

魏广恩, 陈新军. 基于产卵场和索饵场适宜性的西北太平洋柔鱼丰度预测[J]. 海洋学报, 2020, 42(12): 14–25, doi:10.3969/j.issn.0253-4193.2020.12.002

Wei Guang'en, Chen Xinjun. Forecasting Northwest Pacific Ocean neon flying squid abundance based on suitability of spawning and feeding grounds[J]. Haiyang Xuebao, 2020, 42(12): 14–25, doi:10.3969/j.issn.0253-4193.2020.12.002

基于产卵场和索饵场适宜性的西北太平洋柔鱼丰度预测

魏广恩¹, 陈新军^{1,2,3*}

(1. 上海海洋大学 海洋科学学院, 上海 201306; 2. 大洋渔业资源可持续开发教育部重点实验室, 上海 201306; 3. 国家远洋渔业工程技术研究中心, 上海 201306)

摘要: 柔鱼 (*Ommastrephes bartramii*) 是北太平洋重要的经济头足类。短生命周期的特征使其资源丰度主要取决于补充量, 早期生活史阶段的海洋环境将直接影响补充量大小。利用 2004–2015 年我国鱿钓船队在西北太平洋的生产统计数据, 以及产卵场和索饵场海表水温, 将产卵场和索饵场等分成不同数量的海区, 运用相关性分析和随机森林模型, 筛选出产卵期各月产卵场最适海表水温范围占总面积的比值 (Ps) 和索饵期各月索饵场最适海表水温范围占总面积的比值 (Pf) 与单位捕捞努力量渔获量 (CPUE) 显著相关的海区, 将 Ps 值或 Pf 值作为神经网络模型的输入变量, 分别构建基于产卵场、索饵场的资源丰度预报模型, 分析模型的优劣与预报准确度。结果表明: 产卵场划分方案为 5°×5°, 索饵场为 2.5°(经)×4°(纬) 较为合适。随机森林筛选出的海区与相关性分析筛选出的适宜海区范围大致吻合, 且随机森林能够识别与 CPUE 相关的潜在海域。模型的预报结果表明, 其预报准确度均达到 90% 以上, 随机森林筛选出的海区的最适海表水温范围占该海区的比值作为神经网络模型输入因子构建的模型优于相关性分析, 预报准确度更高。基于产卵场海域构建的模型相对于索饵场, 模型精确度更高、更稳定。

关键词: 西北太平洋; 柔鱼; 资源量; 随机森林; 神经网络

中图分类号: S931.1

文献标志码: A

文章编号: 0253-4193(2020)12-0014-12

1 引言

柔鱼 (*Ommastrephes bartramii*) 是一种短生命周期的海洋性头足类, 是北太平洋重要经济种类^[1]。日本鱿钓船最早于 1974 年对该资源进行利用^[2], 我国对该资源的调查始于 1993 年, 1994 年后开始进行商业化生产, 主要捕捞对象为冬春生群体^[3]。历年来的生产情况表明其资源年间变动较大, 因此, 探究其资源丰度波动原因对实际生产有重要意义。

柔鱼是一种“生态机会主义”物种, 其渔场分布和资源丰度都易受海洋环境的影响^[4]。已有的研究表明^[5-7], 柔鱼资源量及分布受产卵场和索饵场海水表层温度和叶绿素浓度等因素的影响, 补充量大小则受产卵场适宜水温范围影响。造成这一现象的主要原因是短生命周期的柔鱼在亲体产完卵后即死亡, 其资源量大小主要取决于补充量^[8-9], 而补充量则主要取决于早期存活率。该阶段的柔鱼(胚胎和仔鱼)自主选择适宜生存环境的能力较弱, 容易产生较高的自然

收稿日期: 2019-07-09; 修订日期: 2019-10-30。

基金项目: 国家重点研发计划(2019YFD0901404); 国家自然科学基金面上项目(NSFC41876141); 上海市科技创新行动计划(10DZ1207500)。

作者简介: 魏广恩(1991—), 男, 安徽省淮南市人, 研究方向为渔业资源。E-mail: wge1991@163.com

* 通信作者: 陈新军(1967—), 男, 教授。E-mail: xjchen@shou.edu.cn

死亡率,海洋环境的变化会影响到仔鱼的生长和存活,进而影响柔鱼资源补充量^[9]。

Murakami等^[11]认为北太平洋柔鱼存在3个产卵场,分别为140°~150°E海域、170°E附近海域和160°W~180°W海域。Nakamura^[12]和Okutain^[13]分别通过性成熟雌性个体的分布和仔稚鱼出现海域验证了这些产卵场的存在。西北太平洋柔鱼冬春生群体的产卵场为130°~170°E^[14]。以往研究中^[15],以单位捕捞努力量渔获量(Catch Per Unit Effort, CPUE)来表征渔业资源丰度的相对指标,以产卵场最适海表面温度的范围来表征柔鱼产卵场环境条件的优劣,通过计算柔鱼产卵期整个产卵场最适海表面温度的范围占总面积的比值(Ps)序列值与对应年份CPUE组成的时间序列值的相关性,来分析产卵场环境对西北太平洋柔鱼资源补充量的影响。西北太平洋柔鱼产卵场是否均匀分布在130°~170°E海域,或者只是分布于该海域内的局部区域尚未定论。故直接使用整个产卵场的Ps来分析柔鱼产卵场环境对柔鱼资源补充量的影响,会因为产卵场经纬度跨度较大,造成产卵场的Ps序列值和CPUE序列值之

间的相关性降低。索饵场最适海表面温度的比例(Pf)的计算亦是如此。故本研究通过对产卵场和索饵场的逐步等分,通过相关性决定等分的空间尺度,通过随机森林得出各个产卵场和索饵场对CPUE的重要性,以此来确定各月份所对应的产卵场和索饵场海域。以相关性和随机森林模型筛选出的最适海表面温度的范围占总面积的比值,作为自变量,以CPUE作为因变量,构建神经网络模型,分析相关性和随机森林两种筛选方式下,神经网络模型预测资源丰度的准确性。

2 材料和方法

2.1 数据来源

2.1.1 渔业数据

西北太平洋柔鱼渔获数据来源于上海海洋大学中国鱿钓技术组,时间为2004–2015年,作业海域主要分布在35°~50°N,140°~160°E范围内。空间分辨率为1°×1°,数据包括日期(年、月)、经度、纬度、产量(t)、年作业船数(表1)等。本研究以CPUE表征西北太平洋柔鱼的资源丰度(图1)。

表1 2004–2015年各年份作业船数

Table 1 The numbers of fishing vessels from 2004 to 2015

年份	2004年	2005年	2006年	2007年	2008年	2009年	2010年	2011年	2012年	2013年	2014年	2015年
作业船数/艘	212	227	327	255	258	273	262	191	225	237	186	116

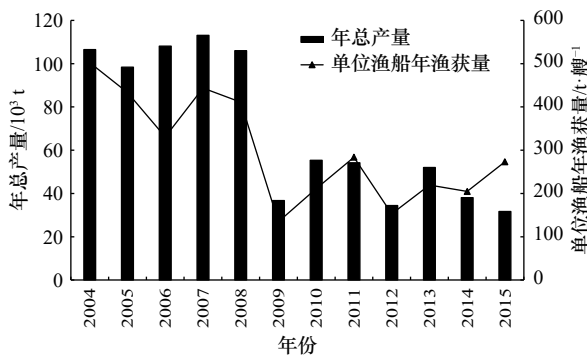


图1 2004–2015年西北太平洋柔鱼产量及CPUE

Fig. 1 The annual total catch and catch per unit effort (CPUE) of *O. bartramii* from 2004 to 2015 in the Northwest Pacific Ocean

2.1.2 环境数据

根据前人研究^[16–17],确定了西北太平洋柔鱼冬春生群体的产卵场分布海域和月份,索饵场分布海域和月份。环境数据包括产卵场(20°~30°N,130°~170°E)(图2)和索饵场(38°~46°N,150°~165°E)(图2)的海

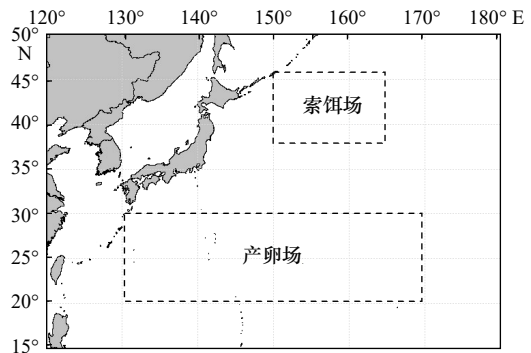


图2 西北太平洋柔鱼产卵场和索饵场分布

Fig. 2 Distribution of spawning and feeding grounds of *O. bartramii* in the Northwest Pacific Ocean

表水温(SST),时间分辨率为月,空间分辨率为0.1°×0.1°,用于计算产卵场Ps值和索饵场Pf值。时间范围为2004–2015年,其中产卵场为1–4月,索饵场为8–11月。SST数据来源于NOAA OceanWatch数据库(http://pifscocceanwatch.irc.noaa.gov/erddap/griddap/OceanWatch_pfgac_sst_monthly.html)。

2.2 实验方法

2.2.1 确定产卵场与索饵场的划分方案

产卵场和索饵场的划分方案遵循逐步等分, 经度和纬度交叉分割的方式。结合产卵场和索饵场海域的实际情况, 进行逐步分割。产卵场海域具体划分方案如图 3 所示, 索饵场具体划分方案如图 4 所示。

分别计算产卵场产卵期(1-4月)和索饵场索饵期(8-11月)各海区各月份对应的 P_s 和 P_f (产卵场产卵期 1-4 月最适海表水温范围为 21~25°C, 索饵场索饵期 8-11 月各月份最适海表水温范围不同, 8 月为 15~19°C; 9 月为 14~18°C; 10 月为 12~15°C; 11 月为 10~13°C^[16-17])。分析各海区各月份的 P_s 和 P_f 时间序列值与 CPUE 时间序列值的相关性。通过相关性分析作为最终划分方案的选择依据。

2.2.2 随机森林法提取产卵场和索饵场海域

1) 随机森林定义及其思想

随机森林就是通过集成学习的思想将多棵树集成的一种算法, 它的基本单元是决策树, 而它的本质属于机器学习的一大分支——集成学习(ensemble learning)方法^[18]。随机森林是一个树型分类器 $\{h(x, \beta_k), k = 1, \dots\}$ 的集合, 其中元分类器 $h(x, \beta_k)$ 是使用 CART 算法构建的没有剪枝的分类决策树; x 是输入向量; β_k 是独立同分布的随机向量, 决定了单棵树的生长过程。通过自助法重采样技术, 从原始的训练样本集 N 中进

行有放回的重复随机取样, 抽取 k 个样本, 生成新的训练集样本集合, 自助样本集生成的 k 个决策树组成随机森林, 新数据的分类结果按决策树投票多少形成

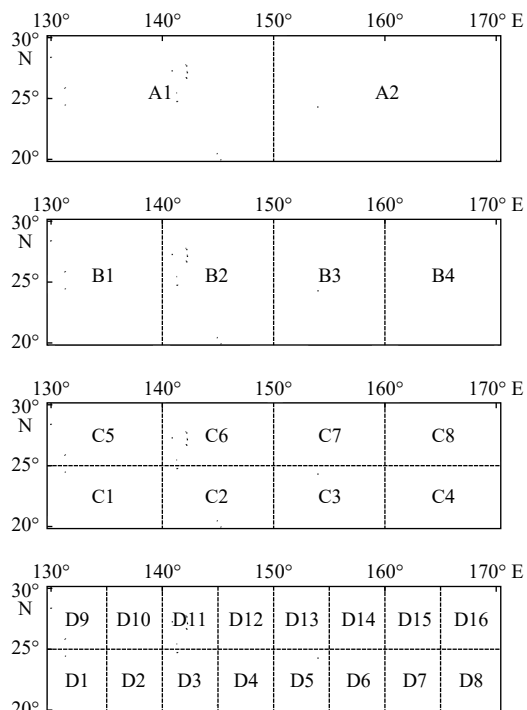


图 3 西北太平洋海域柔鱼产卵场的划分方案

Fig. 3 The division of spawning grounds of *O. bartramii* in the Northwest Pacific Ocean

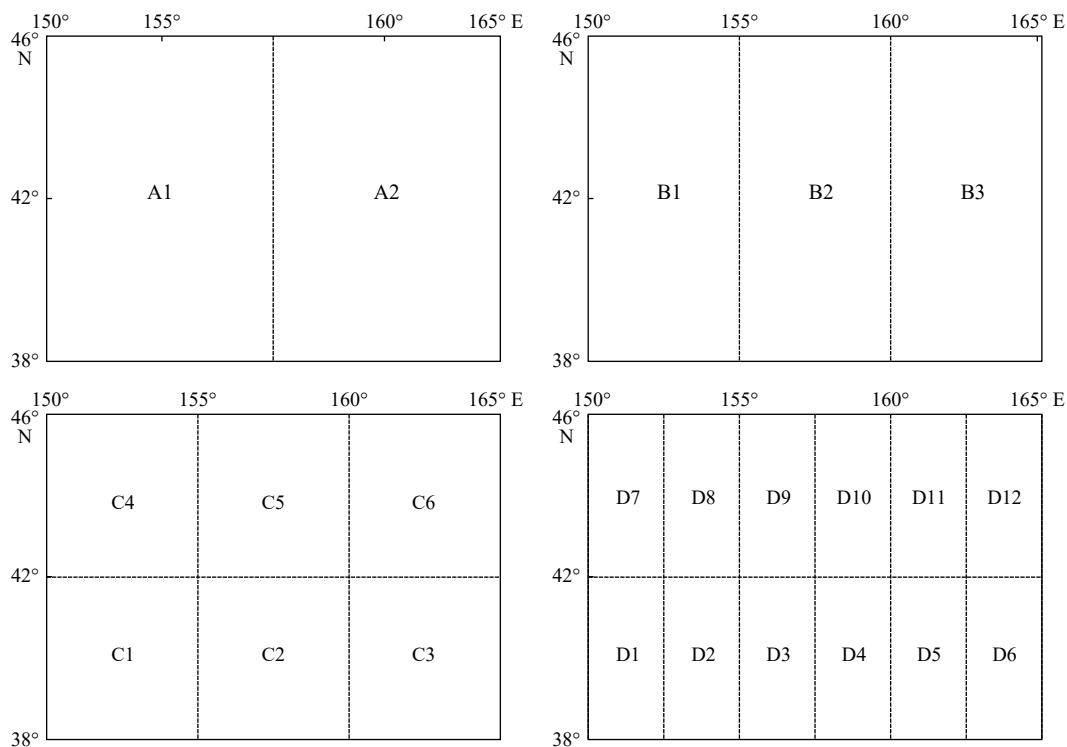


图 4 西北太平洋海域柔鱼索饵场的划分方案

Fig. 4 The division of feeding grounds of *O. bartramii* in the Northwest Pacific Ocean

的分数而定。随机森林就是对数据集的列变量和行观测进行随机化,生成多个分类数,最终将分类树结果进行汇总^[19]。随机森林在降低运算量的同时也能够保证预测精度,而且该算法对多元共线性不敏感,对缺失数据和非平衡数据比较稳健,可以很好地适应多达几千个解释变量的数据集。

2) 随机森林的估计过程

随机森林建立模型以及预测的具体步骤如图5所示。

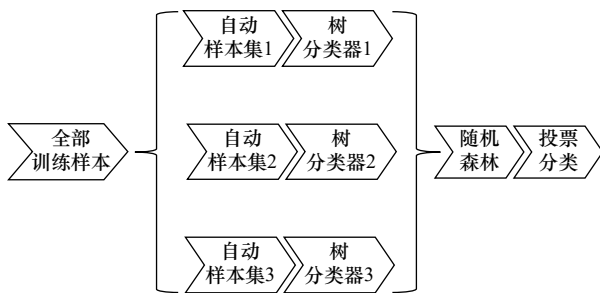


图5 随机森林工作原理

Fig. 5 The working principle of random forest

(1) 首先,用 N 表示原始训练集样本的个数,用 M 表示变量的数目。确定一个定值 m ,用来决定当在一个节点上做决定时,会使用到多少个变量。即随机产生 m 个变量用于节点上的二叉树,二叉树变量的选择仍然满足节点不纯度最小原则。

(2) 应用 Bootstrap 自助法在原数据集中有放回地随机抽取 k 个样本集,组成 k 棵决策树,而对于未被抽取的样本用于单棵决策树的预测。

(3) 每个自助样本集生长成为单棵决策树。在树的每个节点处从 M 个特征中随机挑选 m 个特征 (m 小于 M),按照节点不纯度最小的原则从这 m 个特征中选出一个特征进行分支生长。这棵决策树进行充分生长,使每个节点的不纯度达到最小,不进行通常的剪枝操作。根据 k 个决策树组成的随机森林对待分类样本进行分类或预测,分类的原则是投票法,预测的原则是简单平均^[20]。

3) 随机森林的两个重要参数

树节点预选的变量个数和随机森林中树的个数,这两个参数是构建随机森林模型过程中的两个重要参数,也是决定随机森林预测能力的两个重要参数。其中第一个参数决定了单棵决策树的情况,而第二个参数决定了整片随机森林的总体规模。两个参数分别从随机森林的微观和宏观层面上决定整片随机森林的构造^[21-23]。

4) 随机森林模型评价因素

(1) 森林中单棵树的分类强度:每棵树的分类强度越大,即每棵树生长越茂盛,组成森林的分类性能越好。

(2) 森林中树之间的相关度:树与树之间的相关性越差,或树之间是独立的,则森林的分类性能越好^[24-25]。

2.2.3 神经网络预报模型

误差反向传播神经网络(Error Backpropagation Network, EBP)属于多层前向神经网络,利用误差反向传播的监督算法,学习和存储大量的模式映射关系^[26-27]。该算法主要包括学习过程信号的正向传播与误差的反向传播两个过程。正向传播时,样本从输入层进入,经隐藏层激活函数处理,传向输出层,如果输出层的实际输出与期望的输出不符合误差要求,则转入误差的反向传播阶段。反向传播是将误差以某种形式通过隐藏层向输入层逐层反向传播,将误差分摊给各层所有节点,从而获得各层节点的误差信号,以此误差信号作为修正的依据。这种信号的正向传播与误差的反向传播周而复始地进行,权值不断调整,即为网络学习的过程,直至网络输出的误差减少到可接受程度或进行到预先设定的学习次数为止^[28]。

利用 2.2.1 节和 2.2.2 节所得相关性和重要性海域,按照产卵场和索饵场分别组成样本集,建立神经网络的预报模型。经归一化处理后,将 2004-2013 年的样本,随机化分配 70% 作为训练样本,30% 作为测试样本,2014-2015 年的样本作为验证样本。将最适海表水温范围比值 P_s 和 P_f 作为输入向量,CPUE 作为输出向量。输入层神经元个数根据筛选出的 P_s 或 P_f 海域个数决定。输出神经元为 CPUE,隐含层神经元个数根据经验公式确定:

$$J = \sqrt{n+m} + a, \quad (1)$$

式中, J 为隐含层节点数; m 为输出层节点数; n 为输入层节点数; a 为 1~10 之间的常数。其中,学习速率为 0.1,动量参数为 0.5。最大训练次数为 1000 次,最大误差给定 0.001,经多次训练,取最优结果,防止过拟合出现^[29]。

以均方误差(Mean Squared Error, MSE)作为判断最优模型的标准。均方误差(式(2))根据测试样本的预报值与实际值进行计算,

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2, \quad (2)$$

式中, y_k 为 CPUE 的实际值; \hat{y}_k 为 CPUE 的预报值。比较产卵场和索饵场在两种筛选条件下所得海区的

Ps 和 Pf 作为输入变量时,模型的均方误差。

根据模型的 CPUE 预测值,结合 2014–2015 年的 CPUE 实际值,分别计算各模型的预报准确度,计算公式为

$$P = \frac{1}{N} \sum_{k=1}^N \left(1 - \frac{|y_k - \hat{y}_k|}{y_k} \right) \times 100\%, \quad (3)$$

式中, P 表示预报准确度; y_k 为 CPUE 的实际值; \hat{y}_k 为 CPUE 的预报值。计算产卵场和索饵场在两种筛选条件下所得海区的 Ps 和 Pf 作为输入变量时,以每种模型预报准确度的均值,作为该模型的预报准确度。

3 结果与分析

3.1 产卵场和索饵场划分方案的确定

通过上述划分方案,分别对产卵场和索饵场进行划分,计算每一个划分海区的最适海表水温范围的比值,分析产卵场 Ps 与 CPUE、索饵场 Pf 与 CPUE 序列值的相关性,结果如表 2、表 3 所示。从表中可以看出,产卵场和索饵场随着划分海区的增加,与 CPUE 序列值出现显著相关性的海区随之增加;另一方面,也验证了上文中提及的直接使用整个产卵场或索饵场最适海表水温比例的序列值,分析其与 CPUE 序列值之间的相关性,将造成相关性降低的现象。故确定 D 方案作为产卵场和索饵场的划分方案。

3.2 随机森林确定索饵场和产卵场海域

3.2.1 随机森林的两个重要参数

根据上述各方案下 Ps、Pf 与 CPUE 时间序列相关

表 2 西北太平洋海域柔鱼产卵场 Ps 与 CPUE 时间序列的相关性和显著性

Table 2 The correlation and significance of Ps in the spawning grounds with CPUE time series of *O. bartramii* in the Northwest Pacific Ocean

方案	1月		2月		3月		4月	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
A	-	-	-	-	-	-	-	-
B3	-	-	-	-	-	-	-0.660	<0.05
C5	0.600	<0.05	-	-	-	-	-	-
C6	-	-	0.710	<0.05	-	-	-	-
D3	-	-	-0.670	<0.05	-	-	-	-
D9	0.650	<0.05	-	-	-	-	-	-
D11	-	-	0.780	<0.01	-	-	-	-
D13	-	-	-	-	-	-	-0.600	<0.05

注: -表示该月份研究海域产卵场 Ps 与 CPUE 时间序列值相关性分析结果不显著。

表 3 西北太平洋海域柔鱼索饵场 Pf 与 CPUE 时间序列的相关性和显著性

Table 3 The correlation and significance of Pf in the feeding grounds with CPUE time series of *O. bartramii* in the Northwest Pacific Ocean

方案	8月		9月		10月		11月	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
A1	-	-	-	-	-0.580	<0.05	-	-
B1	-	-	-	-	-0.610	<0.05	0.590	<0.05
C1	-	-	-	-	-0.600	<0.05	-	-
C4	-	-	-	-	-	-	0.840	<0.01
D1	-	-	-	-	-0.610	<0.05	-	-
D6	-	-	-	-	-	-	0.750	<0.01
D7	-0.570	<0.05	-	-	-	-	0.740	<0.01
D8	-	-	-	-	-	-	0.710	<0.05
D10	-	-	-	-	-	-	0.590	<0.05

注: -表示该月份研究海域索饵场 Pf 与 CPUE 时间序列值相关性分析结果不显著。

性出现的规律,确定产卵场和索饵场海区的划分方案。最终选取 D 方案下,产卵场和索饵场所划分的海区作为随机森林的分析对象。分别选取产卵场和索饵场所有划分海区各月份的 Ps 值和 Pf 值作为输入向量,通过随机森林模型,计算各自变量对模型的重要性,筛选出重要性海区。

采用逐一增加变量的方法来进行建模,根据模型的误判率均值最小原则,确定模型中决策树的节点最优变量个数,构建最优模型。从图 6 中可以确定,当构建的 Ps 模型决策树节点所选变量数为 13 时,模型的误判率均值最低,确定 Ps 模型最优变量个数为 13 个。当构建的 Pf 模型决策树节点所选变量数为 12 时,模型的误判率均值最低, Pf 模型最优变量个数为 12 个。

确定模型中决策树的节点最优变量个数之后,根据模型误差是否趋于稳定,确定最优模型中决策树的数量。从图 7 中可以看出, Ps 模型中决策树数量在 300 以内时,模型误差会出现较大的波动,当决策树数量大于 300 以后,模型误差趋于稳定,但仍有少许变化,该模型误差最小值出现在决策树数量为 500 左右,故确定 Ps 模型中最优决策树数量为 500。同样 Pf 模型中决策树数量在 200 以内时,模型误差会出现较大的波动,当决策树数量大于 200 以后,模型误差趋于稳定,进而确定 Pf 模型中最优模型中决策树的数量为 480。

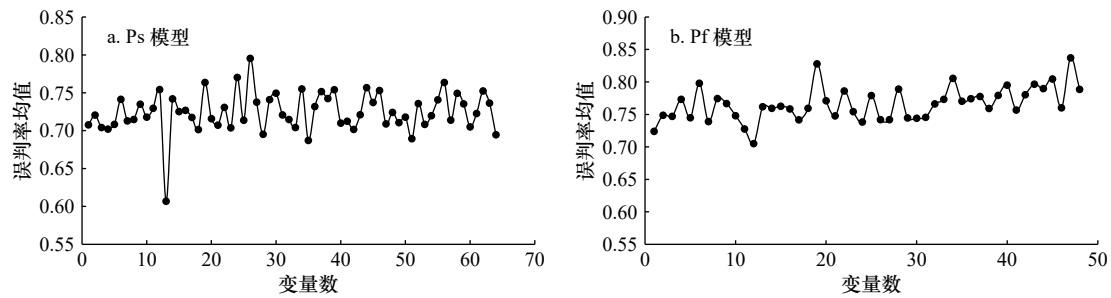


图6 不同决策树节点分支变量数模型的误判率均值

Fig. 6 The mean of misjudgment rate of models with different numbers of branch variable on the decision trees node

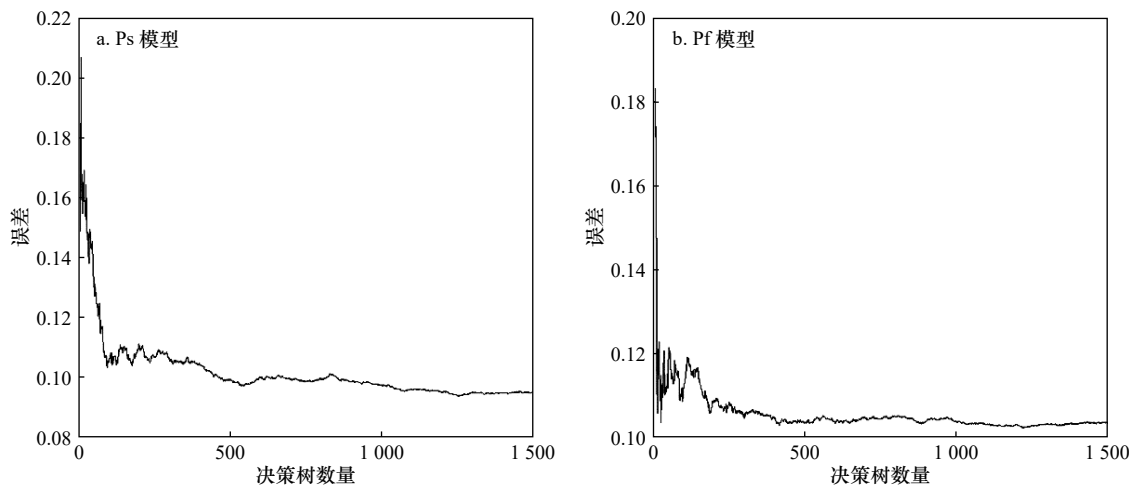


图7 决策树数量和模型误差关系

Fig. 7 The relation between model error and numbers of decision trees

3.2.2 各月份重要海区的选取

通过随机森林模型的构建,可以得到产卵场和索饵场不同月份下各海区对模型影响的重要值。各海区重要值越高,说明该海区对模型的影响越大。本研究以精度平均减少值度量各海区的的重要值。产卵场在产卵期不同海区的重要值如图8所示。

以精度平均减少值15为筛选标准,产卵场重要海区分布:1月为D9海区、D10海区,2月为D11海区,3月无重要性海区,4月为D13海区、D16海区。比较于相关性筛选出的海区,差异在于1月增加了D9的相邻海区D10,2月减少了D11的相邻海区D3(负相关),3月一致,4月增加了D16海区。此结果与相关性海区结果基本吻合。

以精度平均减少值10为筛选标准,产卵场重要海区增加范围:1月份增加D1、D7、D12、D16海区;2月份增加D8、D12、D16海区;3月份增加D6、D8、D15、D16海区,4月份增加D7海区。索饵场在索饵期不同海区的重要性如图9所示。

以精度平均减少值15为筛选标准,索饵场重要海区分布:8月为D5海区、D7海区,9月为D6海区,10月为D1海区,11月为D1海区、D6海区、D7海

区、D8海区。比较于相关性筛选出的海区,差异在于8月增加了D5海区,9月增加了D6海区,10月一致,11月增加了D1海区,减去了D10海区。此结果与相关性海区结果基本吻合。

以精度平均减少值10为筛选标准,索饵场重要海区增加范围:8月份增加D1海区、D3海区、D4海区、D8海区;9月份增加D1海区、D4海区、D5海区、D7海区、D8海区、D12海区;10月份无变化,11月份增加了D9海区。

3.3 资源丰度的预测

3.3.1 神经网络模型的构建

根据上述结果,将相关性作为筛选条件所得海区的最适海表水温范围占该海区面积的比值,以及随机森林模型中所得重要性海区的最适海表水温范围占该海区面积的比值作为输入变量,构造的模型如表4所示,其中,I和II分别是基于相关性筛选的产卵场各海区于产卵期(1-4月)各月份的Ps时间序列值和索饵场各海区于索饵期(8-11月)各月份的Pf时间序列值与CPUE时间序列值显著相关的海区的Ps值和Pf值作为输入因子,构建网络结构(输入层:隐藏层:输出层)分别为4:5:1和6:7:1的神经网络模型;

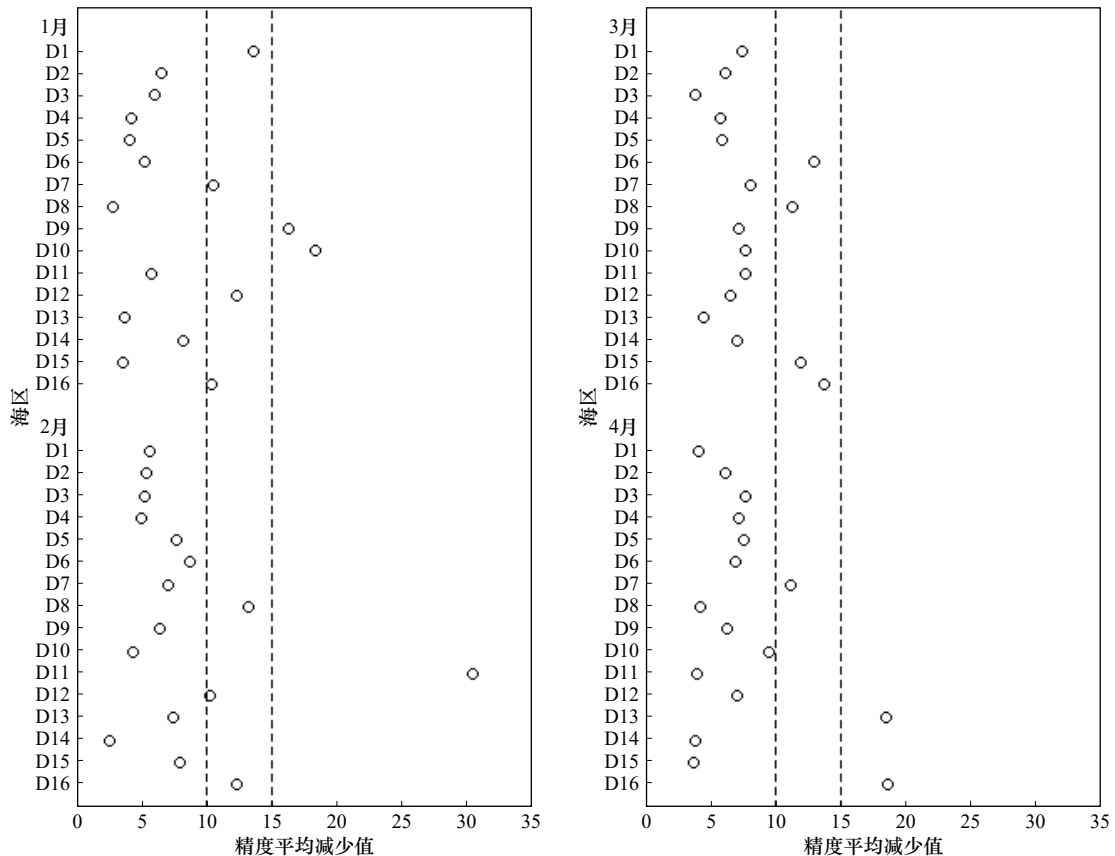


图 8 产卵期各海区重要性分布

Fig. 8 The distribution of each sea area's importance during spawning period

III 和 IV 分别是基于随机森林模型以精度平均减少值 15 为筛选条件获得的产卵场各海区于产卵期(1-4 月)各月份和索饵场各海区于索饵期(8-11 月)各月份重要海区的 Ps 值和 Pf 值作为输入因子,构建神经网络结构分别为 5 : 6 : 1 和 8 : 9 : 1 的神经网络模型。

3.3.2 预报模型实现及结果比较

通过上述构造神经网络预测模型方案,进行模型拟合,模型的均方误差和预报准确度如图 10 所示。D 方案下,产卵场和索饵场划分海区的最适海表水温范围占该海区面积的比值作为输入变量构建的神经网络预测模型均能准确预报柔鱼的资源量,预报准确度在 90% 以上。基于随机森林选取重要性海区的最适海表水温范围占该海区面积的比值构建的模型的预报准确度高于相关性筛选的海区。而基于 D 方案下产卵场划分海区的最适海表水温范围占该海区面积的比值作为输入变量的神经网络预测模型的均方误差,明显小于索饵场。

4 讨论与分析

4.1 相关性海域分析

图 11 中黑色实线包围的海域,为整个产卵场最

适海表水温所占范围。根据 D 方案的划分方式,对产卵场进行划分,可以大致了解划分的各个海区最适海表水温分布情况。从图中可以看出,计算产卵场整片海域的最适海表水温范围所占的比值与按照不同的方案划分情况下计算 Ps 差异较大,不同经度和纬度,最适海表水温空间分布不同,若采用整个产卵场的 Ps 进行分析,极易削弱产卵场海域 Ps 与 CPUE 时间序列的显著相关性,故需要将整个产卵场按照不同方案进行划分,从而找到与 CPUE 序列值显著相关的海域,索饵场亦是如此。

通过产卵场的相关性分析可以看出,随着划分海域不断增多,各海域面积减小,Ps 与 CPUE 时间序列出现显著性的海域也不断增多。主要集中在 1 月、2 月和 4 月。首先从相关性可以肯定,柔鱼资源补充量的变动和各海域最适海表水温的比值存在显著性,相关性却出现差异正负,D 方案下,D3 海区与 D11 海域经度范围一致,纬度范围分别为 20°~25°N 和 25°~30°N。在相同的月份(2 月)却出现了相反的相关性,表明:25°~30°N 海域内 21~25°C 作为最适温度计算所得的 Ps 和 CPUE 的正相关性成立,在 25°N 以南海域,该温度范围下计算的 Ps 与 CPUE 的负相关性表

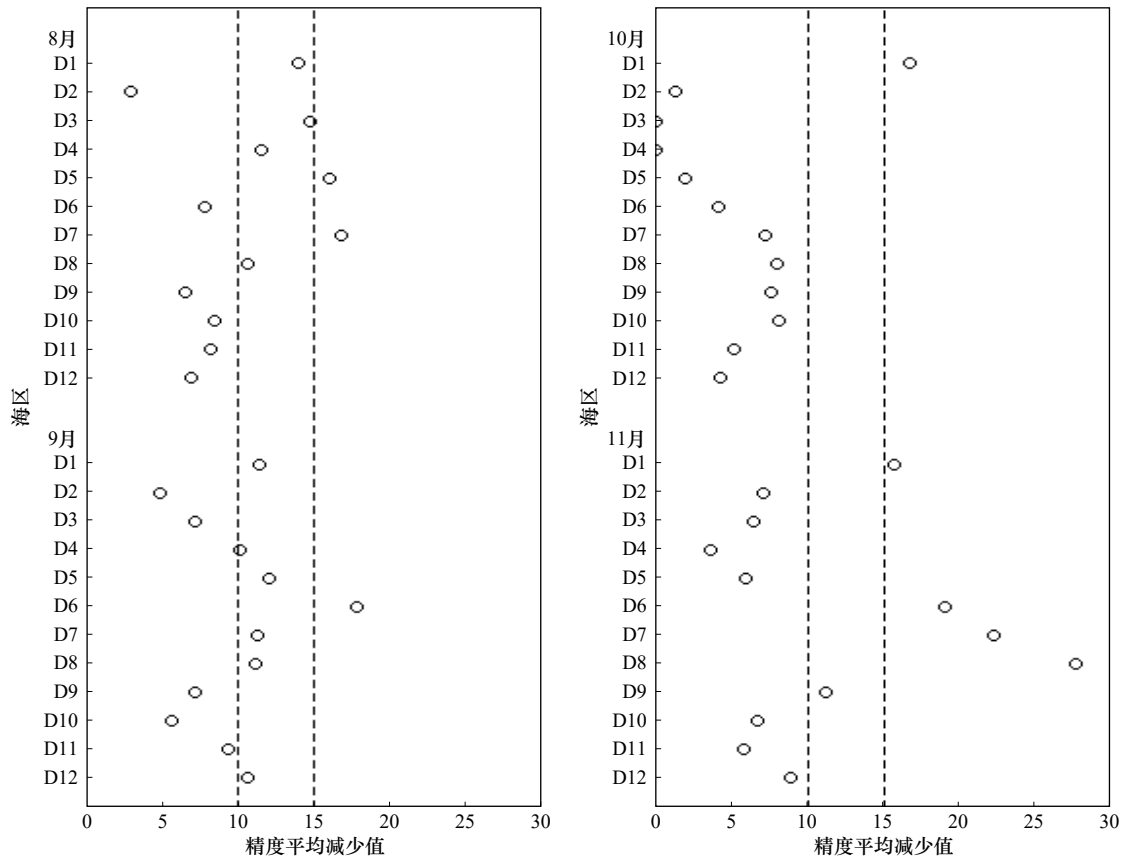


图9 索饵期各海区重要性分布

Fig. 9 The distribution of each sea area's importance during the feeding period

表4 神经网络预报模型构建方案

Table 4 Construct the scenarios of neural network

	海域	编号	输入因子	网络结构
相关性	产卵场海域	I	D9-1、D3-2、D11-2、D13-4	4 : 5 : 1
	索饵场海域	II	D1-10、D6-11、D7-8、D8-11、D8-11、D10-11	6 : 7 : 1
随机森林	产卵场海域	III	D9-1、D10-1、D11-2、D13-4、D16-4	5 : 6 : 1
	索饵场海域	IV	D5-8、D7-8、D6-9、D1-10、D1-11、D6-11、D7-11、D8-11	8 : 9 : 1

明, 该温度范围不是最适温度范围, 随着纬度靠南, 海表水温逐渐升高(图 11), 故最适海表水温范围可能需要作出调整, 而具体最适温度范围有待于后续研究。从4月份来看, B3 和 D13 海域的 Ps 与 CPUE 时间序列的负相关性同样表明最适温度范围不适用于4月份。只结合出现相关性的海域和出现的月份来看, 可预测柔鱼从卵孵化至稚仔鱼的洄游路径: 1月份集中在 D9 海域(25°~30°N, 130°~135°E), 2月份集中在 D3 海区区和 D11 海域(20°~30°N, 140°~145°E), 4月份集中在 D13 海域(25°~30°N, 150°~155°E), 即随着时间的变化, 最适海表水温分布海域发生变化(图 11), 柔鱼幼体从西向东洄游至 150°~155°E 海

域, 后续进入渔场(38°~46°N, 150°~165°E)。

通过索饵场的相关性分析可以看出, 从划分方案 A 至 D, 划分海域不断增多, 各海域面积减小, Pf 与 CPUE 时间序列出现显著性的海域也不断增多。相关性海域出现月份主要集中在 11 月, 海域主要分布在 42°~46°N, 150°~155°E; 38°~42°N, 162.5°~165°E; 42°~46°N, 157.5°~160°E 海域; 8 月和 10 月出现相关性海域极少。从相关性正负来看 8 月和 10 月均为显著负相关, 11 月均为(极)显著正相关。即 8 月的 15~19℃ 和 10 月的 10~13℃ 索饵场最适温度范围所占比值与 CPUE 的关系为显著负相关, 9 月的 14~18℃ 索饵场最适温度范围所占比值与 CPUE 未发现相关

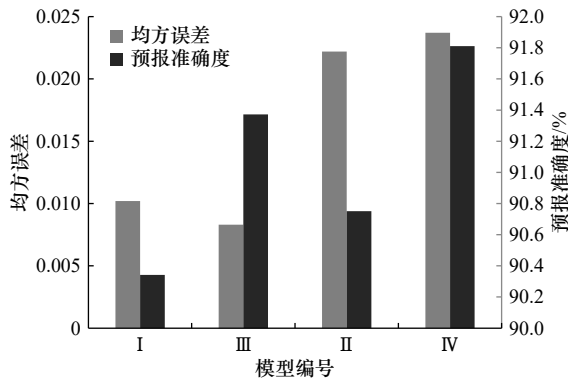


图 10 神经网络模型的模拟结果和预报准确度

Fig. 10 The simulation results and forecast accuracy of neural network

性;故各月份最适温度范围是否准确还需继续验证。因 7 月、8 月和 9 月出现相关性海域较少,很难判断出相关性海域的空间位置随时间的变化。

4.2 随机森林重要性海区分析

随机森林模型得出的产卵场和索饵场重要性海区的分布可以看出,在精度平均减少值 15 作为筛选标准时,产卵场和索饵场重要性海区范围与相关性作为筛选条件所获得的海区基本吻合,故随机森林判定海区重要性结果的可靠性得到验证。若精度平均减少值为 15 所得结果成立,则精度平均减少值 10 作为筛选标准时,重要性海区所增加的范围及月份同样成立。因此,随机森林模型可以探究产卵场与索饵场不同月份各海区最适海表水温比例序列值与 CPUE 序列值之间的潜在关系,从而更加详细的推测柔鱼在产卵期和索饵期洄游路径的变化。根据以上研究结果,结合历年来我国鱿钓船队在西北太平洋的实际作业分布,推测柔鱼的洄游路径,推测结果如图 12 和图 13 所示,其中箭头的方向表示柔鱼在产卵期和索饵期可能存在的洄游路径。以精度平均减少

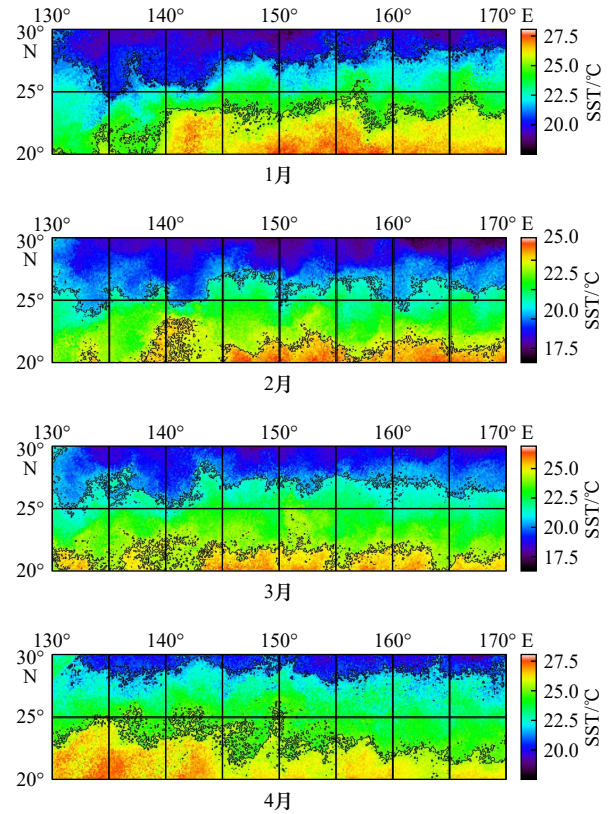


图 11 2015 年产卵场 1-4 月最适海表水温分布情况

Fig. 11 The distribution of the optimal sea surface temperature in the spawning grounds from January to April of 2015

值为 10 作为筛选条件,随机森林结果表明 (图 12),产卵场海区主要集中在两个海域,西部海域主要集中在 140°E 东西海域范围内,随着时间从西向东移动;东部海域主要集中在 160°E 东西海域范围内,随着时间从东向西移动。此结果与 Hayase^[14] 推测的柔鱼产卵场分布在 143°~150.5°E、161°~165.5°E 和 160°~170°E 海域,以及 Young 和 Hirota^[30] 所推断的基本一致。产卵场重要性海区随着时间而出现的空间变化趋势为:西部产卵场 1 月从 20°~30°N, 130°~135°E; 25°~30°N,

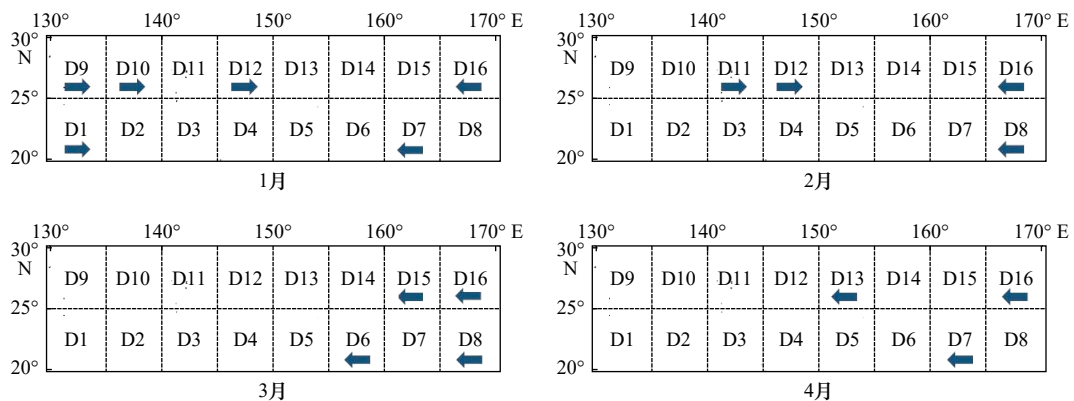


图 12 精度平均减少值为 10 产卵期各月份重要性海区筛选结果

Fig. 12 The results of important sea areas in period of spawning each month under the condition that mean decrease accuracy was 10

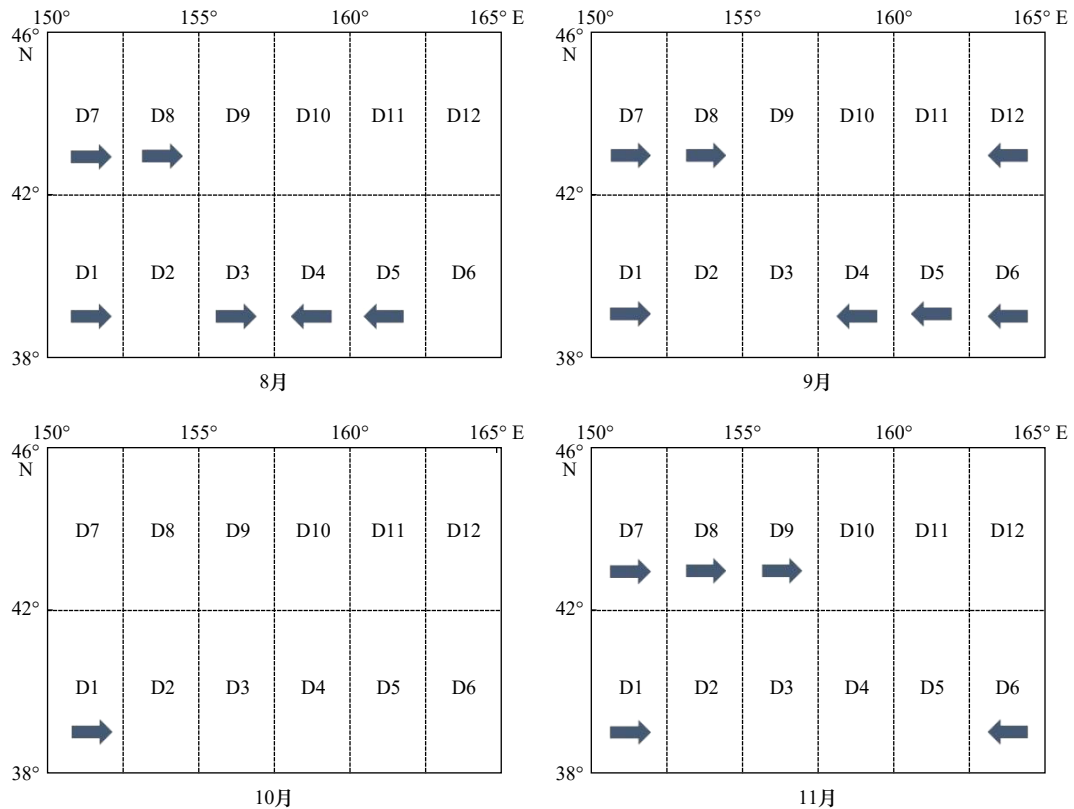


图 13 精度平均减少值为 10 索饵期各月份重要性海区筛选结果

Fig. 13 The results of important sea areas in period of feeding each month under the condition that mean decrease accuracy was 10

135°~140°E 海域, 向东移动至 25°~30°N, 140°~150°E 海域(2月), 至 4月时移动到 25°~30°N, 150°~155°E 海域, 后续进入索饵场时会继续向北移动。东部产卵场 1月和 2月主要集中在 20°~30°N, 165°~170°E 海域, 3月份向西移动至 25°~30°N, 160°~165°E; 20°~25°N, 155°~160°E 海域, 至 4月时移动到 25°~30°N, 150°~155°E 海域, 后续向北移动进入索饵场。此外 160°~170°E 海域仍有柔鱼稚仔鱼分布, 且继续向西洄游。随机森林模型产卵场重要性海区相对于相关性筛选海区, 不仅可以更加详细的展示产卵期的洄游路径的变动, 同时也揭示了东部产卵场的存在, 具有更敏锐的判别能力。

以精度平均减少值为 10 作为筛选条件, 随机森林结果表明(图 13), 索饵场在 8 月份与 9 月份出现的重要性海区, 8 月主要集中在 38°~46°N, 150°~155°E 和 38°~42°N, 155°~162.5°E 海域; 9 月份 38°~46°N, 150°~155°E 和 38°~42°N, 157.5°~165°E。10 月份重要性海区较少, 仅 38°~42°N, 150°~152.5°E 海域, 从 8 月、9 月和 10 月精度平均减少值主要集中在 10~15 之间, 与上述相关性海区较少相验证。近几十年, 随着海洋气候的改变, 柔鱼生存海域环境也发生变化, 各月最适温度范围 8 月为 15~19℃; 9 月为 14~18℃;

10 月为 10~13℃ 可能需要做出微调来适应最新的气候变化所带来的误差。11 月份主要集中在 42°~46°N, 150°~157.5°E 和 38°~42°N, 162.5°~165°E 海域。故索饵场海区各月份主要集中在 42°~46°N, 150°~157.5°E 海域, 部分分布在 38°~46°N, 160°~165°E 海域。

4.3 模型评价

神经网络预报模型结果表明, 基于相关性与随机森林筛选出的产卵场和索饵场海区构建的预报模型可以准确预报柔鱼资源丰度, 可验证两种方法下, 所筛选海区的正确性。通过随机森林模型筛选重要性海区与相关性条件下筛选出的海区比较, 可以看出, 所属海区分布范围大致相同, 部分海区有差异(增加或减少), 基于随机森林构建的神经网络模型的预报准确度优于相关性, 且基于产卵场和索饵场构建的模型预报准确度均有提高。表明随机森林筛选出的重要性海区的的海表水温范围占该海区面积的比值与资源丰度的关系更紧密。从各模型的均方误差来看, 基于索饵场海区构建预报模型的均方误差明显大于产卵场海区构建的模型, 而 MSE 的值越小, 说明预测模型的精确度越高, 且更加稳定。由于北太平洋柔鱼其独特的生物学特性, 短生命周期形成其特有的资源补充模式, 即种群资源量的大小主要取决于补充

量,补充量又主要取决于早期生活史阶段柔鱼幼体的死亡率。从卵孵化至作为补充量进入渔场前,柔鱼个体洄游能力,自主选择合适的生存海域能力较弱,可以看作是被动洄游,极易受到海洋外界因素的影响。进入索饵期后,柔鱼个体生长,游动能力增加,自主选

择合适海域的能力加强,受制于外界环境因素的程度减弱。故基于索饵场最适海表水温的比值 Pf 构建的预测模型精确度较低,模型不稳定。因此,产卵场最适海表水温范围所占的比值 Ps 作为输入变量构建模型更为可靠。

参考文献:

- [1] 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,2005:152-153.
Wang Yaogeng, Chen Xinjun. The resource and biology of economic oceanic squid in the world[M]. Beijing: China Ocean Press, 2005: 152-153.
- [2] 陈新军,田思泉,许柳雄.西北太平洋海域柔鱼产卵场和作业渔场的水温年间比较及与资源丰度的关系[J].上海水产大学学报,2005,14(2):168-175.
Chen Xinjun, Tian Siqun, Xu Liuxiong. Analysis on changes of surface water temperature in the spawning and feeding ground of *Ommastrephes bartramii* and its relationship with abundance index in the Northwestern Pacific Ocean[J]. Journal of Shanghai Fisheries University, 2005, 14(2): 168-175.
- [3] Chen Xinjun, Chen Yong, Tian Siqun, et al. An assessment of the west winter-spring cohort of neon flying squid (*Ommastrephes bartramii*) in the Northwest Pacific Ocean[J]. Fisheries Research, 2008, 92(2/3): 221-230.
- [4] Rodhouse P G. Managing and forecasting squid fisheries in variable environments[J]. Fisheries Research, 2001, 54(1): 3-8.
- [5] Anderson C I H, Rodhouse P G. Life cycles, oceanography and variability: Ommastrephid squid in variable oceanographic environments[J]. Fisheries Research, 2001, 54(1): 133-143.
- [6] 沈新强,王云龙,袁骥,等.北太平洋鱿鱼渔场叶绿素a分布特点及其与渔场的关系[J].海洋学报,2004,26(6):118-123.
Shen Xinqiang, Wang Yunlong, Yuan Qi, et al. Distributional characteristics of chlorophyll a and relation to the fishing ground in the squid fishing ground of the northern Pacific Ocean[J]. Haiyang Xuebao, 2004, 26(6): 118-123.
- [7] Boyle P R, Pierce G J, Hastie L C. Flexible reproductive strategies in the squid *Loligo forbesi*[J]. Marine Biology, 1995, 121(3): 501-508.
- [8] 曹杰,陈新军,刘必林,等.鱿鱼类资源量变化与海洋环境关系的研究进展[J].上海海洋大学学报,2010,19(2):232-239.
Cao Jie, Chen Xinjun, Liu Bilin, et al. Review on the relationship between stock recruitment of squid and oceanographic environment[J]. Journal of Shanghai Oceans University, 2010, 19(2): 232-239.
- [9] Payne A G, Agnew D J, Pierce G J. Trends and assessment of cephalopod fisheries[J]. Fisheries Research, 2006, 78(1): 1-3.
- [10] 余为,陈新军,易倩,等.北太平洋柔鱼早期生活史研究进展[J].上海海洋大学学报,2013,22(5):755-762.
Yu Wei, Chen Xinjun, Yi Qian, et al. Review on the early life history of neon flying squid *Ommastrephes bartramii* in the North Pacific[J]. Journal of Shanghai Ocean University, 2013, 22(5): 755-762.
- [11] Murakami K, Watanabe Y, Nakara J. Growth, Distribution and Migration of Flying Squid (*Ommastrephes bartramii*) in the North Pacific[M]. Research Institute of North Pacific Fisheries, Hokkaido University, 1981: 161-179.
- [12] Nakamura Y. Distribution and maturity of neon flying squid, *Ommastrephes bartramii* (Lesueur), in the surrounding waters of Izu-Ogasawara Islands in spring[J]. Bulletin of Hokkaido National Fisheries Research Institute, 1988, 52: 139-149.
- [13] Okutain T. Systematics and distribution of larvae of Decapod Cephalopods collected from the sea surface on the Pacific Coast of Japan, 1960-1965[J]. Bulletin of the Tokai Regional Fisheries Research Laboratory, 1968, 55(2): 9-57.
- [14] Hayase S. Distribution of Spawning Grounds of Flying Squid, *Ommastrephes bartramii*, in the North Pacific Ocean[J]. Japan International Research Center for Agricultural Science, 1995, 29(1): 65-72.
- [15] Cao Jie, Chen Xinjun, Chen Yong. Influence of surface oceanographic variability on abundance of the western winter spring cohort of neon flying squid *Ommastrephes bartramii* in the NW Pacific Ocean[J]. Marine Ecology Progress Series, 2009, 381(5): 119-127.
- [16] Bower J R. Estimated paralarval drift and inferred hatching sites for *Ommastrephes bartramii* (Cephalopoda: Ommastrephidae) near the Hawaiian Archipelago[J]. Fishery Bulletin, 1996, 94: 398-411.
- [17] Bower J R, Ichii T. The red flying squid (*Ommastrephes bartramii*): A review of recent research and the fishery in Japan[J]. Fisheries Research, 2005, 76(1): 39-55.
- [18] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [19] Gall J, Yao A, Razavi N, et al. Hough forests for object detection, tracking, and action Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2188-2202.
- [20] 曹正凤.随机森林算法优化研究[D].北京:首都经济贸易大学,2014.
Cao Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing: Capital University of Economics and Business, 2014.
- [21] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al. A review of technologies on random forests[J]. Statistics & Information Forum, 2011, 26(3): 32-38.
- [22] 李磊.随机森林及数据可视化在棉蚜等级预测中的应用研究[D].泰安:山东农业大学,2017.

- Li Lei. Application of random forest and data visualization in prediction of cotton aphid grade[D]. Tai'an: Shandong Agricultural University, 2017.
- [23] 张万军. 基于大数据的个人信用风险评估模型研究[D]. 北京: 对外经济贸易大学, 2016.
Zhang Wanjun. Research on individual credit risk evaluation model based on big data[D]. Beijing: University of International Business and Economics, 2016.
- [24] 孟杰, 李春林. 基于随机森林模型分类数据缺失值插补[J]. 统计与信息论坛, 2016, 29(9): 86–90.
Meng Jie, Li Chunlin. Missing data imputation for categorical data based on random forest model[J]. Statistics & Information Forum, 2016, 29(9): 86–90.
- [25] 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应用昆虫学报, 2013, 50(4): 1190–1197.
Li Xinhai. Using “random forest” for classification and regression[J]. Chinese Journal of Applied Entomology, 2013, 50(4): 1190–1197.
- [26] Benediktsson J A, Swain P H, Ersoy O K. Neural network approaches versus statistical methods in classification of multisource remote sensing data[J]. IEEE Transactions on Geoscience and Remote Sensing, 1990, 28(4): 540–552.
- [27] Hush D R, Horne B G. Progress in supervised neural networks[J]. IEEE Signal Processing Magazine, 1993, 10(1): 8–39.
- [28] 汪金涛, 高峰, 雷林, 等. 基于产卵场环境因子的阿根廷滑柔鱼资源补充量预报模型研究[J]. 海洋学报, 2014, 36(12): 119–124.
Wang Jintao, Gao Feng, Lei Lin, et al. Study on forecasting model of recruitment for *Illex argentinus* by using the environmental factors in the spawning ground[J]. Haiyang Xuebao, 2014, 36(12): 119–124.
- [29] 金龙, 况雪源, 黄海洪, 等. 人工神经网络预报模型的过拟合研究[J]. 气象学报, 2004, 62(1): 62–70.
Jin Long, Kuang Xueyuan, Huang Haihong, et al. Study on the overfitting of the artificial neural network forecasting model[J]. Acta Meteorologica Sinica, 2004, 62(1): 62–70.
- [30] Young R E, Hirota J. Description of *Ommastrephes bartramii* (Cephalopoda: Ommastrephidae) paralarvae with evidence for spawning in Hawaiian waters[J]. Pacific Science, 1990, 44(1): 71–80.

Forecasting Northwest Pacific Ocean neon flying squid abundance based on suitability of spawning and feeding grounds

Wei Guang'en¹, Chen Xinjun^{1,2,3}

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory for Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China; 3. National Engineering Research Center for Oceanic Fisheries, Shanghai 201306, China)

Abstract: Neon flying squid *Ommastrephes bartramii* is a cephalopod species of economically importance distributed in the North Pacific Ocean. Because of short lifespan, their abundances mainly depend on the recruitments and the marine environment during their early life stages will directly affect the recruitments. Using fishery data collected by Chinese squid-jigging fleets in the Northwest Pacific Ocean from 2004 to 2015 and the sea surface temperature (SST) of spawning and feeding grounds, which were divided into different numbers of subareas, correlation analysis and random forest model were used to screen out the subareas that CPUE has significant relationships with Ps (the proportion of favorable-SST in spawning grounds) of spawning grounds and Pf (the proportion of favorable-SST in feeding grounds) of feeding grounds during spawning and feeding periods. The Ps and Pf were used as input variables of the neural network model to forecast recruitments based on spawning ground and feeding ground, respectively, and the advantages and disadvantages of the model and forecast accuracy were analyzed. The results show that the schemes of dividing spawning ground into $5^{\circ} \times 5^{\circ}$ and feeding field into $2.5^{\circ}(\text{longitude}) \times 4^{\circ}(\text{latitude})$ were optimal. The ranges of subareas selected by random forests are largely consistent with those selected by correlation analysis, both random forests and correlation analysis can identify potential subareas associated with CPUE, and both have forecasting accuracy of >90%. However, the subareas selected by the random forest are better and the forecast accuracy is higher than those selected by correlation analysis. In addition, the model based on the spawning ground is more accurate and stable than that based on feeding ground.

Key words: Northwest Pacific Ocean; *Ommastrephes bartramii*; abundance; random forests; neural network