

海域油气化探数据处理技术研究

摘要

数据处理是化探找油方法的数据重要组成部分,其开发深度和研究水平直接影响着化探方法的持续发展。

本人应用研究生学习过程中获得的知识,在参与国家高技术研究发展计划(863计划)《船载地球化学快速测试系统研制与应用研究》课题中(注:课题编号:2001 AA 611020/02-02),依据数据处理方法的基本原理,结合海域油气化探实践资料和计算机处理系统新技术对陆地传统的油气化探数据处理方法进行改进,开发出提取指标油气信息的数据处理系统,包括以下五个子系统:

数据使用价值分析的预处理子系统——指标的组合计算、定性数据的量化、变量组合运算、数据标准化、离群点识别与处理;

优选指标组合的子系统——应用相关分析理论对指标进行分类,提取不同级次的找油直接指标、间接指标及辅助指标;

抑制干扰因素子系统——针对影响指标含量的海水(取样)深度、沉积物粒度、碳酸盐含量等因子,建立了定量校正干扰因素的方法;

确定异常子系统——对有监督和无监督两种测量形式,分别采取不同的判别方法确定异常,开创了与传统思维方式及作法截然不同的圈定异常技术;

异常定量评价子系统——首次提出单指标和多指标两级综合指数定量评价异常的方法,使化探异常含油气性优势次序的排列,从定性评价走向定量评价。

论文的主要创新点有以下三点:

1. 针对海域油气化探特点的随机性,同一点不同介质样品(大气、水体和沉积物)分析数据繁杂的特点,建立了数据预处理系统,为指标优化和提取可靠油气信息奠定了基础。

2. 在我国首次提出抑制与校正近代生物地球化学作用干扰的技术方法。

3. 开拓了定量确定异常和评价异常的数据处理方法。

论文通过典型探例说明数据处理系统的有效性,实践证明,该系统是当今海域油气化探比较系统和新颖的方法,具有功能齐全、方法丰富、重点突出、实用性强的特点。为海域油气化探的持续发展与提高提供了技术支撑。

关键词: 数据处理, 油气化探, 指标, 异常评价, 干扰因素, 参数

Study on the Data-Processing Technique in Marine Oil and Gas Geochemical Exploration

Abstract

Data processing is an important component part of oil and gas exploration technique, and its development depth and study level directly influence the sustainable development of oil and gas geochemical technique.

By means of using the knowledge acquired during his graduate school, and based on the basic principle of data processing method, the author has improved the conventional data-processing technique used in continental oil and gas geochemical exploration, and made it suitable for data processing in marine oil and gas geochemical exploration. The newly improved data-processing system consists of the following 5 sub-systems.

(1) Preliminary Processing System for Data-Value Analysis: This sub-system includes calculation of the indicator combination, quantitative analysis of the qualitative data, calculation of combined variables, standardization of the data, and identification and processing of the scattered values.

(2) System for Indicator Optimization: This sub-system can select various levels of direct, indirect and auxiliary indicators for oil and gas geochemical exploration through the classification analysis of all geochemical indicators.

(3) System for Influence-Factor Oppression: This sub-system provides the quantitative correction methods for the influence factors, such as sampling depth under sea, grain of sediment samples, and content of carbonates in the sediment samples, etc.

(4) System for Anomaly Determination: This sub-system utilizes different distinguishing methods to determine oil and gas anomalies, and creates the unique anomaly delineating technique that is totally different from the conventional one.

(5) System for Qualitative Anomaly Evaluation: This sub-system proposes the quantitative method to evaluate oil and gas anomalies by analyzing the integrated evaluation index of the single indicator and the multiple indicators, realizing the quantitative evaluation of the geochemical anomalies.

The paper has brought forth three new achievements in oil and gas geochemical exploration.

(1) It established preliminary data-processing system, which provides the base for extracting the oil and gas information and optimizing the geochemical indicators.

(2) It proposed for the first time in China the oppressing and correcting technique for the influence caused by the Recent biogeochemical reaction.

(3) It initiated the quantitative anomaly determination and evaluation technique.

The paper illustrated the effectiveness of the data-processing system by the typical cases. As a relatively systematic and new data-processing system used in the present marine oil and gas geochemical exploration, this system provides the technique support for the improvement and the sustainable development of marine oil and gas geochemical exploration.

Keywords: Data processing, Oil and gas geochemical exploration, Indicator, Anomaly evaluation, Influence factor, Parameter

独创性声明

本人声明所提交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得合肥工业大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：刘伟

签字日期：2004年5月28日

学位论文版权使用授权书

本学位论文作者完全了解合肥工业大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权合肥工业大学可以将学位论文的全部或部分内 容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：刘伟

导师签名：王

签字日期：2004年5月28日

签字日期：04年5月28日

学位论文作者毕业后去向：

工作单位：北京邮电大学科技发展部

电话：1351109490

通讯地址：北京市海淀区清华路16号学研楼1203室

邮编：100083

致 谢

我是从生产实践中认识到自己知识的不足和计算机技术的重要性，所以在进入硕士研究生学习阶段，我非常珍惜这次学习机会，几年的学习生活，使我更深刻的体会到理论与实践结合，以及多学科结合的必要性和好处，也认识到“学无止境”的真正含意。今后还要付出更大的努力不断地学习和实践，以报答母校和诸位老师的教导与付出的辛勤劳动。

在学习中，我的导师胡学钢教授自始至终给予了精心的指导。胡教授不仅传授给我知识，而且他的严谨作风，实事求是的态度，以及敬业精神深深的影响和教育着我，这将成为我以后学习和工作的宝贵财富。这里我首先向我尊敬的导师——胡学钢教授表示衷心的感谢！

我还要特别感谢研究生院的各位授课老师，是他们给了我知识和技术，对于他们付出的心血，我将永远铭记在心中。

研究生学习的艰辛和欢乐，使我难忘，也为结识许多同窗好友而高兴。在此，我要感谢研究生院的同学以及周围（包括我的家人）所有曾关心过我的人。

作者：刘伟
2004年5月20日

第一章 海域油气化探的基本特点

油气化探是在石油地质学和地球化学基础上发展起来的一门新兴的综合学科。在系统的分析测试自然界中含矿物质过程中，发现与油气有关的化探异常，从而评价区域的含油气远景，寻找油气藏。

上世纪三十年代初，德国的 G.劳伯梅耶（Lanbmeyer）和前苏联的 B.A.索科洛夫（CokoloB）等首次将近地表的烃类气体与地下油气藏联系起来，创建了现在所称的油气化探方法。该方法经过不断完善和发展成为陆地石油、天然气勘探的一种方法，由于它是检测油气的物质成分被视为新的找油气技术与方法。由于化探找油的成功率远远高于地质、物探等方法，而且成本低、速度快，受到勘察家的推崇与欢迎。

石油和天然气是重要的基础工业，在国民经济中占有重要地位。为了争夺油气资源，世界各国（尤其是经济发达的国家）在海域开展油气地质调查，其中化探是普遍应用的方法。

美国、英国、法国、澳大利亚、日本、加拿大、德国、前苏联等在海域油气化探中投入了较大的工作量。综观之有以下几方面的特点：

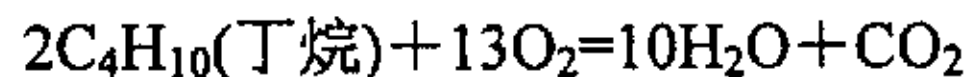
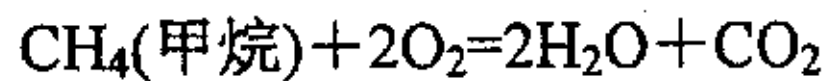
1.1 应用的化探方法、指标多

海域油气化探的初期，应用了烃类气体法，用抓斗采样器采取近海底表层沉积物，在船上晾干后送至陆上化验室，脱出气体分析测试甲烷及其同系物。由于采样过程中，沉积物经过扰动和晾晒，游离烃均已释放，所以测出的烃气含量普遍较低，同时，由于海底表层沉积物中含有大量生物遗体，给测试结果带来外界干扰，影响成果的解释与评价。经过试验研究，逐步改进了取样方法，即用重力取样器，在基本上保持沉积物结构的情况下，通过加压和自身重量将取芯管插下海底，取深度为 1—1.5 米的沉积物（目前多数国家将采样深度定为 2 米）。取出样品后，一分为二，一部分样品立即装入特制的密封管内保存，运回陆地分析测试游离烃气中的甲烷及其同系物；另一部分样品晾干、粉碎后，包装运回陆地分析测试吸附烃类气体中甲烷及其同系物。上述采样的新方法避免了一部分生物干扰和游离烃气的逸散。后来，国外油气化探工作者又对样品采取密封加热的办法，分解出吸留烃，进而完善了从沉积物中全面获得不同赋存状态烃类气体，使测试结果更加逼近客观实际，提高了海域油气化探的应用效果。

前述的游离烃、吸附烃和热释烃都是研究油气组成的烷烃部分，而对油气芳烃组分的研究迟后了十余年。首先由英国的化探技术人员开发了用荧光光谱仪器，测试与分离了海域沉积物中的芳烃组成。提出了用 265nm 激发，在 300

—350nm 区段是轻芳烃产生的荧光；350—400nm 区段是中等分子量芳烃形成的荧光；400nm 以上区段是重分子量芳烃造成的荧光。因此，他们提出了用 320nm、365nm、405nm 等波段作为找油气的指标。L.Horvitz 和 J.Brooks 又进一步证实 365nm/320nm 指标是区分石油与天然气的有效指标。

美国油气化探工作者从上世纪四十年代，研究沉积物蚀变碳酸盐的形成，他们认为油气垂向微运移经氧化是碳酸盐积累的主要因素，其理论依据是：



生成的 CO_2 同沉积物中钙镁盐（或离子）化合生成一种特殊的 CaCO_3 ——蚀变碳酸盐。它具有特殊的物理化学性质，在 580°C 时，又可分解成 Ca 和 CO_2 ，测得 CO_2 含量的高值，是找油气的重要指标。

前苏联油气化探工作者提出了沉积物中微生物、金属微量元素和大气中放射性元素的找油气指标组合。并在一些主要含油气区域进行研究。

德国的地质—地球化学家认为沉积物中烃类的碳同位素是判断气体成因的唯一标志，在油气化探中具有重要的指示意义。

日本的科技工作者对水中溶解芳烃组分的找油气意义做了大量、深入的研究，指出水中苯、酚是找油气的直接指标。

此外，欧美国家的地球化学工作人员还探索海水中溶解烃的找油效果，并且创造性的发明了随船连测水中烃的新技术，不仅提高了工作效率，而且可以同地球物理方法（主要找油气的技术方法）同步、同点进行测量，有较高的应用价值。近年来，随着航天—遥感技术的飞跃发展，在海域油气化探中开始引进空中激光光谱和机载红外光谱等方法。

我国油气化探是解放后从前苏联引进的，上世纪七十年代初开始在海上进行方法试验。由于我国海域辽阔，地质条件优越，蕴藏着丰富的油气资源，在 30 余年的油气化探历史中，吸引了国内、外十余家单位参加了方法实验和应用研究，包括不同精度的工程项目、国家科技攻关研究项目、自然科学基金项目、环境保护项目、综合地质、水文调查项目、油气勘探及国家高新技术研究发展计划（863 计划）项目等，在渤海、黄海、东海、南海四大海区进行了不同程度的地球化学调查。各个国家和单位应用了自己开发的新技术方法和专利技术，争夺我国海域油气化探的市场，所应用的化探方法与指标概括起来有以下四大类：

1.1.1 沉积物中的指标。烷烃部分有：游离烃，主要是甲烷和重烃；吸附烃（也称酸解烃），分析测试 C_1 、 C_2 、 C_3 、 $i\text{C}_4$ 、 $n\text{C}_4$ 、 $i\text{C}_5$ 、 $n\text{C}_5$ 及 C_{2-5} 的烯烃；吸留烃（也称热释烃），主要是甲烷，其次是乙烷、丙烷、丁烷，但这些组分在一

个样品中全部测出者不多。芳烃部分有：荧光光谱（320nm、360nm、405nm、450nm 及有关比值）；紫外吸收光谱（212nm—340nm 六个波段及比值）；金属微量元素（Fe、Mn、Co、Ni、Cu、Pb、Zn 等）；微生物（甲烷菌、乙烷菌、丙烷菌）；蚀变碳酸盐；热释汞；磁化率；三价铁与二价铁；PH—Eh 值；甲烷—丙烷的稳定碳同位素；吸附烃主要有：正戊烷、正己烷、正庚烷、3-甲基庚烷、正壬烷、正癸烷、正十一烷、正十二—十六烷等 48 个组分；放射性一氦。

1.1.2 海水中的指标。在深海区分别取海水表层、中层和底层水样，分析的指标有：溶解烃——甲烷与重烃；苯、酚系列——苯、甲苯、乙苯、邻二甲苯、对二甲苯、间二甲苯，酚、间甲酚、邻甲酚、对甲酚、二甲酚等；三维荧光与同步荧光的主要单体化合是：萘、蒽、菲、茈、芴等；二氧化碳；斯尼法（Sinffer）连续测定与水中总烃及 PAK 技术测试水中痕量芳烃法。

1.1.3 大气指标。取海水表面上的气体，测试的指标有：大气汞、CO₂、H₂S、甲烷。

1.1.4 遥感指标。根据卫星图片解释的油膜、热膜及激光——荧光的特征波长对。

综上所述，在我国海域所用化探方法与指标多达 100 多个，在国际上是罕见的，为海域油气化探研究提供了丰富的资料。

1.2 在同一观测点上取得不同介质的样品

烃类垂向微运移是油气化探的理论依据，这种运移又为运移过程中所处的地质条件、物理——化探环境所制约（影响），因此，油气化探需要查明地球化学空间场和时间场两个方面的变化规律。换句话说，既要在平面上区分背景区和异常区的分布范围和异常类型，又要查明指标在纵向上的聚散特点，为异常的解释提供地球化学依据，正确认识 and 判断指标的晕、源关系。这就要求在野外调查和采样时，不仅要在平面上均匀布点，而且在垂向有点控制，从而构成立体空间的地球化学测量，准确地衬托出与深部油气有关的异常。在陆地油气化探中，由于采样条件与介质的限制，只能侧重于化探指标在平面上分布特征（地球化学空间场）的研究。而对化探指标在垂向上变化规律（地球化学时间场）的研究比较薄弱，因此，化探成果的可信度受到影响。究其原因，主要是在同一采样点难易或不可能采到不同介质的样品，尤其是沉积物和地下水样品做不到同点采取。为了弥补这一缺陷，化探技术人员常在地质解释时，采取异常迭合的办法，即分别作沉积物 and 水的异常图，而后两张图迭加，其结果必然导致异常增多，异常面积加大，降低了化探成果的可靠性。在地质解释困难时，打少量的控制水井，这样不仅增加了工作量和经费，影响了进度，而且不可能在每一个观测点上都打井，化探质量必然受到影响。

海域油气化探的一个突出优点是在同一观测点能取得不同介质的样品，如海

水表面大气样,不同深度(一般分表层、中间及底层)的海水样、海底表面沉积物样、浅层(1-3米)柱状样、海底生物样等,构成了立体空间的地球化学调查,在同一采样点内获得时间场和空间场的地球化学资料,为追踪油气垂向微运移的地球化学形迹奠定了基础。例如1973—1974年在北黄海进行化探调查时,共布观测点126个。每个点在垂向上采到3米深和0.5米深沉积物样各一个,底层水(深度30—54米)、中层水(深度20米左右)和表层水样各一个,并在每个观测点附近采集了大气样。1992年在南海的南沙地区在同一观测点上采到深层(深10米)、浅层(深3米)和近海底表层沉积物样、不同深度的三层水样及大气汞样,并有航空激光荧光、红外及航天遥感资料,为本区油气勘探提供了较详细的地球化学依据。事实证明海域油气化探之所以能全面准确的揭示深部油气信息,成功率比较高与多层次的取得样品,多方位的获得信息有一定的关系。

1.3 勘探程度低,以随机布点为主

油气化探是通过布置的采样点了解区域地球化学特征,每一个化探点代表着一定面积或控制范围内指标含量的平均值。按照油气化探规范要求,以测点的密度为准则,油气化探分为概查、普查、详查、精查四个不同级次阶段。目前陆地的大部分沉积盆地(尤其是我国中、东部)完成了化探普查(1点/KM²),进入油气化探详查阶段(4点/KM²),局部做了每平方公里16个点的化探精查。油气化探是按照先概查、后普查、进一步作详查、最后是精查的级次程序连续进行调查的,即工作的范围由区域到局部,调查的比例尺由小到大,异常的可信度由低到高等。其任务与评价目标是由勘探阶段决定的。在油气勘查的初期阶段,投资具有较大的风险性,利用化探简便、快速、成本低的优势,通过中—小比例尺低精度的化探概查和普查,预测盆地含油气的可能性,指出找油气的有利区带。从战略上为油气勘查提供依据,此时的化探异常反映了油气分布的区域性变化特征。随着油气化探调查精度的不断提高和深入,大比例尺高精度的详查和精查可评价圈闭的含油气性或圈定油气藏的有效范围,化探异常指示了油气聚集的具体特征。然而我国海域油气化探测点布置的随意性很大,许多测区布点密度达不到化探最基本的要求,给资料研究和油气远景评价带来很大的困难。我国海域油气化探虽然起步较早,但由于认识上的误区和资金、设备的不足,进展缓慢,勘探程度很低,处于小比例尺的化探概查阶段,各海区只有很稀的化探点或几条剖面线控制,而且取样点的均匀程度极差,以随机布点为主,在平面上呈现以点带面或以线带区的特点。

近几年来,世界各国为了控制海域油气资源,海域地盘争夺异常激烈(南海尤为突出,黄海和东海次之),我国加快了海洋油气资源调查和勘探步伐,为油气化探带来了生机。我国海域油气化探调查由于跨越年代与间隔时间长,调查单位又各自为战,缺乏可遵循的统一勘探规范和分析标准,因此制约了资料的应用。加之我

国海域辽阔,面积大,地质条件复杂,化探样品的岩性与深度(包括海水深度)变化大,干扰因素甚多,导致各测点化探数据变异大,含量高低相差悬殊,给资料的统一应用和研究带来一定的困难。

1.4 干扰因素集中且强烈

化探检测到的深部油气信息,一般比较微弱,而且研究的课题又是十分复杂多变,化探指标极易受到工业、农业、大气、人文等诸多污染源及地形、地貌、植被、气候及岩性等因素的影响,引起化探指标含量的起伏变化,使化探成果具有多解性。从某种意义上讲,油气化探的地质效果与异常的可靠性,取决于对干扰因素的抑制程度。

与陆地相比,海洋油气化探的干扰因素较为“单一”,其主要影响因素是现代生物地球化学作用和沉积物岩性(包括:颜色、粒度、碳酸盐含量),其中前者影响最为突出。众所周知海洋中生物群极为发育,在低温、缺氧及低硫酸盐环境下,厌氧细菌对其改造、分解形成与油气成份相似的烃类。从分析测试技术上很难区分油气与非油气形成的烃类气体。如何消除或约束干扰,关系到海洋油气化探持续发展的重要课题,也是必须解决的难题。所以在石油天然气地球化学勘探技术规范中,明确规定了干扰因素调查与研究的具体内容。

1.5 本文的研究目标和主要内容

油气地球化学勘查(简称油气化探)是系统地从沉积物、岩石、气体、水体及动植物等介质中检测烃类及其伴生物和蚀变产物,预测盆地的含油气远景,指出油气富集区带,评价区块(构造)的含油气性提供探井试油层段的一种方法。是寻找油气藏的直接方法,但受诸多因素的干扰和影响,所以正演都有唯一的解,而反演具有多解性,这就要求技术人员根据工区具体的地质条件,借助于计算机技术,应用多元统计与处理方法,约制干扰,选取指标,提取油气信息,从而达到寻找深部油气藏的目的。

海域油气化探数据处理技术研究的目的在于从分析测试的大量数据中,剔除各类误差和干扰因子,为油气化探最优提取化探指标,建立有效指标组合、合理的背景值和异常值提供科学依据及软件系统。

作者在参与国家高技术研究发展计划(863计划)《船载地球化学快速测试系统研制与应用研究》课题中(注:课题编号:2001 AA 611020/02-02),结合海域油气化探实践资料和计算机处理系统新技术对陆地传统的油气化探数据处理方法进行改进,实现了多指标、多约束条件下的最优提取指标的新方法,建立了海域油气化探数据处理子系统,该项研究具有较高的应用价值和良好的用户界面。

提出了从分析数据、地质观测参数等原始资料的可利用程度和合成开始,按照化探工作流程及评价程序,分别建立了具有共同理论基础的不同的处理方

法，即：优选指标组合、抑制干扰因素，确定异常及定量评价异常等子系统，它们既是独立的，又是环环相扣，缺一不可的数据处理方法。

提出了数据使用价值分析的多种方法，将前人认为不能参加运算的定性参数（如沉积颜色差异、粒度分级等）赋予定量化数值；将不同含量单位、量纲及数值大小的定量数据标准化后，均可直接参加运算，增强了用于定量研究的可能性，提高了数据的应用范围。

开拓了定量校正干扰因素的处理方法。在这之前只有个别化探工作者对干扰进行过定性研究。例如岩性对指标含量的影响，一般不予考虑，常用主岩性代替全部岩性进行研究；也有的学者采用对岩性进行分类的定性校正方法。而作者提出的定量校正方法，是将全部样品的岩性校正到均一粒度或相同岩性的背景上，求取异常，使化探成果更加科学、合理。

建立了量化确定与评价异常方法。提出了用判别函数值等量化异常；用综合评价指数划分异常级别，进行分级评价。这些技术在油气化探中首次提出，是带有方向性研究课题，并且在实践中展现了良好的应用效果。

第二章 海域油气化探数据处理现状及 建立数据处理方法的意义与原则

2.1 数据处理方法的引进过程

借助于计算机应用数理统计的方法来研究和解决地球化学问题是在长期实践中逐渐发展起来的一门新兴科学。早在 19 世纪初,莱伊尔 (Lyell,1840) 应用统计方法研究地层中的生物种属及其对比。20 世纪 30 年代地学界引进统计检验和方差分析;40 年代回归分析和简单相关分析用来研究化学元素含量之间的相互关系。50 年代电子计算机的发展也使地学界的科技进步进入一个新的阶段。60 年代,一些多变量统计方法解决了地质学中长期以来不能解决的问题——矿床和矿物含量的变化趋势预测等。作为地质学的新领域——地球化学是以大量的分析数据为研究目标,它的兴起与成长更加依赖于数理统计技术的发展。

在油气化探领域中化学分析测试首先应用数理统计方法,如为了判断分析结果是否准确可靠,通常用方差和标准偏差来衡量,为了消除系统误差和偶然误差,除进行重复分析外,也借助于数理统计方法。在油气化探资料整理研究过程中,化探技术人员于上世纪 60 年代开始重视数理统计方法的应用,从最初的对分析测试数据进行简单对比,用其相对的高值确定异常,由于高含量存在许多不确定因素,异常与油气关系不明确;到 70 年代中期应用统计方法确定背景值,将大于背景值者定为异常,提高了化探异常的可信度。后来又引入迭代统计、稳健主分量分析等方法研究地球化学场特征,在背景的基础上加上标准差作为异常下限划分异常,使异常更加逼近油气藏的空间范围,有效地提高了化探的预测成功率。

近二十年来,油气化探分析测试方法基本稳定,一直沿用着已成熟的烃类气体和芳烃等方法,但化探成效及其成功率提高很快,被勘查家视为油气资源调查的重要方法之一,之所以如此,同计算机技术的飞跃发展有直接关系,数据处理技术与方法,已是当前油气化探科研与生产不可缺少的手段。我国油气化探工作者在上世纪七十年代中期开始引进多元统计方法,使化探异常的确定,从依靠指标的绝对含量发展为异常下限的概念。八十年代中后期初步建立起陆地化探数据处理方法,提出了包括背景分析、异常确定、异常识别和解释、质量检查等数据统计方法。九十年代初,数据处理技术与方法被列入了石油天然气地球化学勘查技术规范。从此海域油气化探也就沿用着陆地油气化探的数据处理程序与方法。

2.2 海、陆化探数据处理的背景

已有的陆地化探数据处理方法基本出发点是运用多元统计分析和地质统计

方法,建立区域均一背景值加标准偏差作为异常下限,提取高浓度或局部相对高浓度的异常。其主要方法有:

2.2.1 多母体(含多个异常总体时)混合总体筛分的方法用以确定背景与异常母体并计算各自均值、标偏,从而确定背景与异常下限。一般采用背景母体右拐点,即 $\bar{X} + 1.96S_0$ 为异常下限。

2.2.2 实际工作中,通过作直方图的方法来确定异常母体与背景母体,并依据均值加N倍标准偏差作异常下限,N值一般取0.5-2.0,有时直接从直方图上读取。这在背景母体与异常母体分离很好时尤其适用。

2.2.3 由于以上N值的选取带有经验性,常利用等直线图中异常形态规则及已知区吻合程度,动态确定异常下限并反推N值,从而较好地确定异常下限。

2.2.4 采用迭代切尾法是在求取背景母体的基础上,大致确定异常,但当背景母体不服从正态分布时,有时样本点迭代去了一半仍不服从正态(用偏度、峰度检验),此时应改用其它方法为宜。

2.2.5 也有人利用工区内异常几率10%-20%为准则并参照异常形态大致确定异常。但是在完全背景区(即无油气区)作油气化探时,强制性确定异常(10%-20%)将导致异常解释和评价的误解和困惑。在油气藏大比例尺测量时,异常母体样本数可能占有百分比高于20%。

2.2.6 迭代切尾、均值加标准偏差法,把背景当作一个水平面,然而许多工区因地质背景单元不同,具有不同的背景母体,常采用分区作图拼结法确定异常。分区边界异常拼结有许多处理方法。

2.2.7 当研究区背景为一变化的规则曲面时,常用趋势面或泛克立格法来确定背景与异常。实际应用中,考虑与已知区和经验成果,采用2-4阶次为宜。也有采用剩余值总和与经验系数比值作为异常下限并取得了较好效果。

2.2.8 有些地区采用以上方法均不理想,其化探数据离散度很大(变异系数大于3),数据结构不稳定,为了突出低背景中的异常和分离高浓度区的背景,常用窗口滑动平均或框架滤波的办法提取异常。

2.2.9 依据油气化探异常主要为“边高中低”环状异常模式,用数据空间上的变化梯度来圈定异常的梯度分析法被广泛采用。

2.2.10 依据分形理论、变差函数图(基台值)作为划分异常的准则具有重要意义。

上述系统适合于指标含量高低差别不大,正态程度较高,且测网规则或均匀布点条件下的专门面积化探调查。

但是,海域化探工作,都是随航与其它调查项目同时进行的,调查范围一般很大,测点稀少,布点的随机性很强,多以剖面为主,均匀程度差,一般属于查明区域含油气性的化探概查。总之,海域油气化探从布点、调查精度、目地任务及地质-地球化学背景等均有别于陆地。当前海域油气化探数据处理技术,是沿用

陆地油气化探与金属化探的传统方法，由于缺乏有效的针对性，未能最大限度地抑制干扰，突出化探异常，致使化探应用效果在很多地区不理想，制约了海域油气化探的发展。所以结合海域油气化探特点，建立多参数、多约束条件下的最优提取指标技术，有一定的理论意义与应用价值。

2.3 数据处理的重要性

油气化探是从大量的分析测试数据和诸多地质参数中,通过技术人员的归纳分析,提取油气信息,获得规律性的总结,达到寻找油气藏的目的。显而易见,数据处理在油气化探中起着十分重要的作用。

数据处理是以数学和计算机作为基本工具进行运作的,有利于将地球化学的问题用数学形式表达出来,将定性资料变成定量的概念。油气化探中有较多数据与工作,需要采取多种多样的方案试验比较,或者是多次反复的试验,从中挑选出最佳方案或者计算结果。更有意义的是,利用数理统计技术,可以进行原来技术人员无法进行的计算、模拟及寻找化探数据和参数之间的内在联系,追踪数据“规律”与油气藏的关系,为地球化学工作者的综合研究与地质解释提供新的途径和方法。换句话讲,数据处理技术引入油气化探,将会使该学科及找油气方法发生一个深刻的变革。

我国是世界上仅次于美国的能源消耗大国,油气产量远远不能满足国民经济发展的需要,油气供需矛盾日益突出,面临大量进口石油的严峻形势,发展新的找油气技术,开拓油气新领域,增储上产是每个石油工作者不可推卸的责任。化探是寻找石油与天然气的直接方法,速度快、成本低、效果好。油气生产的日益需要,促使该学科发展和相关技术的日臻完善,建立相应的海域化探数据处理系统、开发实用的数据处理方法(包括软件),具有一定的意义。作者认为化探是以大量分析数据为依据,研究油气聚散规律,传统的经验性和直观性的资料整理方法,已不适应和满足工组需要。我们开发了以计算机为基本工具,以数学作为科学语言的新的数据处理技术与方法,使化探工作者从费时繁杂的手工计算中解脱出来,将更多的精力和时间用在综合地质解释和研究化探异常分布规律上,也为定量校正干扰因素,突出异常及模拟试验等奠定了基础。

2.4 建立数据处理方法的基本原则

2.4.1 数据处理方法必须建立在可靠的数学理论上,将地球化学问题用数学形式合理地表达出来,通过算法语言进行迅速计算;

2.4.2 由于海洋油气化探分析测试的数据在成因上具有多源性,在统计上具有随机性不确定性的特点,同时还有区域化变量的结构性差异,因此,要求提取指标组合或确定异常的数据处理方法需多样化,便于验证和映照(对比),避

免单一方法的片面性和失真，从而找出研究体内地球化学数量特征与内在统计规律。

2.4.3 所选用的数据处理方法，不仅有提取指标组合的定性功能，而且还要有“量”的概念，便于定量校正和定量评价；

2.4.4 数据处理系统的算法及其过程，力求严谨、易行、稳定，计算功能齐全，操作方便，计算结果具有精度高、准确性强；

2.4.5 数据处理系统最终要有利于油气矿床的预测、评价；

适应于海域油气化探数据处理技术与方法设计的具体功能如图 1

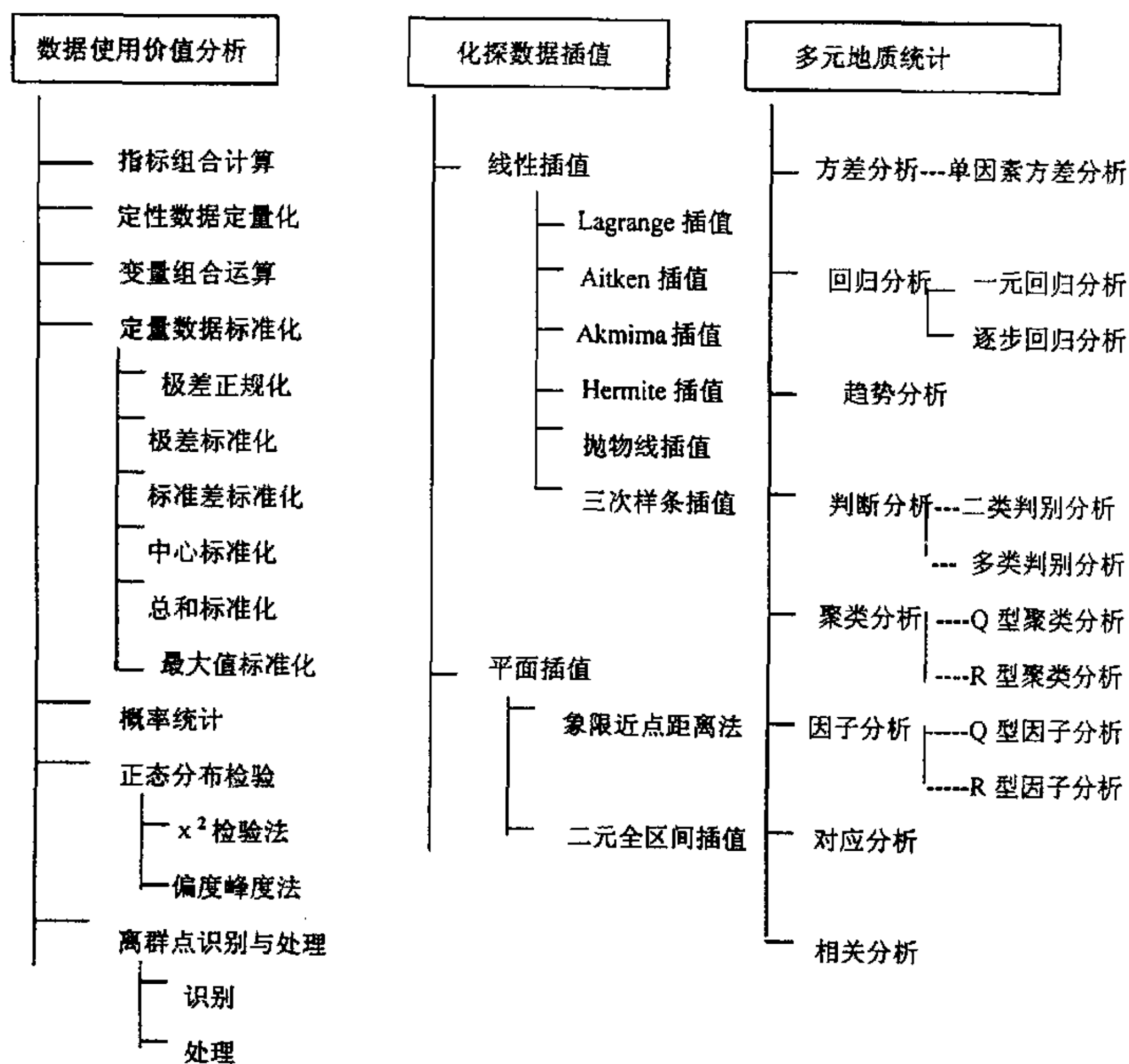


图 1 油气化探数据处理方法设计

2.5 小结

本章主要简述了海域油气化探数据处理的现状及处理方法的引进过程，针对海域油气化探的特殊性，提出了建立海域油气化探数据处理的意義和原则，避免单一方法的片面性和失真，从而找出研究体内地球化学数量特征与

内在统计规律，将地球化学问题用数学形式合理地表达出来，便于定量校正和定量评价，计算结果具有精度高、准确性强；有利于油气矿床的预测、评价。

第三章 化探的数据处理

3.1 数据使用价值分析

油气化探数据资料，一般有两部分组成：一是实验室中各指标的分析测试数据；二是地质—地球化学调查的观测数据。它们不仅数据量大，而且是性质不同的数据对。前者数据中大部分与油气藏有一定的关系（包括成因关系、演变关系、环境关系等），但也混渗有与油气藏无关的干扰因素（如：人文污染、近代生化作用、生物排泄、分析误差等）；而后者数据中大部分是影响或控制前者的环境因子，如样品的质量—沉积物的颜色、粒度、化学成分；采样条件—深度、年份与季节等。这两类数据从性质上讲，分析化验资料是定量的，而地质观测资料是定性的。油气化探技术人员需要或想办法创造统一的尺度和条件，将这两类表面上互不相干、实为因果关系的资料融汇在一起，提取油气信息，达到预测油气藏的目的。要实现或达到这个目的，在数据处理之前，一方面要鉴别数据的使用价值，了解数据的贡献或可信度，对数据去伪存真显得尤为重要。另一方面由于化探指标的含单位不一致，地质观测数据的量纲不同，将原始数据（包括分析化验与地质观测的）直接用于计算往往是不合适的，因此在进行正式计算之前需要对观测数据进行预处理，形成供进一步计算使用的数据。化探数据预处理一般包括变量的组合及参数计算、定量数据的标准化、定性数据的定量化变换、原始数据均匀化和增补、离散数据的判定及处理等步骤。在本次海域的应用研究中的流程如图 2，其中原始数据文件应该为一个行列式排列的数据体。

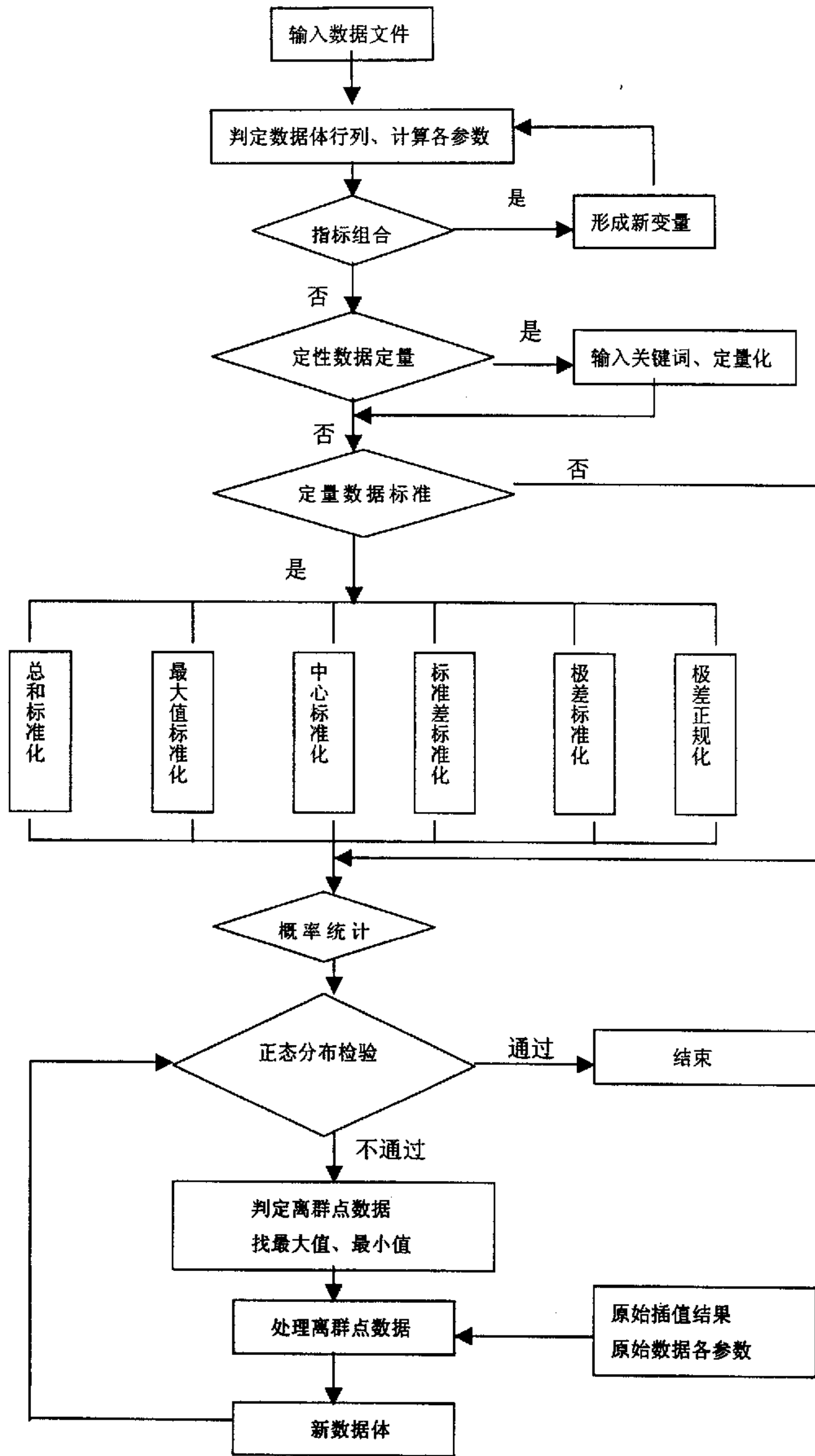


图 2 数据使用价值分析方法与流程

3.1.1 指标的组合计算

根据成因不同，将众多的指标以不同的数学方法组合在一起而构成新的变量。以提高信息的综合性及充分使用。这些算法包括累加指数、比例指数、对数指数、变量乘除等等。计算每个变量的和、均值、均方差、平均差、标准差等参数。例如，在本次的应用中，由 $C_1 \sim C_5$ 计算轻烃总烃，计算甲烷干燥系数 $C_1 / (C_1 + C_2 + C_3 + C_4)$ ；湿度比 $(C_2 + C_3 + C_4) / (C_1 + C_2 + C_3 + C_4) \times 100$ ；平衡比 $(C_1 + C_2) / (C_3 + C_4)$ ；特征比 $(C_4) / (C_3)$ 等等。

3.1.2 定性数据定量化

为了使定性数据能够用于定量研究，直接参加运算，必须将定性数据的符号或代码赋予定量数值。在本次设计中定量化分为二态定量和有序型多态定量两种。

二态即“非此即彼”。例如，可根据对气态烃的吸附性把海底砂质沉积物赋值 0，泥质沉积物赋值 1；有 H_2S 气味道样品赋值 1，没有此气味道赋值 0。此变换后的 0、1 可与极差正规化变换后的定量数据混合使用。极差正规化是将数据均压缩到 $[0,1]$ 内，因此，经 0、1 变换后的定性数据可看作是定量数据的两种极端状态。

多态定量，可以把描述性的关键词根据其地质含义分别赋值。例如，将海底沉积物分为中细粒砂、粉砂、粉砂质粘土、粘土，定量为 0，1，2，3；也可根据需要非等差式赋值为 0，1，3，5 等。这样在程序中可把不同的数据赋不同的颜色或图标，为以后的绘图设计提供方便。

3.1.3 原始数据的均匀化和增补

由于和陆地化探不一样，海洋化探取样往往受局限性大，不可能规则的网格取样，所以对原始数据均匀化，即插值网格化，对原始数据空白区进行增补就十分的重要。同时在进行插值时，往往需要针对不同的点位密度在一个工区内不同局部区域进行不同密度的插值。具体设计方法及应用见（3.4.4）节。

3.1.4 定量数据标准化

定量数据标准化是将具有不同单位、量纲及数值大小、变化范围的各变量的原始观测值，在进行计算之前变换到某种规范尺度之下。避免对原始数据直接使用，可能突出观测值较大地质变量的作用、降低观测值较小地质变量的作用的现象。定量数据的标准化一般在数据矩阵的基础上进行。本次主要采用了以下六种方法，可以从中任选一种。

A.总和标准化：将变量的每个观测值变换为它与该项变量所有观测值总和的比值。在变换后，数据矩阵的元素在[0,1]之间，且每个变量的所有观测值之和等于1。变换公式为：

$$x'_{ij} = x_{ij} / s_j \quad (i=1,2,\dots,n, j=1,2, \dots,m)$$

式中 x'_{ij} 为变换后的数据； x_{ij} 为变换前数据； s_j 为第 j 个变量观测值总和， n 为样品总数。

B.最大值标准化：将每个变量的观测值除以该变量所有观测值中的最大值。标准化后的观测值为1。具体变换公式为：

$$x'_{ij} = x_{ij} / \max_{1 \leq k \leq n} x_{kj} \quad (i=1,2,\dots,m)$$

式中 x'_{ij} 为变换后的数据； x_{ij} 为变换前数据； $\max_{1 \leq k \leq n} x_{kj}$ 为第 j 个变量观测值中的最大者。

C.中心标准化：将变量的每个观测值减去该变量所有观测值得平均值。变换后所有值得平均值为0，即总和为0。

D.标准差标准化：将每个观测值减去该变量所有观测值得平均值，再除以该变量观测值的标准差。每个变量变换后的平均值为0，标准差均为1。具体变换公式为：

$$x'_{ij} = (x_{ij} - \bar{x}_j) / s_j \quad (i=1,2,\dots,n; j=1,2,\dots,m)$$

式中 $\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj}$, $j=1,2,\dots,m$; $s_j = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}$, ($j=1,2,\dots,m$) 为第

j 个变量的标准差。该方法较常用，变换后数据又称规格化数据。

E.极差标准化：将变量的每个观测值减去该变量所有观测值得平均值，再除以该变量观测值的极差。对于每个变量观测值变换后，极差均为1。变换公式

为:

$$x'_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\left(\frac{\max_{1 \leq k \leq n} x_{kj} - \min_{1 \leq k \leq n} x_{kj}}{1} \right)} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

式中 x'_{ij} 为变换后数据; x_{ij} 为变换前的数据; $\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj}$, $j = 1, 2, \dots, m$;

当只有两个样品时, 由于极差标准化后每个点 x 、 y 之差的绝对值为 1, 所以变换后的点必然落在直线 $x-y=1$ 或 $x-y=-1$ 上。 x 、 y 的值取 0.5 和 -0.5 其中之一。

F.极差正规化: 为最常用的标准化方法。将变量的每个观测值减去该变量所有观测值得最小值, 再除以变量观测值得极差。变换后每个变量观测值在[0,1]内。若需将数据变换到任意闭区间[a,b], 变换公式为 (2):

$$x'_{ij} = \frac{x_{ij} - \min_{1 \leq k \leq n} x_{kj}}{\max_{1 \leq k \leq n} x_{kj} - \min_{1 \leq k \leq n} x_{kj}} \quad (1); \quad x'_{ij} = a + (b-a) \frac{x_{ij} - \min_{1 \leq k \leq n} x_{kj}}{\max_{1 \leq k \leq n} x_{kj} - \min_{1 \leq k \leq n} x_{kj}} \quad (2);$$

x'_{ij} 为变换后数据; x_{ij} 为变换前的数据; 其中 $(i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ 。

3.1.5 概率统计

概率统计是油气化探数据有效性分析的方法之一。对诸多指标分别进行概率统计的目的, 一是确认分析数据的分布状态, 了解各指标的可利用程度; 如果数据离散度很大甚至不服从正态分布或对数正态分布, 背景值就不能代表区域性变化特征, 直接影响异常的确定。二是消除分析误差和某些干扰因素的影响, 提高数据质量。

油气化探数据是多种成因的集合体, 因此, 概率分布的初始密度曲线分布比较混杂, 可能出现多种分布类型 (图 3)。当分析测试数据为非正态分布时, 可利用以下统计方法, 求算化探特征参数——平均值、标准偏差等。

| | | | | |
|--------|-----|---------------|---------------|--------------|
| 参 数 | | | | |
| | 均值 | 730.9 (224.6) | 0.94 (0.94) | 2.5 (2.35) |
| | 标准偏 | 143.6 (96.5) | 0.5 (0.39) | 0.94 (0.64) |
| | 偏度 | 0.53 P1=0.41 | -0.12 P1=0.41 | 2.4 P1=0.31 |
| | 峰度 | 1.45 P2=1.24 | 2.34 P2=1.4 | 9.6 P2=0.9 |
| 参 数 | | | | |
| | 均值 | 81 (79.9) | 3.7 (3.36) | 12.8 (11.75) |
| | 标准偏 | 17.7 (13.1) | 2.6 (2.7) | 9.9 (5.3) |
| | 偏度 | 0.44 P1=0.41 | 1.1 P1=0.41 | 5 |
| | 峰度 | -0.03 P2=1.24 | 1.32 P2=1.24 | 41.7 |
| 参 数 | | | | |
| | 均值 | 9.2 (9) | 3.7 (3.39) | 4.2 (2.99) |
| | 标准偏 | 1.8 | 2.7 (2.24) | 7.9 (2.08) |
| | 偏度 | 1.9 P1=0.41 | 1.6 P1=0.31 | 6.3 P1=0.41 |
| | 峰度 | 5.1 P2=1.24 | 3.7 P2=0.9 | 50.9 P2=1.24 |

图3 化探数据的概率分布特征

括号内数字为稳健估计的位置和尺度

A. 分布检验的迭代统计法

该法是在对概率分布是否服从正态分布进行检验的基础上，逐步剔除异常点，即排除异常母体、保留背景母体，然后对背景母体的数据集合进行统计。

概率分布检验的迭代统计检验法，可把非正态的数据逼近于正态分布，利用该方法得到的特征参数，比较接近于客观实际，有较强的代表性。

B. 稳健估计法

该方法主要优点是在处理非正态数据时，适应能力强，输出的结果可靠。它运用加权原理对化探数据进行处理。稳健估计所称的位置和尺度，相当于常规方法中的平均值和标准离差。

表 1 是不同统计方法的结果，从中看出，常规统计方法所求的均值，比迭代统计和稳健估计方法求出的均值偏高，这表明前者统计的数据总体中包含了异常点在内，属于混杂分布类型的数据集合，数据的成因必然是多种多样的，而不是成因较单纯的区域背景数据集合。当用迭代统计和稳健估计方法把数据逼近于正态分布后，即把引进异常的地质因素和自然因素排除之后，对剩下的区域性因素形成的背景母体数据进行统计，所求出的平均值较为接近，具有良好的代表性。当原始数据服从正态分布时，无论用什么方法进行统计，其结果是相似的。

表 1 不同方法求取的化探参数对比表

| 指标 | 分布特征 | 稳健统计 | | 迭代统计 | | 常规方法 | |
|------------------------------|------|--------|-------|--------|--------|--------|--------|
| | | 平均值 | 标准差 | 平均值 | 标准差 | 平均值 | 标准差 |
| C_1 ($\mu\text{l/kg}$) | 近似正态 | 716.60 | 96.50 | 700.00 | 100.00 | 730.90 | 143.60 |
| C_2^+ ($\mu\text{l/kg}$) | 近似正态 | 79.90 | 13.10 | 79.00 | 10.00 | 81.00 | 17.70 |
| C_1 / C_2^+ | 近似正态 | 9.00 | 1.30 | 8.60 | 1.00 | 9.20 | 1.80 |
| iC_1 / nC_4 | 近似正态 | 0.94 | 0.39 | 0.93 | 0.05 | 0.94 | 0.50 |
| He | 近似正态 | 3.36 | 2.70 | 3.20 | 1.90 | 3.56 | 2.80 |
| Hg | 相似伽马 | 3.35 | 2.24 | 3.20 | 2.70 | 3.71 | 2.70 |
| u254nm | 相似伽马 | 11.75 | 5.53 | 11.00 | 5.00 | 12.80 | 9.90 |
| u315nm | 相似伽马 | 2.99 | 2.08 | 2.40 | 1.40 | 4.20 | 7.90 |
| F280nm | 相似伽马 | 6.44 | 6.30 | 4.30 | 3.30 | 8.00 | 8.90 |
| F365nm | 相似伽马 | 40.00 | 17.60 | 37.00 | 14.00 | 53.00 | 160.00 |
| ΔC (%) | 相似伽马 | 2.35 | 0.64 | 2.70 | 0.50 | 2.50 | 0.94 |

从表 1 中看出，稳健统计和迭代统计的离散程度降低，表明在标准离差上较小，说明处理后的数据值域向平均值靠近，这是数据所反映的指标成因趋向单一化的表现。对油气化探而言，离散程度降低是排除异常母体数据之后，对背景母体数据进行统计的必然结果。根据概率统计学的原理，以背景母体数据（近似正态分布）为基础，统计的有关特征参数是合理与可靠的。亦就是说，利用稳健统计和迭代统计求取的异常下限值或异常点是比较恰当的。

3.1.6 正态分布检验

油气化探数据是否服从正态分布（包括处理后的数据）一般通过偏度和峰度检验法来确认，步骤如下：

A. 对观测数据 X_i ($i=1, 2, \dots, n$)，求出其偏度和峰度的矩估计 U_p, U_r 。

$$U_p = U_3/S_3$$

$$U_f = U_4/S_4$$

式中 $u_3 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3$ 为样本三阶中心矩, $u_4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4$ 为样本四阶

中心矩;

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ 为样本数据均值; } S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \text{ 为样本标准差。}$$

B. 对给定的显著检验性水平 α (0.1, 0.05, 0.01, 0.005, 0.001), 求出检验拒绝域临界值。

$$P_p = Z_{\alpha/2} \cdot \sqrt{S_p} \quad P_f = Z_{\alpha/2} \cdot \sqrt{S_f} + 3 - \frac{6}{n+1}$$

式中 $S_p = \frac{6(n-2)}{(n+1)(n+3)}$; $S_f = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$; $Z_{\alpha/2}$ 由正态分布表查得。

若 $|U_p| < P_p$ 且 $|U_f| < P_f$, 则认为服从正态分布。否则, 认为存在离散数据。

表 2 某海区原始数据非正态分布的参数表

| 指标 | 最大值 | 最小值 | 中位数 | 平均值 | 标偏 | 偏度 | 峰度 |
|-------------------|----------|--------|--------|--------|--------|--------|----------|
| C ₁ | 652.316 | 36.613 | 331.99 | 329.47 | 79.98 | -0.253 | 1.506 |
| C ₂ | 48.878 | 3.972 | 21.738 | 22.430 | 6.335 | 0.735 | 1.784 |
| C ₃ | 41.352 | 1.521 | 8.634 | 9.034 | 3.098 | 2.450 | 17.211 |
| iC ₄ | 37.125 | 0.490 | 2.619 | 4.285 | 4.744 | 3.378 | 13.174 |
| nC ₄ | 42.744 | 1.058 | 10.706 | 11.600 | 5.352 | 1.437 | 4.283 |
| iC ₅ | 29.364 | 0.307 | 1.729 | 1.978 | 1.481 | 9.767 | 151.151 |
| nC ₅ | 14.192 | 0.110 | 0.832 | 0.979 | 0.963 | 9.004 | 107.839 |
| C ₂₊ | 145.931 | 9.053 | 47.924 | 50.305 | 16.690 | 1.351 | 4.151 |
| CO ₂ | 90.252 | 0.012 | 5.721 | 7.519 | 6.235 | 2.501 | 13.441 |
| CaCO ₃ | 12.430 | 1.960 | 9.660 | 9.411 | 1.703 | -0.634 | -0.257 |
| Hg | 1471.290 | 0.620 | 1.540 | 2.322 | 24.215 | 60.307 | 3655.215 |

注: 偏度临界值 CO₂、CaCO₃、Hg 为 0.0805, 其它指标为 0.0845;

峰度临界值 CO₂、CaCO₃、Hg 为 0.0161, 其它指标为 0.169。

表 2 的资料看出, 化探原始指标中绝大多数指标的偏度与峰度均大于临界值, 说明指标不服从正态分布。而经过处理, 剔除特异点后, 基本服从正态分布 (表 3)

表 3 各指标经剔除特异点后的地球化学参数表

| 指标 | 背景值 | 标准差 | 偏度 | 峰度 | 偏度临界值 | 峰度临界值 |
|-------------------|---------|--------|---------|---------|-------|-------|
| C ₁ | 332.840 | 67.142 | -0.0251 | -0.0410 | 0.172 | 0.344 |
| C ₂ | 22.113 | 5.619 | 0.206 | -0.133 | 0.171 | 0.341 |
| C ₃ | 8.732 | 2.313 | 0.232 | -0.142 | 0.172 | 0.343 |
| iC ₄ | 2.539 | 0.851 | 0.583 | -0.2167 | 0.188 | 0.376 |
| nC ₄ | 11.035 | 4.290 | 0.399 | 0.0818 | 0.171 | 0.343 |
| iC ₅ | 1.771 | 0.582 | 0.458 | -0.113 | 0.172 | 0.346 |
| nC ₅ | 0.844 | 0.349 | 0.505 | -0.258 | 0.173 | 0.346 |
| C ₂₊ | 48.572 | 13.497 | 0.328 | 0.191 | 0.172 | 0.343 |
| CO ₂ | 6.370 | 3.980 | 0.687 | 0.0369 | 0.083 | 0.166 |
| CaCO ₃ | 9.436 | 1.664 | -0.542 | -0.614 | 0.081 | 0.161 |
| Hg | 1.587 | 0.526 | 0.779 | 0.0028 | 0.084 | 0.168 |

3.1.7 方差分析

方差分析用来判断化探指标的适用性、有效性及优选指标组合，本文重点讨论单因素方差分析。

目的：通过一个指标或变量研究一个因素（母体）的差异性。

输入：因素 A 有 m 个水平（m 个取样层），即 A₁, A₂, …, A_m。每层取 K 个样品，分析甲烷数据组成的矩阵 x 和给定可信度 α（0.1, 0.5, 0.01 等），可信度根据研究精度要求而定，若精度要求高，选较小可信度，反之则选较大的。

$$x = \begin{pmatrix} x_{11}, x_{12}, \dots, x_{1k} \\ x_{21}, x_{22}, \dots, x_{2k} \\ \dots \\ x_{m1}, x_{m2}, \dots, x_{mk} \end{pmatrix} = [x_{ij}]_{m \times k}$$

其中 (i = 1, 2, …, m; j = 1, 2, …, n)

为了检验 A 因素各个水平 A₁, A₂, …, A_m 之间是否有显著性差异，从而对地球化学模型实例中各取样层甲烷参数是否有明显差别做出结论，同时检验单因素方差分析在研究解决具体问题上效果如何。

方法：计算统计量：

$$F = \frac{S_A / (m-1)}{S_e / m(k-1)}$$

其组间差 $S_A = \frac{1}{k} \sum_{i=1}^m (\sum_{j=1}^k x_{ij})^2 - \frac{1}{mk} (\sum_{i=1}^m \sum_{j=1}^k x_{ij})^2$ ，组内差

$$S_e = \sum_{i=1}^m \sum_{j=1}^k x_{ij}^2 - \frac{1}{k} \sum_{i=1}^m (\sum_{j=1}^k x_{ij})^2$$

根据给定可信度 α 查 F 分布表得到临界值 $F_{\alpha}[(m-1), m(k-1)]$ (此过程已把 F 分布表作为内部数组)，比较检验 F 和 F_{α} 。若 $F > F_{\alpha}$ ，则说明 A 总体各水平 A_1, A_2, \dots, A_m 之间有显著性差异；否则，说明无显著性差异 (原理略)。

输出：SA, Se, F, F_{α} , α , m, k 及判定结果。

油气化探是多参数的综合勘探，所获得的分析数据是化探异常研究的基础。然而各方法的分析数据是否有实用价值？这就需要查明与诊断分析数据的起伏变化是地球化学场的自然差异还是分析误差或者是采样误差等因素引起的。只有当分析误差及采样误差显著低于地球化学变差 (或总变差) 时，所获得分析数据才有使用价值。因此，对每次化探调查必须进行采样地点间 (V_1)、重复采样间 (V_2) 及样品重复分析试验间 (V_3) 的变差进行三层套合方差分析加以判别。

上述方差比较，可用 F 检验进行 (表 4)：

① 当 $V_1/V_2 = F_1 > F_{0.05} [(a-1), a(b-1)]$ 时，表示采样及分析变差显著小于各采样点间的地球化学变差；

② 当 $V_2/V_3 = F_2 > F_{0.05} [a(b-1), ab(c-1)]$ 时，表示分析误差显著小于各采样及分析变差；

只有当地球化学变差 (或总变差) 显著大于分析及采样变差 ($F_1 > F_1$ 临界值) 时，所获得分析数据才有使用价值；若地球化学变差 (或总变差) 与分析及采样变差无明显差异时 ($F_1 < F_1$ 临界值) 时，说明采样及分析误差过大，分析数据可靠性差；采样及分析变差显著大于分析误差 ($F_2 > F_2$ 临界值) 时，分析误差不起主导作用；反之；($F_2 < F_2$ 临界值)，说明分析误差起主导作用。

表 4 方差分量计算表

| 变差来源 | 方差和代号 | 自由度 | 均方差 |
|-------|--------|-----------|------------------------|
| 采样地点间 | Ss_1 | $a-1$ | $V_1 = Ss_1 / (a-1)$ |
| 重复采样间 | Ss_2 | $a(b-1)$ | $V_2 = Ss_2 / a(b-1)$ |
| 重复分析间 | Ss_3 | $ab(c-1)$ | $V_3 = Ss_3 / ab(c-1)$ |

注：a—重复采样点位间，一般占 10%；

b—每个点位上重复采样数，一般为 2；

c—每个检查点的样品进行重复分析次数，一般为 2。

3.1.8 离群点的识别与处理

在概率统计中，化探数据的正态分布密度函数可分为四种情况（如图 4）。

A. 一个总体是由正常值组成，正常值的分布密度函数为 $f_N(x)$ ， $f(x) = f_N(x)$ 。

B. 一个总体是由正常值及特异低值组成，正常值的分布密度函数为 $f_N(x)$ ，在正常值左端特异低值的分布密度函数为 $f_{o-}(x)$ ，且特异低值占总体的比例为 P_1 ，则该总体观测值的密度分布函数 $f(x)$ ，分别表示为： $f(x) = (1 - P_1) f_N(x) + P_1 f_{o-}(x)$ 。

C. 一个总体是由正常值及特异高值组成，正常值的分布密度函数为 $f_N(x)$ ，在正常值右端特异高值的分布密度函数为 $f_{o+}(x)$ ，且特异高值占总体的比例为 P_2 ，则该总体观测值的密度分布函数 $f(x)$ ，分别表示为： $f(x) = (1 - P_2) f_N(x) + P_2 f_{o+}(x)$ 。

D. 当特异值同时出现在正常值分布左右两端时，则该总体观测值的密度分布函数为： $f(x) = (1 - P_1 - P_2) f_N(x) + P_1 f_{o-}(x) + P_2 f_{o+}(x)$ 。

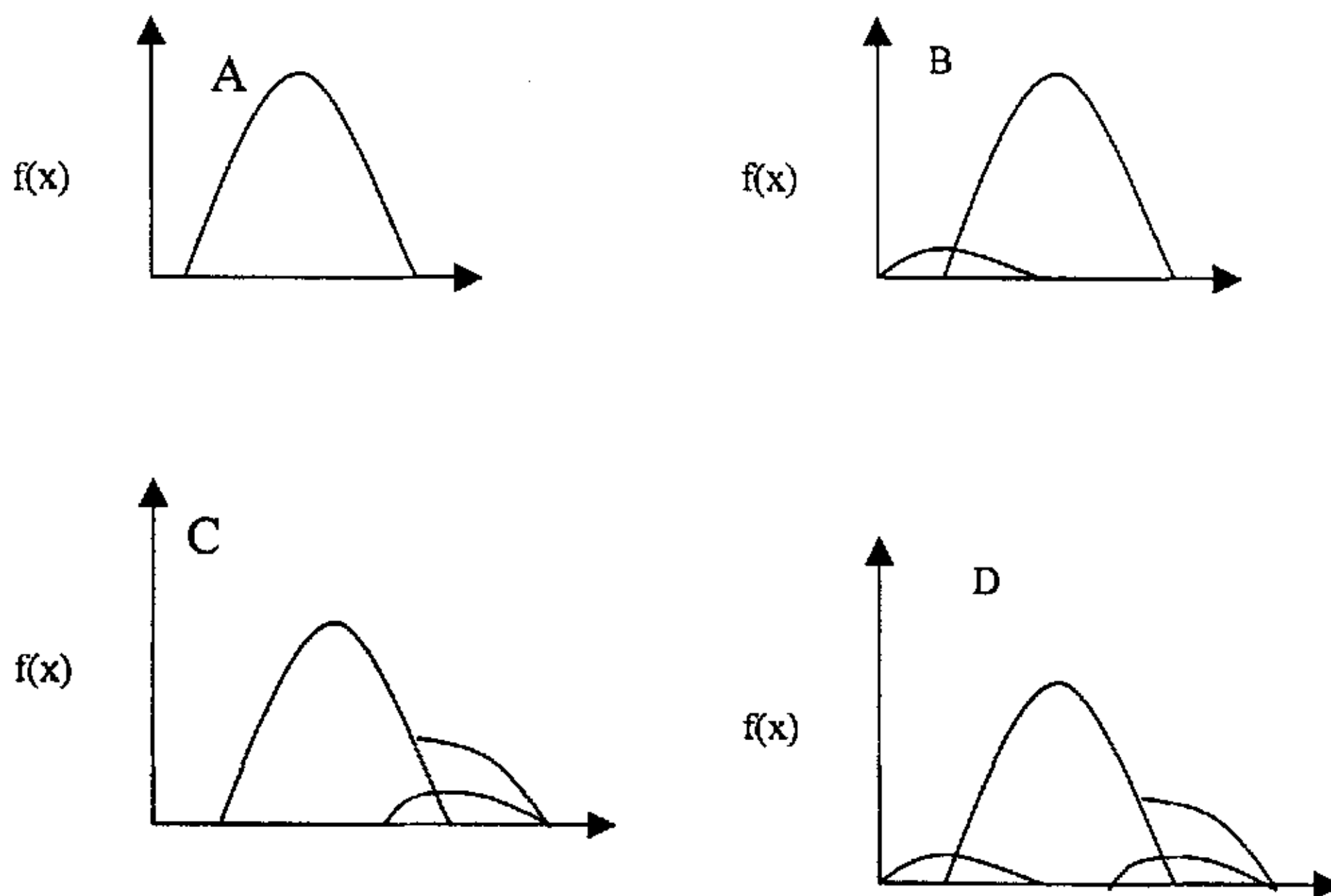


图 4 $f_{o-}(x)$ 、 $f_{o+}(x)$ 与 $f(x)$ 的分布密度函数之间的关系

由于油气运移的复杂性以及多种一般不能人为控制的变化因素（采样的、分析的、环境的和人为污染等）的影响，实际上，化探分析的数据呈现多源母体的分布特征。在数据总体中，那些比全部数值的均值或中位值高得多或低得多的数值（即所谓的特异值），可明显地影响数据总体的统计量。对离群点的看法，一种认为它既非由分析误差所引起，也非由采样方法等人为因素造成，而是实际存在于所研究的母体中；另一种看法是认为它是由多种干扰因素造成的。所以对特异值的处理应持慎重态度，进行多方面识别和处理。

从传统的统计上来讲，对离群点的识别就是从总体观测值分布 $f(x)$ 中将 $f_-(x)$ 、 $f_{0+}(x)$ 与 $f_N(x)$ 区分开来。当认为化探数据来自单一母体时，对特异值的识别方法在统计学上有概率迭代统计法和箱型图法等，通常用均方差的倍数来识别特异值（特异值定为样品均值 M 加三倍均方差 S ： $M+3S$ ）或上下四分位数的差值的 1.5 或 3 倍以上的数据认为是离群点。相应的处理方法是采用删除法或替代法。

当研究区海域景观单元不一，化探数据中离群点的分布与采样数据的位置密切相关，因此，从地质统计学观点看，一个观测值是否是离群点不仅要考虑观测值本身，还应考虑与其相邻的若干样品的观测值（即空间环境）。这里介绍霍金斯（1990）基于地质统计的基本思想提出的估计邻域法。其识别离群点的统计量是：

$$I = \frac{n(G-M)^2}{(n+1)\sigma^2}$$

式中： I 是识别离群点的统计量，它服从自由度为 $1-\infty$ 的 F 分布，当 $I > 3.48$ 时，可疑值 G 被确定为特异值，即表示在 95% 的置信区间的基础上确定 G 为离群点； G 为被研究的可疑值； M 为不包含 G 的邻域内其它样品值的算数平均值； n 为不包含 G 的邻域内的样品数； σ^2 为邻域内观测值的平均方差。其中 n 的大小可按样品中所研究变量的变异性来确定，即用变异函数所确定的变程 a 来确定。式中的计算公式是：

$$\sigma^2 = \frac{1}{m^2} \sum_{k=1}^m \sum_{l=1}^m \gamma_{kl}(h)$$

式中 $\gamma_{kl}(h)$ 为相邻 h 的两个样品 k, l 的观测值 $Z(k)$ 与 $Z(l)$ 的变异函数值, m 为包括可疑值在内的邻域内宽度 (即样品数)。

当识别出离群点的统计量, 识别出离群点后, 可用离群点的下限值 GL 代替离群点:

$$G_L = \sqrt{\frac{3.84 \sigma^2 (n+1)}{n}} + M$$

对于化探方法中的酸解烃法、热释烃法、顶空气法以及荧光光谱法、紫外光谱法等, 分析测试得为一系列的烃气数据或光谱强度数据, 它们之间往往呈一定的关系, 如酸解烃中常呈甲烷 > 乙烷 > 丙烷 > 丁烷, 可以根据其含量或比值进行回归分析识别离群点, 并用回归值代替。

另外, 通过重复采样和重复分析是识别离群点最根本的方法。

3.2 优选指标或指标组合的方法

海域油气化探的基本出发点是采取不同介质的样品, 应用多种方法与指标, 达到获得深部油气信息的目的。但是由于各个指标的成因属性有一定差异, 在复杂的地层构造及沉积岩相等条件下, 它们有一定的局限性, 受应用条件的限制, 只能反映油气的一方面信息。在确定异常与评价异常的含油气性时所起的作用有一定的区别。为了多角度、深层次地揭示油气藏的浅层地球化学效应特征, 需要筛选有效的指标或指标组合, 最大限度地提取综合信息, 提高化探预测成功率。优选指标与指标组合的数据处理方法主要有:

3.2.1 相关分析

相关分析是研究化探指标之间的关系, 一般用相关系数和散点图来表达其相关程度。当相关系数超过某一信度的临界值时, 表示指标间的相关性显著, 这类指标具有共同的成因或相似的地球化学意义。相关分析必须具备两个应用条件, 其一必须是同一总体样品, 二是要有足够的数对。常用的方法有简单相关分析、复相关分析、偏相关分析及典型相关分析。

某海区化探指标的相关矩阵如表 5、表 6 所示。

表5 水化学相关矩阵 (样品数 375 个)

| | C ₁ | F320nm | F360nm | F405nm | U216nm | U222nm | U232nm | 酚 |
|----------------|----------------|--------|--------|--------|--------|--------|--------|---|
| C ₁ | 1 | | | | | | | |
| F320nm | -0.02 | 1 | | | | | | |
| F360nm | -0.01 | 0.82 | 1 | | | | | |
| F405nm | -0.01 | 0.75 | 0.99 | 1 | | | | |
| U216nm | -0.01 | 0.23 | 0.04 | 0.01 | 1 | | | |
| U222nm | -0.00 | 0.30 | 0.04 | 0.00 | 0.27 | 1 | | |
| U232nm | -0.01 | 0.43 | 0.14 | 0.09 | 0.61 | 0.87 | 1 | |
| 酚 | -0.00 | 0.02 | -0.02 | -0.02 | 0.08 | 0.08 | 0.08 | 1 |

表6 沉积物指标相关矩阵

| | C ₁ | C ₃ | C ₂ ⁺ | ΔC | 热释汞 | 碳酸盐 |
|-----------------------------|----------------|----------------|-----------------------------|------|------|-----|
| C ₁ | 1 | | | | | |
| C ₃ | 0.68 | 1 | | | | |
| C ₂ ⁺ | 0.67 | 0.99 | 1 | | | |
| ΔC | -0.36 | 0.27 | 0.27 | 1 | | |
| 热释汞 | -0.16 | 0.18 | 0.18 | 0.45 | 1 | |
| 碳酸盐 | -0.37 | 0.24 | 0.25 | 0.86 | 0.43 | 1 |

水化学指标：芳烃指标间的相关系数高，尤其是荧光光谱 (F) 的三个波段呈显著相关关系 ($r \geq 0.75$)，紫外光谱 (U) 222nm 与 232nm 相关系数高达 0.87。而水溶烃和酚及同其它指标相关性较差。

沉积物指标：吸附烃各指标 (C₁、C₃、C₂⁺) 间相关系数达到 0.67 以上，表明相关性很高；ΔC、热释汞、碳酸盐之间有一定的相关性，而与烃类指标间相关性较差。

3.2.2 因子分析

因子分析是以原始数据间的相关关系为基础，利用数学方法把关系复杂的多个变量归结为少数几个独立的新变量（原始变量的线性组合）即因子。在损失原始信息不多的情况下，揭示原始数据中起主要作用的因素，更合理地解释存在于原始变量之间的相关性，且简化变量结构。

根据研究对象，R 型因子分析研究地质变量，用线性组合的少数几个理论变量（公因子）代替原来多个实际变量，研究变量组合划分、相关性及其优选。Q 型因子分析研究样品，用公因子代替实际样品（如统计预测单元、不同类型地

质体), 研究样品之间的分组、分类、相关性及其优选。公因子不仅具备相关性极小或不相关特点, 而且又能把原始地质变量和样品所包含的不十分明显的差异尽可能多的反映出来。

这里把 R 型和 Q 型因子分析放在一个程序中, 选择实现。程序大致流程见图 5。

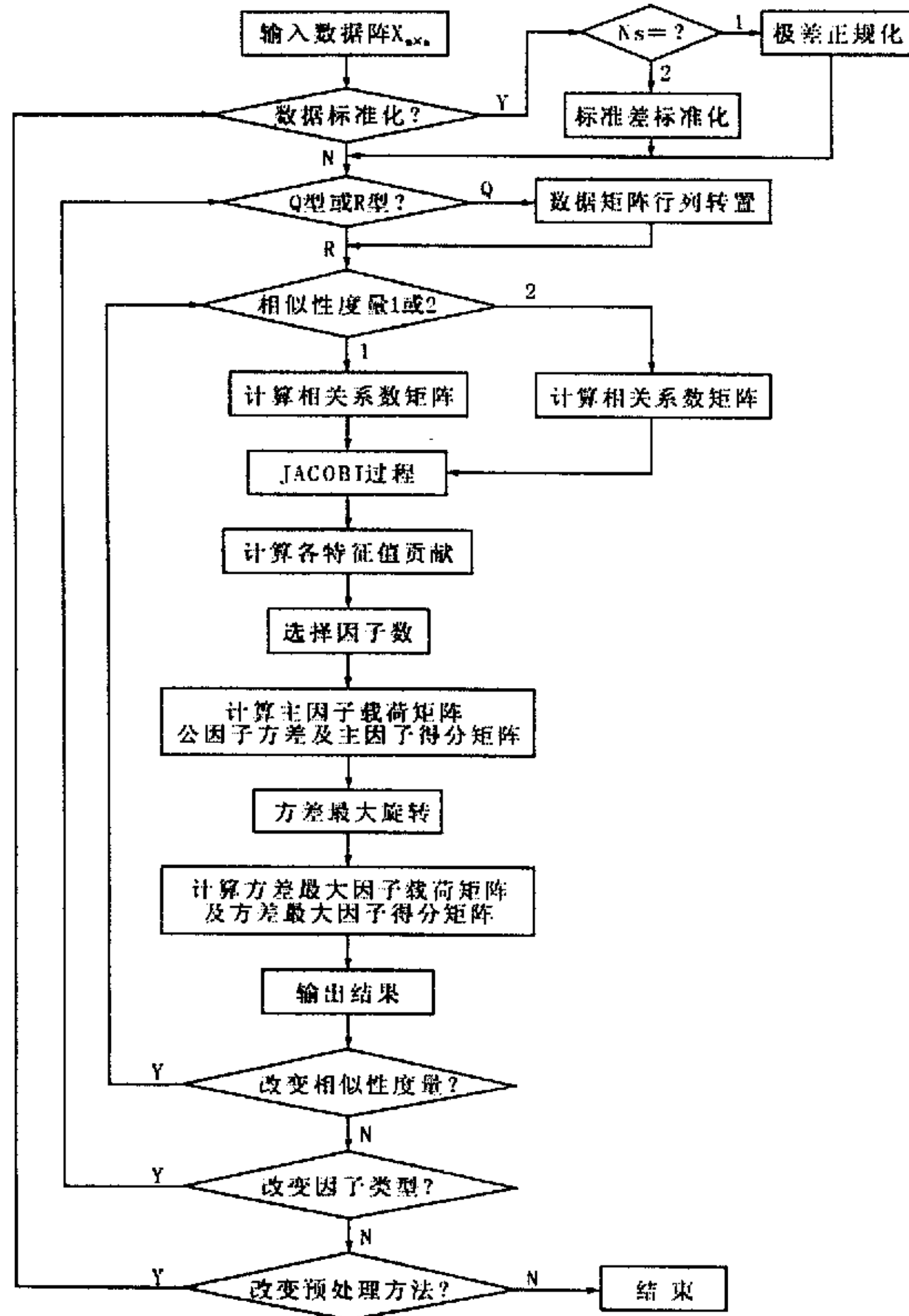


图 5 因子分析程序流程 (据赵永军等, 1998, 有修改)

为反映原始变量之间的差异, 弃去原始变量之间的重叠信息, 并提取新的指标组合, 对海水和沉积物两类指标进行 R 型因子分析, 结果见表 (7、8、9、10), 降维以后, 得出的主要指标组合为:

| | | |
|------|------------|-------------|
| 水化学: | A 荧光光谱 | 贡献值为 37.07% |
| | B 紫外光谱 | 贡献值为 25.46% |
| | C 水溶烃 | 贡献值为 13.50% |
| | D 酚 | 贡献值为 12.31% |
| 沉积物: | A 丙烷、重烃、甲烷 | 贡献值为 40.40% |

B 碳酸盐与 ΔC 贡献值为 38.49%

C 热释汞 贡献值为 11.41%

表 7 水化学指标初始因子载荷

| 指标 | I | II | III | IV | V |
|--------|--------|--------|--------|--------|--------|
| C_t | -0.27 | 0.003 | 0.99 | 0.15 | -0.01 |
| F320nm | 0.91 | -0.17 | -0.002 | 0.01 | 0.01 |
| F360nm | 0.84 | -0.53 | 0.01 | 0.03 | -0.01 |
| F405nm | 0.80 | -0.56 | 0.01 | 0.03 | -0.01 |
| U216nm | 0.39 | 0.57 | -0.01 | 0.01 | -0.72 |
| U222nm | 0.49 | 0.71 | 0.03 | -0.10 | 0.47 |
| U232nm | 0.63 | 0.75 | 0.02 | -0.09 | 0.08 |
| 酚 | 0.05 | 0.17 | -0.15 | 0.97 | 0.07 |
| 累计方差贡献 | 37.70% | 62.53% | 75.03% | 87.34% | 96.70% |

表 8 水化学指标正交因子载荷

| 指标 | I | II | III | IV | V |
|--------|--------|--------|--------|--------|--------|
| C_t | -0.01 | 0.001 | 0.99 | -0.001 | 0.002 |
| F320nm | 0.82 | -0.34 | -0.02 | 0.01 | -0.26 |
| F360nm | 0.66 | -0.64 | -0.01 | -0.02 | -0.37 |
| F405nm | 0.62 | -0.66 | -0.002 | -0.02 | -0.37 |
| U216nm | 0.42 | 0.71 | -0.002 | 0.06 | -0.56 |
| U222nm | 0.72 | 0.37 | 0.003 | 0.05 | 0.57 |
| U232nm | 0.81 | 0.53 | -0.004 | 0.04 | 0.18 |
| 酚 | 0.02 | 0.03 | -0.001 | 0.99 | 0.002 |
| 累计方差贡献 | 35.65% | 59.07% | 71.57% | 84.16% | 96.70% |

表9 沉积物指标初始因子载荷

| 指标 | I | II | III |
|-----------------------------|--------|--------|--------|
| C ₁ | 0.42 | -0.86 | 0.09 |
| C ₃ | 0.91 | -0.39 | -0.06 |
| C ₂ ⁺ | 0.91 | -0.38 | -0.06 |
| ΔC | 0.60 | -0.72 | -0.22 |
| 热释汞 | 0.45 | 0.48 | 0.75 |
| 碳酸盐 | 0.58 | 0.73 | -0.24 |
| 累计方差贡献 | 45.40% | 83.89% | 95.30% |

表10 沉积物指标正交因子载荷

| 指标 | I | II | III |
|-----------------------------|--------|--------|--------|
| C ₁ | 0.25 | -0.80 | 0.47 |
| C ₃ | 0.81 | -0.55 | 0.15 |
| C ₂ ⁺ | 0.81 | -0.54 | 0.14 |
| ΔC | 0.71 | 0.41 | -0.49 |
| 热释汞 | 0.58 | 0.66 | 0.48 |
| 碳酸盐 | 0.69 | 0.42 | -0.51 |
| 累计方差贡献 | 45.02% | 78.65% | 95.30% |

3.2.3 聚类分析

聚类分析又称群分析，它是研究对指标或样品进行分类的一种统计方法，根据化探测量点许多指标（或样品）具体计算它们之间的相似程度，把相似的指标（或样品）归为一类，把不相似的指标（或样品）归为另一类。关系密切的归到一个小的分类单元，关系不密切的归到一个大的分类单元，直到把所有的指标（或样品）都归类完毕，形成一个由小到大的分类系统。最后，再把整个分类系统画成分群图和聚类分析谱系图，用它把所有指标（或样品）的亲疏关系表示出来，提供分类的一方面结果。

根据分类对象的不同，将样品分类称为 Q 型聚类分析，将指标分类称为 R 型聚类分析。

应用聚类分析方法，可以突破传统地球化学所建立的一些定性分类系统，形成具有定量概念的分类关系，借以了解油气化探中各指标间的共同关系。在评价化探异常时，通过对样品进行分类区分不同的异常成因类型。

聚类分析的步骤与方法：

目的：根据需要，对变量或样品按照其亲疏关系进行分类归群。

输入：由 n 个样品、 m 个变量的观测值 x_{ij} 组成的原始数据矩阵 $X_{m \times n}$ ；输入要聚类的方法，是 R 型变量，还是 Q 型样品；在进行聚类统计量之前要进行原始数据阵的标准化或正规化，把所有数据统一量纲，形成预处理后数据阵 $X'_{m \times n}$ 。

$$X_{m \times n} = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1m} \\ x_{21}, x_{22}, \dots, x_{2m} \\ \vdots \\ x_{n1}, x_{n2}, \dots, x_{nm} \end{bmatrix}$$

方法：①研究样品的相似性（Q 型）等价于研究 $X'_{m \times n}$ 矩阵中行与行之间的关系，对矩阵进行归类，研究变量的相关性（R 型）等价于研究 $X'_{m \times n}$ 矩阵中列与列之间关系，对矩阵列进行归类。②在确定 R 型或 Q 型分析后，计算相似系数、相关系数、距离系数各统计量，形成各系数矩阵。③从最短距离聚合法、距离类平均法、相关（相似）系数近邻连接法、相关（相似）系数类平均法等四种方法选择其一，进行逐步的变量（样品）合并，得到分类结果，并做出相应聚类图。

最短距离聚合法中，若类 p 和 q 合并成新类 r ，那么计算类 r 与某类 f 的距离用递推公式：

$$D_{rf} = \frac{1}{2}D_{fp} + \frac{1}{2}D_{fq} - \frac{1}{2}|D_{fp} - D_{fq}|$$

距离类平均法中，定义两类之间的距离为样品（变量）点之间距离的加权重值，即

$$D_{pq} = \frac{1}{npnq} \sum_{\substack{x_i \in p \\ x_j \in q}} d_{ij}$$

若类 p 和 q 合并成新类 r ，那么计算类 r 与某类 f 的距离用递推公式：

$$D_{rf} = \frac{np}{nr} D_{fp} + \frac{nq}{nr} D_{fq}$$

其中 nr 、 np 、 nq 分别为 r 、 p 、 q 中的样品（变量）数，且 $nr = np + nq$ 。

相关（相似）系数法中，设类 p 和 q 合并成新类 r ，那么计算类 r 与某类 f 之间相关（相似）度量的递推公式类似距离法，分别为：

$$R_{rf} = \frac{1}{2}R_{fp} + \frac{1}{2}R_{fq} - \frac{1}{2}|R_{fp} - R_{fq}| \quad R_{rf} = \frac{np}{nr}R_{fp} + \frac{nq}{nr}R_{fq}$$

输出：相似系数、相关系数、距离系数各矩阵数据体、最终聚类图。

R 型聚类分析结果见图 6、7，从谱系图上可知：

水化学指标：相关系数在 0.1 水平上，8 个指标分为三群，即：水溶烃甲烷、酚及光谱；相关系数在 0.4 时，后者将分化为紫外光谱和荧光光谱两组。

沉积物指标：相关系数在 0.4 水平上，8 个指标分为两群，即：

烃类气体为一群，其他指标为另一群；相关系数在 0.6 时，后者将分化为热释汞一组和 ΔC 、碳酸盐一组。

综上所述，三种不同的数据处理方法起到了相互补充、验证的作用，其结果有明显的相似性，表明本区指标组合有较强的稳定性和较高的应用价值。其中：烃类气体（包括水与沉积物）为独立组合，它们是石油与天然气化学组成主要成分，被视为本区找油气的主要指标；荧光光谱和紫外吸收光谱及酚在水化学指标中占有最大的贡献值，它们是石油化学组成中仅次于烷烃的组份，是找油气的次要指标组合； ΔC 、汞和碳酸盐是油气的蚀变和伴生产物，并与沉积环境有一定关系，可作为辅助指标。

利用上述优选的指标组合在我国某海区化探预测中取得良好的地质效果，在综合异常上发现了以天然为主伴有凝析油的油气田（图 8）

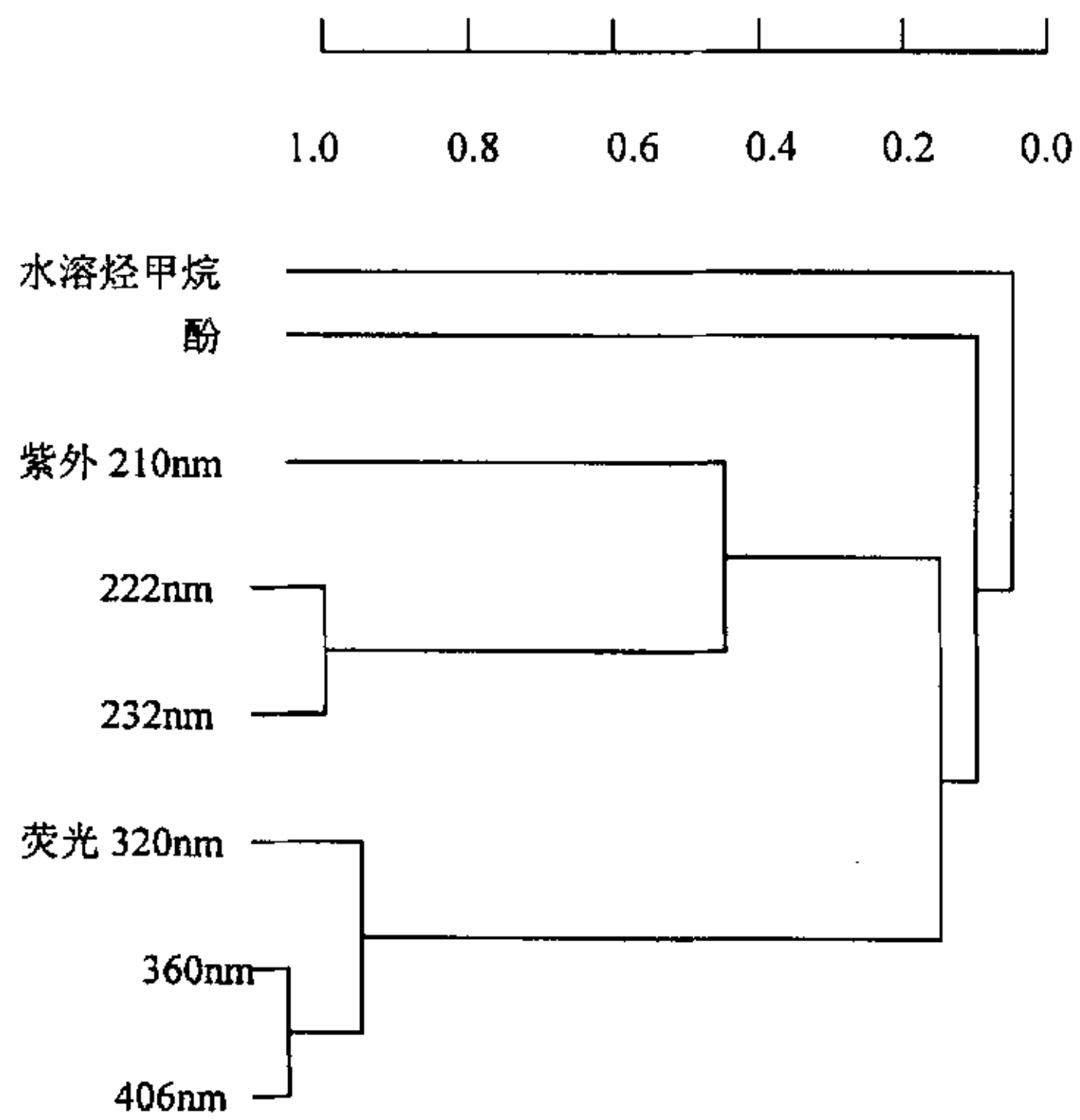


图 6 水化学指标聚类谱系图

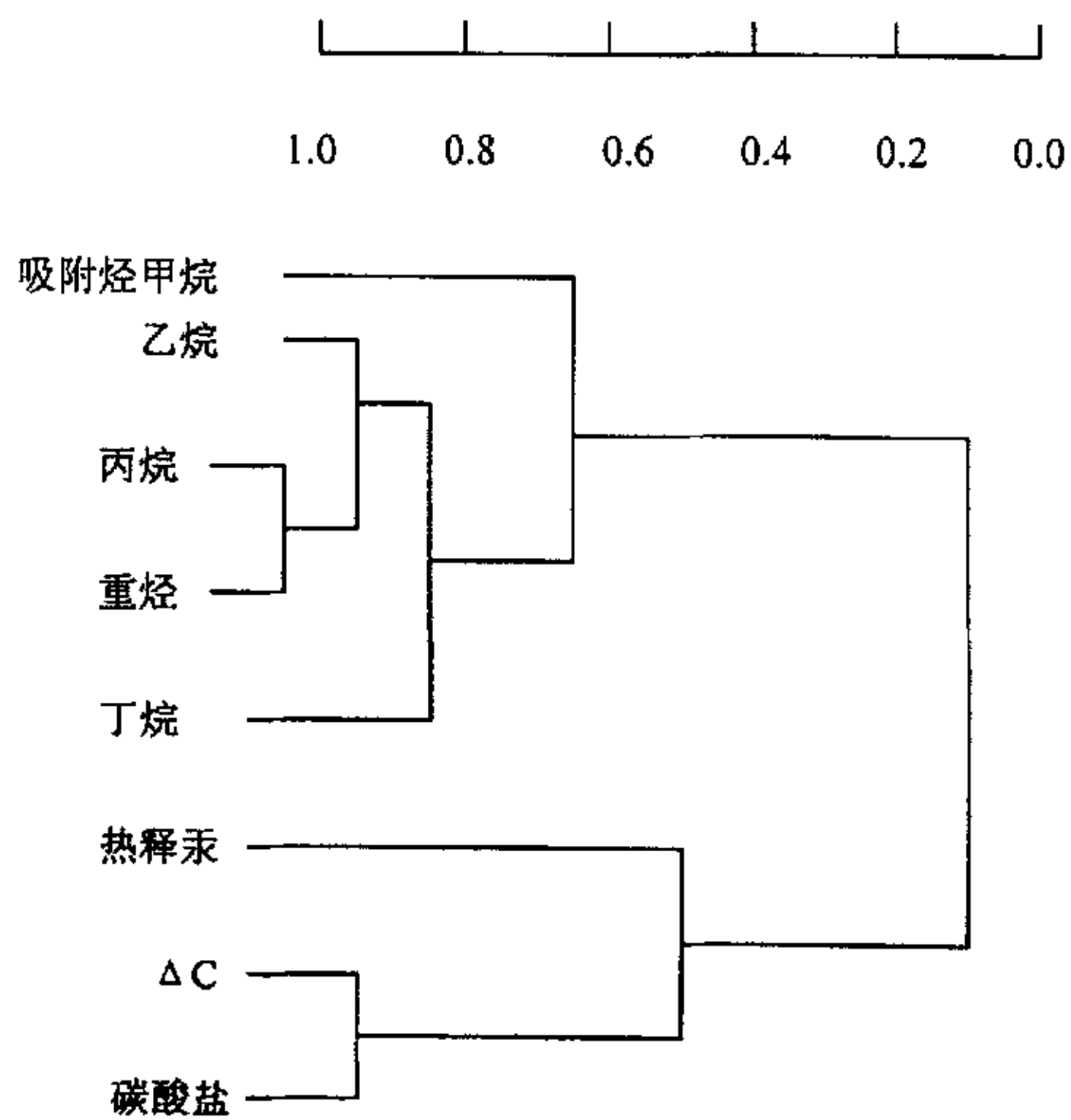
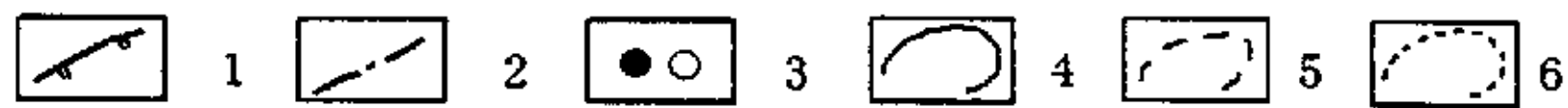
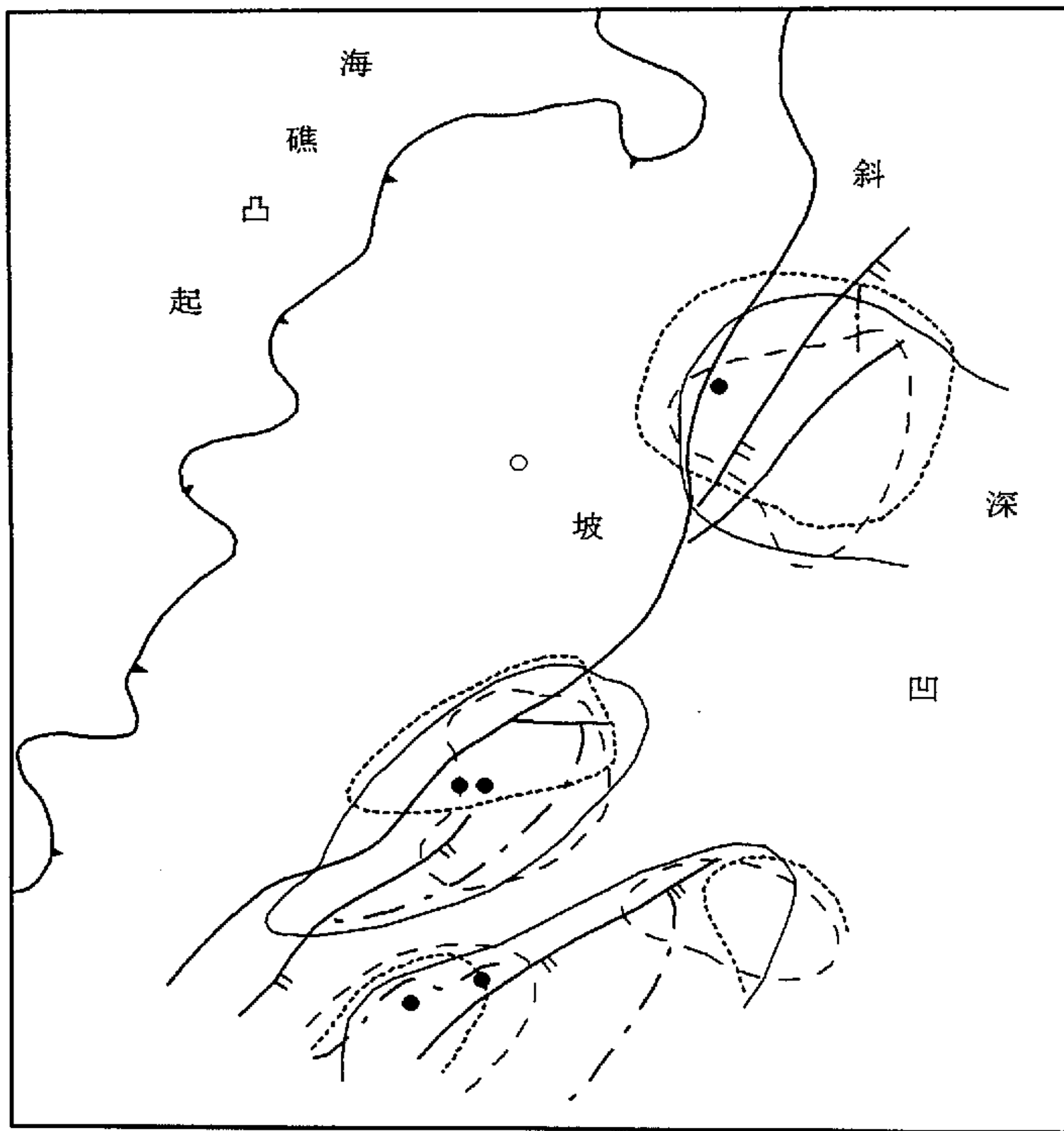


图 7 近海底沉积物指标聚类谱系图



1. 断层 2. 构造 3. 钻井 4. 主要指标 5. 次要指标 6. 辅助指标

图 8: 东海某研究区油气化探主要指标异常范围图

3.3 抑制干扰因素的方法

化探预测的成功与否，在某种意义上讲取决于干扰因素的抑制程度。前已述及，海洋油气化探的主要干扰因素是取样深度（包括沉积物和海水深度）、沉积物粒度、碳酸盐含量及近代生物地球化学作用。这些因素影响着烃类气体含量并与油气藏中运移出来的烃叠加在一起，目前的分析测试技术还不能分离或区分二者的含量，给地质解释带来很大困难，制约着化探方法的应用效果。通过实践证明数据处理技术，尤其是回归分析是抑制干扰的可行方法。

回归分析是根据最小二乘法原理进行最佳估计的一种统计方法，用来研

究变量和变量之间的相关关系，提供变量间相关关系的数学表达式(经验公式)，而且利用概率统计基础知识，对此关系进行分析，来判断所建立的经验公式的有效性。一元线性分析的流程如下：

目的：求一个随机变量 Y (因变量) 与一个普通变量 X (自变量) 之间线性回归方程。即用建立的回归方程，可以用一个指标的含 量估算另一个指标的含 量。

输入：自变量 X 和因变量 Y 自左向右按 x_1, x_2, \dots, x_n, y 排列的数据表。可信度 α 。

方法：用最小二乘法求回归系数，计算残差平方和 Q 、回归平方和 U 、统计量 $F=U(n-2)/Q$ ，查 F 分布表求 $F_{\alpha}(1, n-2)$ 值 (该表可设为内部数据数组，自动查询)。

输出：回归方程的系数、残差平方和 Q 、回归平方和 U 、可信度 α 、统计量 F 、 F_{α} 、最大偏差、最小偏差、平均偏差。

3.3.1 抑制深度影响的方法

由于不同深度的样品所含烃类含量有所变化，在确定化探异常时，必须是在同一深度样品含量的基础上进行。如果深度不同，需要事先设定一个同一深度，根据各点实测的不同深度烃类含量，进行插值，求得设定深度面上的分析值。

由于海域化探取样是不同年度、不同工作单位、多航次完成的，取样深度不统一，在资料整理研究时，首先要统计各采样点的深度，以多数点的深度为基点 (设沉积物样点 1.5 米)，按照这个深度求取插值，即使各点的烃类含量统一在 1.5 米的深度面上。

具体方法是将各个站位的不同深度样品分析测试结果投影到深度—烃含量的二维投影图上，求出一维线性函数 $Y=a+bx$ 的系数 a ，然后进行总体一维线性插值。把在 1.5 米深度无法测试数据的各个点位浅层或深层分析数据做为 b 值代入，求取 1.5 米深度的数值。然后和每个点根据自身纵向上的变化趋势而进行插值的结果比较，取两者的平均值做为最终插值结果，这样将指标含量校正到全区统一深度上进行对比增加了合理性和可靠性。

图 9 是以烃类含量为 Y 坐标，深度为 X 坐标，求取的方程为 $Y=1.82+3.95 X$ 图中为部分点的散点图及其拟合的直线。

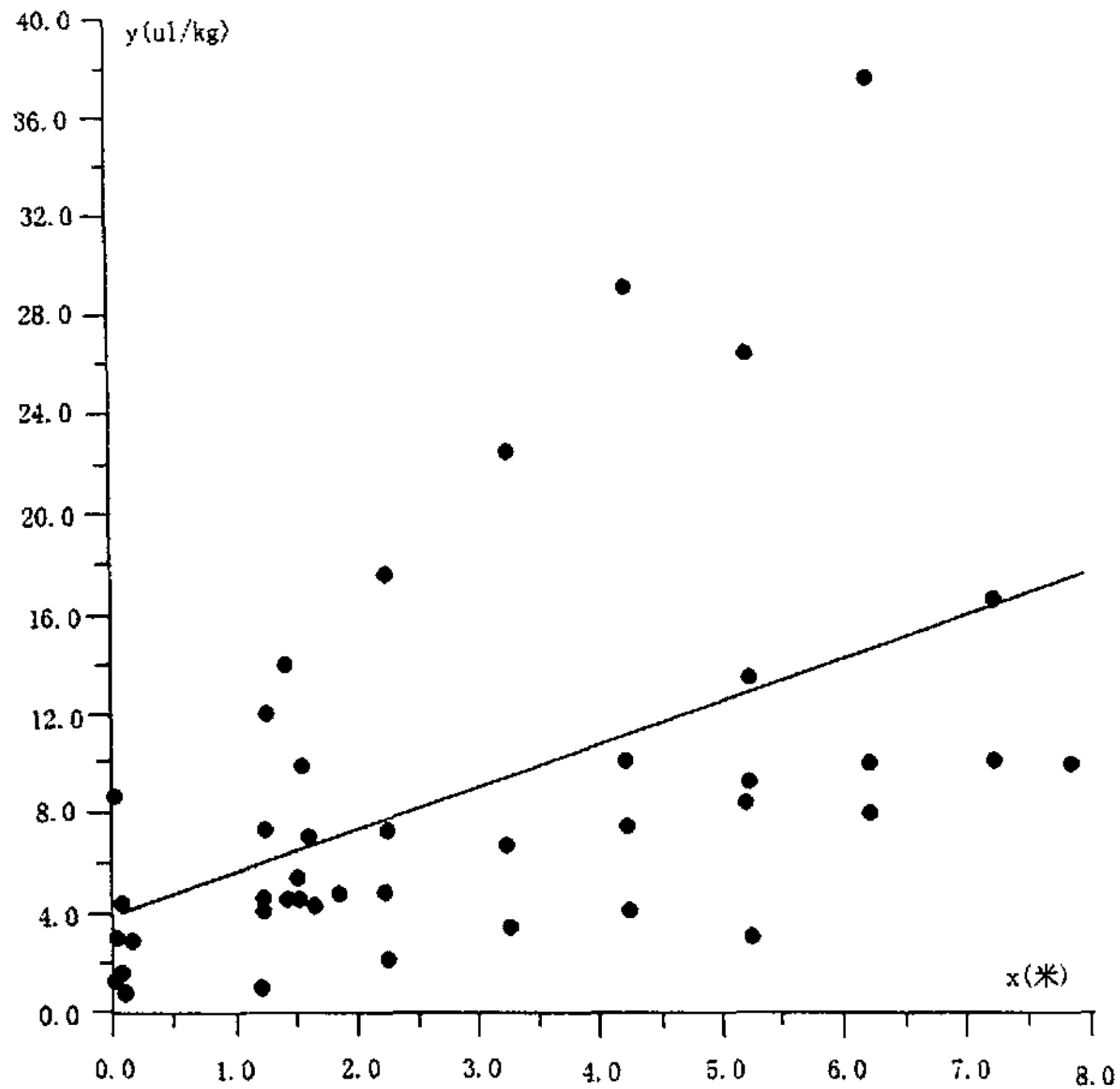


图9 深度校正投影图

应用上述方法也可以校正不同深度海水对烃类含量的影响。

3.3.2 抑制粒度影响的方法

从理论上讲，矿物颗粒越细，比表面积越大，吸附能力越强，所以粘土中烃含量要高于砂性土中的含量。但在自然界中并非完全如此，在许多地区见到烃含量随着沉积物粒度变粗而增高。图10是某海区320多个样品的统计资料，从中看出，甲烷含量随着砂粒含量的增高而增高，二者的相关系数为0.63，对于 $\alpha = 0.05 - 0.01$ 的可信度来讲都是显著相关的，存在着较好的线性关系。表11的资料说明，在样品颜色类似，取样深度相同及样点间距不远的情况下，只是由于粒度大小不一，颗粒粗细不同，烃的含量亦不相同，即砂性土的烃类含量高于粘性土0.1—1.5倍，其中尤以甲烷含量的变化较大。在纵向的一定范围内，同样出现上述现象。

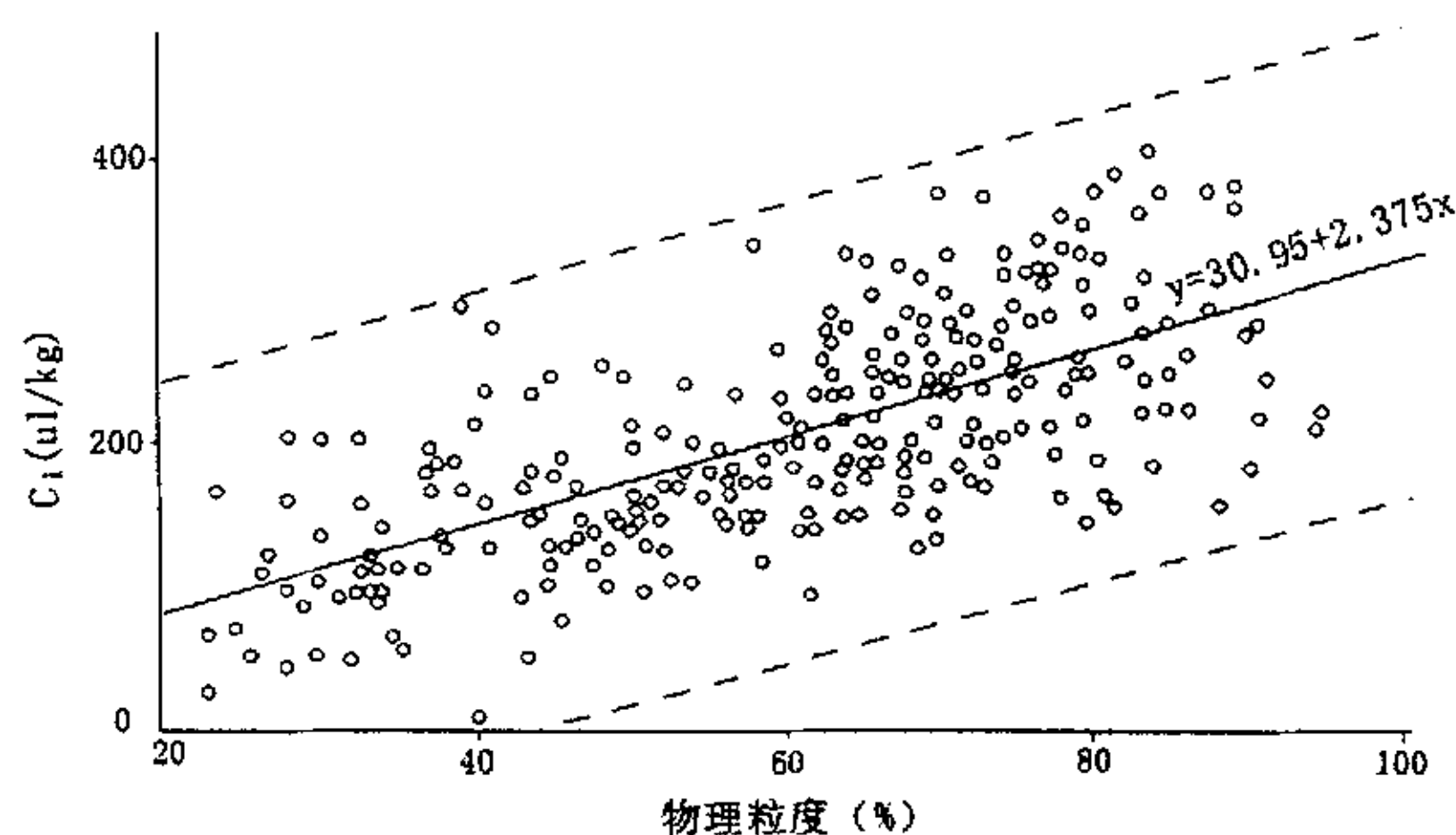


图 10 甲烷与粒度的回归直线

砂性土中的烃类含量高于粘性土的原因，是近代生物地球化学作用生成的烃类气体迭加的结果。

抑制粒度影响的方法是：将甲烷视为因变量 (Y)，粒度作为自变量 (X)，求回归线性方程的方法进行统一校正。如某区烃与沉积物粒度的回归方程是： $Y=30.95+2.89x$

上式表示，砂粒每增加 1%，甲烷含量增加 $2.89 \mu l/kg$ ，据此，可以将本海区内各测量点甲烷恢复到砂粒含量为 61.38% (海区平均值) 的统一水平上，使各测量点的甲烷含量在粒度一致的背景条件下进行比较，消除近代生物地球化学作用的影响。

另外，在化探资料较多的海域，可采用统计方法消除生物成因气的干扰。其理论依据是：生物成因气以甲烷为主，重烃 (乙烷、丙烷、丁烷等) 含量低。而来自油气藏的烃类气体，除含甲烷外，含有较多的重烃组分。可用各测点所含甲烷量与校正后各点重烃的含量之比值来划分异常，避免或消除生物成因气的影响。即：

$$\begin{aligned} \overline{C_1} / \sum \overline{C_{2+}} &= K \\ \sum C_{2+(i)} \times K &= \sum C_{2+(j)} \\ C_{1(i)} / \sum C_{2+(j)} &= C \end{aligned}$$

式中：

- $\overline{C_1}$ —— 砂性样中甲烷含量的均值；
- $\sum \overline{C_{2+}}$ —— 砂性样中重烃含量的均值；
- K —— 校正系数
- $\sum C_{2+(i)}$ —— 各测点重烃的含量
- $\sum C_{2+(j)}$ —— 校正后各测点重烃的含量
- $C_{1(i)}$ —— 各测点甲烷的含量

C——各测点甲烷的含量与校正后各测点重烃含量之比。

表 11 甲烷含量与岩性的关系

| 海区 | 岩性 | C ₁ | 取样深度 (M) | 两点间距离 (KM) |
|------------------|---------|----------------|----------|------------|
| 黄 渤 海 区 | 棕黄色亚粘土 | 38.81 | 1.5 | 1.5 |
| | 棕黄色亚砂土 | 100.00 | 1.8 | |
| | 棕黄色亚粘土 | 353.04 | 1.5 | 0.8 |
| | 灰黄色亚砂土 | 512.82 | 1.5 | |
| | 棕黄色亚粘土 | 280.01 | 1.5 | 2.0 |
| | 棕黄色亚砂土 | 527.85 | 1.5 | |
| | 棕黄色亚粘土 | 551.63 | 1.5 | 1.5 |
| | 棕黄色亚砂土 | 720.52 | 1.5 | |
| | 棕黄色亚粘土 | 301.70 | 1.5 | 1.8 |
| | 棕黄色亚砂土 | 518.70 | 1.5 | |
| | 棕黄色亚粘土 | 251.18 | 1.5 | 2.0 |
| | 棕黄色亚砂土 | 687.04 | 1.5 | |
| 东 海 区 | 黄色亚粘土 | 123.31 | 1.2 | 0.5 |
| | 黄色亚砂土 | 294.11 | 1.2 | |
| | 浅棕黄色亚粘土 | 179.19 | 1.2 | 0.5 |
| | 浅棕黄色亚砂土 | 265.87 | 1.2 | |
| | 浅棕黄色亚粘土 | 170.78 | 1.2 | 0.5 |
| | 浅黄色亚砂土 | 215.46 | 1.2 | |
| | 黄棕色亚粘土 | 106.90 | 1.2 | 0.5 |
| | 浅黄色亚砂土 | 248.93 | 1.2 | |
| | 棕黄色亚粘土 | 169.55 | 1.2 | 0.5 |
| | 浅黄色亚砂土 | 254.33 | 1.2 | |
| | 黄色亚粘土 | 187.71 | 1.2 | 0.5 |
| | 黄色亚砂土 | 222.36 | 1.2 | |

3.3.3 抑制碳酸盐影响的方法

现代沉积物中不同程度地含有多种类型的碳酸盐，其成因有无机碳酸盐和有机碳酸盐两种。

碳酸盐影响烃类含量的原因之一是，在脱气过程中，由于使用酸溶液破坏样品的矿物结构，碳酸盐遇酸分解，生成或释放内部所含的烃类气体，从而导致烃浓度比正常含量增高。也就是，样品中碳酸盐含量越高，烃含量越大。二

者同步增高的现象是各海区普遍见及的。利用前述求回归直线方程的方法，对大于碳酸盐区域平均值的各化探点的烃类含量进行定量校正。如某海区校正方程式如下：

$$\text{甲烷校正方程：} Y_{c1}=144.93+9.27X_{\text{rco3}}$$

$$\text{重烃校正方程：} Y_{c2+}=5.39+1.42X_{\text{rco3}}$$

上述方程式的物理意义与前述甲烷一粒度校正前同。

3.3.4 综合校正干扰因素的方法

在实际工作除上述单一校正干扰因素外，往往要统一校正多种干扰因素，此时，常引用多元逐步回归方法。设 Y 为描述任一地质条件下甲烷的含量， X_i 为影响甲烷浓度的任一因素，利用 Y 与 X_i 的多元线性回归方程，可以对甲烷含量与多种影响因素进行一次性校正，即：

$$Y=a_0+a_1X_1+a_2X_2+\dots+a_nX_n$$

下述回归方程是东海某构造的具体校正结果：

$$C_1=-503.31+8.08_{(\text{粒度})}+72.69_{(\text{碳酸盐})}-89.96_{(\text{水深})}-1.35_{(\text{有机质})}$$

3.4 确定异常的方法

油气化探异常是对于背景而存在的，但背景与异常不是某个具体数值，而是某一特定的分布，如正态分布、对数正态分布、二项分布、泊松分布等多种分布形式。在油气化探中习惯使用正态或对数正态分布理论来研究背景与异常，最重要的参数就是均值与标准差。所谓异常与背景实际上就是不同均方差与均值的背景分布和异常分布样本集合（图 11），从中可以看出，当背景分布与异常分布分离不很大时（最常见情况、本区情况更是如此），若干属于背景分布的样品可能进入异常分布，反之亦然。如果选取 A 为异常下限，则背景分布中大于 A 的部分样品将错误地划为异常样品，这类错误称为第一类错误，其发生的概率等于 A 右面与背景分布所围的面积。随着 A 向右移动，第一类错误的概率就可以减小，但第二类错误发生的概率就增加。因此，从理论上讲，异常下限的确定就是求出使两类错误均为极小的 A 值。对于两种分布互不重叠时，异常下限的确定没有任何困难，但是在油气化探中，异常与背景总是十分接近，需要认真的确定异常，才能取得良好的地质效果。确定异常所要做的就是分离这两个样本集合。

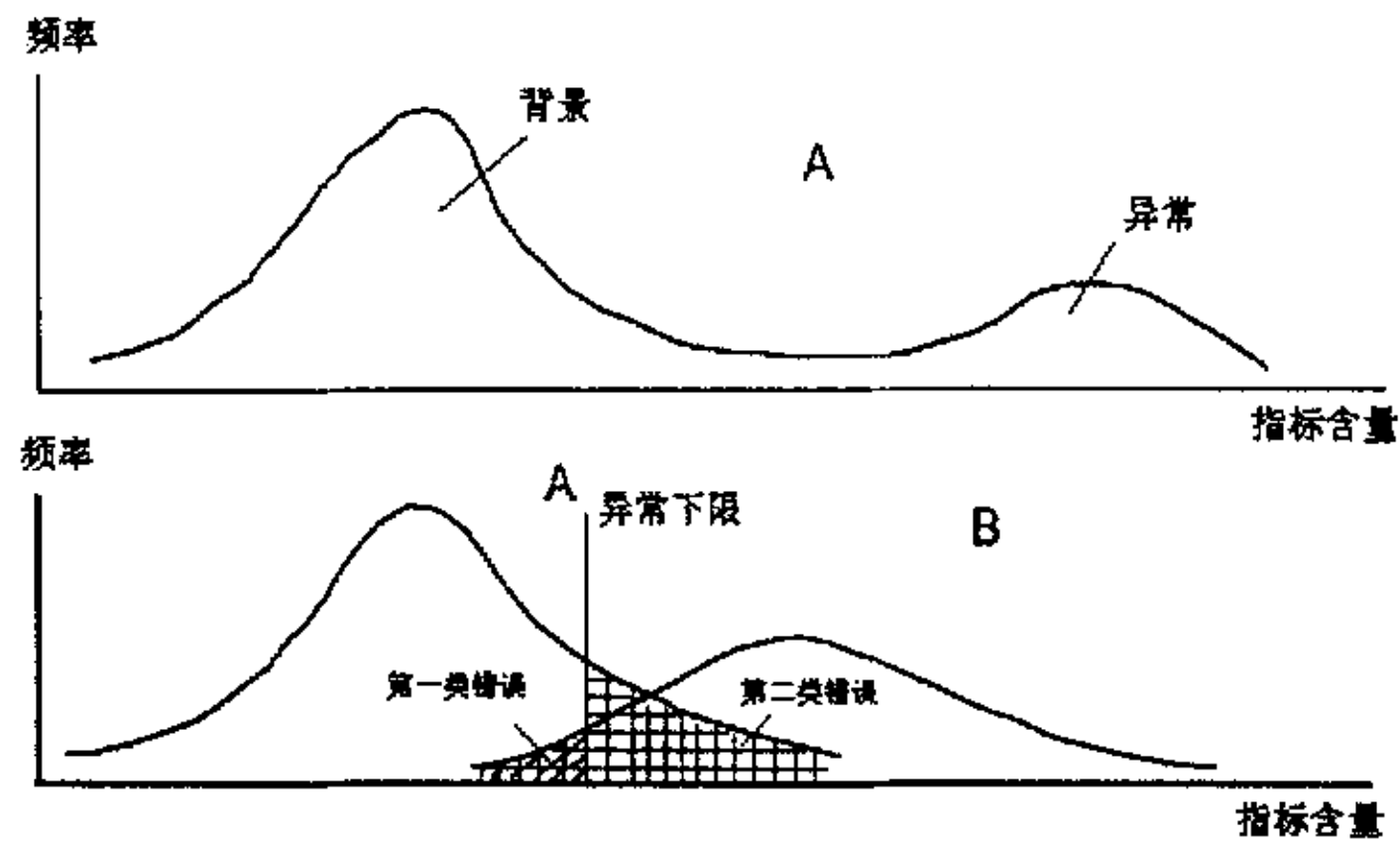


图 11 背景与异常分布

确定异常的有效数理统计方法有：判别分析、对应分析与趋势面分析。

3.4.1 判别分析

油气化探测量仅根据指标含量高低的绝对值确定异常，区分油气藏和非油气藏有一定的困难，而借用判别分析方法，处理指标的分析数据，能够为成因分类、确定异常、筛选目标，提供可靠依据。判别分析可为解决两类问题服务：一是根据已知油田的指标组合特点，判别未知区指标含量与组合同油气藏的关系；二是根据多指标特征对未知区的样品进行分类。在实际工作中主要运用二类判别和多类判别。具体运用过程简述如下：

A：二类判别

目的：根据总体 A、B 的两类样品观测值，建立用于任意判定样品 T（属于 A 或 B）所属总体的线性判别函数，代入任意预测样品可根据其函数值确定其归属。

输入：样品及变量按数据表 12 的顺序排列。二类判别中每个母体取样个数可不相等，但变量个数必须相等。输入判别效果检验显著性水平 α 。

方法：N 个已知样品分属两个母体，它们在变量空间中构成两个点群。线性函数 $Y=b_1X_1+ b_2X_2+ \dots+b_mX_m$ 这个超平面将空

表 12 二类判别分析输入数据表样式

| 类别 | 变量 | | X1 | A. X2 | | Xm |
|--------|-------|--|---------------|---------------|-------|---------------|
| | 样品 | | | | | |
| 母体 A | 1 | | $X_{1,1}^A$ | $X_{1,2}^A$ | | $X_{1,m}^A$ |
| | 2 | | $X_{2,1}^A$ | $X_{2,2}^A$ | | $X_{2,m}^A$ |
| | | | | | | |
| | n_1 | | $X_{n_1,1}^A$ | $X_{n_1,2}^A$ | | $X_{n_1,m}^A$ |
| 母体 B | 1 | | $X_{1,1}^B$ | $X_{1,2}^B$ | | $X_{1,m}^B$ |
| | 2 | | $X_{2,1}^B$ | $X_{2,2}^B$ | | $X_{2,m}^B$ |
| | ... | | | | | |
| | n_2 | | $X_{n_2,1}^B$ | $X_{n_2,2}^B$ | | $X_{n_2,m}^B$ |
| 未知母体 T | 1 | | $X_{1,1}^T$ | $X_{1,2}^T$ | | $X_{1,m}^T$ |
| | ... | | | | | |
| | n_3 | | $X_{n_3,1}^T$ | $X_{n_3,2}^T$ | | $X_{n_3,m}^T$ |

间分成两个区域，为使函数 Y 能把两类样品很好地分开，根据费歇尔准则，应使类内离差尽可能小而类间离差尽可能大，解得系数 b_1, b_2, \dots, b_m 。建立判别指标 Y_0 ，即可将任意未知的样品代入方程进行判别。然后对判别效果进行显著性检验（过程略）。

输出：判别函数方程、各母体（A 和 B）的判别函数、判别指标 Y_0 等检验用参数、两组（A 和 B）模型样

品的回代判别函数值与重新分类检验结果、待判别样品的判别分类结果、判别效果显著性检验参数及结果。

B: 多类判别

目的：对属于多个母体的一批已知样品，每类样品建立一个判别函数，预测时将待判样分别代入判别函数，比较各值大小，进行归类。

输入：输入的原始数据阵和表 12 类似，只不过母体种类可以是多种。

方法：①建立多类判别概率模型；②求取 K 类已知样品的判别函数 $y_k(\bar{x})$ ；③采用贝叶斯（Bayes）准则，确定待判样品 $\bar{x} = (x_1, x_2, \dots, x_m)'$ 可计算出不同的 $y_k(\bar{x})$ 值。 $y_k(\bar{x}) = \max\{y_l(\bar{x})\}, 1 \leq l \leq g$ ；则待判样品 \bar{x} 就属于条件概率最大判别函数的母体 A_k 类。④查自由度位 $(g-l, n-g-l)$ 的 F 分布表，得临界值 F_α ，进行判别能力显著检验。

输出：各类母体判别系数；模型样品判别分类；待判样品判别分类；检验统计量；临界值 F_α ；

例如我们在某海区进行油气化探普查时，采样 253 个，对主要指标的分析数据作了计算，获得八位函数关系式：

$$R = 0.4796(405nm) - 0.0155(425nm) - 0.1316(254nm) + 0.2330(\text{比值}) + 0.30388(228nm) + 0.13010(310/370\text{ nm}) - 0.0130(\text{酚}) - 0.004(\text{甲烷})$$

计算判别函数值为：

$R_0=56.5096$ (分界值)
 $R_1=62.0152$ (已知油区)
 $R_2=52.8352$ (已知非油气区 1)
 $R_1 > R_0 > R_2$

利用上述判别分析结果所圈定的化探异常, 经钻探验证取得较高的成功率, 在化探异常内建成油气田。图 11 是另一个油区的实际资料, 从中看出, 浅层化探指标 R 值的高值带, 基本反映了下伏含油气构造的基本轮廓。

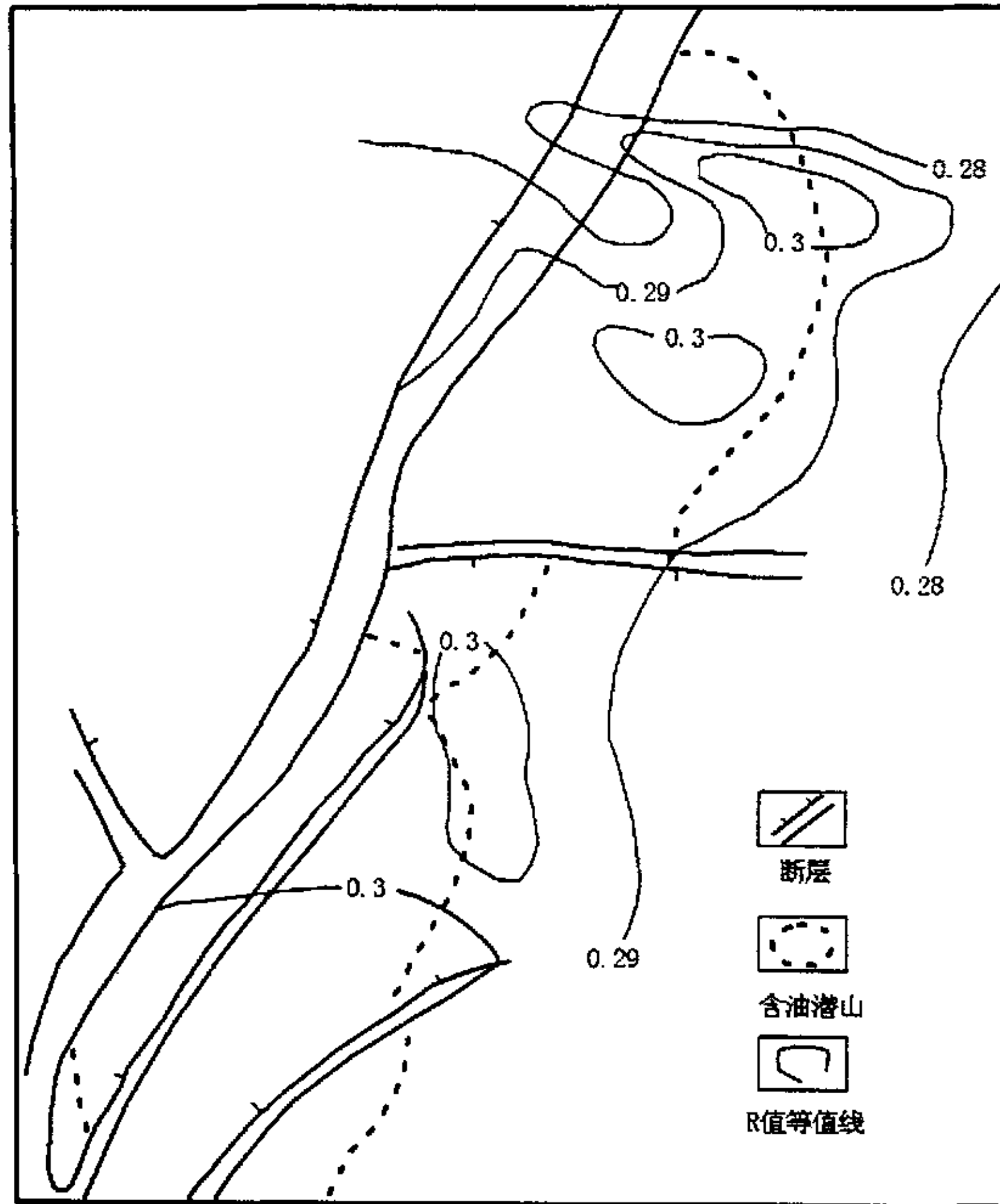


图 12 判别分析综合 R 值等直线图

3.4.2 对应分析

由法国巴黎科学院本泽西 70 年代提出, 该方法把 R 型和 Q 型因子分析结合起来, 对变量和样品一同分析, 同时进行分类、作图。在统一的因子空间内, 同时表达综合变量和统计样品, 并可由 R 型分析很容易地获得 Q 型分析结果。

主要优点和特点: A. 可将样品和变量结合起来分析, 结果能统一反映在一张图上, 便于地质解释和推断。B. 通过变量的变换, 将原始数据 x_{ij} 变成中间变量 y_{ij} , 此过程变量和样品是对称的, 于是可将 R 型和 Q 型因子分析结合起来。

C. 在实际运算中, 只进行 R 型因子分析, 经过简单的换算, 便可得到两种因子分析的结果, R 型分析的因子载荷是 Q 型分析的因子计量, 从而节省计算时间和内存。

输入:

由 n 个样品、每个样品 m 个变量的观测值 x_{ij} 组成的原始数据矩阵 $X_{m \times n}$; 在分析前要把原始数据阵标准化或正规化, 形成预处理后数据阵 $X_{m \times n}$ 。输入绘图参数。

基本原理和算法:

A. 首先把正规化后的矩阵 $X_{m \times n}$ 按下式变换为 Z 矩阵:

$$z_{ij} = \frac{x_{ij} - x_{i.}x_{.j}/T}{\sqrt{x_{i.}x_{.j}}}$$

其中: $x_{i.} = \sum_{j=1}^n x_{ij} \quad (i=1,2,\dots,m);$

$$x_{.j} = \sum_{i=1}^m x_{ij} \quad (j=1,2,\dots,n);$$

$$T = \sum_{i=1}^m \sum_{j=1}^n x_{ij} = \sum_{i=1}^m x_{i.}$$

B. R 型因子分析; 求变量协方差矩阵 $S=ZZ'$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \lambda_m$, 按其累积百分比 $\left(\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^m \lambda_i} \right) \geq 85\%$ 取前 p 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_p$, 并计算它们对应特征向量 u_1, u_2, \dots, u_p , 得到 R 型因子载荷矩阵 U 。

$$U = \begin{bmatrix} u_{11}\sqrt{\lambda_1} & u_{12}\sqrt{\lambda_2} & \dots & u_{1p}\sqrt{\lambda_p} \\ u_{21}\sqrt{\lambda_1} & u_{22}\sqrt{\lambda_2} & \dots & u_{2p}\sqrt{\lambda_p} \\ \dots & \dots & \dots & \dots \\ u_{m1}\sqrt{\lambda_1} & u_{m2}\sqrt{\lambda_2} & \dots & u_{mp}\sqrt{\lambda_p} \end{bmatrix}$$

以因子载荷为坐标, 因子作坐标轴把变量点在因子轴平面上表示出来。

C. Q 型因子分析; 在 R 型因子分析的基础上, 利用 $v_j=Z' u_j$ 求出 Q 型因子载荷 V (形如式 6-3-2)。在 R 型因子分析平面上, 以 Q 型因子载荷为坐标把样品点表示出来。

输出: R 型和 Q 型因子分析的因子载荷数据文件; 因子平面图;

因子平面图上提供的信息:

A. 变量间信息：相近的变量点表示变量紧密相关，即它们具有成因上的联系，指示某一特定的地质作用。

B. 样品间信息：相近的样品点具有相似的性质，属同一类型，是相同地质作用的产物。

C. 变量和样品的关系。同一类型的样品点将为邻近的变量点表征。也就是说，同类样品点为其邻近变量点所指示的地质作用下的产物。

图 13 是南海地区用对应分析确定异常性质的实际探例，其中 1 号异常由四个化探点组成，异常与生油指标有机碳密切相关，是烃源岩发育区的反映；2 号异常由 8 个化探点组成，它们与油气的直接组分—甲烷、乙烷、丙烷、重烃、总烃等指标相关联是油气经过运移富集成藏的结果。上述两个异常就总体而言，远离碳酸盐、二价铁、三价铁等环境指标限定的空间范围，表明异常的形成是深部油气运移的结果。

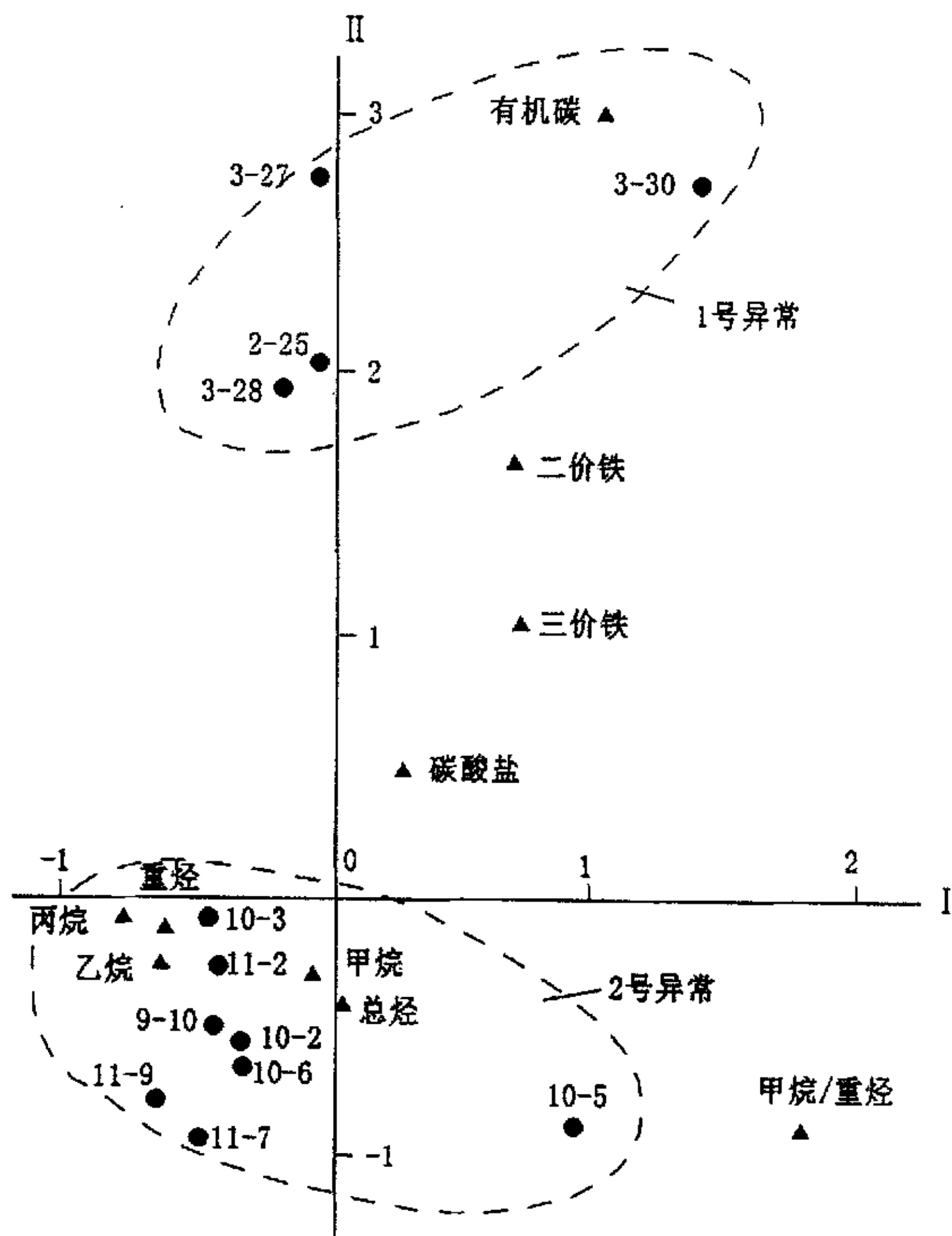


图 13 对应分析因子轴平面聚点图

3.4.3 趋势分析

它是拟合数学面的一种统计方法。油气化探调查结果是用指标含量的数据来表示的，这些数据一般都包括三部的变化，即：

区域性变化：数据中反映地球化学背景的规律性变化的部分，由含油气盆地或凹陷的区域性总因素决定的。

局部性变化：数据中反映油气信息的变化特征，受油气田或含油气圈闭等因素控制的。

随机性变化：由干扰、分析等随机因素造成的偏差。

趋势分析的目的，就是对化探数据中所包含的三部分变化进行分解，排除随机性干扰，通过掌握区域性变化特征，认识地球化学空间场的分布状况，衬托出化探异常，预测油气富集的有利地块和油气藏的位置。

具体的方法就是用数学方法计算出一定数学面来拟合数据内区域上的方向性变化，这个数学面就称为趋势面。通过趋势面方程，算出每个化探测量点上的趋势值，再按一定数值间隔勾绘出趋势面等直线图，从中了解整个区域上的变化总趋势。以趋势面为基础，求出每个化探点的剩余值（各点分析值减去趋势值）；圈出反映局部性变化规律的剩余图，进一步再去掉随机性影响，就得出反映含油气圈闭的化探异常，达到预测油气富集有利区的目的。此外，由于趋势面分析可以提供背景和异常样本，为认识和掌握区域性变化规律和局部变化特征提供了可能。

剩余值的计算公式如下：

$$\Delta Z_i = Z_i - \hat{Z}_i$$

式中： ΔZ_i —剩余值

Z_i —分析测试值

\hat{Z}_i —趋势值

引入剩余值后，原始数据即被分解为趋势值和剩余值两部分，它们各自代表化探指标变化的区域与局部特征。

二维趋势面分析设计的步骤如下：

目的：用给定的数据体，根据输入的趋势面次数，求反映变量在区域范围内的总体变化规律的各点相应的趋势值，求反映地质变量在局部范围内的变化特征的各点的剩余值，求相应趋势面的拟合度。

输入：由各点地理坐标和变量值（ x_i, y_i, z_i ）组成的数据体；趋势面的次数 k （1~7 次）；

方法：平面区域上 N 个观测点及观测值 $M_i(x_i, y_i, z_i)$ ，寻找一个函数 $Z=f(x, y)$ 使它在各点上的函数值与观测值误差都较小。给定次数后，按最小二乘原理求出函数系数。求得趋势方程后将各样品回代得到该点的函数值，它与实测值之间的误差反映了局部异常与随机因素的共同影响。设定异常下限便可分离出局部异常和背景趋势。类似回归分析原理，求总离差平方和残差平方和，进而求取趋势面拟合度，进行拟合度检验。

输出：输入的次数；趋势方程系数；趋势面的拟合度；各点的趋势面值数据；各点的剩余值数据；在计算过程中的各参数；异常下限 R 。

趋势结果的解释及应注意的问题：

A. 选择趋势次数时，一般随地质背景的复杂程度而定，当地层构造条件简单时，趋势次数较低，相反时，次数较高。如果选取的此数过低，会将地球化学背景遗留在残余信息中，失去分解的意义，而次数过高，则会造成趋势的形态不稳定，形成假象或者导致边界效应增强。

B. 选取方程次数越高，拟合度就越大；一般拟合度在 $0.4\sim 0.7$ 之间即可。

C. 异常值应该由包括异常值和随机误差的剩余值中分离，这里直接由剩余值做为异常值。

D. 异常下限 R 这里定义为剩余平均值加上剩余标准差。

E. 样品数越少、点位分布越不均、多项式次数越高，可能引起方程的病态程度就会越高，此时若系统提示方程组无解，可设法改善数据结构或降低多项式次数。

趋势面分析是在油气化探中应用比较多和行之有效的一种数理统计方法。图 14 是某油田水矿化度四次趋势面分析结果，从图中看出趋势值在 $50-70$ 的等值线所限定的异常区基本上同已发现的油田范围一致。

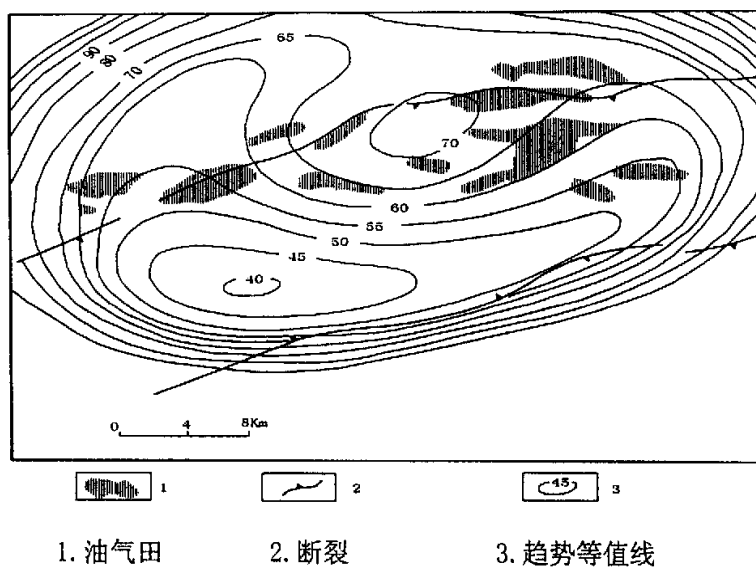


图 14 某油田水矿化度四次趋势面分析

图 15 是水化学主要指标的趋势异常，与油气田面积吻合程度很高。

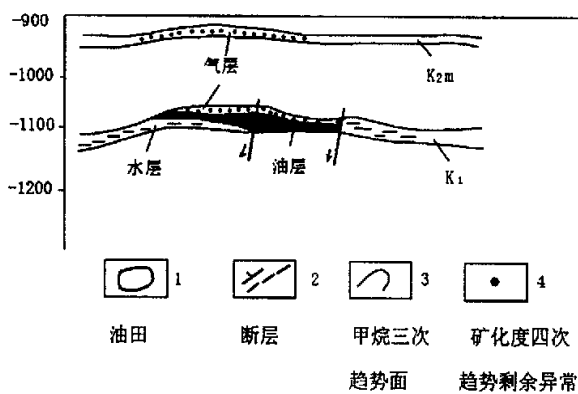
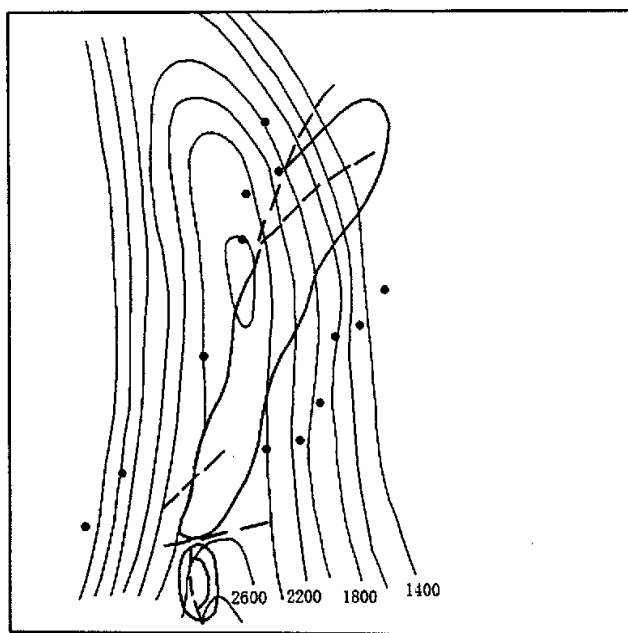


图 15 某油田水化学趋势面图

图 16 趋势面分析结果说明水化学法中酚的高含量（包括剩余值）多分布在油田或深大断裂带的上方，从一方面证实了油田的形成（或化探异常的形成）与深大断裂带的活动、油气以断裂为通道运移富集有关。

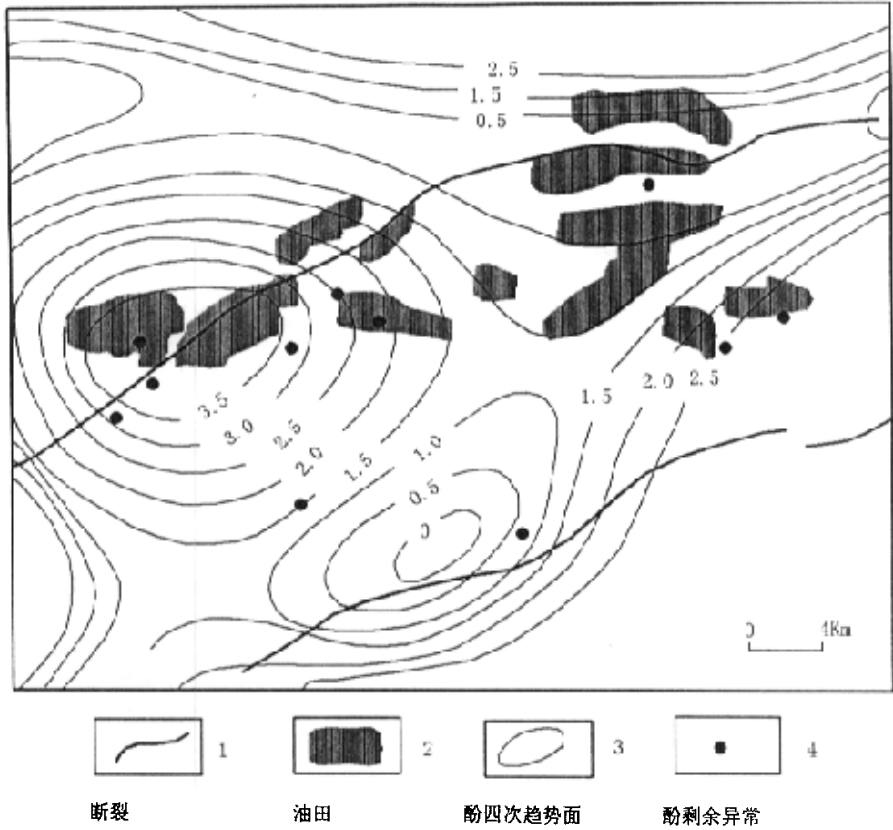


图 16 油田上方海水中酚的四次趋势面分析

图 17 是我国东海陆架盆地 (面积 $27 \times 10^4 \text{KM}^2$)。丙烷/乙烷四次趋势面等值线图。从中分析可看出, 西湖凹陷和台北凹陷中部为趋势高值带所占据, 预示着有良好的含油气远景。目前在这两个凹陷中均发现了大型天然气田和轻质油。此外, 在调查之后 1998 年东海石油公司和英国超准石油公司合作勘探, 在温州东南海域 (距温州 140KM) 的化探异常区内, 钻于天然气 (28 万 $\text{M}^3/\text{日}$), 并伴随少量凝析油。

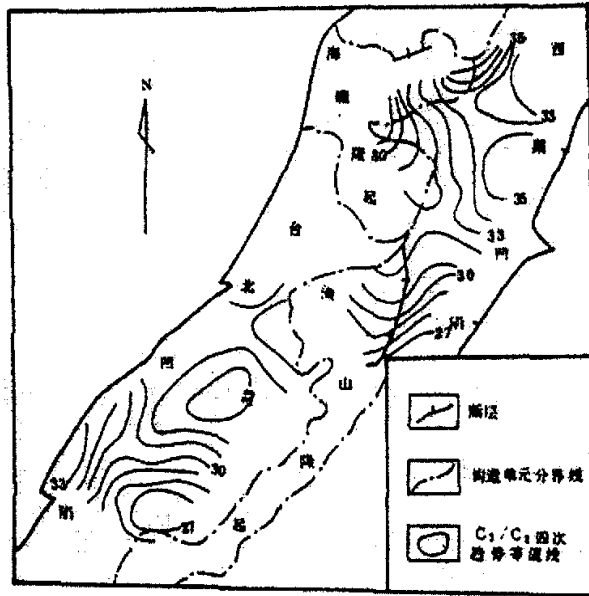


图 17 东海陆架盆地油气化探趋势面分析

典型趋势面分析：是用数学函数拟合第一组变量样本的指标与第二组变量指标的空间分布规律。其优点是能够计算单一母体的多变量空间分布的一套趋势图，叠合成综合趋势图和总剩余异常图。它的最大相关性提供了母体反应面和样本反应之间的较高信度。因此，既可以提供区域性综合异常区（带），又可提供局部综合异常。充分挖掘了埋藏在各个水化学指标中的综合油气信息，从而提高了异常的可信度。图 18 某凹陷水化学指标五次典型趋势面分析结果。它是利用水中可溶气态烃、苯、甲苯、酚、紫外吸收光谱和荧光光谱对七个指标统计而成。从图中看出，水化学异常与剩余异常点，多沿着凹陷中心呈环状分布，显示凹陷是油气富集基本地质单元的总规律。目前已建成多个油气田。

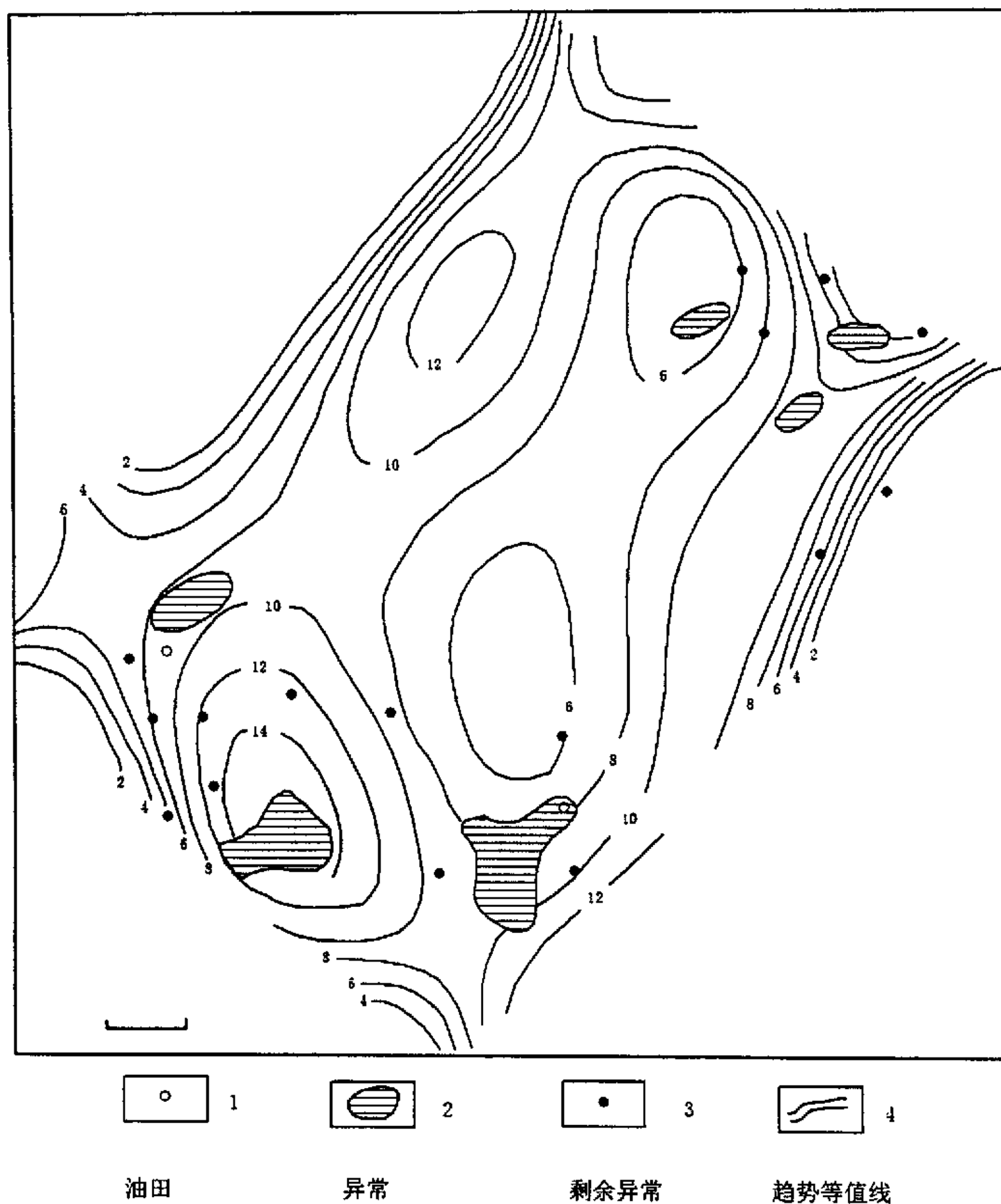


图 18 某凹陷海水中有机组分五次典型趋势面分析

3.4.4 数据插值的方法与意义

前已述及，海域油气化探测量点的布置是随机的，在平面上测点比较稀疏，各点所控制的面积较大，而且分布很不规则，均匀性较差。在趋势面分析之前，必须先进行数据插值研究，即：根据不同密度的少数的离散点值扩充为平面上密度规则分布的数值，或根据需要进行特殊点位的补充、加密等，避免边界效应的影响，达到圈定异常的最佳效果。

A. 线性插值方法：

一维线性插值在进行数据处理时是针对某一个变量，在一维坐标内的任意给定一点进行插值。针对化探数据，笔者这里设计了最常用、最通用的几种方法。

1. 拉格朗日插值 (Lagrange): 可以进行一元全区间等距或不等距插值。在指定插值点 t 前后各取四个节点, 即选取满足: $x_k < x_{k+1} < x_{k+2} < x_{k+3} < x_{k+4} < x_{k+5} < x_{k+6} < x_{k+7}$ 的八个节点 (x_i, y_i) ($i=k, \dots, k+7$), 用七次拉格朗日插值公式计算指定插值点 t 处的函数近似值, 即:

$$z = \sum_{i=k}^{k+7} y_i \prod_{j=k, j \neq i}^{k+7} [(t - x_j) / (x_i - x_j)]$$

这里的 y_i 为某个变量的观测值, x_i 为该变量每个点位的 x 或 y 坐标值。

2. 埃特金逐步插值 (Aitken): 给定 n 个等距或不等距节点 $X_1 < X_2 < \dots < X_{n-1} < X_n$ 及其相应函数值 $Y_1, Y_2, \dots, Y_{n-1}, Y_n$ 。选取最靠近插值点 t 、且尽量使 t 位于其中心的若干节点进行插值。Aitken 插值公式为:

$$P_{k,k+1}(t) = P_{k-1,k}(t) + [(t - x_k) / (x_k - x_{k+1})] \cdot [P_{k-1,k}(t) - P_{k-1,k+1}(t)]$$

3. 阿克玛插值 (Akima): 即我们常说的光滑插值。设给定 n 个等距或不等距节点 $X_1 < X_2 < \dots < X_{n-1} < X_n$ 及其相应函数值 $Y_1, Y_2, \dots, Y_{n-1}, Y_n$ 。用阿克玛方法构造指定子区间上的三次插值多项式与计算指定插值点处的函数值。即若在子区间 $[X_k, X_{k+1}]$ ($k=1, 2, \dots, n-1$) 上的两端处有以下四个条件: $Y_k = f(X_k)$; $Y_{k+1} = f(X_{k+1})$; $Y'_k = g_k$; $Y'_{k+1} = g_{k+1}$; 则可以唯一确定一个三次多项式

$$Y(x) = a + b(x-x_1) + c(x-x_1)^2 + d(x-x_1)^3$$

然后就用该多项式计算该子区间中的插值点 t 处的函数值 $Y(t)$, 具体算法略。

4. 抛物线插值: 可以进行一元三点等距或不等距插值。即给定 n 个等距或不等距节点 $X_1 < X_2 < \dots < X_{n-1} < X_n$ 及其相应函数值 $Y_1, Y_2, \dots, Y_{n-1}, Y_n$ 。选取最靠近插值点 t 的三个节点, 用抛物线公式计算 t 点的函数值。抛物线插值需事先已知抛物线公式。

5. 埃米尔特插值 (Hermite): 根据给定节点上的函数值与一阶导数值, 用埃米尔特插值多项式近似代替函数值, 而指定插值点上的函数值 (具体算法略)。

6. 三次样条插值: 根据给定节点上的函数值及边界条件, 用三次样条函数计算各节点上的数值导数与积分, 并对一元函数进行成组插值与成组微商。根据原始数据提供的端点导数阶数, 边界条件有三种 (具体算法略)。

根据数据特点, 在地质数据中, 特别是油气化探中, 由于前三种插值方法具有不需要事先曲线方程而根据近邻点即可插值的优点, 应用比较广泛。

在本次“863”项目的研究中，由于取样器的不同，各个点位样品的取样深度不一。要在平面上插值、横向比较、解释之前，必须在纵向上进行深度统一，取绝大多数点位的底界面（沉积物 1.5 米，水 60 米）深度做为统一深度。这就需要对全区在上述深度内无分析值的点位的纵向上进行线性插值。笔者应用了 Aitken 插值，其结果和由拟合直线求取的数值相差很小。

B. 二维平面插值方法：

二维平面插值，与计算机绘图过程中网格化处理既有一致性，又有其特征：

(1) 两者数据插值的范围都在数据体 x, y 坐标最大值范围内有效。

(2) 化探数据平面插值的插值点位置、数量在范围内需要有随意性，用户可自行定义。而网格化一般只能是规则网格，用户只能定义其网格密度。

(3) 化探数据平面插值由于每个数据点都有地质意义，所以一般进行参与插值的运算点并非越多越好，一般用邻近点。而普通的网格化可以进行全区数据参与运算。

(4) 化探数据平面插值往往需要根据数据点的稀疏关系、分布均匀性、和地质构造的关系进行分区域不同加权、不同密度的插值。

(5) 在化探数据平面插值形成较为规整的数据体之后，可以直接应用各种搜索法进行等值线图、趋势图等图件的绘制。也可再通过普通网格化各种算法，规则加密数据，再进行绘图。

针对这些特征，这里主要设计了分象限近距倒数加权法。该方法是以要插值的 P 点为中心，将插值区域分为若干个象限，在每个象限中选择距离 P 点最近的 t 个原始数据点参加插值计算。根据其距离插值点的远近，用距离倒数进行加权。这样可以充分考虑区域的地质结构，根据地质构造的复杂程度及点位的稀疏、均匀性确定参加计算的点位。

该方法最为关键的一步是如何确定参加运算的点位及个数。

具体参与计算的原始点的确定以图 19 示例。图中圆点代表随机的原始数据点。插值的有效范围限定在原始数据 x, y 坐标的极大值与极小值之间（即四个顶点组成的矩形区域内）。在矩形区域，根据插值位置随机性、不失一般性，插值点可以分为两类、四种。第一类为边界点，包括四个顶点和此外的四条边上的点。四个顶点（A 类点）在象限划分时只有一个象限，用最小距离的 3 个原始点参加插值计算。四条边上的点（除 A 类点）称 B 类点，插值时只有两个象限。第二类为内部点，其有四个象限，包括图中的 C 类点和 D 类点。C 类点的四个象限内都有原始数据点分布。D 类点四个象限中有三个象限有原始数据点。

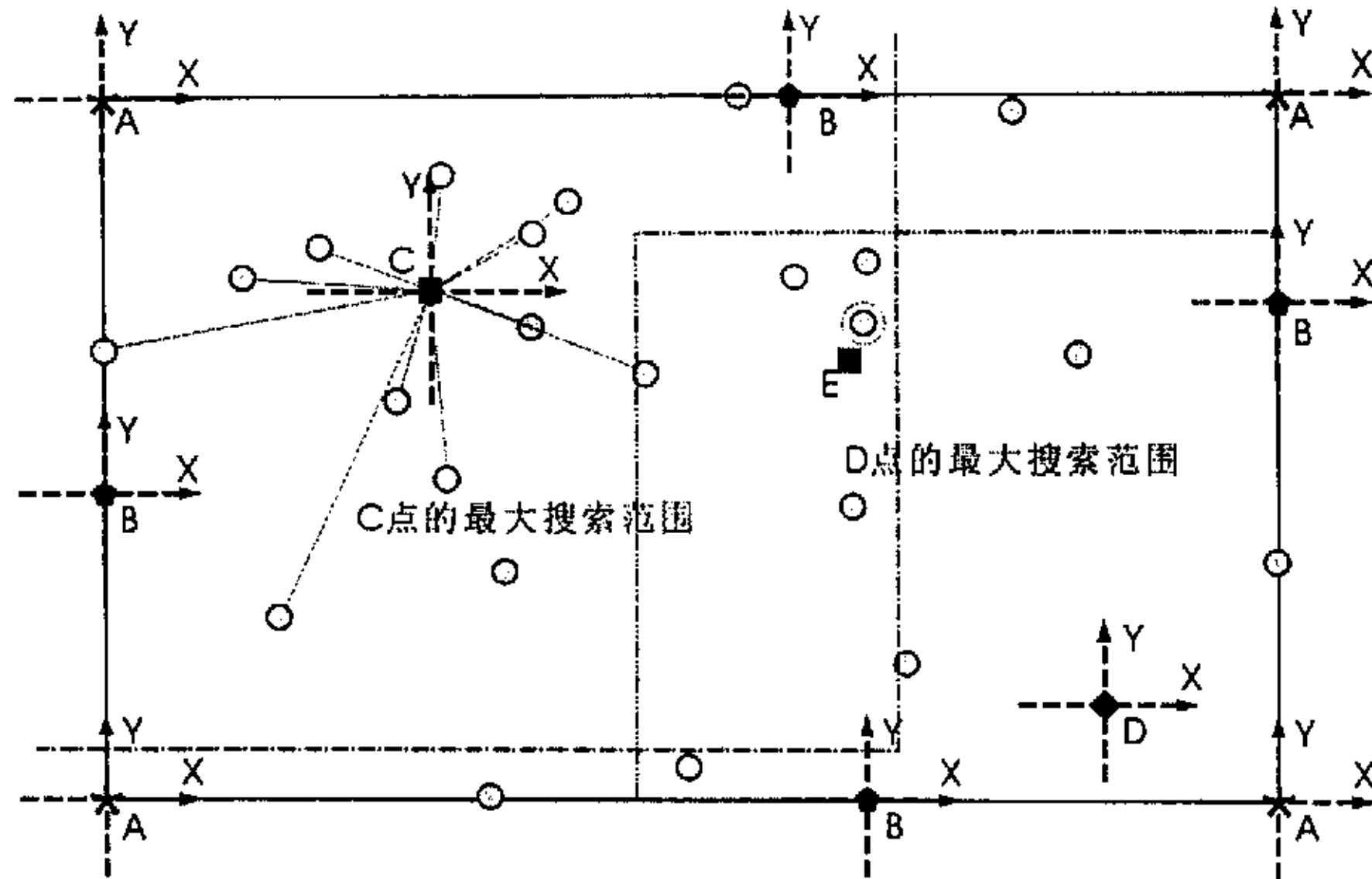


图 19 分象限近距倒数加权法图示

首先定义一个最大搜索范围： $R_0 = k \cdot \{1/20 \cdot [(max_x + min_x) - (max_y + min_y)]\}$ 。根据点的稀疏和搜索的半径大小的需要，可以自行定义系数 k ，范围 $[1, 5]$ ， k 越大，搜索范围越大。一般取值 3.5 左右。寻找参与插值计算的原始点时，在以插值点为中心、以定义的最大搜索范围为半径的正方形区域内每个象限找有 3 个最小距离点参加计算，若某个象限原始数据点数 $k \leq 3$ 时，则 k 个点全部参加计算。具体的计算方法如图 20。

计算插值点的 Z 值公式为：

$$z = \frac{\sum_{i=1}^n (z_i / D_i^2)}{\sum_{i=1}^n D_i^2} ; D_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

该方法的有以下优点：

(1) 可以根据点位的稀疏自己定义搜索范围，尽量保证区域内的点每个象限都有数据点参加运算。从而保证所使用原始数据点在方向上的均匀性，以消除原始数据点分布的不均匀性而造成的方向性干扰。

(2) 保证最少有 3 个、最多 12 个近距离点参加计算，这样避免全点参加计算，减少计算次数，也有利于更加符合实际地质情况。

(3) 在流程中输入插值点可以输入一个单点，也可以是一个数组（外加一层循环即可）。这样比普通插值更具方便性和根据地质情况的随意操作性。

(4) 可以根据点的稀疏和分布规则程度，对不同区域的插值点定义不同的搜索半径，起到分区域加权、分区域规定插值密度的作用。这样可以根据地质构造情况，在各地质构造单元内按不同要求进行定义插值密度、插值点分布。

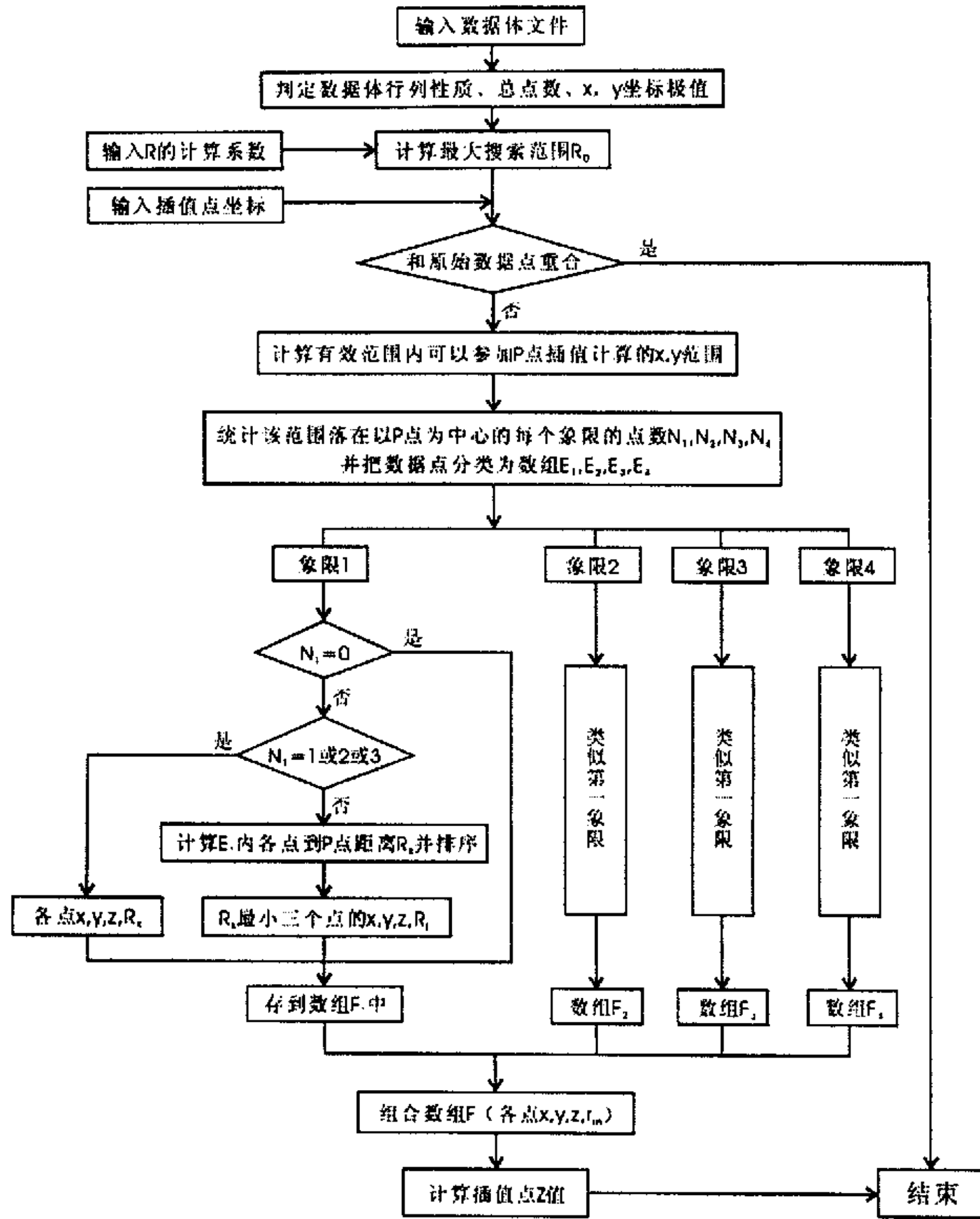


图 20 分象限近距倒数加权法流程简图

3.5 综合异常定量评价的方法

综合异常是油气化探的最终成果，综合异常信息提取技术及解释评价是化探工作者不断探索的重要课题。长期以来沿用着传统的思维方式，对定性评价方法做了大量的研究。定性评价方法主要根据指标含量的高低与各指标异常平面叠合的程度作为异常成因解释及含油气性评价的依据，其优点是简单、快速，但也存在着许多不足，主要有：

3.5.1 各指标绝对含量的高低，包含着许多不确定因素，由于背景值确定方法的不统一和人为性，给不同指标异常之间的对比和评价带来一定的困难；

3.5.2 油气化探一般应用多方法、多指标，达到从不同介质，多层次提取油气信息的目的。由此带来指标间量纲不同，计量单位有别，制约了多指标综合

信息的提取和应用；

3.5.3 定性解释与评价很多方面基于工作者的经验,不可避免的会产生异常的多解性。

针对上述定性解释评价的不足,作者及其同事们在实践中不断探索定量评价的方法,运用“主指标”的衬度组合法提取综合异常的油气信息,进行解释与评价。其基本含义是:在研究指标地球化学性质及其成因的基础上,通过前述优选指标的数据处理方法,筛选出反映油气信息的主要指标组合,作为确定异常和评价异常的主因素,应用下列数学公式将各指标有因次(量纲)的浓度值,变为无因次的评价指数,计算过程如下:

A. 单指标的评价指数:

$$P_i = C_i / C_{si}$$

式中: P_i —单指标的评价指数;

C_i —某指标的实测浓度;

C_{si} —某指标的平均值。

B. 多指标综合评价指数:

$$PI = \sqrt{\frac{(pi)_{\max}^2 + (pi)_{\text{avg}}^2}{2}}$$

PI—多指标综合评价指数

$P_{i \max}$ —各指标中最大的 P_i 值

$P_{i \text{ avg}}$ —各指标中平均的 P_i 值

利用上述公式计算结果,对综合异常的含油气远景进行评价的优点是:在各个主要指标单独评价的基础上,综合应用多指标进行评价,并给出一个量值的概念,根据综合评价指数的大小,划分异常级别具有定量评价的性质;突出了主要指标在异常评价中的作用,直观地表现了各指标在异常评价中的优势程度和顺序,为异常解释评价提供了依据。

例如在我国东海进行的油气化探测量,选用了水化学和沉积物烃类等方法,筛选出7个主要指标,确定了6个综合异常,应用综合评价指数对异常进行定量评价,其结果如表13所示,综合评价指数大于10者为I级综合异常,综合评价指数4—10者为II级综合异常,综合评价指数小于4者为III级综合异常。后经勘探实践证明,这些异常级别的划分与油气富集程度基本吻合,目前,I级综合异常区已突破出油关是油气勘探的重点区带。

表 13 油气化探综合异常评价参数

| 异常名称 | 水化学 | | | 沉积物 | | | | PI 值 | 级别 |
|------|------------------|------------------|------------------|-----|------|-----|------|------|-----|
| | U ₂₂₂ | U ₂₅₄ | F ₄₀₅ | 烃 | 重烃 | 湿度 | 甲烷 | | |
| A | 15.3 | 13.7 | 15.5 | 0.5 | 11.8 | 2.4 | 11.9 | 13.1 | I |
| B | 2.8 | 0.6 | 7.1 | 1.8 | 0.2 | 1.9 | 0.3 | 5.2 | II |
| C | 1.4 | 0.3 | 3.9 | 1.7 | 4.3 | 1.9 | 5.3 | 4.2 | II |
| D | 0.2 | 2.2 | 0.1 | 2.8 | 5.7 | 1.9 | 5.5 | 5.1 | II |
| E | 1.6 | 0.1 | 3.2 | 1.6 | 0.4 | 1.1 | 0.7 | 2.5 | III |
| F | 2.2 | 0.1 | 1.0 | 1.3 | 0.9 | 1.7 | 1.1 | 1.8 | III |

注：各指标为衬度值（含量/均值）

3.5 小结

综上所述，论文将多元统计方法具体的运用到海域油气化探工作中，我的研究是属于创新的内容有以下四个方面：

1. 在海域油气化探中，如何选用合适的数理统计方法，对庞杂的分析数据和地质观测资料整行整理、分析、综合、研究，找出它们的数量特征和内在统计规律，并用于实践中，准确、可靠的预测油气藏，这是论文探索数据处理方法的研究目标。前已述及，我国海域化探走过几十年的里程，在数据处理方面，只在指标提取阶段用了个别方法，基本上属于零散应用的初级阶段，没有形成一个系统有效的数据处理标准程序与系统。论文的创新之一在于结合我国海域油气化探的特点，提出了从分析数据、地质观测参数等原始资料的可利用程度和合成开始，按照化探工作流程及评价程序，分别建立了具有共同理论基础的不同的处理方法，即：优选指标组合、抑制干扰因素，确定异常及定量评价异常等子系统，它们既是独立的，又是环环相扣，缺一不可的数据处理方法。它的系统性、完整性及实用性在国内外海域油气化探文献中至今未见报道。

2. 提出了数据使用价值分析的多种方法，将前人认为不能参加运算的定性参数（如沉积颜色差异、粒度分级等）赋予定量化数值；将不同含量单位、量纲及数值大小的定量数据标准化后，均可直接参加运算，增强了用于定量研究的可能性，提高了数据的应用范围。

3. 开拓了定量校正干扰因素的处理方法。在这之前只有个别化探工作者对干扰进行过定性研究。例如岩性对指标含量的影响，一般不予考虑，常用主岩性代替全部岩性进行研究；也有的学者采用对岩性进行分类的定性校正方法。而作者提出的定量校正方法，是将全部样品的岩性校正到均一粒度或相同岩性的背景上，求取异常，使化探成果更加科学、合理。图 21、22 是同一油田上不同处理方法的结果，从中清楚看出定量校正方法提高了化探预测的准确性。

4. 建立了量化确定与评价异常方法。提出了用判别函数值等量化异常；用

综合评价指数划分异常级别，进行分级评价。这些技术在油气化探中首次提出，是带有方向性研究课题，并且在实践中展现了良好的应用效果。

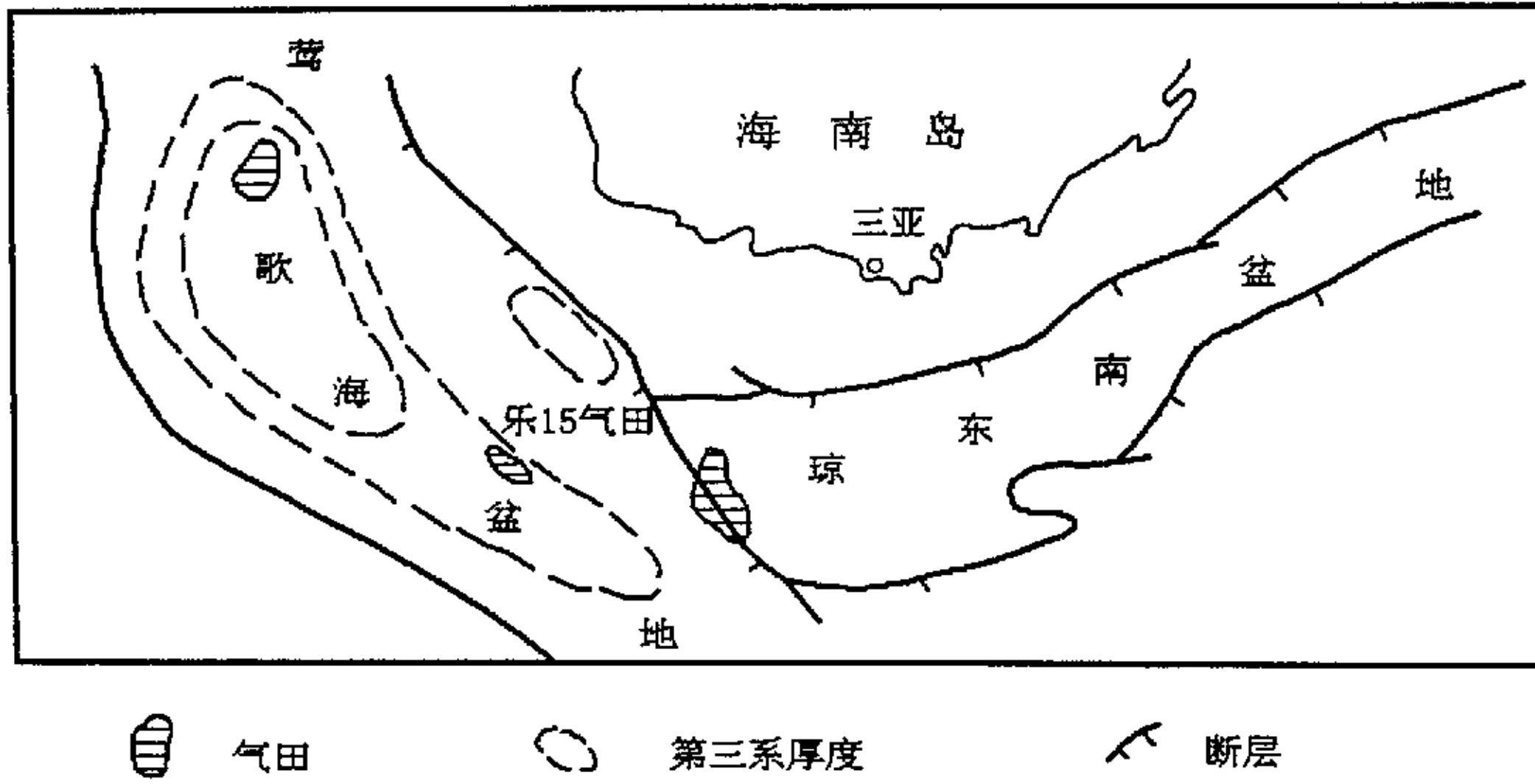


图21 气田位置示意图

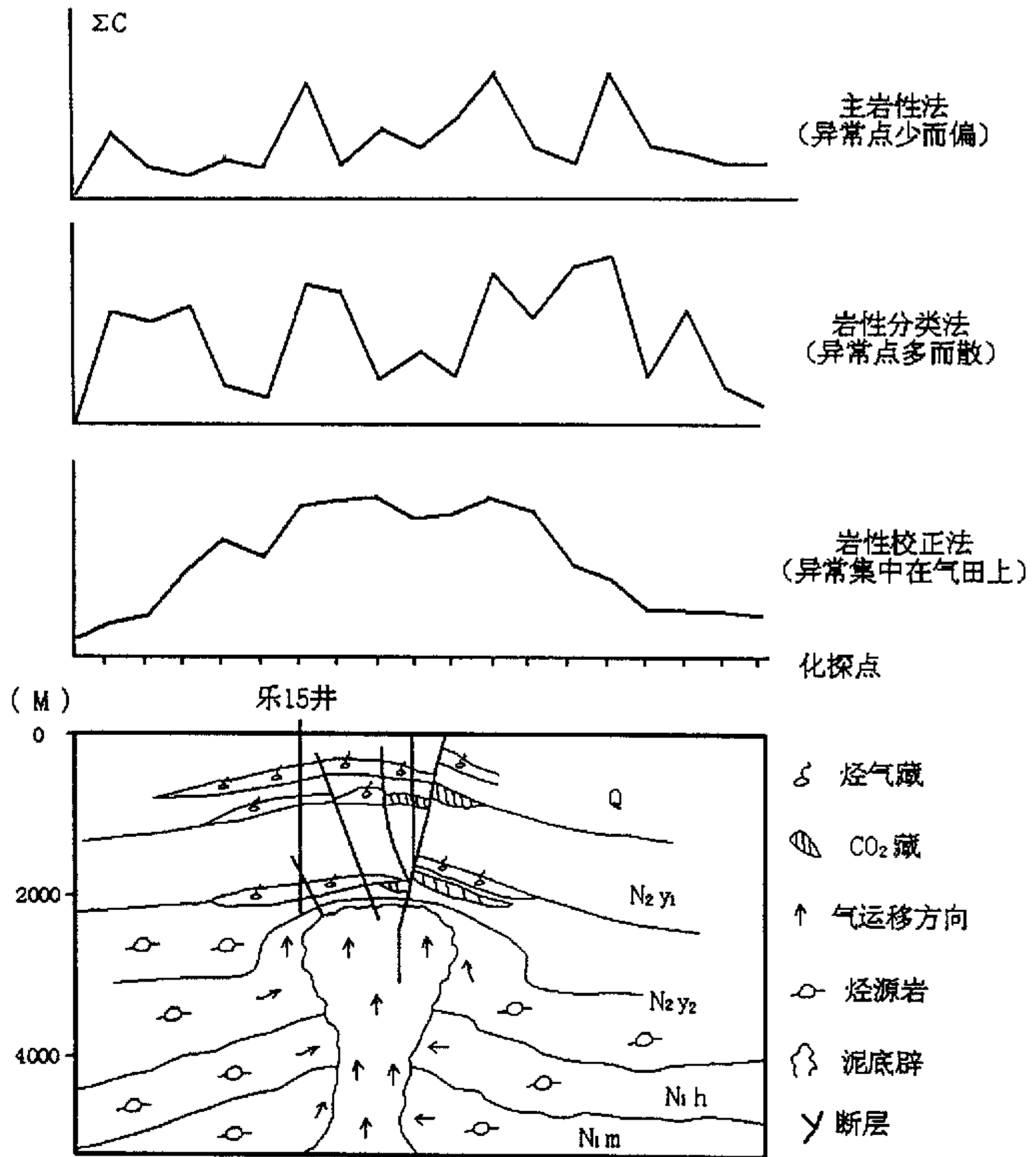


图22 岩性定量校正前后异常对比图

第四章 界面设计与数据输出

海域油气化探多参数、多约束条件数据处理的初步研究，用 windows 界面进行了简单的界面设计。

4.1 界面设计

采用 Fortran90 语言、Visual Fortran6.0 开发平台编写程序，全汉化界面，中文菜单显示，在关键步聚系统都给出了操作提示。

界面包括菜单和对话框两部分。菜单包括一级和二级菜单（图 23）。针对本次设计功能主要是数据处理，所以每个菜单项都给出相应对话框，该对话框主要是包括程序运行的各输入项、输出项及必要的说明，使操作者明白数据体要求和参数设置方法（图 24）。

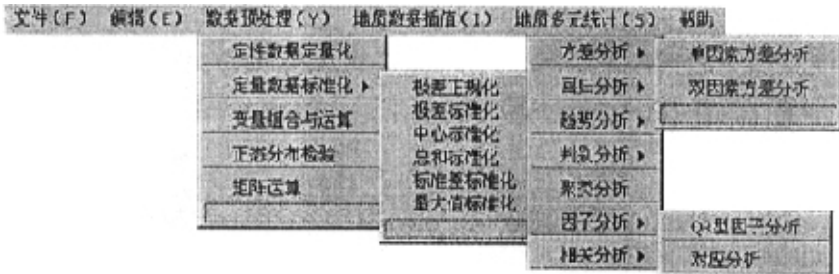


图 23 界面初步设计



图 24 平面插值对话框设计

4.2 数据输入、输出

数据文件事先统一放到指定的目录中。数据的输入都统一存为*.dat 格式，样品按行排列，变量按列排列。在数组之中，不能有 Na 值，即空白的数据。

数据文件的输出，若其为纯数据文件，为了以后作图方便，直接存为*.dat 格式。有汉字或有图形存在的文件可以存为*.txt 格式或*.doc 格式。文件放到

和输入目录相同的固定目录中。

4.3 小结

本章主要用 windows 界面进行了简单的界面设计，采用 Fortran90 语言、Visual Fortran6.0 开发平台编写程序，全汉化界面，中文菜单显示，在关键步骤系统都给出了操作提示。

第五章 总结与展望

化探是寻找石油与天然气的直接方法，由于数理统计学的引入和计算机技术的飞跃发展，化探方法发生了深刻的变化，在油气勘查中的作用引起人们的关注和青睐。

5.1 论文在前人多元统计分析及数学地质等研究成果的基础上，结合我国近 30 年的海域油气化探的实践资料、勘探现状、基本特点及存在的问题，深入研究了在多参数、多约束条件下最优提取油气信息的数据处理技术。它是当今海域油气化探中比较系统和先进的方法，具有功能齐全、方法丰富、重点突出、实用性强的特点，能从繁杂多变的数据中提取和优化指标组合，有效地抑制干扰因素，提供圈定异常的可靠方法，达到定量评价异常的目的，使化探找油技术走向新的平台。该数据处理技术可以融于物探、地质信息的二次叠加处理和综合解释，提高多种方法一体化预测评价油气资源的能力。

5.2 该技术由数据预处理、多元统计、预测评价、图件绘制和系统帮助 5 个功能及一个数据输入界面组成，可在微机内存 64M 内及 WIN98/2000 操作系统中进行。在上述允许范围内，可实现对任意多的化探方法指标和任意多的样本数进行计算，软、硬件环境具有广泛的适用性。对能够通过自动输入的数据（如指标数和样本数）可自动录入，也可自动给出有关参数的参考输入，能方便灵活完成操作。

5.3 论文结合实际探例和海域已有的分析测试资料，具体应用和说明了作者提出的数据处理技术的可靠性，尤其是在多参数条件下，不同提取指标方法的相互映照与对比，校正非油气因素干扰的技术，定量评价异常及划分异常级别的技术等，具有较高的应用价值。提高了化探技术方法预测油气藏的能力。并在生产实践中取得良好的应用效果，展现了一定的推广应用价值。

5.4 由于调查船的租用费很高，海域油气化探工作多伴是同其它地质调查同时进行，受工作环境与条件的限制，采样点布置的均匀程度差或呈非面积性的剖面，给化探的最终成图及区域的面积评价带来很大的困难。针对这一难题，作者采用了内插值与框架滤波等方法填补空缺域，经处理做到了一定调查精度下的面积性区域评价。

综上所述，论文所建立的最优提取信息的数据处理技术是根据海域油气化探特点建立的，具有鲜明的专属性，使海洋油气化探向深度和广度扩展创造了条件，满足了生产需要，受到勘探部门的好评。

本文的主要创新成果如下：

1. 针对海域油气化探特点的随机性，同一点不同介质样品（大气、水体和沉积物）分析数据繁杂的特点，建立了数据预处理系统，为指标优化和提取可靠油气信息奠定了基础。

2. 在我国首次提出抑制与校正近代生物地球化学作用干扰的技术方法。

3. 开拓了定量确定异常和评价异常的数据处理方法。

今后应进一步深化该成果，探索地质、地球物理、地球化学一体化的数据处理方法与模型，形成既有综合性又有化探自身特点的技术方法系列。

研究生期间参加的课题及发表文章

一、参加的课题

1. 海域油气化探综合物探、化探技术研究
2. 船载地球化学快速测试系统研制与应用研究
3. 胜利王庄油田岩石物性三维空间数据体研究
4. 花沟地区 CO₂ 气化探精查技术及综合研究
5. 胜利油田东营凹陷重磁电震联合反演
6. 塔西南-巴楚地区综合物化探研究及勘探方向预测

二、发表的文章

1. 我国重质油的三维荧光特征及其地质意义 《物探与化探》2004. 4
2. CAD 创新设计技术及应用 《组合机床与自动化加工技术》已录用
3. 计算机挖掘技术在海洋油气化探数据处理中的应用 《电脑应用技术》已录用

参考文献

- [1] 刘崇禧: 石油天然气地球化学勘探技术规范, 北京: 中国标准出版社, 1997年
- [2] 程军等: 我国海域油气化探的基本特点, 海洋地质与第四纪地质 2002年, Vol.22 No1 P83-87
- [3] 崔秀荣: 我国渤海—东海表层沉积物中吸附烃的特征, 海洋地质与第四纪特征 1983年, Vol.3 No5 P46-53
- [4] 刘崇禧: 南黄海水中可溶气态烃的分布特征, 海洋科学, 1982年 No4
- [5] 刘崇禧等: 中国油气化探四十年, 北京: 地质出版社, 2001年
- [6] 吴锡生: 化探数据处理方法, 北京: 地质出版社, 1993年
- [7] 郑建禄、张惠等: 南海北部海域表层沉积物中的吸附烃的特征, 热带海洋, 1994年, Vol.13 No2 P93-98
- [8] 祁士华等: 地表地球化学异常的解释与评价, 地质科技情报, 1996年, Vol.15 No3 P49-52
- [9] 程军等: 油气垂向微运移证据及特点, 石油天然气地质, 2000年, Vol.21, No.3, 237~240,
- [10] 刘崇禧等: 油气化探存在的主要问题和对其应用发展的建议, 石油实验地质, 1997年, Vol.20 No.2. P200-203
- [11] 周定成等: 台湾海峡西部海底表层沉积物烃类地球化学特征 台湾海峡, 1997年, Vol.16 No4
- [12] 姚天任、孙洪等: 现代数字信号处理, 北京: 华中理工大学出版社, 1999年
- [13] 蔡崇文: 数学地质的方法与应用, 北京: 冶金工业出版社, 1980年
- [14] 赵鹏大等: 地质勘探中的统计分析, 武汉: 中国地质大学, 1990年
- [15] 孙洪泉等: 地质统计学与应用, 北京: 中国矿业大学出版社, 1998年
- [16] 赵永军等编: 石油数学地质, 北京: 石油工业出版社, 1998年
- [17] 刘安洲: 油气资源多元信息统计预测, 长春: 吉林科学技术出版社, 1991年
- [18] 林玉祥等: 论油气化探数据处理方法与流程, 物探化探计算技术, 1998年 Vol.20 No.3. P274-276
- [19] 张建培等: 海洋油气地球化学勘探—以东海某研究区为例, 石油实验地质, 1997年, No.4. P332-335

- [20] 李武等: 海洋油气化探干扰因素及抑制方法研究, 石油勘探与开发 1995年 ,No1.22 No.3. P51-55
- [21] 吴莹等: 中国海洋有机地球化学研究的若干进展, 地球科学进展, 1999年 No.3. P256-259
- [22] 汤玉平等: 中国主要含油气盆地油气化探数据库及应用, 物探化探计算技术, 2000年 No.4. P296-299
- [23] 侯景儒、黄竞先编译: 地质统计学的理论与方法, 北京: 地质出版社, 1990年
- [24] 罗啸泉: 趋势分析在油气勘探中的应用, 成都理工学院学报增刊, 1996年 No.2. P8-11
- [25] 谢志鹏、刘宗田: 缺值背景中的概念分析与知识获取, 计算机科学, 2000年 Vol.27,No.9, P36~39,
- [26] 王实、高文: 数据挖掘中的聚类方法, 计算机科学, 2000年 Vol.27,No.4, P42~45,
- [27] 刘明吉, 王秀峰等, 数据挖掘中的数据预处理, 计算机科学, 2000年 Vol.27,No.4, P54~57
- [28] 王仁锋、胡光道: 线性地质统计学, 北京: 地质出版社, 1988年
- [29] 陈汉宗等: 南海海域油气藏上方汞异常特征分析, 中国海上油气(地质), 1997年 Vol.11,No.5, P313~319
- [30] 陈汉宗等: 台西盆地某坳陷地球化学异常特征与评价, 热带海洋, 1995年 Vol.14,No.4, P43~50
- [38] 李双林等: 南黄海北部盆地次沉积物烃类地球化学探测, 海洋地质与第四纪特征, 2003年 ,Vol.23 No4 P39-44
- [39] Zhang Jianpei, Wang Fei, 1995, The methods and applications of hydrocarbon geochemical prospecting in the East sea: Proceedings of International Symposium on Petroleum Geology in the East China Sea, 201-206
- [40] Xu Zhong Xiang, et al, 1996; The Sliding grey relational method for delineating regions containing oil and gas; The Journal of Grey System, Vol.8 No.3