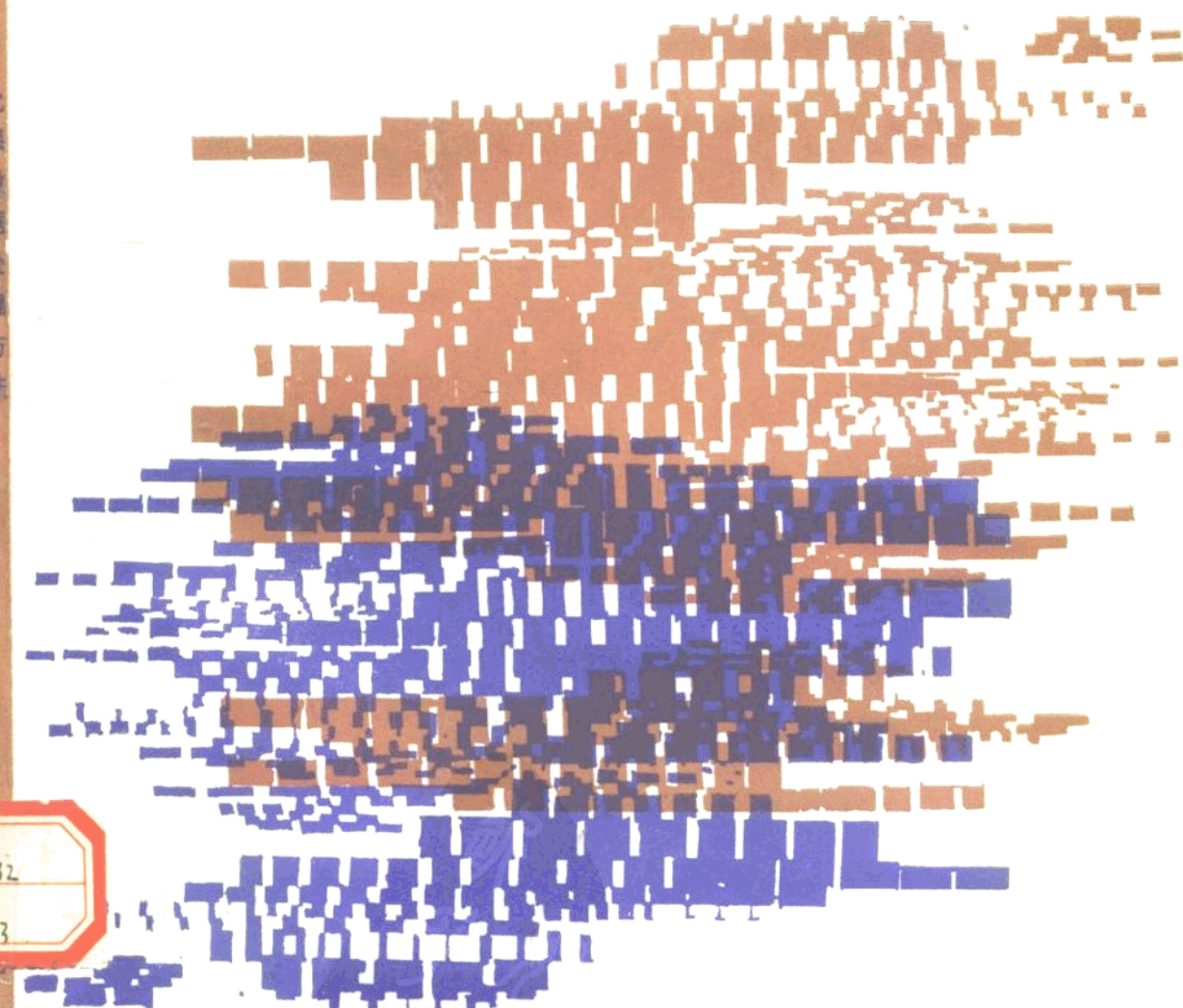




普通高等教育地质矿产类规划教材

化探数据处理方法

吴锡生 主编



地质出版社

- 072205



普通高等教育地质矿产类规划教材

化探数据处理方法

吴锡生 主编

sy50/17



地质出版社

(京)新登字 085 号

内 容 提 要

本书简要介绍了化探数据处理方法的发展过程和分析数据的质量评价与监控方法,系统地阐述了区域化探与矿区化探中的数据处理和编图方法。本书内容以应用为纲,既有方法理论,又有应用实例,基本反映了目前化探数据处理的发展水平,可作为高层次化探专业的教材,也可供从事化探专业的高级技术人员参考。

* * *

本书由林存山、沈镛立主审,经地质矿产部地球化学专业课程教学指导委员会 1991 年 10 月召开的第五次全体会议审定,同意作为高等学校教材出版。

* * *

普通高等教育地质矿产类规划教材

化探数据处理方法

地质矿产部教材编辑室编辑

吴锡生 主编

*

责任编辑:周继荣

地质出版社出版

(北京和平里)

中国地质大学印刷厂印刷

(北京海淀区学院路 29 号)

新华书店总店科技发行所发行

*

开本: $787 \times 1092^{1/16}$ 印张: 8.625 字数: 200,000

1993 年 10 月北京第一版 • 1993 年 10 月北京第一次印刷

印数: 1—1000 册 定价: 4.00 元

ISBN 7-116-01342-3/P · 1111

前 言

《化探数据处理方法》是1989年在广州中山大学召开的地质矿产部地球化学专业课程教学指导委员会上讨论确定的,当时明确为硕士研究生用教材。1991年9月,我们将送审稿分送主审林存山教授与沈镛立教授,承蒙他们在百忙中对初稿进行了全面、仔细的审阅,使书稿中的疏漏与不妥之处得以完善。1991年10月15日至21日在(西安)西北大学召开的地球化学专业课程指导委员会会议上,以中国地质大学阮天健教授为组长组成的评审小组,对本教材又进行了仔细、系统的评审,也提出了许多恳切的修改意见。在此,我们对主审与评审小组给予编写过程中的帮助与指正,表示深切的感谢。本教材由吴锡生主编,其中绪论、第一章由吴锡生编写,第二章的§1、§3、§4、§5和第三章§1、§2、§4、§5、§6由吴锡生、纪宏金编写,第二章的§2和第三章的§3由陈明编写。纪宏金和陈明参加了书稿的校对工作。长春地质学院绘图室的同志们清绘了本书的插图,在此也表示感谢。

由于本课程是在学习《数学地质方法》或《多元统计方法》基础上开设,所以本教材侧重于区域化探与矿区化探的数据处理方面的应用,涉及相应的统计方法内容不再作详细介绍,目的是引导学生开阔思路。教材中的应用实例,大多是我们近几年在生产或科研工作中的成果,少量引用于目前国内外一些新的文献资料。但因作者水平有限,教材中难免有不妥或错误之处,敬请读者不吝指正。

编 者

1992年10月

目 录

绪论	1
一、化探数据处理的发展	1
二、化探数据处理在国内外应用现状	3
三、化探数据处理在勘查地球化学学科中的地位与作用	6
第一章 化探样品分析质量及其监控方法	7
§ 1 分析方法	7
§ 2 分析质量控制的统计学基础	8
一、基本概念	8
二、化探数据中的误差	10
§ 3 分析质量的监控方法	11
一、分析质量监控的目的	11
二、分析质量监控的内容	12
三、分析质量监控的具体作法	12
§ 4 采样与分析质量的评价方法	15
一、分析质量的评价	15
二、采样质量的评价	15
第二章 区域化探数据处理方法	18
§ 1 区域化探中背景与异常的确定方法	18
一、应用多元回归(逐步回归)方法校正背景	18
二、应用稳健多元线性回归分析确定背景与异常	24
三、克立格法	28
四、多元地球化学背景与异常的确定	36
§ 2 区域地球化学图的编制	46
一、国内对区域化探地球化学图的编制的要求与方法	46
二、欧美国家的区域地球化学图件的编制方法	47
三、计算机成图简介	51
§ 3 区域地球化学异常的分类	54
一、聚类分析	54
二、判别分析	56
三、多变量趋势面分析	61
四、对应分析	71
五、多变量场方法	73
§ 4 区域地球化学异常的评价	76
一、多元得分和方法	77
二、特征分析	78
三、RESMA 系统	82
§ 5 应用区域化探异常预测矿产远景储量	86
一、应用回归分析预测矿产远景储量	88

二、应用分步循环回归预测矿产资源量	91
第三章 矿区化探的数据处理方法	93
§ 1 数据的归并	93
§ 2 矿区化探中背景与异常下限的确定	94
一、EDA 法	95
二、稳健统计方法	96
三、应用因子分析分解地球化学场	99
§ 3 矿区地球化学编图	104
一、地球化学断面图	104
二、地球化学剖面图	105
三、地球化学异常剖析图	105
§ 4 异常的分类与评价	105
一、判别分析	106
二、簇群分析	106
三、标准矿化估值法	107
四、成矿能	111
§ 5 矿体剥蚀深度的评价	112
一、相关指数 ρ	112
二、累乘（加）指数（剥蚀深度指数）	112
三、建立在剥蚀深度指数基础上的回归分析方法	113
四、空间延拓法	118
§ 6 预测远景储量	125
一、标准矿化估值法	125
二、回归估计法	125
参考文献	130

绪 论

勘查地球化学开创于 30 年代初。当时原苏联称作金属量测量,主要应用于大比例尺地质找矿,目的是协助物探方法区分矿与非矿异常。1950 年,原苏联在中哈萨克斯坦开展了大规模 1:5 万普查金属量测量,结果发现了很多铅、锌、钨、钼、铜的工业矿床,还扩大了许多已知矿化点的远景。1954 年,美、英、加三国的化探工作者在加拿大新不伦瑞克 8 万平方公里的面积上联合进行了大规模的水系沉积物测量,也取得了良好的找矿效果。可见,由 30 年代至 50 年代初是化探从局部的勘查走向区域性普查,由地质、物探工作的战术辅助手段逐步转变成地质战略布署的主要技术方法之一的时期。如今勘查地球化学已发展成为一门独立的学科。

区域化探的开展,使化探的采样是多介质的,分析是多元素的,研究是多目标的。分析测试的元素可达数十种(我国是 39 至 41 种元素),即使是成矿带的地球化学普查工作,往往测定十余种元素。在矿区化探工作中,无论是采集岩石样品或土壤样品,也都测定 10 至 20 种元素。因此,勘查地球化学所获得的数据量与日俱增,极为庞大,若用手工方法就很难充分发掘隐藏在数据中的大量地球化学信息。从 50 年代开始,国外就将多元统计方法与计算机引入勘查地球化学的数据处理与成图。博依尔(Boyle)和斯密斯(Smith)认为这是加速化探发展的重要成就之一。下面简要介绍化探数据处理的发展与国内外应用现状。

一、化探数据处理的发展

地质、地球化学家最早进行的数据处理是统计化学元素在地壳中的平均含量。1924 年克拉克(F. W. Clarke)首先提出元素在地壳中的分布数据。1932 年根据进一步的统计计算工作又发表了新的克拉克值。1949 年维诺格拉多夫(А. П. Виноградов)、1950 年兰卡玛(K. Rankama)均发表了有关克拉克值的资料。1953 年 8 月苏黎士举行的第一届国际地球化学会议上,专门讨论了地壳中元素平均含量值(“克拉克值”)的修改问题。自 1954 年后美国地调所地球化学与岩石学部每年进行 75000 件样品分析,其中包括 44000 件鉴定,有专门的工作部门来进行大量数据的统计和科学的加工及分类工作,共同研究地壳中元素的平均含量。1954 年阿伦斯(L. H. Ahrens)根据大量统计,发表了“元素对数分布”的论文,引起了世界各国地球化学家的讨论。1963 年我国黎彤也发表了“中国岩浆岩的化学成分”。1964 年泰勒(S. R. Taylor)发表了“地壳中化学元素的丰度”。克拉克值所以引起地球化学家的重视并用数十年加以研究是因为克拉克值不仅有很大的理论意义,而且还有其实用意义。克拉克值是一种度量标准,我们可用它来对比地壳上的某地段、某类型的岩石或矿石中某种元素的含量增高或降低,也可用这一标准来研究化学元素的集中和分散,确定某一成矿区域上的特征元素,即岩浆的成矿专属性等。因此,克拉克值的研究不仅对地球化学的发展有重要意义,而且是勘查地球化学早期发展中地球化学背景值计算的思想与理论基础。

1936 年,索弗朗诺夫(Н. И. Софронов)提出了“分散晕”的概念,涉及到与克拉克值对比下的分散元素富集的分布形式;1937 年,维尔纳茨基(В. И. Вернадский)将地壳平

均含量值与已知岩体中元素的平均含量之比值称作“浓度克拉克值”，为化探中计算衬值提供了理论依据。50年代初，索洛沃夫（А. П. СОДОВОВ）提出了应用概率统计方法确定当地正常场与最低异常（ c_0 ）的方法（即 $c_0 = c_0 + 3\sigma$ ）。从此化探数据处理中的一个基本问题——“背景与异常的划分”就产生了。这个问题至今仍在不断的探讨之中。

化探数据处理发展中的第二个基本问题是研究元素的共生组合规律。地球化学家在30年代就已认识到，任何一个矿床绝非单一元素组成，而往往是有特定的元素共生。他们提出了一些典型的有规律的元素组合，如Li—Be—B, Fe—Ba—Pb, Cr—Ni—Mg—Fe, S—重金属, Zn—Pb—Ag, As—Au, Sn—W—Li, Rb—Cs, I—Br—Cl, Te—Au—Ag—Hg, Hg—Sb—As, Au—Ag—As—(Te), Cd—Zn等等，如今都已被化探工作者作为常用的指示元素组合。研究元素间的共生组合关系，可以采用统计中的相关分析。30年代霍蒂林（Hotelling, H., 1936）就发表了“两组变量间相关性”的论文。对此大家已熟知，这里不多赘述。因子分析作为一种研究多变量间相互成因联系的方法是由斯皮尔曼（Spearman, C., 1904）作为心理学的一种手段提出来的。克鲁巴英（Krumbein, W. C., 1957）首先发表这种数据处理方法在地学方面的应用，如今在国内外已广泛地应用于区域化探与矿区化探之中。

聚类分析是另一种可用于元素分类的方法，它也可用于样品间的分类。这种方法最早是由豪德（Horton, C. W. 1964）与帕克斯（Parks, J. M.）等介绍的。

1970年法国奔斯克立（Benzecri）综合了R型与Q型因子分析，提出了对应分析，使得变量与样品间的相关关系可以同时得到解释。

1941年巴雷脱莱特（Bartlett, M. S.）首先提出典型相关分析，它是研究一组变量与另一组变量间的相互关系的数据处理方法。1967年李（Lee, P. J.）将它发展成典型趋势面分析，研究一组变量在空间的变化规律，随后他又发表了这种方法的FORTRAN IV程序。

总之，研究元素间共生组合或成因联系的数据处理方法自60年代以来，方法愈来愈多。这里不再一一介绍。

化探数据处理发展中的第三个基本问题是研究元素或元素组合在空间的变化规律。1959年克鲁巴英（Krumbein, W. C.）提出了趋势面分析，用趋势值表示元素在空间的背景变化，而用剩余值反映异常的空间分布。1966年克立格（Kriging, D. G.）提出了二维加权移动平均趋势面分析，也是反映背景与异常空间变化的数据处理方法。

化探数据处理发展中的第四个基本问题是分类问题，如样品的分类、异常的分类等。判别分析、聚类分析是常用的分类方法。在安德生（Anderson, T. W., 1958）的“多元统计分析引论”中有判别分析的介绍。1963年浦杜（Potter, P. E.等）用它区分海水与淡水（主要根据水中泥质沉积物的微量元素）。这种方法目前在国内外化探工作中也已广泛应用。

化探数据处理发展中的第五个基本问题是异常的评价。地球化学异常的评价涉及异常的分类、异常的元素组合、异常的规模、异常反映的矿体剥蚀深度，以及根据矿异常预测矿体的远景储量等。这些问题，我们将在第二章、第三章中再作介绍。

化探数据处理发展中的第六个基本问题是地球化学图的制作。在化探开创的初期，主要是用元素的含量等值线图来表达。当用一定间隔网采样时，也可用平面剖面元素含量曲线表达。随着化探的发展与统计方法以及计算机的应用，地球化学图的制作随不同的任务与目的而呈现多种多样，这在以后区域化探与矿区化探的图件编制中再作介绍。

化探数据处理发展中的第七个基本问题是分析数据的质量评价。在化探工作的初期，就

已注意到样品的代表性与分析误差问题。但当时由于原苏联大量采用的是光谱半定量与现场的半定量化学测试方法,其分析灵敏度与准确度都较低;而且金属量测量还只在矿区外围的小面积范围内开展,其测定元素的含量差异较大,故元素含量的浓集地段还是可以圈定的。随着区域化探的开展,化探对样品分析测试的灵敏度与准确度的要求越来越高,而且要求不同批次样品间的分析误差保持在同一个误差范围之内。这样就产生了标样与分析测试的监控方法。同时也应用了检验取样与分析质量的统计方法——方差分析,以此及时检查分析数据质量。

化探数据处理发展中的第八个基本问题是计算机的应用,化探分析数据的增长使得在应用数学统计方法时很难靠手工计算实现。计算机在化探中的应用,使化探数据处理技术与方法得到迅速发展。如今某些国家对野外取样的记录已开始实行卡片化、数值化,并通过计算机直接成图。在70年代初,美国地调所研制了数据处理系统(RASS)。1977年加拿大地调所研制了GAS系统。我国为了区域化探也研制了GC-79数据处理系统。

上述化探数据处理发展中的八个基本问题,只是化探数据处理发展的一个提要,如今国内外涉及化探数据处理的统计学书籍和杂志种类繁多,涉及化探数据处理的论文估计可达千篇以上,我们一时还难以完全收集。故本书不可能包罗化探数据处理发展之全貌,在此仅起抛砖引玉之作用,希望对化探数据处理有志的同事共同努力加以完善。

二、化探数据处理在国内外应用现状

(一) 国外化探数据处理的现状

1988年国际勘查地球化学工作者协会委托R. J. 豪沃思主编了《勘查地球化学手册,第二卷,化探中的统计学和数据分析》。该书请英、加、前苏联、法、德、瑞典、土耳其和挪威等国的著名地球化学家、应用数学家、分析化学家和地质学家撰写了论著,其论题涉及数据库、取样设计、分析质量监控、各种数据处理方法、成图技术以及世界各地化探数据处理的应用状况的综述等,基本概括了国外当前化探数据处理的现状与发展水平。全书已由吴传璧等于1990年翻译出版。在这里我们就不再详述。

(二) 国内化探数据处理的发展与现状

1. 试验与普及时期(1965~1980年)

我国化探数据处理从60年代初开始。1965年谢学锦、邵跃在整理钻孔原生晕资料时,应用了滑动平均方法,使矿体周围的规律更清晰。吴承烈、林存山、伍宗华等在区域化探资料整理中也应用概率统计方法,较合理地划分区域背景与异常;1966年徐道一将趋势面分析用于地层学,发表了“地层数据的趋势面分析”;1972~1974年谢学锦、林存山、杨竹溪等用趋势面分析研究了安徽某岩体磁铁矿中铜含量的分布,发现了铜有明显的定向富集趋势,据此可作为岩体中金属为成矿提供物质来源的依据之一。他们还应用了加权内插的方法,将岩体不规则采样点数据变换成规则网格的数据进行成图,其结果铜含量富集趋势与趋势面分析基本一致;在安徽某地铁帽分类的试验中,应用了聚类分析与因子分析,也取得较满意的效果;1973~1975年吴锡生在吉林某地1:5万区域化探普查中应用趋势面分析划分了区域背景与异常,并根据Cu、Pb、Zn的异常,发现了金矿脉,找到了一个小金矿;并应用聚类分析与因子分析研究了金矿区围岩与矿石中元素的组合特征,为金矿成因提供了地球化学依据。在区域化探异常评价时,应用了已知铜矿床Cu、Pb、Mo的面金属量建立判别函数,对未知异常作了判别与分类,取得了一定成效。1972年冶金部由欧阳宗

圻主持在广西桂林举办了地球化学学习班,聘请了於崇文、张本仁、阮天健、林名章等讲授地球化学等理论与方法。在以后内部出版的学习材料中也已涉及化探数据处理方法的内容。1973年在湖南邵阳地质部物探局举办的激电化探学习班上,物探所介绍了应用回归分析方法从某一标高上矿体中铜含量来预测整个矿体中平均品位;应用判别分析区分有工业价值铁帽与无工业价值的铁帽;1973年冶金系统为推广多元统计方法,欧阳宗圻委托阮天健在《地质与勘探》杂志上系统刊登了“数理统计在地球化学探矿中的应用简介”,为普及国内化探数据处理起到了推动作用;同年地质科学院情报所编印了《化探资料数理统计方法》一书(内部参考),介绍了国外1968~1973年化探数据处理的应用实例与综述性论文;1974年冶金部在保定举办了第二期化探表据处理及电算学习班,规模较大,并有上机实习,聘请了於崇文教授等讲授,为时五个月。以后冶金系统在化探资料整理与异常的解释评价中推广应用了多元统计方法。1975年在福建召开的全国第一次区域化探经验交流会上,黑龙江省、长春地质学院、河南省、广西及物探所的代表都介绍了趋势面分析、移动平均、相关分析等在区域化探数据处理中的应用。1976年地质部物探局在北京举办了地质系统电算程序设计学习班,培养了一批物化探应用计算机与程序设计的技术骨干;1976~1977年陈天与和吴锡生在长春地质学院学报上系统刊登了“数学地质讲座”,介绍了数理统计的基础知识与多元统计方法及其在化探中的应用实例;1977年地质部物探局在无锡召开了物化探电算技术交流会,对应用计算机整理物化探资料起了推动作用。1979年地质部物化探局在邵阳举办地学系统化探数据处理学习班,由吴锡生系统讲授了“多元统计方法”,对各省化探应用多元统计方法处理化探资料起到了普及与推动作用(该讲稿1981年由河北省地矿局科技情报室出版,书名为《地球化学探矿中几种多元统计分析方法》);1979年地质部北京计算中心根据原国家地质总局提出的重新整理1:20万区域金属量测量资料的要求,研制成包括数据存储、数据处理、图象显示、绘图等多功能的化探专用处理系统(简称GC-79)大大提高了上述资料整理的速度,并有22个省局重新整理了这方面资料。GC-79系统也为1978年谢学锦提出的全国区域化探的数据处理与成图作了探索性试验,为全国区域化探成图标准化方面作了先导性工作;1979年地质部物探局创办了《物化探电算计算技术》杂志,为交流与推广应用计算机处理物化探资料起到了积极的推动作用。1980年於崇文等编著的《数学地质的方法与应用》和陈天与、吴锡生合编的《数学地质方法》相继出版。这类书籍对我国化探数据处理起到了普及与深化应用的促进作用。

2. 发展与提高时期(1981~1992年)

我国的化探数据处理在经历了60~70年代的普及、推广与实践之后,在进入80年代时已具有进一步发展和提高的基础,具体表现如下:

(1)微处理机的普及和三级计算机应用体系的形成 据地矿部物化探局1990年统计,为物化探工作配备的微型计算机已有71台,袖珍式计算机(含PC-1500)131台,中小型计算机8台,高档微型计算机5台。计算机从计算中心(站)、研究所、院校逐步普及到野外队,其应用从数据采集、整理、成图到解释,并且开始建立数据库。初步形成了三级计算机应用体系。

①全国级专业数据处理中心(站)。如北京计算中心和武汉、长春、成都地质计算站,北京计算中心在GC-79基础上进一步研制成GC-81第二代区域化探数据处理系统。它包括化探数据的存贮、检索、方法处理和自动成图三大部分。系统有较强的按图幅检索的功能,

成图功能亦有所扩大。至1988年已处理了100多个1:20万区域化探扫面资料。发现了矿产地近百处,大大缩短了资料处理周期。

1985年地矿部物化探研究所建立了物化探软件开发中心。为化探研制开发了地区性的区域化探资料统计、处理、成图自动化系统及相应的数据库,为各省、队独立进行区域化探资料处理与成图起到了促进作用。

②省局或大队级计算站。有9个物探队配备了CS-3微型机,19个物探队配备了DUAL-68000高档微型机。他们在软件开发中心的统一组织协调下,集中对物化探应用软件进行研制、开发和移植;交流和推广应用程序约200个,其中定型40个,涉及重磁、化探、数学地质、电法、绘图、系统管理、数据库、不同机器间的信息通讯等。语句总行数达12万多条,基本上可使物化探资料的常规处理不出省队,而且在云南、陕西、浙江等省取得了较好的地质找矿效果。

③分队级站,普遍配备了袖珍式PC-1500机,已开发了一批化探方法处理程序,实现了PC-1500机与IBM-PC/XT(或长城0520机)、DUAL-68000机之间的信息通讯。这对野外进行数据采集和预处理以及小型计算处理起到重要作用。

在冶金系统的物化探队也都具备了数据处理的软硬件条件。在有色系统的各省物探队也都配备了微机及数字化仪,绘图仪等成套设备,可以独立进行数据处理与成图工作,而且冶金与有色系统各队都已建立或正在建立化探数据库。

由上述可见,现已完成全国微处理机的普及和“三级”计算机应用体系的建立,并且已拥有一大批化探软件开发的骨干力量,我国的化探数据处理、解释、成图等工作定会走向一个新的高度。

(2)新的计算方法不断出现 在背景与异常的划分方面,当对化探分析数据作统计直方图后,发现不服从正态分布时则采用了一些转换方法。如对数转换、广义幂变换以及二步转换等方法使转换后数据接近正态分布,而后确定背景与异常或作其它统计处理。此外,也应用了“总体分解法图识别”、“框架滤波方法”、“加权游动平均法”、“普通克立格法”、“泛克立格法”等直接对全部数据进行处理,并可划分出单元背景与异常。对于多元背景与异常的划分已应用了马氏距离(Mahalanobis Distance)进行“多元背景与异常的确定”,应用因子分析对地球化学场进行分解。稳健统计方法、非参数统计方法(如EDA法)、背景校正方法等在背景与异常的划分中也都有应用。总之,如今背景与异常的确定已不再停留在背景加几倍方差的简单计算的方法上,而是出现了多种多样的计算方法。它对区域化探与矿区化探背景与异常的确定提供了良好的统计计算依据,而且其应用的效果也大有改善。

在分类方面,出现了模糊聚类分析、寻根聚类法、罗吉斯判别分析、多变量场方法、最优分割法、RESMA系统等。这些方法在区分矿与非矿异常和样品的分类等方面具有更多的优点。

在元素和样品的相关性研究方面,除了仍在应用回归分析、因子分析、对应分析等方法外,还采用自相关结构分析与马尔科夫分析。这两种方法有助于研究元素含量的分散和富集规律。

在异常评价方面,应用了成矿能评价法、标准矿化估值法、成矿可能度预测法、成矿概率指数法、成矿相对指数、二维或三维延拓方法评价异常深部含矿性等。这些评价方法

均可使化探异常评价趋于量化。

在成图方面,目前已建立了自动成图系统,内容包括绘制平面剖面图、平面等值线图、彩色等值线图、彩色立体阴影图、黑白立体阴影图、点阵打印机绘制灰度图。物化探研究所在 IBM-PC 机以及 MV 4000 机上研制开发了化探资料数据处理成图系统(包括常用统计、多元分析及多元场法等实用化探资料处理程序),而且物化探研究所和中国地质大学(武汉)计算中心均已引进彩色成图仪。有色系统科研院所也已开发了物化探数据处理、图像处理和成图系统,并正准备引进空间信息系统(SIS)。

以上介绍的我国化探数据处理与成图技术的发展,只是国内近十余年中开拓发展的一部分,但由此也可看出化探数据处理的一些发展趋势。大致有以下几方面:

第一,应用数学领域的发展。最早引入化探数据处理的数学方法是多元统计方法,但近年来,地质统计学、模糊数学、数量化理论、稳健统计学等许多新的分支领域也相继引入化探数据处理。尽管某些应用不一定那么理想,但它开拓了人们的思路。例如数量化理论的应用为化探与地质的结合,并实现计算机处理提供了可能,稳健统计学的应用增加了数据处理的可靠性。

第二,一元方法向多元方法发展。化探工作对象一般属于多元系统,因为世界上尚未发现只有一种元素形成的矿产。所以只有根据多元素的组合特征才能正确识别地质与矿产特征,应用一个以上变量建立数学模型,研究相关变量的集合和个体之间相互关系,进行化探数据处理的方法可统称为多元方法,例如众所周知的多元回归分析、趋势面分析、聚类分析、判别分析、因子分析、对应分析、典型相关分析等都是多元统计方法,在成图中应用的元素累加、累乘制图、因子得分图、判别得分图等也是多元综合性图件,这些图件既充分提取了多元信息,又可使图件简单直观。因此,对多个变量进行观测,汇集丰富的原始信息,全面地反映所研究的地质现象和各个有关侧面,促使化探数据处理继续向多元分析方向发展,仍是今后化探数据处理的发展趋势。

第三,由单独的方法应用向系统化方法发展。前面已提及在许多生产、科研和教学单位,都相继建立与开发了包括数据存贮与检索、质量评价、统计计算、自动成图的数据处理系统,而且有些单位已研制了专家系统。由于化探数据处理是系统工程,故这种发展趋势是必然的。

第四,在异常评价方面正逐步由定性评价向定量评价方向发展。虽然在目前条件下要做到完全定量评价还很困难但这也是化探工作需要研究解决的课题之一。

三、化探数据处理在勘查地球化学学科中的地位与作用

化探数据处理是应用数学方法从地球化学原始数据中提取信息,指示变量(主要是化学元素)与各种地质现象的内在联系,为勘查地球化学寻找矿产资源和解决地质问题等服务的方法技术。由此可见,化探数据处理是勘查地球化学学科的一个组成部分,而且是一个中间性工作。在化探数据处理的前阶段有基础性工作,而且后阶段还有解释评价工作。因此,化探数据处理在勘查地球化学中的地位如下图流程:取样→样品分析→数据处理→异常的解释与评价→矿异常的揭露与矿产发现。

可见,化探的数据处理是建立在取样与分析数据质量可靠的基础上的。至于如何评价取样与分析数据质量的优劣以及怎样保证分析数据的质量,这将在下章进行简要介绍。

第一章 化探样品分析质量及其监控方法

化探是在近代微迹元素分析技术的推动下发展起来的。早期的化探样品分析着重于从背景的起伏中去辨认矿异常。为此强调方法必须简单、快速、具有足够的灵敏度，而往往对分析的精密度作出一定牺牲，至于准确度也很少加以考虑。因为要从背景起伏中去辨认异常，并不需要去求得样品中元素含量的真实数值，而只需获得其相对数值即可。

自从区域化探发展以来，化探分析发生了重大变化。一方面是分析仪器的自动化，不但提高了分析结果的精密度、准确度，而且加快了样品分析的速度；另一方面，化探样品分析的成果，已不单纯是从背景中去辨认异常，而且还要去识别区域性异常与地球化学省，并全面研究元素的地理分布，存在着一个拼图问题。从发展趋势看，今后区域化探资料将成为地质科学中的基础资料，除应用于找矿外，还将为农业、生命科学、环境科学等各方面利用。因而区域化探要求化探样品分析在工作效率、精密度、准确度与灵敏度四个方面全面提高。

§1 分析方法

目前国内各省实验室已广泛使用的化探样品分析方法有：发射光谱方法、原子吸收分光光度方法、比色法、离子选择电极法、原子荧光法、X射线荧光光谱法和等离子焰光量计法等。化探样品分析选用的方法应尽可能采用多元素同时测定的方法，而且必须具有较高

表 1-1 区域化探扫描面样品中元素检出限要求 ($\times 10^{-6}$)

元 素	检出限要求	元 素	检出限要求
Ag	0.1~0.02	Mn	30
As	1	Mo	0.5
Au	0.0003	Nb	5
B	5~10	Ni	1~10
Ba	50	P	100
Be	0.5	Pb	1~10
Bi	0.1	Sb	0.2
Cd	0.1	Sn	1~3
Co	1	Sr	5~50
Cr	15	Ti	100
Cu	1	Th	4
F	100	U	0.2~1
Fe	1000	V	20
Hg	0.01	W	0.5
La	30	Y	5~10
Li	5	Zn	10
		Zr	10

的生产效率,以适应大量化探样品日常分析的需要。其检出下限应达到表 1-1 列出的要求。

选择分析方法的准确度和精密度可用 GSD 标准样的方法进行检验。被选用的方法应对八个 GSD 标准样中的每一个样进行多次分析,作如下两种计算:①平均值与该 GSD 标样的可用值之间的对数偏差 $\Delta \lg c$ 或平均值和可用值之间的相对误差 $RE(\%)$;②相对标准偏差 $RSD(GSD)$ 。其结果应符合表 1-2 的要求。

表 1-2 测定主要、次要、微量及痕量元素的准确度和精密度

表示方法 监控限 含量范围	准 确 度		精 密 度
	$\Delta \lg c(GSD) = \lg \bar{c}_n - \lg c_s$	$RE\%(GSD) = \frac{\bar{c}_n - c_s}{c_s} \times 100$	$RSD\%(GSD) = \frac{\sqrt{\sum_{i=1}^n (c_i - \bar{c}_n)^2}}{\bar{c}_n} \times 100$
检出限三倍以内	$\leq \pm 0.20$	$\leq \pm 45$	35
检出限三倍以上	$\leq \pm 0.13$	$\leq \pm 30$	23
1~5%	$\leq \pm 0.04$	$\leq \pm 10$	8
>5%	$\leq \pm 0.02$	$\leq \pm 4$	3

$\bar{c}_n$ 为 GSD 标样经 n 次实测的平均值; c_s 为 GSD 标样的可用值; c_i 为 GSD 标样第 i 次测定的实测值,测定数 n 应不低于 12 次

目前国内引进的等离子焰光谱计和 X 射线荧光光谱仪,都采用 PDP11/03 计算机进行控制,将分析的数据自动存贮在 8 英寸的软磁盘上,再以通讯方式自动将存在软磁盘上的数据转存在 IBM-PC/XT 计算机的软盘上。利用处理程序进行谱线重叠及元素间的基体校正,高浓度标准曲线的非线性校正,直接打印出化探样品的分析报告,并可进行质量监控,打印出一级标样、二级标样及 5% 检查样的参数统计报告和质量统计报告。

§ 2 分析质量控制的统计学基础

在化探样品分析中,即便用一种固定的方法,有同一个熟练操作人员对同一个样品进行多次重复分析,他每次得到的拟测元素含量结果往往并不相同。之所以出现各种不同的分析结果,原因在于称重、加工、溶样、容量测定等步骤都会产生误差,而且全都综合在最终的分析结果之中。此外,化探样品的采集,即便在同一个采样点上采集两个样品,其分析结果也不会完全一样,这就是采样误差。特别是在区域化探中,往往在一个较大的面积内采集几个样品组合成一个样品,并利用该样品的组分来估量整个单位面积内的平均组分。这也难以避免存在采样误差。化探的数据处理是建立在采样与分析质量可靠的基础上进行的,因此我们必须对如何保证数据质量的可靠性有一个基本了解。

一、基本概念

1. 真值与平均值

真值——在无系统偏差条件下无穷多次测定值的平均值称作真值。由于实际工作中不可能对一个样品测定无穷多次,故严格说来,真值是无法确定的。

平均值——同一种测定方法对同一个样品进行多次测定所得结果进行统计计算求得的

估计值称作平均值。平均值计算可用如下公式：

算术平均值

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

均方根平均值

$$u = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

加权平均值 W

$$W = \frac{w_1 X_1 + w_2 X_2 + \cdots + w_n X_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}$$

几何平均值

$$G = \sqrt[n]{X_1 \cdot X_2 \cdots X_n}$$

$$\lg G = \frac{\sum_{i=1}^n \lg X_i}{n}$$

中位数(中位值) 将观测数据依数值大小次序排列时取其中间值(一般从小至大排列)。

上述式中: X_i ——某次(i)的测定值($i=1, 2, \cdots, n$); n ——测定次数; w_i ——某次观测值对应的权系数。

可用值——对同一个样品分别用多种测试方法测得的估计值,同时对每种方法都赋予与其方法局限性的权而求得的平均值。这是目前勘查地球化学建立标准样时测定标准样中每个元素含量时常采用的方法。

2. 准确度和精密度

准确度——多次测定值的平均值与真值的符合程度。

精密度——多次测定的重现性即它们之间的符合程度。

精密度可以用数据的方差来衡量,也可以用标准偏差来衡量,还可以用变差系数百分数来衡量。

方 差

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

标准偏差

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

变差系数

$$P = \frac{\sigma}{\bar{X}} \times 100$$

$\bar{X}$ ——多次测定的平均值。

准确度和精密度是互不相倚的。一种测试方法或结果,可以既准确又精确,即每次测定

的结果都比较接近,而它们的平均值又非常接近真值。也可以是准确的但不精密,即多次测定的平均值接近真值,但每次测定的数值之间相差较大。也可以是精密的但不准确,即多次测定的数值互相接近,但其平均值则远离真值。或者是既不准确又不精密。

当数据不是由一个样品多次测定的结果,而是多个样品的重复测定结果时,则准确度可用测定数值的总平均值与各个样品的平均值之间的偏差来衡量。精密度可用下式计算:

$$\sigma = \sqrt{\frac{1}{2n} \sum_{i=1}^n (X_{i1} - X_{i2})^2}$$

x_{i1} ——第 i 个样品第 1 次测定结果; x_{i2} ——第 i 个样品第 2 次测定结果; $i=1, 2, \dots, n$ 。

3. 检出限和识辨力

检出限——通常称作灵敏度,是用一定分析测试方法能够测出的元素最低含量。采用不同分析方法分析不同元素,检出限会有很大差异。

识辨力——指分析方法能表达分析元素的浓度级的数目。在一定的含量间隔范围内,表达浓度级数目越多,表明该分析方法识辨力越强。

以上介绍的一些统计参数是化探选择分析方法的评价依据,也是评价化探数据可靠性的重要依据。

二、化探数据中的误差

误差即指某种分析测试方法的测定值与真值之差;偏差(偏倚)是指测定值与平均值之差,但习惯上两者混用,不加区别。

误差可以分为系统误差、偶然误差和过失误差。设误差(偏差)用下式表示:

$$Q_{i1} = X_i - \bar{X}$$

X_i ——第 i 次测量的数值; $\bar{X}$ ——多次测定的平均值; Q_{i1} ——第 i 次测量的误差。若

$$\frac{\sum_{i=1}^n Q_{i1}}{n} \rightarrow 0 \quad n \rightarrow \infty$$

则 Q_{i1} 称作偶然误差。

若

$$\frac{\sum_{i=1}^n Q_{i2}}{n} \rightarrow a \quad (a \text{ 为一常数,可正或负}), \quad n \rightarrow \infty$$

则 Q_{i2} 称作系统误差。

过失误差是人为因素造成,如读数错误、看错谱线等引起。

在化探工作中,特别是区域化探中遇到的偏倚往往是可变的。这在以往应用的半定量光谱分析中特别严重。在当前使用较完善的分析方法时也难以完全避免。在我国开展区域化探的初期,试验结果表明,除了随机误差之外,分析方法与方法之间,实验室与实验室之间,人与人之间,季节与季节之间,甚至月与月,日与日之间都有系统误差存在,这种误差称为可变偏倚。

可变偏倚在采样中也同样存在。不同地区岩石出露程度不同,或某些地区有外来覆盖物分布,就带来了采样偏倚,所采样品在某些地区代表性较好,能比较好地反映基岩中元素含

量变化,另一些地区代表性较差,对基岩含量的估量就可能偏低或偏高。

不同地区水系密度不同,水系沉积物样品反映上游汇水盆地中元素含量平均值时也会发生可变偏倚,特别是当汇水盆地中有几种不同岩石分布时。

采样偏倚还可能由于在不同采样地点岸土物质混入比例不同所造成。采样地点上局部环境的变动也会发生采样偏倚。例如土壤或水系沉积物中Fe、Mn、有机物及pH值的局部变化可以使金属含量变化无规律,甚至出现假异常。

不同性质或不同粒级的样品中金属富集粒度不同。如果没有规定统一的采样方法,不同采样人员的采样方法与习惯不同,所采集的样品性质变化太大,也会发生可变偏倚。

不同地点的岩石、土壤、水系沉积物等的均匀性是不同的。在一个采样地点,由于物质比较均匀,重复采样比较接近,在另一个物质很不均匀的采样地点,重复采样会得出相差较大的结果。这是在区域化探采样中经常遇到的可变精度。

在一个地区或图幅之内出现可变偏倚,对化探的解释推断会带来严重影响,它会在图上出现假的变化趋势和假的异常。

把野外采样操作尽量标准化,可以大大抑制采样人员之间的可变偏倚。不同采样点上,由于物质的差异、局部环境的变化所引起的采样偏倚是难以避免的,但今后可以设法把它们在数据的处理中进行校正。岩石出露程度、覆盖物性质及水系密度不同所造成的可变偏倚也需在有解释推断阶段时给以估量。

不同采样地点局部环境的变化,以及不同地点物质均匀性不同所造成的可变精度,如果严重也会造成假异常。但若在一个地点上采集几个样品混合成一个组合样,有助于抑制可变精度的起伏。以后的数据处理(网格化,移动平均等)还可以进一步压低可变精度。通过这些措施,可以使可变精度减少到不致危害解释推断的程度。

至于随机误差,它在任何采样与分析工作中总是存在的。分析实验室通常使用的检查方法(随机抽若干样品,进行重复分析,计算合格率)主要是为了控制这种误差的。但是化探分析不同于一般矿石分析,在化探分析中对于合格率笼统地规定一种标准来对待不同地区的情况与不同要求显然是不够的。往往同等变化幅度的采样与分析的随机误差,在一个地区或图幅内足以影响或歪曲元素含量的真实变化,而在另一个地区并非如此。这就必须针对具体情况作具体分析。如果在数据中没有严重的可变偏倚与可变精度,则可以使用方差分析方法来检验这一地区范围内随机误差起伏是否有可能掩盖了该地区元素含量变化的真实起伏。对随机误差的检验也可以纳入误差监控系统之中。此外,数据处理(网格化、移动平均等)也可以使随机误差有所降低。

§3 分析质量的监控方法

一、分析质量监控的目的

在化探分析中质量监控的主要目的如下:

第一,及时发现批样之内或批样之间的可变偏倚,以便采取措施,从仪器设备、工作环境、工作条件、操作人员及方法各方面查找原因。

第二,为发现不同图幅或不同实验室之间的恒定偏倚和校正提供依据。

第三,对采样中的随机误差进行估计。