

上海海洋大学
新青年国际渔业组织人才培训计划
总结报告
(2023 年第 1 期)

姓名： 赵士祺

学校： 上海海洋大学

专业： 渔业发展

2023 年 7 月

数据科学在渔业管理中的简单应用—以 2019 年中西太平洋鲣鱼围网渔获分布与环境因子关系浅析为例

摘要:根据联合国粮农组织的预测, 2030 年水产品产量需要翻一番, 才能满足全球不断增长的需求; 合理的渔业管理有助于防止过度捕捞, 实现渔业资源的可持续利用。本文基于 R 语言、广义线性模型和线性方向平均值等科学工具、方法, 探究鲣鱼渔场重心变化, 并分析鲣鱼在不同季节对海洋环境的需求; 构建科学的环境因子适应性指数模型, 为中西太平洋鲣鱼资源管理提供科学依据。有结论如下: (1) 2019 年中西太平洋海域鲣鱼渔场重心整体向西北方向移动; 在北半球夏秋两季太阳直射点在北回归线到赤道地区移动时, 渔场重心出现明显上移; 到北半球春冬两季太阳直射点在南回归线到赤道移动时, 渔场重心出现明显下移。(2) 2019 年 1 月 CPUE 最高时的 SST 为 30°C; 2 月到 4 月 CPUE 最高时 SST 为 29-30°C, 5 月 CPUE 最高时的 SST 为 29.5-30°C, 6 月到 8 月则在 SST 为 29.5-30.5°C 海域; 9 月适宜 SST 有所下降为 30°C 左右, 10 月到 12 月 CPUE 最高时的 SST 为 30-30.8°C。(3) 2019 年 1 月到 5 月 CPUE 最高时的 CHL 为 0.125mg/m³ 左右, 6 月 CPUE 最高时的 SST 有所下降为 0.1mg/m³ 左右, 7 月 CPUE 最高时的 SST 有所上升为 0.127mg/m³ 左右, 8 月恢复为 0.125mg/m³ 左右, 9 月到 12 月 CPUE 最高时的 CHL 均在 0.06-0.125mg/m³。(4) 总结文中数据与前人研究, 可以认为 2019 年中西太平洋研究区域内海表面温度在 28°C-30.5°C 对鲣鱼的活动最具影响, 叶绿素浓度在 0.125mg/m³ 左右对鲣鱼活动最具影响。

关键词: R 语言; 栖息地模型; 地理信息科学; 广义线性模型; 科学渔业管理; 鲣鱼

0 引言

鲣鱼(*Katsuwonus pelamis*)属鲈形总目金枪鱼亚目金枪鱼科鲣属, 是一种暖水性上层洄游鱼类, 具有很高的经济价值; 它属远洋肉食性上层鱼类, 喜欢集群于冷暖水团交汇的水域, 广泛分布于太平洋、大西洋、印度洋的热带、亚热带以及亚寒带海域^[1]。近年来, 鲣鱼在中西太平洋的年产量约占世界总产量的 65% 以上^[2], 研究和掌握鲣鱼渔场分布及其影响因素对于中西太平洋渔业资源管理与保护具有重要意义。王易帆等^[3]和郭爱等^[4]在研究中发现海洋环境显著影响鲣鱼资源空间分布。唐浩等^[5]发现海表温度和叶绿素浓度是影响中西太平洋围网鲣鱼渔场的主要因子。汪金涛等^[6]认为当区域海表温度变化时鲣鱼渔场重

心也逐渐改变,渔场分布受气候环境变化的影响很大,与海表温度存在密切关系。本文根据 2019 年中西太平洋渔业管理委员(Western & Central Pacific Fisheries Commission, WCPFC)公约区域鲉鱼围网作业 $1^{\circ}\times 1^{\circ}$ 网格数据计算鲉鱼渔场重心的同时,结合海洋遥感产品,海表面温度(Sea Surface Temperature, SST)和海表面叶绿素浓度(Sea Surface Chlorophylla, Chl)数据,对鲉鱼渔业资源与海表面温度和海表面叶绿素浓度关系进行分析,为中西太平洋地区鲉鱼管理政策的制定提供科学依据。

1 材料与方法

1.1 研究区域

本文渔业数据来源于 2019 年 WCPFC 公约区域采集的网格化鱼捞数据,所以研究区域限制在 WCPFC 管理区域内,包括以南和以东划一条线为界的太平洋的所有水域,该线从澳大利亚南部海岸向南沿东经 141° 经线向南延伸,一直延伸到与南纬 55° 的交汇处;然后沿南纬 55° 向东,与东经 150° 子午线相交;然后沿着东经 150° 子午线向南,与南纬 60° 相交;然后沿南纬 60° 向东,与西经 130° 子午线相交;然后沿着西经 130° 子午线向北,与南纬 4° 相交;然后沿南纬 4° 向西,与西经 150° 子午线相交;然后沿着西经 150° 子午线向北,管理区域如图 1 所示。

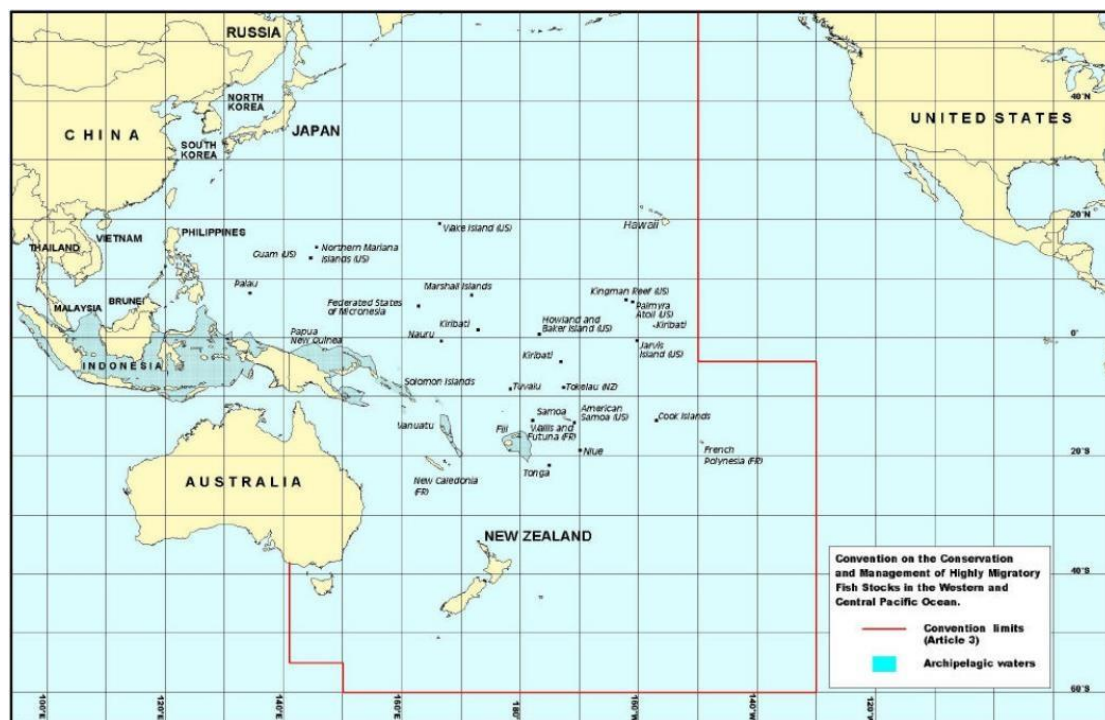


图 1 研究区域

1.2 研究数据

中西太平洋渔业管理委员会提供多种形式的渔业网格数据，这里选择 1950-2020 年的公约区域鲑鱼围网作业 $1^{\circ} \times 1^{\circ}$ 网格数据，时间分辨率选择精确到每月，国家标号选择无国家标号，下载数据后使用 R 语言对数据进行处理，筛选 2019 年的数据，将经纬度设置成标准格式，此处当经纬度数据跨越 180EW 线时，西经数据由于都是负值，用 360+西经，全部转换成正值。目的是为了保证数据空间的连续性，避免分析中出现异常，然后合并不同设备的鲑鱼产量，删除作业天数为 0 的行，根据作业天数和名义捕获量计算得到单位捕捞努力量渔获量(CPUE)，然后筛选需要的数据列组成数据结构准备输出，根据 1-12 月拆分数据，得到每月标准数据 csv 文件共 12 份。在完成数据构造循环的同时顺便计算渔场重心坐标点，单独导出为一份 csv。

海洋生产力网站提供多种的传感器数据，这里海表面温度(SST)与海表面叶绿素浓度(CHL)均选择 viirs 卫星传感器收集到的数据，所有数据均经历过官方的辐射和几何矫正，为二级遥感数据。使用 R 语言对得到的 hdf 数据进行处理，包括重投影坐标系为 WGS84、矩阵仿射变换匹配文件范围、去除-9999 以及导出为 Geotiff 数据等一系列操作，将数据导入 ArcGIS 和 SeaDas 与世界地图进行比对，遥感数据经历一系列操作后仍有完整元数据和较高配准精度，可以进行后续操作。

在完成 hdf 转 tif 后，使用得到的每月标准数据构建点图层数据，将 1-12 个月的 csv 数据导入 R 中，并将 csv 转化为 shp，然后与先前准备好的遥感数据进行数据匹配，投影坐标使用简易圆柱投影 Plate_Carree，中央经线设置为 180EW，使用 R 语言中 raster::extract() 函数提取捕捞位置对应的环境数据，并检查采样后的生产数据和环境数据去除异常值，最后按月将属性表导出为 csv。每份 csv 包含鲑鱼的作业时间、经纬度、捕捞产量、海表面温度(SST)、海表面叶绿素浓度(CHL)和单位捕捞努力量渔获量(CPUE)。

1.3 编写 R 语言程序

本文采用面向过程的编程范式实现代码，通过编写多个可复用模块，尽可能的实现研究过程的自动化。dataPre 模块的功能是鱼捞网格数据预处理模块，包括筛选年、按月份切割文件、经纬度标准化以及计算渔场重心等，通过 dataPre() 函数调用的方式向外暴露接口。csv2shp 模块的功能是根据 csv 文件提供的坐标构建点 shp 文件，通过调用 csv2shp() 函数的方式可以快速构建矢量建采样点。hdf2tif 模块的作用是将 hdf 环境遥感数据批量转化为 tif 数据，调用 hdf2tif() 函数导出 tif 时，默认会进行重投影、范围匹配和-9999.0 值的去除。hdfDraw 模块是由一组相互调用的函数组成，调用 hdfDraw() 可以进行 hdf4 遥感文件的预览，函数在

执行时会忽略-9999.0 值，并且提供高值缺失渲染支持，在部分情况下可以更好的渲染遥感图像的像元值。pointSampling 模块可以自动进行环境因子的点采样，代替繁琐的 ArcGIS 操作，pointSampling()调用一次仅能采集一种环境因子，尚不支持一次多采样。correlationTest 接口模块通过加载 correlationTest_SST 模块和 correlationTest_CHL 模块获得 correlationTest_SST()和 correlationTest_CHL(),两个函数可分别使用 K-S 方法进行 SST 和 CHL 与 CPUE 关系的显著性验证。SI 模块提供 SI_sst()函数和 SI_chl()函数，可以分别利用正态分布模型对 SST 和 CHL 两种环境因子进行曲线拟合，拟合得到的方程为高斯函数的一维形式。最后，main 模块为主模块，所有依赖包的加载、所有子模块的加载、以及大部分函数调用在此处编写。

程序遵守 MIT 开源协议，Github 地址为：<https://github.com/redworld123/SHOUCourse>。

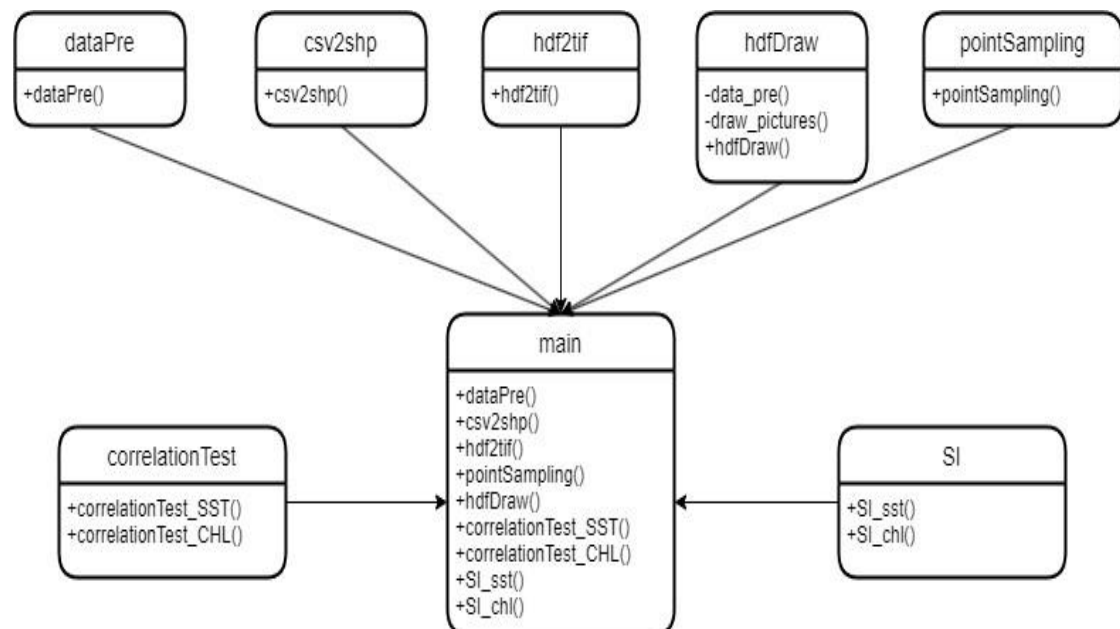


图 2 R 语言程序结构图

1.4 渔场重心

渔获量的分布实际上代表了渔业资源量的空间分布和变化，渔业上常采用渔场重心描述渔场空间位置的变动，其表达式如下所示，X、Y 分别为渔场重心的经度和纬度，Xi 为渔区中心点的经度，Yi 为渔区中心点的纬度，wi为渔区 i 的 CPUE，n 为渔区总个数。

$$X = \sum_i^n \frac{w_i \times X_i}{w_i}; Y = \sum_i^n \frac{w_i \times Y_i}{w_i}$$

1.5 单位捕捞努力量渔获量

单位捕捞努力量渔获量(Catch Per Unit Effort, CPUE)定义为每艘船每天的捕捞产量。随着科技的不断发展,捕捞努力量也在不断增加,因此渔获量也呈现逐年增加的趋势,但是这并不能反映出鳀鱼资源相对丰度的真实情况。而 CPUE 主要是反映单位时间内所获得的渔获量,因此即使捕捞努力量有所提升,CPUE 也不会一直保持增长,而是会呈现波动变化,这能够较真实地反映鳀鱼资源相对丰度的真实情况。

1.6 适应性指数模型

尽管鳀鱼是一种游泳速度较快的鱼类,有着较大的洄游范围,但其资源相对丰度仍然主要是受到周边海洋环境变化的影响,导致其补充量所必须的栖息范围等生存条件发生变化,这直接导致了该年整体资源量发生相应的变化。因此,通过研究环境因子来预测鳀鱼的资源相对丰度有着科学依据,同时也有实际指导意义。利用正态分布模型对 SST 和 CHL 两种环境因子进行曲线拟合,各月利用捕捞努力量与 SST 和 CHL 建立适应性指数(Suitability index, SI)模型如下。 SI_{CHL} 和 SI_{SST} 分别为 CHL 和 SST 的适应性指数, A、B、C 和 D 为常系数, CHL 和 SST 为分阶后的各月 CHL 值和 SST 值。

$$SI_{CHL} = e^{-A(\ln(CHL)-B)^2}$$

$$SI_{SST} = e^{-C(SST-D)^2}$$

2 结果

2.1 渔场中心分布

渔获量的分布实际上代表了渔业资源量的空间分布和变化,渔业上常采用渔场重心描述渔场空间位置的变动。将渔场重心文件导入 ArcGIS 生成点数据,将点转化为线,使用空间统计工具 Linear Directional Mean 度量线性方向平均值;渔场重心位移线性方向平均值是位移轨迹长度和方向的平均值,该值可从整体度量渔场重心位移特征,结果如图 3 所示。由图可知 2019 年鳀鱼渔场重心整体随时间变化向西北方向移动,在北半球夏秋两季太阳直射点在北回归线到赤道地区移动时,渔场重心出现明显上移,到北半球春冬两季太阳直射点在南回归线到赤道移动时,渔场重心出现明显下移,初步假设鳀鱼鱼群受到海表面温度影像较大,会向吸收太阳辐射较多海域移动。对于海表面叶绿素浓度对鳀鱼鱼群影响则需要 ECDF 函数进一步验证。

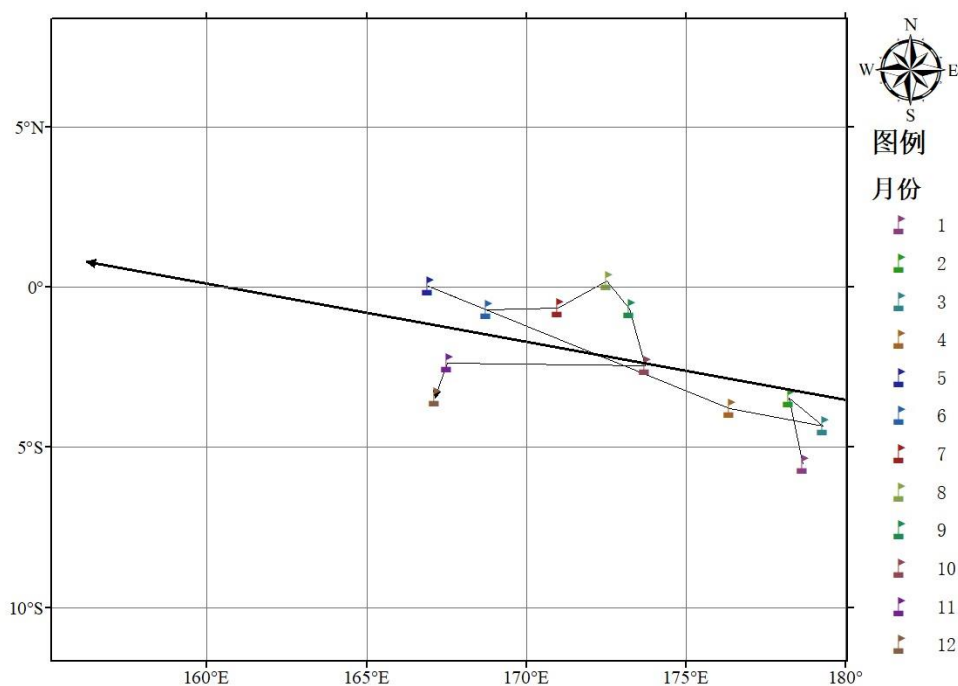
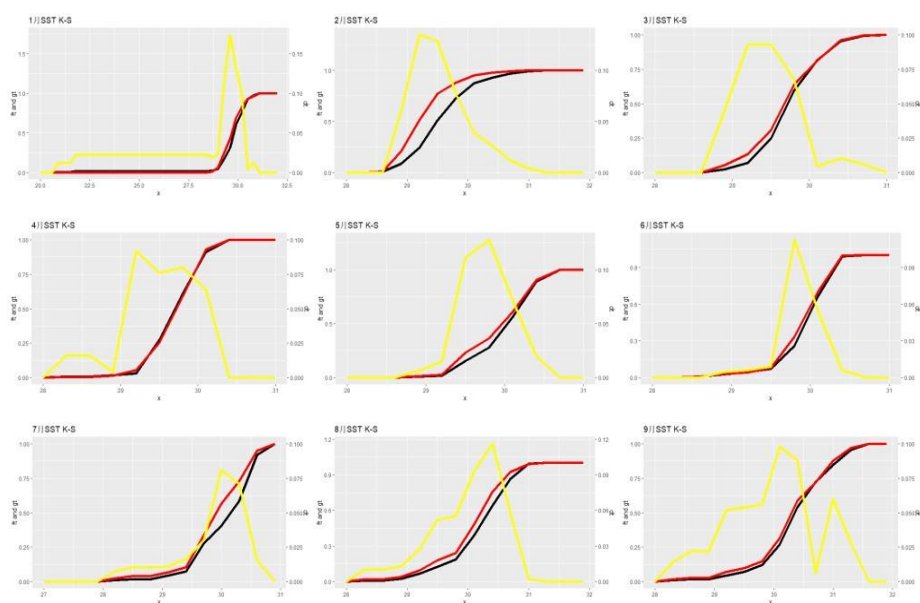


图 3 渔场重心

2.2 捕捞努力量与海表面温度适应性指数模型

使用 R 语言对各月 SST 进行 K-S 检验，由图 4 可知，不同月份最高 CPUE 所对应的 SST 有所差异，1 月 CPUE 最高时的 SST 为 30°C，2 月、3 月和 4 月 CPUE 最高时 SST 为 29-30°C，5 月 CPUE 最高时的 SST 为 29.5-30°C，6 月、7 月和 8 月则在 SST 为 29.5-30.5°C 海域；9 月适宜 SST 有所下降为 30°C 左右，10 月、11 月和 12 月 CPUE 最高时的 SST 为 30-30.8°C。



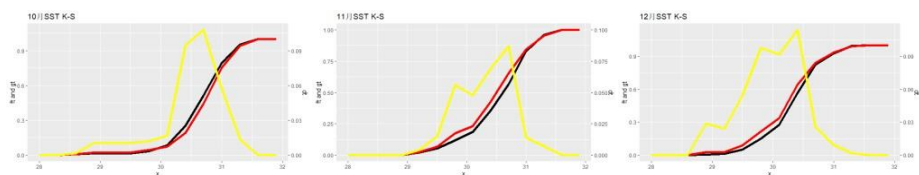


图 4 R 语言对 2019 年各月 SST 进行 K-S 检验

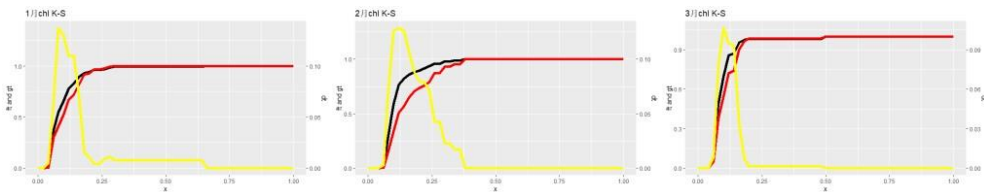
利用 CPUE 与 SST 建立适应性指数模型, CPUE 最高的海域通常为鲢鱼最适应的海域, 该海域 SI 的值为 1, CPUE 最少或为 0 的海域为鲢鱼最不适应的海域, 该海域 SI 值为 0, 结果如表 1 所示。

表 1 基于 CPUE 和 SST 的 SI 曲线函数

月份	模型	P 值
1	$I_{sst} = \exp[-\frac{(t_{sst} - 29.6)^2}{0.6^2}]$	<0.001
2	$I_{sst} = \exp[-\frac{(t_{sst} - 27.9)^2}{1.3^2}]$	<0.001
3	$I_{sst} = \exp[-\frac{(t_{sst} - 29.6)^2}{0.7^2}]$	<0.001
4	$I_{sst} = \exp[-\frac{(t_{sst} - 30.2)^2}{0.8^2}]$	<0.001
5	$I_{sst} = \exp[-\frac{(t_{sst} - 30)^2}{0.6^2}]$	<0.001
6	$I_{sst} = \exp[-\frac{(t_{sst} - 29.9)^2}{0.4^2}]$	<0.001
7	$I_{sst} = \exp[-\frac{(t_{sst} - 31.3)^2}{1.8^2}]$	<0.001
8	$I_{sst} = \exp[-\frac{(t_{sst} - 30.1)^2}{0.9^2}]$	<0.001
9	$I_{sst} = \exp[-\frac{(t_{sst} - 31.1)^2}{1.7^2}]$	<0.001
10	$I_{sst} = \exp[-\frac{(t_{sst} - 30.8)^2}{0.6^2}]$	<0.001
11	$I_{sst} = \exp[-\frac{(t_{sst} - 30.7)^2}{0.5^2}]$	<0.001
12	$I_{sst} = \exp[-\frac{(t_{sst} - 30.2)^2}{0.8^2}]$	<0.001

2.3 捕捞努力量与叶绿素浓度适应性指数模型

使用 R 语言对各月 CHL 进行 K-S 检验, 由图 5 可知, 不同月份最高 CPUE 所对应的 CHL 有所差异, 1 月、2 月、3 月、4 月和 5 月 CPUE 最高时的 CHL 为 0.125mg/m³ 左右, 6 月 CPUE 最高时的 CHL 有所下降为 0.1mg/m³ 左右, 7 月 CPUE 最高时的 CHL 有所上升为 0.127mg/m³ 左右, 8 月恢复为 0.125mg/m³ 左右, 9 月、10 月、11 月和 12 月 CPUE 最高时的 CHL 均在 0.06-0.125mg/m³。



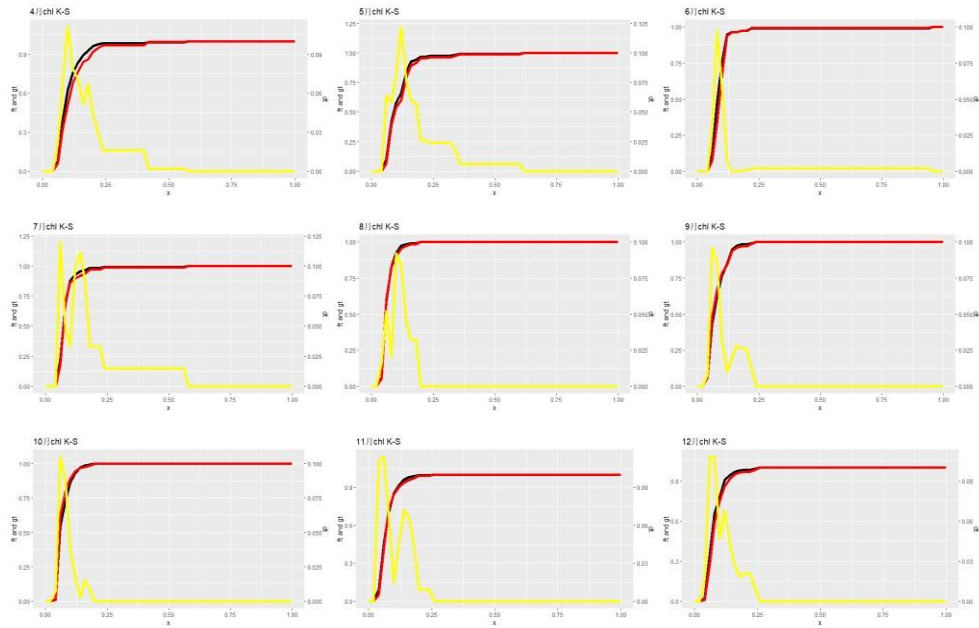


图 5 R 语言对 2019 年各月 CHL 进行 K-S 检验

鲢鱼广泛分布于各大洋热带和亚热带海域，其中以中西太平洋资源量最为丰富。综合评价环境因子对鲢鱼资源量的影响，构建科学的资源预报模型可为持续合理开发该鱼种提供参考。本文利用 2019 年中西太平洋渔获量数据，利用 CPUE 与 CHL 建立适应性指数模型，结果如表 2 所示。

表 2 基于 CPUE 和 CHL 的 SI 曲线函数

月份	模型	P 值
1	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.6)^2}{0.6^2}\right]$	<0.001
2	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.5)^2}{1.2^2}\right]$	<0.001
3	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.5)^2}{0.7^2}\right]$	<0.001
4	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.4)^2}{0.9^2}\right]$	<0.001
5	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.5)^2}{0.5^2}\right]$	<0.001
6	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.6)^2}{0.3^2}\right]$	<0.001
7	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.9)^2}{0.6^2}\right]$	<0.001
8	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 3.2)^2}{0.6^2}\right]$	<0.001
9	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 3.3)^2}{0.9^2}\right]$	<0.001
10	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 3)^2}{0.6^2}\right]$	<0.001
11	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 3)^2}{0.7^2}\right]$	<0.001
12	$I_{chl} = \exp\left[-\frac{(\ln(t_{chl}) + 2.7)^2}{0.6^2}\right]$	<0.001

3 结论与讨论

鲣鱼在洄游过程中具有其自身生物学特性,会受到栖息海域环境影响,如海表温度和叶绿素浓度等^[7]。在中西太平洋热带水域,海表温度常年在 28℃以上,是世界大洋中表面温度最高的海域之一,形成西太平洋暖池^[8]。在信风作用下,东太平洋会产生巨大涌升流,从而会形成具有低温、高盐等特性的冷舌区域。暖池和冷舌的交汇区拥有丰富的浮游植物和微型的浮游动物,是鲣鱼理想的索饵场^[9],因此 SST 和 CHL 是影响鲣鱼渔场分布的重要环境因子。本次研究通过 K-S 方法进行了显著性验证,得出不同月份最高 CPUE 所对应的 SST 和 CHL 值有所差异,但总结本文数据结合过往文献^[10],可以推导出 2019 年中西太平洋研究区域内海表面温度在 28℃-30.5℃对鲣鱼的活动影响最大,叶绿素浓度在 0.125mg/m³左右对鲣鱼活动影响最大,在确认海洋环境因子和产量之间存在正态分布关系后,利用环境影响因子建立了数据模型,但算法在分节值选择和高低值舍去时,仍然需要通过人工辅助,自动计算在部分月份存在较大差异。本次研究编写了大量的 R 语言脚本,基本实现了研究过程自动化,节省了在数据的转化和批量操作上浪费的时间,希望通过数据科学、地理信息科学和统计科学的应用为渔业管理政策的制定提供较为合理的依据。

参考文献

- [1] COLLETTE B, NAUEN C. FAO species catalogue. Volume 2. Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date[R]. Rome: FAO Fisheries Synopsis, 1983: 83-86.
- [2] 靳少非, 樊伟. 鲣鱼资源开发利用研究现状及未来气候变化背景下研究展望[J]. 渔业信息与战略, 2014, 29(4): 272-279.
- [3] 王易帆, 陈新军. 中西太平洋鲣鱼栖息地指数预报模型比较研究[J]. 上海海洋大学学报, 2017, 26(5): 743-750.
- [4] 郭爱, 陈新军. 利用水温垂直结构研究中西太平洋鲣鱼栖息地指数[J]. 海洋渔业, 2009, 31(1): 1-9.
- [5] 唐浩, 许柳雄, 陈新军, 等. 基于 GAM 模型研究时空及环境因子对中西太平洋鲣鱼渔场的影响[J]. 海洋环境科学, 2013, 32(4): 518-522.
- [6] 汪金涛, 陈新军. 中西太平洋鲣鱼渔场的重心变化及其预测模型建立[J]. 中国海洋大学学报, 2013, 43(8): 44-48.
- [7] SUND P N, BLACKBURN M, WILLIAM F. Tunas and their environment in the Pacific Ocean: a review[J]. Oceanography and Marine Biology Annual Review, 1981(19): 443-512.
- [8] 王学昉, 许柳雄, 朱国平. 鲣鱼(Katsuwonus pelamis)生物学研究进展[J]. 生物学杂志, 2009, 26(6): 68-71.
- [9] 胡奎伟, 朱国平, 王学昉, 等. 中西太平洋鲣鱼丰度的时空分布及其与表温的关系[J]. 海洋渔业, 2011, 33(4): 417-422.
- [10] 杨胜龙, 周甦芳, 周为峰, 等. 基于 Argo 数据的中西太平洋鲣鱼渔获量与水温、表层盐度关系的初步研究[J]. 大连水产学院学报, 2010, 25(1): 34-40.