

文章编号:1000-0615(2008)06-0899-07

ARIMA 模型在预测长江靖江段沿岸鱼类 渔获量时间格局中的应用

李辉华¹, 郭弘艺¹, 唐文乔¹, 顾树信², 黄少芳¹, 沈林宏², 魏凯¹

(1. 上海海洋大学省部共建水产种质资源发掘与利用教育部重点实验室, 上海 201306;

2. 江苏省靖江市渔政管理站, 江苏 靖江 214500)

摘要: 长江靖江段处于长江下游及河口段的交汇地带, 是长江河口地区渔业资源养护的重要水域。2002-2006年, 在靖江沿岸用一部定置张网每日采集, 共获得鱼类 3514.84 kg。5年的月平均渔获量为 58.59 kg, 以6月最高, 达 108.61 kg; 12月最低, 仅为 31.57 kg, 呈现出典型的非平稳时间序列。月渔获量经自然对数转换和一次季节差分后, 获得了一组平稳的随机序列。按照残差不相关原则确定模型结构, 依据 Akaike 信息准则(AIC)与 Schwarz 贝叶斯信息准则(SBC)确定模型优度, 用 SPSS V13.0 软件对 2002年1月至2006年12月的月渔获量进行 ARIMA 建模拟合。结果表明, 模型为 ARIMA(1,0,0)(0,1,1)₁₂ (不含常数项), 方程为 $(1-0.327B)(1-B^{12})\text{Ln}y_t = (1-0.825B^{12})e_t$, 的残差是白噪声 ($P>0.05$)。该模型对 2003-2006年渔获量的拟合精度为 71.49%~83.28%, 较好地拟合了既往时段的渔获量。对 2007年逐月渔获量预测的相对精度为 58.64%~99.44%, 年相对精度达 81.60%, 表明该模型能有效应用于长江口沿岸渔获量的预测。

关键词: 鱼类渔获量; 时间格局; ARIMA 模型预测; 沿岸湿地; 长江靖江段

中图分类号: S 934

文献标识码: A

长江下游及其河口地区是我国河流湿地的集中分布区^[1]。这些湿地在涵养水源、保护水质以及减少泥沙输送通量、维系区域水气循环平衡等方面都发挥着十分重要的作用, 也孕育了种类繁多的生物物种, 是重要的物种资源库和基因库^[1-2]。长江河口是许多重要经济水生动物的产卵、育幼和索饵场, 不仅对长江及其近海动物的资源补充和生态平衡起着重要作用, 而且其本身也是我国这个最大河口渔场的重要组成部分^[3-4]。因此它被我国列为湿地和淡水生物多样性保护的关键地区之一, 也被世界自然基金会确立为生物多样性保护的优先区域^[5]。

鱼类是湿地水域经济种类最多、生物量最

大的动物类群和重要的生态健康指标之一, 长江下游及河口地区的渔民也大多集中在这一水域作业^[5-7]。因此, 研究这一水域的鱼类组成及其变动规律, 不仅可为长江下游及其河口地区的渔业提供指导, 也可为这一地区的生境评价提供依据。

1 材料与方法

1.1 研究地点

研究地点位于江苏省靖江市新桥镇长江江段的沿岸滩涂湿地 (31°56'N, 120°01'E)。该江段为长江下游与河口段的交汇水域, 水文状况受径流和潮汐的双重影响, 2002-2007年测得

收稿日期: 2008-01-28

资助项目: 国家自然科学基金重大项目(30490235); 淡水生态与生物技术国家重点实验室开放课题(2004FB04); 上海市重点学科建设项目(J50701)

作者简介: 李辉华(1975-), 男, 湖北公安人, 在职博士研究生, 从事鱼类学研究。E-mail: hhli@shou.edu.cn

通讯作者: 唐文乔, E-mail: wqtang@shfu.edu.cn

的最低水位为 0.0~1.7 m,每次涨落潮的水位波动为 1.1~2.9 m。采样点沿岸上缘为混泥土堤坝,下缘堤脚为宽约 30 m 的大型石块带。堤脚下为宽阔的淤泥质江岸,表面平整,坡度约为 20°。石块带下缘至泥质江岸的内侧常年有宽约 100 m 的密集芦苇带,春秋芦苇高约 3 m,冬季枯萎后割除。

1.2 渔获量采集方法

2002 年 1 月 1 日至 2007 年 12 月 31 日,垂直于江岸设置当地渔民常用规格的丁字形定置张网一部。拦网长 40 m,网片露出江底高约 1.8 m,拦网外缘近江心处布置 2 个笼式网袋。拦网采用 7 号网片,网目为 1.86 cm;网袋为 4 号网片,网目为 0.92 cm。张网的布网位置随季节性的水位涨落而上下移动,一般控制在平水期水面与拦网的上纲齐平。当网片破损或有藻类附着时,更换同一规格的张网。每天由专业渔民在早潮退潮后收集一次渔获物,记录渔获物的总重量,并收集一定频率的样品带回实验室分析。2002 年 1 月至 2006 年 12 月的月渔获量用于建模,2007 年全年的渔获量用于模型验证。

1.3 建模方法

ARIMA(Autoregressive Integrated Moving Average,自回归求和滑动平均)模型是由 Box 和 Jenkins 于 1970 年提出,已成为目前时间序列分析的常用方法^[8-9]。ARIMA 模型应用于时间序列的建模和预报过程,一般可概括为如下五个步骤:

1) 数据预处理:包括零均值化处理和差分平稳化处理,这一步骤主要是使该序列满足时间序列建模的前提条件,此外也有利于提高预测的精度。

2) 模型结构辨识:利用自相关分析和偏自相关分析方法,从 AR(p)模型、MA(q)模型、ARMA(p,q)模型和 ARIMA(p,d,q)模型中,选择一个比较合适的模型作为建模用的模型结构,并确定模型的阶次。

3) 模型参数估计:通过求解方程来确定模型的参数值(变量的系数)。

4) 模型检验:通过对原序列与所建模型之间的误差序列是否具有随机性的检验来实现。若模型检验不能通过,则需回到步骤 2) 重新进行模型结构的辨识。

5) 用所建立的合适模型结构推导出具体的预测模型。

本文建模采用 SPSS V13.0 软件^[9]。

2 结果

2.1 渔获量的年度变化和月间波动

2002 年 1 月 1 日至 2006 年 12 月 31 日的 5 年间,共采集到鱼类 3514.84 kg,年平均渔获量为 702.97 kg,日平均渔获量 1.92 kg。其中以 2005 年最高,为 899.65 kg,其他年份都比较接近,约 650.00 kg(图 1-A)。月平均渔获量为 58.59 kg,其中以 4-8 月最高,6 月高达 108.61 kg,往前和往后均逐月递减,12 月减至最低的 31.57 kg(图 1-B),呈现单峰形的时间格局。

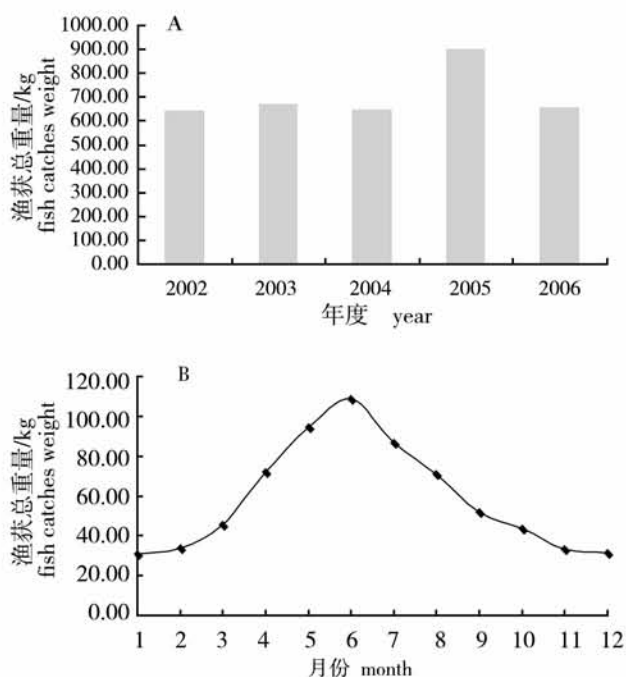


图 1 渔获量的年度变化(A)和月间波动(B)
Fig. 1 Changes of fish catches in each year(A)
and each month(B)

2.2 模型识别

图 2 是 2002 年 1 月至 2006 年 12 月间长江靖江段逐月鱼类渔获量的时间变化图,其特点有:①有周期性,每 12 个月出现一次明显的渔获高峰,峰值出现在每年的 5-7 月份;②序列的方差前后波动较大,如 2003 与 2004 年,渔获高峰与低谷的间距变化较大,表明原始序列不是一个平稳的随机过程。

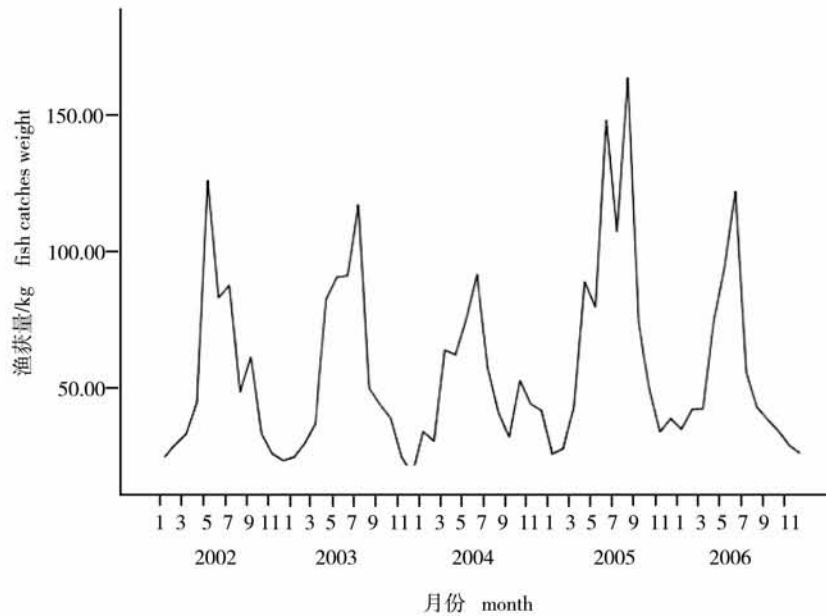


图2 2002-2006年间的月渔获量变动

Fig. 2 Changes of monthly fish catches from 2002 to 2006

一个平稳的随机过程应具备如下特征:均数与方差均不随时间变化,自相关系数只与时间间隔有关而与所处的时间无关^[9]。为使序列符合运用ARIMA模型的条件,先对渔获量数据进行自然对数转换,以平稳序列的方差;并对渔获量序列进行季节差分,以消除序列的周期性。经参数检验($P>0.05$),渔获量经转换与差分后,序列的均值、方差、相关函数等特征均不随时间的推移而变化,即得到了一组平稳的随机序列。据此,可初步确定适用的模型是以12个月为周期的综合了连续模型和季节模型的复合季节模型,即ARIMA($p,0,q$)($P,1,Q$)₁₂。其中, p,q 和 P,Q 分别为连续模型和季节模型中的自回归阶数和移动平均阶数,该阶数通常可由第1和13阶处的自相关系数(ACF)和偏自相关系数(PACF)来确定。观察差分后序列的ACF和PACF图(图3),可以初步确定 $p=1,P=0,q=1,Q=1$,即模型为ARIMA(1,0,1)(0,1,1)₁₂。

2.3 参数计算和模型选择

ARIMA模型的优度可依据Akaike信息准则(Akaike's Information Criterion, AIC)和Schwarz贝叶斯准则(Schwarz's Bayesian Criterion, SBC)进行判定,AIC和SBC值小的模型一般认为相对较好。经多次调试和检验,获得了3种模型的拟合优度统计量及其参数(表1,表

2),表明ARIMA(1,0,0)(0,1,1)₁₂是相对较好的模型。根据简洁原则,将无统计学意义的常数项去除后,发现模型的优度更好(即AIC与SBC更小,表1)。因此模型的方程为: $(1-0.327B)(1-B^{12})\text{Ln}y_t=(1-0.825B^{12})e_t$ 。

该方程表明月渔获量之间存在着一阶关联性,即当月渔获量与前1月的渔获量之间存在着相关性,因此可用当月的渔获量推测下月的渔获量。同时季节平均移动参数表明,渔获量的1个周期(年)波动受前1个周期(年)渔获量波动的影响,即可用当年渔获量的波动推测下一年的渔获量波动情况。两者结合,即可对长江靖江段沿岸鱼类的月渔获量和年渔获量做出预测和预报。

通过对原始数据与拟合数据的误差计算,用ARIMA方程对本模型进行可靠性检验。结果显示,残差统计量均无统计学意义($P>0.05$),即残差是白噪声,也即残差在不同时间上相互独立,这进一步表明所选模型是可行的。

2.4 模型准确性检验

用不含常数项的ARIMA(1,0,0)(0,1,1)₁₂模型对靖江段2002-2006年的月间渔获量进行拟合,并对2007年逐月渔获量进行预测,结果见图4。可见,预测值与实测值基本一致,表现出相似的季节律动(图4)。表3的结果显示,2003-

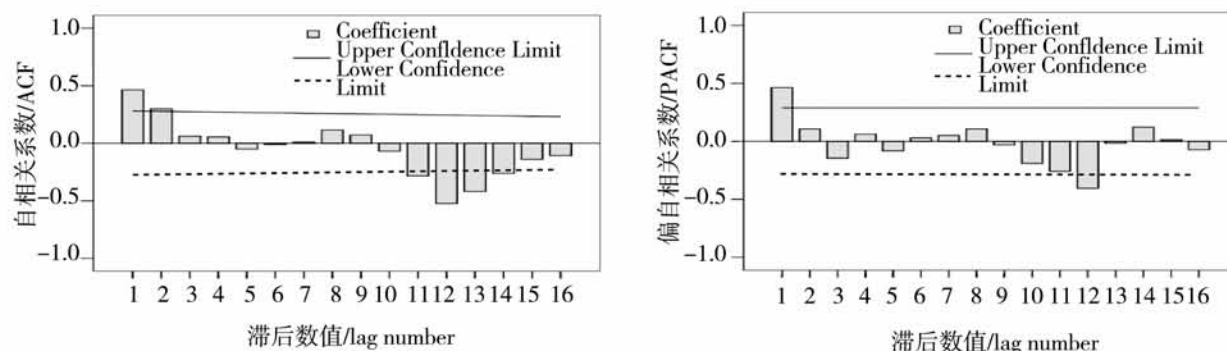


图3 自然对数转换和季节一阶差分后的月渔获量相关图

Fig. 3 ACF and PACF of fish catches with transforming by natural logarithm and seasonal difference

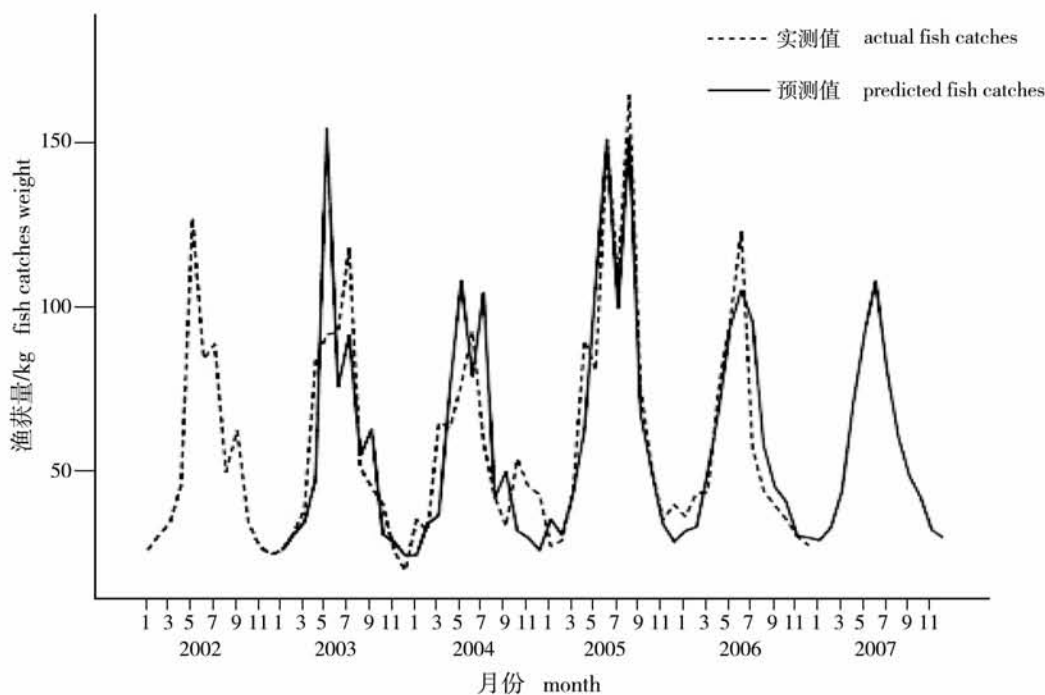


图4 月渔获量的实测值与模型预测值

Fig. 4 Time series of actual fish catches and predicted fish catches of monthly fish catches

表1 3种ARIMA模型拟合优度统计量

Tab. 1 Statistics of goodness fitting for three ARIMA models

统计量 statistics	ARIMA(1,0,1) (0,1,1) ₁₂	ARIMA(1,0,0) (0,1,1) ₁₂ (不含常数) (no constant)	ARIMA(1,0,0) (0,1,1) ₁₂ (含常数) (with constant)
	Schwarz 贝叶斯准则 (Schwarz's Bayesian Criterion, SBC)	38.816	36.469
Akaike 信息准则 (Akaike's Information Criterion, AIC)	44.430	40.212	43.626
log-Likelihood	-16.408	-16.235	-16.006

表 2 3 种 ARIMA 模型的参数估计
Tab. 2 Parameter estimation of the three ARIMA models

参数 parameters	ARIMA(1,0,1) (0,1,1) ₁₂			ARIMA(1,0,0) (0,1,1) ₁₂ (含常数) (with constant)			ARIMA(1,0,0) (0,1,1) ₁₂ (不含常数) (no constant)		
	B	t	P	B	t	P	B	t	P
AR1	0.686	2.767	0.008	.286	2.248	.030	0.327	0.125	0.012
MA1	0.405	1.284	0.206	-	-	-	-	-	-
SMA1	0.984	0.188	0.852	0.920	0.902	0.372	0.825	0.399	0.044
Constant	-	-	-	0.039	1.100	0.277	-	-	-

注: AR1:一阶自回归参数;MA1:一阶平均移动参数;SMA1:一阶季节平均移动参数;B:系数;t:t值,P:P值

Notes: AR1:First order autoregressive parameter;MA1:First order regular moving average parameter;SMA1:First order seasonal moving average parameter;B;Coefficient;t;t value;P:P value

2006 年的模型拟合精度在 71.49%~83.28%,平均为 79.96%,拟合较为准确。

对 2007 年的预测值与实测值的比较结果表明(表 4),年均相对精度为 81.60%,月相对精度在 58.64%~99.44%。虽然个别月份(如 11 月)的相对精度较低,但经 Kolmogrov-Smirnow 双样本检验显示,预测值与实测值之间也并无显著性差异($P>0.05$),表明该模型可用于实际预测。

表 3 ARIMA(1,0,0)(0,1,1)₁₂模型预测的
准确性验证

Tab. 3 The test of the forecast accuracy
of ARIMA(1,0,0)(0,1,1)₁₂

年份 year	2003	2004	2005	2006
相对精度(%) relative precision	82.61	71.49	83.28	82.45

注: 相对精度 = $[1 - |\text{实测值} - \text{预测值}| / \text{实测值}] \times 100$

Notes: Relative precision = $[1 - |\text{actual fish catches} - \text{predicted fish catches}| / \text{actual fish catches}] \times 100$

表 4 2007 年月渔获量实测值与模型预测值比较

Tab. 4 Comparison of actual fish catches and predicted fish catches in 2007

月份 month	1月 Jan.	2月 Feb.	3月 Mar.	4月 Apr.	5月 May	6月 Jun.	7月 Jul.	8月 Aug.	9月 Sep.	10月 Oct.	11月 Nov.	12月 Dec.
实测值(kg) actual fish catches	29.66	33.47	44.67	71.24	92.43	108.28	81.30	61.26	49.22	42.75	32.75	30.44
预测值(kg) predicted fish catches	31.43	29.36	34.00	56.37	75.05	114.03	83.60	60.93	42.80	52.55	55.85	48.20
相对精度(%) relative precision	94.40	85.99	68.61	73.61	76.85	94.96	97.25	99.44	85.01	81.34	58.64	63.15

注: 相对精度计算公式同表 3

Notes: Computing formula of relative precision is listed in Tab. 3

3 讨论

在渔业管理和决策过程中,鱼类渔获量的时间序列分析是一项重要的内容,它能够揭示一些潜在的季节变化趋势。有些统计模型,如 ARIMA 模型、传递函数模型(transfer function models)、干预分析模型(intervention analysis

models)和分解和回归模型(decomposition and regression models)等已被广泛用于鱼类的时间序列预测^[10-13]。Stergiou 等^[14]认为,由于生物时序变量的数据比较匮乏,ARIMA 模型在预测海水鱼类渔获量方面是最有效的^[14]。目前,该方法在我国主要应用于气象、农业产量和疾病等领域的预测预报,而运用于渔业领域的预测研究还未

见报道^[15-17]。

本文通过对长江靖江段沿岸鱼类5年的实测渔获量分析,建立了ARIMA(1,0,0)(0,1,1)₁₂预测模型,较准确地预测了下一年度的渔获量,表明ARIMA模型也能较好地用于开放性淡水水域的渔获量预测。另外,ARIMA模型是线性技术,建模和预测计算均比较简易。

长江靖江段辖区水面1.26万hm²,江段上、下两端均有沙洲将江面分叉,形成了流态复杂的水域环境,是靖江江心洲省级水产种质资源保护区和长江靖江段中华绒螯蟹、鳊国家级水产种质资源保护区的所在地,在长江下游及其河口地区的渔业资源养护中起着重要作用。

根据我们多年的监测与统计,靖江段作业的丁字形定置张网约有800~900部,这样测算的年渔获量在580~720t,在长江年10万t的渔获量中所占比重并不高。另据本研究对2002年1月至2004年12月73份渔获物样本的详细分析^[18],共有鱼类94种,23523尾,128.72kg,表明靖江段沿岸的鱼类物种多样性非常丰富,但渔获个体极小,平均每尾仅5.50g,每天每网也仅有1.76kg,渔业价值和劳动生产率均很低。从资源保护和渔业利用角度看,长江靖江段沿岸不应作为捕捞渔业的场所。

本文承上海海洋大学王武教授提出宝贵意见,江苏靖江市杨秀龙先生6年多来每天协助采集和统计标本,黄平高级工程师、翁志毅高级技师、诸廷俊先生、蔡明刚先生、刘至治副教授和杨金权博士等协助采集或分析标本,特此一并致谢。

参考文献:

- [1] 郎惠卿,林鹏,陆健健. 中国湿地研究和保护[M]. 华东师范大学出版社,1998.
- [2] 国家林业局. 中国湿地保护行动计划[M]. 中国林业出版社,2000.
- [3] 唐文乔,诸廷俊,陈家宽,等. 长江口九段沙湿地的鱼类资源及其保护价值[J]. 上海水产大学学报,2003,12(3):193-200.
- [4] 庄平,王幼槐,李圣法,等. 长江口鱼类[M]. 上海科学技术出版,2006.
- [5] 王利民,王丁. 长江流域的渔业与湿地保护[J]. 人民长江,2004,35(5):37-39.
- [6] 顾树信,沈林宏,戴玉红,等. 长江靖江段鱼虾资源调查与分析[J]. 淡水渔业,2006,36(4):34-36.
- [7] 吴玲玲,陆健健,童春富,等. 长江口湿地生态系统服务功能价值的评估[J]. 长江流域资源与环境,2003,12(5):411-416.
- [8] 顾岚(主译). 时间序列分析:预测与控制[M]. 北京:中国统计出版社,1997.
- [9] 张文彤. SPSS II 统计分析教程[M]. 北京:北京希望电子出版社,2002:250-289.
- [10] Theodoros K, Lazaros I, Georgios K. Time-series modeling of fishery landings using ARIMA models and Fuzzy Expected Intervals software [J]. Environmental Modelling and Software, 2006, 21: 1711-1721.
- [11] Ivone A C, Juan C G, Jos A H. Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches [J]. Fisheries Research, 2007, 86: 120-128.
- [12] Juan C G, Claudio S, Eleuterio Y. Monthly catch forecasting of anchovy *Engraulis ringens* in the north area of Chile: Non-linear univariate approach [J]. Fisheries Research, 2007, 86: 188-200.
- [13] Eleni M S, Demetrios E K, Spyridon K G. Control charts for the toxicity of finished water-Modeling the structure of toxicity [J]. Water Research, 2007, 41(12): 2679-2689.
- [14] Stergiou K, Christou E, Petrakis G. Modeling and forecasting monthly fisheries catches comparison of regression, univariate and multivariate time series methods [J]. Fisheries Research, 1997, 29: 55-95.
- [15] 王志忠,刘秀菊,王树林,等. 黄河流域棉花品种产量性状时间序列的ARIMA模型预测研究[J]. 棉花学报,2007,19(3):220-226.
- [16] 吴家兵,叶临湘,尤尔科. ARIMA模型在传染病发病率预测中的应用[J]. 数理医药学杂志,2007,20(1):90-93.
- [17] 朱继民,汤林华,周水森,等. 非稳定性疟疾发病率的时间序列模型预测可行性研究[J]. 中国寄生虫学与寄生虫病杂志,2007,25(3):232-236.
- [18] 黄少芳. 长江靖江段沿岸鱼类群聚的时间格局[D]. 上海:上海水产大学硕士学位论文,2005.

ARIMA model application to predict temporal pattern of fish catches of coastal area at Jingjiang Reach of the Yangtze River

LI Hui-hua¹, GUO Hong-yi¹, TANG Wen-qiao¹, GU Shu-xin²,
HUANG Shao-fang¹, SHEN Lin-hong², WEI Kai¹

(1. Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources,
Shanghai Ocean University, Ministry of Education, Shanghai 201306, China;

2. Administration of Fishery of Jingjiang, Jingjiang 214500, China)

Abstract: Jingjiang Reach of the Yangtze River, which is an important area for fishery resource conservation, is located at the connected region of lower reaches and estuary of Yangtze River. From 2002 to 2006, there were 3514.84 kg fish collected by stow net at Jingjiang coastal area. Average monthly fish catches added up to 58.59 kg in these 5 years, the maximum was 108.61 kg in June while the minimum was 31.57 kg in December which presented a typical non-stationary time-series. One stable random series could be obtained as monthly fish catches converted by natural logarithm and a seasonal difference. Model structure was determined according to criteria of residual un-correlation and model goodness was determined on the basis of Akaike Information Criteria(AIC) and Schwarz Bayesian Information Criteria(SBC). SPSS V13.0 software was used to construct the ARIMA model based on the monthly fish catches from Jan. 2002 to Dec. 2006 with consideration of residual un-correlation and concision. Akaike's information criterion (AIC) and Schwarz's Bayesian criterion (SBC) were used to confirm the fitness of model. These results showed that ARIMA(1,0,0)(0,1,1)₁₂ model was confirmed and the predicting error for the equation of $(1 - 0.327B)(1 - B^{12}) \ln y_t = (1 - 0.825B^{12})e_t$ was white noise ($P > 0.05$). The fitting precision of ARIMA model was 71.49% - 83.28% during 2003 - 2006, which could ideally fit the past monthly fish catches. Relative precision of the forecasting for gradual month in 2007 was 58.64% - 99.44% as well as the forecasting for 2007 is 81.60%. It indicated that ARIMA model can be well used to forecast the fish catches of coastal wetland at the estuary of the Yangtze River.

Key words: fish catches; temporal pattern; ARIMA model prediction; coastal wetland; Jingjiang Reach of the Yangtze River