基于空间自回归和空间聚类的渔情预报模型

袁红春[1], 谭明星[1], 顾怡婷[1], 陈新军[2]

(上海海洋大学: 1. 信息学院; 2. 海洋科学学院, 上海 201306)

摘要: 作者针对远洋渔场渔情预报精度偏低的问题, 提出一种基于空间自回归和空间聚类的渔情预报模型。该模型利用空间自回归对收集到的渔业历史数据进行预处理, 然后通过空间聚类将所有数据样本根据地理位置分划成若干个区域, 最后研究每个区域中环境数据与渔获数据之间的数学关系, 各自建立栖息地适宜性指数模型(Habitat Suitability Index, HSI), 并以印度洋大眼金枪鱼(*Thunnus obesus*)为例进行验证。结果表明, 本模型的均方差为 0.1742, 与传统线性回归方法的均方差 0.2363 相比, 能更好地表达海洋环境数据与渔获量之间的关系, 预测精度显著提高。

关键词: 大眼金枪鱼(*Thunnus obesus*); 渔情预报; 空间自回归; 空间聚类; 非线性回归

中图分类号: S934 **文献标识码**: A **文章编号**: 1000-3096(2015)12-0165-08

doi: 10.11759/hykx20140705002

传统的远洋渔情预报中, 多数学者采用多元线性回归方法定量地对环境数据及渔获数据进行分析, 获得预报模型。如沈金鳌等[1]通过多元线性回归分析, 将带鱼的资源量指数、各汛期的总捕捞努力量及长江径流量作为预报因子建立模型, 预报崾山冬季带鱼(*Trichiurus lepturus*)鱼汛期的渔获量。陈新军等[2]在分析东黄海鲐(*Scomber japonicus*)的渔场时, 采用广义线性模型(Generalized Linear Model, GLM)和广义可加模型(Generalized Additive Model, GAM), 结合海洋环境因子建立预报模型, 以分析渔场资源的形成机制。陈新军等[3-4]研究了西北太平洋柔鱼(*Ommastrephes batrami*)渔场与环境因子的内在关系, 利用月相对光诱鱿钓作业的影响, 证明月相对日产量的影响显著。但由于环境数据和渔获数据之间的关系并非线性的, 传统的线性回归预报模型不能准确地预测渔情信息。随着人工智能的崛起, 近年来有学者将人工智能技术用于渔情预测。杜云艳等[5]采用空间聚类方法建立关于渔业数据和对应水温间的时空分布模型。袁红春等[6]利用 BP(Back Propagation)神经网络预报西北太平洋柔鱼渔获情况。张月霞等[7]利用案例推理预报东海区鲐鱼中心渔场。但单纯的空间聚类、专家系统、人工神经网络或案例推理方法并不能准确地反应渔业数据的时空分布。单独的空间聚类只考虑空间因素, 忽略了其他因子的影响, 从而导致预测误差较大; 专家系统过于依靠专家知识经验; 人工神经网络的黑箱操作导致训练结果不易理解。上述原因阻碍了人工智能在渔情预报中的应用。其他学者, 如 Masahiko 等[8]、Mohri 等[9]、Daniel 等[10]分别研究了最适宜大眼金枪鱼栖息的水温范围对产量的影响。日本的 Hiroaki[11]通过 GLM 方法标准化印度洋大眼金枪鱼的延绳钓渔获量, 研究海洋环境与金枪鱼渔获量之间的关系。综上, 现有方法仅用单一的模型描述整片海域中的渔情信息, 而同一片海域中不同位置的自然环境不同, 单一的模型不能准确地预报整体渔情。同时, 依据现有统计数据可知, 大量渔获数据缺失, 因数据缺失而导致无法准确描述及研究环境和渔获量之间的关系, 也是渔情预报过程中亟待解决的问题之一[12]。本研究以印度洋大眼金枪鱼为例[13, 14], 提出一种基于空间自回归和空间聚类的动态渔情预测模型, 以丰富预测方法, 提高预测水平。

1 材料与方法

1.1 数据来源

海洋环境数据(海面高度, 海表温度, 叶绿素浓度)来源于美国国家海洋和大气管理局, 渔业作业数据(渔获量)来源于印度洋金枪鱼委员会。

收稿日期: 2014-07-05; 修回日期: 2014-08-06

基金项目: 上海市教育委员会科研创新重点项目(12ZZ162); 上海市科学技术委员会科技支撑项目(12510502000, 14391901400)

作者简介: 袁红春(1971-), 男, 江苏海门人, 教授, 博士, 主要从事智能信息处理研究, 电话: 021-61900604, E-mail: hcyuan@shou.edu.cn; 谭明星, 通信作者, E-mail: 345010018@qq.com

1.2 基于空间自回归的数据预处理

1.2.1 数据的补缺

由于渔获数据缺失严重，若要利用该数据建立预测模型，需进一步补全缺失数据[15]。本研究在传统用于补全数据模型的基础上增加修正项，构建空间自回归模型。首先，给定观察数据为一个 n 维向量 Y，表示渔业产量数据，一个 $n\times m$ 的矩阵 $\boldsymbol{X}$ 为海洋环境数据(m 为环境因子数)。假设 Y 中每一个因变量 y_i 都互相影响，即：$y_i = f(y_j), i \neq j$。回归方程可以修正成如下形式:

$$Y = \rho WY + \beta X + \varepsilon \tag{1}$$

$\boldsymbol{W}$ 是邻接矩阵，在回归模型的空间拓展上起决定性作用。ρ 是解释变量和因变量之间空间独立性强度的参数。残差向量 ε 被假定为服从独立同分布的标准正态分布，β 是系数矩阵。

本研究用八-邻居准则构建邻接矩阵 $\boldsymbol{W}$。一个八-邻居准则构建的矩阵见图 1。

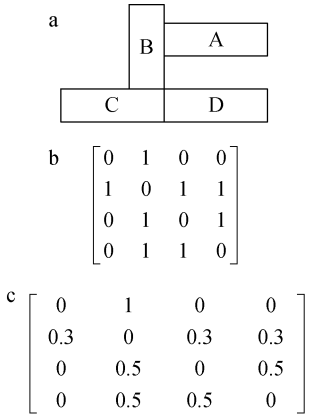

图 1　八-邻居邻接矩阵

Fig.1　An eight-neighborhood its adjacency matrix

a. 地图上位置；　b. 二进制邻接矩阵; c. 归一化后邻接矩阵

a. Position on map; b. Binary adjacency matrix; c. Normalized adjacency matrix

图 1(a)表示空间上 A、B、C、D 4 个单元格的位置关系，图 1(b)表示每个点与其余点位置的关系矩阵。以矩阵第二行为例，第二行表示单元格 B 与其余单元格的位置关系，因 B 与 A 有相邻的边、与 C 共享一条边和一个顶点、与 D 有相接的顶点，所以第一列(A)为 1，第三列(C)为 1，第四列(D)为 1，其余位置为 0。标记每个单元格与其他单元格之间的关系后，得到图 1(b)中以二进制形式表示的邻接矩阵。然后对邻接矩阵中有标记的数值进行归一化。具体方法为：令每一行之和为 1，即归一化后的标记值为 1 除以该行不为 0 的数值的个数。以第二行为例，有标记的个数为 3，1/3 约为 0.3，则每个有标记的值归一化后均为 0.3。

当 $\rho = 0$，式(1)就变成传统回归方程。修正后的模型有以下优点：(1)残差有很少的空间自相关性；(2)如果空间自相关系数在统计学上非常大，就可以确定空间自相关存在的数量。这意味着因变量 y 变化的范围是 y 到相邻观测值的平均值；(3)模型的拟合度很高。

选取产量数据不为零的数据作为数据集，构建空间自回归模型。模型建立后，对空间缺失值进行插补，插补方式为

$$Y_i = \frac{1}{n}\sum_{j=1}^{n} Y_j + \rho W_i Y + \beta X_i \tag{2}$$

其中 Y_i 为待预测的缺失产量数据，$\frac{1}{n}\sum_{j=1}^{n} Y_j$ 表示该预测区域内产量的平均值，W_i 为空间标准加权矩阵 $\boldsymbol{W}$ 的第 i 行，Y 是表示渔业产量的向量，Y 中缺失产量用 0 表示。

1.2.2 数据的归一化

为了研究印度洋大眼金枪鱼延绳钓渔业产量与相应栖息地海域海洋环境因子的关系，需要建立一种客观标准反映该区域的资源丰度。单位捕捞努力量渔获量(Catch Per Unit Effort, CPUE)的大小常被作为资源丰度的相对指数来反映资源丰度的变化，其定义为

$$CPUE_{(i,j)} = \frac{N_{fish_{(i,j)}} \times 1000}{N_{hook_{(i,j)}}} \tag{3}$$

其中，$CPUE_{(i,j)}$ 代表以经纬度(i, j)为起点的范围 5°×5° 区域内钓获率，$N_{fish_{(i,j)}}$ 为该范围内的渔获尾数，$N_{hook_{(i,j)}}$ 为该范围内的下钩枚数。$CPUE_{(i,j)}$ 也可解释为每千钩上钩的大眼金枪鱼数量。再对单位捕捞努力量渔获量进行处理，使用相对资源指数(Relative Abundance Index, RAI)构建栖息地适宜性指数模型。相对资源指数由某一时间地点的单位捕捞努力量渔获量值除以所有单位捕捞努力量渔获量值中的最大值得到，计算方法为

$$RAI_{(i,j)} = \frac{CPUE_{(i,j)}}{CPUE_{\max}} \tag{4}$$

相对资源指数可看作反映栖息地质量的指标，等价于实际的栖息地适宜指数。

1.3 基于空间聚类的渔区划分

根据环境数据及渔获数据的相关性，利用每条数据的地理位置信息进行 K-Means 聚类，把地理位置相对较近、渔获数据之间相关系数较高的区域划分为一类，从而大大降低预测误差。空间中数据对象的记录主要包括空间实体的位置、形状以及对象之间的相互关系，这种关系通常以坐标或者拓扑的形式表示。本研究将每 5°×5° 区域看作空间上的一个实体点，每个实体点包含地理位置数据(纬度，经度)和产量数据(栖息地适宜性指数 HSI)。其中，地理位置数据中的纬度和经度为实体点的空间数据，产量数据 HSI 为实体点的属性数据。结合空间实体点的空间数据和属性数据等多方面因素考虑，并参考大量有关聚类分析的文献后[16-18]，设计如下适用于渔情预报的空间聚类算法：

算法 1 空间聚类算法

输入: 地理位置数据和产量数据

输出: 聚类结果

步骤如下:

(1) 确定待聚类数据集 D，聚类数 K，预设迭代次数 m 和收敛条件;

(2) 初始化聚类重心 C_i;

(3) 计算每个数据 D_i 到各个 C_i 的距离，选取最近的距离归并到该类中;

(4) 更新重心 C_i;

(5) 计算所有 C_i 值的变化;

(6) 直到满足收敛条件或达到迭代次数 m，停止迭代。

其中聚类重心的计算方法如下:

$$X=\frac{\sum_{i=1}^{k}M_i\cdot X_i}{\sum_{i=1}^{k}M_i},Y=\frac{\sum_{i=1}^{k}M_i\cdot Y_i}{\sum_{i=1}^{k}M_i} \tag{5}$$

X 为聚类重心的经度，Y 为聚类重心的纬度，M_i 为渔区 i 每个月的产量，X_i 为渔区 i 重心点的经度，Y_i 为渔区 i 重心点的纬度，渔区的个数为 k。通过该聚类方法可以把地理位置相对较近、渔获数据之间相关系数较高的区域划分为一类，从而减少因不同地理位置或不同环境所引起的误差。

1.4 基于非线性回归的栖息地适宜性指数模型

HSI 模型在 20 世纪 80 年代由美国鱼类和野生动物保护委员会提出[19-21]，被用来定量地描述野生动物的栖息地质量。

考虑到渔获量的高低不仅与地理位置有关，还与鱼类生存的环境因子，如海表温度(SST)、海面高度(SSH)和叶绿素浓度(CHL-a)等有关。因此，本研究针对每个环境因子，利用其和栖息地适宜性指数之间的关系，分别计算其对大眼金枪鱼产量影响的适宜性指数(Suitability Index, SI)；最后，通过回归分析关联各种 SI 值得到综合 HSI 模型(图 2)。

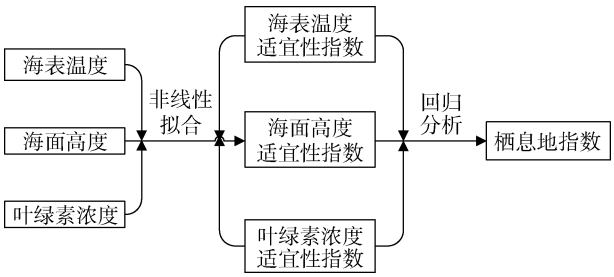

图 2 栖息地适宜性指数模型构建方法

Fig.2 Method for constructing the habitat suitability index model

利用综合优化分析计算软件平台——First Optimization 的公式拟合工具箱，对每个环境因子和其对应的 RAI 分别作非线性拟合，把最佳拟合公式作为栖息地适宜性指数计算公式中的单项栖息地适宜性指数 SI，通过对 3 种环境因子在不同月份的拟合结果比较可知，3 种环境因子与栖息地适宜性指数的拟合形式可用如下形式表示:

$$SI=\sum A\cdot X^B \tag{6}$$

其中，$A=(a_1,a_2,\cdots,a_{10})$，$B=(0,1,\cdots,9)$，$X=(x_1,x_2,\cdots,x_m)'$ 为环境因子。因此，在计算参数 A 时，可先计算 X^B，令 $Y=X^B$，将非线性回归形式转为线性回归形式:

$$SI=\sum A\cdot Y \tag{7}$$

计算完每一项 SI 之后，本研究将栖息地适宜性指数模型设置为:

$$HSI=\mathrm{a}\times SI_{\mathrm{sst}}+b\times SI_{\mathrm{ssh}}+c\times SI_{\mathrm{chla}}+d \tag{8}$$

SI 是单项栖息地适宜性指数，d 为回归方程中的常数项。根据每个月不同的环境数据和栖息地适宜性指数数据，利用回归方程(8)计算出参数 a、b、c 和 d 即可得到每个月的综合栖息地适宜性指数模型。

2 实验过程与实验结果

2.1 实验过程

本研究的整个实验流程见图3，在数据预处理部分加入空间自回归对缺失数据进行补全；然后利用空间聚类对补全后的数据进行渔区划分；再对每个渔区分别建立栖息地适宜性指数模型；最后通过捕捞点的地理位置和环境信息，确定该点所属渔区，结合该渔区的预测模型对该点产量进行预测，并与真实数据对比，从而验证该模型。

利用 2005 年 1 月~2010 年 12 月印度洋大眼金枪鱼的数据进行建模。首先，根据本研究提出的空间自回归对缺损的数据进行补全。从 IOTC 官方网站下载印度洋 40°S~15°N，40°E~120°E 区域内大眼金枪鱼渔获数据。根据文献[13]中显示，印度洋大眼金枪鱼除 45°S 以上的高纬度区域之外均有产量。但获取的作业数据中，许多地理位置没有记录相应的产量数据。通过选取产量不为零的数据作为建立模型的数据集，对产量的空间缺失值进行插补，扩充后续建模数据样本。

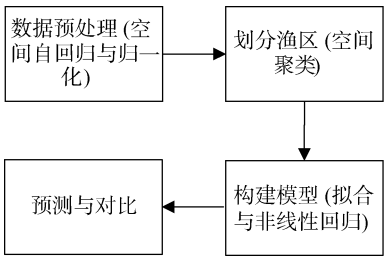

图 3　实验流程及对应方法

Fig.3　Experimental process and corresponding methods

以 2009 年 1 月数据为例，先对东经取正值，西经取负值，北纬取 90–纬度值，南纬取 90+纬度值，对环境数据进行归一化，与渔业作业数据进行匹配即可获得适合后续聚类、回归的数据样本。表 1 为 2009 年 1 月经过组织之后的数据样本。

表 1　2009 年 1 月数据样本(部分)

Tab.1　Data sample in Jan 2009(partial)

年	月份	纬度	经度	**SST (℃)**	SSH (m)	CHL-a (μg/L)	HSI
2009	1	85	75	28.04	38.53	0.42	0.63
2009	1	85	60	28.39	40.50	0.45	0.44
2009	1	85	65	28.40	43.74	0.29	0.43
2009	1	90	75	28.60	36.40	0.21	0.65
2009	1	90	85	28.77	39.73	0.18	0.57
2009	1	90	65	28.86	42.88	0.15	0.46
2009	1	90	80	28.86	44.27	0.13	0.41
2009	1	90	70	28.75	46.69	0.12	0.36
2009	1	90	90	28.82	49.30	0.13	0.31
2009	1	90	50	28.60	21.59	0.21	0.20
2009	1	90	60	28.07	22.43	0.07	0.18
2009	1	90	55	28.13	25.86	0.08	0.15
2009	1	95	95	28.39	34.77	0.12	0.62
2009	1	95	90	28.78	43.39	0.15	0.52
2009	1	95	50	28.84	45.53	0.18	0.41
2009	1	95	85	28.93	44.09	0.18	0.40
2009	1	95	45	29.03	44.46	0.17	0.40
2009	1	95	80	29.46	60.35	0.91	0.39

然后根据算法 1，得到聚类结果。聚类数量根据实际应用而定，它的值会直接影响最终聚类结果，在实际应用中 K 值通常取为 2~5。因此，本研究将 K 分别设置为 2、3、4、5，利用 Matlab2012b 进行仿真实验。根据不同 K 值对数据进行聚类，得到所有数据点归属类别后，逐类别进行相关系数计算，设 $x_i=(x_{i1},x_{i2},...,x_{ip})$ 和 $x_j=(x_{j1},x_{j2},...,x_{jp})$ 是第 i 个和 j 个数据点观测值，两个数据点之间相关系数为:

$$\gamma_{ij}=\frac{\sum_{k=1}^{p}(x_{ik}-\bar{x_i})(x_{jk}-\bar{x_j})}{\sqrt{\left[\sum_{k=1}^{p}(x_{ik}-\bar{x_i})^2\sum_{k=1}^{p}(x_{jk}-\bar{x_j})^2\right]}} \tag{9}$$

总体相关系数为:

$$\overline{r_{ij}}=\frac{\sum r_{ij}}{k} \tag{10}$$

通过算术平均方法得到不同 K 值下的相关系数，表 2 为相关系数计算结果。

根据计算结果，当 K=3 时，得到的相关系数最大。因此，本研究将 K 值确定为 3。

表 2　K 取不同数值的相关系数
Tab.2　Correlation coefficients of different K values

K 取值	2	3	4	5
相关系数	0.446	0.567	0.526	0.431

在 K 均值算法中，聚类重心的确定对后续聚类结果有很大影响。由于聚类的目的是找到地理位置上相对集中的几个区域，从而减少回归的误差，而聚类的指标是以距离为度量值，因此，本研究通过均匀划分经纬度选取初始聚类重心。由于研究区域为印度洋 40°S~15°N，40°E~120°E，其经向范围较纬向范围大，因此，本研究以赤道 0°为纬度基线，以此基线平分经度跨度，东经取正值，西经取负值，北纬取 90–纬度值，南纬取 90+纬度值，则初始中心的坐标以经纬度表示约为(0, 54), (0, 80)和(0, 106)。

令阈值为 0.5，将最终聚类迭代次数限定为 1000 次，利用前文提出的空间聚类算法，以表 1 的数据为例根据经纬度数据以及 HSI 数据，对 2009 年 1 月份聚类结束后，经过组织得到表 3。

表 3　2009 年 1 月数据聚类结果(部分)
Tab.3　Results of data clustering in Jan 2009 (partial)

年	月份	纬度	经度	SST (℃)	SSH (m)	CHL-a (μg/L)	HSI	聚类组号
2009	1	85	75	28.04	38.53	0.42	0.63	3
2009	1	85	60	28.39	40.50	0.45	0.44	3
2009	1	85	65	28.40	43.74	0.29	0.43	3
2009	1	90	75	28.60	36.40	0.21	0.65	3
2009	1	90	85	28.77	39.73	0.18	0.57	1
2009	1	90	65	28.86	42.88	0.15	0.46	3
2009	1	90	80	28.86	44.27	0.13	0.41	1
2009	1	90	70	28.75	46.69	0.12	0.36	3
2009	1	90	90	28.82	49.30	0.13	0.31	1
2009	1	90	50	28.60	21.59	0.21	0.20	3
2009	1	90	60	28.07	22.43	0.07	0.18	3
2009	1	100	65	27.91	76.09	0.09	0.48	3
2009	1	100	70	27.67	75.62	0.04	0.39	3
2009	1	100	45	27.67	75.21	0.04	0.14	2
2009	1	100	60	27.23	79.74	0.04	0.08	3
2009	1	105	105	26.61	77.90	0.05	0.65	1
2009	1	105	95	25.86	59.03	0.04	0.39	1
2009	1	110	45	27.89	59.75	0.04	0.16	2
2009	1	110	40	26.17	59.58	0.03	0.04	2
2009	1	120	35	27.33	62.47	0.16	0.11	2

根据 2005 年 1 月至 2010 年 12 月的聚类结果，类别 1 为赤道~35°S、80°E~110°E 的东印度洋区域；类别 2 为 10°N~10°S、50°E~85°E 热带印度洋区域；类别 3 为赤道~35°S、30°E~55°E 的西印度洋区域。

将 2005 年 1 月~2010 年 12 月每一个月的数据进行聚类，然后对所有数据进行类别标号。选取 2005 年 1 月~2010 年 12 月的金枪鱼延绳钓产量数据和相关海洋环境因子作为建模数据。对每一个环境因子和产量分别作非线性拟合，将最佳拟合公式作为栖息地适宜性指数计算公式中的单项 SI，再利用公式(8)计算出参数 a、b、c 和 d 即可得到每个月的综合栖息地适宜性指数模型。

2.2　实验结果

本研究通过聚类过程发现：当海表温度在 26℃~29℃，叶绿素浓度在(0.1~0.3)mg/m^3，海面高度较高时，栖息地适宜性指数较高，这与文献[13, 22~24]中的表述一致。

本研究以 2005 年至 2010 年的 1 月份数据为例，将栖息地适宜性指数大于 0.4 的网格点在地图上标出，结果发现：印度洋大眼金枪鱼延绳钓的产量分布面较广，除 45°S 以上的高纬度海域外，几乎整个印度洋海域，均有其产量分布；主要渔场集中在 10°N~20°S，50°E~85°E 的海域；高产渔区以西印度洋为主；东印度洋也有分布，但产量明显稀少。以上均与文献[13]和文献[25]中吻合。

由于在构建栖息地适宜性指数模型之前对数据进行了空间聚类，将地理位置相对较近而且栖息地适宜性指数与环境因子相关性较高的数据样本聚集到一起，因此在模型检验及后续预测中心渔场中，需要根据数据样本到每个聚类重心的距离选取栖息地适宜性指数模型。

本研究采用 2011 年 1~12 月的数据对模型进行验证。针对每月数据，根据每条记录对应的经纬度，计算其到 3 个聚类重心的距离，确定其所属类别。再将该记录的环境数据代入相应的栖息地适宜性指数模型，得到对应的栖息地适宜性指数估计值。最后在同一张图中对比预测值与真实值得到图 4。

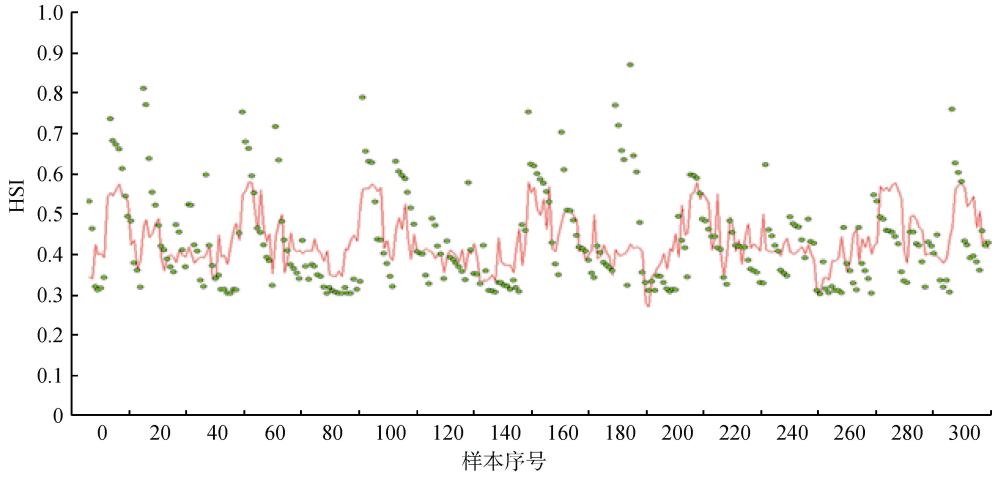

图 4　预测与真实数据比较

Fig.4　Comparison of predicted and actual data

在图 4 中横坐标表示每组数据的序号，圆点是实际HSI值，折线为预测模型得到的HSI值，由图可知真实 HSI 最高值接近 1.0，但多数集中在 0.4~0.6，而预测值最高约为 0.5，且出现频率较高，与事实符合。

3　讨论与结论

得到每个月的栖息地适宜性指数估计值之后，计算其均方误差(MSE)，图 5 为本研究提出的基于空间聚类的渔情预测方法(Spatial Clustering Based Fishery Forecasting, SCBFF)和传统线性回归方法的误差对比。

由图 5 可知，本研究提出的方法要优于传统的线性回归方法，这是因为传统的线性回归预测方法并未考虑到因为数据缺失而导致预测精度偏低的问题，作者通过空间自回归模型补全数据，从数据本身的角度出发，为之后的预测奠定了良好的基础。此外传统的预测方法是直接对整个渔区进行建模预测，并未考虑不同地理位置或不同环境所引起的偏差，作者充分考虑该偏差所引起的预测误差较大的问题，根据环境数据随着地理位置变化而变化的特点，采用基于空间聚类的渔区划分方法，将渔获量之间相关性较高并且相对地理位置较近的数据，聚集在同一个类中，对每个渔区分别建立模型，有效地提高了预测精度。

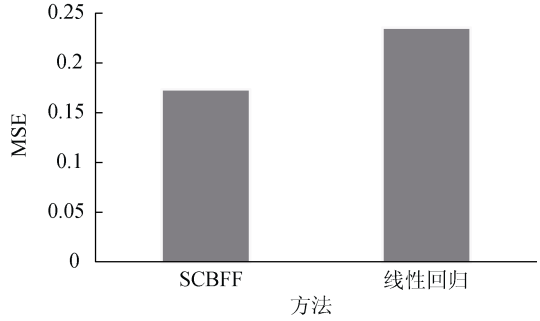

图 5　SCBFF 与线性回归的均方误差比较

Fig.5　Comparison of MSE between SCBFF and linear regression

由于环境数据采集困难，本研究仅从海表温度、

海面高度和叶绿素浓度等 3 个方面考虑海洋环境因子对金枪鱼产量的影响，但实际上还有许多影响印度洋金枪鱼分布的因素，包括海水盐度、饵料生物分布等环境因子，这为本课题的后续研究提供空间。同时，本研究能够根据现有的 3 种环境因子与栖息地适宜性指数之间的关系构建出栖息地适宜性指数模型，且平均预测性能良好，但对于部分极值，如栖息地适宜性指数非常高的区域并不能很好地预测出来，这也将成为后期深入研究方向之一。

本研究针对传统渔情预报方法由于精度偏低的问题，利用空间自回归、*K* 均值空间聚类、非线性回归等技术的优点，提出一种基于空间自回归和空间聚类的渔情预报模型。首先，利用空间自回归根据空间位置插补数据的能力，插补缺损数据，从而补全数据。再利用基于空间聚类的渔区划分方法，对渔区进行划分。最后根据每一类中环境数据和产量数据之间的数学关系，通过函数拟合与非线性回归分析，得到每个不同渔区块各自的栖息地适宜性指数模型。通过与传统的金枪鱼预测方法进行对比试验，实验结果表明，在提高模型拟合程度和减小预测误差方面，本方法要优于传统的渔情预报方法。

参考文献:

[1] 沈金鳌，方瑞生. 浙江近海冬汛带鱼渔获量预报方法的探讨[J]. 水产科技情报, 1982, 5: 1-5.

[2] 陈新军，郑波，李纲. GLM 和 GAM 模型研究东黄海鲐资源渔场与环境因子的关系[J]. 水产学报，2008, 32(3): 379-386.

[3] 陈新军. 西北太平洋柔鱼渔场与水温因子的关系[J]. 上海海洋大学学报, 1995, 3: 181-185.

[4] 陈新军，田思泉，钱卫国. 月相对北太平洋海域柔鱼钓获率的影响[J]. 海洋渔业, 2006, 28 (2): 136-140.

[5] 杜云艳，周成虎，崔海燕，等. 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.

[6] 袁红春，顾怡婷，汪金涛，等.西北太平洋柔鱼中长期预测方法研究[J]. 海洋科学, 2013, 37(10): 65-70.

[7] 张月霞，丘仲锋，伍玉梅，等. 基于案例推理的东海区鲐鱼中心渔场预报[J]. 海洋科学, 2009, 33(6): 8-11.

[8] Masahiko M, Yasuaki T. Vertical distribution and optimum temperature of bigeye tuna in the eastern tropical India Ocean based on deep tuna longline catches[J]. Journal of National Fisheries University, 1997, 46(1): 13-20.

[9] Mohri M, Hanamoto E, Takeuchi S. Optimum water temperatures for bigeye tuna in the Indian Ocean as seen from tuna longline catches[J]. Nippon Suisan Gakkaishi, 1996, 62: 761-764.

[10] Daniel G, Francis M. Comparative analysis of the exploitation of bigeye tuna in the Indian and eastern Atlantic oceans with emphasis on purse seine[R]. Victoria: IOTC Proceedings, 1999, 2: 158-171.

[11] Hiroaki O, Naozumi M, Takayuki M. GLM analyses for standardization of Japanese longline cpue for bigeye tuna in the indian ocean applying environmental factors[R]. Japan: IOTC Proceedings, 2001.

[12] Menard F, Marsac F, Bellier B, et al. Climatic oscillations and tuna catch rates in the Indian Ocean: a wavelet approach to time series analysis[J]. Fisheries Oceanography, 2007, 16(1): 95-104.

[13] 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: 24-26.

[14] 李军，李志凌，叶振江. 大眼金枪鱼渔业现状和生物学研究进展[J]. 齐鲁渔业, 2005, 22(12): 35-38.

[15] Herrera M L, Pierre M J.Review of the statistical data available for the tropical tuna species[R].Maldives: Indian Ocean Tuna Commission, 2011.

[16] Lin C R, Liu K H, Chen M S. Dual clustering: integrating data clustering over optimization and constraint domains[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(5): 628-637.

[17] Zhou J G, Guan J H, Li PX. A Dual clustering algorithm for distributed spatial databases[J]. Geo2spatial Information Science, 2007, 10(2): 137-144.

[18] 柳盛，吉根林. 空间聚类技术研究综述[J]. 南京师范大学学报(工程技术版), 2010, 10 (2): 57-62.

[19] 王家樵，朱国平，许柳雄. 基于 HSI 模型的印度洋大眼金枪鱼栖息地研究[J]. 海洋环境科学, 2009, 28(6): 739-742.

[20] 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, 15(2): 269-278.

[21] 胡振明. 利用栖息地适宜指数分析秘鲁外海茎柔鱼渔场分布[D]. 上海: 上海海洋大学, 2009.

[22] 杨胜龙，张禹，樊伟，等. 热带印度洋大眼金枪鱼渔场时空分布与温跃层关系[J]. 中国水产科学，2012, 19(4): 679-689.

[23] 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(4): 466-471.

[24] 陈雪冬，崔雪森. 卫星遥感在中东太平洋大眼金枪鱼渔场与环境关系的应用研究[J]. 遥感信息，2006, 1: 25-28.

[25] 杨胜龙，马军杰，伍玉梅，等. 印度洋大眼金枪鱼和黄鳍金枪鱼渔场水温垂直结构的季节变化[J]. 海洋科学, 2012, 36(7): 97-103.

A fishery forecasting model based on a spatial auto-regressive model and spatial clustering

YUAN Hong-chun[1], TAN Ming-xing[1], GU Yi-ting[1], CHEN Xin-jun[2]
(1. College of Information Technology; 2. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China)

Received: Jul., 5, 2014
Key words: *Thunnus obesus*; fishery forecasting; spatial auto-regressive model; spatial clustering; nonlinear regression

Abstract: In order to improve the predictive accuracy of pelagic fishing grounds, we have proposed a fishery-forecasting model based on a spatial autoregressive model and spatial clustering. In our model, the spatial autoregressive method is employed to first preprocess historical fishery data. Using the spatial clustering method, all data samples are then divided into several regions based on their geographical locations. By analyzing the mathematical relationships between environmental data and fishing data in the same region, a habitat suitability index model was built, with a follow-up experiment on bigeye tuna (*Thunnus obesus*) in the Indian Ocean. The results of the experiment showed that compared with the mean square error of 0.2363 in a traditional linear regression method, the model proposed in this paper had a mean square error of 0.1742. Therefore, our model can better demonstrate the relationship between marine environmental data and fishing quantity, and the predictive accuracy has been significantly improved.

(本文编辑: 谭雪静)