

基于差分自回归移动平均法预测苏格兰鲱鱼的迁移

王睿¹, 梁一迪², 张广运¹, 白羽^{3,4,5*}

¹北京建筑大学环境与能源工程学院, 北京

²北京建筑大学电气与信息工程学院, 北京

³北京建筑大学理学院, 北京

⁴北京建筑大学建筑结构与环境修复功能材料北京市重点实验室, 北京

⁵北京建筑大学大数据建模理论与技术研究所, 北京

Email: *buceaby@163.com

收稿日期: 2021年3月3日; 录用日期: 2021年5月10日; 发布日期: 2021年5月17日

摘要

作为苏格兰渔业经济支柱的鲱鱼, 正在随着全球变暖而向北迁移。为研究鲱鱼种群分布位置的变化, 把北大西洋苏格兰海域划分成28个区域, 将过去167年间捕鱼季的海洋表面温度视为时间序列, 建立了基于差分自回归移动平均法的海洋表面温度预测模型, 得到了未来50年海表温度变化情况。结合鲱鱼的适宜生存条件, 分析了鲱鱼的迁移路径及未来最有可能聚集的地点。研究结果为苏格兰地区的渔业经济管理与规划提供了参考和指导。

关键词

鲱鱼, 时间序列, 差分自回归移动平均法, 海洋表面温度

Prediction of Scottish Herring Migration Based on Autoregressive Integrated Moving Average Method

Rui Wang¹, Yidi Liang², Guangyun Zhang¹, Yu Bai^{3,4,5*}

¹School of Environment and Energy Engineering, Beijing University of Civil Engineering and Architecture, Beijing

²School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing

³School of Science, Beijing University of Civil Engineering and Architecture, Beijing

⁴Beijing Key Laboratory of Functional Materials for Building Structure and Environment Remediation, Beijing University of Civil Engineering and Architecture, Beijing

*通讯作者。

⁵Institute of Big Data Modeling and Technology, Beijing University of Civil Engineering and Architecture, Beijing
Email: *buceaby@163.com

Received: Mar. 3rd, 2021; accepted: May 10th, 2021; published: May 17th, 2021

Abstract

Herring, the mainstay of Scottish fishing economy, is moving north in response to global warming. The Scottish North Atlantic is divided into 28 regions to study the change of herring population distribution. The sea surface temperature (SST) of the fishing season in the past 167 years is taken as a time series and the prediction model of SST based on the Autoregressive Integrated Moving Average (ARIMA) method is established to get the SST changes in the next 50 years. The migration routes and the most likely aggregation sites of herring in the future are analyzed based on the optimal survival temperature of herring. The results provide reference and guidance for fishery economic management and planning in Scotland.

Keywords

Herring, Time Series, Autoregressive Integrated Moving Average Method, Sea Surface Temperature

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

鲱鱼是苏格兰渔业的主要捕获品种之一[1]。然而，随着全球海洋温度的升高，未来鲱鱼可能会向北迁移到温度更适宜自身繁衍生息的地方。研究海洋温度变化情况和鲱鱼的迁移路径，对渔业经济管理而言意义重大。

时间序列是指对某种统计指标依次按照时间顺序进行排列而形成的数值序列[2]。与其他的评估模型相比，通过分析时间序列构造的模型，不仅更简单，而且拟合效果也更好[3]。差分自回归移动平均法 (ARIMA) 是 George Box 和 Gwilym Jenkins 所创建的 Box-Jenkins (B-J) 方法[4]的进一步发展和改进[5]。该方法将预测对象随时间推移而形成的数据序列视为一个随机序列，通过数学模型近似描述这个序列，模型一旦被识别，就可以从时间序列的过去值及现在值预测出未来值[6]。由于 ARIMA 结合了一组复杂的参数，如自回归 (AR)、积分 (I) 和移动平均 (MA) [7]，它对未来趋势的预测较为精确。

此前已有一些预测海洋温度和鱼群迁移的研究，李昊达等[8]通过反演得到了过去 21 年的海洋表面温度 (SST) 间断数据，并以此为基础建模预测未来 50 年的 SST，此方法误差较大；范昊林等[9]建立了逐步灰色 GM (1, 1) 模型，处理后的温度数据中存在异常值，对后续预测造成了一定的干扰，不能准确表示出鱼群迁移的实际情况。

在前人研究的基础上，本文提取了过去 167 年鲱鱼捕鱼季的平均海洋表面温度，将其视为时间序列，通过 ARIMA 建立了 SST 预测模型。结合苏格兰鲱鱼的适宜生存条件，推测了未来 50 年鲱鱼最有可能出现的位置。

2. 数据处理

鲱鱼是浅层鱼类，在一段相当长的时间内，鲱鱼在经纬度上的水平迁移范围，远大于深度上的纵向迁移范围，因而海洋表层的水温变化是影响鲱鱼游动的主要因素。海洋表层温度数据来源于美国国家海洋和大气局在其网站上提供的 *NOAA ERSST V5 data* [10]数据集。该数据集给出了海表温度和经纬度、时间(月)的关系。

目前鲱鱼主要分布在欧洲西北部的大不列颠岛周围和挪威海海岸附近。根据鲱鱼的分布情况，把苏格兰附近的海洋按经度线和纬度线划分为网格状，一共得到了 28 个区域，如图 1 所示，每个区域选择一个观测点，观测点位置见表 1，并用这一观测点的 SST 代表此区域的温度。

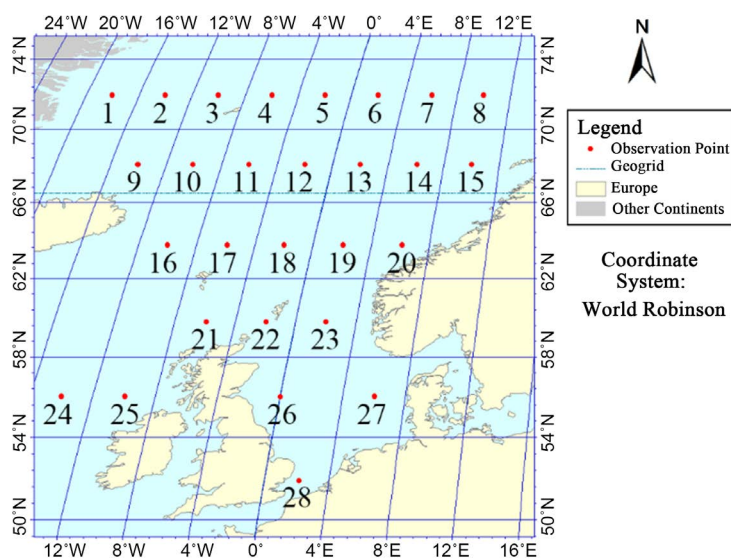


Figure 1. Distribution of observation points

图 1. 观测点分布图

Table 1. Position of observation points

表 1. 观测点的位置

序号	位置	序号	位置	序号	位置	序号	位置
1	(72°N, 18°W)	8	(72°N, 10°E)	15	(68°N, 10°E)	22	(60°N, 2°W)
2	(72°N, 14°W)	9	(68°N, 14°W)	16	(64°N, 10°W)	23	(60°N, 2°E)
3	(72°N, 10°W)	10	(68°N, 10°W)	17	(64°N, 6°W)	24	(56°N, 14°W)
4	(72°N, 6°W)	11	(68°N, 6°W)	18	(64°N, 2°W)	25	(56°N, 10°W)
5	(72°N, 2°W)	12	(68°N, 2°W)	19	(64°N, 2°E)	26	(56°N, 0°)
6	(72°N, 2°E)	13	(68°N, 2°E)	20	(64°N, 6°E)	27	(56°N, 6°E)
7	(72°N, 6°E)	14	(68°N, 6°E)	21	(60°N, 6°W)	28	(52°N, 2°E)

另一方面，虽然鲱鱼在整个生命阶段中进行着南北洄游，其资源量会发生季节性变化，但鲱鱼在捕鱼季的迁移情况与渔业经济有着更直接且紧密的联系。考虑到苏格兰鲱鱼的捕鱼季主要集中在每年的 6 月至 9 月，因此提取了 28 个观测点在 1854~2020 年间每一年 6~9 月的平均温度数据，并将每个观测点这 167 年间的 SST 视为一个时间序列。目标海域内标记捕鱼点的第 i 个 SST 序列数据编号为 $\{u_{it}\}$

($l=1,2,\dots,28;t=1,2,\dots,167$)。图 2 绘制出了各区域过去 167 年间的 SST 变化曲线,可以看出海洋表面温度的整体变化趋势呈现弱平稳性。

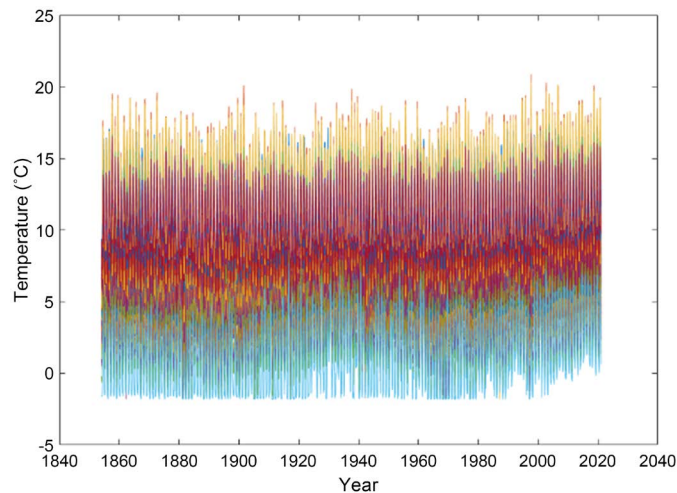


Figure 2. Sea surface temperature in past 167 years
图 2. 过去 167 年的海洋表面温度

为了便于分析,做出以下假设。

- 1) 假设北大西洋的水质将保持在一个相对稳定的区间内, 28 个不同区域的海水盐度差异不大。
- 2) 假设未来 50 年不会爆发大规模致死鱼类的传染病, 也不存在外来物种的入侵或天敌的影响。因为以上任何一种情况都可能导致鲱鱼的灭绝。
- 3) 假设鲱鱼在迁移过程中, 种群整体数量相对稳定, 不存在其他新的物种与鲱鱼构成捕食关系致使鲱鱼数量激增或锐减。

3. 海洋表面温度预测模型

一般地, 差分自回归移动平均法记作 ARIMA(p, d, q), 其中 p 为自回归项数, d 为时间序列成为平稳序列时所做的差分阶数, q 为移动平均项数[11]。ARIMA (p, d, q)的数学表达式为

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d y_t = c + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (1)$$

式中, y_t 是 t 时刻观测值; $L^i y_t = y_{t-i}$ 是滞后的时间序列, c 是常数项, $\phi_i (i=1,2,\dots,p)$ 是自回归系数, $\theta_i (i=1,2,\dots,q)$ 是移动平均系数, $\phi_p \theta_q \neq 0$, ε_t 是随机误差, 满足 $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$, $E(\varepsilon_s \varepsilon_t) = 0, \forall s < t$ 。为了使用该方法建立海洋表面温度预测的数学模型, 必须先确定 p, d, q 的值。

3.1. 平稳性检验

为了确定 d 的值, 还需要采用平滑法进一步验证 SST 时间序列的平稳性, 即对弱平稳时间序列进行差分。由于每进行一次差分运算, 都会损失掉一部分的有效信息, 所以对差分的阶数有所控制, 一般情况下不会超过二阶差分。

分别取 $l=1,2,\dots,28$, 对原始序列 $\{u_{lt}\}$ ($t=1,2,\dots,167$) 进行了一阶差分和二阶差分, 发现时间序列均呈现出了二阶差分波动幅度大, 一阶差分更为平稳的现象。比如, 任意取四个观测点 3、15、17 和 28, 它们的时间序列的一阶差分和二阶差分如图 3 所示。所以 ARIMA 选定一阶差分, 即 $d=1$ 。

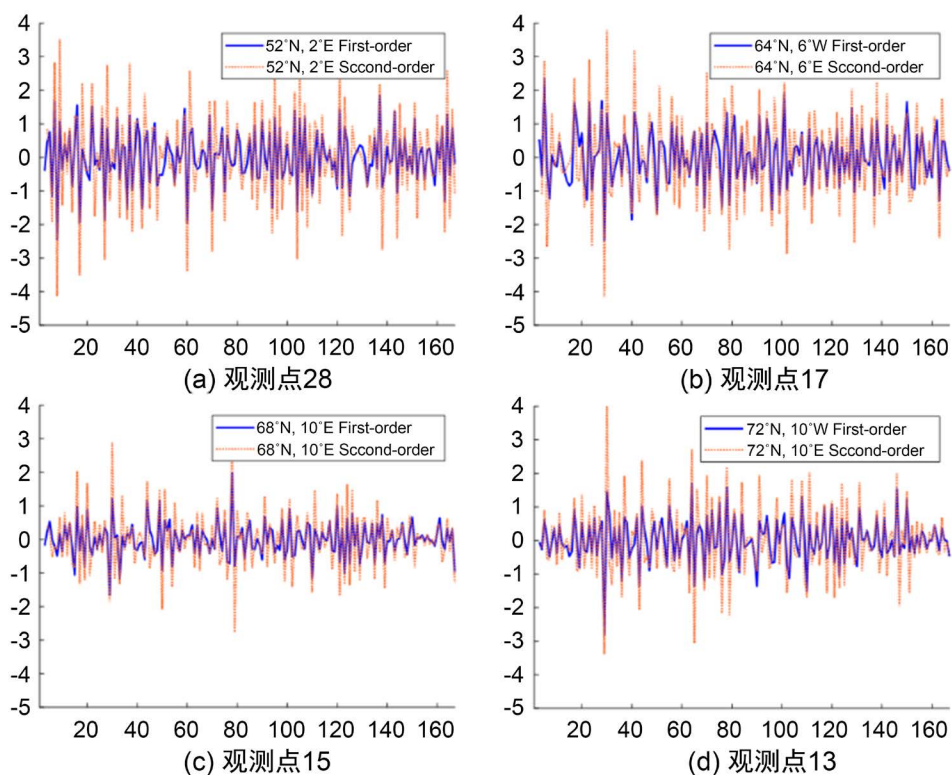


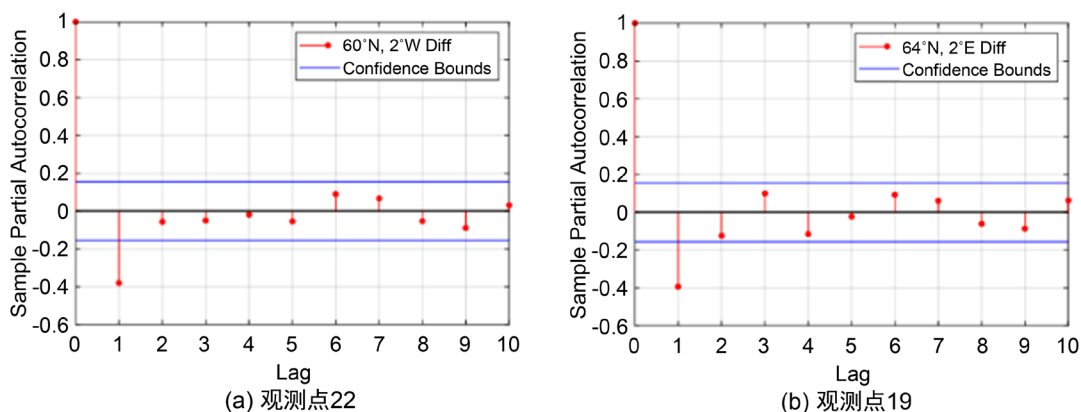
Figure 3. Time series of first-order difference and Time series of second-order difference
图 3. 一阶差分和二阶差分时间序列

3.2. 模型定阶

海洋表面温度预测模型中 q 通过观察自相关系数(ACF)确定, p 通过观察偏自相关系数(PACF)确定。利用差分后的数据生成 ACF 和 PACF 图像, 其中, 在四个观测点 8、18、19 和 22 的图像如图 4 和图 5 所示。不难发现, ACF 和 PACF 分别在最初的一阶和五阶内明显大于 2 倍标准差范围, 而后全部处于 2 倍标准差范围内。观察所有的观测点, 结果显示 95% 的观测点的 ACF 和 PACF 具有相同现象, 从而确定 $q = 1, p = 5$, 即海洋表面温度预测模型为 ARIMA (5, 1, 1), 任取 $l = 1, 2, \dots, 28$, 表达式为

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4 - \phi_5 L^5)(1 - L)u_t = c_l + (1 + \theta_1 L)\varepsilon_t \quad (2)$$

满足 $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_{\varepsilon}^2, E(\varepsilon_s \varepsilon_t) = 0, \forall s < t$ 。



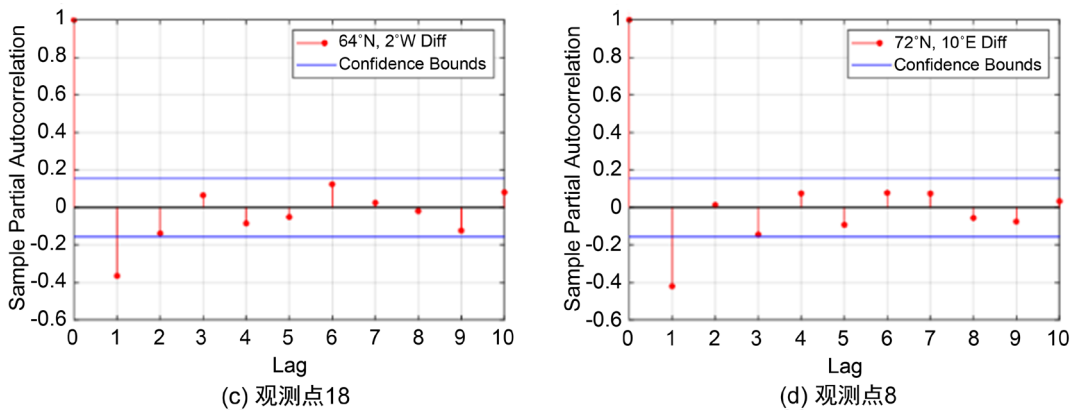


Figure 4. Sample autocorrelation function

图 4. 自相关系数

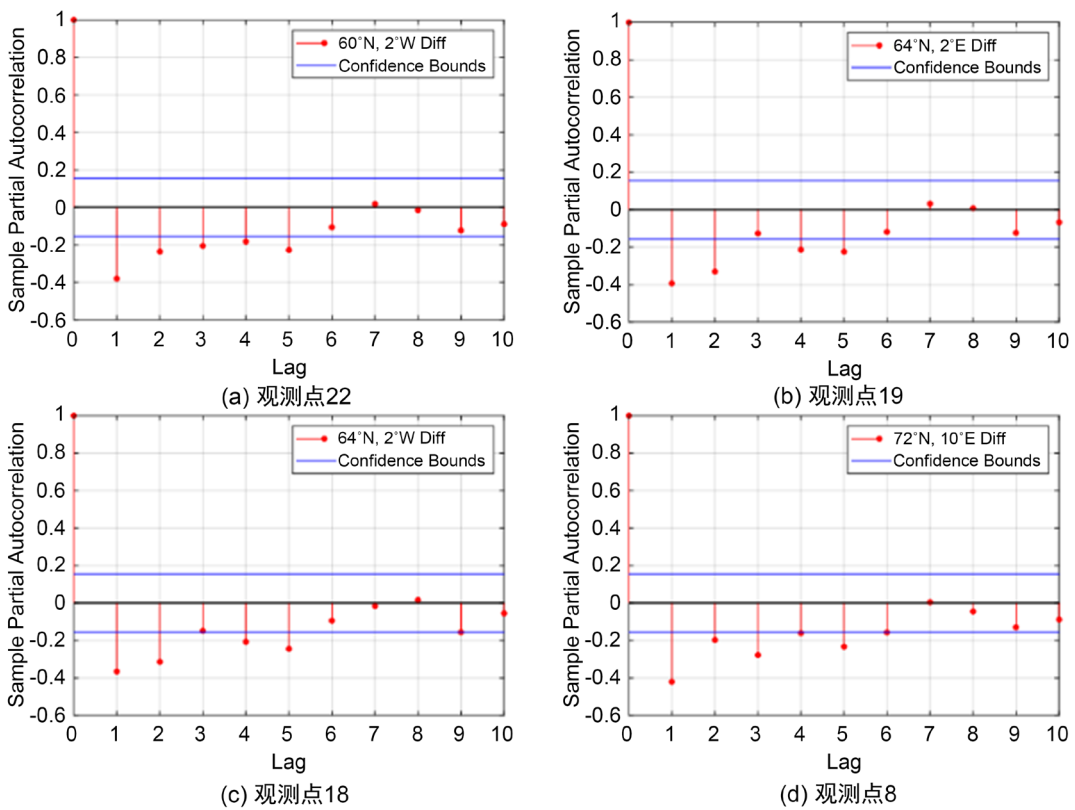


Figure 5. Sample partial autocorrelation function

图 5. 偏自相关系数

比如，在观测点 7 处，模型参数取值如表 2 所示，其中自回归系数 $\phi_i (i = 1, 2, \dots, 5)$ 和移动平均系数 θ_{7i} 对应的 P 值都比较小，说明了采用 ARIMA (5, 1, 1)模型是可行的。

Table 2. Model parameters of the 7th observation point

表 2. 观测点 7 处的模型参数

参数	值	标准差	t 统计值	P 值
c_7	0.0329	0.0418	0.7875	0.4310

Continued

ϕ_{71}	-0.4151	0.1740	-2.3855	0.0171
ϕ_{72}	-0.3716	0.1385	-2.6838	0.0073
ϕ_{73}	-0.3061	0.1130	-2.7081	0.0068
ϕ_{74}	-0.2625	0.0945	-2.7777	0.0055
ϕ_{75}	-0.2794	0.0778	-3.5909	0.0000
θ_{71}	-0.4371	0.1646	-2.6554	0.0079
σ_{7e}	0.7254	0.0665	10.9092	0.0000

3.3. 白噪声检验

上面建立的海洋表面温度预测模型为 ARIMA (5, 1, 1)。为判断模型建立的是否准确, 对残差序列进行白噪声检验。如果残差序列通过了白噪声检验, 就说明时间序列中有用的信息均提取完毕, 可以使用模型进行预测, 否则需要进一步提取信息。取横坐标为正态分布的分位数, 纵坐标为残差的分位数, 对每一个观测点画出残差分位数图。任取四个观测点 8、11、16 和 23, 它们的残差分位数图如图 6 所示。可见, 残差序列基本满足正态分布, 可以通过白噪声检验, 由此表明海表温度预测模型 ARIMA (5, 1, 1) 构建良好, 精度很高, 在此基础上进行的分析较为可靠。

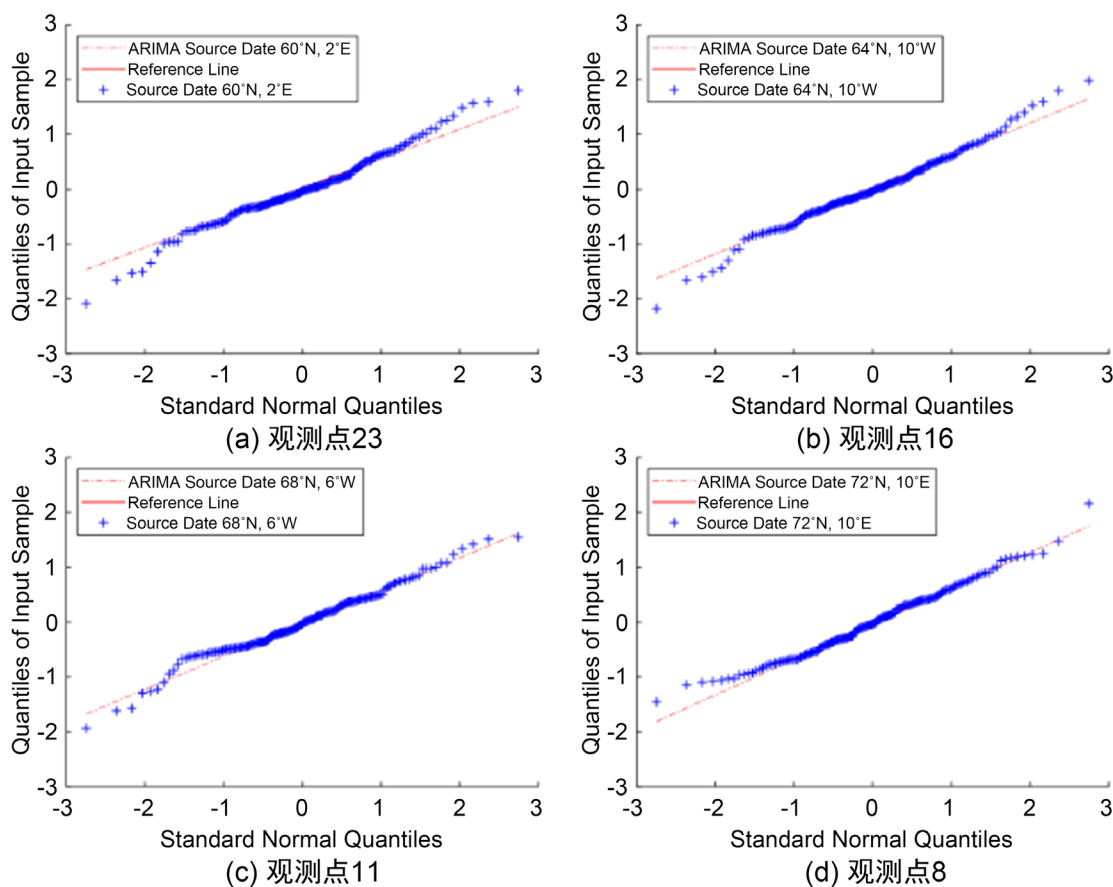


Figure 6. Residual quantile-quantile plot

图 6. 残差分位数

4. 鲱鱼迁移预测

4.1. 海洋表面温度变化情况

利用 ARIMA (5, 1, 1)模型对海洋表面温度进行预测，在未来 50 年内，28 个观测点的温度大部分呈现出了升高的趋势，往往随着纬度的升高，温度逐渐降低。部分观测点的温度变化如图 7 所示，50 年后 28 个观测点海洋表面温度热力图如图 8 所示。但是也有例外的点，比如在观测点 7 处，50 年后的温度要比现在的温度低，推测这可能是受到了洋流的影响。

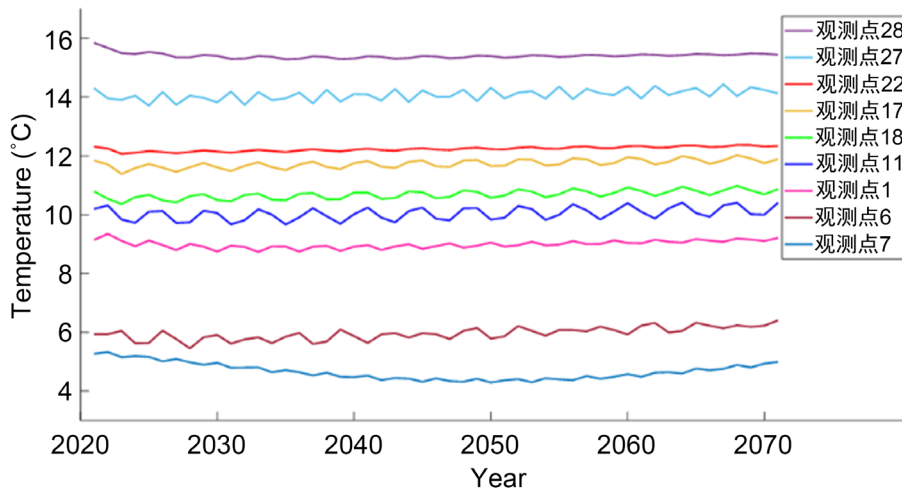


Figure 7. Sea surface temperature projections for the next 50 years
图 7. 未来 50 年海洋表面温度预测

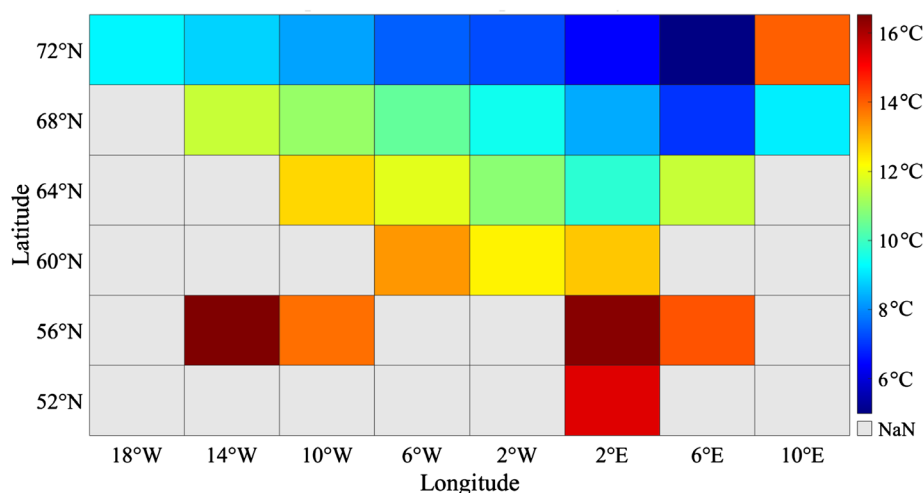


Figure 8. Heat map of sea surface temperature after 50 years
图 8. 50 年后海洋表面温度热力图

4.2. 鲱鱼迁移区域

根据全球鱼类信息系统的资料[12]，鲱鱼的最低生存温度为 1°C，最高生存温度为 18°C，虽然保持在这一生存范围内鲱鱼就可以存活下来，但是还是存在一个最佳的生存温度区间。考虑到自然界尤其是生物界中，大多数事物的发展与因果规律都符合正态分布，因而假定鲱鱼的最佳生存温度区间为 9°C~11°C。

根据预测的海洋表面温度未来 50 年内的变化情况, 28 个观测点中温度变化在 $9^{\circ}\text{C}\sim 11^{\circ}\text{C}$ 范围内的为观测点 1、10、11、12、15、18 和 19, 所以苏格兰鲱鱼将从目前的栖息地逐步迁移到最佳生存区域。鲱鱼的迁移路径最有可能为, 先向北移动到挪威海 $62^{\circ}\text{N}\sim 66^{\circ}\text{N}$ 、 $4^{\circ}\text{E}\sim 4^{\circ}\text{W}$ 区域, 然后一部分鲱鱼向 $66^{\circ}\text{N}\sim 70^{\circ}\text{N}$ 、 $8^{\circ}\text{E}\sim 12^{\circ}\text{E}$ 迁移, 另一部分鲱鱼向 $66^{\circ}\text{N}\sim 70^{\circ}\text{N}$ 、 $0^{\circ}\sim 12^{\circ}\text{W}$ 迁移, 并最终迁移至靠近格陵兰岛的 $70^{\circ}\text{N}\sim 74^{\circ}\text{N}$ 、 $16^{\circ}\text{W}\sim 20^{\circ}\text{W}$ 区域, 如图 9 所示。因此建议苏格兰的渔业公司将现有资产逐步转移到苏格兰港口的北岸, 从迁移路径可见未来那里的鱼类资源会更加丰富。

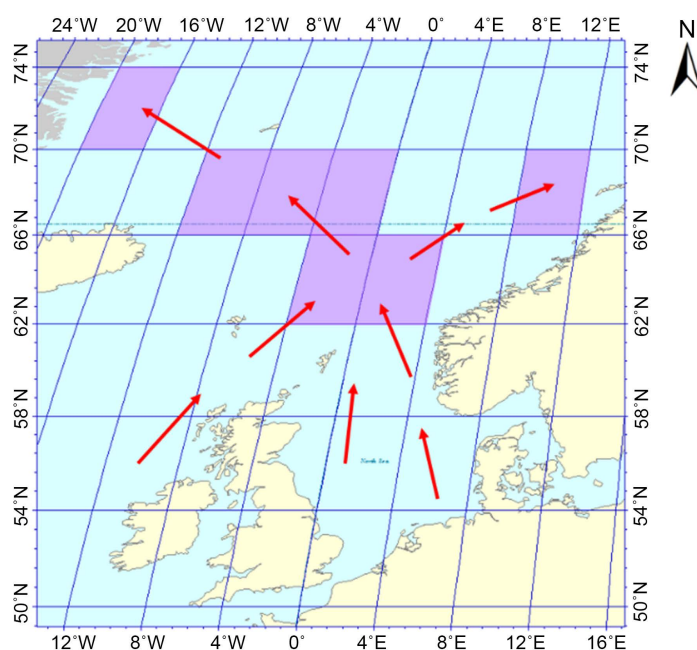


Figure 9. The migration route of herring

图 9. 鲱鱼迁移路径

5. 结论

1) 基于 28 个观测点在 1854~2018 年 6~9 月的平均海洋表面温度数据, 采用时间序列分析法建立了 ARIMA (5, 1, 1) 模型。预测海洋表面温度未来 50 年内整体呈现上升趋势, 且纬度越高, 温度越低。但也有些例外的观测点, 推测这些观测点的温度变化不符合一般规律可能是由于受到了洋流的影响。

2) 在足够长的时间内, 鲱鱼会有向北迁移到挪威海的趋势。根据鲱鱼的适宜生存温度区间和海洋温度变化情况, 预测了未来 50 年鲱鱼最有可能向 $66^{\circ}\text{N}\sim 70^{\circ}\text{N}$ 、 $8^{\circ}\text{E}\sim 12^{\circ}\text{E}$ 或 $66^{\circ}\text{N}\sim 70^{\circ}\text{N}$ 、 $0^{\circ}\sim 12^{\circ}\text{W}$ 迁移, 还有一部分鲱鱼会迁移至靠近格陵兰岛的 $70^{\circ}\text{N}\sim 74^{\circ}\text{N}$ 、 $16^{\circ}\text{W}\sim 20^{\circ}\text{W}$ 区域。

3) 海洋表面温度预测模型的建立, 是以过去 167 年的温度数据为基础的, 庞大的数据量能够支撑起模型的精确度和可靠性。模型的优势在于使用了差分自回归移动平均法, 从而有效消除 SST 预测中出现的随机波动。此模型具有一定的普世价值, 将其应用到鱼群迁移和渔业经济预测中, 有助于渔民更好的掌握鱼群分布情况以及未来的业务前景。

基金项目

北京市属高校基本科研业务费专项资金项目(X20142), 北京建筑大学教育科学研究项目(Y2015), 北京建筑大学研究生教学质量提升项目(J2020004)。

参考文献

- [1] 2020 COMAP's Mathematical Contest in Modeling.
<https://www.comap.com/undergraduate/contests/mcm/contests/2020/problems/>
- [2] 刘清华. 时间序列模型在水文预报中的应用[J]. 湖南水利水电, 2007(3): 42+67.
- [3] 孙秋霞, 王镇岭, 高齐圣. 我国人口时间序列拟合模型的比较与选择[J]. 数学的实践与认识, 2010, 40(2): 20-25.
- [4] Box, G.P.E. and Jenkins, G.M. (1978) *Times Series Analysis: Forecasting and Control*. San Francisco Press, San Francisco.
- [5] Brus, D.J. and de Gruijter, J.J. (1997) Random Sampling or Geostatistical Modelling? Choosing between Design-Based and Model-Based Sampling Strategies for Soil (with Discussion). *Geoderma*, **80**, 1-44.
[https://doi.org/10.1016/S0016-7061\(97\)00072-4](https://doi.org/10.1016/S0016-7061(97)00072-4)
- [6] 石美娟. ARIMA 模型在上海市全社会固定资产投资预测中的应用[J]. 数理统计与管理, 2005(1): 69-74.
- [7] Mbah, T.J., Ye, H.W., Zhang, J.H. and Long, M. (2021) Using LSTM and ARIMA to Simulate and Predict Limestone Price Variations. *Mining, Metallurgy & Exploration*, **38**, 913-926. <https://doi.org/10.1007/s42461-020-00362-y>
- [8] 李昊达, 王雪飞, 任咏琪. 基于 SST 反演的鱼类迁徙预测[J]. 江西水产科技, 2020(4): 3-5.
- [9] 范昊林, 苏柏豪, 杨文玉, 等. 基于灰色马尔可夫预测模型的鱼类活动位置研究[J]. 数码世界, 2020(4): 117.
- [10] NOAA Extended Reconstructed Sea Surface Temperature (SST) V5.
<https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v5.html>
- [11] 李标, 李雪梅, 古丽斯. ARIMA 模型在手足口病预测预警中的应用[J]. 中国卫生产业, 2014, 11(23): 26-27.
- [12] Whitehead, P.J.P. (1985) *FAO Species Catalogue. Vol. 7. Clupeoid Fishes of the World (Suborder Clupeoidei). An Annotated and Illustrated Catalogue of the Herrings, Sardines, Pilchards, Sprats, Shads, Anchovies and Wolf-Herrings. Part 1—Chirocentridae, Clupeidae and Pristigasteridae*. United Nations Development Programme, Rome, 470-471.