

DOI: 10.16562/j.cnki.0256-1492.2018100802

基于粒度分布曲线的邻近传播聚类算法在沉积环境识别中的应用——以白洋淀地区为例

刘祥奇^{1,2}, 宋磊¹, 吴奇龙^{1,3}, 李国民², 毛欣¹

1. 中国地质科学院水文地质环境地质研究所, 石家庄 050061
2. 中国地质大学(北京)工程技术学院, 北京 100083
3. 中国地质大学(武汉)地球科学学院, 武汉 430074

摘要:以白洋淀地区出露的22个沉积剖面中沉积物的粒度数据为研究对象,利用邻近传播聚类算法(AP聚类算法)进行聚类分析,并与已知典型沉积环境形成的沉积物粒度频率分布曲线进行对比,探讨了基于沉积物粒度特征的邻近传播聚类算法在沉积环境识别中应用的可行性以及白洋淀地区不同沉积环境的粒度特征。结果表明:邻近传播聚类算法能够将相同或者相近动力条件下形成的沉积物聚为一类,挖掘沉积物粒度数据中蕴含的沉积动力信息;所有样品的沉积物粒度频率分布曲线划分为11类簇;考虑到同一沉积环境中动力条件的变化,将聚类结果得到的分布范围、形状相近的11簇曲线进一步合并为4组曲线,并与已知典型沉积环境粒度曲线进行对比,识别出湖沼相、湖相、河流相、洪积相等4种主要的沉积相。其中,湖沼相与湖心相沉积环境、湖滨相与河漫滩相沉积环境形成的沉积物粒度组分相近。湖沼相与湖心相、湖滨相与漫滩相、河床相的沉积物粒度主要组分分别为细粉砂、粗粉砂、细砂或粗砂,洪积物中粗粉砂、中砂、粗砂含量相近,呈多峰态。邻近聚类传播算法可为沉积环境动力条件反演、分区等提供潜在的新手段。

关键词:粒度分布曲线;邻近传播聚类算法;沉积环境;识别;白洋淀地区

中图分类号:P512.2

文献标识码:A

Application of the affinity propagation clustering algorithm based on grain-size distribution curve to discrimination of sedimentary environment——A case study in Baiyangdian area

LIU Xiangqi^{1,2}, SONG Lei¹, WU Qilong^{1,3}, LI Guomin², MAO Xin¹

1. Institute of Hydrogeology and Environmental Geology, Chinese Academy of Geological Sciences, Shijiazhuang 050061, China
2. China University of Geosciences, Beijing 100083, China
3. China University of Geosciences, Wuhan 430074, China

Abstract: In the paper, 22 sedimentary sections are selected from the Baiyangdian region as research carriers. The affinity propagation (AP) clustering algorithm is adopted to cluster the similar grain size distribution of sediment into clusters. The results are then compared with the characteristics of grain size distribution which are known in typical sedimentary environments. The grain size distribution patterns in different sedimentary environments in Baiyangdian area are concluded and the feasibility of application of the affinity propagation clustering algorithm based on sediment grain characteristics to environmental interpretation discussed. The results suggests that the AP clustering algorithm can gather the sediments formed under the same or similar dynamical conditions into groups, and dig out the sedimentary dynamical information contained in the grain size data; The distribution curves of grain size for all samples are subdivided into 11 clusters. Considering the change in dynamic conditions in the same sedimentary environment, the 11 cluster curves are further categorized into 4 sets of curves, which could be compared with the grain size curves from known environment. 4 sedimentary facies i.e. the lacustrine-swamp facies, lacustrine facies, fluvial facies and alluvial facies are recognized. Among them, in terms of sedimentary dynamics, the lacustrine-swamp facies are similar with the central lake facies, while the lake shore facies similar with alluvial flat facies. The lacustrine-swamp facies and central lake facies are mainly composed of fine silty sand, whereas the lake shore facies and alluvial flat facies composed of coarse silt. River bed facies consist of sand or coarse sand. The contents of coarse silt, medium sand and coarse sand are similar in the flooding deposits, characterized by multi-peak curves.

资助项目:中国地质调查局项目“中国北方晚全新世水文-生态环境与粉尘活动的湖泊沉积记录”(SK201503),“雄安新区第四纪地质调查”(Q201801-7),“过去200年白洋淀环境演化过程及其驱动因素”(SK201706)

作者简介:刘祥奇(1994—),男,硕士研究生,研究方向为第四纪地质学、工程地质学。

通讯作者:宋磊(1983—),男,助理研究员,研究方向为湖泊沉积与气候变化, E-mail: sdsonglei@163.com

收稿日期:2018-10-08; **改回日期:**2019-01-27. 周立君编辑

The performance proves that the AP clustering algorithm can provide a new mean for inversion and zonation of sedimentary environment conditions.

Key words: grain-size distribution curve; the affinity propagation (AP) clustering algorithm; sedimentary environment; discrimination; Baiyangdian

沉积物粒度组成及其分布模式是沉积物的基本特征,能够揭示沉积物形成的动力条件,是判别沉积环境和沉积物源的重要指标。不同沉积环境形成的沉积物受其搬运方式、动力条件等的约束,可分选出特定的粒度组合,因此粒度分布特征分析具有重要的沉积学意义^[1]。沉积物粒度频率分布曲线可以定量反映某一粒径分布含量、次总体类型、分选程度等沉积过程的细节信息^[2]。这些信息的提取有助于理解沉积物的来源、成因及形成过程,尤其是粒度分布子体分离法^[3]、Weibull分布函数拟合法^[4]、端元模型法^[5]等统计方法的应用,对于研究不同粒度组成端元或者组分的气候环境意义具有重要的推动作用。上述统计学方法重点关注某一沉积类型沉积物形成的动力过程和沉积成因,对于沉积环境复杂、沉积类型多样的沉积区,其沉积物成因相对复杂,如何快速、准确判别沉积物的沉积相显得尤为重要。聚类分析作为数据挖掘领域重要的手段之一,直接比较数据之间的性质差异,将数据性质相同或相近的聚为一类,在沉积物粒度数据分析与处理中具有较好的适用性^[6]。然而,常规的聚类分析方法需要预设聚类数量,不能真实地反映根据沉积物粒度分布曲线类型反演出研究区存在的沉积动力类型,而且以往研究应用聚类分析大多以粒度参数表征粒度频率分布曲线作为研究对象,不能精确地表示粒度分布曲线形态以及各粒径分布含量^[7]。值得注意的是,AP聚类算法(Affinity Propagation Clustering Algorithm)作为新兴的聚类方法,具有快速、高效、不需要预设聚类数量的优点,对于高维数据聚类效果明显,在图像、文本聚类以及基因表达数据分析等领域应用广泛^[8-10],将AP聚类算法方法引入沉积学研究中,对于粒度数据的深度挖掘与分析具有重要的推动作用。

白洋淀地区分布深厚的第四纪沉积物,沉积环境较为复杂,以冲积、洪积和湖积作用为主,是开展不同沉积环境条件下沉积物粒度分布曲线判别与划分研究的理想区域。区内主要侧重于利用气候环境代用指标进行古环境演化方面的研究,缺乏沉积环境动力条件的研究^[11-13]。通过采集研究区内沉积剖面中已知沉积相的沉积物,确定不同沉积环境的沉积物粒度特征,研究AP聚类算法在沉积学研

究中的可行性与适用性,尝试建立沉积物粒度特征与沉积动力条件的内在联系。

1 自然地理概况

研究区位于华北平原白洋淀及其周边区域,是永定河冲积扇与滹沱河冲积扇的前缘低洼区,位置为38.7°~39.2°N、115.6°~116.5°E。研究区属于温带大陆性半湿润半干旱气候,年平均温度为7.3~12.7℃,多年平均降水量564mm,主要集中在6—8月份^[14]。区内地势平坦,由西向东地势略有降低。白洋淀有拒马河、府河、唐河等九条河流汇入以及赵王新河、大清河等出淀河流。第四纪以来,该地区不断沉降,并普遍接受沉积,而且区内水文条件的变化对第四纪沉积物的形成也具有重要的影响,逐渐形成了一套以冲洪积和湖积为主的沉积地层^[15]。

2 材料与方法

2.1 样品采集与测试

本研究对白洋淀周边地区露头剖面的沉积地层进行了详细考察,选取其中22个剖面作为研究对象(图1,剖面地层描述见表1),对各剖面典型沉积相层位进行样品采集,共获得85个沉积样品。现代河流漫滩相沉积样品来源于白沟引河,共采集2个沉积样品,沉积物为棕黄色黏土质粉砂,具有典型的二元结构。现代湖泊表层沉积样品来源于白洋淀湖泊,共采集12个沉积样品,沉积物为灰黑色粉砂质黏土、黑色黏土,有机质含量高(图2)。

粒度测试在中国地质科学院第四纪年代学与水文地质演变重点实验室完成,采用mastersizer 2000激光粒度仪进行测试,该仪器的测试范围0.2~2000μm。测试前加入足量的10%双氧水并加热,去除有机质。待烧杯中不再冒泡之后,加入足量的10%盐酸,直至不再冒气泡才停止加热,然后加入蒸馏水静止24h。利用虹吸法抽取上层清液,加入蒸馏水,静止24h,反复几次直至呈中性。加入10mL 10%的六偏磷酸钠,用超声波振荡10min后上机测试。

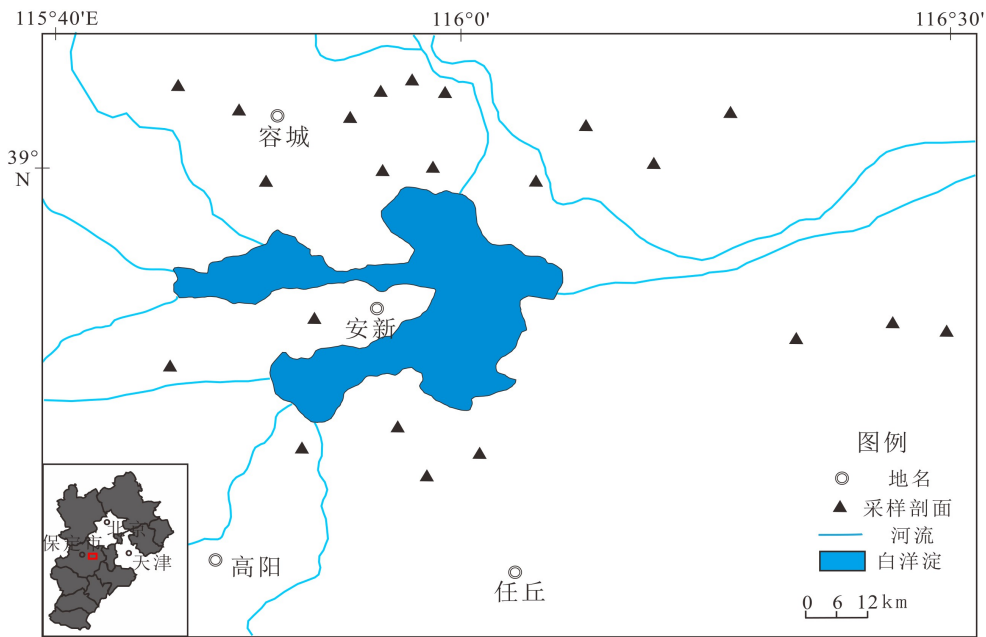


图1 研究区域采样点

Fig.1 study area and sampling sites

2.2 邻近传播算法 (AP 聚类算法)

邻近传播聚类算法又称 AP 聚类算法, 是 B J Frey 和 D Dueck 于 2007 年首次提出的一种新的聚类算法^[16]。它是一种不需要预设聚类数, 通过数据点之间的“信息传递”来寻找自己的类代表点 (exemplar), 将属于同一类代表性点的样品聚成一类, 从而实现聚类过程^[17]。

AP 聚类算法将所有的样品点都看作为潜在的类代表性点, 以数据之间的相似度系数为基础, 两两之间进行“信息传递”。在信息传递的过程中, 利用因子图(图 3)中 The Max-Sum algorithm 准则进行信息的迭代更新, 迭代收敛时, 使得公式(1)在约束条件下网络相似度 $S(c)$ 达到最大。

$$S(c) = \sum_{i=1}^N s(i, c_i) + \sum_{k=1}^N \delta_k(c)$$

$$\text{约束条件: } \delta_k(c) = \begin{cases} -\infty, & \text{当 } c_k \neq k \text{ 但是 } \exists i: c_i = k \\ 0, & \text{否则} \end{cases} \quad (1)$$

式中, c 为由类代表性点构成的向量, i 为样品点, $s(i, c_i)$ 表示样品点与其对应的类样品点之间的相似度; $\delta_k(c)$ 是约束条件, 表示数据点 i 选取 k 作为它的类代表性点, 但是数据点 k 并不是类代表性点, 因此使其值为 $-\infty$, 用于限制没有类代表性点的类簇产生。

相似性数据矩阵 $s(n, n)$ 是由 n 个数据之间的相似度系数构成, 对角线上的元素由人为设定的偏向

参数所构成, 它是用于评价各数据点作为类代表性点可能性的程度。数据之间传递的信息包括: 从潜在的类代表性点传递到数据点的归属度和从数据点传递到类代表性点的吸引度^[16]。在因子图中, 基于 The Max-Sum algorithm 准则信息更新的公式:

$$r(i, k) = s(i, k) - \max_{k' \neq k} (s(i, k') + a(i, k')), \forall i, k \quad (2)$$

$$a(i, k) = \begin{cases} \sum_{i': i' \neq k} \max\{0, (i', k)\}, & \forall k = i \\ \min\left\{0, r(k, k) + \sum_{i': i' \neq i, k} \max\{0, r(i', k)\}\right\}, & \forall k \neq i \end{cases} \quad (3)$$

当信息迭代到更新收敛时, 点 i 所对应的类代表点 c_i :

$$c_i = \arg \max \{a(i, k) + r(i, k)\}, \forall i \quad (4)$$

基于 AP 聚类算法的基本思想及其原理, 可得算法的基本步骤:

第 1 步: 根据样品点之间的相关距离, 计算相似的矩阵 S 以及设置各个样品点对应的偏向参数。

第 2 步: 设置各个样品点的初始归属度 $a(i, k) = 0$ 。

第 3 步: 利用公式(2)、(3)进行信息更新, 在更新的过程中, 为了防止迭代震荡, 引入阻尼因子 λ , 范围(0, 1)。

第 4 步: 当各个点对应的类代表性点 c_i 在迭代过程中不再发生改变或者迭代次数达到设定的最大值时, 迭代终止。将相同类代表性点的样品点归为一个类簇。

表 1 采样点描述
Table 1 Description of sampling sites

采样剖面名称	剖面位置	剖面描述	采样数/个
新庄窠村	38°59'31.56"N, 115°50'54.04"E	沉积物以灰黄色黏土质粉砂、灰黑色或灰色粉砂质黏土和浅黄色细砂为主,沉积环境主要为湖滨相、湖沼相和河床相	3
段庄村剖面	39°03'20.44"N, 115°49'30.28"E	沉积物为灰黑色粉砂质黏土、灰黄色黏土质粉砂,沉积环境为湖沼相、湖滨相	3
北张村剖面	39°04'40.23"N, 115°46'32.13"E	沉积物以灰黑色、棕黄色粉砂质黏土和灰黄色细砂为主,沉积环境为湖沼相以及河床相	4
向村剖面	39°00'05.42"N, 115°56'46.51"E	沉积物以灰黄色、浅灰色黏土质粉砂、灰褐色粉砂质黏土、灰色细砂以及灰黄色粉砂质中砂为主,沉积环境主要为湖滨相、湖沼相和河床相、洪积相	4
大阳村	39°00'18.32"N, 115°59'19.28"E	沉积物为灰黑色粉砂质黏土、浅黄色黏土质粉砂以及具有水平层理的灰黄色粉砂,沉积环境主要有湖沼相、漫滩相以及河床相	3
李庄头村	38°59'33.60"N, 116°04'48.67"E	沉积物以细砂、中砂为主,沉积环境主要为河流相	3
瞭马台村	39°04'14.65"N, 115°59'05.65"E	沉积物具有典型的“二元结构”,沉积环境为漫滩相	1
北刷村	09°04'56.42"N, 115°58'16.81"E	沉积物以浅灰色中砂为主,沉积环境为河床相	1
南文村	39°04'19.98"N, 115°56'42.30"E	沉积物以黄褐色、灰黄色黏土质粉砂以及含有黑色碳屑的灰黄色粉砂质粗砂为主,沉积环境主要有洪积相、湖相	4
马家庄村	39°02'53.77"N, 115°55'07.67"E	沉积物为灰黄色黏土质粉砂,沉积环境为湖滨相	2
胡家台村	39°02'25.71"N, 116°07'02.92"E	沉积物以灰黄色粗粉砂、细砂为主,发育交错层理,沉积环境为漫滩相	3
大步村	39°00'31.04"N, 116°10'27.99"E	沉积物为灰褐色、黑色黏土、褐黄色黏土质粉砂以及浅灰色中细砂,沉积环境为湖沼相、湖滨相、河床相	8
邢岗村	39°03'08.65"N, 116°14'18.92"E	沉积物以棕红色、灰黄色粉砂质黏土以及灰色、灰黄色黏土质粉砂为主,沉积环境为湖沼相、湖滨相	10
袁郭村	38°51'03.40"N, 116°25'15.25"E	沉积物以灰黑色粉砂质黏土、浅黄色黏土质粉砂、灰黄色中细砂为主,沉积环境为湖沼相、湖滨相	4
胡屯村	38°51'59.62"N, 116°22'34.97"E	沉积物以灰褐色、灰黑色粉砂质黏土以及灰黄色粉砂为主,沉积环境主要为湖沼相、湖滨相	5
小务村	38°51'8.86"N, 116°17'39.73"E	沉积物以棕黄色亚砂土、细粉砂、灰黑色粉砂质黏土为主,沉积环境为湖滨相、湖沼相	6
双塔村	38°45'26.20"N, 116°1'44.33"E	沉积物以灰黄色细粉砂、灰黑色黏土质粉砂为主,沉积环境为湖滨相、湖沼相	4
东良淀村	38°44'5.70"N, 115°59'7.99"E	沉积物为棕红色粉砂质黏土,沉积环境为湖相	2
北队村	38°46'22.53"N, 115°57'27.94"E	沉积物为浅灰色细砂土、棕黄色黏土质粉砂,沉积环境为漫滩相、湖滨相	3
韩堡村	38°46'23.07"N, 115°53'24.07"E	沉积物为灰黄色黏土质粉砂,沉积环境为湖滨相	2
东垒头村	38°52'16.08"N, 115°54'02.73"E	沉积物以灰黑色、棕红色粉砂质黏土、灰黄色黏土质粉砂为主,沉积环境为漫滩相、湖滨相、湖沼相	7
董庄村	38°49'32.73"N, 115°46'06.36"E	棕红色、灰黑色粉砂质黏土,沉积环境为湖沼相	3

3 沉积物粒度分布曲线的 AP 聚类算法

沉积物粒度分布曲线特征是判断沉积作用类型的重要手段之一。通常,研究者采用粒度参数代替粒度频率分布曲线,近似地作为沉积作用类型的代用指标,常常会丢失曲线中蕴含的重要信息并且具有多解性。为了充分描述曲线的分布特征,在 $-1\sim 11.3\Phi$ ($0.4\sim 2\ 000\ \mu\text{m}$) 范围内按间距 0.16Φ 划分 75 个粒级。同时,考虑到某些粒级变量的含量较低并

且在不同样品之间存在显著差异,使用对数比转换的数据,可能会掩盖住数据的结构信息,本文选用经过标准化的原始数据^[18-20]。采用欧式距离作为沉积物粒度频率分布曲线之间的相似性的度量,则样品 x 、 y 之间的相似度:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (5)$$

式中: x_{ik} , x_{jk} 分别表示沉积物样品 i , j 的第 k 项变量含量。

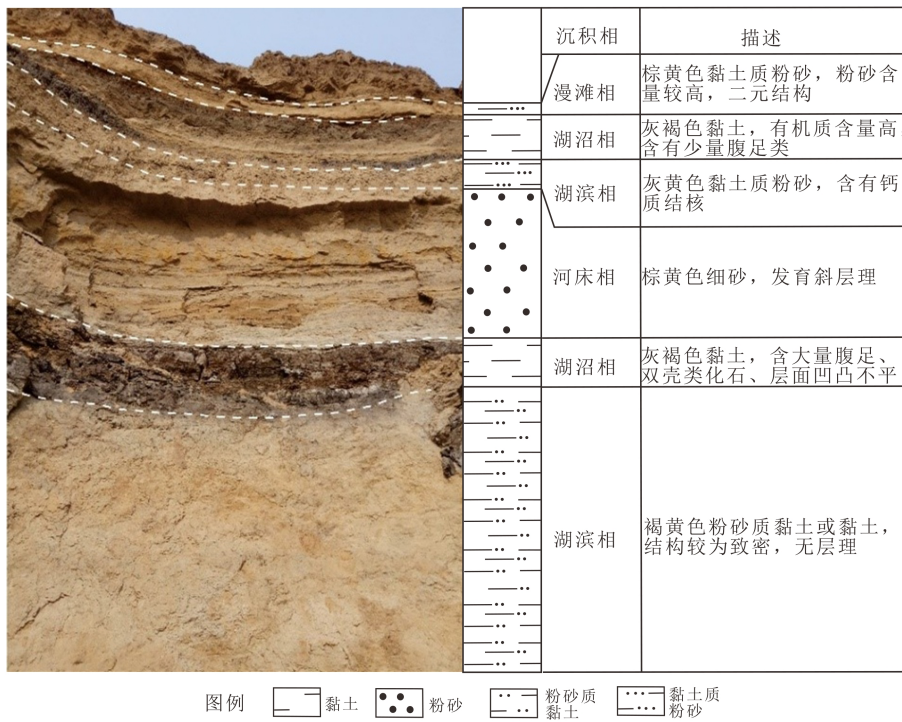


图2 大步村剖面及其地层描述

Fig.2 Dabu Village section and its stratigraphic description

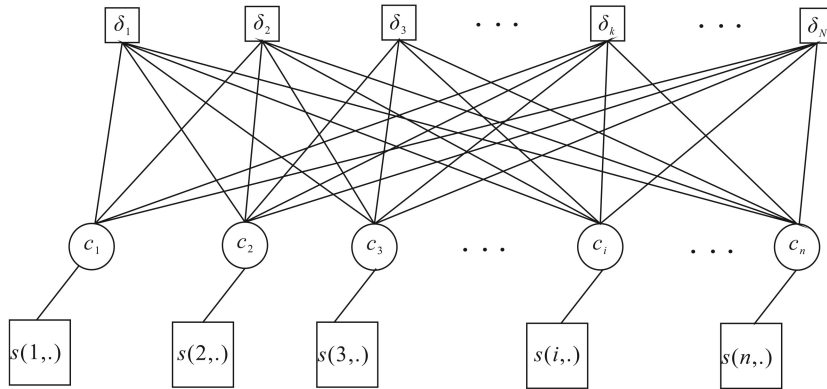


图3 邻近传播因子图

Fig.3 Factor GrAPh for affinity propagation

经过上述数据的前处理程序，AP 聚类算法将 85 个剖面典型沉积相沉积物的粒度频率曲线划分为 11 簇(图 4)。

11 簇聚类结果显示，每一簇的粒度频率分布曲线分布范围比较集中，各粒级所对应的含量之间的差异很小，粒度频率分布曲线相似度较高。根据 AP 聚类算法的聚类结果采用 silhouette 函数绘制每一个点的轮廓值。轮廓值作为聚类分析有效性的常用内部指标，它被定义为：

$$Sil(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (6)$$

式中， $a(i)$ 是第 i 个点与同簇的其他点之间的平均距离； b 为一个向量，其元素是第 i 个点与属于不同类的各个点之间的平均距离。轮廓值的取值范围为 $[-1, 1]$ ，此值越大说明该点的分类越合理^[21]。本文中 11 簇样品点的轮廓图如图 5 所示。

其中，各簇内样品点的轮廓值大部分大于 0.3，仅有 5 个点的值小于零，最小为 -0.08。考虑到划分聚类的数量以及各簇之间数据差异性相对较小，采用 AP 聚类算法对于沉积物粒度频率分布曲线的聚类划分是合理的，并且聚类效果显著。值得一提的是，AP 聚类算法可以通过调整偏向参数能够增加

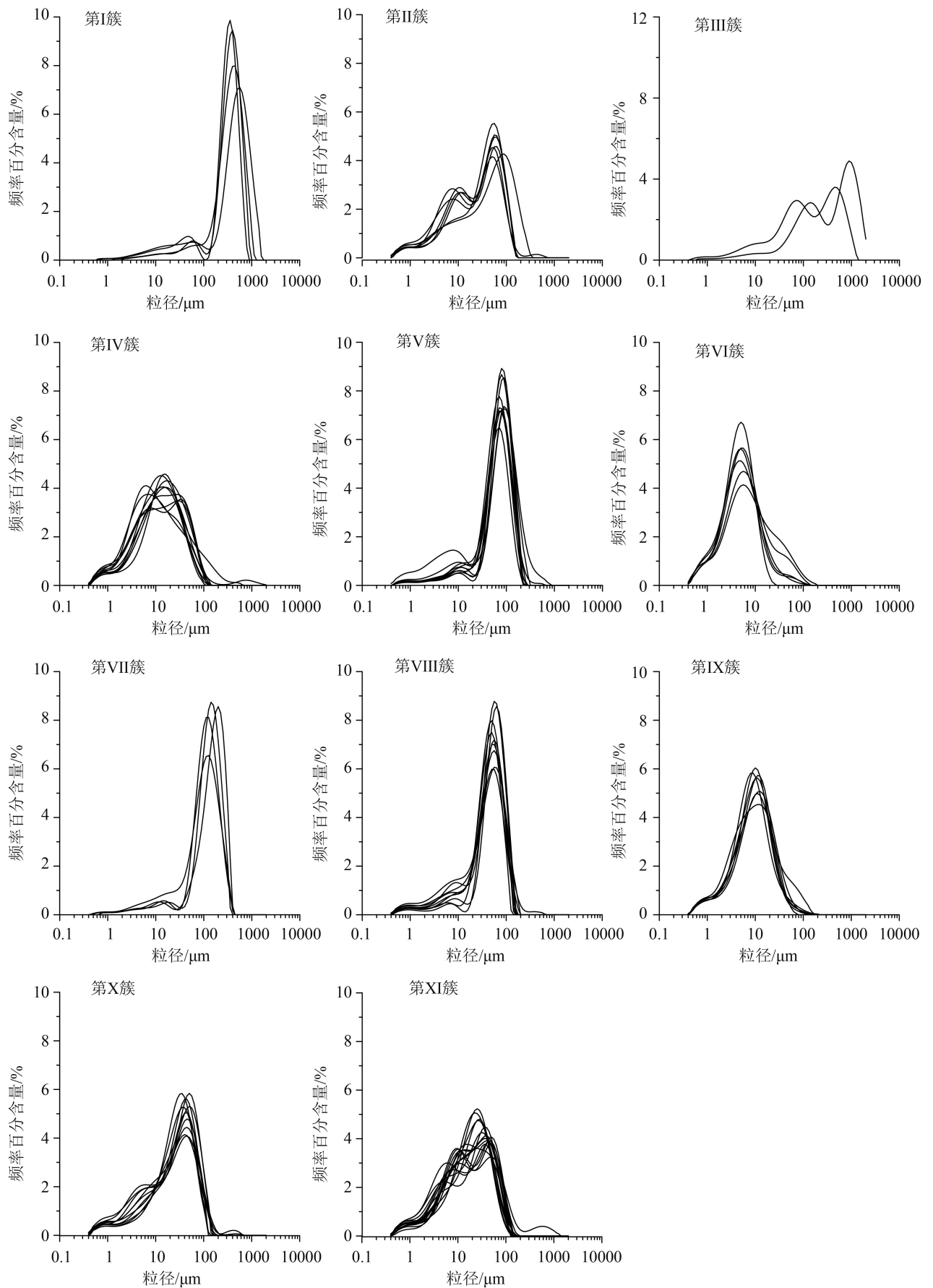


图 4 划分的 11 类簇曲线
Fig.4 11-group cluster curves

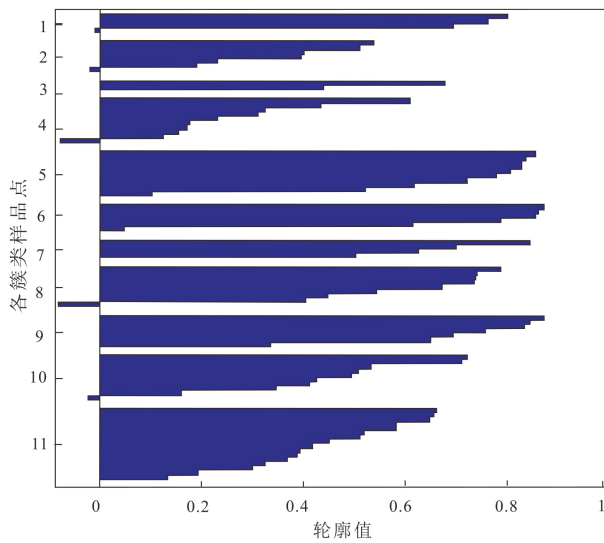


图5 各簇样品点的轮廓值

Fig.5 Silhouette value of the sample points of each cluster

或者减小聚类数量进而改变各簇内数据之间的差异大小,使得根据研究目的获得适合的聚类效果。本文选用数据点相似系数最大值作为偏向系数,以此保证各簇内数据差异很小,进而将动力条件差异不明显的沉积环境形成的粒度样品聚为一类。

4 分析与讨论

4.1 基于 AP 聚类算法的沉积物沉积相识别

不同沉积类型的粒度分布曲线各粒级对应的百分含量不同,沉积类型对应的动力条件、运输介质等差异越大则粒度分布曲线分布区间、形态差异越明显^[22]。当沉积物所受的搬运介质、搬运方式一定,并且介质动力大小稳定时,沉积物粒度组分的总体服从于某种分布的单因子随机事件的集合,在频率分布曲线和概率累计曲线形态上表现为单峰和一段式,否则呈多峰和多段式^[23-24]。如前文所述,AP 聚类算法将研究区沉积物的粒度频率分布曲线划分为 11 类簇,各簇内曲线形态、分布范围相近。但是,考虑到沉积物颗粒对于动力条件变化十分灵敏,同一沉积环境条件下粒度频率曲线会随着流体动力条件的波动在曲线形态、分布范围上产生相应的变化。因此,需要对曲线形态相似、分布范围发生轻微变化的曲线进一步归纳分析。

从 11 类簇粒度分布曲线可知,每一簇曲线与该簇的代表性点之间的差异很小,利用类代表性曲线作为每簇曲线的典型样品曲线,并按曲线峰态、

分布范围以及搬运方式将 11 簇曲线进一步划分为 4 组(图 6)。第一组包含 IV、VI、IX 三簇对应的沉积粒度频率分布曲线,曲线形式表现为单峰、一段式(图 4、6a),峰值分布于 5~20 μm,分布范围较为集中,仅存在悬浮搬运方式,表明沉积物形成于动力条件较弱的静水环境;第二组包含 II、V、VIII、X、XI 五簇沉积物粒度频率分布曲线,曲线形态为双峰、两段式(图 4、6b),主峰在 20~200 μm,次峰在 5~20 μm,跳跃和悬浮总体的搬运截点在 4Φ 左右,各总体的含量有所差异,表明沉积环境较第一组动力强度有所增加,并且水动力相对变化。第三组包括 I、VII 类簇粒度分布曲线,为双峰、两段式(图 4、6c),主峰在 100~500 μm,次峰在 10~100 μm,主要为跳跃总体存在,搬运截点变化较大,表明沉积物形成于强动力的沉积环境;第四组为 IV 类簇,为双峰、两段式(图 4、6d),跳跃与悬浮总体含量相当,表明沉积物形成于动力条件强并且变化明显的沉积环境。

根据上述分析,4 类粒度频率分布曲线的组合可能代表了 4 种较为一致的沉积环境,将各组中代表性样品粒度频率分布曲线与稳定湖相(湖心相)、湖滨相、河流相等典型沉积相粒度频率分布曲线进行对比(图 7)^[23]。图中可以看出聚类结果得到的曲线与已知沉积环境的粒度曲线主次峰位置、曲线形态、分布范围较为一致。结果表明,第一组沉积物形成于稳定湖相,包括动力条件相似的湖心相、湖沼相;第二组沉积物形成于湖滨相或漫滩相;第三组沉积物形成于河流相;第四组为洪水沉积物。由此可见,AP 聚类算法确定的 4 个组合分别代表了 4 种沉积环境。为了进一步验证上述分析的准确性与可行性,将上述结果与野外判别的 4 类 85 个沉积样品的沉积相进行对比发现,AP 聚类算法确定的 4 个组合可以将野外识别的沉积相类别聚集到一起,两者具有较好的一致性。

前人对大量已知沉积相的沉积物粒度参数进行定量描述,建立了多种沉积环境之间沉积物粒度参数判别公式。利用张平等根据典型稳定湖相与典型风成黄土沉积物粒度参数建立了稳定湖相与风成黄土粒度判别式(式 8)^[25]。本文利用此判别式对采集的所有样品进行沉积环境的判别分析。

$$F = 20.363Mz - 56.371Sd - 67.922SK + 23.516Ku - 55.626 \quad (7)$$

式中, F 表示判别值,若 $F > 0$ 为稳定湖相沉积物; Mz 、 Sd 、 SK 以及 Ku 分别是利用矩值法计算以 Φ 为单位的平均值、分选系数、偏态以及峰态。

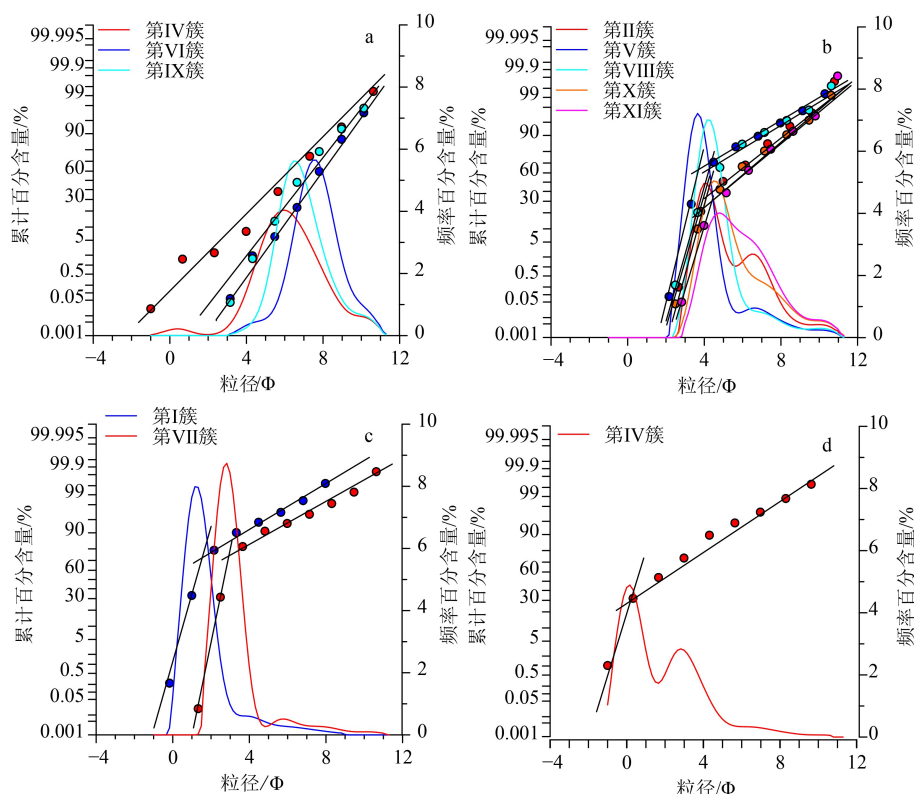


图 6 四组代表性样品频率曲线、概率曲线

Fig.6 Frequency distribution and probability cumulative curves of four groups of representative samples

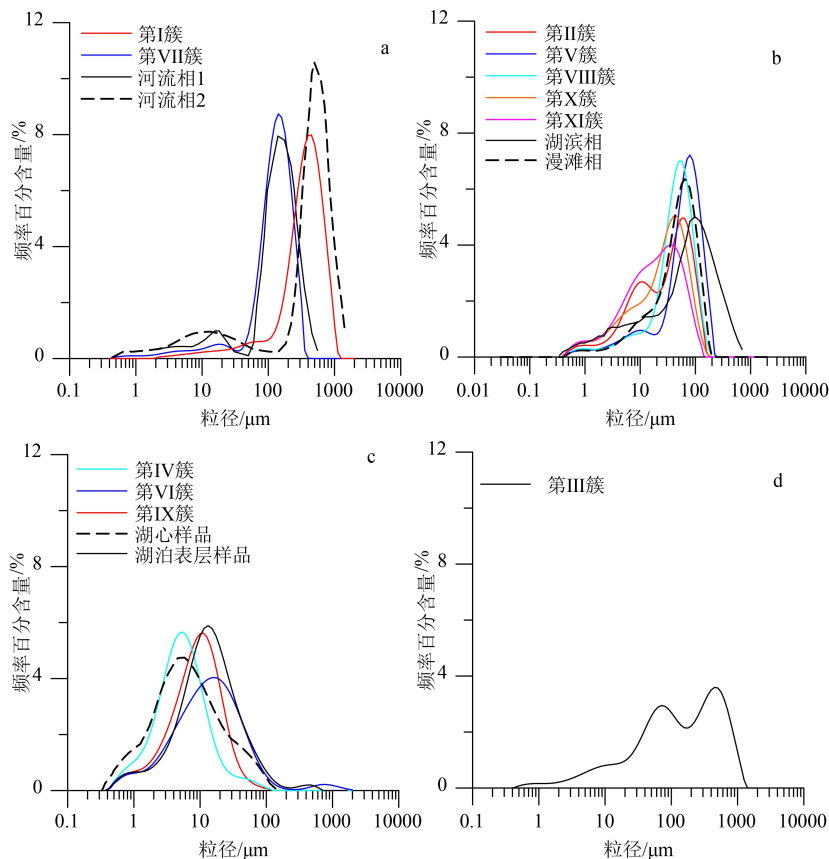


图 7 各簇代表性与典型沉积相粒度频率分布曲线对比

Fig.7 Comparison of frequency distribution curves of typical sedimentary facies and representative sediments in each cluster

其中第Ⅳ、Ⅵ、Ⅸ簇为正值,代表稳定湖相,判别结果与对比法得出的结果一致。

4.2 研究区不同沉积环境的沉积物粒度特征

基于粒度频率分布曲线的邻近传播聚类算法将全部85个粒度样品聚集成4类,并与已知沉积相粒度频率曲线进行对比,确定4类曲线对应的沉积环境分别为稳定湖相、湖滨相或漫滩相、河流相以及洪积相四种沉积相。各类沉积物之间在粒度频率分布曲线形态、分布范围以及组成成分上存在差异,通过定量描述这些差异能够为沉积环境、类型的区分提供依据^[26]。粒度频率分布曲线利用粒度参数进行定量描述,粒度参数主要有粒度平均值、分选系数、偏态和峰态。为了更加充分揭示粒度参数对于沉积环境的指示意义,选用物理意义明确的Folk-Ward图解法进行计算(表2)^[27]。表3给出了聚类后类代表性样品的粒度组成成分、粒度参数统计。

从表3中可以看出聚类结果的11簇样品中,第Ⅱ、Ⅴ、Ⅷ、Ⅹ、Ⅺ簇样品之间粒度参数和组成成分相近,第Ⅳ、Ⅵ、Ⅸ簇样品之间粒度参数和组成成分相近,第Ⅰ、Ⅶ簇样品与第Ⅲ簇样品粒度分布形成于高能环境。4类典型沉积相的粒度组成与粒度参数之间存在显著差异。

第一类包括Ⅳ、Ⅵ、Ⅸ簇沉积物,粒度组分以细粉砂为主,黏土、细粉砂和粗粉砂三者含量之和占95%以上,平均粒径 $6.227\sim 7.537\Phi$,在4类沉积环境中粒度成分最细,分选系数为 $1.301\sim 1.722$,分选

差,除Ⅸ簇中部分样品呈正偏外,其余全为近对称分布,峰态变化于 $1.066\sim 1.158$,第Ⅳ簇呈常峰态,其余全为尖峰态。

第二类包括Ⅱ、Ⅴ、Ⅷ、Ⅹ、Ⅺ簇沉积物,其中Ⅱ、Ⅷ、Ⅹ、Ⅺ簇沉积物以粗粉砂为主,占40%以上,黏土、细粉砂、细砂含量均大于10%,四者之和占98%以上,平均粒径为 $5.255\sim 5.841\Phi$,第Ⅴ簇沉积物以细砂(52%)为主,粗粉砂(34%)次之,平均粒径为 4.213Φ ,较Ⅱ、Ⅷ、Ⅹ、Ⅺ簇沉积物颗粒稍粗,表明强度略大于Ⅱ、Ⅷ、Ⅹ、Ⅺ簇沉积物对应的动力环境。该组分选系数为 $1.343\sim 1.799$,分选差,偏态为 $0.209\sim 0.467$,属于正偏、极正偏,峰态为 $0.887\sim 1.561$,第Ⅱ簇沉积物粒度分布曲线呈宽峰态,第Ⅴ、Ⅷ簇粒度分布曲线呈尖峰态或极尖峰态,其余两簇为常峰态。

第三类包括Ⅰ、Ⅶ簇沉积物,该组粒度组分较粗并且含量差异相对较大。其中,第Ⅰ簇沉积物以

表2 Folk&Ward图解法公式

Table 2 Formulas of Folk-Ward graphic methods

粒度参数	计算公式
平均粒径 (Mz)	$Mz = \frac{\Phi_{16} + \Phi_{50} + \Phi_{84}}{3}$
分选系数 (Sd)	$Sd = \frac{\Phi_{84} - \Phi_{16}}{4} + \frac{\Phi_{95} - \Phi_5}{6.6}$
偏态 (SK)	$SK = \frac{1}{2} \left[\frac{\Phi_{16} + \Phi_{84} + 2\Phi_{50}}{\Phi_{84} - \Phi_{16}} + \frac{\Phi_5 + \Phi_{95} + 2\Phi_{50}}{\Phi_{95} - \Phi_5} \right]$
峰态 (Ku)	$Ku = \frac{1}{2.44} \left[\frac{\Phi_{95} - \Phi_5}{\Phi_{75} - \Phi_{25}} \right]$

式中,粒径为 Φ 值粒径, $\Phi = -\log_2^d$, d 为毫米直径; Φ_x 为 $x\%$ 累积含量的 Φ 值粒径。

表3 各簇代表性样品粒度组成成分及粒度参数

Table 3 The size composition and size parameters of representative samples in each cluster

簇类	类代表性样品	沉积相	粒度成分/%						平均粒径/ Φ	分选系数	偏态	峰态
			黏土	细粉砂	粗粉砂	细砂	中砂	粗砂				
I	SITE01-01	河流相	1.27	3.72	6.95	18.33	45.34	24.38	1.462	1.155	0.344	1.584
II	SITE03-03	湖滨相	12.72	25.35	40.23	21.47	0.2	0.02	5.255	1.75	0.32	0.887
III	SITE04-06	洪积相	2.85	6.15	15.43	29.62	16.16	29.8	1.582	1.976	0.319	0.908
IV	SITE12-02	湖沼相	19.67	41.22	34.45	4.33	0.14	0.13	6.227	1.722	0.096	1.066
V	SITE13-08	漫滩相	4.72	7.81	34.25	52.54	0.58	0.09	4.213	1.515	0.467	1.561
VI	SITE13-09	湖沼相	40.67	47.21	10.66	1.46	0	0	7.537	1.301	0.019	1.158
VII	SITE14-04	河流相	2.58	5.38	12.08	72.54	7.42	0	2.919	1.108	0.343	1.797
VIII	SITE17-04	漫滩相	6.68	10.46	55.59	27.18	0.08	0.01	4.577	1.343	0.363	1.485
IX	SITE18-01	湖心相	20.74	55.49	22.38	1.4	0	0	6.819	1.329	0.175	1.123
X	SITE21-06	湖滨相	12.75	22.99	50.4	13.71	0.13	0.02	5.507	1.799	0.374	1.014
XI	SITE18-02	湖滨相	15.1	33.35	43.32	7.98	0.11	0.12	5.841	1.745	0.209	0.981

中砂(45%)为主,细砂(18%)、粗砂(22%)次之,平均粒径为1.462,样品分选差,呈极正偏,为尖峰态,代表强动力沉积环境;第Ⅶ簇沉积物中细砂含量占70%以上,平均粒径为2.919 Φ ,颗粒较第Ⅰ簇沉积物偏细,表明介质动力条件强并且稳定,相比于Ⅰ簇沉积物对应的沉积环境动力条件较低。

第四类中第Ⅲ簇沉积物以细砂为主,粗粉砂、中砂、粗砂含量均大于15%,平均粒径为1.582 Φ ,分选系数1.976,样品分选差,极正偏,宽峰态,表明沉积动力复杂多变且动力较强。

4.3 AP 聚类算法在沉积环境识别中应用的可行性

根据上述方法对粒度分布曲线直接进行聚类分析,避免人为干扰聚类效果,有效地揭示研究区内的沉积环境差异是此方法的优点,也是能够利用

此方法对粒度分布曲线进行研究区所存在的沉积环境类型反演的出发点。本文以白洋淀地区作为研究对象,尝试利用AP聚类算法建立粒度分布曲线与沉积环境之间的内在联系,通过识别结果与野外识别进行对比,发现所得结果较为一致,85个样品中有6个不一致。如下给出了沉积环境识别过程中存在差异的沉积物样品信息(表4)以及样品对应的粒度分布曲线(图8)。

其中, SITE01-03、SITE14-01、SITE14-02 三个样品的粒度曲线具有明显双峰,分布范围和曲线形态与典型湖滨相或漫滩相较为符合,野外观察结果存在识别错误(图8)。SITE03-04、SITE05-01、SITE13-02 样品峰值粒径比典型湖沼相粒径较粗,但是没有明显次峰(图8)。对比其他簇曲线形态及分布范围,此类样品更为可能形成于沉积环境相对稳定,

表4 沉积环境识别过程中存在差异的沉积物样品

Table 4 Sediment samples with differences in the process of sedimentary environment identification

样品编号	簇类	野外人为识别	粒度分布曲线对比识别
SITE01-03	II	灰黑色粉砂质黏土, 湖沼相沉积	湖滨相或漫滩相
SITE03-04	II	灰黑色黏土, 湖沼相沉积	湖滨相或漫滩相
SITE05-01	XI	灰黄色粉砂, 河流相沉积	湖滨相或漫滩相
SITE13-02	XI	灰黄色粉砂, 河流相沉积	湖滨相或漫滩相
SITE14-01	XI	灰黑色粉砂质黏土, 湖相沉积	湖滨相或漫滩相
SITE14-02	II	湖沼相沉积	湖滨相或漫滩相

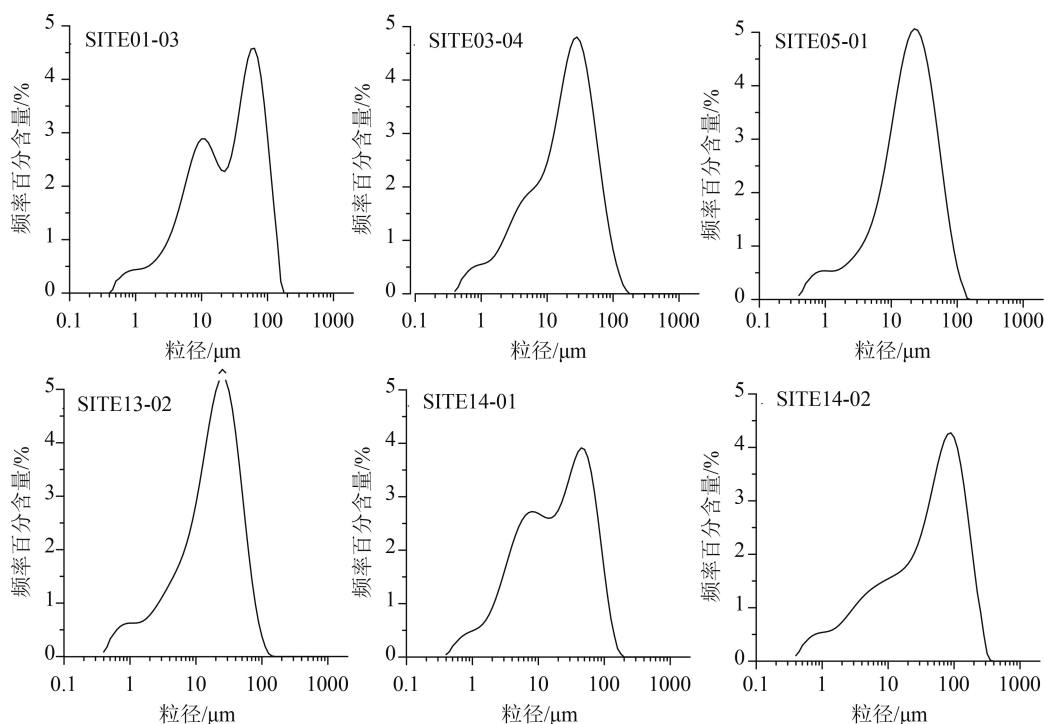


图8 沉积环境识别过程中存在差异的样品粒度曲线

Fig.8 Sediment samples with differences in sedimentary environment identification

但动力强度较强的湖沼相沉积环境。AP聚类算法进行聚类分析时,由于样品相对典型湖沼相样品粒径较粗,被划分到湖滨相沉积物一类,出现了偏差。因此,对于85个样品中有3个样品沉积环境识别出现误差,错误率为3.53%。

鉴于聚类方法是尽可能将相似的数据点聚集成一类,沉积物的形成受物源、介质等多个因素影响,同一沉积环境条件下,粒度分布曲线会在一定范围内发生波动,因此采用邻近传播聚类分析时需要合理选取偏向系数控制各簇之间数据的差异。本文为了充分挖掘粒度数据中蕴含的沉积环境类型,选取了相似系数最大值作为偏向系数,尽可能减小数据之间差异,导致出现同一沉积环境条件下形成的粒度曲线可能被划分到不同簇中的现象。

5 结论

(1) AP聚类算法具有不需要预设聚类数量,能够准确地确定研究区沉积环境中存在的动力类型,将曲线形态、分布范围相近的沉积物聚成一类。每一簇的曲线形态和分布范围都能反演出研究区对应的沉积环境存在的动力类型,并且建立沉积物粒度特征与沉积动力学之间的内在联系,为沉积物类型划分提供依据。

(2) 根据粒度概率累积曲线以及分布曲线特征,将11簇曲线中形态、分布范围相近的曲线进一步合并,得到4类曲线,并与已知沉积环境样品曲线进行对比,确定出白洋淀地区剖面地层中出露着湖沼相、湖相、河流相以及洪积相四种沉积相。

(3) 白洋淀地区存在的4类典型沉积环境中湖心相、湖沼相的动力环境相近,以细粉砂为主,分选差,动力强度最低;湖滨相以粗粉砂为主,双峰明显,分选差。河流相颗粒较粗,以细、中砂为主,沉积物粒度特征反映漫滩相沉积环境与湖滨相相似;洪积相具有明显双峰,含量相近,分选差,水动力条件动荡。

由此可见,基于粒度频率分布曲线的邻近聚类传播算法可以作为挖掘研究区存在的沉积类型的新方法,为沉积环境动力条件反演、分区等提供潜在的新手段。

参考文献 (References)

[1] Weltje G J, Prins M A. Muddled or mixed? Inferring palaeoclimate from size distributions of deep-sea clastics [J]. *Sedimentary Geology*, 2003, 162(1-2): 39-62.

[2] 范天来, 范育新. 频率分布曲线和概率累积曲线在沉积物粒度数据分析中应用的对比[J]. *甘肃地质*, 2010, 19(2): 32-37. [FAN Tianlai, FAN Yuxin. A comparison of grain size expression methods: A case study [J]. *Gansu Geology*, 2010, 19(2): 32-37.]

[3] 曲政. 沉积物粒度数据表征方法的研究[J]. *中国粉体技术*, 2001, 7(4): 24-31. [QU Zheng. A study on characterization methods of grain-size data of sediment [J]. *China Powder Science and Technology*, 2001, 7(4): 24-31.]

[4] 孙东怀, 鹿化煜, David R, et al. 中国黄土粒度的双峰分布及其古气候意义[J]. *沉积学报*, 2001, 18(3): 327-335. [SUN Donghuai, LU Huayu, David R, et al. Bimode grain-size distribution of Chinese loess and its paleoclimate implication [J]. *Acta Sedimentologica Sinica*, 2001, 18(3): 327-335.]

[5] Weltje G J. End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem [J]. *Mathematical Geology*, 1997, 29(4): 503-549.

[6] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. *软件学报*, 2008, 19(1): 48-61. [SUN Jigui, LIU Jie, ZHAO Lianyu. Clustering algorithms research [J]. *Journal of Software*, 2008, 19(1): 48-61.]

[7] 李玉中, 陈沈良. 系统聚类分析在现代沉积环境划分中的应用——以崎岖列岛海区为例[J]. *沉积学报*, 2003, 21(3): 487-494. [LI Yuzhong, CHEN Shenliang. Application of system cluster analysis to classification of modern sedimentary environment: A case study in Qiqu Archipelago Area [J]. *Acta Sedimentologica Sinica*, 2003, 21(3): 487-494.]

[8] 杨传慧, 吉根林, 章志刚. AP算法在图像聚类中的应用研究[J]. *计算机与数字工程*, 2012, 40(10): 119-121. [YANG Chuanhui, JI Genlin, ZHANG Zhigang. Research on application of algorithm AP in images clustering [J]. *Computer and Digital Engineering*, 2012, 40(10): 119-121.]

[9] 吴娱, 钟诚, 尹梦晓. 基因表达数据的分层近邻传播聚类算法[J]. *计算机工程与设计*, 2016, 37(11): 2961-2966. [WU Yu, ZHONG Cheng, YIN Mengxiao. Gene expression data clustering algorithm using hierarchical affinity propagation [J]. *Computer Engineering and Design*, 2016, 37(11): 2961-2966.]

[10] 肖宇, 于剑. 基于近邻传播算法的半监督聚类[J]. *软件学报*, 2008, 19(11): 2803-2813. [XIAO Yu, YU Jian. Semi-supervised clustering based on affinity propagation algorithm [J]. *Journal of Software*, 2008, 19(11): 2803-2813.]

[11] 许清海, 陈淑英, 孔昭宸, 等. 白洋淀地区全新世以来植被演替和气候变化初探[J]. *植物生态学与地植物学学报*, 1988, 12(2): 143-151. [XU Qinghai, CHEN Shuying, KONG Zhaochen, et al. Preliminary discussion of vegetation succession and climate change since the Holocene in the Baiyangdian lake district [J]. *Acta Phytocologica et Geobotanica Sinica*, 1988, 12(2): 143-151.]

[12] 王永, 闵隆瑞, 董进, 等. 河北白洋淀全新统沉积特征与地层划分[J]. *地球学报*, 2015, 36(5): 575-582. [WANG Yong, MIN Longrui, DONG Jin, et al. Sedimentary characteristics and stratigraphic division of Holocene series in Baiyang Dian, Hebei Province [J]. *Acta Geoscientica Sinica*, 2015, 36(5): 575-582.]

[13] 杨慧君, 王永, 迟振卿, 等. 河北白洋淀老河头剖面25.5 kaBP以来气候环境变化的沉积记录[J]. *现代地质*, 2015, 29(2): 291-298.

- [YANG Huijun, WANG Yong, CHI Zhenqing, et al. Sedimentary record of climate change during the past 25.5 ka of Laohetou profile from Baiyangdian, Hebei Province [J]. *Geoscience*, 2015, 29 (2): 291-298.]
- [14] 高彦春, 王金凤, 封志明. 白洋淀流域气温、降水和径流变化特征及其相互响应关系[J]. *中国生态农业学报*, 2017, 25 (4): 467-477. [GAO Yanchun, WANG Jinfeng, FENG Zhiming. Variation trend and response relationship of temperature, precipitation and runoff in Baiyangdian Lake Basin [J]. *Chinese Journal of Eco-Agriculture*, 2017, 25 (4): 467-477.]
- [15] 何乃华, 朱宣清. 白洋淀地区近3万年来的古环境与历史上人类活动的影响[J]. *海洋地质与第四纪地质*, 1992, 12 (2): 79-88. [HE Naihua, ZHU Xuanqing. Palaeoenvironment changes since 30000 aBP and effects of human activities in the Baiyangdian area [J]. *Marine Geology and Quaternary Geology*, 1992, 12 (2): 79-88.]
- [16] Frey B J, Dueck D. Clustering by passing messages between data points [J]. *Science*, 2007, 315 (5814): 972-976.
- [17] Brusco M J, Köhn H F. Comment on "Clustering by passing messages between data points" [J]. *Science*, 2008, 319 (5864): 726.
- [18] Nelson P A, Bellugi D, Dietrich W E. Delineation of river bed-surface patches by clustering high-resolution spatial grain size data [J]. *Geomorphology*, 2014, 205: 102-119.
- [19] Baxter M J, Beardah C C, Cool H E M, et al. Compositional data analysis of some alkaline Glasses [J]. *Mathematical Geology*, 2005, 37 (2): 183-196.
- [20] Ordóñez C, Ruiz-Barzola O, Sierra C. Sediment particle size distributions apportionment by means of functional cluster analysis (FCA) [J]. *Catena*, 2016, 137: 31-36.
- [21] 朱连江, 马炳先, 赵学泉. 基于轮廓系数的聚类有效性分析[J]. *计算机应用*, 2010, 30 (2): 139-141. [ZHU Lianjiang, MA Bingxian, ZHAO Xuequan. Clustering validity analysis based on silhouette coefficient [J]. *Journal of Computer Applications*, 2010, 30 (2): 139-141.]
- [22] 张晋, 李安春, 万世明, 等. 南海南部表层沉积物粒度分布特征及其影响因素[J]. *海洋地质与第四纪地质*, 2016, 36 (2): 1-10. [ZHANG Jin, LI Anchun, WAN Shiming, et al. Grain size distribution of surface sediments in the southern South China Sea and influencing factors [J]. *Marine Geology and Quaternary Geology*, 2016, 36 (2): 1-10.]
- [23] 殷志强, 秦小光, 吴金水, 等. 中国北方部分地区黄土、沙漠沙、湖泊、河流细粒沉积物粒度多组分分布特征研究[J]. *沉积学报*, 2009, 27 (2): 343-351. [YIN Zhiqiang, QIN Xiaoguang, WU Jinshui, et al. The multimodal grain-size distribution characteristics of loess, desert, lake and river sediments in some areas of Northern China [J]. *Acta Sedimentologica Sinica*, 2009, 27 (2): 343-351.]
- [24] 孙东怀, 安芷生, 苏瑞侠, 等. 古环境中沉积物粒度组分分离的数学方法及其应用[J]. *自然科学进展*, 2001, 11 (3): 269-276. [SUN Donghuai, AN Zhisheng, SUN Ruixia, et al. Mathematics method and its application of grain size distribution of Paleoenvironment Sediments [J]. *Progress in Natural Sciences*, 2001, 11 (3): 269-276.]
- [25] 张平, 宋春晖, 杨用彪, 等. 稳定湖相沉积物和风成黄土粒度判别函数的建立及其意义[J]. *沉积学报*, 2008, 26 (3): 501-507. [ZHANG Ping, SONG Chunhui, YANG Yongbiao, et al. The significance and establishment of discriminant function with grain size of stable lacustrine sediment and eolian loess [J]. *Acta Sedimentologica Sinica*, 2008, 26 (3): 501-507.]
- [26] 卢连战, 史正涛. 沉积物粒度参数内涵及计算方法的解析[J]. *环境科学与管理*, 2010, 35 (6): 54-60. [LU Lianzhan, SHI Zhengtao. Analysis for sediment grain size parameters of connotations and calculation method [J]. *Environmental Science and Management*, 2010, 35 (6): 54-60.]
- [27] Blott S J, Pye K. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments [J]. *Earth Surface Processes and Landforms*, 2001, 26 (11): 1237-1248.