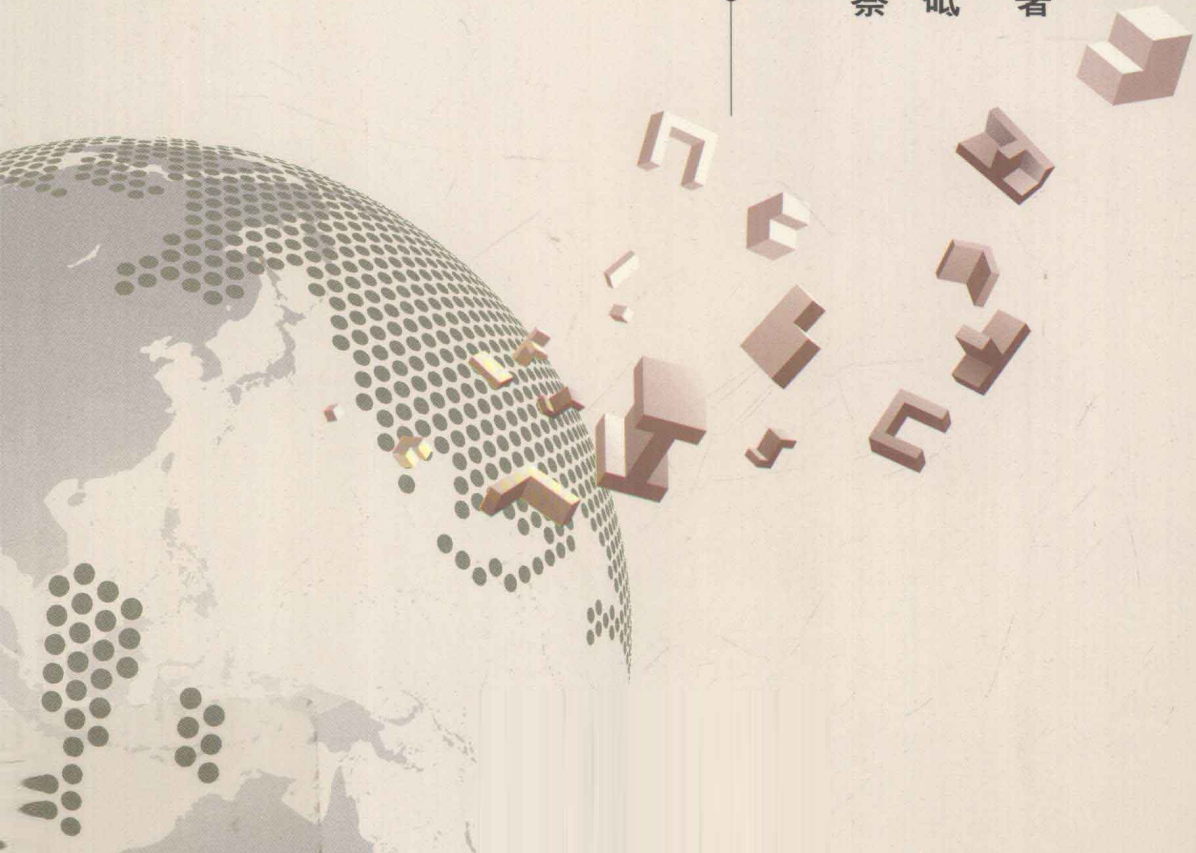
网格环境上的地理计算模式

蔡砥 著



電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

网格环境上的地理计算模式

蔡砥 著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

地理计算是以计算科学和计算技术为依托的，解决复杂的地理学问题的一种研究方法。地理计算的产生和发展得益于信息技术的迅速发展。网格作为一种跨组织边界的计算资源共享和管理技术，为地理空间信息的共享和互操作、分析和建模模拟等都提供了更有深度和广度的可能。本书一方面深入地论述了地理计算的内容，并从计算的角度出发进行了分类。以此为基础，广泛地收集各种与地理计算有关的应用网格平台项目的设计、实施的案例，论述了在网格平台上开展地理计算的多种模式。

计算模式的研究是科学计算系统的一种基础性研究，本书可供从事地理信息系统、计量地理和地理计算领域的科研人员、系统分析设计人员以及高校教师和研究生参考使用。

图书在版编目（CIP）数据

网格环境上的地理计算模式/蔡砥著. —北京：电子工业出版社，2011. 3
ISBN 978-7-121-12412-9

Ⅰ. ①网… Ⅱ. ①蔡… Ⅲ. ①网格—计算—应用—地理信息系统—研究 Ⅳ. ①P208

中国版本图书馆 CIP 数据核字（2010）第 231368 号

责任编辑：张云怡
印　　刷：
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：720×1000　印张：8.75　字数：157 千字
印　　次：2011 年 3 月第 1 次印刷
定　　价：56.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

人类对于其赖以生存的自然环境的知识积累过程几乎与人类一样久远，而作为一门学科的地理学的出现，也可以追溯到古希腊。然而这种古老的特性并不会阻碍其接纳新的研究范式，也不拒绝现代观测和分析手段的使用。

技术的发展是推动近代地理学发展的一个重要因素。起源于二战期间的遥感技术为我们提供了全新的、大范围的地球观测视角，卫星和航天遥感的快速发展更使得人们每天都可以获得几乎遍及整个地球表面的陆地、海洋、大气、人类、生物等各个全层的海量信息。20 世纪 60 年代发展起来的地理信息系统（Geographic Information System，GIS）和遥感影像处理系统为各种地理空间数据的获取、处理、存储、管理、分析和分发提供了重要的信息技术支撑。相比以前，地理工作者能够轻松地获得并处理丰富的地理数据和资料。20 世纪 70 年代，美国逐步建立的 GPS（Global Positoning System，全球定位系统）使得人们能够准确地确定自己的地理位置。2000 年，美国总统克林顿签署 GPS 民用化法案，使得民用 GPS 能够提高到 20m 分辨率的精度。而后进一步开放，使得目前静态定位精度可以达到毫米级，而即使是公众手持的 GPS 也可以达到亚米（<1m）级的空间精度。除了地理空间信息技术的长足进步以外，高性能计算技术、各种问题求解环境以及各种新算法的产生，更是使得复杂的、高时空分辨率的地理问题的求解成为可能。采用计算来对地理空间数据进行分析，对地理系统进行模拟已经成为新的研究范式，地理计算由此产生。

在地理空间信息技术和计算技术迅速发展的今天，一种新的计算环境也得到了广泛的重视和发展，这就是网格。网格计算环境作为一种全新的分布式计算环境，能够将地理上大范围分布的各种计算资源，包括处理器、存储空间、数据库、软件、网络等大量计算资源相互联通和协调，整合为一个有机的、可靠的系统以供科学研究和工程计算使用。

网格计算环境以大范围的资源共享、服务提供和互操作为重要特征。在这样的计算环境中，必须考虑的一个重要问题是各种计算资源如何有机地组合以完成特定的地理计算问题。为此，需要从两个重要方面进行细致的考察。一个是地理计算问题在各种计算资源的利用上有什么特点；另一个则是这些计算资源是如何联结、聚合和调度的。这就构成了计算模式的问题。地理计算模式的深入探讨，

对于构建具有应用意义的地理计算问题的求解环境非常重要，或者说，对于构建网格的地理应用非常重要，这是本书写作的出发点。

本书的写作尽量避免对复杂的技术问题的探讨，这是因为，计算模式是针对网格的特点而进行的各种计算资源的配置和调度以便顺利完成各种地理计算工作，因此探讨的重点偏向于资源的联结和管理问题，而非系统的设计和实施等细节问题，写作过程中更多地采用的是文献综述、地理网格项目案例分析以及用例分析等方法。

全书分为8章。第1章对自1996年以来10届地理计算国际会议的论文进行了全面的回顾，梳理论文的内容和历届会议的关注重点，目的是理清地理计算的提出、发展和特点，并对其技术基础进行探讨。第2章从计算的角度出发讨论地理计算的分类体系，通过对各种内容的地理计算问题所需要的数据量、计算时间复杂性和所涉及的组织机构的多方性等三个方面的分析，提出了基于计算复杂性考虑的地理计算分类体系。第3章讨论计算和计算环境的概念，从计算模型以及保障计算模型的实施所涉及的各种计算资源及其管理出发，认为计算环境是特定的各种计算资源通过网络相互联结，能够根据给定目的对这些计算资源可靠地进行调度以完成计算任务的系统。本章还论述了主要的计算环境类型的构成和特点，并引出了网格计算环境。第4章从网格的特点、体系结构、服务体系等方面对网格进行了较为详细的讨论，重点突出网格是一种跨管理域的、地理上广域分布的弱耦合性分布式计算环境。本章还对主流的网格平台Globus Toolkit做了简要的介绍。对网格在科学研究和工程计算的应用，特别是在地理计算中的应用做了讨论。第5章对网格上的地理信息共享技术进行了研究，首先讨论的是Web GIS中最重要的三种地理信息服务：WMS、WFS和WCS；进一步通过两个重要的地理信息网格项目论述了实施地理信息网格服务的关键技术；最后介绍了ESG和上海TIG中的地理信息服务的开发和特点。第6章论述网格上的汇聚型地理计算，以一个空间决策问题的求解作为用例，详细地讨论了地理数据以及相应的处理、分析过程的汇聚过程，提出了若干重要的地理信息网格服务以及相应的客户端设计。第7章以一个计算复杂性较高的P－Median问题的求解为例，介绍了类似问题的区域分解以及在计算网格上的并行化策略实施，从而论述了Master－Worker计算模式在网格环境上实现的关键问题。第8章探讨了网格上的协同计算，介绍了一个应用广泛的网格协同系统——Access Grid，并以用例的形式来探讨协同计算的作用和特点。

本书是以作者的博士学位论文《网格计算环境中的空间分析计算模式》为

基础完成的，该博士学位论文的完成得到了导师王铮研究员的悉心指导。而本书主要在地理计算、网格技术以及计算模式方面都做了扩展。在本书的完善和出版过程中，得到广州市教育局研究生学位点扶持建设项目经费支持。广州大学地理科学院陈健飞院长在本书的写作过程中给予了充分地支持，这是本书得以顺利完成的重要保障，借此对他表示衷心的感谢！另外，在本书的写作期间，作者所在学院的许多同事都在工作安排、实验条件等方面给予了诸多便利，在此一并表示感谢！

蔡砥（caidi@gzhu.edu.cn）
广州·广州大学地理科学学院

目　录

第1章 地理计算的兴起

1.1 地理计算起源

最早将计算机应用于地理学相关领域的是1950年美国人口调查局。该局购买了UNIVAC计算机用于当年的人口调查。20世纪60年代所发生的被称为计量革命的运动（Haggett，Chorley，1967），则促使了数学手段在地理学研究中的应用，在计量地理发展的阶段中，计算机起到了显著的作用。20世纪70年代，由于有计算机的辅助，使得统计中许多单调乏味，用人工很难完成的计算得以简便地完成，从而促进了多元统计分析在地理学中的应用（徐建华，1996）。

然而，能够与地理学所关注的空间结合起来进行数据管理和操作的尝试，是1965年加拿大森林与农村开发部门用于土地资源分类和辅助制图的一套系统，该系统被称为加拿大地理信息系统（CGIS，Canada Geographic Information System），并被认为是第1个GIS。GIS的出现，开辟了计算机应用于地理学的新的时代。

有了GIS，人们可以方便地将原来绘制在纸质地图上的很多信息转移到计算机中，变成矢量数据或栅格数据进行管理、显示和传输。同时，GIS在两个方面大大地降低了研究人员的工作量：一是地理要素的查询检索，在GIS的空间查询和属性查询的辅助下，研究人员能够快速地找到任何符合条件的数据项；二是图形运算和属性运算，以往费时、费工的面积量算、叠置分析、缓冲区分析等，都可以通过图形运算和属性运算迅速自动地完成。GIS的强大功能很快被土地管理、规划、军事等各方面所认同，并得到广泛地应用。随着桌面个人计算机的产生和普及，在20世纪80年代到90年代，掀起了GIS开发和应用的高潮，同时也对GIS的作用和性质产生了各种理解，认为GIS是一种软件工具、信息系统或者新的科学研究范式。

部分人认为GIS只是在地理空间数据的采集、获取、存储、处理、管理、分析和传输方面起到代替人工的作用，本身并不能对地理的思想和原理起到促进作用。另一部分人则认为，GIS在地理空间数据方面具有强大的功能，因此应该成

为与地理空间有关的组织机构的管理信息系统的一个重要组成部分，许多人致力于开发各种基于 GIS 的应用系统，用于各种管理，典型的如设施管理信息系统、土地管理信息系统。而许多地理学研究人员则逐步认识到，由于 GIS 以及各种对地球的观测和野外测量、探测、调查等方面的技术进步，必然形成一种新的科学研究范式，GIS 及相关技术将对地理学的发展起到突出的作用。因此，他们大力提倡发展 GIS 的空间分析功能、地理建模功能，认为 GIS 应该是继地图以后最为重要的地理学方法，有利于地理学家增强对于地理现实的理解，可以包含以下三个方面（Goodchild，2003）：

- 作为助手的 GIS。用于代替手工处理那些单调乏味的、复杂的、易出错的又代价高昂的任务。
- 作为交流媒介的 GIS。代替以前的媒介，让人们可以共享关于地球表面的知识。
- 作为手段的 GIS。GIS 数据库代替了地理现实，构建数字地球，辅助野外地理考察，通过对历史的、现在的和未来的信息的数字表达以增强感知。

对 GIS 是地理学方法的这种理解，源于地理学家对 GIS 乃至整个信息技术的运用经验。在 20 世纪 90 年代和 21 世纪初，地理学迎来了海量数据时代，每天都从卫星、航空器、观测站、地面传感器、野外测量和调查、社会统计等大量的来源产生着关于地球的各个地区的数据；与此同时，利用各种高性能的计算设备，基于这些数据开展分析、建模等工作，使得人们可以更深入广泛地对地球开展多种尺度的认知和理解。在这个时期，地理计算（Geocomputation）作为一种地理学研究范式，被提了出来。

倡导地理计算的是 Openshaw 和 Abrahart（1996），他们意识到在新的时期，地理学对计算机的应用已经不同于计量运动时期将计算机当做一种辅助运算的工具，而是一种科研方法。地理计算的提出，不再仅仅将计算机看做一种系统，其重点不是如何更好地采集、获取、管理和传输地理空间数据，而是着重于利用先进的计算技术，包括各种先进的算法、高性能计算等来解决地理问题。

1.2 国际 GeoComputation 大会

1.2.1 第 1 届地理计算大会

1996 年，在 Openshaw 等人的提议下，第 1 届地理计算国际会议在英国的利兹大学召开，这届会议得到了来自世界各地地理学者的广泛支持，与会人员共发表论文 97 篇，主要涉及地理系统模拟、空间统计分析与参数估计、地理知识发

现、模糊聚类、人工神经网络算法、非线性系统、空间推理、分形、元胞自动机建模（CA，Cellular Automata）、ABS（Agent－Based Simulation）等方面。

在大会上，Openshaw 和 Abrahart（1996）对地理计算的主要内容进行了阐述，认为地理计算是包括了高性能计算、人工智能和 GIS 的综合技术，应用范围有以下几个方面：

- 采用更多的样本点、更小的分区、更好的内插网格，消除由于计算性能的限制而必需的简化和近似，从而提高模型的分辨率和精度。
- 采用计算强度高的统计方法，如非参数检验的折叠刀方法、自举法或者蒙特卡罗显著性检验。
- 采用进化策略和遗传算法等来避免传统非线性优化方法中关于全局凸连续的假设。
- 采用计算强度更高的替代方法来获得更高质量的结果，例如，采用无监督的人工神经网络分类器对大规模的空间数据集进行分类，或者利用模拟退火算法来提高大规模空间区位优化问题解的质量。
- 利用基于人工智能的方法来替代传统的建模工具，比如，采用人工监督的神经网络算法来进行非线性关系的建模，而无须给出方程式。
- 通过发展新的混合（hybrid）技术将知识结合到已有的技术中，例如，空间数据的模糊建模可以将已有的定性知识和通过自组织模糊适应训练器在数据库里发现的知识结合起来。
- 处理那些被忽略的，或者被认为是不适合求解、难以求解的老问题，例如，利用人工生命和分布式自主体（Distribuited Agent）从数据中发现模式（Pattern）而不用预先指定在何处、何时去发现何物。
- 采用机器视觉技术从 GIS 数据库中捕获反复出现的高度抽象和高度概括的模糊模式。

1.2.2　历届地理计算大会

地理计算提出的十几年来，得到了人们广泛地认同，世界上先后召开了 10 届国际大会。进入新世纪以来，地理计算更是和 GIScience 国际大会隔年轮替召开，共同推动了地理学计算范式的快速发展。历届地理计算大会的简况见表 1－1。对历届的论文进行统计汇总，可以发现近年来地理计算大会所关注的内容包括地理数据分析、地理系统模拟、CA、ABS、空间运筹、地理研究可视化、自动制图、地理信息组织等方面。下文就最主要的几个方面做简单的回顾，包括地理数据分析、地理系统模拟、CA、ABS 和空间运筹。

表 1-1 地理计算国际会议简况

序 号	年 份	地 点	论 文 数
1	1996	英国利兹大学	97
2	1997	新西兰奥塔哥大学	28
3	1998	英国布里斯托尔大学	52
4	1999	美国马里华盛顿学院	59
5	2000	英国格林威治大学	47
6	2001	澳大利亚昆士兰大学	64
7	2003	英国南安普敦大学	69
8	2005	美国密歇根大学	129
9	2007	爱尔兰大学	103
10	2009	澳大利亚新南威尔士大学	58

1. 地理数据分析

对地理数据的分析，无论是包含了空间、时间还是时—空的变异和解释，一直是地理学的重要内容。在地理计算国际会议上，除了传统的统计、空间探索性统计、地统计等统计方法之外，在人工神经网络方法、遗传算法、Swarm 方法、模糊集合等新方法方面都获得广泛的重视。

人工神经网络（ANN，Artificial Neural Network）方法是一种模拟简单神经元信号传输和变换的原理而构造的用于多个输入信号向多个输出信号进行映射的方法，在无须进行先验统计分布假设的前提下就可以形成输入—输出的非线性映射，从而在模式识别、预测等方面有良好的应用前景，因此成为地理计算的一种重要的新方法。ANN 方法的这种输入—输出的映射特点可以用于要素之间关系的建立，其作用类似于统计上的回归，如 ANN 用于地表能量交换过程的模拟（Pokrovsky，1999）。在水文系统的建模方面，ANN 用于预测水文网络的发展，被称为神经网络水文学（Abrahart，1996；Abrahart、See 和 Kneale，1998），也用于特定水文站的水位预测并与传统用于分析时间序列的 ARMA 模型进行比较（Abrahart、See，1998）或其他预测（如 Abrahart，2001）。模式识别、类型识别也是 ANN 的一个重要应用，如植被类型的划分（Foody、Boyd，1998；Gahegan、Takatsuka，1999；Bosch，1999；Gibson、Cameron－Jones，2001；Huang，2001）。研究表明（Gahegan、West，1998），和决策树等其他方法相结合，有利于处理复杂的地理数据集，提高分类的精度。人工神经网络的应用难点，在于选择恰当的网络层级结构以及相关的参数如权重、阈值等，当给定一个 ANN 结构后，参数的确定可以视为一个优化问题，由于非线性特征的存

在，该优化问题可能会陷入局部最优，解决这个问题的一个新途径是采用遗传算法去优化参数（如 Fischer、Reismann，1999；Dawson，2000）。另外，ANN 的一个弱点是缺乏显式的函数来描述输入与输出的关系，这使得 ANN 看起来像是难以捉摸的黑箱，有研究着力于建立可视化的 ANN 训练系统（Laffan，1998）。

空间数据挖掘是在空间数据库或空间数据仓库的基础上，综合利用多门学科的理论技术，从大型空间数据库中挖掘事先未知、潜在有用、最终可理解的可信新知识，揭示蕴涵在空间数据中的客观世界的本质规律、内在联系和发展趋势，实现知识的自动获取，提供技术决策与经营决策的依据（Lu、Han 和 Ooi，1993）。该定义非常广泛，并没有限定具体的方法，因此前面所论及的各种研究，只要涉及空间数据，即可认为属于空间数据挖掘。不过，空间数据挖掘更着重于数据的规模很大，需要大量的计算。例如，在大型数据库或者遥感影响中发现局部相似性、不同空间尺度的相似性（Holt、MacDonell 和 Benwell，1998），或者对大规模空间数据库中发现地理现象之间的关联关系（如 Estivill－Castro、Lee，2001），典型地被认为是属于空间数据挖掘。

空间数据挖掘经常采用各种新的方法。在 Alvanides，Openshaw（1996）在他们开发的分区设计系统 ZDES3 中，将一个禁忌搜索（Tabu Search）算法和一个模拟退火算法嵌入到罚函数的优化框架中，来获得最佳的分区方案。他们还以 1991 年的英国人口普查数据对上述方法和其他人工方法进行了比较（Openshaw，Alvanides，2001），发现其对消除小概率影响是有效的。分区可以看做是一种包含了空间维度的聚类过程，在聚类中，遗传算法也可以发挥重要作用，研究发现，当确定适当的优化目标时，遗传算法能够比神经网络方法获得更小的类内差异（Painho、Baçăo，2000）。聚类研究中也引入专家系统，采用经验的描述性规则对那些可能属于不同类的对象进行划分（如 Pakiarajah、Crowther 和 Hartnett，2000）。由于数据库很庞大，全局的假设或先验条件将影响空间分析的结果，因此高效的局部算法成为必需。一种称为 Swarm 的人工智能方法逐步被引入地理学，特别是在聚类以及空间相似性的发现上。该方法的基本思路是利用大量的小型模式识别器来搜索整个“地理赛博空间”（GeoCyberspace），或者说是整个地理数据集，在发现可能的模式的同时，这些模式识别器会相互交换信息，以印证潜在模式的可靠性（Macgill、Openshaw，1998）。在空间数据挖掘中，模糊集合非常适用于主观评价和定性描述的资料处理。例如，常常会使用模糊集来定义土地单元的环境适宜性（Baja、Chapman 和 Draqgovich，2001）或者灾害的敏感度（Schernthanner，2007）。另外，通过模糊集合的定义，可以将边界具有不确定性的“地点”（Place）这个传统地理概念和有明确定量特点的点、线、面等 GIS 要素结合在一个系统中（Albrecht、Guesgen，1998），地点可以被表述为模糊足迹

（Fuzzy Footprint），或者模糊边界（Fuzzy Boundary）、对象（Harada、Sadahiro，2005）。模糊理论也可以应用于地理事物的分类（如 Mason、Jacobson，2007）；也包括遥感影像分类上（如 Atkinson，1999），对于混合像元来说，其属于何种分类，本身就可看做一个模糊集合问题。空间数据挖掘并非都采用新的方法，也可以利用基本的 GIS 工具进行创造性的使用，例如，Voronoi 图、四叉树（Emerson、Chinniah、Lam 等，2005，）以及统计和地统计方法（如 Makido、Shortridge，2005，用于混合像元的亚元分类）、马尔科夫链（如 Li、Zhang，2005）等。

当空间数据集庞大到算法无法单独处理时，需要采用其他策略，例如，将数据集先按照某些特征相似性分割为小的子集再进行分析（Kechadi、Bertolotto 和 Martino 等，2007）。

2. 地理系统模拟

动力学模型常常用于自然现象的模拟，例如，WEPP 和 EUROSEM 用于土壤侵蚀，SAKE、TOPMODEL、TELEMAC－2D 用于水文模拟，GCMs 用于大气模拟，人口变动要素合成模型用于人口迁移模拟等。由于空间数据获取技术的进步，高分辨率的空间数据成为提高模拟精度的途径，例如，LIDAR 的高程数据用于河漫滩的水流模拟（Marks、Bates，1998）。

然而必须意识到，由于全面采集数据的困难、对潜在的各种机制进行全面模拟的困难以及各种误差的存在，模型的简化是必然的。有时候这种简化会形成蝴蝶效应，使得模拟结果与实际相差甚远。典型的如坡面漫流形成细流的形状基本是难以预见的，因此在一定的尺度下引入自组织系统来进行模拟，细流被认为是相互竞争的，使得某种熵最大化的过程。将小尺度的自组织和大尺度的动力学模型结合起来进行模拟是一个重要的方法（如 Favis－Mortlock，1998；Boer，1999）。

将地理模拟软件与 GIS 结合起来，或者直接将模型编写为 GIS 的一个扩展模块，是 20 世纪 90 年代以来的一个明显趋势。GIS 可以用图层的方式来存储和操作多种空间上非均质的地理要素，从而方便了模拟各种参数、初始条件、外生变量的管理。例如，河流盆地的地形、土地覆盖类型、气象气候条件等可以作为空间数据存储在 GIS 数据库中，利用 GIS 做恰当的插值处理等就可以用于流域动力学模型的输入变量，从而模拟盆地的形成和发展（De Roo，1998），也可以用于管理水文模型的各种局部参数（如 Schmidt、Hennrich 和 Dikau，1998）。到 21 世纪初，GIS 和其他模拟软件的耦合已经成熟，许多模拟软件具有和 GIS 结合的接口，甚至直接可以利用 GIS 数据，因此利用模拟的结果进行显示、分析和相应的优化、假设分析、政策分析等成为常规的手段。例如，用基于道路网络数据的交通事故模拟结合 GIS 来优化应急响应部门的配置（Huang、Pan，2005）。

Alvanides、Boyle 和 Duke – Williams 等（1996）根据 1981 和 1991 年的人口普查数据，建立了到选区之间的人口迁移模型，该模型涵盖了英格兰到威尔士之间的 99 302 个选区，构建了选区间的迁移流矩阵。

与人口有关的另一个需要海量数据和计算的模拟是全球 1km 分辨率的人口分布估计计划——LandScan，该计划收集了全球各个来源的尽可能详细的人口调查数据，每个地区的人口数据作为总量约束，同时利用其他相关的空间数据来估计每个空间单元的人口“似然”（Likelihood）系数，作为空间单元人口数的估计依据。其中，与人口分布密度相关的其他空间数据包括土地覆盖类型、坡度、道路邻近度以及夜间灯光辐射强度的遥感数据，进一步的工作是将空间分辨率提高到 90m（Bhaduri、Bright 和 Coleman，2005）。

3. CA

CA 是一种自下而上的建模方法，能够将空间和时间有机结合起来，模拟地理现象的时空过程。CA 包含 5 个基本要素（见 Nara、Torrens，2005）：元胞（Cells）、状态（States）、栅格阵列（Lattice）、邻域（Neighborhoods）和转换规则（Transition Rules）。空间被划分为若干元胞，这些元胞紧密排列构成了阵列。元胞具有可以改变的属性，以描述元胞所在空间范围的自然或人文属性，同时，在每个系统时间间隔内，元胞根据全局（Global）和邻域（Neighborhood）的关系以及相应的转换规则来改变自身的属性。通过这样的机制，模拟期被划分为若干时间间隔（称为期），每期内所有的元胞都会改变自身的属性，称为一次迭代，通过多期的迭代使得整个空间的每个局部都能够体现动态变化的特征，因此受到地理计算国际大会的关注。表 1 – 2 为历届 CA 论文情况。

表 1 – 2　历届 CA 论文情况

年　份	CA 论文篇数	主 要 内 容
1998	1	地形的形成
1999	3	用于城市模拟的 CA 系统原型研究；基于图形的 CA 系统；高分辨率的人口表面模拟
2000	1	CA 城市建模原则的理论探讨
2001	2	CA 引入模糊集合；CA 用于地图融合
2003	3	城市模拟的 CA 模型的应用和比较；人工神经网络用于 CA 的知识库模型；模型校准方法比较

续表

年 份	CA 论文篇数	主 要 内 容
2005	10	CA 应用例子（城市、森林害虫传播）；CA 与 ABS 结合模拟疾病传播；老市区衰落模拟；地理自动机（GA）的提出和 OBEUS 系统原型开发；利用自组织人工神经网络辅助城市模拟；CA 模型校准方法的比较；可持续的土地分配
2007	5	非规则 CA 和 ABS 结合进行城市土地利用变化模拟；GA 用于动物疾病扩散；基于马尔科夫链的 CA 用于城市土地变化模拟；矢量化的 CA，用于土地利用变化模拟；城市增长模拟
2009	1	随机 CA 模型用地变化模拟，ABS + CA 模拟城市增长

CA 常用于城市演化模拟，CA 通过简单的局部转换规则，可以模拟出复杂的城市空间结构（黎夏、叶嘉安和刘小平等，2007）。Clark 等（1997）最初成功地建立了城市土地演化 CA 并用于旧金山海湾和华盛顿特区扩张研究，其模拟结果和真实数据非常接近。Colonna、Di Stefano 和 Lombardo 等（1999）尝试开发了一个和 Clark 等人不同的 CA 模拟原型系统，该系统在元胞的形状和阵列方式、对边界的处理、元胞反映的土地利用类型、地理要素的属性化处理、元胞之间的空间相互作用等方面均做了探索性研究，并利用该系统模拟了罗马的城市发展。在城市演化过程中，用地的转化规则有时候是以自然语言的形式表达的，具有显著的模糊特征，为此模糊集合也被引入到 CA 中（Liu、Phinn，2001），用于定义某些模糊地理边界，典型的如城市的边界、CBD（Central Business District，中央商务区）的边界，都是模糊的。Sun（2005）将生态、经济、社会、地理和环境等基础理论集合为一个层次框架，用于开发一个用于研究区域土地利用演化和影响估计模型（LEAM，Land - use Evolution and impact Accessment Model），该模型的空间单元不但能够获取局部的邻域信息，而且能够获取整个区域的全局信息，如社会环境和经济的发展趋势及综合。

另一个广泛应用 CA 的领域是传染病研究（如 Malanson，2007；Green、Ahearn 和 Carney 等，2007；Laffan、Lubarsky 和 Ward 等，2007）。通过 CA 模拟，可以阐明在给定的地理环境中是如何产生患病风险的空间差异的；也可以探讨个体行为如何影响疾病的传播过程（Malanson，2007）。CA 也应用于生物扩散研究，如结合模糊集合定义“虫害 - 树木”的约束来模拟虫害对森林地带的群袭过程（Bone、Dragicevic，2005）。

CA 的一个基本问题是邻域的规定。由于常用的元胞是正方形，因此邻域可以是 4 格的 von Neumann 邻域，或者是 3 ×3 的 Moore 正方形邻域（如 Okwuashi、McConchie和 Nwilo 等，2009）。Moore 邻域经常被使用，甚至比 3 ×3 更大范围的

Moore 邻域也被使用，较大的邻域可以提高局部规则的参数估计的统计确定性，有的研究甚至包含 112 个元胞的邻域（White、Engelen，1994，1993），但邻域的大小对模拟结果的影响并没有系统的研究（Liu、Phinn，2001）。邻域的形状被确认是有影响的，对城市扩张的研究发现，Moore 邻域会导致城市呈现指数增长，和实际的增长模式不符，von Neumann 邻域则可以降低城市的增长率（Yeh，Li，2003）。Von Neumann 的 4 个邻域（东、南、西、北向）可以看做是最小的离散化圆形邻域，圆形邻域会降低相应的失真现象。

邻域转换规则是 CA 的另一个基本问题，影响着 CA 的模拟结果。为了让规则驱动的 CA 能够更好地符合实际的结果，必须利用样本数据对规则所涉及的参数进行估计，这个过程可以使用空间统计模型，或者使用其他方法，如人工神经网络算法等。Yeh、Li（2003）总结了城市模拟中的若干种方法，包括：

- 采用生存博弈的概念来模拟元胞的产生、存在和消亡。
- 利用 AHP（Analytical Hierarchy Process，层次分析法）来估计概率。
- 利用模糊集合定义转换规则。
- 运用预先定义的参数矩阵来计算转换的可能性。
- 利用灰度值法模拟城市的转变。
- 利用人工神经网络来校正（Calibrating）和模拟城市的发展。

一种称为 SVM（Support Vector Machine，支撑向量机）的方法也用于参数校正。SVM 是一种机器学习的典型方法，在利用拉格朗日乘数寻求输入 x 的函数 $f(x)$ 以及输出 y 所构成的最优超平面时，非零拉格朗日乘数构成了支撑向量，这些向量可以用于估计 $f(x)$ 的参数。Okwuashi、McConchie 和 Nwilo 等（2009）的研究表明，SVM 比人工神经网络方法具有更高的精度。但由于 SVM 需要更高的计算量，因此在实际应用中比较少用。

CA 在非空间的表述方面比较欠缺，使得其在反映系统的反馈以及社会经济因素对决策的影响等方面的能力受到制约（Wu、Silva，2009），这促使了 ABS 的发展。2003 年以后，CA 逐渐被 ABS，或者结合了 CA 和 GIS 的 ABS 所代替。

4. ABS

另一种自下而上的建模方法是 ABS，即基于自主体（Agent）的建模和模拟。ABS 最初用于人的社会行为的研究，将每个个体的决策能力、学习能力用软件单元进行模拟，这种软件模块叫做自主体，一个系统则由大量的自主体以及一定的环境变量所构成。随着 ABS 建模思想的推广，ABS 已经不仅仅用于研究人的社会行为，在经济模拟、生态系统模拟、城市模拟等领域都出现了 ABS 的热潮。由于 ABS 是一种全新的利用计算机进行建模的研究方法，并且能够将个体决策