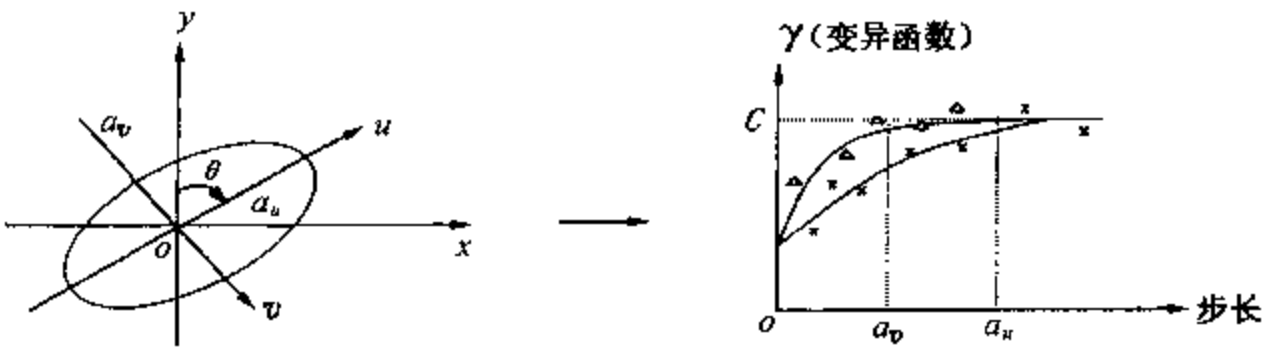


图 2.19　二维实验变异函数的拟合

值得指出的是，根据实验变异函数选择合适的理论变异函数模型是一个交互过程和反复试验过程。结合已知的各种理论变异函数模型的形状及实验变异函数散点分布特点，多次拟合并最终选取合适的理论变异函数模型。一旦获得了理论变异函数模型 $\gamma(h)$，主方向方位角 θ，主方向变程 a_u，次方向变程 a_v 及块金常数 C_0，拱高 C 等参数，则可利用前面讲过的方法将其转化成一个各向同性的理论变异函数模型。

2.5.2　三维实验变异函数的拟合

在三维情形，一般计算三个方向的实验变异函数，即主方向（此方向变程最长），次方向（变程次之），第三个方向（变程最短）的实验变异函数。这三个方向是相互垂直的，它们的变程相应于椭球的三个半轴。在几何各向异性的前提下，对这三个方向分别拟合选定的合适的变异函数理论模型和参数并转化为一个各向同性的变异函数模型。

图 2.20 中，符号×表示 u 方向的实验变异函数，符号△表示 v 方向的实验变异函数，符号○表示 w 方向的实验变异函数。

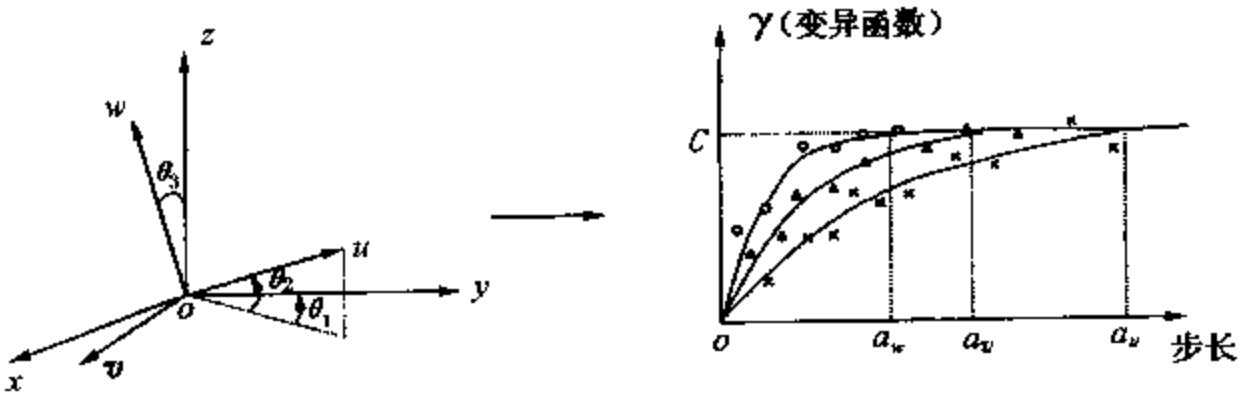


图 2.20　三维实验变异函数的拟合

若获得了新坐标系 $ouvw$ 下的三个角度 $\theta_1, \theta_2, \theta_3$，同时拟合了相应三个方向的理论变异函数模型及变程参数 a_u, a_v, a_w 及拱高 C，那么就有 $a_v/a_u = k_1(k_1<1)$，$a_w/a_u = k_2(k_2<1)$ 为两个各向异性比。可根据前述的方法，将它们转化为一个各向同性的三维空间的理论变异函数模型。

2.6　结构分析的一般步骤

根据已知的井位信息和数据，计算实验变异函数并拟合得到合适的理论变异函数模型过程称之为结构分析。本小节就结构分析的具体步骤以及一些参数选取的经验给予介绍。

1）选定二维或三维空间的研究区域，并了解指定区域的井位分布情况以及已知的井位信息，明确研究目标：依据已知信息对储层参数进行预测。若依据已知的单一井位属性推断未知位置的油藏属性，那么经常是求变异函数模型。如果是根据已知井位的若干个不同属性来推断未知的油藏位置的某个属性，那么经常是寻求交互的变异函数或协方差函数模型，此时，解决问题的方法是一种协同方法。

2）利用已知区域的地质知识，了解该区域的主要物源方向从而简化计算实验变异函数的求解过程，只需求解两个（二维情形）和三个（三维情形）主要方位的实验变异函数。一般来讲，单个步长可选为指定方位上的平均井距，步长数目 15～25 个。步长容限可选为1/2该方位上的井距，角度容限可选 $\pi/8$，带宽可选为两倍井距。根据这些参数可计算各个方位上的实验变异函数。

3）根据已经计算的两个或三个方向上的实验变异函数散点分布情况（以步长为横坐标，实验变异函数值为纵坐标），选取合适的理论变异函数模型及相应参数。最后，得到各向同性的变异函数模型。

在拟合过程中，对于短距离的实验变异函数值应给予重视，使拟合模型形状尽量照顾到这些点。因为这些点所提供的信息对空间变异性及预测结果有重要影响。另外需要注意的是，在大多数情形下，由于单个步长一般不会太小，从而对选取块金常数不能提供精确的信息。通常的做法是，若研究目标为区域上的物性参数变化情况，那么小的块金常数会告诉我们该物性参数具有很好的连续性。因此多数情形下，可置块金常数为零。变程的选取依据是当步长达到何值时，实验变异函数值基本稳定地在某个值范围邻近摆动，这个值就是拱高，而相应的步长就是变程。

4）当研究区域的井数目较少的时候，参与实验变异函数计算的点对数目也会很少，这些都影响到选取合适的理论变异函数模型。此时，可借用已有相似沉积特征的储层，若可类比的储层井位数目较多且容易选取理想的变异函数模型情况下，此时就可用该变异函数模型为模板加以使用。有时，露头信息或数据对于建立地下储层变异函数模型有很好地指导作用。

5）最后特别需要指出的是，结构分析是一个反复验证和交互的过程，它也是地质统计学分析的核心。如果预测结果不能够和地质认识很好相符的话，那么反过来检查和修改相应的变异函数模型也是必需的。因此，对变异函数模型及其每一个参数对预测结果的影响，了解得愈深刻，那么愈能得心应手地选取合适的理论变异函数模型。

第 3 章　随机模拟与地质统计学的基本原理

变量空间分析、克里金估计和随机模拟构成了地质统计学的三个主要基本内容。第 2 章关于协方差函数和变异函数的内容，目的就是研究地质统计学中的变量空间分析。克里金估计是一种空间变量的线性最佳无偏估计，在油藏描述的地质图件绘制中有很多应用。油气储层随机建模，则是地质统计学中的随机模拟方法在油藏描述和储层评价中的一种应用。

本章的主要任务是讨论随机模拟的基本概念。随机抽样是随机模拟的一个最基本的特点。在叙述随机模拟概念时，一定会涉及到随机抽样的理论。随机模拟和克里金估计在本质上有很大的差别，本章就此专门进行了讨论。对沉积相空间分布进行研究是储层随机建模的一个主要功能。对用离散变量表示的沉积相空间分布进行建模的方法，不同于一般孔隙度、渗透率等连续变量的随机建模。因此，本章开始就讨论了连续性地质变量和离散性地质变量的问题。

3.1　地质变量的分类

3.1.1　连续型地质变量

根据已知资料来预测储层属性在三维空间的分布和井间分布，是油藏描述和储层表征的目的。储层属性可以是多方面的，诸如利用测井资料求取的物性参数：孔隙度、渗透率、含油饱和度、泥质含量等，地震资料反映的波阻抗、声阻抗、振幅、反射时间等参数，地质资料反映的如地层厚度、矿岩厚度、有效厚度、断层数据、沉积相带的分布等信息，以及动态数据反映的信息。储层表征就是根据研究的问题和目的，结合已有的各种资料，从井筒及其地质信息或属性出发来推断储层属性在空间（井间）的分布。

储层的属性又称为地质变量。在地质统计学框架之内，其模型为随机函数（或随机场）。连续型地质变量是指在储层空间中分布具有连续性或持续性的地质变量。它们的取值范围可以充满一个区间，例如孔隙度、渗透率、含油饱和度、地震属性和地层厚度等，可归入连续型地质变量的范畴。该连续性随机变量或连续性随机函数可表示为 $\{Z(x)\mid x\in D\}$，其中 D 为研究的储层区域。此时，对任意 $x\in D$，$Z(x)$ 相应于一个连续型随机变量。连续型地质变量的空间变异性可以用前章中的变异函数来度量。

对连续型地质变量的研究在地质统计学中是比较早的，相对来讲方法多样且应用广泛。

3.1.2　离散型地质变量

离散型地质变量是相对连续型地质变量而言的。离散型地质变量是指只取有限个值的随机变量，又称之为范畴变量。例如储层空间的沉积相带分布，此时，由于可以用一个特定的整数值来指代一个特定的带相，那么该地质变量早期储层范围内只有有限个值。沉积相分布常用取值为离散的随机函数（或随机场）$\{L(x)\mid x\in D\}$ 来刻画，其中 x 是空间参照域 D 中的点，参照域 D 代表了储层的空间展布的范围，$L(x)$ 则代表了相的类型，也称为相的标签，常取值 $\{l_1,l_2,\cdots,l_n\}$。

例如对于砂泥岩两种相类型，此时 $L(x)$ 可取值 0 或 1，当 $L(x)=0$ 时，表明该位置

x 对应于砂岩，而当 $L(x)=1$ 时，表明该位置 x 对应于泥岩。整个研究区域 D 内相的分布可以用该离散型变量 $L(x)$ 来刻画。

又比如对于五种相带组成的储层，可以用离散的地质变量 $\{L(x)\mid x\in D\}$ 来刻画，$L(x)$ 可取值 1、2、3、4 和 5，分别表示泥岩、河道砂、点砂坝、决口扇和天然堤。

对于沉积相带这种离散型地质变量模型又常分为两大类。

1. 事件现象模型（Omre，1990）

当起控制作用的背景相支配下出现许多小尺度的另外一种相的单元时，应该采取事件现象模型。在以砂岩为主的储层中常伴随着小尺度泥岩单元，可称为砂岩背景下的事件现象。在以泥岩为主的储层中常伴随着许多小尺度的透镜状砂岩，可以称为泥岩背景下的事件现象。

2. 镶嵌现象模型（Tipper，1990；Omre，1990）

当各种不同沉积相互相重叠或互相镶嵌时，需要采用镶嵌模型。此时，没有背景相，例如，河流三角洲沉积中砂泥相的互相镶嵌。

3.1.3　地质变量的指示化方法

将地质变量进行指示化处理是非参数地质统计学的一大特点，特别值得指出的是美国 Stanford 大学的 Journel 教授在这方面作出了卓有成效的工作。随机建模主要分为两大类，一类是基于目标的模拟方法，另一类是基于象素的模拟方法。指示方法在基于象素的模拟方法中居重要地位。

1. 连续地质变量的指示化

当某一连续的地质变量在空间不满足高斯场条件时，此时对该随机函数建模就必须采用非参数化的方法。一般的处理步骤就是对随机函数的分布进行直接估计，设 $\{Z(x)\mid x\in D\}$ 是一连续的随机函数，给定若干个截断值 $z_1<z_2<z_3<\cdots<z_k$，那么对任一 $z_k(1\leq k\leq K)$，定义一个指示变量

$$I(x,z_k)=\begin{cases}1, & Z(x)\leq z_k\\ 0, & Z(x)>z_k\end{cases}\tag{3.1}$$

那么 K 个指示变量 $I(x,z_k)(1\leq k\leq K)$ 实际上是定义在 D 上的 K 个随机函数。对于空间研究区域 D 上任一位置 $x\in D$，则经验分布函数 $F(z)$，可由 $F(z_k)=P[Z(x)\leq z_k](1\leq k\leq K)$ 来确定，但是不难推知

$$P[Z(x)\leq z_k]=E[I(x,z_k)]\qquad(1\leq k\leq K)\tag{3.2}$$

因此 $F(z_k)$ 可由指示变量的数学期望来推断，这一点是相当重要的，也是指示化方法的核心。

指示化方法之所以称为非参数化方法，就在于它直接对随机函数分布模型进行估计，而不必强加其他假设。而高斯场模型是一种参数化模型，它只由两个参数（均值函数及协方差函数）来确定，它具有运算速度快，处理方便的特点，但是它又有适应而不广泛的特点。而指示化方法正好与高斯场模型形成互补。

2. 离散型地质变量的指示化

前面已讲过离散型地质变量 $\{L(x)\mid x\in D\}$，其中 $L(x)$ 可能取值只有有限个 l_1，$l_2,\cdots,l_n$。对应于离散型地质变量可以相应地定义 K 个指示变量

$$I_k(x)=\begin{cases}1, & L(x)=l_k\\ 0, & L(x)\neq l_k\end{cases}\qquad(1\leq k\leq K)\tag{3.3}$$

此时，一个指定变量 $I_k(x)$ 相应于对范畴 l_k 的刻画。例如在沉积相的情形，则 $I_k(x)$ 描述了位置 $x \in D$ 处是否存在相 l_k。当 $I_k(x) = 1$ 时，则位置 x 存在相 l_k，否则不存在相 l_k。

因此，将离散型随机变量进行指示化处理后，可得定义在研究区域 D 上的 K 个随机函数 $\{I_k(x) \mid k = 1,2,\cdots,K\}$。对这 K 个随机函数的研究就相当于对离散随机函数进行处理。特别地有

$$P[L(x) = l_k] = P[I_k(x) = 1] = E[I_k(x)] \tag{3.4}$$

该式亦建立了位置 x 处存在相 l_k 的概率与相应指示化变量 $I_k(x)$ 数学期望之间的关系。

3.2　地质统计学基本工具

随机函数 $\{Z(x) \mid x \in D\}$ 的建模及模拟最终常归结为空间研究区域 D 内任一指定位置 x 处的单变量 $Z(x)$ 的分布。因此对单变量（称为边缘随机变量）的统计分析是基本而且重要的。本节就给出单个随机变量统计分析的基本工具。

3.2.1　均值、方差、极差和变异系数

设 Z 是一个随机变量，其分布函数 $F(z)$，那么

$$EZ = \int_{-\infty}^{+\infty} z \mathrm{d}F(z) \cong m \tag{3.5}$$

称为总体数学期望或均值，而

$$\mathrm{Var}(Z) = E[Z - EZ]^2 = \int_{-\infty}^{+\infty} (z - EZ)^2 \mathrm{d}F(z) \cong \sigma^2 \tag{3.6}$$

称为总体 Z 的方差，σ 称为 Z 的标准差。

设 $Z_{\max}$ 和 $Z_{\min}$ 分别为表示随机变量 Z 的最大、最小值，那么

$$R = Z_{\max} - Z_{\min} \tag{3.7}$$

称为 Z 的极差。而

$$V = \frac{\sigma}{EZ} \tag{3.8}$$

称为随机变量 Z 的变异系数。

若已获得容量为 N 的一组样本 $\{z_1, z_2, \cdots, z_N\}$，则总体 Z 的数学期望 m、方差 σ^2、极值 R 和变异系数 V 可进行相应估计。

$$\hat{m} = \left(\sum_{i=1}^{N} z_i\right)/N \tag{3.9}$$

称为样本均值。它是对总体 Z 均值 m 的估计。

$$\hat{\sigma}^2 = \left[\sum_{i=1}^{N} (z_i - m)^2\right]/(N-1) \tag{3.10}$$

称为样本方差。其是对总体 Z 方差 σ^2 的估计。

令 $z_{\max} = \max\{z_1, z_2, \cdots, z_N\}$，$z_{\min} = \min\{z_1, z_2, \cdots, z_N\}$，那么

$$\hat{R} = z_{\max} - z_{\min} \tag{3.11}$$

称为样本极差。它是对总体极差 R 的估计。而

$$\hat{V} = \hat{R}/\hat{m} \tag{3.12}$$

称为样本变异系数。它是对总体变异系数的估计。

3.2.2　直方图、比例曲线和经验分布函数

1. 直方图

直方图是地质统计学中的一个最基本的工具，它实际上是对随机函数一维边缘分布概率密度的估计。设 $\{z_1, z_2, \cdots, z_N\}$ 是 $Z(x)$ 在 D 内的 N 个采样，在平稳假设下，将 $[Z_{\min}, Z_{\max}]$ 均分成 M 类，每个子类区间长度为 $\Delta = (Z_{\max} - Z_{\min})/M$，而 $[Z_{\min} + (m-1)\Delta, Z_{\min} + m\Delta]$ 为第 m 个子类区间，计算

$$P_m = P\{[Z_{\min} + (m-1)\Delta, Z_{\min} + m\Delta]\}, (1 \leq m \leq M) \tag{3.13}$$

其中，$P\{[Z_{\min} + (m-1)\Delta, Z_{\min} + m\Delta]\}$ 代表样本落入第 m 个子类区间的比例。P_m 实际上是对随机变量取值落入第 m 个子区间概率的估计。

以 Z 为横坐标 P 为纵坐标可得直方图（图 3.1）。从直方图可以直观地分析地质属性值 Z 分布的总体特征，如均值、方差、极差等其它变异情况。从直方图形态可大概了解 Z 分布的类型，如是否接近正态、对数正态等其它分布类型。值得指出的是直方图分析仅提供随机函数 $\{Z(x) \mid x \in D\}$ 的一维边缘分布信息。

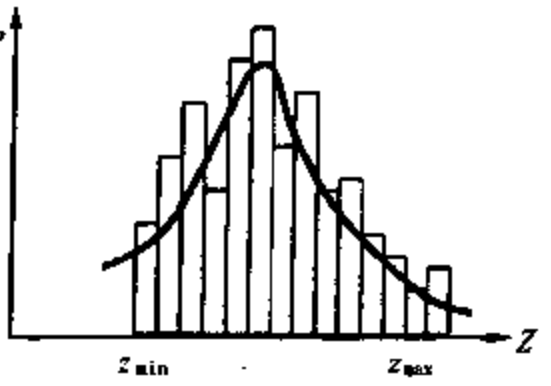


图 3.1　直方图

对离散型地质变量 $\{L(x) \mid x \in D\}$，$L(x)$ 可能取值 $\{l_1, l_2, \cdots, l_n\}$，那么根据 N 个采样的样本，此时选取分类数目 $M = N - 1$，借用直方图可获知每个范畴 $l_k (1 \leq k \leq K)$ 在空间 D 所占据的比例。对于沉积相的研究，该比例就是对应相占据比例的估计。

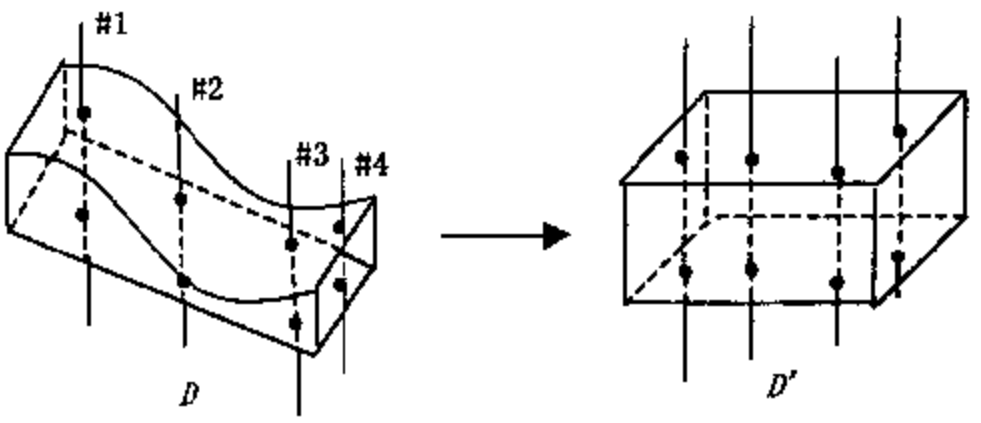


图 3.2　层模型的归一化处理

2. 比例曲线

比例曲线是直方图的推广（Matheron，1987）。为了介绍比例曲线的概念，先介绍实际空间区域 D 到模拟空间区域 D' 的变换。

（1）区域坐标变换

对三维储层的地质统计建模与模拟过程中，经常需要进行一种坐标变换。其物理背景就在于真实的储层由于受到地质过程，如压实、剥蚀、构造等作用的影响，经常变成弯曲的层状模样，称之为层模型。而直角坐标系下，三维空间具有箱形结构。通常的变异函数分析或其它统计分析是在直角坐标系下进行的。因此，将层状模型恢复成箱形的模拟区域，对于真实地分析和统计地质变量的相关结构至关重要。这一过程相应于层面恢复过程（见图 3.2）。

如不考虑剥蚀而只考虑压实和构造作用的影响，那么坐标变换为直井情形

$$x' = x, \quad y' = y, \quad z' = (z - z_{\text{Top}})/(z_{\text{Bot}} - z_{\text{Top}}) \tag{3.14}$$

在此 (x, y, z) 为某个直井轨迹中相应的一个空间位置 (x, y, z)，而 Z_{Top}，Z_{Bot} 为相应层面在该直井中对应的顶底深度，则 (x', y', z') 为变换后新的坐标。由于 $0 \leq z' \leq 1$，因此这种变换又常称为层模型的归一化处理。

如考虑剥蚀和斜井情形，坐标变换要进行适当的修改。

一旦完成了模拟区域 D 内的统计分析后，可经过相应的反变换得到实际区域 D 内的地质统计分析结果。

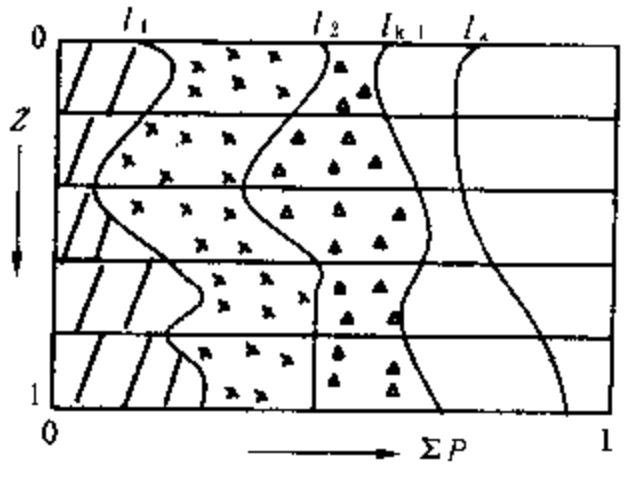


图 3.3　比例曲线

（2）比例曲线的计算

在研究区域 D 内，根据地质研究获知相带（岩相或沉积相）在空间不同深度位置占据的比例是变化的。为了再现这种变异性，通常做法是在模拟区域 D' 内，用水平的且相互平行的若干平面将 D 分割成层状的块段，分别计算每一小块段内各种相占据的比例（直方图分析）。这样，每一小块段内部可以计算出各种相占据的比例，并累加显示。此时以水平方向作为比例值，竖直方向作为深度值，则可以得到 K 条比例曲线。每条曲线相当于指定相在不同深度空间位置占据的比例变化趋势（如图 3.3）。

值得指出的是一种相占据的比例是用相应深度处相邻两条比例曲线之差来确定的。

由比例曲线的变化可以显示不同相带在不同深度位置分布的非平稳性。本书后面将介绍的截断高斯模拟方法以比例曲线作为其主要的输入信息。

（3）经验分布函数

设 Z 为一单个随机变量，其分布函数为 $F(z)$，则由定义

$$F(z) = P(Z \leq z), (-\infty < z < +\infty) \tag{3.15}$$

可知分布函数是累积概率。概率分布 $P(z)$ 和分布函数 $F(z)$ 之间是一一对应的。在离散型地质变量情形，用概率分布 $P(z)$ 描述其概率通常是方便的。对连续地质变量，用分布函数 $F(z)$ 来描述其概率特征比较方便。

当总体 Z 的分布函数 $F(z)$ 未知时，此时应该用采样 $\{z_1, z_2, \cdots z_N\}$ 对函数 $F(z)$ 进行估计，定义

$$\hat{F}_N(z) = \frac{\sum \{z_i \leq z\}}{N} \tag{3.16}$$

为经验分布函数。其中 $\sum \{z_i \leq z\}$ 表示 N 个样本中不超过 Z 的样本的个数。$F_N(z)$ 是总体分布函数 $F(z)$ 估计量。以图 3.4 为经验分布示意图。

3.2.3　分位数、P-P 图、Q-Q 图

1. 分位数

设 Z 是一个随机变量，$F(z)$ 是其分布函数。任给一个数 $p(0 < p < 1)$，若能确定一个数 z_p，使

$$P = P(Z \leq z_p) = F(z_p) \tag{3.17}$$

则称 z_p 为随机变量 Z 的 p 分位数。由此可见，p 分位数 z_p 是这样一个数，随机变量 Z 取值小于该值的概率为 p。

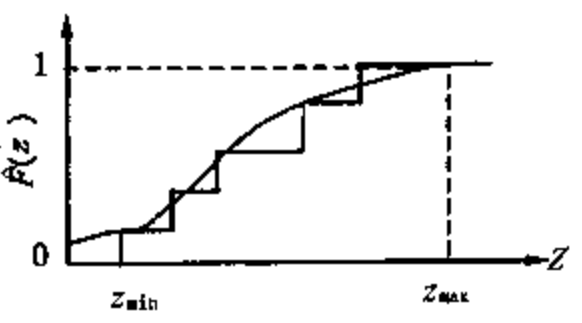


图 3.4　经验分布函数示意图

从而亦写成 $z_p = F^{-1}(p)$。特别地，当 $p = 1/2$ 时，$z_{0.5}$称为中位数。$z_{0.25}, z_{0.75}$也是重要的分位数。图 3.5 是分位数示意图。

由于对任一个 $p(0 < p < 1)$，可对应相应的分位数 z_p，由此可定义分位数函数 $q(p) = F^{-1}(p)$。因此分位数函数 $q(p)$ 是分布函数 $F(z)$ 的反函数（图 3.6）。

因此，可定义中位数：$M = q(0.5)$，偏态符号：$\text{sign}(m - M)$。

在此 m 为数学期望，当偏态符号为正，说明分布是正偏的，当偏态符号为负，说明分布是负偏的。中位极差：$IR = q(0.75) - q(0.25)$。

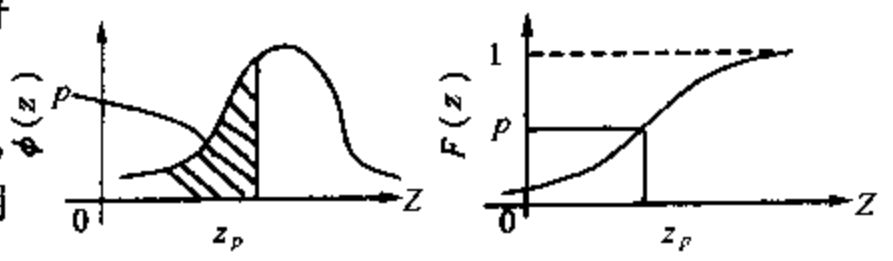


图 3.5　分位数示意图

2.P－P 图和 Q－Q 图

对两个单变量分布 $F_1(z)$ 和 $F_2(z)$，若要比较它们是否相同时，经常使用 P－P 图和 Q－Q 图进行比较。

P－P 图制作过程为，取一系列值 $z_k(k = 1,2,\cdots,K)$，以 $F_1(z_k)$ 为横坐标，$F_2(z_k)$ 为纵坐标，作散点图。若该散点图近似于一条直线，那么说明两个分布 $F_1(z)$ 和 $F_2(z)$ 具有相同的形状。该散点图称为 P－P 图（图 3.7）。

Q－Q 图制作过程为，取一系列值 $p_k(k = 1,2,\cdots,K), 0 < p_k < 1$，以 $q_1(p_k)$ 为横坐标，$q_2(p_k)$ 为纵坐标，作散点图。若该散点图近似于一条直线，那么说明两个分布 $F_1(z)$ 和 $F_2(z)$ 具有相同的形状，该散点图称为 Q－Q 图（图 3.8）。

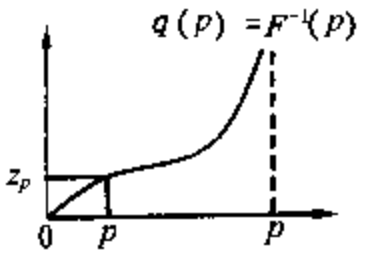


图 3.6　分布函数 $F(z)$ 的反函数

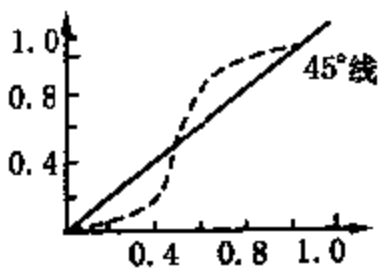


图 3.7　P－P 图

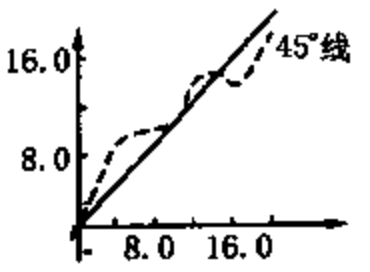


图 3.8　Q－Q 图

3.3　估计与模拟

3.3.1　先验分布与后验分布

地质统计学关心的是“对于时空中不规则波动现象的研究”。地质统计学提供一系列工具，目的在于理解空间变异性并对其进行建模。

设考虑一个分布在空间区域 D 内的储层属性是 $Z(x)$，即 $\{Z(x) \mid x \in D\}$。用前面谈到的概念，即考虑一个随机函数（Random Function）。若没有任何采样数据的话，则可以根据地质概念模型或其他可以类比的信息，对某个特定的位置 x，建立相应的该位置 $x \in D$ 的属性分布模型。

$$F(x,z) = \text{Prob}\{Z(x) \leq z\} \tag{3.18}$$

该式（3.18）实际上是位置 x 的随机变量（Random Variable）RV 的分布函数。有时，它又称为累积概率分布函数 cdf。

式（3.18）表达了对位置 x 处，储层属性不确定性的度量。若获得了采样数据集，$\{z(x_\alpha) \mid \alpha = 1,2,\cdots,n\}$，则该数据集就提供了对采样位置 x 的信息。此时分布（3.18）将要重新修正为

$$F[x,z \mid (n)] = \text{Prob}\{Z(x) \leq z \mid (n)\} \tag{3.19}$$

此时式（3.19）称为条件化的已知数据集的分布函数，或称为条件累积概率分布函数，记为 $ccdf$。式（3.19）中的（n）表示“条件到已知数据集 $\{z(x_\alpha) \mid \alpha = 1,2,\cdots,n\}$”。(3.19)

表达了在获得采样数据后对未抽样位置不确定性的新认识。

(3.18) 称为未抽样位置 x 处，储层属性 Z 的先验概率模型，而相应的 (3.19) 称为后验概率模型。对未抽样位置处储层属性的统计推断经常基于后验概率模型 (3.19)。上述思想可以表达成：先验信息＋数据⇒后验信息。

这也是贝叶斯 (Bayes) 统计推断的核心。先验信息常来自于地质概念模型或地质学家的经验及其它有关信息，又称之为软信息，而数据又常称为硬信息。上述思想反映了软信息和硬信息互相结合在地质统计学推断中所起的作用 (Journel，1986；Omre，1989)。

需要指出的是，地质统计学更关心地质变量在空间不同位置的相互依赖和相互影响。因此，单个未抽样位置的概率分布的推断应基于空间多个位置联合概率模型，即整个随机函数 $Z(x)$ 概率模型。所以，一个随机函数 $Z(x)$ 可用任何数目 K 以及 K 个位置 $x_k(k=1,2,\cdots,K)$ 的所有 K 元条件累积概率分布函数集合来表达：

$$F(x_1,x_2,\cdots,x_k;z_1,z_2,\cdots,z_k)=\text{Prob}\{Z(x_1)\leq z_1,\cdots,Z(x_k)\leq z_k\} \tag{3.20}$$

该式是对 K 个随机变量 $Z(x_1),Z(x_2),\cdots,Z(x_k)$ 联合不确定性的度量。同样地，若已知采样数据集 $\{z(x_\alpha)\,|\,\alpha=1,2,\cdots,n\}$，那么 (3.20) 可以条件化到后验的联合条件累积概率分布函数

$$F[x_1,x_2,\cdots,x_k;z_1,z_2,\cdots,z_k\,|\,(n)]=\text{Prob}\{Z(x_1)\leq z_1,\cdots,Z(x_k)\leq z_k\,|\,(n)\} \tag{3.21}$$

很显然，(3.21) 是位置 $x_1,x_2,\cdots,x_k$，样本容量 n，样本几何构型 (采样位置为 $x_\alpha,\alpha=1,2,\cdots,n$) 及样本值 $\{z(x_\alpha)\,|\,\alpha=1,2,\cdots,n\}$ 的函数。

在范畴变量的情况下，此时 $RFZ(x)$ 可能的取值是有限集 $\{1,2,\cdots,L\}$，那么可以有类似的情况，如此时 (3.18) 可以写成

$$F[x,k\,|\,(n)]=\text{Prob}\{Z(x)=k\,|\,(n)\} \tag{3.22}$$

这是未抽样位置 x 处，$RFZ(x)$ 的离散的条件概率分布。(3.21) 可以写成：

$$F[x_1,x_2,\cdots,x_k;z_1,\cdots,z_k\,|\,(n)]=\text{Prob}\{Z(x_1)=l_1,\cdots,Z(x_k)=l_k\,|\,(n)\} \tag{3.23}$$

在此，$l_1,l_2,\cdots,l_k$ 在集合 $\{1,2,\cdots,L\}$ 中取值。

地质统计学中，大多数与未抽样位置 $Z(x)$ 有关的信息来自于邻近位置 x 的样本，这些样本可以是同一属性或其它有关的属性。因此，更一般的情形是对于多种属性的相关性建模。这是协同建模的思想。

3.3.2　随机建模与随机模拟

储层随机建模就是建立储层属性的概率分布模型 (3.21) 或 (3.23)。该模型反映了属性空间分布的概率特征及联合的不确定性。有了这种模型，就可以对储层任何未抽样位置 x 的属性 $Z(x)$ 的分布不确定性进行刻画。它可用如下分布来描述：

$$F(x,z)=\text{Prob}\{Z(x)\leq z\} \tag{3.24}$$

$$F[x,k\,|\,(n)]=\text{Prob}\{Z(x)=k\,|\,(n)\} \tag{3.25}$$

它们反映了未抽样位置 x 属性 $Z(x)$ 分布的不确定性。

(3.24) 所反映的概率密度曲线如图 3.9。

该曲线形态刻画了未抽样位置 x 处属性不确定性的大小。若方差越小，不确定性越小。在极端情况下，若整个油藏任何位置处属性皆已知的情况下，则曲线 (3.9) 退缩成一点，

从而不确定性为零，即完全确定。

随机建模方法，一般分为参数化和非参数化两种。参数化方法指假定模型类别，仅推断模型参数。如可假定模型为高斯模型，此时仅推断均值函数和协方差函数。高斯模型具有稳定和计算速度快的特点，因此经常作为连续型地质变量随机建模的首选模型。当然，在使用高斯模型前，还应对其能否使用该模型做适当的检验。非参数方法指直接推断分布模型。如随后讲到的指示模拟方法就是一种非参数化的随机建模方法，它直接推断模型（3.24）和(3.25)。该方法的优点是适应面广，但缺点就是计算量大，且在具体应用中还应做一些其他的处理。根据建模本身的技巧和特点，又可将随机建模方法分为高斯建模、指示建模、截断高斯建模和示性点过程建模等方法。

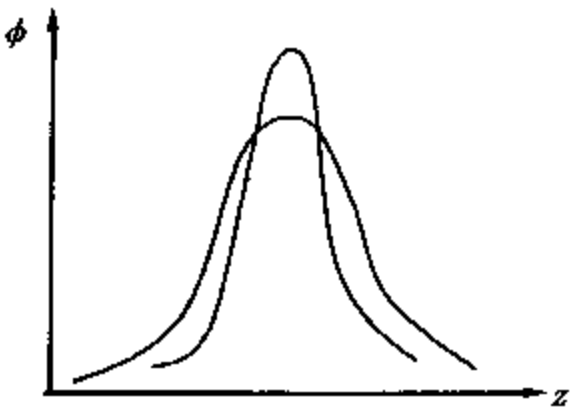


图 3.9　概率密度曲线

一旦建立了模型（3.23）和（3.24），就可采用相应的随机模拟方法（Stochastic Simulation）产生来自于模型的储层属性空间分布各种变化的等可能实现。随机模拟是一个抽样过程，该过程抽取等可能的，等概率的来自随机模型的各个部分的联合实现，一系列实现（通常用网格化数据表达）$\{Z^{(l)}(x)\,|\,x \in D\}, l = 1,2,\cdots,L$，表示区域 D 中属性值 $Z(x)$ 空间分布的 L 个可能图像。每个实现也被称为随机图像（Stochastic Imaging），它们反映了强加在 $Z(x)$ 随机模型的某些性质。

随机模拟通常分为条件模拟和非条件模拟。非条件模拟所产生的随机图像仅要求再现随机模型所要求的关于储层属性空间分布的相关结构，而条件模拟还要求条件化到已知的井位观察位置必须与已知数据相一致。通常讲的随机模拟一般指条件模拟。

3.3.3　估计与模拟的本质区别

1. 估计

估计方法就是，利用已知数据 $\{Z(x_i)\ \ i = 1,2,\cdots,n\}$ 进行滑动平均，给出未抽样位置处相应的内插值。因此估计方法又常称为内插方法。传统的估计方法有距离反比加权法与三角剖分法等。自从 60 年代初期，法国的 G.Matheron 教授提出区域化变量理论并创立地质统计学以来，克里金估计方法已得到了广泛的应用。以下以普通克里金方法为主讨论克里金估计的基本思想。

设 $\{Z(x)\,|\,x \in D\}$ 是研究的油藏空间属性，将其看成是分布在空间范围的二阶平稳随机场。若已获得了观察数据 $\{Z(x_i)\,|\,i = 1,2,\cdots,n\}$，则可利用已知数据估计未知位置 $x \in D$ 处的值 $Z(x)$。若利用线性平均

$$Z^*(x) = \sum_{i=1}^{n} \lambda_i Z(x_i), \lambda_i (i = 1,2,\cdots,n) \tag{3.26}$$

λ_i 是权系数。它依赖于不同位置 x。这相当于用随机变量 $Z^*(x) = \sum_{i=1}^{n} \lambda_i Z(x_i)$ 来预测随机变量 $Z(x)$。为了得到一个好的预测，必须采取一种衡量标准。那就是无偏性

$$E[Z^*(x) - Z(x)] = 0 \tag{3.27}$$

和方差最小

$$E[Z^*(x) - Z(x)]^2 = \min \tag{3.28}$$

由式（3.27）知 $\sum_{i=1}^{n} \lambda_i = 1$。利用（3.28）并使用拉格朗日乘数法可得克里金方程组

$$\begin{cases} \sum_{i=1}^{n} C(x_i - x_j)\lambda_i - \mu = C(x - x_j), (j = 1,2,\cdots N) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{3.29}$$

在此，$C(h)$ 为随机场的协方差模型，μ 为拉格朗日乘数因子。求解克里金方程组（3.29）可得到权系数 $\lambda_i(i = 1,2,\cdots,n)$，从而可给出未知位置属性的克里金估计。

通过分析，克里金估计原理具有如下特点：

1）克里金方法区别于其他传统估计方法关键之处，就在于它不仅考虑了已知点与待估计点的影响，而且也考虑了已知点之间的相互影响，即强调数据构型的作用。不同位置的相互影响大小是用协方差（或变异函数）来定量描述的。

2）从克里金方程组可以看出克里金估计通过已知数据点，故克里金估计方法是严格的内插方法。

3）利用最小均方误差（3.28）知，克里金估计实际上是对条件数学期望 $E\{Z(x)|Z(x_i), i = 1,2,\cdots,n\}$ 的估计，或者讲是对（3.25）式均值的估计，在线性无偏类中是最优的。若将高斯场模型强加给 $\{Z(x)|x \in D\}$，则易知克里金估计在所有无偏估计类中是最优的。

因此，克里金估计实际上是一种线性平滑的低通滤波器，实质上是对数学期望的估计，从而具有光滑效应。对变化平缓的地质参数，使用克里金估计，并以等值图形式来直观表达，结果会合理美观，可以反映平均变化趋势。

2. 模拟

随机建模和随机模拟技术是地质统计学中继克里金估计技术之后，迅速发展的一个新工具。它的起源和发展主要是由于油藏工程领域越来越重视对油藏非均质的认识和评价，认为油藏非均质性是影响提高采收率的重要因素。这种非均质性必须用新的方法来刻画和定量描述。克里金估计方法由于是一种光滑的内插方法而不能定量评价储层空间分布的非均质性，克里金估计给出的是一种单一的数值模型，是一种对平均值的估计，所以不能反映隐含在随机场概率模型中整体的相关结构。而随机模拟通过建立油藏属性的概率模型，再施行抽样过程，抽取各种相互独立的，等概率的来自模型各个部分的联合实现，一系列实现的差异就反映了储层属性空间分布的非均质性和不确定性。

模拟的结果强烈地依赖于建立的概率模型。随机建模过程必须充分利用各种模型的地质数据并强调以地质概念模型为基础。正由于此，随机模拟不象克里金估计那样，有统一的公式和算法。实际上，不难发现克里金方法本身不与模型发生关系，它仅利用协方差或变异函数这个矩信息。对于不同地质问题，需要建立不同的概率模型。随机建模本身是一门艺术，寻找一种快速适用的模拟算法对于特定的问题是至关重要的。

因此，随机模拟有以下特点：

1）模拟算法是依赖于模型的，不同模型要采用不同模拟算法。

2）随机模拟结果强调结果的整体相关性。它从整体上对储层属性空间分布提供了不确定性的度量。克里金方法不能保证条件化到统计量。如用克里金方法得到的结果，其直方图和协方差与原始数据计算结果会有很大偏差。

3）克里金方法只给出一个数值结果，而随机模拟能给出多种数值结果。这些结果的差异正好反映了隐含在概率模型之中的不确定性。

3.4　抽样基本理论

随机模拟最终归结为对抽样位置储层属性的单个随机变量的抽样或样本的产生。本节就讲述常用的单个随机变量样本的产生方法。

3.4.1　[0，1] 上均匀分布随机数的产生

产生来自于区间 [0，1] 上的均匀分布的随机变量的样本是一切随机模拟的基础。设 U 是一个随机变量，它服从 [0，1] 上的均匀分布，常记为 $U \sim U[0,1]$。均匀分布的分布函数 $F_U(u)$ 如下定义

$$F_U(u) = \begin{cases} 0, u < 0 \\ u, 0 \leq u \leq 1 \\ 1, u > 1 \end{cases}$$

其概率密度函数及其分布函数示意如图 3.10。

产生 [0，1] 上均匀随机数的方法常用线性同余发生器

$$X_i = (aX_{i-1} + c) \bmod (2^n);$$

$$u_i = X_i / 2^n; \qquad (i = 1,2,\cdots) \tag{3.30}$$

在上述公式中 a，c 为两个常数且为两个整数，a，c，n 决定了随机数发生器的质量，常取 $n=30$ 或 $n=31$。

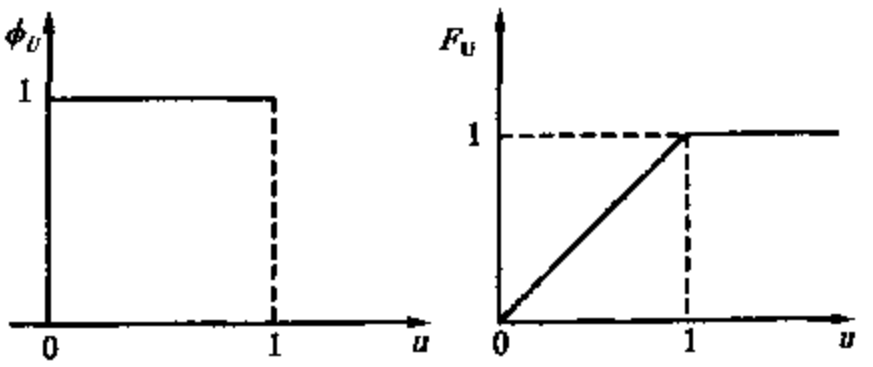


图 3.10　均匀分布概率密度函数 $\varphi_U(\cdot)$ 及其分布函数示意图 $F_U(\cdot)$

初始值 X_0 常称为随机种子，一般取为奇整数。由此可见，当指定随机数发生器种子，可产生相应一列 [0，1] 上随机数 $\{u_i | i = 1, 2, \cdots\}$。关于这种随机数发生器的使用，已有常见的算法和子程序供选择（Ripley，1988）。

3.4.2　产生来自于连续型随机变量的样本

设 Z 是一个连续型随机变量，Z 的分布函数 $F_Z(z)$。则 Z 经过其分布函数 $F_Z(z)$ 变换后新的随机变量 $U = F(Z)$ 服从 [0，1] 上的均匀分布。事实上当 $u < 0$ 时，有 $F_U(u) = P(U \leq u) = P[F_Z(z) \leq u] = 0$。

当 $0 \leq u \leq 1$ 时，有

$$F_U(u) = P(U \leq u) = P[F_Z(z) \leq u] = P[Z \leq F_Z^{-1}(u)] = F_Z[F_Z^{-1}(u)] = u。$$

当 $u > 1$ 时，有 $F_U(u) = P(U \leq u) = P[F_Z(Z) \leq u] = 1$。从而 $U = F(Z)$ 服从 [0，1] 上的均匀分布。

这样得到关系 $U = F(Z)$ 和 $Z = F^{-1}(U)$，只要产生来自于均匀分布 [0，1] 上的随机样本 $u_i (i = 1,2,\cdots)$，就可得到来自于 Z 的样本 $z_i = F^{-1}(u_i), i = 1,2,\cdots$。二者的关系如图 3.11。

在实际应用中，有时用非参数化方法获得了经验分布函数 $F_n(z)$，它是对总体分布函数 $F(z)$ 的估计，而 $F_n(z)$ 是一个阶梯函数。为此，将 $F_n(z)$ 内插成一个连续的函数。常用的内插方法是幂函数内插法，即当 $i = 1,2,\cdots,n, z_0 = z_{\min}, z_n = z_{\max}$ 时。

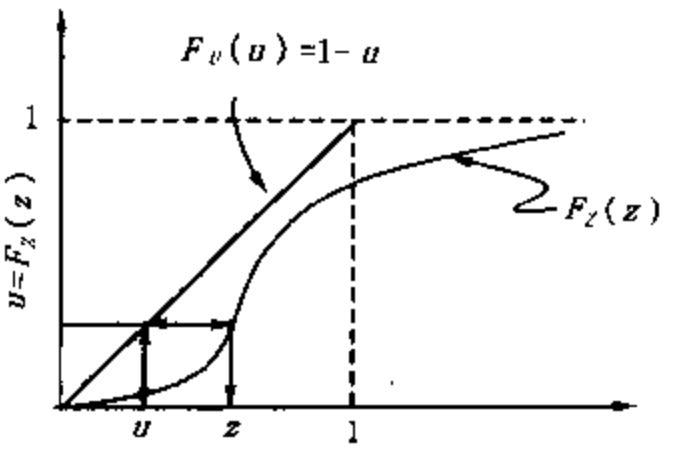


图 3.11　连续型随机变量分布与均匀分布的对应

令 $\lambda_{z_{i-1},z_i}^{(\omega)}=\left[\dfrac{z-z_{i-1}}{z_i-z_{i-1}}\right]^{\omega}$，$\omega>0$，则可置

$$\hat{F}_n(z)=\hat{F}_n(z_{i-1})+\hat{F}_n(z_i)\cdot\lambda_{z_{i-1},z_i}(\omega) \tag{3.31}$$

$\omega=1$ 相应于线性内插（图 3.12）。

3.4.3　产生来自于离散型随机变量的样本

设 Z 是一个离散型随机变量，可能的取值为 1,2，…，K。Z 的概率分布为 $P(Z=i)=p_i$，则为了产生来自于 Z 的一个随机样本，首先产生一个来自于［0，1］上的均匀随机数 u，考察 u 落入下列 K 个区间中的一个：

$$[0,p_1],(p_1,p_2+p_1),\cdots,\left(1-\sum_k^{K-1}p_k,1\right)$$

若落入第 k 个区间，则获得 Z 的一个样本 k。这样通过产生一系列均匀随机数，可得一系列来自于 Z 的样本。

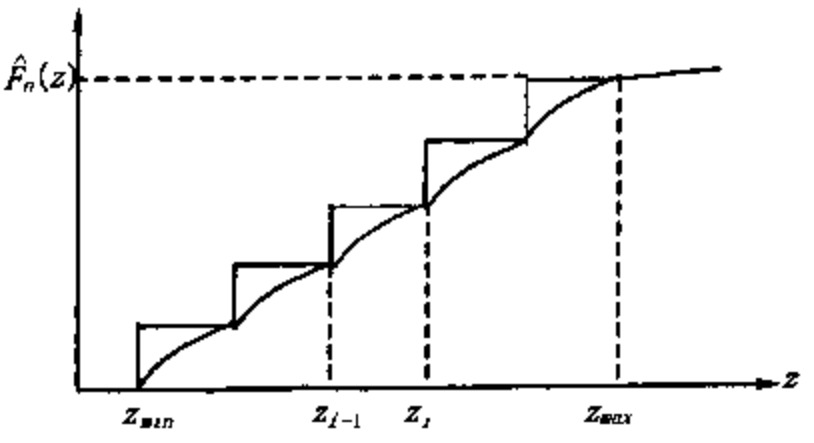


图 3.12　经验分布函数的内插

3.4.4　产生来自于正态分布 $N(\mu,\sigma^2)$ 的样本

正态分布 $N(\mu,\sigma^2)$ 作为常用连续型随机变量的分布，经常需要产生来自于该分布的样本。设 $Z\sim N(\mu,\sigma^2)$，μ 为分布的均值，σ^2 为分布的方差。

设 $Z_1\sim N\ (0,\ 1)$，即服从标准正态分布的随机变量，那么 $Z=\mu+\sigma\cdot Z_1$。因此只需要来自于标准正态分布随机变量 Z_1 的样本，就获得了 Z 的样本。利用 3.4.1 的结论设 Z_1 的分布函数为 $G\ (\cdot)$，则 $Z_1=G^{-1}\ (U)$。其中，U 是服从［0，1］均匀分布的随机变量。因此，先产生一个来自均匀分布的随机样本 u，再求 $G^{-1}\ (u)$ 可得到来自于 Z_1 的样本，从而可得到 Z 的样本 $Z=\mu+\sigma\cdot Z_1$。通常求 $G^{-1}\ (\cdot)$ 是用数值逼近方法来解决。

3.5　随机模拟基本原理

3.5.1　误差模拟原理

随机建模目的在于建立随机场 $\{Z(x)|x\in D\}$ 的概率分布模型。实际应用中，是建立随机场的离散化后的联合概率模型（3.21）或（3.23）。通常的做法是将研究区域离散成网格。在二维情形，网格共有 $N_x\times N_y$ 个结点，随机建模就是建立 $N_x\times N_y$ 个随机变量的联合概率模型。在空间情形，建立 $N_x\times N_y\times N_z$ 个随机变量的联合概率模型。再使用抽样理论获得每个网格结点处随机场模拟值（可称之为实现），这样就得到整个随机场的一个实现 $\{Z(i,j,k),i=1,2,\cdots,N_x;j=1,2,\cdots,N_y;k=1,2,\cdots,N_z\}$。因此，随机模拟值就是获得的随机场的数值结果。

误差模拟总的思路是从非条件模拟入手，通过模拟每个网格结点误差，并将误差加回到每个网格结点的克里金估计值之中，从而得到整个区域随机场的条件模拟（Journel and Huijbregts，1978）。具体过程如下。

设 $Z(x)$ 是一个平稳的零均值函数。若在已知数据条件下，获得了未知的抽样位置 x 的克里金估计值（简单克里金）$z^*(x)$，则

$$Z(x) = Z^*(x) + [Z(x) - Z^*(x)] \tag{3.32}$$

要直接对误差 $Z(x) - Z^*(x)$ 进行模拟是困难的，这是因为

$$\begin{aligned} &E\{[Z(x) - Z^*(x)][Z(y) - Z^*(y)]\} \\ &= E\{Z(x)[Z(y) - Z^*(y)]\} - E\{Z^*(x)[Z(y) - Z^*(y)]\} \end{aligned} \tag{3.33}$$

由于克里金估计正交于克里金误差，从而

$$E\{Z^*(x)[Z(y) - Z^*(y)]\} = 0$$

$$\begin{aligned} E\{Z(x)[Z(y) - Z^*(y)]\} &= E[Z(x)Z(y)] - E[Z(x)Z^*(y)] \\ &= C(x - y) - E\sum_{i=1}^{n}\lambda_i(y)Z(x_i)Z(x) \end{aligned}$$

在此，$\lambda_i(y)$ 是克里金估计权系数，$\{Z(x_i)\,|\,i = 1,2,\cdots,n\}$ 是已知数据集。从而

$$E\{Z(x)[Z(y) - Z^*(y)]\} = C(x - y) - \sum\lambda_i(y)C(x_i - x) \tag{3.34}$$

显然说明了直接模拟估计误差是不可行的。这是因为估计误差协方差（3.33），结构（3.34）是复杂的，甚至是非平稳的。

为此，采取的思路是选取一个独立于 $Z(x)$ 的关于 $Z(x)$ 的非条件模拟 $Z_s(x)$，即 $Z_s(x)$ 是 $Z(x)$ 的非条件模拟，且 $Z_s(x)$ 独立于 $Z(x)$。此时，$Z(x)$ 和 $Z_s(x)$ 具有同样的协方差结构，且

$$Z_s(x) = Z_s^*(x) + [Z_s(x) - Z_s^*(x)] \tag{3.35}$$

在此，$Z_s^*(x)$ 是对于数据 $Z_s(x)$ 进行的克里金估计。若在进行 Z 和 Z_s 估计时，使用的是来自于同一采样位置的数据集，那么 $Z_s(x) - Z_s^*(x)$ 和 $Z(x) - Z^*(x)$，具有同样的协方差结构。

这就提供了用非条件模拟的误差来替代（3.32）中的误差，$Z(x) - Z^*(x)$，得到

$$Z_{sc}(x) = Z^*(x) + [Z_s(x) - Z_s^*(x)] \tag{3.36}$$

这样（3.36）就给出了位置 x 处的条件模拟结果。

事实上，$E[Z_{sc}(x)] = E[Z(x)]$，并且

$$E[Z_{sc}(x)Z_{sc}(y)] = E[Z^*(x)Z^*(y)] + E\{[Z_s(x) - Z_s^*(x)][Z_s(y) - Z_s^*(y)]\}$$

而又有 $Z_s^*(x) \perp [Z_s(y) - Z_s^*(y)]$，因此：

$$\begin{aligned} E[Z_{sc}(x)Z_{sc}(y)] &= E[Z^*(x)Z^*(y)] + E\{Z_s(x)[Z_s(y) - Z_s^*(y)]\} \\ &= E\{\sum_{i=1}^{n}\lambda_i(x)Z(x_i)\sum_{j=1}^{n}\lambda_j(y)Z(x_j)\} + C(x - y) - E\{\sum_{j=1}^{n}\lambda_j(y)Z_s(x)Z_s(x_j)\} \\ &= C(x - y) + \sum_{j=1}^{n}\lambda_j(y)\sum_{i=1}^{n}\lambda_i(x)C(x_i - x_j) - \sum_{j=1}^{n}\lambda_j(y)C(x - x_j) \end{aligned}$$

由简单克里金方程组［相当于（3.29）中 $\mu=0$］知，$\sum_{i=1}\lambda_i(x)C(x_i - x_j) = C(x - x_j)$，从而 $E[Z_{sc}(x)Z_{sc}(y)] = C(x - y)$。因此（3.36）给出了 $Z(x)$ 的条件模拟结果。

在应用中，误差模拟需要一个预先的条件模拟去产生来自于非条件模拟 $Z_s(x)$ 的实现 $Z_s(x)$ 以及经过两次克里金产生的 $Z^*(x)$ 和 $Z_s^*(x)$。实际上，由于两次克里金估计中数据构型是一样的，因而只需计算一次克里金权系数。这样就很容易获得条件模拟的实现

$$z_{sk}(x) = z^*(x) + [z_s(x) - z_s^*(x)]$$

3.5.2　序贯模拟原理

一旦建立随机场 $\{Z(x)|x \in D\}$ 的概率模型，产生来自于概率模型实现的另一类方法就是序贯模拟方法（Journel and Alabert，1988）。将研究区域 D 网格化，设共有 N 个网格结点。若研究的是 N 个网格结点上储层的同一属性，则考虑 N 个随机变量 Z_i 的联合分布，更一般地若考虑共有 N 个网格结点区域上 K 种不同属性的组合，则 $N = K \times N$。这相当于协同模拟的情形。

N 个随机变量的条件联合概率模型记为

$$F_N[z_1, z_2 \cdots, z_N|(n)] = \text{Prob}\{Z_i \leq z_i, i = 1,2,N|(n)\} \tag{3.37}$$

表达式（3.37）完全是一般化的，并无特别的限制。变量 Z_i 中的全部或一部分可以是范畴变量。

为了得到来自于（3.37）的 N 元样本，可以由 N 个相继的步骤来完成。每一步是从一个单变量 *ccdf* 中抽样，且条件数据不断增加。由概率论知识知道有如下关系

$$F_N[z_1, z_2, \cdots z_N|(n)] = \text{Prob}\{Z_1 \leq z_1|(n)\} \cdot$$
$$\text{Prob}\{Z_2 \leq z_2|(n+1)\} \cdot \text{Prob}\{Z_3 \leq z_3|(n+2)\}$$
$$\cdot \cdots \cdot \text{Prob}\{Z_N \leq z_N|(n+N-1)\}$$

因此联合分布（3.37）完全由一系列如下 N 个单变量条件累积分布函数唯一确定：

$$Z_1 \sim \text{Prob}\{Z_1 \leq z_1|(n)\};$$
$$Z_2 \sim \text{Prob}\{Z_2 \leq z_2|(n+1)\};$$
$$Z_3 \sim \text{Prob}\{Z_3 \leq z_3|(n+2)\};$$
$$\cdots$$
$$Z_N \sim \text{Prob}\{Z_N \leq z_N|(n+N-1)\}$$

给定原始数据 (n) 时，从 Z_1 的单变量条件累积分布函数中抽取一个样本 z_1，将 z_1 作为新的条件数据加入到原始数据集 (n) 中，此时利用新的信息集 $(n+1) = (n) \cup \{Z_1 = z_1\}$ 为条件化数据。

给定更新的信息集 $(n+1)$ 时，再从 Z_2 的条件累积分布函数中抽取一个样本 z_2，到更新后新的信息集 $(n+2) = (n+1) \cup \{Z_2 = z_2\}$。这样序贯地考察 N 个随机变量，可以得到 N 个相依随机变量 $\{Z_i|i = 1,2,\cdots,N\}$ 的联合实现 $\{z_i|i = 1,2,\cdots,N\}$。如果需要 L 个这样的联合实现的话，那么整个序贯抽样过程就重复 L 次，可以得到 L 个来自于概率模型（3.37）的联合实现 $\{Z_i^{(l)}|i = 1,2,\cdots,N\}, l = 1,2,\cdots,L$。

值得指出的是，序贯模拟思想独立于建立（3.37）中单个变量条件累积分布函数的模型与算法。在大多数情形下，所有（3.37）中单个随机变量假设为服从高斯分布的。这相当于强加（3.37）为高斯模型。此时，单个变量的均值和方差可通过 N 个简单克里金来确定。在不假定（3.37）为高斯模型的前提下，单个变量的条件累积分布函数可直接通过指示克里金系统来确定。

在具体操作过程中应注意下列问题。

严格使用序贯模拟方法需要确定越来越复杂的条件累积分布函数，即条件数据从 n 增加到 $(n+N-1)$。因此，通常使用邻域内的数据参与条件化以避免数据过大问题，然而当前面被模拟过的网格结点数目从 1 增加到 $N \gg n$ 时，可以出现邻域中靠近未抽样的位置的

条件数据大部分是前面已模拟的网格数据，从而屏蔽了较远的原始数据的影响。解决该问题的一个办法是多步（或多阶段）网格概念，此时可通过两步或多步方法模拟 N 个网格结点的值。

1）首先对一个粗网格，如每 10 个网格为一组，每一次每组的第 10 个网格结点使用大半径邻域来模拟。大邻域允许再生大尺度变异函数结构。

2）其次，剩余的网格结点用小邻域来模拟。

理论上讲，无须指定 N 个结点被访问的次序。然而，实践已经证明应使用网格结点的随机访问序列。确实如果 N 个结点按行来访问，则与理论模型的任何偏差会导致模拟结果呈现人为的沿着行展布的痕迹。

第 4 章　高斯场模型

众所周知，正态分布在实际应用中起着十分重要的作用。许多实际模型都假定数据来自于多元正态分布。然而，实际数据绝不会严格地服从多元正态分布。一般来讲，正态分布仅是真实总体分布的很有用的近似。

正态分布的优点是数学上易于处理，并能获得好的结果。然而，对其它分布，情况往往不是这样。正态分布在实际中应用广泛主要有两个原因。首先，正态分布在许多情况下确实能作为真实总体的分布。其次，根据中心极限原理，无论总体分布如何，当样本容量足够大时，许多多元统计量的分布是近似正态的。

在地质统计学框架内，高斯场模型同样具有广泛的应用。它可以作为许多连续的随机函数的总体模型。高斯场模型实际上是多元正态分布在空间范围内的推广。因此，为了介绍高斯场模型，有必要先对多元正态分布的一些基本理论给予讲解。

4.1　基 本 理 论

4.1.1　条件高斯分布的一般理论

1. 多元正态分布及其性质

多元正态分布是一元正态分布向二元以上的情况的推广。设 Z 是一个连续型随机变量，它服从一元正态分布，记为 $Z\sim N(\mu,\sigma^2)$，μ 称为总体均值，σ^2 称为总体方差。一元正态分布的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2],-\infty<x<+\infty \tag{4.1}$$

n 元正态分布，指 $(Z_1,Z_2,\cdots,Z_n)$ 为 n 元随机变量。它服从 n 元正态分布，记为 $(Z_1,Z_2,\cdots,Z_n)^T\sim N_n(\vec{\mu},\Sigma)$。在此，$\vec{\mu}$ 为均值向量，$\vec{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T$，$\Sigma$ 为 $n\times n$ 的协方差矩阵。假定$\sum$是正定的，则 n 元正态分布的联合密度函数为

$$f(z_1,z_2,\cdots,z_n)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp[-\frac{1}{2}(\vec{z}-\vec{\mu})^T\Sigma^{-1}(\vec{z}-\vec{\mu})] \tag{4.2}$$

在此，$\vec{z}=(z_1,z_2,\cdots,z_n)^T$，$-\infty<z_i<+\infty$，$i=1,2,\cdots,n$。特别地，当 $n=2$ 时，二元正态分布的联合密度函数可写成

$$f(z_1,z_2)=\frac{1}{2\pi\sqrt{\sigma_1\sigma_2(1-\rho_{12}^2)}}\times\exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{z_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{z_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{z_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{z_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\} \tag{4.3}$$

在此

$\mu_1=E(Z_1),\mu_2=E(Z_2),\sigma_{11}=\mathrm{Var}(Z_1),\sigma_{22}=\mathrm{Var}(Z_2)$,

$\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \mathrm{Cov}(Z_1, Z_2)/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}})$；

ρ_{12}称为相关系数。协方差矩阵为$\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$。

在 n 元正态分布中，当 $\vec{z} = \vec{\mu}$ 时，正态密度达最大值。因此，$\vec{\mu}$ 常称为众数，也是 $\vec{Z} = (Z_1, Z_2, \cdots, Z_n)^T$ 的期望值或均值。

多元正态分布有如下性质：

1）$\vec{Z} - (Z_1, Z_2, \cdots, Z_n)^T$ 分量的线性组合仍服从正态分布。若 $\vec{Z} \sim N_n(\vec{\mu}, \Sigma)$，则随机变量的任意线性组合 $\vec{a}^T\vec{Z} = a_1Z_1 + a_2Z_2 + \cdots + a_nZ_n \sim N(\vec{a}^T\vec{\mu}, \vec{a}^T\Sigma\vec{a})$，在此，$\vec{a}^T = (a_1, a_2, \cdots, a_n)^T$ 为线性组合系数向量。

2）$\vec{Z}$ 分量的任一子集，仍服从多元正态分布，即 n 元正态分布的任一 m（$m \leqslant n$）元的边缘分布仍是正态分布。事实上，若将 $\vec{Z}$ 及其均值向量 μ 和协方差阵 $\sum$ 进行分块，

$$\underset{(n\times 1)}{\vec{Z}} = \begin{pmatrix} \underset{(m\times 1)}{\vec{Z}_1} \\ \underset{((n-m)\times 1)}{\vec{Z}_2} \end{pmatrix}, \underset{(n\times 1)}{\vec{\mu}} = \begin{pmatrix} \underset{(m\times 1)}{\vec{\mu}_1} \\ \underset{((n-m)\times 1)}{\vec{\mu}_2} \end{pmatrix}, \text{和 } \underset{(n\times n)}{\Sigma} = \left(\begin{array}{c|c} \underset{(m\times m)}{\Sigma_{11}} & \underset{m\times(n-m)}{\Sigma_{12}} \\ \hline \underset{(n-m)\times m}{\Sigma_{21}} & \underset{(n-m)\times(n-m)}{\Sigma_{22}} \end{array}\right), \text{则 } \vec{Z}_1 \sim N_m(\vec{\mu}_1, \Sigma_{11})。$$

3）协方差为零的两个分量相互独立。

4）多元正态分布的任一条件分布仍是正态的。

设 $\vec{Z} = \begin{pmatrix} \vec{Z}_1 \\ \vec{Z}_2 \end{pmatrix} \sim N_n(\vec{\mu}, \Sigma)$，其中，$\vec{\mu}\begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$，且 $|\Sigma_{22}| > 0$，

则 $(\vec{Z}_1 \mid \vec{Z}_2 = \vec{z}_2) \sim N(\vec{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{z}_2 - \vec{\mu}_2), \Sigma_{11} - \Sigma_{12}^T\Sigma_{22}^{-1}\Sigma_{21})$。

即条件分布服从均值为，$\vec{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{z}_2 - \vec{\mu}_2)$ 和协方差为 $\Sigma_{11} - \Sigma_{12}^T\Sigma_{22}^{-1}\Sigma_{21}$ 的正态分布。注意在此，协方差不依赖于条件变量 $\vec{z}_2$。

特别地，对于二元分布来讲，条件分布服从一元正态分布：

$$(Z_1 \mid Z_2 = z_2) \sim N\left[\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2), \sigma_{11}(1 - \rho_{12}^2)\right]$$

此时，条件均值是条件变量 z_2 的线性函数 $\mu_1 + (\sigma_{12}/\sigma_{22})(z_2 - \mu_2)$，而条件方差 $\sigma_{11}(1 - \rho_{12}^2)$ 不依赖于条件变量 z_2。二元正态分布的条件分布示意如下（图 4.1）。

从图 4.1 可以看出，二元正态分布的条件分布皆是正态的。它们的方差是相同的，即 $\sigma_{1|2}^2 = \sigma_{11}(1 - \rho_{12}^2)$，条件分布的均值是条件变量的线性回归

$$\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2)$$

这些条件分布具有相同的形态。通常的估计方法就是对条件期望的估计，因此是一种线性回归模型。这种线性估计常具有无偏和估计误差最小的特征。地质统计学中的克里金估计理论，可以放在线性回归的框架内加以讨论，也可放在高斯场模型框架内加以讨论。

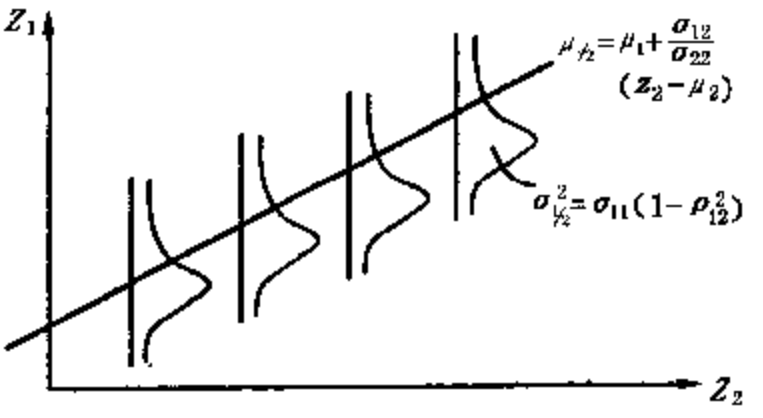


图 4.1　二元正态分布条件分布的示意图

2．高斯场模型

高斯场模型 $\{Z(x) \mid x \in D\}$，$D \subset R^d$，$d=1, 2, 3)$，可以看成是定义在 R^d 空间子区域 D 上的随机函数。它对任意 $n \in N$ 和数值 $(x_1, x_2, \cdots, x_n)$，$x_i \in D$，$i=1, 2, \cdots, n$，n 元随机变量 $[Z(x_1), Z(x_2), \cdots, Z(x_n)]$ 是服从 n 元正态分布的。或者讲，高斯场模型 $\{Z(x) \mid x \in D\}$ 的任何边缘分布皆是多元正态的。当 $d=1$ 时，高斯场模型相当于一维高斯过程，$d=2$ 时，高斯场模型相当于平面上的高斯过程，当 $d=3$ 时，高斯场模型相当于空间上的高斯过程。

定义在 R^d 上的高斯场模型，完全由均值函数 $m(x)$ [在平稳假设下，通常假定 $m(x)$ 为零] 和协方差函数 $C(x, y)$ 来确定，在此 $x, y \in D$。因此，高斯场模型可记为

$$Z(x) \sim \text{Gauss}[m(\cdot), C(\cdot, \cdot)] \tag{4.4}$$

若 $C(x, y) = C(x-y, 0)$ 则称高斯场模型 $Z(x)$ 是平稳的。进一步，若，$C(x-y, 0) = C(|x-y|) = C(h)$，则称高斯场模型是各向同性的。各向同性的高斯场模型，其协方差函数仅依赖于空间两点的距离。

在实际应用中，通常将研究区域 D 离散化成网格系统，如在三维空间中，可将 D 离散成网格数为 $N_x \times N_y \times N_z$ 元的多元正态随机变量模型。因此，高斯场模型的模拟转化为对 n 元正态随机变量的模拟（$n = N_x \times N_y \times N_z$）。在三维空间中，一般地 n 可以达到 10^4 或 10^6 数量级。此时会遇到庞大的 $n \times n$ 协方差矩阵的分解。因而应寻找有效的模拟方法。

一般情况下，极少有直接对具有指定协方差函数 $C(\cdot, \cdot)$ 的高斯场模型进行模拟的，而是采用模拟具有指定协方差函数的其它随机过程。这种过程一般是非高斯的。然后，依据中心极限定理，通过平均若干这种非高斯过程实现方法可得到高斯场模拟的结果。

法国的 Matheron 教授在 1973 年提出的模拟具有各向同性协方差函数的平稳高斯场模型方法——旋转带法，是产生平稳各向同性高斯场模型实现的又一个有用的方法。对于该方法，本章随后会给予充分讨论。

4.1.2　连续随机曲面的高斯场模型

油藏的物性参数常具有连续变化之特点。因此，通常采用高斯场模型对之进行建模。此时，选取的模型是一个连续的曲面外加一个具有空间相关性的高斯噪声模型。该模型具有广泛的适应性，例如泛克里金法的理论就可以放在该模型框架内加以讨论。下面就是基本模型的框架。

设 $z=z(x)$ 是一个连续的或近似连续的曲面，它定义在区域 D 内。该曲面在点 x_1，x_2，$\cdots$，x_n 处观测数据记为 $z(x_i)$，$i=1, 2, \cdots, n$。利用这些观测数据再现整个曲面的一种空间统计方法，就是把 $z(x)$ 看成是空间随机过程 $Z(x)$ 的一个实现。为此，需要把该曲面的有关信息与 $Z(x)$ 的一个适当的模型联系起来。先把这个模型设置为一个光滑的趋势面外加具有空间相关性的噪声，再利用观测数据去估计该模型的参数，回答在该模型假设下关于 $Z(x)$ 的问题，获取所有可用的信息。

为此，设 $Z(x)$ 是一个高斯过程，且为一个回归趋势面外加一个零均值的高斯噪声，即

$$Z(x) = m(x, \beta) + \varepsilon(x) = \sum_{j=1}^{p} \beta_j f_j(x) + \varepsilon(x) \tag{4.5}$$

$$\gamma(x, y) = \frac{1}{2} E\{[Z(x) - Z(y)]^2\}$$

在此，$\beta=(\beta_1, \beta_2, \cdots, \beta_p)^T$ 是一个向量。f_j 是回归函数，常取 $f_1(x)=1$。β_j 是回归参数，$\gamma(\cdot, \cdot)$ 是变异函数。协方差函数 $\mathrm{Cov}[Z(x), Z(y)]=\mathrm{Cov}[\varepsilon(x), \varepsilon(y)]=\sigma^2 K(x, y)$ 可用来度量空间噪声的连续性程度。

如果欲在 x 点处进行内插，从以上高斯假设知

$$\begin{Bmatrix} Z(x) \\ Z_{\mathrm{dat}} \end{Bmatrix} \sim N_{n+1}\left\{\begin{pmatrix} f(x)^T\beta \\ F\beta \end{pmatrix}, \sigma^2\begin{pmatrix} K(x,x) & k^T \\ k & K \end{pmatrix}\right\}$$

这里，K 是一个 $n\times n$ 的以 $K(x_i, x_j)$ 为元素的矩阵，称为协方差矩阵。还有，$k=[K(x,x_1),K(x,x_2),\cdots,K(x,x_n)]^T$，$Z_{\mathrm{dat}}=[Z(x_1),Z(x_2),\cdots,Z(x_n)]^T$。而 F 是一个 $n\times p$ 的矩阵，第 i 行为 $f(x_i)^{\mathrm{T}}=[f_1(x_i), f_2(x_i), \cdots, f_p(x_i)]$。

从而在观测数据给定的条件下，随机变量 $Z(x)$ 的分布可写成如下的高斯分布：$[Z(x)\,|\mathrm{data}] \sim N\{m(x, \beta)+k^{\mathrm{T}}K^{-1}[Z_{\mathrm{dat}}-m_{\mathrm{vee}}(x, \beta)], \sigma^2(k(x, x)-k^{\mathrm{T}}K^{-1}k)\}$。这里 m_{vee}是一个向量，其分量 $m(x_i, \beta)$。由于 $m(x, \beta)=F\beta$，所以一般来讲这个向量要随着回归函数 $f_i(x)$ 的改变而变化。

利用最小二乘法对 β 进行估计，使得 $(Z_{\mathrm{dat}}-F\beta)^TK^{-1}(Z_{\mathrm{dat}}-F\beta)$ 达到极小。当协方差矩阵 K 已经确定时，β 的最小二乘估计 $\hat{\beta}_{\mathrm{LS}}$可写成如下形式 $\hat{\beta}_{\mathrm{LS}}=HF^TK^{-1}Z_{\mathrm{dat}}$。在此，$H=(F^TK^{-1}F)^{-1}$，$\hat{\beta}_{\mathrm{LS}}$为 β 的无偏估计，且协方差矩阵为 σ^2H。最后，在 x 处的内插结果可写成：

$$\begin{aligned}\hat{Z}(x) &= m(x,\hat{\beta})+k^TK^{-1}[Z_{\mathrm{dat}}-m_{\mathrm{vee}}(x,\hat{\beta})] \\ &= f^T(x)\hat{\beta}+k^TK^{-1}(Z_{\mathrm{dat}}-F\hat{\beta})\end{aligned}$$

对应的内插方差或估计方差可写成

$$\begin{aligned}\hat{\sigma}^2_{(x)} &= E\{[\hat{Z}(x)-Z(x)]^2\} \\ &= \sigma^2\{k(x,x)-k^TK^Tk+[f(x)-F^TK^{-1}F]^TH[f(x)-F^TK^{-1}F]\}\end{aligned}$$

由于讨论的目的是利用研究的曲面去估计 $Z(x)$，而不是研究趋势面本身。因此以上讨论的估计方差是关于随机变量 $Z(x)$ 的，而不是关于趋势面 $f^T(x)\beta$ 的。所以应该特别注意，$\hat{\sigma}^2(x)$ 与 $\mathrm{Var}\hat{Z}(x)$ 是不同的。因此，以上的分析与传统的趋势面分析是有明显差异的。顺便指出，上面的结果与通过泛克里金估计进行空间内插的结果是一致的。这里如果取

$$f^T(x)\beta=\beta_1+\beta_2x+\beta_3y$$

$\hat{Z}(x)$ 则对应平面上的一次漂移泛克里金估计结果，如果取

$$f^T(x)\beta=\beta_1+\beta_2x+\beta_3y+\beta_4x^2+\beta_5xy+\beta_6y^2$$

则 $\hat{Z}(x)$ 相应于平面上的二次漂移的泛克里金估计。

4.2　高斯场模型的检验

4.2.1　一维和二维边缘分布的正态性检验

1. 一维边缘分布的正态性检验

设 $\{z(x_\alpha)\,|\alpha=1, 2, \cdots, n\}$ 是来自于研究区域 D 的一个观测数据集。当样本容量 $n\geqslant 25$ 时，直方图能够显示出一维分布的一个尾部是否比另一个长得多。如果直方图近似对称，那么还可通过计算落在某些区间上观测值的数目作进一步验证。对一元正态分布 $Z\sim N$

(μ, σ^2)，$P\{|Z-\mu|<\sigma\} \approx 0.68$，$P\{|Z-\mu|<2\sigma\} \approx 0.95$，当样本容量 n 较大时，可以预期观测值落入区间 $(\bar{x}-s, \bar{x}+s)$ 和 $(\bar{x}-2s, \bar{x}+2s)$ 内的比例 $\hat{p}_1$ 和 $\hat{p}_2$ 分别约为 0.68 和 0.95。在此，$\bar{x}$ 和 s 分别为样本均值和样本标准差。根据 $\hat{p}_1$ 和 $\hat{p}_2$ 近似正态分布，则当

$$\hat{p}_1 - 0.68 | > 3\sqrt{\frac{0.68 \times 0.32}{n}} = \frac{1.4}{\sqrt{n}}$$

或

$$| \hat{p}_2 - 0.95 | > 3\sqrt{\frac{0.95 \times 0.05}{n}} = \frac{0.65}{\sqrt{n}}$$

表示 Z 与正态假设有偏离。当观测值的比例 $\hat{p}_1$ 和 $\hat{p}_2$ 很小时，Z 的分布比正态分布有较厚的尾部。

用图形进行数据分析是另一种有用的方法，这就是 Q-Q 图检验方法。它可用来验证正态性假设。Q-Q 图是样本分位数和假定数据来自正态总体时所期望得到的分位数的散点图。当这个图形大致呈直线形状时，正态假设就认为是可以接受的，不然，正态假设的合理性就值得怀疑。

为了方便起见，将样本从小到大排列 $z_{(1)} \leqslant z_{(2)} \leqslant \cdots \leqslant z_{(n)}$。其中，$z_{(i)}$ 是样本分位数。若 $z_{(i)}$ 都不相等，则恰有 i 个观测值小于或等于 $x_{(i)}$。由于假定观测值来自连续型分布，这个断言总是成立的。为了分析上的方便，常用 $(i-1/2)/n$ 近似代替落在 $z_{(i)}$ 左边的样本比例 i/n。用点 $[Z_{(i)}, (i-1/2)/n]$ 作图可给出经验分布函数，而 Q-Q 图提供了 $[Z_{(i)}, (i-1/2)/n]$ 和从正态分布导出的相应理论值一致性的度量。

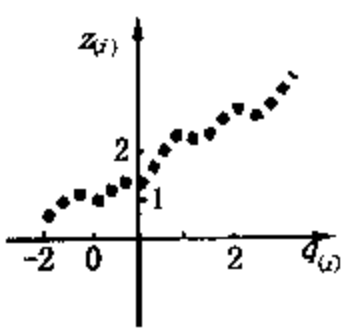


图 4.2　一元正态性检验的 Q-Q 图

利用 $P(i) = (i-1/2)/n$，可以得标准正态分位数 $q_{(i)}$。很显然，如果经验分布函数和正态分布函数比较接近，则点 $[q_{(i)}, z_{(i)}]$ $(i=1, 2, \cdots, n)$ 应大致是一条直线，如图 4.2 所示。

2. 二维边缘分布的正态性检验

观测数据集 $\{z(x_\alpha) | \alpha=1, 2, \cdots, n\}$ 是否来自于一个高斯随机场，常常事先是不知道的。通常的作法是正态刻度变换（详见下节），使变换后的数据 $\{y(x_\alpha) | \alpha=1, 2, \cdots, n\}$ 来自具有一维标准正态边缘分布的一个随机场 $\{Y(x) | x \in D\}$。然而，它是否为标准的高斯场模型，还应进行二维以及二维以上的边缘分布联合标准正态性检验。通常作二维边缘分布标准正态性检验就可以了。

$\{Y(x) | x \in D\}$ 二维边缘联合标准正态性检验可采用如下方法来作。利用关系式 (Deutsh and Journel, 1992)

$$P\{Y(x) \leq y_p, Y(x+\vec{h}) \leq y_p\} = P^2 + \frac{1}{2\pi}\int_0^{\arcsin \rho_Y(\vec{h})} \exp\left(-\frac{y_p^2}{1+\sin\theta}\right) d\theta \quad (4.6)$$

在此，$y_p = G^{-1}(p)$ 是标准正态分布的 p 分位数，$\rho_Y(\vec{h})$ 是标准高斯随机场 $Y(x)$ 的相关系数函数。由于 $P\{Y(x) \leq y_p, Y(x+\vec{h}) \leq y_p\} = E[I(x; p) \cdot I(x+\vec{h}; p)] = p - \gamma_I(\vec{h}; p)$，在此，

$$I(x; p) = \begin{cases} 1, \text{当 } Y(x) \leq y_p \\ 0, \text{其它情形} \end{cases}$$

是一个指示变量，而 $\gamma_I(\vec{h}; p)$ 是相应的指示变异函数。

这样，通过(4.6) 式，就建立了指示变异函数 $\gamma_1(\bar{h};p)$ 和二元联合正态表达式之间的关系。当(4.6) 式近似成立时，表明数据 $\{Y(x_\alpha)\mid \alpha=1,2,\cdots,n\}$ 来自于具有二维边缘标准正态分布的随机场。

值得指出的是公式 (4.6) 是对不同的 $\vec{h}$ 来分别计算的。因此，二维联合正态性要求对所有 $\vec{h}$，式 (4.6) 都应该满足。

4.2.2　正态刻度变换

设 $\{z(x_\alpha)\mid \alpha=1,2,\cdots,n\}$ 来自于一个随机场模型 $\{Z(x)\mid x\in D\}$，一般来讲，$\{Z(x)\mid x\in D\}$ 并不满足是一个高斯场。为此，可使用一种非参数化的变换方法，将已知观测数据集 $\{Z(x_\alpha)\mid \alpha=1,2,\cdots,n\}$ 转换成服从标准正态分布的数据集 $\{y(x_\alpha)\mid \alpha=1,2,\cdots,n\}$。即 $Z(x)\xrightarrow{\text{变换}}Y(x)$。值得指出的是，这种变换仅保证随机场模型 $\{Y(x)\mid x\in D\}$ 的一维边缘分布是标准正态的。

为了验证 $\{Y(x)\mid x\in D\}$ 是标准高斯场模型还需要对其进行二维边缘分布的正态性检验（见 4.2.1)。下面就具体讨论这种非参数化的变换方法，称之为正态刻度变换。

由概率论与数理统计基本理论，设 Z 和 Y 是两个随机变量，它们的分布函数分别记为 $F_Z(z)$ 和 $F_Y(y)$，则当 y_p 和 z_p 分别表示 Y 和 Z 的 p 分位数时

$$F_Y(y_p)=F_Z(z_p)=p,y_p,z_p\in[0,1]$$

因此

$$y=F_Y^{-1}[F_Z(z)]$$

如果 Y 是标准正态分布的随机变量，则 $F_Y=G(y)$，从而

$$y=G^{-1}[F_Z(z)] \tag{4.7}$$

就是正态刻度变换。在此 $G(\cdot)$ 为标准正态分布函数。

公式 (4.7) 给出了如何将一个随机变量转化为服从标准正态分布随机变量的方法。在实际应用中，将已知观测数值集 $\{Z(x_\alpha)\mid \alpha=1,2,\cdots,n\}$，排成次序 $z_{(1)}\leqslant z_{(2)}\leqslant\cdots z_{(n)}$，则相应于第 k 个 z 数据的累积频率为 $\hat{F}_Z(Z_{(k)})=k/n$。这样应用公式 (4.7) 可知 $z_{(k)}$ 正态刻度变换值就是标准正态分布的 k/n 分位数，即

$$y_{(k)}=G^{-1}\left(\frac{k}{n}\right) \tag{4.8}$$

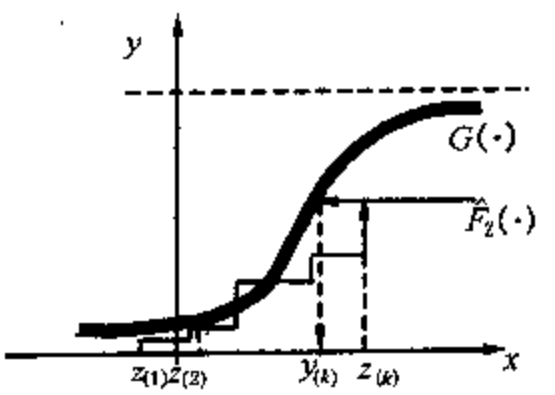


图 4.3　正态刻度变换示意图

这样就可用公式 (4.8) 将原来数据 $z_{(k)}$ 转化为服从标准正态分布的数据 $y_{(k)}$，值得指出的是，实际应用中，求 $G^{-1}(\cdot)$ 是用数值逼近方法来获取的。

图 4.3 是公式 (4.8) 的示意图。

4.3　模拟方法

4.3.1　序贯高斯模拟方法

序贯高斯模拟 (Sequential Gaussian Simulation) 是一种直接产生来自高斯场模型实现的方法。它基于序贯模拟思想（参见第 3 章 3.5.2 节序贯模拟原理)。该方法是先将研究区域离散化成网格系统，然后序贯地处理每一个网格结点。由于每个网格结点处随机变量是服从

条件化的正态分布，因此网格结点值完全由均值和方差两个参数确定。通过求解克里金方程组就可给出该网格节点处的均值和方差，从而将该结点处变量的正态分布完全确定下来，并采用相应的抽样方法得到该网格结点处一个样本。需要指出的是，求解克里金方程组时的条件数据是包括原始数据，和先前已模拟的、落在模拟邻域内所有被模拟的网格结点处的值。

设 $\{Z(x) \mid x \in D\}$ 是一个高斯场模型，其序贯高斯模拟步骤如下：

1）确定原始数据 $\{Z(x_\alpha) \mid \alpha = 1, 2, \cdots, n\}$ 的单变量分布函数 $F_Z(z)$，通常是获取经验分布函数 $\hat{F}_Z$；

2）利用 $\hat{F}_Z$，进行原始 Z 数据的正态刻度变换，得到服从单变量正态分布的数据 y。此时可认为 $\{y(x_\alpha) \mid \alpha = 1, 2, \cdots, n\}$ 是来自于一个随机场模型 $\{Y(x) \mid x \in D\}$，该随机场模型具有一维的边缘正态分布；

3）对 y 数据进行二元正态性检验，若满足二元正态性，则可认为 $\{Y(x) \mid x \in D\}$ 是近似服从高斯场模型。否则，使用其它模拟方法，如第 5 章将要讨论的指示模拟方法；

4）如果 $\{Y(x) \mid x \in D\}$ 认为服从高斯场模型，或者讲认为数据 y 来自于某个高斯场模型，则可进行如下序贯模拟步骤：

——指定一个网格系统结点的随机访问路径，在每个网格结点 $u \in D$，保留指定数目的特定邻域内的条件数据。其中，包括原始 y 数据和先前已模拟过的网格结点 y 值。

——利用简单克里金估计（simple kriging）来确定在位置 u 处高斯场 $\{Y(x) \mid x \in D\}$ 的条件分布函数的两个参数：均值和方差。

——从上述确定的条件累计分布函数中抽取一个模拟值 $y^{(l)}(u)$。

——将模拟值 $y^{(l)}(u)$ 追加到已知条件数据集中。

——再接着处理下一个网格结点，直至所有网格结点都被模拟完毕。

——将模拟的正态网格结点值 $\{y^{(l)}(u) \mid u \in D\}$ 进行反变换，得到原始 z 数据的模拟值 $\{z^{(l)}(u) \mid u \in D\}$。

若原来正态刻度变换为 $\varphi(\cdot)$：$Y = \varphi(Z)$，则反变换为 $Z = \varphi^{-1}(Y)$，从而 $z^{(l)}(u) = \varphi^{-1}[y^{(l)}(u)]$，$u \in D$。

上述序贯高斯模拟是一种条件模拟。它保证原始 z 数据和直方图及变异函数都被条件化。

4.3.2　旋转带模拟方法

旋转带方法（Turning Bands）是产生具有零均值和指定的各向同性协方差函数 $C(h)$ 的高斯场实现的一种非条件模拟方法。该方法是 1973 年由法国的 Matheron 教授首先提出来的。它是产生高斯场非条件模拟的经典方法，具有快速实用的特点。众所周知，非条件模拟是条件模拟的基础。旋转带方法在二维和三维空间中实施有一定的差异，但它们都最终转化为一维空间中的非条件模拟的实施。因此，本节将分别讨论三维旋转带方法，二维旋转带方法和一维非条件模拟。

1．三维旋转带方法

设 D_1 是三维空间中过原点的一条直线，其方向由单位向量 $\vec{u}_1$ 所确定。在 D_1 上考察一个二阶平稳的随机函数 $Y(\vec{u}_1)$，即有

$$E[Y(\vec{u}_l)] = 0$$

$$\mathrm{Cov}(Y(u_1), Y(u_1 + h_{u_1})) = C^{(l)}(h_{u_1})$$

设（$\vec{u}_1$，$\vec{v}_1$，$\vec{w}_1$）表示三个相互垂直的单位向量，那么就形成了三维空间中的一个坐标系。

设（h_{u_1}，h_{v_1}，h_{w_1}）表示向量 $\vec{h}$ 在该坐标系中的三个分量，按下式可定义一个三维随机函数 Z_1：Z_1（u_1，v_1，w_1）$=Y$（u_1），那么，该随机函数也是二阶平稳的，即有：

$$E[Z_1(u_1,v_1,w_1)]=0$$

$$C_1(h)=C_1(h_{u_1},h_{v_1},h_{w_1})=C^{(1)}(h_{u_1})$$

利用以下步骤可获取 Z_1 的一个实现，即首先在直线 D_1 上生成一个一维实现 y（u_1），然后对以平面 $\{u_1=$常数$\}$ 为中心平面，以平面 $\{u_1+\delta_1=$常数$\}$ 和平面 $\{u_1-\delta_1=$常数$\}$ 分别为顶面和底面的区域内的任意一点处，随机函数 Z 的实现定义为 y（u_1）。这里，δ_1 是 D_1 上模拟点之间距离的一半。

考虑在单位球面上均匀分布的 N 个单位向量 $\vec{u}_1$，$\vec{u}_2$，…，$\vec{u}_N$，以 y（u_i）及以它们为方向的通过坐标原点的 N 个直线 D_1，D_2，…，D_N。这样，在每条直线 D_i 上可产生同构于随机函数 Y 的另一个随机函数 Y_i 的一个新的实现 y（u_i），且假定这 N 个随机函数 $\{Y(\vec{u}_i)_i$，$i=1$，2，…，$N\}$ 是相互独立的。利用每一个一维实现 y（u_i），可生成三维随机函数 Z_i 的一个实现 z_i（u_i，v_i，w_i）$=y$（u_i），那么可得如下求和式：

$$z=\left(\sum_{i=1}^{N}z_i\right)/\sqrt{N}$$

可将 z 看成一个具有零均值的二阶平稳的三维随机函数 Z（x_1，x_2，x_3）的实现。因此，随机函数 Z 和 Z_i（$i=1$，2，…，N）之间建立了关系：

$$Z=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i \tag{4.9}$$

进一步知，当 $N\to\infty$ 时，Z 的协方差函数趋于一个各向同性的协方差函数 C（h）。C（h）和 $C^{(1)}$（h）之间有如下关系：

$$C(\vec{h})=c(h)=\int_{\Omega}C^{(1)}(\langle\vec{h},\vec{u}\rangle)\mathrm{d}u$$

在此，Ω 是 1/2 单位球面，$\langle\vec{h}$，$\vec{u}\rangle=h_u$ 表示向量 $\vec{h}$ 在坐标轴 $\vec{u}$ 上的投影，而

$$h=|\vec{h}|=(h_u^2+h_v^2+h_w^2)^{1/2}$$

上述积分可用球坐标形式写成

$$C(h)=\int_0^{2\pi}\mathrm{d}\theta\int_0^{\pi/2}C^{(1)}(|h\cdot\cos\varphi|)\sin\varphi\mathrm{d}\varphi=\frac{2\pi}{h}\int_0^h C^{(1)}(s)\mathrm{d}s$$

这里 C（h）是三维空间随机场 $\{Z(x)\,|x\in D\}$ 的协方差函数，而 $C^{(1)}$（h）是一维随机函数 Y（u）的协方差函数。在实际应用中，往往是给定 C（h），需要确定所对应的 $C^{(1)}$（h）。这时，从上式两端求导，可导出如下关系式：

$$C^{(1)}(s)=\frac{1}{2\pi}\frac{\partial}{\partial s}[s\cdot C(s)] \tag{4.10}$$

这样，首先按（4.10）用 C（s）计算出 $C^{(1)}$（s），并生成在直线 D_1 上具有协方差函数 $C^{(1)}$（s）的一维随机函数 Y（u）的实现，然后旋转该直线，且在其上不断地产生 Y（u）的实现。再利用公式（4.9）可获得三维空间随机场 $\{Z(x)\,|x\in D\}$ 的实现，并保证它具有指定的协方差函数 C（s）。

应该注意的是，当 N 趋向无穷大时，由中心极限定理可知，用前面的方法模拟所得是一个高斯场 $\{Z(x)\,|x\in D\}$。其具有零均值和指定的协方差函数 C（s）。因此，旋转带

方法是一种非条件化的高斯场模拟方法，而且是一种各向同性的高斯场非条件化模拟方法。

2. 二维旋转带方法

二维旋转带方法和三维旋转带方法相类似，只是相应的一维随机函数 $Y(x)$ 的协方差函数 $C^{(1)}(h)$，和二维空间随机场 $\{Z(x) \mid x \in D\}$ 的协方差函数 $C(h)$ 的关系，和三维时的情况有所不同。设一维随机函数 $Y(x)$ 二阶平稳，具有零均值，考虑穿过模拟区域中心的一条直线 D_1（图 4.4）。

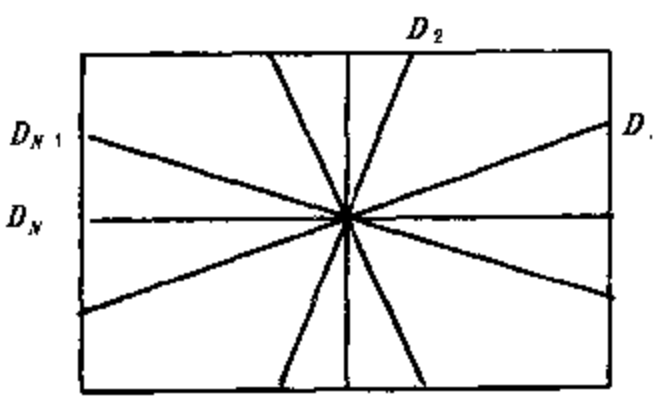


图 4.4　二维旋转带方法示意图

对该区域中任意一点 x，用 x_{D_1} 表示它在直线 D_1 上的投影，类似于三维情形，可定义一个二维随机函数如下：

$$Z_1(x) = Y(x_{D_1})$$

且有：

$$E[Z_1(x)] = 0$$

$$E[Z_1(x)Z_1(x+\vec{h})] = C^{(1)}(h_{D_1})$$

这里 h_{D_1} 为向量 $\vec{h}$ 在直线 D_1 上的投影。为了对 $Z_1(x)$ 进行模拟，需要考虑一系列和 D_1 垂直的条带，其宽度为 $Y(x_{D_1})$ 在 D_1 上模拟值所在相邻点之间的距离。这样在整个区域上就形成了若干个条带。每一个条带中都包含着其上存在 $Y(x_{D_1})$ 模拟值的一个点，可让 $Z_1(x_{D_1})$ 在条带内任一点处的实现取成 $Y(x_{D_1})$ 的实现 $y(x_{D_1})$。进一步考虑 N 条直线 D_1，D_2，…，D_N，它们的方向 $\vec{u}_1$，$\vec{u}_2$，…，$\vec{u}_N$ 是单位圆周上均匀分布的单位向量。在直线 D_i 上可生成随机函数 $Y(x)$ 的实现 $y(x_{D_1})$，且使这些实现是相互独立的。因此，对于 N 个互相独立的一维实现 $y(x_{D_i})$，$i=1, 2, \cdots, N$ 可得到 N 个二维随机场 $\{Z(x) \mid x \in D\}$ 的实现 $Z_i(x)$。在上述每一个条带内，$Z(x)$ 的实现都是相同的，于是用以下求和式可得到 $Z(x)$ 的实现 $z(x)$

$$z(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} z_i(x)$$

这就是二维平稳随机场 $\{Z(x) \mid x \in D\}$ 的一个实现，并有

$$E[Z(x)] = 0$$

$$C(h) = E[Z(x)Z(x+\vec{h})] = \frac{1}{N} \sum_{i=1}^{N} C^{(1)}(\langle \vec{h}, \vec{u}_i \rangle)$$

这里 $\langle \vec{h}, \vec{u}_i \rangle$ 表示向量 $\vec{h}$ 在直线 D_i 上的投影。当 N 很大时，$Z(x)$ 协方差函数趋向于一个各向同性的协方差函数 $C(h)$，

$$C(h) = \frac{1}{\pi} \int_0^{\pi} C^{(1)}(h\cos\theta) \mathrm{d}\theta$$

作变换 $x = h \cdot \cos\theta$，则有

$$C(h) = \frac{1}{\pi h} \int_{-h}^{h} \frac{C^{(1)}(x)\mathrm{d}x}{[1-(x/h)^2]^{1/2}}$$

对此积分方程求解便有：

$$C^{(1)}(s) = C(0) + \int_0^s s(s^2-h^2)^{\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}h}[C(h)]\mathrm{d}h \tag{4.11}$$

比较公式（4.10）和（4.11）不难发现，在二维和三维空间中 $C(h)$ 和 $C^{(1)}(s)$ 之间的

关系是不同的。三维空间中的关系较简单，而二维空间中的较复杂。

特别地，若 $C(h)$ 为球状模型，即：

$$C(h) = \begin{cases} C[1 - \frac{3}{2}(h/a) + \frac{1}{2}(h/a)^3] & \text{当 } h < a \\ 0 \end{cases}$$

那么，进行二维旋转带模拟时，相应地有

$$C^{(1)}(s) - \begin{cases} C(1 - 3\pi/4)[(s/a) - (s/a)^3/2], \text{当 } s < a \\ C(1 - \frac{3}{2}[(s/a) - (s/a)^3/2]\arcsin(a/s) - (3/4)(s/a)^2[1 - (a/s)^2]^{1/2}, \text{当 } s \geq a \end{cases}$$

而进行三维旋转带模拟时，相应地有

$$C^{(1)}(s) - \begin{cases} C[1 - 3(s/a) + 2(s/a)^3], \text{当 } s < a \\ 0, \text{当 } s \geq a \end{cases}$$

图 4.5 表明了 $C^{(1)}(s)$ 在二维和三维旋转带模拟时的差异。

3. 一维非条件模拟

从以上讨论可知，利用旋转带方法进行模拟的基础，是对具有零均值的二阶平稳的一维随机函数 $Y(x)$ 进行非条件模拟，并使其具有指定的协方差函数 $C^{(1)}(s)$。然后，再可利用公式（4.10）或（4.11）得到相应二维或三维随机场 $\{Z(x) \mid x \in D\}$ 的模拟结果。而要进行 $Y(x)$ 的一维非条件模拟，谱分析方法是一种经典的方法。其优点是不受 $C^{(1)}(s)$ 形式的限制。在实际应用中，谱分析方法需要较多的富立叶变换。以下先讨论当 $C^{(1)}(s)$ 具有特殊结构时，利用谱分析方法进行一维非条件模拟，然后再讨论 $C^{(1)}(s)$ 为一般形式时的情况（Riplay，1987）。

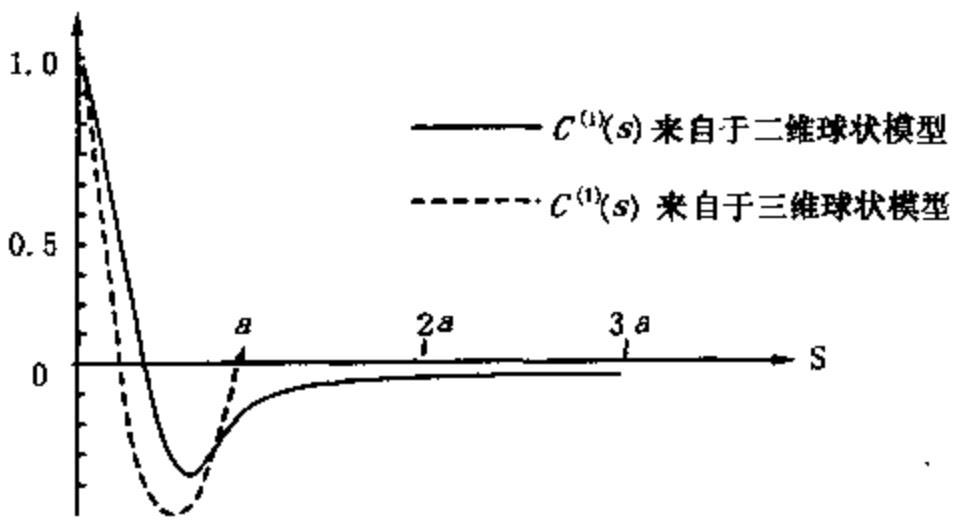


图 4.5　来自于二维、三维球状模型 $C^{(1)}(s)$ 图形

(1) $C^{(1)}(s)$ 为特殊结构

若 $C^{(1)}(s)$ 可表示成已知函数 $f(u)$ 与其转置 $\check{f}(u) = f(-u)$ 的褶积时，即

$$C^{(1)}(s) = f * \check{f} = \int_{-\infty}^{+\infty} f(u) f(u + s) \mathrm{d}u$$

则 $Y(x)$ 的模拟过程如下。

设 $T(\mathrm{d}r)$ 为一个平稳的一维随机测度，且具有狄拉克协方差测度，如泊松过程或布朗运动（布朗运动是增量为高斯分布的，且相互独立的随机过程）的微分形式，即

$$E[T(\mathrm{d}r)T(\mathrm{d}r')] = \begin{cases} 0 & \text{当 } r \neq r' \\ \sigma^2 \mathrm{d}r & \text{当 } r = r' \end{cases}$$

在实际应用中，随机测度 $T(\mathrm{d}r)$ 是用沿着一条直线，位于离散区间内的一系列随机变量 T_i 来逼近的。这些 T_i 是相互独立且有相同的分布，期望值为零，方差为 $E[T_i^2] = \sigma^2$。令

$$Y(u) = \int_{-\infty}^{+\infty} f(u + r) T(\mathrm{d}r) = T * \check{f}$$

可以证明 $Y(u)$ 是一个平稳随机函数，其协方差函数等于 $C^{(1)}(s)$。为了得到 $Y(u)$ 的一个实现，需要将各随机变量 T_i 的实现 t_i 以规则间隔 b 分配到直线 D 上的各个点上，如图

4.6所示。t_i 值可以取为具有均匀分布的随机变量的独立实现。

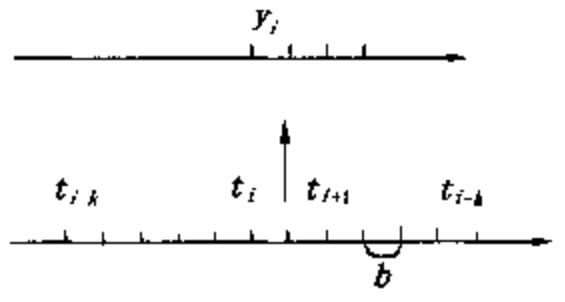


图 4.6　$Y(u)$ 实现示意图

若令 t_{i-k}，…，t_i，…，t_{i+k} 是点 $i-k$，…，i，…，$i+k$ 处实现，对于每一点 i，可令

$$y_i = \sum_{k=-\infty}^{k=+\infty} t_{i+k} \cdot f(kb)$$

y_i 这个表达式具有实用意义的一个必要条件是，函数 $f(kb)$ 随着 $|k|$ 的增大迅速减少到零。因此，不必对 $k=-\infty$ 到 $+\infty$ 求和。易证，可将 y_i 看成是具有零均值，且协方差为

$$C_d^1(s) = \sigma^2 \sum_{-\infty}^{+\infty} f(kb) f(kb - bs)$$

的一维随机函数 $Y(u)$ 的实现。$C_d^{(1)}(s)$ 可看成是以前给定的 $C^{(1)}(s)$ 的一个离散逼近，两者之间的偏差可通过参数 σ^2 和 b 的适当选取而予以校正。

下面将讨论两种具体的情况。在地质统计学中，球状模型和指数模型是两种经常使用的协方差结构，在使用旋转带方法时，相应于它们的一维协方差 $C^{(1)}(s)$ 也可以表示成已知函数的褶积形式。

如果把球状协方差结构用下式表示

$$C(s) = \begin{cases} k[1 - 3s/(2a) + s^3/(2a^3)], \text{当 } 0 \leq s \leq a \\ 0, \text{当 } s \geqslant a \end{cases}$$

而指数协方差结构用下式表示：

$$C(s) = ke^{-as}, \text{当 } s \geq 0$$

这里，k，a 都是正常数，那么对于三维球状模型就有：

$$C^{(1)}(s) = \begin{cases} (k/2\pi)[1 - (3s/a) + 2(s^3/a^3)], \text{当 } 0 \leq s \leq a \\ 0, \text{当 } s \geq a \end{cases}$$

在此 k 是一个适当选取的常数。当 $C^{(1)}(s)$ 表示成如下的褶积形式时：$C^{(1)}(s) = f * \check{f}$ 则有

$$f(u) = \begin{cases} (6k/\pi a^3)^{1/2} - u, \text{当 } -\frac{a}{2} \leq u \leq \frac{a}{2} \\ 0, \text{其它} \end{cases}$$

指数模型的协方差函数也可表示成如下的褶积形式：$C^{(1)}(s) = f * \check{f}$。只是 $f(u)$ 的形式和球状模型时不同，应为：

$$f(u) = \begin{cases} (2ka/\pi)^{1/2}(1 - au)e^{-as}, \text{当 } u \geq 0 \\ 0, \text{当 } u < 0 \end{cases}$$

由于离散逼近引起的 $C_d^{(1)}(s)$ 和 $C^{(1)}(s)$ 之间的差别，需要进行相应的校正，最后才能得到所需的一维模拟结果。

(2) 一般的谱分析模拟方法

对于二维球状模型协方差函数，相应的一维协方差函数 $C^{(1)}(s)$ 难以表示成褶积的形式。为此，以下讨论一维直线上的一般谱分析模拟方法。

如果先在直线 R^1 上模拟一个平稳的随机过程 $Z_1(\cdot)$，它具有协方差函数 $C^{(1)}(s)$，则可在 R^d 选择一个随机旋转 $\vec{O}$，并定义如下的过程：

$$Z(x) = Z[(\vec{O}x)_1]$$

那么就可以把旋转带方法推广到任意维的欧氏空间 R^d 上去。为此，以下讨论一般的处理方法。

设 X（x）为直线上的平稳随机过程，协方差函数为 C（x），不失一般性可将 X（x）表示成如下的随机积分形式

$$X(x) = \int K(x - y)\mathrm{d}B(y)$$

这里，B（y）为布朗运动过程，K（$x-y$）需要待定，那么其协方差函数应为 C（x）= $\int K$（$x-u$）K（u）$\mathrm{d}u$。所谓布朗运动，植物学家 Robert Brown 于 1828 年观察到悬浮在水中的花粉不规则运动而得名的（Karatzas and other, 1988)。简单地说，布朗运动过程是一个连续的随机过程 $B=\{B_t\}$，且满足如下条件：$B_0=0$，又当 $0<s<t$ 时，$B_t - B_s$ 是相互独立且满足正态分布。它对于以上随机积分进行离散逼近，得到 X（x）$\approx\sum K$（$x - n\delta$）ε_n。这里 $\varepsilon_n \sim N$（0，δ），且相互独立。为了通过快速富立叶变换来计算 X（x）在 $\{x = r\delta\}$ 处的值，考察 $\{0, 1, \cdots, N-1\}$ 上的平稳过程：$X_j = X$（$j\delta \mathrm{mod} N\delta$）。如果将 X_j 表示成如下形式

$$X_j = \sum_{r=0}^{N-1} K(j\delta - r\delta)\varepsilon_r$$

再令 $K_j = K(j\delta \mathrm{mod} N\delta)$，那么就有

$$X_j = \sum_s K_{j-s}\varepsilon_s$$

设 $\tilde{X}_r$ 和 $\tilde{\varepsilon}_r$ 是 X 和 ε 的离散富立叶变换，则有：

$$\begin{aligned}\tilde{X}_r &= \sum_{j=0}^{N-1} e^{2\pi ijr/N} X_j \\ &= \sum_j e^{2\pi ijr/N}\Big(\sum_s K_{j-s}\varepsilon_s\Big) \\ &= \sum_s \varepsilon_s \sum_j e^{2\pi ijr/N} K_{j-s} \\ &= \sum_s \varepsilon_s \sum_j e^{2\pi ir(s+k)/N} K_k \\ &= \tilde{K}_r \tilde{\varepsilon}_r\end{aligned}$$

在此，$k = j - s$，$i = (-1)^{1/2}$。

进一步，可通过下式从 $\tilde{X}_r$ 得到 X_j

$$X_j = N^{-1}\sum e^{-2\pi ijr/N}\tilde{X}_r$$

由此可见，为了计算 X_j，首先要生成 $\tilde{K}_r$，接着生成（ε_0，ε_1，$\cdots$，ε_{N-1}）和 $\tilde{\varepsilon}_r$，然后再生成 $\tilde{X}_r$ 和 X_j。

为了直接生成 $\tilde{\varepsilon}_r$，设

$$\tilde{\varepsilon}_r = U_r + iV_r$$

这里 $V_0=0$，$U_{N-r} = U_r$，$V_{N-r} = V_r$，N 为偶数，N 个随机变量 U_0，$\cdots$，$U_{N/2}$，V_1，$\cdots$，$V_{N/2-1}$是相互独立的正态分布：

$$U_0 \sim N(0, N\delta), U_1, \cdots, U_{N/2} \sim N(0, N\delta/2)$$

$$V_1, \cdots, V_{N/2-1} \sim N(0, N\delta/2)$$

对函数 K 的选择有一定的自由度，只要使 X_j 的谱密度满足 $\tilde{C}_s = |\tilde{K}|^2$ 即可，不失一般

性，可选 $\tilde{K}_r=\tilde{K}_{-r}=\tilde{C}^{1/2}$。为了生成 $\tilde{C}_s$ 和 $\phi_s=\tilde{C}_s^{1/2}$，需要先生成 C_0，…，C_{N-1}。为此可考虑到 $C_r=C\ (r\delta)$，$r\leq N/2-1$，以及关系式

$$C_{N-r}=C_r$$

注意，这里 N 为偶数，而 $r=0$，…，$N/2-1$。

总上所述，可通过以下步骤生成 $X\ (0)$，…，$X\ [\ (N-1)\ \delta]$ 的模拟值。

1）对随机函数 U_0，U_1，…，$U_{N/2}$，V_1，…，$V_{N/2-1}$进行抽样，所得实现记为 u_0，u_1，…，$u_{N/2}$，v_1，…，$v_{N/2-1}$；

2）用快速富立叶变换求取以下和式

$$N^{-1}\sum_{r=0}^{N-1}[e^{-2\pi ijr/N}(u_r+iv_r)\varepsilon_r]$$

即得 $X\ (j\delta)$ 的一个模拟值。

利用以上方法可生成 $\{0,1,\cdots,N-1\}$ 上任何平稳高斯过程。这样的过程可逼近 R^1 上任何一个平稳的一维高斯过程。

4.4　随机分数维场及其模拟方法

4.4.1　随机分数维场模型

非均质多孔介质流体流动的理论表明，流体迁移性质受到渗透率分布空间相关结构的影响。当渗透率作为一个空间变量，其变异函数的变程大于流体流动的路径时，迁移一扩散方程的解不可能提供对介质迁移的特征的精确预测。

随机模拟技术可用来认识储层渗透率的空间分布，而且还可要求模拟结果的空间相关性与观测数据提供的渗透率的空间相关性相符合。虽然地质统计学的一些方法已经用来刻划储层的性质，然而在建立储层动态模型的过程中，非均质多孔介质的迁移统计理论已受到明显挑战。各种沉积相储层岩性和物性的空间分布，不仅具有某种尺度的空间相关性，同时具有分数维特征，因此需要进行分数维条件模拟（Hewett，1986）。所以，有必要讨论分数维分布的几何特征和空间相关结构。区域数据分数维特征的度量以及具有分数维特征的空间变量的模拟是本节所要讨论的内容。

1．分数维分布

美国 IBM 公司的数学家曼德尔布罗特的分数维几何学，对于自然界中不规则的复杂图形赋予特殊的数学描述而开创了--门新的学科。分数维几何学的研究对象在自然界中广泛存在。从多孔岩石的微观结构，手指纹的形态，到沉积矿床中矿石的分布，海岸线的形状，以及山峦地貌等，都可用分数维几何学对其特征进行建模。

在所有尺度上都具有变异性（即处处不可微）和自相似性，是分数维分布的特点之一。把分数维图形划分成较规则的局部的企图是不可能的。分数维图形的任何一部分，无论是多么的细小，其复杂程度丝毫不会减弱。数学家最初研究的分数维结构就是具有自相似性的几何图形，如 Kock 曲线和 Sierpinski 垫片等。它们提供了揭示分数维性质极好的例子。然而，这些例子对于地质建模而言显得过于“典型”。这是因为描述自然性状最适合的分数维结构应该是具有统计自相似性的随机函数，而不是严格的自相似性。

分数维分布的几何性态是利用几何图形中不能充填的空间空隙来刻划的，且可用分数维数这个参数来定量描述。这个参数反映了图形填充空间分布的性质。

在 d 维欧氏空间中，刻度为 rL 的物体落入刻度为 L 的空间单元中的数目可表示为

$$N = r^{-d}$$

例如当 $r=1/3$ 时，那么在一维空间中有三条长度为 $L/3$ 的线段落入长度为 L 的线段，此时 $N=3$，上述等式成立；同样，当 $r=1/3$ 时，在二维空间中有 9 个边长为 $L/3$ 的正方形落入边长为 L 的正方形，这时 $N=9$，因此上述等式也成立。此外，在三维空间中有 27 个边长为 $L/3$ 的立方体落入边长为 L 的立方体，从而表明上式也成立。显然，当 $r=1/4$ 等等时，上述等式也是成立的。

一个分数维图形不会充满其嵌入的空间 R^d，可以用如下形式的刻度关系来描述

$$N = r^{-D}$$

这里，D 即为分数维图形的维数，且有 $D<d$。Sierpinski 垫片就是一个典型的分数维图形。把单位边长的等边三角形分成四个边长为 1/2 的小三角形再取掉中心的那一个，并连续做下去，即可得到 Sierpinski 垫片（如图 4.7）。

这样，上式中 N 是 3，而 r 是 1/2，再根据以上等式有

$$D = -\frac{\ln N}{\ln r} = \frac{\ln 3}{\ln 2} = 1.585$$

因此，Sierpinski 垫片的分维数 $D=1.585$，且有 $1<D<2$。显然，该图形还有自相似性，即它的任何一个小三角形和原始的那个三角形有相同的结构。

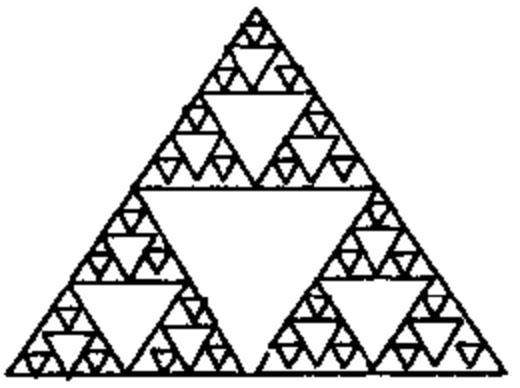


图 4.7　Sierpinski 垫片示意图

以上已经指出，在实际中，许多不规则的具有复杂结构的图形并不具有严格的自相似性，并和典型的分数维图形有明显的差别，因此需要用统计方法进行处理。这就是所谓的随机分数维统计学。此时，需要使用分数维的概率定义：在刻度为 L 的分数维图形中任意一点落入刻度为 l 的一小部分的概率，等于总空间被尺度为 l 的那部分所占据的比例，即 $\mathrm{Prob}(l) = (l/L)^{d-D}$。这里 $d-D$ 称为协同维数，它度量了分数维图形占据空间的程度。

虽然较小的分数维单元在个数上是相当多的，然而和较大单元相比，它们对分数维图形分布密度的贡献是比较小的。变异性的网格结构分析并未考虑到这种空间填充的程度，常常过高估计了小尺度的变异性对大尺度上总的变异性的贡献。

具有分数维分布的油藏参数 $Z(x)$ 的变化具有如下形式：

$$\mathrm{Prob}\left\{\frac{Z(x+l)-Z(x)}{l^H} \leq y\right\} = F(y) \tag{4.12}$$

这里 $F(y)$ 是分布函数，H 依赖于协同维数和 $F(y)$ 的形式。当 $F(y)$ 是误差分布函数，即具有零均值与单位方差的正态分布函数，以及 $H=1/2$ 时，关系式 (4.12) 定义了熟知的布朗运动过程。这个过程可看成是不相关的高斯随机变量之和。当 $H\neq 1/2$，且 $F(y)$ 是误差分布函数时，该关系定义了一类随机函数，称为分数维布朗运动；$H=d-D$。当 $F(y)$ 是偏态分布函数时，那么 H 与 $d-D$ 的关系可通过矩生成函数来确定。特别地，当 $F(y)$ 是对数正态分布函数时，有 $H/2=d-D$。关于分数维布朗运动的性质已有许多文献进行了讨论。分数维布朗运动可看成随机分数维函数的原型。在一些特定的情况下，例如在描述渗透率分数维分布的时候，需要将 $F(y)$ 推广到非高斯分布的情况。

随机分数维变异函数可定义如下：

$$2\gamma(l) = E\{[Z(x+l)-Z(x)]^2\} = V_H \cdot l^{2H} \tag{4.13}$$

可把增量［$Z(x+l)-Z(x)$］满足以上形式的分布称为自相似的。这是因为在任意尺度 rL 上的变异性与尺度 l 上的变异性之间的关系可表示如下：

$$\gamma(rl)=r^{2H}\cdot\gamma(l)$$

这说明在任何尺度上的方差可以从其它尺度上测量的方差而求取。在不同尺度上函数的样本，经过放大或缩小后，会变得相互类似。(4.13) 式揭示了分数维随机函数其变异函数具有幂指数函数的特征，这也是分数维随机函数的显著特点。

在 d 维空间中，随机分数维函数可写成如下的一般形式：

$$2\gamma(|\vec{r}_2-\vec{r}_1|)=E\{[Z(\vec{r}_2)-Z(\vec{r}_1)]^2\}=V_H\cdot|\vec{r}_2-\vec{r}_1|^{2H}$$

这说明在 d 维空间中，分数维布朗运动具有相似性指数 H，其分布在 $d+1$ 维空间中（指 $Z(x)$ 关于 x 的图形），分布的分数维维数为 $D=d+1-H$。

通过考察相继增量的相关性可以看出：对 $0<H<1/2$，$H=1/2$，$H>1/2$ 三种情况，分数维图形有完全不同的性质。令 $Z(0)=0$，可以定义过去增量为 $-Z(-l)$，将来增量为 $Z(l)$，而相继增量的相关性定义为：

$$\frac{E[-Z(-l)Z(l)]}{E\{[Z(l)]^2\}}=2^{2H-1}-1$$

显然此相关性独立于 l。当 $H=1/2$，那么相关性为零，就相当于布朗运动具有独立增量。当 $H<1/2$，该相关性为负。当 $H>1/2$，相关性为正。这表示了局部区域变异的持续性，其强弱可用 H 度量。当 $H>1/2$ 时，分数维布朗运动所显示的变异的持续性可以在水文地质或沉积环境的垂向变化中观测到。通过观测可知，储层物性参数值具有聚类特征，而并非在空间中是完全随机分布的。在砂泥岩序列中平均孔隙度和渗透率是聚集于砂体中的高值和泥岩中低值的平均。虽然孔隙度和渗透率的平均值很大程度上依赖于高值，但对于预测地层流体性质而言仅有这点认识是远远不够的，重要的是了解高值和低值是如何在储层中聚集在一起的。H 值恰好提供了高值区空间填充情况的一种度量，以及该种区域的规模。最后，需要特别强调的是：分数维随机函数是一种连续的，仿射相似的随机过程，且具有高斯增量，但其本身并不是高斯过程。

2. 分数维维数的确定

首先对研究对象进行结构分析，确定其变异函数：$2\gamma_Z(l)=E\{[Z(x+l)-Z(x)]^2\}$。该变异函数本身就是一个随机函数，可以用实验变异函数进行估计：

$$2\gamma_Z^*(l)=\frac{1}{N}\sum_{i=1}^{N}[Z(x_i+l)-Z(x_i)]^2。$$

这里 N 是相距为 l 的数据点对的数目。相应的估计方差

$$E\{[\gamma_Z(l)-\gamma_Z^*(l)]^2\}=4\gamma_Z(l)\cdot\sigma_z^2/N \tag{4.14}$$

这里 σ_z^2 为数据总的方差。

当 l 较小，N 较大时，变异函数的估计较为准确。当 l 达到所研究区域尺寸的一半时，利用实验变异函数进行估计就显得不可靠。这主要是由于仅使用了随机函数的单个实现，从而会出现涨落方差。当与 l 超过区域尺寸的一半时，这种涨落方差可能变得很大。由于分数维分布揭示了各种尺度上的相关性，可以把实验变异函数的利用限制在刻划小尺度的相关结构的范围内，这样不仅可保证效果，又不失一般性。

确定分数维维数的方法并不唯一。其中一种就是通过确定相似性指数 H 来求取。可先把 $\gamma^*(l)$ 用对数坐标描绘成散点图，然后通过确定相应直线的斜率而得到 H 值。然而由

由于 γ^*（l）不能正确反映出大尺度上的相关性，即当 l 较大时，γ^*（l）不甚准确。因此用这种方法得到的 H 值是近似的。

4.4.2　分数维场的模拟

在二维空间中对分数维分布的随机函数的模拟可采用序贯窗口屏蔽方法（Omre, 1990）。该方法能较好地模拟分数维分布，且快速实用。具体操作时，对（d^n+1）×（d^n+1）的网格系统的节点进行二维的分数维模拟，步骤如下：

1）确定一系列主方向 $\{\phi_i,\ i=1,\ 2,\ \cdots,\ N\}$，例如可选取 0，$\pi/4$，$\pi/2$，$3\pi/4$。如图 4.8。

2）采用传统的序贯方法对区域 D_2 的边界进行初始化，先选定 z_{01}，其次取 $z_{02}\sim N(\mu_{2|1},\ \sigma^2_{2|1})$；然后再取 $z_{03}\sim N(\mu_{3|2,1},\ \sigma^2_{3|2,1})$，最后是取 $z_{04}\sim N(\mu_{4|3,2,1},\ \sigma^2_{4|3,2,1})$，其中

$$\mu_{i|i-1,\cdots,1}=E[z_i\mid z_{i-1},\cdots,z_1]$$

$$\sigma^2_{i|i-1,\cdots,1}=Var(z_i\mid z_{i-1},\cdots,z_1)$$

3）用折半方法加密网格后，每 t 步网格系统（$2'+1$）×（$2'+1$）的节点模拟值，是从正态分布 $N(\mu_{\cdot|\cdot},\ \sigma^2_{\cdot|\cdot})$ 中进行抽样而得到的。对 $\mu_{\cdot|\cdot\cdot}$和 $\sigma^2_{\cdot|\cdot\cdot}$需要分两种情况进行确定，首先，中心位置处的数值是由该中心周围的四个角点上的值来确定，由于这四个角点都是前一步网格系统的节点，故其值是已知的；其次，各边中点处的数值是由所在那条边的两个端点处的值来确定。例如在图 4.8 中，中心 o 处的 $\mu_{\cdot|\cdot}$和 $\sigma^2_{\cdot|\cdot}$是从 z_{01}，z_{02}，z_{03}和 z_{04}四个位置处的模拟值所确定的，而 A，B，C，D 四个点的 $\mu_{\cdot|\cdot}$和 $\sigma^2_{\cdot|\cdot}$的值是通过离该点最近的两个角点来确定的。因而，可用下式来确定第 t 步网格系统节点处的模拟值

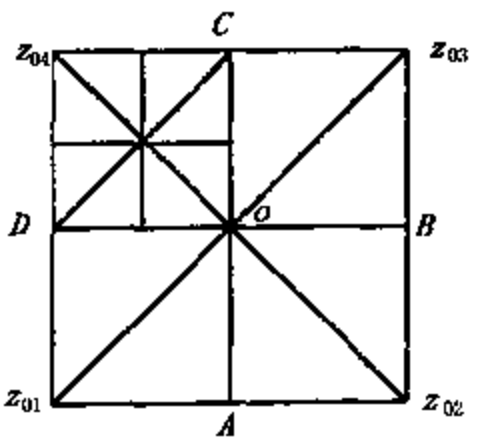


图 4.8　序贯层次窗口模拟示意图

$$Z_{ij}=z_{ij}\sim N(\mu_{\cdot|\cdot},\sigma^2_{\cdot|\cdot})$$

4）当 t 增大到 n 时，可得到分数维场在所有网格节点处的模拟值 z_{ij}。

第5章　指示模型

指示模型是随机建模理论的一个主要组成部分。该理论最早是由美国斯坦福大学石油工程系的 Journel 教授在 80 年代提出的。随后，指示化方法及其模型得到了快速发展。指示模型的基本思想就是将地质信息进行离散的编码，通常编码成 0 与 1 两值的指示变量，然后将克里金的基本思想用于指示变量，最终得到指示变量的克里金估计。而指示变量的克里金估计恰好给出了未知位置变量的概率分布的估计。该概率模型就是对未知位置变量不确定性的度量。基于该后验概率概型，可以模拟变量的分布。因此，基于指示模型的各种模拟方法，核心是各种指示克里金估计模型。不同的克里金估计模型，使用的信息类型不同，从而导致不同的后验概率模型。随后的模拟就基于这些后验概率模型。模拟的过程和方法基本上是一样的。通常使用的模拟思想是序贯指示模拟思想。下面首先具体讨论三种基本的指示模型，它们分别是简单指示克里金模型（Simple IK），普通指示克里金模型（Ordinary IK）和中位指示克里金模型（Median IK）。

5.1　单变量指示模型

5.1.1　指示模型的基本理论

在第 3 章中，讨论了地质变量的指示化方法，对于地质变量 $\{Z(x) \mid x \in D\}$，无论它是连续的还是离散的，都可以对其离散化而得到一系列指示变量。若 $\{Z(x) \mid x \in D\}$ 是连续的地质变量，那么通过 K 个截断值 z_1，z_2，…，z_k，将 $Z(x)$ 可能的取值区间截断后得到 K 个指示变量：

$$I(x;z_k)=\begin{cases}1,\text{若 } Z(x)\leq z_k\\0,\text{其他}\end{cases}$$

在此，$k=1$，2，…，K。

指示克里金（IK）方法可以提供地质变量 $\{Z(x) \mid x \in D\}$ 在一个待估位置 x_0 处的条件累积概率分布函数（$ccdf$），在不同截断值 z_k 处的最小均方误差估计为

$$\begin{aligned}[i(x;z_k)]^*_{IK}&=E\{I(x;z_k)\mid(n)\}^*\\&=\text{Prob}^*\{Z(x)\leq z_k\mid(n)\}\quad k=1,2,\cdots,K\end{aligned}\tag{5.1}$$

指示克里金过程可以对一系列截断值 z_k，$k=1$，2，…，K 重复进行，这相应于求解 K 个克里金方程组。来自于（5.1）的 K 个值，相当于对未抽样位置 x 处分布函数 $F(z)$ 在 K 个点处 z_1，z_2，…，z_k 的离散估计，从而可得到关于未抽样位置处 $Z(x_0)$ 的不确定性的概率模型。

如果 $\{Z(x) \mid x \in D\}$ 本身是一个离散的地质变量，则此时无须进行截断处理，只对其进行 0－1 编码标定就可以了。若研究区域 D 内有 K 种岩相，则可定义 K 个指示变量：

$$I_k(x)=\begin{cases}1,\text{若位置 } x \text{ 处有指定的岩相 } k\\0,\text{其它}\end{cases}$$

在此，$k=1$，2，…K。相应的直接指示克里金估计就可以给出未抽样位置 x 处岩相的概率

估计：

$$[i_k(x)]^*_{IK} = E\{I_k(x) \mid (n)\}^* = \text{Prob}^*\{I_k(x) = 1 \mid (n)\} \tag{5.2}$$

而来自于（5.2）的 K 个概率值就是岩相的概率模型。

由此可见，对于指示模型来讲，指示克里金方法起核心作用。通过它可以得到后验的指示变量的概率（5.1）或（5.2），而后验概率是模拟的基础。所以，通常讲克里金估计理论是随机建模的工具。在随机建模过程中，克里金方法常贯穿于其中。通过求解克里金方程组可以得到后验的累积分布函数，并为模拟作准备。对照第四章也会发现，对高斯场模型，通过求解克里金方程组可得到高斯模型两个重要函数，均值和方差。而这两个函数完全可以确定高斯场分布。

对单变量指示模型来讲，确定其后验概率模型，通常用下面三种常用的指示克里金方法。

1. 简单指示克里金

首先，需要指出的是指示变量 $I(x; z)$ 的平稳均值是随机函数 $\{Z(x) \mid x \in D\}$ 的累积分布函数。这是由于

$$\begin{aligned} E\{I(x; z)\} &= 1 \cdot \text{Prob}\{Z(x) \leq z\} + 0 \cdot \text{Prob}\{Z(x) > z\} \\ &= \text{Prob}\{Z(x) \leq z\} = F(z) \end{aligned} \tag{5.3}$$

因此，在平稳假设下 $I(x; z)$ 的简单指示克里金估计（简记为 SIK）可写成：

$$\begin{aligned} [i(x; z)]^*_{SIK} &= [\text{Prob}\{Z(x) \leq z \mid (n)\}^*]_{SIK} \\ &= \sum_{\alpha=1}^{n} \lambda_\alpha(x; z)\, i(x_\alpha; z) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha(x; z)\right] F(z) \end{aligned} \tag{5.4}$$

在此，$\lambda_\alpha(x, z)$ 是相应于截断值 z 的简单克里金的加权系数。这些权系数可以通过求解简单指示克里金方程组相应给出：

$$\sum_{\beta=1}^{n} \lambda_\beta(x; z)\, C_I(x_\beta - x_\alpha; z) = C_I(x - x_\alpha; z) \qquad (\alpha = 1, 2, \cdots, n) \tag{5.5}$$

在此，$C_I(\vec{h}; z) = Cov\{I(x; z), I(x + \vec{h}; z)\}$ 是在截断值 z 处指示协方差。当然，在（5.5）中，亦可用指示变异函数 $\gamma_I(h; z)$ 代替指示协方差 $C_I(h; z)$。如果使用 K 个截断值 z_k，那么简单指示克里金需要 K 个指示协方差 $C_I(h; z_k)$ 和 K 个累积分布函数值 $F(z_k)$。

2. 普通指示克里金

正如简单克里金的定义，简单指示克里金依赖于平稳假设，即认为累积分布函数值 $F(z)$是平均值。当数据较丰富时，普通克里金采用滑动邻域方法来处理。此时可认为在不同位置对其先验累积分布函数值 $F(z)$ 进行重新估计。

此时，(5.4) 式成为

$$\begin{aligned} [i(x, z)]^*_{OIK} &= [\text{Prob}\{Z(x) \leq Z \mid (n)\}]^*_{OIK} \\ &= \sum_{\alpha=1}^{n} \lambda_\alpha(x; z)\, i(x_\alpha; z) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha(x; z)\right] m^*(x) \end{aligned} \tag{5.6}$$

在此，$m^*(x)$ 是利用邻域中数据对邻域中平均值进行估计所得的值，从而（5.6）又可写成：

$$[i\ (x,\ z)]^{*}_{\mathrm{OIK}}=\sum_{\alpha=1}^{n}\gamma_{\alpha}\ (x;\ z)\ i\ (x_{\alpha};\ z) \tag{5.7}$$

在此，$\gamma_{\alpha}\ (x;\ z)$ 为普通指示克里金的加权系数，满足 $\sum_{\alpha=1}^{n}\gamma_{\alpha}\ (x;\ z)\ =1$，该权系数可通过下列普通指示克里金方程组求解得到：

$$\begin{cases}\sum_{\beta=1}^{n}\gamma_{\beta}\ (x;\ z)\ C_I\ (x_{\beta}-x_{\alpha};\ z)\ -\mu\ (x;\ z)\ =C_I\ (x-x_{\alpha};\ z), & \alpha=1,\ 2,\ \cdots,\ n\\ \sum_{\beta=1}^{n}\gamma_{\beta}\ (x;\ z)\ =1\end{cases} \tag{5.8}$$

在此，$\mu\ (x;\ z)$ 是与限制 $\sum_{\beta-1}^{n}\gamma_{\beta}\ (x;\ z)\ =1$ 相联系的拉格朗日乘数因子。比较（5.5）和（5.8）可知，简单指示克里金和普通指示克里金的权系数是不一样的。

3．中位指示克里金

在指示克里金中，K 个截断值 z_k 的选取常使得相应的指示协方差 $C_I\ (\vec{h};\ z_k)$ 互不相同。然而也有这样的情况，那就是样本指示协方差或变异函数似乎相互成比例，即样本指示相关函数几乎具有相同的形状。此时，称连续随机函数模型 $[Z\ (x)\ |x\in D]$ 具有镶嵌结构，即

$$\rho_z\ (\vec{h})\ =\rho_I\ (\vec{h};\ z_k)\ =\rho_I\ (\vec{h};\ z_k,\ z_k'),\ \text{对任何}\ z_k,\ z_k' \tag{5.9}$$

在此，$\rho_z\ (\vec{h})$ 和 $\rho_I\ (\vec{h};\ z_k,\ z_k')$ 都是相关函数，它们分别相应于连续随机函数 $|Z\ (x)\ |x\in D|$ 和其指示变换的相关函数。

因此，如果直接通过样本相关函数或通过样本指示相关函数在中位截断 $z_k=M$，$F\ (M)\ =0.5$ 就可以对单个相关函数给予更好的估计。的确，在中位截断可以使指示数据更好地展布在 0 与 1 之间，无异常值出现。

在模型（5.9）之下的指示克里金称为中位指示克里金，可简记为 MIK。在这种情况下，可以使指示克里金过程更简单快速。这由于对于 K 个截断值只须调用一个单一的易推断的中位指示变异函数。另外，如果指示数据构形对所有截断值是相同的，那么可通过求解一个单一的指示克里金。得到的系数对所有其它截断值也可以使用。例如，在简单指示克里金的情形：

$$[i(x;z_k)]^{*}\ =\ \sum_{\alpha=1}^{n}\lambda_{\alpha}(x)i(x_{\alpha};z_k)+[1-\sum_{\alpha=1}^{n}\lambda_{\alpha}(x)]F(z_k) \tag{5.10}$$

在此，$\lambda_{\alpha}\ (x)$ 是简单指示克里金的加权系数。它们对所有截断值 z_k 是相同的，并通过下面单一的简单克里金的加权方程组确定

$$\sum_{\beta=1}^{n}\lambda_{\beta}(x)C(x_{\beta}-x_{\alpha})\ =\ C(x-x_{\alpha}),\quad \alpha\ =\ 1,2,\cdots,n$$

协方差 $C\ (\vec{h})$ 可通过 z 的样本协方差建模，或更好的办法是通用样本中位指示协方差建模。

5.1.2　连续变量的指示模型

前面已对连续变量 $[Z\ (x)\ |x\in D]$ 指示化及相应的 IK 方法给予了讨论，本小节将

对连续变量指示化优点及它模拟方法给予进一步讨论。为了得到一个连续变量 $[Z(x) \mid x \in D]$ 的空间分布，可以将其离散成 K 个互相排斥的类 $S_k = \{x \mid Z(x) \in (z_{k-1}, z_k]\}$，$k=1, 2, \cdots, K$，并将分布解释和模拟成 K 个类指示值的空间分布。类间分辨率的损失可以通过一些预先指定的类间分布，如指数模型进行部分恢复。

把一个连续变量 $Z(x)$ 变成 K 个镶嵌类的优点就在于通过不同指示变异函数对每个类的空间分布进行建模，这样的模型会具有更强的适应性。前面一章谈到了在一些情况下可利用高斯场对连续变量 $Z(x)$ 进行建模，但是高斯场模型是一种参数化的模型。该模型完全由均值和方差两个系数完全确定。在一些情况下，高斯模型并不能揭示变量 $Z(x)$ 空间分布结构，必须用指示模型给予刻画。指示模型的优点在于它是直接估计未知位置处的地质变量的分布，而不假定高斯性，从而具有更强的适应性。因此，指示化方法是一种非参数化的建模方法。但是需要指出的是，该模型会引起另外一些问题，如需要对 K 个指示协方差函数进行推断，从而过程要复杂一些。另外一个问题就是利用后验分布在若干个截断点处的估计值要恢复其形状，会出现次序校正问题。

纯粹的范畴总体不具有次序性，与之形成对照的是一个连续变量 $Z(x)$ 的类 $\{x \mid Z(x) \in (z_{k-1}, z_k]\}$，$k=1, 2, \cdots, K$ 必须序贯地排列。由于这个理由，刻画这些类最好使用累积指示值。这也就是前面讲过的连续变量的指示化方法。

$$I(x, z_k) = \begin{cases} 1, \text{如果 } Z(x) \leq z_k \\ 0, \text{其它} \end{cases}$$

类 $S_k = \{x \mid Z(x) \in (z_{k-1}, z_k)\}$ 通过乘积 $I(x; z_k)[1 - I(x, z_{k-1})] = 1$ 来指定。

除去第一个和最后一个类，累积的类指示变异函数比类指示变异函数更易于推断。特别是当类具有较小的边缘概率时更是如此。使用累积指示值（称为 cdf 型指示值）代替类指示值（称为 pdf 型指示值）可以避免使用协同克里金。最后，累积指示值是直接与研究的连续变量的累积分布函数相联系的（Journel，1992）。

若使用 SIK 或 OIK 可得到每个待估位置处的累积分布函数估计值，从而得到后验的累积分布函数，那么可使用序贯指示模拟思想（Sequential Indicator Simulation，简记为 SIS）得到研究区域 D 内每个网格结点模拟值。具体方案如下：

在每个要模拟的网格结点沿着随机路径，通过指示克里金利用 K 个概率估计出累积分布函数的值。$F^*[x; z_k(n)] = \text{Prob}^*\{Z(x) \leq z_k \mid (n)\}$，$k=1, 2, \cdots, K$。通过类间内插，可对所有的 $z \in [z_{\min}, z_{\max}]$ 的累积分布函数提供连续的光滑处理。在此，条件数据 (n) 包括指定的邻域中原始指示数据和已模拟过的网格结点处的指示数据。

网格结点处一个实现 $z^{(l)}$ 可以这样得到：抽取一个均匀随机数 $p^{(l)} \in [0, 1]$，然后求取累积分布函数的 $p^{(l)}$ 分位数

$$z^{(l)}(x) = F^{*(-1)}[x; p^{(l)} \mid (n)], \tag{5.11}$$

使得

$$F^*[x; z^{(l)}(x) \mid (n)] = p^{(l)}$$

然后对指示数据集更新模拟值。对另外的网格位置 x'，沿着随机路径再使用指示克里金用同样方法得到模拟值。一旦所有的位置都已模拟，就可得到随机图像 $[z^{(l)}(x) \mid x \in D_L]$。在此，$D_L$ 为研究区域 D 的网格系统。

5.1.3　离散变量的指示模型

设 $\{L(x) \mid x \in D\}$ 是离散化地质变量，其中 $L(x)$ 可能取值只有有限个整数 $\{l_1,$

l_2，…，l_k|。这相当于区域 D 内定义了 K 个互相排斥的范畴 S_k，$k=1$，2，…，K。其中，$S_k=\{x|L(x)=l_k\}$。这 K 个范畴代表了区域 D 内某种离散地质变量（如岩相、沉积相等）的分布，是完备的。即任何位置 $x\in D$ 只能属于这 K 个范畴中的一个。

利用指示化方法，可以定义类 S_k 的指示变量

$$I_k(x)=\begin{cases}1,\text{当 } x\in S_k\\0,\text{其它}\end{cases}\qquad k=1,2,\cdots,K$$

显然，这 K 个指示变量具有相互排斥性和完备性

$$I_k(x)\cdot I_{k'}(x)=0\quad \text{当 } k\neq k' \text{ 时}$$

$$\sum_{k=1}^{K} I_k(x)=1$$

对指示变量 $I_k(x)$ 直接使用克里金方法就可得到位置 x 属于范畴 S_k 的概率估计。如使用简单指示克里金，相应于（5.4）在离散情形下的简单指示克里金估计可写成：

$$\text{Prob}^*\{I_k(x)=1|(n)\}=\sum_{\alpha=1}^{n}\lambda_\alpha(x)i(x_\alpha)+\left[1-\sum_{\alpha=1}^{n}\lambda_\alpha(x)\right]\cdot p_k \qquad (5.12)$$

在此，$p_k=E\{I_k(x)\}$ 是范畴变量 S_k 的边缘概率。加权系数 $\lambda_\alpha(x)$ 的确定是通过使用具有范畴 S_k 的指示协方差模型的简单指示克里金给出的。

当边缘概率 p_k 变化比较光滑的时候，可考虑使用 OIK 系统。利用序贯指示模拟思想，可给出离散的范畴变量的模拟结果。具体过程如下：

1）对每一网格结点 x，沿着随机路径，由通过次序关系校正的克里金提供了 K 个估计概率 $P_k^*[x|(n)]$，$k=1$，2，…，K。条件信息（n）由原始的 i_k 数据和先前模拟的 i_k 值（关于范畴 S_k）组成。

2）确定任意次序的 K 个范畴，如 S_1，S_2，…，S_k，这个次序决定了概率区间［0，1］上的累积分布函数刻度，即存在 K 个区间。

$$[0,P_1^*],(P_1^*,P_2^*+P_1^*],\cdots,(1-\sum_{k=1}^{K-1}P_k^*,1]$$

3）从［0，1］区间中抽取一个均匀分布的随机数 p。p 所属的区间就决定了在位置 x 处模拟的范畴。把所有 K 个指示数据集更新成附加的新的模拟信息，并沿着随机路径处理下一个位置 x'。这样依次处理会得到研究区域 D 内网格系统上的模拟实现。

需要指出的是，在指示克里金方法中，范畴 S_k 的条件概率仅考虑 i_k 指示数据，即保留的信息就是一个数据在位置 x_α 属于或不属于 S_k。常常有这样的情况，与数据 $i_k(x_\alpha)=1$ 比较接近的位置会导致在该位置范畴 S_k 以更高的概率出现。但交互相关性在指示克里金中并没有加以考虑。因此，为了更充分利用 K 个范畴信息，需要考虑它们之间的相互影响和关系，这就是协同模拟思想。下一节就讨论这方面的内容。

5.2　协同指示模型

5.2.1　一般的协同指示模型

指示克里金并没有完全使用包含在原始数据 Z 中的信息。在连续变量的情形，指示克里金只考虑数据 Z 是否超过某个截断值 z_k。在 k 个范畴变量的情形，指示克里金方法只能一次考察一种范畴的信息。

为了克服上述缺点，一个可供选择的方法就是对每一个累积分布函数值的估计同时考虑 k 个指示数据 i（x_α；z_k），$k=1$，2，…，K。此时，相应的估计称为协同指示克里金（Indicator Cokriging），简记为 COIK。该估计值定义为：

$$[i(x,z_{k0})]^*_{\mathrm{COIK}} = \sum_{k=1}^{K}\sum_{\alpha=1}^{n}\lambda_{\alpha,k}(x,z_{k0})\cdot i(x_\alpha,z_k) \tag{5.13}$$

相应的协同克里金方程组需要 K^2 个直接的和交互的指示协方差组成的矩阵。它们的表达式为

$$C_I(h;z_k,z_{k'}) = \mathrm{Cov}\{I(x;z_k),I(x+h;z_{k'})\}$$

当 $k=k'$时，为直接指示协方差，当 $k\neq k'$时，为交互指示协方差。对 K^2 个协方差进行推断和联合建模，当 K 较大时，是难以实现的。

使用协同克里金经验表明，当主变量与次变量在空间具有相同抽样位置的时候，它并不比克里金结果有太多的改进。如对定义在各个截断值上的指示数据使用协同指示克里金的时候就会出现这种情况。另外，当使用连续变量的时候，从一个截断值变到另一个截断值，相应的累积指示数据确实带走了很多本质的信息。在这种情况下，使用指示克里金或协同指示克里金损失的信息差不多。最后需要指出的是使用协同指示克里金也存在次序校正问题（关于次序校正问题见第 5.4 节）。

在实际问题中，更常使用的并不是指示协同克里金方法，而是指示主成分克里金（Indicator Principal Component Kriging），简记为 IPCK。该模型具有快速实用的特点，在建立岩相或沉积相带联合概率模型方面有独到之处。它与指示克里金相比，使用了更多的包含在原始数据中的信息。

5.2.2　指示主成分模型

指示主成分克里金 IPCK 是通过定义在某个分隔向量 $\vec{h}_0$（常选为 0 或较小的模向量）上的指示协方差阵 $[C_I(\vec{h}_0;z_k,z_{k'}),k,k'=1,\cdots,K]$ 的主成分进行克里金估计而得到。这 K 个主成分 $Y_k(x)$，$k=1$，2，…，K 是原始 K 个指示变量的线性组合：

$$Y_k(x) = \sum_{k'=1}^{K} a_{k,k'}I(x;z_{k'}) \tag{5.14}$$

即 $Y=A^T\cdot I$。在此，$K\times K$ 的矩阵 $A^T=[a_{k,k'}]$ 是正交矩阵 A 的转置。A 来自于协方差矩阵的谱分解

$$[C_I(\vec{h}_0;z_k,z_{k'})] = A\,\Lambda\,A^T \tag{5.15}$$

Λ 是一个由指示协方差阵的 K 个特征值组成的对角矩阵。

这 K 个主成分，通过以上的构造方法可知是正交的（相应于分隔向量 h_0），即

$$\mathrm{Cov}\{Y_k(x),Y_{k'}(x+\vec{h}_0)\} = 0,\text{当 } k\neq k'$$

指示主成分克里金方法需要将正交性推广到任意的分隔向量 $\vec{h}$ 上去，即 $\mathrm{Cov}\{Y_k(x),Y_{k'}(x+\vec{h})\}\cong 0$，对任意 $\vec{h}$，$k\neq k'$。这样，K 个主成分 $Y_k(x)$ 的协同克里金就归结为它们单个的独立的克里金，大大减少了推断和计算的时间。

从 K 个指示主成分 $y_k^*(x;z_k)$，$k=1$，2，…，K 的克里金估计开始，然后使用线性逆变换可得到相应的关于指示值的近似无偏估计。这些估计实际上为 ccdf 模型。

$$[i(x;z_k)]^*_{\text{IPCK}} = \sum_{k'=1}^{K} a_{k',k} \cdot y^*_{k'}(x) \tag{5.16}$$

即

$$I^* = A \cdot Y^*$$

使用该模型的经验表明，当变量 Z 是连续变量的时候，结果并不比直接使用指示克里金会有明显改善。然而计算花费的CPU时间会大大减少。

指示主成分克里金的主要应用就在于评价 K 个范畴变量（$K>1$）的累积分布函数值。通过实例分析研究表明，指示主成分克里金在评价储层空间相带分布方面显示出自身的优点，那就是快速实用。模拟结果确实能较好地反映相带的空间分布及接触关系。

使用指示主成分模型需要形式为（5.15）的正交矩阵的分解，从而提供关于主成分变换（5.14）和逆变换（5.16）。这种分解在一般实用程序库中都可找到。

最后，需要指出的是指示主成分克里金模型本身也具有两个缺点。其一就是使用主成分 y_k（x）而不用原始范畴指示值。在变异函数建模阶段，主成分的实际意义往往不清楚，从而导致难以抓住空间范畴变量的连续性和各向异性信息。其二就是对指示主成分各向异性分析的偏差会导致对各种范畴累积分布函数模型各向异性的偏差。

5.3　软硬信息相结合的指示模型

5.3.1　软信息与硬信息

指示模型的主要优点在于其可产生后验的条件累积分布函数，这使得指示模型可以考虑软数据。在建立定量地质模型的过程中，有一部分数据来自于较为准确的测井测量和取心分析数据，称为硬数据。而另一部分信息却来自于不甚准确测量和地质经验的数据。它们往往可以提供研究变量的可能的取值范围，走势以及与其它信息的相关情况。这种类型的数据称之为软数据。建立油藏定量模型常常需要软硬数据充分结合，硬数据为主要骨架，而软数据为硬信息的必要补充。只有这样，才能使地质建模过程更趋于合理（Journel，1992）。

1. 硬信息的类型

硬信息是相应于特定的地质变量而言的。所研究的地质变量为 $\{Z(x) \mid x \in D\}$，则来自于抽样位置 $\{x_\alpha \mid \alpha = 1, 2, \cdots, n\}$ 的较为准确的关于变量 Z 的观测数据 $\{z(x_\alpha) \mid \alpha = 1, 2, \cdots, n\}$ 称为关于 Z 变量的硬信息。这种观测数据可以是测井解释数据或岩心分析数据，也可以是地震数据。另外，其它较为准确的辅助变量观测值 $\{z_A(x_\beta) \mid \beta = 1, 2, \cdots, m\}$，若其它辅助地质变量 $\{Z_A(x) \mid x \in D\}$ 和所研究的地质变量 $\{Z(x) \mid x \in D\}$ 之间有较为明确的互换关系，使得来自于辅助变量 Z_A 的较为准确的观测信息能转换成变量 Z 的硬信息，那么来自于辅助变量 Z_A 的信息也称为硬信息。由此可见，只要能准确标定所研究地质变量取值的信息都称为硬信息。来自于硬数据的指示信息称为硬指示信息。

2. 软信息的类型

软信息也是相应于特定的地质变量 $\{Z(x) \mid x \in D\}$ 而言的。所有不能提供精确的关于变量 Z 值，而只能提供关于变量 Z 的一些概率分布信息的数据都称为软信息。众所周知，概率分布是一种不确定性的度量，所以软信息是一种以概率方式描述的不确定性信息。例如来自于地质经验或其它辅助变量所提供的关于变量 Z 的信息。它们不能准确地告知变量 Z 在一些特定位置的值，但是可以提供数据 Z 的可能的分布信息。软信息包括地质人员所概括的一些统计规律、经验认识。

图 5.1 和图 5.2 给出了硬信息和软信息的示意图。对一个特定位置 x_a，硬信息提供精确观测值 $Z(x_a)$，从图形上看是一条直线。而软信息提供的是关于 Z 值的概率分布密度曲线，该曲线形态揭示了信息强度。当曲线较为平坦时，说明软信息不确定性很大。当曲线较为陡峭时，说明软信息不确定性很小，更接近于硬信息。特别极端的情况是概率曲线退化成一条直线，说明软信息成了硬信息。从这两幅图可以看出硬信息与软信息的区别与联系。

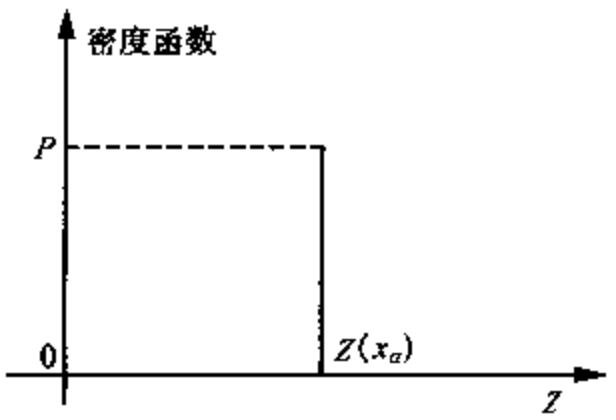


图 5.1　硬信息示意图

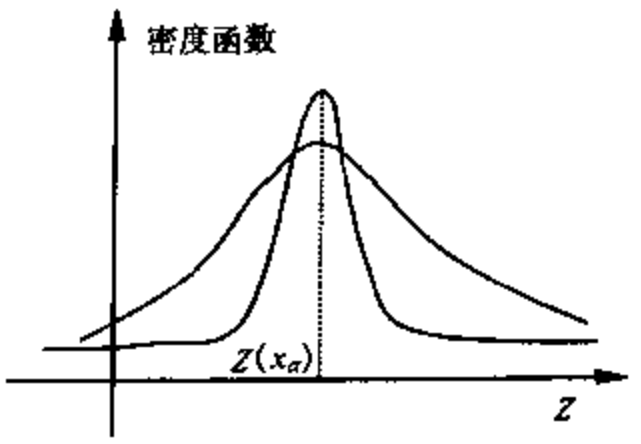


图 5.2　软信息示意图

5.3.2　马尔柯夫—贝叶斯模型

1. 模型引入

该模型的优点在于它可以将硬信息和软信息结合起来作为先验概率，利用指示协同克里金方法得到后验概率模型（Journer and Zhu, 1990）。利用前面关于软硬信息含义，常用的先验信息可以为下列四种形式之一：

第一，来自于局部硬数据 $Z(x_a)$ 的局部硬指示数据 $i(x_a; z)$：

$$i(x_a;z) = \begin{cases} 1, \text{当 } z(x_a) \leq z \\ 0, \text{其它} \end{cases} \tag{5.17}$$

或在范畴变量情形

$$i(x_a) = \begin{cases} 1, \text{当 } x_a \in \text{范畴 } S_k \\ 0, \text{其它} \end{cases} \tag{5.18}$$

第二，来自于辅助信息的局部硬指示数据 $j(x_a; z)$，可提供关于局部 $z(x_a)$ 的不等式约束，如果 $z(x_a) \in (a_a, b_a]$，则

$$j(x_a;z) = \begin{cases} 0 & \text{当 } z \leq a_a \\ \text{不定(缺失)} & \text{当 } z \in (a_a, b_a] \\ 1 & \text{当 } z > b_a \end{cases} \tag{5.19}$$

第三，来自于辅助信息的局部软指示数据 $y(x_a; z)$，可提供关于值 $z(x_a)$ 的先验概率

$$y(x_a;z) = \text{Prob}\{Z(x_a) \leq z \mid \text{局部信息}\} \in (0,1] \neq F(z) \tag{5.20}$$

在此，$F(z)$ 为整体先验信息，$E[Y(x; z)] = F(z)$。

第四，在平稳域 D 内所有位置 x 处整体的先验信息

$$F(z) = \text{Prob}\{Z(x) \leq z\} \quad x \in D \tag{5.21}$$

在任何位置 $x \in D$，关于值 $z(x)$ 的先验信息可由上述四种先验信息之一刻画。指示克里金过程可以看成是应用局部累积分布函数提供的信息，使用贝叶斯方法将局部累积分布函数更新成后验的条件累积分布函数的过程

$$[\text{Prob}\{Z(x) \leq z \mid (n+n')\}]^* = \lambda_0 F(z) + \sum_{\alpha=1}^{n} \lambda_\alpha (x; z)\, i(x_\alpha; z) + \sum_{\alpha'=1}^{n'} \lambda_{\alpha'} (x; z)\, y(x_{\alpha'}; z) \tag{5.22}$$

在此，$\lambda_\alpha (x; z)$ 是与 n 个近邻硬指示数据类型（5.17）、（5.18）或（5.19）相联系的加权系数，$\lambda_{\alpha'} (x; z)$ 是与类型（5.20）相联系的 n' 个数据的加权系数，λ_0 是与整体先验 cdf 相联系的加权系数。为保证无偏性，λ_0 常选为

$$\lambda_0 = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha (x;z) - \sum_{\alpha'=1}^{n'} \lambda_{\alpha'} (x;z)$$

用来估计条件累积分布函数的（5.22）可看成是收集各种不同类型信息的指示协同克里金，即收集了硬指示数据 i 或硬指示数据 j 和用先验概率表达的软数据 y 的信息。当不考虑软信息（$n'=0$）时，表达式（5.22）就成为简示单指示克里金的形式。

当在同一位置有不同的局部信息可供使用时，则可以导出类型（5.20）的不同先验概率 $y_1 (x_\alpha; z) \neq y_2 (x_\alpha, z)$，则（5.22）要作相应的改写。

下面进一步分析（5.20）式是怎样将先验累积分布函数模型更新成后验的条件累积分布函数模型。

如果软数据 y 的空间分布是通过硬指示数据的协方差 $C_I (\vec{h}; z)$ 进行建模的话，即

$$\begin{aligned} C_Y(\vec{h};z) &= \text{Cov}\{Y(x;z), Y(x+\vec{h};z)\} \\ &= C_{IY}(\vec{h};z) = \text{Cov}\{I(x;z), Y(x+\vec{h};z)\} \\ &= C_I(\vec{h};z) = \text{Cov}\{I(x;z); I(x+\vec{h};z)\} \end{aligned}$$

那么先验的累积分布函数值 $y(x_\alpha'; z)$ 在它们的位置 x_α' 处没有更新，即

$$[\text{Prob}\{Z(x_\alpha) \leq z \mid (n+n')\}]^*_{IK} = y(x_\alpha, z),$$

但是更常见的情况是软数据 y 来自于与硬数据有关的，然而是不同的协方差模型：

$$C_Y(\vec{h};z) \neq C_{IY}(\vec{h};z) \neq C_I(\vec{h};z)$$

从而，指示协同克里金估计（5.20）式归结为对所有的先验的累积分布函数的完全更新。

在具体操作中，要使用马尔柯夫—贝叶斯模型，有两个关键问题应很好地解决。其一就是协方差 C_I，C_{IY}，C_Y 的推断。其二就是软数据的推断。下面就具体讨论这两个问题的解决。

2. 协方差推断

如果数据充分的话，可以对所有截断值直接对协方差进行建模：$C_Y(\vec{h};z) \neq C_{IY}(\vec{h};z) \neq C_I(\vec{h};z)$。然而，这种直接的方法是比较麻烦的。一个可供选择的解决方法是建立协方差的马尔柯夫—贝叶斯模型，即

$$C_{IY}(\vec{h};z) = B(z) C_I(\vec{h};z), \text{对一切 } \vec{h}$$

$$C_Y(\vec{h};z) = \begin{cases} B^2(z) C_I(\vec{h};z), \text{对一切 } \vec{h} \neq 0 \\ \mid B(z)\ C_I(\vec{h};z), \text{当 } \vec{h} = 0 \end{cases} \tag{5.23}$$

系数 $B(z)$ 可以通过软数据到硬数据的标定而得到。更精确地讲：

$$B(z) = m^{(1)}(z) - m^{(0)}(z) \in [-1,1]$$

其中，

$$m^{(1)}(z) = E\{Y(\vec{h};z) \mid I(\vec{h};z) = 1\}$$

$$m^{(0)}(z) = E\{Y(\vec{h};z) \mid I(\vec{h};z) = 0\}$$

考虑一个标定的数据集 $\{y(x_\alpha;z), i(x_\alpha;z), \alpha=1, 2, \cdots, n\}$。在此，软概率 $y(x_\alpha;z)$ 在 $[0,1]$ 内取值并与硬数据 $i(x_\alpha;z)$ 来比较。$m^{(1)}(z)$ 是相应于 $i=1$ 时 y 的平均值。一种极端的情形是 $m^{(1)}(z)=1$。此时，相应于所有 y 值可精确地预测出 $i=1$。类似地，$m^{(0)}(z)$ 是对应于 $i=0$ 时 y 值的平均。极端情形为 $m^{(1)}(z)=0$。

参数 $B(z)$ 度量了 y 数据分隔两种情形 $i=1$ 和 $i=0$ 的优劣。最优的情形为 $B(z)=\pm 1$。最差的情形为 $B(z)=0$，即 $m^{(1)}(z)=m^{(0)}(z)$。

当 $B(z)=1$，表达式 (5.22) 中软先验概率数据 $y(x_\alpha';z)$ 可看成是硬指示数据。它们不必更新。相反，当 $B(z)=0$ 时，软数据 $y(x_\alpha';z)$ 不必考虑，因为它们在 (5.22) 中权系数为零。

不用软自协方差 $C_Y(\vec{h};z)$ 的一种近似方法是只考虑单个 y 数据。该数据与待估计位置 x 距离最近。这种简化了的协同指示克里金模型称为同位协同克里金（collocated cokriging)。此时，(5.22) 式简化为

$$[\text{Prob}\{Z(x) \leq z \mid (n+1)\}]^* = \lambda_0 F(z) + \sum_{\alpha=1}^{n} \lambda_\alpha(x;z) i(x_\alpha;z) + \lambda' y(x_\alpha';z) \tag{5.24}$$

在此，x_α'，与 x 距离最近。此时，只需要 $C_I(\vec{h};z)$ 和 $C_{IY}(\vec{h};z)$ 的推断，且在马尔柯夫—贝叶斯假设下有 $C_{IY}(\vec{h};z)=B(z)C_I(\vec{h};z)$，对一切 $\vec{h}$。这样可以减少计算量。

特别需要指出的是，协方差模型 (5.23) 是在马尔柯夫假设下分析导出的。那就是“硬信息 $I(x;z)$ 屏蔽了所有同位的软信息 $Y(x;z)$”，即 $\text{Prob}\{Z(x') \leq z \mid i(x;z), y(x;z)\} = \text{Prob}\{Z(x') \leq z \mid i(x;z)\}$，对于一切 $y(x;z)$，x，x'和z，因此，能否用马尔柯夫—贝叶斯模型 (5.22) 关键在于检验协方差模型 (5.23) 的正确性。

3. 软数据的推断

在具体应用过程中，经常遇到考虑一个主连续变量 $\{Z(x) \mid x \in D\}$ 通过一个有关的次变量 $\{V(x) \mid x \in D\}$ 推断的情形。

首先可利用硬数据 $\{z(x_\alpha) \mid \alpha=1, 2, \cdots, n\}$ 导出一系列硬指示数据 $\{i(x_\alpha;z_k) \mid \alpha=1, 2, \cdots, n; k=1, 2, \cdots, K\}$。软指示数据，$y(x_\alpha', z_k) \in [0,1]$，$k=1, 2, \cdots, K$，相应于次变量 $V(x_\alpha')$，可以通过 Z 值关于同位的 V 值散点图得到（如图 5.3)。

把 Z 值的变化区间分成 L 个类 $(V_{l-1}, V_l]$，$l=1, 2, \cdots, L$。对于类 (V_{l-1}, V_l)，y 的先验概率累积分布函数可以通过一些主数据值 $Z(x_\alpha')$ 累积直方图来建模，其中这些主变量数据必须使得同位的次变量数据 $V(x_\alpha')$ 落入类 $(V_{l-1}, V_l]$，即 $y(x_\alpha';z) = \text{Prob}\{Z(x_\alpha') \leq z \mid V(x_\alpha') \in (V_{L-1}, V_l)\}$。

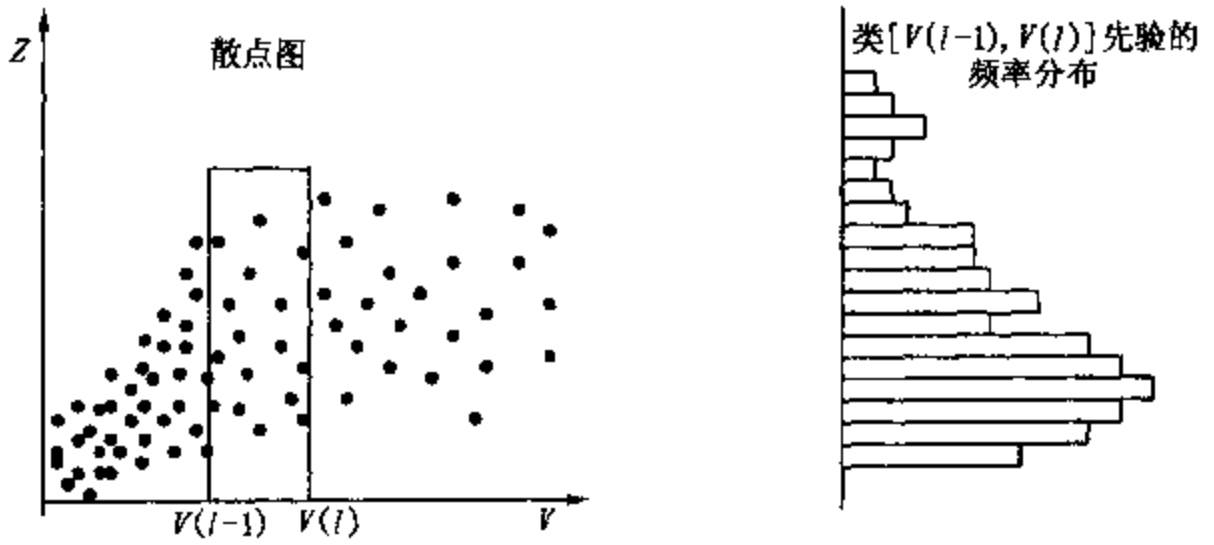


图 5.3　通过散点图对软先验概率的推断

次变量 V（x）并不必是连续的，类（V_{l-1}，V_l] 可用 V 值的范畴变量来代替，例如当 V 是与岩相和沉积相带有关的数据情形时，就属于这种情况。

提供先验的 y 概率值的标定散点图可以从类似的数据较为充分的研究区域借用而来，这些数据可以不同于标定（5.23）中协方差系数 B（z）的数据。

5.4　后验概率的次序关系校正及内插和外推

5.4.1　后验概率条件累积分布函数模型的次序关系校正

前面已经指出，指示模型具有很大的灵活性，它是一种非参数化的方法。在不假定后验概率模型的形式前提下，通过利用原始数据和先验概率信息，直接获取后验的条件累积分布函数模型。从而，使得指示模型能更充分利用所得到的各种信息。但是，这种灵活性必须以解决次序问题为代价（见图 5.4，Deutsch and Journel，1992）。由前面讨论已知，所有指示模型最终归结为对后验条件累积分布函数模型的推断。通过对后验条件累积分布函数模型的抽样可得到指示模拟结果。无论何种条件累积分布函数模型，都应满足常规概率的次序关系。具体来讲，对连续变量 $\{Z(x) \mid x \in D\}$ 的条件累积分布函数应满足

$$\text{Prob}\{Z(x) \leq z \mid (n)\} = F[x;z \mid (n)] \in [0,1]$$

和

$$F[x;z'_k \mid (n)] \geq F[x;z_k \mid (n)], \text{对一切 } z'_k > z_k \tag{5.25}$$

对于来自范畴变量的指示变量 $\{I_k(x) \mid k=1, 2, \cdots, K\}$ 其概率函数应满足

$$\text{Prob}\{I_k(x) = 1 \mid (n)\} = F[x;k \mid (n)] \in [0,1]$$

$$\sum_{k=1}^{K} F[x;k \mid (n)] = 1 \tag{5.26}$$

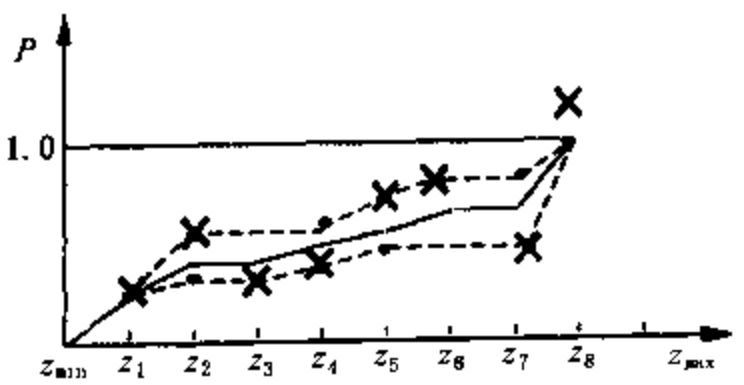


图 5.4　次序关系调整示意图
（Deutsch and Journel，1992）

指示模型导出的条件概率不会自行验证次序关系（5.25）和（5.26）。在范畴概率的情形，(5.26) 的第一个限制条件容易满足，这只须当估计值 F^* [x; k| (n)] 超出区间 [0, 1] 时，可将 F^* [x; k| (n)] 校正到最近的界 0 或 1。

(5.26) 的第二个限制条件处理较为困难，因为 K 个独立的克里金方程组是独立求解来完成的。

解决该问题的方法之一，就是求解（$k-1$）个克里金方程组而留下一个范畴变量 k_0，该范畴变量的选择应使它具有较大的先验概率 P_{k_0}，以便

$$F^*[x;k_0 \mid (n)] = 1 - \sum_{k \neq k_0} F^*[x;k \mid (n)]$$

连续变量条件累积分布函数的次序关系的校正具有更大的灵活性，这主要是由于累积指示值次序容易处理。

图 5.4 给出了一个例子（Deutsch and Journel，1992），具有如下次序关系

$$F[x;z_3 \mid (n)] < F[x;z_2 \mid (n)];$$

$$F[x;z_7 \mid (n)] < F[x;z_6 \mid (n)];$$

$$F[x;z_8 \mid (n)] > 1$$

此时，次序关系可这样调整。取上、下关系折线的平均。当某个点处值大于 1 时，将其校正到 1。在该图中，符号“×”表示已知数据点。显然，要进行次序关系的调整，虚线表示上、下界，实线表示最后调整的结果（Deutsch and Journel，1992）。

为何在指示模型中会产生次序关系问题。这主要有两个原因，其一是负的指示克里金加权系数引起。其二是某些类中数据的缺失。实际应用表明，次序问题的产生主要归结于数据的缺失。更确切地讲，在某个截断值 z_k 之上界，不存在数据 z 时，用指示克里金会出现次序问题。在这种情况下，指示数据集对于截断值 z_{k-1} 和 z_k 都是相同的。然而相应的指示变异函数是不同的。这最终会导致条件累积分布函数值不同而可能引起次序问题。

为了尽可能避免次序问题有很多处理办法，主要包括：

1）通过用在［0，1］上连续取值的先验的累积分布函数 F（x_a；z）代替每一套指示数据｛i（x_a；z_k），$k=1$，2，…，K｝，从而把原始数据 z 离散成的分布函数光滑成更为连续的分布函数。指示克里金步骤仍保持不变，这样表达式（5.4）中硬指示数据 i（x_a；z_k）被核值 F（x_a；z_k）简单地代替。如图 5.5 中，为了得到更为连续的尽可能避免次序问题的解决，付出的代价是损失了精确性。这归结于核函数 F（x_a；z）只是展布在硬数据 z（x_a）的周围。

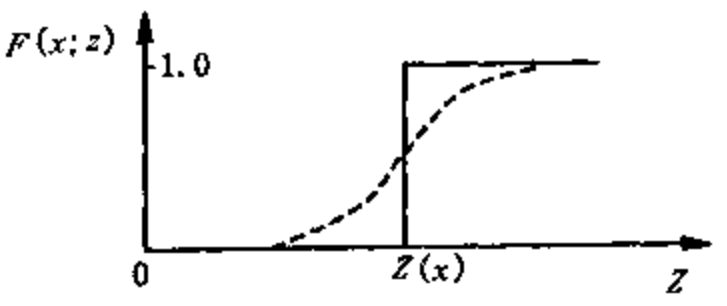


图 5.5　二值指示阶梯函数 i（x；z）（实线）被连续的核 $cdfF$（x；z）（虚线）代替（Deutsch and Journel，1992）

2）随着截断值的递增或递降，对于指示变异函数系数的变化可以定义一种连续性，使得这些变异函数的系数当截断值从一个变到另一个的时候不会有太太的改变。一种较好的办法，就是用同一种基本结构的套合来对所有指示变异函数进行建模，即

$$\gamma_I(\vec{h};z_k) = C_0(z_k) + C_1(z_k)Sph[|\vec{h}|/a(z_k)]$$

在此，Sph（$*$）是一个具有单位基台与变程 a 的球状变异函数模型。每一套系数（C_0，C_1，a）关于截断值变化的光滑性可保证函数的光滑变化。一种极端的情形是对所有截断值保留同一变异函数模型。如中位指示克里金可以减少产生次序问题的机会，但付出的代价是缺乏灵活性。

3）仅保留某些 z_k 值，使得类 F［x_a；z_1｜（n）］中至少包含一个数据。

以上三种方法只能减少，但不能从根本上消除次序问题。对保留的指示克里金的条件累

积分布函数值必须进行最后的校正。下面介绍通过上、下调整的平均进行次序校正的方法（Deutsch and Journel，1992）。

1）相应于图 5.4 上部线型的调整方法

从最低的截断值 z_1 开始，如果指示克里金的条件累积分布函数值 $F[x_a; z_1|(n)]$ 不在 [0，1] 中，那么把它置为最近的界 0 或 1。然后考察下一个截断值 z_2，如果保留的指示克里金累积分布函数值 $F[x_a; z_2|(n)]$ 不在区间 $[F[x_a; z_1|(n)], 1]$ 之内，把它置于最近的界。这样依次处理余下的截断值 z_k，$k=3, \cdots, K$。

2）相应于图 5.4 下部线型的调整方法

从最大截断值 z_k 开始，如果指示克里金保留的条件累积分布函数值 $F[x_a; z_k|(n)]$ 不在 [0，1] 中，置之为最近的界 0 或 1。然后考察下一个较低的截断值 z_{k-1}。如果指示克里金保留的条件累积分布函数值 $F[x_a; z_{k-1}|(n)]$ 不在 $[0, F[x_a; z_k|(n)]]$ 之内，置之为最近的界。这样依次处理余下的截断值。

(3) 对上、下部调整值进行平均，可得到图 5.4 中的实线。这就是最后的条件累积分布函数次序校正结果。

5.4.2　后验概率条件累积分布函数的内插与外推

前面讨论了后验概率条件累积分布函数的次序校正问题。但是，值得注意的是，虽然对后验条件累积分布函数模型进行了次序调整使之满足次序关系，但是如果样本容量 n 较小的话，必须有一些假设条件以便保证能够在最小的 z 数据（下尾），最大的 z 数据（上尾）之外进行外推，以及在相应 z 数据之间进行内插，从而得到连续分布在（$-\infty$，$+\infty$）内的完整的条件累积分布函数模型。由于利用数据展开技术，随机模拟将产生比原始数据多得多的值。因此，由样本结合指示克里金技术得到的条件累积分布函数所提供的最大分辨率可能是不充分的。特别在尾巴处，问题更为突出。

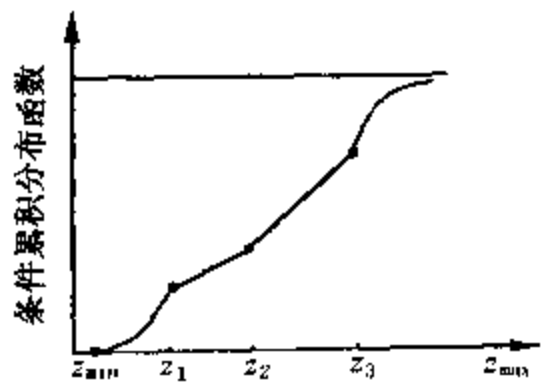


图 5.6　条件累积分布函数值的内插与外推示意图

当截断值较少时，分辨率缺乏将更严重。在这种情形，由指示克里金导出的条件累积分布函数值之间的内插，特别是条件累积分布函数在尾巴的外推显得尤为重要。

图 5.6 给出了一个条件累积分布函数 $F(x; z)$ 的例子（Deutsch and Journel，1992），$F(x; z)$ 在三个截断值（z_1，z_2，z_3）处值若已知，这三个截断值也许是边缘累积分布函数的三个中位数（即下中位数、中位数、上中位数）。中间两个类 $(z_1, z_2]$ $(z_2, z_3]$ 间进行线性内插，这归结为假定类间是均匀分布的。尾巴（z_{min}，z_1）外推至一个固定值 z_{min}，使用幂函数模型（见 5.27 式）。这相应于负偏态分布假设。上尾（z_3，$+\infty$）外推至一个无限上界，使用指数模型（见 5.28 式）。这相应于假定正偏态双曲模型。

下面具体讨论常用的条件累积分布函数内插与外推模型。（Deutsch and Journel，1992）

1. 幂函数模型

对于有限的类区间（z_{k-1}，z_k）和正数（指数）$w>0$，条件累积分布函数可以写成

$$F^{w}_{z_{k-1},z_k}(z)=\begin{cases}0,\text{当 } z\leq z_{k-1}\\ \left[\dfrac{z-z_{k-1}}{z_k-z_{k-1}}\right]^{w},\text{当 } z\in(z_{k-1},z_k]\\ 1,\text{当 } z\geq z_k\end{cases}\tag{5.27}$$

$w<1$ 时，相应于正偏态。$w=1$，相应于均匀分布。$w>1$，相应于负偏态，见图 5.7。

2. 指定界值之间的线性内插模型

该方法考虑在每个类（z_{k-1}，z_k］中给定界值时，得到固定数目的若干子类，然后在若干子类中以指定的概率进行线性内插。例如，给出三个界值 a_{k1}，a_{k2}，a_{k3}，定义了四个子类（z_{k-1}，a_{k1}］，（a_{k1}，a_{k2}］，（a_{k2}，a_{k3}］，（a_{k3}，z_k］。这样做可以使得使用者利用界值提高类间分辨率，从而恢复由于 z_k 值离散化而损失的部分细节。

3. 双曲模型

该模型是一种可供选择的用于上尾外推的正偏态分布模型。上尾条件累积分布函数的确定是至关重要的，因而必须使模型具有更强的适应性，包括可能给出一个较长的尾巴。双曲条件累积分布函数上尾模型是一个双参数分布模型。

$$F_{\omega,\lambda}(z)=1-\frac{\lambda}{z^{\omega}},\omega\geq 1,z^{\omega}>\lambda>0\tag{5.28}$$

刻度函数 λ 可以辨认任何预先计算的分位数。如 p—分位数 z_p，使得 $F_{\omega,\lambda}(z_p)=p$，从而 $\lambda=z_p^{\omega}(1-p)$。

函数 $w>1$ 控制条件累积分布函数达到上限值的快慢。w 愈小，分布的尾巴愈长。图 5.8 给出了双曲模型示意图。

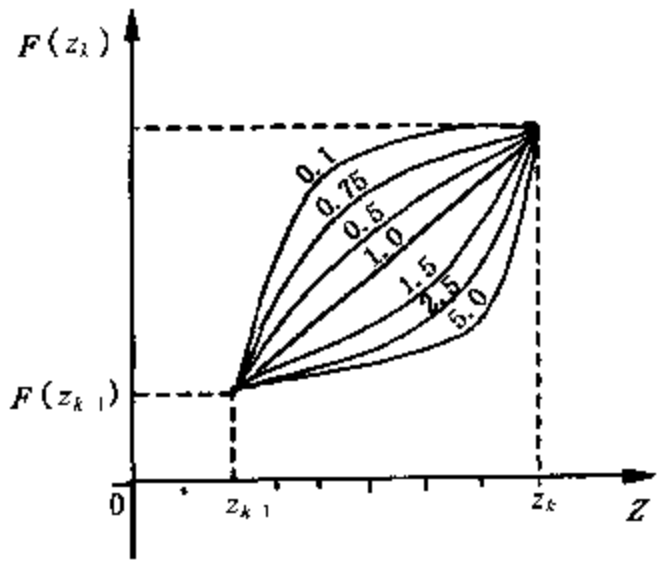


图 5.7　一些幂函数模型
$w=0.1$，0.25，0.5，1.0，
1.5，2.5，5.0

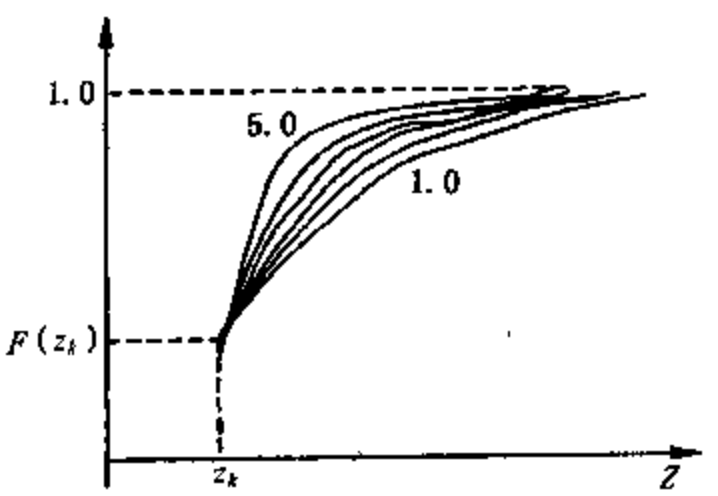


图 5.8　双曲模型示意图，
$\omega=1.0$，1.25，1.5，2.0，2.5，5.0

5.5　DQ 油田的实例研究

利用指示主成分模型来求取 DQ 油田沉积相分布的时候，有时有些层已经被分得很细，因此可以视为在二维空间中的沉积相分布的问题。彩图 1 是 SB 区 SS 油层组第 8 小层的水下枝状三角洲内前缘沉积微相分布。彩图 2 是用指示主成分模型得到的沉积微相分布图。由于是在二维空间中，人们可以根据所在地区的地质特点，设想有一些有特定沉积相的点，并加

入原始数据后，再进行模拟。彩图 3 是彩图 1 的原始数据，加上一些点，进行模拟得到的。彩图 3 看起来比彩图 2 连续性要更好一些。

彩图 4 和彩图 5 是 BY 区西区 P2 油层组第 10 小层的沉积微相的三维显示图，分别是用没坐标系和有坐标系表示的，而彩图 6 和彩图 7 分别是相应的渗透率和孔隙度的三维空间分布。彩图 8 则是沉积微相的栅状图，通过它可以观察储层内部的沉积微相分布。彩图 9 是无构造的沉积微相分布图。

外围区内有井 400 口，用指示主成分模型对它的各油层区的沉积微相分布进行预测。彩图 10、11 是外围区的第 2 油层组的沉积微相分布图，是不同的模型参数做出的。彩图 12 和彩图 13 是第 3、5 油层组的沉积微相分布图。

5.5.2　BN 油田多达 8 种沉积微相的随机建模

为了建立三维储层地质模型，对 BN 油田 SS 段的第 3、第 4 和第 5 砂层组，在精细地质研究与精细储层参数研究的基础上，主要对 3^{1^1}，3^{1^2}，4^{1^1}，4^{1^2}，4^{1^3}，4^{2^1}，4^{2^2}，5^1，5^2 这 9 个沉积时间单元为研究层位。首先，为了对这些时间单元的数据进行处理，利用了 132 口井的单井分层数据、测井解释数据及沉积相数据。然后，选取的三维网格系统为：$50\times50\times11=27500$。在这些网格结点上对沉积相、物性参数的空间分布进行建模，绘出时间单元构造图、沉积相三维立体图、剖面图、孔隙度、渗透率、含油饱和度三维立体图、剖面图，共计 30 多幅图件。最终，给出了相应沉积时间单元单井的物性、沉积相解释数据体及构造数据体。经过各种验证，该区块三维随机建模的结果是符合地质沉积特征的。

随机建模的内容包括地层—构造格架模型、沉积相模型和储层物性参数模型，分别讨论如下。

1. 地层——构造格架分析

地层——构造格架研究，是沉积相模型与储层物性参数模型研究的前提，从建立的三维地层——构造格架结果分析（见彩图 14），构造东南方向高，最高部位在 118－52 井附近，向西北逐渐变低，最低点在 118－61 井附近，高低差异大，工区构造为东南高西北低的构造格局。

2. 沉积相模型分析

BN 油田的 SS 层的各油层组含有 8 种沉积亚（微）相：主河道、水道区、叶状体、道间、扇缘、外扇、滑塌、深湖。用指示主成分模拟对该区块进行处理。

将这 8 种亚（微）相用数字代码表示如下：

$$I_1=\begin{cases}1,\text{位置 }\vec{u}\text{ 属于主河道}\\0,\text{其它}\end{cases}$$

$$I_2=\begin{cases}2,\text{位置 }\vec{u}\text{ 属于水道区}\\0,\text{其它}\end{cases}$$

$$I_3=\begin{cases}3,\text{位置 }\vec{u}\text{ 属于叶状体}\\0,\text{其它}\end{cases}$$

$$I_4=\begin{cases}4,\text{位置 }\vec{u}\text{ 属于道间}\\0,\text{其它}\end{cases}$$

$$I_5=\begin{cases}5,\text{位置 }\vec{u}\text{ 属于扇缘}\\0,\text{其它}\end{cases}$$

$$I_6 = \begin{cases} 6, \text{位置 } \vec{u} \text{ 属于外扇} \\ 0, \text{其它} \end{cases}$$

$$I_7 = \begin{cases} 7, \text{位置 } \vec{u} \text{ 属于滑塌} \\ 0, \text{其它} \end{cases}$$

$$I_8 = \begin{cases} 8, \text{位置 } \vec{u} \text{ 属于深湖} \\ 0, \text{其它} \end{cases}$$

在各时间单元沉积亚（微）相三维立体展布图上，以不同的颜色表示不同的亚（微）相。彩图 14 是沉积时间单元 3^{1^2} 的微相空间分布图。从图中可以看出，主水道微相，呈分支分叉状（色标条 1 号颜色：墨绿色）；水道区微相为辫状水道纵横交错的区域（色标条 2 号：绿色）；滑塌亚相为深湖亚相（色标条 3 号：蓝色）中单独的一堆沉积（色标条 4 号：洋红色）；扇缘扇前微相（色标条 5 号颜色），是中扇的外环边缘；外扇亚相（见色标条 6 号颜色），是扇体与深湖相泥岩相邻的扇体边缘部分，环绕中扇呈条带状，为砂泥岩互层。

从沉积相展布图上同时可看出，主水道方向为东南—西北方向，而各时间单元的沉积亚（微）相的位置、形状、宽窄都是不同的，它反映了浊积岩沉积的特点，即在不同的沉积时期，携带碎屑物质的强水流通过辫状水道将碎屑物带入深—半深湖中沉积，形成浊积扇，如从 5^1、5^2 沉积相分布图中可看出，在沉积早期，物质供给稳定，浊积扇从东南向西北延展沉积形成；从 4^{2^2} 沉积相分布图可看出，中扇在东南形成条带状，而在西北则出现滑塌浊积岩体。

3. 储层参数模型分析

建立孔、渗、饱等物性参性三维模型，可直观地显示储层物性参数在三维空间的展布特征。选用序贯高斯模型对该区块储层物性参数进行建模和模拟。

在各时间单元的物性参数空间展布图上，渗透率在 $0 \sim 100 \times 10^{-3} \mu m^2$ 范围内分级别表示，孔隙度在 0～25% 范围内分级别表示，饱和度在 0～75% 范围内分级别表示（见彩图 15，16，17）。

（1）孔隙度分布特征

从孔隙度分布图上看出，3^{1^1} 时间单元主水道的孔隙度为 20%～25%，水道区、叶状体、道间微相为 18%～20%，扇前、扇缘微相，外扇亚相的控制区孔隙度相对较低，为 13%～18%，滑塌亚相为 15%～18%。彩图 15 中可看出，3^{1^2} 时间单元主水道孔隙度为 20% 左右，水道区为 18% 左右，滑塌为 15%～18%。4^{1^1} 时间单元主水道上 79－1、102－21 两口井孔隙度在 20% 左右，水道区在 15%～20%，滑塌在 15%～20%，外扇在 10%～15%。4^{1^2} 时间单元主水道，水道区在 20%，扇前、扇缘、外扇在 10%～15%。4^{1^3} 时间单元主水道在 18%～20%，扇前、扇缘、外扇在 13%～15%，4^{2^1} 时间单元主水道，水道在 18%～20%，扇前、扇缘、外扇在 13%～15%，滑扇在 13% 左右，4^{2^2} 时间单元主水道、水道区在 18%～20%，滑塌在 13%～15%。5^1，5^2 时间单元基本上在主水道、水道区孔隙度为 15%～20%，扇前、扇缘、外扇为 13%～15%。

分析各时间单元孔隙度的分布情况，中扇亚相控制区域，包括主水道、水道区、叶状体、道间微相，孔隙性较好，孔隙度在 18% 左右，主水道上个别井在 20%～25%；滑塌亚相控制区在 13%～18%；扇前、扇缘微相及外扇亚相孔隙性较差，孔隙度在 13%～15%。

（2）渗透率分布特征

分析渗透率空间展布图，可看出整个工区渗透性较低。3^{1^1}时间单元中扇区基本在$5\times10^{-3}\mu m^2$以上，在东南方向102－10、102－41井附近渗透性较好，在$40\times10^{-3}\mu m^2$以上，外扇、滑塌在$1\sim5\times10^{-3}\mu m^2$。彩图15可以看出，$3^{1^2}$时间单元渗透性较好，主水道在$40\times10^{-3}\mu m^2$以上，水道区在$10\sim20\times10^{-3}\mu m^2$，外扇、滑塌在$1\sim10\times10^{-3}\mu m^2$。$4^{1^1}$时间单元主水道上79－1井渗透率在$15\times10^{-3}\mu m^2$以上，水道区在102－44井周围渗透率在$15\times10^{-3}\mu m^2$左右，滑塌区为$5\times10^{-3}\mu m^2$。$4^{12}$时间单元在主水道上渗透性较好，$25\times10^{-3}\mu m^2$左右，扇前、扇缘、外扇在$1\sim5\times10^{-3}\mu m^2$。对$4^{1^3}$，$4^{2^1}$，$4^{2^2}$，$5^1$，$5^2$等时间单元的渗透率分布，主水道、水道区为$15\times10^{-3}\mu m^2$以上，扇前、扇缘、外扇为$1\times10^{-3}\sim5\times10^{-3}\mu m^2$，滑塌区域渗透率为$1\times10^{-3}\sim15\times10^{-3}\mu m^2$。

（3）含油饱和度分布特征

分析9个时间单元的含油饱和度展布图，3^{1^1}，4^{1^1}，4^{1^2}，4^{1^3}，4^{2^1}，4^{2^2}，5^1，5^2等时间单元的含油饱和度在30%以上面积较大。而这些含油饱和度较高的区域基本处于主水道、水道区。如3^{1^1}时间单元含油饱和度展布图上，102－21、102－10井附近在主水道上的含油饱和度50%～60%，118－34、118－35井所在的滑塌区域在40%～60%。

4. 储层参数与沉积相的关系

沉积相展布及储层参数的分布特点，可以看出储层参数的变化主要是受沉积相带的控制。这是体现了储层沉积学的基本原理，同时也是由储层随机建模的原理所决定的。彩图14和彩图15、彩图16、彩图17的对比可以充分说明这个性质。

辫状主水道及叶状体微相部位一般孔隙性渗透性相对较好，含油饱和度值较高，这是由于主水道粒度较粗、孔隙较大；而叶状体，砂体均质程度较好，厚度较大。

水道区微相，沉积以砂岩为主，一般砂岩厚度大于4m，储层物性在该沉积区域分布相对较好。

扇前、扇缘微相，砂岩厚度小，在中扇外环，反映出物性较差。

外扇亚相由于与深湖相接壤，一般为砂泥岩互层，在物性参数分布图上为物性较差部位。

滑塌亚相，受其沉积成因的影响，物性反映中等。

从沉积相与储层物性参数的关系看，该区块的进一步挖潜，应以主水道、叶状体、道间、水道区微相控制区域为主挖潜对象。

5. 沉积相模拟结果的过滤处理

由于模型参数在提取过程中出现的误差及8种沉积相的复杂空间接触关系，使得初始空间沉积相模拟结果会出现如下不合理现象：

1）由于沉积相可用的数据太少且分布不均匀时，可能导致空间预测结果在远离已知数据点时未能出现相应的沉积相预测结果。

2）由于模型误差，在整个模拟结果中会出现一些面积不大的相夹杂于另外一些背景相之中。根据地质知识可知，这种相接触关系显然是不合理的。这种现象称之为“跳相”现象。

3）从地质学家角度来看，在一个特定的地区，相应特定的沉积环境和沉积类型，沉积

相的分布具有一定的接触关系，称之为相序规律。初始模拟结果显然不能完全服从一定的相序规律。

基于以上考虑，为了更合理地模拟空间沉积相分布，必须对初始模拟结果进行必要的修正，我们称之为过滤处理，整个思路和目的就在于滤掉不必要的“噪声”（跳相现象），从而达到相应的相接触关系（相序规律）。

通过对BN油田具体研究区块仔细分析，首先建立了如下相序关系模式：

1）递变接触的依次为：主水道——水道区——扇缘——外扇——深湖，或：1－2－5－6－8。

2）非递变接触的相：叶状体－3，道间－4，滑塌－7。

3）上述两种类型相之间为可包含关系：深湖相包含滑塌，水道区包含叶状体及道间。

考虑到上述相序关系，在具体设计过滤程序过程中，对初始模拟结果进行过滤，使过滤后的结果能同时满足相序关系模式1）、2）、3）。实际的操作过程是对初始模拟的三维数据体进行过滤处理，最后给出合理的三维数据体。所以这样过滤是真实三维过滤，而不是仅仅对图形进行直接编辑，从具体结果分析来看，这一步骤是很必要而且是成功的。

在此需要特别指出的是，过滤处理只是基于初始模拟结果的。初始模拟结果是基础。只有初始模拟结果基本反映了相带的主要分布趋势和空间展布，再进行过滤处理才能取得较好的效果。过分依赖于过滤而忽视了原始模拟结果的相对正确性，这其实是本末倒置，不可能得到满意的结果。其次就是对不同地区建立相应的相序变化模式是至关重要的，在具体的相编码过程中，尽可能使接触和递变的相具体相互靠近地编码，这样有利于实施过滤方法。总之，找出合理的相序关系是过滤成败的关键因素。

6．三阶段模拟方法

在给出三维地层——构造格架模型、沉积相三维空间的模拟结果后，方可进行物性参数的模拟。在具体研究过程中，采用三阶段模拟方法，首先先建立三维地层——构造格架模型，接着进行沉积相三维空间的模拟，即建立三维沉积相模型，再在沉积相三维空间模拟结果控制下，进行物性参数模拟。这样做，才能达到物性参数模拟与沉积相模拟结果的一致性。

具体过程如下：

1）将物性按不同相带分类，分类的原则是将物性参数分布较为接近的若于相带分为一组，然后对不同组内的物性参数分别进行变异函数分析，得到相应的理论模型参数。相带分类过程称之为相组提取。在研究的具体区块内可将主河道和深湖相单独分为一组，水道区、叶状体、道间为一组，扇缘、外扇和滑塌为一组，共四组相。值得指出的是划分的组要尽可能区分参数的不同分布，但组数又不可能过多，否则，每组中可用的数据量太少，从而统计分析和模拟结果不可靠。

2）利用沉积相模拟网格数据，结合已得到的不同的相组的变异函数拟合参数，按不同相组分别独立地进行序贯高斯物性参数模拟，这样可得到整个区域的物性参数模拟结果。

储层的多阶段随机建模技术具有诸多优点。已成为当前国际上一项先进的储层表征技术。它可以揭示不同尺度下储层的非均质性以及这些非均质的相互关系，使得渗透率等物性参数空间分布的模拟结果，受到沉积相空间分布的模拟结果的控制，并与之保持一致。

第 6 章　截断高斯模型

截断高斯模型（Trancated Gaussian Model）是由法国地质统计学中心的马特隆教授（G.Matheron）等提出的。该模型首先建立高斯随机函数模型，然后通过对连续变化的高斯随机函数模型进行截断处理，最后获得离散的地质变量（如沉积相带空间分布）的模拟结果。该模型的优点在于它可以反映地质沉积相的空间接触关系（或称之为相序规律），同时可通过比例曲线来反映沉积相带在空间的非均质变化。本章重点讨论该模型的基本思想。

6.1　比 例 曲 线

截断高斯模型主要用来解决沉积相带空间的分布问题。沉积相（或岩相）在储层空间不同位置分布是不均匀的，或者讲在储层不同位置，不同沉积相占据的比例是不同的。因此，根据已知井位数据获取不同相带在空间不同位置占据的比例是在进行沉积相带模拟过程中至关重要的信息。而比例曲线可以反映不同深度和不同平面位置处不同沉积相带占据的比例，它是截断高斯模型具有特色而且非常重要的内容。本节就深入阐述比例曲线的获取过程。

6.1.1　垂直比例曲线

垂直比例曲线主要反映不同沉积相（或岩相）在空间不同深度位置占据的比例。其具体求取过程如下：首先将研究区域归一化，即将地层转化成归一化的等厚模型，以便地质统计分析使用。拉伸的等厚模型其厚度可以选取为整个层位观察的最大厚度。在等厚模型中，依据不同深度计算落入该范围内所有井位观察的不同相占据的比例，并依次累加可得到相应深度位置处不同相占据的比例。很显然，当自上而下来观察这张累积比例图形时，会发现不同相带占据的比例是变化的，从而不同相带交接位置形成若干曲线，比例曲线由此得名。

假设一个区块有五口井位的观测资料，共有三种相带，则归一化处理后井位观察和比例曲线范围如下图 6.1：

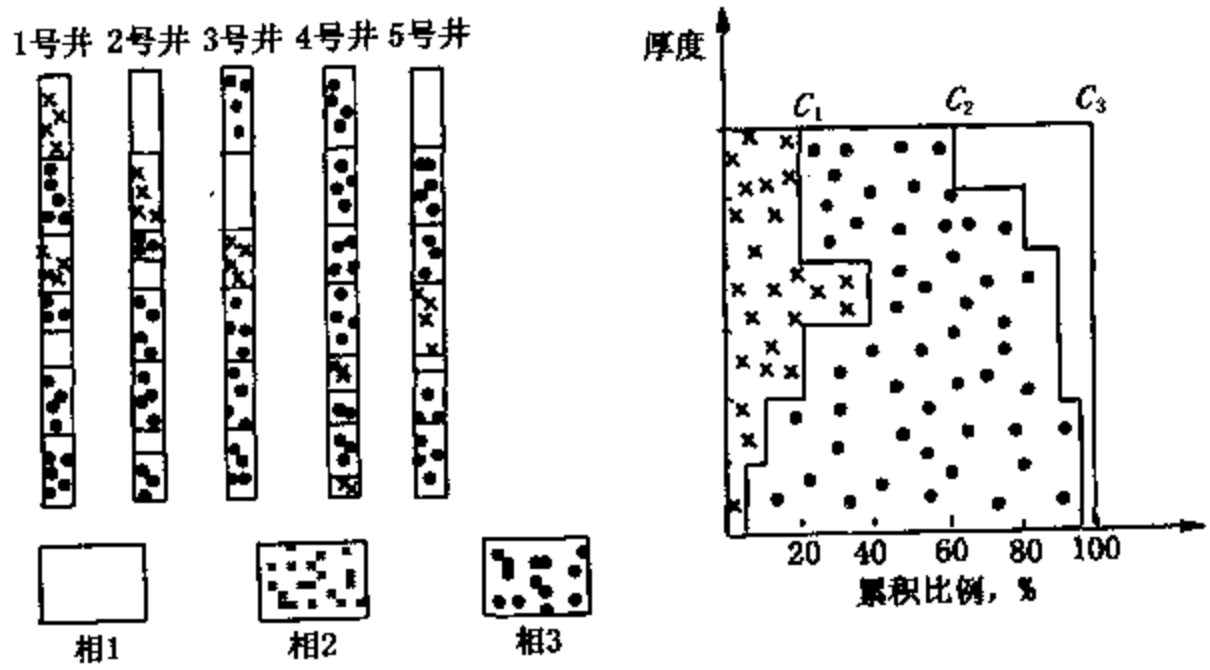


图 6.1　比例曲线示意图

分析上图能得到三条曲线，C_1 反映相 1 的比例变化，C_2 反映相 1 相 2 累积比例变化，C_3（100%）反映了三种累积比例为 100%。曲线相减 $C_2 - C_1$ 反映了相 2 真实比例变化，$C_3 - C_2$ 反映了相 3 真实的比例变化，三种相真实的比例变化记为 P_i（Z）（$i=1$，2，3）。Z 表示深度。

6.1.2　水平比例曲线

为了反映相带在平面上的变化，有必要引入水平比例曲线。该曲线是通过统计归一化后的井位相观察在垂直于已知平面不同位置占据比例而得到的。在此需要指出的水平比例曲线在统计过程中是利用每个单井整个资料统计的，而不考虑垂直方向。与之形成对照的是，垂直比例曲线是在某一深度位置所有井的资料都参与统计。

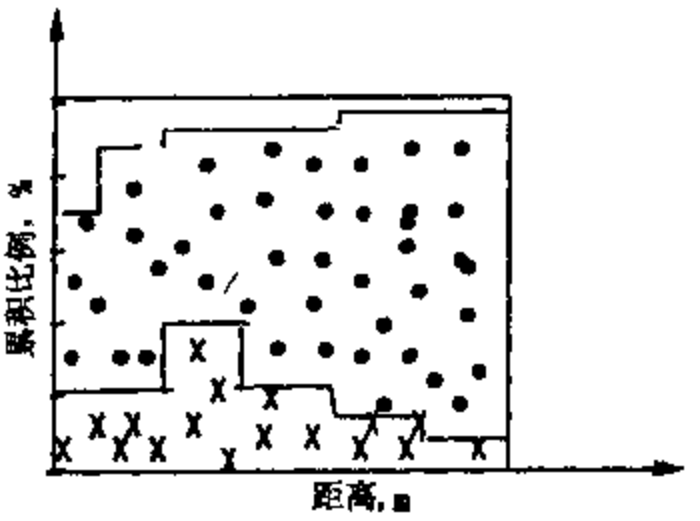


图 6.2　水平比例曲线示意图

同样，若利用前面 5 口井资料可以获得水平比例曲线，见图 6.2。

6.2　相序规律与高斯函数的截断

6.2.1　相序规律

在一个特定的地区，相应于特定的沉积环境和沉积类型，沉积相的分布具有一定的接触关系，称之为相序规律。沉积相分布在纵向上与横向上都具有一定的关系，而且二者之间也是有密切关系的。对于大的区域性沉积来讲，相带之间递变的接触关系更为明显。

例如对于湖泊三角洲沉积体系来讲，可以划分为三个相带，即三角洲平原，三角洲前缘和前三角洲。而且三者之间是逐渐过渡的。

对于曲流河河流沉积环境来讲，可以分为五种相带，河道充填，点砂坝，天然堤，决口扇和河漫滩。而且这五种微相带也具有明显的接触关系。点砂坝是由河流侧向沉积而形成的，其他河道内砂质沉积称为河道充填，从而这两种微相构成主河道沉积。河道边沉积相是在河水溢岸时，在近河道内侧形成的沉积物，细分为天然堤和决口扇两种微相。河漫滩是指曲流河道及其边缘区相邻处的广大平原沉积物。因此，沉积过程控制着相带之间的接触关系。

对于扇三角洲平原分流河道沉积，可分为河道砂，河间薄层砂，河间泥三种微相。三者之间是渐变接触的。

因此，在建立沉积相分布定量模型时，符合相序规律是建模的先决条件。截断高斯模型的最大特点就是该模型可以严格反映沉积相的接触关系。

6.6.2　高斯函数的截断处理

设 $|Y(x)|x\in D|$ 是一个定义在研究区域 D 上的平稳的高斯随机函数，均值为 0，方差为 1，且相关函数为 $\rho(h)$。若要模拟在空间 D 上的 n 种相 F_1，F_2，…，F_n 的分布，可通过对 $Y(x)$ 进行截断而获得。此时，可定义 $F_i = \{x\in D | t_{i-1} < y(x) \leq t_i\}$。这也等价于定义了 n 种指示函数：

$$I_{F_i}(x) = \begin{cases} 1, \text{当 } t_{i-1} < y(x) \leq t_i \\ 0, \text{其他} \end{cases}$$

在此 t_i 被称为截断值，$i=1$，2，…，n，且 $t_0=-\infty$，$t_n=+\infty$。

很显然，通过这样截断处理后，可以保证相序关系，依次为 $F_1\rightarrow F_2\rightarrow\cdots\rightarrow F_n$。这主要由于高斯随机函数 $Y(x)$ 是连续变化的缘故。

截断值 t_i（$i=1$，2，…，n）需要通过一定的方法加以确定。确定的原则主要为了保证不同相在研究区域中应占据的比例 P_i（$i=1$，2，…，n）。如果沉积相的分布在空间具有平稳性，则每种相占据的比例 P_i 不随位置的变化而变化，此时可用常规的直方图统计每种相占据的比例 P_i（$i=1$，2，…，n）。在这种平稳情况下，截断值 t_i（$i=1$，2，…，n）也不随位置的变化而变化。此时，截断值可通过下面方法获取。

1）首先，对第一种相 F_1，选取 t_1，使 $P_1=F_1$ 占据的比例，即

$$P(x\in F_1)=P[-\infty<Y(x)\leq t_1]=G(t_1)。$$

在此 G 是累积标准高斯分布函数，从而 $t_1=G^{-1}(P_1)$。

2）对于其他相 F_i（$i>1$），则选取 t_i，使 $P_i=F_i$ 占据的比例，即

$$P(x\in F_i)=P[t_{i-1}<Y(x)\leq t_i]=P[Y(x)\leq t_i]-P[Y(x)\leq t_{i-1}]=G(t_i)-G(t_{i-1})。$$

这样，递推依次可获得截断值 t_1，t_2，t_3，…，t_n。比例值 P_i 与截断值 t_i 之间关系，如图 6.3。

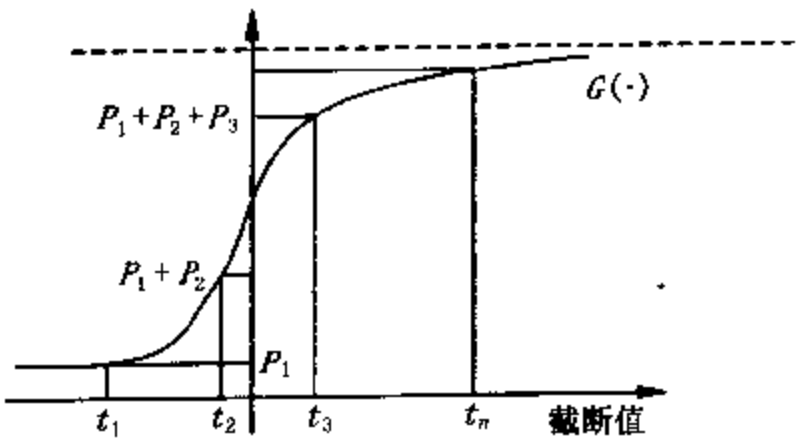


图 6.3　截断值的获取

由图 6.3 可以知道，截断值的获取实际上等于累积比例对应的分位数。

对于更一般的情形，当比例值随位置面变化的时候，即相分布具有非平稳性，则此时，相应的截断值也依赖于位置，即 t_i 成为 $t_i(x)$，$x\in D$，此时有下列对应关系

P_i 在空间为常数$\Leftrightarrow t_i$ 在空间为常数；

P_i 随深度变化，即有 $P_i(h)\Leftrightarrow t_i$ 随深度变化，即有 $t_i(h)$；

P_i 随位置变化，即有 $P_i(x)\Leftrightarrow t_i$ 随位置变化，即有 $t_i(x)$。

图 6.4 给出了高斯随机函数是如何用常数截断值 t_1，t_2 截断后而得到三种相（图 a），以及被随深度变化的截断值 $t_1(h)$，$t_2(h)$ 截断后而得到的三种相（图 b）。

从图中可以看到三种相之间的递变接触关系，相 1 过渡到相 2，相 2 过渡到相 3，而相 1 与相 3 不接触。但是，需要指出的是，利用随深度变化的截断值也可以反映不递变的相接触关系。如图 6.5 井柱的上半段反映了相 1 与相 3 之间也可以具有接触关系，这样使得截断高斯模型具有更强的灵活性。

由于井柱前半段 $t_1(h)$ 与 $t_2(h)$ 重合，从而说明缺失第二种相，从而相 1 与相 3 直接接触，因此，随位置而变化的截断值的使用使截断高斯模型既可以反映相之间递变接触关系，也可有突变接触。

6.3　截断高斯模型

6.3.1　截断高斯模型的实施步骤

1）首先，选取研究区域 D，并对研究区域内沉积环境有明确认识，应用层序地层学或

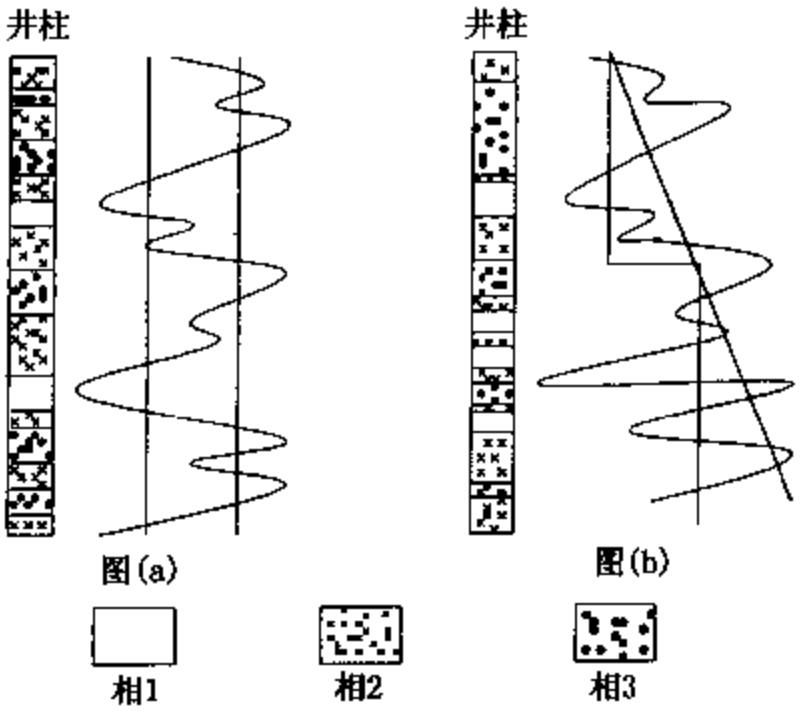


图 6.4　高斯函数截断与相关系示意图

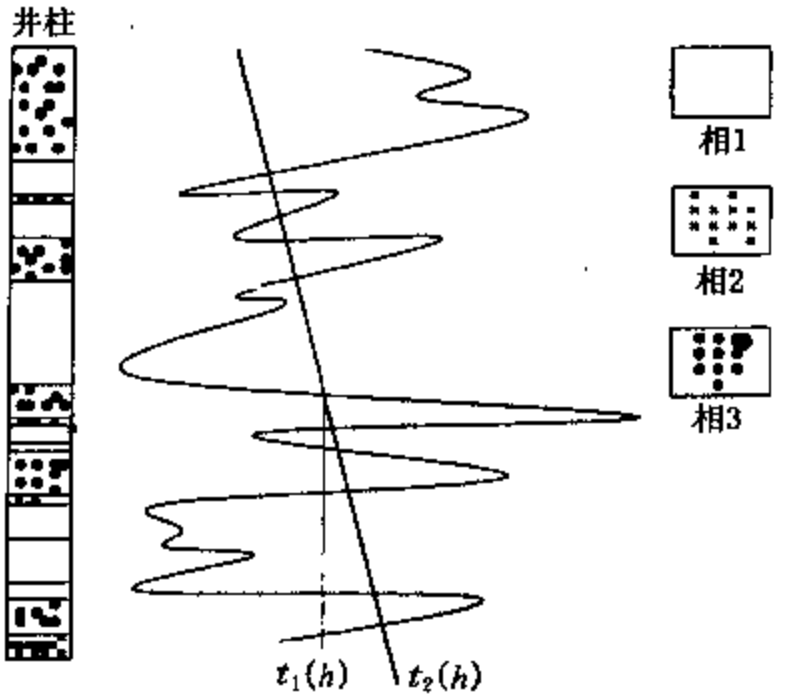


图 6.5　截断高斯函数与相关系示意图

其它地质认识划分合理的地层时间单元，对区块内相接触关系有足够认识，并将相进行编码，使编码符合相的递变接触关系。接着计算相应的比例曲线 $x \in D$，$P_i(x)$，$i=1$，2，$\cdots n$。一般来讲，比例曲线依赖于空间位置，得到比例曲线后，利用相的累积比例得到相应的截断值 t_i（x）（$i=1$，2，$\cdots$，n），$x \in D$。

2）在研究区域 D 内作一个非条件的平稳高斯场 $\{Y(x) \mid x \in D\}$ 的模拟，该高斯场具有均值 0，方差 1，相关系数函数 $\rho(h)$ 的选择要保证在该高斯函数经过截断处理后，得到的理论交互指示变异函数模型与原始相数据的交互指示变异函数相容。

$\{Y(x) \mid x \in D\}$ 的非条件模拟可以通过旋转带方法来完成，同时除了获得网格结点处非条件模拟结果外，应获取原始井位数据处非条件高斯模拟结果。

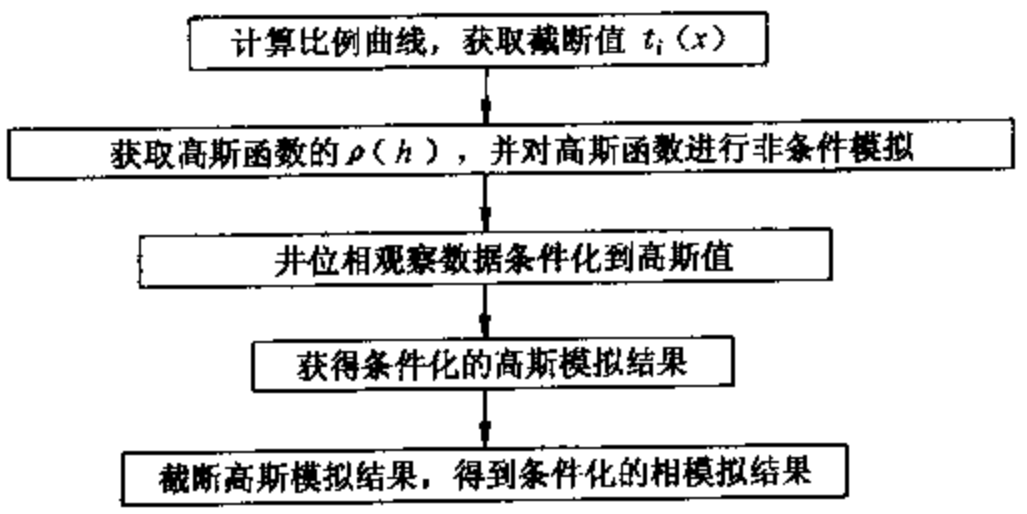


图 6.6　截断高斯模拟流程示意图

3）将已知井位的相观察数据条件化成高斯值，以保证最后截断处理后，模拟结果条件化到已知的井位数据。这一步是通过克里金方法来完成的。

4）利用（2）所得到的结果结合（3）所给出的条件化井位高斯值，使用误差模拟方法可以得到条件化的高斯模拟结果。

5）将（4）得到的条件模拟高斯结果进行截断处理，可以得到最终的沉积相条件模拟结果。上述完成截断高斯模型的五个步骤可用下面的框图 6.6 来描述。

关于上述步骤（1），前节已经给出，下面将讨论其余四个步骤

6.3.2　高斯函数模型的确定

这一步在截断高斯模型中是关键。该步骤主要是如何选择一个平稳的具有零均值和单位方差的高斯随机函数模型，使具有指定的相关函数 $\rho(\vec{h})$。即选取 $\{Y(x) \mid x \in D\}$，使

$$Y(x) \sim \text{Gauss}[0, \rho(\vec{h})] \tag{6.1}$$

在此，核心问题是 $\rho(\vec{h})$ 的选取，$\rho(\vec{h})$ 选取的原则是使得相应高斯函数截断后所得到的指示变量的非中心化交互协方差与原始相数据计算的指示变量实验交互协方差相容，具体过程如下

设有 n 种相 F，F_2，…，F_n，则可定义 n 个指示变量

$$I_{F_i}(x)=\begin{cases}1, & x\in F_i\\ 0, & \text{其它}\end{cases}\qquad i=1,2,\cdots,n \tag{6.2}$$

计算每种相所对应的指示变量的非中心化协方差

$$\hat{C}_{F_iF_j}(x,y)=\hat{E}[I_{F_i}(x)I_{F_j}(y)]\qquad i\neq j,i,j=1,2,\cdots,n \tag{6.3}$$

式（6.3）是实验非中心化协方差。

从理论上讲，非中心化协方差模型可如下计算（Galli and Armstrong，1993）

$$\begin{aligned}C_{F_iF_j}(x,y)&=E[I_{F_i}I_{F_j}]=P[I_{F_i}(x)=1,I_{F_j}(y)=1]\\&=P[t_{i-1}<Y(x)\leq t_i,t_{j-1}<Y(y)\leq t_j]\end{aligned}$$

从而

$$C_{F_iF_j}(x,y)=\int_{j-1}^{j}\int_{i-1}^{i}g(u,v)\mathrm{d}u\mathrm{d}v \tag{6.4}$$

在此，g（u，v）为二元联合正态分布密度函数

$$g(u,v)=\frac{1}{2\pi|c|^{\frac{1}{2}}}\exp[-\frac{1}{2}(u,v)c^{-1}\begin{pmatrix}u\\v\end{pmatrix}]$$

其中：

$$c=\begin{pmatrix}1 & \rho(x-y)\\ \rho(x-y) & 1\end{pmatrix}$$

在平稳假设下，(6.4) 可写成 $C_{F_iF_j}(\vec{h})=\int_{j-1}^{j}\int_{i-1}^{i}g(u,v)\mathrm{d}u\mathrm{d}v$。在此，$g$（$u$，$v$）为二元联合正态分布密度函数，且

$$c=\begin{pmatrix}1 & \rho(\vec{h})\\ \rho(\vec{h}) & 1\end{pmatrix}$$

需要指出的是，(6.3) 中的截断值可以是随位置而变化的。

(6.4) 式建立了高斯随机函数的协方差函数与被截断的指示值之间非中心化协方差函数之间的关系。在具体应用中，就是通过调节 $\rho(\vec{h})$，使通过（6，4）获得的 $C_{F_iF_j}(h)$ 可以作为相应指示变量的实验非中心化协方差（6.3）的模型，在这里 $i\neq j$，i，$j=1$，2，…，n。

二元联合正态分布密度函数具体展开式为

$$\begin{aligned}g(u,v)&=\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\exp[-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)]\\&=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp[-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)]\end{aligned} \tag{6.5}$$

在具体实施中，(6.4) 右端的积分可用数值方法来计算。

一旦确定了 $\rho(\vec{h})$ 模型 (6.1) 也随之确定。此时，可得到 (6.1) 的非条件模拟结果。如果是各向同性协方差模型，则 $\rho(h)=\rho(\vec{h})$，$h=\|\vec{h}\|$。此时可使用旋转带方法来模拟，否则可选用序贯高斯方法来模拟。

6.3.3　井位观察数据的条件化

设在已知的井位观察位置 $\{x_1, x_2, \cdots, x_N\}$ 处有 N 个观察数据 $\{F_{i_1}, F_{i_2}, \cdots F_{i_N}\}$，其中，$i_1, i_2, \cdots, i_N$ 为 1，2，…，N 中的一个，研究区块共有 n 种相 $F_1, F_2, \cdots, F_n$。井位观察数据的条件化步骤的主要目的在于获得高斯随机函数 (6.1) 的一个实现 $\{\overline{Y}(x) \mid x\in D\}$，使得在已知的井位观察位置有

$$\overline{Y}(x_j)\in(t_{i_j-1}, t_{i_j}) \tag{6.6}$$

在此，$j=1, 2, \cdots, N$。

满足 (6.6) 的 $\overline{Y}(x)$ 是条件化到井位数据的，这是因为

$$I_{F_{ij}}(x)=\begin{cases}1, t_{ij-1}<Y(x)\leq t_{ij}\\ 0, \text{其它}\end{cases}$$

从而 $Y(x_j)$ 落入 (t_{ij-1}, t_{ij}) 就保证 x_j 位置相观察为 F_{ij}。完成上述过程是通过克里金方法来实现的（*Galli and Armstrong*，1997），具体过程如下：

(1) 设 $x_1\in F_{i_1}$，即 x_1 位置具有相 F_{j_1} 观察，则产生一个来自于 $Y(x_1)\sim N(0, 1)$ 的样本 $\overline{Y}(x_1)$，使得 $t_{i_1-1}<\overline{Y}(x_1)\leq t_{i_1}$，

(2) 假设已经产生了 $\overline{Y}(x_1), \cdots, \overline{Y}(x_K)$，且满足 K 个限制，则建立 $\overline{Y}(x_{K+1})$ 可采用下面方法。令：$\overline{Y}(x_{K+1})=\overline{Y}^*(x_{K+1})+\sigma\cdot R(x_{K+1})$

在此 $\overline{Y}^*(x_{K+1})$ 是用 $\overline{Y}(x_i)$，$i=1, 2, \cdots K$，对 $\overline{Y}(x_{K+1})$ 进行的克里金估计，σ 为克里金估计的标准差。

$R(x_{K+1})\sim N(0, 1)$ 且与 $\overline{Y}(x_i)$，$i=1, 2, \cdots, K$，相互独立。如果已知 $\overline{Y}(x_{k+1})\in[t_{i_{k+1}-1}, t_{i_{k+1}}]$，则 $\dfrac{t_{i_{k+1}-1}-\overline{Y}^*(x_{k+1})}{\sigma}<R(x_{k+1})\leq\dfrac{t_{i_{k+1}}-\overline{Y}^*(x_{k+1})}{\sigma}$，从而 $R(x_{k+1})$ 是一个截断的高斯随机变量。这样，可产生 $R(x_{k+1})$ 的一个样本，并乘以 σ 且加到 $\overline{Y}^*(x_{k+1})$ 上去就可得到一个 $\overline{Y}(x_{k+1})$。它满足指定的区间限制。

在具体实施时，还可直接产生一个来自于 $N[\overline{Y}^*(x_{k+1}), \sigma^2]$ 的样本 $\overline{Y}(x_{k+1})$。若 $\overline{Y}(x_{k+1})\in(t_{i_{k+1}-1}, t_{i_{k+1}})$，则保留。否则，反复使用过程 (2)，就可获得条件化的高斯值 $\{\overline{Y}(x_i)\mid i=1, 2, \cdots, N\}$。

6.3.4　条件化的高斯函数模拟

该过程是通过 6.3.1 中步骤 (2) 获得的来自于高斯随机函数 (6.1) 的非条件模拟结果 $\{Y_S(x)\mid x\in D\}$，以及通过其中的步骤 (1) 获得的来自于原始观察数据的条件化的井位高斯值 $\{\overline{Y}(x_i)\mid i=1, 2, \cdots, N\}$，来最终得到条件高斯结果 $\{Y_{CS}(x)\mid x\in D\}$。主要思想是使用误差模拟原理，即用如下公式

$$Y_{CS}(x) = Y_C^*(x) + [Y_S^*(x) - Y_C^*(x)] \tag{6.7}$$

获得条件模拟 $Y_{CS}(x)$。其中 $Y_C^*(x)$ 是利用已知的 $\{\bar{Y}(x_i) \mid i=1, 2, \cdots, N\}$ 进行克里金估计得到。而 $Y_S^*(x)$ 是利用 $\{Y_S(x_i) \mid i=1, 2, \cdots, N\}$ 进行克里金估计获得。

需要指出的是，由于两次克里金是在同样的井位构形下完成的，从而克里金权系数是相等的，因此克里金方程组只需求一次解。

6.3.5　条件化高斯函数模拟的截断处理

当得到了条件模拟结果 $\{Y_{CS}(x) \mid x \in D\}$ 之后，可用截断值对该结果进行截断处理，

$$I_{F_i}(x) = \begin{cases} 1, t_{i-1} < Y_{CS}(x) \leq t_i \\ 0, \quad \text{其它} \end{cases}$$

这等价于

$$x \in F_i \Leftrightarrow t_{i-1} < Y_{CS}(x) \leq t_i \tag{6.8}$$

这样的截断处理就获得了相的模拟结果，需要指出的是，如果截断值依赖于空间位置 x，则（6.8）式仍适用。

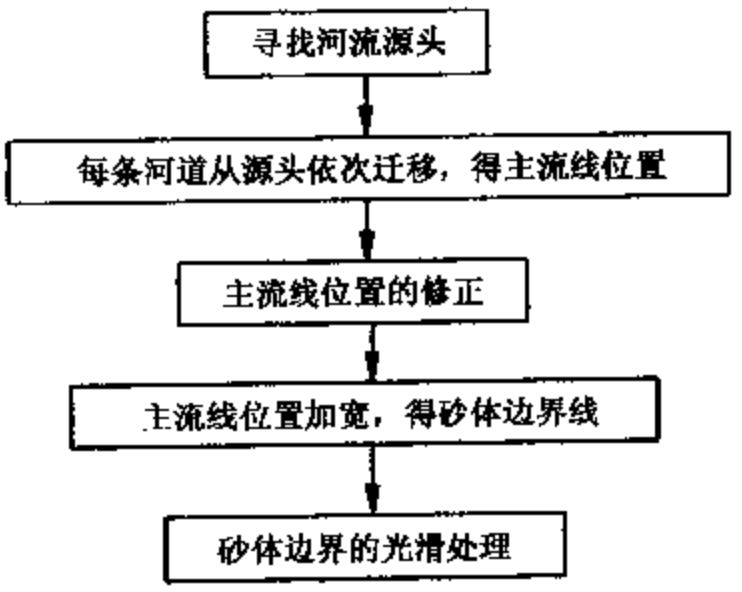
第 7 章　随机游走模型

一个随机过程 $\{S_n, n\geq 0\}$ 称为随机游走模型：设 X_1，X_2，…为独立同分布的随机变量，且有 $E(|X_i|)<\infty$；令 $S_0=0$，$S_n=\sum_{i=0} X_i$，$n\geq 0$（Ross，1983）。一维的随机游走模型一般抽象地定义如下：以时间 T 为间隔，掷一次硬币，若结果是头像，则往左走一步（步长为一常数）。若结果是面值，则往右走一步。这样，一个一维随机游走过程是从 $t=0$ 开始，随着 t 增加，x 呈显为一个阶梯形。二维的随机游走模型 $\{x(t), y(t)\}$，是由两个相互依赖的一维随机游走模型组成。实践证明，随机游走模型的理论可用来描述河道走向，也可用在场论方面的研究（Ambrahamsen，1988），还在可用在分析排队和破产系统的研究（Ross，1983）。有关随机游走模型的理论可参考文献：Ross，1983；Spitzer，1964；Cox and Miller，1965；Feller，1966。

随机游走模型在油藏描述和油藏工程中应用的文献已发表了不少，主要是应用在流体的描述、地质构造的描述、渗透率粗化和随机图像的选取等等方面（King and Scher，1985；Deutsch and Srinivasan，1996；McCarthy，J.，1993；Lenormand and others，1990）。但是，在本书作者的文章以前，还未见到随机游走模型用于储层沉积相的随机建模方面的文献。

随机游走模型可用于刻画河流沉积环境中河道砂体的分布（Wang and Zhang，1997）。其优点在于可以描述河道砂体的连续性形态，河道宽度、河道弯曲程度以及河道分叉与汇合程度。主要思路在于将研究区域离散成网格系统，然后寻找区域内河道源头位置的分布，并在此基础上依次获取河道主流线位置，并加宽得到砂体分布形态，由于砂体主要位置是由格点依次迁移而得到的，且由迁移概率给予定量描述，因此，该模型被称为随机游走模型。通过在每个结点利用抽样方法可以获取河道向邻近哪一个结点迁移，进而得到砂体走向空间分布的结果。

随机游走模型主要框架如下：



7.1　区域网格化

假定物源方向是从北向南的，若不具有北南方向的物源，那么可利用坐标旋转变换，使得在变换后的坐标下物源近似北南方向。

将研究区域离散成网格。令 nx，ny 分别表示 x 方向和 y 方向的网格数，$\mathrm{d}x$，$\mathrm{d}y$ 分别表示相应方向的网格跨度，一般地假定 $\mathrm{d}x$ 和 $\mathrm{d}y$ 小于平均井距之半，如图 7.1。

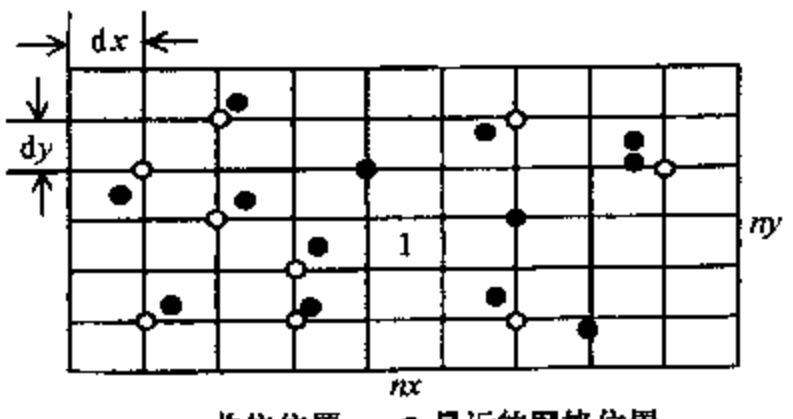


图 7.1　网格化示意图［左上角的网格系统模型（1，1）］

并将已知井位的相观察数据赋于最近的网格结点，这样做有助于格点的迁移。

设研究区域共有 N 种相 F_1，F_2，…，F_N，不失一般性，可令：

$$F_i = \{x \mid \tau_{F_i}(x) = 1\}$$

这里，$\tau_{F_i}(x)$ 为相 F_i 的指示函数，即

$$\tau_{F_i}(x) = \begin{cases} 1, x \in F_i \\ 0, x \notin F_i \end{cases}$$

若用整数对相进行编码，可令 $L(x)$ 为定义在平面区域 D 上的离散型变量的随机函数，$L(x)$ 可能的取值为 1，2，…，N，且 $F_i = \{x \mid L(x) = i\}$ $(i = 1, 2, \cdots, N)$，则在位置 x，若 $L(x) = i$，则表示该位置属于相 F_i。该相观察被赋于邻近网格结点后，则所有网格结点处具有指定的值：

$$L(i,j) = \begin{cases} k, \text{位置}(i,j)\text{处有第}\ k\ \text{种相观察} \\ 0, \text{其它情形} \end{cases}$$

$$i = 1,2,\cdots,nx, \quad j = 1,2,\cdots,ny$$

7.2　河道源头

设 $k=1$ 时，指明河道砂体，而 $k=2$，…，N 时为其余相（非河道），那么可用下面方法寻找可能的河道源头。

在网格系统 y 方向上端，选取一个指定宽度 Δy（Δy 用若干个网格点来表示），依次从北向南 $ny \to ny - \Delta y$，从南向东 $1 \to nx$ 搜索，若找到一个 $k=1$ 位置网格结点 m_1，而紧随其后有观察 $k \neq 1$，则该位置，就是第一条河道可能的源头（m_1，1），依次类推可以得到若干条河道源头（m_1，1），…，（m_u，u）。总之，认为两条河道源头之间肯定有非河道相观察，如图 7.2 所示。

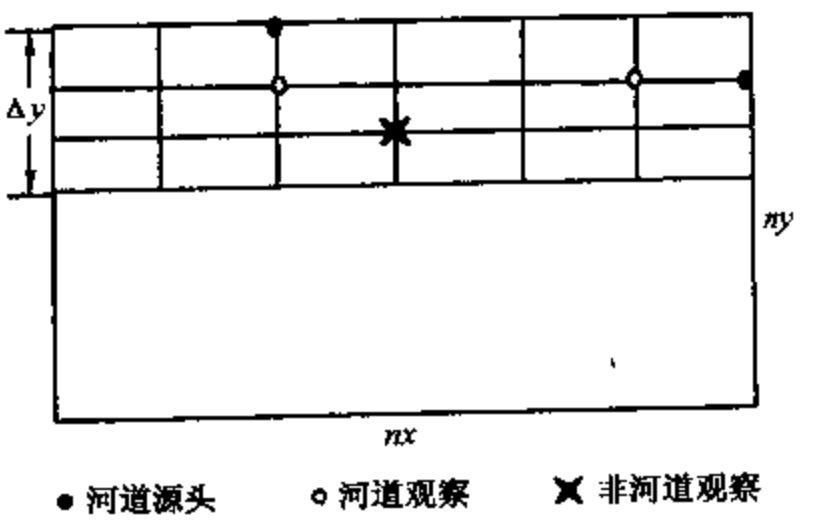


图 7.2　河道源头寻找示意图

7.3　河道主流线位置

第一河道及其随后河道主流线的位置寻找，是由源头开始的，并依次向近邻的某个网格结点移动。如图 7.3 所示。

在流程图中，P_1 为当前网格位置。若令 P_1 (i, j)，则有 P_a $(i-1, j)$，P_b $(i-1, j+1)$，P_c $(i, j+1)$，P_d $(i+1, j+1)$，P_e $(i+1, j)$。依次判别 P_1 向 P_a，P_b，P_c，P_d，P_e 5 个方向的邻近网格结点的哪一个迁移，步骤如下：

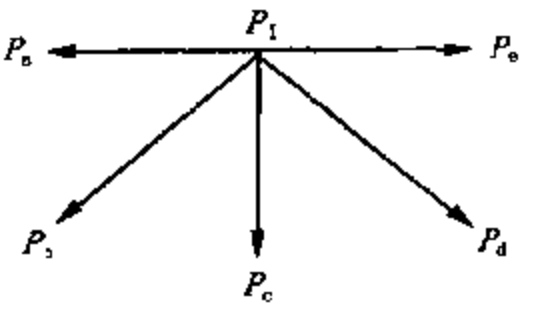


图 7.3　网格位置迁移示意图

1）在 a 方向寻找是否有最近的河道观察位置，若有，则 $P_1 \rightarrow P_a$，即 P_1 向 P_a 迁移。

2）在 a 方向和 b 方向范围$(\overset{\wedge}{a,b})$内，寻找是否有最近的河道观察位置，若有，则有 $P_1 \rightarrow P_b$。若$(\overset{\wedge}{a,b})$范围内无最近的河道观察，如有河道观察，则以一定的概率由 P_1 向 P_b 迁移，移迁概率为：

$$P(P_1 \rightarrow P_b) = \frac{\lambda_{ab} d_{ab}}{\sqrt{(nx)^2 + (ny)^2}}$$

在此 d_{ab}为 P_1 至$(\overset{\wedge}{a,b})$最近的非河道观察位置距离。值得指出的是，该距离是用整数点坐标计算的。λ_{ab}为一数乘因子（人工选取）。由该迁移概率可以知道，d_{ab}增大时，迁移概率亦增大，d_{ab}减少时迁移概率亦减小。

3）在$(\overset{\wedge}{b,c})$及范围内寻找最近的河道观察位置时，若存在，则 $P_1 \rightarrow P_c$。若$(\overset{\wedge}{b,c})$范围内无最近的河道观察，但有河道观察，那么 P_1 以一定的概率向 P_c 迁移，迁移概率为：

$$P(P_1 \rightarrow P_c) = \frac{\lambda_{bc} \cdot d_{bc}}{\sqrt{(nx)^2 + (ny)^2}}$$

在此 λ_{bc}为数乘因子，d_{bc}为最近观察距离，易知当 d_{bc}增大时，P $(P_1 \rightarrow P_c)$ 也增大。

4）在$(\overset{\wedge}{c,d})$范围内有最近的河道观察，则有 $P_1 \rightarrow P_d$，若在$(\overset{\wedge}{c,d})$范围内无最近的河道观察，但有河道观察，则迁移概率为：

$$P(P_1 \rightarrow P_d) = \frac{\lambda_{cd} d_{cd}}{\sqrt{(nx)^2 + (ny)^2}}$$

显然，亦有 d_{cd}增大，P $(P_1 \rightarrow P_d)$ 也增大。

5）若 P_1 没有向 P_a，P_b，P_c，P_d 等方向迁移，则一定有 $P_1 \rightarrow P_e$。为了保证河道主方向为从北至南方向，一般地规定：$\lambda_{bc} \geq \lambda_{ab}$和 $\lambda_{bc} \geq \lambda_{cd}$。如此不断变更。当迁移至边界位置时，停止操作。这样就获得了第一条主河道位置。

7.4　分支河道位置

一般来讲，根据沉积持点，每条河道可能有若干条分支河道。因此，仍沿第一条河道源

头搜索，找其分支河道位置。寻找方法与 7.3 节中的类似，只不过此时每个范围搜索至第一条河道主流线位置为终结。如图 7.4 所示。

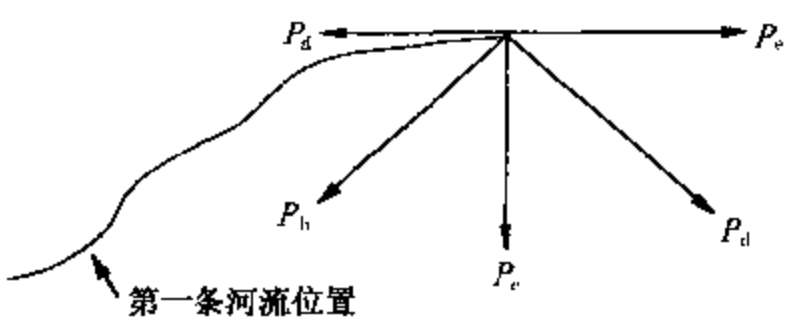


图 7.4　第一条主河道的第二条分支河道位置搜索示意图

另外需指示的是，每个范围搜索的最边位置除观察位置处，还可能是第一条河道已走过的位置。通过调整相应的数乘因子与最近距离种类（是否为第一条河道位置），使第二条分支河道与第一条分支河道具有某种汇合与分叉的可能性。即改变汇合或分叉的可能性。

这样，通过不断从源头出发可以模拟不同的分支河道，一般来讲，分叉河道条数愈多，则出发次数应增加。通过对辫状分流河道模拟实践研究发现，分叉主河道可以出发三条就可获得满意效果。

7.5　河道主流线位置的修正

一般情形下，河道主流线可能出现下述几种情形，图 7.5 所示。

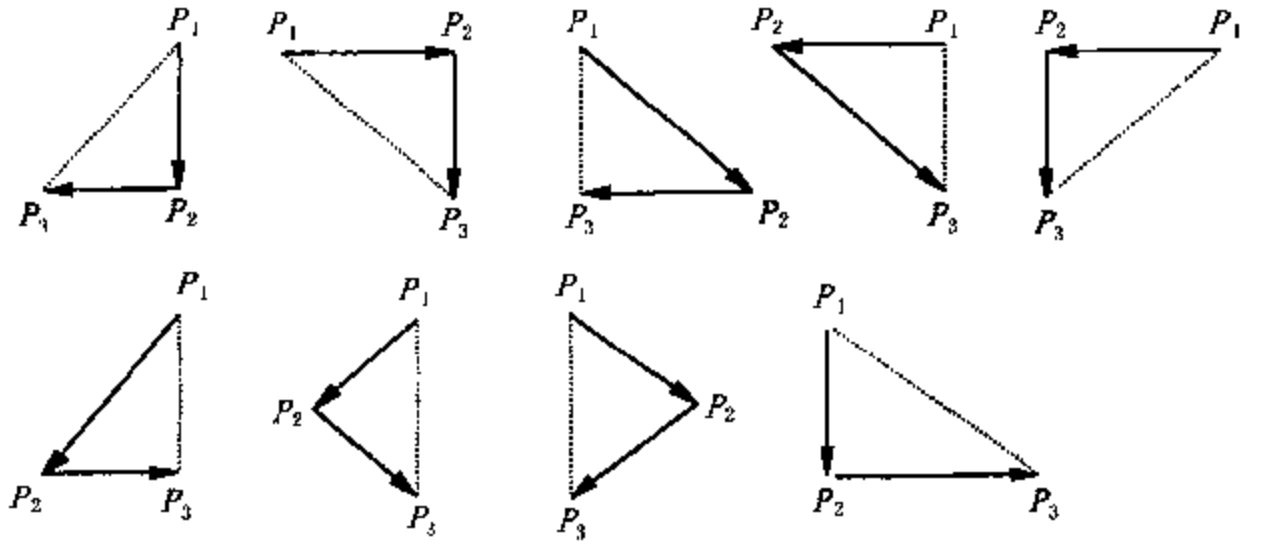


图 7.5　河道位置的修正

此时，可以剔除中间位置 P_2，使 $P_1 \xrightarrow{\text{直至}} P_3$，这样可以简化加宽河道计算，同时不影响河道主流线。

7.6　河道加宽与光滑

对每条河道主流线可以相应向两侧加宽，加宽示意图为图 7.6 所示。

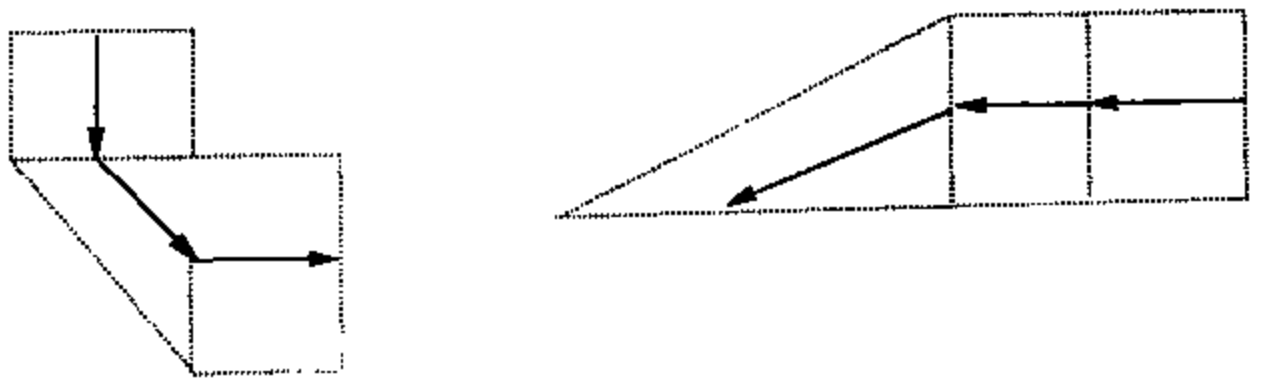


图 7.6　河道主流线加宽示意图

值得注意的是，在加宽过程中，应将邻近的非河道观察位置剔除掉，并确定向两侧加宽的宽度。

$$\Delta L_1 = \min(\Delta D_1, \text{width}/2), \Delta L_2 = \min(\Delta D_2, \text{width}/2)$$

在此，width是河道平均宽度，ΔD_i = 相应一侧最近邻的非河道观察距已知网格点的距离，$i=1$，2，$\Delta L_1 + \Delta L_2 \leq \text{width}$。这样可以获取每条河道主流线的两个边界位置（用浮点数的存贮）。再进行插值光滑可以获得光滑的每个河道的边界位置。

7.7　实例研究

大庆油田的主要产油区：萨尔图、葡萄花、高台子油层，形成于下白垩纪中期，为一套典型的大型河流—三角洲沉积物，具有埋藏深度浅，油层厚度大，平面及纵向非均质严重的特点。多年的开发地质研究成果说明大庆油田的河流相储层砂体类型包括：砂质辫状砂体、曲流河砂体、高弯度分流河道砂体、低弯度分流河道砂体和顺直型分流河道砂体。河流相储层主要分布于油田北部的萨葡油层中，在大庆油田的高产稳产中起着举足轻重的作用。然而，由于其内部结构的复杂性，严重地影响了采收率。油田开发后期要充分挖掘这部分潜力，首先必须进行精细的储层描述工作。其中，利用密井网数据进行平面连续性的精细描述，对于深入认识河流相储层剩余油的平面分布状况和建立精细地质模型具有重要的意义（隋军等，2000）。河流相储层内划分河道的流向和位置，对于储层平面连续性的精细描述至关重要。

仍以第 5 章的 DQ 油田为例，彩图 18 和彩图 19 是利用随机游走模型做出的两张沉积相图。彩图 18 是 BY 区西区 SS 油层组第 9 小层的水下分流河道或窄分流河道。彩图 19 则是第 10 小层的沉积相图。彩图 20 也是第 10 小层的沉积相图，只是不同的实现。河道走向为北南向，河道比较窄。

应该指出的是，由于其河道狭窄，分叉会合频繁，上述地区的这种水下分流河道在 200～300m 的米井网数据之下，要确定河道的走向和位置比较困难。一般的随机建模方法都不奏效。这就是本章提出的随机游走模型的意义所在。

第 8 章　示性点过程模型

示性点过程的储层随机建模方法，最早是在十多年前由挪威学者提出的，并一直是由他们热衷研究的方法。目前，人们称这种方法为基于目标（Object － Based）的方法，而把指示模型、截断高斯模型等称为面向网格点的方法。

8.1　点过程和示性点过程

本节首先引出二项式点过程和泊松点过程，然后讨论点过程和示性点过程基本原理。这些概念都是属于随机几何学的范畴。

随机几何学的现代理论是由 D.G.Kendall，K.Krickberg 和 R.E.Miler 创立的。该理论在较复杂概率分布下考虑了随机几何模式（Stoyan，1987）。

一般的几何学研究的基本成分是点。用随机几何学的术语，点的模式称为点过程。点过程在随机几何学中发挥着基本的作用。在随机几何学中，除了描述点的点过程以外，还包括有研究不规则随机模式的随机闭集，研究纤维和表面的纤维过程、表面过程和描述晶体结构，断裂模式和泡沫结构的随机拼嵌模型，以及包括随机几何学的若干应用，其中有利用一维和二维截面上提供的信息来推断三维几何体的性质的方法等。泊松点过程是最简单的点过程。布尔模型是最简单的随机闭集。泊松点过程中随机分布的点被换成了一个一个以具有固定半径 r 的圆盘，泊松点过程就变成了布尔模型。泊松点过程中的各个点被称为萌芽，而布尔模型中的每个圆盘叫做基本颗粒。

点过程提供了各种模型来研究点的不规则空间分布模式。由于物理学，生物学和排队论等理论的需要，促进了点过程的理论的发展。这个理论的先驱者是 Neyman（1939）。这些点的空间分布的随机模式可以显示为这些点的相互作用。它们可以一组一群地分布在空间中，也可是非常规则地排列，或是其距离具有最小值；也有可能以上特征同时出现在点的分布模式中。点过程的统计理论就在于刻画和定量描述这些相互作用。点过程理论的模型可以用于比较原始的点分布模式，也可用来表达这些模式。

如果不考虑点的相互作用，那么点空间分布的随机模式就是完全随机的。这种模式的理论模型对这些点的比较具有很大的重要性。

二项式点过程和泊松点过程是最简单和最重要的点过程。它们是作为完全随机的点过程出现。然而，示性点过程则是一种特殊的点过程。以上几种过程中，点过程是核心。以下对它们分别进行讨论。因为以下的有关数学讨论比较严密，涉及了集合及其运算和测度等概念。对此，请参考有关集合论和实变函数论的专著和教科书。

8.1.1　理论准备

为了更好理解点过程的定义，以下对波雷尔（Borel）集、σ 代数系统、可测空间［X，χ］、测度空间［X，χ，μ］和概率空间［Ω，A，P］进行有关解释。

首先，波雷尔集，就是用闭集经过有限多次或可数多次的并，交和补的运算，得到的集合。

一个 σ 代数系统是指某一个背景的集合 X 的子集构成的一个系统，记为 χ。它满足如下三个条件：

1）$X\in\chi$。

2）若 $A\in\chi$，则其补集 $A^c\in\chi$。

3）若 A_1，$A_2\cdots\in\chi$，则 $\bigcup_{k=1}^{\infty}A_k\in\chi$。

如果空间 X 和一个 σ－代数系统 χ 结合，那么 $[X,\ \chi]$ 被称为可测空间。一个函数 f：$X\to R$ 称为 χ—可测，意味着，如果对于 B^1 中的所有的波雷尔集 B，该函数的反函数 $f^{-1}(B)=\{x\in X:\ f(x)\in B\}$ 属于和 X 联系的 σ 代数系统 χ。当基本的代数运算应用于可测函数时，产生的仍是可测函数。

在可测空间 $[X,\ \chi]$ 上的一个测度 μ 定义了三个元素的组合 $[X,\ \chi,\ \mu]$，称为测度空间。和每一个测度空间相伴的是一个定义在 χ—可测函数构成的系统上的一个函数 f：$x\to[0,\ \infty]$，给每一个函数 f 分配一个落在 $[0,\ \infty]$ 的值。

$$\int f\mathrm{d}\mu。$$

这个值称为 f 的积分。

如果一个测度空间 $[\Omega,\ A,\ P]$ 使得 $P(\Omega)=1$，即可被称为一个概率空间，而且测度 P 被称为概率，或者称为概率测度。

8.1.2　二项式点过程

最简单的点过程是只包含 1 个点。在一个紧致集 $W\subset R^d$ 均匀分布的一个随机点，点落在 W 的一个子集的概率为该子集和 W 的体积之比：

$$P(\zeta\in A)=v_{\mathrm{d}}(A)/v_{\mathrm{d}}(W)\tag{8.1}$$

这里，A 是 W 中的波雷尔集。研究均匀分布的随机点及其分布概率，构成了随机几何学的许多研究方向。从一定程度上讲，现代随机几何学的进步可以归结为，均匀分布的随机点被更为一般的随机点模式所代替。

一个随机点的均匀分布是一个简单的随机模式。然而，n 个互相独立的均匀分布的随机点能互相叠加形成一个新的点过程。这就是含有 n 个点的二项式点过程。正式地讲，由 n 个互相独立的、均匀分布的在同一个紧致集 W 中的点 ζ_1，…，ζ_n 构成的点过程就是一个二项式点过程。根据（8.1）式，可得：

$$\begin{aligned}P(\zeta_1\in A_1,\cdots,\zeta_n\in A_n)&=P(\zeta_1\in A_1)\cdots P(\zeta_n\in A_n)\\&=v_{\mathrm{d}}(A_1)\cdots v_{\mathrm{d}}(A_n)/v_{\mathrm{d}}(W)^n\end{aligned}$$

这里，A，…，A_n 是 W 中的 n 个波雷尔集。这相当于一个等价的定义：如果随机向量 $(\zeta_1,\ \cdots,\ \zeta_n)$ 是均匀分布在 W^n 上，那么点构成了 R^d 子集上的一个二项式点过程。图 8.1 展示了在 $W=[0,\ 1]^2$ 上的一个二项式点过程在 $n=50$ 之下的模拟结果。

如果 ζ_1，…，ζ_n 构成了在 W 中的一个二项式点过程，那末这个随机模式是由 $\Phi_W^{(n)}$ 表示的这些点所构成。对最一般的用途，这些点的排列是不成比例的，$\Phi_W^{(n)}$ 能被视为一个随机集。它能被视为一个随机计数测度。对一个波雷尔集 A，用 $\Phi_W^{(n)}(A)$ 表示 $\Phi_W^{(n)}(A)$ 落在 A 中的点的数目，那末

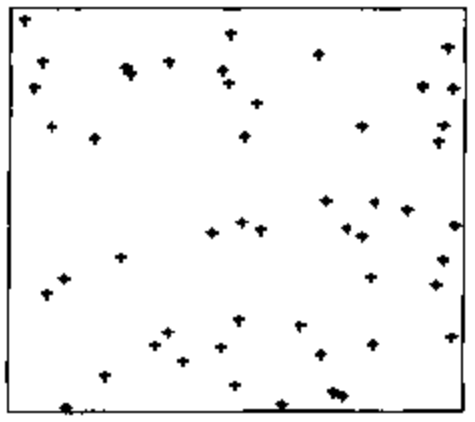


图 8.1　一个均匀分布在一个单位内的二项式点过程（Stoyan and others，1987）

$$\Phi_W^{(n)}(\phi) = 0, \quad \Phi_W^{(n)}(W) = n;$$

$$\Phi_W^{(n)}(A_1 \cup A_2) = \Phi_W^{(n)}(A_1) + \Phi_W^{(n)}(A_2);$$

这里，A_1，A_2 是 W 中互不相交的子集。这时，σ—可加性的性质存在。所以，$\Phi_W^{(n)}$ 确实是一个随机测度。

二项式点过程具有一般点过程的典型的双重意义：它可作为一个随机集，又可作为一个随机测度。这个典型的双重意义是所有点过程的一个重要性质。

8.1.3　泊松点过程（Poission Point Process）

泊松点过程可以分为是平稳的和广义的两种。一个平稳的泊松点过程 Φ 具有以下两个特点：

1）点数的泊松分布。在一个有限的波雷尔集中，Φ 中点的数目是具有泊松分布的随机变量。其均值为 $\lambda_{vd}(B)$，其中 λ 是一个常数；

2）独立的分散性。在 k 个互不相交的波雷尔集中，Φ 中的点数目构成了 k 个独立的随机变量。

以上第 2 个特征反映了“完全地随机”即“纯粹的随机”的性质。在第 1 条性质中的 λ 是平稳的泊松点过程的一个特征参数。它给出了在单位体积内点的数目的均值：

$$\lambda_{vd}(B) = E\Phi(B)。$$

对所有的有界的波雷尔集，它被称为平稳的泊松点过程的强度或密度。其实平稳的泊松点过程的以上这两个性质并不是独立的。这个命题已被 Renyi（1967）证明了。

在实践中往往出现显示出震动的点的分布模式。例如，它们在某一个方向上呈现出明显的增加的趋势。这种现象能在某些森林中的树的位置中看出。可以假设相对一个方向森林相对于海平面的高度在增加，或土壤逐渐变得干燥。这样，沿着这个方向，单位区域内树的平均数将减少。这时，就需要用广义的泊松点过程来描述。

广义的泊松点过程用一个散开的随机测度 Λ 来表征，这个随机测度称为在 R^d 上的广义泊松点过程 Φ 的强度测度，且对应于平稳泊松点过程的强度测度。广义的泊松点过程 Φ 具有如下两条性质：

1）点数的泊松分布，在一个有界的波雷尔集合 B 中的 Φ 中点的数目是一个均值为 $\Lambda(B)$ 的泊松分布的随机变量：

$$P[\Phi(B) = m] = [\Lambda(B)]^m \cdot \exp[-\Lambda(B)]/m!$$

对于：$m = 0, 1, 2, \cdots$。

2）独立的分散性，在 k 个互不相交的波雷尔集中，Φ 的点的数目是 k 个独立的随机变量。

如果随机测度 Λ 具有关于勒贝格测度的密度，然而它可写成，

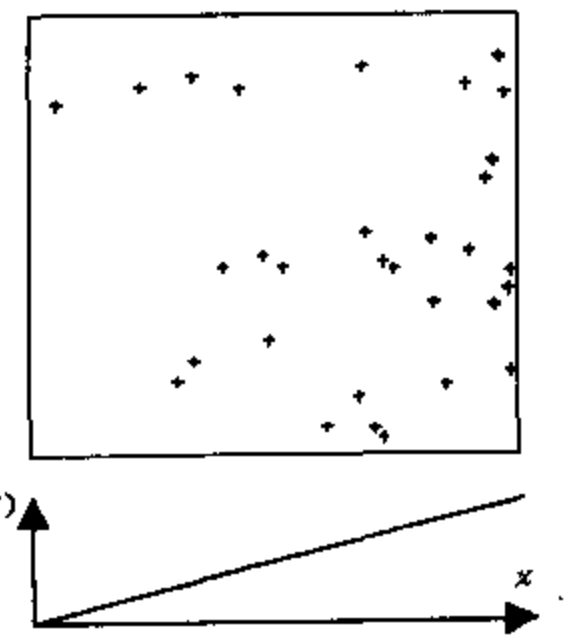


图 8.2　具有线性变化函数 $\lambda(\chi)$ 的泊松点过程（Stoyan and others，1987）

$$\Lambda(B) = \int_B \lambda(x)dx$$

对于波雷尔集合 B。密度 λ（x）称为泊松点过程的强度函数。由此可见，广义泊松点过程的点数的泊松分布是依赖于所取的波雷尔集合 B。

8.1.4　点过程（Point Process）

在讨论了两种具体的点过程：二项式点过程和泊松点过程的基础上，以下讨论一般的点过程的定义。

d 维空间 R^d 上的一个点过程，定义为取值于一个可测空间［$\boldsymbol{N}$，$\mathcal{N}$］上的一个随机变量。这是点过程的第一种定义。这里，$\boldsymbol{N}$ 是 R^d 的所有子集 φ 的一个组合。这些子集 φ 满足以下的条件：

1）子集 φ 是局部有限的，它必须包含，且仅包含有限多个元素 x_n；

2）集合 φ 是简单的，于是当 $i \neq j$ 时，有 $x_i \neq x_j$。为简便起，φ 有时可写成 φ $\{x_n\}$。

在以上的两个条件中，$\boldsymbol{N}$ 中的每个 φ 可视为 R^d 中的一个闭子集。N 中的一个元素 φ 可视为在 R^d 上的一个测度，使得 φ（B）是 B 中的 φ 所含点的数目，B 是一个有界的波雷尔集合。$\boldsymbol{N}$ 可定义为 $\boldsymbol{N}$ 上的最小的，使得所有的映射 $\varphi \to \varphi$（B）是可测的 σ－代数系统。

更正式地，一个点过程 Φ 是从概率空间［Ω，A，P］到可测空间［$\boldsymbol{N}$，$\mathcal{N}$］的可测映射。更直观地说，它是在 $\mathcal{N}$ 中的所有的 φ 中的随机地选择的一个。它产生一个定义在［$\boldsymbol{N}$，$\mathcal{N}$］上的分布，即为 Φ 的分布 P。

作为点过程的其它两种定义，点过程可看成是在一个空间区域内的离散点的随机集合，也可看成是该随机集合落在这个空间区域内的点的数目的随机测度。这里 $x \in \Phi$ 的含意是点 x 属于随机序列 Φ，而 Φ（B）$= n$ 的含意是集合 B 包含着属于 Φ 的 n 个点。

一个点过程 Φ 的分布由下列概率确定：

$$P(Y) = P(\Phi \in Y) = P(\{\omega \in \Omega; \Phi(\omega) \in Y\})$$

对 $\mathcal{N}$ 中的 Y。分布的有限维是特别重要的。它们具有如下形式的概率：

$$P\{\Phi(B_1) = n_1, \cdots, \Phi(B_k) = n_k\}$$

这里 B_1，…，B_k 是有界的波雷尔集合，以及 n_1，n_2，…，$n_k \geqslant 0$。Φ 在［$\boldsymbol{N}$，$\mathcal{N}$］上的分布函数由这样的系统唯一地确定，该系统在所有的值 $k = 1$，2，…上有定义。事实上，它是由成分 B_i 两两互不相交组成的子系统确定的。

一个更小的子系统是概率为零的子系统：

$$v_B = P(\{\varphi \in N; \varphi(B) = 0\}) = P[\Phi(B) = 0] = P(\Phi \cap B \text{ 为空})$$

对于波雷尔集 B。如果一个点过程是简单的，那么 P 由值为 v_K 值组成的系统确定，K 是 R^d 上的一个紧集。

8.1.5　示性点过程（Marked Point Processes）

一个点过程，对其上的每一个点被赋予一个特征（或称为一个属性，或示性）时，就称为示性点过程。这样，在 R^d 上的一个示性点过程是以［x_n；m（x_n）］为元素的一个随机序列 $\Psi = \{[x_n; m(x_n)]\}$。这里，点 x_n 构成了在 R^d 上的一个非示性的点过程，且是不

含任何特征，而 m（x_n）就是对应于 x_n 的特征。这些特征可以含有相当复杂的结构。这些特征属于一个包含特征的给定的空间 M^*，可以假定是一个 Polish 空间。M^* 的波雷尔 σ－代数系统可记为 μ^*。

Matthes（1963）和他的同事们建立了示性点过程的理论。其最初的应用在排队论方面（König，等人，1967）。这方面的应用也可在 Franken 等人（1981）的专著中见到。

以下是示性点过程的一些例子：当 x 是一个粒子的中心时，m（x）可以是该粒子的体积；当 x 是一棵树的位置时，m（x）可以是该树干的直径；当 x 是一个原子的中心时，m（x）可以是该原子的类型。

在以上前二个例子中，这些特征（示性）可以是连续变量。在以上的第三个例子中，这些特征可以是类型。相应的示性点过程可称为多元点过程或多类型点过程。图 8.3 和图 8.4 和是示性点过程的实例。

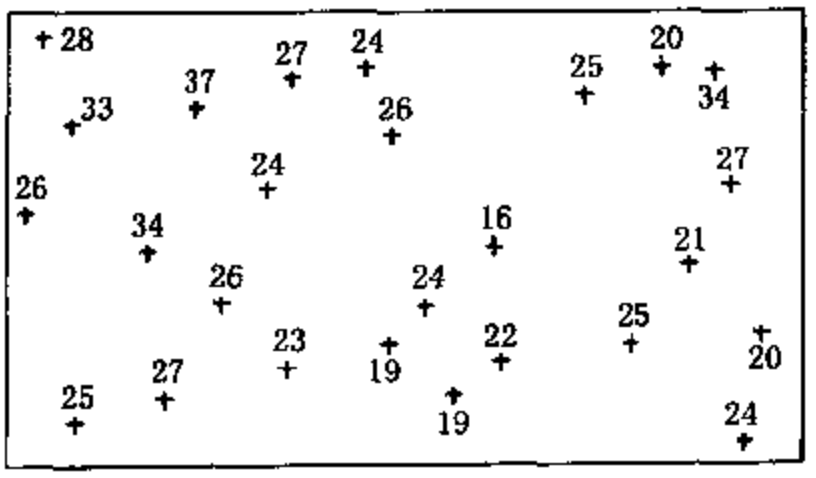


图 8.3　森林中松树的位置。数值表示树干的直径，可以被解释为作为位置的点过程的特征。（Stoyan and others，1987）

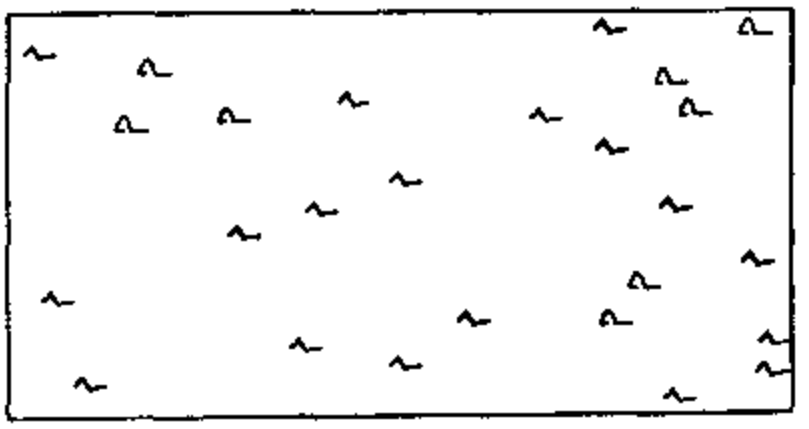


图 8.4　森林中云杉和白桦的位置。该示性点过程具有特征 0 和 1。（Stoyan and others，1987）

Ψ 可以经常被解释为 $R^d \times \mathrm{M}^*$ 上的一个非示性的点过程。若 $\chi \in \Psi$，则 $\chi = [x, m(x)]$，即 χ 由 x 和 m（x）两部分组成。这里，$x \in R^d$，m（x）$\in M^*$。对于一个波雷尔集和 μ^* 的子集，在 B 中 Ψ 的点的数目，加上 L 中的特性被记为 Ψ（$B \times L$）

8.2　模拟退火算法

模拟退火算法可以在储层随机建模示性点过程中使用，也可以作为单独一种算法在储层随机建模中使用。它在本章的以下部分将被应用。

模拟退火算法近来在模拟地球科学现象中受到相当的重视。这是因为这种方法能提供大量广泛的条件化信息。它不象其地质统计学的模拟方法，只能要求满足少量经过选择的统计量，如直方图和变异函数等。模拟退火算法可以要求满足几乎可以是任意的可以想象出的统计量。

退火模拟相对来讲是一个新方法，可用来产生无论是连续或范畴变量的各种条件随机图像，并且以数值计算为主要目的。该技术具有这样的能力，可综合再现两点统计量和隐含的复杂多元空间统计量。该技术将花费大量的 CPU 时间和磁盘空间。对此，一种解决途径是先用快速的模拟算法生成粗糙的随机图像，然后用采退火模拟算法对该图像进行修正。可选择的快速模拟算法有序贯高斯算法或中位指示算法。

8.2.1　算法概要

考虑一个简单的问题，要求能生成一个取值为 0 或 1 的指示变量的空间分布。利用条件模拟，该指示变量的空间分布可满足通常所用的条件信息：直方图、变异函数以及变量在若干点处的值。模拟退火算法可用以下步骤解决这个问题。

1）对一个有观测数据的网格节点，按以下规定分配一个指示值：如果这个观测数据是 1，那么就赋予 1；如果是 0，就赋 0。

2）对观测数据已经作为条件化信息使用后，仍然空着的那些所有的网格节点随机地用一定的概率赋予 0 或 1。在任何一个点处以概率 p_0 选择指示值为 0，而选 1 的概率则为 $p_1 = 1 - p_0$。p_0 和 p_1 值的选取可根据 0 和 1 的全局分布性质而定。

3）选择一个“能量函数”E 作为训练图像的统计质量和初始网格节点的值的相对应的质量之间的偏离的质量。该能量函数的选取在稍后提及。

4）选取一个开始温度 t。初始温度的选取，温度下降之前交换的次数，和温度下降的速度，三者都是在有关退火技术的文献中所指的“冷却进度表”中所经常提到的。

5）选择和观测点不相关的任意两个点，并交换它们的指示值。计算交换后的能量函数 E_{new}，并把它和交换前的能量函数 E_{old}相比较。如果新的能量函数小于旧的，那么就采纳这次交换。如果大于，那么就以下列概率 P_{accept}来采纳这次交换：

$$P_{accept} = \exp(-\frac{E_{new} - E_{old}}{t})$$

6）重复以上的步骤，并轻微地减低温度 t。

7）重复以上的两个步骤，直到能量趋于 0。

退火算法的关键是网格上数据的叠代更新，而用能量函数改变的大小来作为判断一次交换是否成功的依据。倘若指示变量的空间分布希望满足给定的指示变异函数的要求，能量函数应有如下的形式：

$$E = \sum_{i=1}^{n} [\gamma_{grid}(h_i) - \gamma_{model}(h_i)]^2$$

这里 $\gamma_{grid}(h_i)$ 是实验变异函数，$\gamma_{model}(h_i)$ 是相应的理论变异函数。这个能量函数是用一个数值反映了两者之间的差别。

当指示值在几个网格节点上被交换后，新的网格节点的值对应的实验变异函数可能真正地接近于理论变异函数，人们所希望的是某种侥幸的出现：能偶尔发现一组交换碰巧能把某种空间连续性赋予网格节点。在第 5 步的接受/拒绝过程描述了上述的保证。人们经常接受朝着最终目标的进步，同时在某些时候是要忍受拒绝。在初始时，温度 t 被置成某一个非常高的数值，拒绝的概率将是相当高的。作为一个叠代算法的过程，t 将随着时间越来越减小，拒绝的概率也将减少。

通过一系列的试验和误差对比，算法最终将发现一系列的交换能将较好的空间连续性赋予网格。最终，正如人们所希望的那样，有可能取得的 E 值接近于零。在这些点，这网格节点上的指示值最终具有正确的变异函数。在每一个滞后，E 的计算对于评价能量函数都是有用的。该网格节点的实验变异函数将匹配变异函数的理论值。对计算能量函数没能考虑的那些滞后，模拟退火算法并不保证实验变异函数在那些滞后值处是正确的。

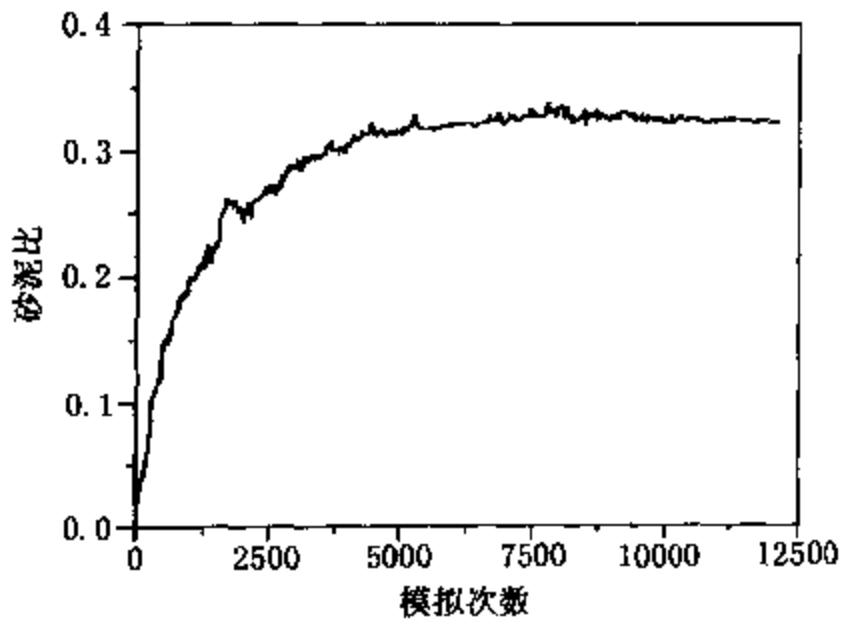


图 8.5　能量函数为砂泥比的退火实例

图 8.5 是 GD 油田的一个应用模拟退火的实例。它在采用了示性点过程模拟河道的基础上，以砂泥比作为能量函数进行退火的结果。根据该储层内各井的数据统计，即利用各井中砂岩的长度与该层全部井身的长度之比，可得到该层的砂泥比是 0.3。每作一次模拟，获得一个河道的空间分布，可以计算河道的体积占全部储层的体积之比，将此作为相应的砂泥比。模拟退火的目标是该砂泥比应该最终达到 0.3。该实例共进行了大约 12500 次模拟，最后砂泥比到达 0.3 上下，离 0.3 已经很接近。在这进行 12500 次模拟时，后一次的结果和前一次的结果相比较，有时可能只是减少了一条河道，有时可能只是增加了一条河道，而有时可能只是交换了上一次模拟结果中已有二条河道的位置。到底是哪一种可能，根据随机抽样等概率地决定。显然，模拟退火的结果使得河道空间分布更加符合实际，更加合理。

8.2.2　优点

相对进行条件模拟的其他地质统计学方法，退火最大的优点是：人们可以将所期望的任何统计量组合进入能量函数。在上述的例子中，所做的能量函数仅仅依赖变异函数。如果，为了更满足这个变异函数，需要进一步肯定，有 2×2 组的网格节点完全是 1，那么可以定义能量函数如下：

$$E = w_1 \cdot \sum_{i=1}^{n} [\gamma_{\text{grid}}(h_i) - \gamma_{\text{model}}(h_i)]^2 + [N_{2\times2,\text{grid}} - N_{2\times2,\text{model}}]^2$$

人们可以发现在条件模拟中，一些相当复杂的统计量可以被结合进来，成为能量函数的一部分，如试井分析，示踪测试和地震资料。

8.2.3　缺点

模拟退火算法作为一门专门的技术是最近才发展起来，还不甚成熟。这是它的主要缺点。它还没有积累足够的智慧和经验，也没有很多的文献可以提供更多的技巧，以设计合适的能量函数和冷却计划。

另外模拟退火算法的一个缺点是，当能量函数比较复杂时，算法收敛很慢。因为退火过程要数以千万次的交换，来发现一个网格的统计量和理想的图像的统计量相匹配。退火的实际的应用应包括容易更新的能量函数。如果能量函数不能被更新，那么在每次交换后，新的能量函数就必须放弃，而要重新计算。这样模拟退火算法所付出的计算量就会过高。

退火的另一个缺点是理论上的。它缺乏统一的数学工具。其基本算法是一个优化过程，它会引起许多问题。例如使用其算法会导致什么样的不确定性，以及是否能产生相同可能的现实，等等。

8.3　示性点过程建模的原理和算法

随机几何学中的点过程理论在对地质体进行随机建模时是极有用的，地质沉积相空间分

布问题在石油勘探开发中有着极其重要的意义。由于地质事件，地质过程的复杂性，以及获取资料的局限性，这类问题面临着如何用有限的信息来恢复地质沉积相的空间形态。这时，预测的不确定性是经常发生的，因此使用统计方法建立地质沉积相空间分布的随机模型是合适的。在国际上，使用随机点过程理论对地质体进行建模已引起足够的重视。

8.3.1　布尔模型与示性点过程模型

1. 布尔模型

设 $\phi=\{x_1, x_2, \cdots x_n, \cdots\}$ 是 R^d 空间中参数为 λ 的泊松点过程，$\Sigma=\{\Sigma_1, \Sigma_2, \cdots \Sigma_n, \cdots\}$ 为 R^d 中独立同分布的随机紧集序列，且 Σ 与 ϕ 相互独立，又假设 $E_{\gamma d}(\Sigma_0 \oplus K)<\infty$ 对所有紧集 $K \in R^d$ 成立，至此，$E_{\gamma d}$ 是 R^d 空间中的 Lebesque 测度，Σ_0 是与 Σ_n（$n=1, 2, \cdots$）同分布且与 Σ_n（$n=1, 2, \cdots$）和 ϕ 独立的随机紧集，那么

$$B=\bigcup_{n=1}^{\infty}(\Sigma_n+x_n)=(\Sigma_1+x_1)\cup(\Sigma_2+x_2)\cup\cdots$$

称为具有生长颗粒 Σ_0 的布尔模型。

在实际问题中，生长颗粒 Σ 可选取成随机半径的圆盘，随机长度和方向的线段，或有限数目随机聚类的点等等。生长颗粒的形状用它们的分布 M 来刻画。由于 Σ 是随机集，因而它们的分布是在一个参照空间 K 上的概率测度，也就是示性点过程的属性分布（Mark Distribution）。

2. 示性点过程

示性点过程是这样一个随机过程，它把一种特征或属性赋予过程的每个点上，因此，R^d 上的一个示性点过程是一个随机序列 $\Psi=\{X_n, m(X_n)\}$，$\{X_n\}$ 是一个点过程，$\{m(X_n)\}$ 是相应的属性。属性 $\{m(X_n)\}$ 可以是相当复杂的，它们对应于一个指定的属性空间 M^*，M^* 常假定为 Polish 空间，M^* 的波雷尔 $\sigma-$代数系统记为 μ^*。

布尔模型可以看成是示性点过程的简单情形，那就是种子点过程是泊松过程，种子的位置是相互独立的，均匀分布于研究区域之中的。而一般的示性点过程，其种子点过程可以是一般的点过程。使用简单的点过程（如泊松过程），产生复杂点过程的方法主要有四种：

1）薄化技术（Thinning）。

2）聚类技术（Clustering）。

3）叠加技术（Superposition）。

4）直接改变分布的概率密度。

使用上述方法 4）可以得到在统计物理中非常有用的点过程——吉布斯点过程（Gibbs Point Processes）。这种点过程与吉布斯分布有关。吉布斯分布描述了巨大而封闭的物理系统内部子系统的平稳态。这种过程在实际应用中非常广泛。在本文谈及的地质应用中所使用的示性点过程中，种子位置就是由吉布斯点过程产生。

8.3.2　示性点过程建模的原理

地质沉积相的分布常用取值为离散的随机场 $\{L(x): x \in D\}$ 来描述，其中 x 是参照域 D 中的点，参照域 D 代表了储层在空间分布的范围，$L(x)$ 则代表位置 x 处相的类型，常取值于 $\{l_1, l_2, \cdots l_n\}$。在地质现象中经常遇到如下情形，那就是在起控制作用的背景相下出现小尺度的另外一种相单元，这常称为事件现象模型。如在砂岩为主的储层中常伴随着小尺度的泥岩单元，这可称为砂岩背景下的事件现象。大范围的地层断裂带往往是由许多

小的断层组成，那么这些小断层可看成是组成主断裂带的事件。在潮汐滩背景下河流冲积扇可视为事件现象，而海相或三角洲沉积相中的河道砂也可以看成是事件现象。下面就具体论述对地质沉积相带的空间分布的示性点过程建模。

1. 模型选取

可选示性点：$\vec{U}=\{x, L, \vec{S}\}$，其中，$x$ 是区域中随机参照位置，L 是相的类型，$\vec{S}=(S_1, S_2, \cdots, S_n)$ 是多元随机变量。$\vec{S}$ 表示研究对象的形状、尺寸、方向等，对于河道砂体则代表其长、宽、厚和方向。

设有 n 个相单元，则它们相互之间的关系可通过如下联合概率来描述：

$$P\{\vec{U}_1=\vec{u}_1,\cdots,\vec{U}_n=\vec{u}_n\}$$

这就指定了 D 中 n 个示性点所有可能展布情况的概率，显然这个联合概率分布函数的维数达 $n\times(3+1+m)$，处理起来比较困难，在实际应用中，常采用如下形式的参数化联合概率：

$$P\{\vec{U}_1=\vec{u}_1,\cdots,\vec{U}_n=\vec{u}_n\}=C_0\exp\{\sum_{i=1}^{n}\beta(\vec{u}_i)+\sum_{1\leqslant i<j\leqslant n}\gamma(\vec{u}_i,\vec{u}_j)\} \tag{8.2}$$

在此，C_0 是一个正则化常数，$\beta(\cdot)$ 是一个指定的函数。$\gamma(\cdot,\cdot)$ 是成对交互作用函数，反映了参照点和示性点属性之间的关系。

利用所研究区块的储层数据，可以得到某些指定位置或整个区域内的一些特征值，如对某种相占据的体积比例和物性参数的变异性进行估计。此时，从随机模型中所产生的实现必须能再生这些指定的特征，这就是条件模拟。所以，有必要对模型（8.2）进行拓广，以便迎合这些限制性的条件：

$$\begin{aligned}P\{\vec{U}_i=\vec{u}_i;i=1,2,\cdots,n\}&=C_0\exp\{\sum_{i=1}^{n}\beta(\vec{u}_i)\\&+\sum_{1\leqslant i<j\leqslant m}\gamma(\vec{u}_i,\vec{u}_j)\}\times\exp\{-\sum_{j=1}^{r}\sigma_j[\omega_j-g_j(\vec{u}_k,k=1,\cdots,n)]^2\}\end{aligned} \tag{8.3}$$

这里最后一个因子是一个限定值，它反映了 r 个限制 ω_i，$i=1, 2, \cdots, r$，与函数 $g_i(\cdot)$，$i=1, 2, \cdots, r$，所指定的实现的相应性质之间的偏差程度，且每个限制都有一个强度参数 σ_i 与之联系。这些强度参数 σ_i 增加时，所产生的限制影响将增加。如果这些参数为零，那么就相应于不考虑限制条件的情况。

2. 模型参数的确定

一旦建立了沉积相分布的示性点过程模型，下一步就是确定模型的参数，并利用随机模拟再现各种沉积相的可能分布。模型参数的确定，一方面，依赖于观察数据，另一方面也依赖于地质认识和经验。为了进行条件模拟，下列信息作为输入模型的参数是必须预先指定的：

1）确定性数据，如来自井中有关沉积相的数据。

2）各种沉积相对应的地层的中心位置，形状等参数的概率分布。

3）每种相所占据的频率。

4）相与相之间的空间关系。

5）沉积相的地质成因。

利用模型（8.3）可对研究区块分流河道砂体剖面进行建模和模拟，如选取：

$$\gamma(\cdot,\cdot) = 0$$

$$\beta(m) = \beta_1(x) + \beta_2(t) + \beta_3(w)$$

在此：β_1（x）表示位置 x 处河道发育的程度，服从［0，1］上的均匀分布；β_2（t）$\sim N$（$5m$，$1m^2$）表示河道厚度的概率分布，服从正态分布 $\mu=5m$，$\sigma=1m^2$；β_3（w）$\sim N$（$15m$，16）表示河道宽度的概率分布，服从正态分布 $\mu=15m$，$\sigma=4m^2$。

以下描述用示性点过程进行河流相储层建模的模型参数。这个过程基于以下的假设：

1）两种相：储层由两种沉积相组成，河道砂和背景相。

2）成组：河道聚集成河道带或河道组。

3）多重观测：当几个河道在同一口井中被观测到时，用户必须确定哪些河道属于同一组。此外，在不同井中观测到的河道属于同一组的概率也应该被确定。

4）主轴：每个河道都有一个主轴，而该组每个河道都有一个主轴，且平行于整个河道的主轴。

5）河道形式：相对于河道主轴的水平偏离和垂直偏离，以及河道的宽度和厚度都可用高斯过程进行建模。

6）参数：一个示性点过程是对于若干河道而定义的，在此，每个点都是代表一组河道。

在各种模型中，以下各种参数是需要的：

1）在油藏的范围内河道组的数目；各个河道组主轴方向的分布。

2）河道组主轴和各河道主轴之间的距离的分布。

3）确定从河道主轴的偏离的参数。

4）确定河道的厚度和宽度的分布的参数。

5）确定不同河道相对主轴的偏离，以及宽度和厚度的相关的相关系数。

6）确定不同河道组之间互相作用的参数。

7）砂泥比，即河道的体积占储层整个体积的比例。

8）有两类输入数据能够被确定：

第一，井中关于河道的观测值，且将在模拟中被重现；第二，井位——除在井中观测到的河道以外的河道都不穿过这些井。

3．算法

用来进行模拟的算法是 Ripley/Kelly 生灭过程和 Metropolis 算法（Metropolis，1953）的一个综合。在以下将给出这些算法的详细叙述。简单地说，所用的模拟算法可以描述如下：

1）抽取互相独立的河道组的数目；

2）进行 Ripley/Kelly 生灭过程，迭代的次数由用户确定。这时不同组之间是互相排斥，但不考虑砂泥比。一次迭代包括如下过程：第一，随机地抽一个组；第二，用一个潜在的新的组替代这个组，接受这个组的概率与给定所有的其他条件下，这个潜在的新组的概率成比例。如果这个可能的新组没有被接受，那么就抽取另外一个可能的新组，且继续这个过程，直到有一个可能的新组被接受。

3）利用 Metropolis 算法进行若干次迭代。利用各河道之间的排斥，如果砂泥比和用户

给定的砂泥比不一样，则要进行惩罚。随着迭代次数的增加，这个惩罚的概率是增加的。继续这种迭代，直到取得用户所指定的砂泥比，或直到迭代到一定的数目。这时，每次迭代包括如下内容：第一，随机地抽出一组河道：第二，利用一个新的组来代替这组河道，从而以确定一个可能的新的实现，如果这个可能的新的实现比旧的更合适，那么就接受这个新的实现。如果这个可能的新的实现的概率比旧的要小，那么接受这个新实现的概率就等于这个可能的新实现的概率和旧的实现的概率的比值。如果这个可能的新的实现没有被接受，那么就继续进行下一步迭代。

8.3.3　具体算法

这个模型包括有两种沉积相，第一种是：河道砂，及其位置和形式，另一种是背景相。在示性点过程中的每一个点表示一组河道。各个点的位置表示河道组的主轴，而示性值表示主轴为该方向的河道组中的各个河道的位置、宽度和厚度。更精确地说，一组河道 X_i 由以下的变量所确定：

1）y_0^i，z_0^i，各个河道组的主轴和 $x=x\min$ 平面的交点，这里 $x_{\min}$是模拟值的 x 坐标的最小值。

2）φ_h^i，φ_v^i 分别是河道组主轴和直线 $\{(x,\ y,\ z)=(t,\ y_0^i,\ z_0^i):\ t\in R\}$ 的夹角，和河道组主轴和直线 $\{(x,\ y,\ z)=(x_{\min}+t\cos\varphi_h^i,\ y_0^i+t\cdot\sin\varphi_h^i,\ z_0^i):\ t\in R\}$ 的夹角。

3）N_i 对应于河道组中河道的数目。

对第 i 个河道组，第 j 个河道由以下的函数所确定：

1）$d_h^{ij}(t)$：水平方向上，河道中心到河道组主轴的偏离。

2）$d_v^{ij}(t)$：垂直方向上，河道中心到河道组主轴的距离。

3）$T^{ij}(t)$：河道的厚度。

4）$W^{ij}(t)$：河道的宽度。这里，t 是从点 $(x_{\min},\ y_0^i,\ z_0^i)$ 沿着河道组主轴的长度。

以上的 y_0^i，z_0^i，φ_h^i，φ_v^i，N_i 将会在以下的讨论中提到。

河道组的全部 $X=(X_1,\ \cdots,\ X_N)$ 的联合概率函数是：

$$P(X)=C\cdot\exp\{\sum_{i=1}^{N}V_1(X_i)+\sum_{i=1}^{N}\sum_{j=i+1}^{N}V_2(X_i,X_j)\}$$

这里：C 是正则化常数；X_i 代表第 i 个河道组；N 是河道组的数目；$V_1(\cdot)$ 是确定一个河道组分布的函数；$V_2(\cdot,\ \cdot)$ 是确定河道组之间相互关系的函数。

函数 V_1 具有以下的形式：

$$\begin{aligned}V_1(X_i)=&\ln f_{\varphi h}(\varphi_h^i)+\ln f_{\varphi v}(\varphi_v^i)+\ln f_{y0}(y_0^i)\\&+\ln f_{z0}(z_0^i)+\ln f_N(n^i)+\ln f_{dh}(d_h^i)+\ln f_{dv}(d_v^i)\\&+\ln f_\tau(\tau^i)+\ln f_w(w^i)\end{aligned}\tag{8.4}$$

这里：$f_{\varphi n}(\cdot)$ 是 φ_h^i 的概率分布，是均值为 $\mu_{\varphi h}$，标准离差为 $\sigma_{\varphi n}$的高斯分布；$f_{\varphi h}(\cdot)$ 是 φ_v^i 的概率分布，是均值为 $\mu_{\varphi h}$，标准离差为 $\sigma_{\varphi v}$的高斯分布；$f_{y0}(\cdot)$ 是 y_0^i 的概率分布，是一个均匀分布；$f_{z0}(\cdot)$ 是 z_0^i 的概率分布，是一个均匀分布；

(8.4) 式中，$f_N(\cdot)$ 是一个河道组的河道数目的概率分布。河道数的最小值 $N_{\min}$加上

一个双正态分布变量和 $n = N_{max} - N_{min}$次试验，概率为 $P = (\mu_N - N_{min}) / (N_{max} - N_{min})$。这里，$N_{max}$是河道数目的最大值，而 μ_N 是河道数的平均值。

还是在 (8.4) 式中，$f_{dh}(\cdot)$ 是 $d_h^{i1}(t)$，$d_h^{i2}(t)$，…，$d_h^{iN}(t)$ 的联合概率分布。它们被视为相关的高斯过程；其均值和协方差函数为

$$E[d_h^{ij}(t)] = \delta_h^{ij}$$

$$\mathrm{Cov}[d_h^{ij}(t), d_h^{ik}(s)] = \sigma_{dh}^2 \cdot \exp\{\alpha_h \cdot |t - s|\} \cdot \rho_h^{jk} \tag{8.5}$$

这里：

δ_h^{i1}，δ_h^{i2}，…，δ_h^{iN} 是 N 个独立的高斯分布的变量，其均值为零，而标准离差为 $\sigma_{\delta h}$。

$$\delta_\tau^{i1} = \delta_\tau^{i2} = \cdots = \delta_\tau^{iN} = \mu_\tau$$

$$\delta_w^{i1} = \delta_w^{i2} = \cdots = \delta_w^{iN} = \mu_w$$

(8.5) 式中，δ_{dh}是水平方向上河道中心到河道组主轴的偏离的标准离差；α_h 是 d_h 的变异函数的参数；ρ_h^{jk} 是 1，当 $j = k$ 时，否则是 ρ_h，这里 ρ_h 是在同一个河道组内不同河道的相关系数；$f_{d_v}(\cdot)$ 是 $d_v^{i1}(t)$，…，$d_v^{iN}(t)$ 的联合概率分布。按照 $d_h^{i1}(t)$，$d_h^{i2}(t)$，…，$d_{h(t)}^{iN}$，对它们进行建模；$f_\tau(\cdot)$ 和 $f_w(\cdot)$ 分别是 $\tau_v^{i1}(t)$，$\tau_v^{i2}(t)$，…，$\tau_v^{iN}(t)$ 和 $\omega_v^{i1}(t)$，$\omega_v^{i2}(t)$，…，$\omega_v^{iN}(t)$ 的；联合概率分布。他们的建模是依据 $d_h^{i1}(t)$，$d_h^{i2}(t)$，…，$d_h^{iN}(t)$，但是除去 δ_τ^{i1}，δ_τ^{i2}，…，δ_τ^{iN} 和 δ_w^{i1}，…，δ_w^{iN}。

它们不是高斯分布，但是有：这里 μ_τ 和 μ_w 分别是一个河道的厚度的均值和宽度的均值，由用户确定。

函数 V_2 具有如下的形式

$$V_2(X_i, X_j) = \int_{S_1}^{S_2} \Lambda(S)\mathrm{d}S / (S_2 - S_1)$$

$$\Lambda(S) = \begin{cases} -\infty & \text{如果 } \mathrm{d}(s) \leqslant \Delta \\ \ln \dfrac{\mathrm{d}^2(s) - \Delta^2}{\rho^2 - \Delta^2}, & \text{如果 } \Delta < \mathrm{d}(s) < \rho \\ 0, & \text{如果 } \mathrm{d}(s) \geqslant \rho \end{cases}$$

$$\mathrm{d}(s) = \| (X_i^0 + r_i \cdot s) - (X_j^0 + r_j \cdot s) \|$$

$$\Delta = \sqrt{\Delta x^2 + \Delta y^2 + \left(\frac{\Delta z}{\alpha_z}\right)^2}$$

其中，X_i^0 和 X_j^0 分别是河道组 i 和河道组 j 的主轴上的点，且使得这两点间的模最小。r_i 和 r_j 分别是河道组 i 和河道组 j 的主轴的单位向量；(S_1, S_2) 是一个区间，使得 $(X_i^0 + r_i \cdot s)$ 和 $(X_j^0 + r_j \cdot s)$ 落在模拟值区间 (S_1, S_2) 内部。

从这个模型所用的模拟算法首先是使用修正的 Ripley/Kelly 的生灭过程，以获取各河道组之间的正确相互作用，然后是使用修正的 Metropolis 算法和模拟退火算法，以获取要求的砂泥比。关于这两种算法进一步的叙述可分别参见 Stoyan 等人（1987）和 Ripley（1987）的文章。所用的模拟算法可叙述如下：

1）初始化，抽取一个数，它表示相互独立的河道组的数目。

2）实施 Ripley/Kelly 生灭过程的 $n_{R/K}$叠代过程，其中每一次叠代由以下三步组成：第一步，随机地抽取一个河道组，并移走这个河道组；第二步，抽取一个潜在的河道组；第三步，以下面的概率 P 来接受这个潜在的河道组。

$$P = \exp\{V_1(X_i) + \sum_{j\neq i} V_2(X_i, X_j)\}$$

这里 X_i 是这个潜在的新的河道组。如果这个潜在的新的河道组没有被接受，那么控制就转向上述的步骤 2）。

3）实施 Metropolis 算法叠代过程。在这个叠代过程中，用到一个稍加修改的概率分布：

$$P_{T(t)} = C \cdot \exp\{\sum_{i=1}^{N} V_1(X_i) + \sum_{i=1}^{N}\sum_{j=i+1}^{N} V_2(X_i, X_j) - \frac{(ng - ng^*)^2}{T(t)}\}$$

这里，ng 是砂泥比，即由河道所占的体积和整个油藏的体积之比；ng^* 是用户确定的砂泥比；$T(t) = T_0 \cdot a^t$ 是叠代数 t 的温度，T_0 和 $a<1$ 是由用户确定的参数。

利用 Metropolis 算法进行叠代，一直到砂泥比落入目标砂泥比 ng^* 为中心的一个范围，这个范围的宽度是由用户指定的，或者一直到叠代数目达到预先指定的数目。叠代数 t 由以下组成：

1）随机地抽取一个河道组。

2）定义一个潜在的新的实现替代这个河道组。接受这个河道组的概率 P 定义如下：

$$P = \min(1, \frac{P_{T(t)}(X_{new})}{P_{T(t)}(X_{old})})$$

8.4　先验模型和后验模型

储层随机建模方法，甚至地质统计学中的基本推断都是基于后验模型的。这种后验模型是通过先验模型加上数据得来的。

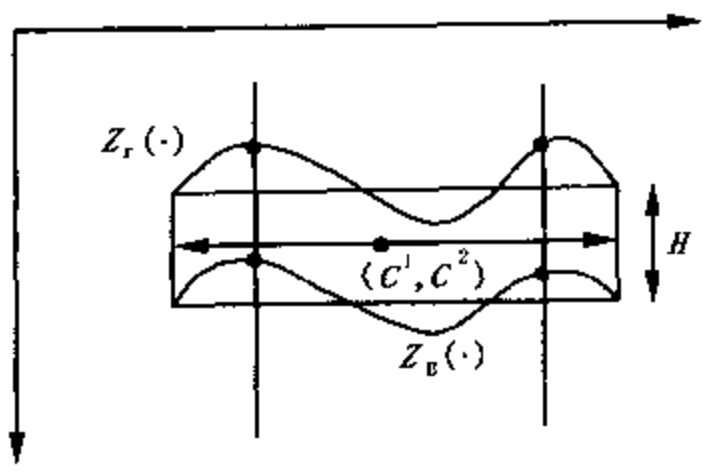


图 8.6　河道及其几何参数
(Syversveen and Omre, 1996)

以下以井中数据为条件，讨论利用示性点过程对沉积相带空间分布进行建模的一种模型(Syversveen and Omre, 1996)。该模型的特点是包含了沉积相带间的空间相互作用，然而，并不要求井和井之间的联系。

在图 8.6 中，每条河道的截面可用中心在 $C=(C^1, C^2)$，长和厚为 $S=(L, H)$ 的一个长方体表示。其顶底由一对随机函数 $Z(\cdot) = [Z_B(\cdot), Z_T(\cdot)]$ 表示。对此可写成

$$U = [C, S, Z(\cdot)]$$

此时，储层含有的河道条数 N 为未知。河道互相之间是排斥作用。

8.4.1　先验模型

令 $\theta_i = (C_i^1, C_i^2, l_i, h_i)$ 表示第 i 条河道的长方形，$z(\cdot)$ 表示河道 i 顶面和底面

位置。n 条河道的联合概率密度函数可写成如下：

$$f[\theta, n, z(\cdot)] = f(\theta, n) f[z(\cdot) \mid \theta, n] \tag{8.6}$$

这里 $\theta = (\theta_1, \cdots, \theta_n)$，以及 $z(\cdot) = [z_1(\cdot), \cdots, z_n(\cdot)]$。其中第一项 $f(\theta, n)$ 是一个示性点过程，其对象互相作用具有如下形式（Stoyan 和 Stoyan，1994）：

$$f(\theta, n) = \text{const} * \exp\{\sum_{i=1}^{n}[b(C_i^1, l_i) + b(C_i^2, h_i)] + \sum_{j=1}^{n}\sum_{i<j} d(C_i, C_j) + na\}$$

这里 $b(C_i^1, l_i) + b(C_i^2, h_i)$ 是河道的边缘特征。由于假定了 C^1，C^2，L 和 H 是相互独立的，于是就有

$$\exp[b(C^1, l)] = \exp[b(C^1) + b(l)]$$

和

$$\exp[b(C^2, h)] = \exp[b(C^2) + b(h)]$$

$d(C_i, C_j)$ 是在河道中心点之间的点对之间相互作用，其最后一项和河道的稠密强度有关。进一步，$b(C_i^2, h_i)$ 被定义为截断的双高斯概率密度函数，其目的是为了解析地计算先验密度函数。这个截断为了使 C^2 仅仅在储层范围内取值。H 是正的。C^2 的均值选取为储层的 x_2 坐标的中心值，其方差选取为较大，以便在该储层上近似于均匀分布。对 H 的均值和方差可以从观测数据和经验贝叶斯方法来确定（Syversveen 和 Omre，1997）。$\exp[b(C_i^1, l_i)]$ 可以地质经验来确定。交互作用函数 $d(C_i, C_j)$ 可以通常由河道之间的排斥情况来确定。密集强度参数 a 可以从数据用蒙特卡洛方法来估计（Geyer 和 Thompson，1992；Syversveen，1996）。

方程（8.6）式的右端的第二项 $f[z(\cdot) \mid \theta, n]$ 是和河道 i 的顶界和底界的位置有关。函数 $Z_i(\cdot)$ 被 $\Theta_i = \theta_i$ 条件化，是二阶平稳的高斯随机函数，其均值和协方差如下：

$$E[Z_i(x_i) \mid \Theta_i = \theta_i] = \begin{Bmatrix} C_i^2 + h_i/2 \\ C_i^2 - h_i/2 \end{Bmatrix}$$

和 $\text{Cov}(Z_i(x'_1), Z_i(x''_1) \mid \Theta_i = \theta_i) = \begin{bmatrix} \sigma^2 R(x'_1, x''_1) & \tau R(x'_1, x''_1) \\ \tau R(x'_1, x''_1) & \sigma^2 R(x'_1, x''_1) \end{bmatrix}$。这里 $x_1 \in [C_i^1 \pm l_i/2]$。矩阵 $R(x'_1, x''_1)$ 由下式给出：$R(x'_1, x''_1) = \begin{pmatrix} 1 & \rho(x'_1 - x''_1) \\ \rho(x'_1 - x''_1) & 1 \end{pmatrix}$

这里 σ^2 是方差，τ 是顶底之间的协方差，$\rho(\cdot)$ 是空间相关函数。为了保证协方差函数的正定性，必须要有 $|\tau| < \sigma^2$。注意到顶底面相对 C^2 是对称的。根据地质经验，来选取相关函数，方差和协方差函数。假定随机场 $Z_i(\cdot) \mid \Theta_i = \theta_i$ 在不同河道内是条件独立的，可以得到 N 个河道的储层的概率密度函数

$$f[\theta, n, z(\cdot)] = f(\theta, n) \prod_{i=1}^{n} f[z_i(\cdot) \mid \theta_i] \tag{8.7}$$

关于 $\exp[b(C^1)]$，$\exp[b(l)]$，交互作用函数 $d(C_i, C_j)$，空间相关函数 $\rho(\cdot)$，方差 σ^2，顶底之间的协方差 τ，以及 a 的确定，参见（Syversveen and Omre，1996）。

8.4.2　观测值和似然函数

观测值由形为长方体的储层的尺寸，以及其中的 m 口井的观测值所组成。河道的顶，底的深度的观测值是向量 o。各井的位置用 w_i，$i=1$，…，m 表示。假定在这些井中观测值没有显示河道的重叠，也不知道这些观测值是否属于相同的河道。所以，该似然函数可写成一个指示函数，当条件化被满足时，值为 1，否则值为 0。

通过这种形式的似然函数，要解析地计算后验分布函数是不可能的。其原因是井和井之间的联系是不知道的。因此，将引入一个新的随机变量 T，它是由 Θ 隐含地定义的。它告诉人们哪些井中的观测数据是属于同一河道。T 被称为河道的结构，它们具有不同的结构。例如，这三个河道的结构属于真实地层中的同一对象的三个观测值，可以作为后验的实现中的一个，二个或三个对象。引入了 T 后，似然函数可写成如下：

$$f[o\mid\theta,n,z(\cdot)]=\text{const}*\prod_{i-1}^{n}I(o_i\mid\theta_i,z_i(\cdot),t) \tag{8.8}$$

这里 $I\ [o_i|\theta_i,\ z_i\ (\cdot),\ t]$ 等于 1 时，条件化在河道 i 上获得满足。否则，它等于 0。t 出现在（8.8）式的右端的理由，是关于一个河道的知识对确定结构 t 是不充分。

后验模型由表达式（8.6）的先验模型和（8.8）式的似然函数所确定。然而，人们希望分别模拟（θ，n）和 z。因此，将以表达式（8.9）的形式来定义后验概率密度函数 pdf，来分析对（Θ，N）的似然函数：

$$\begin{aligned}f(o|\theta,n)&=\text{const}*\prod_{i=1}^{n}f(o_i|\theta_i,t)\\&=\text{const}*\prod_{i=1}^{v_t}f(o_i|C_i^2,h_i,t)\prod_{j=1}^{n}I(o_j|C_j^1,l_j,t)\end{aligned}$$

这里 $f\ (o_i|C_i^2,\ h_i,\ t)$ 是在结构 t 之下，在穿过河道 i 的井中的高度的观测值的似然函数，而 $I\ (o_j|C_j^1,\ l_j,\ t)$ 则是保证侧向条件化被满足的一个指示函数。数 v_t 是在结构 t 内所穿过的河道个数。因为观测值是没有误差的，$f\ (o_i|C_i^2,\ h_i,\ t)$ 是直接地和 $f\ (z_i\ (\cdot)\ |C_i^2,\ h_i)$ 有关，因此是一个多元高斯概率密度函数。这第二个等式的成立是基于这样的事实：在先验分布函数中每个河道的水平尺寸和垂直尺寸是相互独立的。

为了能计算后验模型的解析式，似然函数能分解成条件互相独立的部分是很重要的。这将在以下可以看到。

8.4.3　后验模型

和表达式（8.7）相似，对［Θ，N，$Z\ (\cdot)$］的后验模型能被写成：

$$f[\theta,n,z(\cdot)\mid o]=f(\theta,n|o)\prod_{i-1}^{n}f[z_i(\cdot)\mid\theta_i,o] \tag{8.9}$$

对其右端的第一项，经推导可得：

$$\begin{aligned}f(\theta,n\mid o)&=\text{const}*f(\theta,n)f(o|\theta,n)\\&=\text{const}*\exp\{\sum_{i=1}^{n}[b(c_i^1,l_i|o,t)+b(c_i^2,h_i|o,t)]\\&+\sum_{j=1}^{n}\sum_{i<j}d(c_i,c_j)+na\}I(o|t,n)\end{aligned} \tag{8.10}$$

其中，exp $[b\ (c_i^2,\ h_i|o,\ t)]$ 和一个双高斯概率密度函数只差一个常数（Syversveen and Omre，1997）。exp $[b\ (c_i^1,\ l_i|o,\ t)]$ 等于 exp $[b\ (c_i^1,\ l_i)]$ $I\ (o|c_i^1,\ l_i,\ t)$。

第 i 个河道的几何尺寸的条件概率密度函数 *pdf*——$f\ (z_i(\cdot)|\theta_i,\ o)$ 由双高斯随机函数组成。这些函数具有给定的参数以及被一些位置处的观测值所条件化（Journel and Huijbregts，1978）。

8.4.4　模型的模拟

所做的模拟可以分为两步。首先，参数 Θ 和河道的数目 n 可以从表达式（8.10）定义的密度函数中用 Metropolis－Hastings 算法来抽取（Metropolis 等人，1953；Hastings，1970）。然后，条件高斯随机场 $z_i(\cdot)|\theta_i,\ o$ 则利用条件模拟来采样而求取（Journel 和 Huijbregts，1978）。

Metropolis 和 Hastings 算法的转移内核可由三种不同的转移方式进行：

1）增加一个没有观测值的河道。

2）删去一个任意的没有观测值的河道。

3）交换有观测值经过的河道，包括交换一个结构函数。

当增加一个新的河道，推荐新的河道时，应从对应的各向同性的示性点过程的边缘先验分布函数中抽取。当交换有观测值经过的河道时，一个结构从可能的结构中均匀地抽取。给定了这个结构，由在结构方面进行交换所影响的每一个河道，是从用数据来进行条件化的一个边缘分布函数中来抽取的（Syversveen 和 Omre，1997）。从这个条件边缘分布函数中抽取河道，对于这个算法的收敛是至关重要的。

Metropolis－Hasting 算法中的接受概率时由从中进行抽取的后验概率密度函数和这些接受核来确定，并且它们的计算是根据 Syversveen and Omre 的论文（1997）。因为这个模型定义在维数相对低的空间中，算法的收敛并不构成严重的问题。

8.5　应用实例

从 80 年代中期以来，在河流相河道结构的模拟中广泛地使用了许多不同类型的基于目标的随机模型（Clementsen and others，1990；Gundeso and other，1990；Hirst and others，1993；Deutsch and other，1996）。这些模型主要应用在处于开发早期阶段油井数据很少的油田中。它们的应用帮助人们提高了对在复杂河流相储层中流体流动过程的理解，它们的实现已经用来评价储层的连通性、设计井网、预测总产量（Stanley and others，1990；Nybraten and others，1990；Tyley and others，1995）。

在北海，许多较大的河流相油藏现已到了开发中后期，油藏管理的最重要的任务就是发现和开发死油带。以前的河道模型就不能够应用在油井数据很丰富的这种油田中，需要使用复杂的多井条件。

从奥斯陆的挪威计算中心发展起来的河流相模型已经过了不断的修改，以适应挪威大陆架油田的晚期开发。人们已采取了几条措施成功地获得了多井条件，这些措施包括：

1）使用更具弯曲性的河道几何体模型。

2）使用河道间相互关系的先验信息。

3）使用油井测试数据。

这里，采用了一个全新的随机建模算法。

本节描述了这种算法的原理和应用过程。测试这种算法的重点是评价弯曲的、逼真的河道几何体在获取多井条件中的作用。这种测试是以从中国东部地区沿海油田——孤东油田的密井网所得到的复杂数据组为基础的。在这以前，基于目标的随机模型还从未被应用到有如此小而规则的井间间隔的数据组中。

8.5.1　GD 数据组

GD 油田位于中国东部地区，它主要的产油层段来自于馆陶组的河道砂体。这个油田可扩展到大约 10km^2 的区域。在这个以注水来提高压力的油田中已钻了 230 多口井。井中出水量已十分之高，因此，油藏管理的一个挑战性任务就是确定死油带位置。

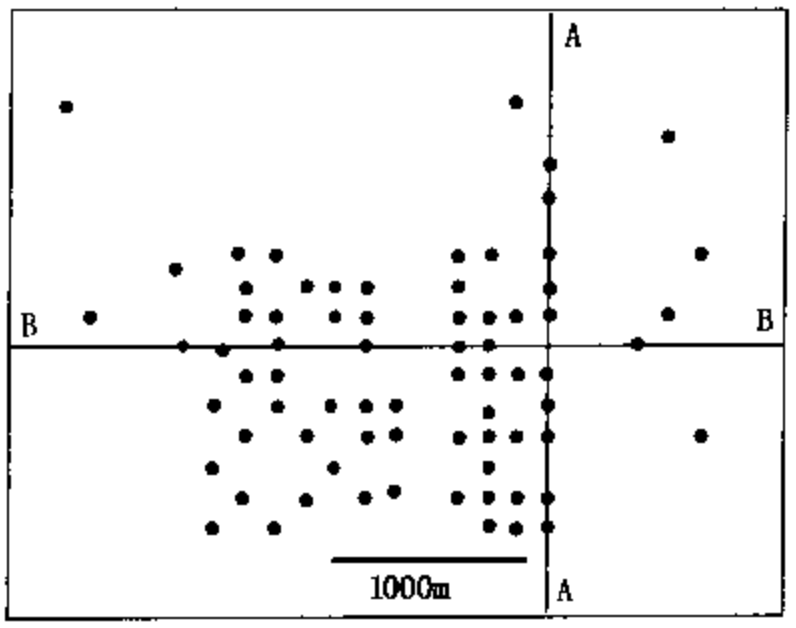


图 8.7　井位图

在此以前，已经用指示模拟方法对 GD 油田的储层非均质性进行了建模（Wang and Zhang，1995）。

大量的小间距的油井提供了极好的数据。这些数据被用来测试为描述已到开发后期油田的非均质性而建立的基于目标模型的适用性，并建立了一个厚度为 43m 的储层层段模型。这个层段的有效厚度与总厚度比值是 0.29。由于本储层层段的局部钻开油层和一些数据的质量问题，仅使用了 230 口井中的 75 口井来进行测试（图 8.7）。这些井是以规则井网排列的，且相互间隔 150m。

通过该储层的横截面（图 8.8）可看到河道砂体的复杂几何形态。在相邻井间，一些砂岩很明显地是连续的，另一些有可能是连续的，还有一些很明显是不连续的。用必要的一维油井数据产生三维模型不是一个小问题。

8.5.2　随机模型和随机算法

随机模型描述了在漫滩相背景下河道储层体的分布。假设储层体是河道沉积的产物，典型的河流相储层就是由几百个河道储层体组成的。

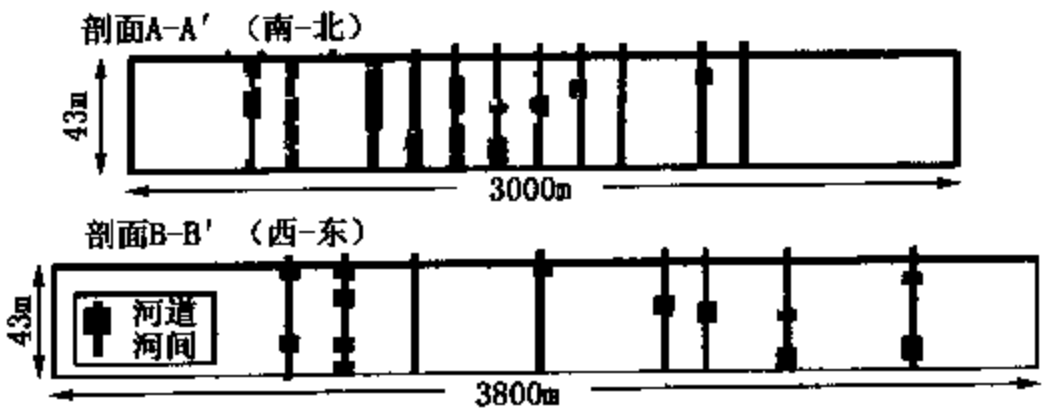


图 8.8　剖面图

早期的模型方案是在 80 年代中期发展起来的。在这些方案中，所要建模的河道就象在不渗透基岩中的简单平行六面体，并在早期阶段引入了弯曲性。

这些十分简单化的、“不具弯曲性”的模型不能适应大量油井，并且当许多大的北海河流相储层中的数据变得更为丰富的时候，就有必要修改最初的模型，以产生一个能够描述逼真几何体的更为“弯曲”模型。这一模型也就扩展到考虑了井间河道的先验相互关系方面。

在现有的模型中，每个砂体是用位置和方向来定义的。宽度、厚度和弯曲起伏都可沿着每个砂体的轴变化。厚度和垂直位置也会对应于每个砂体的轴线变化。这种随机模型具有很大的弯曲性并能描述河道状几何体的各种变化。

迭代模拟退火算法可用来产生模型的实现。每次迭代都涉及单个河道体的增添、删除或者交换。退火方法可以用来确保精确的油井条件，并且可以获得正确的不同相的体积比。

1．工作流程

随机建模过程的基本工作流程如图 8.9 所示。第一步涉及大规模的储层带的井间相关、构造图的产生，以及井中各相的解释。

储层层段的顶与底可以解释成馆陶组内部的等时面。这一储层带的厚度从 30m 到 55m，平均 43m。构造图是通过对油井数据进行克里金处理而产生的。在模拟所需要的可逆变换中要用到这些图（Deutsch and other，1996；MacDonald and others，1992）。首先把建模区域变换为恒定厚度为 43m 的盒子以用于模拟，在对相结构模拟之后，再把它变回到原来的真实坐标。

这里使用了两种相来分类：河道相和漫滩相。如果还有其他的相也提供了有意义的非均质性，比如裂缝和不同的条形，也可以同时对这些相建模，以适应于河道的分布（Egeland and others，1993）。

2．参数估计

在准备工作完成之后，就可以开始参数估计（图 8.9）。这要涉及到油井数据的统计分析、几何参数的估计以及退火参数的调整。

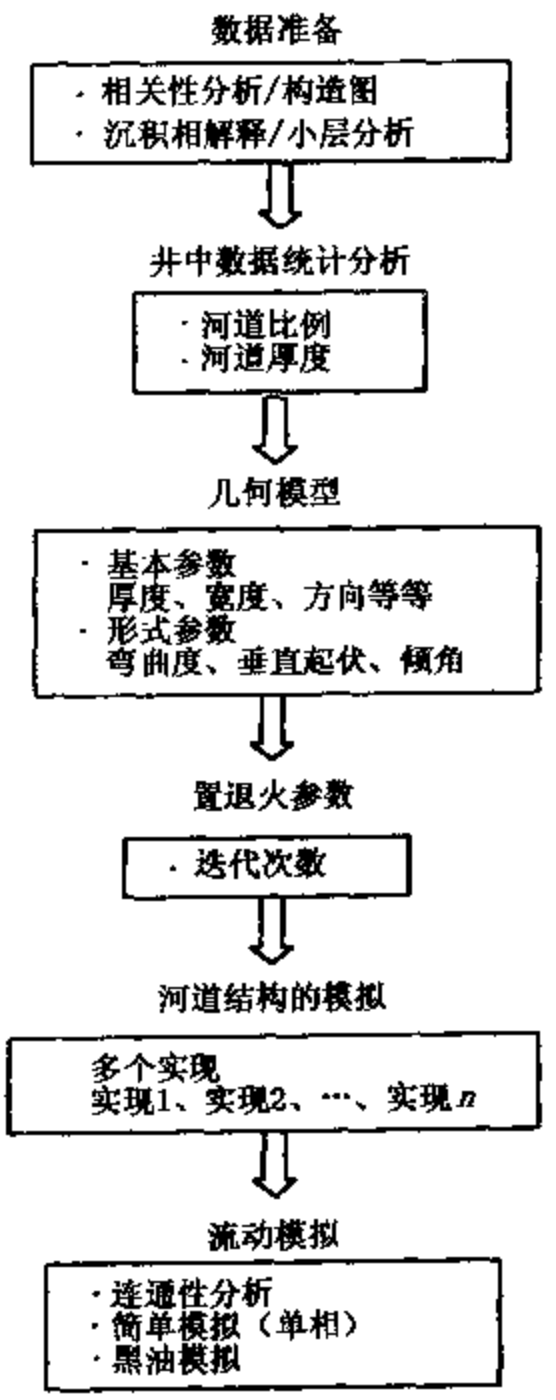


图 8.9　随机建模流程图

(1) 油井数据的统计分析

利用油井数据可以估计的主要参数有河道厚度和相的比例。75 口井中观察到的 144 个

河道的直方图可以利用油井数据作出（图 8.10）。单河道砂体的厚度从 1m 到 16m，平均 6.31m，标准偏差 2.96m（图 8.11）。

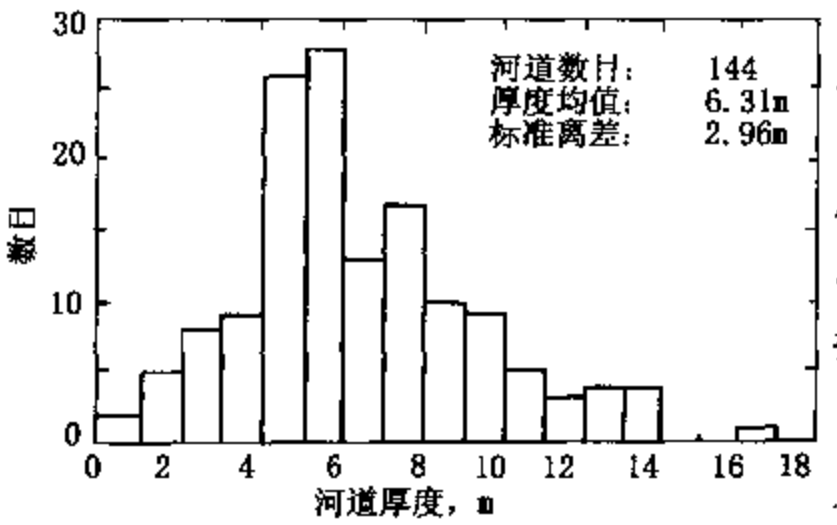


图 8.10　75 口井中河道厚度的直方图

在各井处，各河道总厚度占储层厚度的比例是 0.29，河道部分的垂直走向是很明显的（图 8.11），43m 厚的模拟层段包括两个河道比例较大的层段（>0.45）和两个富泥岩的层段（河道比例<0.2）。这些变化定义了 3 个在模拟层段内的地层亚组（图 8.10）。

（2）几何参数

除了河道比例之外，还必须对大量的其它几何参数进行估计。大部分这种参数，如河道宽度和走向，不能只用油井参数来估计，即使有很多井也不行。它们的估计必须使用古地球物理信息和从露头或地下岩石类比所得到的数据。

当随机模型变得越来越逼真和复杂的时候，需要估计的、用来描述模型的参数就会显著增加。这些几何参数已被分成两组主要类型，从而使参数估计过程简单化。对于在河流相储层中的连通性和流动性最为主要的参数可划分为“基本参数”，在模型中为复杂的多井条件提供弯曲性的、已被合并的参数可划分为“形式参数”。

1）基本参数。基本参数与河道比例、走向、宽度、厚度以及起伏相关。它们如下所列：河道比例，平均河道厚度，在不同河道间的河道厚度变化，主河道走向，河道走向的变化，平均河道宽度，在不同河道间的河道宽度变化，河道起伏的幅度，河道起伏的波长。

在此，地质学家一个重要的任务就是，估计以所有油井数据、古地球物理信息、露头和现存储层类比为基础的参数值。

2）有关河道的形式参数。河道形式参数与河道体的外部形态相关（图 8.11）。油藏工程师们常把它们的作用误解为可有可无。这些由油井条件决定参数的主要功能就是允许产生河道的弯曲。这些形式参数有如下四种：在单河道内的宽度变化，在单河道内的厚度变化，垂直起伏，和倾角。

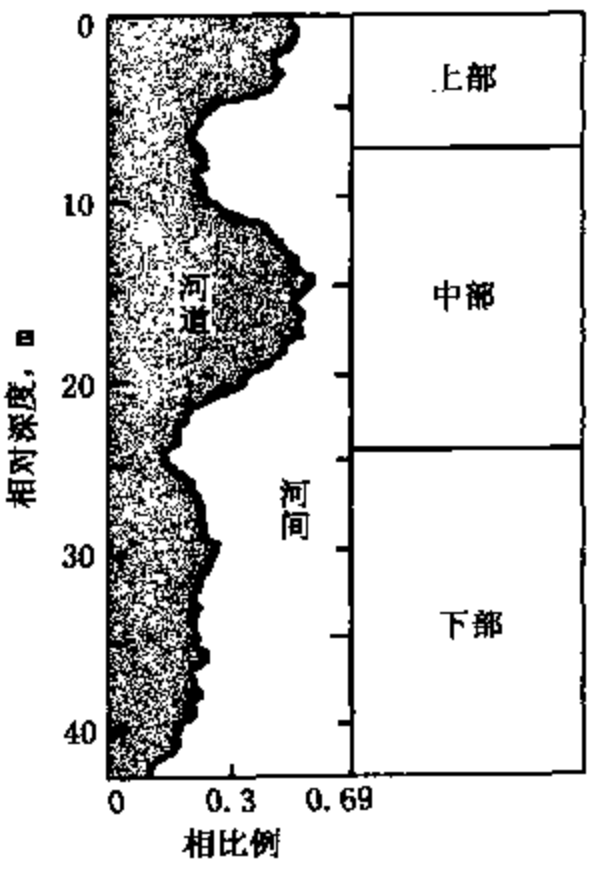


图 8.11　河道总厚度占储层厚度的比例随深度变化

宽度和厚度能够沿着单河道砂体及其法线变化。这些参数能够用来捕获自然的，并且是逼真的河道几何形态中的局部变化（彩图 21）。

严格地从沉积学的观点来看，垂直起伏和河道倾角是没有意义的参数。然而，为了允许大规模压实效应的和局部差异压实效应的，及油井相关的错误，这些参数中有一些灵活性是必要的，也是合乎需要的（彩图 21）。

3）退火参数。在分配了几何参数之后，就应当调节几个“退火参数”，以确保从模型中产生满意的实现（图 8.7）。可以用这些退火参数来定义迭代模拟算法中的退火方案。其中，最重要的参数就是迭代次数。

当给定迭代次数时，有两个目标会相互冲突：第一，迭代次数应当足够大，以确保模拟能达到大量的不同“接受标准”；第二，迭代次数应当尽可能的小，以节省模拟时间。最重要的接受标准是：第一，所有的油井观察值应被模拟实现精确地通过；第二，模拟实现达到河道比例的目标。

用来产生模型实现、有精确油井条件和所期望的河道比例的迭代次数紧紧地依赖于要被估计的平均河道厚度。总之，较宽的河道需要较多次的迭代。在平均河道宽度小于 100m 的地方，一次成功的模拟实现可用少于 5000 次迭代。然而，为了一个平均厚度为 150m 的河道的实现，有必要使用 15000 次迭代来确保正确的油井条件和相的比例。

8.5.3　模拟的河道结构分析

因为要用油井数据条件下的随机模型来模拟孤东储层，几个输入参数已经进行了系统意义上的改变，以评价河道结构的排列。在下面，我们要更为详细地讨论形式参数及河道厚度所带来的影响。

1. 形式参数

当估计平均河道宽度是 150m 时，河道的形式参数会被系统意义上的改变。除了平均河道宽度之外，有必要给定不同井间的河道宽度的变异性（比如，所有的河道有相同的尺寸，或者是储层既有宽河道又有窄河道）。此变异性是个标准偏差，它被定为 0.5 乘以在所有模型中的平均河道宽度。

当形式参数设置得太小以描述“无弯曲性的”河道地貌时（图 8.11），产生一个正确的条件模拟实现就不可能。无弯曲性的河道与油井数据产生了急剧的冲突。

当引入有限的灵活性的时候，就有可能正确地产生一个模拟的实现，这需要一个很大的迭代次数（>50000）。由大量的河道（120～130 条）和极窄的平均河道宽度（大约 90m）可以描述模型的实现。大量的河道表明，在相邻井间，所观察到的 144 个河道中有少数几个是相互关联的。因为在一个储层中放置一个窄的河道比放置一个宽的河道更为容易，所以就会出现对用户指定的河道宽度的修改。在迭代模拟过程中，较宽的无弯曲性的河道经常会与油井数据发生冲突，并且会比窄河道更为频繁地遭到拒绝。

当引入形式参数中有意义的灵活性时，就有可能产生一个河道数少于 70、宽度大于 130m 的实现。随着灵活性的增加，即使它们出现在稍有不同的等级上，在相邻的观察点间也会有更多的河道相关（图 8.12）。垂直起伏是由差异压实效应和影响坐标变换的次要相关错误引起的。

2. 平均河道宽度

虽然井间间隔十分小，但是要以确定性的方式描绘出单个河道砂体的形态仍是很困难的。在 GD 数据组中（图 8.7），有与河道尺寸及河道结构相关的大量的不确定性。因此，被估计的平均河道宽度（“先验分布”）在系统的意义上从 100m 到 300m 被改变了（表 8.1）。在模拟中，就会使用形式参数中有意义的灵活性。



图 8.12　相邻的观察点间会有更多的河道相关

虽然单个的被估计的河道可以宽于 700m，但是要用大于 170m（见表 8.1）的平均宽度（“后验分布”）来产生一次实现也是很困难的。这与油田中密井网区域的最小井间距是相对应的。当河道宽度增加时，要正确地产生一个条件模拟实现所需要的迭代次数也急剧地增加了（表 8.1）。

表 8.1　宽度分布的更新

宽度平均值（先验）	100M	125M	150M	200M	300M
宽度平均值（后验）	96M	108M	127M	152M	167M
迭代次数	5000	10000	15000	30000	250000

根据油井数据对输入的平均河道宽度的估计值所作的修改表明，GD 储层的平均河道宽度是不可能超过 150～200m 的。这种现象部分地是由建模的人为因素造成的。这种人为因素和如下的事实有关：宽河道更经常地与油井数据产生冲突，且比窄河道更容易被拒绝。

3. 典型的模型实现分析

一个典型的模型实现包含 55 到 70 个河道储层体。彩图 22 与彩图 23 展示了两个河道砂体的实现。图 8.13 表示了上、中、下三个亚组的河道结构平面图。这个实现是用河道宽度的大范围变化及河道走向的大幅变化（图 8.12）来表征的。这种大幅度的变化在地质上是逼真的，并且与详细的类比露头研究互相一致（Nass and others，1993）。

两个实现的平均河道宽度大约都是 140m。这两个实现都是用 30000 次迭代而产生的，在运算能力极强的工作站上需要运算大约 20min。

通过两个实现的横截面，展现了河道砂体侧向堆积的有效程度和局部地更连续的储层的产生。在中部和上部亚组中，侧向堆积发育得特别好（彩图 23），储层的连续性在南北方向比东西方向稍好一些。

在两个随机实现之间的实质性的差别应该给予重视（彩图 22 与彩图 23）。尽管大量的井之间的间距很近，仍然有考虑河道结构不确定性的必要。

4. 井之间的相关

100 次随机建模的结果可用来评价所观察到的相邻河道有多少个是相关的/相连的。其平均结果是，所观察到的 144 个河道几乎有 2/3 在井间是相关的。例如，这些河道可在两个或更多的井中观察到。一些河道可在十多口井间是相关的。

沿着河道砂体的长度方向相关是特别普遍的。它表明这样相关的模型中允许有足够的弯曲性。河道轴法线方向的相关不是很普遍。这可能反映出 GD 油田河道的狭窄，但也可能与在河道轴法线方向上比沿着河道轴方向更不具弯曲性的模型有关。在河道横截面几何体中增加更多的弯曲性可以允许模拟更宽的河道（Hektoen and other，1997；MacDonald and others，1997）。

8.5.4　结论

根据以上的分析，可得到如下的结论：

1）在拥有丰富密井网的油田开发后期可以使用基于目标的随机模型；

2）任何模型都是现实的简化。然而，过于简单的、无弯曲性的河道几何体不能符合有大量油井的条件；

3）现有的从挪威计算中心发展起来的河流模型，可用来模拟在密井网储层平均河道宽度与井间距相近的 GD 油田的河道结构；

4）在孤东数据组河道结构的多井条件下，河道形式参数（皱曲、垂直起伏以及河道倾角）中的弯曲性是必要的；

5）河道模型有足够的灵活性以考虑如此的多井条件，那里的油井近似平行于砂体走向的。如果要估计更宽的河道，就需要更多的河道轴的法线方向的灵活性。

第 9 章　渗透率粗化

在油藏工程中使用地质统计学中的克里金估计和随机建模等油藏描述技术可以描述各井间的油藏物性参数（即孔隙度、渗透率、饱和度等），从而产生了这些物性的地质模型。随着油藏描述技术向着更为精细的方向的发展，可以得到描述油藏参数空间分布的三维定量精细地质模型，用以描述储层非均质性。三维定量精细地质模型一般包括数百万到数千万个三维网格体数据。这一组数据也会被用到油藏动态数值模拟研究中，作为油藏数值模拟系统的输入数据。但由于计算机能力的限制（Galli and others，1996；Almeida，and others，1996；Baker and other，1996；Pickup and other，1994；Gomez－Hernndez and other，1994；Aasum and others，1993；Lee and others，1995），在计算机上运行流体动态数值模拟系统是极其复杂且消耗 CPU 时间的，直接把数据量太大的精细地质模型输入到数值模拟系统中，会遇到各种困难。目前，人们正在研究利用平行算法解决具有几百万甚至几千万个节点的超大规模的油藏数值模拟的问题。然而，对精细地质模型进行粗化处理，减小其数据量而尽量保持地质信息也是一种解决问题的有效途径。这种要求意味着在绝大多数情况下必须加大模拟网格块的尺度，从而减少网格块的数目，使一组相邻的块只有一个值。对其它油藏参数，如孔隙度和含油饱和度等，可以用简单的数学平均方法进行处理。而由于渗透率是一个不可简单叠加的量，就必须对它进行特殊处理，这就产生了渗透率粗化技术（Galli and others，1996；Dimitrakopoulos and other，1993）。

人们不禁要问，既然对渗透率要进行粗化，为何一开始不把网格设得粗一些。问题是，这样就会忽略一些层薄的，水平方向分布范围较小的一些非渗透层或渗透率特高的层，给建模所得的整个渗透率的分布造成很大的畸变。这是目前油藏描述技术越来越精细的一个主要原因。因此，建模所用的网格的大小一定要根据具体的地质情况和所要解决的问题来决定，不能过大。

在进行油藏数值模拟时，一般要用到的孔隙度、含油饱和度等物性参数的性质比较简单。对它们的粗化可以用一般的平均算法来完成。所以，在本书中对它们的粗化问题不进行专门的讨论。渗透率是一个相对复杂的参数，它的粗化不能简单地用平均方法来解决。因此，在本章中进行专门的讨论。

9.1　引　　言

9.1.1　意义

把许多小尺寸的网格块合并成一个大尺寸的网格块。该大尺寸网格块上的各物性参数是通过各小网格块的值经过一定计算而得到的，这种技术就称为网格粗化技术。渗透率在油藏工程中具有十分重要的意义。它是描述流体流动的最主要的参数。相对渗透率能够反映油层的束缚水饱和度，残余油饱和度，以及在开发过程中随着注入水饱和度的增加，油井含水的增长情况。因此，渗透率在石油工程中具有特殊的意义，在油藏数值模拟中也是核心参数。因此，对渗透率的粗化方法特别受到重视和专门的研究。

渗透率粗化方法已经有数十种之多。但不管什么方法，渗透率粗化的最主要目标就是提供一种等效渗透率值。这个等效渗透率值要尽可能地保持原有精细尺寸渗透率地质模型的地质非均质性。

加大网格尺寸会产生一系列问题：如何计算以地质模型为基础的较大尺寸网格上的渗透率。由于渗透率不是一个可以简单累加的量，即不可能通过简单的代数平均来得到等效渗透率（Galli and others，1996；Dimitrakopoulos and other，1993）。

这一事实使这个问题变得复杂化。在过去二十多年中，国外有数百篇文章讨论了这一问题（Galli and others，1996）。人们已提出了数十种渗透率粗化的方法（Galli and others，1996）。早期的方法认为渗透率在各个方向上是相同的，即认为渗透率是各向同性的。这样粗化得到的渗透率在各个方向上就都是相同的，因此这样的粗化值就不能很好地保持渗透率地质模型的非均质性。更为完善一些的新技术要考虑渗透率的各向异性，在各个不同的方向上采用不同的方法，可以得到渗透率粗化后的张量值。因此这些新技术能较好地保持地质模型的非均质性，但在方法上相对也较复杂。

在国内，油藏数值模拟已普遍应用于油田开发方案的制订和调整。有的油田还引进了大型的并行机和百万节点的数值模拟软件。因此，对油藏数值模拟的输入资料有很高的要求。油藏描述所产生的地质模型也广泛地作为数值模拟器的输入。表 9.1 是 DQ 油田使用油藏数值模拟的情况。

表 9.1　DQ 油田使用油藏数值模拟的情况

网格体的规模	大规模级	中规模级	小规模级
具体输入的网格数目	100×100×100	50×50×50	40×40×10
平均每个输入的网格块	x 方向：20～70m，y 方向：20～70m，z 方向：1～20m		
使用计算机的机型	工作站或并行机	工作站	工作站
计算机的 CPU 数目	2～8 个	1～2 个	1 个
CPU 速度	300MIPs		
内存容量	3G～8GB	150～250MB	120MB
使用的数值模拟软件	VIP 并行	DQHY 黑油	
运算一次所需时间	100 小时	20 小时	8 小时

由表 9.1 可见，油藏数值模拟所要求输入的数据体的网格数目也受到计算机能力的限制，一般从几万到近百万个。而用目前国内油藏描述软件所生成的网格数据体的网格数目最大可达数百万个。只能处理十几万个网格数据的小型工作站是不可能对数百万个网格数据进行处理的。因此对运用这些软件生成的网格数据体进行粗化是必要的。

本章的目的就是，研究对储层随机建模所生成的三维定量精细地质模型进行粗化的技术，主要是研究渗透率的粗化技术。在此基础上，比较已有的渗透率粗化技术，讨论几种较合理的方法。

9.1.2　渗透率粗化的现状

渗透率粗化技术是 20 世纪 70 年代在国外出现的一种新技术。它是应动态油藏数值模拟对油藏描述结果的使用要求而发展起来的。

作为早期的方法，算术平均是渗透率粗化的最简单方法，但它也是最不精确的方法。由

于渗透率不是一个可简单叠加的量（Galli and others，1996；Dimitrakopoulos and other，1993）。因此在进行粗化时，必须考虑渗透率的实际物理意义及储层的非均质性，并利用渗流理论的思想进行网格渗透率合并。

早期的方法有调和平均、几何平均、加权平均等。由于大部分方法没有考虑局部渗透率的空间分布，且认为渗透率是各向同性的，因此它们都是不够精确的。

为了更合理地进行渗透率粗化，人们又提出了新的粗化方法。比如，有基于地质统计学的加权平均法（Dimitrakopoulos and other，1993）。用有限差分法求解流动方程的粗化。还有基于自相容近似的方法，结合了几何平均与调和平均的代数公式法及基于重整化的粗化等（Galli and others，1996；Christie and others，1995；Peaceman，1996）。

最初的重整化方法都要求粗化前的微观块和粗化后的宏观块是矩形，并且一个宏观块中应有 $N\times N\times N$ 个微观块，而 N 必须是2的幂次方。而较新的重整化方法可以克服这些限制。基于重整化的方法有标准实空间重整化、张量重整化、简化重整化等（Galli and others，1996；Dimitrakopoulos and other，1993）。它们都可以考虑渗透率的各向异性。其中，张量重整化方法采用了一种新的有限差分“直接”方案和周期性边界条件，它允许计算渗透率全张量。最新的各种重整化方法大都考虑了渗透率的各向异性，可以计算渗透率的张量，能够更合理地反应渗透率的物理性质，它还具有计算速度快的优点。

还有一大类渗透率粗化方法是利用达西定律，且以质量守恒为基础的求解流动方程的粗化方法（Almeida and others，1996；Barker and other，1996）。这类方法充分利用了储层渗透率的物理性质，可以较好地解决渗透率粗化问题。这类方法中，大部分都可计算渗透率张量。在早期，能够计算渗透率张量的方法大都只能计算二维张量或对角线张量。而较新的方法则可计算三维张量（Gomez－Hernndez and other，1994；Aasum and others，1993；Lee and others，1995；Lee and others，1994）。

较新的方法还有两相粗化法（Pickup and other，1994），动态粗化技术（Thibeau，and others，1995），解析粗化方法（Aasum and others，1993；Lee and others，1995；Lee and others，1994）等。为了更好地提高精度，目前众多研究者都利用了考虑了储层渗透率各向异性且带有边界条件的新的粗化方法。

9.2　基本理论

渗透率是本章研究的核心内容，在研究渗透率粗化之前，有必要了解渗透率的性质及它在油藏工程中所处的位置。下面的讨论就从渗透率的概念开始，介绍了渗透率的含义，以及它与地质统计学、油藏数值模拟等的关系，渗透率粗化方法和求解流动方程的关系及渗透率的分类。

9.2.1　渗透率

渗透率是储层岩石通过流体能力的最重要量度（秦同洛等，1989）。法国科学家达西（Darcy）于1856年公布了利用水通过自制的钢管砂子滤器，进行稳定流实验研究的结果。后人把他的成果进行归纳和推广，称之为达西定律，并将渗透率的单位命名为达西。一个达西单位的渗透率表示，长度为1cm和截面为1cm^{2}的岩样，在压力梯度为0.101MPa（兆帕）的作用下，能通过粘度为10^{-3}Pa·s流体的流量1cm^3/s。在国际标准（SI）系统中，渗透率的单位是面积，常以微米平方表示。

岩石渗透率的测定，通常采用达西的线性流动方式，其微分关系式为：

$$q = -\frac{AK\mathrm{d}p}{\mu \mathrm{d}L} \tag{9.1}$$

式中　q——流体通过岩心的稳定流量，m^3/s；

A——岩石的横截面积，m^2；

K——岩石的渗透率，μm^2；

μ——流体粘度，Pa·s；

$\mathrm{d}p/\mathrm{d}L$——压力梯度，Pa/s。

将（9.1）式分离变量后进行积分得渗透率的表达式为：

$$K = \frac{q\mu L}{A\Delta p} \tag{9.2}$$

式中　Δp——测试时作用于岩石两端的压力，Pa；

L——岩心长度，m。

其它符号的单位同（9.1）式所注（秦同洛等，1989）。

岩石孔隙中只有单相流体存在时，所测得的渗透率就称为绝对渗透率，它不随测试气体和测试压力而改变。岩石的渗透率是一个张量。一般来说，储油岩层不是均质的、各向同性的，而是各向异性的。这一点碳酸盐岩储层特别突出。对于砂岩储层，除了水平和垂直方向的渗透率有差异外，在平面上各方向的渗透率也往往是有差别的。因此，渗透率的各向异性问题应是油田开发中应考虑的一个因素。但有时，储层砂岩在各方向上的水平渗透率可能差不多，可以近似地看做各向同性（秦同洛等，1989）。

当有两相或三相介质同时在多孔介质中流动时，相渗透率就是对每一流动相，通过多孔介质能力的直接量度。相对渗透率定义为，有效渗透率与特定渗透率的比值，以%或 f 表示。有效渗透率，是在多相同时流动的条件下，多孔介质对某一相的渗透率，有时又称之为相渗透率。特定渗透率是指岩石在特定的饱和条件下，测定的有效渗透率，以 μm^2 或 $10^3\mu m^2$ 表示。由于它是量度相对渗透率的基准，故又有基准渗透率之称。

渗透率在油藏工程中具有十分重要的意义。相对渗透率能够反映油层的束缚水饱和度，残余油饱和度，以及在开发过程中随着注入水饱和度的增加，油井含水的增长情况[17]。因此，渗透率在油藏数值模拟中是最重要的参数。同时，由于渗透率具有特殊的物理性质，对它的粗化不能用简单的数学平均。因此，对渗透率的粗化特别地受到重视。

本章中所研究的渗透率粗化方法都是针对有效渗透率的。

9.2.2　拟渗透率粗化技术和求解流动方程

对网格渗透率进行粗化的技术可分为拟渗透率粗化技术（或快速渗透率粗化技术）和基于达西定律求解流动方程的渗透率粗化技术。

通过求解流动方程，也就是求解基于达西定律的质量守恒方程，可以得到一个渗透率的粗化值。由于这种方法利用了渗透率的物理性质、流体的流动性质和达西定律，它所得到的渗透率值最接近于宏观网格块的真实渗透率值。实际上是不可能得到宏观块上的真实渗透率值的，因此就可以把利用求解流动方程所得到的渗透率值看做宏观块真实渗透率的等效值。

当无论通过数值的或解析的求解压力的方法不能或难以计算出渗透率的粗化值时，就必须用拟渗透率粗化方法，即快速渗透率粗化方法来得到等效渗透率的一个逼近值。这个值可以用于初始化迭代的渗透率粗化方法，其目的是使该粗化方法的迭代次数最少。为了验证所

选择的拟渗透率粗化方法的优劣，就有必要用众多的拟渗透率粗化方法的结果与等效渗透率的真实值相对比，找到一个最优的结果。而在大部分问题中，通过解析地求解流动方程的方法是不可能获得等效渗透率的真实值的。此时，就可以用通过数值地求解流动方程的方法所得到的等效渗透率值，来近似地代替渗透率的真实值。为此，可把这个值定义成“参考渗透率值”，将它视为渗透率的真实值，用以评价诸多拟渗透率粗化方法的优劣。

而数值地求解流动方程的渗透率粗化方法有很多，比如，有限差分法、有限元方法和求解基于达西定律质量守恒方程的方法等。在这里，选择了后者。这是由于该方法的概念明确，CPU时间比较节省。

9.2.3　渗透率粗化技术和油藏数值模拟（陈月明，1989）

在油气采收的生产及优化过程中，对油藏的性能的理解、预测和历史拟合都基于对油藏流动模拟的研究。油藏数值模拟就是用油藏数值模型来研究油藏的各种物理性质和流体在其中的流动规律，以便更好地认识油层，作出正确的评价，确定合理开发方案和提高采收率的措施。具体来说，就是通过求解油藏物理过程的数学方程组来研究这个物理过程的变化规律。随着计算机软硬件技术的发展，20世纪60年代以后，大多数油藏数值模拟采用数值方法来求解数学方程组。

油藏数值模拟要输入的资料包括油藏描述产生的静态资料和生产井/注入井的数据。油藏描述资料包括地质静态描述，如油藏构造、油层厚度、孔隙度、渗透率、油层深度等。利用地质统计学方法进行的油藏描述，再利用不同的粗化方法就可得到这些数据。

由于渗透率粗化的方法有多种，将不同渗透率粗化方法的结果，分别输入油藏数值模拟，得到的产油量、含水率、饱和度分布和压力可以有多个。将不同渗透率粗化方法的结果，依油藏数值模拟的结果不同进行排队，选取出一个或几个最佳的结果，以确定最佳的渗透率粗化方法。这就是渗透率粗化和油藏数值模拟的双向迭代作用。

总的说来，渗透率粗化技术就是把利用地质统计学方法进行油藏描述所得到的网格渗透率值，转换到适合于油藏数值模拟系统使用的渗透率值的一个接口技术，是油藏描述和油藏数值模拟之间的一座桥梁。

9.2.4　渗透率粗化技术的分类

渗透率粗化技术有许多，可以按不同的标准进行不同的分类。

可以按油藏描述所使用的地质模型把渗透率粗化技术分为三类：确定性方法、随机方法、启发式方法。确定性方法假定地质模型已完全精确地被掌握；而随机方法却与之相反，它假定近似地了解了这个模型，并带有一种概率的观点。启发式方法就是提出一组规则来计算似乎真实的等效渗透率。

在商品化软件“STORM”（Geometric，1996）中提出了5种渗透率粗化方法，可以分为三类：

1）在宏观块中基于达西定律及质量守恒条件下求解流动方程的粗化；

2）以重整化原理为基础的粗化；

3）算术调和平均。

按复杂度与精度从低到高来划分，可把渗透率粗化分为：

1）认为块渗透率在各个方向上是均匀的，利用各种平均的方法，如算术平均、调和平均、几何平均等；

2）认为块渗透率在各个方向上是不均匀的，在水平和垂直方向上分别用平均的方法计

算块渗透率；

3)最复杂、且最精确的方法就是当运用非流动边界条件到所有的块上时，求解流动方程。

此外，还有其它的分类方法。

9.3 粗化方法

在下面的叙述中，用到一些专用术语，首先说明如下。微观块，是指网格粗化前的小尺寸网格块。宏观块，是指网格粗化后的大尺寸网格块。一组微观网格块可以粗化合并成为一个宏观网格块。微观压力是网格粗化前微观块上所受到的压力。宏观压力，是指网格粗化前宏观块上所受到的压力。微观流量，是指流体通过微观块的流量。宏观流量，则是流体通过宏观块的流量。

符号系统说明：k：细体小写符号：代表标量；$\boldsymbol{K}$，k：黑体或大写符号，代表张量；E ()：数学期望；σ^2，Var ()：方差；C ()，Cov ()：协方差；γ ()：变异函数；D：空间维数；Q：流量；μ：流体粘度；ω：权系数；p：压力；K：各向同性或各向异性的局部渗透率。

9.3.1 基于平均的粗化

渗透率粗化的最简单方法就是数学平均，包括算术平均、几何平均、调和平均等。它们都假设储层是各向同性的，即渗透率张量的各个分量具有相同的值 $K_x = K_y = K_z = k$，并且不考虑介质中流体的流动方向，不考虑介质受到的压力。其粗化方法就是把一组微观尺度网格块上的渗透率分别进行算术平均、几何平均、调和平均或加权平均而得到相应宏观尺度网格块上的渗透率。

假设在如图 9.1 所示平面上的一组微观块，各微观块的渗透率分别是 $k_{1,1}$，$k_{1,2}$，$k_{2,1}$，$k_{2,2}$。使用不同的粗化方法后，它的的等效渗透率 k_{ef}分别应该是：

$$k_{ef} = (k_{1,1} + k_{1,2} + k_{2,1} + k_{2,2})/4 \tag{9.3}$$

$$k_{ef} = \sqrt[4]{k_{1,1} \cdot k_{1,2} \cdot k_{2,1} \cdot k_{2,2}}, k_{1,1}, k_{1,2}, k_{2,1}, k_{2,2} \text{ 均大于 } 0 \tag{9.4}$$

$$k_{ef} = \frac{2 \times (k^1 \times k^2)}{k^1 + k^2} \tag{9.5}$$

其中，$k^1 = \dfrac{2 \times (k_{1,1} \times k_{1,2})}{k_{1,1} + k_{1,2}}$，$k^2 = \dfrac{2 \times (k_{2,1} \times k_{2,2})}{k_{2,1} + k_{2,2}}$。(9.3)，(9.4)，(9.5) 分别对应的是算术平均，几何平均，调和平均等三种方法进行粗化后得到的等效渗透率。

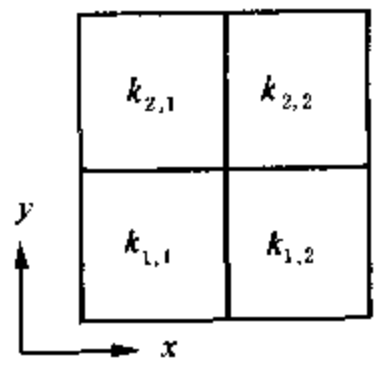


图 9.1　平均粗化的例子

假设，流动方向是在 x 方向上。这样，会产生两种特殊的情况。第一种情况，若 $k_{1,1}=0$，$k_{2,1}=0$，那么在整个宏观块上是没有流动的，但用算术平均得到的结果却不为 0。第二种情况，若 $k_{1,1}=0$，$k_{1,2}=0$，则整个宏观块上是有流动的，但用几何平均得到的结果却等于 0。然而，此时因为分母为零，调和平均根本就不可能运算。由此可见，这几种普通平均的结果常常是不真实的。通常，算术平均要大于真实值，而调和平均和几何平均要小于真实值。这些方法的优点是算法简单，易于实现，运算速度快。

但是，从另外一种角度来看此问题，可以发现，若对平行于流动方向的一组微观块使用调和平均方法进行粗化，得到一个粗化值 μ_h，对垂直于流动方向的一组微观块使用算术平均方法进行粗化，得到一个粗化值 μ_a，联合使用这两种方法，就可避免出现前面所述的两种情况。由这一思路，就可得到另一种很有效的粗化方法——简化重整化方法，这一方法将在 9.3.6 节讨论。

9.3.2　基于达西定律质量守恒的求解流动方程的粗化（Almeida，1996）

描述多孔介质中流体流动的最基本的两个方程就是：达西定律和质量守恒方程。

渗透率的本质就是储层岩石通过流体的能力的量度，其规律通过达西定律来描述。而在活塞式流动过程中，流入的流体和流出的流体的质量是相等的，这也是活塞式流动的一个基本性质。因此，用基于达西定律和质量守恒方法才能很好地解决渗透率粗化问题。

1. 理论分析（Almeida，1996）

若将一组微观块合并为一个宏观块时，不考虑这一个宏观块周围其它微观块对它的影响，则称此合并为不带有边界条件。首先，分析一个不带边界条件的算法。此算法还假定在水平方向上渗透率是各向异性的。

假设每个储层的网格块的尺寸是 Δx、Δy、Δz，并且已知每个网格块的渗透率。假定渗透率是一个张量型变量，在每个块中已知 x，y，z 三个方向的分量为 K_x，K_y，K_z。若微观块的初始渗透率为一个标量 k。这时，可以在微观块上给渗透率张量的三个分量分配以相同的值 $K_x = K_y = K_z = k$。

为了简便起见，这里只考虑了在二维空间上的粗化方法。此方法的目标是计算每个宏观块的等效渗透率张量的分量（并设在 x、y 方向上各有 n_x、n_y 个微观块）K_x、K_y（见图 9.2）。

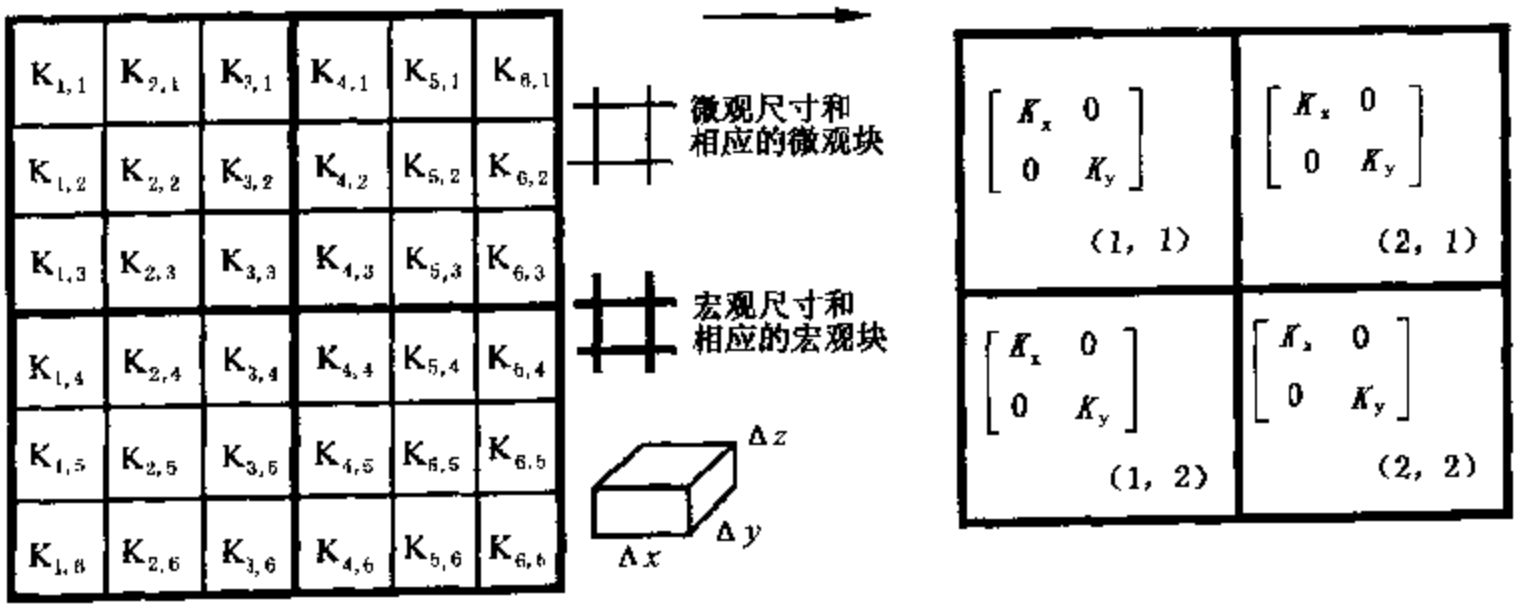


图 9.2　一组微观网格块中的渗透率粗化

在与一个宏观块对应的每一组微观块上，考虑了两种活塞式流动情况（见图 9.3），每种情况都分别平行于块的坐标轴，并且是从无流动边界条件的压力梯度导出的。为了计算等效渗透率的分量 k_x，模拟了一个平行于 ox 轴的流动，并遵从如下条件：

1）属于第一列的块都有一个恒定压力，均为 p_{inp}；

2）属于第 n_x 列的块的压力都比 p_{inp} 小；利用等于 $p_{inp} - p_{out}$ 的压力梯度可以导出 p_{out}；

3）边界上的块没有流动能力，即在 oy 方向上没有流动。考虑不可压缩流体的流动，对于第（i，j）个微观块，由于稳态条件所产生的质量守恒方程如下

$$(Q_{x_{i+\frac{1}{2}},j} - Q_{x_{i-\frac{1}{2}},j}) + (Q_{y_{i,j+\frac{1}{2}}} - Q_{y_{i,j-\frac{1}{2}}}) = 0 \tag{9.6}$$

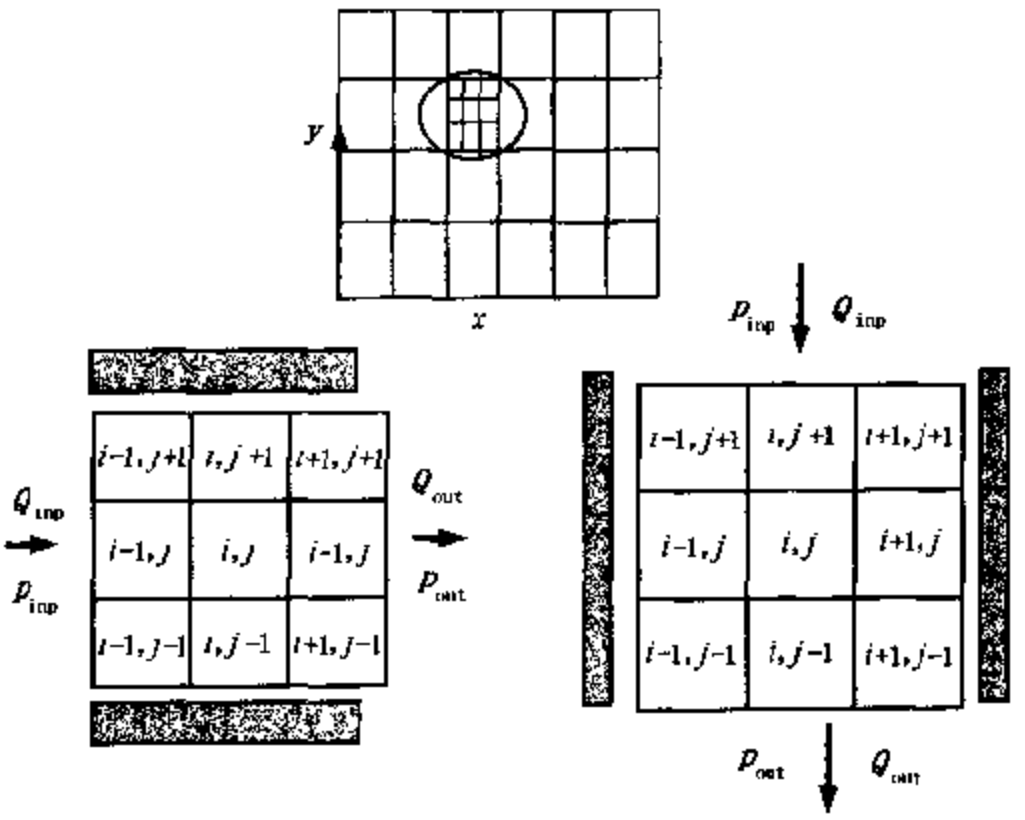


图 9.3　对每个宏观块确定有效渗透率的两个流动条件

(Almeida, 1996)

上式中，下标 $(i+\frac{1}{2}, j)$，$(i-\frac{1}{2}, j)$，$(i, j-\frac{1}{2})$，$(i, j+\frac{1}{2})$ 分别代表 (i, j) 块与相邻块［即 $(i+1, j)$，$(i-1, j)$，$(i, j-1)$，$(i, j+1)$］的交面，Q 表示两块间的流量。

假定只考虑一种流体和由达西定律的适用性条件所产生的流动，由

$$Q_x = \frac{-K_x \Delta y \Delta z}{\mu} \frac{\mathrm{d}p}{\mathrm{d}x} \text{ 和 } Q_y = \frac{-K_y \Delta x \Delta z}{\mu} \frac{\mathrm{d}p}{\mathrm{d}y} \tag{9.7}$$

可产生：

这里，$n_x = n_y = 3$。

$$\begin{aligned}
Q_{x_{i-\frac{1}{2},j}} &= -\frac{K_{x_{i-\frac{1}{2},j}} \Delta y \Delta z}{\mu} \frac{(p_{i-1,j} - p_{i,j})}{\Delta x} \\
Q_{x_{i+\frac{1}{2},j}} &= -\frac{K_{x_{i+\frac{1}{2},j}} \Delta y \Delta z}{\mu} \frac{(p_{i,j} - p_{i+1,j})}{\Delta x} \\
Q_{y_{i,j-\frac{1}{2}}} &= -\frac{K_{y_{i,j-\frac{1}{2}}} \Delta x \Delta z}{\mu} \frac{(p_{i,j-1} - p_{i,j})}{\Delta y} \\
Q_{y_{i,j+\frac{1}{2}}} &= -\frac{K_{y_{i,j+\frac{1}{2}}} \Delta x \Delta z}{\mu} \frac{(p_{i,j} - p_{i,j+1})}{\Delta y}
\end{aligned} \tag{9.8}$$

上式中，$K_{x_{i-\frac{1}{2},j}}$ 表示块 $(i-1, j)$ 和块 (i, j) 交面处的平均渗透率，$K_{x_{i+1,j}}$ 表示块 (i, j) 和 $(i+1, j)$ 交面处的平均渗透率；$K_{y_{i,j-\frac{1}{2}}}$ 表示块 $(i, j-1)$ 和块 (i, j) 交面处的平均渗透率，$K_{y_{i,j+\frac{1}{2}}}$ 表示块 (i, j) 和 $(i, j+1)$ 交面处的平均渗透率。以上四个渗透率可用两个相邻块的调和平均值来计算：

$$K_{x_{i-\frac{1}{2},j}} \cong 2 \frac{K_{x_{i-1,j}} \cdot K_{x_{i,j}}}{K_{x_{i-1,j}} + K_{x_{i,j}}}$$

$$K_{x_{i+\frac{1}{2},j}} \cong 2\frac{K_{x_{i,j}} \cdot K_{x_{i+1,j}}}{K_{x_{i,j}} + K_{x_{i+1,j}}}$$

$$K_{y_{i,j-\frac{1}{2}}} \cong 2\frac{K_{y_{i,j-1}} \cdot K_{y_{i,j}}}{K_{y_{i,j-1}} + K_{y_{i,j}}} \tag{9.9}$$

$$K_{y_{i,j+\frac{1}{2}}} \cong 2\frac{K_{y_{i,j}} \cdot K_{y_{i,j+1}}}{K_{y_{i,j}} + K_{y_{i,j+1}}}$$

把（9.8）式代入（9.6）式，并提出因子可得下式：

$$\frac{\Delta y\Delta z}{\Delta x}\left[K_{x_{i+\frac{1}{2},j}}(p_{i,j} - p_{i+1,j}) - K_{x_{i-\frac{1}{2},j}}(p_{i-1,j} - p_{i,j})\right] +$$

$$\frac{\Delta x\Delta z}{\Delta y}\left[K_{y_{i,j+\frac{1}{2}}}(p_{i,j} - p_{i,j+1}) - K_{y_{i,j-\frac{1}{2}}}(p_{i,j-1} - p_{i,j})\right] = 0$$

为简化上式，两边同乘以 $\Delta x\Delta y/\Delta z$ 可得：

$$\Delta y^2\left[-K_{x_{i+\frac{1}{2},j}}p_{i+1,j} - K_{x_{i-\frac{1}{2},j}}p_{i-1,j} + (K_{x_{i+\frac{1}{2},j}} + K_{x_{i-\frac{1}{2},j}}) \cdot p_{i,j}\right] +$$

$$\Delta x^2\left[-K_{y_{i,j+\frac{1}{2}}}p_{i,j+1} - K_{y_{i,j-\frac{1}{2}}}p_{i,j-1} + (K_{y_{i,j+\frac{1}{2}}} + K_{y_{i,j-\frac{1}{2}}}) \cdot p_{i,j}\right] = 0 \tag{9.10}$$

在每个微观网格块上都可以有一个这样的质量守恒公式，这样可得到一个有个方程的方程组，通用表达式如下所示：

$$A_{i,j}p_{i,j} - B_{i,j}p_{i-1,j} - C_{i,j}p_{i+1,j} - D_{i,j}p_{i,j-1} - E_{i,j}p_{i,j+1} = 0 \tag{9.11}$$

式中，$A_{i,j} = (K_{x_{i+\frac{1}{2},j}} + K_{x_{i-\frac{1}{2},j}}) \cdot \Delta y^2 + (K_{y_{i,j+\frac{1}{2}}} + K_{y_{i,j-\frac{1}{2}}}) \cdot \Delta x^2$；

$B_{i,j} = K_{x_{i-\frac{1}{2},j}} \cdot \Delta y^2$；$C_{i,j} = -K_{x_{i+\frac{1}{2},j}} \cdot \Delta y^2$；$D_{i,j} = K_{y_{i,j-\frac{1}{2}}} \cdot \Delta x^2$；$E_{i,j} = -K_{x_{i,j+\frac{1}{2}}} \cdot \Delta x^2$。

此方程可用迭代法求解。在这里使用了逐次超松弛因子迭代法。在每次迭代中，块（i，j）上的一个新的压力可用下式解出：

$$p_{i,j}^{k+1} = p_{i,j}^{k} + \omega(R_{i,j}^{k}/A_{i,j}) \tag{9.12}$$

上式中，ω 为松弛因子，它的优化值在 1 到 2 之间，其目的是使迭代次数减到最少。在这里它的取值是 1.8（Almeida，1996），并且，

$$R_{i,j}^{k} = -A_{i,j}p_{i,j}^{k} + B_{i,j}p_{i-1,j}^{k+1} + C_{i,j}p_{i+1,j}^{k} + D_{i,j}p_{i,j-1}^{k+1} + E_{i,j}p_{i,j+1}^{k}$$

方程组（9.11）式的解就是块（i，j）中心的压力值 $p_{i,j}$。

为了在 ox 方向得到相同流动方向的渗透率张量的分量（在宏观网格上的），可以用达西定律和质量守恒定律得到：

$$Q_{\text{inp}} = -\frac{\Delta y\Delta z}{\Delta x}\sum_{j=1}^{n_y}\left[K_{x_{i+\frac{1}{2},j}}(p_{2,j} - p_{\text{inp}})\right] \tag{9.13}$$

$$Q_{\text{out}} = -\frac{\Delta y\Delta z}{\Delta x}\sum_{j=1}^{n_y}\left[K_{x_{n_x-\frac{1}{2},j}}(p_{\text{out}} - p_{n_{x-1,j}})\right] \tag{9.14}$$

给定 $Q_{\text{inp}} = Q_{\text{out}}$，利用（9.7）式及（9.13）式和（9.14）式可得：

$$\frac{\Delta y\Delta z}{\Delta x}\sum_{j=1}^{n_y}\left[K_{x_{i+\frac{1}{2},j}}(p_{2,j} - p_{\text{inp}})\right] = -\frac{n_y\Delta y\Delta z}{(n_x - 1)\Delta x}K_x(p_{\text{out}} - p_{\text{inp}}) \tag{9.15}$$

由此可得：

$$K_x = \sum_{j=1}^{n_y} [K_{x_{i+\frac{1}{2},j}}(p_{2,j} - p_{inp})] / \frac{n_y}{(n_x - 1)}(P_{out} - P_{inp}) \tag{9.16}$$

同理考虑平行于 oy 方向的流动，可以得到下面的解：

$$K_y = \sum_{i=1}^{n_x} [K_{y_{i,j+\frac{1}{2}}}(p_{i,2} - p_{inp})] / \frac{n_x}{(n_y - 1)}(P_{out} - P_{inp}) \tag{9.17}$$

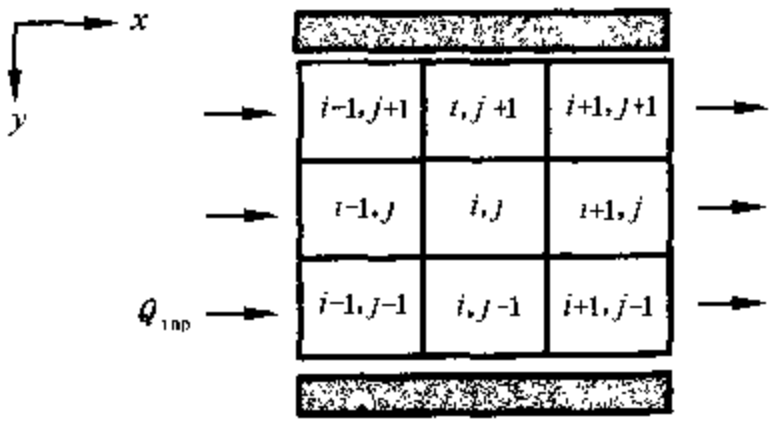


图 9.4　仅在 x 方向有流动的情况

2. 具体计算方法

(1) 计算 x 方向的宏观块渗透率

如图 9.4 所示，在 x、y 方向上各有 n_x、n_y 个微观块。假设 y 方向为不渗透流动边界，只有 x 方向有流动。

设第一列微观块都有同一压力梯度 p_1，第二列微观块都有同一压力梯度 p_2，依此类推，第 n_x 列微观块都有同一压力梯度 p_n。这样，对每一列微观块，都有相同的压力梯度。因此，在 y 方向的流量为 0，即 $Qy_{i,j+\frac{1}{2}} = Qy_{i,j-\frac{1}{2}} = 0$，而在 x 方向的流量不为 0，即 $Qx_{i+\frac{1}{2},j} = Qx_{i-\frac{1}{2},j} \neq 0$。于是，式 (9.6) 就可简化为：

$$(Qx_{i+\frac{1}{2},j} - Qx_{i-\frac{1}{2},j}) = 0 \tag{9.18}$$

把 (9.8) 式代入上式，并提出因子可得下式：

$$\frac{\Delta y \Delta z}{\mu \Delta x}[K_{x_{i+\frac{1}{2},j}}(p_{i,j} - p_{i+1,j}) - K_{x_{i-\frac{1}{2},j}}(p_{i-1,j} - p_{i,j})] = 0$$

$$K_{x_{i+\frac{1}{2},j}}(p_{i,j} - p_{i+1,j}) - K_{x_{i-\frac{1}{2},j}}(p_{i-1,j} - p_{i,j}) = 0$$

即

$$(K_{x_{i+\frac{1}{2},j}} + K_{x_{i-\frac{1}{2},j}})p_{i,j} - K_{x_{i+\frac{1}{2},j}}p_{i+1,j} - K_{x_{i-\frac{1}{2},j}}p_{i-1,j} = 0$$

对于每个微观块都可有这样一个质量守恒方程。这样可得到一个有 n_x，n_y 个方程的方程组，通用形式为：

$$A_{i,j}p_{i,j} - B_{i,j}p_{i-1,j} - C_{i,j}p_{i+1,j} = 0 \tag{9.19}$$

其中 $A_{i,j} = K_{x_{i+\frac{1}{2},j}} + K_{x_{i-\frac{1}{2},j}}$；　$B_{i,j} = K_{x_{i-\frac{1}{2},j}}$；　$C_{i,j} = K_{x_{i+\frac{1}{2},j}}$。

此方程组可以用松弛因子法来解。在每次迭代中，块 (i, j) 上的一个新的压力可用下式解出：

$$p_{i,j}^{k+1} = p_{i,j}^{k} + \omega(R_{i,j}^{k}/A_{i,j}) \tag{9.20}$$

上式中，ω 为松弛因子，在这里其取值是 1.8 (Almeida, 1996)，并且，

$$R_{i,j}^{k} = -A_{i,j}p_{i,j}^{k} + B_{i,j}p_{i-1,j}^{k+1} + C_{i,j}p_{i+1,j}^{k}$$

这样就可求得块 (i, j) 中心的压力值 $p_{i,j}$，然后就可利用 (9.16) 式求出 x 方向的粗化渗透率值来。

(2) 计算 Y 方向的宏观块渗透率

如图 9.5 所示，在 x、y 方向上各有 n_x、n_y 个微观块。假设 x 方向为不渗透流动边界，

只有 y 方向有流动。要计算 y 方向的宏观渗透率，所用方法与在 x 方向使用的方法类似。

设第一行微观块都有同一压力梯度 p_1，第二行微观块都有同一压力梯度 p_2，依此类推，第 n_y 行微观块都有同一压力梯度 p_n。这样，对每一行微观块，都有相同的压力梯度。因此，在 x 方向的流量为 0，即

$$Qx_{i+\frac{1}{2},j} = Qx_{i-\frac{1}{2},j} = 0,$$

而在 y 方向的流量不为 0，即

$$Qy_{i,j+\frac{1}{2}} = Qy_{i,j-\frac{1}{2}} \neq 0。$$

于是，式（9.6）就可简化为

$$(Q_{y_{i,j+\frac{1}{2}}} - Q_{y_{i,j-\frac{1}{2}}}) = 0 \qquad (9.21)$$

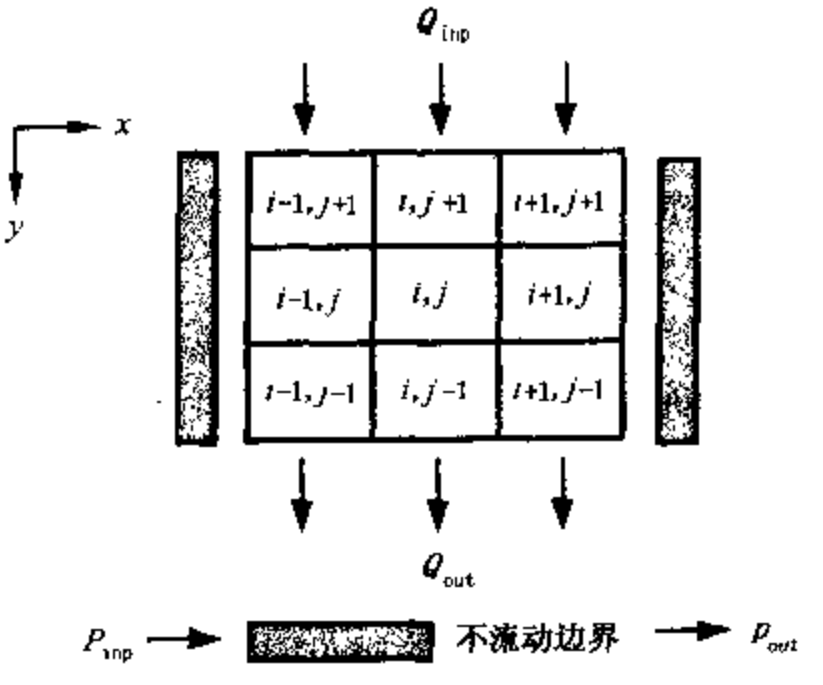


图 9.5　仅在 y 方向有流动的情况

把（9.8）式代入上式，并提出因子可得下式：

$$\frac{\Delta x \Delta z}{\mu \Delta y}[K_{y_{i,j+\frac{1}{2}}}(p_{i,j} - p_{i,j+1}) - K_{y_{i,j-\frac{1}{2}}}(p_{i,j-1} - p_{i,j})] = 0$$

即

$$(K_{y_{i,j+\frac{1}{2}}} + K_{y_{i,j-\frac{1}{2}}})p_{i,j} - K_{y_{i+\frac{1}{2},j}}p_{i+1,j} - K_{y_{i-\frac{1}{2},j}}p_{i-1,j} = 0$$

同样，对于每个微观块都可有这样一个质量守恒方程。这样可得到一个有 $n_x \cdot n_y$ 个方程的方程组，通用形式为：

$$A_{i,j}p_{i,j} - B_{i,j}p_{i-1,j} - C_{i,j}p_{i+1,j} = 0 \qquad (9.22)$$

其中 $A_{i,j} = K_{y_{i,j+\frac{1}{2}}} + K_{y_{i,j-\frac{1}{2}}}$；　$B_{i,j} = K_{y_{i,j-\frac{1}{2}}}$；　$C_{i,j} = K_{y_{i,j+\frac{1}{2}}}$。

与 x 方向相同，此方程组同样可以用松弛因子法来解。最后可以求得每个块中心的压力值 $p_{i,j}$，然后就可利用（9.17）式求出 y 方向的粗化渗透率值来。

考虑边界条件的基于达西定律和质量守恒的粗化方法也就是考虑宏观块周围的网格块对其粗化渗透率的影响（王家华，张团峰，1999）。这里所讨论的方法是以上的方法的一种改进。在前面所述的基于求解压力的方法中，计算了渗透率张量的等效分量 K_x、K_y（分别用 9.16 式和 9.17 式）。这是基于平行于坐标轴的流动和用来求解方程组时所强加的非流动边界条件的。这个经典方法有一个重要缺点：由于它孤立地考虑了一组块，相邻的其他块的渗透率值就被忽略了。而相邻块中的渗透率值可能产生不同于假设的垂直于 ox、oy 轴方向的流动模式。在等效渗透率的计算中，应当考虑这些不同的流动模式。由于这个缺点，提出了这种新方法。这个方法基于对每个宏观块的渗透率张量的分量进行迭代和优化计算，并且要利用到相邻的宏观块的渗透率和用前面的粗化方法所产生的粗化值。

如上面所述的求解流动方程的方法在一些情况下，可以成功地用来计算宏观块渗透率。但它不适用于宏观块尺寸上有非常大的非均质性的情况。同时，用这种技术获得的宏观块渗透率依赖于在每个块上求解流动方程时所使用的边界条件。一个宏观块渗透率不可能对于任何边界条件都是有效的。然而，假如微观块渗透率的协方差的变化范围是有限的，那么就会有一个最小的宏观块尺寸，超过它，宏观块渗透率将会和求解它时所用的边界条件相独立（Gomez－Hernandez and Journel，1994）。

用有限差分法求解流动方程的粗化方法（Gomez - Hernandez and Journel，1994），就是试图减少在计算宏观块的等效渗透率时边界条件所带来的影响，用不一定平行于网格块面的主要轴来抓住块渗透率的张量特征。它属于随机类的粗化方法。它也利用了达西定律求解压力的思想来计算网格块的粗化渗透率。限于篇幅，在此对于该种方法不再详述。

9.3.3　重整化方法

重整化方法为一种较新的渗透率粗化方法（Christie and others，1995；Peaceman，1996）。

如前所述的简单平均的方法不够精确，并未考虑储层渗透率的各向异性，也只能计算标量。而如求解流动方程等解析的方法虽较精确，但计算开销比较大。为了既考虑储层非均质性又要求计算速度较快，就提出了渗透率粗化的标准重整化方法。它是前两类方法的折衷。

重整化的基本思想就是，把本来只须一步的从精细网格到粗网格合并计算，变换为一系列的从精细网格到越来越粗的网格的多步合并。

标准重整化是一种递归的算法。宏观网格块的等效渗透率可由对微观网格块的一系列合并而得到。例如，在 D 维空间中，可合并 2^{nD} 个网格到精细度较低的 $2^{(n-1)D}$ 个网格。重复这一过程直到找到所期望的尺寸的网格。一个极端的例子如图 9.6 所示。

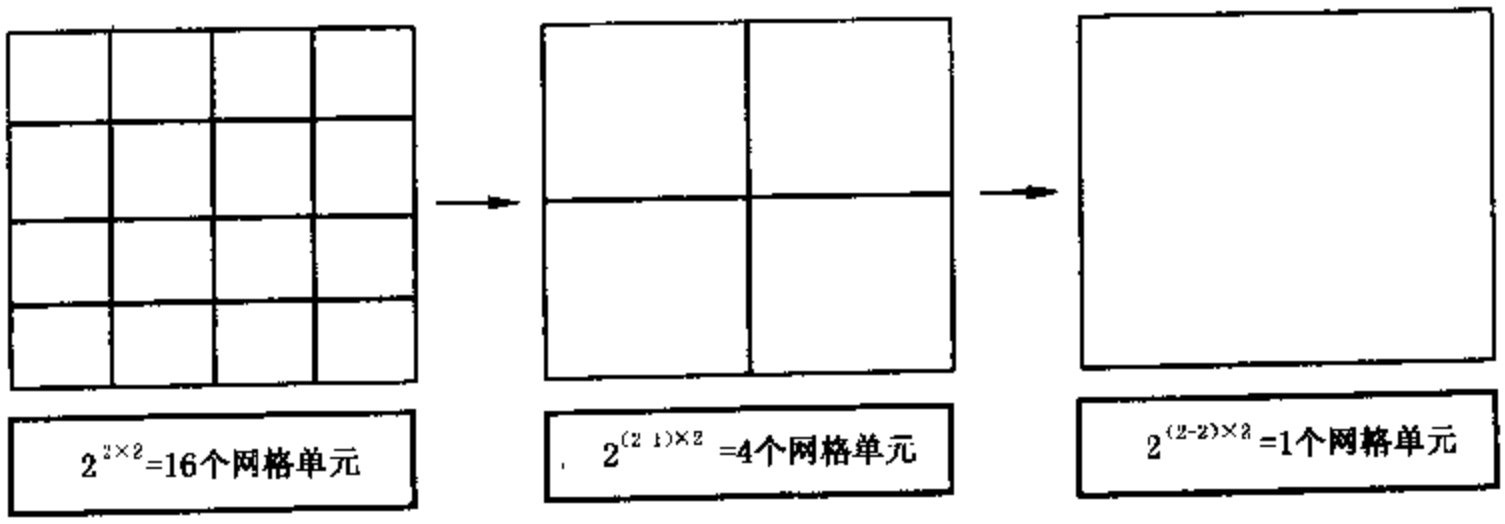


图 9.6　重整化的一般原理

标准重整化的基本操作是在 D 维空间中，计算由 2^D 个微观网格块组成的宏观网格块的等效渗透率。由于在 2 维和 3 维空间中没有精确的公式可用，等效渗透率的计算就不得不用近似计算。例如，在二维空间中，King（Galli and others，1996）用电网络模拟来表示多孔介质，可以通过连续的星——三角形变换获得渗透率。还可以通过列出一个有限差分类型的线性方程来获得渗透率。上面两个过程是等效的，并可产生下式（Galli and others，1996）。

$$K_b^{xx}=\frac{4(k_1+k_3)(k_2+k_4)[k_2k_4(k_1+k_3)+k_1k_3(k_2+k_4)]}{[k_2k_4(k_1+k_3)+k_1k_3(k_2+k_4)][k_1+k_2+k_3+k_4]+3(k_1+k_2)(k_3+k_4)(k_1+k_3)(k_2+k_4)}$$

式中，K_b^{xx} 是 x 方向的宏观块渗透率，k_1,k_2,k_3,k_4 是水平面上四个微观块渗透率。

上式的结果是一种张量型渗透率值。由此可见，只要运用不同的合并格式，就可以产生不同的重整化方法。重整化方法既可以求取渗透率标量，也可以求取渗透率张量。

9.3.4　简化重整化方法（Galli and others，1996）

这是一种新的重整化方法，是对上面的标准重整化方法的改进。这一方法由两部分组成：首先是可以产生两种值的标准重整化；接下来运用一个考虑了非均质性的公式来计算最终渗透率值。

1. 简化重整化

假设要计算在方向 $i\in\{x,y,z\}$ 上的渗透率 K^{ii}。

所有的重整化算法的通用原则就是循环地对网格单元分组进行合并。例如，在 2 维空间上可把 4（$=2^2$）个网格单元分成一组，在 3 维空间上可把 8（$=2^3$）个网格单元分成一组。当这样分组时，需要一个公式来计算这些网格单元基本分组的粗化渗透率。通常，可通过使用在网格单元分组上的有限差分方法求解流动方程来求得粗化渗透率。还可以用解析的或数值的方法来计算网格单元的渗透率。

一个更简单的方法就是把网格单元两个两个地分为一组。要计算网格单元在一个方向上的粗化渗透率，有两种可能性：如果这两个网格单元是并行的，理论上粗化渗透率就等于这两个网格单元渗透率值的算术平均；如果这两个网格单元是串行的，则理论上粗化渗透率就等于这两个网格单元渗透率值的调和平均（见图 9.7）。

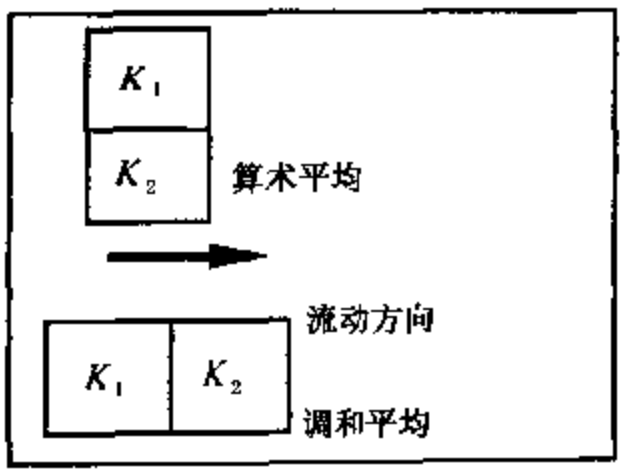


图 9.7　计算两个微观块为--组的粗化渗透率的局部公式

从这个思想来看，可以用两个迭代过程来计算渗透率粗化值（见图 9.8）：

首先把网格单元按平行于假定的流动方向 i 两个两个地分组，然后沿着垂直于流动的方向两个分组（根据介质是 2 维的还是 3 维的，在垂直于流动的方向上可有一个或两个分组）。可以一次次地重复上面这两步至到最终获得唯一的一个块渗透率值，称此值为 $c^{ii}_{\min}$。接下来，先在垂直于流动方向的 2 维空间的每个线上或 3 维空间的每个面上对网格单元分组，然后沿着流动方向对网格单元分组。正如前面一步，迭代过程的最后一步会得到一个值，这个值总比前面的值 $c^{ii}_{\min}$大，称之为 $c^{ii}_{\max}$。

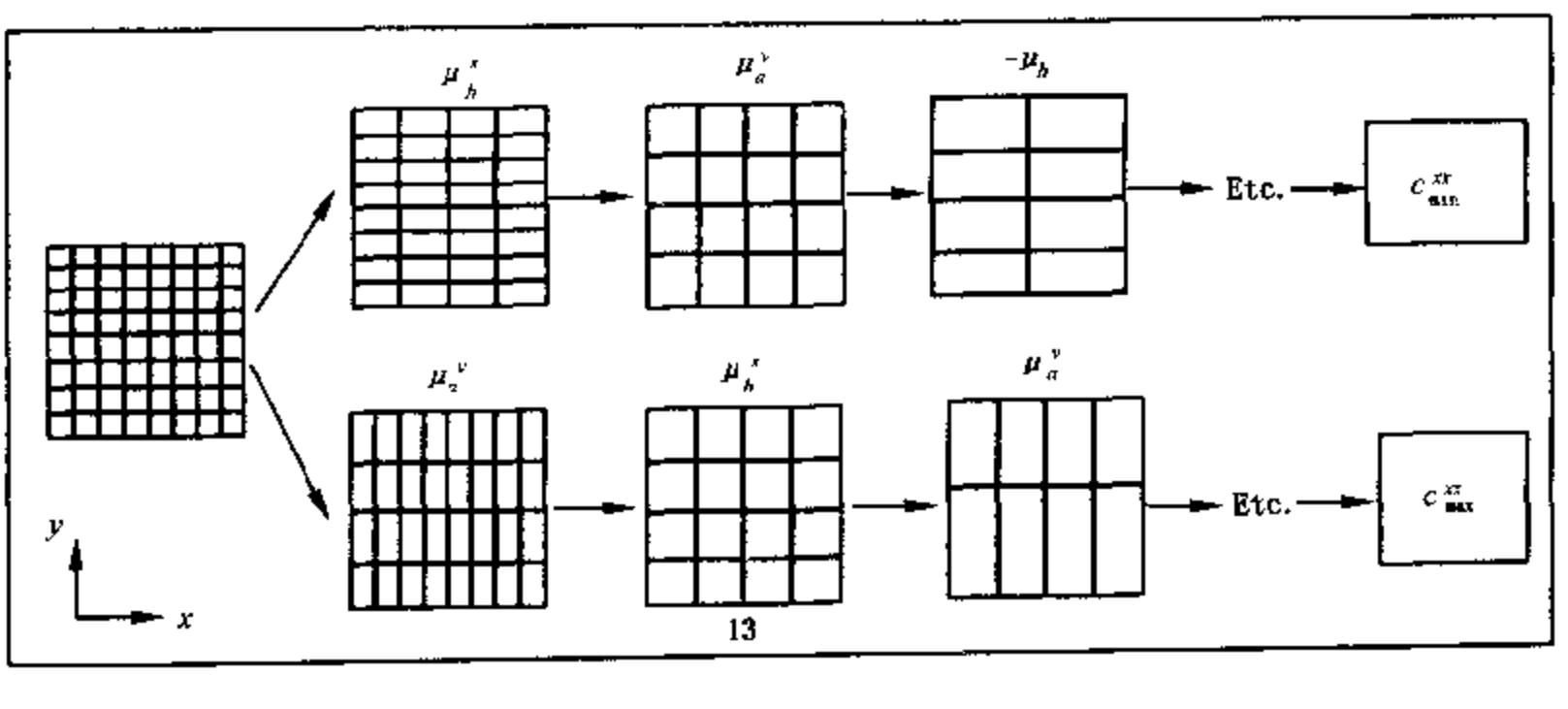


图 9.8　用两个迭代过程来计算渗透率粗化值

如果只考虑渗透率是各向同性的，计算到最后就可得到这两个值的几何平均值 $K=\sqrt{c_{\max}\cdot c_{\min}}$。这就是最后要求的粗化渗透率值。

2. 考虑各向异性

简化重整化和参考渗透率值比较的一个结果就是：当流动垂直于平面网格单元时，$c^{ii}_{\min}$看上去是渗透率的最佳估计值，而当流动平行于平面网格单元时，$c^{ii}_{\max}$看上去是渗透率的最佳估计值。

因此，这里的思想就是找到一个公式，给出考虑介质非均质性的由 $c^{ii}_{\min}$与 $c^{ii}_{\max}$组合成的

单一的渗透率估计值。为了作出这种组合，一个标准的技术就是用指数方法：

$$K = c_{\max}^{\alpha} c_{\min}^{1-\alpha}$$

若指数 α 取值为 1/2，则有 $K = \sqrt{c_{\max} \cdot c_{\min}}$。

下面这部分的目的就是求出作为非均质性函数的指数 α。

（1）基本公式

此处的思想就是要把两个或更多的估计值组合到一个综合公式中。

Kruel－Romeu（Galli and others，1996）给出了下面的公式，它是以作为非均质性比率函数的 Cardwell 及 Parsons 边界的组合为基础的。

$$K_{ge}^{xx} = K_1^{(\theta_{x2}\theta_{y3}+\theta_{x2}\theta_{y3})} K_2^{(1-\theta_{y3}-\theta_{x3})} K_3^{(1-\theta_{x2})\theta_{y3}} K_4^{(1-\theta_{y2})\theta_{x3}}$$

此式中，K_1 与 K_2 是 Cardwell 和 Parsons 边界，K_3、K_4 是两个中间值：

$$K_1 = \mu_h^x[\mu_a^y(\mu_a^z)] = \mu_h^x[\mu_a^z(\mu_a^y)]$$

$$K_2 = \mu_a^y[\mu_a^z(\mu_h^x)] = \mu_a^z[\mu_a^y(\mu_h^x)]$$

$$K_3 = \mu_a^y[\mu_h^x(\mu_a^z)]$$

$$K_4 = \mu_a^z[\mu_h^x(\mu_a^y)]$$

$$\theta_{y2} = \frac{\arctan\sqrt{a_y}}{\pi/2}; \quad \theta_{z2} = \frac{\arctan\sqrt{a_z}}{\pi/2}; \quad \theta_{y3} = \frac{\theta_{y2}(1-\theta_{z2})}{1-\theta_{y2}\theta_{z2}};$$

$$\theta_{z3} = \frac{\theta_{z2}(1-\theta_{y2})}{1-\theta_{y2}\theta_{z2}};$$

$a_y = \frac{k^{yy}}{k^{xx}}\left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)^2$ 及 $a_z = \frac{k^{zz}}{k^{xx}}\left(\frac{\mathrm{d}x}{\mathrm{d}z}\right)^2$ 表示由于拉展网格 $\mathrm{d}x$、$\mathrm{d}y$ 及 $\mathrm{d}z$（代表每个方向上的网格单元的尺寸）和局部渗透率（假定它在全部微观块上是恒定的）的非均质性所产生的各向异性因素。

（2）新公式

这里的思想就是令 $\alpha = \theta_{y3} + \theta_{z3}$。在改变了变量名且经过一些代数变换后，就得到了一个紧凑的 α 表达式。下面的公式总结了其结果：

$$K^{ii} = (c_{\max}^{ii})^{\alpha} \cdot (c_{\min}^{ii})^{1-\alpha} \tag{9.23}$$

其中，

$$\alpha(a_j^i, a_k^i) = \frac{u(a_j^i) + u(a_k^i) - 2u(a_j^i)u(a_k^i)}{1 - u(a_j^i)u(a_k^i)}, \quad i \neq j \neq k \in \{x, y, z\};$$

$$a_i^j = \frac{k^{ii}}{k^{jj}}\left(\frac{d_j}{d_i}\right)^2;$$

$$u(t) = \frac{\arctan\sqrt{t}}{\pi/2}。$$

上式中，非均质性比率 a_i^j 和基本函数 u（t）的定义同（9.23）式，$d_{[i]}$ 代表 i 方向网格单元的尺寸。

由 Ababou 提出而被 Kruel－Romeu（Galli and others，1996）引用了的另一基本函数是 ν（t）。根据这个函数，就有可能定义一个指数 β 来模拟 α，而且易于计算。

$$\nu(t) = \frac{t}{t+1}; \quad \beta(a_j^i, a_k^i) = \frac{\nu(a_j^i) + \nu(a_k^i)}{1 + \nu(a_j^i) + \nu(a_k^i)}$$

不幸的是，即使当描绘这两个函数的图时，α 与 β 间的差别看起来很小，并且即使 β 的表达

式比 α 的简单，但在大部分情况下，α 比 β 更好一些。

9.3.5　基于地质统计学的加权平均法（Dimitrakopoulos，1993）

这是一种用地质统计学的方法来对三维微观渗透率进行粗化的方法。此方法的目标就是开发一种新的实际的方法来解决把渗透率值分配到大尺寸的网格上的粗化问题。它是在地质统计学框架里开发的。在地质统计学里，渗透率代表一个空间上平稳的、各态历经的随机函数的输出。结果，油藏中微观块上渗透率的方差可用概率分布函数来描述，而随机函数的空间协方差即变异函数可从可用的微观块尺寸估计得来[8]。

在二维油藏研究中，宏观块渗透率可定义成微观块数据的空间几何平均。下面的方法就把此方法扩展到三维油藏的情形。这里，绝对宏观块渗透率可定义为在一个油藏块上的微观块渗透率值的空间加权平均。宏观块渗透率是渗透率变异函数、平均块体积及权常数的函数。

1. 渗透率空间平均

考虑微观块上的标量渗透率 $k\ (x)$，它可在三维非均质油藏的任意一点 $x = (u, v, w)$ 处测量获得。在油藏中任意一点 x 处测到的样本值可以看作从与特定位置相关的随机变量所得的结论。对所有的位置 x 的这组随机变量被称为空间随机函数 $k\ (x)$。实际上，空间随机函数由它们的单变量频率分布和其一阶及二阶统计矩来表征。这些数量可以用样本数据来估计。在下面的过程中，考虑了由变量 $Y\ (x) = \ln k\ (x)$ 所得到的 $Y\ (x)$。在这里，假定随机函数 $Y\ (x)$ 是二阶平稳，各态历经的，且是多变量正态的。$Y\ (x)$ 的一阶和二阶矩定义如下：

$$E[Y(x)] = \alpha;\ \mathrm{Var}[Y(x)] = \sigma^2;\ \mathrm{Cov}[Y(x), Y(x+h)] = \sigma(h)$$

及

$$\gamma(\bar{h}) = \sigma^2 - \sigma(\bar{h})。$$

其中，$\sigma(\bar{h})$ 及 $\gamma(\bar{h})$ 分别是 $Y(x)$ 的空间协方差和变异函数。

在这里，宏观块渗透率 K_r 被定义为在宏观块体积 V 上的微观块渗透率的一个空间幂次平均。这样定义的理由在后面给出。于是就有：

$$K_V = [1/V\int_V k(x)^\omega \mathrm{d}x]^{1/\omega},\quad \text{当 } \omega \neq 0 \text{ 时} \tag{9.24}$$

及

$$K_V = \exp[1/V\int_V \ln k(x)\mathrm{d}x] = k_g,\quad \text{当 } \omega = 0 \text{ 时} \tag{9.25}$$

当 ω 分别取 -1，0，$+1$ 时，就可得到空间调和、几何、算术平均。根据地质统计模型，K_V 是一个随机变量，因为它是在 V 内的所有位置 x 上的微观块的随机变量 $k\ (x)$ 的一个函数。从（9.24）式开始，并用二阶近似值 $k(x)^\omega \approx 1 + \omega \ln k(x) + [\omega \ln K(x)]^2/2$，便得到 $\ln k_V$ 的平均值和方差：

$$E[\ln k_V] \approx \alpha + (\omega/2)\bar{\gamma}(V, V) \tag{9.26}$$

及

$$\mathrm{Var}[\ln k_V] \approx \sigma^2 - \bar{\gamma}(V, V) \tag{9.27}$$

此处 $\bar{\gamma}\ (V,\ V)$ 表示块 V 内 $Y\ (x)$ 的平均变差函数：

$$\bar{\gamma}(V, V) = (1/V^2)\int_V\int_V \gamma(|r - s|)\mathrm{d}r\mathrm{d}s \tag{9.28}$$

当偏置向量 $\bar{h}$ 的末端独立地掠过块 V 时，项 $\bar{\gamma}$（V，V）可作为 γ（$\bar{h}$）的平均值用数值的方法计算出来。

假设 $\ln k$ 是正态分布的，就可以用（9.26）、（9.27）两式及对数正态分布的性质来确定 k_V 的期望值：

$$E[k_V] = E[\exp(\ln k_V)] = \exp(E[\ln k_V] + 1/2\mathrm{Var}[\ln k_V])$$
$$\approx k_\omega \exp\{(1/2 - \omega/2)[\sigma^2 - \bar{\gamma}(V,V)]\} \tag{9.29}$$

此式中 k_ω 是对一个对数正态随机变量进行的全体“加权”平均值：

$$k_\omega = E[k(x)^\omega]^{1/\omega} = \exp(\alpha + \omega\sigma^2/2) \tag{9.30}$$

式（9.29）有几个有趣的特征。$E\cdot[k_V]$ 是指数 ω 和宏观块尺寸 V 的一个函数。假如块尺寸缩小到一个点上，则 $\bar{\gamma}(V,V) = 0$，且 $E[k_V]$ 等于呈对数正态分布的微观块渗透率期望值，即 $E[k(x)] = \exp(\alpha + \sigma^2/2)$。当宏观块尺寸变得十分大时，$E[k_V]$ 向全体幂次平均 k_ω 的方向减少。因此，平均宏观块渗透率的均值依赖于所考虑的宏观块尺寸。由于几何空间平均的特殊情况，很早就有文献(Desbarats and Dimitrakopoulos，1990）描述这些重要现象。可将式(9.29) 重写成下面的形式以强调整体几何平均 $k_G = e^\alpha$ 及考虑了几何影响 g 的项：

$$g = [(1-\omega)^{\bar{\gamma}(V,V)}]\sigma^2/2$$

$E[k_V] \approx k_G \exp[\sigma^2(1/2 - g)]$。此式中的几何项 g 由下式确定：

此处，g 包含了指数 ω 和项 $\bar{\gamma}$（V，V），且 $\bar{\gamma}$（V，V）考虑了块 V 的有限平均体积。当体积 V 增加时，

$$\lim_{V\to\infty} g = (1 - \bar{\omega})/2 \tag{9.31}$$

这就提供了一个 ω 和 g 之间的关系。这些相似之处的重要意义虽并未完全被理解，但它表明，空间加权平均可对有效网格块渗透率提供一个有用的平均。

提出空间加权平均的理由是由于经验。经观察发现，可以找到一个权值 ω，使得在砂/泥序列中，用加权平均可以很好地逼近由数值计算所确定的真正的宏观块有效渗透率 k_e。进一步说，目前的研究表明，由（9.24）式所定义的空间加权平均也是 k_e 的一个很好的近似。

2. 具体计算方法

已知幂公式，最关键的问题是求指数 ω。ω 可以用下面的图形方法，对每一组条件用经验来确定。由式（9.26）的偏差知，$\ln k_v$ 与 ω 有近似的线性关系：

$$\ln k_v \approx \omega[\bar{\gamma}(V,V)/2] + \ln k_g$$

上式中，k_g 是在式（9.25）中与 k_V 一致的空间几何平均。$\bar{\gamma}$（V，V）由（9.28）式定义。k_V 是未知的。但由前面的讨论知，块渗透率 k_V 是等效渗透率 k_e 的一个近似值。k_e 可用数值地求解基于达西定律质量守恒的流动方程来求得。因此，可以先计算出 n 个宏观块的 k_e，即

$$k_{e_1}, k_{e_2}, \cdots, k_{e_n},$$

令这 n 个宏观块的块渗透率分别等于它们的等效渗透率，即

$$k_{V_1} = k_{e_1}, k_{V_2} = k_{e_1}, \cdots, k_{V_n} = k_{e_1}。$$

再分别计算出这 n 个宏观块的几何平均值

$$k_{g_1}, k_{g_2}, \cdots, k_{g_n},$$

这样就可以利用（9.31）式分别计算出 n 个 ω 值 ω_1，ω_2，…，ω_n，最后取这 n 个 ω 值的平均值，就得到了全域的 ω 值。

9.4　实例研究和方法比较

利用快速渗透率粗化方法得到的粗化值仅是个近似结果。此结果所产生的粗化渗透率值与真实渗透率值相比较的偏差取决于原始渗透率地质模型和所选用的渗透率粗化方法。当按相对偏差从小到大的标准对粗化方法进行排序时，分类就依赖于所考虑的渗透率地质模型，及在此地质模型内所要获得的渗透率的方向。这一章就要讨论一下如何选择最有效的快速渗透率粗化方法的一些可能的标准。

9.4.1　参考值

快速渗透率粗化方法的目的是，当不可能用数值地或解析地求解流动方程的方法来得到等效渗透率值时，就可用快速粗化方法来获得等效渗透率的值。同时，还可用它来对迭代的渗透率粗化方法进行初始化，其目的就是尽量减少迭代的次数。为了验证所选择的粗化方法，有必要把快速渗透率粗化方法所得到的结果与等效渗透率的真实值进行比较。此真实值本应通过解析地求解流动方程的方法来获得，但在大多数问题中，是无法用解析的方法求解流动方程的。因此，就可以不采用等效渗透率的真实值，而用通过数值方法求解流动方程而得到的等效渗透率值来代替真实值。此等效值接近于真实值，其接近程度还取决于所选择的求解方法。在此，把用通过数值方法求解流动方程而得到的等效渗透率值称为“参考值”。

9.4.2　评价标准

假设每一个渗透率地质模型都有一个等效渗透率值，以及一些要对其进行排序评价的快速渗透率粗化结果。在这里，有以下一些评价标准或比较方法。

1. 比较偏差

比较两个值所采用的最自然的办法就是，把一个值看做参考值，计算另一个值与参考值之间的差，称之为误差。相对误差就是误差和参考值的比值。很明显，在一个单个的渗透率地质模型上比较两个粗化方法，较好的粗化方法就是其相对误差的绝对值较小的那种方法。当测定一种粗化方法时，人们期望找到一种对大多数渗透率地质模型都可产生一个较好的值的粗化方法。这时，就有必要计算误差的统计值。误差的实现期望（实际上就是平均值）叫做偏差。相对偏差的定义是偏差和参考值的数学期望的比值。当一个粗化方法可以产生与参考值相同的结果时，它的偏差就是0。一个好的粗化方法将不会产生大大的相对偏差。但反之，一个产生很小偏差的粗化方法也不一定是一个好的粗化方法。这是因为，一个粗化方法可以产生很大的正的或负的偏差，如果正偏差的平均值与负偏差的平均值在绝对值上是相等的话，那么此粗化方法的偏差就会十分小。在这种情况下，可能会意外地得到很糟的值。很明显，小偏差可以作为一个可能是好方法的标准，但它并不完善。因此，有必要寻找其它的标准来补充或取代它。

2. 比较相关系数

如果对粗化方法的结果进行相同的线性变换，并且所有的渗透率地质模型有一个参考值的话，很显然，就可以很容易地修正粗化方法以获得一个更完美的结果。当在两个变量之间存在这样一种关系时，这些变量之间的相关系数就等于1。假如两变量之间的关系用图形表示近似于一条直线时，相关系数就给出了直线附近点的分散性：当所有的点都位于直线上

时，相关系数就等于 1（完美的线性关系），当没有直线关系时，相关系数就等于 0。当然，相关系数接近于 1 是一个优质粗化方法的标准，但是，相关系数很小并不一定表示此方法是一个糟糕的粗化方法。当两个变量间有非线性关系时，它们间就不是线性关系，此时，可能会发现它们的相关系数很小。然而，也可以运用一个代数变换把一个变量变换成另一个变量。在这种情况下，对相关系数的研究就不可能提供一个好的标准，需要其它的方法来对粗化方法进行评价。

3．比较散点图

对粗化结果和参考值的比较所得到的信息可用散点图进行图形化的表示。在这种图上，可以看到，两个值是否相等，它们间总体上是否有线性关系或非线性关系，或者它们之间是否有不可见关系。假如它们之间有某种关系存在，就可看到在表示它们之间关系的曲线周围的点具有较小的分散性。散点图还包含和相关系数及方差有关的信息。当要对粗化方法和参考值进行比较时，研究散点图是进行这种比较的最通用的方法。

9.4.3　实例研究

下面就是利用 D 油田的 P12－13 和 P210 两小层数据来测试“网格粗化”模块所实现的七种渗透率粗化算法的结果。在这里选取了三组数据。一组数据是 P12－13 小层的数据，粗化前渗透率地质模型的数据网格块的数目是 $50\times50\times10=25000$，对 $2\times2\times1=4$ 个微观网格块进行粗化，粗化后的网格块数目是 $25\times25\times10=6250$。另一组数据是 P210 小层数据，粗化前渗透率地质模型的网格块数目是 $100\times100\times20=200000$，分别进行两种粗化运算，一种是对 $2\times2\times1$ 个微观网格进行合并，粗化后的网格块数目是 $50\times50\times20=50000$，另一种是对 $4\times4\times2$ 个微观网格进行合并，粗化后的网格块数目是 $25\times25\times10=6250$。还有一组数据也是 P210 小层数据，不过粗化前渗透率地质模型的网格块数目是 $50\times50\times20=50000$，对 $2\times2\times1=4$ 个微观块进行合并，粗化后的微观块数目是 $25\times25\times20=12500$。

所用的“储层地质统计分析系统 GASOR”所使用的硬件环境是：SUN SPARC 4 工作站，CPU 速度是 60MIPS（兆指令每秒），32MB 内存。

1．方法的选择

在这里所选择的待比较的粗化方法是本软件中所实现的七种方法。这些方法是已有渗透率粗化技术中最为常见的方法，其中部分方法是较容易实现的，而还有一些方法已被证明是很有效的。表 9.2 列出了这些方法。

表 9.2　所选择的渗透率粗化方法

渗透率粗化方法的名称	表示符号
算术平均	μ_a
几何平均	μ_g
调和平均	μ_h
基于地质统计学的加权平均算法	μ_p
简化重整化的最大范围	c_{max}
简化重整化的最小范围	c_{min}
简化重整化的几何平均	c_{geo}
标准重整化	r_s
基于达西定律质量守恒的求解压力的方法	K_{ref}

其中，最后一种方法 K_{ref}是作为渗透率的参考值，以它作为标准来对其它方法所产生结果的精确度进行衡量。在这些值中 μ_a，μ_g，μ_h，μ_p，c_{geo}，r_s 都是粗化后的最终估计值，而 c_{max}，c_{min}是渗透率的简化重整化算法的“中间值”，之所以称它们为“中间值”，是因为不能把它们当作最终的粗化渗透率。把 μ_a，μ_g，μ_h 这几个用数学平均而得到的值称为代数估计值。

2. 执行时间和计算复杂度的比较

为了比较各算法在执行时间和计算复杂度方面的差异，表 9.3 是每一种粗化算法的平均 CPU 执行时间。

表 9.3　所选择的渗透率粗化算法在 SUN 5 工作站上的平均 CPU 执行时间

渗透率粗化算法	CPU 时间，s
算术平均	45
几何平均	45
调和平均	46
基于地质统计学的加权平均算法	48
简化重整化的几何平均	48
标准重整化	46
基于达西定律质量守恒的求解压力的方法	151

(a) PI2－13 小层，粗化前：50×50×10，粗化后：25×25×10。

渗透率粗化算法	CPU 时间，s
算术平均	308
几何平均	304
调和平均	300
基于地质统计学的加权平均算法	343
简化重整化的几何平均	349
标准重整化	313
基于达西定律质量守恒的求解压力的方法	899

(b) P210 小层，粗化前：100×100×20，粗化后：50×50×20。

渗透率粗化算法	CPU 时间，s
算术平均	89
几何平均	86
调和平均	85
基于地质统计学的加权平均算法	102
简化重整化的几何平均	95
标准重整化	90
基于达西定律质量守恒的求解压力的方法	368

(c) P210 小层，粗化前：100×100×20，粗化后：25×25×10。

续表

渗透率粗化算法	CPU 时间，s
算术平均	76
几何平均	77
调和平均	76
基于地质统计学的加权平均算法	84
简化重整化的几何平均	88
标准重整化	78
基于达西定律质量守恒的求解压力的方法	173

(d) P210 小层，粗化前：50×50×20 粗化后：25×25×20。

由上表可见，最快的算法是几种代数平均算法：算术平均、几何平均以及调和平均。接下来是两种重整化方法：标准重整化和简化重整化的几何平均。基于地质统计学的加权平均算法所耗费的 CPU 时间较长一些。最耗时的算法就是基于达西定律质量守恒的求解压力的方法。

基于地质统计学的加权平均算法之所以较耗时是因为它首先要求解压力方程从而求出幂值 ω。基于达西定律质量守恒的求解压力的方法由于对每组网格块都要求解流动方程，所以最为费时。对快速渗透率粗化算法和基于达西定律质量守恒的求解压力的方法的平均 CPU 执行时间进行比较可以看到，后者的平均 CPU 执行时间几乎是前者的数倍。

从编程的角度来考察一下它们的复杂度。由于要用迭代的方法求解流动方程，基于达西定律质量守恒的求解压力的方法的复杂度最高。基于地质统计学的加权平均算法的复杂度次之。两种重整化方法相对来说较容易编程，而三种代数平均算法的复杂度最小。

从存储分配的角度来看，这些算法都允许对有数百万个微观网格块的三维数据体进行粗化，并可以把数千个微观网格块粗化成一个宏观块。这些情况要视计算机的硬件配置而定，若计算机的内存和硬盘足够大，也可以对上千万个微观网格块数据进行粗化。

3. 相对偏差

在下面几部分，对前面所述的三组具体的三维渗透率地质模型网格数据体运用不同的粗化算法进行粗化，然后对粗化结果进行比较。比较的项目主要有相对偏差，线性相关系数，散点图等。通过对以上这些项目的比较，可以根据不同的标准对快速渗透率粗化技术进行分类。

对一个给定的快速粗化算法，相对偏差 e 可定义如下（Galli and others, 1996）：

$$e = | \overline{K}_{\text{quick}} - \overline{K}_{\text{ref}} | / \overline{K}_{\text{ref}}$$

式中，$\overline{K}_{\text{quick}}$是用快速渗透率粗化技术计算出来的渗透率平均值，$\overline{K}_{\text{ref}}$是粗化渗透率的参考标准值。

表 9.4 分别给出了 P12 − 13 和 P210 小层的粗化渗透率的相对偏差。参考渗透率值是基于达西定律质量守恒的求解流动方程的方法求出的粗化渗透率值。表中的数据已按相对偏差 e 从小到大的顺序排列。

表 9.4　所选择的渗透率粗化算法的相对偏差

序号	渗透率粗化算法	相对偏差 e
1	简化重整化的几何平均 c_{geo}	0.004
2	几何平均 μ_g	0.009
3	标准重整化 r_s	0.020
4	基于地质统计学的加权平均算法 μ_p	0.036
5	简化重整化的最小范围 c_{min}	0.037
6	简化重整化的最大范围 c_{max}	0.055
7	调和平均 μ_h	0.161
8	算术平均 μ_a	0.280

(a) P12 − 13 小层，粗化前：50×50×10，粗化后：25×25×10。

序号	渗透率粗化算法	相对偏差 e
1	简化重整化的最大范围 c_{max}	0.014
2	几何平均 μ_g	0.033
3	简化重整化的几何平均 c_{geo}	0.043
4	标准重整化 r_s	0.064
5	简化重整化的最小范围 c_{min}	0.065
6	调和平均 μ_h	0.168
7	基于地质统计学的加权平均算法 μ_p	0.357
8	算术平均 μ_a	0.560

(b) P210 小层，粗化前：100×100×20，粗化后：50×50×20。

序号	渗透率粗化算法	相对偏差 e
1	简化重整化的最大范围 c_{max}	0.024
2	标准重整化 r_s	0.124
3	简化重整化的几何平均 c_{geo}	0.192
4	简化重整化的最小范围 c_{min}	0.281
5	几何平均 μ_g	0.581
6	调和平均 μ_h	0.637
7	基于地质统计学的加权平均算法 μ_p	0.713
8	算术平均 μ_a	0.924

(c) P210 小层，粗化前：100×100×20，粗化后：25×25×10。

序号	渗透率粗化算法	相对偏差 e
1	简化重整化的几何平均 c_{geo}	0.003
2	标准重整化 r_s	0.015
3	简化重整化的最小范围 c_{min}	0.016
4	简化重整化的最大范围 c_{max}	0.027
5	几何平均 μ_g	0.047
6	调和平均 μ_h	0.096
7	基于地质统计学的加权平均算法 μ_p	0.216
8	算术平均 μ_a	0.244

(d) P210 小层，粗化前：50×50×20，粗化后：25×25×20。

由表 9.4 可以看到，相对偏差的范围从小于 1% 到大于 90%。综合表 9.4 可以看到，最好的算法是标准重整化（$e<13\%$），简化重整化的几何平均（$e<20\%$），次之，基于地质统计学的加权平均算法的相对偏差（$3\%<e<72\%$）相对来说较大。最糟的方法是三种代数平均算法。在这几种代数平均算法中，几何平均是最好的方法（$1\%<e<60\%$），调和平均次之（$10\%<e<64\%$），最差的就是算术平均，其相对偏差从 25% 到 92.4%。

4. 线性相关系数

表 9.5 分别给出了 P12－13 和 P210 小层的粗化渗透率的线性相关系数。参考渗透率值同样是基于达西定律、质量守恒的求解压力的方法求出的粗化渗透率值。表中的数据已按线性相关系数从大到小的顺序排列。

表 9.5　所选择的渗透率粗化算法的线性相关系数

序号	渗透率粗化算法	相关系数
1	几何平均 μ_g	0.934
2	基于地质统计学的加权平均算法 μ_p	0.934
3	简化重整化的最小范围 c_{min}	0.921
4	简化重整化的几何平均 c_{geo}	0.921
5	标准重整化 r_s	0.921
6	简化重整化的最大范围 c_{max}	0.915
7	算术平均 μ_a	0.899
8	调和平均 μ_h	0.893

(a) P12－13 小层，粗化前：50×50×10，粗化后：25×25×10。

序号	渗透率粗化算法	相关系数
1	标准重整化 r_s	0.942
2	简化重整化的几何平均 c_{geo}	0.938
3	简化重整化的最小范围 c_{min}	0.936
4	简化重整化的最大范围 c_{max}	0.936
5	几何平均 μ_g	0.924
6	调和平均 μ_h	0.896
7	基于地质统计学的加权平均算法 μ_p	0.854
8	算术平均 μ_a	0.806

(b) P210 小层，粗化前：100×100×20，粗化后：50×50×20。

序号	渗透率粗化算法	相关系数
1	标准重整化 r_s	0.744
2	简化重整化的几何平均 c_{geo}	0.730
3	简化重整化的最大范围 c_{max}	0.726
4	简化重整化的最小范围 c_{min}	0.675
5	基于地质统计学的加权平均算法 μ_p	0.656
6	算术平均 μ_a	0.628
7	几何平均 μ_g	0.492
8	调和平均 μ_h	0.426

(c) P210 小层，粗化前：100×100×20，粗化后：25×25×10。

续表

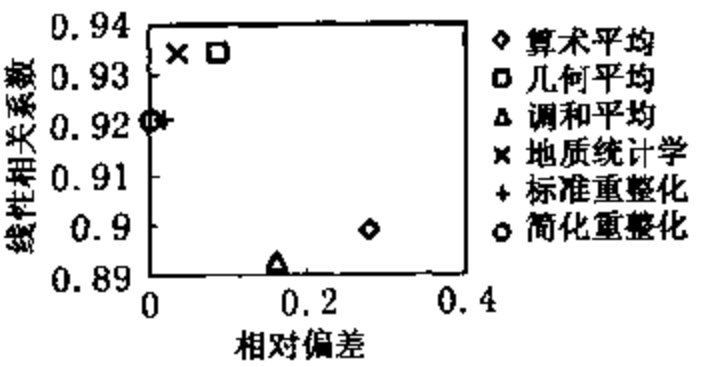
序号	渗透率粗化算法	相关系数
1	标准重整化 r_s	0.969
2	简化重整化的几何平均 c_{geo}	0.968
3	简化重整化的最大范围 c_{max}	0.968
4	简化重整化的最小范围 c_{min}	0.966
5	几何平均 μ_g	0.960
6	调和平均 μ_h	0.944
7	算术平均 μ_a	0.934
8	基于地质统计学的加权平均算法 μ_p	0.931

(d) P210 小层，粗化前：50×50×20，粗化后：25×25×20。

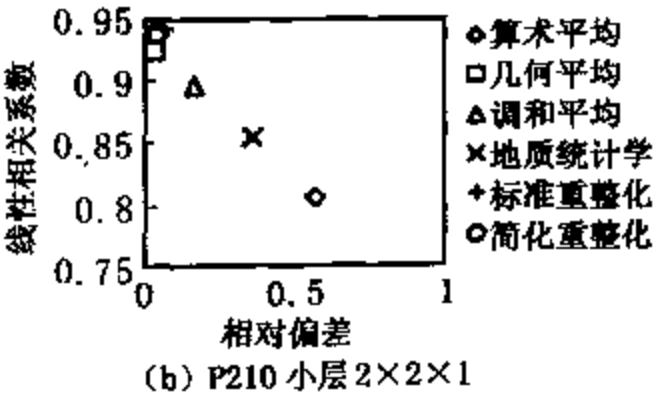
总体上，从表9.5可以观察到标准重整化 r_s 和简化重整化的几何平均 c_{geo} 这些技术具有较好的线性相关系数。基于地质统计学的加权平均算法 μ_p 的线性相关系数相对来说就较小一些，最糟的是几种代数平均方法 μ_a，μ_g，μ_h，它们的线性相关系数最小。

对于这种情况的解释是很显然的。由于代数平均并不依赖于局部渗透率的空间分布，而仅仅依赖于局部渗透率的值，因而两个具有相同局部渗透率值和不同空间分布的介质，将会有相同的代数平均值和不同的粗化渗透率。而重整化方法则考虑渗透率的空间分布，所以具有较好的线性相关系数。

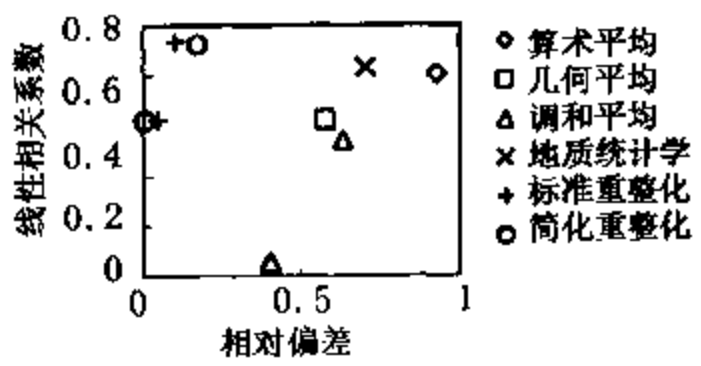
图9.9中是各种渗透率粗化算法的线性相关系数和相对偏差的对比的散点图。在图中，每一个点代表了在精确性参数空间中的一种粗化技术。每个点的坐标就是相对偏差和线性相关系数。因此点越接近左上角，它所代表的粗化算法就越精确。



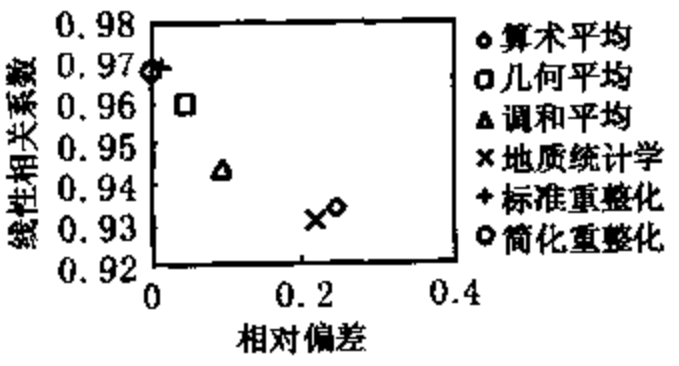
(a) P12-13 小层 2×2×1



(b) P210 小层2×2×1



(c) P210 小层 4×4×2



(d) P210 小层2×2×1

图9.9　各种粗化算法的线性相关系数和相对偏差的比较

由图9.8可看到，最可靠的粗化技术是 r_s 和 c_{geo}。在三种数学平均算法中，几何平均的

精确度较好，基于地质统计学的加权平均法居中。

5. 散点图

图 9.9 是用 P210 小层的数据粗化后所做的散点图。参考渗透率值同样是基于达西定律质量守恒的求解流动方程的方法求出的粗化渗透率值。散点图的坐标是，x 为用快速渗透率粗化技术算出的粗化渗透率，y 为粗化渗透率的参考值 K_{ref}，即基于达西定律质量守恒的求解压力的方法算出的粗化渗透率值。

散点图很清楚地表明了精确的和不精确的粗化技术之间的区别。从散点图中可以看到，r_s 和 c_{geo}同样是一种较好的技术。它们都接近直线 $y=x$。几种数学平均相对来说比较差，而基于地质统计学方法的加权平均居中。更进一步，从散点图我们可以清楚地看到每一种粗化渗透率的理论范围。可以看到，μ_a 的范围主要落在直线 $y=x$ 的下方，而 μ_h 的范围主要落在直线 $y=x$ 的上方，因此有 $\mu_a > K_{ref} > \mu_h$。这也与 9.3.1 节所指出的一致。并且正如前面 9.3.5 节所指出的，可以看到 $c_{max} > c_{geo} > c_{min}$。基于地质统计学的加权平均算法 μ_p 的范围主要落在直线 $y=x$ 的上方，即 $\mu_p < K_{ref}$。这主要是因为 μ_p 是加权平均，与调和平均类似，其值总是小于参考值。

6. 三维渗透率立体图

利用 GASOR 3.0 软件的三维立体图显示模块，可以把粗化前后的渗透率三维立体图显示出来。下面彩图 25～32 就是一组渗透率的三维立体图。所选用的数据是 DQ 油田的 P210 小层的数据。粗化前的网格数是 100×100×20，粗化后的网格数是 50×50×20。

在测试的七种粗化技术中，基于达西定律质量守恒的求解流动方程的粗化结果是作为标准参考渗透率的。在这些技术中，三种数学平均算法虽然 CPU 执行时间最短，且程序的复杂度低，但在每一次不同的测试中都显示出了明显的不精确性。因此，它们不能作为主要的渗透率粗化算法。基于地质统计学的加权平均类似于几何平均，它的各项性能接近于或高于几何平均。

最为精确的粗化方法就是基于达西定律质量守恒的求解压力的方法。但它要对每一组网格块求解流动方程组，因此程序复杂度最高，存储要求也大。若要处理大的数据体，则要求计算机的性能较好。且它的 CPU 执行时间最长。

较好的方法就是两种重整化方法。它们是对前两类方法的折衷。在各个不同的测试中，两种重整化方法都表现了较好的性能。它们的 CPU 执行时间不长，程序复杂度和存储要求也适中。它们的各种指标都比较令人满意。因此，它们可以作为主渗透率粗化算法。

在本章研究中，所使用的各种渗透率粗化方法都认为渗透率是各向同性的，所以每种粗化方法都假定在水平和垂直三个方向上渗透率是相等的，即 $K_x = K_y = K_z = k$。在每种渗透率粗化方法中，都只采用了渗透率的一个标量值 k。这是本次研究的一个主要的不足。由于地层的非均质性及流体受到的各种不同的压力，渗透率在各个方向上是不同的，即 $K_x \neq K_y \neq K_z$。因此，在对渗透率粗化之前，原始渗透率网格数据体就要包含有三个不同方向的渗透率。在进行渗透率粗化时，要考虑三个方向的渗透率的不同和地层的非均质性，可以使用三维的全张量渗透率粗化方法（Galli and others, 1996; Almeida and others, 1996; Pickup and other, 1994; Aasum and others, 1993; Lee and others, 1995, Peaceman, 1996; Guerillot and other, 1995; Lee and others, 1994; and Thibeau and others, 1995）。

在许多较完善的粗化方法中，不但要考虑渗透率在不同的方向上有不同的值，而且还要考虑一个宏观块周围的介质对它的影响，也就是要考虑此宏观块的边界的流动对它的影响。

这是因为一个网格块总是处于一个介质之中，它不是孤立的，它的流动要受到周围流动的干扰。在本文所研究的方法之中，都是把一个宏观块当作一个孤立的块来考虑，没有考虑边界流动带来的影响。要把这些因素考虑进去，一个可能的办法就是使用完全渗透率张量的粗化方法。Tegnander 和 Gimse（1998）提出利用油藏数值模拟的一些结果，来评价渗透率粗化的方法。其中包括累计原油产量、在某几个位置处的含油饱和度、井中压力等。

有意义的工作应放在研究和实现完全渗透率张量且考虑边界条件影响的粗化方法上。最终目标是使粗化后的三维渗透率地质模型尽量保持其粗化前的非均质性。

参 考 文 献

1 Abrahamsen, P..Random walk and quantum mechanics.Master thesis, University of Oslo, 1988

2 Aasum Y., Kasap E., Kelkar M..Analytical Upscaling of Small－Scale Permeability Using a Full Tensor.SPE 25913, 1993

3 Abu－elbashar O., Daltaban T., Wall C.and Archer J., Application of Analytical methods in predicting waterflood performance of reservoir with stochastic sand bodies, 2nd European conference on the Mathematics of Oil－Recovery, Paris, 65～73, 1990

4 Alabert F..The practice of fast conditional simulations through the LU decomposition of the covariance matrix, Math Geology, 19 (5), 369～386, 1987

5 Almeida J.A., Soares A., and Pereira M.J..Upscaling of Permeability: Implementation of A Conditional Approach to Improve, The Performance in Flow Simulation.European 3－D Reservoir Modeling Conference, 16～17, April, Stavanger, Norway, SPE 35490, 1996

6 Anderson J..An introduction to multivariate statistical analysis.New York: John Wiley & Sons, 1958

7 Armstrong M..Improving the estimation and modeling of the varogram.Geostatistics for Natural Resources Characterization, Part I, 1～19, 1984

8 Augedal H.O., Stanley K.O.and More H..SISABOSA, a program for stochastic modeling and evaluation of reservoir geology.Proc., Conference on Reservoir Description and Simulation with Emphasis on EOR, Oslo, Institute for Energy Technology, 1986

9 Barker J.W., Thibeau S., A Critical Review of the Use of Pseudo Relative Permeabilities for Upscaling, European 3－D Reservior Modelling Conference.16～17 April, Stavanger, Norway, SPE 35491, 1996

10 Bashor W.and Araktingi U..The importance of the geological model for reservoir characterization using geostatistical techniques and the impact on subsequent fluid flow.SPE 26474, 601～616, 1993

11 Beier R.and Hardy H..Comparison of 2D and 3D factual distribution in reservoir simulations.SPE Reservoir Engineering, 195～200, August, 1994

12 Berteig V., Halvorsen K.and H.Omre H..Prediction of hydrocarbon pore volume with uncertainties.SPE 18325, 633～643, 1988

13 Boufassa A.and Armstrong M..Comparison between different kriging estimators.Math Geology, Vol 21, No.3, 1989

14 Bridge J.and M.R.Leeder M..A simulation model of alluvial stratigraphy.Sedimeuto logy,26, 617～644, 1979

15 Brooker.P..Two－dimensional Simulation by turning bands.Math.Geology, Vol.17, No.1, 1985

16 Bui－Tran V., Pajon J－L.and Joseph P..1991, three－dimensional reservoir visualization.SPE (22301), 129～138, 1991

17 Christakos G..On the problem of Permissible covariance and variorum models.Water

Resources Research. Vol.20, No, 2, 251~265, February, 1984

18 Christakos.G..Stochastic simulation of spatially correlated geo - processes.Jour.Math Geology, Vol.19, No.8, 807~831, 1987

19 Christie M.A., Mansfield M., King P.R., Baker J.W.and Cuiverwell 1.D..A Renormalisation - Based Upscaling Technique for WAG Floods in Heterogeneous Reservoirs.SPE 29127, 1995

20 Clementsen R., Hurst A.R., Knarud R.and Omre H..A Computer Program for Evaluation of Fluvial Reservoirs.North sea Oil & Gas Reservoirs I1, A.t.Buller, et al. (eds.), London: Graham & Trotman, 373, 1990

21 Clemetsen R., Hurst A., Omre H.and H.Knarcid H..A computer program for evaluation of fluvial reservoirs: Proceedings of International Conf..On North Sea Oil and Gas Reservoirs, Norwegian Inst.of Technology, Trondheim, 1989

22 Christie M.Mansfield M., King P., Bake J.and Cuiverwell 1.A renormalisation - based upscaling technique for WAG floods in heterogeneous reservoir.SPE 29127, 1995

23 Cox D.and Miller H..Theory of stochastic processes.London: Methuen, 1965

24 Damsleth C., Tjølsen, Omre H.and Haldorsen H..A two stage stochastic model applied to a North sea reservoir.SPE 20605, 1990

25 Davies D., William 's B.and Vessel R., Reservoir geometry and internal permeability distribution in fluvial, tight, gas sandstone' s, Travis peak formation.Texas: SPE Reservoir Engineering, 7~12, February, 1993

26 de Fouquet, Beucher H., Galli, A.and Ravenne C..Conditional simulation of random sets.Application to an argilaceous sandstone reservoir, M.Armstrong (ed.) .Geostatistics, Vol.2, 517~530, 1989

27 Desbarats A.J.and Dimitrakopoulos R..Geostatistical Modeling of Transmissibility for 2D Reservoir Studies.SPEEF (Dec.) 437~443; *Trans*., AIME, 235, 1990

28 Deutsch C.and Journel A..GSLIB : geostatistical software library and users guide, New York: Oxford University Press, 1992

29 Deutsch C..Direct assessment of local accuracy and precision..the 5th International Geostatistics Congress, WollonG' 96, 22~27, September, 1996

30 Deutsch C., and Journel A..The application of simulated annealing to stochastic reservoir modeling.Advanced Technology series, SPE, Richardson, TX April, 222, 1994

31 Deutsch C.and Srinivasan S..Improved reservoir management through ranking stochastic reservoir models, SPE Symposium on Improved Oil Recovery.Apr 21 ~ 24, Society of Petroleum Engineers (SPE), 105~113, 1996

32 Deutsch C.V.and Wang L..Quantifying object - based stochastic modeling of fluvial reservoirs.Mathematical Geology, 827, 857, 1996

33 Dimit R., 陈霞 . 油藏描述与生产预测的随机方法 . 国外石油地质, (3), 33~42, 封三, 1998

34 Dimitrakepoulos R..Conditional simulation of intrinsic random function of order k.Math Geology, Vol.22, No.3, 1990

35　Dimitrakopoulos R. and a Desbarats A.. Geostasistical modeling of gridlock permeability's for 3D reservoir simulators. pages 13～18, SPE Reservoir Engineering, February, 1993

36　Dimitrakopoulos R. and Desbarats A.J..Geostatistical Modeling of Gridblock Permeabilities for 3D Rescrvoir Simulators.SPE Reservoir Engineering, 13－18, February, 1993

37　Ding L., Mebra R.and Dannelly J..Stochastic modeling in reservoir simulation.SPE (18431), 303～320, 1989

38　Doob J..Stochastic Processes.New York: John Wiley & Sons, 1953

39　Doud P..A review of recent developments in geostatistics.Computers & Geosciences, Vol.17, No 10, 1481～1500, 1982

40　Doyen and Guidish T..Seismic discrimination of lithology: a Bayesian approach.In Geostatistis Symposium, Calgary, AB, May, 1990

41　Doyen P.and Psaila D..Basin sequeatial indicator simulation of channel sands from 3D seismic data in the Oseberg tiled, Norwegian North Sea.SPE 28382, 197－211, 1984

42　Doyen P..Porosity from seismic data: A geostatistical approach.Geophysics, 1263～1275, October, 1988

43　Dubrule O.and Haldorsen H..Geostatistics for permeability estimation.Reservoir Characterization, 233～247, 1986

44　Dubrule O..A review of stochastic models for petroleum reservoirs.Geostetistics, Voll.2, 493～506, 1989

45　Duquerroix J－P.L., Lemouzy P., and Noetinger B., Romeu R.K..Influence of the Permeability Anisotropy Ratio on Large－Scale Properties of Heterogeneous Reservoirs.SPE 26612, 1993

46　Egeland T., Georgsen F., Knarud R.and Omre H..Multifacies Modeling of Fluvial Reservoirs.Paper SPE 26502 presented at the 1993 SPE Annual Technical Conference and Exhibition, Houston, Oct.3～6, 1993

47　Ekrann S., Dale M., K.Langaas K., and Mykkeltveit J..Capillary Limit Effective Two Phase Properties for 3D Media.SPE 35493, 1996

48　Emanuel A., Alameda G., Behrens R.and Hewlett T..Reservoir performance prediction methods based or factual geostatistics.SPE 16971, 405～418, 1987

49　Feller W..An introduction to probability theory and its applications, Vol.II, New York: Wiley, 1966

50　Fielding C.and R.C.Crang R..An application of statistical modeling to the prediction of hydrocarbon recovery factors in fluvial reservoir sequences.The Society of Economic Paleontologists and Mineralogists, 321～327, 1987

51　Franken P., König D., Arndt U.and Schmidt V..Queues and point processes.Akademie～Verlag, Berlin and J.Wiley & Sons, Chichester, 1981

52　Galli A., Goblet P., Griffin D., Ledoux E., Le Loc' h G., Mackay R.and Renard P..Technical Report : Quick Upscaling of Flow and Transport Related Parameters.Ecole des Mines de Paris Centre de Géostatistique Centre d 'Informatique Géologique Fontainebleau, France, University of Newcastle Water Resource Systems Research Unit Newcastle Upon Tyne,

England, May 20, 1996

53 Galli A.and Armstrong M..Petroleum geostatistic technical report, Center de Getatistique, Fontainebleau, 1996

54 Gao H., Wang J., and Zhao P..The updated Kriging variance and optimal sample design.Mathematical Geology, Vol.28, No.3, 1996

55 Gao H., Zhao P., Wang J.And Galli A..Updating of population parameters and credibility of discriminant analysis.Mathematical Geology, Vol.31, No.3, 1999

56 Georgsen F.and Omre H..Combining Fibre Processes and Gaussian andom Functions for Modeling Fluvial Reservoir.Geostatistics Troia '92, Proceedings of the 4th International Geostatistical Congress, Soares. A. (ed.), Kluwer Academic Publishers, DorDrecht, 425, 1993

57 German S.and D.German D..Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (6), 721~741, November, 1984

58 Geyer C.and Thompson E..Constrained Monte Carlo maximum likelihood for dependent data.J.R.Stat.Soc.B, 54 (3): 657~699, 1992

59 Gómez-Hernandez J.and Srivastava R..ISIM3D: an ANSI-C three dimensional multiple indicator conditional simulation program.Computers & Geosciences, 16 (4), 395~440, 1990

60 Gómez-Hernandez J.and Journel A..Stochastic characterization of grideblock permeability.SPE Formation Evaluation, June, 1994

61 Governs P..Spatial orthogonality of the principal components computed from regionalized variables.Math.Geology, Vol.25, No.3, 1993

62 Guérillot D.R.and Sophie Verdier S..Different Pressure Grids for Reservoir Simulation in Heterogeneous, SPE 29148.

63 Gundcso.R.and Egeland O..SESIMIRA - a new geological tool for 3D modeling of heterogeneous reservoirs.North sea Oil & Gas Reservoirs II, A.t.Buller, et al. (eds.), London: Graham & Trotman, 363, 1990

64 Hϕiberg J., Omre H.and Tjelmeland H..Large scale barriers in extensively drilled reservoirs.Proceedings of European Conf, On the Mathematics of Oil Recovery, Arles, 1990

65 Haldorsen H.and Damsleth E..Stochastic modeling.JPT, 404~412, 1990

66 Haldorsen H.and Lake L..A new approach to shale management in field scale simulation models.SPE 10976, 1~8, 1984

67 Haldorsen H.and MacDonald C..Stochastic modeling of underground reservoir facies.SPE 16517, 575~ 588, 1987

68 Hard J., Yang C.and Chopra A..Ability of geostatistical simulations to reproduce geology: A critial evaluation.SPE 28414, 533~545, 1994

69 Hasting W..Monte Carlo samping methods using Markov chain sand their applications. Biometrika, 57 (1), 97~109, 1970

70 Hektoen A.and Holden L..Bayesian modeling of sequence stratigraphic bounding surfaces.The 5th International Geostatistics Congress, Wollong' 96, 22~27, September, 1996

71　Hektoen A.－L.and Holden L..Bayesian modeling of sequence stratigraphic bounding surfaces.Geostatistics Wollongong '96: Proceedings of the 5th International Geostatistical Congress, E.Baafi, (ed.), Kluwer Academic Publishers, Dordrecht, 1997

72　Hewlett, T..Fractal distributions of reservoir heterogeneity and their influence on fluid transport.SPE 15386, 1986

73　Hewlett T.and Behrens R..Conditional simulation of reservoir heterogeneity with fractals, SPE 18326, 1998

74　Hird K.and Kelkar M..Conditional simulation method for reservoir description using spatial and well－performance constraints.SPE Reservoir Engineering, 145～152, May, 1994

75　Hirst J.P.P., Blackstock C.R.and Tyson S..Stochastic Modeling of Fluvial Sandstone bodies, The Geological modelling of Hydrocarbon Reservoirs and Outcrop Analogues.S.S.Flint and I.D.Bryant (eds.), Special Publications of the International Association of Sedimentologists 15, 237, 1993

76　Hjort N.and Omre H..Topics in spatial statistics, Proceedings from 13, Nordic Conference in Mathematical Statistics, Odessa, Denmark, 11 － 15, June, 1990

77　Holden L..Make decisions using stochatic models, EAGE/SPE International Symposium on Petroleum Geostatistics.20～23, April 1999, Toulouse, France, 1999

78　Holding S..3D geoscience modeling: computer techniques for geological characterization, Springer－Verlay Beilin Heidelberg, 1994

79　Huber P..Robust Statistics.New York: John Wiley & Sons, 1981

80　Isaaks and Srivastava R..An Introduction to Applied Geostatisties.New York: Oxford University Press, 1989

81　Isaaks E., and Srivastava.R..Spatial continuity measures for probability and deterministic geostatistics Mach.Geology, 20 (4), 313～341, 1988

82　Jaedong Lee J., Kasap E., and Mohan G..Kelkar M.G..Development and Application of a New Upscaling Technique.SPE 30712.1995

83　Johnson C.and Jones T..Putting geology into reservoir simulations: A three－dimensional modeling approach.SPE (18321), 585～594, 1988

84　Jones T..Problems in using geostatistics for petroleum applications.Geostatistics for Natural Resources characterization, Part 2, 651～674, 1984

85　Journel A.and Huijbregts C..Mining Geostatistics.New York: Academic Press, 1978

86　Journel A..Geostatistics for conditional simulation of forebodes.Economic Geology, 69, 673～680, 1974

87　Journel A..The deterministic side of geostatistics.Math.Geology, 17 (1), 1～16, 1985

88　Journel A.and Alabert F..Focusing on spatial connectivity of extreme－valued attributes: stochastic indicator modes of reservoir heterogeneity 's.SPE 1832, 1988

89　Journel A.and Alahert F..New method for reservoir mapping. JPT, 212～218, February, 1990

90　Journel A.and Gomez－Hernandez J..Stochastic imaging of the Wilmington clastic se-

quence.SPE Formation Evaluation, 33～40, March, 1993

91 Journel A.and Isaaks E..Conditional indicator simulation: application to a Saskatchewan uranium deposit. Math Geology, 16 (7), 655～718, 1984

92 Journel A.and Posa D..Characteristic behavior and order relations for indicator variograms.Mach.Geology, Vol.22, No.81, 1990

93 Journel A.and Rossi M..When do we need a trend model in kringing? . Mach.Geology, 715～750, 1989

94 Journel A.and Zhu H..Integrating soft seismic data: Markov－Bayes updating, an alternative to cokriging and traditional regression.Stanford: In Report 3, Stanford Center for Reservoir Forecasting, CA, May, 1990

95 Journel A..The place of non－paramatic geostatistics.In G.Verily et al., editors, Geosta.for Natural Resorces characterization, 307～355, Reidal, Dordrecht, Holland, 1984

96 Journel A..Geostatistics: Models and tools for the earth sciences.Math.Geology, Vol.18, No.1, 1986

97 Journel A..Fundamentals of geostatistics in five lessons, Volume 8 short course in geology.Washington, D.C.: Ameriean Geophysical Union, 1989

98 Journel A..Non－parametric estimation of spatial distributions.Mach.Geology, 15 (3), 445～468, 1989

99 Journel A..Deterministic geostatistics: A new visit. The 5th International Geostatistics congress, Wollong '96, 22～27, September, 1996

100 Journel A..Past, Present and Future of Petroleum Geostatistiocs.EAGE/SPE International Symposium on Petroleum Geostatistics, Toulouse, France, 1999

101 Karatzas I.and Shreve S..Brownian motion and stochastic calculus.Springer－Verlag, 1988

102 Keijzer J.and Kortekaas T..Comparison of deterministic and probabilistic simulation models of channel sands in the Statfjord reservoir, Brent field.SPE 20947, 255～261, 1990

103 King M., and Scher H..Probahilistic Stability Analysis of Multiphase Flow in Porous Media.at the 60th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers held in Las Vegas, NV September 22～25, SPE 14366, 1985

104 King P..The connectivity and conductivity of overlapping sand bodies.North Sea Oil and Gas Reservoirs Ⅱ, 1990

105 König D., Matthes K., Nawrotzki K..Verallgemeinerugen der Erlangschen und Engsetschen Formeln.Berlin: Akademie－Verlag, 1967

106 LeBlane R.Sr..Distribution and continuity of sandstone reservoirs Part 1.JPT, 776～792, July, 1997

107 Lee J., Kasap E., Kelkar M.G..Analytical Upscaling of Permeability for Three－Dimensional Grid Block.SPE 27875, 1994

108 Leloc'h G.and Galli A..Truncated plurigaussion method: theoretical and practical points of view.The 5th International Geostatistics Congress, Wollong '96, 22～27, September, 1996

109 Lenormand R., Kalaydjian F., Bieber M－T.and Lombard J－M..Use of a Multifractal Approach for Multiphase Flow inHeterogeneous Porous Media: Comparison With CT－Scanning Experiment.at the 65th AnnualTechnical Conference and Exhibition of the Society of PetroleumEngineers held in New Orleans, LA, September 23～26, SPE 20475, 1990

110 Lucia F., and Fogy G..Geologic/stochastic mapping of heterogeneity in a carbonate reservoir.JPT, 1298～1303, October, 1990

111 MacDonald A., HΦye T., Lowry T., Jacobean P., Aasen J.and Grindheim A..Stochastic flow unit modeling of a North Sea coastal－deltaic reservoir.124/FIRST BREAK, Vol.10, No.4, April, 1992

112 MacDonald A.C., Falt L.M.and Hektoen A.－L.Stochastic modeling of Sequence Boundary Geometries. American Association of Petroleum Geologists Bulletin (Accepted for Publication), 1997

113 MacDonald A.C., Høye T.H., Lowry P., Jacobsen T., Aasen J.O.and Grindheim A.O., Stochastic Flow Unit Modeling of a North Sea Coastal－deltaic Reservoir.First Break 10, 124, 1992

114 Mantoglou A., Digital simulation of two and three dimensional stochastic processes with a spectral turning bands method.Math.Geology, Vol.19, No.2, 1987

115 Martinez R., Cornish B.and Seymour R..An Integrated approach for reservoir description using seismic, borhole and geologic data.SPE 19581, 99～599, 1989

116 Matheron G..The theory of regionalied variables and its applications.Occludes Mines, Fontaineblear, France, 1971

117 Matheron, G.The intrinsic random functions and their applications, A dunces in Applied Probability, 5, 439～468, 1973

118 Matheron, G..The internal consistency of models in geostatisties.M.Armstrong (ed.) Geostatistics, Vol.1, 21～38, 1989

119 Matheron G., Beucher H., de Fouguet C., Gall A., Guerillal P.and Ravenne C..Conditional Simulation of the geometry of fluridelta ic resole.Dallas, Texas: SPE Annual Conference, 591～589, 1987

120 Matter C.and Dalton R..Reservoir simulation.JPT, 692～695, June, 1990

121 Matt es K..Stotionäre zufällige Punktolgen.Jahresbricht Deutche Math, I ' erein, 66～79, 1963

122 Matropolis N., Rosenbluth A.Rosenbluth M., Teller A.and Teller E..Equation of state calculation by computing machines.J.of Chemical Physics, 21: 1087～1092, 1953

123 McCarthy J..Reservoir Characterization: Efficient Random－Walk Methods for Upscaling and Image Selection.at the SPE Asia PacificOil and Gas Conference and Exhibition held in Singapore, 8～10 February, SPE 25334, 1993

124 Miall A..Artichectral－Element Analysis: A new method of faces analysis applied to fluvial deposits.Earth Science Reviews, 22, 261～308, 1985

125 Myers E..To be or not to be...stationary? That is the question.Math.Geology, 347～362, 1989

126 Nass A., Falt L.M., Corregidor J.and Lagreid U..3D－models from outcrop analogues: Fluvial 3D architecture of the Escanilla Fm., Spanish Pyrenees, Proc..Dallas: American Association of Petroleum Geologists 1997 Annual Convention, April 6～10, 1993

127 Neyman J..On a new class of contagious distributions applicable in entomology and bacteriology, AMS 10, 35～57, 1939

128 Nybraten G., Skolem E.and ø stby K..Reservoir Simulation of the Snorre Field.London: Graham & Trotman, North sea Oil & Gas Reservoirs II, A.t.Buller, et al. (eds.), 103

129 Omre H..The variogram and its estimation, Geostatistics for Natural Resources Characterization.Pare I, 107～125, 1984

130 Omre H.and K Sølna..Stochastic modeling and simulation of fault zones, Proceedings and codate Conf.On Geomastics and Geostatistics

131 Omre H.Sølna and H.Tjelmeland..Simulation of calcite cemented barriers in the Troll field.SAND/18/1989, December, 1989

132 Omre H., Halvorse K.and Berteig V..A Bagesian approach to kriging, Armstrong, M. (ed.), Geostatistics, Vol.1, 109～126

133 Olea R..Geostatistical glossary and multilingual dictionary.Oxford: Oxford University Press, 1991

134 Panda N.and Lake L..Parallel simulated annealing for stochastic reservoir modeling.SPE 26418, 9～18, 1993

135 Peaceman D.W..Effective Transmissibilities of a Gridblock by Upscaling－Why Use Renormalization? .SPE 36722, 1996

136 Philip M.and Watson D..Matheronian geostatisti－quo vadis? .Math.Geolog, Vol.18, No.1, 1986

137 Pickup G.and Sorbie K..Development and application of a new two－phase scale up method based on tensor permeability 's.SPE 28586, 217～230, 1994

138 Pickup G.E.and Sorbie K.S..Development and Application of a New Two－Phase Scaleup Method Based on Tensor Permeabilities.SPE 28586, 1994

139 Posa D..Limiting stochastic operations for stationary spatial processes. Math.Geology, Vol.23, No.5, 1991

140 Qiu Y., P.X, and J.Xiao..Fluvial sandstone bodies as hydrocarbon reservoir in lake basin.The society of Economic of Paleontologists and Mineralogists

141 Ravenne R., Eschard A., Gall A.and Matthew J..Heterogeneity 's and geometry of sedimentary bodies in a fluvial－deltaic reservoir.SPE Formation Evaluation, June, 1989

142 Renyi A..Remarks on Poisson process. Studia Sci.Math.Hung.2, 119～123, 1967

143 Richardson J., Harris G., Rosier R.and Van Hee G..The effect of small, discontinuous shale' s on oil recovery.JPT, 1531～1537, October, 1997

144 Ripley B..Stochastic Simulation.New York: John Wiley & Sons, 1987

145 Ripley B..Statistical inference for spatial processes.New York: Cambridge University Press, 1988

146　Ross S..Stochastic processes.John Wiley & Sons, Inc., 1983

147　Silverman B..Density Estimation for statistics and date analysis.New York: Chapman and Hall, 1986

148　Sneider R., Tinker C.and Mocked L..Deltaic environment reservoir types and their characteristics.JPT, 1538－1545, October, 1977

149　Srivastava R..Philip and Watson－quo vadunt? .Math.Geology, Vol.18, No.1, 1986

150　Spritzer F..Principles of randomWalk.New Jersey: Nostrand, 1964

151　Stanley K.O., Jorde K., Rastad N.and Stockbridge C.P..Stochastic Modeling of Reservoir Simulation, Snorre Field, Northern North sea, Norway.North sea Oil & Gas Reservoirs II, A.t.Buller, et al. (eds.), 91 London: Graham & Trotman, 1990

152　Stoyan D.and Stoyan H..Fractals, random shapes and point fields.John Wiley & Sons, 1994

153　Stoyan W., Kendall W.and Make J..Stochastic Geometry and its Applications.New York: John Wiley & Sons, 1987

154　Suro－Perez V.and Journel A..Indicator principal component kriging. Math. Geology, 759～788, 1991

155　Switzer P..Non－stationary spatial covariance's estimated fro m monitoring data. M.Armstrong (ed.), Geostatistics, Vol.1, 127～138, 1989

156　Syversveen A.and Omre H..Marked point models for facies units conditioned on well data.The 5th International Geostatistics Congress, Wollong '96 22～27 September, 1996

157　Syversve en A.and Omre H..Conditioning of marked point processe s within a Bayessian framework.Scandinavian Journal of Statistics, 1997

158　Thibeau S., Barker J.W., Souillard P..Dynamical Upscaling Techniques Applied to Compositional Flows.SPE 29182, 1995

159　Tegnander C. and Gimse T.. Flow simulation to evaluate upscaling of permeability.Math.Geology, Vol.30, No.6, pp 717～731, 1988

160　Tipper J..Correlation in mosaic environments.Journal of Geological Society, Vol. 147, 487～494, 1990

161　Tyler J., Tarald Saves and Henriquez A..Heterogeneity modeling used for production simulation of lunar reservoir, SPE Formation Evaluation, 85～92, June, 1994

162　Tyler K., Henriquez A.and Svanes T..Modeling Heterogeneities in Fluvial Domains: a Review of the Influence on Production Profiles.Stochastic Modeling and Geostatistics : Principles Methods and Case Studies, J.M.Yarus, and R.L.Chambers (eds.), American Association of Petroleum Geologists Computer Applications in Geology 3, 77～90, 1995

163　Van B., Pajon J.and Joseph P..3D reservoir visualization, JPT, 1310～1314, November, 1991

164　Wang J., and MacDonald A..Modeling channels arcgitecture in a densely drilled oilfield in east China.the 1997 SPE Annual Technical Conference and Exhibitionheld in San Anrtio, SPE 38678, 1997

165 Wang J., and Zhang T..Three－stage Stochastic Modeling method to Characterize Reservoir with Fluvial Facies.paper SPE 29965 presented at the 1995 SPE International Meeting on Petroleum Engineering, Beijing, P.R.China, November 14～17, 1995

166 Wang J., Zhang T.and Huang C..The simulation of braided channels in two dimensions with random walk model.Proc.of The 30th Int' t Geol.Congr., Vol.25, 115～124, 1997

167 Weber K.and B.V.Meats choppy, B.V.How heterogeneity affects oil recovery.Reservoir characterization, 487～544, 1986

168 Weber, K.and Geun, L..Framework for construes classic reservoir simulation models.SPE 19582, 1989

169 Weber K.and Van Genus L..Framework for constructing clastic reservoir simulation models.JPT, 1248～1253, October, 1990

170 Wolcott D.and Chopre A..Incorporating reservoir heterogeneity with geostatistics to investigate waterflood recoveries.pages 26～32, SPE Formation Evaluation, March, 1993

171 Yarus J.and Chambers R..Stochastic modeling and geostatistics.Tulsa, Okla.U.S.A.: Published by the American Association of Petroleum Geologists

172 Zhu H.and Journel A..Mixture of populations.Math.Geology, Vol.23, No.4, 1991

173 陈丽华、王家华、李应暹、田崇鲁等编著．油气储层研究技术．北京：石油工业出版社，2000

174 陈霞，荆克尧译．油藏描述与生产预测的随机方法．国外石油地质，1998，33～42，封三

175 陈希孺，王松桂．近代回归分析——原理方法及应用．安徽：安徽教育出版社，1987

176 陈永生．油田非均质对策论．北京：石油工业出版社，1993

177 陈月明．油藏数值模拟基础．山东：石油大学出版社，1989，1～5

178 杜启振，侯加根．储层随机建模综述．世界石油科学，(5)，1997

179 高海余，赵鹏大，王家华．确定勘探井位的识别概率和伪熵标准．第30届国际地质大会论文集（25卷）．北京：地质出版社，1999

180 郭福星、戴俭华．概率论与数理统计．上海：上海科学技术出版社，1991，112～142

181 纪发华，熊琦华．序贯指示建模方法在枣南油田储层非均质研究中的应用．石油学报，15（增），1994

182 李龙滟，张振顺．用随机模拟方法对井间砂体预测．石油与天然气地质，20（2），1999，179～181

183 林壬子，张金亮．陆相储层沉积学进展．北京：石油工业出版社，1996.2

184 吕晓光，王德发，姜洪福．储层地质模型及随机建模技术．大庆石油地质与开发，第19卷，第1期，2000

185 穆龙新，贾文瑞．油藏早期评价阶段储层建模技术的发展方向．石油勘探与开发，21（5），1994

186　穆龙新，贾爱林，陈亮，黄石岩．储层精细研究方法．北京：石油工业出版社，2000.3

187　聂昌谋，纪发华．储层参数随机建模方法在胡状集油田储层非均质研究中的应用.断块油，3（1），1996

188　隋军，吕晓光，赵翰卿，许运新．大庆油田河流—三角洲储层研究．北京：石油工业出版社，2000

189　钱敏平．随机过程引论．北京：北京大学出版社，1990

190　秦同洛，李汤，陈元千．实用油藏工程方法．北京：石油工业出版社，1989，40～45

191　裘怿楠．中国陆相碎屑岩储层沉积学的进展．沉积学报，No.3，1982，16～24

192　裘怿楠．储层沉积学研究工作流程．石油勘探与开发，No.1，1990a，86～90

193　裘怿楠．烃类储层地质模型的建立，参加第十三国际沉积大会专题报告．江苏油气 No.4，1990b，2～7

194　裘怿楠．储层地质模型．石油学报，No.4，1991，55～62

195　裘怿楠．储层描述中的几个问题．江苏油气，No.1，1993，2～9

196　裘怿楠．裘怿楠石油开发地质文集．北京：石油工业出版社，1997a

197　裘怿楠．储层沉积学展望，石油学会 2000 年展望学术会议上的报告，1987.7．裘怿楠，石油开发地质文集，石油工业出版社，1997b

198　裘怿楠．石油开发地质方法论（Ⅰ），（Ⅱ），（Ⅲ），石油勘探与开发，No.2－4，1996c

199　裘怿楠．油藏精细描述技术——油田开发领域中一项基础技术的发展动向．裘怿楠，石油开发地质文集，北京：石油工业出版社，1997d

200　裘怿楠，陈靖海．陆相沉积与油藏．裘怿楠，石油开发地质文集．北京：石油工业出版社，1997e

201　裘怿楠．储层沉积相研究现状．中国科学院、石油工业部科技协作交流材料，1982.12.27．裘怿楠，石油开发地质文集，北京：石油工业出版社，1997f

202　裘怿楠，薛叔浩等．油气储层评价技术．北京：石油工业出版社，1994

203　裘怿楠，陈子琪．油藏描述．北京：石油工业出版社，1996a

204　孙洪泉．地质统计学及其应用．北京：中国矿业大学出版社，59～69，1990

205　文健，王军等．埕岛油田馆陶组上段储集层随机模型．石油勘探与开发，25（1），1998，69～72

206　王家华，高海余，周叶．克里金地质绘图技术——计算机的模型算法．北京：石油工业出版社，1999

207　王家华，张团峰．辫状河道的二维随机游走模型模拟．第 30 届国际地质大会论文集（25 卷），数学地质与地质信息，北京：地质出版社，1999

208　王仁铎，胡光道．线性地质统计学．北京：地质出版社，1989

209　王仁铎，胡光道．线性地质统计学．北京：地质出版社，1984

210　王学仁，王松桂．实用多元统计分析．上海：上海科学技术出版社，1984

211　王志章，石占中等．现代油藏描述技术．北京：石油工业出版社，1999.12

212　魏宗舒，吕乃刚（译）．统计思想．上海：上海翻译出版公司，1987

213 吴胜和，熊琦华等编著．油气储层地质学．北京：石油工业出版社，1998

214 吴胜和，熊琦华等．储层建模中随机模型的地质适用性分析．石油地球科学文集，北京：石油工业出版社，1998

215 吴胜和，金振奎，黄沧钿，陈崇河．储层建模．北京：石油工业出版社，1999

216 徐宁余．随机模拟在石油地质研究中的应用与面临的挑战．石油实验地质，21（增刊），1999，72～75

217 俞启泰．河流相、三角洲相渗透率平面分布数值抽象地质模型．新疆石油地质，20（4），314～319，1999

218 张金亮，沈凤，赖志云等．早期油藏地质研究及油藏表征．西安：西北大学出版社．1993.12

219 张金亮，赵英，赖伟庆．河间油田砂岩油藏剩余油分布研究．西安：陕西科学技术出版社，1995.12

220 张团峰，王家华．随机几何学及其在地质沉积相方面的应用．第3届中国工业与应用数学学会论文集，清华大学出版社，1994

221 张团峰，工家华，景平，阎汉杰．三维储层随机建模与随机模拟技术研究．中国数学地质（7），1996

222 张团峰，王家华．储层表征中的随机模拟算法．西安电子科技大学学报，Vol.22，1995

223 张团峰，王家华．储层评价中基于变异函数的模拟方法．中国数学地质（6），1995

224 张团峰，王家华．定量评价储层特征空间分布的不确定性．西安石油学院科技大会论文集，西安：陕西科技出版社，1996

225 张团峰，王家华．利用随机模拟提高油藏数值模拟的效果．西安：西安石油学院学报，Vol.11，No.3，1996

226 张团峰，王家华．试论克里金估计与随机模拟的本质区别．西安：西安石油学院学报，Vol.12，No.2，1997

227 张团峰，王家华．油气储层随机模拟的基本原理．测井技术，No.4，1995

228 张团峰．关于回归函数核估计的叠对数律．纯粹数学与应用数学，Vol.12，No.2，1996

229 张团峰．关于回归函数核估计的联合渐近正态性．西安石油学院学报，Vol.8，No.3，1993

230 张团峰．关于回归函数核估计的正态逼近速度．应用数学学报，Vol.14，No.3，1991

231 张一伟．熊琦华，纪发华．地质统计学在油藏描述中的应用．山东：石油大学出版社，1992

232 张一伟，熊琦华等．陆相油藏描述．北京：石油工业出版社，1997

233 张永贵，陈明强．模拟退火组合优化算法在油气储层随机建模中的应用．西南石油学院学报，19（3），1997

234 张永贵，李允．储层地质统计随机模拟．石油大学学报，22（3），1998

235 赵鹏大，王家华，高海余．钻探位置的最优化和空间抽样概率．第30届国际地质

大会论文集（25 卷），数学地质与地质信息，北京：地质出版社，1999

236　赵永胜．储层三维地质模型难使数值模拟摆脱困境．石油学报，22（3）.113～116，119，1998

237　钟宝荣，李龙滟等．储层随机建模与条件模拟的计算机实现．油气成藏机理及油气资源评价国际研讨会，北京：石油工业出版社，1996

后　记

我和张团峰副教授有幸一起工作了十年，其成果写成了《油气储层随机建模》。这是继《克里金地质绘图技术》之后，我们关于地质统计学在石油勘探开发方面应用的第二本论著。我们希望这本论著的出版，有利于我国油藏描述技术和储层随机建模技术的发展。同时，由于我们的水平有限，该书一定会有许多错误。我们期望着广大的读者对该书提出批评和指正。

由于国内外油气田开发的需要，储层随机建模发展得很快，正在进入一个全面发展的时代。在继续探索不同沉积特征储层的沉积相空间分布的随机建模方法的同时，利用各井中资料分析沉积相带空间分布的传统的基本格局正在被突破。随机地震反演、随机深度转换、随机建模方法在分析断层系统和裂缝系统方面的应用，等等，已是储层随机建模的新发展。一个集成系统将会包括地质模型、断块模型、裂缝模型、波形估计和地震反演模型、流体流动模型和油藏风险评价模型。近来，人们又提出了油藏描述和油藏数值模拟的双向迭代的研究课题。油藏描述的结果会影响油藏数值模拟的结果，已是人所共知的了。反过来，人们也可以从油藏数值模拟的结果中，去按某一种标准对油藏描述的结果进行排队，并从中选优。这样，对油藏描述和油藏数值模拟两者的目的及其相互关系的认识就比较全面了。

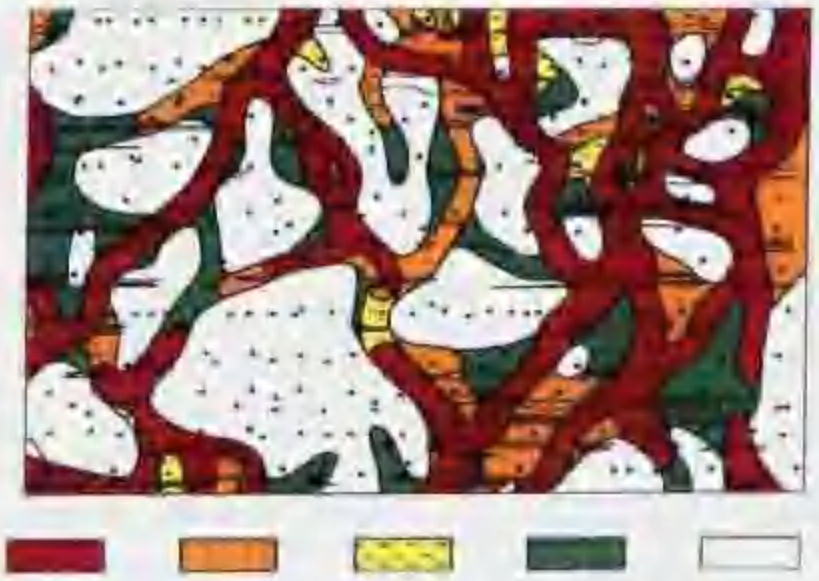
外国公司的储层随机建模软件系统纷纷在国内登场。有的软件系统的功能涵盖了全部油藏描述的内容，储层随机建模只是它的一个部分。有的软件系统则是专门提供储层随机建模的功能。有的软件系统含有较多的随机建模的模型，有的则只有少数一、二种。丰富多彩的三维图像显示和处理的软件功能，被越来越多的储层随机建模的软件系统所采用。

随着我国改革开放的形势的深入，在国内越来越多的油田开始熟悉、研究和使用储层随机建模这项先进的技术。一大批有水平的论著和论文相继出版、发行。我们认为，在发展储层随机建模的过程中，软件开发、方法研究和生产问题的解决三者应当有机地结合起来。在软件开发方面，应很好地向国外软件学习。许多国内专家在看了国外软件后，对齐全的软件功能和逼真生动的地质图件十分感慨。我们坚信，只要认真明确地找出在技术和管理方面和国外的差距，迎头赶上，国产软件受到大家欢迎的日子一定不会很遥远的。在对国外随机建模方法广泛深入了解的基础上，方法研究应从解决国内油气田实际问题的需要出发，以针对中国储层研究的难点和热点问题为主。例如，陆相储层非均质性严重的问题，剩余油的空间分布的问题，精细的时间单元的划分引起的问题，多个随机建模的实现的最佳选择问题，等等。解决油气田开发中的生产问题，是发展储层随机建模的最终目的，是促进软件开发和方法研究最大的动力。这个问题也是各级油气生产部门的领导最关心的。

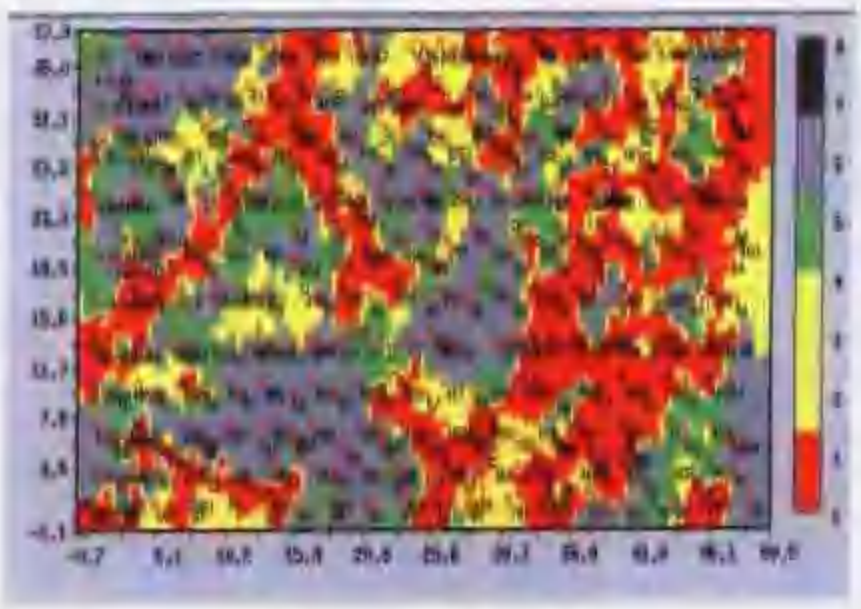
正如裘怿楠教授在本书序言中指出的那样，油气储层随机建模技术，是储层地质工作本身高新技术化的体现。促进和发展储层地质工作的高新技术化的繁重而光荣的任务放在我们大家的面前。我们愿意和广大从事储层地质学、油藏描述和油藏数值模拟的朋友们、同行们相互精诚团结合作，为解决我国的复杂油气开发生产问题，发展我国的石油工业，共同努力。

王家华

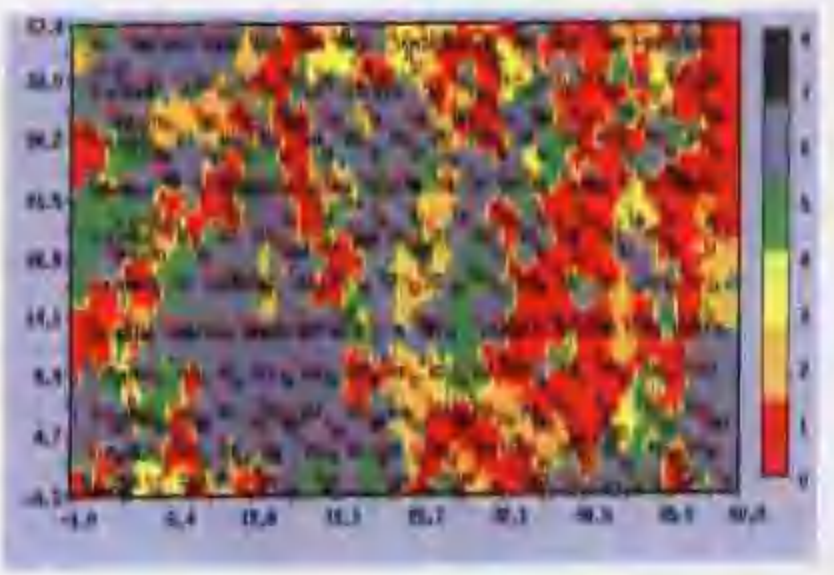
2000 年 9 月



彩图1 SS油层组第8小层枝状三角洲前缘沉积微相分布图



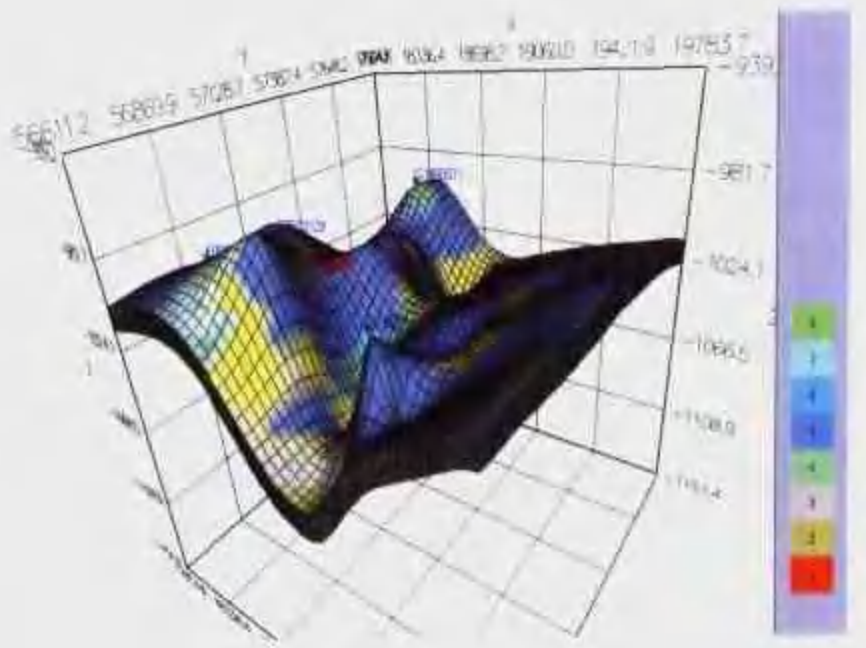
彩图2 用指示主成分模型得到SB区SS油层组第8小层的沉积



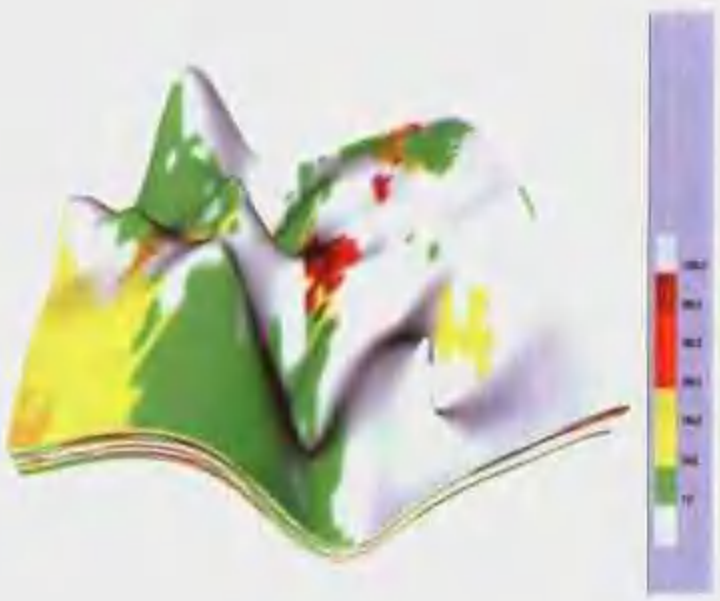
彩图3 加点后用指示主成分模型得到SB区SS油层组第8小层的沉积微相



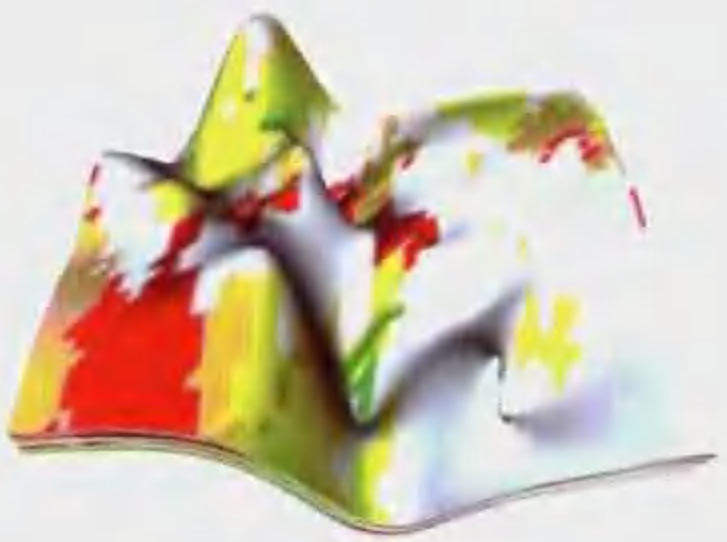
彩图4 BY区西区P2油层组第10小组的沉积微相的三维显示图



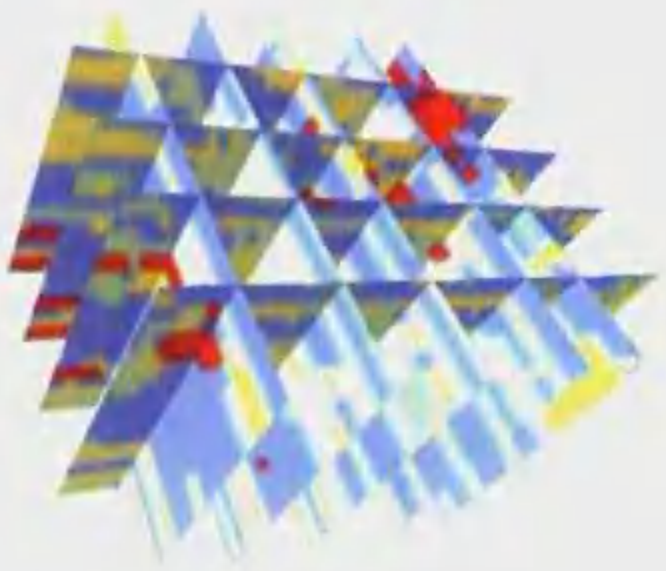
彩图5　BY区西区P2油层组第10小组的沉积微相的三维显示（有坐标系）



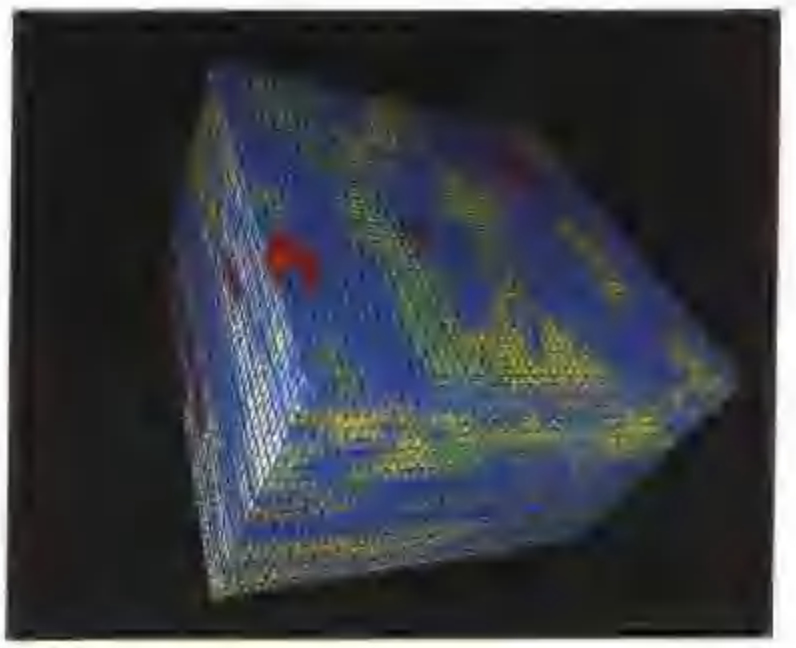
彩图6　BY区西区P2油层组第10小组的渗透率的三维显示



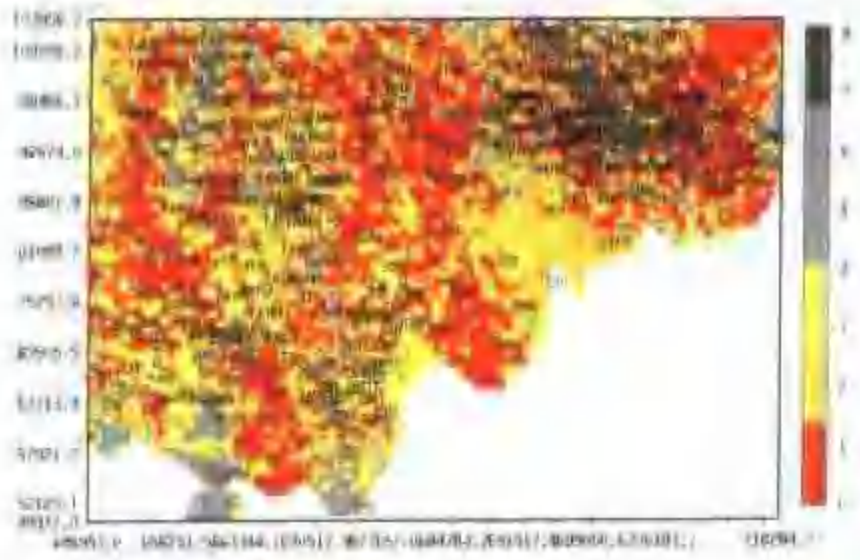
彩图7　BY区西区P2油层组第10小组的孔隙度的三维显示



彩图8　BY区西区P2油层组第10小组的沉积微相删状图



彩图9 BX16西区P2油层组第10小组的流程微相三维显示图（无构造）



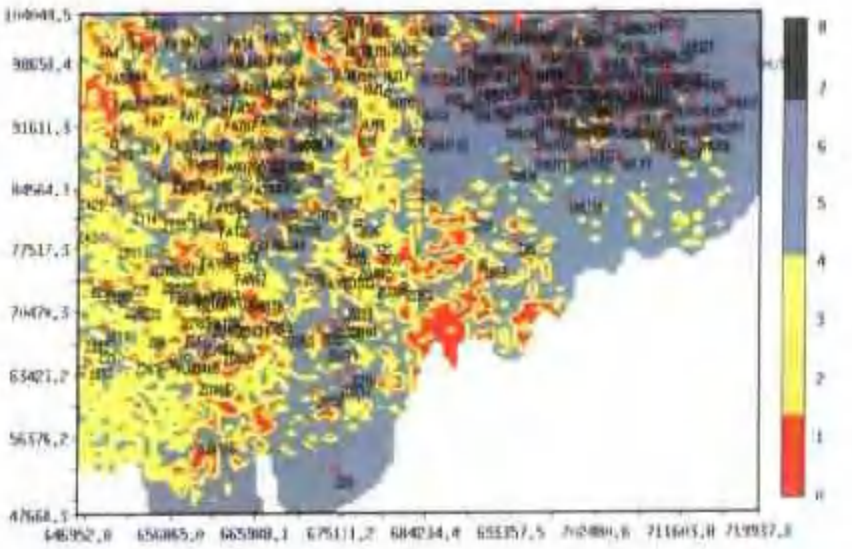
◀彩图10 外围区的第2油层组的沉积微相分布图（用不同的模型参数）



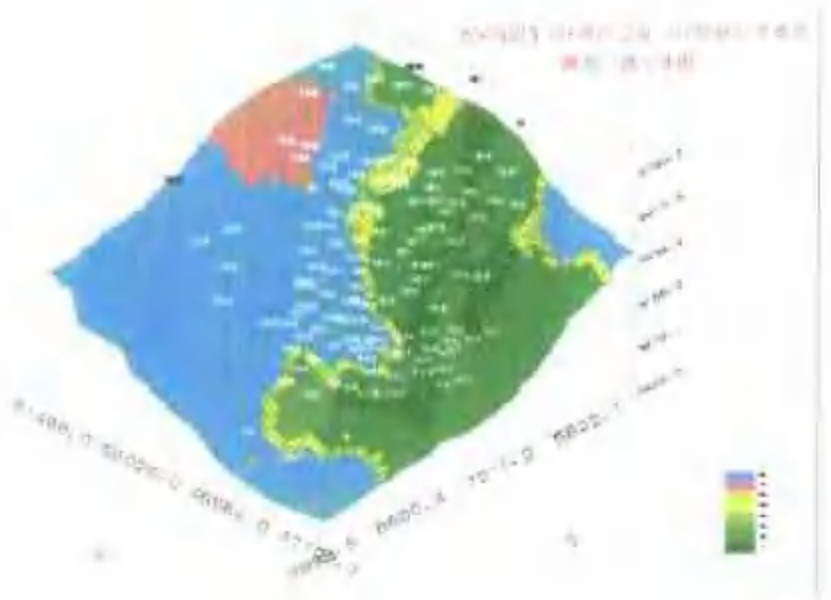
彩图11 外围区的第2油层组的沉积微相分布图（用不同的模型参数）▶



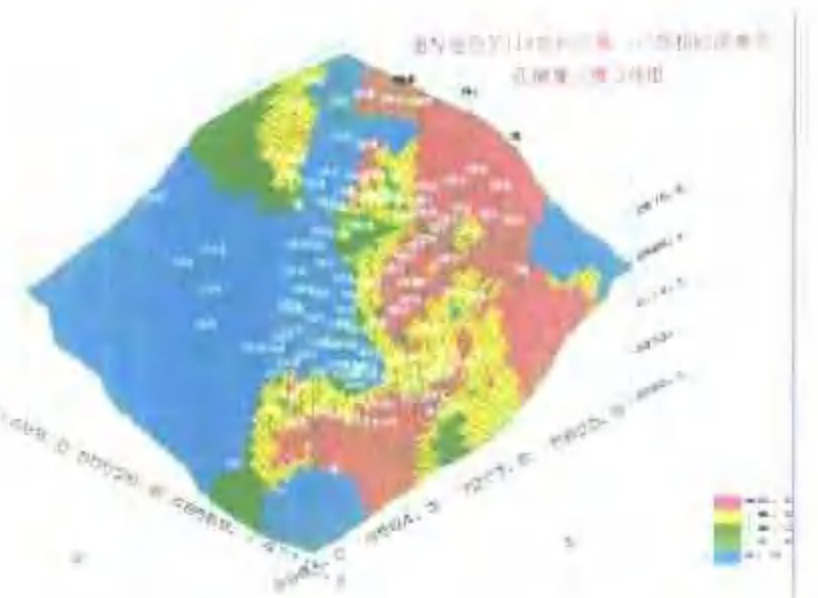
◀彩图12 外围区的第3油层组的沉积微相分布图（用不同的模型参数）



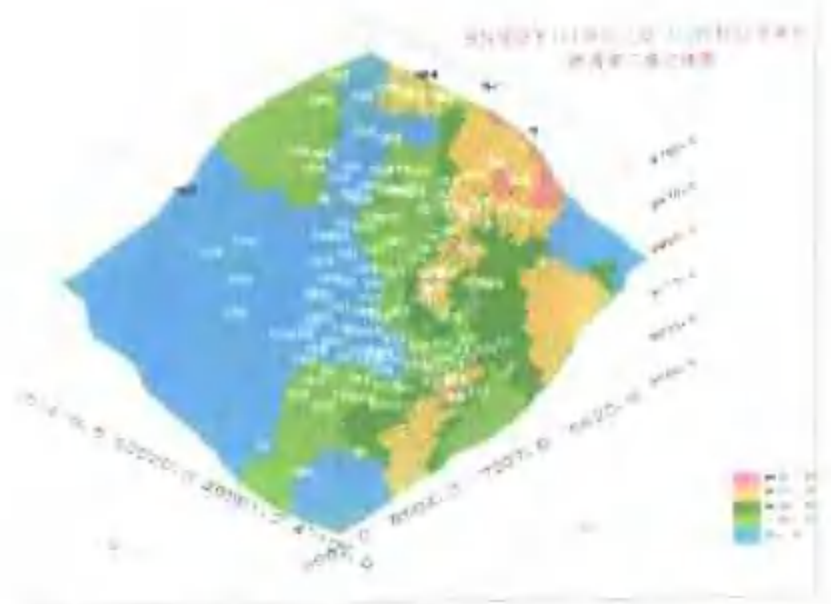
彩图 13 外围区的第 5 油层组的沉积微相分布图 ▶



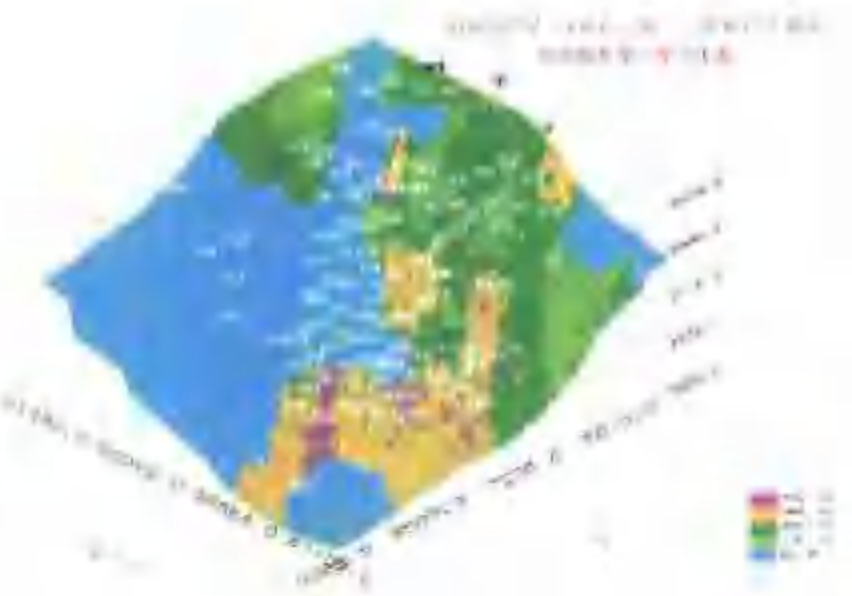
◀ 彩图 14 沉积时间单元 $3^{1^{7}}$ 的微相空间分布



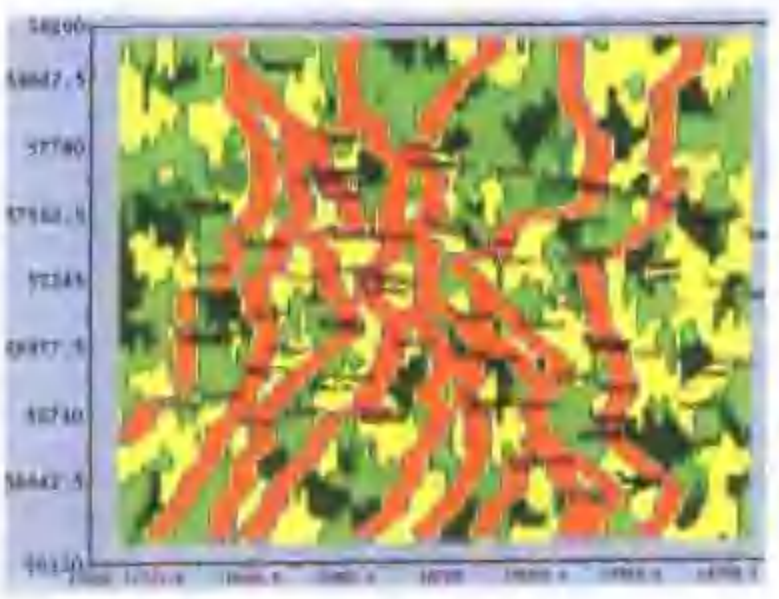
彩图15 沉积时间单元$3^{1^{7}}$的孔隙度空间分布 ▶



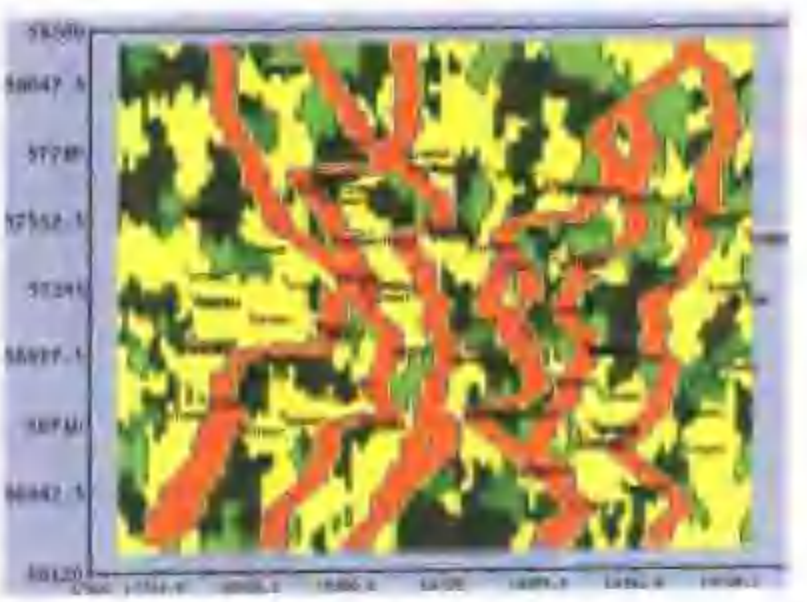
◀ 彩图 16 沉积时间单元 $3^{1^{7}}$ 的渗透率空间分布



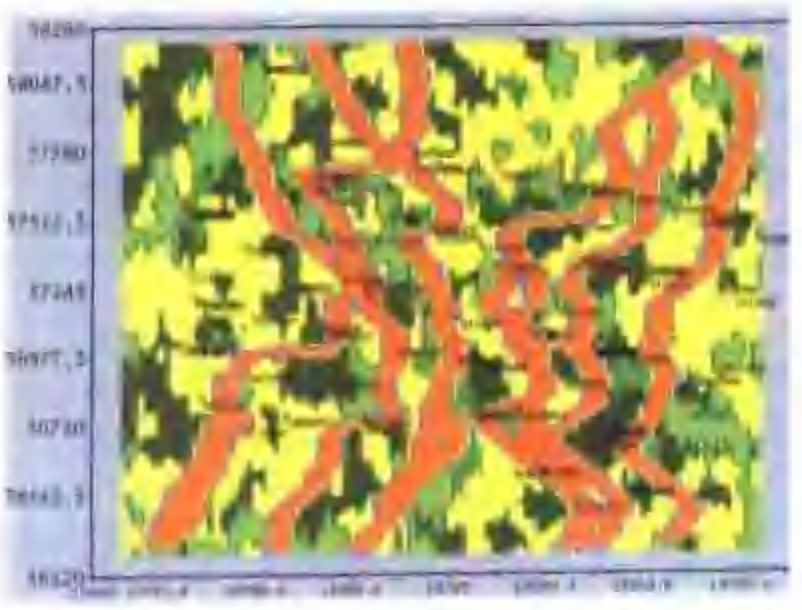
彩图17 沉积时间单元3¹的含油饱和度空间分布 ▶



◀ 彩图18 利用随机游走模型做出的BY区西区SS油层第9小层的水下分流河道沉积相图



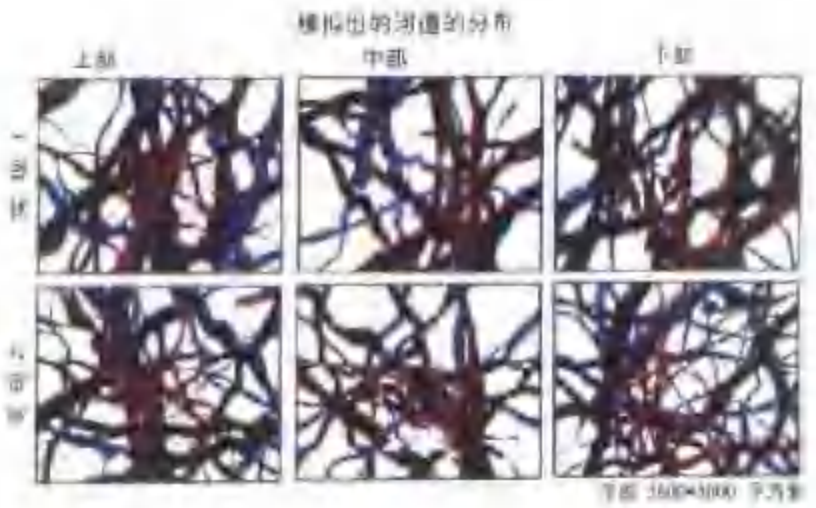
彩图19 利用随机游走模型做出的BY区SS油层第10小层的水下分流河道沉积相图 ▶



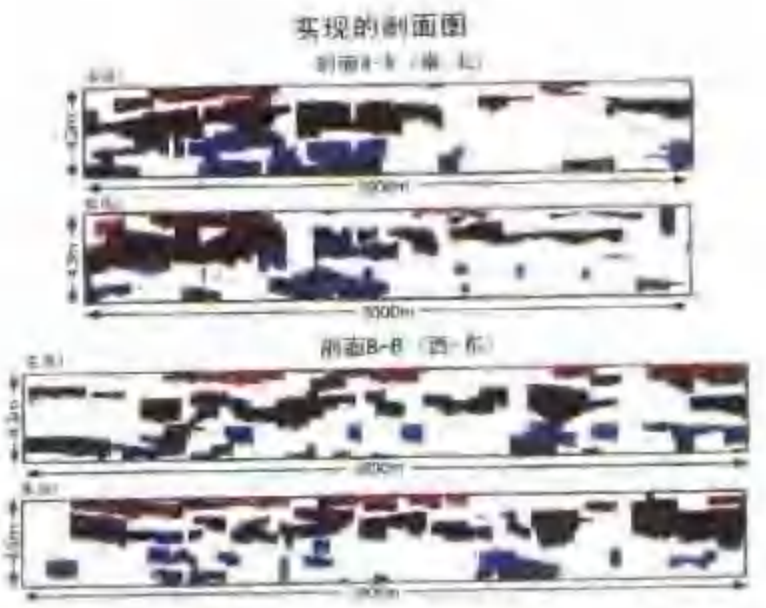
◀ 彩图20 利用随机游走模型做出的BY区西区SS油层第10小层的水下分流河道沉积相图（不同实现）



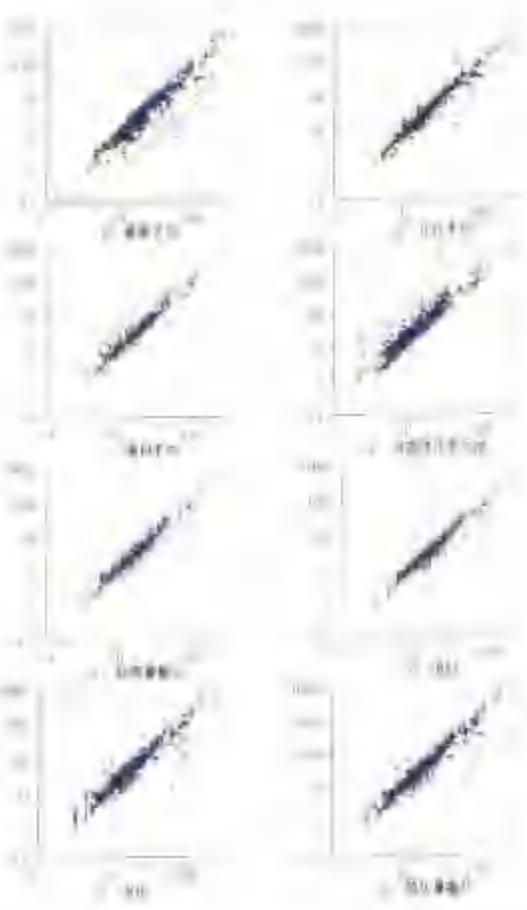
彩图21 河道几何形态的逼真的变化 ▶



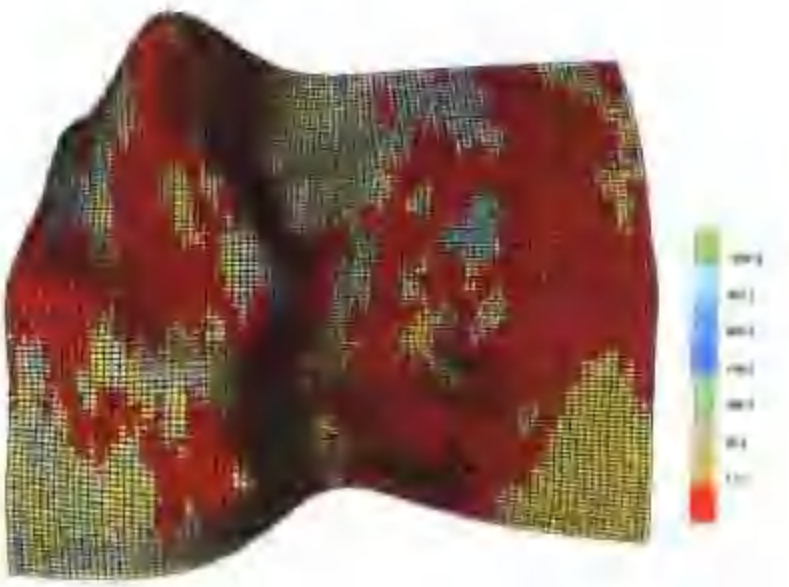
◀ 彩图22 上、中、下三个亚组的河道结构平面图



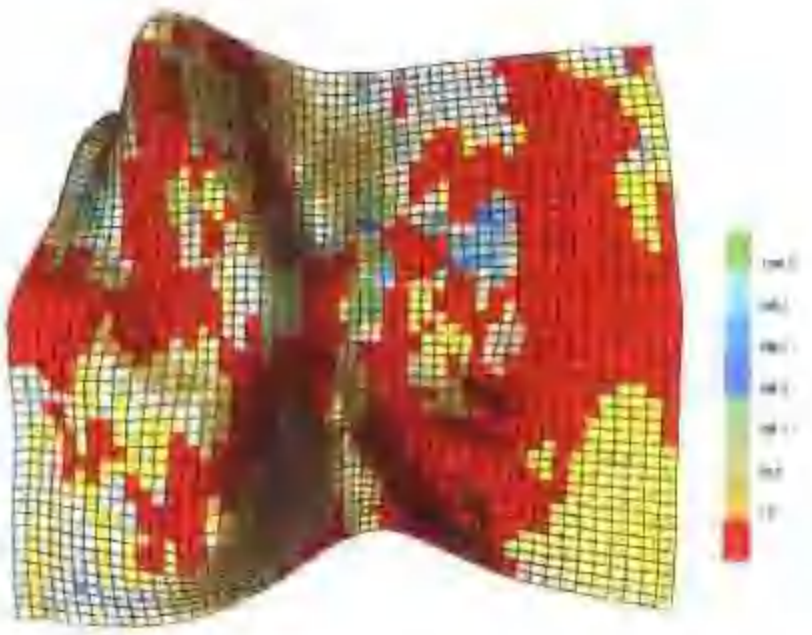
彩图23 河道砂体实现的纵剖面图 ▶



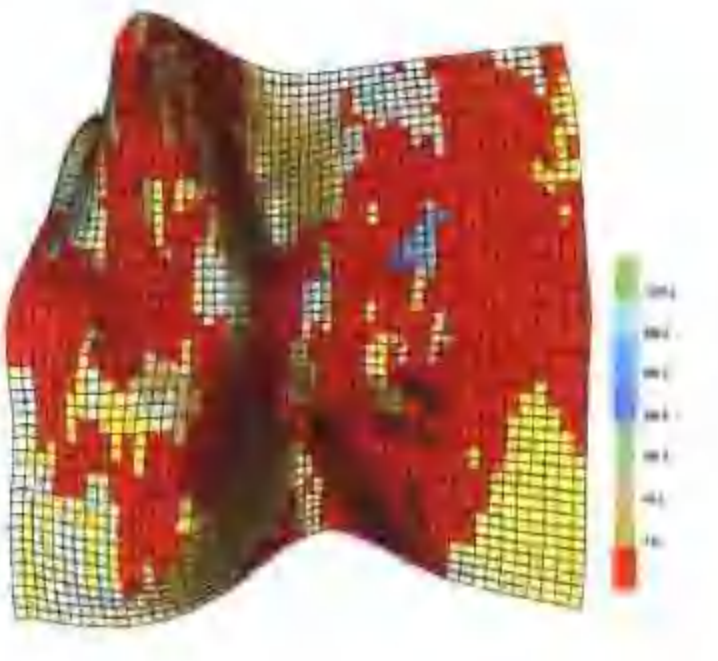
◀ 彩图24 P210小层粗化渗透率对比散点图
粗化前后网格数目分别为100 × 100 × 20和50 × 50 × 20



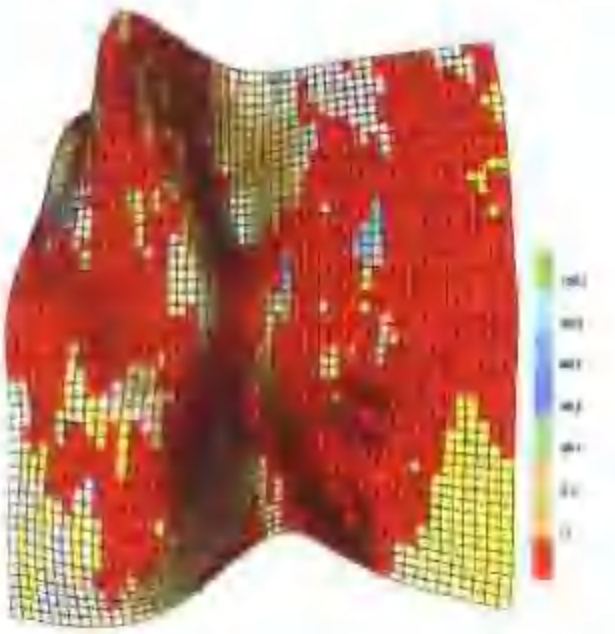
彩图25 P210小层粗化前的渗透率(100 × 100 × 20)



彩图26 P210小层的算术平均粗化后渗透率(50 × 50 × 20)

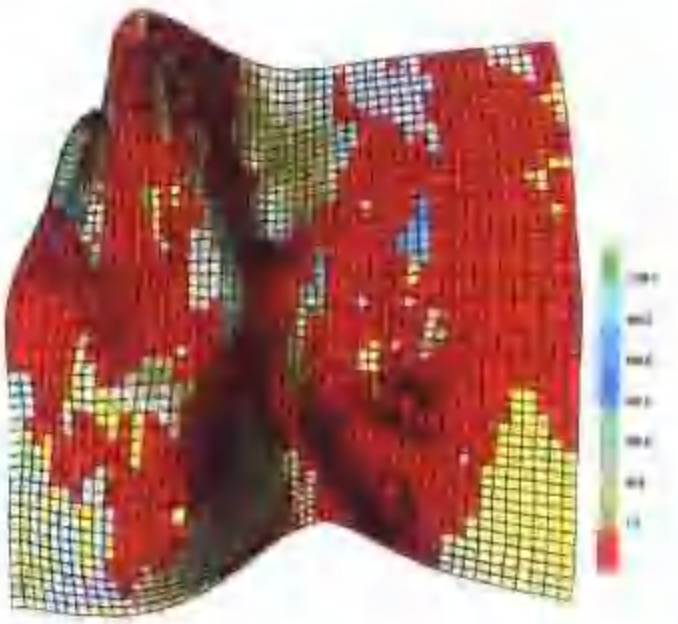


彩图27 用几何平均粗化后渗透率(50 × 50 × 20)

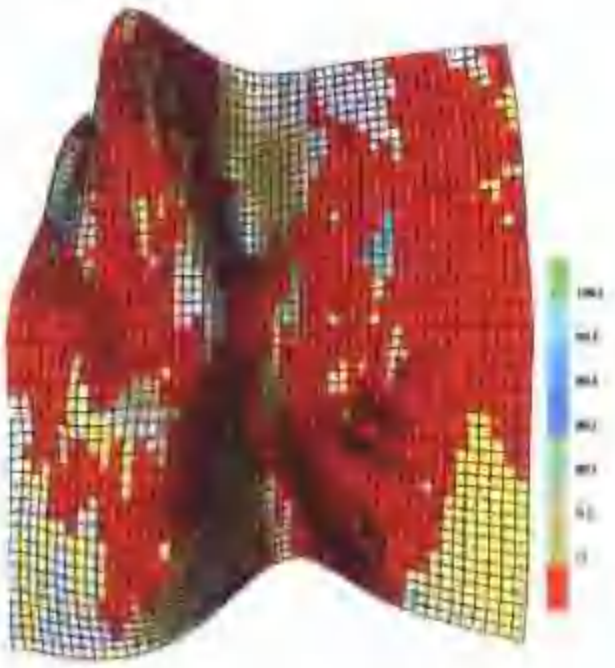


彩图28 用调和平均粗化后的渗透率(50 × 50 × 20)

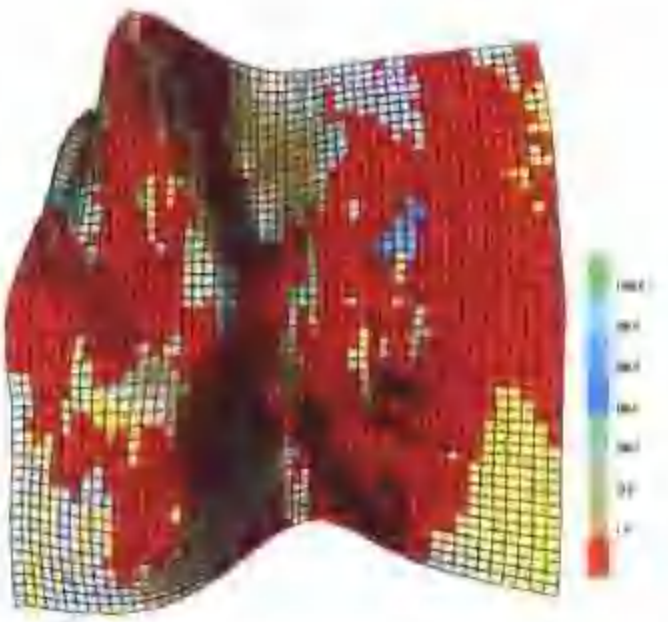
164282



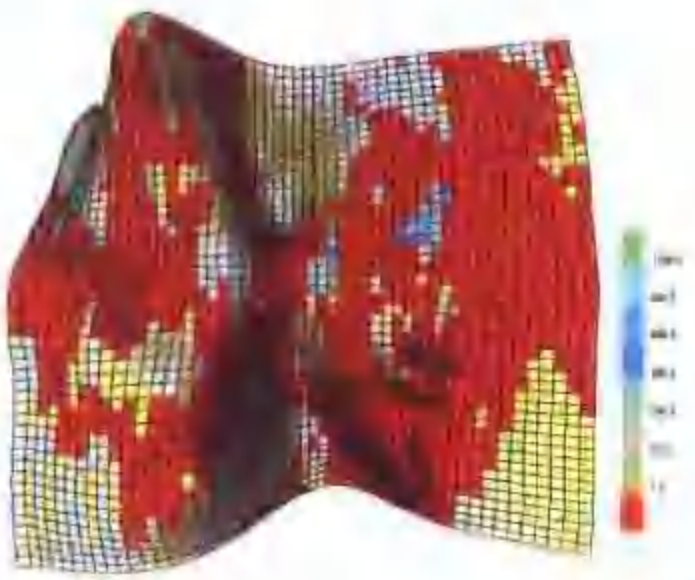
彩图29 用基于地质统计学的加权平均粗化后渗透率（50 × 50 × 20）



◀彩图30 用标准重整化粗化后渗透率（50 × 50 × 20）



彩图31 用简化重整化粗化后渗透率（50 × 50 × 20）



◀彩图32 用求解流动方程粗化后渗透率（50 × 50 × 20）

石油大学 0168157