

基于 Fisher 算法对现代黄河三角洲叶瓣垂向环境演变的初步验证

张程瑜¹, 冯秀丽¹, 刘 杰², 刘 潇¹

(1. 中国海洋大学 海洋地球科学学院, 山东 青岛 266100; 2. 国家海洋局第一海洋研究所, 山东 青岛 266061)

摘要: 黄河三角洲沉积环境复杂, 自 1855 年以来形成多个三角洲叶瓣, 为了验证 Fisher 算法在黄河三角洲沉积环境划分中的积极作用, 在对取自黄河三角洲的柱状沉积物进行粒度分析的基础上, 使用 Fisher 算法并借助 DPS 软件, 对钻孔沉积物垂直剖面进行初步划分, 同时与通过其他传统方法划分的沉积层序进行对比。结果表明: 通过 Fisher 算法可以将研究区的沉积物划分为 3 层, 分别为 1.1~5.7, 5.7~18.1 及 18.1 m 以下, 地层的分段结果与他人的研究结果相吻合。因此, 采用 Fisher 算法对黄河三角洲沉积物进行层序划分是可行的。

关键词: 黄河三角洲; Fisher 算法; 粒度; 沉积层序

中图分类号: P736.21 文献标识码: A

文章编号: 1000-3096(2015)10-0080-05

doi: 10.11759/hyxx20131012001

黄河三角洲滨海湿地是在黄河所携带的泥沙的物质基础上, 在陆海动力相互作用下, 渐渐发育、演化形成的滨海湿地系统^[1]。黄河三角洲的垂向沉积序列基本上可分为两大类型: 河口沙坝型和烂泥湾型。叶瓣的垂向发展是叶瓣在平面上向海延伸的垂向表现, 其结果将记录在地层序列结构上。现阶段主要是通过通过对沉积物的敏感粒级组分、古生物分布、岩性、光释光等方法来分析沉积环境的演变, 本文中尝试使用一种新的方法, 即 Fisher 算法来对沉积环境的演变进行分析, 对沉积序列进行分层。

Fisher 算法如今在各个科研领域开始被大家接受, 尤其是在地质学研究中是一个非常有用的手段, 只要地质体的某些特征(粒度、元素含量、古生物含量等等)存在规律性的差异, 采用最优分割的数学处理方法, 就能按照顺序进行理想的分段, 如根据岩石单轴强度计算归纳出某一地区各深度地层的岩石强度变化规律, 作为钻头优选, 出砂预测等工程应用的参考^[2]、对油田火山岩岩性的识别, 初步评价油田石炭系火山岩油藏难采储量^[3]、对岩土层亚层进行划分, 从而安全地进行建筑场地工程地质勘查^[4]等等。同时最优分割法现阶段已经广泛地应用于声学、物理学、古生物学等多门学科, 如归纳其在地学方向的发展, 特别是沉积学、层序学, 也一定有更可观的前景。如在语音声学模型分析中的研究与应用^[5]、划

分第四纪孢粉带^[6]等。

当实验数据具有一定的规律性时, 即可考虑根据 Fisher 算法对数据进行分段, 沉积物的粒度参数数据也满足一定的规律性, 因此初步判断, Fisher 算法可以应用于通过粒度参数分析沉积环境的演变, 对沉积层序进行划分。因此, 本文使用 Fisher 算法并借助 DPS 软件对黄河三角洲柱状沉积物粒度数据进行初步分层, 同时与他人通过其他传统方法划分的沉积层序进行对比, 验证 Fisher 算法在黄河三角洲沉积环境划分中的有效性。

1 数据来源

本次研究所用黄河三角洲柱状沉积物样品采集于黄河三角洲湿地, 采集方式是浅钻, 钻孔坐标为 37°44'27"N, 118°55'50"E, 长度约为 21 m。样品主要为黄河三角洲近代沉积, 土质较软, 上部 10 m 左右为粉砂, 易液化。下部为约 10 m 左右的淤泥质黏土。

为更好地反映陆源悬浮物搬运沉积时的粒度特征, 在粒度分析的前处理过程中只加入 10% H₂O₂ 溶

收稿日期: 2013-10-12; 修回日期: 2013-11-11

基金项目: 海洋公益科研专项(201005009)

作者简介: 张程瑜(1987-), 男, 山东青岛人, 硕士, 主要从事海洋地质环境与工程方面研究, E-mail: zhangchengyu_love@126.com; 冯秀丽, 通信作者, E-mail: fengxiuli@ouc.edu.cn

液去除有机质的影响,而保留沉积物中的碳酸盐。本次实验是在中国海洋大学海洋地球科学学院的土工实验室,运用了英国的 Malvern 2000 型激光粒度仪进行测试并进行数据处理。仪器测量范围为 0.02~2 000 μm , 粒径分辨率 ϕ 为 0.01, 重复测量的相对误差<3%, 粒径误差小于 0.5 μm 方可合格。样品间隔为 10 cm, 共计 186 个样品。

2 研究方法

在运用沉积物粒度分析进行沉积序列中的环境分析时,现在常用的方法是,将沉积序列中不同物源或不同沉积动力条件下的粒度组分进行分离,然后再详细研究不同组分所代表的环境地质意义,从而根据其在沉积序列中的变化推断海洋气候环境的演变历史。现在较为常用的是粒级-标准偏差的变化对钻孔粒度组分进行分离。

近年来,在电子计算机应用的促进下,聚类分析在地质地球化学领域中得到了迅速发展,对地球科学的定量化、地质数据的自动化处理、地质过程的计算机模拟以及地质思维的智能化等方面起到了很大的促进作用,有力地推动了古老的地球科学与现代科学的结合^[7]。因此,本文将采用一种新方法对钻孔粒度组分进行分离。使用 Fisher 算法将沉积带划分之后,对通过粒级-标准偏差算法划分的沉积带进行验证。

Fisher 算法又称为最优分割法,是 Fisher 于 1958 年提出的^[8]。其原理是采用有序聚类的方法,把连续的相邻的相异系数较小的层位聚在一个带中,从而划分出不同的分段。Fisher 算法的一般形式是: $\{X_i, X_{i+1}, \dots, X_j\}$, $(1 \leq i < j \leq n)$

其计算步骤如下^[9]:

假设样本依次为 $x_1, x_2, \dots, x_n$, Fisher 算法具体的计算步骤如下。

第 1 步,数据正规化处理。设原始数据矩阵为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}_{n \times m}$$

将矩阵 X 中的 x_{ij} 变为 $Z_{ij} = (x_{ij} - \min_{1 \leq j \leq n} \{x_{ij}\}) /$

$(\max_{1 \leq j \leq n} \{x_{ij}\} - \min_{1 \leq j \leq n} \{x_{ij}\})$, $(i=1, 2, \dots, p; j=1, 2, \dots, n)$, 得到矩阵

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{p1} & z_{p2} & \dots & z_{pm} \end{bmatrix}_{p \times m}$$

第 2 步,计算矩阵 Z 的变差,从而得到变差矩阵 D 。假设 $\{z_1, z_{i+1}, \dots, z_j\}$ 是一类 $(j \geq i)$, 其平均值为

$\bar{z}_{ij} = (\sum_{a=i}^j z_a) / (j-i+1)$, 样本的变差(离差平方和)为

$d_i = \sum_{a=i}^j (z_a - \bar{z}_{ij})^2$, 则整个矩阵的变差为:

$$d_{ij} = \sum_{b=i}^j \sum_{a=1}^p (z_{ab} - z_{a_{ij}})^2, i, j=1, 2, \dots, n$$

因为 $d_{ij} = \begin{cases} 0, & i=j \\ d_{ji}, & i \neq j \end{cases}$, 因此计算 $n(n-1)/2$ 个

$d_{ij} (1 \leq i, j \leq n)$ 即得到矩阵 D 。

第 3 步,计算误差函数。将第 i 号样本 z_i 设为 i 。如将 n 个样本分为 k 类, 设某分法为 $p(k, n): \{i_1=1, i_1+1, \dots, i_2-1\}, \{i_2, i_2+1, \dots, i_3-1\}, \dots, \{i_k, i_k+1, \dots, n\}$, 其中 $i=1 < i_2 < \dots < i_k < n$ 。将误差函数定义为:

$$\psi[p(k, n)] = \sum_{j=1}^k (D_{i_j, i_{j+1}} - 1)$$

即某种分段方法的误差函数为各分段的变差和。当 n 和 k 固定时, $\psi[p(k, n)]$ 越小, 表示类内的离差平方和越小, 分类就越合理。因此, 需要找到一种分段方法使 $\psi[p(k, n)]$ 达到最小。

现在假设误差函数达到最小的分法是 $p_0(k, n)$, 此时误差函数的值为 $\psi[p_0(k, n)]$ 。

第 4 步, $\psi[p_0(k, n)]$ 的递推公式。当 $k=2$ 时, 要找出一个分界线使全部样本分成两类, 而 $p_0(2, n)$ 是所有可能的分界线中使误差函数达到最小的分法。于是,

$$\psi[p_0(2, n)] = \min_{2 \leq j \leq n} \{D(1, j-1) + D(j, n)\}$$

对 n 施行归纳法, 可使递推公式 $\psi[p_0(k, n)] =$

$$\min_{m \leq j \leq n} \{\psi[p_0(k-1, j-1)] + D(j, n)\} \text{ 成立。}$$

该式表示 n 个样本分成 k 类的最优分法, 可以看成 $j-1$ 个样本分成 $k-1$ 类的最优分法再与最后 $(n-j+1)$ 个样本形成的第 k 类合并而成。而 j 由 k 一直变到 n , 从中可选择一个最优的 j , 从而形成 n 个样本的最优分法。

第 5 步, 聚类。由上可知, 采用最优分割法主要是计算两个表: 第一个表是各种可能类的变差矩阵 D ; 第二个表是最小误差函数表 $\psi[p_0(r, q)]$ 。首先计算变差矩阵 D , 再由 D 计算 $\psi[p_0(2, q)]$, 然后计算 $\psi[p_0(3, q)]$, 再计算 $\psi[p_0(4, q)]$ 等, 一直持续计算到 $\psi[p_0(k, q)] (q=k, k+1, \dots, n)$ 。有了上述两个表就可进行分类了。首先找出 j_k (最后一个类的第一个样本号), 使得 $\psi[p_0(k-1, j_k-1)] + d(j_k, n)$ 达到最小, 即 j_k 应满足关系式 $\psi[p_0(k, n)] = \psi[p_0(k-1, j_k-1)] + d(j_k, n)$, 然后找 j_{k-1} , 使它满足关系式 $\psi[p_0(k-1, j_{k-1}-1)] = \psi[p_0(k-2, j_{k-1}-1)] + d(j_{k-1}, j_k-1)$ 。以此类推, 找出 $j_{k-2}, j_{k-3}, \dots, j_2, j_1=1$ 。于是得到分类: $G_1=\{j_1=1, \dots, j_2-1\}$, $G_2=\{j_2, \dots, j_3-1\}, \dots, G_k=\{j_k, \dots, n\}$ 。

3 结果

徐元芹等人^[1]曾经在文章中, 通过对粒度敏感组分、古生物特征、光释光测年等数据给出该钻孔较为详细准确的钻孔地层柱状图, 详细地分析了该钻孔的沉积环境演变历史。

本文根据岩性地层学(粒度组成和粒度参数)等方法, 通过粒级-标准偏差算法来获得粒度组分的个数和分布范围, 对钻孔粒度组分进行分离, 得出钻孔的敏感组分, 结合其他已有的资料和文献, 对所得样品进行分析, 对调查区域内的钻孔岩心进行了

地层划分, 由于人为取样等误差可能造成实验数据有些许不同, 但所得结论与已有结果基本相符, 由于缺少测年及古生物等资料, 因此通过粒度简单的将钻孔共分为 3 个地层, 地层的分段与趋势跟徐元芹已有分层相吻合, 代表了黄河三角洲沉积体叶瓣的不同时期的环境演变过程。

研究所用的聚类分析方法借助 DPS 软件完成^[10]。样品间隔为 10 cm。首先依次将分析样品按由上到下的次序编号为 1, 2, 3, ..., 186, 把全部用于分析的粒度参数及编号数据导入 DPS 软件, 利用其中的最优分割法对全部样品进行有序聚类, 分段误差函数、分段数及聚类方案表 1。

从计算方法可知, 有序聚类中划分出的段数越多时, 其误差函数就越小, 分辨这种数量优劣的原则是观察某点其误差函数曲线下降是否特别大, 该点是否是这个函数曲线的拐点, 这个点后面的曲线是否开始变得平缓, 函数曲线的斜率是否变小^[11]。由样品的误差函数统计图(图 1)可看出, 当分段数为 3

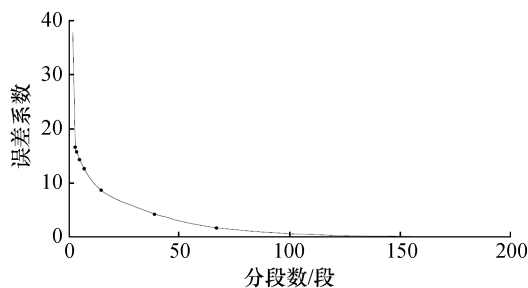


图 1 误差函数统计及分段数函数图

Fig. 1 Statistics of error function and segmentations of the core samples

表 1 钻孔有序聚类方案

Tab. 1 Sequential clustering program of the core samples

分类数	误差函数	最优分割结果
2	37.7243	1~41, 42~181
3	16.5557	1~41, 42~151, 152~181
4	15.7210	1~22, 23~41, 42~151, 152~181
5	14.2735	1~11, 12~13, 14~41, 42~151, 152~181
6	13.6083	1~11, 12~13, 14~41, 42~151, 152~154, 155~181
7	12.5659	1~11, 12~13, 14~41, 42~151, 152~156, 157~167, 168~181
8	11.9894	1~11, 12~13, 14~41, 42~151, 152, 153~154, 155~167, 168~181
9	11.2143	1~11, 12~13, 14~33, 34, 35~41, 42~151, 152~156, 157~167, 168~181
10	10.6378	1~11, 12~13, 14~33, 34, 35~41, 42~151, 152, 153~154, 155~167, 168~181
...		

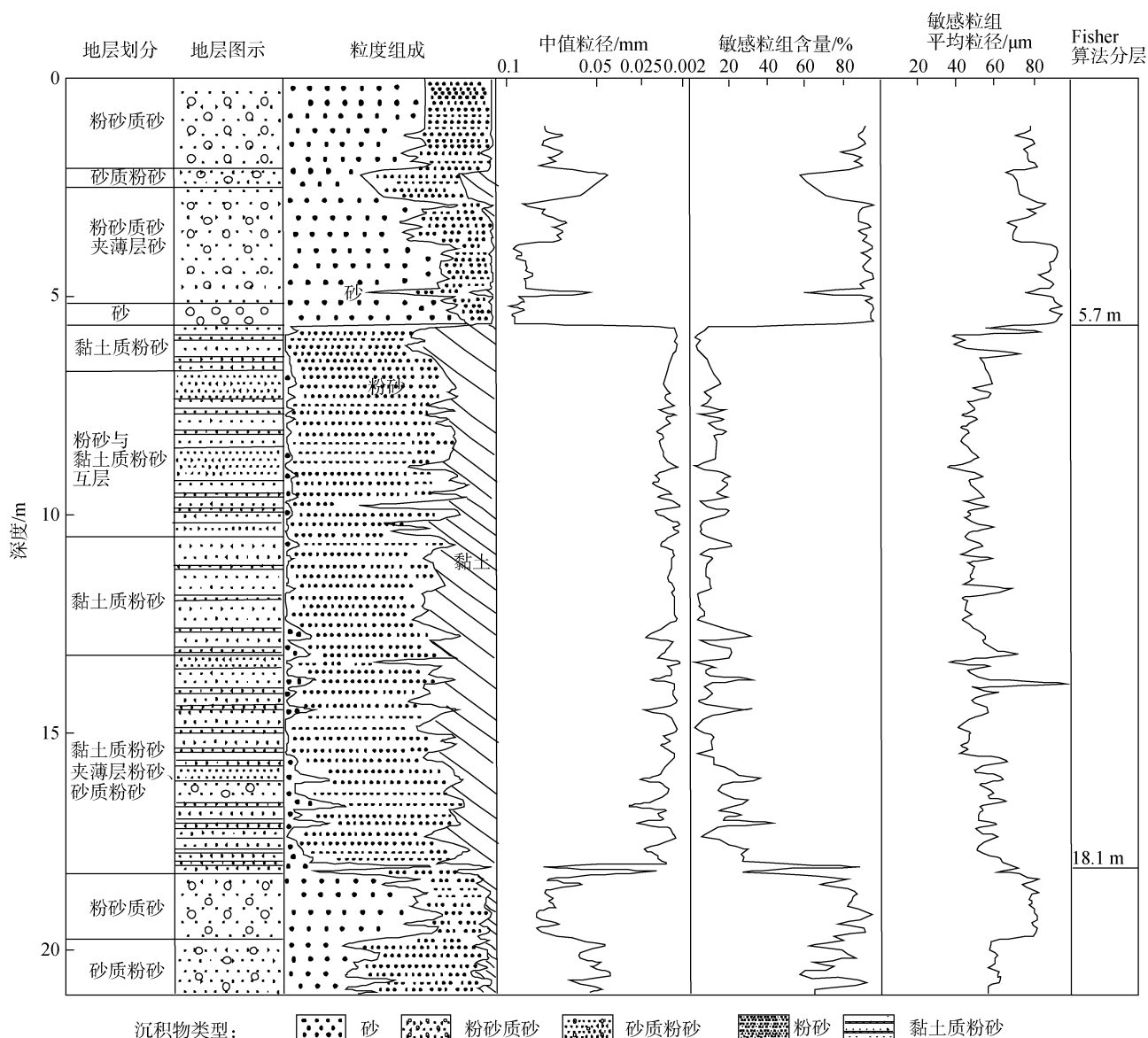


图 2 Fisher 算法分层与柱状图对比

Fig. 2 Division of sediments stages of the core based on Fisher algorithm

时,后面的曲线开始变得相对平缓,据此可以将钻孔柱状样初步分为3段,分别为1.1~5.7, 5.7~18.1及18.1 m以下。Fisher 算法分层结果见图2。

本文运用数学方法,计算沉积物粒度参数,得出一个初步的分段,粒度变化呈现的沉积环境演变历史基本跟已证明的结果相吻合,因此可以证明通过 Fisher 算法分析粒度参数来对地层的分段进行佐证可行。可见基于 Fisher 算法的聚类分析在钻孔柱状样分层具有一定的指导意义和参考价值。

4 讨论

1) 采用最优分割法进行处理,使其计算过程不

掺杂人为因素,从而减少了主观性的干扰,大大提高其准确性。Fisher 算法在沉积学层序划分中的应用仍比较少,而且目前仅仅可以对某些具有规律性的参数进行聚类分层,但是地层在垂向上的划分往往需要结合多种方法(测年、孢粉、古生物等等)综合划分,所以, Fisher 算法对地层划分具有一定的参考意义,可是稍显局限。但是可以预见,其在今后的研究中可能会占据越来越重要的位置,并且随着研究的深入和科学的发展, Fisher 算法将成为一种主流的方法。

2) Fisher 判别法相较于粒级-标准偏差曲线等传统的划分方法,优势在于:对样品数据的分布和方差等都没有什么限制,应用范围较广,适用于多种

沉积环境;使用数学方法的计算,更能方便地应用于沉积序列的建立与解释,消除人为影响因素,这在实际研究中是非常重要的。

5 结论

1) 通过钻孔粒度参数分析,将钻孔初步分为了三段: (1)1.1~5.7 m; (2) 5.7~18.1 m; (3)18.1~21 m, 粒度分层呈现的沉积环境的演变历史,基本与已有结论相符,因此通过 Fisher 算法来对沉积序列进行划分是可行的,对沉积环境的演变可以进行一定的佐证。

2) 误差函数统计及分段数函数图中由于样品原因,实验处理原因等等,在图像中可能会出现多个变化较大的拐点,每个变化较大的拐点对应的段数可能对样品来源地的物源变化或沉积环境演变均有指示意义。

参考文献:

- [1] 徐元芹,李萍,李培英,等.黄河三角洲 HS908 孔处滨海湿地发育与演变[J].海洋科学进展, 2011, 29(3): 346-354.
- [2] 吴惠梅,李忠慧,朱亮,等.有序样品聚类的最优分割法在地层特性评价中的应用[J].江汉石油学院学报, 2008, 130(12): 460-462.
- [3] 张家政,关泉生,谈继强,等.Fisher 判别在红山嘴油田火山岩岩性识别中的应用[J].新疆石油地质, 2008, 29(6): 761-764.
- [4] 王生全,唐亦川,薛喜成.多元有序地质数据的最优分割法在工程中的应用[J].西安科技学院学报, 2002, 22(2): 147-150.
- [5] 蔡骅,张杰,黄志同.Fisher 算法在语音声学模型分析中的研究与应用[J].数据采集与处理, 2000, 15(4): 471-475.
- [6] 刘会平.利用有序样品最优分割法划分第四纪抱粉带的初步研究[J].华中师范大学学报(自然科学版), 1995, 29(3): 391-399.
- [7] 范迪富.有序样品聚类的最优分割应用程序开发[J].江苏地质, 2004, 28(1): 22-26.
- [8] 盛骤,谢式千,潘承毅.概率论和数理统计[M].北京:高等教育出版社, 2001.
- [9] 唐启义,冯明光.DPS 数据处理系统-实验设计、统计分析 & 数据挖掘[M].北京:科学出版社, 2006.
- [10] 甘桂兰,程新,李昆太.DPS 数据处理软件在发酵培养基优化中的应用[J].中国酿造, 2008, 11: 69-71.
- [11] 杨向东,秦乃岗.使用有序样品聚类分析法划分东南沿海地震活跃幕[J].华南地震, 2004, 24(2): 55-61.

The preliminary verification of the modern Yellow River Delta's vertical environment evolution based on Fisher algorithm

ZHANG Cheng-yu¹, FENG Xiu-li¹, LIU Jie², LIU Xiao¹

(1. College of Marine Geosciences, Ocean University of China, Qingdao 266100, China; 2. The First Institute of Oceanography, State Oceanic Administration, Qingdao 266061, China)

Received: Oct., 12, 2013

Key words: the Yellow River Delta; Fisher algorithm; grain size; sedimentary sequence

Abstract: The Yellow River Delta has a complex sedimentary environment, forming 12 delta lobes since 1855. In order to verify the effect of Fisher algorithm on the division of sedimentary environment in the Yellow River Delta, this paper divided the sedimentary environment of the core drilling from the Yellow River Delta by using the Fisher algorithm and Data Processing System (DPS) software, and compared the results with those divided by other traditional methods. The results showed that the deposits could be divided into three layers: 1.1~5.7 m, 5.7~18.1 m, and below 18.1 m, which coincided well with the results of other reports. Therefore, using Fisher algorithm to divide the sedimentary strata in the Yellow River Delta is feasible.

(本文编辑: 刘珊珊)