·公共卫生与预防医学研究·

基于SARIMA模型的南京地区蚊虫侵害的预测研究

孙燕群[1,2*]，张守刚[1]，陆墨原[1,3]，张　艳[1]，潘衍宇[4]，王　冲[1]，吴起新[1]，姚美雪[5]，李成国[1]

[1]南京市疾病预防控制中心，南京医科大学附属南京疾病预防控制中心，江苏　南京　210003；[2]军事科学院军事医学研究院微生物流行病研究所，病原微生物生物安全国家重点实验室，北京　100071；[3]东南大学公共卫生学院，江苏　南京　210009；[4]中国疾病预防控制中心传染病预防控制所，北京　102206；[5]徐州医科大学公共卫生学院流行病与卫生统计学系，江苏　徐州　221004

[摘　要]　**目的**：基于南京地区蚊虫侵害密度，建立季节性差分自回归移动平均模型预测方法，为进一步防控蚊媒病和开展爱国卫生运动提供新的思路和方法。**方法**：应用季节性差分自回归移动平均模型对2019年蚊虫密度进行预测。**结果**：拟合后的预测模型为ARIMA(2,1,0)(1,1,0)$_{12}$，模型残差序列为白噪声，模型预测拟合后R^2=0.907。**结论**：模型预测的拟合效果较好，说明ARIMA模型适用于开展蚊虫侵害预测研究。

[关键词]　蚊虫侵害；季节性；差分自回归移动平均；预测

[中图分类号]　R384.1　　[文献标志码]　A　　[文章编号]　1007-4368(2022)01-108-04

doi：10.7655/NYDXBNS20220120

蚊类（Mosquito）属于昆虫纲（Insecta）、双翅目（Diptera）、长角亚目（Nematocera）、蚊科（Culicidae）。蚊虫最重要的生态习性是刺叮吸血，从而传播多种传染病[1-3]。由于全球气候、环境、交通、城市化的变化以及昆虫自身抗药性等因素，蚊虫等病媒生物对人类的威胁不断上升，重大病媒生物传播疾病总共约占全球传染性疾病负担的17%，每年造成70多万人死亡，其中以蚊虫造成的疾病负担最重[4]。蚊虫种类多，分布广，适应能力强，易产生抗药性，其侵害和控制已成为世界关注的焦点。

差分自回归移动平均（differential autoregressive moving average，ARIMA）模型是由美国统计学家Box和英国统计学家Jenkins于20世纪70年代初提出的时间序列分析、预测和控制的方法，又称Box-Jenkins法，主要用于拟合具有平稳性或者可以被转换为平稳序列的时间序列，结合了自回归和移动平均的长处，具有不受数据类型束缚和适用性强的特点[5-6]，广泛应用在公共卫生领域中的疫情预警预测等[7-9]，在病媒生物侵害密度领域应用不多。本研究利用南京地区长期、连续、系统开展的蚊虫侵害横断面调查数据，运用季节性差分自回归移动平均（SARIMA）模型进行数据拟合，对本地区蚊虫侵害数据开展预测研究，为进一步防控蚊媒病和开展爱国卫生运动提供新的思路和方法。

1　资料和方法

1.1　资料

蚊虫侵害数据来源于江苏省疾病预防控制综合业务集成平台中的病媒生物监测网络直报系统，包括监测点编号、监测时间、蚊虫数量等。

1.2　方法

1.2.1　研究变量

以蚊虫密度（D）为主要研究对象，其计算公式为：D=Nm/(Nl·T)，式中：Nm为蚊虫数量，单位为只；Nl为灯的数量，单位为灯；T为诱蚊小时数，单位为h。

1.2.2　数据处理

以月为时间段，统计2015年1月—2019年12月南京地区南京蚊虫密度，其中每年的1月、2月和12月为非蚊虫活动时间，蚊虫密度以0填充。选择2015年1月—2018年12月数据为训练集，2019年1—12月数据为验证集。

1.2.3　模型预处理

纯随机性检验：纯随机性检验又称白噪声检验，是专门用来检验序列是否为纯随机的方法，使用Box.test中的Box-Pierce函数进行检验。单位根检验是对序列是否平稳的检验，一般使用ADF检验

[基金项目]　南京市卫生科技发展专项资金（YKK17200，YKK18178）；南京市第十周期医学重点专科（传染病预防控制）

*通信作者（Corresponding author），E-mail：623528991@qq.com

进行判断。

1.2.4 模型构建

SARIMA是ARIMA中的一种，主要分为季节乘积和季节相加两种，以季节乘积为多，主要处理序列的季节效应、长期趋势效应和随机波动之间存在复杂的交互影响关系。SARIMA乘积模型公式为：$\Phi_P(B^s)\varphi(B)\nabla_S^D\nabla_{xt}^d=\delta+\Theta_Q(B^s)\theta(B)wt$，也可以简化为SARIMA(p,d,q)×(P,D,Q)s，其中，p、d、q分别表示短期相关模型的自回归、趋势差分、移动平均的阶数，P、D、Q分别表示季节趋势的自回归、趋势差分、移动平均的阶数，s表示周期时间间隔。使用Box.test中的Ljung-Box函数进行残差白噪声检验。经过残差诊断合格后可以对模型进行预测。

1.2.5 模型评价

选择平均误差(mean error，MSE)、均方根误差(root mean squared error，RMSE)、平均绝对误差(mean absolute error，MAE)、平均绝对比例误差(mean absolute sacled error，MASE)和决定系数R^2进行模型拟合效果的评价。

1.3 统计学方法

研究人员将蚊密度按照时间2015年1月—2019年12月进行整理，用Excel2013进行汇总，建立时间序列，利用R x64 4.0.3软件及下载的forecast、tseries包进行统计学分析，$P<0.05$为差异有统计学意义。

2 结 果

2.1 序列概况

绘制2015—2018年蚊虫侵害数据的时间序列分解图(图1)，可以看出蚊虫侵害数据时间序列季节效应较明显，而长期趋势不明显。

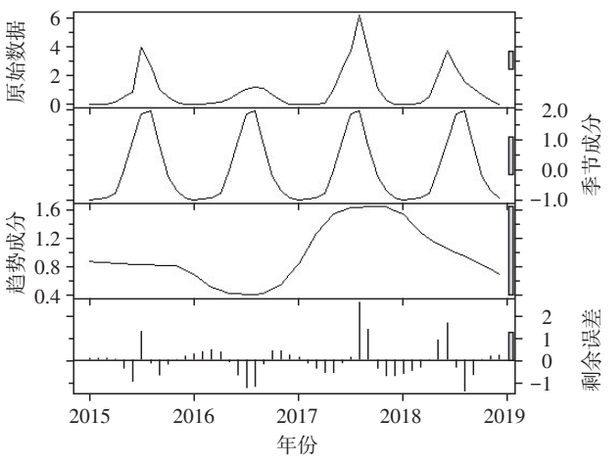

图1 2015年1月—2018年12月时间序列分解图

2.2 SARIMA模型构建

2.2.1 模型准备阶段

经Box-Pierce函数进行检验，序列为非白噪声序列($P<0.01$)，经ADF检验不平稳(Dickey-Fuller=−2.275，P=0.466)，后经一阶差分后ADF检验平稳(Dickey-Fuller=−3.853，$P<0.05$)。

2.2.2 SARIMA模型预测

参数判断：分别对2次差分绘制自相关函数(auto correlation function，ACF)图和偏自相关函数(partial ACF，PACF)图，差分后的ACF图和PACF图见图2。1阶差分的自相关图非截尾，q考虑取0，偏自相关图为2阶截尾，p考虑可取1和2；季节差分之后的自相关图呈现显著非零，Q考虑取0，偏自相关图在lag12处有spike，P考虑取1。利用可能的参数，分别建立ARIMA$(1,1,0)(1,1,0)_{12}$或者ARIMA$(2,1,0)(1,1,0)_{12}$模型，计算最小信息量准则(AIC)值，其中前者AIC=114.57，后者AIC=113.54，综合考虑ARIMA$(2,1,0)(1,1,0)_{12}$模型。

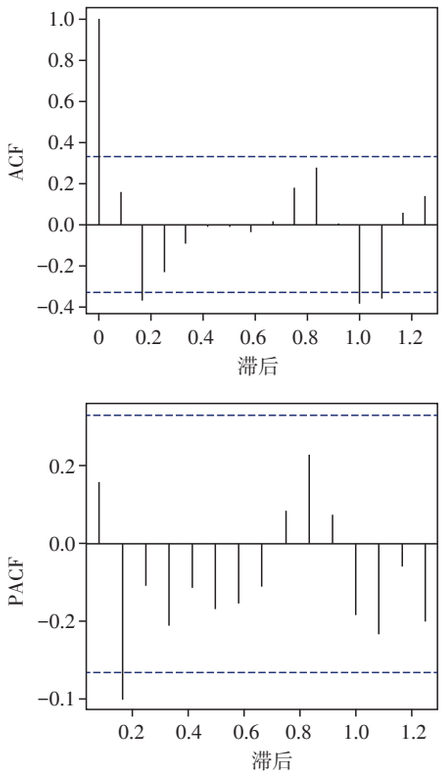

图2 经1阶和1次季节差分后时间序列自相关图和偏自相关图

模型诊断：对ARIMA$(2,1,0)(1,1,0)_{12}$进行Ljung-Box残差检测，可以认为该模型残差序列不存在自相关，为白噪声序列(χ^2=0.079，P=0.778)。

模型预测：用forecast函数预测2019年1—12月

的蚊虫侵害密度，预测结果见图3。

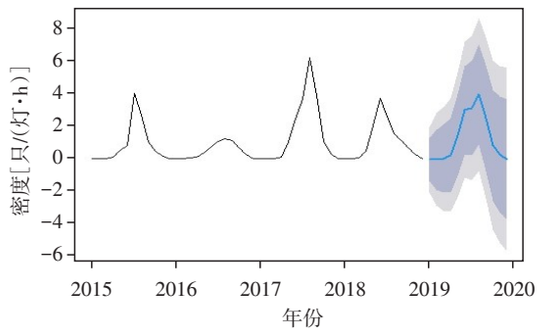

图3 2015年1月—2018年12月实际时间序列及2019年1—12月预测序列

2.2.3 SARIMA模型评价

SARIMA预测的2019年1—12月结果与实际监测结果比较见表1。

表1 2019年1—12月时间序列预测与实际监测比较

月份（月）	预测密度［只/(灯·h)］	实际密度［只/(灯·h)］	绝对误差［只/(灯·h)］	相对误差
1	−0.062	0	0.062	—
2	−0.064	0	0.064	—
3	−0.013	0.144	0.157	1.093
4	0.224	0.548	0.324	0.591
5	1.439	1.658	0.219	0.132
6	2.987	2.590	−0.398	−0.154
7	3.098	3.940	0.842	0.214
8	3.959	3.046	−0.913	−0.300
9	2.499	1.988	−0.512	−0.258
10	0.829	0.846	0.017	0.020
11	0.245	0.681	0.437	0.641
12	−0.050	0	0.050	—

根据预测密度和实际密度，计算ME、RMSE、MAE、MASE和R^2。计算如下：ME=0.029 074，RMSE=0.441 683，MAE=0.332 771，MASE=0.517 030，R^2=0.907。结果表明预测精度较高，预测与监测曲线拟合见图4。

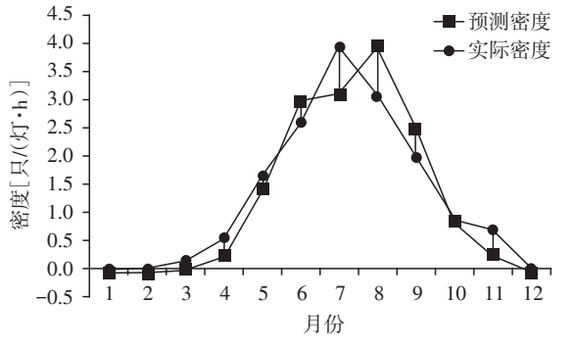

图4 2019年1—12月时间序列预测和监测结果拟合曲线

3 讨 论

蚊虫侵害受到多种因素，如气温、湿度、气压等气象因素以及植被、水网、土地利用模式等地理因素的影响。由于不断进行的城市化进程和人员交通物流持续增强的社会化活动，各种社会性因素也在影响着蚊虫侵害问题。人民对于日益增长的美好生活的要求，对于人居环境改善的迫切需求等，使得蚊虫侵害显然不仅仅是一个疾病预防的问题，而是逐渐演变成一个公众关注的公共卫生问题，因此要更加关注蚊虫侵害的现状和未来，更加注重蚊虫控制的效果评价，更加注重蚊虫和蚊媒病的相关性研究[10-13]。

蚊虫侵害密度监测是国内开展较为成熟的横断面研究，方法较为系统和成熟，有长时间的工作积累和分析方法，但大部分都局限在横断面分析，如对蚊虫的密度、种群、群落结构和季节消长进行描述性研究，仅有少部分开展前瞻性的预测预警研究，方法较为单一。SARIMA方法是ARIMA模型中对于季节变化趋势较为明显的时间序列进行研究的方法，相对于简单的AR或者MA模型，SARIMA加入了季节性变化因素，使模型分析更为精准。有学者利用ARIMA模型对三带喙库蚊或白纹伊蚊密度开展过类似研究，预测的拟合效果均较好，说明ARIMA模型适用于开展蚊虫侵害预测研究[14-16]。从本研究来看，SARIMA对于本地区蚊虫侵害的预测精度较高，决定系数R^2达到了0.9以上，预测的成功率大大提升，对提前进行蚊媒病防制动员和爱国卫生运动具有前瞻性意义。

但是也要看到SARIMA模型仍是一种传统的线性时间序列模型，预测时仅仅考虑从历史看未来，而没有加入可能影响未来的各种因素（如上述所说的气象环境、社会因素等），因此本研究预测的高峰值与实际高峰值有所差别，所以这是本研究之后要考虑开展的方向。如气象因素对于蚊虫密度也有滞后效应，参照肖扬[13]的研究可以使用分布滞后非线性模型（DLNM）开展类似研究。对于时间序列中的非线性部分本文未涉及，参照国内类似研究可以在ARIMA的基础上加入支持向量机（SVM）或神经网络模型等人工智能技术[17-20]，使得预测拟合效果更加贴近真实值，这样才能真正服务于现实公共卫生工作需要。

[参考文献]

[1] 徐承龙，姜志宽. 蚊虫防制(一)——蚊虫的危害与形态分类[J]. 中华卫生杀虫药械，2006，12(4)：289-293
[2] 瞿逢伊. 我国蚊虫种质资源现状及其共享利用[J]. 中国寄生虫学与寄生虫病杂志，2006，24(z1)：13-16
[3] 张菊仙，龚正达. 中国蚊类研究概况[J]. 中国媒介生物学及控制杂志，2008，19(6)：595-599
[4] 张 仪. 新发媒传疾病及其防控[J]. 中国血吸虫病防治杂志，2012，24(5)：501-504
[5] 王 燕. 时间序列分析-基于R[M]. 北京：中国人民大学出版社，2015：142-161
[6] KABACOFF R I. R语言实战[M]. 北京：电子工业出版社，2015：315-341
[7] 游楠楠，刘 巧，李忠奇，等. 基于ARIMA模型的江苏省不同地区肺结核发病趋势的预测[J]. 南京医科大学学报(自然科学版)，2020，40(6)：909-914，919
[8] 运 玲，王福才，张秋芬. 差分自回归移动平均模型在蚊密度分布特征预测中的应用[J]. 中国媒介生物学及控制杂志，2020，31(1)：21-26
[9] 黄玉萍，傅伟杰，熊长辉，等. ARIMA模型在江西省布鲁氏菌病发病数预测中的应用[J]. 中国人兽共患病学报，2020，36(3)：202-205
[10] 何亚明. 三峡库区万州段蚊虫生态、气象因素及蚊媒病的相关性研究[D]. 重庆：陆军军医大学，2018
[11] 王利亚. 我国流行性乙型脑炎时空分布特征及风险预测研究[D]. 北京：军事医学科学院，2014
[12] 余向华. 蚊媒传染病流行特征及气象影响因素研究[D]. 杭州：浙江大学，2007
[13] 肖 扬. 广州白纹伊蚊分布及与气象因素和登革热发病的关联性研究[D]. 广州：广东药学院，2017
[14] 肖 珊，陈立章，龙建勋，等. 基于R语言自回归积分移动平均模型在长沙市三带喙库蚊密度预测中的应用[J]. 医学动物防制，2020，36(3)：278-281
[15] 赵克昌，杨金煜，卢岩松，等. 基于自回归积分滑动平均模型的西部某市白纹伊蚊幼虫密度预测[J]. 中华卫生杀虫药械，2018，24(1)：36-39
[16] 潘衍宇，吴海霞，国 佳，等. 基于R语言自回归积分移动平均模型的广州市白纹伊蚊密度预测研究[J]. 中国媒介生物学及控制杂志，2018，29(6)：545-549
[17] 张雪凝，施学忠，赵 浩，等. SARIMA和SARIMA-GRNN模型在流行性腮腺炎发病率预测中的应用对比[J]. 中国卫生统计，2020，37(4)：489-492
[18] 李文瀚. 山西省人间布病的流行特征及基于ARIMA-ERNN组合模型预测效果研究[D]. 太原：山西医科大学，2019
[19] 尤玉玲，李 娟，高孙玉洁，等. ARIMA模型和SVM模型联合在感染性腹泻发病预测中的应用[J]. 医学动物防制，2020，36(5)：432-435
[20] 黄国宝，黎衍云，吴 菲，等. ARIMA模型和ARIMA-SVM模型对上海市2型糖尿病患者肺结核发病的预测效果[J]. 复旦学报(医学版)，2020，47(6)：899-905

[收稿日期] 2020-10-26
(本文编辑：蒋 莉)