

赵鹏大等著

地质勘探中的

统计分析

中国地质大学出版社

目 录

前 言	赵鹏大 (1)
-----	-----------

第一篇 地质体的数学特征

第一章 地质体数学特征概论	赵鹏大 (5)
第二章 地质数据分析	(11)
第一节 地质数据的统计分析	李紫金 (11)
第二节 变量及变量组合的统计评价——BIBD方法	袁 丁(张光前整理) (21)
第三章 地质体数学特征研究	(31)
第一节 矿体数学特征的意义及研究内容和方法	魏 民 (31)
第二节 Cu、Ni、Sn、Mo矿体数学特征之一——统计特征与结构特征	魏 民 (33)
第三节 Cu、Ni、Sn、Mo矿体数学特征之二——空间特征与几何特征	魏 民 (41)
第四节 矿产勘查中地质体数学特征应用实例	池顺都 (55)

第二篇 地质勘探中的数学模型

第四章 地质体数学模型	(71)
第一节 矿体变化性的成因分类模型和勘探分类模型	魏 民 (71)
第二节 马尔柯夫链分析在沉积型磷矿成因分类中的应用	张光前 (77)
第三节 自回归-判别模型与含煤性的定量分析	金友渔 (81)
第四节 多元统计分析与地质统计学相结合进行定量地层划分的研究	王仁铎 (86)
第五节 研究地质体空间特征的地质统计学方法	王仁铎 (93)
第六节 陡倾斜地质-构造面的趋势面分析	冉宗培 (102)
第七节 自典型相关分析的数学模型及在研究煤质参数多元空间 相关性中的应用	金友渔 (108)
第八节 残差图分析在研究煤厚与砂体定量关系中的应用	金友渔 (113)
第九节 目标反索模式抽取法及其应用	梁 波(张光前整理) (120)
第十节 区域地球化学数学模型研究	胡光道 (123)
第五章 地质过程数学模型	(135)
第一节 马尔柯夫链地层剖面模拟——以大峪口剖面为例	张光前 (135)
第二节 详细沉积断面的半马尔柯夫过程分析	金友渔 (139)
第三节 负二项最近点距离分布理论及其在我国铜矿资源潜力评价中 的应用	金友渔 (148)
第四节 最可预测面方法在沉积环境模拟中的应用	张光前 (157)
第五节 构造应力场的数学模拟——有限元法及其应用	李紫金 (165)
第六章 地质工作方法(矿产预测及矿床勘探)数学模型	(171)
第一节 安徽铜陵1:20万图幅铜、铁、硫、石膏、明矾石资源预测	李紫金 (171)
第二节 中国南方晚震旦世小比例尺磷矿资源评价	张光前 (181)

第三节	矿化有利度克立格统计预测法	· · · · · 骆俊峰(王仁铎整理)	(185)
第四节	矿化富集中心的数学模型及在某锡—多金属远景区矿床统计 预测中的应用	· · · · · 金友渔	(193)
第五节	大比例尺矿床统计预测研究中的序贯预测法	· · · · · 袁 丁(张光前整理)	(200)
第六节	模糊回归分析及其应用	· · · · · 郭敬东(张光前整理)	(207)
第七节	地质信息与勘探信息定量评价研究综述	· · · · · 魏 民	(212)
第八节	地质勘探中的统计信息分析——勘查信息量数学模型	· · · · · 魏 民	(214)
第九节	初勘决策分析数学模型	· · · · · 池顺都	(224)
第十节	勘探工程地质统计信息法	· · · · · 胡光道	(231)
第十一节	勘探精度评价和预测数学模型	· · · · · 魏 民	(241)
第十二节	用普通克立格法进行储量预测的精度问题	· · · · · 冉宗培	(244)

第三篇 计算机在地质工作中的应用

第七章	计算机在矿产预测中的应用	· · · · ·	(257)
第一节	单元地质变量自动取值原理及计算机程序	· · · · · 周有武	(257)
第二节	地质统计学化探数据处理程序	· · · · · 胡光道	(262)
第八章	计算机在矿床勘探中的应用	· · · · ·	(266)
第一节	矿床勘探初级程序系统	· · · · · 胡光道	(266)
第二节	地质统计学储量计算程序系统	· · · · · 胡光道	(273)
第九章	专家系统在沉积环境分析中的应用	· · · · · 梁 波(张光前整理)	(277)
英文目录及英文摘要	· · · · ·		(284)

前 言

《地质勘探中的统计分析》是国家教委资助的博士点科研基金项目的成果。它反映了近年来我校“数学地质”博士学科点部分教师和研究生的研究成果。由于定稿和出版的周期较长,所以未能将1987年以后的工作包括在内。

中国地质大学(武汉)数学地质研究室的研究工作围绕着矿床统计预测;矿床勘探中的统计分析;地质统计学及计算机地质应用四大方面进行。本专著是以这四个方面研究成果为基础,按照数学地质的基本任务:查明地质体数学特征,建立地质体的数学模型;查明地质过程各因素及其相互关系,建立地质过程的数学模型;查明地质数据及工作方法特征,建立地质工作方法的数学模型以及地质矿产数据库、专家系统、信息系统以及计算机制图等计算机其它地质应用而编著的。可以说,本专著是数学地质理论和方法在矿产普查与勘探中的应用研究成果。反过来,在这个领域的应用研究又进一步丰富了数学地质理论和方法。

“数学地质”一词是60年代初期由苏联学者A.B.维斯捷利乌斯(1962)首先提出并在以后为学术界所接受,所以,“数学地质”作为一门新兴学科其历史不足30年。按照维斯捷利乌斯的定义:“数学地质是研究在具体工作中建立一分析和利用地质现象数学模型的科学”。在我们随后的研究中认为有必要把“地质现象数学模型”具体化为“地质体、地质过程和地质工作方法”的数学模型。这正是我们编写本专著的基本纲要。

有人认为,“数学地质”和其它地质学科不同,它只能加工和浓缩地质信息而不能提供新的信息,如物探可以提供地球物理场特征和揭示物探异常,化探可以提供地球化学场特征并揭示化探异常,而数学地质则不能产生自己特有的信息。对于这样一种观点已经被越来越多的事实所否定。E.H.T.惠顿曾不止一次地指出,数学地质在岩石学中的应用揭示了完全意想不到的结构特性,而且具有深刻的成因意义;发展了原有的对交代现象的认识,建立了新的可以解决很多实际问题的交代模式;创造了对侵入岩体进行特殊制图的方法,揭示了用一般地质方法不能查明的特征。这指的是苏联维斯捷利乌斯用马尔科夫模型研究花岗岩所取得的一系列重要成果。此外,在沉积学、地球化学等领域也有通过数学地质的应用而引入完全新的概念或更正、查明某些老概念的实例。这里要强调的是:本书前面一些章节所论述的“地质体数学特征”正是数学地质在研究各种地质对象时所提供的完全新型的信息。地质体数学特征不仅可以通过数量规律性的查明更全面、深刻和具体地描述和定义地质体,而且可以通过不同地质体所具有的特定的数学特征而达到分析、对比和鉴定地质体的作用。还可以根据数学特征,例如分布律特征,查明地质体的成因和形成过程。众所周知,很多地质体、地质现象、地质概念缺乏严格的定义;人们对同一地质对象有时理解各异,例如,在矿床勘查中最常用的像“异常”,“成矿远景区”等等这样的名词概念都缺乏准确的定义。因而,人们在确定什么是“异常”,什么是“远景区”时常常发生难以决断或因人而异的现象,甚至就是“矿”与“非矿”也是需要建筑在对含矿体各种数学特征(表现为各类工业指标)查明的基础上才能加以区分和鉴别。所以,地质体数学特征是“数学地质”提供的最重要的信息。

近年来,地质文献中大量使用“模型”或“模式”这一概念。有人还把“模型概念的引

入”作为与板块构造理论、计算机地质应用相提并论的近代地质科学三大重要进展之一。对模型的分类有多种多样，对地质现象的模拟也有多种途径。但这里强调的是“数学模型”和“数字模拟”。本书所论述的地质过程数学模型和更大篇幅所阐述的地质工作方法数学模型正是“数学地质”对地质科学所作的独特贡献。发生在广阔空间、漫长时间、复杂介质和极端条件下的地质过程是目前无法在实验室中加以再现，更不是野外实际观察所能及的，唯一可行的途径是通过建立数学模型利用计算机进行数字模拟。例如近年来国内外所进行的沉积过程计算机模拟，沉积盆地发育和演化数字模拟等。这些都是其它地质学科无法实现的。

地质工作方法数学模型的建立和应用可以说是数学地质中最普及和最重要的一个领域。地质生产和研究工作有许多特有的或常用的方法，如历史分析、成因分析和综合分析；鉴定、对比和分类；取样、测试和解释；预测、评价和决策等等。为了更科学、正确、广泛、客观和迅速地进行这类工作，发展一套数学模型并利用计算机加以实现是十分有益和重要的。可以毫不夸张的说，利用各种地质工作方法的数学模型（简称“数学地质方法模型”或“方法模型”）已成为现代地质科学研究不可缺少的组成部分。牢靠而充分的野外实际资料，精密而准确的室内测试数据，合理而定量的科学分析和计算是构成现代地质科研成果的三项基本要素。应该说，目前往往“一头一尾”较为欠缺，这正是数学地质大有作为的地方。

本书从矿床勘查这一个侧面概述数学地质三个最重要应用方面的某些研究成果。尽管是不系统而远非完善的，但在这方面坚持不懈的努力是完全必要和十分有意义的。

第一篇

地质体的数学特征

第一章 地质体数学特征概论

开展地质工作，首先是对研究对象——各种地质体进行观察、测量、取样和试验，取得第一性资料或原始信息。其次，在取得原始数据或观察结果的基础上进行综合分析，如进行横向综合对比或纵向成因分析等。在进行地质观察和度量时可能出现下面一些问题：

(1) 观测的局限性。观察可能只是作用于地质体的一小部分范围，或取样化验的结果只代表个别或少数观测点，在这种情况下很难根据局部推断总体。根据局部观察结果来描述地质体总体的作法有可能导致严重的失真。

(2) 观测的表面性。在观察或描述某种地质体时，未必能经常确定其最本质的特征。即使作最详细的描述，也常常与真实情况有一定距离或有某种歪曲。

(3) 观测误差。由于人为的或技术的原因，由于观测条件的限制及影响，对地质体的观察可能发生误差，从而影响对地质体的正确描述和认识。

(4) 观测尺度的差异性。对同一地质体，可以进行宏观的、中观的和微观的观测；对不同规模的地质体，其观测尺度更有明显差异。这些尺度水平和量纲各异的观测结果在对地质体的描述和认识上以及相互对比和机制分析上都会带来一些困难和问题。

(5) 观测结果的变异性。地质体的变化性是各种地质体最重要的特征。在同一观测点，不同时间可以获得不同的观测值而构成一个时间序列。在不同观测点，地质体各种属性或标志的观测值具有空间变化性而构成区域化变量。同一地质体的不同标志值间具有不同程度的相关联系而构成协变化或协区域变化。某些观测值还可具有特异变化而构成特异值或奇异值。

(6) 观测现象对比和成因解释的复杂性。解决各类地质任务有两种命题方式。其一是先提出假说或某种理论，然后以实际观测结果与之对照并考查其符合程度。如果符合，则任务得到解决。如果不符则调整或修改假说，或改变条件再进行观测，如此反复进行直至问题解决。这就是先建立地质体的某种模型，再以此模型为依据研究解决地质问题，这是所谓“正向解题”法。另一方式是先进行地质体的观测，然后根据不同观测结果进行相应的推断和解释。这是所谓“反向解题”法。由于前面叙述的有关观测中的问题，使这种对比解释具有不同程度的不确定性、模糊性和多解性，而造成此项工作的复杂性。

凡此种种问题，使我们在研究和解决有关地质任务时，必须首先选择一种正确认识地质体和有效区分不同地质体的方法和途径。

查明地质体的数量规律性或地质体的数学特征，便是克服上述很多困难的一种可能的有效方法和途径。

一、地质体数学特征概念和基本内容

我们通常熟悉的是地质体的地质构造特征、岩矿地球化学特征、地球物理特征等等，而地质体“数学特征”则是比较陌生的概念。地质体数学特征是指地质体各种属性的数量特征或数量规律性。可以从以下几个方面考查地质体数学特征：

(1) 几何特征: 描述地质体规模、形态、产状、空间位置、高程、埋深、边界清晰程度、轮廓复杂程度等。

(2) 统计特征: 揭示地质体各种属性的统计规律性。最重要的是统计分布特征或分布律, 各种统计特征数: 平均数、标准差、方差、协方差、相关系数、偏倚系数、峰凸系数、复相关系数、偏相关系数、混合分布类型、不同成分总体比例、截尾分布等。

(3) 空间特征: 描述地质体空间变化特征。趋势函数、剩余、自相关、互相关、变异函数、变程、拱高、基台、块金值、变化梯度、变化速度、变化周期、变化频率、功率谱等。

(4) 结构特征: 描述地质体的具体结构和抽象结构。均质性、非均质程度、各向异性、同向异性、连续程度、熵函数、马尔科夫性、时空协调性、因子结构、典型相关等。

可以区分单个地质体数学特征和地质群体数学特征; 地质体单个属性或单个标志数学特征和多属性、多标志组合数学特征。在考查多体、多标志数量特征或数量规律性时, 通常需建立地质体的数学模型。

可以区分概率模型和确定模型。前者是包括随机因素或由随机变量构成的数学模型, 目前在地质应用中, 在地质体数学特征研究中更具有普遍意义和重要作用。后者是在确定条件下由具有必然因果联系的参数和关系所构成的数学模型, 它在描述某些受确定性法则支配的简单(单一)地质过程, 如受重力法则支配的单一岩层的沉积过程时具有良好的效果。但有时也可以同时应用两类不同的数学模型, 如地质统计学通常应用概率模型描述地质体的空间连续性, 无论半变异函数或自协方差函数都看成是概率模型的一切可能现实的总和(积分)的平均数(期望值), 然而还可以用确定性模型来评价空间连续性, 此时半变异函数和自协方差函数被定义为在一个有限固定域内的积分。在概率模型条件下, 根据二阶平稳假说, 自协方差与半变异函数线性相关。这样一种关系使得在大多数地质统计学研究中既可使用自协方差函数, 也可以使用半变异函数。在确定模型条件下, 两者的关系是非线性的, 两个函数是不可以互换的。

这里应强调的是: 对地质体的简单定量描述并不是地质体的数学特征, 而只是它的定量表征或它的数学特征的一种显示或一个现实, 如对地质体规模的局部测定值、地质体成分的少量取样化验值等。只有当揭示出地质体的数量规律性或当各种数学表征能反映地质体的本质特征和总体特征时才能称其为该种地质体的数学特征。数学模型亦然, 只有当数学模型能揭示地质体成因本质时, 才能称其为地质体数学特征。例如苏联学者维斯基捷利乌斯1972年建立的“理想花岗岩”随机模型是正长石、斜长石及石英3种矿物颗粒转移必定具有一种简单的马尔科夫性, 即一重马氏链模型, 而交代作用成因的花岗岩则具二重马氏链性质。这两种马尔科夫模型代表了两种不同成因的花岗岩, 因而它们是花岗岩体的重要数学特征。

单一地质体和地质体群的数学特征可能有不同的内容和不同的表达方式, 如单一地层和地层序列的数学特征表达方式则不同。后者是一个时间序列, 即各种测定值或观测值按时间顺序构成的数据系列。连续进行了十年之久(1976—1986)的IGCP第148号专题“定量地层对比”既要对单个地层的某种特征(如岩性, 化石含量和组成、时代等)进行对比, 又要对群体特征, 即地层序列, 或由某种特征构成的时间序列进行对比, 这就需要查明地层和地层序列的数学特征。在对地层时间序列进行分析时, 可以从不同角度, 不同条件, 不同对象等等分别进行研究, 这就需要查明并依据不同的地层时间序列数学特征, 如表1-1-1所示。

表1-1-1 地层时间序列分析特征表

1. 连续及不连续
A. 连续时间序列 B. 不连续时间序列 C. 由于离散观测而使连续系列表现为不连续系列 D. 由于离散组合而使连续系列表现为不连续系列
2. 均质和非均质
A. 均质时间序列 B. 非均质时间序列 C. 具有长期趋势的非均质序列 D. 显示各种旋回模式的非均质序列
3. 单元和多元时间序列
A. 单元时间序列 B. 多元时间序列
4. 时间序列变量类型
A. 名义型, 包括“事件”类型变量 B. 有序型 C. 间隔型 D. 比例型
5. 时间序列变量的地层意义
A. 岩性地层对比或岩性地层分带 B. 生物地层对比或生物分带 C. 时代对比 D. 相对生物地层值 (RBV) E. 混合变量: 岩性及生物地层数据的混用
6. 地层时间序列分析的对象
A. 单一时间序列内部结构 B. 二个或更多时间序列之间关系 C. 按单位进行的时间序列对比 (相当于“固定间隔”对比) D. 按样品集进行的时间序列对比 (相当于“逐点”对比) E. 成对对比 F. 同时多重对比 G. 固定区间化对比 H. 分带-综合时间序列的构置 I. 样品-综合时间序列的构置
7. 确定性和概率性角度
A. 确定性角度进行地层对比或分带 B. 概率性角度进行地层对比或分带

(据J.C.Tipper, 1988, 简化)

由表1-1-1所列各种内容可见,进行不同类型的定量地层对比,需要查明和依据的数学特征显然是不同的。

二、地质体数学特征的作用和意义

地质体的定量表征是查明地质体数学特征的基础。地质体各种属性、各种标志的定量化是需要研究和解决的首要问题。例如对矿体而言,可分为内部标志和外部标志。内部标志如矿石品位、体重、物质成分、结构构造、矿化连续性等。外部标志如规模、形态、厚度、产状等。这些标志有些可以获取定量观测值,有些只能获取名义型或有序型数据。近年来,有人采用模糊集概念量化所有的描述性变量。

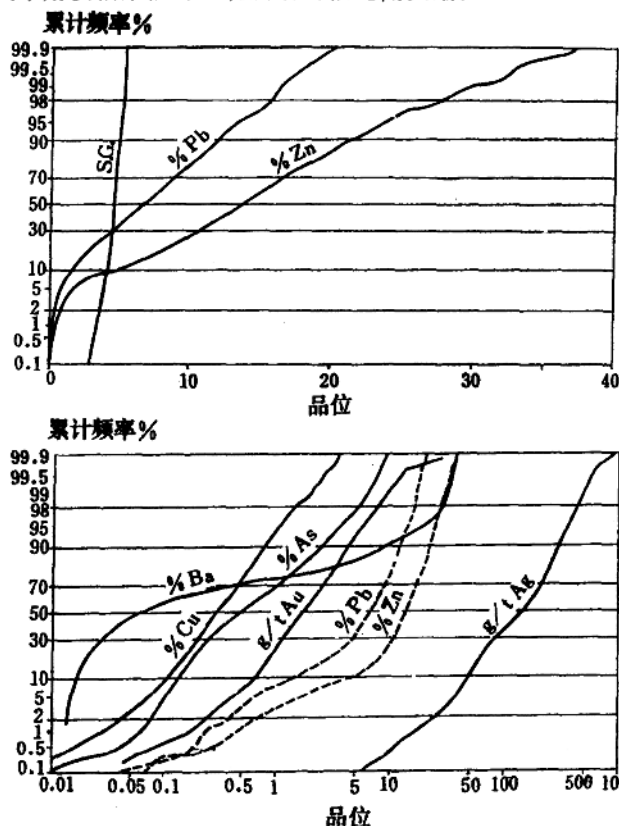


图1-1-1 矿石相对密度(S.G.)与元素的累积频率
(据G.J. McArthur)

通过地质体数学特征的研究可以提供很多有益的,有时是具有成因意义的信息。例如对矿体内部标志、各种有用组分品位数学特征的研究可以提供矿化阶段和一些有实用价值的信息。

澳大利亚G.J.李克阿瑟详细研究了西塔斯曼尼亚块状硫化物多金属矿床Hellyer矿体中各种组分的数学特征,见图1-1-1至图1-1-3。

图1-1-1为矿床中Cu、Ag、Au、As、Pb、Zn等元素含量的概率图。其中Cu、Ag、Au、As显示对数正态分布,而Pb、Zn则近于正态。Ba元素却呈具明显拐点的、显示为两种成分总体混合的概率曲线。其一是作为矿化背景而广泛分布的重晶石矿化,另一总体则来自构成矿体上盘顶盖的块状重晶石矿化。矿石相对密度(S.G.)的概率曲线则显示负偏分布特征。

显示对数正态性的元素,如Cu,具有十分明显的比例效应,见图1-1-2。图中每个点表示一个钻孔截穿矿体所得长度为1m的样品化验值的平均数及标准差。Cu的平均数与其标准差 σ 呈明显正相关联系。而近于正态分布的Zn则不具有上述的比例效应。

在考查元素间相关特征时发现,对于有两个或以上共生组合相的金属在许多情况下与其它金属之间不具线性相关关系。如图1-1-3所示Pb对Ag的相关关系图有两个明显的趋势,一个从原点向上的正相关点子趋势是由Pb与方铅矿密切共生的黝铜矿中之Ag之间的相关

而构成的。而图中的水平趋势则是黄铁矿品格中的Ag与Pb毫无关系的表现。这一特征也表明黄铁矿中之Ag没有回收的经济价值。图1-1-4是As对Au作图，图上显示了点在一个趋势带内集中和点的散布并存的现象。这是由于毒砂有两个不同的共生相，其中一个明显含Au的，而另一个则与Au无关。电子探针研究表明，围绕黄铁矿颗粒生长的次生毒砂含有300—500ppm的Au，而块状和脉状毒砂则显著贫Au。对矿石数学特征的这类研究不仅具有成因意义，而且对采、选、冶等工作也有实际意义。

我们在“试论地质体数学特征”一文中曾指出过地质体数学特征的区分和鉴定意义。特别是在研究大量地质体本质特征及关键标志或其组合的数量规律性质，利用数学特征值鉴别和区分地质体将是一个十分有效的途径。

利用地质体数学特征进行预测的

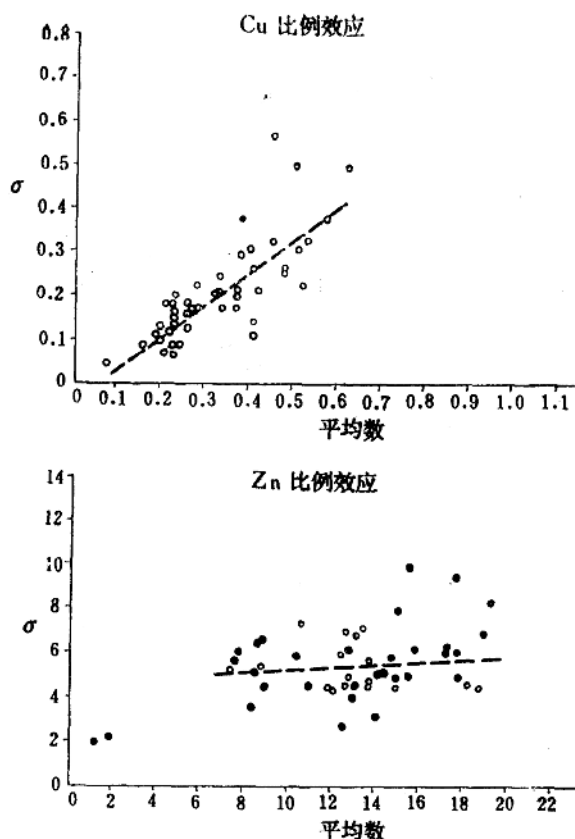


图1-1-2 Cu、Zn比例效应图
(据G. J. McArthur)

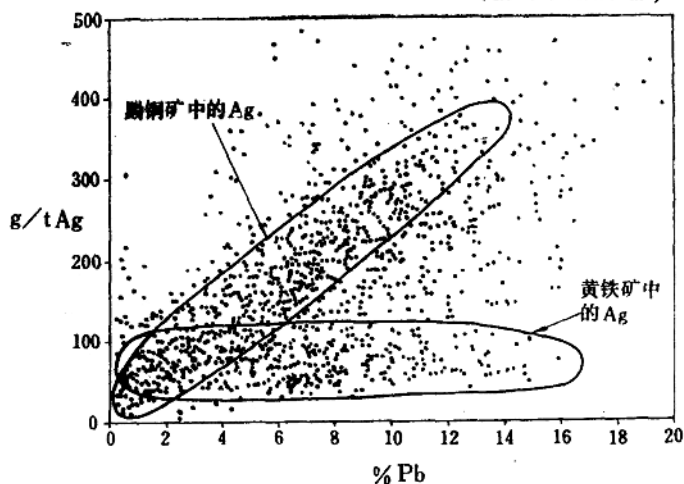


图1-1-3 Ag对Pb的散点图

潜力也是很大的。例如H. V. 帕利瓦等，曾利用矿床规模的数学特征——吉波夫律预测印度Pb-Zn矿资源取得一定效果。印度已知的21个前寒武纪Pb-Zn矿累积储量为

21.346Mt。当把已知的最大矿床 (Agucha 矿床, 其储量为 9.265Mt) 作为第一号最大矿床在双概率纸上作图时未显示任何规律性。由于还估计有 15.96Mt 的新增储量, 所以试验把 Agucha 矿床作为第二号大矿床排序, 并用吉波夫律加以拟合时, 则所有已知矿床恰可在双概率纸上排成一条直线, 并指明: 1号, 3—8号, 11—17号及 19—21号矿床缺失。如果吉波夫律确实是矿床规模的一

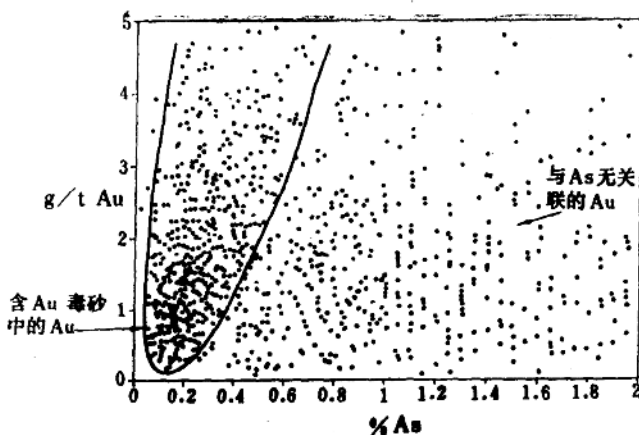


图1-1-1 Au对As的散点图

(据G.J.McArthur)

种数学特征, 则可指望上述暂时缺位的矿床将在前寒武纪岩石中发现, 对于用吉波夫律法预测出的总共 87Mt 金属储量来说, 现在仅仅发现了 25% 的矿量。

上面举的例子都说明地质体数学特征的查明可以提供具有成因、区分、鉴别、预测以及其它实际意义的信息, 这种信息是其它学科所不能取代的, 因而有其特殊意义。尽管如此, 地质体数学特征问题仍是一个有待进一步研究的新课题。

主要参考文献

- 赵鹏大, 1982, 试论地质体的数学特征, 《地球科学》, №1.
- Isaaks E.H. and Mohan Srivastava R., 1988, Spatial Continuity Measures for Probabilistic and Deterministic Geostatistics: Math Geol., V.20, №4, P.314.
- Paliwal H.V., Bhatnagar S.M., Halder S.K., 1983, Lead-Zinc Resources Prediction in India: An Application of Zipf's Law: Math. Geol., V.18, №4, P.543—544.
- Tipper J.C., 1968, Techniques for Quantitative Stratigraphic Correlation: a Review and Annotated Bibliography, Geol. Mag., 125 (5), P.475—494.

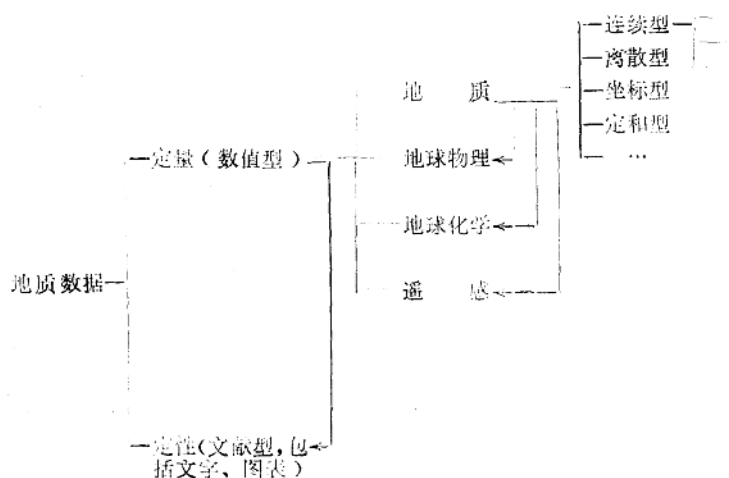
第二章 地质数据分析

第一节 地质数据的统计分析

一、地质数据的分类及预处理

对于地质数据,各研究者从不同的角度出发进行了不同的分类。如从统计学观点通常将数据分为定性和定量数据;从信息系统出发将数据分为数值数据和文献数据;从数据的意义及数量概念的完整程度分名义型数据、有序型数据、间隔型数据及比例型数据;从数学定量方法分为纯量、向量、定和及坐标数据;按数据与研究对象之间的关系分地址性、归属性、准则性及因素性数据;从地球科学各专业学科又分为地质、地球化学、地球物理、卫星影像…等数据;还有人从数据的应用出发将地质数据分为原始数据及方法数据。

在这里,我们对地质数据的分类,主要考虑两方面:一方面考虑地质数据的特点,另一方面应考虑地球科学研究中便于计算机对数据的储存、信息提取和数据的加工处理。采用以下构思分类:



在同一类型的数据中,还应有尺度水平的概念,如同是带有坐标的化探元素分析数据,应有比例尺的区别,1:20万、1:5万、1:1万、1:2000等化探数据,因其研究目的及意义不同,应分别归类。

地质数据不等于地质变量,一般习惯用法,地质数据是指地质科学研究中所需的大量地质信息(包括数字、文字、图表),而地质变量是指参与建立数学模型的成分和参数,地质变量来源于地质数据。反过来说,地质数据是构置地质变量的基础。有的地质数据可以直接作为地质变量去使用,但多数地质数据则需要经过加工处理方可构置为地质变量,将地质数据构置为地质变量时,首先要对地质数据进行预处理。

地质数据的预处理:这里所谈的预处理是指用地质数据构置地质变量前首先必须排除或压低数据中所包括的随机干扰(噪音)而突出有用信息,从而提高数据的可利用程度以增加

所构置变量的可靠性。通常包括对数据的各种变换、筛分、可疑观测值的剔除、奇异值的稳健处理、缺失数据的补齐、过密数据的抽样、数据的网格化、以及对不同时间、不同技术条件下所获得的资料水平的不一致性和专有性数据所进行的分析处理等。

数据的变换及筛分：通常是对所获取的各类数据查明其分布律及各种特征数，对偏离正态的数据进行变换如对数变换、平方根、反正弦变换；对非线性相关数据的线性转换；对不同单位不同数量级数据的统一量纲的变换；对于呈混合分布的数据进行筛分。数据的变换及筛分的预处理方法在各种有关书籍中叙述较多，这里不一一列举。

数据的校正：这方面的预处理在各类数据中出现情况甚多。如，对地球化学测量数据，为了消除或压低由于地形、河流条件、及季节性气候变化等因素对背景值的影响，在构置或使用化探数据作为变量时，必须对数据进行背景值的环境校正。同理，对激发极化测量的电阻率数据，为了排除土壤、地形、环境因素的影响，亦必须通过对远景区数据的环境校正预处理，排除环境异常而突出矿化异常。这里列举一个位于湖泊及森林地带中的铜远景区，557个样品的土壤 Cu 的等值线图及电阻率数据等值线图（图2-1-1、2-1-2、2-1-3），由图表明，进行环境校正后异常有明显变化，用校正后数据构置变量进行预测效果会有明显提高。

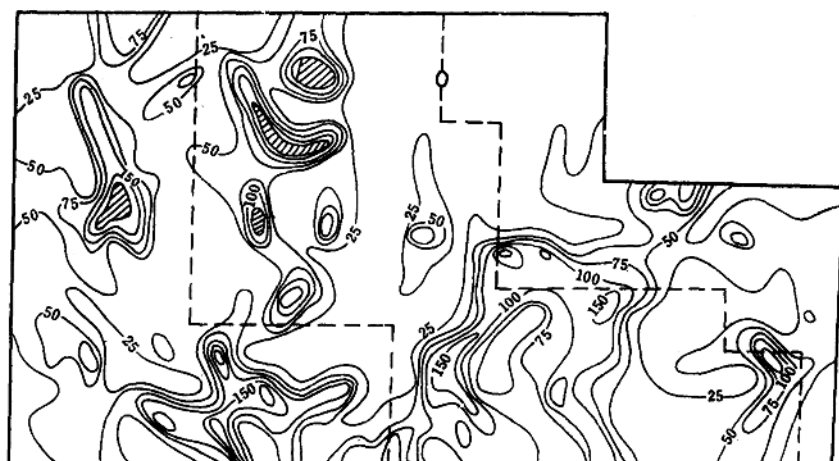


图2-1-1 土壤中铜的等值线图（原始数据）

（斜线面积 > 200 ppm）

（据 R. R. 克尔伯特，1976 年）

一般航磁资料给出的通常是磁场总强度的异常 (ΔT)，而地面磁测资料多数是磁场的垂直分量异常 (ΔZ)，因而在研究区特别是位于低纬度的研究区内，由于斜磁化的影响会造成 ΔT 及 ΔZ 异常形态复杂化且异常中心发生偏移，致使解释困难且掩盖了有用信息，如果直接用此异常作为变量参加建立综合找矿模型，必然导致错误的结果，为此需要对原始数据进行二维频率滤波方法的预处理，通过处理进行位场转换将原始观测数据场转换成为能够直观地反映地质事件或更便于解释的各类位场，如将 ΔT 转换为 ΔZ 、延拓、求导、化极等等。用转换后的异常数据作为变量或构置新的变量，将大大提高预测的效果。同样，重力数据通常也需要进行类似的预处理，方能获得令人满意的结果。

为了克服由于卫片图象的失真畸变而造成的图象分辨率和对比度的下降，可对卫片数字

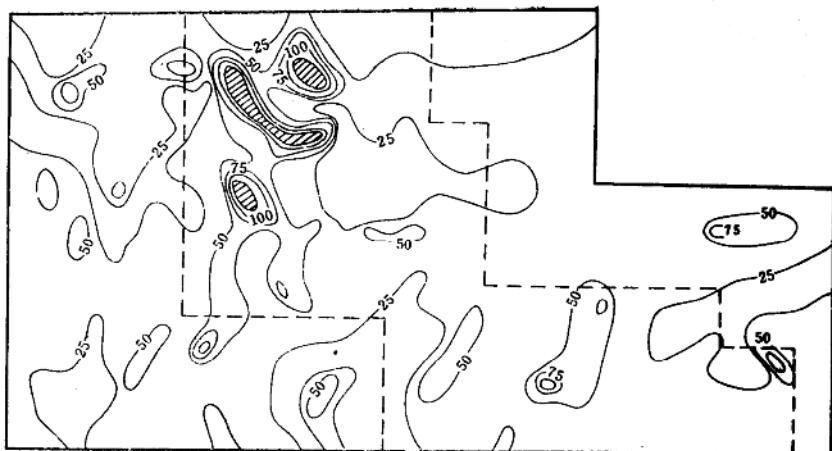


图2-1-2 进行环境校正后铜的等值线图
(斜线面积 $> 150\text{ppm}$)
(据R.R.克尔伯特, 1976年)

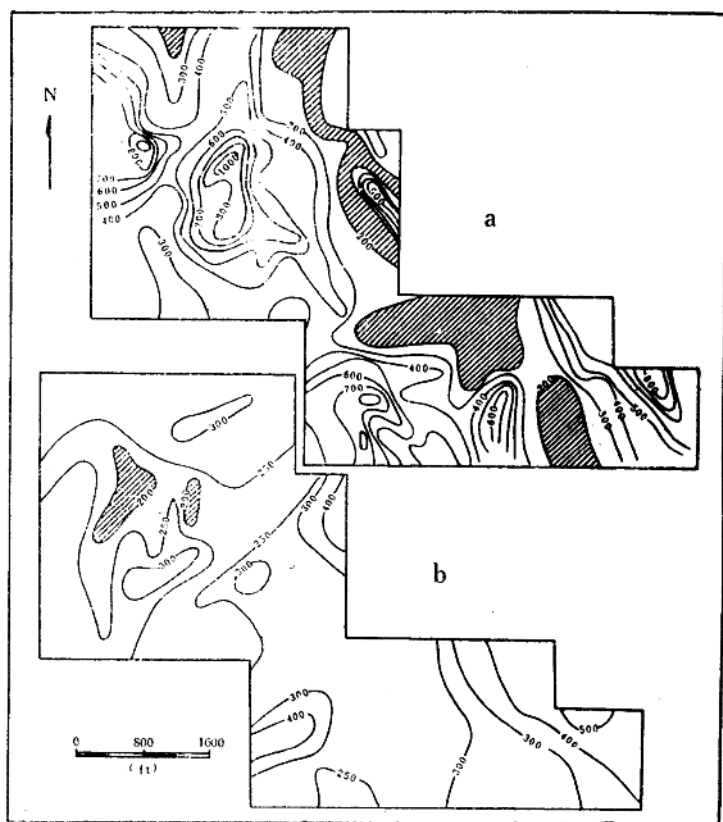


图2-1-3 从激发极化法测量的电阻率数据中排除环境控制的效果
a. 电阻率等值线图(原始数据); b. 进行环境校正后的电阻率等值线图, 斜线面积 $< 200\Omega$
(据R.R.克尔伯特, 1976年)

图象进行空间量化和灰度级量化的处理,用被空间量化和灰度级量化的图象(二维象元矩阵)构置变量参加建立模型。再如,对同一研究对象所采集的样品体积不同,但是由于样品数量的限制又必须利用这些不同体积样品所提供的信息,一般情况下,样品体积越大会导致组分含量的方差越小,且含量级别降低,而形成不同的频率分布,因而使用样品体积不等的组分含量数据时,首先应进行条件校正的预处理,通常采用体积加权法或双因素方差分析。

数据的均匀化:地质数据在空间分布上由于数据的不齐全经常分布很不均匀,在进行统计分析研究时为了不丢失信息,一般采用插值法、统计法及计算机模拟方法对缺失数据进行补齐;对过密数据进行抽样;或者用滑动窗及单元算术平均值的办法对空间分布不均匀的数据进行网格化处理。预处理后的数据构成均匀分布的二维数据矩阵,可直接作为变量或构置新的变量。

可疑观测值的剔除及奇异值的稳健处理:一批数据中的个别数据在数组中与其它数据相差非常大,如对此值不加处理直接参与建立模型,必然造成错误的结果。对此种数据如果判断为可疑值即可剔除,通常用格拉布斯法进行判定,令 α 为危险率,即实际上不是可疑数据而误判为可疑数据所犯错误的概率,地质上一般取 $\alpha=0.05$,令 X_k 为被检验的可疑数据,计算 $T^* = \frac{X_k - \bar{X}}{S}$ 式中 $\bar{X}$ 为样本平均值, S 为样本的标准差,查表得检验临界值 T (给定 α 及样品数 n 即可查表),若计算值 $T^* > T$ (表列值)则 X_k 即为可疑数据应予以剔除。若某数据与数组数据相差很大,但又不愿剔除,可视为奇异值,可以进行稳健处理以消除其影响。

其它方面的预处理,如不同时间、不同技术水平条件下所获得的化探数据,在利用时必须消除不同元素因分析方法灵敏度不一而造成的含量不可比性,使之变成大范围内的可比参数。为此,可以将某元素的原始分析数据转换为浓度克拉克值

浓度克拉克值 = 异常平均含量 / 区域性(或全球性)克拉克值

浓度克拉克值不但增加了大范围内元素的可比性,而且其本身具有元素富集或贫化程度的度量意义,可以作为或构置变量参加预测模型。

以上列举的各类数据预处理方法可以大大提高地质数据的可利用程度(包括精度和数据容量),是进一步构置地质变量的基础,必须给予足够的重视。

二、地质变量的研究

地质变量的研究主要是为了保证研究成果取得最佳效果的前提下,最大限度的简化模型成分。所以,对经过预处理的大量地质数据在进入数学模型之前,必须经过反复研究——进行若干必要的统计分析和进行地质变量的构置(即地质变量的取值、变换和筛选)。同时,地质变量研究的过程,亦是深化地质认识的过程,通过研究变量可为解决某些地质问题直接提供信息。

在“1:20万铜陵幅资源总量预测方法研究”中,对地质变量的研究主要作了以下两方面的工作:

1. 研究区内找矿地质条件的统计分析

目的在于运用统计分析方法,从查明反映矿床的数量规律性入手,研究控矿地质条件及成矿规律,为建立该区矿床成因模式及找矿模型提供信息,进而指导选择及构置参与预测模型的变量。