

文章编号: 1674 - 5566(2013)05 - 0770 - 08

## 中大西洋延绳钓渔业大眼金枪鱼体长频率时空分布

汪文婷<sup>1</sup>, 田思泉<sup>1,2,3,4</sup>, 戴小杰<sup>1,2,3,4</sup>, 杨晓明<sup>1,2,3,4</sup>, 吴峰<sup>1</sup>

(1. 上海海洋大学 海洋科学学院, 上海 201306; 2. 上海海洋大学 大洋渔业资源可持续开发省部共建教育部重点实验室, 上海 201306; 3. 上海海洋大学 农业部大洋渔业资源环境科学观测实验站, 上海 201306; 4. 上海海洋大学 国家远洋渔业工程技术研究中心, 上海 201306)

**摘要:** 鱼类体长数据是渔业资源评估和管理的基础信息之一, 来自渔业非独立调查 (fishery dependent survey) 的体长频率数据可以用来反映目标鱼种亲体和补充量的分布情况, 并为基于体长结构的渔业资源评估模型重要输入信息。多元回归树是一种用来分析生物数据与环境特征数据间关系的挖掘技术。本文根据2007年12月-2009年12月中国远洋金枪鱼渔船渔捞日志所记录的大眼金枪鱼生物学数据, 利用多元回归树方法并结合地理信息系统分析了大西洋大眼金枪鱼的空间分布及季度变化情况。结果表明: 空间分布上大型大眼金枪鱼主要集中在 $7.5^{\circ}\text{N} \sim 15^{\circ}\text{N}$ ,  $17.5^{\circ}\text{W} \sim 45^{\circ}\text{W}$ ; 中型大眼金枪鱼主要集中在 $12.5^{\circ}\text{S} \sim 5^{\circ}\text{N}$ ,  $17.5^{\circ}\text{W} \sim 45^{\circ}\text{W}$ ; 小型个体主要分布在 $7.5^{\circ}\text{S} \sim 5^{\circ}\text{N}$ ,  $5^{\circ}\text{W} \sim 17.5^{\circ}\text{W}$ ; 经K-S检验, 各空间尺度内体长分布差异显著。2009年1、2季度体长分布相对同质; 3、4季度体长分布相对同质。

**研究亮点:** 利用多元回归树数据挖掘技术, 对中大西洋大眼金枪鱼体长频率时空分布进行分析。大西洋中部大眼金枪鱼体长分布空间上有显著差异, 且体长分布有显著的季节性差异。研究旨在为我国远洋金枪鱼渔业服务, 为合理评估和管理大西洋大眼金枪鱼资源提供科学参考。

**关键词:** 体长频率; 时空分布; 多元回归树; 延绳钓; 大西洋; 大眼金枪鱼;

**中图分类号:** S 932.4

**文献标志码:** A

鱼类的体长信息对渔业资源评估至关重要, 体长数据的优点在于容易获得, 且克服了大多数热带海域鱼类鳞片或耳石辨别年龄时存在误差的问题<sup>[1-2]</sup>。此外, 从鱼类的体长数据中可以获得很多信息, 如鱼类生长参数的估算、自然死亡率的估算、鱼类年龄结构组成和摄食等<sup>[3-5]</sup>。尽管来自渔业非独立调查数据 (fisheries dependent data, 通常指商业性渔业数据) 在资源评估时存在一定的偏差, 但体长频率数据可以用来反映目标鱼种亲体和补充量的分布情况, 并可以作为基于体长结构的渔业资源评估模型的重要输入信息<sup>[1]</sup>。大西洋大眼金枪鱼 (*Thunnus Obesus*) 广泛分布于 $50^{\circ}\text{N} - 45^{\circ}\text{S}$  间的海域, 是金枪鱼延绳钓渔业中最主要捕捞对象<sup>[5-6]</sup>。养护大西洋金枪鱼国际委员会 (The International Commission for the

Conservation of Atlantic Tunas, ICCAT) 每年的资源评估报告<sup>[4]</sup>中根据大眼金枪鱼体长频率划分大西洋捕捞方式。国内对大眼金枪鱼研究主要集中在生长与年龄<sup>[7]</sup>、渔场分布与海洋环境关系<sup>[8-10]</sup>等方面, 基于体长数据的相关时空分布尚无报道。国外学者对大西洋大眼金枪鱼体长空间分布表明<sup>[4]</sup>, 大西洋大眼金枪鱼为一个鱼类种群, 其小型幼鱼多与其他金枪鱼如黄鳍金枪鱼、鲣鱼等混合成群<sup>[4]</sup>, 几内亚湾是大西洋大眼金枪鱼主要的繁育场所; 随着大眼金枪鱼生长, 幼鱼有由热带海域向温带海域扩散的趋势。本文根据中国渔船的渔捞日志数据, 利用多元回归树 (multivariate regression tree, MRT) 结合地理信息系统方法对大西洋大眼金枪鱼体长频率空间分布及季度变化进行研究, 了解中大西洋不同体型

收稿日期: 2013-03-05 修回日期: 2013-04-11

基金项目: 农业部三大洋金枪鱼观察员项目(08-54); 上海市自然科学基金(11ZR1415500)

作者简介: 汪文婷 (1987—), 女, 硕士研究生, 研究方向为渔业资源评估和管理。E-mail: peggywangas@sina.com

通信作者: 田思泉, E-mail: sqtian@shou.edu.cn

大眼金枪鱼分布状况。研究旨在为合理管理与评估大眼金枪鱼资源提供参考。

## 1 材料与方法

### 1.1 数据来源

大西洋大眼金枪鱼体长数据均来自于中国远洋渔业协会金枪鱼技术组的中国远洋金枪鱼渔业渔捞日志数据库,本文分析的大眼金枪鱼体长(叉长)数据由渔捞日志中记录的加工重量乘 1.13 得到总重,再根据式 1<sup>[4]</sup> 换算得到。数据的时间跨度为 2007 年 12 月 - 2009 年 12 月,体长范围为 40 ~ 240 cm,年龄包括 1 龄幼鱼及 7 龄以上的成鱼样本。空间范围见图 1,共计 65 286 尾大西洋大眼金枪鱼体长数据,共分 29 个样本点 (5° × 5°)。

$$W_t = 2.396e^{-5} \times L_F^{2.9774} \quad (1)$$

式中:  $W_t$  表示体重;  $L_F$  表示体长。

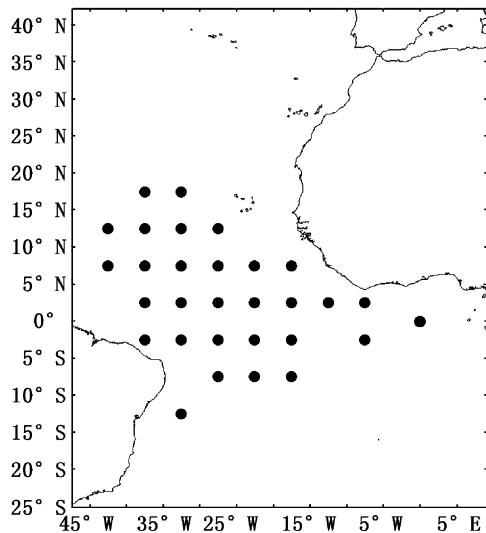


图 1 大眼金枪鱼样本分布水域

Fig. 1 Distribution of *Thunnus obesus* samples

### 1.2 分析方法

#### 1.2.1 数据的预处理

首先对体长数据进行预处理,按空间分辨率为 5° × 5° 和时间分辨率为季度的尺度对所有体长数据进行分类,根据公式 2<sup>[11]</sup> 确定每个时空尺度内体长组距为 11.8 m,为便于计算取体长组距 10 cm 进行分组,共计 19 组: ≤ 70 cm, 70 ~ 80 cm, ..., 210 ~ 220 cm, 220 ~ 230 cm, 230 ~ 240 cm, 240 ~ 250 cm。

$$S = R / (1 + 3.322 \times \lg N) \quad (2)$$

式中:  $S$  表示体长组距;  $R$  表示体长全距,即最大体长与最小体长的差值;  $N$  表示总尾数。

#### 1.2.2 数据分析方法

(1) 使用 ArcGIS 软件绘制 1 ~ 4 季度大眼金枪鱼不同体长组距尾数百分比空间分布图,分析其体长频率时空变化情况;并利用 K-S (Kolmogorov-Smirnov) 拟合优度检验每个 5° × 5° 空间尺度内大眼金枪鱼体长分布差异的显著性。

(2) 使用多元回归树对大眼金枪鱼的体长空间及季度分布进行分析。

(3) 大眼金枪鱼各季度体长比较的统计学分析:绘制各季度体长样本尾数百分比,得到各季度优势体长组,统计各季度平均体长频率的变异系数 (Coefficient of Variation, CV)、峰度 (Kurtosis) 及偏度系数 (Skewness) 数据分布情况。

由于本文的数据来源于商业性渔船,因此各季度渔获样本在空间上分布势必不均衡,为解决这样的问题,本文在研究中使用 Bootstrap 方法<sup>[15]</sup> 对每个空间尺度的样本体长数据进行再抽样,再抽样的前提是 Bootstrap 数据分布与原数据分布相一致,本文使用再抽样方法使得每个空间尺度内样本数达到 1 000 个,以达到数据分析要求。

本文分析中根据 ICCAT 对大西洋大眼金枪鱼的体型定义<sup>[4]</sup> 进行分类:体长小于 100 cm 为未成熟小型幼鱼体长小于 100 cm;体长范围在 100 ~ 140 cm 为中型大眼金枪鱼;体长大于 140 cm 为大型大眼金枪鱼;而体长大于 200 cm 的个体相对罕见。

多元回归树以相对熵 (Relative Entropy) 或称 K-L 散度 (Kullback-Leibler Divergence, KLD) 作为度量体长频率的不纯度 (Impurity) 准则<sup>[16]</sup>, 见式 2 中  $D(P|Q)$  表示 K-L 散度。其物理意义是:对于两个概率分布  $P$  和  $Q$  而言, KL 散度是用来度量使用基于  $Q$  的编码来编码来自  $P$  的样本平均所需要的额外的比特 (信息量单位 bit) 个数 (本文中可以理解为概率事件  $P(x)$  与  $Q(x)$  的相似程度,即聚类上两子集合的相似程度)。典型情况下,  $P$  表示数据的真实分布,  $Q$  表示数据的理论分布或模型分布。当两个概率事件完全相同时,即  $P(x) = Q(x)$ , 其相对熵为 0。

$$D(P|Q) = \sum_j p(j) \log \left( \frac{p(j)}{q(j)} \right) \quad (3)$$

对于一组数据  $P_1 \cdots P_n$ , 那么其算术平均数  $\bar{p}$  为:

$$\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i \quad (4)$$

根据公式 3 和公式 4, 本文中数据共分为 a 组 (a = 19、分组样方 b = 45), 体长频率样本的 K-L 散度  $I_{KLD}$  的计算公式:

$$I_{KLD} = \sum_{i=1}^a D(p_i | \bar{p}) = \sum_{i=1}^a \sum_{j=1}^b P_i(j) \log \left( \frac{P_i(j)}{\bar{P}(j)} \right) \quad (5)$$

式中:  $D(p_i | \bar{p})$  为  $P_i$  的 KL 散度;  $i$  为体长频率样本  $i = 1, \cdots, a$  (a = 45);  $j$  为分组情况  $j = 1, \cdots, b$  (b = 19, )。

依据 BREIMAN 等在分类回归树中<sup>[11]</sup> 将根节点数据 G 分割为左右子节点  $G_L$  和  $G_R$ , 那么  $I_{KLD}$  改为  $\text{Im}P(G_L, G_R)$ :

$$\begin{aligned} \text{Im}P(G_L, G_R) &= n_{all} H(\bar{p}_{all}) - n_L H(\bar{p}_L) - n_R H(\bar{p}_R) = \\ & n_L D(\bar{p}_L | \bar{p}_{all}) + n_R D(\bar{p}_R | \bar{p}_{all}) = n_L \sum_{j=1}^b \bar{P}_L(j) \log \\ & \left( \frac{\bar{P}_L(j)}{\bar{P}(j)} \right) + n_R \sum_{j=1}^b \bar{P}_R(j) \left( \frac{\bar{P}_R(j)}{\bar{P}(j)} \right) \end{aligned} \quad (6)$$

式中:  $n_{all}$  表示根节点样本数量;  $n_L$  表示左子节点样本数量;  $n_R$  表示右子节点样本数量;  $P_1 \cdots P_n$  的左右节点的熵值用  $H(p) = - \sum_{j=1}^b p(j) \log p(j)$  表示, 那么  $H(\bar{p}_{all})$  为根节点的平均熵值;  $H(\bar{p}_L)$  为左节点的平均熵值;  $H(\bar{p}_R)$  为右节点的平均熵值;  $D(\bar{p}_L | \bar{p}_{all})$  表示左节点熵的比重;  $D(\bar{p}_R | \bar{p}_{all})$  表示右节点熵的比重。

通过交替验证 (Cross Validation) 的结果根据 1 - 标准差 (1-se) 准则对多元回归树进行剪枝 (Prune), 同时兼顾树的规模和预测误差的大小, 1-se 准则为保证预测误差的值在最小的误差加减一个相应的标准差内, 选择此范围中最小值决定多元回归树分类数目, 是数据挖掘中常用的剪枝标准<sup>[12]</sup>。回归树根节点 (末端节点) 以体长频率平均数  $P_i(j)$  表示。

本文的数据 K-S 检验和分析采用了统计软件 SPSS 和 R 语言中的 mvpart 程序包, 绘图软件采用了地理信息系统软件 ArcGIS, 再抽样 Bootstrap 通过 Resample 软件实现。

## 2 结果

### 2.1 大眼金枪鱼体长时空分布状况

因为本文分析的体长数据来源于商业性渔

船的渔捞日志, 渔船作业位置在空间上肯定存在不均匀的现象, 渔船总是倾向于获得高产量的区域作业, 但渔船的空间变化情况通常也反映了目标鱼种的变动情况<sup>[16]</sup>, 因此本文使用的渔捞日志数据可以用来分析大西洋大眼金枪鱼体长时空变化情况。使用 ArcGIS 9.0 软件绘制大西洋大眼金枪鱼 1-4 季尾数百分比分布图 (图 2a-d), 图中柱形标识越高则表明该体长组距的尾数百分比值越大, 具体数值范围见图 2 备注。可见不同体型大眼金枪鱼 1 季度到 4 季度在空间分布存在差异。根据上述大眼金枪鱼不同体型体长分类, 1 季度中型大眼金枪鱼分布在  $0 \sim 15^\circ \text{N}$ ,  $35^\circ \text{W} \sim 45^\circ \text{W}$ , 大型大眼金枪鱼分布在  $0 \sim 15^\circ \text{N}$ ,  $25^\circ \text{W} \sim 35^\circ \text{W}$ ; 2 季度大型大眼金枪鱼主要分布在  $20^\circ \text{W} \sim 45^\circ \text{W}$ ,  $10^\circ \text{N} \sim 15^\circ \text{N}$ ; 3 季度分布规律不明显; 4 季度分布规律显著, 中型大眼金枪鱼主要分布于  $35^\circ \text{W} \sim 40^\circ \text{W}$ ,  $0 \sim 20^\circ \text{N}$ 。但仅凭目测难以得到亲体与补充量的准确分布情况, 利用多元回归树能得到更为精确的划分。

每个样本尺度内 Bootstrap 再抽样 1 000 尾, 统计得到 1~4 季度平均体长频率分布柱状图及统计参数 (表 1 和图 3), 表明 1 季度优势体长组为 130~150 cm 为双峰分布, 峰度值 (Kurtosis) 大于 0 呈扁平分布; 2 季度优势体长组为 120~150 cm, 呈双峰扁平分布; 3 季度优势体长组为 140~170 cm, 由峰度值小于 0 呈尖峰分布; 4 季度优势体长组为 120~140 cm 为单峰尖峰分布。由 1~4 季度平均体长频率变异系数 (Coefficient of Variation, CV) 得 1、2、4 季度的离散程度较为接近, 数据较为离散, 3 季度数据最为集中; 偏度系数 (Skewness) 均大于 0 呈右偏分布, 1 季度的尾数分布偏斜程度最大。

### 2.2 K-S 检验

上述分析结果中得到体长频率在空间分布上存在显著季度性及空间分布规律。为了验证大西洋大眼金枪鱼体长频率在空间分布差异是否显著, 使用 K-S 统计检验方法对其进行检验, 经检验, 各个空间尺度的体长频率空间分布存在显著性差异 (表 2), 在显著水平  $\alpha = 0.10$  的水平下, 统计量 D 均大于  $P(\alpha/2)$ , 说明大眼金枪鱼在每个空间尺度 ( $5^\circ \times 5^\circ$ ) 的体长频率分布差异显著。

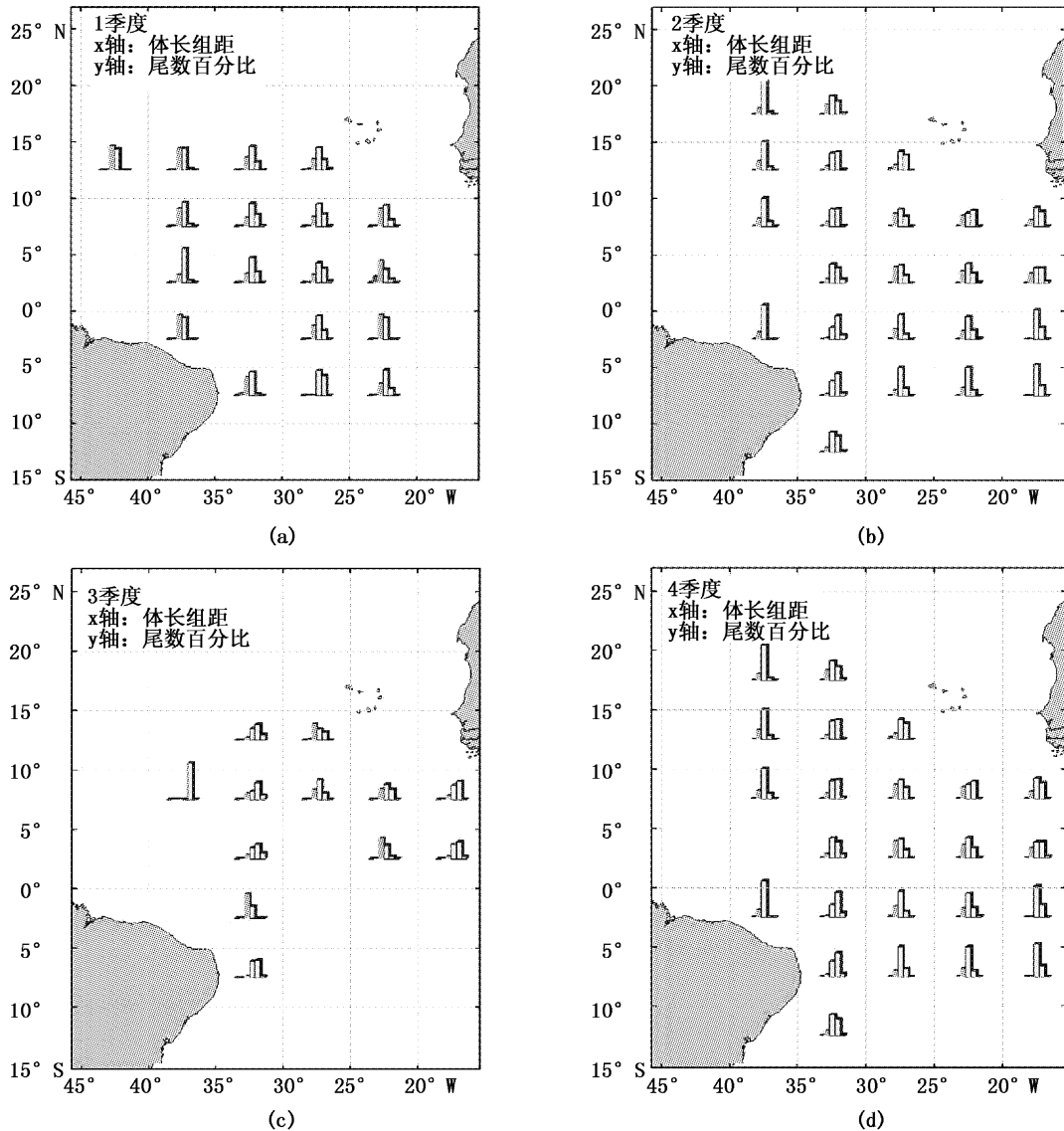


图 2 1-4 季度大西洋大眼金枪鱼尾数百分比分布图

Fig. 2 Seasonal size distribution of *Thunnus obesus* in the Atlantic Ocean

a - d 柱状图中:横坐标均为体长组距,从左到右依次表示  $\leq 70$  cm,  $70 \sim 100$  cm,  $100 \sim 130$  cm,  $130 \sim 160$  cm,  $160 \sim 200$  cm,  $> 200$  cm; 纵坐标为尾数百分比值,范围为  $0 \sim 100\%$ 。

表 1 季节体长频率直方图的统计参数

Tab. 1 Statistics parameters of seasonal length-frequency

季节	Bootstrap 样本量	平均体长/cm	离散系数	峰度	偏度
季度 1	20 000	144.26	1.24%	0.06	1.12
季度 2	18 000	143.58	1.26%	0.06	1.10
季度 3	11 000	163.41	0.91%	-0.95	0.44
季度 4	20 000	143.14	1.21%	-0.57	0.93

### 2.3 多元回归树分析结果

由于大西洋大眼金枪鱼体长频率的时空分布呈显著差异,利用多元回归树方法来拟合体长频率数据以进一步分析。在多元回归树模型中,将体长频率数据作为解释变量,以 1~4 季度、每

个空间尺度( $5^\circ \times 5^\circ$ )经纬度作为响应变量生成的多元回归树(根据交叉验证结果进行剪枝)结果见图 4(a),根据拟合的多元回归树在空间上进行的划分结果见图 4(b),拟合结果见表 3。

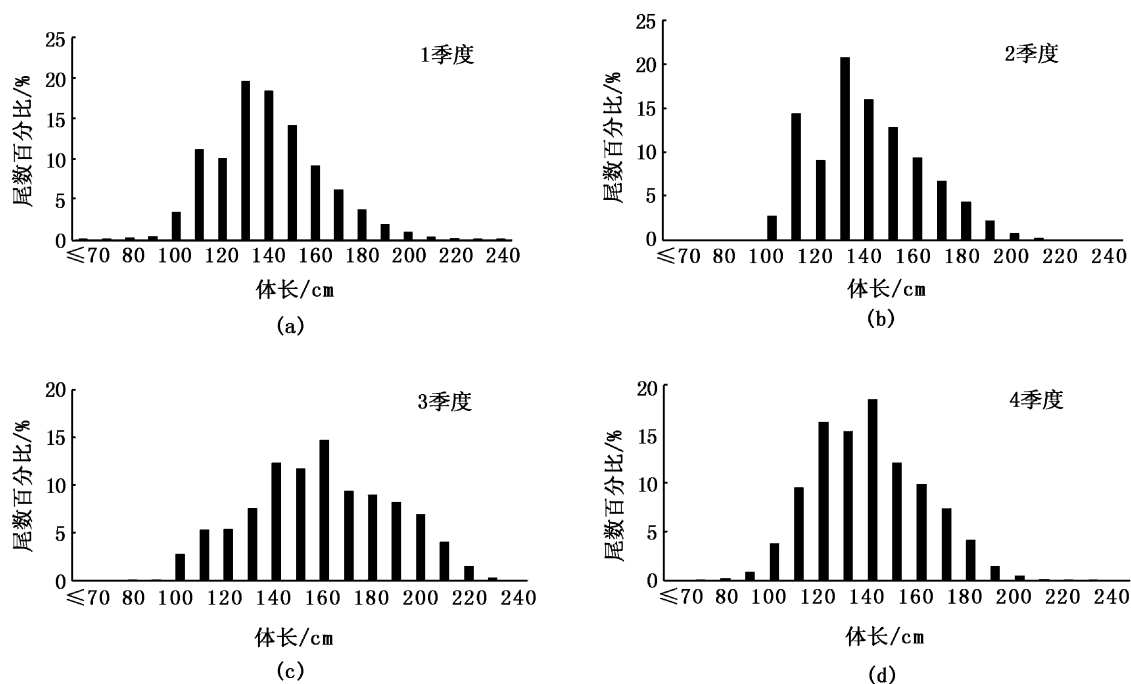


图3 1~4季度平均体长频率柱状图

Fig. 3 Histogram of seasonal average length frequency of *Thunnus obesus*表2 各渔区 K-S 检验统计量 D、P( $\alpha/2$ )

Tab. 2 Statistics parameters of K-S test

样本点	1	2	3	4	5	6	7	8
统计量 D	1	0.972	1.298	0.973	0.811	0.794	0.794	1.136
P( $\alpha/2$ )	0.324	0.487	0.069	0.3	0.526	0.649	0.649	0.152
样本点	9	10	11	12	13	14	15	16
统计量 D	0.811	1.136	1.784	1.460	0.794	1.136	1.298	1.298
P( $\alpha/2$ )	0.526	0.152	0.003	0.028	0.649	0.152	0.069	0.069
样本点	17	18	19	20	21	22	23	24
统计量 D	1.622	1.136	0.794	0.973	0.973	0.793	0.794	0.794
P( $\alpha/2$ )	0.01	0.152	0.649	0.3	0.3	0.649	0.649	0.649
样本点	25	26	27	28	29			
统计量 D	0.811	0.794	0.973	0.973	0.811			
P( $\alpha/2$ )	0.526	0.649	0.3	0.3	0.526			

表3 体长数据的多元回归树拟合结果

Tab. 3 Result of multivariate regression tree

节点	分裂	样方数	偏差值	拟合体长频率值
1	root	45	23 536.33	30.381
2	lat < 5	26	7 070.092	17.539
4	lon > = -20	6	157.094	2.665 *
5	lon < -20	20	5 187.74	22 *
3	lat > = 5	19	12 984.36	47.954 *

注: \* 表示回归树根节点。

由图 4a 和表 3 多元回归树拟合结果可见 2007 年 12 月 ~ 2009 年 12 月 45 个体长频率样方

(标准误差  $se = 0.49$ ) 分为 3 个类别。根节点平均体长频率值(图 4a 及表 3)、多元回归树的拟合结果得到体长空间划分图见图 4b 及多元回归树体长频率拟合结果表 3 可见大西洋大眼金枪鱼渔获体长频率分布特征,大型大眼金枪鱼主要分布在海域 1(7.5°N ~ 15°N, 17.5°W ~ 45°W); 中型大眼金枪鱼主要集中在海域 2(12.5°S ~ 5°N, 17.5°W ~ 45°W); 小型大眼金枪鱼主要集中在海域 3(7.5°S ~ 5°N, 5°W ~ 17.5°W)。多元回归树第一次划分以纬度 5°N 为分割点,左节点 5°N 以南 26 个样方(图 4b 中海域 2 和海域 3,右节点 5°

N 以北共 19 个样方(图 4b 中海域 1);第二次划分以经度 20°W 为分割点,右节点为 20°W 以西

海域(图 4 中海域 2),共 6 个同质样方,左节点为 20°W 以东(图 4 中海域 3),共 20 个同质样方。

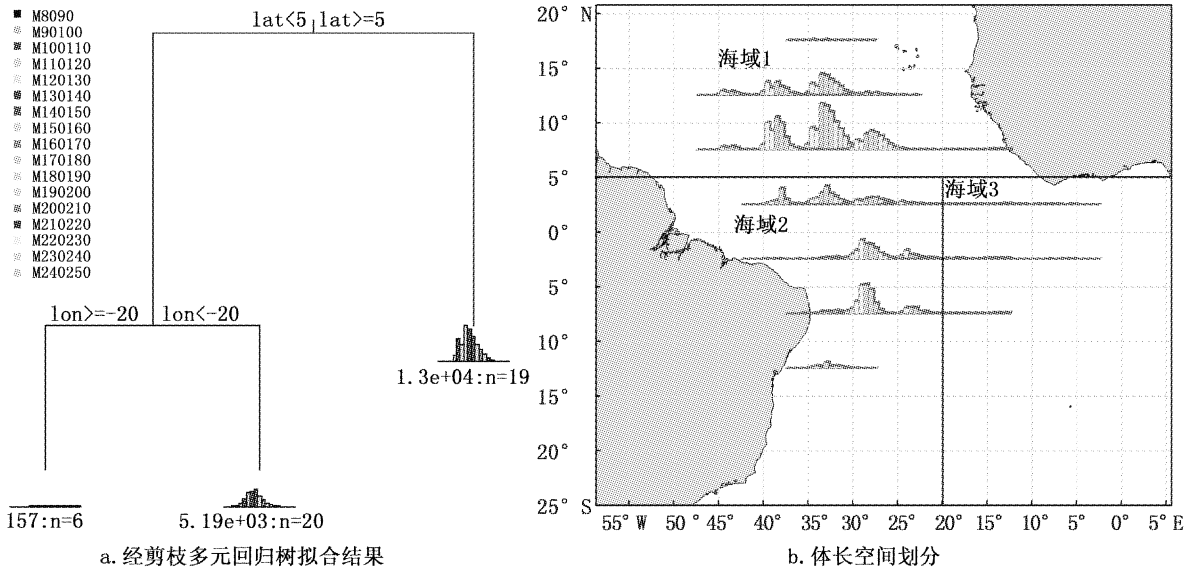


图 4 多元回归树结果

Fig. 4 The result of multivariate regression tree

lon. 经度; lat. 纬度; M70 表示体长组距 ≤ 70 cm; M7080 表示体长组距为 70 ~ 80 cm, 以此类推; 末端节点数值表示每个子节点中平均体长频率值  $P_i(j)$ 。

多元回归树分析结果所得到的不同体型大眼金枪鱼的空间分布特征与之前应用 ArcGIS 对 1~4 季度体长分布的分析结果相一致: 大型大眼金枪鱼主要集中在 7.5°N ~ 15°N, 17.5°W ~ 45°W; 中型大眼金枪鱼主要集中 12.5°S ~ 5°N, 17.5°W ~ 45°W; 小型个体主要分布在 7.5°S ~ 5°N, 5°W ~ 17.5°W。

回归树剪枝参照表 4, 得到最小误差 (xerror) 是 1.049, 最小误差对应的标准差 (xstd) 为 0.498, 1-se 准则为保证预测误差的值在最小的误差加减一个相应的标准差之间, 那么预测误差在 0.551 ~ 1.547 之间, 所以剪枝准则所采取的阈值复杂性参量 (complexity parameter, cp) 为 2 及阈值的最小误差均大于 1.0596, 所以由图 4 可见拟合的多元回归树分裂 2 次, 如分裂超过 2 次使

得数据划分过分同质而失去实际分析意义, 而分裂次数过少导致获得的拟合回归树结果不准确。

### 3 结论与讨论

#### 3.1 体长空间分布

本文选用的大眼金枪鱼样本体长范围 40 ~ 250 cm, 范围已经覆盖 1 龄至 7 龄以上的成鱼<sup>[13]</sup>, 且通过常用的 Bootstrap 再抽样方法合理的“扩大”“缩小”空间尺度的样本量, 达到数据分析需要。

本文利用多元回归树对大西洋中部大眼金枪鱼体长频率数据进行拟合, 反应了大西洋大眼金枪鱼在空间尺度上及季度上不同体型大眼金枪鱼的分布情况。本文多元回归树拟合的结果及通过 ArcGIS 绘制 1~4 季度体长尾数百分比分布图及多元回归树结果精确得到中西大西洋不同体型大眼金枪鱼个体在时空上的分布特征: 2009 年 1、2 季度体长分布相对同质; 3、4 季度体长分布相对同质。小型多集中 5°S ~ 5°N 的低纬度海域; 中型及大型个体多集中在纬度相对较高的 5°N ~ 15°N 海域。这与 ICCAT 小型幼体与成

表 4 拟合多元回归树复杂性参量表

Tab. 4 Statistics parameters of complexity of Multivariate Regression Tree

cp	nsplit	rel	error	xerror	xstd
1	0.148	0	1	1.052	0.517
2	0.073	1	0.852	1.049	0.498

鱼纬向分布描述相一致<sup>[4]</sup>,幼体由低纬度的热带海域(几内亚湾)向较高纬度的温带海域扩散现象。上述分布情况可能与大眼金枪鱼索饵的需要或大西洋中部的北赤道流、赤道逆流及加纳利洋流或是海表面温度<sup>[14-17]</sup>差异有关。此外,垂直分布研究表明成鱼与幼鱼的昼夜垂直分布不同,大西洋大眼金枪鱼成鱼较幼鱼白天栖息于更深的水层,大型个体向上洄游时间较小型个体更长<sup>[18]</sup>。

作为高度洄游性的金枪鱼种类,通过对大眼金枪鱼体长频率时空分布的研究,可以掌握中大西洋大眼金枪鱼的亲体与补充量个体的分布情况。下一步工作将进一步探究该海域鱼群性比分布<sup>[14]</sup>比例及海洋环境因子关系<sup>[16]</sup>,更好的推知大眼金枪鱼的洄游路线及分布情况,从而有效的养护开发大西洋大眼金枪鱼渔业资源,对渔船作业范围、作业网目选择性提出建议,有效的保护大眼金枪鱼幼鱼资源。

### 3.2 不足与展望

为确保数据的准确性,所有渔捞日志输入数据库前,首先由人为判断是否造假,随机选择不同页码,判断渔获体重及渔获海域是否符合实际情况。此外,结合该渔船汇报给中国远洋渔业协会的月报数据中的经纬度判断与该渔捞日志记录范围是否一致。数据库输入端,设置关键参数,如最大体重、最小体重等,如发现如果的数据超过极限值,则显示错误信息;每本渔捞日志输入完后,对其进行空间分布制图,判断该渔船作业位置是否合理。本文数据选用时,如发现疑似偏差记录,那么整本渔捞日志数据均在本文数据分析中意义排除。即便如此,录入系统只能判断严重的偏差值,对于偏差较小的数值,无法判别。且本文使用的体长数据来源于中国渔船的渔捞日志,渔捞日志记录的是加工后重量数据,本文在分析中将其转换为体长数据,可能转换中或多或少存在一些偏差。综上,体长样本可能存在一定误差。

多元回归树优点在于比单一回归树,能处理大量多类别的生物样本与环境间的关系,而对数据的要求低,能处理不完整或是浮点数据,在多元回归树方法中可以选择节点间的不纯度(impurity)测量方法<sup>[19-20]</sup>:基尼指数(Gini-index)、信息增益(Information Gain)、信息增益比

(Gain Ratio)、距离度量(Distance Measure);且不需要在自变量(一般为环境因子数据)与因变量(物种数据)建立参数估计的回归关系。因此在处理分析复杂的渔业信息(如渔获数据与海洋环境因子)上,多元回归树有很好的应用空间。

在研究方法上,本文利用多元回归树选用1-标准差(1-se),这是最常用剪枝标准,在聚类中尝试其他的剪枝标准,如Kernel评估值可能具有更好的分类结果,在今后渔业数据挖掘研究中有待进一步探究。

### 参考文献:

- [1] PAULY D, MORGAN G R. Length-based methods in fisheries research [M]. International Center for Living Aquatic Resources Management Kuwait Institute for Scientific Research, 1987: 1-7.
- [2] GULLAND J A, ROSENBERG A A. A review of length-based approaches to assessing fish stocks [R]. Food and Agriculture Organization, 1992: 2-11.
- [3] BAELD P. Growth mortality and yield-per-recruit of deep-water royal red prawns (*Haliporoides sibogae*) off eastern Australia, using the length-based MULTIFAN method [J]. Marine Biology, 1994, 118(4): 617-625.
- [4] ICCAT. Report of the 2010 ICCAT Bigeye tuna stock assessment session [R]. Pasaia, Gipuzkoa, Spain, 2010: 1-2.
- [5] 刘群,任一平,王艳君. 大眼金枪鱼的资源现状[J]. 海洋湖沼通报, 2003(2): 74-78.
- [6] 沈海学,梁振林. 大西洋大眼金枪鱼渔业概况[J]. 海洋湖沼通报, 2000(1): 68-74.
- [7] 陈新军,许柳雄,宋利明,等. 大西洋海域大眼金枪鱼年龄与生长初步研究[J]. 海洋渔业, 2006(1): 28.
- [8] 宋利明,许柳雄,陈新军. 大西洋中部大眼金枪鱼垂直分布与温度、盐度的关系[J]. 中国水产科学, 2004, 11(6): 562-566.
- [9] 戴小杰,马超,田思泉. 印度洋中国大眼金枪鱼延绳钓渔业 CPUE 标准化[J]. 上海海洋大学学报, 2011, 20(2): 276-282.
- [10] 曹晓怡,周为峰,樊伟,等. 印度洋大眼金枪鱼、黄鳍金枪鱼延绳钓渔场重心变化分析[J]. 上海海洋大学学报, 2009, 18(11): 467-471.
- [11] 陈国宝,李永振,陈丕茂,等. 鱼类最佳体长频率分析组距研究[J]. 中国水产科学, 2008, 15(4): 659-666.
- [12] VOKOUN J C, RABENI C F, STANOVICK J S. Sample-size requirements for evaluating population size structure [J]. North American Journal of Fisheries Management, 2001, (21): 660-665.
- [13] BROWN C. An overview of Atlantic bigeye tuna (*Thunnus obesus*) growth studies and implications for age-structured stock assessment [J]. ICCAT Collective Volume of Scientific

- Papers,2005(57):94-114.
- [14] HANAMOTO E. Effect of oceanographic environment on bigeye tuna distribution[J]. Bulletin of the Japanese Society of Fishery Oceanography,1987,51(3):203-216.
- [15] 樊伟,崔雪森,周魁芳. 太平洋大眼金枪鱼延绳钓渔获分布及渔场环境浅析[J]. 海洋渔业,2004,26(4):261-265.
- [16] 宋利明,高攀峰,周应祺. 基于分位数回归的大西洋中部公海大眼金枪鱼栖息环境综合指数[J]. 水产学报,2007,31(6):798-804.
- [17] 樊伟,沈新强,林明森. 大西洋大眼金枪鱼渔场、资源及环境特征的研究[J]. 海洋学报,2003,25(s2):167-176.
- [18] HOLLAND K N,SIBERT J R. Physiological thermoregulation in bigeye tuna, *Thunnus obesus*[J]. Environmental Biology of Fishes, 1994, 40(3):319-327.
- [19] DE'ATH G,FABTICOUS K E. Classification and regression trees;a powerful yet simple technique for ecological data analysis[J]. Ecology,2000,81(11):3178-3192.
- [20] DE'ATH G. Multivariate regression trees;a new technique for modeling species-environment relationships [J]. Ecology, 2002,83(4):1105-1117.

## Analysis of spatio-temporal distribution of size for *Thunnus Obesus* in the Central Atlantic Ocean based on logbook data

WANG Wen-ting<sup>1</sup>, TIAN Si-quan<sup>1,2,3,4</sup>, DAI Xiao-jie<sup>1,2,3,4</sup>, YANG Xiao-ming<sup>1,2,3,4</sup>, WU Feng<sup>1</sup>

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China; 3. Scientific Observing and Experimental Station of Oceanic Fisheries Resources and Environment, Ministry of Agriculture, Shanghai Ocean University, Shanghai 201306, China; 4. National Engineering Research Center for Pelagic Fisheries, Shanghai 201306, China)

**Abstract:** Size data is the basic information in fishery stock assessment and management. The spatio-temporal distribution of length-frequency derived from fisheries dependent data can be used to estimate the distribution of stock and recruitment. Multivariate regression tree (MRT) is a data-mining methodology that can analyze the relationships between multispecies data and environmental characteristics. In this study, based on the dressed weight data in the logbooks collected by Tuna Technology Group of China Distant-water Fisheries Association from December 2007 to December 2009, Multivariate Regression Tree and Geographical Information System (GIS) were applied to analyze spatial distribution and seasonal variation of bigeye tuna size in the Central Atlantic Ocean. The result showed that the large-sized bigeye tuna are mainly distributed in 7.5°N-15°N, 17.5°W-45°W area; Medium-sized bigeye tuna are mainly distributed in 12.5°S-5°N, 17.5°W-45°W area; Small-sized individuals are mainly distributed in 7.5°S-5°N, 5°W-17.5°W area. The results of K-S test showed that size distribution had seasonal differences apparently; Kolmogorov-Smirnov test proved that there were significant differences in size distribution among spatial areas. Quarters 1 and 2 were relatively homogeneous, and quarters 3 and 4 were relatively homogeneous.

**Key words:** length-frequency; spatio-temporal distribution; multivariate regression tree; longline; Atlantic Ocean; bigeye tuna