

## 渔情预报技术及模型研究进展

陈新军<sup>1,2</sup>, 高峰<sup>1,2\*</sup>, 官文江<sup>1,2</sup>, 雷林<sup>1,2</sup>, 汪金涛<sup>1,2</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. 上海海洋大学国家远洋渔业工程技术研究中心, 上海 201306)

**摘要:** 渔情预报是渔场学的重要研究内容,对渔业科学生产和渔业资源管理具有重要的意义。近年来,随着现代统计理论、数值计算方法、数据挖掘和人工智能等理论和技术的发展,使渔情预报技术和模型的发展焕发出了新的活力,为此本文对渔情预报技术及模型研究进展进行了回顾,并对今后发展提出了展望。本文简要概述了渔情预报建模的理论和方法,包括渔情预报相关的渔场学基础、数据模型和预报模型,重点介绍了基于统计和机器学习、人工智能方法的渔情预报模型的应用和研究现状,并对各种模型在渔情预报应用中的优势与缺陷进行了综合分析,针对存在的问题提出了建议。主要建议包括:建立专为渔业服务的海洋环境预报系统;进行长期系统的渔业资源调查,针对不同鱼种和海区对数据获取和处理方法进行标准化和规范化;借助随机模拟方法降低模型不确定性,提高预报精度。本总结与分析将为国内的渔情预报模型研究工作提供参考。

**关键词:** 渔情预报; 预报模型; 统计学; 机器学习

**中图分类号:** S 934

**文献标志码:** A

渔情预报是渔场学研究的重点<sup>[1]</sup>。准确的渔情预报可以指导企业合理安排渔业生产,缩短寻找渔场的时间,减少成本、提高渔获产量。我国于 20 世纪 50 年代开始对近海主要经济鱼种进行渔情预报工作,积累了丰富的经验。20 世纪 80 年代以来,地理信息系统的发展为渔情分析和渔场预报研究提供了强大的分析工具。海洋卫星遥感技术的普及使得人们能快速获取大范围内的海况信息,实时船位监控及海事卫星通讯技术使得远洋渔船能有效地接收渔情预报机构的实时预报。借助这些技术和手段,国内外多个机构已对部分远洋鱼种进行了渔情预报的业务化运行<sup>[2-3]</sup>。

随着我国远洋渔业的发展,其生产规模在不断扩大扩大的同时,生产成本也不断升高,远洋渔业企业对渔情预报准确性的要求也越来越高,这也为

渔情预报技术和预报模型的研究提出了新的挑战。近年来,随着现代统计理论、数值计算方法、数据挖掘和人工智能等理论和技术的发展,使传统的统计预报模型焕发出了新的活力,各类机器学习方法也为渔情预报模型的开发提供了新的思路。本文在分析渔情预报理论和方法的基础上,对国内外渔情预报技术及其预报模型进行了概括和总结,以为今后的研究工作提供参考。

### 1 渔情预报的理论和方法

#### 1.1 渔情预报

渔情预报是对未来一定时期和一定水域内水产资源状况各要素,如渔期、渔场、鱼群数量和质量以及可能达到的渔获量等所做出的预报。按照预报内容的不同,可将渔情预报分为三种类型,即关于资源状况的预报、关于时间的预报和关于空

收稿日期:2012-08-29 修回日期:2013-03-31

资助项目:国家“八六三”高技术研究发展计划(2012AA092303);国家科技支撑计划(2013BAD13B01);国家发改委产业化专项(2159999);上海市科技创新行动计划(12231203900)

通信作者:高峰,E-mail:gaofeng@shou.edu.cn

间的预报<sup>[1]</sup>。每种预报的侧重点不同,相应的预报原理和模型也不同。

关于资源状况的预报,即预报鱼群的数量、质量以及在一定捕捞条件下的渔获量,这种预报主要是中长期的。准确的中长期预报对于渔业管理和生产都具有重要意义,不但渔业管理部门可以将预报结果作为制订渔业政策的参考信息,渔业生产企业也可以根据这些预报合理安排捕捞努力量,适应负责任捕捞和渔业可持续发展的要求。目前,关于渔业资源状况的预报模型主要以鱼类种群动力学为基础,数学上则主要采用统计回归、人工神经网络和时间序列分析等方法<sup>[4-9]</sup>。

关于时间的预报主要包括预报渔期出现的时间和持续的时间等。这类预报不但要求预报者对目标鱼类的洄游和集群状况有较深入的了解,而且需要采用有效的观测手段,实时地了解目标区域的天气、海流、水温结构以及饵料生物情况,结合渔民和渔业研究者的经验来进行预报。随着国内渔业生产模式的改变,渔情预报研究者已从渔业生产一线脱离,因此目前这类预报主要以有经验的渔业生产者的现场定性分析为主,其原理很难进行明确的量化解,已有的定量研究一般也仅采用简单的线性回归<sup>[10-11]</sup>。

关于空间的预报,即预报渔场出现的位置或鱼类资源的空间分布状况,即通常所说的渔场预报。由于渔业资源的逐渐匮乏以及燃油、入渔等成本的不断升高,渔业生产过程中渔场位置的预报变得越来越重要,企业对其实时性、准确性的要求也越来越高。因此渔场位置的预报模型研究相当活跃,国内外大多数渔情预报模型都是渔场的位置预报模型<sup>[12]</sup>。

## 1.2 渔情预报模型的组成

从生物分布预测的角度来看,一个合理的渔情预报模型应考虑三个方面的内容,即渔场学基础、数据模型和预报模型<sup>[13]</sup>。其中,渔场学部分主要包括鱼类的集群及洄游规律、环境条件对鱼类行为的影响以及短期和长期的环境事件对渔业资源的影响。数据模型部分主要包括渔业数据和环境数据的收集、处理和应用的方法以及这些方法对预报模型的影响。预报模型部分则主要包括建立渔情预报模型的理论基础和方法以及相应的模型参数估计、优化及验证,以及其不确定性分析<sup>[14]</sup>。

**渔场学基础** 鱼类在海洋中的分布是由其自身生物学特性和外界环境条件共同决定的<sup>[15]</sup>。首先,海洋鱼类一般都有集群和洄游和习性,其集群和洄游的规律决定了渔业资源在时间和空间的大体分布。其次,鱼类的行为与其生活的外界环境有密切的关系。鱼类生存的外界环境包括生物因素和非生物因素两类。生物因素包括敌害生物、饵料生物、种群关系。非生物因素包括水温、海流、盐度、光、溶解氧、气象条件、海底地形和水质因素等。最后,各类突发或阶段性、甚至长期缓慢的海洋环境事件,如赤潮、溢油、环境污染、厄尔尼诺现象、全球气候变暖,对渔业资源也会产生短期和长期的影响,进而引起渔业资源在时间、空间、数量和质量上的振荡。只有综合考虑这三方面因素的影响,才能建立起合理的渔情预报模型<sup>[1]</sup>。

**数据模型** 渔场预报研究所需要的数据主要包括渔业数据和海洋环境数据两类<sup>[16]</sup>,这些数据的收集、处理和应用的策略对渔情预报模型具有重要影响<sup>[17]</sup>。在构建渔情预报模型时,为了统一渔业数据和环境数据的时间和空间分辨率,一般需要对数据进行重采样。由于商业捕捞的作业地点不具备随机性,空间和时间上的合并处理将使模型产生不同的偏差;与渔场形成关系密切的涡流和峰面等海洋现象具有较强的变化性,海洋环境数据在空间和时间尺度上的平均将会弱化甚至掩盖这些现象。因此在构建渔情预报模型时应选择合适的时空分辨率,以降低模型偏差、提高预测精度。另外,渔情预报模型的构建也应充分考虑渔业数据本身的特殊性,如渔业数据都是一种类似“仅包含发现”(presence-only)的数据,即重视记录有渔获量的地点,而对于无渔获量的地点的记录并不重视。最后,低分辨率的历史数据<sup>[18]</sup>、空间位置信息<sup>[19]</sup>等数据的应用也应选择合适的策略。

**预报模型** 渔情预报模型主要可分为三种类型,即经验/现象模型、机理/过程模型和理论模型<sup>[14]</sup>。总的来说,现有的渔情预报模型还是以经验/现象模型为主。这类模型常见的开发思路有两种:一种以生态位(ecological niche)或资源选择函数(resource selection function, RSF)为理论基础,主要通过频率分析和回归等统计学方法分析出目标鱼种的生态位或者对于关键环境因子的

响应函数,从而建立渔情预报模型<sup>[10-11]</sup>。另一种是知识发现的思路,即以渔业数据和海洋环境数据为基础,通过各类机器学习和人工智能方法在数据中发现渔场形成的规律,建立渔情预报模型<sup>[12,20]</sup>。

总的来说,基于统计学的渔情预报模型以回归为中心,其模型结构是预先设定好的,主要通过已有数据估计出模型系数,然后用这些模型进行渔场预测,可以称之为“模型驱动”(model-driven)的模型。而基于机器学习和人工智能方法的预测模型则以模型的学习为中心,主要通过各种数据挖掘方法从数据中提取渔场形成的规则,然后使用这些规则进行渔场预报,是“数据驱动”(data-driven)的模型。近几十年来,传统统计学和计算方法都发生了很大的变化,统计学方法和机器学习方法之间的区别也已变得模糊<sup>[14]</sup>。

### 1.3 渔情预报模型的构建

借鉴 Guisan 等<sup>[21]</sup>关于生物分布预测模型的研究,可将建立渔情预报模型的过程分为四个步骤:(1)研究渔场形成机制;(2)建立渔情预报模型;(3)模型校正;(4)模型评价和改进。

渔情预报模型的构建应以目标鱼种的生物学和渔场学研究为基础,力求模型与渔场学实际的吻合<sup>[13]</sup>。如果对目标鱼种的集群、洄游特性以及渔场形成机制较清楚,可选择使用机理/过程模型或理论模型对这些特性和机制进行定量表述。反之,如果对这些特性和机制的了解并不完全,则可选择经验/现象模型,根据基本的生态学原理对渔场形成过程进行一种平均化的描述。除此之外,无论构建何种预测模型,都应充分考虑模型所使用的数据本身的特点,这对于基于统计学的模型尤其重要<sup>[22]</sup>。

模型校正(model calibration)是指建立预报模型方程之后,对于模型参数的估值以及模型的调整<sup>[23]</sup>。根据预报模型的不同,模型参数估值的方法也不一样。例如对于各类统计学模型,其参数主要采用最小方差或极大似然估计等方法进行估算;而对于人工神经网络模型,权重系数则通过模型迭代计算至收敛而得到。在渔情预报模型中,除了估计和调整模型参数和常数之外,模型校正还包括对自变量的选择。在利用海洋环境要素进行渔情预报时,选择哪些环境因子是一项比较重要也非常困难的工作。周彬彬<sup>[24]</sup>在利用回归

模型进行蓝点马鲛渔期预报研究时认为,多因子组合的预报比单因子预报要准确。Harrell 等<sup>[25]</sup>研究表明,为了增加预测模型的准确度,自变量的个数不宜太多。另外,对于某些模型来说,模型校正还包括自变量的变换、平滑函数的选择等工作<sup>[26]</sup>。

模型评价(model evaluation)主要是对于预测模型的性能和实际效果的评价。模型评价的方法主要有两种,一种是模型评价和模型校正使用相同的数据,采用变异系数法或自助法评价模型<sup>[27-28]</sup>;另一种方法则是采用全新的数据进行模型评价,评价的标准一般是模型拟合程度或者某种距离参数<sup>[29]</sup>。由于渔情预报模型的主要目的是预报,其模型评价一般采用后一种方法,即考查预测渔情与实际渔情的符合程度。

## 2 渔场预报模型

### 2.1 统计学模型

线性回归模型 早期或传统的渔情预报主要采用以经典统计学为主的回归分析、相关分析、判别分析和聚类分析等方法<sup>[12]</sup>。其中最具有代表性的是一般线性回归模型。通过分析海表面温度(sea surface temperature, SST)、叶绿素 a(*Chl. a*)浓度等海洋环境数据与历史渔获量、单位捕捞努力渔获量(catch per unit effort, CPUE)或者渔期之间的关系,建立回归方程:

$$CPUE = \beta_0 + \beta_1 \cdot SST + \beta_2 \cdot chl + \dots + \varepsilon \quad (1)$$

式(1)中; $\beta_i$ 为回归系数, $\varepsilon$ 为误差项。一般线性回归模型采用最小二乘法对系数 $\beta_i$ 进行估计,然后利用这些方程对渔期、渔获量或 CPUE 进行预报。如陈新军<sup>[30]</sup>认为,北太平洋柔鱼日渔获量 CPUE (kg/d)与 0~50 m 水温差  $\Delta T$ ( $^{\circ}\text{C}$ )具有线性关系,可以建立预报方程  $CPUE = -880 + 365\Delta T$ 。

一般线性模型结构稳定,操作方法简单,在早期的实际应用中取得了一定的效果<sup>[10-12]</sup>。但一般线性模型方差小、偏差大,用于预报时存在一定的局限性。一方面,渔场形成与海洋环境要素之间的关系具有模糊性和随机性,一般很难建立相关系数很高的回归方程。另一方面,实际的渔业生产和海洋环境数据一般并不满足一般线性模型对于数据的假设,因而导致回归方程预测效果较差<sup>[14]</sup>。目前,一般线性回归模型在渔情预报中的应用已比较少见,而逐渐被更为复杂的分段线性

回归<sup>[31]</sup>、多项式回归<sup>[32]</sup>和指数(对数)回归<sup>[33-34]</sup>、分位数回归<sup>[35]</sup>等模型所取代。

**广义回归模型** 广义线性模型 (generalized linear model, GLM) 通过连接函数  $g$  对响应变量进行一定的变换,将基于指数分布族的回归与一般线性回归整合起来,其回归方程如下:

$$g[E(Y)] = \beta_0 + \sum_{i=1}^p \beta_i \cdot X_i + \varepsilon \quad (2)$$

GLM 模型可对自变量本身进行变换,也可加上反映自变量相互关系的函数项,从而以线性的形式实现非线性回归。自变量的变换包括多种形式,如多项式形式的 GLM 模型,方程如下:

$$g[E(Y)] = LP = \beta_0 + \sum_{i=1}^p \beta_i \cdot (X_i)^p + \varepsilon \quad (3)$$

**广义加性模型** (generalized additive model, GAM) 是 GLM 模型的非参数扩展。其方程式如下:

$$g[E(Y)] = LP = \beta_0 + \sum_{i=1}^p f_i \cdot X_i + \varepsilon \quad (4)$$

GLM 模型中的回归系数  $\beta_i$  被平滑函数局部散点平滑函数  $f_i$  所取代。与 GLM 模型相比, GAM 更适合处理非线性问题<sup>[36]</sup>。

自上世纪 80 年代开始, GLM 和 GAM 模型相继应用于渔业资源研究中。特别是在 CPUE 标准化研究中,这两种模型都获得了较大的成功<sup>[37-38]</sup>。在渔业资源的空间分布预测方面, GLM 和 GAM 也有广泛的应用<sup>[36,39-41]</sup>。如 Chang 等<sup>[40]</sup>利用两阶段 GAM (2-stage GAM) 模型研究了缅因湾美国龙虾的分布规律。但在渔情分析和预报应用上,国内研究者主要还是将其作为分析模型而非预报模型。如牛明香等<sup>[42]</sup>在研究东南太平洋智利竹筴鱼中心渔场预报时,使用 GAM 作为预测因子选择模型。GLM 和 GAM 模型能在一定程度上处理非线性问题,因此具有较好的预测精度。但它们的应用较为复杂,需要研究者对渔业生产数据中的误差分布和预测变量的变换具有较深的认识<sup>[43]</sup>,否则极易对预测结果产生影响。

**贝叶斯方法** 贝叶斯统计理论基于贝叶斯定理,即通过先验概率以及相应的条件概率计算后验概率。其中先验概率是指渔场形成的总概率,条件概率是指渔场为“真”时环境要素满足某种条件的概率,后验概率即当前环境要素条件下

渔场形成的概率。贝叶斯方法通过对历史数据的频率统计得到先验概率和条件概率,计算出后验概率之后,以类似查表的方式完成预报。已有的研究表明,贝叶斯方法具有不错的预报准确率。如樊伟等<sup>[44]</sup>对 1960—2000 年西太平洋金枪鱼渔业和环境数据进行了分析,采用贝叶斯统计方法建立了渔情预报模型,综合预报准确率达到 77.3%。

贝叶斯方法的一个显著优点是其易于集成的特性,几乎可以与任何现有的模型集成在一起应用,常用的方法就是以不同的模型计算和修正先验概率<sup>[45-47]</sup>。目前渔情预报应用中的贝叶斯模型采用的都是朴素贝叶斯分类器 (simple Bayesian classifier),该方法假定环境条件对渔场形成的影响是相互独立的,这一假定显然并不符合渔场学实际。相信考虑各预测变量联合概率的贝叶斯信念网络 (Bayesian belief network) 模型<sup>[48]</sup>在渔情预报方面也应该会有较大的应用空间。

**时间序列分析** 时间序列 (time series) 是指具有时间顺序的一组数值序列<sup>[49]</sup>。对于时间序列的处理和分析具有静态统计处理方法无可比拟的优势,随着计算机以及数值计算方法的发展,已经形成了一套完整的分析和预测方法。时间序列分析在渔情预报中主要应用在渔获量预测方面<sup>[4-9,50]</sup>。如 Grant 等<sup>[4]</sup>利用时间序列分析模型对墨西哥湾西北部的褐虾商业捕捞年产量进行了预测。Georgakarakos 等<sup>[6]</sup>分别采用时间序列分析、人工神经网络和贝叶斯动态模型对希腊海域枪乌贼科和柔鱼科产量进行了预测,结果表明时间序列分析方法具有很高的精度。

**空间分析和插值** 空间分析的基础是地理实体的空间自相关性,即距离越近的地理实体相似度越高<sup>[51]</sup>,距离越远的地理实体差异性越大。空间自相关性被称为“地理学第一定律” (first law of geography),生态学现象也满足这一规律。空间分析主要用来分析渔业资源在时空分布上的相关性和异质性,如渔场重心的变动、渔业资源的时空分布模式等<sup>[52-55]</sup>。但也有部分学者使用基于地统计学的插值方法(如克里金插值法)对渔获量数据进行插值,在此基础上对渔业资源总量或空间分布进行估计<sup>[52,56-57]</sup>。如 Monestieja 等<sup>[57]</sup>使用地统计学方法对地中海西北部长须鲸

的空间分布进行了预测。需要说明的是,渔业具有非常强的动态变化特征,而地统计学方法从本质上来讲是一种静态方法<sup>[52]</sup>,因此对渔业数据的收集方法具有严格的要求。

## 2.2 机器学习和人工智能方法

关于空间的渔场预测也可以看成是一种“分类”,即将空间中的每一个网格分成“渔场”和“非渔场”的过程。这种分类过程一般是一种监督分类(supervised classification),即通过不同的方法从样本数据中提取出渔场形成规则,然后使用这些规则对实际的数据进行分类,将海域中的每个网格点分成“渔场”和“非渔场”两种类型。提取分类规则的方法有很多,一般都属于机器学习方法。机器学习是研究计算机怎样模拟或实现人类的学习行为,以获取新的知识的方法。机器学习和人工智能、数据挖掘的内涵有相同之处且各有侧重<sup>[58]</sup>,这里不作详细阐述。机器学习和人工智能方法众多,目前在渔情预报方面应用最多的是人工神经网络、基于规则的专家系统和范例推理方法。除此之外,决策树、遗传算法、最大熵值法、元胞自动机、支持向量机、分类器聚合、关联分析和聚类分析、模糊推理等方法都开始在渔情分析和预报中有所应用<sup>[12,59]</sup>。

**人工神经网络模型** 人工神经网络(artificial neural networks, ANN)模型是模拟生物神经系统而产生的。它由一组相互连接的结点和有向链组成<sup>[58]</sup>。人工神经网络的主要参数是连接各结点的权值,这些权值一般通过样本数据的迭代计算至收敛得到,收敛的原则是最小化误差平方和。确定神经网络权值的过程称为神经网络的学习过程。结构复杂的神经网络学习非常耗时,但预测时速度很快。人工神经网络模型可以模拟非常复杂的非线性过程,在海洋和水产学科已经得到广泛应用<sup>[60-61]</sup>。在渔情预报应用中,人工神经网络模型在空间分布预测和产量预测方面都有成功应用<sup>[62-67]</sup>。

人工神经网络方法并不要求渔业数据满足任何假设,也不需要分析鱼类对于环境条件的响应函数和各环境条件之间的相互关系,因此应用起来较为方便,在应用效果上与其它模型相比也没有显著的差异。但人工神经网络类型很多,结构多变,相对其它模型来说应用比较困难,要求建模者具有丰富的经验<sup>[68]</sup>。另外 ANN 模型对于知识

的表达是隐式的,相当于一种黑盒(black box)模型,这一方面使得 ANN 模型在高维情况下表现尚可<sup>[69]</sup>,一方面也使得 ANN 模型无法对预测原理做出明确的解释。当然目前也已经有方法检验 ANN 模型中单个输入变量对模型输出贡献度<sup>[70]</sup>。

**基于规则的专家系统** 专家系统是一种智能计算机程序系统,它包含特定领域人类专家的知识 and 经验,并能利用人类专家解决问题的方法来处理该领域的复杂问题<sup>[71]</sup>。在渔情预报应用中,这些专家知识和经验一般表现为渔场形成的规则。目前渔情预报中最常见的专家系统为环境阈值法和栖息地适宜性指数模型。

**环境阈值法 (environmental envelope methods)** 是最早也是应用最广泛的渔情空间预报模型之一。鱼类对于环境要素都有一个适宜的范围,环境阈值法假设鱼群在适宜的环境条件出现而当环境条件不适宜时则不会出现。这种模型在实现时,通常先计算出满足单个环境条件的网格,然后对不同环境条件的计算结果进行空间叠加分析,得到最终的预测结果,因此也常被称为空间叠加法。空间叠加法能够充分利用渔业领域的专家知识,而且模型构造简单,易于实现,特别适用于海洋遥感反演得到的环境网格数据,因此在渔情预报领域得到了相当广泛的应用<sup>[72-74]</sup>。

**栖息地适宜性指数 (habitat suitability index, HSI)** 模型是由美国地理调查局国家湿地研究中心鱼类与野生生物署提出的用于描述鱼类和野生动物的栖息地质量的框架模型<sup>[75]</sup>。其基本思想和实现方法与环境阈值法相似,但也有一些区别:首先,HSI 模型的预测结果是一个类似于“渔场概率”的栖息地适应性指数,而不是环境阈值法的“是渔场”和“非渔场”的二值结果;其次,在 HSI 模型中,鱼类对于单个环境要素的适应性不是用一个绝对的数值范围描述,而是采用资源选择函数来表示;最后,在描述多个环境因子的综合作用时,HSI 模型可以使用连乘、几何平均、算术平均、混合算法等多种表示方式<sup>[76]</sup>。HSI 模型在鱼类栖息地分析和渔情预报上已有大量应用<sup>[31,34-35,76]</sup>。但栖息地适应性指数作为一个平均化的指标,与实时渔场并不具有严格的相关性,因此在利用 HSI 模型来预测渔场时需要非常地谨慎<sup>[31]</sup>。

范例推理 范例推理 (case-based reasoning, CBR) 模拟人们解决问题的一种方式, 即当遇到一个新问题的时候, 先对该问题进行分析, 在记忆中找到一个与该问题类似的范例, 然后将该范例有关的信息和知识稍加修改, 用以解决新的问题。在范例推理过程中, 面临的新问题称为目标范例, 记忆中的范例称为源范例。范例推理就是由目标范例的提示, 而获得记忆中的源范例, 并由源范例来指导目标范例求解的一种策略<sup>[77]</sup>。这种方法简化了知识获取, 通过知识直接复用的方式提高解决问题的效率, 解决方法的质量较高, 适用于非计算推导, 在渔场预报方面有广泛的应用<sup>[78-81]</sup>。范例推理方法原理简单, 并且其模型表现为渔场规则的形式, 因此可以很容易地应用到专家系统中。但范例推理方法需要足够多的样本数据以建立范例库, 而且提取出的范例主要还是历史数据的总结, 难以对新的渔场进行预测<sup>[12]</sup>。

### 2.3 机理/过程模型和理论模型

前面提到的两类模型都属于经验/现象模型。经验/现象模型是静态、平均化的模型, 它假设鱼类行为与外界环境之间具有某种均衡。与经验/现象模型不同, 机理/过程模型和理论模型注重考虑实际渔场形成过程中的动态性和随机性。在这一过程中, 鱼类的行为时刻受到各种瞬时性和随机性要素的影响, 不一定能与外界环境之间达到假设中的均衡<sup>[21]</sup>。渔场形成是一个复杂的过程, 对这个过程的理解不同, 所采用的模型也不同。部分模型借助数值计算方法再现鱼类洄游和集群、种群变化等动态过程, 常见的有生物量均衡模型、平流扩散交互模型、基于三维水动力数值模型的物理-生物耦合模型等。如 Doan 等<sup>[82]</sup>采用生物量均衡方程进行越南中部近海围网和流刺网渔业的渔情预报研究, Rudorff 等<sup>[83]</sup>利用平流扩散方程研究大西洋低纬度地区龙虾幼体的分布, 李曰嵩<sup>[84]</sup>利用非结构有限体积海岸和海洋模型建立了东海鲈早期生活史过程的物理-生物耦合模型。另外一些模型则着眼于鱼类个体的行为, 通过个体的选择来研究群体的行为和变化。如 Dagorn 等<sup>[85]</sup>利用基于遗传算法和神经网络的人工生命模型研究金枪鱼的移动过程, 基于个体的生态模型 (individual-based model, IBM) 也被广泛地应用于鱼卵与仔稚鱼输运过程的研究<sup>[86]</sup>。

## 3 分析与展望

综上所述, 随着现代统计学、人工智能和数据挖掘技术、数值计算以及物理海洋等学科的发展, 许多模型和方法都显示出了巨大的潜力。但也应看到, 这些方法也存在一些固有的局限性。例如各类统计学方法应用较为简单, 但对建模者的渔业专家知识要求较高, 同时由于模型结构固定, 因此预测能力有限。机器学习和人工智能方法对建模者的专家知识没有太多要求, 但大多数模型结构复杂多变, 因此对建模者的实际应用经验要求较高。各类机理/过程模型和理论模型能充分考虑渔场形成过程中的动态性和随机性, 但由于缺乏实时数据的支持, 主要还是只是用于鱼类生态过程研究。

成功的渔场预报系统需要强大的渔场环境预报系统的支持。国内由于技术条件的限制, 渔情预报软件系统一般只能使用近实时的海洋环境数据代替预测数据进行预报, 这种处理方式对模型预报精度有一定影响。针对这种情况, 建议以三维水动力数值模型为基础, 研究实时海洋遥感和 Argos 浮标观测数据的有效获取和融合方法, 建立专为渔业服务的海洋环境预报系统, 为渔情预报系统提供时空分辨率较高的海洋环境数据。

准确的渔情预报建立在对目标鱼种生活史过程、种群状况和渔场形成机制的了解和把握的基础上。要获得这些知识, 必须对目标鱼种进行长期和系统的资源调查, 了解目标鱼种的产卵场、育肥场和越冬场的空间分布、幼体和成体的生长和移动过程、种群内不同群体的组成、渔业过程对种群本身的影响、渔场形成的条件以及渔汛各阶段的特点等, 从整体上把握目标鱼种的生活史过程和种群状况。海洋鱼类空间分布范围广、变化性强, 不同鱼种之间也具有较大差异, 调查中需要针对不同鱼种和海区对数据获取和处理方法进行研究, 对采样点的空间分布、采样时间、调查项目、数据重采样方式等进行标准化和规范化, 以保证数据的有效性和易用性。此外, 渔捞日志作为目前渔场预报模型建模中渔业数据的主要来源, 也应针对商业捕捞的特点进行标准化和规范化。

最后, 渔场预报模型构建过程中也有许多需要完善的地方。首先, 由于多个环境因子的限制作用、相互作用和条件作用, 确定鱼群对于单一环

境因子的响应曲线时只能依赖于某些生态学假设<sup>[87]</sup>,而这些假设可能与实际情况有偏差。其次,出于现实的考虑,研究者在建立渔情预报模型时一般使用容易获取的环境因子,而某些对渔场形成较关键但获取或表达困难的环境因子也有可能被忽略。另外,目前建立预报模型所用到的遥感数据反演精度有限,渔捞日志数据也常受到人为因素的影响,因此在准确性和代表性上有一定欠缺。这些问题都会增加渔情预报模型的不确定性,从而对预报精度产生影响。要解决这些问题,建议从渔场形成的大背景出发选择重要的环境因子,以严格的渔业海洋学、物理海洋学、生态动力学等调查与实验为基础来确定关系模型。此外,应以实际渔业生产数据和随机模拟方法相结合建立渔场随机模拟系统,通过计算机模拟的方式寻找合理的建模方法,尽可能减少数据本身的误差和不确定性,提高渔情预报模型的精度。

#### 参考文献:

- [1] 陈新军. 渔业资源与渔场学[M]. 北京:海洋出版社,2004:86-215.
- [2] 陈新军. 先进的海洋遥感与渔情预报技术[J]. 实验室研究与探索,2007,26(8):153.
- [3] 陈荣发. “东海、黄海海况速报及其中中心渔场测报”已通过鉴定[J]. 海洋开发与管理,1988(4):61.
- [4] Grant W E, Matis J H, Miller W. Forecasting commercial harvest of marine shrimp using a Markov chain model[J]. Ecological Modelling,1988,43(3-4):183-193.
- [5] Stergiou K I, Christou E D, Petrakis G. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods [J]. Fisheries Research, 1997, 29 (1): 55-95.
- [6] Georgakarakos S, Koutsoubas D, Valavanis V. Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters [J]. Fisheries Research,2006,78(1):55-71.
- [7] 程炎宏,樊伟. 东海区海洋捕捞产量的时间序列分析研究[J]. 中国水产科学,2001,8(3):31-34.
- [8] 刘中华,王庆,黄伟建. 武汉东湖鱼产量时间序列模型及回归分析[J]. 生态科学,2006,25(1):52-55.
- [9] 董江水,诸英富. 时间序列分析模型在江苏省河蟹总产量预测中的应用[J]. 金陵科技学院学报,2008,24(3):102-105.
- [10] 王为祥,朱德山. 黄海鲈鱼渔业生物学研究 II. 黄、渤海鲈鱼行动分布与环境关系的研究[J]. 海洋水产研究,1984(6):59-76.
- [11] 韦晟. 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.
- [12] 樊伟,崔雪森,沈新强. 渔场渔情分析预报的研究及其进展[J]. 水产学报,2005,29(5):706-710.
- [13] Austin M P. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling [J]. Ecological Modelling, 2002,157(2-3):101-118.
- [14] Franklin J. Mapping species distributions: spatial inference and prediction[M]. New York: Cambridge University Press,2010:105-112.
- [15] Sundermeyer M A, Rothschild B J, Robinson A R. Assessment of environmental correlates with the distribution of fish stocks using a spatially explicit model [J]. Ecological Modelling, 2006, 197: 116-132.
- [16] Zainuddin M, Saitou K, Saitou S. Albacore (*Thunnus alalunga*) fishing ground in relation to oceanographic conditions in the western North Pacific Ocean using remotely sensed satellite data [J]. Fisheries Oceanography,2008,17(2):61-73.
- [17] Goodchild M F, Steyaert L T, Parks B O, et al. GIS and environmental modeling: progress and research issues[M]. Hoboken: John Wiley & Sons,1996:35-38.
- [18] Reside A E, Waston I, van DerWal J, et al. Incorporating low-resolution historic species location data decreases performance of distribution models [J]. Ecological Modelling, 2011, 222 (18): 3444-3448.
- [19] Dormann C F, McPherson J M, Araujo M B, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review [J]. Ecography,2007,30:609-628.
- [20] 王海峰,张健,黄晓亚. 数据挖掘技术及其在渔情预报中的应用[J]. 计算机时代,2007(11):52-53.
- [21] Guisan A, Zimmermann N E. Predictive habitat distribution models in ecology [J]. Ecological Modelling,2000,135(2-3):147-186.
- [22] Venables W M, Ripley B D. Modern applied statistics with S-Plus [M]. New York: Springer-Verlag,1994.
- [23] Rykiel E J. Testing ecological models: the meaning of validation [J]. Ecological Modelling, 1996, 90

- (3):229-244.
- [24] 周彬彬. 应用自由度调整复相关系数方法进行渔期预报的研究[J]. 海洋学报(中文版), 1987, 9(6):774-779.
- [25] Harrell F E, Lee K L, Mark D B. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors[J]. *Statistics in Medicine*, 1996, 15(4):361-387.
- [26] Nelder J A, Wedderburn R W M. Generalized linear models[J]. *Journal of the Royal Statistical Society, Series A*, 1972, 135(3):370-384.
- [27] Houwelingen V J C, Cessie S L. Predictive value of statistical models[J]. *Statistics in Medicine*, 1997, 9(11):1303-1325.
- [28] Manel S, Dias J M, Ormerod S J. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions; a case study with a Himalayan river bird[J]. *Ecological Modelling*, 1999, 120(2):337-347.
- [29] Fieldings A H, Bell J F. A review of methods for the assessment of prediction errors in conservation presence/absence models [J]. *Environmental Conservation*, 1997, 24(1):38-49.
- [30] 陈新军. 西北太平洋柔鱼渔场与水温因子的关系[J]. 上海水产大学学报, 1995, 4(3):181-185.
- [31] 陈新军, 冯波, 许柳雄. 印度洋大眼金枪鱼栖息地指数研究及其比较[J]. 中国水产科学, 2008, 5(2):269-278.
- [32] 邵全琴, 马巍巍. 西北太平洋黑潮路径变化与柔鱼 CPUE 的关系研究[J]. 海洋与湖沼, 2005, 36(2):111-122.
- [33] 刘传楨, 严隽箕. 渤海秋汛对虾数量预报方法的研究[J]. 水产学报, 1981, 5(1):65-79.
- [34] 陈新军, 刘必林, 田思泉, 等. 利用基于表温因子的栖息地模型预测西北太平洋柔鱼 (*Ommastrephes bartramii*) 渔场[J]. 海洋与湖沼, 2009, 40(6):707-713.
- [35] 冯波, 田思泉, 陈新军. 基于分位数回归的西南太平洋阿根廷滑柔鱼栖息地模型研究[J]. 海洋湖沼通报, 2010, (1):15-22.
- [36] Guisan A, Edwards T C, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene [J]. *Ecological Modelling*, 2002, 157(2-3):89-100.
- [37] Venables W N, Ripley B D. GLMs, GAMs and GLMMs; an overview of theory for applications in fisheries research[J]. *Fisheries Research*, 2004, 70:319-337.
- [38] Chang S K, Hoyle S, Liu H I. Catch rate standardization for yellowfin tuna (*Thunnus albacares*) in Taiwan's distant-water longline fishery in the Western and Central Pacific Ocean, with consideration of target change [J]. *Fisheries Research*, 2011, 107(1-3):210-220.
- [39] Francis M P, Morrison M A, Leathwick J, et al. Predictive models of small fish presence and abundance in northern New Zealand harbours [J]. *Estuarine, Coastal and Shelf Science*, 2005, 64(2-3):419-435.
- [40] Chang J H, Chen Y, Holland D, et al. Estimating spatial distribution of American lobster *Homarus americanus* using habitat variables [J]. *Marine Ecology Progress Series*, 2010, 420:145-156.
- [41] Lehmann A, Overton J M, Leathwick J R. GRASP: generalized regression analysis and spatial prediction [J]. *Ecological Modelling*, 2003, 160(1-2):165-183.
- [42] 牛明香, 李显森, 徐玉成. 基于广义可加模型和案例推理的东南太平洋智利竹筴鱼中心渔场预报[J]. 海洋环境科学, 2012, 31(1):30-33.
- [43] 官文江, 陈新军. 应用一般线性模型估算鲈、鲈大型灯光围网渔业的捕捞效率[J]. 水产学报, 2009, 33(2):221-227.
- [44] 樊伟, 陈雪忠, 沈新强. 基于贝叶斯原理的大洋金枪鱼渔场速预报模型研究[J]. 中国水产科学, 2006, 13(3):426-431.
- [45] Pereira J M, Itami R M. GIS-based habitat modeling using logistic multiple regression; a study of the Mt. Graham Red Squirrel [J]. *Photogrammetric Engineering and Remote Sensing*, 1991, 57(11):1475-1486.
- [46] Osborne P E, Alonso J C, Bryant R G. Modelling landscape-scale habitat use using GIS and remote sensing; a case study with great bustards[J]. *Journal of Applied Ecology*, 2001, 38:458-471.
- [47] Tarmansen M, McClean C J, Preston C D. The use of genetic algorithms and Bayesian classification to model species distributions [J]. *Ecological Modelling*, 2006, 192(3-4):410-424.
- [48] Dlamini W M. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland [J]. *Environmental Modelling and Software*, 2010, 25(2):199-208.
- [49] Brockwell P J, Davis A R. Introduction to time series and forecasting[M]. 北京:人民邮电出版社, 2009.



- [50] Corbineau A, Rouyer T, Cazelles B, *et al.* Time series analysis of tuna and swordfish catches and climate variability in the Indian Ocean(1968 – 2003) [J]. *Aquatic Living Resources*, 2008, 21 (3): 277 – 285.
- [51] Tobler W. A computer movie simulating urban growth in the Detroit region [J]. *Economic Geography*, 1970, 46(2): 234 – 240.
- [52] Rivoirard J, Simmonds J, Foote K G, *et al.* Geostatistics for estimating fish abundance [M]. London: Blackwell Science Inc., 2000: 1 – 216.
- [53] 苏奋振, 周成虎, 邵全琴. 东海区鱼类资源时空变化[J]. *中国水产科学*, 2001, 8(3): 15 – 19.
- [54] 杜云艳, 周成虎, 崔海燕. 遥感与 GIS 支持下的海洋渔业空间分布研究—以东海为例[J]. *海洋学报*, 2002, 24(5): 57 – 63.
- [55] 苏奋振, 周成虎, 史文中, 等. 东海区底层及近底层鱼类资源的空间异质性[J]. *应用生态学报*, 2004, 15(4): 683 – 686.
- [56] Hengla T, Sierdsemab H, Radovi A, *et al.* Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging [J]. *Ecological Modelling*, 2009, 220(24): 3499 – 3511.
- [57] Monestieza P, Dubrocab L, Bonnin E, *et al.* Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts [J]. *Ecological Modelling*, 2006, 193(3 – 4): 615 – 628.
- [58] Tan P N, Steinbach M, Knmar V. Introduction to data mining [M]. 北京: 人民邮电出版社, 2011: 1 – 11.
- [59] 易倩, 陈新军. 基于信息增益法选取柔鱼中心渔场的关键水温因子[J]. *上海海洋大学学报*, 2012, 21(3): 425 – 430.
- [60] 楼文高. 人工神经网络在水产科学中的应用[J]. *上海水产大学学报*, 2001, 10(4): 347 – 352.
- [61] 吴凤霞, 李纯厚, 戴明. 人工神经网络在海洋科学中的应用[J]. *南方水产*, 2009, 5(1): 75 – 80.
- [62] 邵帼英. 基于神经网络的东南太平洋公海智利竹筴鱼渔情预报的研究[D]. 上海: 上海水产大学, 2007: 1 – 73.
- [63] Brosse S, Guegan J F, Tourenq J N, *et al.* The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake [J]. *Ecological Modelling*, 1999, 120(2 – 3): 299 – 311.
- [64] Brosse S, Giraudel J L, Lek S. Utilization of non-supervised neural networks and principal component analysis to study fish assemblages [J]. *Ecological Modelling*, 2001, 146(1 – 3): 159 – 166.
- [65] Iglesias A, Dafonte C, Arcay B, *et al.* Integration of remote sensing techniques and connectionist models for decision support in fishing catches [J]. *Environmental Modelling and Software*, 2007, 22(6): 862 – 870.
- [66] Suryanarayana I, Braibanti A, Rao R S, *et al.* Neural networks in fisheries research [J]. *Fisheries Research*, 2008, 92(2 – 3): 115 – 139.
- [67] Yanez E, Plaza F, Gutierrez-Estrada J C, *et al.* Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) abundance forecast off northern Chile: A multivariate ecosystemic neural network approach [J]. *Progress in Oceanography*, 2010, 87(1 – 4): 242 – 250.
- [68] Moisen G G, Frescino T S. Comparing five modelling techniques for predicting forest characteristics [J]. *Ecological Modelling*, 2002, 157(2): 209 – 225.
- [69] Carpenter G A, Gopal S, Macomber S, *et al.* A neural network method for efficient vegetation mapping [J]. *Remote Sensing of Environment*, 1999, 70(3): 326 – 338.
- [70] Olden J D, Lawler J J, Poff N L. Machine learning methods without tears: a primer for ecologists [J]. *Quarterly Review of Biology*, 2008, 83(2): 171 – 193.
- [71] 尹朝庆. 人工智能与专家系统 [M]. 北京: 水利水电出版社, 2009: 1 – 14.
- [72] 杨胜龙, 周为峰, 伍玉梅, 等. 西北印度洋大眼金枪鱼渔场预报模型建立与模块开发 [J]. *水产科学*, 2011, 30(11): 667 – 672.
- [73] Sady M, Hendiarti N, Sachoemar S I, *et al.* Fishing ground prediction using a knowledge-based expert system geographical information system model in the South and Central Sulawesi coastal waters of Indonesia [J]. *International Journal of Remote Sensing*, 2009, 30(24): 6429 – 6440.
- [74] Sady M, Hendiarti N, Sachoemar S I, *et al.* Application of knowledge-based expert system model for fishing ground prediction in the tropical area [C]. Jakarta: The Second APEC Workshop of SAKE, 2007: 1 – 7.
- [75] U. S. Fish and Wildlife Service. Standards for the development of habitat suitability index models [R].

- Washington, D. C; U. S. Fish and Wildlife Service, 1981;1 - 81.
- [76] 龚彩霞,陈新军,高峰,等. 栖息地适宜性指数在渔业科学中的应用进展[J]. 上海海洋大学学报, 2011,20(2):260 - 269.
- [77] 叶施仁,史忠植. 基于 CBR 的中心渔场预报[J]. 高技术通讯,2001,11(5):64 - 68.
- [78] Ichiro A. 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.
- [79] 崔雪森,樊伟,沈新强. 西北太平洋柔鱼渔情速报系统的开发[J]. 水产学报, 2003, 27(6): 600 - 606.
- [80] 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,2006,7(2): 69 - 72.
- [81] 张月霞,丘仲锋,伍玉梅,等. 基于案例推理的东海区鲈鱼中心渔场预报[J]. 海洋科学,2009,33(6): 8 - 11.
- [82] Doan B, Le H C, Nguyen D T. Fishing ground forecast in the offshore waters of Central Vietnam (experimental results for purse-seine and drift-gillnet fisheries) [J]. VNU Journal of Science, Earth Sciences,2010,26(2):57 - 63.
- [83] Rudorff C A G, Lorenzetti J A, Gherardi D F M, et al. Modeling spiny lobster larval dispersion in the Tropical Atlantic [J]. Fisheries Research, 2009, 96(2 - 3):206 - 215.
- [84] 李曰嵩. 东海鲈鱼 (*Scomber japonica*) 早期生活史过程的生态动力学模拟研究[D]. 上海:上海海洋大学,2011:1 - 157.
- [85] Dagorn L, Petit M, Stretta J M. Simulation of large-scale tropical tuna movements in relation with daily remote sensing data; the artificial life approach [J]. BioSystem,1997,44(3):167 - 180.
- [86] 陈新军,李曰嵩. 基于个体生态模型在渔业生态中应用研究进展[J]. 水产学报, 2012, 36(4): 629 - 639.
- [87] Austion M P. Species distribution models and ecological theory: a critical assessment and some possible new approaches [J]. Ecological Modelling, 2007,200(1 - 2):1 - 19.

## Review of fishery forecasting technology and its models

CHEN Xinjun<sup>1,2</sup>, GAO Feng<sup>1,2\*</sup>, GUAN Wenjiang<sup>1,2</sup>, LEI Lin<sup>1,2</sup>, WANG Jintao<sup>1,2</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. National Distant-water Fisheries Engineering Research Center, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** Fishery forecast is an important research content of oceanography, which has the vital significance to the scientific production and management of fishery resources in fishery science. In recent years, with the development of modern statistics theory, the numerical calculation method, and data mining and artificial intelligence theory and technology, the development of fishery forecasting technology and model has displayed a new vitality. Therefore, the studies on the fishery forecasting technology and model development are reviewed, and the future development of fishery forecasting was put forward. In this paper, the theory and methods of fishery forecasting are summarized, including fishery oceanography, data models and prediction models related to this subject. Prediction models based on statistics methods and machine learning and artificial intelligence methods are emphasized, as well as the advantages and drawbacks of each kind of the forecasting model. Some research perspectives of fishery forecasting models are also proposed, i. e. developing ocean environments forecasting system, conducting systematic fishery resources survey of long standing and the standardization and normalization of fishery data acquisition and processing, reducing the uncertainty of prediction models with stochastic simulation methods and improving the prediction accuracy.

**Key words:** fishery forecasting; prediction models; statistics; machine learning

**Corresponding author:** GAO Feng. E-mail: gaofeng@shou.edu.cn