

## 实验技术与方法

## 季节性自回归差分移动平均模型在牡蛎中诺如病毒检出率预测上的应用

杨明树<sup>1,2</sup>,董蕾<sup>1,2</sup>,贾添慧<sup>1,2</sup>,喻勇新<sup>1,2</sup>(1. 上海海洋大学食品学院,上海 201306; 2. 农业农村部水产品贮藏保鲜  
质量安全风险评估实验室,上海 201306)

**摘要:**目的 基于季节性自回归差分移动平均(ARIMA)模型分析并预测上海市售牡蛎中诺如病毒(NoV)的检出率,为水产品中NoV的污染规律提供参考。方法 2016年6月—2019年11月,从上海芦潮港海鲜市场定期采购牡蛎样品共531只,通过巢式聚合酶链式反应(Nest-PCR),对其进行了NoV检测,按季度分析检出率。采用季节性ARIMA模型对牡蛎中NoV的检出率数据拟合建立模型,经过数据平稳化、模型选择和拟合、模型诊断得到最优模型,并运用最优模型对未来四个季度牡蛎中NoV的检出率进行预测。结果 拟合出的季节性ARIMA(0,1,1)(0,1,0)<sub>4</sub>为最优模型,赤池信息量准则的修正值(AICc)最小(58.70),残差经Ljung-Box检验为白噪声序列。模型拟合牡蛎中NoV的阳性率趋势与实际检出率趋势基本吻合,平均绝对误差(MAE)为4.85,平均绝对百分比误差(MAPE)为30.25。用最优模型预测的未来四个季度牡蛎中NoV阳性检出率分别为31.89%、12.80%、9.47%、6.14%。结论 季节性ARIMA模型(0,1,1)(0,1,0)<sub>4</sub>能较好地拟合牡蛎中NoV的阳性检出率趋势,对NoV污染的牡蛎等水产品的风险评估及NoV流行的防控具有一定的意义。

**关键词:**季节性自回归差分移动平均模型;诺如病毒;检出率;时间序列分析;预测

中图分类号:R155 文献标识码:A 文章编号:1004-8456(2021)04-0430-05

DOI:10.13590/j.cjfh.2021.04.005

**Application of seasonal ARIMA model in prediction of detection rate of norovirus in oyster**YANG Mingshu<sup>1,2</sup>, DONG Lei<sup>1,2</sup>, JIA Tianhui<sup>1,2</sup>, YU Yongxin<sup>1,2</sup>(1. College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China;  
2. Laboratory of Quality and Safety Risk Assessment for Aquatic Products on Storage and  
Preservation (Shanghai), Ministry of Agriculture, Shanghai 201306, China)

**Abstract: Objective** The seasonal autoregressive integrated moving average (ARIMA) model was used to predict the detection rate of norovirus in oysters sold in Shanghai, which provided a reference for the prevalence of norovirus in aquatic products. **Methods** Oyster samples were regularly purchased from the Shanghai Luchaogang seafood market. A total of 531 oyster samples were tested for norovirus by nest-polymerase chain reaction (Nest-PCR), and the positive detection rate was calculated every quarter. The seasonal ARIMA model was used to fit the norovirus detection rate data in oysters from June 2016 to November 2019 to construct the model. After data stabilization, model selection and fitting and model diagnosis, the optimal model was obtained and the optimal model was used to predict the detection rate of norovirus in oysters in 2020. **Results** The seasonal ARIMA (0, 1, 1) (0, 1, 0)<sub>4</sub> was the optimal model. Akaike's information criterion and the finite corrections (AICc) (58.70) was the smallest. The residual error was a white noise sequence by Ljung-Box test. The trend of norovirus positive rate in oysters fitted by the model was basically consistent with the trend of actual detection rate, the mean absolute error (MAE) was 4.85 and the mean absolute percentage error (MAPE) was 30.25. The positive detection rates of norovirus in oysters predicted by the optimal model in the next four quarters were 31.89%, 12.80%, 9.47%, and 6.14%, respectively. **Conclusion** The seasonal ARIMA model (0, 1, 1) (0, 1, 0)<sub>4</sub> can fit the trend of positive detection rate of norovirus in oysters. This model has certain significance for the risk assessment of aquatic products such as oysters contaminated by norovirus. It also has certain significance for the prevention and control of the norovirus epidemic.

**Key words:** Autoregressive integrated moving average model; norovirus; detection rate; time series analysis; prediction

收稿日期:2020-09-30

基金项目:“十三五”国家重点研发计划重点专项(2017YFC1600703);国家自然科学基金(31601570)

作者简介:杨明树 男 博士研究生 研究方向为食源性病毒 E-mail: yangmingshu5266@qq.com

董蕾 女 硕士研究生 研究方向为食源性病毒 E-mail: donglei0668@163.com

通信作者:喻勇新 男 高级工程师 研究方向为食源性病毒 E-mail: yxyu@shou.edu.cn

诺如病毒 (Norovirus, NoV) 是一种感染哺乳动物的无囊膜包被的单链 RNA 病毒,属于杯状病毒科<sup>[1]</sup>,是引起非细菌性急性肠胃炎的主要病原,也是食源性疾病的主要病原之一<sup>[2]</sup>。根据其完整的主要结构蛋白(VP1)基因和依赖 RNA 的 RNA 聚合酶(RdRP)基因的核苷酸多样性,被分为 10 个基因类群(Genogroup)、60 种基因型(Genotype)<sup>[3]</sup>,其中 GI, GII, GIV, GVIII 和 GIX 可感染人。患者临床表现包括恶心、呕吐、腹痛、腹泻和发烧等症状<sup>[1]</sup>,目前,仍缺乏获得批准的针对 NoV 的疫苗或抗病毒药物<sup>[4]</sup>。

牡蛎被认为是病毒和病原细菌的载体<sup>[5]</sup>。其通过滤食将随动物或人的粪便进入环境中的 NoV 富集于鳃和消化腺等组织上。NoV 疫情暴发经常与人们食用了被病毒污染的牡蛎等贝类有关<sup>[6]</sup>。Cheng 等<sup>[7]</sup>利用反转录聚合酶链式反应(Polymerase chain reaction, PCR)技术检测了 11 个国家的 507 份牡蛎样品,结果显示 53 份样品检出 NoV,检出率为 10.5%。基于欧盟食品和饲料快速预警系统(Rapid Alert system for Food and Feed, RASFF)分析,欧洲 2000—2010 年间的 33 例 NoV 引发的食源性疾病中,有 22 例与食用被 NoV 污染的贝类有关,占总数的 66.7%<sup>[8]</sup>。NISHIDA 等<sup>[9]</sup>收集了日本 1 512 个生食牡蛎样品,发现约有 75 份牡蛎检出 NoV,属于不同 NoV 基因型。显然, NoV 污染已成为贝类食品安全的主要隐患,需进行常规性监控及预警分析。

目前,鲜有根据牡蛎中检测出的 NoV 阳性率对 NoV 的暴发进行预测分析的研究报道。季节性自回归差分移动平均(Autoregressive integrated moving average, ARIMA)模型是一种时间序列的预测方法,其基本思想是将一个随时间推移而形成的数列视为一组随机序列,用数学模型对其进行描述,从而根据已发生的既往序列值预测未来值<sup>[10]</sup>。本研究基于 ARIMA 模型分析并预测了上海市售牡蛎中 NoV 的检出率,研究结果有助于对牡蛎中 NoV 的污染进行风险评估,为预防和控制因牡蛎中 NoV 的污染而引起的疫情暴发提供理论依据。

## 1 材料与方法

### 1.1 数据采集

2016 年 6 月—2019 年 11 月,每半月在芦潮港水产市场购买牡蛎样品。2016 年 6 月—2019 年 5 月,购买牡蛎样品 451 只,2019 年 6 月—2019 年 11 月,购买牡蛎样品 80 只。采用巢式聚合酶链式反应(Nest polymerase chain reaction, Nest-PCR)<sup>[11]</sup>

检测牡蛎样品中的 NoV,将检测结果按照季度(3—5 月,6—8 月,9—11 月,12 月—次年 2 月)进行划分并计算 NoV 阳性率,将计算得到的结果用于模型预测。

### 1.2 方法

季节性 ARIMA 模型可将时间序列中的季节部分和非季节部分相结合,表达公式如下:

$$ARIMA(p, d, q)(P, D, Q)_m^{[10]}$$

大写字母来标记模型中季节性的部分,小写字母来标记非季节性的部分。模型的季节性部分中含有和非季节性非常类似的项,但包含了季节性时段的回溯。模型中  $d$  为非季节差分阶数,  $p$  和  $q$  分别为非季节自回归阶数和非季节移动平均阶数,  $D$  是季节差分的阶数,  $P$  和  $Q$  分别是季节自回归阶数和季节移动平均阶数,  $m$  是季节长度。

建模的过程如下:

(1)数据平稳化。ARIMA 模型建立的前提条件是要求时间序列具有平稳性。通过做出 2016 年 6 月—2019 年 5 月牡蛎中 NoV 阳性检出率的原始序列图对其进行直观判断,并通过 ADF (Augmented Dickey-Fuller) 检验来验证序列的平稳性。若序列不平稳,则使用数据变换或差分的方法使数据平稳。

(2)模型选择和拟合。结合数据平稳化处理,对预处理后达到平稳性要求的序列,绘制自相关(ACF)图和偏相关(PACF)图,根据 ACF 图和 PACF 图截尾或拖尾的情况对模型进行初步识别和定阶。

若 ACF 图为拖尾, PACF 图为截尾,则适用 AR( $p$ )模型, PACF 截尾的值则为  $p$  值。反之,则选择 MA( $q$ )模型, ACF 图截尾的值是  $q$  值。若 ACF 和 PACF 图均拖尾,则考虑 ARIMA( $p, d, q$ )模型。参数  $P$  和  $Q$  的值较难判断,一般采用从低阶到高阶逐步尝试的方法,取值通常不超过 2。

在 R 软件中用 `arima` 函数拟合 ARIMA 模型。函数可返回移动平均项的系数,以及赤池信息量准则(AICc)系数,再通过对比备选模型的 AICc 值确定 ARIMA 中的参数,选择值最小的模型为最佳 ARIMA 模型。

(3)模型诊断。对最优模型进行 Ljung-Box 检验,可知残差之间是否有相关性。然后,利用最优模型对 2019 年 6—11 月牡蛎中 NoV 的检出率进行预测。比较 2019 年 6—11 月实际牡蛎中 NoV 检出率是否在预测检出率范围内。

(4)模型预测。利用最优模型对 2019 年 12 月—2020 年 11 月未来四个季度上海市售牡蛎中 NoV 的检出率进行预测。

### 1.3 统计学分析

使用 Excle 2019 汇总 2016 年 6 月—2019 年 11 月每个季度牡蛎样品中 NoV 的检出率,运用 R 语言软件的相关模块建立 ARIMA 模型。通过 ADF 检验序列的平稳性,以  $P < 0.05$  为差异有统计学意义;通过 Box-Ljung 检验残差之间的相关性,以  $P < 0.05$  为差异有统计学意义。

## 2 结果与分析

### 2.1 牡蛎阳性检出率时间序列趋势

对牡蛎中 NoV 阳性率进行分析(图 1),发现每年的冬季均是 NoV 检出率的高峰,序列以年为单位有明显的周期性。阳性检出率为 3.3%~35.2%,数据的波动范围较大。经 ADF 检验, $P = 0.7158$  拒绝原假设,故序列为非平稳序列。

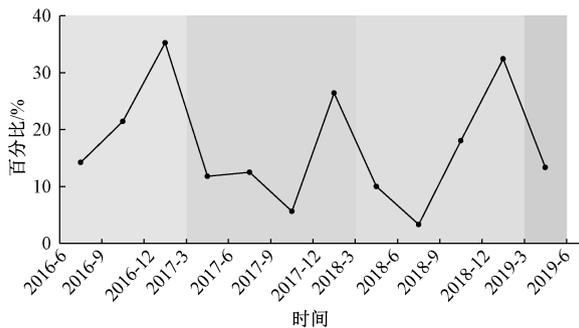


图 1 2016 年 6 月—2019 年 5 月牡蛎中 NoV 的阳性率  
Figure 1 The positive rate of norovirus in oysters from June 2016 to May 2019

### 2.2 数据平稳化

选取 2016 年 6 月—2019 年 5 月的数据作为序列的原始数据。由于数据稳定性较差,为使时间序列达到平稳性要求,对该序列进行非季节性 1 阶差分见图 2。

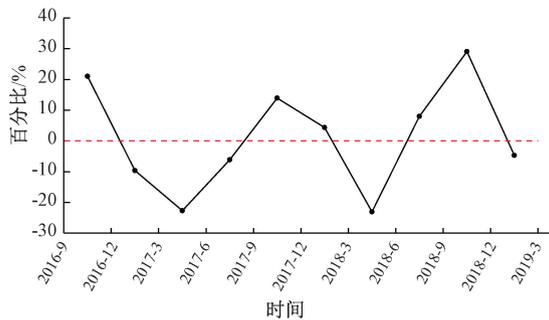


图 2 2016 年 6 月—2019 年 5 月牡蛎中 NoV 阳性率经一次差分后的时间序列

Figure 2 Time series of the positive rate of norovirus in oysters from June 2016 to May 2019 after a difference

牡蛎中的检出率呈现一定的季节性,因此再进行周期为 4 的季节差分(图 3)以消除季节性,结果如下。

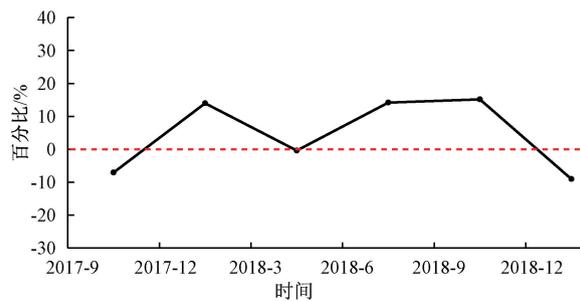


图 3 2016 年 6 月—2019 年 5 月牡蛎中 NoV 阳性率经数据平稳化后的时间序列

Figure 3 Time series of NoV positive rate in oysters from June 2016 to May 2019 after data stabilization

经 ADF 检验  $P < 0.001$  有统计学意义,认为数据平稳差分后的数据满足平稳性要求,可以进行建模。

### 2.3 模型的选择和评价

绘制数据平稳化后的 ACF 和 PACF,如图 4 所示。

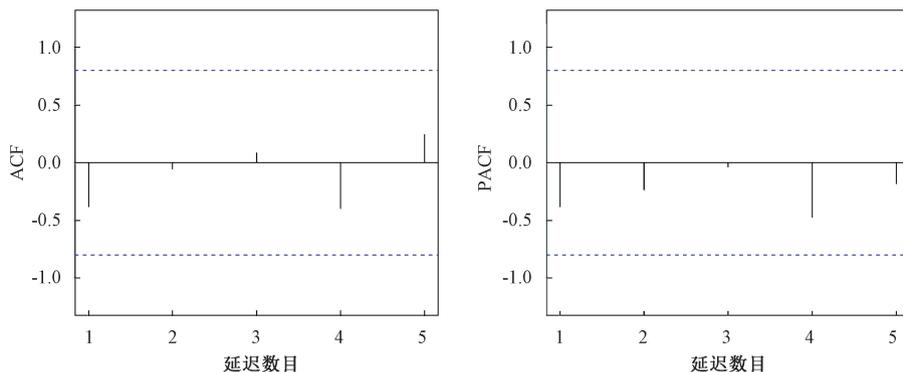


图 4 平稳化序列的自相关图和偏自相关图

Figure 4 ACF and PACF of stationary series

对于非季节项和季节项,各做了一阶差分,所以  $d=1, D=1$ 。在滞后项为一阶时有明显的自相关,而当滞后阶数逐渐增加时,偏相关逐渐减小,所以选择

ARIMA(0,1,1)(P,1,Q)<sub>4</sub> 模型。因为 P 和 Q 不好判断,但原则上不大于 2,所以构建了 7 个备选模型(表 1),选择 AICc 值最小的模型来作为最优模型。

表1 备选 ARIMA 模型的拟合参数及 AICc

Table 1 Fitting parameters and AICc of alternative

ARIMA models	
备选 ARIMA 模型	AICc
(0,1,1)(0,1,0)	58.70
(0,1,1)(1,1,0)	61.65
(0,1,1)(0,1,1)	63.09
(0,1,1)(1,1,1)	75.65
(0,1,1)(0,1,2)	76.07
(0,1,1)(1,1,2)	117.65
(1,1,1)(0,1,0)	65.56

由表可知, ARIMA(0,1,1)(0,1,0)<sub>4</sub> 模型的 AICc 值为 58.70,是备选模型中 AICc 最小的。因此,最后选择了 ARIMA(0,1,1)(0,1,0)<sub>4</sub> 模型为最优模型。

### 2.4 模型评价

对 ARIMA(0,1,1)(0,1,0)<sub>4</sub> 模型进行 Ljung-Box 检验,统计量为 4.15, P=0.245,故本研究中的最优模型未通过显著性检验,差异无统计学意义,残差之间不存在自相关性,为白噪声序列。利用最优模型对 2019 年 6—8 月和 9—11 月牡蛎中 NoV 的检出率进行预测(图 5)。

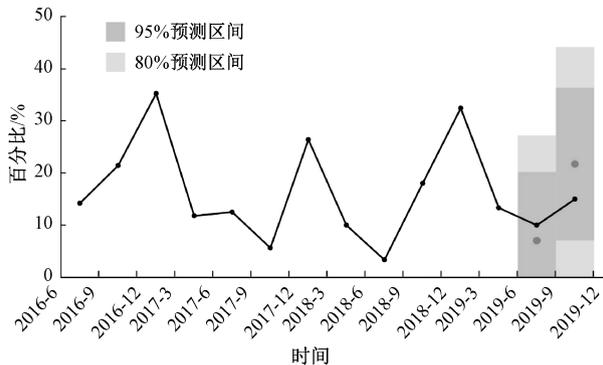


图5 牡蛎中 NoV 的检出率 ARIMA 模型预测值与实际值拟合效果图

Figure 5 ARIMA predicted value and actual value of NoV positive rate in oyster

预测的 2019 年 6—8 月牡蛎中 NoV 检出率为 7.05%,2019 年 9—11 月的牡蛎中 NoV 检出率为 21.74%。与 2019 年 6—11 月实际牡蛎中 NoV 检出率比较发现,都在预测检出率 95% 的预测区间以内,且平均绝对误差(MAE)为 4.85,平均绝对百分比误差(MAPE)为 30.25,说明该模型的预测值与实际值基本吻合。

### 2.5 模型预测

运用 ARIMA(0,1,1)(0,1,0)<sub>4</sub> 对 2019 年 12 月—2020 年 11 月上海市售牡蛎中 NoV 的检出率进行预测(图 6)。未来四个季度预测的牡蛎中 NoV 阳性检出率结果分别为 31.89%、12.80%、9.47%、6.14%。

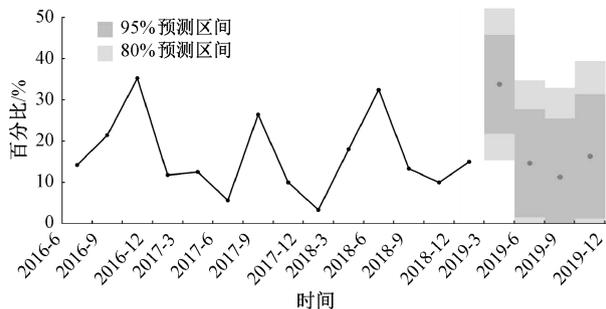


图6 牡蛎中 NoV 检出率 ARIMA 模型预测拟合图

Figure 6 Predictive fitting of ARIMA for NoV positive rate in oyster

### 3 讨论

牡蛎是 NoV 在环境中的主要传播载体, NoV 在人群中暴发常与人们误食了含有 NoV 的牡蛎有关,且具有很强的季节性<sup>[5]</sup>。若能找到牡蛎中的 NoV 在时间上的分布规律,并提前预测,可避免部分由牡蛎引起的 NoV 腹泻事件的暴发。因此,本文通过 ARIMA 模型预测上海市售牡蛎中 NoV 的检出率。本研究发现每个季节均检出含有 NoV 的牡蛎,冬季检出率最高。而 NoV 在人群中的暴发高峰也是在冬季,由此也可看出 NoV 的暴发与误食了含 NoV 的牡蛎有很大的关系。

ARIMA 模型是由 Box 与 Jenkins 于 20 世纪 70 年代提出的一种著名的时间序列预测方法,该模型已广泛应用于传染病的预测,如艾滋病<sup>[12]</sup>、猩红热<sup>[13]</sup>、疟疾<sup>[14]</sup>等疾病。用牡蛎中 NoV 的检出率这一时间序列去预测未来市售牡蛎中 NoV 的阳性率是可行的。用时间序列预测牡蛎中 NoV 的检出率只需考虑时间因素,而无需分析其他影响因素,大大降低了收集资料的成本,提高了该方法的可实现性及便捷性。

当然,牡蛎中 NoV 的检出率还可能与其生活环境的温度、盐度、降水、太阳辐射、人类活动等因素有关<sup>[5]</sup>,本研究的拟合过程中未纳入这些因素。其次,本研究中牡蛎的采集周期为半个月,由于商家无货、实验操作损失等不可抗拒因素,导致样品采集周期内存在样品量缺少等情况,因此本研究采用一个季度为周期来降低数据的不连贯或不完整性对预测结果带来的影响,可能会影响预测的精确度。另外,ARIMA 模型预测只有基于足够多的时间序列数据才能得到较满意的预测效果,且该模型更适用于短期预测,因此后期需不断更新数据,进行动态修正。除文中提到的 7 个备选模型,本研究还尝试了其他参数的模型,但并未找到比 ARIMA(0,1,1)(0,1,0)<sub>4</sub> 更

小的 AICc 值。

除 ARIMA 模型,还可尝试其他建模方法用于预测牡蛎中 NoV 的检出率,例如贝叶斯模型、回归模型、人工神经网络模型等,可能会提高市售牡蛎中 NoV 检出率预测的精准度。本研究构建的模型作为此类研究的一次尝试,相关研究结果对进行 NoV 污染的牡蛎等水产品的风险评估及 NoV 流行的预防控制具有一定的价值。

## 参考文献

- [ 1 ] ROBILOTTI E, DERESINSKI S, PINSKY B A. Norovirus [ J ]. *Clinical Microbiology Reviews*, 2015, 28( 1 ): 134-164.
- [ 2 ] MOORE M D, GOULTER R M, JAYKUS L. Human norovirus as a foodborne pathogen: challenges and developments [ J ]. *Annual Review Food Science and Technology*, 2015, 6: 411-433.
- [ 3 ] CHHABRA P, GRAAF M, PARRA G I, et al. Updated classification of norovirus genogroups and genotypes [ J ]. *Journal of General Virology*, 2019, 100( 10 ): 1393-1406.
- [ 4 ] NATALIE N E, ENOSI TUIPULOTU D, WHITE P A. Norovirus antivirals: Where are we now? [ J ]. *Medicinal Research Reviews*, 2019, 39( 3 ): 860-886.
- [ 5 ] WANG J, DENG Z Q. Detection and forecasting of oyster norovirus outbreaks: Recent advances and future perspectives [ J ]. *Marine Environmental Research*, 2012, 80: 62-69.
- [ 6 ] HALL A J, EISENBART V, ETINGÜE A L, et al. Epidemiology

- of foodborne norovirus outbreaks, United States, 2001-2008 [ J ]. *Emerging Infectious Diseases*, 2012, 18( 10 ): 1566-1573.
- [ 7 ] CHENG P K C, WONG D K K, CHUNG T W H, et al. Norovirus contamination found in oysters worldwide [ J ]. *Journal of Medical Virology*, 2010, 76( 4 ): 593-597.
- [ 8 ] DICAPRIO E, MA Y M, PURGIANTO A, et al. Internalization and dissemination of human Norovirus and animal caliciviruses in hydroponically grown romaine lettuce [ J ]. *Applied and Environmental Microbiology*, 2012, 78( 17 ): 6143-6152.
- [ 9 ] NISHIDA T, NISHIO O, KATO M, et al. Genotyping and quantitation of noroviruses in oysters from two distinct sea areas in Japan [ J ]. *Microbiology and Immunology*, 2007, 51 ( 2 ): 177-184.
- [ 10 ] HYNDMAN R, ATHANASOPOULOS G. Forecasting: principles and practice, 2nd edition [ M ]. Melbourne, Australia: Monash University, 2018: 1-384.
- [ 11 ] 贾添慧,董蕾,王永杰,等. 牡蛎中 G II 型诺如病毒巢式 PCR 检测方法的优化与评价 [ J ]. *上海海洋大学学报*, 2020.
- [ 12 ] 王淑霞,李明阳,刘希波,等. ARIMA 模型在全国艾滋病新发现病例数预测中的可行性研究 [ J ]. *中国艾滋病性病*, 2020, 26( 7 ): 705-708.
- [ 13 ] 孔德川,潘浩,郑雅旭,等. ARIMA 模型在上海市猩红热发病率预测中的应用 [ J ]. *实用预防医学*, 2020, 27 ( 8 ): 1011-1013.
- [ 14 ] 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.

## · 专家寄语 ·

《中国食品卫生杂志》是中华预防医学会、中国卫生信息与健康医疗大数据学会主办的中文核心期刊,是全国食品卫生、安全工作者的高层次交流平台,也是反映我国食品安全领域前沿科学进展的重要窗口,影响因子在预防医学领域名列前茅。杂志以习近平新时代中国特色社会主义思想为指导,把保护人民健康,保障食品安全放在首位,秉承正确办刊理念,大力宣传习近平总书记“最严谨的标准、最严格的监管、最严厉的处罚、最严肃的问责”的要求,及时报道食品安全风险监测、风险评估和食源性疾病预防等方面的学术新成就、新经验,带动全国卫生健康、农业、市场监管、进出口管理等食品安全管理部门水平的提升,也是相关科研机构、高等院校、食品相关行业研究和发展的风向标。其鲜明的风格特色和多年办刊铸就的权威性、指导性和影响力,获得了业界的高度评价。按照党的十九届五中全会以来习近平总书记提出的“把握新发展阶段,贯彻新发展理念,构建新发展格局”的指示精神,希望《中国食品卫生杂志》继续服务于健康中国 and 食品安全战略,聚焦新发展阶段食品安全领域的前沿进展和后疫情时代产业创新调整变化带来的挑战,面对新时代人民群众对食品安全和营养健康不断增长的需求,坚持创新、协调、绿色、开放、共享的发展理念,牢记习近平总书记对新闻宣传工作的总要求,把《中国食品卫生杂志》办成导向正确、技术先进、信息量大、覆盖广泛、影响力强的一流杂志,为保障人民舌尖上的安全做出新的更大的贡献!