

求和自回归移动平均模型在通辽市成蚊密度预测中的应用

邵华, 布仁巴图, 秦忠良, 商娜, 倪晓娜, 张志平, 李莹盈

通辽市疾病预防控制中心, 内蒙古 通辽 028000

摘要: 目的 分析通辽市不同年份成蚊密度, 通过拟合求和自回归移动平均模型 (autoregressive integrated moving average, ARIMA) 对未来蚊虫密度进行预测。方法 选用诱蚊灯法监测通辽市 2017—2021 年不同生境成蚊密度, 根据监测结果, 建立 ARIMA 模型, 对 2022 年成蚊密度进行预测。结果 2017—2021 年通辽市各监测点平均蚊密度为 7.91 只/(灯·夜)。其中淡色库蚊为优势蚊种。在五类生境中, 除 2017 年农户蚊密度较高外, 其他年份都是牲畜棚密度较高。每年成蚊密度均为单峰曲线, 除 2017 年高峰出现在 7 月份外, 其余年份高峰均出现在 8 月, 根据 2017—2021 年蚊虫密度结果, 拟合 ARIMA(1, 1, 1) × (1, 1, 0)₁₂ 模型, 残差序列为白噪声序列 ($Q=14.498, P=0.488$), 用此模型预测 2022 年的成蚊密度, 5—10 月份分别为 8.12、7.48、13.79、29.31、22.08 和 12.37 只/(灯·夜)。结论 利用 2017—2021 年的数据建立 ARIMA 模型, 能够预测 2022 年的成蚊密度和季节消长趋势, 为进一步蚊媒传染病风险评估提供理论数据支持。

关键词: 蚊密度; 预测; 求和自回归移动平均模型

中图分类号: R184.31 文献标识码: A 文章编号: 1006-3110(2023)02-0242-04 DOI: 10.3969/j.issn.1006-3110.2023.02.030

Application of autoregressive integrated moving average model to the prediction of mosquito density in Tongliao City

SHAO Hua, BUREN Ba-tu, QIN Zhong-liang, SHANG Na, NI Xiao-na, ZHANG Zhi-ping, LI Ying-ying

Tongliao Municipal Center for Disease Control and Prevention, Tongliao, Inner Mongolia 028000, China

Abstract: **Objective** To analyze the mosquito density in Tongliao City in different years, and to predict the future density of mosquito by autoregressive integrated moving average (ARIMA) model. **Methods** The mosquito trap lamp method was selected to monitor the adult mosquito density in different habitats in Tongliao City from 2017 to 2021. According to the monitoring results, an ARIMA model was established to predict the adult mosquito density in 2022. **Results** From 2017 to 2021, the average mosquito density at each monitoring site in Tongliao City was 7.91 mosquitoes/(light·night). Among them, *Culex pipiens pallens* was the dominant mosquito species. Among the five habitats, except for the high mosquito density of peasant household in 2017, the density of livestock sheds in other years was high. The annual adult mosquito density was a single peak curve. Apart from the peak in 2017 occurred in July, the peak in other years occurred in August. ARIMA (1, 1, 1) × (1, 1, 0)₁₂ model was fitted according to the results of mosquito density from 2017 to 2021, and the residual sequence was white noise sequence ($Q=14.498, P=0.488$). This model was used to predict the adult mosquito density in 2022, which were 8.12, 7.48, 13.79, 29.31, 22.08 and 12.37 mosquitoes/(light·night) from May to October, respectively. **Conclusion** The ARIMA model established based on the data from 2017 to 2021 can predict the density and seasonal fluctuation of adult mosquitoes in 2022 so as to provide theoretical data support for further risk assessment of mosquito-borne infectious diseases.

Keywords: mosquito density; forecast; autoregressive integrated moving average model

利用统计分析方法, 对时间序列的数据进行分析, 能够预测变量或事物的变化趋势。在公共卫生领域已经得到有效的应用, 并取得了较好效果^[1]。求和自回归移动平均模型 (autoregressive integrated moving average,

ARIMA) 通过捕捉序列数据中时间序列趋势来控制非平稳性 (non-stationarity) 或季节性 (seasonality), 以及控制序列的自相关^[2], 从而根据建立的模型来预测未来^[3]。蚊分布广, 种类多, 是最重要的医学昆虫类群。蚊类不仅吸血骚扰, 更重要的是传播多种疾病, 蚊子传播的疾病达 80 种之多^[4]。蚊媒传染病的发生及发病率与蚊虫密度存在着正相关性, 通过对蚊种群信息和密度进行监测, 能够为蚊媒传染病的发生提供预险预

基金项目: 内蒙古自治区科技计划项目 (2021GG0312)

作者简介: 邵华 (1984-), 男, 内蒙古赤峰人, 硕士研究生, 副主任医师, 研究方向: 流行病与卫生统计学。

警^[5]。本研究收集通辽市 2017—2021 年的蚊虫密度和季节消长监测数据,根据数据类型特点,选择建立 ARIMA 模型,对下一年度的蚊虫密度进行预测,从而对蚊媒传染病的发生提供预险预警,为提前制定相关防控措施提供科学依据。

1 材料与方法

1.1 监测地点和方法 2017—2021 年,按照《内蒙古自治区病媒生物监测实施方案》,采用诱蚊灯法,每年选择通辽市主城区及两个不同的旗县开展成蚊密度监测。每个监测区城镇选择居民区、医院、公园各 2 处,农村选择农户和牲畜棚(牛棚、猪圈、羊圈、养殖场等)各 2 处。每处监测点布设诱蚊灯 1 台。在成蚊活动期即每年的 5—10 月,每月上下旬各开展 1 次密度监测,相邻监测时间的间隔至少为 10 d,诱蚊灯悬挂在避风、无干扰光源的场所,光源离地 1.5 m,波长 2 537,功率为 8 W。诱蚊灯布放时间为日落前 1 h 至次日日出后 1 h,收集结束后封存集蚊袋,带回实验室将捕获的成蚊采用冰箱冷冻法处理,根据形态学特征,鉴别蚊虫种类、性别并计数。蚊虫密度计算公式为:蚊虫密度 [只/(灯·夜)] = 捕获蚊虫数量(只)/(诱蚊灯数×监测夜数)。

1.2 统计学分析 采用 Excel 2019 软件建立 2017—2021 年原始监测数据库,用 SPSS 19.0 软件进行统计分析。根据数据特点,建立 ARIMA 模型对下一年度成蚊密度进行预测。ARIMA(p, d, q)×(P, D, Q)_s 模型中:AR 是“自回归”, p 为自回归项数;MA 为“滑动平均”, q 为滑动平均项数, I 为“差分”, d 为使之成为平稳序列所做的差分次数。 P, D, Q 为季节性的阶数, s 表示季节性的周期。ARIMA 模型的建模过程:①序列特征及平稳化:应用时间序列图、自相关系数函数(autocorrelation function, ACF)图、偏自相关系数函数

(partial autocorrelation function, PACF) 图分析序列特征(趋势性、季节性)及平稳性。若序列平稳,进行平稳序列的时间序列分析;若序列不平稳,结合序列特征,采用合适的的数据预处理方法,包括数据转换、普通差分、季节差分等,实现序列平稳化后进行分析;②模型的识别:根据 ACF 图、PACF 图和模型拟合结果,进行模型的初步识别和定阶;③参数估计和模型诊断:利用非线性最小二乘法估计模型参数。参数估计后,对模型的残差序列进行白噪声检验,用以判断 ARIMA 模型的适合性。根据标准化的 BIC 准则判定模型的拟合优度(BIC 值相对较小的模型较好)。如不恰当则返回第二阶段,重新选定模型。通过调试比较,确定残差序列为白噪声,标准化的 BIC 值较小,模型较为简洁的为最佳模型;④预测应用:利用 2017—2021 年各月总蚊虫密度数据建立的模型,对下一年度各月的蚊虫密度进行预测,见图 1。

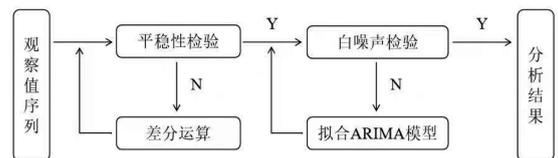


图 1 ARIMA 模型建模流程图

2 结果

2.1 蚊虫种类构成 2017—2021 年通辽市各监测点共布幼蚊灯 1 211 盏,捕获成蚊 9 579 只,平均蚊密度为 7.91 只/(灯·夜)。其中淡色库蚊 9 527 只(占 99.46%),三带喙库蚊 27 只(占 0.28%),日本伊蚊 22 只(占 0.23%),特别是 2020—2021 年连续两年均捕获到日本伊蚊,分别为 9 只、13 只。

2.2 不同年份不同生境成蚊数量及密度 在五类生境中,除 2017 年农户蚊密度较高外,其他年份都是牲畜棚密度较高,见表 1。

表 1 不同年份不同生境成蚊密度[只/(灯·夜)]

年度	捕获只数 (只)	布灯数 (盏)	总密度	居民区		公园		医院		农户		牲畜棚	
				数量	密度	数量	密度	数量	密度	数量	密度	数量	密度
2017	1 222	228	5.36	172	3.58	208	4.33	80	2.22	440	9.16	322	6.70
2018	1 419	228	6.22	169	3.52	273	5.69	13	0.36	465	9.69	499	10.40
2019	1 658	330	5.02	235	3.26	429	5.96	88	1.47	431	6.84	475	7.54
2020	1 408	134	10.51	325	9.29	298	8.51	29	1.04	269	15.82	487	30.44
2021	3 872	291	13.31	430	6.83	861	18.30	334	6.07	1 002	18.91	1 245	19.76

表 2 通辽市 2017—2021 年不同月份蚊密度[只/(灯·夜)]

年度	5 月	6 月	7 月	8 月	9 月	10 月
2017	2.58	4.93	14.26	9.26	0.68	0.45
2018	4.47	5.08	8.13	8.82	7.18	3.66
2019	0.79	3.5	7.76	8.35	6.69	3.43
2020	1.28	2.95	9.36	27.15	11.74	4.32
2021	5.81	4.61	10.80	25.78	20.10	9.84

2.3 成蚊密度季节消长 2017—2021 年,每年成蚊密度均为单峰曲线,除 2017 年高峰出现在 7 月份外,其余年份高峰均出现在 8 月,见表 2。

2.4 ARIMA 模型的建立及预测

2.4.1 平稳性检验 将 2017—2021 年通辽市蚊密度

绘制成时间序列图,每年蚊密度呈明显的周期性,季节周期为 12 个月,为非平稳时间序列,见图 2。进一步对数据进行一阶差分和一阶季节差分,差分后序列变得平稳,见图 3。

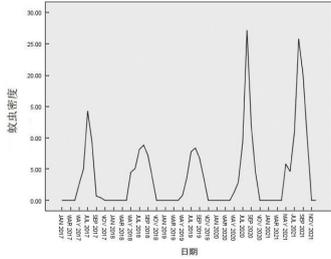


图 2 2017—2021 年通辽市蚊密度时间序列图

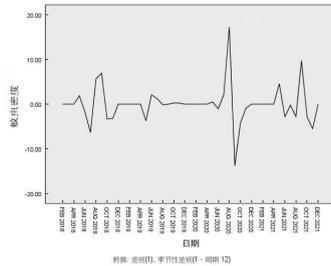


图 3 2017—2021 年通辽市蚊密度一阶差分和一阶季节差分后的时间序列图

2.4.2 模型的建立 通过一阶差分和一阶季节性差分,得出 $ARIMA(p,1,p) \times (P,1,Q)_{12}$ 模型,12 表示季节性,12 个月为一个周期,对差分后的时间序列做 ACF 图和 PACF 图,见图 4、图 5。根据 ACF、PACF 图, p, q, P, Q 从低到高选择,选择 $ARIMA(1,1,1) \times (1,1,1)_{12}$ 、 $ARIMA(1,1,1) \times (2,1,0)_{12}$ 和 $ARIMA(2,1,1) \times (1,1,0)_{12}$ 为备选模型,对此 3 个模型进行拟合优度检验(残差序列检验),结果显示 Ljung-Box Q 值分别为 14.498、14.490、10.093, P 值均 >0.05 ,表明 3 个模型的残差序列均为白噪声序列,说明 3 个模型对序列的特征提取完全。根据平稳 R^2 和标准化 BIC 值。按照 BIC 值筛选模型,BIC 值最小者为最优模型。最终选择了 $ARIMA(1,1,1) \times (1,1,0)_{12}$ 为最优模型,见表 3。对残差的 ACF、PACF 进行检验均落入 95% 可信区间(95% CI) 范围内,见图 6。

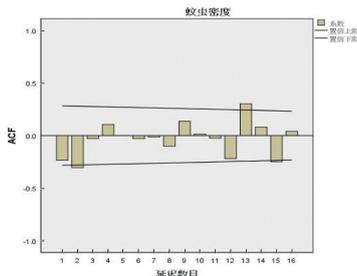


图 4 2017—2021 年通辽市蚊密度一阶差分和一阶季节差分后的 ACF 图

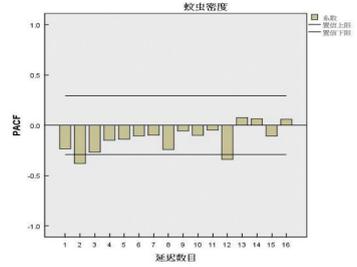


图 5 2017—2021 年通辽市蚊密度一阶差分和一阶季节差分后的 PACF 图

表 3 备选模型拟合优度检验

模型识别	BIC 值	平稳 R^2	Ljung-Box $Q(18)$		合计 P 值
			统计值	自由度	
$ARIMA(1,1,1) \times (1,1,1)_{12}$	2.844	0.371	14.498	15	0.488
$ARIMA(1,1,1) \times (2,1,0)_{12}$	2.949	0.371	14.490	14	0.414
$ARIMA(2,1,1) \times (1,1,0)_{12}$	2.867	0.421	10.093	14	0.755

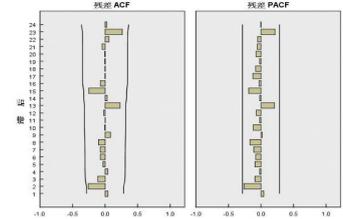


图 6 $ARIMA(1,1,1) \times (1,1,0)_{12}$ 模型残差序列的 ACF 和 PACF 图

2.4.3 模型的预测 根据建立的模型预测 2022 年的成蚊密度,5—10 月份预测值分别为 8.12、7.48、13.79、29.31、22.08 和 12.37 只/(灯·夜),见表 4。预测值时间序列图见图 7。

表 4 利用模型预测 2022 年通辽市成蚊密度[只/(灯·夜)]

时间	5 月	6 月	7 月	8 月	9 月	10 月
预测值	8.12	7.48	13.79	29.31	22.08	12.37
95%CI 上限	15.49	14.85	21.16	36.68	29.45	19.74
95%CI 下限	0.75	0.11	6.42	21.94	14.71	5.00

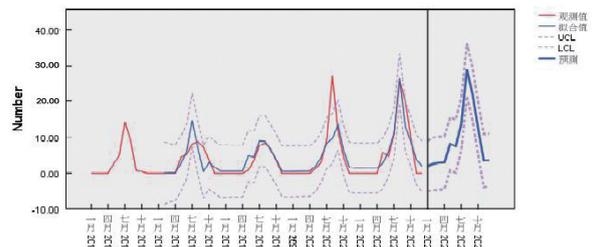


图 7 利用模型通辽市蚊虫密度预测时间序列图

3 讨论

时间序列分析就是对时间序列进行观察、研究,寻找变化发展的规律,预测将来的走势,是一种根据动态数据揭示系统动态结构和规律的统计方法。其基本思想是根据系统的有限长度运行记录(观察数据),建立能够比较精确地反映序列中所包含的动态依存关系的