

基于贝叶斯原理的大洋金枪鱼渔场速预报模型研究

樊伟¹, 陈雪忠², 沈新强²

(1. 中国水产科学研究院 渔业资源遥感信息技术重点开放实验室, 上海 200090; 2. 中国水产科学研究院 东海水产研究所 农业部海洋与河口渔业重点开放实验室, 上海 200090)

摘要:高度洄游的大洋性金枪鱼类(Scombridae)是世界远洋渔业的重要捕捞对象, 中国金枪鱼生产仍处于初期发展阶段, 因此研究和预测金枪鱼渔场具有重要的现实意义。本研究利用美国 NASA 提供的卫星遥感反演海表温度(SST)三级数据产品和太平洋共同体秘书处(SPC)提供的有关国际金枪鱼历史捕捞产量资料, 分析金枪鱼同 SST 等海洋渔场环境要素之间的统计关系, 建立了金枪鱼渔场的贝叶斯概率预报模型。通过对历史数据进行模型回报试验, 结果表明太平洋鲣鱼渔场综合预报的准确性达到 70% 以上, 对渔业捕捞生产具有一定的指导意义。[中国水产科学, 2006, 13(3): 426-431]

关键词:金枪鱼; 渔场; 贝叶斯概率; 预报模型**中图分类号:**S973.1 **文献标识码:**A **文章编号:**1005-8737-(2006)03-0426-06

金枪鱼是世界上最重要的渔业产业之一, 因而成为各远洋渔业国家的争相捕捞对象。其全球总产量从 20 世纪 50 年代末的约 100 万 t 迅速发展到 1988 年总产量超过 400 万 t, 近年来全球金枪鱼类总产量更是维持在 600 万 t 左右^[1]。中国大洋金枪鱼渔业起步晚, 技术落后, 2003 年捕捞总产量仅 4 万 t 左右, 同日本、韩国及中国台湾地区差距甚远。开展金枪鱼渔场渔情速预报技术的研究, 可为中国快速发展大洋金枪鱼渔业提供科学决策和生产指导。

渔场渔情速预报模型及技术的研究, 国内外已经有 40 余年的历史^[2-10], 但由于海洋渔场及其环境的区域性、复杂性、易变性 & 渔业资源时空分布的波动性等多种原因, 人们对渔场渔情速预报模型及方法的研究仍缺少成熟统一的模型及方法。目前应用较为成功的理论与方法主要有数理统计预报模型、渔场预报专家系统和人工神经网络模型(ANN)等。统计模型如刘树勋对东海拖网渔场的预报研究和韦晟对蓝点马鲛的预报研究^[5,6]。专家系统如 Ichiro 和沈新强等分别建立的资源评估和中心渔场专家系统^[7,8], 但建立专家系统需要依赖不同的专业分析模型或建立较完备的渔业知识库及规则库才能实现。人工神经网络模型如 Laurent 对印度洋金

枪鱼渔场的研究^[9], 如 ANN 对知识的表达、存储和推理(计算)都是隐式的, 不能解释自己的推理依据和推理过程。目前国际上关于金枪鱼渔场渔情速预报业务应用的研究鲜见, 金枪鱼渔场的预报模型及方法研究仍处于探索试验阶段。本研究依据所收集到的渔场环境和渔获资料, 采用以概率统计为基础的贝叶斯概率, 构建了金枪鱼渔场的贝叶斯概率预报模型, 并成功进行了回报试验。贝叶斯概率的渔场预报已见于 Nieto 等人^[10]对秘鲁鳀鱼的预报研究, 其优点主要是充分考虑了人们已有的捕捞经验和对渔场的认识。

1 数据与方法

1.1 数据

金枪鱼捕捞数据采用太平洋共同体秘书处(SPC)提供的 1960~2000 年太平洋海域全球各国捕捞的金枪鱼产量资料, 其中捕捞产量为分月、分鱼种的 5°×5°经纬度单元网格统计产量(这里把该单元格称为渔区, 即一个 5°×5°经纬度网格为一个渔区)。由于金枪鱼围网作业主要集中在西太平洋区域, 故该组织对金枪鱼围网捕捞产量的统计范围限于 160°W 以西的太平洋海域。渔场环境资料主要为美国 NASA 提供的 1981~2001 年卫星遥感反演

收稿日期: 2005-04-07; 修订日期: 2005-11-07.

基金项目: 国家科技“863”高技术研究发展计划项目(项目编号: 2003AA637030).

作者简介: 樊伟(1971-), 男, 副研究员, 主要从事渔场海洋学及卫星遥感渔业应用研究. E-mail: dhypzh@public4.sta.net.cn

全球区域月平均海表温度(SST)数据,数据空间分辨率为 $1^{\circ} \times 1^{\circ}$,其他还包括有1997~2003年的遥感反演叶绿素数据,用于渔场环境的可视化分析。

1.2 方法

1.2.1 模型 根据所收集到的金枪鱼历史捕捞产量数据周期长和金枪鱼接近充分开发阶段的特点,金枪鱼渔场渔情分析速预报模型采用考虑人们所掌握先验知识的贝叶斯概率预报模型,金枪鱼贝叶斯概率预报模型如下式:

$$P(h_0/e) = \frac{P(e/h_0) \times P(h_0)}{\sum_{i=0}^1 P(e/h_i) \times P(h_i)} \quad (1)$$

其中: h_0 指假设为真的情况,即渔场为真的情况, h_1 为假设非真情况。 $P(h_0/e)$ 为在给定条件下为渔场的概率,也即后验概率。 $P(e/h_0)$ 为条件概率, $P(h_0)$ 为不考虑给定环境条件时渔场的先验概率。

由公式(1)可见,贝叶斯渔场概率预报模型需要给出渔场的先验概率和相应的条件概率,其预报计算过程如图1所示。依据所收集的渔获产量及渔场环境数据等情况,渔场先验概率和条件概率的计算方法主要说明如下。

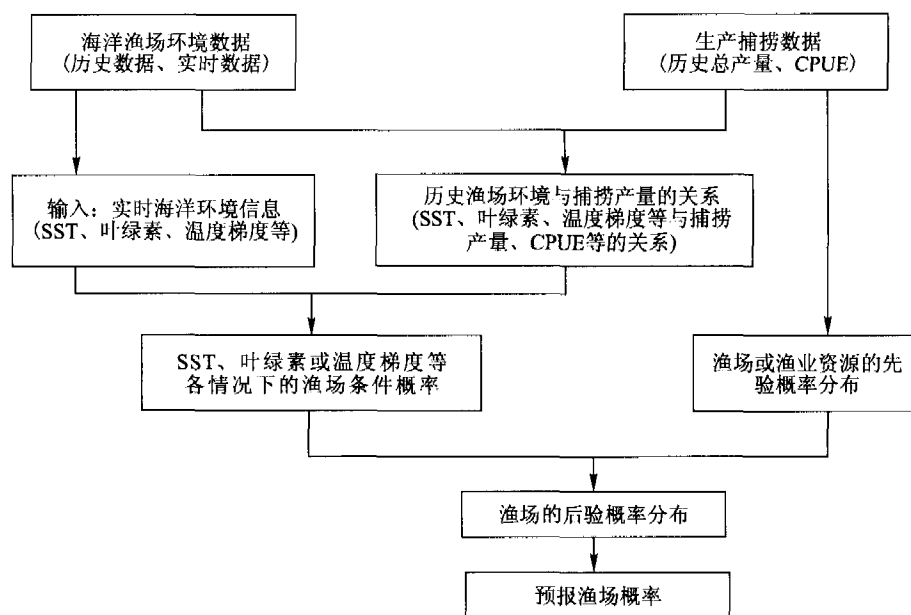


图1 金枪鱼渔场渔情速预报模型及流程

Fig.1 Model and flow chart of tuna fishing grounds forecasting

1.2.2 先验概率计算 金枪鱼渔场先验概率的计算主要利用全球金枪鱼的历史捕捞产量数据,其基本假设是渔区历史捕捞总产量(或单位努力渔获量,简称CPUE)和有渔获产量的渔区出现频率高的海域,资源丰度分布较高,形成渔场的概率就高。因此,金枪鱼渔场先验概率的计算考虑两方面的因素:

①渔区累计的渔获产量(或CPUE),也即累计渔获产量(或CPUE)高的渔区,形成渔场的概率就高,公式(2)为依据累计产量(或CPUE)的先验概率 P_1 计算公式。

$$P_1 = \frac{Y_i}{Y_{\min} + Y_{\max}} \times 100\% \quad (2)$$

式中 Y_i 为某 i 渔区的累计产量(或CPUE), Y_{\min} 为所有计算渔区中的最小产量(或最小CPUE), Y_{\max} 为所有计算渔区中的最大累积产量(或最大CPUE)。

②考虑在过去的40年(1960~2000年)期间,将渔区产量(或CPUE)达到或超过某阈值的渔区记为渔场渔区(即渔场为“真”,否则称渔场为“否”),历史上该渔区渔场为“真”的次数越多,其渔场的先验概率也较高,公式(3)为渔场先验概率 P_2 的计算方法。

$$P_2 = \frac{N_i}{N_{\text{total}}} \times 100\% \quad (3)$$

其中 N_i 为某 i 渔区的渔场为“真”的次数, N_{total} 为 40 年某一特定月份所有的统计期数。

因此,通过对累计产量(或 CPUE)和出现频次给予不同的权值,可得到综合两方面因素的渔场先验概率 P_p , 见公式(4):

$$P_p = P_1 \times (1-a) + P_2 \times a \quad (4)$$

式中 a 为 P_1 的权重, $(1-a)$ 为 P_2 的权重。

1.2.3 条件概率计算 贝叶斯概率预报模型中,除了计算得到先验概率外,还要计算出相应的条件概率。渔场条件概率的计算主要通过分析金枪鱼渔获产量或 CPUE 数据与渔场环境要素(SST、叶绿素、温度梯度等)之间的关系得到。对渔场预报来说,渔场条件概率主要指渔场的形成同渔场环境 SST、叶绿素等环境要素之间的相关性,也就是当某渔区渔场为“真”时某个渔区内的 SST 或叶绿素等达到某一条件的可能性。这里以 SST 为例说明渔场条件概率的计算方法。

通常,人们依据捕捞产量和温度数据研究鱼类的适温范围和最适温度,进而分析判断渔场最可能所处的位置。以 1 月份太平洋鲣鱼为例,将温度范围进行区间划分,可以统计出渔场为“真”时,某温度条件出现的概率(如图 2),即条件概率 $P(e/h_0)$, 见公式(5)。同理,也可统计出渔场为“否”时,出现某温度条件的概率 $P(e/h_1)$ 。见公式(6)。

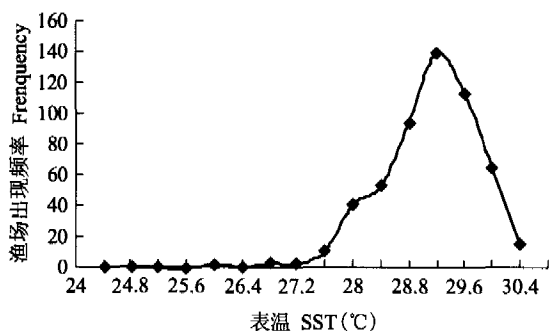


图 2 1 月份鲣鱼渔场区域的温度分布频率

Fig. 2 Temperature distribution of skipjack fishing grounds in Jan

$$p(e/h_0) = M_0/N_0 \times 100\% \quad (5)$$

$$p(e/h_1) = M_1/N_1 \times 100\% \quad (6)$$

式中: N_0 为渔场为“真”的渔区总数, M_0 为在渔区为“真”前提下,某给定的温度条件出现的总次数。 N_1 为渔场为“否”的渔区总数, M_1 为在渔区为“否”前提下,某给定的温度条件出现的总次数。

1.2.4 渔场预报概率 在计算获取渔场条件概率和先验概率分布的基础上,采用贝叶斯概率分析原理[公式(1)],便可计算预报每个渔区的后验概率,也就是所要预报的渔场形成概率。

除了 SST 外,渔场的时空分布还受其他环境因子的影响,当考虑多个环境要素时,同理可以计算得到不同环境因子的渔场预报概率,再根据不同环境因子在渔场形成中的重要性程度大小,采用加权方法得到综合的渔场预报概率。

2 结果与分析

根据上述的贝叶斯模型计算方法,本研究对太平洋金枪鱼围网渔场和延绳钓渔场分别进行了回报试验,因篇幅关系,这里以西太平洋金枪鱼围网鲣鱼渔场为例,给出有关回报试验结果,回报检验时,对公式(4)中的 a 分别取不同的权重,逐月进行回报统计。

设 P 为渔区预报概率。对于渔场为真的渔区,其预报的平均概率为

$$\overline{P_f} = \left(\sum_{i=1}^n P_i \right) / n \quad (7)$$

其中 n 为渔场渔区个数, $i=1, 2, \dots, n$ 。

对于渔场为否,也即非渔场的渔区,有

$$\overline{P_f'} = \left(\sum_{j=1}^m P_j \right) / m \quad (8)$$

其中 m 为非渔场渔区个数, $j=1, 2, \dots, m$ 。则预报误差指数用下式表示:

$$E_p = (1 - \overline{P_f})^2 + \overline{P_f'}^2 \quad (9)$$

统计显示:误差指数分布与所定义的产量(或 CPUE)阈值 x 和频度权重 a 有关,在 $x-a$ 平面直角坐标系中,其分布呈马鞍形。用多项式 $E_p = A + B \cdot x + C \cdot a + D \cdot x^2 + E \cdot x \cdot a + F \cdot a^2$ (其中 A, B, C, D, E, F 均为常数, x 表示渔获产量(或 CPUE)阈值, a 表示频度所取权重)对误差指数进行拟合。当 x 变化时, (即对不同的阈值),要使 E_p 一直保持最小,必须使 $a = (-C - E \cdot x) / (2 \cdot F)$, 在无限限制条件的情况下,这是一条直线。由于 $0 \leq a \leq 100\%$, 所以 a 的取值就随 x 值不同而在一条折线上变化,这样才能使预报的误差指数达到最小,该折线上的点以下简称最优点,折线以外的点称非最优点。例如对 1 月份鲣鱼渔场进行回报统计,误差指数及最优点(在折线 KLMN 上)分布如图 3 所示。

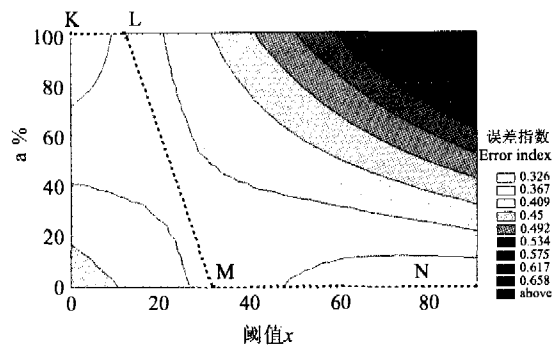


图 3 1 月份鲣鱼渔场概率回报结果的误差指数分布
注:折线 KLMN 为 a 的最优点取值
Fig.3 Differential index distribution of SKJ fishing ground forecast in Jan
Note: Curve KLMN is optimal point values of parameter a

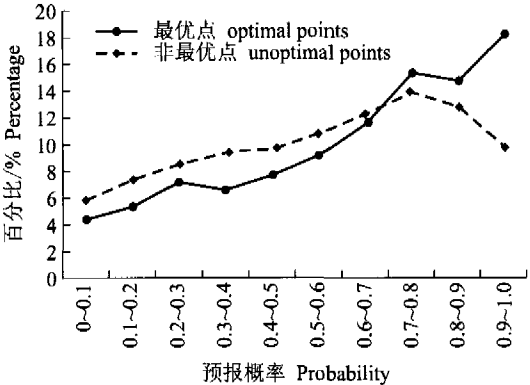


图 5 鲣鱼渔场为“真”的渔区预报概率
Fig.5 Predicting probability of skipjack fishing grounds

对于渔场为“否”的渔区来说,无论是利用最优还是利用非最优点所获得的先验概率进行回报检验,预报效果都无明显差异。如图 4 所示。而图 5 中对于渔场为“真”的渔区,考虑最优点与否,其预报效果有明显不同,即 a 取最优点情况下的预报效果显著优于非最优点。从表 1 的渔场预报结果分析中可见,对渔场为“真”的渔区来说,回报的渔场概率大于 50% 的渔区数量比例达到 68.8%,大于 60% 的渔区比例达到 59.7%。而对渔场为“否”的渔区,回报的渔场概率小于 10% 渔区比例达到 78%,小于 60% 的渔区比例高达 94.9%。如果把回报概率大于 60% 作为渔场进行划分,回报渔场和回报非渔场的综合预报准确度可以达到 77.3%。

表 1 太平洋鲣鱼渔场回报结果

Tab.1 Hindcasting results of skipjack fishing grounds

概率范围 % Scope of probability	预报结果 forecasting results			
	渔场 Fishing grounds		非渔场 Non-fishing grounds	
	渔区个数	比例	渔区个数	比例
< 10	577	4.5 %	55 789	78.0 %
10 - 20	690	5.3 %	3 701	5.2 %
20 - 30	928	7.2 %	2 857	4.0 %
30 - 40	856	6.6 %	2 084	2.9 %
40 - 50	994	7.7 %	1 831	2.6 %
50 - 60	1 182	9.1 %	1 589	2.2 %
60 - 70	1 495	11.5 %	1 357	1.9 %
70 - 80	1 976	15.3 %	1 151	1.6 %
80 - 90	1 903	14.7 %	716	1.0 %
> 90	2 355	18.2 %	449	0.6 %

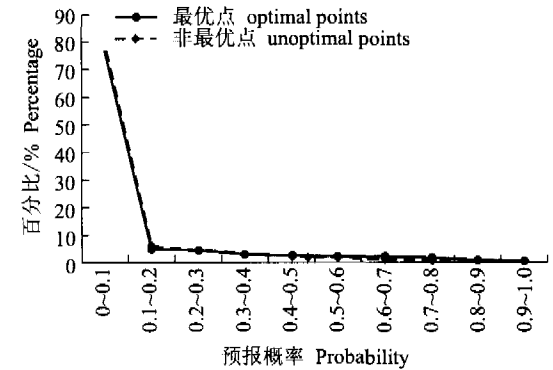


图 4 鲣鱼渔场为“非”的渔区预报概率
Fig.4 Predicting probability of skipjack non-fishing grounds

3 讨论

海洋渔场分析预报采用的模型方法不同,预报结果也呈现出不同的形式。如韦晟给出的是渔期出现时间^[6],沈新强给出了渔场位置等^[7]。本研究中采用贝叶斯概率模型进行金枪鱼渔场预报的结果用百分比概率来表示,类似气象预报中的概率预报,比较符合渔业资源分布的客观特点,与海洋现象的模糊性特点也相吻合,易于理解和为渔民和渔业用户所接受。

渔场由于渔船作业生产有淡季、旺季之分和渔业资源的年际波动变化,划分渔场区的阈值标准也应不同。回报统计结果显示(图 3),阈值 x 的取值不同,渔场先验概率计算时所考虑的因素重要性也大不一样。当阈值 x 取值较大时,渔场先验概率着

重考虑渔场出现频率因素,即当产量(或 CPUE)较高时,着重依据捕捞作业次数多少来判定是否为渔场;反之则较多地考虑产量(或 CPUE)因素,即捕捞产量(或 CPUE)成为判定渔场的重要依据。

为满足渔船实际捕捞进行渔场分析而实施渔场预报时,是以每周的 SST 数据作为环境条件,而由于历史金枪鱼捕捞数据统计精度的限制,本研究中渔场先验概率是以每月为周期进行计算的,因此,周期的不一致性使得渔场预报的精度提高存在一定困难,如何根据实际调查资料,对渔场的先验概率进行修正,进而得到每周的渔场先验概率是需要进一步探索解决的问题,如可以考虑根据前一个预报周期的渔场生产位置修正渔场先验概率。此外,渔场条件概率的计算也是以每月的捕捞产量和月平均 SST 进行统计的,由于月平均 SST 相当于经过平滑处理后的信息,损失了短周期的信息,对于统计计算获取准确的渔场条件概率也有一定影响。从预报结果精度来看,若不考虑非渔场区的预测结果,预报概率超过 60% 的渔场区域的预测精度仅 59.7%,并且由于所定义预报渔区网格单元较大($5^{\circ} \times 5^{\circ}$),这对于渔船瞄准渔场捕捞仍显得不够。

海洋渔场分析预报作为渔场海洋学的重要研究方向之一,对于海洋生态动力学的深入研究也大有裨益,因此,以卫星遥感等获取的多源海洋环境数据为基础,建立以海洋生态动力机制为框架的渔场动态预报模型是需要进一步探索的课题,以提高渔场预报精度和实际应用效果。

致谢:美国 NASA 和南太平洋共同体秘书处(SPC)为本研究提供了 SST 和渔场生产资料,谨致谢忱!

参考文献:

- [1] 苗振清,黄锡昌. 远洋金枪鱼渔业[M]. 上海:上海科学技术文献出版社, 2003.
- [2] 宇田道隆. 海洋渔场学[M]. 东京:恒星社厚生阁发行所, 1963. 20-22.
- [3] 毛志华,朱乾坤,潘德炉,等. 卫星遥感速报北太平洋渔场海温方法研究[J]. 中国水产科学, 2003, 10(6): 502-505.
- [4] Lehodey P, Andre J M, Bertignac M, et al. Predicting skipjack tuna forage distributions in the equatorial Pacific using a coupled dynamical bio-geochemical model[J]. Fish Oceanogr, 1998, 7: 3/4, 317-325.
- [5] 刘树勋,韩士鑫,魏永康. 判别分析在渔情预报中应用的研究[J]. 海洋通报, 1988, 7(1): 63-70.
- [6] 韦 晟,周彬彬. 黄渤海蓝点马鲛短期渔情预报的研究[J]. 海洋学报, 1988, 10(2): 216-221.
- [7] 沈新强,樊 伟,韩士鑫,等. 中心渔场智能预报系统的设计与实现[J]. 中国水产科学, 2000, 7(2): 69-72.
- [8] Ichiro A, Tadashi I, Isamu M, et al. A prototype expert system for predicting fishing condition of anchovy (Engraulidae) off the coast of Kanagawa Prefecture[J]. Nippon Suisan Gakkaishi, 1989, 55(10): 1777-1783.
- [9] Laurent D, Michel P, Stretta J M. Simulation of large scale tropical tuna movements in relation with daily remote sensing data: the artificial life approach [J]. Biosystems, 1997, 44: 167-180.
- [10] Nieto K, Y nez E, Silva C. Probable fishing grounds for anchovy in the northern Chile using an expert system [A]. I-GARSS, International Geoscience and Remote Sensing Symposium[C]. Sidney, IEEE, 2001. 9-13.

Tuna fishing grounds prediction model based on Bayes probability

FAN Wei¹, CHEN Xue-zhong², SHEN Xin-qiang²

(1. Key laboratory of fisheries resources remote sensing and information technology, Chinese Academy of Fishery Sciences, Shanghai, 200090, China; 2. Key laboratory of marine & estuarine fisheries, Ministry of Agriculture, East China Sea Fisheries Research Institute, Shanghai, 200090, China)

Abstract: Highly migration tuna is one of the most important fishing objects of high sea fisheries of the world. At present, the tuna fishery of China is at the primary development period, and it is practically significant to engage in forecasting and studying on tuna fishing grounds. Based on three-level satellite data of SST supplied by NASA and historical tuna catch data supplied by SPC, using a geographical information system (GIS), relations between the catch of tuna and SST were studied. With this information and using the Bayesian theory approach, a tuna fishing grounds forecasting expert system was set up and was developed generating probable fishing grounds charts. Bayes models are different from generic statistical method, in which not only the model information and data information, but also transcendental information is used adequately. The result of 40-years-hindcasting experiments shows that the predicting accuracy of skipjack fishing grounds in West Pacific ocean is over 70%, which is significant to guide fishermen's fishing operation. However, because the computation period of fishing grounds transcendental probability and conditional probability is every month, it must be modified according to field survey data for future fishing grounds forecasting ever week. [Journal of Fishery Sciences of China, 2006, 13(3): 426 - 431]

Key words: Tuna; Fishing grounds; Bayes probability; Prediction model