

# 数学地质简介

徐冬国 罗秋霞

地质学是一门复杂的科学，它历史悠久、因素复杂、情况多变。地质工作者通过各种途径、多种手段研究它、认识它并企图掌握它。但是由于科学水平的限制以及它本身的局限性，以前大部分尚行尚在肉眼观文和定性描述阶段。随着各种学科的迅速发展，各种新仪器新方法相继引入地质领域，宇宙地质、深部地质、遥测技术和各种新的化探物探方法等相继问世，这时依靠传统的经验整理资料已远不能适应形势发展。如何及时地科学地整理各种资料，充分提取其中信息，以进一步认识并掌握地质规律，这已经是广大地质工作者颇感兴趣并亟待解决的课题之一。数学地质——地质学和数学之间的一门新的边缘学科就是在这样的形势下诞生和发展起来的。

不同性质的事物必然在数量上有所反映，有所差别。反之，量变引起质变，数量的变化亦必然是反映了事物的不同性质。因此研究这些数量之间的关系是有意义的，它不是纯粹的数学游戏，而是深入认识事物本质的方法之一。

马克思曾经指出：“科学仅当它成功地利用数学时才达到完善的程度。”应用数学的理论和方法，借助于电子计算机，定量地研究地质问题的一套理论和方法——数学地质的兴起与发展，它为数学的应用，为地质学的发展开辟了崭新的前景。特别是近年来由于电子计算机的迅速发展和普遍应用，使以前人工计算无法胜任的数学方法可能付诸实现，从而大大促进了数学地质的发展。虽然数学地质目前还处于新生发展阶段，但是它已经渗透到地质学的各个领域，如沉积、地层、岩石、构造、矿床、古生物、地壳、水文、化探物探等各个方面，并在国内外一些单位中取得了可喜的成果，因此它的普遍应用将是指日可待。

电子计算机是本世纪重大科学发明之一。它的出现对科学技术和国民经济各个领域产生了深远的影响，在地质学中亦不例外。国外目前计算机主要向巨型、微型、网络化发展。巨型计算机速度越来越快，每秒运算可达一亿五千万次。内存越来越大，可达几十万、几百万个单元，并大量采用半导体存储器。外存大量发展各种磁带和磁盘。小型计算机只有手提包那么大，有的微型机仅如钮扣大小。电子计算机的外部设备花样越来越多，稳定性和精度越来越高。如最快的绘图仪每秒可走1.5米，同时用四种颜色绘图。用通信线路把多个分布在各地的计算机联接起来组成计算机网络，以便共享网络中所有的软、硬件和数据。为了充分发挥计算机功能以及方便用户，各种软件包括操作系统、管理系统、各种程序设计语言（例如Fortran、Algol及Basic等），编译程序以及检查和诊断程序等近年来也迅速发展。

从五十年代开始，国外地质逐步引进电子计算机技术，到目前为止，几乎渗透到地质工作的各个方面，诸如石油勘探、开发、集输以及存储资料等。据统计美国应用于这一方面的电子计算机有104台，西德16台，加拿大15台，英国10台，意大利8台，中东5台，其它如澳大利亚、委内瑞拉、荷兰、法国、瑞典、挪威、比利时、巴西和印尼等也都使用了电子计算

机。在我国国内，已经发展了第二代、第三代电子计算机。几个大油田也都已建立了电子计算机站。介绍和引进了国外的先进技术和统计方法。对软件以及地质上常用的计算机程序做了大量的工作。但总的说来，国内目前还处于初创和摸索阶段，计算机网络和终端还是空白。

数学地质应用电子计算机作为基本工具，从而已现了它飞速发展的前景。没有电子计算机，数学地质的进一步发展是难以想象的，或者说有些方法诸如多变量统计分析以及数学模拟等也难以付诸实现。

数学地质涉及的内容和范围很广，从方法的角度来看，大约可以分为三方面内容：应用概率论数理统计以及其它一些数学理论对数据进行统计分析、地质过程的数学模拟、数据资料的处理和存贮。

一、数据资料的处理和存贮就是把各种类型的资料按地区、井种、类别进行编码、存入穿孔卡片、磁带或磁盘中。索取时填写一定格式的命令，通过电传打字机或穿孔纸带输入计算机，由机器自动检索并按要求以图表、文字、数字等形式输出。美国在研究落基山地区时，七天内从井史控制系统档案中对一万口探井资料进行了分析解释，并提供了四张构造图和五张等值线图。这些工作量如果用人工则需三个人搞一年。我国大庆油田用计算机建立了勘探、地质、实验等五个方面资料库，存贮了一百万数据。

二、在地质历史的动态发展过程中，地质体经历了许多次变化，使地质现象显得很复杂。通过对地质现象的特征及其变化过程的观研究建立数学模型，通过计算机进行计算和逻辑运算，模拟地质过程和地质特征，并考核结果，在实践中对比检验，修正模型，以得出正确的数学模拟公式。也就是说用数学模型去模拟该系统的历史形成发展过程，诸如构造模拟、沉积模拟、生油模拟等。法国石油研究所蒂索在研究巴黎盆地托尔辛阶时建立了动力学方程模拟有机物转化的历史过程，用来计算原始生油量，计算值与实际观测值之间吻合良好。美国阿拉斯加州普鲁德霍湾油田的第一口探井发现油气后，结合该井有关数据，选用密西西比河的三角洲模型，对普鲁德霍湾油田进行了模拟计算，对地质条件及油气水边界进行了详细研究。在此基础上对面积约2000平方公里的构造进行详查，效果很好，共钻详探井53口，井距3~7公里，在一年时间内基本搞清情况，拿下可采贮量13亿吨的大油田。在美国的一些院校、研究部门，包括石油公司正在把运算模拟实验作为新的资源勘探的手段加以利用。

三、地质数据资料一般是指标多容量大。如何根据地质工作的要求，选用合适的数理统计方法进行整理、分析和综合，找出所研究的地质体的数量特征和内在统计规律性，以得出较为正确的结论，这就是多元统计分析的内容，也是数学地质中所占比例较大、应用较多的重要内容之一。一些单位在应用中已见成效。如美国落基山粉河盆地白垩系莫迪砂岩贮集层用趋势面分析对其分布进行研究，找出其规律性。通过勘探后共发现石油可采贮量3380万吨。而在未使用趋势面分析之前，该盆地某些含油气的地质特征是无法知悉的。加拿大阿尔伯达盆地南部地区通过趋势面分析研究前白垩系剥蚀面与上伏砂岩（贮集层）的关系，发现已知油气聚集与剥蚀面的异常符合很好，从而予报了一些尚未勘探的可能油气田的位置，使它们成为今后的重点工作区。在墨西哥卡比多灰岩碳酸盐微相的研究中，用点群分析在一个采集了大量岩样的地层剖面中为客观地划分微相提供了一个简易、快速、半定量的方法和初步看法，为进行更详尽的研究打下了基础。国外在用因子分析研究沉积相方面也取得了良好的

效果。在我国，大庆油田应用趋势分析研究油层分布规律，用逐步回归筛选因子提高了油田重要参数（例如含油饱和度）的预测精度，用聚类分析和主成份分析研究油层沉积背景，对油田的开发起了积极作用。胜利油田用回归分析研究影响渗透率的因素，提高了利用测井资料编制渗透率解释图版的精度。四川石油管理局用趋势分析研究构造，确定井位；四川二普用趋势分析研究须二段厚度变化规律；三普用判别分析在渭河地区应用重矿孢粉资料划分地层；成都地院石油系用逐步多组判别分析研究苏北高邮凹陷阜三段沉积相；武汉地院石油教研室岩相科研组、甘肃石油地质研究所等单位用因子分析研究岩心粒度分析等都取得了一定的成效。在寻找金属矿方百有关单位也做了大量工作，取得了不少成果。

有关多元统计分析的内容很多，现仅就应用较多的几个方法作一简要介绍。

### （一）聚类分析和判别分析

这是划分和判别样本类型的一种统计方法。“物以类聚”。根据样本所具有的各种指标，定量地确定各样本相互间的亲疏程度并据此进行分类，这就是聚类分析的内容。如果已知类型，而要判断未知样本归属其中那一类，这属于判别分析的范畴。地质学中分类的问题是很多的，如岩石分类、矿物分类、化石分类、辨认堆积物来源、确定地层划分、区分含油层与含水层等。

聚类分析根据分类对象的不同一般可分为Q型群分析和R型群分析。其方法是一样的，只是一个是对样品分类，一个是对指标分类而已。把观测对象看作是  $n$  维空间的向量，若他们间或夹角越小或距离越近或相关性越大，则这两观测对象就越接近，它们就可能属于同一类别，因此可根据样本的相似系数或距离系数或相关系数的大小来对样本逐级进行分类。

如果要在两组地质体之间求出一个判别函数以便将它们合理区分，对于未知样本可根据观测数据判断其归属，这是两组判别。判别函数的一般形式是

$$y = c_1 x_1 + c_2 x_2 + \dots + c_p x_p$$

$C_i$  为待定参数。它可根据费歇准则即两类间的区别最大，而每类内部的离散性最小的原则化为求极值的问题，通过解线性方程组而求得。由判别函数就可确定判别标准，然后进行判别。

对于多组判别则不是建立组间的判别函数，而是计算新样本属于各组的概率  $P_i$ ，比较各  $P_i$ ，将样本归入概率最大的一组。所以关键是要求出概率  $P_i$ ，而  $P_i$  是建立在贝叶斯公式基础上的

$$P_i = \frac{e^{f_i}}{\sum_{i=1}^G e^{f_i}} \quad (G \text{ 为组别数})$$

$f_i$  可由样本观测数据通过计算得到。

判别分析是根据样本的多个变量（指标）来进行判别的，而在这众多的变量中是否每个都对判别有效？应加以甄别。因为变量太多不仅繁琐而且有时也会降低判别效果，使判别不稳定。而变量太少又不能包含变量所提供的所有信息或大部分信息。因此适当地选择变量是很重要的。选择变量的标准应根据每个变量的判别能力的大小，也就是组间离差和组内离差之比的增长情况，挑出判别效果最大的变量使其进入判别式，剔除作用不显著的变量将其退

出判别式。从一个变量开始，每次引进或删除一个变量，反复进行，直到不能删除也没有引入变量为止。选定变量以后，再根据多组判别方法进行判别。

## （二）回归分析

回归分析是定量地研究各变量间相关关系的一种统计方法。地质工作者常用来进行预测和控制。

如果各变量间存在着一定的相关关系，这些关系如果是已知其函数形式，而未知具体参数，这时只要根据函数形式适当估计未知参数；如果是未知函数形式，这时就不仅是估计参数的问题，而是首先要确定变量间的函数形式；这些都是回归分析要解决的问题。一般可根据大量观测和实际经验大致确定变量间关系的方程式，估计其中参数值，然后再对方程进行检验，也就是检验诸变量间是否存在这样的关系。

在许多实际问题中，变量间不一定呈线性关系，但是通过一些数学变换可以使某些非线性关系化为线性关系，因此讨论多元线性回归模型是有意义的。假定变量间具有线性关系

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p$$

$a_i$  为待定参数。可由最小二乘法原理，使观测值与回归值间的偏差平方和达到最小；从而求得  $a_i$ ，便得回归方程  $y$ 。

与逐步判别相似，在建立回归方程时，应对各变量进行筛选。挑选重要的对因变量  $y$  起主要作用的一些变量将它们引入回归方程中，而把一些次要的对因变量  $y$  不起显著作用的变量剔除出回归方程，从而使回归方程只包含影响显著的变量，既具有较高精度又比较简单。挑选变量的原则是根据每个变量对回归方程的贡献大小。从一个变量开始，每次剔除或引进一个变量，逐步进行，直到没有引入也不能剔除时为止。逐步回归分析是在多元回归基础上派生出来的一种计算技巧。

趋势面分析是多元回归分析中的一种特殊情况。它是一种多维（二维或三维）高次非线性回归分析。趋势面分析在整理化探、物探和航测数据，了解矿物分布趋势、富集部位、指示找矿方向等方面是一种行之有效的方法。计算趋势面的数学表达式最常用的是多项式函数。根据多项式的次数称为一次趋势面、二次趋势面……。方程形式为

$$Z = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + \dots$$

其中  $x, y$  为地理坐标。与回归分析同样，用最小二乘法估算待定参数  $b_i$ ，便得出趋势面方程  $Z$ 。用该方程所表示的趋势面来逼近（或拟合）地质体某些特征的趋势变化，从而可把观测数据分为三个部分：受大区域因素控制的区域性变化；受局部因素控制的异常；由随机因素造成的偏差。由观测值减去趋势值和偏差值后便为异常值，并可作出相应的等值趋势图和剩余异常图。

## （三）因子分析

因子分析是研究系统分类和成因分类的重要手段之一。在研究样品之间、变量之间或物质成份与地质作用之间的相互关系时，借助因子分析能从众多的数据中提取少数有代表性的独立的新因子来反映研究对象所提供的大部分信息，从而把握住了它们间的相互关系，简化了观测系统，减少了研究对象。同时主因子本身往往揭示了研究对象在成因上或空间上的联

系，可以直接用于地质解释和推断。这是一种抓主要矛盾的统计方法。

因子分析根据研究对象是样品还是指标又分为主因子分析（Q型分析）和主成份分析（R型分析）二种，其方法完全类似。

假设一组相关样品 $X_1, X_2, \dots, X_n$ 可用少数互不相关的主因子 $f_1, f_2, \dots, f_m$  ( $m < n$ )来表示，即

$$\begin{aligned} X_1 &= a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m \\ X_2 &= a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m \\ &\vdots \\ X_n &= a_{n1}f_1 + a_{n2}f_2 + \dots + a_{nm}f_m \end{aligned}$$

系数矩阵中各元素 $a_{ij}$ 称为因子载荷，它反映了样本与主因子间的相关关系；每行平方和表示全部主因子对样本 $X_i$ 总差方的贡献；每列平方和是表示每一主因子对诸样本所提供的方差贡献，它是衡量主因子相对重要性的指标。主因子以及因子载荷是根据各因子方差贡献，由大到小逐个提取，直至样本的主因子方差被分解完毕为止。具体计算可由样本相关矩阵求特征值和特征向量而得到。因子的个数 $m$ 的大小可根据主因子方差贡献的累积百分数达到80%以上而确定。

有时为了使因子载荷矩阵结构简单，各因子所代表的地质意义明确，可对因子载荷矩阵施行方差极大正交旋转。如果研究的对象不是互相正交，而是斜交的，这时也可以研究斜交因子模型。

#### （四）典型相关分析

典型相关分析是研究两组变量间相关关系的一种方法。例如生物群与生活环境之间；两条区域剖面的对比结果；矿体化学成份与围岩化学成份之间关系等等，这些都是两组变量“集团”之间的关系。而研究两组变量间关系比较复杂，若逐一研究不易抓住本质又较繁琐，典型相关分析就是在两组变量间各选择有若干代表性的综合指标，然后通过研究这两组综合指标的相关关系来“揭露”两组变量间的内部联系。它的基本思路与主因子分析中提取主因子的思想相类似，主因子是按照方差贡献的大小，由大到小逐个提取，直到变量的公共因子方差被全部分解完毕为止，而典型相关分析则是按照两组变量间相关性由大到小逐对提取，直到两组变量间的相关性被分解完毕为止。

数字地质虽然在我国运用不久，但已初见成效。随着科学技术的不断发展，通过广大地质工作者的辛勤努力，数字地质一定会在地质科学各个方面发挥更大的作用。