

DOI: 10.3724/SP.J.1118.2020.20087

基于鱼类体长频率分布和平均体长的采样站位数优化

孟新翔¹, 薛莹^{1, 2, 3}, 张崇良^{1, 2, 3}, 纪毓鹏^{1, 3}, 任一平^{1, 2, 3}, 徐宾铎^{1, 2, 3}

1. 中国海洋大学水产学院, 山东 青岛 266003;
2. 青岛海洋科学与技术试点国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266237;
3. 海州湾渔业生态系统教育部野外科学观测研究站, 山东 青岛 266003

摘要: 鱼类体长数据较易获得, 常用于数据有限的渔业资源评估和管理, 且基于鱼类体长的指标可以作为渔业生态系统的生态监测指标。本文根据山东半岛南部海域底拖网季度调查收集的方氏云鲷体长数据, 应用计算机模拟重采样方法, 研究了调查站位数对估计方氏云鲷(*Enedrias fangi*)体长频率分布和平均体长的影响。结果表明, 在各个季节, 方氏云鲷体长频率分布离散指数(DI)变化范围为 0~0.91, 大部分站位的 DI 值为 0.75~0.90, 说明调查站位间体长频率分布较为相似, 可以对调查站位数进行优化, 提高采样效率; 调查站位越多, 体长频率分布和平均体长估计精确度越高; 由于对不同体长数据指标估计的精度不同, 用以估计体长频率分布和平均体长的样本量存在差异; 20 个和 30 个调查站位分别为估计方氏云鲷体长频率分布和平均体长的可接受样本量。本研究结果可为基于不同体长指标的渔业资源调查设计提供一定的理论支撑, 有助于对基于体长指标有效站位数的初步了解, 为山东半岛南部海域的鱼类资源科学调查和科学管理提供参考。

关键词: 体长频率分布; 相对估计误差; 样本量优化; 方氏云鲷; 山东半岛

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2020)11-1333-08

鱼类体长数据常包括体长频率分布、比例资源密度(proportional stock density, PSD)和平均体长等^[1-3]。鱼类体长频率分布表示不同体长组的鱼类个体数量或者比例^[1], 其常用于渔业资源评估和管理, 如用于估算鱼类生物量^[4]和构建种群动态模型^[5]等。因此, 鱼类体长频率分布数据质量可能影响资源评估结果, 进而影响渔业管理的有效性。

关于鱼类体长频率分布估计精度的影响因素已有较多研究。体长分组组距影响鱼类体长频率分布估计的有效样本量, 例如, Miranda^[6]研究表明不同体长分组组距影响估计鱼类体长频率分布的样本量, 组距 2.5 cm 比组距 1 cm 时需要更小样本量; Wang 等^[7]研究表明, 在相同精度下, 不同

体长分组间隔时估计鱼类体长频率分布所需的样本量不同。底拖网调查方式对估计鱼类体长频率分布精度会产生影响, 底拖网调查采样可认为是鱼类种群大小特征的整群抽样方法, 利用“整群”捕捞获取鱼类体长数据; 某一网次鱼类个体通常比整个海域的鱼类个体具有更相似的体长组成。Pennington 等^[8]指出由于鱼类集群行为, 底拖网调查采样获得的体长频率分布数据精度差, 用于估算鱼类种群体长频率分布的有效个体数远远小于实际调查采样鱼类个体数量, 通过减少各站位采样鱼类个体数和增加体长分布差异较大的采样站位数, 能提高鱼类体长频率分布数据的采样效率。Zhang 等^[9]对 6 种底栖鱼类的体长数据分布有效样本量进行研究, 结果表明, 增加采样站位数

收稿日期: 2020-03-31; 修订日期: 2020-05-21.

基金项目: 国家重点研发计划项目(2019YFD0901204); 山东省支持青岛海洋科学与技术试点国家实验室重大科技专项(2018-SDKJ0501-2).

作者简介: 孟新翔(1994-), 男, 硕士研究生, 主要从事渔业资源与生态学研究. E-mail: mengxinxiang@stu.ouc.edu.cn

通信作者: 徐宾铎, 副教授, 主要从事渔业资源与生态学研究. E-mail: bdxu@ouc.edu.cn

量比增加每个站位内测定的个体数量可以更好地提高采样估计精度。Pennington 等^[10]估算了基于平均体长指标所需要的样本量,表明提高精度的唯一方法是增加捕捞船数量,而不是增加每条船采集的鱼类个体数量。

本文根据山东半岛南部海域底拖网季度调查收集的方氏云鲷体长(*Enedrias fangi*)数据,利用计算机模拟重采样方法,研究采样站位数对方氏云鲷体长频率分布和平均体长估计的影响,以期 为山东半岛南部海域鱼类资源调查设计及其优化提供理论参考。

1 材料与方 法

1.1 数据来 源

方氏云鲷体长数据来自 2016 年 10 月、2017 年 1 月、5 月、8 月在山东半岛南部海域进行的渔业资源底拖网调查。本次调查区域为 35°N~37°N、124°E 以西水域,共设置 63 个站位(图 1)。调查船为 220 kW 单拖网渔船,囊网网目 20 mm,拖曳时网口扩张宽度约 15 m,高度 7.5 m,拖网调查均在白天进行,拖速为 3.0 kn,每站拖网时间为 1 h。根据《海洋调查规范 第 6 部分:海洋生物调查》(GB/T12763.6-2007)^[11]进行渔获样品的保存、生物学测定和分析处理。在进行调查目标鱼种、主要经济鱼种和优势种的渔业生物学测定时,每个站位随机留取生物学测定样品 50 尾,不足 50 尾时全测,并记录体长信息,统计其体长范围、平均体长等信息,调查样本量信息见表 1。

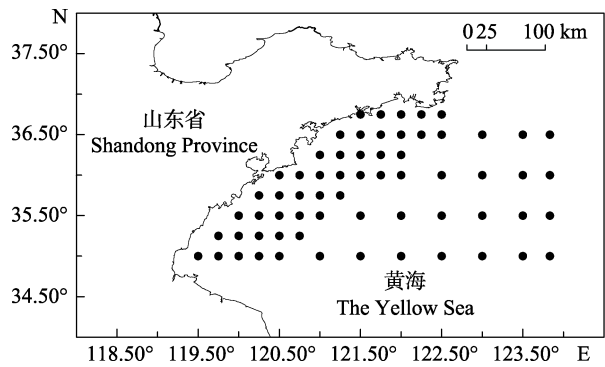


图 1 山东半岛南部海域渔业资源底拖网调查站位
Fig. 1 Sampling stations of bottom trawl survey in the southern waters off Shandong Peninsula

表 1 山东半岛南部海域方氏云鲷体长组成及样本量信息

Tab. 1 The information of body length composition and sample size for *Enedrias fangi* in the southern waters off Shandong Peninsula

季节 season	测定个体数量 number of individuals measured	出现站位数 number of occurrence stations	体长范围/ mm standard length range	平均体长 /mm mean length
春季 spring	1919	61	53–184	124.1±0.4
夏季 summer	1254	48	75–168	122.3±0.5
秋季 autumn	483	47	87–167	129.3±0.5
冬季 winter	614	55	79–168	101.0±0.7

1.1.1 体长分布离散指数 应用体长分布离散指数(dispersion index, DI)衡量各站位中方氏云鲷体长频率分布变异程度^[12]。将方氏云鲷体长按 5 mm 间隔进行分组,计算各调查站位的体长分布离散指数 DI,公式如下:

DI=∑ⁿ_{i=1}(L_i×(1-L_i))

式中, *n* 体长组数量, *L_i* 是体长频率分布图中第 *i* 体长组所占的比例。当站位内鱼类个体体长属于同一体长组时,体长频率分布比较集中,DI 值最小,为 0;当各体长组分布均匀时,DI 值最大,即该站位中体长频率分布有较大变异,随着体长组数量 *n* 的增加,则 DI 值也增加^[12]。

1.1.2 体长频率分布估计精度与站位数的关系 本次计算机模拟研究中,假设原始调查获得的方氏云鲷体长数据能够反映研究海域方氏云鲷体长大小结构,暂不考虑各站位实际测定的尾数。通过比较不同样本量下模拟数据体长频率分布与原始调查数据体长频率分布差异大小,分析体长频率分布估计精度与站位数的关系。模拟研究步骤主要包括: (1)分季节将出现方氏云鲷的调查站位(表 1)的方氏云鲷体长数据进行汇总,以汇总数据中方氏云鲷体长最大、小值作为上下界限,按 5 mm 体长间隔分组,计算调查原始体长频率分布数据“真值”; (2)应用 Bootstrap 重抽样方法从调查站位中有放回地随机选取一定数量的调查站位,对抽取站位的体长数据进行汇总,按步骤(1)的相同方法获得模拟体长频率分布数据,将模拟体长频率分布数据在各体长组内的体长频率减去相对

应的原始调查数据在各体长组内的体长频率, 计算二者在所有体长组的体长频率的绝对差值和 (sum of absolute difference, SAD); (3) 上述步骤(2) 重复 1000 次, 求得各站位数下重采样体长频数分布与原始体长频数分布的 1000 次绝对差值和 SAD 分布, 以解释模拟抽样数据体长频率分布偏离原始数据体长频率分布的程度; (4) 依次降低站位数, 重复步骤(2) 和(3), 计算不同站位数时的 SAD, 分析不同站位数(1, 5, 10, ..., 55, 60, 63) 对体长频率分布估计精度的影响。

1.1.3 平均体长估计与站位数的关系 以方氏云鲷平均体长为估计目标, 分析不同站位数对平均体长估计精度的影响。主要步骤包括: (1) 分季节, 筛选出现方氏云鲷的调查站位(表 1), 然后计算各站位的方氏云鲷平均体长, 作为各站位的平均体长原始数据; (2) 从原始数据中有放回地随机选取 n ($n=1, 5, 10, \dots, 55, 60, 63$) 个调查站位, 获得 n 个站位的平均体长模拟数据, 根据该模拟数据计算研究海域的方氏云鲷平均体长模拟值 $L_i^{\text{estimated}}$, 每个站位数时重复抽样 1000 次, 通过比较平均体长模拟值 $L_i^{\text{estimated}}$ 与平均体长真值 L^{true} , 计算不同站位数时的平均体长估计值的相对估计误差 (relative estimation error, REE) 值; (3) 重复步骤(2) 100 次, 得到 100 次 REE 值的分布。

研究海域的方氏云鲷平均体长“真值” L^{true} 计算公式如下^[13]:

$$L^{\text{true}} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^k y_{ij}$$

式中, N 为方氏云鲷的测定总尾数, n 为调查站位数, k 为各个调查站位中测定的方氏云鲷尾数, y_{ij} 为第 i 个调查站位中测量的第 j 尾鱼的体长。

1.2 评价指标

应用绝对差值和 SAD, 评价重抽样模拟数据体长频率分布偏离原始数据体长频率分布的程度。

绝对差值和 SAD 计算公式如下^[14]:

$$\text{SAD} = \frac{1}{b} \sum_{y=1}^b |P_{y_i} - P_{y_j}|$$

式中, P_{y_i} 为原始调查数据体长频率分布中第 y 个

体长组的频率, P_{y_j} 为模拟数据体长频率分布中第 y 个体长组的频率, b 为体长组个数。

应用相对估计误差 REE 评价平均体长估计值的精确度和准确度^[15],

$$\text{REE} = \frac{\sqrt{\sum_{i=1}^R (L_i^{\text{estimated}} - L^{\text{true}})^2 / R}}{L^{\text{true}}} \times 100\%$$

式中, L^{true} 是根据原始调查数据计算的该海域方氏云鲷平均体长“真值”, $L_i^{\text{estimated}}$ 是根据第 i 次再抽样模拟数据计算的平均体长“模拟值”, R 是模拟次数(本研究为 1000 次)。

2 结果与分析

2.1 体长分布离散指数

在各个季节中, 方氏云鲷体长分布离散指数分布范围为 0~0.91, 其变化范围在不同季节有所不同, 除 DI 值为 0 的站位外, 春季方氏云鲷体长分布离散指数较高且集中在 0.90 附近, 即站位内方氏云鲷个体体长分布较均匀, 该站位内体长频率分布有较大的变异, 个体不存在集中分布于个别体长组的趋势。秋季、冬季和夏季方氏云鲷体长分布离散指数值分布在 0.5~0.91 之间, 部分站位内方氏云鲷个体体长分布较集中(图 2)。

2.2 站位数与体长频率分布估计精度的关系

各季节方氏云鲷体长频率分布估计的绝对值和 SAD 值的变化均较大。随着站位数的增加, 各季节的 SAD 值均呈下降趋势且趋于稳定(图 3)。在各个季节, 调查站位数增加到 20 后, 随着调查站位数的继续增加, SAD 值下降较小, 体长频率分布估计的精度提高不明显。

方氏云鲷体长频率分布估计 SAD 值随站位数的变化存在季节差异。在站位数较少时, 方氏云鲷体长频率分布估计 SAD 值秋、夏季变化范围较大, 冬季较小, 春季最小。在各个季节, 方氏云鲷体长频率分布 SAD 值随站位数的变化其变化范围不同, 例如在春季调查站位数从 1 增加到 5, SAD 平均值从 8.1 下降为 4.3, 降幅为 47.14%; 在秋季, 调查站位数从 1 增加到 5, SAD 平均值从 11.2 降低到 6.5, 降幅为 41.50%。

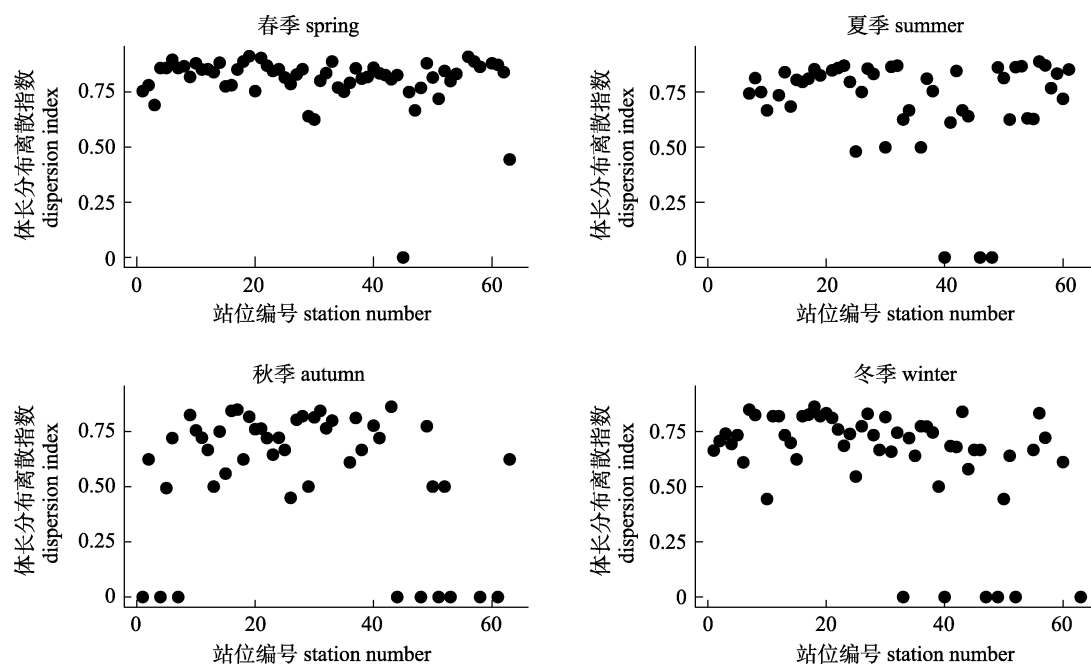


图 2 方氏云鲷体长分布离散指数

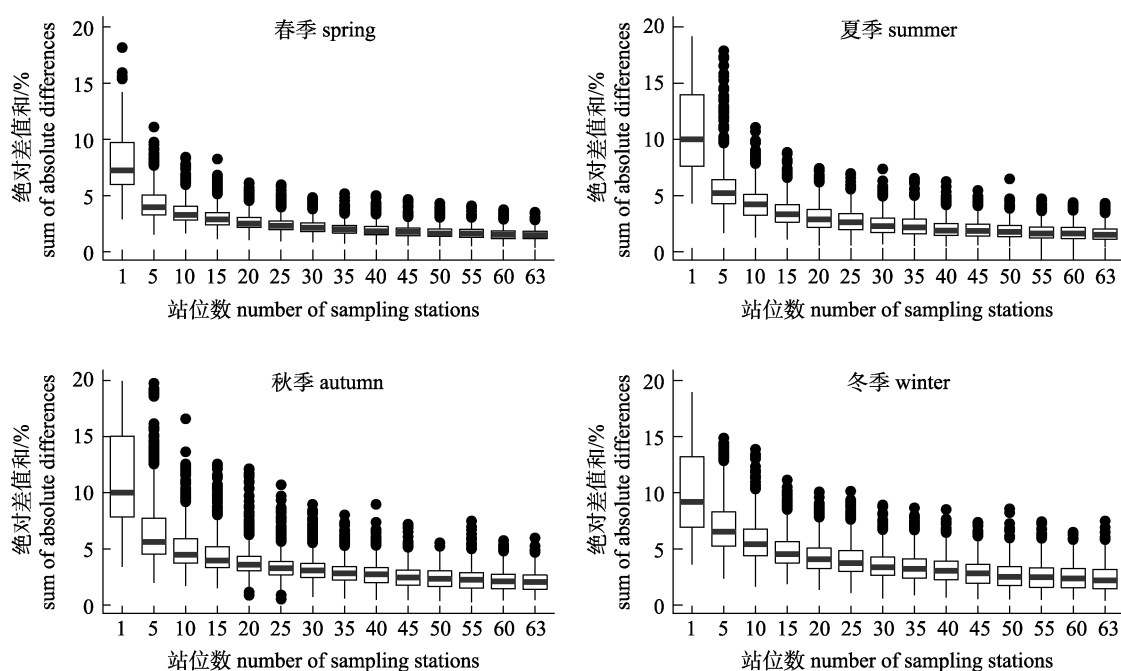
Fig. 2 The dispersion index of the length frequency distribution of *Enedrias fangi*

图 3 方氏云鲷体长频率分布估计 SAD 值与站位数的关系

Fig. 3 Relationship between the sum of absolute difference of estimation of length frequency distribution and the sample size for *Enedrias fangi*

2.3 站位数与平均体长估计精度的关系

平均体长估计值的 REE 值范围较大, 但总体变化趋势一致, 随着站位数的增加 REE 值逐渐减小; 在各个季节, 调查站位数增加到 30 时, 随着

调查站位数的继续增加, 平均体长估计的精度提高不明显。方氏云鲷平均体长估计值的 REE 值随站位数的变化存在季节差异, 夏、秋季变化范围较大, 冬季次之, 春季最小。在各个季节, 方氏云

鲮平均体长估计值的 REE 值随站位数的变化其变化范围也不同。例如在春季调查站位数从 1 增加到 5, REE 值平均值从 20.66% 降低到 9.26%, 降幅

为 55.15%; 在秋季调查站位数从 1 增加到 5, REE 值平均值从 59.51% 降低到 26.71%, 降幅为 55.11% (图 4)。

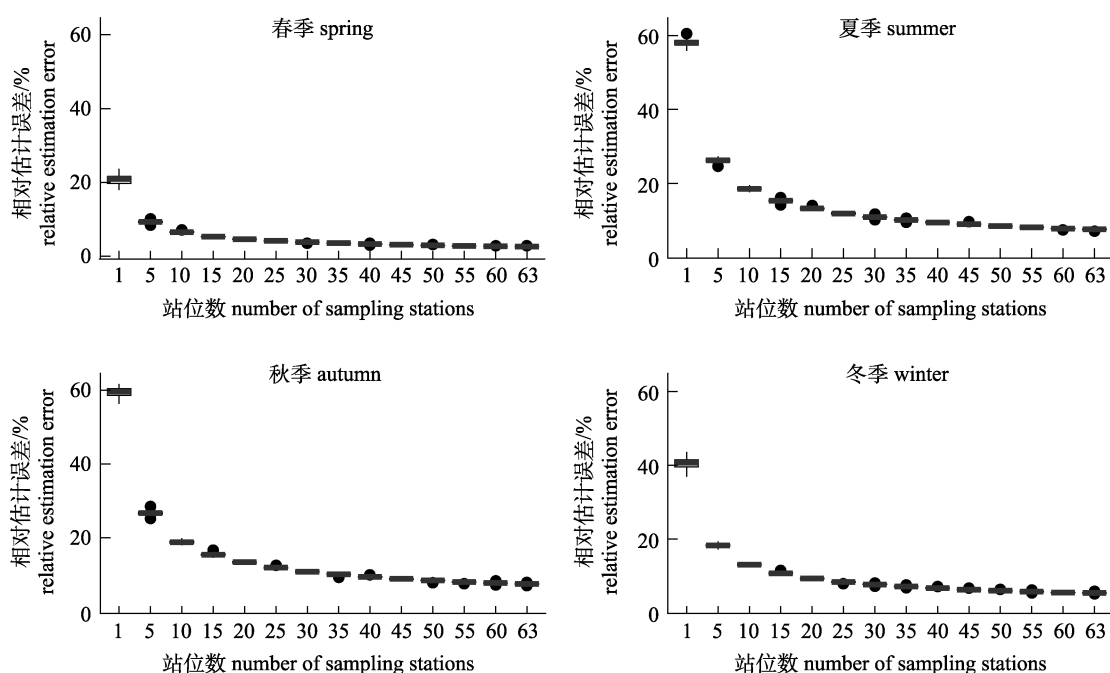


图 4 方氏云鲮平均体长估计 REE 值与站位数的关系

Fig. 4 The relationship between the relative estimation error of estimation of mean length and the sample size for *Enedrias fangi*

3 讨论

本文研究了采样站位对方氏云鲮不同体长指标(体长频率分布、平均体长)估计的影响,并探究了站位内方氏云鲮体长频率分布变异大小对体长频率分布估计所需样本量的影响。在各个季节,方氏云鲮体长分布离散指数 DI 值为 0~0.91,大部分站位 DI 值为 0.75~0.90,表明调查站位内体长频率分布有较大的变异,调查站位间体长频率分布较为相似,可以对调查站位数进行优化,提高采样效率。调查站位数与平均体长和体长频率分布估计的关系表明,调查站位数越多,其估计精度就越高,用以估计体长频率分布的有效站位数和估计平均体长的有效站位数不同。

评价指标 SAD 值和 REE 值体现了根据重采样模拟数据与原始调查数据估计目标鱼种体长频率分布/平均体长间的偏差程度。其值与站位数的关系可用于评估给定站位数的相对准确度,并可以为资源监测计划设计提供参考^[14,16]。SAD 值和

REE 值越小,表明方氏云鲮模拟数据和调查数据越相似,模拟效果越好。当站位数较小时,随着站位数的增加, SAD 值和 REE 值下降较快并逐渐趋于平缓(图 3 和图 4);这是由于在站位数较少时,随机取样采集到 DI 值较低(站位内体长样本集中分布)站位的概率较大,容易产生较高的估计误差。例如,在秋季,大部分调查站位的方氏云鲮体长分布离散指数 DI 值在 0.5~0.75 之间,总体来说,站位内方氏云鲮个体较为集中地分布在某些体长组,站位间方氏云鲮体长频率分布存在较大差异,当调查站位数较少时,模拟研究表明易产生较大的 SAD 值(图 3),需要增加采样站位数量以减少物种体长大小分布差异对采样精度的影响。

随着站位数的继续增加,评价指标 SAD 值和 REE 值逐渐趋于平缓,表明增加样本量对体长数据指标估计的精度影响不大,例如,在春季方氏云鲮体长分布离散指数 DI 值分布在 0.75~0.90 之间,各站位内方氏云鲮体长各体长组分布均匀,

站位间方氏云鲷体长频率分布较相似,增加调查采样站位数对采样精度的影响较小,较少的站位数仍可获得相对较高的体长频率分布估计准确度和精确度。

假设当前的调查数据能够真实反映该海域目标鱼类种群体长结构的前提下,在各个季节,用以估计体长频率分布的样本量和估计平均体长的样本量存在差异,增加相同的采样站位数,平均体长估计精度提高得更高。例如,在春季站位数从 1 增加到 5 个, SAD 降幅为 47.14%, REE 降幅为 55.15%, 提高相同的精度,体长频率分布估计需要更多的站位数,平均体长估计则受样本量的影响较小。在调查站位数分别增加到 20 时和 30 时,随着调查站位数的继续增加,体长频率分布和平均体长估计的精度提高不明显,故 20 个调查站位为估计方氏云鲷体长频率分布的可接受样本量,30 个调查站位可以视为估计方氏云鲷平均体长的样本量。Schultz 等^[17]对小型鱼类(最大长度小于 200 mm)体长频率分布有效样本量大小研究也发现类似规律,研究表明,当样本量小于 50 尾时,随样本量增加准确度提高最大,体长频率分布估计的有效样本量为 40~130 尾。

底拖网调查可认为是二阶段抽样,在第一阶段,从调查海域中随机选取采样站位;在第二阶段,从每个调查站位随机选取鱼类个体^[8,18]。由于生物栖息地偏好,大多数鱼类具有集群性^[19],底拖网调查采集的样本数据通常是从“集群”中收集的,相邻区域站位内方氏云鲷体长组成相似,不同区域站位间体长大小组成明显不同。虽然平均体长指标受调查站位数的影响较小,但是平均体长指标不能提供全面的体长信息,必须考虑整体体长分布,体长频率分布直方图可以直观、详细地展示目标种类体长分布信息,以准确地识别目标物种体长频率分布结构^[20],但体长频率分布直方图容易受调查数据规模的影响,在实际调查中需要选择合理的调查设计方案。用以估计体长频率分布的样本量和估计平均体长的样本量存在差异,可能由于计算机模拟使用的“真值”(体长频率分布和平均体长)的变异情况引起,其关系有待进

一步的研究。同时在以后的研究中应该考虑系统性采样误差和不同采样设计等对体长频率分布估计的影响,如 Breton 等^[21]对电网采样的鱼类体长频率数据进行了校正。本文的研究结果可为基于不同体长指标的渔业资源调查设计提供一定的参考,有助于对基于体长指标有效站位数的初步了解,为山东半岛南部海域的鱼类资源科学调查和科学管理提供参考。

参考文献:

- [1] Anderson R O, Neumann R M, Murphy B R, et al. Length, weight, and associated structural indices[M]// Fisheries Techniques. 2nd Edition. Bethesda: American Fisheries Society, 1996: 447-482.
- [2] Anderson R O. Management of small warm water impoundments[J]. Fisheries, 1976, 1(6): 5-7, 26-28.
- [3] Gabelhouse D W Jr. A length-categorization system to assess fish stocks[J]. North American Journal of Fisheries Management, 1984, 4(3): 273-285.
- [4] Froese R, Winker H, Coro G, et al. A new approach for estimating stock status from length frequency data[J]. ICES Journal of Marine Science, 2018, 75(6): 2004-2015.
- [5] Dzul M C, Dinsmore S J, Quist M C, et al. A simulation model of the Devils Hole pupfish population using monthly length-frequency distributions[J]. Population Ecology, 2013, 55(2): 325-341.
- [6] Miranda L E. Approximate sample sizes required to estimate length distributions[J]. Transactions of the American Fisheries Society, 2007, 136(2): 409-415.
- [7] Wang J Q, Xu L L, Li B, et al. An evaluation of the effects of sample size on estimating length composition of catches from tuna longline fisheries using computer simulations[J]. Aquaculture and Fisheries, 2020, 5(3): 122-130.
- [8] Pennington M, Burmeister L M, Hjellvik V. Assessing the precision of frequency distributions estimated from trawl-survey samples[J]. Fishery Bulletin, 2002, 100(1): 74-80.
- [9] Zhang Y Y, Cadrin S X. Estimating effective sample size for monitoring length distributions: A comparative study of Georges Bank groundfish[J]. Transactions of the American Fisheries Society, 2013, 142(1): 59-67.
- [10] Pennington M, Helle K. Evaluation of the design and efficiency of the Norwegian self-sampling purse-seine reference fleet[J]. ICES Journal of Marine Science, 2011, 68(8): 1764-

- 1768.
- [11] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Specifications for Oceanographic Survey-Part 6: Marine Biological Survey, GB/T 12763.6-2007[S]. Beijing: Standards Press of China, 2007. [中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 海洋调查规范第 6 部分: 海洋生物调查 GB/T 12763.6-2007[S]. 北京: 中国标准出版社, 2007.]
- [12] Chih C P. Incorporating effective sample sizes into sampling designs for reef fish[J]. Fisheries Research, 2010, 105(2): 102-110.
- [13] Jin Y J, Du Z F, Jiang Y. Sampling Technique[M]. 4th Edition. Beijing: China Renmin University Press, 2015: 107-132. [金勇进, 杜子芳, 蒋妍. 抽样技术[M]. 第 4 版. 北京: 中国人民大学出版社, 2015: 107-132.]
- [14] Li B. Modeling and simulation of American lobster in the Gulf of Maine: Spatial distribution, monitoring, and conservation[D]. Maine: The University of Maine, 2018.
- [15] Wang J, Xu B D, Zhang C L, et al. Sample size optimization for cluster design of bottom trawl fish surveys in the Yellow River estuary[J]. Journal of Fishery Sciences of China, 2017, 24(5): 931-938. [王晶, 徐宾铎, 张崇良, 等. 黄河口鱼类底拖网调查采样断面数的优化[J]. 中国水产科学, 2017, 24(5): 931-938.]
- [16] Chen Y. A Monte Carlo study on impacts of the size of sub-sample catch on estimation of fish stock parameters[J]. Fisheries Research, 1996, 26(3-4): 207-223.
- [17] Schultz L D, Mayfield M P, Whitlock S L. Sample sizes needed to describe length-frequency of small-bodied fishes: An example using larval Pacific lamprey[J]. Journal of Fish and Wildlife Management, 2016, 7(2): 315-322.
- [18] Waters J R, Chester A J. Optimal allocation in multivariate, two-stage sampling designs[J]. The American Statistician, 1987, 41(1): 46-50.
- [19] Kingsley M C S, Carlsson D M, Kanneworff P, et al. Spatial structure of the resource of *Pandalus borealis* and some implications for trawl survey[J]. Fisheries Research, 2002, 58(2): 171-183.
- [20] Stefánsson G, de Croos M D S T. A case study of sampling strategies for estimating the length composition of commercial catches: The Sri Lankan shrimp trawl fishery[J]. Crustaceana, 2011, 84(12-13): 1581-1591.
- [21] Breton A R, Hawkins J A, Winkelman D L. Correcting length-frequency distributions for imperfect detection[J]. North American Journal of Fisheries Management, 2013, 33(6): 1156-1165.

Optimization of number of sampling stations in fishery-independent surveys for estimating fish length frequency distribution and mean length

MENG Xinxiang¹, XUE Ying^{1, 2, 3}, ZHANG Chongliang^{1, 2, 3}, JI Yupeng^{1, 3}, REN Yiping^{1, 2, 3}, XU Binduo^{1, 2, 3}

1. College of Fisheries, Ocean University of China, Qingdao 266003, China;

2. Laboratory for Marine Fishery Science and Food Production Processes; Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China;

3. Field Observation and Research Station of Haizhou Bay Fishery Ecosystem, Ministry of Education, Qingdao 266003, China

Abstract: Length frequency distribution and mean length data are often used to estimate parameters of growth and mortality of fish stocks, as they are relatively easy to obtain from fishery-independent surveys for most fish species and can provide insights into the dynamics of fish populations, especially for data-poor or -limited fisheries. A lack of smaller size classes of fish populations may represent recruitment deficiencies, whereas a lack of larger size classes might indicate high mortality of adult individuals of fish stock. Precise and accurate estimates of population parameters are critical to ensure rational fish stock assessment and management. The performance of a survey sampling design might vary with different size structure indices, including length frequency distribution and mean length. To provide high quality length data for supporting fishery resource assessment and scientific management, the performances of the number of survey stations for estimating *Enedrias fangi* length frequency distribution and mean length were evaluated using computer simulations based on the original data collected from bottom trawl surveys in October 2016, and January, May, and August 2017 in the southern waters off Shandong Peninsula. The dispersion index (DI) was calculated to reflect *E. fangi* size composition variability within survey stations to identify the factors that influence sampling efficiency. Large DI values implied that length samples were evenly distributed among all size bins, whereas a DI close to zero indicated all observations were skewed into relatively few size bins. The sum of absolute difference (SAD) and relative estimation error (REE) were used to measure the performances (accuracy and precision) of different numbers of stations in estimating the target size structure indices. The results showed that the *E. fangi* length frequency distribution DI ranged from 0 to 0.91 in four seasons, whereas the DI of most stations was between 0.75 to 0.91. The number of sampling stations could be reduced so that the sampling efficiency could be improved. In general, SAD and REE showed similar trends, decreasing with sample sizes initially and then becoming stable after certain sample sizes during all four seasons, which implied that the length data estimation precision improved with an increased number of sampling stations. The optimal number of sampling stations varied with target indices. The optimal sample size was 20 stations for estimating length frequency distribution and 30 stations for estimating mean length. This study identified the accepted optimal sample sizes to estimate length frequency distribution and mean length and could provide a technical reference for fishery-independent surveys to estimate the size structure index.

Key words: length frequency distribution; relative estimation error; sample size optimization; *Enedrias fangi*; Shandong Peninsula

Corresponding author: XU Binduo. E-mail: bdxu@ouc.edu.cn