

现场海域人工鱼礁分布状态聚类分析

沈天跃, 章守宇*, 沈蔚, 汪振华

(上海海洋大学海洋科学学院, 上海 201306)

摘要: 人工鱼礁的流场营造、鱼类诱集等建设目标, 一般是通过单位鱼礁、鱼礁群等不同规模的形式实现的; 实际投放的成千上百个人工鱼礁处于何种分布状态、它们是否满足单位鱼礁等规模要求, 是非常值得关注的课题。因此, 如何对人工鱼礁的实际分布状态进行合理的划分归类或删除, 以最大程度地贴合设计方案的配置组合方式, 对于正确评价鱼礁投放的准确与否是非常必要并具有现实意义的。本研究基于 C3D 侧扫声呐系统采集的高清晰水下影像, 结合 ArcGIS 的数据矢量化功能, 提取出各个鱼礁单体在水下的空间位置以及各鱼礁之间的相互距离与方位等空间关系。在此基础上, 利用空间聚类分析, 借助基于划分、层次和约束的算法对投放后的鱼礁进行划分、归类或删除, 并选择单位鱼礁的重心、影响面积、鱼礁单体数量以及礁体间距 4 个指标进行比较分析, 探明人工鱼礁实际的组合聚类模式。结果显示, 3 种空间聚类算法误差的排列顺序为约束算法 < 划分算法 < 层次算法, 其值分别为 0.093, 0.203, 0.264。通过对比分析, 在基于约束的聚类算法下, 最能反映人工鱼礁的实际集聚情况。

关键词: 人工鱼礁; 单位鱼礁; 空间聚类

中图分类号: S 931.4

文献标志码: A

人工鱼礁的流场营造, 鱼类诱集等效应往往是在特定布局和规模条件下, 由多个单体形成一定空方数的单位鱼礁、以及单位鱼礁构成的鱼礁群等实现。单位鱼礁是指由若干鱼礁单体在一定的范围内随意堆叠或有规则组合而成的具有实质效应的最小鱼礁规模^[1]; 每个单位鱼礁在海域可以形成局部的、相对独立的人工复合生态系统^[2]。

实际海域投放的人工鱼礁, 通过 C3D 测扫声呐等现代仪器调查发现, 这些鱼礁单体在表象上往往呈现不规则的凌乱状态, 主要是由于海况条件和投放技术的限制以及其他人为等因素造成的^[3]。这样的分布是否满足单位鱼礁乃至鱼礁群等不同规模的要求, 现阶段研究尚难以进行定量判断。怎样从成百上千个人工鱼礁实际分布状态中, 合理区分各个单位鱼礁和鱼礁群, 是量化人工鱼礁投放误差、评价礁体实际投放是否准确、是否满足礁区建设要求的前提。但是, 目前国内外对投放后的人工鱼礁如何进行划分归类, 揭示人

工鱼礁实际的分布状态, 还未见有相关研究报道。

尽管实际投放的人工鱼礁在表象上呈现散乱分布状态, 但各个鱼礁单体之间在实际空间上存在空间自相关性; 距离较近的鱼礁属于同一类的可能性较大, 即满足地理学第一定律“空间上距离近的实体间相似性比距离远的实体的相似性大”^[4]。而空间聚类主要是通过空间数据的空间自相关性, 以某个特定标准(如距离准则)把一个数据集分割成不同的簇^[5], 使得同一个簇内数据对象的相似性尽可能大, 不同的簇中数据对象的差异性尽可能大^[6-7], 因而发现空间实体自然的空聚集模式^[8], 揭示空间实体的分布规律, 提取空间实体的群体空间结构特征, 预测空间实体的发展变化趋势^[4,9-16]。目前, 存在许多空间聚类方法, 不同的聚类方法由于主要思想的差异, 聚类结果各不相同。

本研究利用 C3D 声呐系统, 获取每个鱼礁的三维状态, 运用划分聚类、层次聚类和约束聚类分

收稿日期:2014-11-11

修回日期:2015-04-21

资助项目:公益性行业(农业)科研专项(201203063, 201003068)

通信作者:章守宇, E-mail: syzhang@shou.edu.cn

析鱼礁之间的集聚模式,对投放后的鱼礁进行划分、归类或剔除,比较每种聚类结果的精度,以期获得人工鱼礁实际分布状态最好的集聚模式,为后续的鱼礁投放误差分析提供依据。

1 材料与方 法

1.1 数据获取

利用 C3D 声呐系统对东海某一人工鱼礁区(调查区域的大小为 1 236 m × 917 m)进行现场测量,快速获取水下每个鱼礁的三维位置和状态,获得大范围水深和水下声呐图像信息(图 1)。海上 C3D 声呐调查是在该海域人工鱼礁投放后一个月进行的,在鱼礁投放与声呐调查期间未经历台风等重大海况影响,因此,获得的影像资料可被认为是未经发生滑坡、倾覆、损毁等情况的投放现状。基于 C3D 提供的高清晰水下声呐数据,结合 ArcGIS 矢量化功能,人工提取图像中每个鱼礁点在水下的空间位置以及它们之间的相互距离与方位等空间关系,对数据集中参与聚类的对象进行数据清理(图 2)。

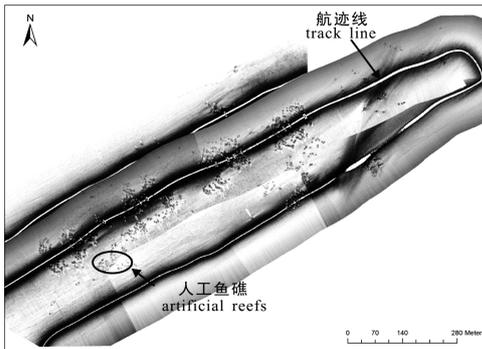


图 1 C3D 鱼礁成像

Fig.1 The C3D measured image for artificial reefs

1.2 空间聚类方法

基于划分的聚类算法是历史最为悠久的一种聚类算法,也是当前应用最为广泛的聚类算法之一,其中最经典的划分方法是 K-means^[4,17-20]。K-means 算法的核心内容在于迭代优化过程,不断优化平方误差准则,最终获得 k 个划分^[21]。K-means 算法的设计过程如下:

(1) 选定 k 个实体,每个实体视为簇(类)的质心。

(2) 按照距离最近原则,将剩余实体赋给最近的簇,重新计算每个簇的质心^[4,22-23]。

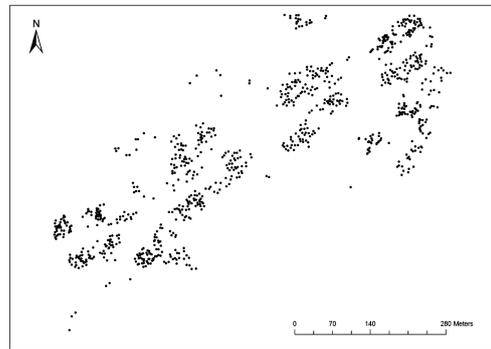


图 2 人工鱼礁分布

Fig.2 The distribution of artificial reefs

(3) 不断重复步骤(2),直到下式的平方误差准则收敛为止。

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

式中, E 表示平方误差准则, p 为空间实体, m_i 为簇 C_i 的质心, k 为簇的数目, $i=1,2,3,\dots,k$ 。

基于层次的聚类方法也是一类比较常用的空间聚类算法,通过反复分裂或聚合操作来获得满足一定要求的空间聚类结果,分为 2 种形式:凝聚法和分裂法^[4]。

基于约束的聚类是指在满足用户某种特定条件下的聚类。可以看作是在非监督分类中加入了知识,成为半监督分类方法^[24]。对于任意给定的平面散乱点集,经过 Delaunay 三角剖分得到的三角网是唯一的^[25],并且空外接圆和最大最小角这 2 个性质保证 Delaunay 三角网中的三角形近似等边三角形^[26],因此得到的 Delaunay 三角网是公认具有整体最优性质的三角网^[27]。实体间空间邻近关系的定义是空间聚类的核心问题, Delaunay 三角网是空间聚类中实体间邻近关系表达的一种有效方法^[28]。Delaunay 三角网生成时不仅考虑到距离,生成后又包含了拓扑关系,所以可以采用 Delaunay 三角网来表达空间数据的空间特征。事物主要受两方面的影响:整体因素和局部因素。据此,为了得到人工鱼礁的分布情况,基于约束的算法借助 Delaunay 三角网,通过施加不同层次、不同类型的约束准则与相应的参数,删除三角网中不符合约束条件的边,消除整体因素和局部因素的影响,形成相对稀疏的簇(即群体),再依据簇间邻近原则重新合并簇。依据空间自相关性,对原始数据集(图 3 - a)构造

Delaunay 三角网(图 3-b),施加整体约束条件和局部约束条件 2 种不同层次,以及距离和个数 2 种不同类型的约束准则,将 Delaunay 三角网划分

为多个簇(图 3-c),再根据距离准则合并簇,最终将数据集划分为 2 个簇(图 3-d)。

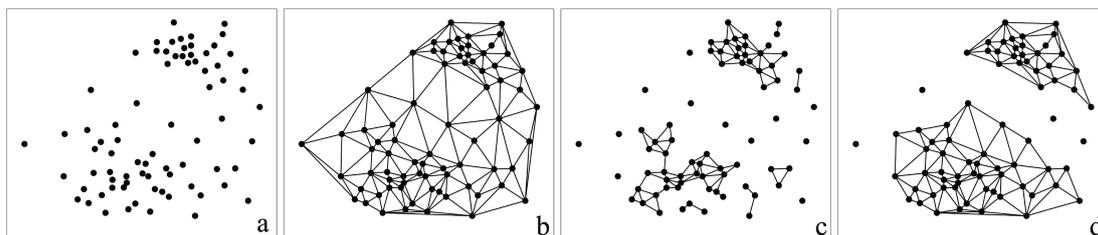


图 3 聚类算法步骤图

(a)数据集;(b)构造 Delaunay 三角网;(c)生成稀疏图;(d)合并簇

Fig.3 Clustering algorithm steps

(a) data set;(b) constructing delaunay;(c) generating sparse graph;(d) merging clusters

1.3 检验标准

聚类结果的重叠区域为 3 种聚类算法的共有区域,处于共有区域中的空间实体之间自相关性高,分布相对紧密,不会根据聚类算法的不同导致聚类结果存在差异,因此,重叠区域反映出单位鱼礁的实际分布状况。而各种聚类算法与重叠区域之间的差距则反映出各种算法的差别。计算 3 种聚类结果与重叠区域的差异,根据差异的大小来判断 3 种聚类方法结果的优劣。

2 结果

各种方法得到的聚类结果大体相似但又存在差异,总体上分为 24 个单位鱼礁,大部分鱼礁呈集聚分布,存在部分分布较为离散的单位鱼礁(图 4)。利用 ArcGIS 中的区域叠置功能得出 3 种聚类算法的重叠区域,重叠区域的个数达 35 个(图 5)。

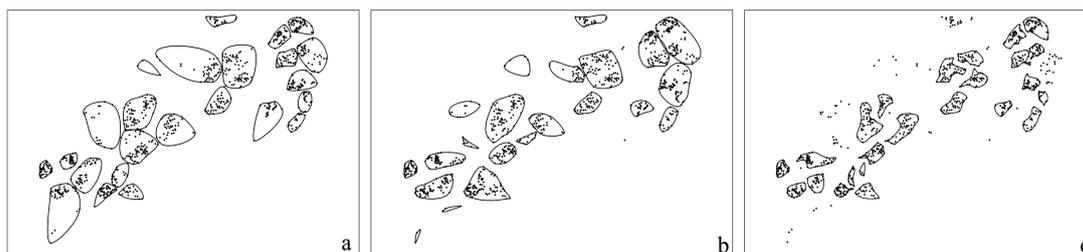


图 4 3 种聚类算法结果

(a)划分算法;(b)层次算法;(c)约束算法

Fig.4 The results for 3 clustering algorithms

(a) partition algorithm;(b) hierarchical algorithm;(c) constrained algorithm

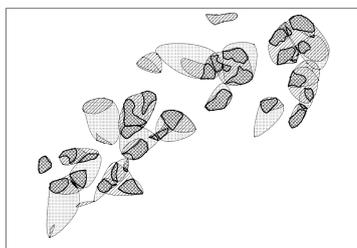


图 5 3 种聚类算法重叠区域

Fig.5 The overlap regions for 3 clustering algorithms

2.1 单体数量误差

用 3 种方法对人工鱼礁聚类结果显示,层次聚类误差波动较大,其次为划分聚类,约束聚类波动最小,误差的排序为约束 < 划分 < 层次,其值分别为 0.187,0.221,0.329(图 6)。基于约束的算法具有最小的标准差,即基于约束的算法与重叠区域之间的差异最小,相似性最高(图 6)。

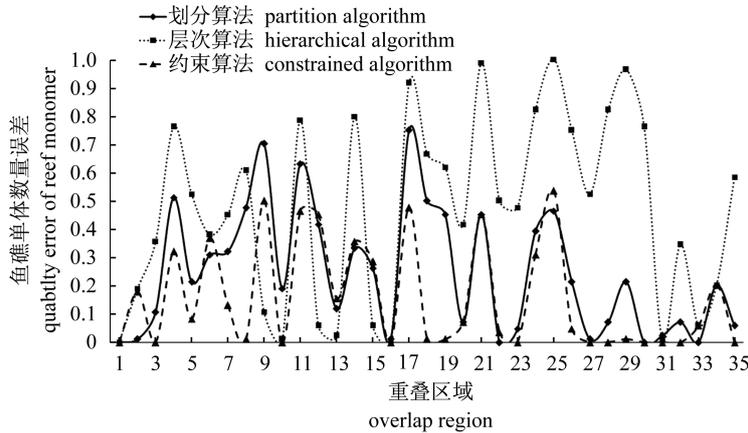


图 6 3 种聚类算法单体数量误差

Fig. 6 The quantity errors of reef monomer for 3 clustering algorithms

2.2 单位鱼礁重心误差

利用 3 种方法对人工鱼礁进行空间聚类,并 利用标准差对 3 种聚类方法获取的聚类结果进行 误差检验。3 种聚类方法的误差波动差异较小, 划分聚类误差波动略高于层次聚类和约束聚类。

误差的排序为层次 < 约束 < 划分,其值分别为 0.220,0.243,0.251。层次聚类标准差最小,因此,该方法得出的聚类重心与重叠区域的重心最 接近(图 7)。

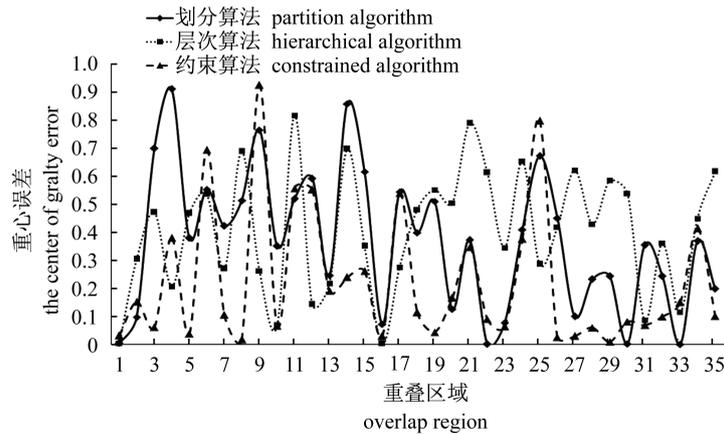


图 7 3 种聚类算法重心误差

Fig. 7 The center of gravity errors for 3 clustering algorithms

2.3 单位鱼礁面积误差

结果显示,约束聚类误差曲线较平缓,最大误差 不超过 0.4,总体误差值明显小于其他 2 种方法,划分聚类和层次聚类误差较大(图 8)。误差的 排序为约束 < 划分 < 层次,其值分别为 0.118, 0.301,0.313。约束聚类得到最小的标准差,因此,从面积因素考虑,约束聚类得到的聚类结果最 接近鱼礁实际分布情况。

体分布的凝散程度,礁体之间间距小,说明鱼礁 分布集中,反之鱼礁分布分散。用 3 种聚类方法 得出的礁体间距误差差异比较小,在第 9 个重叠 区域,3 种聚类算法的误差值都达到最大值,说明 第 9 个重叠区域中鱼礁单体之间的间距较小(图 9)。误差的排序为层次 < 约束 < 划分,其值 分别为 0.220,0.225,0.266。因此,从鱼礁单体 间距误差方面考虑,层次聚类最能反映鱼礁分 布的凝散程度。

2.4 鱼礁单体间距误差

单位鱼礁中鱼礁单体间距的大小反映出礁

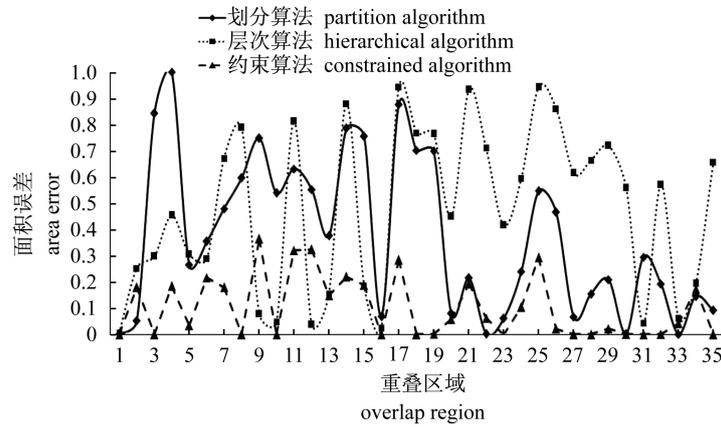


图 8 3 种聚类算法面积误差
Fig. 8 Area errors for 3 clustering algorithms

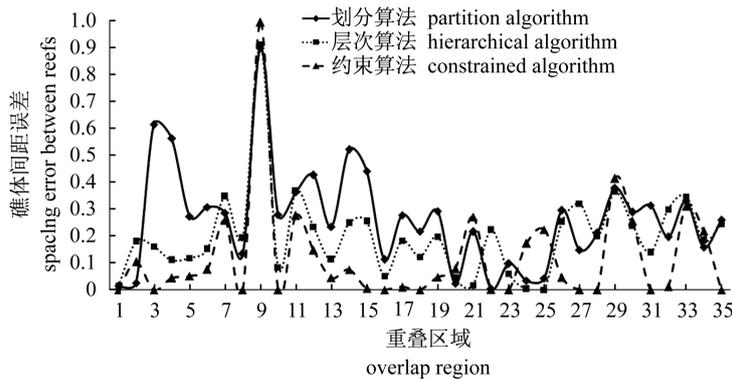


图 9 3 种聚类算法礁体间距误差
Fig. 9 Spacing errors between reefs for 3 clustering algorithms

2.5 总体误差分析

对 3 种聚类算法的影响面积、重心、礁体间距、鱼礁单体数量进行主成分分析,确定各个指标的影响程度。结果显示,在划分聚类中,4 个成分的贡献率分别为 65.958%、27.187%、4.870%、1.985% (表 1)。第一主成分与重心存在明显的正相关,相关性达到 0.929,因此第一主成分可以被认为重心的代表。同理,第二主成分代表礁体间距,第三主成分代表单体数量,第四主成分代表面积。层次聚类中第一主成分代表单体数量,第二主成分代表重心,第三主成分代表礁体间距,第四主成分代表面积。4 个成分的贡献率分别为 65.878%、22.478%、9.594%、2.050%。约束聚类中第一主成分代表面积,第二主成分代表礁体间距,第三主成分代表重心,第四主成分代表单体数量。4 个成分的贡献率分别为 76.351%、20.198%、2.314%、1.136%。

根据主成分分析得到的各指标的贡献率,建立误差模型, x_1, x_2, x_3, x_4 分别表示面积、单体数量、重心、礁体间距, Y_1, Y_2, Y_3 分别表示划分聚类总体误差、层次聚类总体误差以及约束聚类总体误差,如下公式所示。

划分聚类:

$$Y_1 = 0.0198 \cdot x_1 + 0.0487 \cdot x_2 + 0.6595 \cdot x_3 + 0.2718 \cdot x_4$$

层次聚类:

$$Y_2 = 0.0205 \cdot x_1 + 0.6587 \cdot x_2 + 0.2247 \cdot x_3 + 0.0959 \cdot x_4$$

约束聚类:

$$Y_3 = 0.7635 \cdot x_1 + 0.0113 \cdot x_2 + 0.0231 \cdot x_3 + 0.2019 \cdot x_4$$

根据模型得出 3 种聚类算法的总体误差,约束聚类总体误差低于划分聚类和层次聚类,约束聚类误差曲线相对比较平缓,误差最大值低于 0.4

(图 10)。层次聚类在第 11 以及 17~30 重叠区域中误差值都远高于其他 2 种聚类方法,在第 22 个重叠区域误差值达到最大。标准差的排序为约束 < 划分 < 层次,其值分别为 0.093,0.203,

0.264。因此,在综合考虑面积、单体数量、重心、礁体间距这 4 个指标时,约束聚类误差值最小,精度最高。使用约束聚类得到的聚类结果最能反映鱼礁实际分布状态的集聚模式。

表 1 3 种聚类算法 4 个主成分负荷值和贡献率
Tab.1 Loadings and contribution rates of four components for 3 clustering algorithms

算法 algorithm	指标 index	成份 component			
		1	2	3	4
划分算法 partition algorithm	重心 centre of gravity	0.929	0.089	0.348	-0.092
	面积 area	0.825	0.207	0.428	0.305
	礁体间距 spacing between reefs	0.111	0.991	-0.069	0.020
	单体数量 quantity of reef monomer	0.509	-0.139	0.849	0.035
	贡献率/% contribution rate	65.958	27.187	4.870	1.985
层次算法 hierarchical algorithm	重心 centre of gravity	0.403	0.911	0.081	0.013
	面积 area	0.884	0.355	0.179	0.244
	礁体间距 spacing between reefs	0.121	0.069	0.990	0.011
	单体数量 quantity of reef monomer	0.929	0.321	0.088	-0.161
	贡献率/% contribution rate	65.878	22.478	9.594	2.050
约束算法 constrained algorithm	重心 centre of gravity	0.849	-0.328	0.415	0.006
	面积 area	0.980	-0.138	0.031	-0.143
	礁体间距 spacing between reefs	-0.179	0.982	-0.062	-0.004
	单体数量 quantity of reef monomer	0.968	-0.179	0.067	0.160
	贡献率/% contribution rate	76.351	20.198	2.314	1.136
	累计贡献率/% cumulative contribution rate	76.351	96.549	98.864	100.000

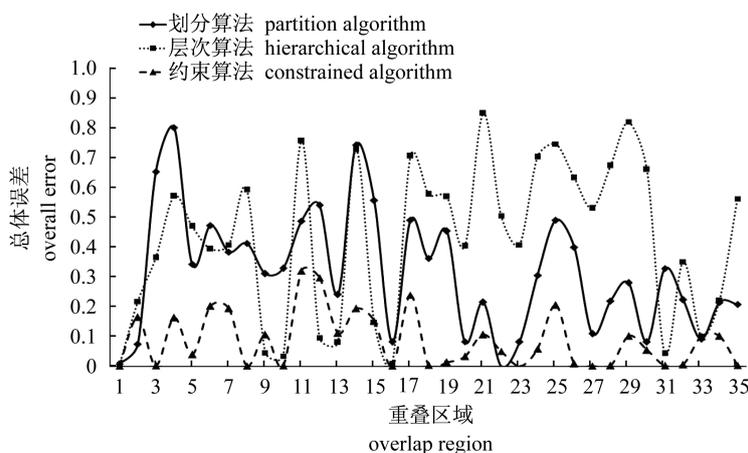


图 10 3 种聚类算法总体误差
Fig.10 Overall errors for 3 clustering algorithms

3 讨论

人工鱼礁投放是人工鱼礁建设的重要环节,是联系鱼礁设计与实际建设情况的纽带。由于人

工鱼礁投放后的实际分布位置与设计方案位置不吻合,常常发生偏移,存在误差,其中人工投放施工技术是最主要的原因。目前,在人工鱼礁投放过程中经常出现随意投放的现象,导致人工鱼礁

建设无法按照设计方案实施。通过侧扫声呐图像获得的人工鱼礁水下分布情况只能得到鱼礁单体实际的位置信息,主观定性判断鱼礁实际分布与设计方案的差异。而如何客观准确判断由若干鱼礁单体组成的单位鱼礁在实际海域中处于何种分布状态,目前尚未见有相关研究报道,这无疑成为评价人工鱼礁投放施工技术的阻碍。本研究利用空间聚类算法对人工鱼礁进行划分、归类,提取出单位鱼礁实际分布情况,是对人工鱼礁分布规律探讨的一个新尝试,从而使人工鱼礁在空间聚类方面的研究进一步深入,为今后定量描述人工鱼礁投放误差特性、评价人工鱼礁投放质量和建设效果提供了前提条件。

空间聚类算法主要以空间实体的空间自相关性为基础,根据各自的判断准则(例如距离)进行划分归类,不同的算法由于准则的差异,聚类的结果各不相同。Kaufman等^[29]指出划分聚类易受异常点或噪声点的干扰,而Guha等^[30]也提出层次聚类对异常数据敏感,无法剔除零散点,因此在对鱼礁聚类时,无法排除距离相对较远的鱼礁单体,得到的聚类结果中由于个别零散单体的存在,单位鱼礁面积较大。此外,单位鱼礁中礁体之间的距离也会出现偏大现象。而当鱼礁之间的间距较远时,鱼礁的协同作用降低,因此在聚类过程中应排除零散鱼礁单体。而约束聚类通过设置不同层次的约束条件,避免了聚类结果中面积和礁体间距过大的问题。郭景峰等^[31]指出层次聚类的伸缩性较差,对数据输入顺序敏感,层次聚类得到的单位鱼礁中单体数量差异较大,分布不均匀。在单位鱼礁边界区域,鱼礁分布较分散,密度较低,层次聚类无法区分出不同密度的空间簇,容易将多个单位鱼礁划分到同一个簇中,进而一些单位鱼礁的单体数量过多,与此同时,另外一些单位鱼礁的单体数量随之改变。约束聚类在进行划分时,添加了单体数量这一约束条件,使得各个单位鱼礁的单体数量差距不显著。划分聚类通过迭代过程反复计算聚类的重心,得到的重心位置不稳定,最终的重心位置与初始聚类的重心位置可能存在较大的差异。约束聚类根据不同层次和不同类型的约束条件进行单位鱼礁的划分、归类,划分的结果更为合理。

本研究利用基于划分的算法、基于层次的算法以及基于约束的算法,对实际的人工鱼礁进行

聚类分析,得到东海人工鱼礁的实际分布现状。从尽可能全面表征单位鱼礁实际存在状态的几个要素,如重心、影响面积、鱼礁单体数量以及礁体间距作为评估几种聚类方法结果的标准。重心反映单位鱼礁主要分布的位置;影响面积反映出单位鱼礁的影响范围;日本有关专家认为单位鱼礁作为鱼礁能产生效应的最小单位规模至少需要达到 400 m^3 ^[1],在给定鱼礁单体的尺寸时,单位鱼礁的单体数量进而就反映了鱼礁的规模,鱼礁单体数量的增多或减少,鱼礁的规模也随之改变;而礁体间距则反映鱼礁单体之间分布的紧密程度,鱼礁之间间距具有一定的要求,较大的间距,礁体之间的协同作用降低,间距较小,鱼礁之间分布相对集中,又会导致鱼礁形成的面积缩小。最终得出约束聚类总体较好,精度最高,最能反映实际人工鱼礁集聚模式。

本研究数据处理得到了上海海洋大学林军老师的大力帮助,在此表示感谢!

参考文献:

- [1] Yin Z Q. Study on the theory and method of effect evaluation of the artificial reef [D]. Shanghai: Shanghai Ocean University, 2010. [尹增强. 人工鱼礁效果评价理论与方法的研究. 上海:上海海洋大学, 2010.]
- [2] Wang L H. Research on the layout and adaptive management of artificial reef area in Dongshan sea area [D]. Xiamen: Xiamen University, 2012. [王麟鹤. 东海海域人工鱼礁区的布局与适应性管理研究. 厦门:厦门大学, 2012.]
- [3] Xia Z Y. Artificial reef engineering [M]. Beijing: China Ocean Press, 2011: 112 - 119. [夏章英. 人工鱼礁工程学. 北京:海洋出版社, 2011: 112 - 119.]
- [4] Liu Q L. A methodology of adaptive spatial clustering analysis [D]. Changsha: Central South University, 2011. [刘启亮. 自适应空间聚类方法研究. 长沙:中南大学, 2011.]
- [5] Yu L. Research of Clustering algorithm based on improved minimum spanning trees [D]. Harbin: Harbin University of Science and Technology, 2008. [于亮. 基于改进的最小生成树聚类算法研究. 哈尔滨:哈尔滨理工大学, 2008.]
- [6] Feng X P, Zhang T F. Comparison of four clustering methods [J]. Microcomputer & Its Applications, 2010, 29(16): 1 - 3. [冯晓蒲, 张铁峰. 四种聚类方

- 法之比较. 微型机与应用, 2010, 29(16): 1-3.]
- [7] Ma H Y. Research on clustering algorithm based on ant colony optimization [D]. Beijing: Beijing University of Technology, 2009. [马鸿雁. 基于蚁群优化的聚类算法的研究. 北京: 北京工业大学, 2009.]
- [8] Zhang L. Research on GML data clustering algorithm based on spatial topological relations [D]. Nanjing: Nanjing Normal University, 2010. [张丽. 基于拓扑关系的 GML 空间面对象聚类算法研究. 南京: 南京师范大学, 2010.]
- [9] Deng M, Liu Q L, Li G Q, *et al.* Analysis and application of spatial clustering [M]. Beijing: Science Press, 2011: 10-76. [邓敏, 刘启亮, 李光强等. 空间聚类分析及应用. 北京: 科学出版社, 2011: 10-76.]
- [10] Ma N N. Research on fuzzy clustering analysis algorithm based on density [D]. Baotou: Inner Mongolia University of Science & Technology, 2012. [马娜娜. 基于密度的模糊聚类分析算法研究. 包头: 内蒙古科技大学, 2012.]
- [11] Xiao W. EM algorithm and its initialization research for parameter estimation of Gaussian mixture models [J]. Electronic Test, 2011, (6): 26-30. [肖维. 用于高斯混合模型参数估计的 EM 算法及其初始化研究. 电子测试, 2011, (6): 26-30.]
- [12] Wang H. Research on improved C-means clustering algorithm [D]. Kaifeng: Henan University, 2011. [王慧. C-均值聚类算法的改进研究. 开封: 河南大学, 2011.]
- [13] Wang X H, Shen J, Wang R B, *et al.* A study on removing duplication using N-garm terms for Chinese text [J]. Journal of Hangzhou Dianzi University, 2010, 30(1): 26-29. [王小华, 沈杰, 王荣波, 等. 一种新的基于蚁群和凝聚的混合聚类算法. 杭州电子科技大学学报, 2010, 30(1): 26-29.]
- [14] Ma C. A survey of spatial clustering research [J]. Computer Technology & Development, 2009, 19(4): 134-137. [马程. 空间聚类研究. 计算机技术与发展, 2009, 19(4): 134-137.]
- [15] Xie J Y, Jiang S, Wang C X, *et al.* An improved global K-means clustering algorithm [J]. Journal of Shanxi Normal University: Natural Science Edition, 2010, 38(2): 18-22. [谢娟英, 蒋帅, 王春霞, 等. 一种改进的全局 K-均值聚类算法. 陕西师范大学学报: 自然科学版, 2010, 38(2): 18-22.]
- [16] Li Z J. An improved K-means clustering algorithm [J]. Pioneering With Science & Technology Monthly, 2012, (1): 134-135. [李卓君. 一种改进的 K-means 算法. 科技创业月刊, 2012, (1): 134-135.]
- [17] Huang Y J, Tang D Y, Hu H W, *et al.* Analyzing gene sequence data by dynamically iterative clustering algorithms [J]. Journal of University of South China: Science & Technology, 2004, 18(4): 57-60. [黄元江, 汤德佑, 胡红武等. 动态迭代聚类算法分析基因序列数据. 南华大学学报: 自然科学版, 2004, 18(4): 57-60.]
- [18] Li W C, Zhou Y, Xia S X, *et al.* A novel clustering algorithm based on hierarchical and K-means clustering [C] // Proceedings of the 26th Chinese Control Conference. Beijing: Beihang University Press, 2007: 605-609. [李文超, 周勇, 夏士雄, 等. 一种新的基于层次和 K-means 方法的聚类算法第 26 届中国控制会议论文集. 北京: 北京航空航天大学出版社, 2007: 605-609.]
- [19] Lloyd S. Least squares quantization in PCM [J]. Information Theory, 1982, 28(2): 129-137.
- [20] MacQueen J. Some methods for classification and analysis of multivariate observations [C] // Proceedings of the 5th Berkeley symposium on mathematical statistics and probability. California: University of California Press, 1967, 1(14): 281-297.
- [21] Chen K. Target clustering based on cluster analysis applied research [D]. Xi'an: Xidian University, 2013. [陈奎. 基于聚类分析的目标分群问题的应用研究. 西安: 西安电子科技大学, 2013.]
- [22] Xu X Y. Application of K-means algorithm in relational database [J]. Journal of Guilin University of Electronic Technology, 2008, 28(4): 313-316. [徐向阳. K-均值聚类算法在关系数据库中的应用. 桂林电子科技大学学报, 2008, 28(4): 313-316.]
- [23] Qu X H, Gao W L, Ding B R, *et al.* The K-means algorithm improvement based on the number of clustering K and initial centroid [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2011, (4): 42-46. [屈新怀, 高万里, 丁必荣, 等. 基于聚类数和初始值的 K-means 算法改进研究. 组合机床与自动化加工技术, 2011, (4): 42-46.]
- [24] Huang Z M. Research on interactive spatial clustering algorithm with constraints [D]. Beijing: China Agricultural University, 2007. [黄志敏. 带约束条件

- 的交互式空间聚类算法研究. 北京: 中国农业大学, 2007.]
- [25] Liu G J. Research and implementation of spatial clustering analysis system SDCAS based on SADBS [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2005. [刘国俭. 基于 SADBS 的空间聚类分析系统 SDCAS 的研究与实现. 南京: 南京航空航天大学, 2005.]
- [26] Fu G X. Spatio-temporal 3D terrain modeling based on measuring data of AUV [D]. Harbin: Harbin Engineering University, 2011. [傅桂霞. 基于 AUV 测量信息的时空 3D 数据地形构建. 哈尔滨: 哈尔滨工程大学, 2011.]
- [27] Zang B. 3D Reconstruction based on laser data of mobile robot [D]. Dalian: Dalian University of Technology, 2005. [臧波. 基于移动机器人激光测距数据的物体三维重建. 大连: 大连理工大学, 2005.]
- [28] Liu Q L, Deng M, Shi Y, *et al.* A novel spatial clustering method based on multi-constraints [J]. *Acta Geodaetica et Cartographica Sinica*, 2011, 40 (4): 509 - 516. [刘启亮, 邓敏, 石岩, 等. 一种基于多约束的空间聚类方法. 测绘学报, 2011, 40 (4): 509 - 516.]
- [29] Kaufman L, Rousseeuw P J. Finding groups in data: An introduction to cluster analysis [M]. New York: Wiley - Interscience, 2005: 68 - 72.
- [30] Guha S, Rastogi R, Shim K. CURE: An efficient clustering algorithm for large databases [C] // Association of Computing Machinery Special Interest Group on Management of Data Record. Association of Computing Machinery, 1998, 27 (2): 73 - 84.
- [31] Guo J F, Zhao Y Y, Bian W, *et al.* A Hierarchical clustering algorithm based on improved cluster cohesion and separation [J]. *Journal of Computer Research and Development*, 2008 (Suppl. 1): 202 - 206. [郭景峰, 赵玉艳, 边伟峰, 等. 基于改进的凝聚性和分离性的层次聚类算法. 计算机研究与发展, 2008 (增刊 1): 202 - 206.]

Clusteranalysis on distribution status of artificial reefs in field sea area

SHEN Tianyue, ZHANG Shouyu*, SHEN Wei, WANG Zhenghua
(College of Marine Science, Shanghai Ocean University, Shanghai 201306, China)

Abstract: The flow field variation and fish attraction effects in artificial reefs are usually realized through different scales of forms such as unit fish reef and fish reef cluster. What distribution status of the hundreds and thousands of artificial reefs actually launched is and if such reefs meet the requirement of unit reef worth very much attention. So, how to conduct reasonable classify or eliminate reefs and take the maximum extent to fit design scheme, is very necessary and has practical significance for evaluating whether ARs are launched accurately. In this paper, based on high-definition underwater images collected by C3D side-scan sonar system, combining with ArcGIS, we can extract the submarine space position of each fish reef monospace as well as spatial relationships i. e. mutual distance and azimuth between them etc. On this basis, a trial on spatial cluster analysis based on partition, hierarchical and constrained algorithms, was compared in order to find a relatively suitable solution for interpretation of distribution. Centre of gravity for deployed reefs, their potential influenced area, number of unit reefs and distance between each of the two reefs are defined as four basic indices for further comparison. Results show that the error of three spatial clustering algorithms in an ascending order is: constrained algorithm < partition algorithm < hierarchical algorithm, the values were 0.093, 0.203, 0.264. Comparative analysis shows that the constrained cluster algorithm can reflect the actual distribution pattern of the artificial reefs in a best way.

Key words: artificial reefs; unit fish reef; spatial clustering

Corresponding author: ZHANG Shouyu. E-mail: syzhang@shou.edu.cn