

文章编号: 1674-5566(2010)02-0240-06

不同名义 CPUE 算法对 CPUE 标准化的影响

田思泉^{1,2,3}, 陈新军^{1,2,3}

(1 上海海洋大学海洋科学学院, 上海 201306

2 上海海洋大学大洋生物资源开发和利用上海市高校重点实验室, 上海 201306

3 上海海洋大学大洋渔业资源可持续开发省部共建教育部重点实验室, 上海 201306)

摘要: CPUE 标准化是渔业资源评估的重要研究内容。不同名义 CPUE 计算方法通常会影响到 CPUE 的标准化, 进而影响到对资源丰度的评价。在 CPUE 标准化过程中, 首先对名义 CPUE 进行计算, 通常有以下 3 种方法: (1) 直接将每个记录的产量除以其对应的捕捞努力量计算出 CPUE 值, 然后输入模型; (2) 先对所有的记录计算出 CPUE, 然后以一个空间尺度 (小渔区, 如 $0.5^\circ \times 0.5^\circ$) 统计数据, 对每个小渔区内所有的 CPUE 直接求平均值 (所有 CPUE 求和再除以记录的数量) 输入模型; (3) 以一个空间尺度统计数据, 用小渔区内的所有记录的总产量除以总捕捞努力量计算的 CPUE 输入模型。本文以我国在西南大西洋的阿根廷滑柔鱼鱿钓渔业为例, 分别使用上述 3 种 CPUE 计算方法, 运用广义加性模型 (Generalized Additive Model, GAM) 模型进行 CPUE 标准化 (对应的模型分别为 GAM_a , GAM_b 和 GAM_c), 比较不同模型的结果。研究发现, 3 种模型得出的年标准化 CPUE 和月标准化 CPUE 存在差别, GAM_a 与 GAM_b , GAM_c 的结果相差较大, GAM_b 和 GAM_c 所得的结果相差较小, 而 GAM_a 与 GAM_b , GAM_c 的标准化 CPUE 值、变异系数以及模型的各因子方差贡献率等相差较大。3 种模型之间的差别主要由于样本数、捕捞努力量的假设、数据记录的时空尺度和模型中因子的选择等因素影响。因此在使用商业性渔业数据分析渔业资源状况时, 需要考虑由于 CPUE 计算方法的不同带来的不确定性。

关键词: CPUE 标准化; 广义加性模型; 西南大西洋; 阿根廷滑柔鱼

中图分类号: S932.8 **文献标识码:** A

Impacts of different calculating methods for nominal CPUE on CPUE standardization

TIAN Si-quan^{1,2,3}, CHEN Xin-jun^{1,2,3}

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China;

2. Key Laboratory of Shanghai Education Commission for Oceanic Fisheries Resources Exploitation, Shanghai Ocean University, Shanghai 201306, China;

3. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China)

Abstract: CPUE standardization is a role in fisheries stock assessment. Different calculating methods for nominal CPUE can affect standardized CPUE which usually is used to index fisheries abundance. In CPUE

收稿日期: 2009-12-08

基金项目: 国家科技“八六三”计划 (2007AA092202); 国家科技“八六三”计划 (2007AA092201); 国家科技支撑计划 (2006BAD9A05); 上海市捕捞学重点学科 (S0702)

作者简介: 田思泉 (1978-), 男, 博士, 主要从事渔业资源与渔场学方面的研究。E-mail: sqtian@shou.edu.cn

通讯作者: 陈新军, Tel: 021-61900340, E-mail: xjchen@shou.edu.cn

standardization modeling, nominal CPUE is as a response variable which usually can be calculated in three methods: (1) for every record, CPUE is calculated by catch dividing the corresponding effort, and the all CPUEs were input the model; (2) all fisheries data records are first grouped by one spatial scale (fishing grid, such as $0.5^{\circ} \times 0.5^{\circ}$), then for every grid, average CPUE is calculated by all CPUE dividing the number of records, all average CPUEs of fishing grids are as the model input; (3) the fisheries data grouping and the model input are same as (2), but the average CPUE for every grid is calculated by total catch dividing total fishing effort. For evaluating impacts of different nominal CPUE inputs on CPUE standardization modeling, Chinese *Illex argentinus* fishery in the South Atlantic Ocean was as a study case, the three above-mentioned nominal CPUEs calculated from this fisheries data were input generalized additive models which were used to standardize these CPUEs. The corresponding models were GAM_a , GAM_b and GAM_c , respectively. The results derived from the GAMs were compared. From these analysis, there were differences among these standardized CPUEs derived from the three GAMs. GAM_b and GAM_c showed similar trends in standardized CPUEs. But there were significant variances between GAM_a and the latter two GAMs in the values of standardized CPUE, coefficient of variance and the variance contribution rate of each explanatory variable in the models. The differences among these GAMs were mainly owing to sample numbers, the assumption for calculating fishing effort, the spatio-temporal scale for grouping fisheries data and model selection, etc. On the summary, the certainties due to different calculating methods for CPUE should be considered when commercial fisheries data were used to analyze the status of fisheries stock.

Key words: CPUE standardization, generalized additive model, Southwest Atlantic Ocean, *Illex Argentinus* fishery

渔业资源评估中,单位捕捞努力量渔获量(Catch Per unit effort, CPUE)通常被认为与渔业资源丰度成比例,被作为资源相对丰度指数来反映渔业资源丰度的大小^[1]。渔业数据按来源的独立性可以分为两类:一种是独立调查数据(Independent survey data),主要来自于科学调查的数据;另一种是非独立数据(Dependent survey data),主要来自渔业观察员、港口调查数据或渔民汇报的数据(即商业性渔业数据)^[2-4]。在许多渔业(尤其是大洋性渔业)数据收集集中^[5-8],由于独立调查数据获取的高费用问题和难以操作性,商业性渔业数据常被用来分析渔业资源状况,商业性渔业数据由于常受到各种因素(如时间、空间、渔船参数、环境等)的影响,常采用统计模型对其CPUE进行标准化,使用标准化后的CPUE来反应渔业资源丰度的情况^[1]。在使用统计模型对商业性渔业CPUE进行标准化中,CPUE作为模型中的应变变量,其输入方式通常有3种:(1)直接将每个记录的产量除以其对应的捕捞努力量计算出CPUE值,然后输入模型;(2)先对所有的记录计算出CPUE,然后以一个空间尺度(小渔区,如 $0.5^{\circ} \times 0.5^{\circ}$)统计数据,对每个小渔区内

所有的CPUE直接求平均值(所有CPUE求和再除以记录的数量)输入模型;(3)以一个空间尺度统计数据,用小渔区内的所有记录的总产量除以总捕捞努力量计算的CPUE输入模型。CPUE输入方式的不同,可能会影响CPUE标准化的结果,进而影响人们对渔业资源形势的判断。本文以我国渔船在西南大西洋的阿根廷滑柔鱼渔业^[9]为例,分别上述3种方法计算CPUE,运用广义加性模型(Generalized Additive Model, GAM)方法^[10]对CPUE进行标准化。通过比较标准化后的CPUE从而评价不同CPUE输入方法及其对评价渔业资源的影响。

1 材料和方法

1.1 数据来源

1.1.1 商业性渔业数据

商业性生产数据来源于中国远洋渔业分会上海海洋大学鱿钓技术组建立的西南大西洋阿根廷滑柔鱼生产数据库,时间为2000—2007年。数据字段包括日期、经度、纬度、产量、作业次数。

1.1.2 海洋环境数据

海洋环境数据(包括温度、盐度和海平面高

度)来源于 NOAA (<http://oceanwatch.pifsc.noaa.gov/las/servelets/database/>), 时间为 2000—2007年, 环境因子分别为表温 (SST)、叶绿素 (chl_a)和海面高度 (SSH)。

1.2 数据分析方法

1.2.1 CPUE计算

根据前言中所提的 3 种方法分别计算 CPUE, 其计算公式分别为

$$CPUE_a = \frac{C}{E} \quad (1)$$

式中: C表示一艘渔船一天的产量; E表示其对用的作业次数。

$$CPUE_b = \frac{\sum CPUE_{a_i}}{n} \quad (2)$$

式中: CPUE_b表示平均 CPUE; CPUE_a表示一个月内、一个 0.5°×0.5°小渔区内每个记录的 CPUE, 计算方法同(1); n表示记录的个数。

$$CPUE_c = \frac{\sum C_i}{\sum E_i} \quad (3)$$

式中: CPUE_c表示平均 CPUE; $\sum C_i$ 表示一个月内、一个 0.5°×0.5°内总产量; $\sum E_i$ 表示对应的总作业次数。

将上述计算的 3 种 CPUE 分别根据经纬度和时间一致性与环境数据进行匹配。

1.2.2 使用 GAM 模型对 CPUE 进行标准化

GAM是一种非参数化的多元线性回归模型。GAM模型相对于传统的回归模型在分析资源丰度与环境的空间关系上提供了更多的信息, 可以更好地描述 CPUE与其它变量之间的非线性关系^[8], 近年来被越来越多的运用到 CPUE标准化研究中^[11-13]。本文所用的 GAM模型对 CPUE进行标准化, 其一般表达方式为

$$\ln(CPUE + c) = y + m + \varphi(\text{longitude}) + \varphi(\text{latitude}) + \varphi(\text{SST}) + \varphi(\text{SSH}) + \varphi(\text{Chl}) + \varepsilon \quad (4)$$

式中: c为常数项(取 1), y表示年, m表示月, longitude为经度, latitude为纬度, SST(sea surface temperature)为海表温度(°C), SSH(sea surface height)为海面高度(m), Chl(chlorophyll a)为叶绿素(mg/m³), ε为误差, 为样条平滑(spline smother)。

模型的因子选择和拟合度根据因子的显著

水平(P)和赤池信息准则(Akaike information criterion, AIC)^[14]的值来判断。分别用上述 3 种 CPUE数据输入 GAM模型进行 CPUE标准化, 对应的 GAM模型分别表示为 GAM_a、GAM_b和 GAM_c, 标准化的 CPUE分别表示为 S-CPUE_a、S-CPUE_b和 S-CPUE_c。模型的计算通过 S-Plus 8.0 软件来实现。

1.2.3 标准化 CPUE比较

对上述 3 个模型标准化后的结果, 分别计算年标准化 CPUE和月标准化 CPUE的值; 计算不同模型的年标准化 CPUE和月标准化 CPUE的均值和变异系数(coefficient of variance, V_c), 进行统计学比较。变异系数计算的方法为

$$V_c = \frac{D_s}{m} \quad (5)$$

式中: D_s为每个模型的所有年标准化 CPUE或月标准化 CPUE的标准差, m为每个模型所有年标准化 CPUE和月标准化 CPUE的均值。

2 结果

2.1 CPUE标准化结果

本文在使用 GAM模型对不同 CPUE标准化建模中, 首先根据 P值, 将模型中不显著的因子 (P<0.05)排除, 在这 3 个模型中, 只有 GAM_a中叶绿素(chl)被排除出模型(见表 1), 然后根据模型的结果, 分别计算各年(2000—2007)的标准化 CPUE和各月(1—5)的标准化 CPUE。各模型对应的年标准化 CPUE和月标准化 CPUE分别见图 1和图 2。从图中可以看出, 使用 3 种模型得出的年标准化 CPUE和月标准化 CPUE均不一致。从年标准化 CPUE看, S-CPUE_a和 S-CPUE_b的最高值出现在 2006年, 而 S-CPUE_c的最高值出现在 2007年; S-CPUE_a最小值出现在 2001年, S-CPUE_b出现则 2000年, S-CPUE_c而则出现则 2004年; 从图中看, 三者的各年标准化 CPUE变动趋势存在着明显的不同。从月标准化 CPUE看, S-CPUE_b和 S-CPUE_c基本上呈相同的变化趋势, 从 1月到 5月标准化 CPUE值一直呈下降趋势。而 S-CPUE_a的值 1—3月一直呈上升趋势, 然后 3—5月再呈下降趋势。

表 1 GAM 模型中各因子的统计检验

Tab 1 Statistical test of each GAM for model selection

因子	GAM _a		GAM _b		GAM _c	
	P	方差贡献率 (%)	P	方差贡献率 (%)	P	方差贡献率 (%)
Year		1.032 9		0.967 0		0.634 1
month		0.448 3		2.687 5		2.198 1
ξ longitude	0.000 0	5.721 0	0.000 1	3.806 8	0.000 3	4.019 2
ξ latitude	0.000 0	1.036 9	0.000 0	4.045 5	0.000 0	4.315 0
ξ ssj	0.000 0	2.821 8	0.001 9	1.150 2	0.001 6	1.195 4
ξ sshj	0.000 1	0.356 5	0.018 0	1.127 5	0.026 8	1.014 2
ξ chj	0.000 0	1.123 4	0.041 6	0.911 2	0.057 7	

注: 在本文的 GAM 中, Year 和 month 作为“factor”, 始终存在于模型中。

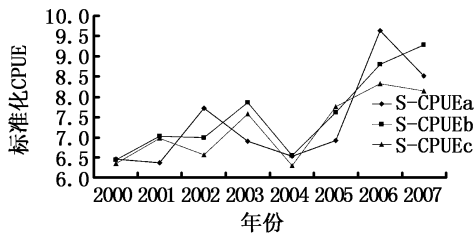


图 1 不同 GAM 模型的年标准化 CPUE
Fig 1 Annual standardized CPUEs derived from different GAMs

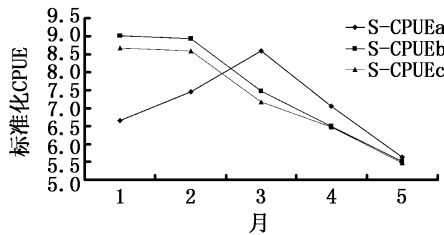


图 2 不同 GAM 模型的月标准化 CPUE
Fig 2 Monthly standardized CPUEs derived from different GAMs

2.2 标准化 CPUE 的比较

表 2 是分别对 3 个模型输出的年标准化 CPUE 和月标准化 CPUE 之间的统计学比较分析, 从各模型的年标准化 CPUE 和月标准化 CPUE 平均值看, 各模型的标准化 CPUE 平均值差别不大, 但 GAM_b 的平均值都是最大。变异系数可以反映样本值变化情况, 在年标准化 CPUE 中, GAM_a 的变异系数最大, 该模型的结果反映资源的年际变化波动最大, 而 GAM_c 的变异系数最小, 反映资源年际变化波动最小; 而从月标准化 CPUE 看, GAM_b 的变异系数最大, 该模型的结果反映月资源变化波动最大, GAM_c 的变异系数最小, 该模型的结果反映月资源变化波动最小。从模型的 AIC 值看, 由于 AIC 值得比较需要模型的输入样本数相同, 因此只能比较 GAM_b 和 GAM_c, GAM_b 的 AIC 值更小, 反映 GAM_b 模型的拟合度更好。

表 2 3 个 GAM 模型结果的统计学比较

Tab 2 Statistical comparing the standardized CPUEs derived from the GAMs

模型		GAM _a	GAM _b	GAM _c
年标准化 CPUE	平均值	7.384 8	7.575 3	7.251 1
	变异系数	0.1584	0.137 0	0.111 3
月标准化 CPUE	平均值	7.084 84	7.492 3	7.284 9
	变异系数	0.152 84	0.204 2	0.189 9
AIC		6.042 33	2.067. 03	2.082 17

3 讨论

对商业性生产数据进行 CPUE 标准化的研究最早是从对渔船努力量的标准化开始, 主要是通过渔船的捕捞能力与标准船的捕捞努力量的效率比而对生产数据进行标准化, 但随着渔业现代

化技术的发展, 各船之间的捕捞效率基本相差不大。阿根廷滑柔鱼属于短生命周期的中上层种类头足类, 其资源极易受到环境因子的影响^[9], 而且其生命周期中的资源的时空分布也是不同的。因此本文采用的 GAM 模型对西南大西洋阿根廷滑柔鱼渔业 CPUE 进行标准化时, 只考虑时

间、空间、海洋环境等因子对 CPUE的影响,而实际上影响 CPUE的因子有很多,但由于数据的可获取性,在 CPUE标准化模型中,忽略了其它因子的影响。

在本文的分析中, GAM_b 和 GAM_c 分析结果比较一致,而 GAM_a 与二者的区别较大。从模型中各因子的方差贡献率看(见表 1),在 GAM_a 模型各因子中,经度模型所解释的方差贡献率中为最高,达到 5.72%,随后为 SSI 和纬度,分别可解释 2.8% 和 1.04% 的方差贡献率; GAM_b 模型和 GAM_c 模型比较相似,纬度、经度和月的方差贡献率位都居前 3 位,总共可解释 10% 以上的模型方差贡献率。因此,从模型各因子所解释的方差贡献率看, GAM_b 和 GAM_c 比较相似,而 GAM_a 明显不同,各模型中因子的方差贡献率不同,对其分析的结果必然产生影响。此外, GAM_b 和 GAM_c 的输入的 CPUE 值是经过一个空间尺度 ($0.5^\circ \times 0.5^\circ$) 和时间尺度 (月) 进行统计求的平均值,每个时空尺度的都有一定数据样本数,经过求平均值后 CPUE 数据之间的极差明显变小,一些 CPUE 记录的异常值也被弱化,而且两个模型的解释变量的值都是一样的,两个模型数据输入的区别就是由于不同计算方法的 CPUE 输入值,在鱿钓渔业的研究^[7-8]中,由于数据获取难度的原因,一般假设每艘船的捕捞能力是等效的(而实际上是不等效的),CPUE 受捕捞努力量的影响被忽略,导致这两种方法计算的 CPUE 相差较小(差异的程度主要与每个时空尺度内记录的样本数相关),因此, GAM_b 和 GAM_c 标准化 CPUE 差异不大。而 GAM_a 的 CPUE 输入采用的是每艘船的原始数据记录,原始记录的 CPUE 值之间往往存在较大的极差,可能存在一些异常值,也就是说 CPUE 值的波动较大,其数据量远大于 GAM_b 和 GAM_c 的 CPUE 输入数据量,模型的解释变量的样本数也不同,可能导致 GAM_a 的结果与 GAM_b 和 GAM_c 差异较明。此外,渔业数据统计的时空尺度对 CPUE 标准化的结果产生影响^[15-16]。因此,总结这 3 个模型的 CPUE 标准化结果产生差异的原因,主要受样本数、捕捞努力量的假设、数据记录的时空尺度和模型中因子的选择等因素影响。

从统计学角度看, GAM_a 模型的 CPUE 数据来自于每个渔船的实际生产记录数据,因此更适

合用于分析渔业资源状况, GAM_b 的 CPUE 数据相比 GAM_c 的 CPUE 数据更符合统计的误差分布假设,但由于商业性 CPUE 受到很多因素的影响,因此很难评价哪个 CPUE 计算方法更适合,需要对科学调查数据来进行验证,以评价不同计算方法的合适度。在对商业性渔业数据进行 CPUE 标准化中,模型的不同会导致结果的不同^[11]。本文的分析计算结果表明,在 CPUE 标准化模型中,使用不同方法计算的 CPUE 作为模型输入,其标准化后 CPUE 在反映资源的丰度变化时也是不同的。如本文的 GAM_a 模型与 GAM_b 、 GAM_c 模型的年标准化 CPUE 和月标准化 CPUE 相差较大,采用不同的 CPUE 计算方法评估的资源状况明显不同,因此不同的 CPUE 输入所获得的结果可能影响到人们对资源状态的判断,从而影响人们对资源的开发和保护策略的制定。在对商业性渔业数据进行 CPUE 标准化时,需要考虑由于 CPUE 计算方法的不同,对资源的评估结果可能带来不确定性,一个不合适的模型输入,可能导致对资源形势的错误判断。因此在使用商业性渔业数据进行标准化时,不但需要考虑模型的不同对结果的影响,还要考虑相同模型不同的模型输入对结果的影响,需要进行类似本文的比较分析,从而选择最合适的模型和方法。

国内外很多对西南大西洋阿根廷滑柔鱼渔业的研究^[5, 17-18]所采用的数据都是截止到 2003 年,本文采用的数据为 2000—2007 年,因此在年际变化上很难进行比较。本文 GAM_b 和 GAM_c 模型计算的月标准化 CPUE 显示资源丰度在 1 月份最高,然后逐月开始下降,这与陈新军等^[18]研究的结果基本一致,Saca 等^[5]使用 GAM 模型分析在西南大西洋生产的西班牙渔船的 CPUE 与影响因子的关系显示,CPUE 主要受 SST 纬度和月份影响较大,这与本文 GAM_b 和 GAM_c 模型各因子方差贡献率的结果(表 1)基本相似。

本文使用 GAM 模型对 CPUE 进行标准化,GAM 能非常好的表达 CPUE 与其影响因子的非线性关系,而且对数据的误差分布要求不高,并具备分析时空数据的功能,因此非常合适被用来进行渔业生态数据的分析,近年来被越来越多运用到 CPUE 标准化研究上^[19]。

参考文献:

- [1] Maunder M N, Punt A E. Standardising catch and effort data: a review of recent approaches [J]. Fisheries Research, 2004, 70: 141—159.
- [2] Booth A J. Incorporating the spatial component of fisheries data into stock assessment models [J]. ICES Journal of Marine Science, 2004, 57: 858—866.
- [3] Scheiter K, Chen Y, Wilson C. Comparing two sampling programs for the Maine lobster fishery [J]. Fisheries Research, 2004, 68: 343—350.
- [4] Steizermüller V, Ehrlich S, Zauke G P. Impact of additional small scale survey data on the geostatistical analyses of demersal fish fish species in the North Sea [J]. Scientia Marina, 2005, 69: 587—602.
- [5] Sacau M, Pierce G J, Wang J et al. The spatio-temporal pattern of Argentine shortfin squid *Illex argentinus* abundance in the southwest Atlantic [J]. Aquatic Living Resources, 2005, 18: 361—372.
- [6] Tian S Q, Chen X J, Chen Y et al. Standardizing CPUE of *Onnastrephes banranii* for Chinese squid jiggling fishery in Northwest Pacific Ocean [J]. Chinese Journal of Oceanography and Limnology, 2009, 27(4): 729—739.
- [7] 陈新军, 赵小虎. 鲁外海茎柔鱼产量分布及其与表温关系的初步研究 [J]. 上海水产大学学报, 2006, 15(1): 65—70.
- [8] 田思泉, 陈新军, 冯波, 等. 西北太平洋柔鱼资源丰度与栖息环境的关系及其时空分布 [J]. 上海海洋大学学报, 2009, 18(5): 586—592.
- [9] 王尧耕, 陈新军. 世界大洋性经济柔鱼类资源及其渔业 [M]. 北京: 海洋出版社, 2005: 189—215.
- [10] Hastie T J, Tibshirani R J. Generalized Additive Models [M]. London: Chapman & Hall, 1990: 335.
- [11] Bigelow K A, Boggs C H, He X. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery [J]. Fisheries Oceanography, 1999, 8: 178—198.
- [12] Campbell R A. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models [J]. Fisheries Research, 2004, 70: 209—227.
- [13] Punt A E, Walker T J, Taylor B L, et al. Standardization of catch and effort data in a spatially-structured shark fishery [J]. Fisheries Research, 2000, 45: 129—145.
- [14] Akaike H. A new look at the statistical model identification [J]. IEEE Trans Autom Control, 1974, AC-19: 716—723.
- [15] Piet G J, Quirjns F J. The importance of scale for fishing impact estimations [J]. Canadian Journal of Fisheries and Aquatic Science, 2009, 66: 829—835.
- [16] Tian S Q, Chen Y, Chen X J et al. Impacts of spatial scales of fisheries and environmental data on catch per unit effort standardization [J]. Marine and Freshwater, 2009, 60: 1273—1284.
- [17] Chen C S, Chiu T S. Standardising the CPUE for the *Illex argentinus* fishery in the Southwest Atlantic [J]. Fisheries Science, 2009, 75: 265—272.
- [18] 陈新军, 刘金立. 巴塔哥尼亚大陆架海域阿根廷滑柔鱼渔场分布及与表温的关系分析 [J]. 海洋水产研究, 2004, 25(6): 19—24.
- [19] Denis V, Lejeune J, Robin J P. Spatio-temporal analysis of commercial trawler data using general additive models: patterns of Loliginid squid abundance in the north-east Atlantic [J]. ICES Journal of Marine Science, 2002, 59(3): 633—648.