

DQHXZDDYFX

地球化学中的多元分析

中国地质大学出版社

胡以铿 编著

绪 论

近三十年来,在电子计算机应用的促进下,多元分析在地质地球化学领域中得到了迅速发展,对地球科学的定量化、地质数据的自动化处理、地质过程的计算机模拟以及地质思维的智能化等方面起到很大的促进作用,有力地推动了古老的地球科学与现代科学的结合。可以说多元分析方法的应用已经远远超出数据处理的范围,已发展成为地质和化探工作中的一种富有生命力的新的研究手段,从而使地球科学在理论上和方法上产生了重大的变革。

一、地球化学中的多元分析的涵义

地球化学中的多元分析属于数学地质的一个分支。它以多元统计分析为主要手段,以电子计算机为工具来研究地球化学及化探问题,将数学与地球化学密切结合,使地球化学数学化、定量化,以便用计算机进行研究。

这里说的多元分析即多变量统计分析,它研究的对象是多变量的数据集合。这种数据集合一般由 n 个个体,每一个个体都有 p 个变量的观测值构成。形式上,可以把多元分析作为数理统计学在研究客观事物中多种因素间相互依赖的统计规律性的一个分支。多元分析主要讨论以下几方面问题:

1. 讨论数据集合中变量与变量、样品与样品、变量与样品之间的相互关系或依赖关系。
2. 对变量或样品进行分组、归类或判别。
3. 研究数据集合的空间展布和时间序列变化规律。
4. 简化数据结构。把具有内在联系的几个变量变换为一个新的独立变量或者约简数据集合的空间维数,用较简单的形式来表示所研究的复杂的多元数据集合。
5. 提出假设和检验假设。主要是探讨以上四个方面所得结果的有效性、可靠性或可信性。

较常用的多元分析方法有:相关分析与回归分析、聚类分析、判别分析、趋势面分析、因子分析、对应分析、典型相关分析、马尔科夫概型分析和地质统计学。其中,回归分析、因子分析、对应分析、典型相关分析是研究变量和样品属性及依赖关系,聚类分析、判别分析等是对样品进行分组和归类,而趋势面分析、地质统计学及马尔科夫概型分析是探讨变量的时空展布。

地球化学中的多元分析的任务除了介绍这些多元分析的数学方法、原理以外,还要讨论它们在地球化学工作中的应用条件,所能解决地球化学问题的范围以及所计算得到数学结果的地球化学解释等问题。

二、迅速发展的原因

地球化学中的多元分析是现代地球化学、数学和电子计算机技术迅速发展的必然结果。从现代地球化学和地球化学勘查的发展趋势来看,主要表现为以下3个方面:

1. 从经验到理论

在地球化学特别是地球化学探矿工作中,首先得到大量的事实规律。这些从实践中归纳出来的事实规律属于经验,需要上升为理论,用理论来解释它们,并反过来用于指导地球化学找矿的实践,目前很多经验性的事实规律已经被地球化学的理论所解释,但也有不少问题还有待于理论上的深化。可以说现在正面临着地球科学飞跃的前夕。在这个飞跃中,数学也将起到它应有的作用。

2. 由定性走向定量

地球科学与其他自然科学相比,更多地使用定性的描述。研究人员根据自己所观察到的地质现象作人工的归纳、演绎和推理。长期以来地质工作者较多地使用文字和各种图件来记录、反映地质现象。这样使得地质工作的野外工作量大,而且人为性强。对同一地质现象在不同研究者的文字描述中不容易作数量上的对比,因而也不利于地质工作的自动化。随着科学的发展,人们愈来愈强烈地认识到“一种科学仅当它成功地运用数学时,才算达到了完善的程度”。数学这门研究现实世界数量关系与空间形式的科学近年来很快地渗透到地质科学中,人们愈来愈渴望用简明的数学关系式或图表来表征各种地质现象和地质过程。正如S·C·鲁宾逊所说:“地质数据若用数学方法整理、核对、综合、计算和检验,则对地质解释、分类、概念检验、问题解答以及概率预报,能提供较之只限于任何个人的经验和记忆更为全面和客观的基础。而且,这样的基础可以作为许多地质人员的共同基础,他们每个人再加上自己的经验和知识去完成他们自己的解释工作。在推进地质学这门科学向前发展的过程中,这种共同基础正好可以成为一个主要因素。”由此看来,数学和地球科学的结合已成为地质科学向前发展的一个必然方向。

3. 从单变量到多变量发展

客观世界中任何事情都是与其周围的事物或者与其所处的环境存在着千丝万缕的联系,在地球化学工作中也是一样,仅仅根据单变量的观测数据来分析问题已经远远不能满足研究工作的要求,而必须同时对多个变量进行观测、汇集与研究,与问题全部有关的原始资料,较全面地研究它们之间的相互关系和相互影响,去粗取精、去伪存真,才能得出正确的结论。

地球化学和地球化学探矿不同于其他地质类学科的另一特点是它具有大量可测定的数据。仅从测试化学元素方面来看,从40、50年代只分析几个主要成矿元素,到60、70年代发展为既分析成矿元素又分析控矿元素和运矿元素。80年代以来,则向测试更多元素方面发展。如地质矿产部1:20万化探扫面工作规定每个样品必须分析40多个元素,占元素周期表中能分析元素的大部分。

数学地质发展十分迅速的原因,从数理统计方面来看,各种统计方法越来越成熟,新的统计分析方法不断涌现,特别是多元统计分析发展很快,稳健统计分析和成分数据的统计分析方法的提出,使多元统计分析无论在理论上还是在应用方面都提高到一个新的水平。

50年代以来,电子计算机技术得到迅速发展,这为多元统计从数学方法的理论研究推

广到实际应用提供了必要的条件。如果没有电子计算机,要对一些数量较大的数据进行某些多元统计分析,如趋势分析、因子分析、判别分析等几乎是不现实的。电子计算机的使用对于多元分析在地质上的应用提供了一个必要的条件。

三、发展史及研究现状

早在1840年,英国地质学家莱伊尔(C. Lyell)运用统计的方法,根据近代海洋生物的相对含量,把第三纪地层中的生物种属与现代海生生物种属对比,成功地对第三纪地层作了进一步划分。

1890年,皮尔逊(K. Pearson)编写了《数学进化论贡献》一书,其中有古生物化石的统计分析。

1914—1934年间,俄国学者列文森——列辛格(Лёвнсон—Лёсницк)通过研究岩石的岩浆酸度系数的频率分布,探讨了安山岩、玄武岩、流纹岩的分类问题。

1936—1945年数学方法的应用由地质学的个别问题逐渐扩展到地质学的一些分支。在岩石学中引入了双变量或三变量分析,出现了三角岩比图、百分率图等。在矿物学和结晶学领域中,还运用了简单回归分析和方差分析,以了解矿物性质与化学成分之间的关系。1944年,苏联学者维斯捷列乌斯(А. Б. Вистеллус)在苏联科学院报告集上发表了《分析地质学》一文,提出用定量方法研究地质问题的初步思想。他从事了三十多年的数学地质工作,成为苏联数学地质的创始人和国际数学地质学会的第一任主席。

1945—1960年间,数学地质迅速发展,电子计算机也开始在地质工作中应用。数学统计学的方法普遍应用于地质学的所有分支。1946年,苏联学者研究金属矿床元素统计分布特点。1954年,绍(D. M. Shaw)等人应用统计方法研究地球化学问题。1956年,切叶思(F. Chayes)应用均值、方差于岩石学研究。1956年,克伦宾(Graybill)在研究岩石的矿物和化学成分时,应用了多元统计分析方法。1958年,美国的克鲁拜因(W. C. Krumbein)从事区域地质统计分析方面的工作。同年,克鲁拜因首次在地质杂志上公布电子计算机地质计算程序。从此,地质学开始进入了应用电子计算机的新时代。

1961—1970年,数学方法在地质学中应用更为广泛,数学地质进入了发展昌盛时期,多元统计分析在地质学中大量应用,数学地质发展成为一门独立学科。自1961年开始,美国亚利桑那大学召开了一系列“电子计算机在矿产工业中的应用”讨论会。1966年,美国堪萨斯地质调查所召开了“电子计算机在地球科学中的应用”讨论会。1967年,国际地质科学联合会成立了“地质数据存储、自动处理和索取委员会”(COGEO DATA)。1968年在巴黎召开的国际地质会议上成立了国际数学地质协会(IAMG),并开始出版国际数学地质协会杂志和地质计算机程序公报。美国地质调查所首次公布其电子计算机贡献文集——《电子计算机在地球科学中应用》一书出版。

近年来,数学地质在国内外发展很快。尤其在地质数据的存储、索取、自动处理和显示方面取得了长足的进步。据报导,到1980年止,世界上已建成了约500个大型地质数据库,涉及到地质学的各个重要领域。例如,瑞典设计了一个野外数据的存储索取系统(GEOMAP),用于野外地质填图工作。芬兰设计一种野外填图用的数据存储、索取和处理系统(GEOKU),具有自动查错和处理功能。加拿大戴维(P. David)和列比依斯(J. Lebluis)设计了一套全部地层信息标准记录系统(LED A),可用于区域地质填图

和地层对比。此外,还设计出Geolog、Geosystem和野外数据处理系统(DASCH),用于地质、勘探、地球化学和地球物理野外调查等方面。DASCH系统可用于回归分析、因子分析、聚类分析、趋势面分析及自动绘图等。丹麦设计了一种称为System-I的实验性数据处理系统,可以处理野外数据和实验室数据。

除上述记录野外数据的系统外,近年来在数据处理系统和计算机绘图方面又有新的进展,许多国家已经或正在建立一些更高级的地质数据的综合处理系统。其中,比较重要的有:

法国高等矿业学校发展了SIGMI系统,配备有较完整的软件,可直接用于地质数据的存储、索取和处理。

英国地球科学研究所用FORTRAN-IV语言编写的综合数据处理系统(G-EXEC系统),能在大多数计算机上应用,系统地处理任何类型的地质数据。英国“矿产矿山数据库”(MINTMINES)、“不列颠采矿数据库”(BRITPITS)和“矿产统计数据库”(MINSTATS)等,大都使用G-EXEC系统。此外,英国工程计算服务公司(ECS)的GEDNET地质应用软件系统,也可以处理各种地质数据。

加拿大西安略大学地质系建立了SAFRAS地质数据系统。这是一个以COBOL语言编写的系统,它可以存储、编辑、索取和处理各种地质数据。

美国地质调查所设计一套数据存储和索取系统(GRASP),最近又扩充为PACER和GANET两个子系统。它可用来直接输入地质数据,并能进行更新、检索和各种数据处理。美国俄克拉何马大学的通用信息处理系统(GIPSY),及由该系统所支持的美国“计算机化矿产资源信息库”(CRIB),是世界上最大的一个矿产资源数据库。库内存有美国40 000个矿床和矿产地、国外6000个矿床和矿产记录。文件内容包括矿床位置、地质特征、储量、产量等数据。用户可以通过电子计算机网络在世界500多个城市用电话查阅或索取CRIB数据库中的数据。除上述综合系统外,还有一些专门数据系统。如“北美石油数据系统”(PDS)、“矿产可得性系统”(MAS)、“全国水文处理数据存储和检索系统”、“全国铀资源数据库”、“地球化学数据库”等。

捷克和斯洛伐克设计了一个数据处理系统,它包括10个数据文件。这些文件用统一的计算机系统联系起来,可处理环境地质数据,如处理矿床、野外地质填图、地下水、矿泉水、滑坡、水文学和工程地质的数据等。

加拿大地质调查所与主管地形测量的单位合作,应用电子计算机自动显示,实现了绘制地形图自动化和地质图半自动化。最近,以激光为基础的光学数据处理系统,将逐渐用于地质学中。

综上所述,国外数学地质方法在地球科学各个领域中得到广泛的运用,并推动着地质学研究进一步走向量化。地质过程的数学抽象、数学模型的建立及数学模拟等数学地质理论研究更加深入。在电子计算机两大功能(科学计算和信息处理)基础上建立起来的各种地质数据自动处理系统,已经开始从地质数据库向地质知识库,进而向具有人工智能的专家系统发展。

在我国数学地质大致经历了以下三个发展阶段。

1969—1978年为酝酿和引进阶段。在此期间,各科研单位和地质院校纷纷成立数学地质小组或数学地质研究室,举办各种数学地质短训班。在内容上,主要是引进国外比较成

熟的多元统计方法。1978年10月27日,在杭州召开了第一届全国数学地质学术讨论会,与会代表115人,提交了105篇论文资料。这是对该阶段工作的总结。

1979—1981年,为数学地质创立阶段。第一届全国数学地质学术讨论会之后,我国数学地质工作犹如雨后春笋,蓬勃发展。从事数学地质的专业人员不断增加,一支由老、中、青原数学工作者和地质工作者组成的数学地质队伍基本形成。1981年4月22日,在长沙召开了第二届全国数学地质学术讨论会,与会代表269人,提交论文203篇。会议期间成立了“全国数学地质专业委员会”,它标志着我国数学地质作为一门独立的学科正式创立。

1982—至今为稳定发展阶段。我国数学地质经过一段蓬勃发展之后,人们发现许多数学地质方法在使用时要受一定条件的限制。数学地质决不是简单的数学加上地质,数学地质的过程是一个研究和实践的过程。它要求数学地质专业人员既懂得数学的理论、方法以及应用条件,又要熟悉地质思维和各方面的专业知识,要在数学和地质的结合上下功夫。为了使数学地质更好地与生产实践相结合,迫使数学地质工作者对数学地质的方法内容、使用条件和解决实际问题的程度进行深入研究。这个阶段在数学理论方法上引进消化吸收了稳健多元统计分析和成分数据的统计方法,使对数学理论了解较少的地质工作者在数学的应用条件上降低了要求,便利了数学地质的推广,使数学分析方法的理论水平也提高了一步。在应用方面,对已有的数学地质方法的实用性进行了深入研究,新的数学地质模型的创立,适合我国国情的地质数据系统的建立,使数学地质在生产实践中逐渐形成生产能力。地质矿产部计算中心筹建了“全国化探扫描面数据库”。中国地质大学(武汉)开始筹建“区域地球化学数据库”。

数学地质,尤其是多元统计分析发展很快,应用也十分广泛,但也有它的局限性。多元统计分析的结果直接依赖于原始数据,更明确地说多元统计分析只能起到对原始数据的综合、加工和提炼的作用。因而在讨论多元统计的成果时,应该考虑原始数据中是否提供了与研究问题有关的信息,如果原始数据没有包含某种信息,那么在解释其结果时不能牵强附会。因此,数学地质工作者对于样品的采集和变量的选择是一个必须认真研究的问题。

四、多元分析方法的工作步骤

地球化学中的多元分析是以数学为工具,电子计算机为手段,解决地质地球化学问题为目的,它解决问题的一般模式是:地质地球化学问题转化为数学问题,然后对数学问题的解进行地质解释。它是以地质地球化学为基础和出发点,经过变地质模型为数学模型,最后达到揭示地质问题的数学规律,或查明地质体和地质过程的数学特征,建立地质体和地质过程的数学模型,从而达到认识地质事件的客观特征,了解地质事件的运动规律的目的。

地球化学中的多元分析方法的工作步骤应当是:

1. 了解地质背景,收集并分析前人工作资料;
2. 制定地球化学样品的采集方案,研究对样品的测试项目(包括测试仪器、分析方案和分析灵敏度、精确度的确定),并对样品进行分析测试;
3. 收集地球化学样品各种资料(包括样品的采集位置、地质产状、各项分析测试的原始数据等),建立地球化学数据库;
4. 根据工作目的选择适当的多元统计分析方法,构成相应的数学模型;
5. 确定合理的计算步骤和方法,并将计算步骤和方法编写出计算机程序或调用数据

处理系统中标准计算程序，选择地球化学数据库中适当样品的适当变量进行计算，

6. 对计算的结果进行地质解释；

7. 若对计算结果的地质解释不满意，可以调整计算方法、样品和变量的选择，然后再作计算和结果解释，通过反复实践，以获取较满意的结论。

目 录

绪论	(1)
第一章 地球化学变量	(7)
第一节 地球化学变量的特点及分类	(7)
第二节 变量的选择	(9)
第三节 变量的变换	(13)
第四节 数据集合的两种空间表示	(17)
第二章 聚类分析	(19)
第一节 一般概念	(19)
第二节 相似性度量	(19)
第三节 点群的谱系簇分	(23)
第四节 计算实例	(24)
第五节 逐步计算成图法	(28)
第六节 聚类分析在地质上的应用	(29)
第七节 聚类分析中一些问题的讨论	(34)
第八节 其它几种聚类分析方法	(36)
第三章 多元回归分析	(47)
第一节 基本概念	(47)
第二节 多元线性回归方程的建立	(47)
第三节 各自变量对回归效果的贡献	(52)
第四节 回归分析的一般步骤	(54)
第五节 最优回归方程的选择	(56)
第六节 几个问题讨论	(65)
第七节 回归分析在地球化学中的应用	(67)
第四章 趋势分析	(70)
第一节 概述	(70)
第二节 多项式趋势分析的数学方法	(70)
第三节 趋势方程的拟合度及显著性检验	(75)
第四节 几个问题的讨论	(76)
第五节 规则测网趋势面分析的数学方法	(78)
第六节 趋势分析在地球化学中的应用	(81)
第五章 判别分析	(84)
第一节 概述	(84)
第二节 费歇准则下的两类线性判别	(85)
第三节 两类判别计算实例	(91)

第四节	贝叶斯准则下的多类线性判别	(95)
第五节	逐渐判别分析	(103)
第六节	判别分析在地质上的应用	(110)
第七节	判别分析中应注意的问题	(112)
第六章	因子分析	(114)
第一节	概述	(114)
第二节	因子分析的数学模型	(115)
第三节	主成分分析	(121)
第四节	正交旋转——正交因子解	(123)
第五节	斜交旋转——斜交因子解	(126)
第六节	因子计量	(131)
第七节	因子分析的计算步骤与实例	(133)
第八节	Q型因子分析	(137)
第九节	因子分析在地球化学工作中的应用	(139)
第十节	因子分析的若干问题和讨论	(141)
第七章	对应分析和非线性映射	(143)
第一节	什么是对应分析	(143)
第二节	对应分析的数据变换方法	(143)
第三节	对偶定理	(144)
第四节	对应分析因子解	(145)
第五节	对应分析计算步骤	(146)
第六节	应用实例*	(147)
第七节	非线性映射简介	(150)
第八章	典型相关分析	(155)
第一节	概述	(155)
第二节	总体中的典型相关和典型变量	(155)
第三节	典型相关系数的显著性检验	(159)
第四节	用相关矩阵代替协方差矩阵的情形	(160)
第五节	应用实例*	(160)
第六节	问题与讨论	(163)
第九章	马尔科夫链型分析	(165)
第一节	基本概念	(165)
第二节	马尔科夫链的转移概率	(166)
第三节	连续参数非平稳马尔科夫过程	(169)
第四节	多元时空序列转移矩阵U	(173)
第五节	转移矩阵U的谱分解	(177)
第六节	过程的演化趋势	(179)
第十章	地质统计学(简介)	(184)
第一节	概述	(184)

第二节	区域化变量理论.....	(185)
第三节	变异函数及结构分析.....	(188)
第四节	线性克里格法.....	(194)
第十一章	稳健统计学*	(198)
第一节	概述.....	(198)
第二节	稳健统计学中的几个概念.....	(198)
第三节	位置和尺度的稳健估计.....	(200)
第四节	方差的稳健估计.....	(206)
附录 I	线性代数初步知识.....	(209)
附录 II	数理统计常用表.....	(224)
主要参考文献		(234)

绪 论

近三十年来,在电子计算机应用的促进下,多元分析在地质地球化学领域中得到了迅速发展,对地球科学的定量化、地质数据的自动化处理、地质过程的计算机模拟以及地质思维的智能化等方面起到很大的促进作用,有力地推动了古老的地球科学与现代科学的结合。可以说多元分析方法的应用已经远远超出数据处理的范围,已发展成为地质和化探工作中的一种富有生命力的新的研究手段,从而使地球科学在理论上和方法上产生了重大的变革。

一、地球化学中的多元分析的涵义

地球化学中的多元分析属于数学地质的一个分支。它以多元统计分析为主要手段,以电子计算机为工具来研究地球化学及化探问题,将数学与地球化学密切结合,使地球化学数学化、定量化,以便用计算机进行研究。

这里说的多元分析即多变量统计分析,它研究的对象是多变量的数据集合。这种数据集合一般由 n 个个体,每一个个体都有 p 个变量的观测值构成。形式上,可以把多元分析作为数理统计学在研究客观事物中多种因素间相互依赖的统计规律性的一个分支。多元分析主要讨论以下几方面问题:

1. 讨论数据集合中变量与变量、样品与样品、变量与样品之间的相互关系或依赖关系。
2. 对变量或样品进行分组、归类或判别。
3. 研究数据集合的空间展布和时间序列变化规律。
4. 简化数据结构。把具有内在联系的几个变量变换为一个新的独立变量或者约简数据集合的空间维数,用较简单的形式来表示所研究的复杂的多元数据集合。
5. 提出假设和检验假设。主要是探讨以上四个方面所得结果的有效性、可靠性或可信性。

较常用的多元分析方法有:相关分析与回归分析、聚类分析、判别分析、趋势面分析、因子分析、对应分析、典型相关分析、马尔科夫概型分析和地质统计学。其中,回归分析、因子分析、对应分析、典型相关分析是研究变量和样品属性及依赖关系,聚类分析、判别分析等是对样品进行分组和归类,而趋势面分析、地质统计学及马尔科夫概型分析是探讨变量的时空展布。

地球化学中的多元分析的任务除了介绍这些多元分析的数学方法、原理以外,还要讨论它们在地球化学工作中的应用条件,所能解决地球化学问题的范围以及所计算得到数学结果的地球化学解释等问题。

二、迅速发展的原因

地球化学中的多元分析是现代地球化学、数学和电子计算机技术迅速发展的必然结果。从现代地球化学和地球化学勘查的发展趋势来看,主要表现为以下3个方面:

1. 从经验到理论

在地球化学特别是地球化学探矿工作中,首先得到大量的事实规律。这些从实践中归纳出来的事实规律属于经验,需要上升为理论,用理论来解释它们,并反过来用于指导地球化学找矿的实践,目前很多经验性的事实规律已经被地球化学的理论所解释,但也有不少问题还有待于理论上的深化。可以说现在正面临着地球科学飞跃的前夕。在这个飞跃中,数学也将起到它应有的作用。

2. 由定性走向定量

地球科学与其他自然科学相比,更多地使用定性的描述。研究人员根据自己所观察到的地质现象作人工的归纳、演绎和推理。长期以来地质工作者较多地使用文字和各种图件来记录、反映地质现象。这样使得地质工作的野外工作量大,而且人为性强。对同一地质现象在不同研究者的文字描述中不容易作数量上的对比,因而也不利于地质工作的自动化。随着科学的发展,人们愈来愈强烈地认识到“一种科学仅当它成功地运用数学时,才算达到了完善的程度”。数学这门研究现实世界数量关系与空间形式的科学近年来很快地渗透到地质科学中,人们愈来愈渴望用简明的数学关系式或图表来表征各种地质现象和地质过程。正如S·C·鲁宾逊所说:“地质数据若用数学方法整理、核对、综合、计算和检验,则对地质解释、分类、概念检验、问题解答以及概率预报,能提供较之只限于任何个人的经验和记忆更为全面和客观的基础。而且,这样的基础可以作为许多地质人员的共同基础,他们每个人再加上自己的经验和知识去完成他们自己的解释工作。在推进地质学这门科学向前发展的过程中,这种共同基础正好可以成为一个主要因素。”由此看来,数学和地球科学的结合已成为地质科学向前发展的一个必然方向。

3. 从单变量到多变量发展

客观世界中任何事情都是与其周围的事物或者与其所处的环境存在着千丝万缕的联系,在地球化学工作中也是一样,仅仅根据单变量的观测数据来分析问题已经远远不能满足研究工作的要求,而必须同时对多个变量进行观测、汇集与研究,与问题全部有关的原始资料,较全面地研究它们之间的相互关系和相互影响,去粗取精、去伪存真,才能得出正确的结论。

地球化学和地球化学探矿不同于其他地质类学科的另一特点是它具有大量可测定的数据。仅从测试化学元素方面来看,从40、50年代只分析几个主要成矿元素,到60、70年代发展为既分析成矿元素又分析控矿元素和运矿元素。80年代以来,则向测试更多元素方面发展。如地质矿产部1:20万化探扫面工作规定每个样品必须分析40多个元素,占元素周期表中能分析元素的大部分。

数学地质发展十分迅速的原因,从数理统计方面来看,各种统计方法越来越成熟,新的统计分析方法不断涌现,特别是多元统计分析发展很快,稳健统计分析和成分数据的统计分析方法的提出,使多元统计分析无论在理论上还是在应用方面都提高到一个新的水平。

50年代以来,电子计算机技术得到迅速发展,这为多元统计从数学方法的理论研究推

广到实际应用提供了必要的条件。如果没有电子计算机,要对一些数量较大的数据进行某些多元统计分析,如趋势分析、因子分析、判别分析等几乎是不现实的。电子计算机的使用对于多元分析在地质上的应用提供了一个必要的条件。

三、发展史及研究现状

早在1840年,英国地质学家莱伊尔(C. Lyell)运用统计的方法,根据近代海洋生物的相对含量,把第三纪地层中的生物种属与现代海生生物种属对比,成功地对第三纪地层作了进一步划分。

1890年,皮尔逊(K. Pearson)编写了《数学进化论贡献》一书,其中有古生物化石的统计分析。

1914—1934年间,俄国学者列文森——列辛格(Лёвнсон—Лёсницк)通过研究岩石的岩浆酸度系数的频率分布,探讨了安山岩、玄武岩、流纹岩的分类问题。

1936—1945年数学方法的应用由地质学的个别问题逐渐扩展到地质学的一些分支。在岩石学中引入了双变量或三变量分析,出现了三角岩比图、百分率图等。在矿物学和结晶学领域中,还运用了简单回归分析和方差分析,以了解矿物性质与化学成分之间的关系。1944年,苏联学者维斯捷列乌斯(А. Б. Вистеллус)在苏联科学院报告集上发表了《分析地质学》一文,提出用定量方法研究地质问题的初步思想。他从事了三十多年的数学地质工作,成为苏联数学地质的创始人和国际数学地质学会的第一任主席。

1945—1960年间,数学地质迅速发展,电子计算机也开始在地质工作中应用。数学统计学的方法普遍应用于地质学的所有分支。1946年,苏联学者研究金属矿床元素统计分布特点。1954年,绍(D. M. Shaw)等人应用统计方法研究地球化学问题。1956年,切叶思(F. Chayes)应用均值、方差于岩石学研究。1956年,克伦宾(Graybill)在研究岩石的矿物和化学成分时,应用了多元统计分析方法。1958年,美国的克鲁拜因(W. C. Krumbein)从事区域地质统计分析方面的工作。同年,克鲁拜因首次在地质杂志上公布电子计算机地质计算程序。从此,地质学开始进入了应用电子计算机的新时代。

1961—1970年,数学方法在地质学中应用更为广泛,数学地质进入了发展昌盛时期,多元统计分析在地质学中大量应用,数学地质发展成为一门独立学科。自1961年开始,美国亚利桑那大学召开了一系列“电子计算机在矿产工业中的应用”讨论会。1966年,美国堪萨斯地质调查所召开了“电子计算机在地球科学中的应用”讨论会。1967年,国际地质科学联合会成立了“地质数据存储、自动处理和索取委员会”(COGEO DATA)。1968年在巴黎召开的国际地质会议上成立了国际数学地质协会(IAMG),并开始出版国际数学地质协会杂志和地质计算机程序公报。美国地质调查所首次公布其电子计算机贡献文集——《电子计算机在地球科学中应用》一书出版。

近年来,数学地质在国内外发展很快。尤其在地质数据的存储、索取、自动处理和显示方面取得了长足的进步。据报导,到1980年止,世界上已建成了约500个大型地质数据库,涉及到地质学的各个重要领域。例如,瑞典设计了一个野外数据的存储索取系统(GEOMAP),用于野外地质填图工作。芬兰设计一种野外填图用的数据存储、索取和处理系统(GEOKU),具有自动查错和处理功能。加拿大戴维(P. David)和列比依斯(J. Lebluis)设计了一套全部地层信息标准记录系统(LED A),可用于区域地质填图

和地层对比。此外,还设计出Geolog、Geosystem和野外数据处理系统(DASCH),用于地质、勘探、地球化学和地球物理野外调查等方面。DASCH系统可用于回归分析、因子分析、聚类分析、趋势面分析及自动绘图等。丹麦设计了一种称为System-I的实验性数据处理系统,可以处理野外数据和实验室数据。

除上述记录野外数据的系统外,近年来在数据处理系统和计算机绘图方面又有新的进展,许多国家已经或正在建立一些更高级的地质数据的综合处理系统。其中,比较重要的有:

法国高等矿业学校发展了SIGMI系统,配备有较完整的软件,可直接用于地质数据的存储、索取和处理。

英国地球科学研究所用FORTRAN-IV语言编写的综合数据处理系统(G-EXEC系统),能在大多数计算机上应用,系统地处理任何类型的地质数据。英国“矿产矿山数据库”(MINTMINES)、“不列颠采矿数据库”(BRITPITS)和“矿产统计数据库”(MINSTATS)等,大都使用G-EXEC系统。此外,英国工程计算服务公司(ECS)的GEDNET地质应用软件系统,也可以处理各种地质数据。

加拿大西安略大学地质系建立了SAFRAS地质数据系统。这是一个以COBOL语言编写的系统,它可以存储、编辑、索取和处理各种地质数据。

美国地质调查所设计一套数据存储和索取系统(GRASP),最近又扩充为PACER和GANET两个子系统。它可用来直接输入地质数据,并能进行更新、检索和各种数据处理。美国俄克拉何马大学的通用信息处理系统(GIPSY),及由该系统所支持的美国“计算机化矿产资源信息库”(CRIB),是世界上最大的一个矿产资源数据库。库内存有美国40 000个矿床和矿产地、国外6000个矿床和矿产记录。文件内容包括矿床位置、地质特征、储量、产量等数据。用户可以通过电子计算机网络在世界500多个城市用电话查阅或索取CRIB数据库中的数据。除上述综合系统外,还有一些专门数据系统。如“北美石油数据系统”(PDS)、“矿产可得性系统”(MAS)、“全国水文处理数据存储和检索系统”、“全国铀资源数据库”、“地球化学数据库”等。

捷克和斯洛伐克设计了一个数据处理系统,它包括10个数据文件。这些文件用统一的计算机系统联系起来,可处理环境地质数据,如处理矿床、野外地质填图、地下水、矿泉水、滑坡、水文学和工程地质的数据等。

加拿大地质调查所与主管地形测量的单位合作,应用电子计算机自动显示,实现了绘制地形图自动化和地质图半自动化。最近,以激光为基础的光学数据处理系统,将逐渐用于地质学中。

综上所述,国外数学地质方法在地球科学各个领域中得到广泛的运用,并推动着地质学研究进一步走向量化。地质过程的数学抽象、数学模型的建立及数学模拟等数学地质理论研究更加深入。在电子计算机两大功能(科学计算和信息处理)基础上建立起来的各种地质数据自动处理系统,已经开始从地质数据库向地质知识库,进而向具有人工智能的专家系统发展。

在我国数学地质大致经历了以下三个发展阶段。

1969—1978年为酝酿和引进阶段。在此期间,各科研单位和地质院校纷纷成立数学地质小组或数学地质研究室,举办各种数学地质短训班。在内容上,主要是引进国外比较成

熟的多元统计方法。1978年10月27日,在杭州召开了第一届全国数学地质学术讨论会,与会代表115人,提交了105篇论文资料。这是对该阶段工作的总结。

1979—1981年,为数学地质创立阶段。第一届全国数学地质学术讨论会之后,我国数学地质工作犹如雨后春笋,蓬勃发展。从事数学地质的专业人员不断增加,一支由老、中、青原数学工作者和地质工作者组成的数学地质队伍基本形成。1981年4月22日,在长沙召开了第二届全国数学地质学术讨论会,与会代表269人,提交论文203篇。会议期间成立了“全国数学地质专业委员会”,它标志着我国数学地质作为一门独立的学科正式创立。

1982—至今为稳定发展阶段。我国数学地质经过一段蓬勃发展之后,人们发现许多数学地质方法在使用时要受一定条件的限制。数学地质决不是简单的数学加上地质,数学地质的过程是一个研究和实践的过程。它要求数学地质专业人员既懂得数学的理论、方法以及应用条件,又要熟悉地质思维和各方面的专业知识,要在数学和地质的结合上下功夫。为了使数学地质更好地与生产实践相结合,迫使数学地质工作者对数学地质的方法内容、使用条件和解决实际问题的程度进行深入研究。这个阶段在数学理论方法上引进消化吸收了稳健多元统计分析和成分数据的统计方法,使对数学理论了解较少的地质工作者在数学的应用条件上降低了要求,便利了数学地质的推广,使数学分析方法的理论水平也提高了一步。在应用方面,对已有的数学地质方法的实用性进行了深入研究,新的数学地质模型的创立,适合我国国情的地质数据系统的建立,使数学地质在生产实践中逐渐形成生产能力。地质矿产部计算中心筹建了“全国化探扫描面数据库”。中国地质大学(武汉)开始筹建“区域地球化学数据库”。

数学地质,尤其是多元统计分析发展很快,应用也十分广泛,但也有它的局限性。多元统计分析的结果直接依赖于原始数据,更明确地说多元统计分析只能起到对原始数据的综合、加工和提炼的作用。因而在讨论多元统计的成果时,应该考虑原始数据中是否提供了与研究问题有关的信息,如果原始数据没有包含某种信息,那么在解释其结果时不能牵强附会。因此,数学地质工作者对于样品的采集和变量的选择是一个必须认真研究的问题。

四、多元分析方法的工作步骤

地球化学中的多元分析是以数学为工具,电子计算机为手段,解决地质地球化学问题为目的,它解决问题的一般模式是:地质地球化学问题转化为数学问题,然后对数学问题的解进行地质解释。它是以地质地球化学为基础和出发点,经过变地质模型为数学模型,最后达到揭示地质问题的数学规律,或查明地质体和地质过程的数学特征,建立地质体和地质过程的数学模型,从而达到认识地质事件的客观特征,了解地质事件的运动规律的目的。

地球化学中的多元分析方法的工作步骤应当是:

1. 了解地质背景,收集并分析前人工作资料;
2. 制定地球化学样品的采集方案,研究对样品的测试项目(包括测试仪器、分析方案和分析灵敏度、精确度的确定),并对样品进行分析测试;
3. 收集地球化学样品各种资料(包括样品的采集位置、地质产状、各项分析测试的原始数据等),建立地球化学数据库;
4. 根据工作目的选择适当的多元统计分析方法,构成相应的数学模型;
5. 确定合理的计算步骤和方法,并将计算步骤和方法编写出计算机程序或调用数据

处理系统中标准计算程序，选择地球化学数据库中适当样品的适当变量进行计算，

6. 对计算的结果进行地质解释；

7. 若对计算结果的地质解释不满意，可以调整计算方法、样品和变量的选择，然后再作计算和结果解释，通过反复实践，以获取较满意的结论。

第一章 地球化学变量

第一节 地球化学变量的特点及分类

一、地球化学变量的特点

各种地球化学特征在不同的研究对象(样品或测点)上进行测试,可以得到各自的数值。这种在不同对象上可以取值的地球化学特征称为地球化学变量。例如,元素含量、矿石的相对密度、矿体的厚度、薄片某种副矿物的颗数、溶液的 pH、Eh 值、成矿成岩的温度和压力、热液包体溶液的盐度等等,都是地球化学变量。

地球化学变量具有 3 个特点。即它具有随机变量的性质,又具有确定性变量的特征,还具有区域化变量的结构性。

地球化学变量的随机性表现在以下几方面:

1. 地球化学样品的采取具有抽样的性质。地质工作所研究的对象是天然物质。因此,不可能把整个大自然或一个区域的地质体都搬到实验室里来,而只能抽取其中一部分样品,这部分样品与所研究的对象相比,只占很少很少的数量。这样的工作方法本身就决定了地球化学变量具随机变量的性质。

2. 从成因上看,地球化学变量的取值受多因素的控制。如某样品中某种元素的当今含量,它取决于漫长地质历史中的各种地质作用,在地质历史长河中,作用到某个研究对象上的各种因素,人们往往不能全部准确地推算到。常常遇到这样的情况,对于两个不同的研究对象,虽然它们有十分相似的地质地球化学环境,或者甚至是同一地点采取的,但它们的某些地球化学特征的测试结果却可以是不相同的,表现为随机变量的特征。

3. 地球化学变量的随机性还表现在测试过程中,由于测试者的主观因素,测试仪器的精度不同,测试分析方案不同,测试时周围环境的影响,测试时间不同等等,对同一样品的同一变量的不同次测试结果就不一定完全相同。这可以归结为测试误差,误差就带有随机性特点。实际上,地球化学变量在某个个体上的真实值是不知道的,每一次测试只是对其真实值的一次估计。

地球化学变量又具有确定性的一面。这是因为每一个研究对象在指定的时刻,某种地球化学特征的真实数值是客观存在的,是有确定数值的。这个值不依赖于人们是否已经测试到了。

地球化学变量还具有结构性。所谓结构性是指地球化学变量由于受地质因素的控制,在不同方向上变化率的不同,具有不同的自相关的特性。当一个变量分布在空间内时(即赋予其空间位置时),称为区域化变量。区域化变量是一个函数 $f(x)$,它在三维空间内的