

数学地质数据处理系统

谭光国 林守江 黄维纲 王秀惠

(天津地质研究院)

(冶金部自动化研究所)

随着地质科学向定量化发展,地质物化探工作中应用多元统计分析方法,取得了丰硕成果。运用多元统计分析方法处理地质物化探数据,对成矿规律研究、矿产统计预测、构造解析推断、勘探工程布置等提供了有力依据。地质物化探工作中所取得的丰富数据,急需一个较完整的数据处理系统予以处理。由于冶金地质系统计算机资源不足,尚缺乏一个较完整的数学地质数据处理系统,给资料处理工作带来许多困难,特别是对大数据量的处理,过去那种单个程序的处理方法远远不适应于今天地质工作的发展。为解决这一问题,冶金部天津地质研究院与冶金部自动化研究所合作,在日产 ACOS—400 型计算机上,研制了命名为 MGDPS (Mathematical Geology Data Processing System) 的数学地质数据处理系统。为地质物化探工作者提供了功能齐全、使用方便、处理数据量大、输出灵活多样、计算效率高的程序系统。

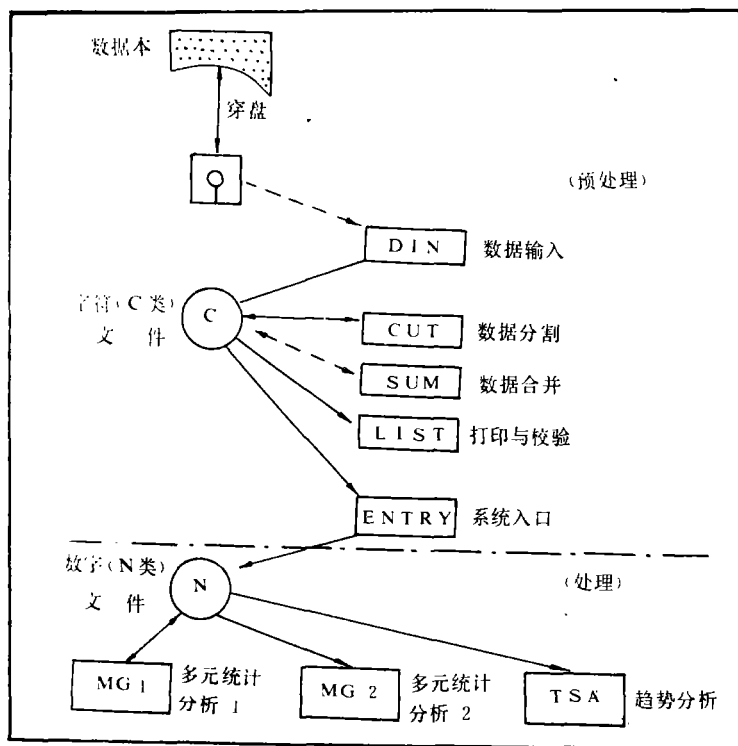


图1. MGDPS系统结构示意图

一、系统结构

1. MGDPS 系统结构如图1所示。系统中各程序模块是按照数据处理流程的需要而安排的。对每一程序模块,又按其功能划分成相互独立的命令块。

2. 系统中所有模块均采用如图2所示的两种标准结构之一构筑而成。

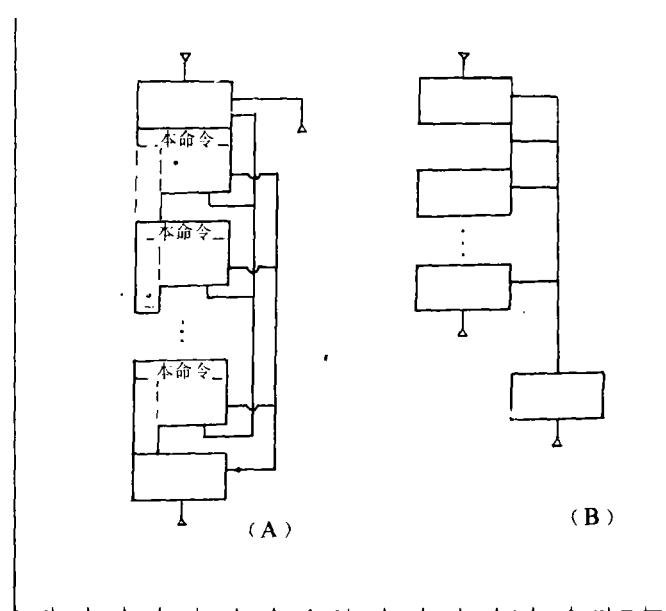


图2. 两种标准结构

3. 系统设计基于顺编文体，所有数据均以顺编文体的形式存放在计算机中。在数据预处理期间，数据文件属C类字符文件，用户可以在终端上显示、修改数据。当数据校核无误后，启动 ENTRY 程序模块，将C类字符文件转换成N类数字文件，供系统处理。

4. 该系统可以处理三种类型的数据：D型（离散型）、J型（非规则网型）和R型（规则网型）。且三种数据类型可以相互转换（见图3）。

5. 系统程序设计是基于数据结构的。文件的数据结构如图4所示。对每一程序模块程序设计的先后次序是参数校核、功能实现、出错处理，并配有必要的注释，使程序结构更加清晰。

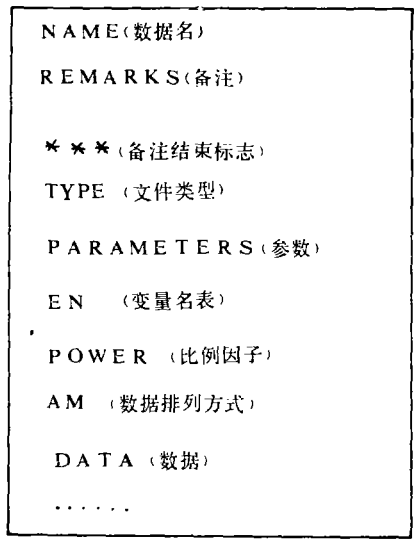


图4. 文件的数据结构

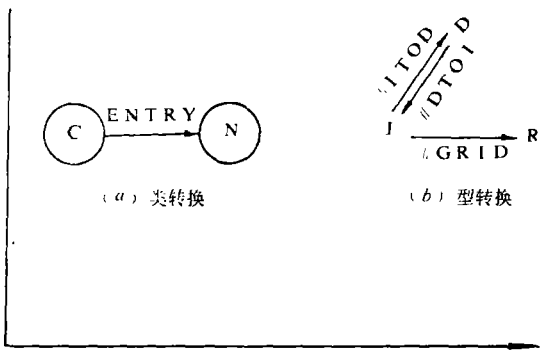


图3. 文件的类型转换

二、系统功能

系统中各命令块如下表 现将系统两部分的功能分述如下:

命令块名表

	MG 1	MG 2	TSA
命令	# MONRGA (一元回归分析) # MULRGA (多元回归分析) # GRARGA (逐步回归分析) # DCMAA2 (两组判别分析) # DCMAAB (多组判别分析B) # DCMAAF (多组判别分析F) # DCMAAG (逐步判别分析) # QMDETA (Q型因子分析) # RMDETA (R型因子分析)	# QMDCSA (Q型聚类分析) # RMDCSA (R型聚类分析) # KMDCSA (K均值聚类分析) # OTMDVS (最优分割) # CSPDA1 (对应分析1) # CSPDA2 (对应分析2) # TPCCRA (典型相关分析) # ULNIGA (非线性映象) # MPBMGA (马尔可夫概型分析)	# POLTSA (多项式趋势面分析) # KOPTSA (正交多项式趋势面分析) # HARTSA (调和趋势面分析) # TPCTSA (典型趋势面分析)
辅助命令	# LOAD (取文件) # SAVE (存文件) # PRINT (打印文件) # SELECT (选择变量) # DELECT (删除变量) # VALUE (按值选取样品)	# RELINE (调整样品顺序) # ITOD (I型变D型) # DTOI (D型变I型) # RENAME (更改文件名) # REMARKS (更换备注)	
辅助命令			# GRID (网格化) # AREA (取子区) # DIFF (求差值) # IBAP (打印线状等值线图) # BMAP (打印带状等值线图) # VMAP (打印平面数值图)

1. 数据预处理部分的主要功能

1 系统能对数据文件进行分解、合并、打印和按用户指定的取值范围检查数据, 交换样品顺序, 利用终端的CPT显示、修改数据。

2 经过校对的数据可以利用 ENTRY 程序转换成系统可以处理的U类数据。

2. 数据处理部分的主要功能

1 在数学地质方法上, 可提供给用户 22种不同的处理方法, 系统可以完成回归分析、判别分析、因子分析、聚类分析、对应分析、典型相关分析、非线性映象分析、马尔可夫概型分析、趋势面分析、剩余的 trend 面分析和典型趋势面分析。而且每一类分析中又有不同的算法, 例如判别分析中有两组判别分析、Bayes 准则的多组判别分析、Fisher 准则的多组判别分析和逐步判别分析等。

2 系统还具有存取文件、选变量、删变量、按值选取样品、更换文件名、添加备注、网格化、取子区、求差值等辅助命令, 使得数据处理灵活方便, 充实了系统功能。

3 各种多元统计分析方法的输出结果均配有图解。对于趋势面分析、因子分析等的结

果以文件形式保存起来的数据,用户可以根据不同情况以清单形式输出数据,或提供给其它模块使用,或以不同比例尺打印输出全区或部分子区的线状或带状等值线图或平面数值图。

4. 系统具有很强的档案处理能力。自动地为各种输出数据和图表表明它们的类型、参数、处理方法、处理日期、处理时间等,同时还给出被处理数据的履历。

5. 系统具有较强的错误检测能力。它可以准确地为用户判明由于误操作、命令或参数填写不当以及数据的特殊性质所形成的各种错误,及时中断作业的执行,并输出错误信息。

三、系统特点

1. 本系统一次最大可处理50万数据。在大数据量处理时,可以避免因分区逐一处理出现的重复处理和处理结果的拼接等问题。

2. 本系统数学地质方法较为齐全,而且具有较丰富的辅助命令,因而用户仅需准备一次数据,就可以进行各种方法的处理。

3. 系统对各种数学地质方法的输出结果均配有图解,如分类图、聚类图等,使计算结果更为清晰。丰富的输出命令使用户可以得到各种原始数据、中间结果或最后结果的数据清单、线状等值线图、带状等值线图或平面数值图。

4. 由于具有网格化命令,系统不但可以处理离散(D型)数据和规则网(R型)数据,而且可以处理非规则网(I型)数据。

5. 系统具有取子区的功能,可以对任意指定的子区范围作进一步处理。子区网格可以平行于原测区网格,也可以偏转一定的角度。

6. 趋势分析是耗费机时较多的方法,系统采用快速算法计算系数矩阵,同时产生多个变量的趋势分析结果,明显地减少了趋势分析的计算量。对非规则数据,采用“自动搜索选点法”进行网格化,取点合理,选点效率高,计算速度快。系统对扇入较高的低层次模块,均采用计算速度快、精度高的优化算法。

7. 系统具备很强的档案处理能力。因而对数据文件所经历的处理过程全都准确地记录下来,便于用户查核。

8. 系统中各程序模块均配置有自检能力,因而当操作出现误动作,或用户提供的命令或参数不合理时,系统拒绝执行,并打印出错误信息,避免给用户带来不必要的机时损失,而且便于纠错。

9. 该系统不仅适合于离散数据的数学地质处理,同样也适合于物化探数据的数学地质处理。

10. 本系统全部用 FORTRAN 语言编写,使用作业控制语言操作,系统的所有功能可以互相穿插使用,处理手段灵活,操作简单易学,使用十分方便。

该系统85年通过冶金部组织的鉴定,并正式投入运行。由于该系统采用了现代程序设计方法,因此系统结构清晰、可靠性高、可读性好,而且功能的扩充和删改,以及功能块的各

种算法的追加和变更均十分方便,易于维护。由于系统的应变能力,本系统不仅是一个良好的应用系统,而且也可以作为一个研究系统,利用它来开展一些数学地质新方法的研究工作。

主要参考文献

- [1] 於崇文等:《数学地质的方法与应用(地质与化探工作中的多元分析)》,冶金工业出版社,1980年
- [2] 王学仁:《地质数据的多变量统计分析》,科学出版社,1982年
- [3] 中国科学院计算中心概率统计组:《概率统计计算》,科学出版社,1979年
- [4] 南京大学数学系计算数学专业:《概率统计基础和概率统计方法》,科学出版社,1979年
- [5] K·安斯伦等:《数字计算机上的数学方法(统计分析)》,上海科学技术出版社,1981年
- [6] 刘贵德等:《FORTRAN 算法汇编》,国防工业出版社,1983年
- [7] Davis, J. C.: 《Statistics and Data Analysis in Geology》, Chap. 6, 1973.

THE GEOMATHEMATICAL DATA PROCESSING SYSTEM

Tang Guangguo

Huang Weigang

Lin Shoujiang

Wang Xiuhuei

(Tianjin Geological Academy, M.M.I.)

(Automation Institute, M.M.I.)

Abstract

The geomathematical data processing system MGDPS consists of three parts: data pre-processing, data processing and output. MGDPS is multi-functional and high-efficient in processing a great amount of data. It also has powerful managing and checking functions for documents.

MGDPS is modular structure, written in programming language FORTRAN. It is developed by using the method of modular programming based on various data structures, so it being easy to be read, maintained and augmented.