

对数比统计分析及粒度数据中沉积水动力环境信息的萃取

周 蒂

(中国科学院南海海洋研究所, 广州 510301)

提 要 沉积物粒度研究是沉积水动力环境研究的重要手段之一。由于闭合效应的影响, 经典统计方法难以揭示粒度数据的真实特征。对数比统计分析是克服闭合效应的有效途径。本文介绍了波罗的海西部和南海北部的两个研究实例, 在原始数据的经典统计分析不能揭示样品的自然分类的情况下, 以粒度数据的对数比变换为基础的 RQ 型主分量分析揭示出各沉积物类型之间的明显差异, 而根据这种沉积物分类进行的沉积分区清晰地反映了沉积水动力环境的空间变化特征。

关键词 对数比统计分析 RQ 型主分量分析 沉积物粒度数据 沉积水动力环境

第一作者简介 周蒂女 52 岁 研究员 (博士) 数学地质、海洋地质和地球化学

不同粒级沉积物的搬运和沉淀受沉积水动力条件所控制, 因此实测的沉积物粒度数据反过来可作为沉积水动力环境研究的重要依据。沉积学研究从第一步起 (即沉积物粒度分类及沉积分区) 就贯穿着体现沉积水动力环境因素的思想, 力求揭示在不同水动力环境下形成的自然分类和分区, 而不是按人为给定的某种标准来进行主观分类和分区。因此, 自计算机普及以来, 采用多元统计方法来辅助沉积物粒度数据研究便成了自然的选择。

但是, 粒度数据是一种成分数据, 不可避免地要受到闭合效应的影响^[1]。本文将用两个研究现代海洋沉积的实例表明闭合效应有可能强烈掩盖沉积物的自然分类和分区特征, 而采用对数比统计方法则是克服这一困难的有效途径。

1 成分数据的闭合效应及对数比统计分析

成分数据指每行 (代表一个样品) 中各变量 (代表各成分) 的值都非负且和为 1 的这类数据。由于各成分的和为常数, 至少有两个成分之间会显出负相关关系, 而不论这些成分间的真实相关性如何。由于成分数据是一种综合指标, 其引入可带来符号可正可负的所谓“伪相关”。此外, 由于成分的取值在 $[0, 1]$ 范围之内, 成分数据不符合正态分布, 也就不满足经典统计方法的基本假设。这三个问题统称闭合效应。大量实例表明, 不考虑闭合效应的影响, 盲目采用经典统计方法来处理成分数据有可能导致谬误的结果。

Aitchison^[2] 提出用对数比变换来避免闭合效应的问题, 设计了一整套统计方法, 可称为对数比统计方法。简单来说, 对于含 D 种成分的某成分数据

$$X = [x_j], i=1, 2, \dots, n; j=1, 2, \dots, D$$

其对数比变换为

$$Y = [y_{ij}] = [\log (x_{ij}/x_{iD})], i=1, 2, \dots, n; j=1, 2, \dots, D-1$$

而中心化对数比变换为

$$Z = [Z_{ij}] = [\log (x_{ij}) / g (x_i)]$$

其中 $g (x_i) = (x_{i1} \cdot x_{i2} \cdot \dots \cdot x_{iD})^{1/D}; i=1, 2, \dots, n; j=1, 2, \dots, D$ 。

可证明上述两种对数比变换不受闭合效应的影响,因而能代表成分间的真实关系。此外,对数比变换取值范围 $(-\infty, \infty)$,且经常符合多元正态分布,因而可应用经典统计方法(参阅〔1〕)。

2 波罗的海达斯浅滩沉积水动力环境研究

波罗的海是北欧的一个半封闭陆间海。达斯浅滩为其西南部狭窄出口的东段,宽仅 10—12 km,水深平均 16 m,最深(卡得水道) 32 m。波罗的海与北海的海水交换有 73% 经过达斯浅滩,因而其沉积水动力环境研究有重要意义。

在达斯浅滩及其东约 2000 km² 海区采得 1340 个海底沉积物样,经筛分获得八个粒级,即砾 (RGL)、极粗砂 (VCSS)、粗砂 (CSS)、中粗砂 (CMSS)、中砂 (MSS)、细砂 (FSS)、极细砂 (VFSS)、泥 (MUD)。这些粒级的百分含量为本研究的基本数据。研究步骤如下:

(1) 通过 RQ 型主分量分析来探讨各粒级间的相关性及沉积物样品的群聚性,选出最能分辨样品自然类型的主分量;

(2) 用当选主分量的 Q 载荷进行聚类分析,得出沉积物分类;

(3) 绘制沉积物类型平面分布图,据以进行沉积分区;

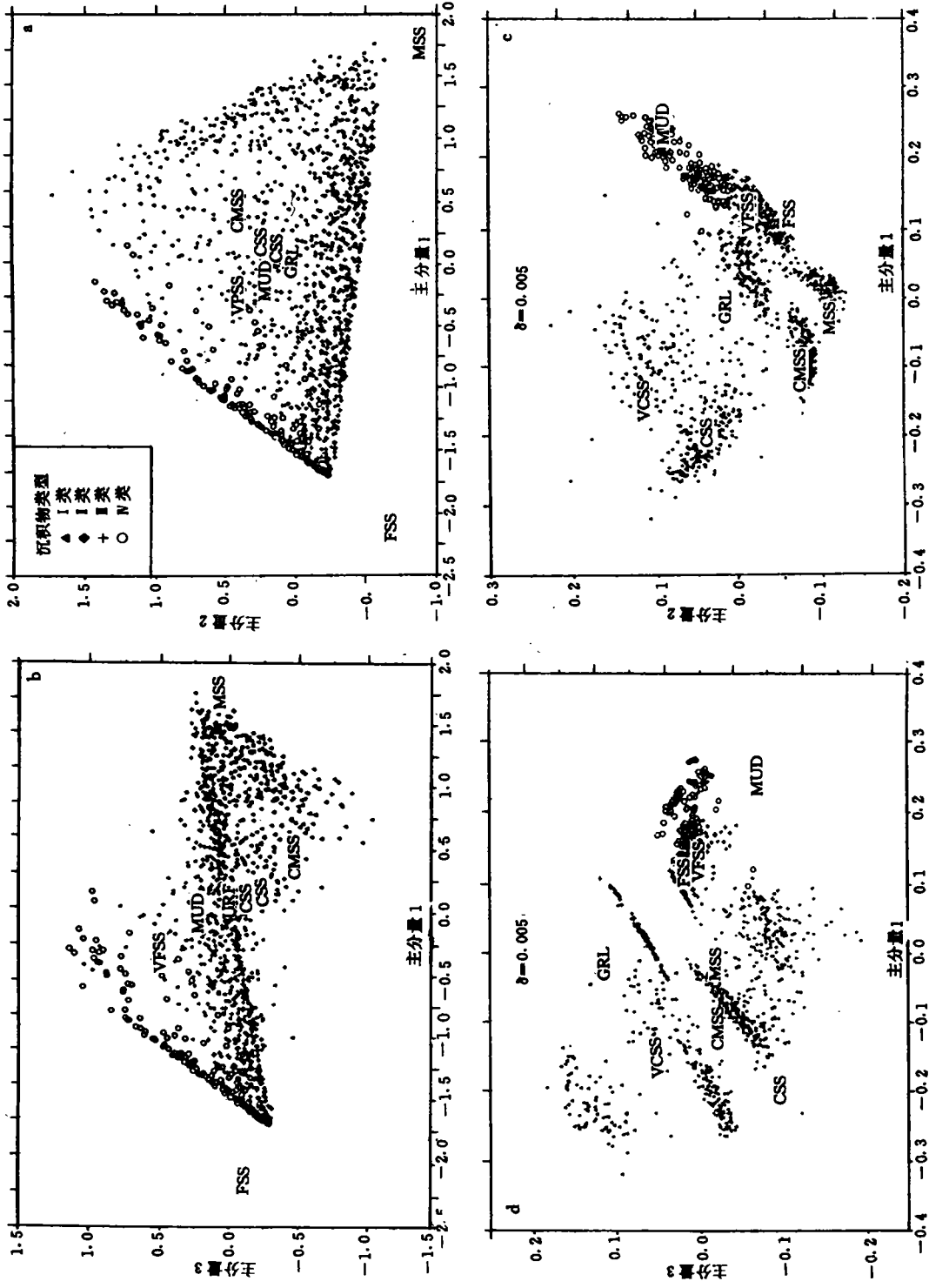
(4) 利用沉积物分选系数的梯度图及 McLaren 和 Bowles (1985)^{〔6〕} 的概率模型研究沉积物来源及运移方向;

(5) 利用沉积物搬运门限流速的经验公式,根据沉积物粒级的空间分布估计海底海水流速,进而分析研究区的沉积水动力环境。

本文将根据 1—3 步的研究结果侧重说明对数比统计分析的作用和效果。

RQ 型主分量分析 (RQPCA) 是探索性统计分析的一种重要方法。它利用 R 型和 Q 型主分量分析的对偶性,以 R 载荷图反映变量间的相关性,以 Q 载荷图反映样品的群聚性,以二者的叠合或对照揭示样品点群的变量特征^{〔2〕}。本文用 RQPCA 作为揭示样品粒度自然分类的主要手段。在作 R 和 Q 载荷叠合图 (图 1, 2) 时,将 R 载荷乘以归一化系数 $\sqrt{D/n}$,以使图上的 R 载荷与 Q 载荷尺度相当。

为了效果对比,首先对原始数据进行常规 RQPCA。图 1 a、b 为其协方差矩阵前三个主分量的 RQ 载荷叠合图,代表了数据总方差的 91%。在图上样品未显示任何群聚现象;变量间的相关关系也不合理。如在第一、二主分量散点图 (图 1a) 上,R 载荷象个三元图,其中上方端元由六个粒级组成,包括最粗的 GRL 和最细的 MUD,呈现高度正相关。这些都是闭合效应造成的假象。



a、b. 根据原始数据的协方差矩阵作出 c、d. 根据对数比变换的协方差矩阵作出 样品阵列作出样品分类同图2所示, 粒级代号见正文
 图1 达斯浅滩沉积物粒度数据的RQ型主分量分析前三个主分量的RQ载荷图。

Fig. 1 RQ-loadings of the first three principal components for the grain-size data from the Darss Sill

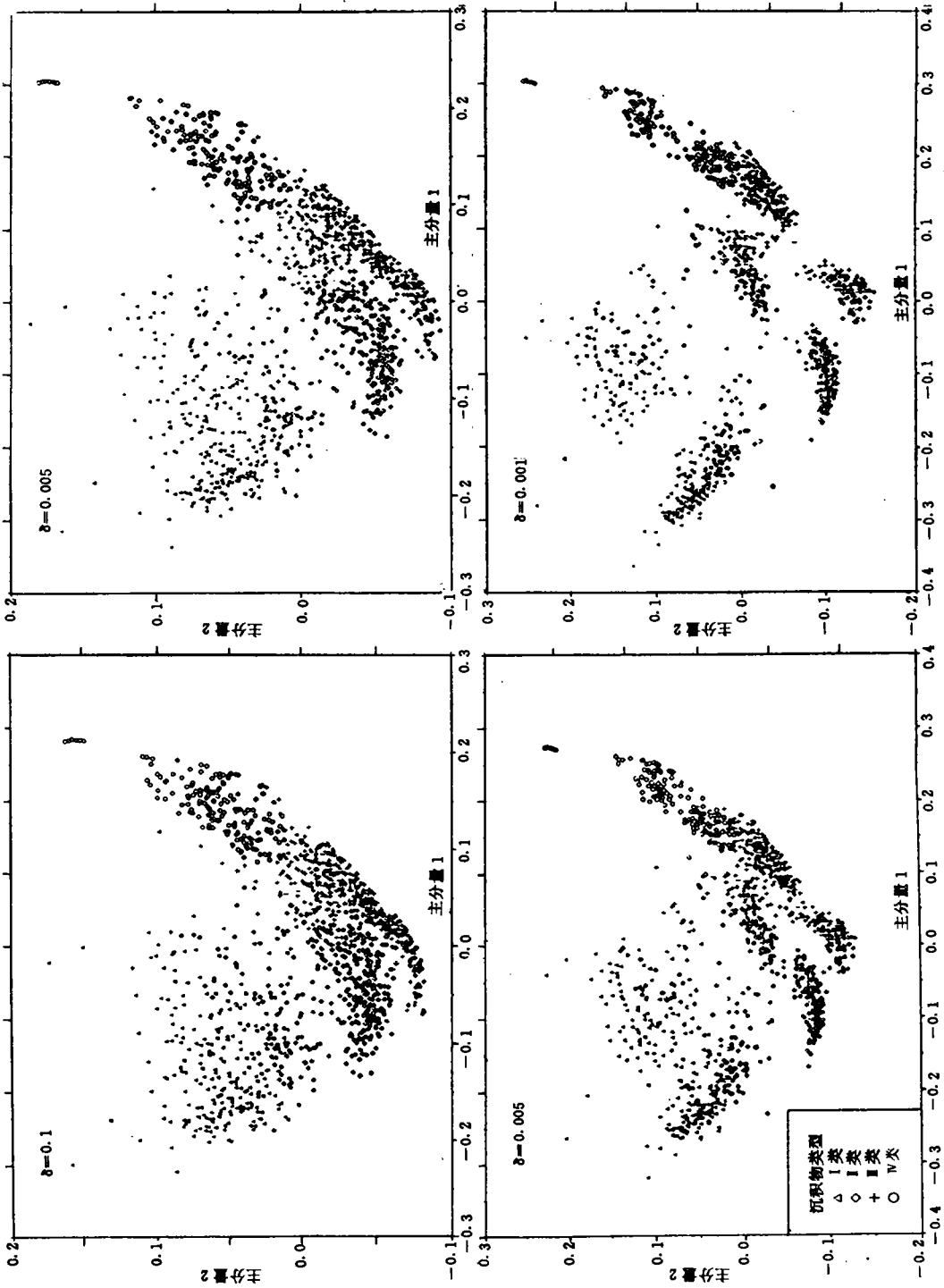


图2 对数比主分量分析中零替换的灵敏度分析, 示不同 δ 值所得出的前两个主分量的Q载荷图

Fig. 2 Sensitivity analysis of zero substitution in logratio principal components analysis, showing Q loadings of the first two principal components with given δ values.

对粒度数据进行中心化对数比变换,然后以协方差矩限为基础进行RQPCA,结果见图2c、d。沉积物样品出现十分清晰的分群。第一主分量代表了总方差的一半以上,反映了细粒级(MUD、VFSS、FSS)与除GRL以外的粗粒级(CSS、VCSS、CMSS)之间的分异,这是该区水动力条件下主要处于沉积状态的组分与主要处于剥蚀、搬运状态组分间的分异。第二主分量(代表16%总方差)显示了MSS与VCSS间的分异,而第三主分量(代表11%总方差)主要显示GRL含量的变化,为残积分量。

前三个主分量对样品群聚性反映最好,因而被选作聚类分析的基础。由于对群聚性反映差的主分量被删去,聚类分析的效果提高了,结果将沉积物划分为四大类若干亚类(表1),进而将该区划分为七种沉积区(图3)。

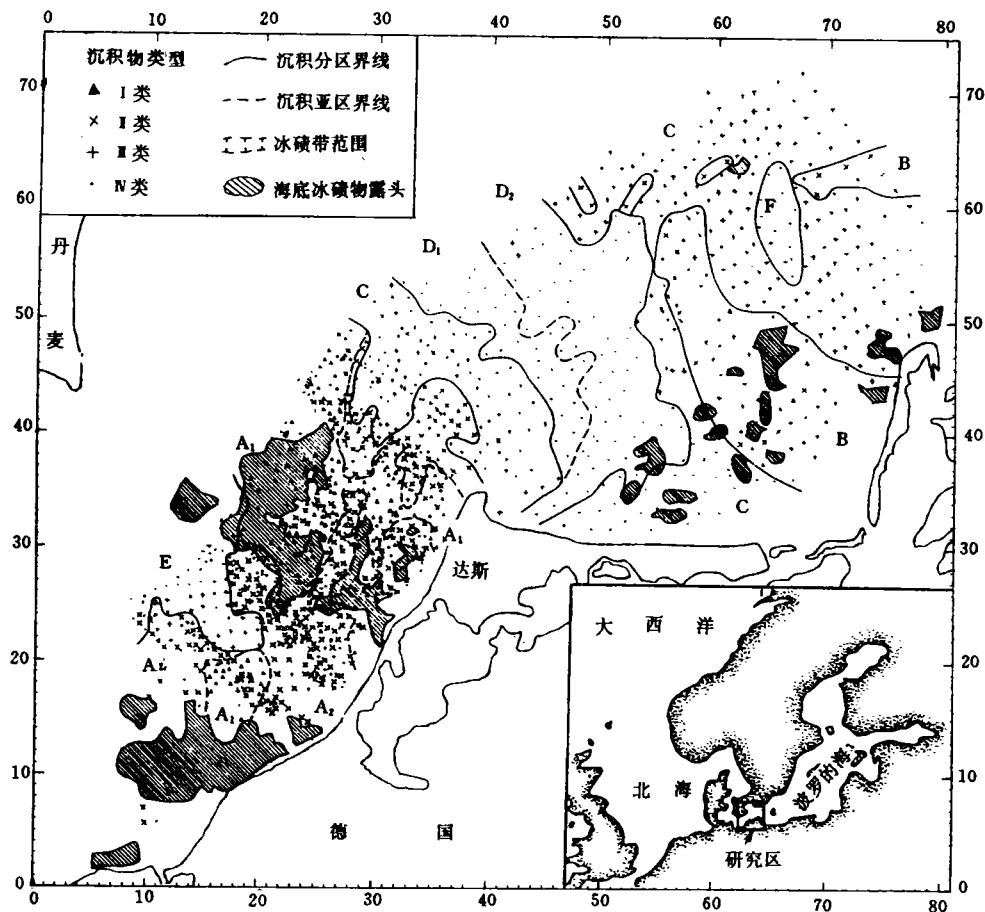


图3 达斯浅滩的区域位置、各类沉积物的空间分布及沉积分区图

Fig. 3 Map of the distribution of sediment types and sedimentary regionization for the Darss Sill.

The inset shows the geographic location of the study area.

从数学角度对沉积物分类的效果进行了检验。用费歇线性判别函数进行四类判别, Wilks $\lambda = 0.05$, 类间均值相等的概率为零, 大多数样品的归类概率为1, 正判率高达

84.7%，说明分类效果十分显著。

对零替代的影响进行了分析。由于零不能取对数，我们按 Aitchison (1986)^[5] 的建议用一个小正数 $\epsilon = \delta (C+1) (D-C) / D^2$ 来代替零；式中 C 为该样品中零含量的个数， δ 为给定正数，可取数据精度的一半，本例中为 0.005。我们分别用 $\delta = 0.001, 0.005, 0.05, 0.1$ 进行了 RQPCA (图 2)，发现样品点群间的分离程度随 δ 值的增大而降低，**晰可辨**，而与图 1a 截然不同。据此可认为图 1 c、d 所反映的群聚现象并不是由零替代产生的假象。

从沉积水动力学角度对分类和分区的效果进行了考查。合理的分类将使 (1) 类间在粒度分布上有显著差异，(2) 在空间分布上有一定的范围和组合，(3) 能反映研究区沉积环境的特点并符合沉积水动力学的一般规律。由表 1 可见本文的分类符合条件 (1)，由图 3 可见条件 (2) 也可满足，以下主要讨论条件 (3)。

与研究区的地形、地貌、地质、海流资料进行对照，可见本文的沉积分区极好地反映出在水动力条件控制下沉积物的来源、分散、沉积的特征和规律。

研究区的主要地质地貌特征是一条魏克塞尔冰期的前缘冰碛带，呈北西—南东向出露于达斯浅滩的海底。图 3 中 A 区出现于冰碛带上方，沉积物以粗粒及分选差为特征，是研究区的主要物源区。其中以 I 类沉积物为主的 A₁ 亚区为残积区，而有 I、II 类沉积物混合分布的 A₂ 亚区代表切割冰碛带的狭窄水道及冰碛带下游（东侧）的短程拖曳搬运产物。

B 区为次要物源区，地貌上为一北西向的楔形浅滩，卫星照片显示为上升流区。该区也有一些冰碛物露头，由这些露头以及相邻浅滩的侵蚀产生的 I、II 类沉积物与从达斯浅滩远程搬运来的 III 类沉积物的混合是本区的特征。

C 区面积较大，位于 A 区以东及区北东，为 III 类中细砂的主要分布区，代表以跃移和间歇悬浮方式大量搬运的沉积物。

D 区被 C 区环绕，由 IV₂ 亚类分选极佳的细砂组成，是研究区内的主要远程沉积区。它的边界清楚，表明在该区水动力条件下由搬运向沉淀的转变发生在中砂和细砂粒级之间，对应于粒径 0.2 mm。

E 区位于研究区西北角，沉积物以泥含量极高的 IV₁ 和 IV₃ 亚类为主，推测这是冰湖期富有机质的细粒沉积物出露海底的地区；这些细粒沉积未被冲走与该区终年存在一个涡流有关^①。

F 区为与 B 区有关的另一细粒沉积区，性质与 D 区类似。

在上述沉积分区的基础上进一步根据沉积物粒度特征分析了研究区的沉积物剥蚀、搬运、沉积规律及其所反映的近海底海流优势流向和流速的空间变化^[7]。

表 1 达斯浅滩海底沉积物分类及主要特征

Table 1 Sediment classification for the Darss sill, nestern Baltic Sea.

类型	描述	样品数	中值粒径 (φ)	分选度 (φ)
I	分选差的粗细砂、含砾和泥	260	$\frac{1.3}{1.0 \sim 2.1}$	$\frac{1.15}{0.85 \sim 1.48}$
II	中-细砂	432	$\frac{2.1}{1.9 \sim 2.5}$	$\frac{0.73}{0.63 \sim 0.92}$
III	分选良好的细砂	470	$\frac{2.6}{2.1 \sim 2.7}$	$\frac{0.60}{0.60 \sim 0.86}$
IV	细砂和泥	178	$\frac{3.1}{3.0 \sim 3.9}$	$\frac{0.69}{0.48 \sim 1.00}$

* 线上为该类的值，线下为亚类的变化范围。

① 波罗的海海洋研究所内部资料。

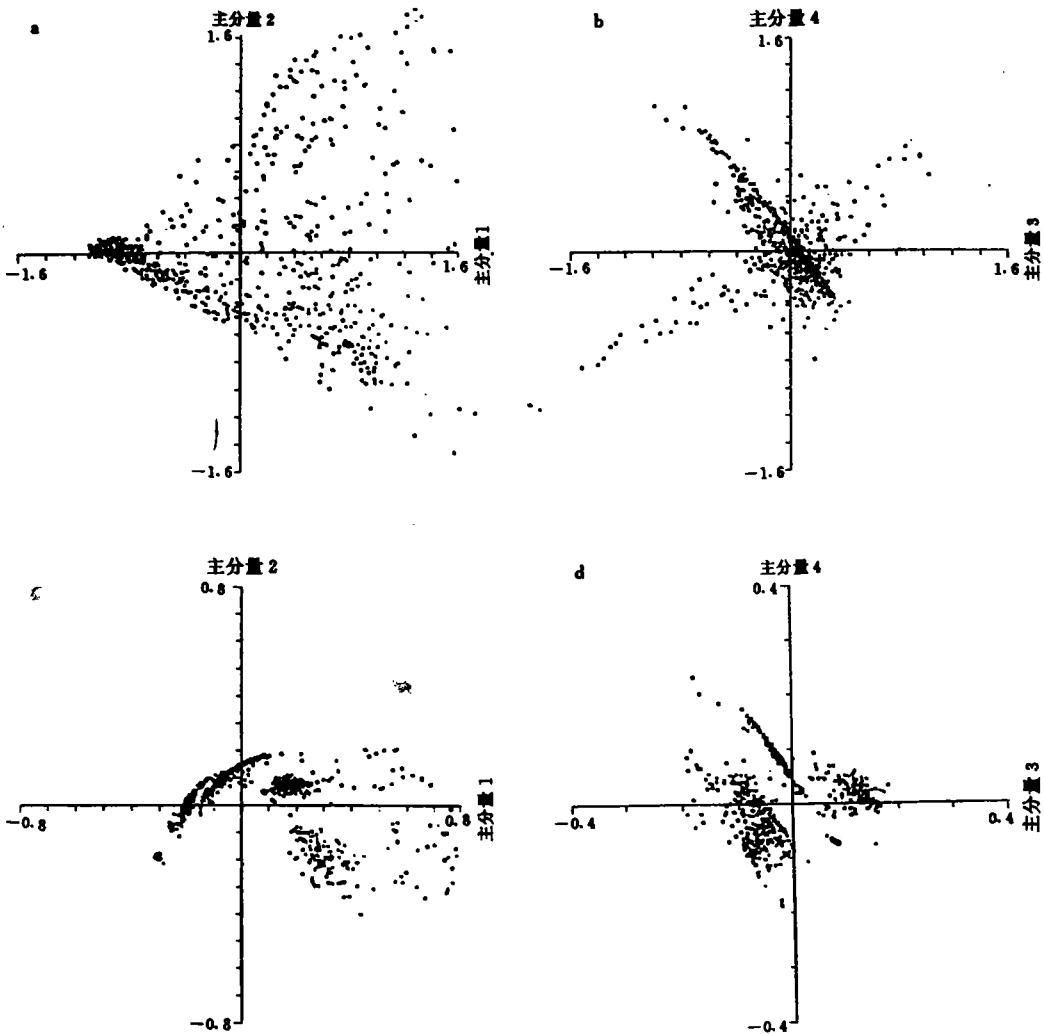


图4 南海北部陆架现代沉积物粒度数据的常规主分量分析(a、b)和对数比主分量分析(c、d)的前四个主分量的Q载荷图

Fig. 4 Scatter plots of Q-loadings for the first four principal components for the grain-size data from the northern shelf of the South China Sea, computed by ordinary principal components analysis (a and b) and logratio principal components analysis (c and d)

3 南海北部陆架沉积水动力环境研究

南海北部陆架自晚第三纪以来一直位于华南大陆边缘,在全球海平面变化及构造变动影响下发生多次海水进退及相应的浅海至陆坡的环境交替。晚更新世玉木冰期(2—1.5万年前)低水面时海岸线曾退到现今水深70—130 m处;全新世海侵高峰时海面曾高出现今海面3—5 m^[4]。陆架沉积物以来自华南的陆缘碎屑为主,南部有一些化学-生物化学沉积。

根据在陆架西部采集的723个海底沉积物样品的粒度数据,采用前述方案进行对数比统计分析,获得了沉积物分类及沉积环境分区^[3]。这里侧重报导对数比变换在该研究中所起的作用。

图 4 a、b 为对这些样品从 -1ϕ 到 10ϕ 这十二个粒级的重量百分含量进行常规 RQPCA 的前四个主分量 Q 载荷图, 可见样点不显示任何群聚。形成鲜明对照的是, 对同样数据进行对数比 RQPCA 获得的前四个主分量揭示出样点的清晰群聚现象 (图 4 b、c)。根据这四个主分量进行聚类分析, 将沉积物划分成三类共七个亚类^[10], 进而将该区划分成四种沉积区 (图 5), 分别对应于不同的沉积水动力环境: A 区为内陆架现代陆源细粒沉积区, B 区为中-外陆架经叠加改造的玉木冰期残留沉积区, C 区为高能环境下的现代或残留粗粒沉积区, 而 D 区为珠江口外由上升流保护的残留占三角洲堆积体。

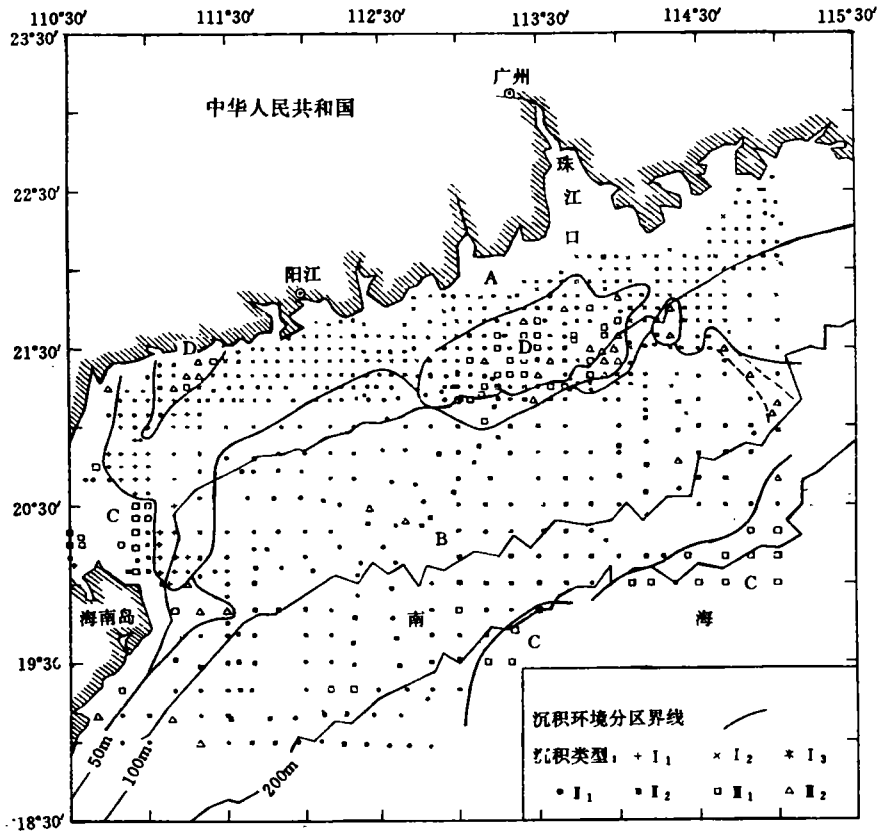


图 5 南海北部陆架各类沉积物的空间分布及沉积分区图

Fig. 5 Map of the distribution of sediment types and sedimentary regionization for the northern shelf of the South China Sea

结 论

本文从数学原理和实践效果两方面说明了对数比统计分析对于从沉积物粒度数据中萃取沉积水动力环境信息的重要性。无论在水流湍急多变、地质地貌复杂的达斯浅滩, 还是在开阔的南海北部陆架, 用原始粒度数据进行 RQ 型主分量分析时闭合效应严重, 无法反映样品的自然分类特征。而经对数比变换后的粒度数据的 RQ 型主分量分析则有效地揭示出沉积物在水动力条件控制下的自然分类, 以及沉积水动力环境的自然分区, 从而为进一步

研究区域内沉积物的剥蚀、搬运和沉积规律打下了基础。

本文是作者在德国波罗的海海洋研究所学术访问的成果之一。衷心感谢 J. Harff 教授及其同事的热情邀请和大力支持协助。

参 考 文 献

- [1] 周蒂. 成分数据的对数比统计方法简介. 地质科技情报, 1988, 7 (2): 107—114.
- [2] 周蒂. 对应分析的局限性及 RQ 式因子分析. 物化探计算技术, 1985, 7 (4): 332—341.
- [3] 周蒂, 罗又郎, 陈汉宗. 珠江口外现代沉积物粒度类型的对数比统计分析及其沉积水动力环境的讨论. 南海科学集刊 (第 10 集), 1992, 52—62.
- [4] 薛万俊等. 南海中北部晚第四纪古气候及古海洋学. 南海地质研究 (4), 1991, 1—96.
- [5] Aitchison J. The statistical analysis of compositional data. Chapman and Hall, London, 1986, 416.
- [6] McLaren P and Bowles D. The effects of sediment transportation on grain-size distributions. J. Sed. Petrol., 1985, 55 (4): 457—470.
- [7] Zhou Di, W Lemke, F Tauber and J Harff. Pattern and dynamics of modern sedimentation in Darss sill area, western Baltic Sea, as reflected by grain-size data. unpubl. manus, 1994.

Logratio Statistics and the Extraction of Environmental, Signals from Grain-Size Data

Zhou Di

(South China Sea Institute of Oceanology, Academia Sinica, Guangzhou 510301)

Abstract

Grain-size distributions of sediments contain important information on sedimentary source, dynamics and environment. The correct use of statistical techniques to extract as much as possible useful signals from grain-size data is of crucial significance. When applying multivariate analysis to grain-size data, however, the author encountered the problem of "closure", which may mask true patterns significantly. The logratio statistics is an effective way to avoid this problem. This paper presents two case studies respectively in the western Baltic Sea and northern South China Sea. While the classical statistics failed in representing the natural clustering of sediment samples, the RQ-mode principal components analysis based on logratio transforms of grain-size data made a clear separation among sediment classes. The sedimentary regionalization based on this classification successfully reflects the spacial variation of sedimentary hydrodynamic environments in these areas.

Key words logratio statistics RQ-mode principal components analysis grain-size data sedimentary hydrodynamic environment.