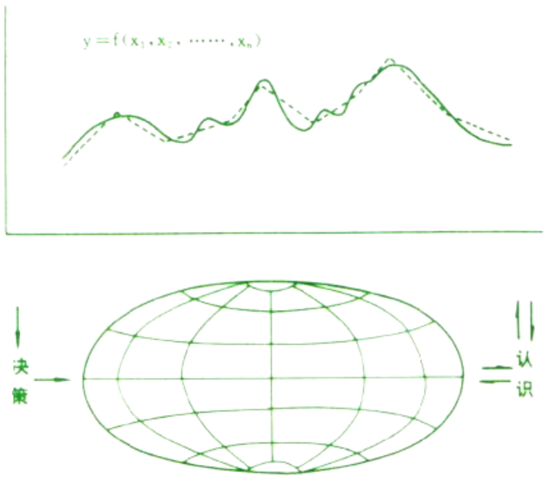
现代地学数学模拟

李钜章 著



气 象 出 版 社

现代地学数学模拟

李钜章著

气　象　出　版　社

1994

(京) 新登字 046 号

内容提要

本书针对地学的学科特点，从定性与定量的关系和数学模拟的实质出发，通过基本数据分析整理、分类、分区、过程模拟、规划与决策、判断与推理等方面的实例，通俗地阐述现代地学数学模拟的基本步骤，同时介绍若干最基本的便於灵活运用的或可以借鉴的数学方法.

本书可供地学工作者、应用数学工作者以及生物学等相邻学科工作者参考使用.

现代地学数学模拟

李钜章　编著

责任编辑：庞金波　　终审　纪乃晋
封面设计：李钜章　责任技编：朱立平　责任校对：李钜章

气象出版社出版·发行
（北京西郊白石桥路 46 号　邮政编码：100081）

科远印刷厂　　印刷
开本：1／16　印张　10.2　字数：230 千字
1994 年 9 月第一版　　1994 年 9 月第一次印刷
印数　1—1000 册　　定价　9 80 元
ISBN7—5029—1835—3／P·0712

序

李钜章教授根据他从事自然地理学——特别是地貌数学分析的长期实践，结合若干地学工作中典型、具体的实例，包括自然过程、区划、规划与决策、判断与推理等经常遇到的科学问题，茹苦含辛，锲而不舍，探讨地学研究工作者如何运用形式语言表述和概括传统地学知识；如何运用数学思维探索新的地学规律．他的大量研究工作，向地学与数学相嫁接、向定性与定量相结合的方向，迈出了很大的一步．《现代地学数学模拟》一书的编著，汇集了他数十年辛勤耕耘的所得．以简洁朴实，深入浅出的语言，无私地奉献给读者．这种春蚕吐丝的忘我精神，使我很感动．德不孤则必有邻，正如培根所说，“知识只有通过传播与交流，才能产生价值”．我想这本书的读者，将会从中吸取营养，得到启示，将会成为李钜章教授的知音，共同为地球信息科学的发展，为地学适应信息社会的需求，作出更多的贡献，李钜章教授也会感到精神上的回报和学科发展的欣慰．因此，本书的问世，是值得庆幸的．

钱学森教授指出：从定性向定量化的集成是现代科学发展的大势所趋．在信息社会中，信息的获取、传输、存储、处理过程几乎都是以数学方式进行；而分析、模拟、预测、预报更离不开数学方法．地学界对数学方法的应用长期存在着截然不同的看法，也经历过多次历史的反复．李钜章教授以切身的体会和认识，曾在《地理学定量研究中的若干问题》（地理研究，1988，第7卷，第2期）一文中，作过比较冷静的分析．阐明了他自己的见解，也指出了存在的局限和问题．这些客观上的局限性和主观上的认识问题，在地学方面有着相当广泛的普遍性，只是由于学科发展阶段的不平衡，程度参差不齐，地理学更为滞后而已．

其实，定量研究在地理界起步并不太晚，计量地理学曾在欧美盛极一时，在我国亦不乏先例，例如40年代遵义、浙江、福建等地的相对地势制图．内蒙古侵蚀面的分析与东北人口重心迁移的研究，50年代东南沿海水系的分形分维研究与黄土高原地貌制图，都显示出定量分析的萌发，可惜只是初试锋芒，浅尝辄止，没有持之以恒，形成气候．而在其它地学领域则有所不同，我国大气、海洋科学中，流体力学方面的数学分析与模拟，经过近半世纪的努力，已跃居世界先进行列，地震等地球物理场的数学分析与模拟也毫不逊色；数学地质、生物数学也取得可喜的进展．总结经验教训，我们认为成败关键在于以下几个方面：一是正视地球系统科学与人地关系的复杂性与不确定性．对其中内外循环、演化过程机理的认识还很模糊．用简单的数学方法表达地学的知识规则和逻辑推理比较困难．数学方法不能生搬硬套，只有选择确有物理含义的适当数学模型，并经过区域参数的订正，才有可能逼近地学规律．近年来采用模糊数学和分形分维方法，比单纯采用数理统计，开始有了可喜的进展．二是吸收信息科学和空间技术的最新成就．由于传感仪器、卫星平台和台站网络的进步，对地观测的广度和深度迅速延伸到高空、海洋和地球深部，全球范围的立体、同步观测成为现实，数据的自动采集与标定趋向国际标准化，大大增强了时空变量的可比性．三是运用地理信息系统作为多维分析、模拟与综合集成的有效载体，发挥它多层次结构与泛目标开发的功能．事实雄辩地告诉我们，顺应时代潮流、加速自身现代化的学科，从而得到了较快的发展，背离这种趋势的学科往往陷入难以自拔的困境．地理学亡羊补牢，急起直追，还是很有基础，很有潜力的．

面临空间时代与信息社会的机遇与挑战，地学信息来源极大丰富，数据处理技术日新月异；地学工作者责无旁贷的当务之急，是致力于加速地学数据的标准化和指标体系，建立地学分析、模拟、预测、预报模型，以利于地学规律的表述与运作，为地学战略决策与地学工程设计，提供切实有效的信息服务。李钜章教授在现代地学数学模拟研究工作中，表现出来的知难而进，坚持不懈，讲求实效的顽强精神，今天格外值得我们学习和提倡。

陈述彭

1994年劳动节日

目　　录

第一章 绪 论

第一节 概 述

一、建立地学信息系统的关键

在现代科学研究及其成果的推广中，信息系统的建立与运用，是科学研究现代化的一个重要标志．由于地学学科特有的极强之空间地域性，需要各种图形的处理功能，它给计算机工作者提出了一个用计算机进行图形处理以及图形与文字数据的联接、转换等工作的课题．为解决这一课题建立的计算机系统一般称为"地理信息系统"（Geographic Information System，简称 GIS）．现已有多种 GIS 系统可供选择，故可以说在计算机技术上已基本能满足需要．所以计算机技术已不是建立实用的地学信息系统的重要障碍．

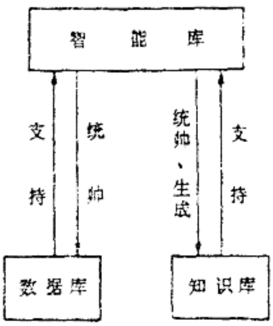


图 1.1 信息系统结构图

信息系统一般包括智能库、数据库、知识库三部分，其关系如图 1.1 所示．由图可见：1）智能库是反映分析、判断、推理等逻辑关系的软件之集合；其功能主要是根据存贮在数据库中的资料数据以及存贮在知识库中的知识，对实际或研究中的问题作出回答．2）数据库是将所需资料数据存贮及管理，主要包括录入、存贮、检索、查询、分类、排序、追加、插入、删除、修改、显示，打印、拷贝以及一些基本统计等多种功能．3）知识库是记录、存贮智能系统运行中所需的基本知识（诸如各种参数、判据等）及其管理与更生程序．这些基本知识是人们通过对客观世界的观测、分析归纳、研究总结出的某些规律及有关信息．有时它们需要在数据库的支持下由智能软件生成或改进，但更多的是前人的研究成果或从其它信息系统获得的结果．

智能库是解决问题的关键，数据库和知识库是信息系统的基础，没有它们的支持，再好的智能系统也不可能解决任何实际问题．就象一个人若不让他了解实际情况、不掌握有关信息，再有经验、能力再高，也不可能作出正确的判断一样．知识库和数据库都是为支持智能系统而存在，所以其中应包括的项目、详细程度、数据的精度乃至存贮的方式、数据库的结构等都应服从于智能系统的要求，必需满足又不应该超过智能库的需要．超出要求必然造成人力、物力的浪费，而不满足要求信息系统便无法运转．当然，这里必需考虑信息系统发展之需要，因为任何一个信息系统都有一个逐步发展过程，而一个智能软件的建立一般都需要大量的资料数据．一个信息系统的建立，首先应该充分地分析、研究所研究的对象和目的，确定智能系统的总体框架；然后在这个总体框架的统帅下，搜集（包括由实际观测获得）资料数据和有关的基本知识，开始建立数据库和知识库；再按总体框架之设计，利用已有的资料数据和知识逐步建立和发展智能系统、改进知识库、充实、扩展数据库．因此，应该说：智能库是信息系统的灵魂．地学信息系统的建立，关键在于信息系统的核心——智能系统的建立上，或者说是信息系统的总

设计思想上．智能系统的实质就是形式化的思维规则的集合，是对人的思维之模拟．由于目前使用的主要是数值计算机，所以这种模拟主要是数学模拟，也就是说地学信息系统的建立关键在于数学模拟．

二、数学模拟与定量化

任何数学模拟都是对人的思维之模拟．象综合评判、分类、分区等问题本身就是人的思维活动，而象土壤水分的变化，农作物产量的形成等自然过程（现象）的数学模拟，也只能是通过模拟人对该问题的认识进行间接的模拟．从本质上说，数学本身就具有很高的抽象性，不经过抽象就不可能变为形式化的数学语言，而抽象就是人的一种思维．所以即使象牛顿力学的基本公式 $F=ma$ 这样反映确定性物理规律的数学式，也只有通过对人的认识的模拟才能建立起来．它能够正确反映客观规律，以人对该问题的认识正确反映客观规律为前题．并在通过物理实验的验证，证明这种认识和模拟的正确性之后，才能最终确立．还必须指出：数学的抽象形式表现力是有局限性的，抽象概念本身总是带有某种片面性和僵化性，所以绝不能迷信数学方法，以为运用数学方法，经过数学计算或经过计算机处理就绝对可靠．这一点是在进行数学模拟以及运用 GIS 研究、解决问题时必须十分明确的．既然任何数学方法都是对人的思维的模拟，则无论运用什么数学方法都必须明了它所模拟的思维实质、了解它的作用．在了解某种数学方法实质之前就拿来套用，以为只要获得之结果合理或者可以解释，便说明该方法可用，否则将其弃之一旁，显然是不正确的．当然，数学的形式逻辑系统有助于正确逻辑思维的形成，但应该特别指出的是，这种作用主要是在应用数学方法，即建立数学模型的过程之中．而且要以充分了解所用的数学方法，正确处理定性与定量、定性思维与数学方法的关系为前题．而模拟结果的正确性，归根结蒂要通过与物质世界相联系的实践去验证．

应用数学方法首先必须定量化，进行数学模拟建立信息系统当然也必须首先定量化．因而定量与定性的关系是进行数学模拟时必须十分明确的另一个重要问题．定量与定性是密切相关的，它们之间并没有截然的分界线．质和量是不可分割的，任一种性质总和一定的数量相关联．例如一个山地包含有若干自然垂直带，各带有着不同性质，每个垂直自然带就与一个海拔区间数相关联；水的流体性质与一定的温度区间数相关联，等等．人们在了解事物的性质之后就要进一步了解与之相关联的量．譬如在了解一个山地含有什么垂直自然带之后，自然就要求了解它们的分布高度，也就是要了解与各自然带相关联的海拔区间数．又如在了解了水的流体性质及其与温度的关系之后，会要求了解在什么温度下水能保持其流体性质，也就是要了解与之关联的温度区间数．也就是说，研究的深化自然要求定量，即要求研究那些引起质变的量．而这种决定事物性质的量（界限、阈值）在绝大多数情况下，并不是确定不变的，而且往往不能精确地确定下来，即使较简单，较分明的与水的流体性质相关联的温度区间数，也不是固定不变地总是[0, 100]（摄氏度）．这个区间数至少会因大气压不同而发生变化．而且即使在同一环境下，在一般日常测量精度下，这个区间数可以被看作是分明的，精确的，但当测量精度足够高时，这个区间的边界仍然存在着一定的模糊性．至于与自然垂直带相关联的海拔区间数就更是这样，两个垂直带之间一般是一个过渡带而不是一条截然的分界线，而这个过渡带在不同地点往往会有不同的宽度和不同的海拔高度．所以这种量可以称之为定性的模糊量．用一个模糊集来表示一个自然垂直带的分布空间显然是适宜的．从另一角度来说，以往地学研究中常用的定性分等

(分级）方法实际上也是一种定量，而且是分明的（确定性的）定量．虽然这种方法存在着较明显的缺点，但它毕竟是属于某一等级的事物就仅属于该等级而完全不属于其它任何别的等级，所以排中律在这里是完全的，其中没有一点模糊性．例如对土壤盐碱化程度分为无盐碱、轻度盐碱、中度盐碱和重盐碱（盐碱土）四个等级，就是对土壤盐碱化程度的一种量的刻度．这当然也算是一种定量化．而且任一土壤都必属于一个且仅属于一个等级，所以是一种分明的（确定的）定量化．只不过其精度较粗，并且没有用数字形式加以表达而已．由此可见，用自然语言描述，侧重定性的研究中，也有定量的内容，存在着定量化因素．其实，定量化是研究深入的必然结果，而且随着研究的深入，定量的精度会逐步提高．再从分析研究方法上看，定性与定量同样也不是截然分离，毫不相关的，更不是互相对立的．读者可以找到大量定性思维与数学定量方法互相渗透，互相依赖的例证．事实上，数学方法本身就是形式化的人类思维，其中大多数都是定性分析思维的总结归纳．人的认识本来就总是始于感性认识，经过定性阶段才能达到定量化．就以数理统计方法而言，它实质上就是一般定性经验的积累，形成过程思维活动的归纳总结之后的形式化抽象，而在有了数理统计分析方法之后人们常常运用数理统计分析方法来帮助分析、总结已掌握的资料数据，并从而获得经验性的认识．这在以往地学基本上是定性的研究中也是常用的，只不过大多数仅限于诸如平均值、变率等简单的方法．由于数理统计方法仅以所提供的样本之数据为依据，而不问它们的实际含义和代表性如何，对每个数据一律一视同仁，因而在一定程度上有利于排除主观因素的影响．这是数学方法的优点，同时也是数学方法的缺点．因为由于不以定性分析为基础，仅凭数理统计方法从仅有的数据中得出结论，有时会与客观规律相违背．特别是当提供用作数理分析统计的数据（样本）数量不多时，更易发生这种情况．综上所述，定性与定量是相对的，互相渗透的，两者绝非截然不同，更不是相互对立的．定量分析研究，特别是应用抽象的数学方法时，必须以定性研究为基础，定量研究是定性研究深化的必然要求和必然结果．

任何数量都有一个精度，而任何精度都有其相对性．所以数量化不是绝对的．而且一个数也并不是越精确越好，有时一个精确度较差、定性的模糊量（任何一个数对更高的精度而言都可以看作是一个模糊量），比一个刻板的过于精确的量，能够更深刻地反映事物的本质，因而更合理、更准确、更实用，当然也应该算更好．比如在日常生活中，人们根据没有严格地精确定量的定性印象可以确认多年不见的亲人或老友，而如果要严格地按精确的刻板的量来识别一个人，则恐怕连刚刚分手的熟人也不能确认．因为任何一个人每时每刻都在发生变化．所以严格地说，同一个人前一瞬间和后一瞬间都不可能完全绝对相同．也就是说，即使对同一个人也不可能两次测出极精确的完全相同的刻板数据来．当然也就不可能刻板地按极精确的量来识别任何一个人．事实上，任何学科的任何规律都有其相应的精度，超过这个相应的精度其结果也就不再正确．这个精度与研究深度和详细程度密切相关．地学是一门综合性很强的高层次学科，其研究结果之精度显然不可能超过其中各有关因素及有关的基本规律的精度．而且一般比其中精度最低的精度要低．实际上由于地学问题的多元性，在研究中一般都不能把所有有关的因素及其交互作用都一一加以考虑．这些未被考虑的因素必然大大降低研究结果（规律、模型等）的精度．所以一般地说：综合地学的研究精度低于部门地学的研究精度，部门地学研究的精度低于物理、化学等基本学科的研究精度．不同对问题的研究深度、研究目的和研究之详细程度，单从数学

方法上提高数值上的精度是没有意义的，有时甚至是错误的．也就是说学科的性质本身决定了地学的定量化不可能有很高的精度．如果从实用的角度来衡量定量化的好坏，则显然还必须考虑经济效益．而一般来说获得较精确的数据都必须付出较多的花费，那么提高研究成果的精度当然也要花费更多的时间、精力和经费．所以，有时虽然从学科上说完全可以获得较高一些的精度，但从实际需要和经济效益考虑还可能选择精度较低的研究方案．因为，只要能够满足实际需要（能够解决实际问题）当然以花费较少的方案为好．

三、地学与数学模型

数学模型方法是处理科技领域中各种实际问题的一般数学方法，"数学模型"的含义很广．粗略说来，数学模型乃是针对或参照某种事物系统的特征或数量相依关系，采用形式化数学语言，概括地或近似地表述出来的一种数学结构．当然这种结构应该是借助数学概念和符号刻划出来的某种系统的纯关系结构．所谓纯关系结构是指已经扬弃了一切与关系无本质联系的属性后的系统而言，所以，数学模型的形成过程中已经用了抽象分析方法．也可以说，抽象分析法是构造数学模型的基本手段．从广义上讲数学中各种基本概念，如实数、向量、集合、群、环、域、范畴、线性空间、拓扑空间等都是数学模型．而反映、模拟特定问题或特定具体事物系统的数学结构是数学模型的狭义的解释，也是在应用数学中数学模型的一般含义．本书中均采用这种解释．

数学模型种类很多，而且可以按不同的观点进行分类．例如，按所用的数学方法进行分类，诸如确定性数学模型、统计数学模型、模糊数学模型等；也可以按所解决问题的性质分为综合评判模型，动态仿真模型、优选模型、预测预报模型等．下面仅就按模型对问题反映的深度划分的白箱模型、灰箱模型和黑箱模型略加说明．白箱模型是通过对所研究的系统内部结构及各因素的关系进行模拟，即从系统机制人手模拟系统之功能，从而得出各种输入所对应的输出．构造白箱模型必须对所研究的对象有较深人透彻的认识，对研究对象的物理机制缺乏认识是不可能进行抽象并用数学方法模拟的．白箱模型的构造过程主要是进行物理机制的分析，几乎不依靠样本数据，已知的样本数据主要用于对模型进行可靠性检验．黑箱模型则不同所研究系统的内部结构及其中各因素的关系，仅根据用于建模的样本数据确定输人与输出的对应关系，因而对样本数据的要求较高，不但要有准确和足够的数量而且要包括各种可能出现的情况．黑箱模型建模所用的数学方法，一般是数理统计，建立的多是统计数学模型．一般用于对所研究的对象了解不多的情况．一般地说，完全的白箱模型和完全的黑箱模型都是很少的，绝大多数模型是灰箱模型．特别是在地学中，一方面由于研究深度不够而不能构造完全的白箱模型，另一方面又由于样本数量不足而不能建立起可靠的黑箱模型．因而只能取长补短建立介于两者之间的灰箱模型．这有两种情况：其一是既对研究对象的系统结构和各因素之间的关系有所了解，但还不能把它们完全确定下来，因而就要以建模样本数据提供的输人、输出的关系作为建立模型的重要依据；另一方面，资料数据又不足以建立可靠的黑箱模型，故必须同时考虑所研究对象的系统结构和各因素之间的关系，才有可能建立一个实用的数学模型．另一种情况是模型中有一部分是白箱模型而另一部分却是黑箱模型．实际中更多的情况是两者都有的灰箱模型．

地学是一门高度综合的科学，它是建立在多种基础学科之上的，这就决定了地学问题的研究深度和精度都受到基础学科的制约．这是地学模型极少白箱模型的一个重要原因，因素的多元性和变异性又带来了问题的复杂性和不确定性．所谓多元性是指一个地学系统

所包含的或影响一个地学事物（现象）的因素很多．例如土壤的性质除了受气候、植物、微生物、土壤母质、地貌条件等多种自然因素的影响外还受到人为活动的影响，而且其中不少方面含有多个因素．例如气候因素中至少应包括辐射、温度、降水、湿度等，地貌条件显然至少应包括海拔高度、坡度、地下水等，人为活动则显然应包括利用方式、灌溉、施肥等．它们之间的关系十分复杂是已为多数学者所公认的．如果对所论对象了解不够，需考察那些因素与所研究的对象有关，那样所涉及的因素就会更多．这样在很多情况下，甚至要获得起码数量的有关资料数据都不是易事．此外，有些因素对目标的影响是多途径的，例如地貌条件中的海拔高度既影响降水等气候因素，也影响植物、微生物等因素，而间接影响土壤之性质，再加上各因素的交互作用，问题就更加复杂了．所谓变异性是指所研究的对象（例如土壤性质）及与之有关的各因素（如气候、植物、地貌等）在空间上都存在着差异，在时间上都在不断地变化着，要获得完全同步的所有有关资料数据，可以说是不可能的（至少目前是极难的）．当然实际上在研究地学这样多元性问题时，总是仅考虑主要因素而忽略其它不太重要的因素．但是即使仅是主要因素完全同步的资料数据也是不容易取得的，而未被考虑的因素的差异还会引起对象或某些主要因素的变异．更主要的是，对象（如土壤性质）和因素（如地貌、气候等）几乎都有一个发展过程，而这些过程的发展速度可以相差很大，以致不可能用同一时间尺度来量度．例如土壤性质的变化就远比气温、降水等气候因素的变化慢得多．因此在研究气候条件与土壤性质的关系时，往往采用气候要素的平均值，而（实际上）当所取的统计年限不同时，所得到的平均值当然也就存在着差别．所有这些都会使得研究对象及其有关因素存在着变异性．这里存在的不确定性包括两个方面：一是由于变异性带来的资料数据以及研究对象本身的不确定性，这种不确性主要是地学学科特点所决定的；另一则主要是由多元性带来的数学解的不确定性，这种不确定性是数学局限性的一种表现．

地学又是研究事物在时间和空间中分布及变化规律的学科．时间、空间的连续性必然带来中介过渡性．排中律的破缺必然带来模糊性．实际研究中，当然不可能（对任何地区）逐点逐时地进行研究，只能是分类，分区选用典型样本进行研究．然而实际上，无论划分为多少类都不能完全避免出现既象甲类又象乙类，或者任何一类都不象等过渡性现象．仍以土壤性质而言，实际上从某类土壤到另一类土壤，无论空间还是时间，都不是跳跃式的，大多数都是逐渐过渡的，而且划分类型所依据的因素（即划分的指标项）越多，这种客观存在的中介过渡性就越复杂．那些处在中介过渡的样本，不遵循非此即彼的排中律，这种排中律的破缺所带来的不确定性乃是模糊性．它显然有别于由随机性引起的不确定性．精确数学是解决确定性问题的有力工具．分明，确切，是其主要特点与长处．而上述两种不确定性的客观存在必然限制了精确数学的应用．

四、地学与数理统计

数理统计是处理具由随机性引起的不确定性问题的有力工具．但是由于地学研究对象具有极强的地域性和时序性，而且大多数过程是不可逆的，使得自然界中同类对象（事物、现象、过程等）发生的次数是有限的，被观察记录下来的就更少．加之因素的多元性，能够用作数理统计的样本数量一般都远不能满足大数定律的要求．就以有专门观测队伍的气象和水文资料而言，虽然有些站的观测记录在一百年以上，但多数站的观测年分都不足百年，在区域分异研究中，为了各站资料能够互相比较，所有统计量都必须取同一年

限，也就是说只能按最短的观测年限进行统计，而且客观上存在着多年变化，所以在研究空间变化时不能简单地用增加统计年分来消除或减少随机性的影响．当研究时间变化时，若试图用空间多点的同时观测来减少随机性的影响，则又受到地域空间的变化所限制．即增加观测点数不单要增加工作量（对有些地学现象或地学要素所要增加的工作量是相当大的，往往已经不易办到），更重要的是客观上存在着的空间地域差异，随着测点增加，有关区域随之扩大，它们的差异也就愈大；这样相当一部分资料数据也就没有重复性意义．实验（方法）是弥补自然对象（事物、现象、过程等）资料数据不足的另一种办法．但地学过程很多是十分缓慢的，而且规模十分巨大，这类巨大而缓慢的过程显然是很难甚至是不可能进行实验的．特别是综合性过程，因素的多元性本身就是实验的一大障碍．这不单因为多元性本身带来的复杂性（它大大增加实验设计的难度），更由于多种因素中只要有一个因素不能用人工方法模拟就使得整个综合过程都无法进行实验．从上述可见，数理统计方法在地学中的应用受到很大的限制．严格地说，由于地学问题可用于分析研究的资料数据在数量上常常不能使概率达到稳定，即不能满足数理统计的基础——大数定律，应用数理统计方法得到的结果不能认为一定是确实可靠的．这是在使用数理统计方法时不可忘记的．当然，数理统计方法确是一种经常使用的数学方法．因为虽然由于样本数量不足而得不到稳定可靠之结果．但人们无论用什么方法（包括数学的与非数学的）都只能根据已有的资料数据认识世界和作出判断，而数理统计方法则是一种从已有资料数据中获取较客观、较全面、较集中的综合信息的方法．不过人的思维与抽象后的数理统计方法是有区别的，至少有以下区别：1）人的思维总是自觉不自觉地以前期的认识为基础的．例如一个人对一个剖面的孢粉分析数据的认识是以对各种孢粉的一般可能出现的概率（它与各种孢粉的产出量、保存率、散布和沉积情况等有关）为基础的，即某种孢粉出现的多少及其意义是与这个头脑中已有的一般概率相比较而言．而在数理统计中则一般地仅从所分析的那个样本中各种孢粉含量的百分数或与整个剖面各样本的比较来衡量．所以无论由谁来作数理统计结果都是相同的，因此比较客观，而按人的思维，则一个有经验的孢粉工作者与一个对孢粉不熟悉的人或两个有不同经验的人会得出不同的结论．2）在人的思维中，总是把数值与其实际含义相联系，而数理统计则只问其数值而不考虑其含义．例如在土地评价等级中“9”级与“10”级和“1”级与“2”级，在人的头脑中都是相差一个等级．而在数理统计中，9 与 10 的差别和 1 与 2 的差别则只有仅考虑绝对差时才是相同的．3）在人的思维中，对每一个数据都会考虑它们的可信程度．而在数理统计中则不考虑它们的可信程度．仍以孢粉分析为例，在数理统计中，在计算两个孢粉含量的绝对差时，100 与 101 和 1 与 2 同样相差 1，当计算比值时 100 与 200 和 1 与 2 之比同是 1／2，但在人的思维中绝不会把 1 与 2 的差别和 100 与 101 的差别相提并论，也不会把 1 与 2 的差别和 100 与 200 的差别等同看待．4）在人的思维中不单考虑各因素的重要程度．而且对每个样本的代表性也常常需要考虑（虽然这些考虑有时是不自觉地进行的）；而数理统计中，除有意识地用权重反映各因素的重要程度外，对任何样本和因素都是等同对待的，从不考虑各样本的代表性．5）上述思维中的种种考虑常常是不自觉地进行的，所遵循的规则更多是模糊的，不完全统一的，因而较难避免片面性和主观性．而用数学方法分析，则把整个分析过程都用形式语言表达出来，因而可以对整个分析过程作客观的评判、讨论和修正，从而有可能减少主观性和片面性，还可以起到集思广益的作用．

五、地学与模糊数学

模糊数学是描述和解决带有模糊性问题的数学工具．其关键在于隶属函数的建立．而隶属函数的建立主要（甚至几乎完全）依靠所涉及的专业知识．数学方法一般只在诸如专家咨询方法等专业知识统计分析上起到一些作用．不过这可能正是模糊数学的生命力之所在，因为这使得一下把问题全部推给数学成为不可能，在避免过分依赖数学上起到较大的作用，而建立的隶属函数则在保证定量与定性结合上起到重要的作用．所以虽然模糊数学至今尚没有为数学提出一个新的基石（正因为如此，至今仍有人不承认模糊数学是数学的一个分支）．但模糊数学确实在解决实际问题中起到重要作用，并在应用中得到发展．当然这还在于模糊数学所描述的确实是一种与随机性完全不同的另一种不确定性．虽然至今对这两种不确定性描述和处理的抽象形式语言十分相似（可以说表面完全相同），但却有着完全不同的意义．从另一角度来说，这种主要依靠专业知识建立隶属函数的方法，使得高度抽象的数学向实际靠拢了一步．模糊数学正是靠这一点得以在应用中发挥作用和不断发展的．对存在着大量模糊性问题的地学，模糊数学当有较大的用武之地．当然地学中的模糊性有些是很复杂的．例如反映植物适应（或生存）环境的生态龛，不但在各因素组成的环境空间中是一个模糊区域，而且由于任一种植物都是在对环境的不断适应中发生变化，不断演化的，因而在时间上这个模糊区域又是不断变化着的．这种变化又是与环境条件及其变化相关联的．而这种变化的幅度和速度又是有一定极限的，这种极限同样也是一个模糊量．在地学研究中显然还有比这更复杂的模糊性问题．此外地学的综合性要求广博的知识，随着广度的增加，广与深的矛盾就越加突出．这就限制了地学研究的深度．而只有较深的理性认识才有可能用形式语言加以表达．所有这些都会给建立隶属函数增加难度，也就使模糊数学在地学中的应用受到了限制．

综上所述，已可以看到以往地学研究侧重于描述性的定性研究，主要使用自然语言，较少运用数学方法，是地学学科特点以及数学的局限性所决定的，是科学发展之必然．计算机技术的发展，模糊数学等新的数学方法的问世，减少了数学的局限性，开拓了数学的应用范围，而各有关学科研究的深化和地学学科本身研究的进展，使得对问题的认识逐步深化．这样一个由定性向定量转化的新时期已经到来。当然在研究工作定量化上，地学仍然面临着艰巨的任务．定量研究与定性研究是不可能截然区分的．更不能设想把构造数学模型的工作交给对所论问题缺乏必要认识而仅精通数学的工作者去做．事实上，一个数学模型是否成功，效果如何，既受模型构造者的数学水平的影响，更由构造者的专业水平所决定．因而地学研究的定量化，需要地学工作者和数学工作者的共同努力．了解定性与定量的关系，了解专业知识和数学知识在构造数学模型中的作用，对研究工作的定量化是有帮助的．

第二节　数学模型的建立

数学模型是对人的思维本身或反映客观事物、客观规律的思维的模拟．如果把象综合评判一类的思维本身和人的思维所反映的客观事物或规律都称为现实原型，则数学模型与现实原型的关系是反映与被反映的关系．因为数学模型的构造必须经过抽象分析过程，即经过对现实原型扬弃次要环节的过程．故数学模型只能与现实原型具有相对一致性，事实

上往往只能在基本环节上（这些环节或关系结构是能用数学形式表现出来的）有近似的一致性。这有点象美术中的写生或文学中人物典型的塑造那样，都是某些现实原型的本质或特征的近似刻划与集中反映而已。

一、确定数学模型类别

对所研究的实际问题，即对现实原型要分析其对象与关系结构（包括量变因果关系）的本质属性，以便确定其数学模型的类别。由于地学中不少问题不经过认真分析就不可能掌握其本质属性，因而在建立地学数学模型中，这一步骤更为重要。例如：某个自然带（譬如北亚热带）的北（或南）界之类的"自然地带划界问题"，这在地学中是常会遇到的。曾有人试图运用数学方法来解决，但如果缺乏认真的分析，是很难凑效的。因为这类问题实质上包含着两种性质完全不同的问题。其一是确定界线的具体走向（位置）的问题；另一则是确定所论概念（如某个自然带）的内函（或外延）的问题。曾经有过的一些争论表面上是具体界线位置问题，而实际上是双方对所争论的问题的关键概念（例如"北亚热带"）有着不同的认识（解释、定义）。当然，由于各自然带的概念含义，常常与它们的具体界线的确定同时形成。即在自然带区域范围确定时，区内的共同特性即为其含义，并可据之给出其定义。所以这两类问题常会被混在一起进行讨论。但当试图应用数学方法解决时，则必须把两者明确分开。因为它们需要用不同的数学方法来解决。若是要通过确定区域来确定概念的含义，则应该用"区划"方法划定区域范围（也就是确定其界限），然后归纳出区域内的特点作为该自然带的含义（定义）。这实质上是在一个研究区域内找出相对突变的带（线）。但不同区域按突变带（线）划分出的区域的性质显然是有差异的，而自然带的概念应该具有全球一致性，不应因研究区域不同而各异。所以当研究的区域不大，掌握的资料仅在一个不大范围内时，显然不能用这样的方法来解决划界问题。这就是另一类问题，即根据已有的对概念的认识，也就是根据有关自然带的内函（或外延）进行判别空间各点的归属。这显然应该应用模式识别方法来处理。若需要既考虑研究区内各有关要素的区域变异特点，又需考虑与一般的（全球的）概念的一致性，问题就复杂得多，恐怕不是现有单一的数学方法所能解决的。此外，自然地带界线是一种"地理界线"，而"地理界线"有自己的特点。首先，"地理界线"往往不是一条截然的分界线，而是一个（过渡）带。也可以说"地理界线"常常具有一定的模糊性，而且这种模糊性又是相对的。在大比尺小尺度研究中的一些界线显然具有模糊性，例如在研究山麓带时，山地与平原的界线一般是模糊的，即使山地与平原以悬崖相接，悬崖崖脚的堆积物仍然使山地与平原之间的界线具有一定的模糊性。但在小比尺大尺度的研究中，例如在研究全国（全省）的地貌（或土地）类型时，不少山地与平原界线则可以忽略其模糊性而视为分明的。其次"地理界线"还具有变异性（包括时间和空间的变异）。最常见的是界线随着时间的不同而变动。因为许多因素都是随时间而变的。例如气温就是在不断变化着的。对这种不断变化的因素，人们一般地总是从相对稳定平衡中去把握它。例如要研究气温区域差异及其分布特征，就经常用年平均气温之类的统计数值作为研究的依据。这种相对稳定的统计值显然地会因统计年限不同而异。它们的变化当然会引起含有这类要素的模型所确定的界线发生相应的变化。此外，"地理界线"既可以是客观上确实存在的一个突变带（线）的反映（不同的地段发生突变的各因素值可能有所不同，这是导致"地理界线"空间变异性的一个原因），又可以是在基本均匀连续渐变的区域中为了某种目的人为地划出的界线。更多的是一条界线中

某一段是客观存在突变的反映，而另一段则是在相对均匀渐变中人为划定（界线不反映客观存在的变异的空间分布特点）．在建立确定自然地带界线的数学模型时，至少必须充分考虑以上所述的各点．譬如考虑到“地理界线”的模糊性，采用模糊数学方法来建立这类数学模型显然是比较合理的．

二、确定主要因素和主要关系

确定所研究的系统并抓住主要矛盾，即必须考虑问题所属系统并选择具有关键性作用的变量和对其量的关系进行考察，也就是说要抓住主要因素、主要关系，因为数学模型所反映的应该是主要因素的关系结构．这种分析研究理应主要地从所研究问题的专业角度，依靠有关的专业知识，而不是从数学角度．虽然某些数学方法会有助于这种分析研究，但仅应作为一种工具，用于帮助人们获得某种有用的信息．例如计算刻划各因素值在样本范围（一般要求这些样本应该能代表研究范围）内的分异程度的变差系数，当样本确实能反映研究范围情况时，可以为选择因素提供有用的信息．因为分异性无疑是主要因素的必要条件，一个在研究范围内没有差异的因素当然不会是影响结果的主要因素．但显然并不是分异程度大的因素就一定是主要因素．更不能说因素的重要程度与分异程度一定存在正相关关系．况且计算获得的变差系数仅是数值上的分异，它并不等同于因素对结果影响大小在所研究范围内的差异．例如黄土高原地区降雨量的变差系数远小于人口密度的变差系数，但降雨量对土壤侵蚀强度的影响则远大于人口密度对土壤侵蚀强度的影响，因此绝不能仅仅根据变差系数的大小来选择建立模型所用的因素．其它数学方法也有类似情况．以建立生产潜力模型来说，所讨论的是在不同环境下，太阳能通过地球表面的植物（主要是农作物）被固化的最大可能值．这是一个太阳–地球的能量转化系统．其中太阳辐射是能量的来源，作物是能量转换的关键．其它环境因素是通过对作物生长的影响产生作用，可以说其它环境因素是影响生长潜力的间接因素．因此，在分析这些环境因素时，应该从它们对作物生长的影响来考察．譬如，当把水作为另一单个因素考虑时，对土壤这个因素，就应主要从供给作物养分的能力来考虑．一般可选取，有机质、氮、磷、钾含量，代换量，pH 值，盐碱含量以及土层厚度，障碍层，质地等与作物生长有明显关系的物理、化学特性等作为构造模型的要素，而不应选取土壤中的矿物组成，以及象 SiO_2 的含量等对作物生长影响不大的因素．虽然其中有的因素（在一些区域中）也有较大的变差系数．相反，若在研究范围内，因某种微量元素含量存在差异而影响作物的生长，即使其含量很小，变差系数也不大，也应该选作数学模型的因素．由于地学问题所涉及的因素很多，而获取信息又比较困难，所以有时要获取所有有关因素的信息也不可能，当然更不可能依靠数学方法选择主要因素．事实上，在选取主要因素时，还必需考虑其信息获得的可能性，因此有时要选取替代信息，间接地反映某个主要因素的况状．

三、进行正确抽象

正确抽象在这里指使用恰当的数学概念、数学符号和数学表达式，形式化地表达所研究的问题．如果遇到已知的数学工具不够用，即不能恰当地模拟人对所研究问题的思维，则需要大胆地创造，根据实际情况提出新的数学概念和数学方法，用以建立数学模型，达到进行正确数学抽象之目的．这里所说的正确的数学抽象包括从各因素的选取、标定到整个数学模型建立的每个步骤．这就必须对所论问题及所用的数学方法都有足够的认识．例如在进行模糊聚类分析中，若生硬地模仿分等评分方法对各因素进行标定，以致对定量因

素也作如下标定：将定量因素值（X）转化为分值（Y），令：

$$Y = i / a \qquad \text{当} \quad X_{\min} + (i-1)\,b \leqslant X \leqslant X_{\min} + ib$$

式中 $X_{\min}$ 为所有样本中该变量的最小值，$b = (X_{\max} - X_{\min}) / a$，$a$ 为划分的等级级数，$X_{\max}$ 为所有样本中该变量的最大数值．其结果不但降低了原始数据所提供的信息的精度，而且还会造成某些不合理的现象．例如当$X_1 = X_{\min}$，$X_2 = X_{\min} + b - \varepsilon$，$X_3 = X_{\min} + b + \varepsilon$（$\varepsilon$为一极小值）时，显然 X_1 与 X_2 的差别远大于 X_2 与 X_3，但 Y_1 与 Y_2 均取值 $1/a$，分值完全相同．而 Y_3 则取值 $2/a$，明显有别于 X_2．这当然不太合理，所以会作如此规定，很可能是由于不了解模糊聚类中仅要求各因素标定后，数值均落在[0，1]区间中，并不一定都要用分等评分．实际上分等评分一般仅适用于对定性因素的标定．对定量的因素完全可以采用诸如：$Y=(X-X_{\min})/(X_{\max}-X_{\min})$ 这样的方法进行标定．这样的标定就不会出现上述那种不甚合理的现象．正确的标定方法，应该是对各因素分别根据各自具体情况及其与目标的关系各自采用最适宜的标定方法，完全不必强求用同一方法进行标定．又譬如数量积法：

$$r_{ij} = \begin{cases} 1 & i = j \\ \left(\sum\limits_{k=1}^{n} X_{ik} \cdot X_{jk} \right) / n & \text{其它} \end{cases}$$

式中 $X_i=(x_{i1}, x_{i2}, \cdots, x_{ik}, \cdots, x_{in})$，$X_j=(x_{j1}, x_{j2}, \cdots, x_{jk}, \cdots, x_{jn})$ 是刻划 X_i 与 X_j 两个样本的相似程度的一种方法．但如果不了解或者忽略了它一般要求 $\sum\limits_{k=1}^{n} X_{ik}$ 为常量，而在不满足这个条件时也用它来计算，就会得出错误结论．例如若用人口密度、单位面积工业产值和单位面积农业产值三个因素衡量各县之间的相似程度也用数量积计算，则由于落后山区县这三个因素的指标值均小于平原县，使两个落后山区县的数量积小于它们与平原县的数量积，从而导出两个落后山区县之间相似程度小于它们与平原县之间的相似程度的错误结论．

四、数学模型的检验

数学模型建成，一般还必须把从数学模型分析得出的数学推论返回到现实中去，看看能否正确地回答实际问题．而且还不能仅仅简单地根据建模所用的样本数据本身是否得出合理的结论或是否与已知结果一致，来判定模型的合理程度或是否有效．因为这样所得到的仅是拟合程度，它反映的是所建的数学模型使用建模样本中各因素作输入得到之结论（输出）与相应的已知结果的一致．实际上建模的过程就是根据建模样本的输入与输出的关系不断调整的过程．而这样建立的数学模型所反映的关系既可以是所论问题的本质关系，也可以仅仅是建模样本的纯数值的（不反映实质的）关系．特别是在建模样本数量不多或种类单一时，找出这样的（纯数值关系的）数学结构是并不十分困难的．所以应该尽可能地用未被用于建模（即在建模过程中没有参考和据以调整模型）的样本来检验所建的数学模型，考查所获得结果与已知结果的一致程度，即所建数学模型的符合程度．因为只

有符合程度才能作为判断数学模型的可靠程度的依据．从理论上说，一个有效的数学模型必须保证在其适用范围内的任何一种情况（即各种因素任何可能的组合）都能够得到确定的合理结论．因此用以检验数学模型符合程度的样本应该包含各种可能遇到的情况，而且严格地说各种情况的比例应与实际发生的概率相同并有足够的数量，才能使得出的符合率与实际应用时的情况一致．但对地学研究而言，要满足这种要求可以说是完全不可能的．因为在地学研究中，由于资料数据极为有限，故一般都只能把已有的资料的绝大部分以至全部都用于构造数学模型．即使这样，对构造数学模型而言，也还是不够的，特别是在要从中取得必要的统计信息时，更难以保证统计结果的稳定可靠．所以对地学数学模型的检验往往需要采用其它方法．而不应停留在仅给出拟合率的状况．检验数学模型是否可靠的一种简单有效的方法是：假设各种可能出现的因素组合之典型数值，并将其输入数学模型中进行运算，再检查所推出之结果是否正确合理．只要这种假设的理想组合，确实能够概括所有可能遇到的各类情况，便确实可以达到检查数学模型是否有效可靠之目的（虽然这种方法一般不可能得到具体的符合率值）．例如当选定水质、埋藏深度和贮量三个因素对地下水开采条件进行综合评价时，需要检验综合评价模型，可以假定三个因素分别取：1）很好，2）中等，3）很差三种状态．这样就共有 27 种组合．若最初所建的综合评判模型是最简单的平均评判模型（即综合评判值就取三因素单项评判值之平均值），当把 27 种组合输入模型并算出其综合评判结果时，便会发现：水质评价很差，埋藏深度评价和贮量评价都很好的组合，其综合评价优于三项因素单项评判均为中等的组合．这当然是不合理的．因为水质很差即使水量很大，提取很方便也没有开采的价值．由此便可以说明，最初建立的模型不能有效地解决所论的问题，遇到这种情况就必须对问题抽象的正确性和数学方法表达的正确性两个方面进行检查．对问题抽象是所论问题的专业知识，研究水平和逻辑思维问题．譬如对地下水开采，取这三个因素是否正确，对各因素的单因素评判应如何进行以及三个因素如何进行综合评判等．为了简化问题，假定取这三个因素能够对地下水开采条件进行综合评价，各因素的单项评判已经解决而且各因素之间没有交互作用，而水质很差，埋藏深度与贮量评价都很好的综合评价应该差于三者都是中等的状况，因为埋深和贮量的优良是不能弥补水质差的弱点的．而平均评判模型刻划的是各因素之间可以互相弥补的综合评判问题．也就是说对这种不能互相弥补的综合评判模型选用平均模型是不对的．也可以说这里所用的数学方法是不正确的，应该改用其它综合评判方法来表达这类问题．这就是在选用已有的数学方法时需要检查的内容．若所用的方法是自己创造的，则除了要检查所用的方法是否恰当（正确）地表达所要表达的内容外，还需要对所创造的方法的推导过程进行检查，以保证所建立的数学模型能够得出确定的可用的结论．

最后，数学模型还应该具有化繁为简，化难为易的作用．也就是说，所建的数学模型不单必须较之现实原型简单、明了，而且应该是越简单越好．在用数学模型解决实际问题时，总会有一定的误差；一个好的数学模型应该具备估计误差范围的功能．这些误差一般应该主要由未包括在数学模型之内的因素所引起，特别是当模型建得较好（即所用的数学方法确实能正确地表达所论之问题）时．当然，建模型中所用的数学方法不够理想、或参数的确定不够正确等也是数学模型产生误差的原因之一．不过这种误差可以在建造模型中或建成模型之后通过对模型的改进而加以克服或减小．由于数学模型，特别是综合性极强的地学数学模型，不可能把所有的有关因素都包括在内，因此由未被包括在数学模型中的