

一种基于时间序列分析的赤潮预测新方法研究 ——以浙江海域为例

徐丽丽^{1,2}, 余骏³, 高鑫鑫^{1,2}, 车助镁³, 何雯^{1,2}, 邱婷^{1,2}

(1. 国家海洋局东海预报中心, 上海 200081; 2. 海洋生态监测与修复技术重点实验室, 上海 200081; 3. 浙江省海洋监测预报中心, 浙江 杭州 211800)

摘要: 根据1989—2018年《中国海洋灾害公报》及浙江省海洋监测预报中心提供的1981—2019年监测数据, 统计分析了浙江海域赤潮的发生次数长期变化和季月波动特征。基于月发生次数的时间序列, 建立了一种以数据驱动为核心、简易实用的赤潮预测模型, 并对2019年进行后报检验。结果表明: (1) 近40 a浙江海域赤潮发生次数呈“低频发生-爆发增长-缓降波动”3段式年际变化规律, 同时97.5%集中爆发在春夏季, 其中5月爆发次数最多; (2) 通过数据差分处理和参数检验, 最终确定的最优时间序列预测模型为 $ARIMA(1, 1, 1)(1, 1, 0)_{12}$, 且残差通过白噪声检验, 拟合优度系数为0.68, 绝对误差控制在两次以内的样本占总数的84%, 具有可预测性; (3) 模型经浙江海域2019年逐月赤潮后报预测检验, 其年相对误差为14%, 月绝对误差在两次以内, 预测结果理想。在目前赤潮发生机制尚不完全清楚的情况下, 利用ARIMA模型进行赤潮发生次数的年际和季节特征预测是一种经济且有效的途径, 也是赤潮灾害年度趋势预测方法和业务化预警技术的新尝试和有效补充。

关键词: 浙江海域; 赤潮发生次数; ARIMA模型; 年度趋势预测

中图分类号: X55 **文献标识码:** A **文章编号:** 1003-0239(2020)05-0095-09

1 引言

赤潮又称红潮, 国际上也称为“有害藻华”或“红色幽灵”。赤潮已成为当今世界普遍关注的海洋生态问题^[1]。我国于2005年制定了《赤潮海洋灾害应急预案》并实施至今, 赤潮灾害预测及评估工作被列为政府工作报告内容之一; 各级海洋减灾相关部门和预报业务机构建立了赤潮年度预测和月及周会商业务制度, 作为我国政府赤潮灾害应急决策和处置的工作依据和技术支撑。我国浙江海域受南下长江冲淡水、北上台湾暖流、钱塘江和甬江等径流注入以及外海上升流系的综合影响, 其环境极有利于赤潮生物的生长繁殖。国内最早的赤潮记录是1933年费鸿年记载的浙江镇海至台州-石浦

一带的夜光藻和骨条藻赤潮^[2]。根据1989—2019年《浙江省海洋环境公报》和相关文献资料^[3], 浙江海域在1981—2019年间共记录到赤潮事件744次, 累计面积超过 $1.74 \times 10^6 \text{ km}^2$ 。尤其进入21世纪以来, 超过上千平方公里的浙江海域内, 大型赤潮和有毒赤潮都呈明显增长趋势。这不但对人民生命健康与海洋生态环境造成了极大威胁, 同时也成为浙江省海洋经济高质量和可持续发展的一个重要制约因素, 因此引起我国学者和各级政府的高度关注。

至今, 学术界已对赤潮发生过程的模拟预报和赤潮灾害趋势或概率预测等做了大量研究。但因赤潮爆发原因复杂, 各藻种及种间发展的生态机制尚不完全清楚, 现有的预测研究从开始的主要基于气象水文等因子的定性分析和条件预报^[4-5], 逐渐发

收稿日期: 2020-03-12; 修回日期: 2020-05-14。

基金项目: 国家重点研发计划(2016YFC1401906)。

作者简介: 徐丽丽(1984-), 女, 工程师, 硕士, 主要从事海洋生态灾害和海洋环境预警研究。E-mail: xulili@ecs.mnr.gov.cn

展到利用连续赤潮现场监测或遥感数据^[6-8]、基于主成分和多元回归等统计模型做预测^[9-10]。但在实际业务应用中,这些方法受限于样本数据、变量因子的敏感性等因素,从而影响预报精度乃至难以做出有效预测。随着监测数据的丰富和某些赤潮种生态规律机制研究的深入,部分学者开始利用物理-化学-生物耦合的生态动力学数值模型进行赤潮模拟研究^[11-17]。如夏综万等^[16]考虑生物动力学和环境动力学因素,建立了大鹏湾夜光藻赤潮生态仿真模型;李雁宾^[17]对长江口及邻近海域季节性的赤潮生消过程控制机理进行了研究。生态动力学数值模型可以对机理明确、变量初始值及边界条件来源详尽且连续的局部海域做某个具体赤潮过程的模拟预测,但在以年为尺度的业务化预测上的应用效果有限。近年来,以神经网络和深度学习为代表的大数据预报技术飞速发展,其在非线性模式识别方面具有独特的信息处理和解算能力,非常适用于赤潮这种机制尚不清楚的高维非线性系统^[18-19],但其对数据要求量大、质量严格且建模过程复杂。以上是赤潮预测研究取得的众多成果,在业务化应用中也取得一定成效。但由于赤潮爆发受到水文气象条件、海水理化因子变化以及船舶带来的外来浮游物种入侵等众多因素的影响,加之赤潮生态系统各因子间表现出的高度非线性和不确定性,以及连续监测数据获取困难等问题,目前在一线业务预报机构中能真正使用,且满足实际业务工作需求的赤潮预报工具箱的选择还不够多。尤其是针对东海区浙江海域的业务化赤潮预测研究主要以定性分析为主^[20-23],业务一线应用的定量预测方法基本空白。

为满足自然资源部东海分局及浙江省监测预报中心赤潮年度预测和月会商的业务需求,并规避数据限制和机理研究要求,本文拟基于1981—2018年浙江海域赤潮月发生次数构建时间序列;考虑赤潮长期的年代变化特征又兼顾其季节性生态变化规律,提出一种基于自回归移动平均(Auto Regressive Integrated Moving Average, ARIMA)模型的时间序列模型,以及易于实现且便于使用的赤潮预测新方法,为浙江海域的赤潮灾害年度预测、业务化赤潮预警和灾害评估工作提供新的技术工具。

2 数据与方法

2.1 数据来源

赤潮原始数据来源于自然资源部海洋预警监测司发布的1989—2019年《中国海洋灾害公报》及浙江省海洋监测预报中心提供的1981—2019年的监测数据。具体参数包含赤潮发生时间、发生海域、经纬度、最大记录面积、分布形态、优势藻种、密度和水色,时间跨度为1981—2019年。因赤潮发生具有年际、季节和月等不同时间尺度变化,如果直接利用原始时间序列数据构造模型,容易因非平稳特性产生虚假回归,因此整理并建立38a浙江赤潮发生频率的时间序列后,先对其进行平稳化处理。

2.2 分析方法

时间序列分析法是根据一组相依有序的离散数据,建立反映时间序列中所包含的动态依存关系的数学模型,并进行未来状态预测^[24]。对于非平稳时间序列,主要运用ARIMA模型,亦称Box-Jenkins模型。指定3个参数,即描述自回归阶数(p)、差分次数(d)和移动平均阶数(q),模型通常被写作ARIMA(p,d,q)。其数学表达式为:

$$\varphi_p(1-B)^d Z_t = \theta_q(B) a_t \quad (1)$$

式中: Z_t 为原序列; a_t 为白噪声序列, 是一列相互之间无关、均值为0、误差的方差为 σ^2 的随机变量序列; B 为后移算子即 $BZ_t = Z_{t-1}$; φ_p 为自回归算子, $\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)$, P 为模型的自回归除数; θ_q 为移动平均算子, $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, q 为模型的移动平均阶数; θ_0 为参数, $\theta_0 = \mu(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$, μ 为平均数。

建立模型前需根据时间序列的特性(平稳性、非平稳性和季节性)确定建模类型。若序列非平稳且有季节性,则模型函数被记作ARIMA(p,d,q)(P,D,Q)_s。它可以用于分析不仅含有季节性成分、还混有非季节性成分的时间序列资料。其中(p,d,q)和(P,D,Q)分别为非季节性和季节性自回归(Auto Regressive, AR)、差分(I)和移动平均(Moving Average, MA)的阶数,s代表季节周期。本文中模型

的原始序列平稳化、模型参数估计、模型诊断和预测均借助 SPSS 25.0 统计分析软件,采用编程法(语句)分析处理。建模过程见图 1。分 3 个关键阶段^[25]:(1)模型参数的确定。利用自相关和偏自相关分析时间序列原始数据的随机性、平稳性和季节性,初步确定模型参数 p 、 d 、 q 及 P 、 D 、 Q 的取值;(2)模型参数的检验。首先采用 Box-Ljung 检验模型白噪声判断模型的拟合优度;若有两个或两个以上的模型通过检验,则根据贝叶斯信息准则(Bayesian Information Criterion, BIC)判断模型类型和最优阶次^[26],具体算法见式(2)和(3);(3)预测应用。通过对比模型预测值与实际值的差值,评价模型预测的准确性。

$$\sigma^2 = \left[\sum_{i=1}^N (x_i - \bar{x}_i)^2 / N \right] \quad (2)$$

$$BIC = \ln(\sigma^2 + \frac{2k \ln N}{N}) \quad (3)$$

式中: N 为样本个数, k 为参数估计的数量, σ^2 为误差的方差。在 ARIMA 模型中取 BIC 最小值作为

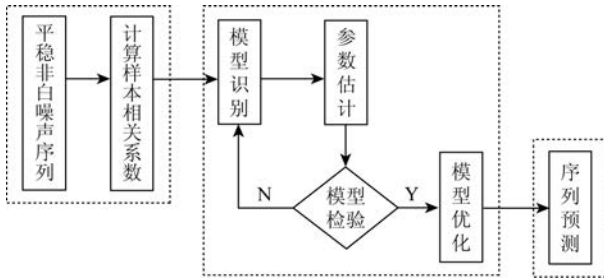


图1 时间序列模型建模具体流程

评价指标确定 p 、 q 以及 P 、 Q 值。

3 分析与讨论

3.1 序列平稳化处理

根据 1981—2019 年共 744 起赤潮发生事件可知(见图 2),浙江海域赤潮发生次数具有明显的“低频发生-爆发增长-缓降波动”3 段式年际变化特征。20 世纪为低频发生阶段:其中 80 年代发生频次寥寥无几,年均两次;90 年代尤其是前期略有增加,年均 4.5 次。进入 21 世纪后,赤潮发生次数呈现爆发性增长,高达 45 次/a;2003 年到达了历史峰值 79 次后逐渐回落。2010 年后进入第三阶段,发生次数缓降后呈稳定波动状态,比前 10 a 明显减小,年发生数维持在 18 次/a。

从季节分布来看,一年四季皆有赤潮发生,但集中爆发于春夏两季(发生次数占 97.5%)。由图 3 可知,除 11、12 月以外,其余各月份均有赤潮发生且主要在 4—8 月。其中 5 月发生最多(287 次),其次为 6 月(188 次),分别占全年的 46.7% 和 29.7%;发生最少月份为 10 月,共发生 4 次,约占 0.3%。从持续天数来看,浙江海域持续天数 1~3 d 的短期赤潮出现次数最多,占 64%;20 d 以上的超长周期赤潮事件共出现 11 次,占 1.8%,其中最长持续天数为 31 d。值得注意的是两次超长持续时间的赤潮事件发生时间段分别为 2011 年 2 月 9 日—3 月 7 日和 2017 年 2 月 7 日—3 月 9 日,均发生在冬季的象山港港底海

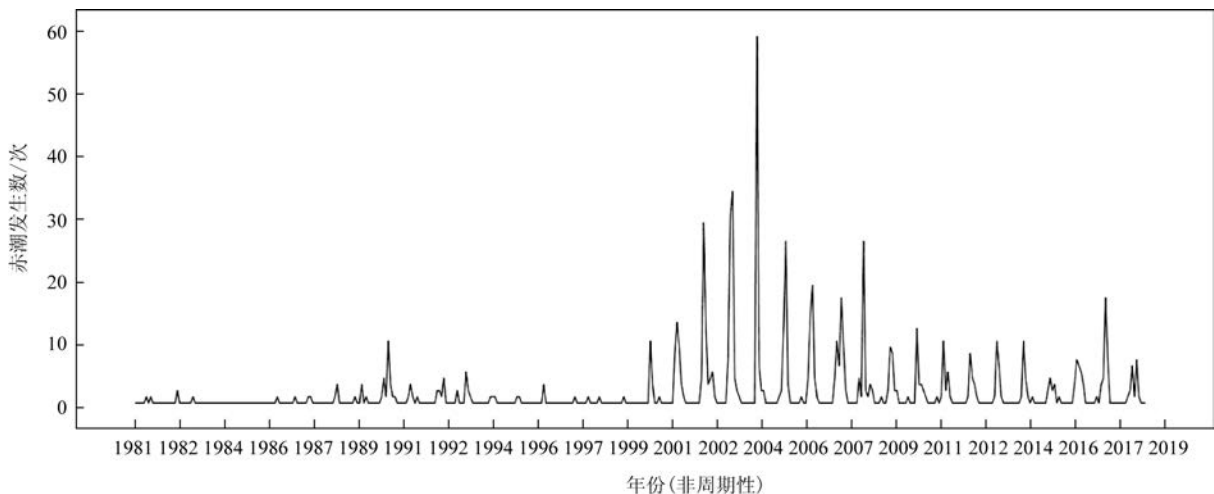


图2 近 40 a 浙江海域赤潮逐月发生次数时序图

域。相关研究已表明^[27],象山港在电厂建成前冬季平均水温约为8~9℃。2005年底宁海国华电厂和2006年底大唐乌沙电厂相继投产后,温排水使得附近海域的水温升高,4℃温升包络线范围不断扩大使水温达到15℃左右的可能性大大增加,而且冬春季温排水的热效应使浮游植物量增加约5%。因此推测由于温排水的热效应导致生物量增加,从而使得象山港海域冬末春初的赤潮呈现低温期爆发、持续时间长、影响面积小等特点。因此,必须重视特殊海域“低温期”的赤潮事件,以防在赤潮预警报业务中“漏报事件”发生。

由图2、3可见,浙江海域赤潮发生次数既存在

明显的年际变化,同时也存在典型的季节波动,因此导致基于赤潮发生次数逐月原始数据建立的时间序列非平稳且具有季节性。当变量不平稳时,若直接构造ARIMA模型容易产生虚假回归,因此须先对原始数据进行一阶普通差分。差分后虽然没有上升和下降趋势,但是随着时间的增加周期性一直存在,因此还需继续做季节差分。图4显示,一阶季节差分后,序列的长期趋势和季节性趋势基本消失,数值围绕0上下随机波动。自相关图呈现逐渐衰减的趋势,自相关系数能够趋于0,且延迟16阶后在0值附近波动,统计量的相伴概率小于0.05,通过单位根检验(ADF)^[28],故可认为处理后的时间序

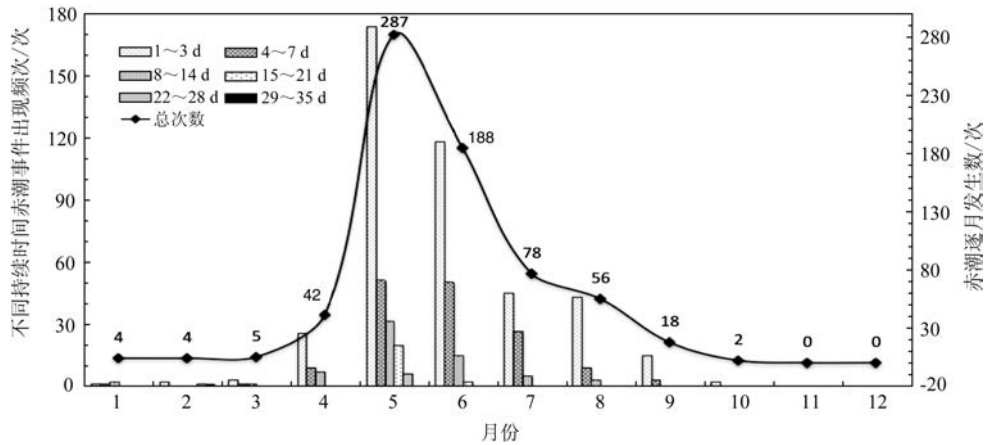


图3 近40 a浙江海域赤潮月发生次数分布特征图

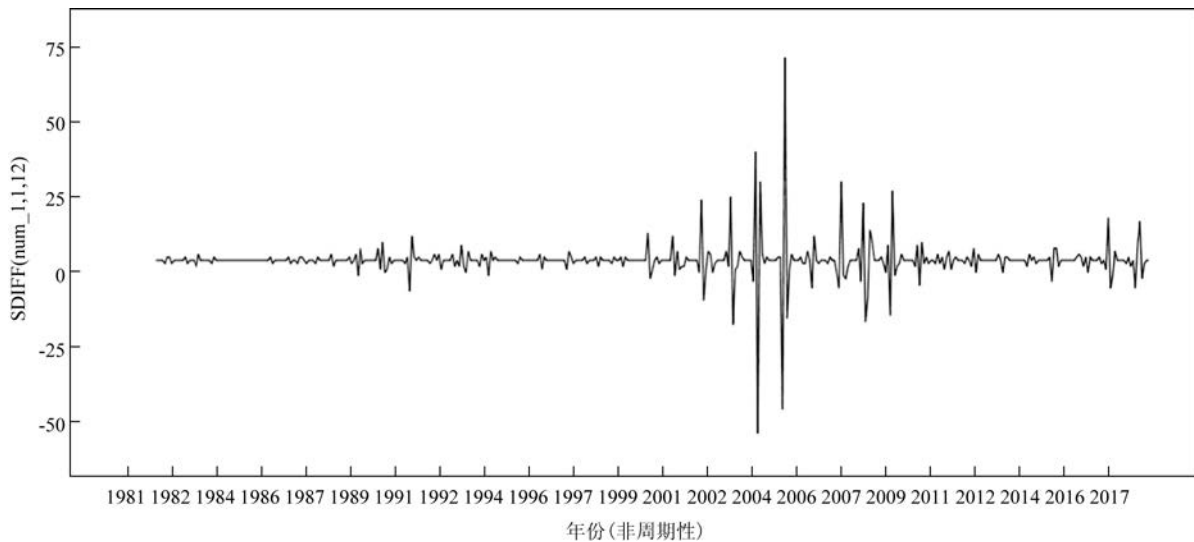


图4 近40 a浙江海域逐月赤潮发生数差分序列图

列平稳,符合建模的条件。

3.2 模型拟合及参数确定

上文处理后的平稳时间序列,可依据 Box-Jenkins 的理论方法直接构建 ARIMA 模型。

根据赤潮发生次数时间序列的差分情况及季节性特征,可以初步判别该时间序列是周期为 12 的复合季节模型 $ARIMA(p,d,q)(P,D,Q)_s$,故 S 取值为 12;差分后序列自相关系数 (Auto-Correlation Function, ACF) 和偏相关系数 (Partial Correlation Function, PACF)^[29]表明 (见图 5),ACF 和 PACF 均在 P 参数大于 1 后骤减,故初步判断连续模型为 RIMA (1,1,1)。季节模型的参数 P 、 Q 判断较复杂,一般情况下超过二阶的情况很少见,可以分别取 0、1、2 由低阶到高阶逐个实验。本文基于 SPSS 25.0 统计软件“时间序列预测”模块中的专家建模器,通过极大似然法^[30]进行估计,初步拟定模型参数,剔除系数不显著的模型,并对剩下模型的残差进行 Ljung-Box 非线性检验^[31],选出 $P>0.05$ 的模型;再从选出的模型中遵从 BIC 准则^[26]筛选出最优模型。由表 1 结果

可知,Ljung-Box 的 Q 检验显示残差目前并未违反白噪声的假设,也没有出现离群值,选取的赤潮发生次数时间序列的最优预测模型为 $ARIMA(1,1,1)(1,1,0)_{12}$,拟合优度系数为 0.68,结果在可接受范围内。图 5 可见模型残差的 ACF 和 PACF 均 ≤ 0.5 ,残差序列各数值间没有相关性,这说明建立的预测模型已充分提取了序列信息,是合适且可信的。

3.3 模型检验及预测讨论

一般情况下,为评估预测模型的稳定性和适应性,会选择拟合优度、平均绝对误差和相对误差来评价模型的整体拟合度^[32-33]。因为赤潮逐月发生次数样本的特殊性,大多数样本为 0 值或数值较小,从而导致绝对误差和相对误差较大,但在实际业务中却属于可接受范围,因此本文采用绝对误差值范围出现概率作为评价指标。基于上文建立的最优模型 $ARIMA(1,1,1)(1,1,0)_{12}$,对 1981 年 1 月—2018 年 12 月的赤潮逐月发生次数进行模拟计算,模型拟合优度系数为 0.68,模拟值与实测值较吻合 (见图 6)。

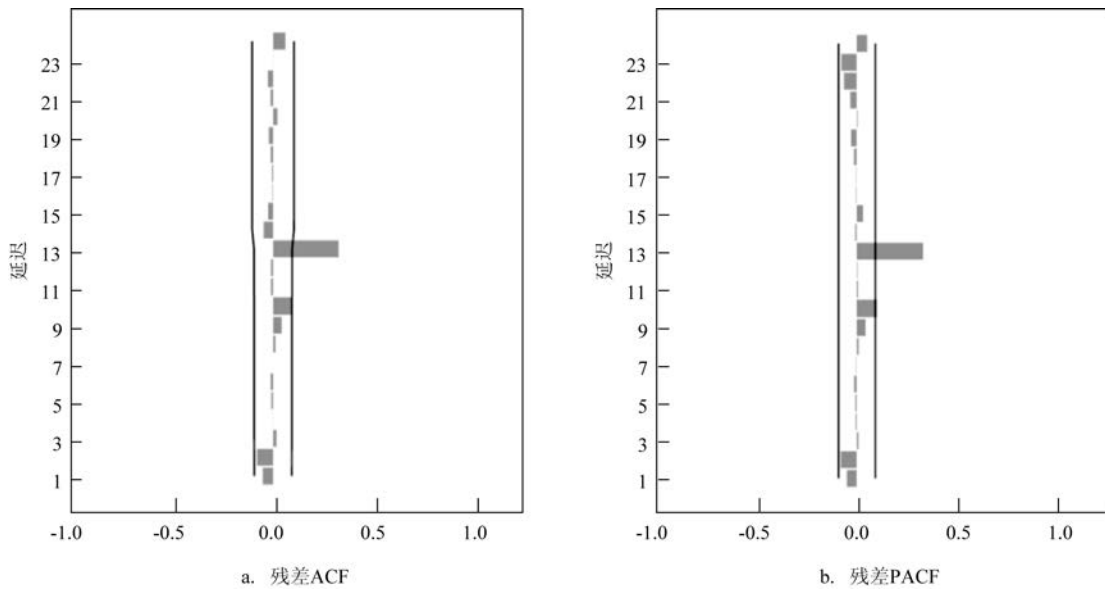


图5 残差的 ACF 和 PACF 图

表1 浙江海域赤潮发生次数 ARIMA 预测模型的相关参数

模型	模型拟合度统计	Ljung-Box Q (18)			离群值数
	平稳 R ²	统计	自由度	显著性	
ARIMA (1,1,1)(1,1,0) ₁₂	0.68	59.525	15	0	0

ARIMA 模型中,数据样本的时间尺度以月为单位进行计算。744次赤潮发生事件按照实际发生时间归入各月中,形成 ARIMA 模型的 457 个样本。由表 2 可知,在 457 个样本中,绝对误差控制在两次以内的样本共 382 个,占总数的 84%;其中绝对误差为 0 次即模拟值与实测值完全吻合的月份有 292 个,占比 64%。但也存在个别月份绝对误差较大的情况,主要出现在 2000—2005 年浙江海域赤潮爆发

性增长阶段,尤其是 2004 年 5 月赤潮发生次数爆发性增长至峰值 59 次,ARIMA 模型未能准确拟合这种超历史极值的小概率异常情况。

利用上文建立的模型对浙江海域 2019 年赤潮逐月发生次数进行预测(见图 7)。2019 年浙江海域赤潮实际发生次数为 22 次,预测值为 19 次,赤潮年发生次数的相对误差为 14%。其中 4 月、5 月和 8 月的绝对误差都控制在两次以内,其余月份预测值与

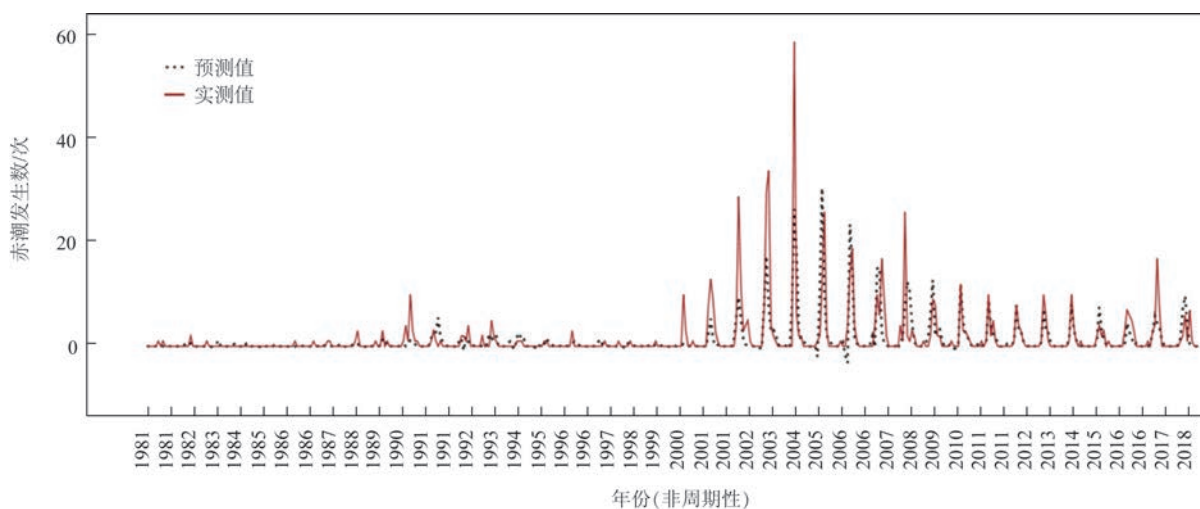


图 6 赤潮逐月发生数 ARIMA 模型预测值与实测值对比

表 2 ARIMA 模型模拟值与实测值的绝对误差统计

绝对误差/次	0	1	2	3~5	6~10	11~15	16~20
月/个	294	90	25	28	11	5	4
占比/%	64	20	5	6	2	1	1

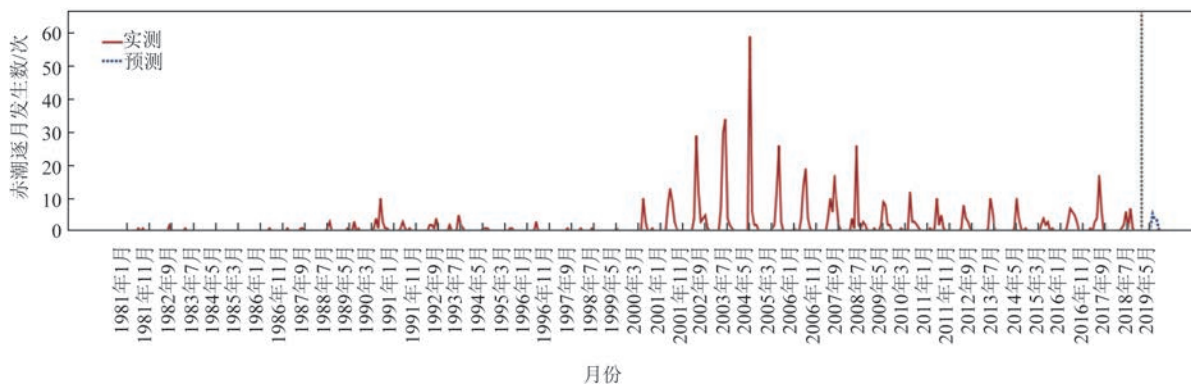


图 7 浙江海域 2019 年赤潮逐月发生次数预测

表3 浙江海域2019年赤潮逐月发生次数预测值
与实测值的绝对误差统计

月份	模拟值	实测值	绝对误差
1	0	0	0
2	0	0	0
3	0	0	0
4	2	4	-2
5	5	4	1
6	8	8	0
7	4	4	0
8	0	2	-2
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0

实测值完全吻合(见表3)。可见本文建立的ARIMA模型能够较准确地进行赤潮发生次数的年度趋势预测。

4 结论

考虑赤潮系统的高维非线性机制和年度预测业务需求,本文利用近40 a浙江海域赤潮逐月发生次数的时间序列,分析其在年际、季和月3个不同时间尺度的变化特征;通过对原始序列的差分处理,基于时间序列分析方法建立了ARIMA预测模型,并对2019年浙江海域赤潮逐月发生次数进行后报和检验。结果如下:

(1)浙江海域赤潮发生次数存在明显年际变化特征,经历了“低频发生-爆发性增长-缓降后波动”3个明显的阶段,其中2000年和2010年为转折点。同时也存在典型的季节波动,97.5%集中爆发在春、夏季,其中每年的5月发生最多,共298次,占全年的46.7%。从持续天数来看,1~3 d的短期赤潮出现次数最多,占64%。

(2)通过对原始序列的差分处理和参数检验,最终建立ARIMA(1,1,1)(1,1,0)₁₂模型且残差通过白噪声检验,拟合优度系数为0.68,绝对误差控制在两次以内的样本占总数的84%。用此模型对浙江海域2019年赤潮发生次数进行预测,后报检验显示年发生总次数相对误差为14%,各月绝对误差均

在两次以内,预测结果与实际较吻合。

(3)时间序列预测法需要的仅是序列本身的历史数据,对机理研究没有高要求,在赤潮年度预测等业务中具有简易实用且经济性好的优势;同时在具备长时间序列数据的基础上,其精度在业务应用可接受范围内。在今后的相关研究中,可探索线性与非线性模型的最优组合模型,如ARIMA与非线性自回归神经网络模型组合,加强赤潮相关影响因子的收集并纳入时间序列模型中。这样既保留模型本身的线性预测能力,又弥补了其在非线性预测方面的不足,以期提高模型预测精度。

(4)本文所建立的ARIMA模型具有简捷经济实用、业务性强的特点。但同时值得注意的是,因其仅考虑时间序列上的依存性和随机波动的干扰性,对于超历史极值会有无数据依存导致拟合不佳的情况。如2004年5月出现历史第一极值“59次”,分析其原因主要是海水养殖面积、沿海城市集聚发展导致近岸海域4类及劣4类海水比例在2004年达到峰值,同期海温也处于快速上升期。在营养盐充沛供给、水文气象条件适宜等多种因素共同作用下导致了赤潮频繁爆发,出现超历史极值。因此,下一步研究中需将时间序列模型的预测结果与其他关联因素相结合开展综合统计分析,进一步改进超历史极值情况的预测精度。

参考文献:

- [1] 齐雨藻. 中国沿海赤潮[M]. 北京: 科学出版社, 2003.
- [2] 周名江, 朱明远, 张经. 中国赤潮的发生趋势和研究进展[J]. 生命科学, 2001, 13(2): 54-59, 53.
- [3] 周燕. 浙江海域赤潮灾害及防治[M]. 杭州: 浙江大学出版社, 2017.
- [4] 张福星, 姚玉娟, 马林芳. 温州沿海赤潮发生的水文气象条件及赤潮特征分析[J]. 海洋预报, 2016, 33(5): 89-94.
- [5] 林祖亨, 梁舜华. 大鹏湾盐田海域夜光藻赤潮形成与潮汐的关系[J]. 海洋通报, 1993, 12(2): 35-38.
- [6] 赵冬至, 张丰收, 杨建洪, 等. 表征赤潮水体太阳激发的叶绿素荧光高度的波段优化——归一化荧光高度法[J]. 海洋学报, 2005, 27(6): 146-153.
- [7] 赵冬至, 张丰收, 赵玲. 近岸海域叶绿素和赤潮的AVHRR归一化差值探测方法研究[J]. 海洋技术, 2003, 22(3): 30-33.
- [8] 陈晓翔, 邓孺孺, 何执兼, 等. 赤潮相关因子的卫星遥感探测与赤潮预报的可行性探讨[J]. 中山大学学报(自然科学版), 2001, 40(2): 112-115.

- [9] 洪君超, 黄秀清, 蒋晓山, 等. 嵊山水域中肋骨条藻赤潮发生过程主导因子分析[J]. 海洋学报, 1993, 15(6): 135-141.
- [10] 黄秀清, 蒋晓山, 陶然, 等. 长江口海区一次骨条藻赤潮发生过程的多元分析[J]. 海洋环境科学, 2000, 19(4): 1-5.
- [11] 苏纪兰, 唐启升. 中国海洋生态系统动力学研究 II 渤海生态系统动力学过程[M]. 北京: 科学出版社, 2002: 445.
- [12] 张书文, 夏长水, 袁业立. 黄海冷水团水域物理-生态耦合数值模式研究[J]. 自然科学进展, 2002, 12(3): 315-319.
- [13] 刘浩, 尹宝树. 渤海生态动力过程的模型研究 I. 模型描述[J]. 海洋学报, 2006, 28(6): 21-31.
- [14] 林军. 长江口外海域浮游植物生态动力学模型研究[D]. 上海: 华东师范大学, 2011.
- [15] 林军, 章守宇, 龚甫贤. 象山港海洋牧场规划区选址评估的数值模拟研究: 滨海电厂温排水温升的影响[J]. 上海海洋大学学报, 2012, 21(5): 816-824.
- [16] 夏综万, 于斌, 史键辉, 等. 大鹏湾的赤潮生态仿真模型[J]. 海洋与湖沼, 1997, 28(5): 468-474.
- [17] 李雁宾. 长江口及邻近海域季节性赤潮生消过程控制机理研究[D]. 青岛: 中国海洋大学, 2008.
- [18] 冯剑丰, 王洪礼, 李胜朋. 基于支持向量机的浮游植物密度预测研究[J]. 海洋环境科学, 2007, 26(5): 438-441.
- [19] 王洪礼, 葛根, 李悦雷. 基于模糊神经网络(FNN)的赤潮预警预测研究[J]. 海洋通报, 2006, 25(4): 36-41.
- [20] 徐丽丽, 龚茂珣, 徐婷婷, 等. 东亚气候异常与东海赤潮发生频次的关系研究[J]. 海洋预报, 2013, 30(5): 8-14.
- [21] 杨红, 何春良. 我国海域赤潮发生与海温及厄尔尼诺的相关分析[J]. 海洋湖沼通报, 2009(2): 1-6.
- [22] 邓邦平, 张昊飞, 何彦龙, 等. 1981-2011年间东海赤潮发生与厄尔尼诺关系分析[J]. 生态科学, 2017, 36(6): 161-164.
- [23] 何春良, 杨红. 我国海域赤潮时空分布及其与厄尔尼诺的关系[J]. 上海海洋大学学报, 2009, 18(2): 206-211.
- [24] 王振龙, 胡永宏. 应用时间序列分析[M]. 北京: 科学出版社, 2007.
- [25] 夏菁, 张华勋, 林文, 等. ARIMA模型在疟疾发病率预测中的应用[J]. 中国血吸虫病防治杂志, 2016, 28(2): 135-140.
- [26] 冯双磊, 王伟胜, 刘纯, 等. 风电场功率预测物理方法研究[J]. 中国电机工程学报, 2010, 30(2): 1-6.
- [27] 徐丽丽. 象山港温排水海域赤潮发生预报模型研究[D]. 南京: 南京信息工程大学, 2015.
- [28] 余辉, 袁晶, 于旭耀, 等. 基于 ARMAX 的 PM2.5 小时浓度跟踪预测模型[J]. 天津大学学报(自然科学与工程技术版), 2017, 50(1): 105-111.
- [29] 刘珊, 陈幸荣, 蔡怡. 基于 ARIMA 模型的台风频次预测[J]. 海洋预报, 2016, 33(1): 53-58.
- [30] 张文彤. SPSS11 统计分析教程[M]. 北京: 北京希望电子出版社, 2002: 557-582.
- [31] 唐玉娜, 李启会. ARMA 模型在预测问题中的应用[J]. 嘉兴学院学报, 2006, 18(S1): 183-187.
- [32] 徐静. ARIMA 模型及其应用[J]. 立信会计高等专科学校学报, 2001, 15(3): 21-24.
- [33] 杜勇宏, 王健. 季节时间序列理论与应用[M]. 天津: 南开大学出版社, 2008.

Study on the red tide prediction based on time series analysis —A case study in Zhejiang sea area

XU Li-li¹, YU Jun², GAO Xin-xin¹, CHE Zhu-mei², HE Wen¹, QIU Ting¹

(1. East China Sea Marine Forecasting Center of State Oceanic Administration, Shanghai 200081 China; 2. Key Laboratory of Marine Ecological Monitoring and Restoration Technology, Shanghai 200081 China; 3. Marine Monitoring and Forecasting Center of Zhejiang Province, Hangzhou 211800 China)

Abstract: Based on the historical data collected and compiled from China Marine Disaster Bulletin and the Marine Monitoring and Forecasting Center of Zhejiang Province from 1981 to 2019, the long-term change and seasonal variation characteristics of the red tide occurrence in Zhejiang sea areas are analyzed in this paper. We establish a red tide prediction model based on time series of the monthly occurrence and verify the hindcast results for 2019. The results show that the frequency of red tide occurrence in Zhejiang sea area presents three distinct inter-annual variation features of “low frequency-explosive growth-decline fluctuation” over the past 40 years. At the same time, 97.5% of the red tide outbreaks are concentrated in spring and summer with the most frequent occurrence in May. Based on data differential processing and parameter verification, ARIMA(1, 1, 1)(1, 1, 0)₁₂ is identified as the optimal time series prediction model, and the model residual passes the white noise

test with a fit coefficient of 0.68. The absolute error within 2 times accounts for 84% of the total samples, which proves the predictability of the model. The hindcast verification on the monthly red tide occurrence is conducted for 2019 in Zhejiang sea area. The relative error is 14% and the absolute error is within 2 times, which is reasonably well. As a result, considering the mechanism of red tide is not completely clear at present, it is an economic and effective way to predict the occurrence of red tide using ARIMA model, which is also a new effort and effective supplement to the existing technology of the annual trend prediction and operational early warning of red tide disaster.

Key words: Zhejiang sea area; the occurrence of red tide; ARIMA model; annual-trend prediction