

· 论 著 ·

纳入气象因素的ARIMAX模型预测流行性感胃流行趋势

吕晓丽¹, 朱一², 竹军伟¹

1. 余杭区疾病预防控制中心公共卫生应急科, 浙江 杭州 311100; 2. 杭州市疾病预防控制中心传染病防治所

摘要: **目的** 评价纳入气象因素的多元差分自回归移动平均模型 (ARIMAX) 预测流感样病例 (ILI) 流行趋势的效果, 为流行性感胃监测和预警提供参考。**方法** 收集杭州市余杭区2014年第1周—2018年第26周4家监测哨点医院上报的ILI和同期气象资料; 利用余杭区2014年第1周—2017年第52周ILI资料以及经Lasso回归模型筛选的气象变量, 建立ARIMAX模型, 预测余杭区2018年第1—26周流感样病例占门诊急症病例的比例 (ILI%), 并与实际情况比较以验证模型的预测效果。**结果** 2014年第1周—2018年第26周, 余杭区监测哨点医院共报告ILI病例60 419例, ILI%为1.29%。Lasso回归分析结果显示, 周平均绝对湿度与ILI%呈正相关 ($r=27.769$), 周平均气温与ILI%呈负相关 ($r=-0.117$)。纳入周平均绝对湿度和周平均气温建立最优模型ARIMAX (1, 0, 0) (1, 0, 0)₁₂, 贝叶斯信息准则 (BIC) 值为81.30, 平均绝对误差百分比 (MAPE) 为15.77%。采用最优模型预测余杭区2018年第1—26周ILI%, MAPE值为43.75%。**结论** 纳入气象因素的ARIMAX模型可对ILI流行趋势进行预测, 但预测精度有待提高。

关键词: 流感样病例; 气象因素; 多元差分自回归移动平均模型; 预测

中图分类号: R181.3 **文献标识码:** A **文章编号:** 2096-5087 (2021) 08-0780-04

Influenza incidence prediction based on ARIMAX model including meteorological factors

LÜ Xiaoli*, ZHU Yi, ZHU Junwei

*Department of Public Health Mergency, Yuhang Center for Disease Control and Prevention, Hangzhou, Zhejiang 311100, China

Abstract: Objective To evaluate the feasibility of autoregressive integrated moving average with explanatory variables (ARIMAX) model including meteorological factors on the prediction of influenza-like illness (ILI), so as to provide a basis for the monitoring and early warning of influenza. **Methods** The ILI data reported by four sentinel hospitals in Yuhang District of Hangzhou from the 1st week of 2014 to the 26th week of 2018 was collected, as well as the meteorological data during the same period. The ARIMAX model was established using the percentage of ILI cases in total outpatients (ILI%) data from the 1st week of 2014 to the 52nd week of 2017 and the meteorological factors selected by Lasso regression model. The ILI% from the 1st to 26th week of 2018 was predicted and compared with the actual values to verify the ARIMAX model. **Results** From the 1st week of 2014 to the 26th week of 2018, a total of 60 419 cases of ILI were reported by the four sentinel hospitals of Yuhang District, with ILI% of 1.29%. Lasso regression analysis showed that there was a positive correlation between weekly average absolute humidity and ILI% ($r=27.769$), and a negative correlation between weekly average temperature and ILI% ($r=-0.117$). The ARIMAX (1, 0, 0) (1, 0, 0)₁₂ with weekly average temperature and absolute humidity was selected as the optimal model, with the Bayesian information criterion (BIC) value of 81.30 and the mean absolute percentage error (MAPE) value of 15.77%. The MAPE value of the ARIMAX model predicting the ILI% from 1st to 26th week of 2018 were 43.75%. **Conclusion** The ARIMAX model including meteorological factors can be used to predict the prevalence of ILI, but the accuracy needs

DOI: 10.19485/j.cnki.issn2096-5087.2021.08.006

基金项目: 浙江省医药卫生科技项目 (2019YK148); 杭州市卫生科技计划 (一般) 项目 (2018B048)

作者简介: 吕晓丽, 本科, 主管医师, 主要从事传染病防控工作

通信作者: 朱一, E-mail: zhuyilxb@yeah.net

to be promoted.

Keywords: influenza-like illness; meteorological factors; autoregressive integrated moving average with explanatory variables model; prediction

流行性感胃(流感)是由流感病毒引起的急性呼吸道传染病,传播速度快且人群普遍易感,在世界范围内发病率和病死率均较高^[1-2],已成为重大公共卫生问题之一。我国于2004年建立流感监测系统,通过流感样病例(influenza-like illness, ILI)占门急诊病例的比例(ILI%)^[3]和流感确诊病例占ILI的比例(流感阳性标本检出率)2个指标反映流感的流行强度^[3]。杭州市余杭区自2010年开展流感监测工作,至2014年监测哨点医院已扩展至4家,覆盖全部片区。已有的监测体系虽可实时反映流感流行情况,但在流感发病趋势的预测预警方面仍有不足。差分自回归移动平均模型(autoregressive integrated moving average model, ARIMA model)在传染病预测预警领域应用广泛,适用于如ILI%等具有季节周期性,但时间序列特征不典型的指标预测^[4],是目前常用的流感预测方法^[5]。既往研究提示气温、相对湿度、气压、风速等气象因素与ILI相关^[6-8]。在考虑时间因素的基础上加入气象因素建立多元差分自回归移动平均模型(autoregressive integrated moving average with explanatory variables model, ARIMAX model),对短期ILI的预测精度比一元ARIMA模型更高^[9]。本研究利用杭州市余杭区2014年第1周—2017年第52周ILI%和气象监测资料建立ARIMAX模型,并利用2018年第1—26周ILI%资料评价模型预测效果,为流感发病趋势预测预警提供理论依据。

1 资料与方法

1.1 资料来源 余杭区2014年第1周—2018年第26周ILI资料来源于余杭区4家监测哨点医院。ILI诊断标准为:发热,体温 $\geq 38\text{ }^{\circ}\text{C}$,伴有咳嗽或咽痛^[10]。ILI% $(\%) = (\text{流感样病例数}/\text{门急诊病例总数}) \times 100\%$ 。余杭区2014年第1周—2017年第52周气象资料来源于余杭区气象局,收集每日平均大气压、平均气温、最高气温、最低气温、平均水汽压、平均相对湿度、最小相对湿度、降水量、平均风速和最大风速,计算得出:日平均绝对湿度=日平均水汽压 $(461 \times \text{日平均气温})$;周降水量为一周内每日降水量累加值;周平均大气压、周平均气温、周平均水汽压、周平均相对湿度、周平均风速、周平均绝对湿度为一周内该气象因素的日平均值;周最高气温、周最大风速、周最低气温、周最小相对湿度为一周内该气

象因素的极大或极小值;周变温=本周平均气温-上一周平均气温^[11]。

1.2 方法 以2014年第1周—2017年第52周余杭区ILI%和气象资料为建模数据,拟合ARIMAX模型。建模过程如下:(1)采用ADF检验(augmented Dickey-Fuller test)验证周ILI%时间序列的平稳性,结合残差序列的自相关系数函数(autocorrelation function, ACF)图和偏自相关系数函数(parital autocorrelation function, PACF)图建立ARIMA乘积季节模型。(2)采用Lasso回归模型筛选气象因素,剔除系数与0接近,选择系数绝对值较大的气象因素;采用ADF检验验证平稳性,对非平稳序列进行差分使其平稳化,结合互相关函数(cross-correlation function, CCF)图查找最佳滞后阶数^[12]。以延迟数目大于0,CCF值超过2倍标准差为可能存在相关性。(3)将ILI%作为响应序列,不同滞后阶数的气象因素作为输入序列建立ARIMAX模型。选取残差序列为白噪声,贝叶斯信息准则(Bayesian information criterion, BIC)最小和平均绝对误差百分比(mean absolute percentage error, MAPE)在10%~20%的模型为最优模型^[13]。(4)利用ARIMAX模型预测2018年第1—26周的ILI%,比较实际值与预测值并评价预测误差。

1.3 统计分析 采用Excel 2007软件建立数据库,采用SPSS 22.0软件描述数据,采用R.3.6.0软件建立模型并进行统计分析。检验水准 $\alpha=0.05$ 。

2 结果

2.1 余杭区周ILI%分布 2014年第1周—2018年第26周,余杭区监测哨点医院共报告门急诊病例4 682 823例,ILI为60 419例,ILI%为1.29%。2018年第3周ILI%最高,为3.93%;2015年第9周ILI%最低,为0.35%。2014—2017年余杭区周ILI%呈明显周期性,每隔12周或24周出现一个高峰。

2.2 余杭区气象资料分析 Lasso回归分析结果显示,周平均绝对湿度与ILI%呈正相关($r=27.769$),周平均气温与ILI%呈负相关($r=-0.117$)。滞后0~5阶(lag0~lag5)的周平均绝对湿度和滞后0~2阶(lag0~lag2)的周平均气温对ILI%有影响,见图1和图2。ADF检验结果显示,周平均绝对湿度为平稳序列(ADF=-4.469, $P=0.010$),周平均气温为非平

稳序列 (ADF=-3.098, P=0.116), 对周平均气温进行 1 阶拆分后得到平稳序列 (ADF=-4.469, P=0.010)。

2.3 ARIMAX 模型建立 余杭区 2014 年第 1 周—2017 年第 52 周 ILI% 时间序列为平稳序列 (ADF=

-4.404, P=0.010), 结合 ACF 和 PACF 图 (见图 3、图 4) 建立以 12 为周期的 ARIMA 季节乘积模型。其中 ARIMA (1, 0, 0) (1, 0, 0)₁₂ BIC 值最小, 为 90.00, 残差符合白噪声序列 ($\chi^2=0.767, P=0.381$), MAPE 值为 16.14%, 该模型为最优模型。

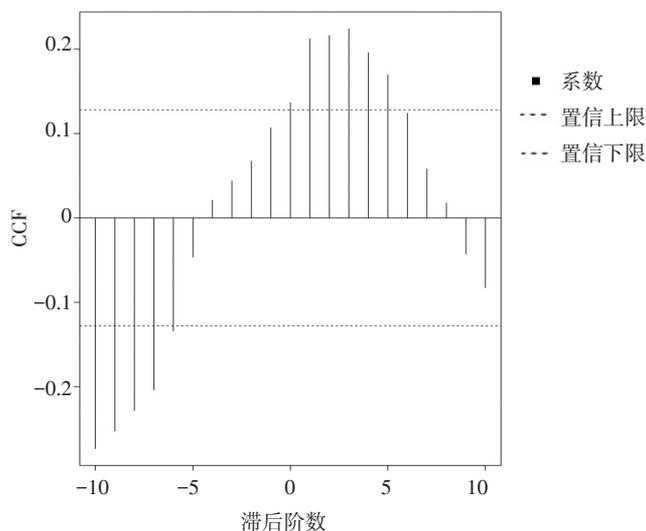


图 1 周平均绝对湿度与 ILI% 的 CCF 图

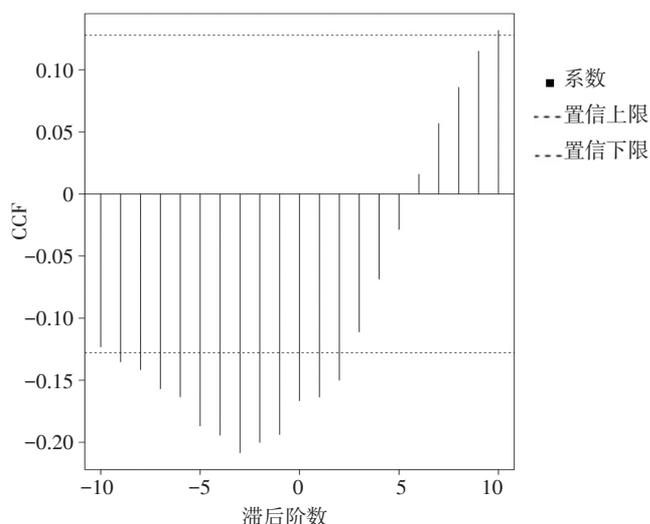


图 2 周平均气温与 ILI% 的 CCF 图

以周平均绝对湿度、1 阶拆分后的周平均气温分别作为输入序列, 以不同滞后阶数进行组合, 建立 ARIMAX 模型。其中, 周平均气温 (lag0) + 周平均绝对湿度 (lag0) 的 ARIMAX (1, 0, 0) (1, 0,

0) 模型 BIC 值最小, 为 81.30, 残差符合白噪声序列 ($\chi^2=0.631, P=0.416$), MAPE 值为 15.77%, 该模型为最优模型。

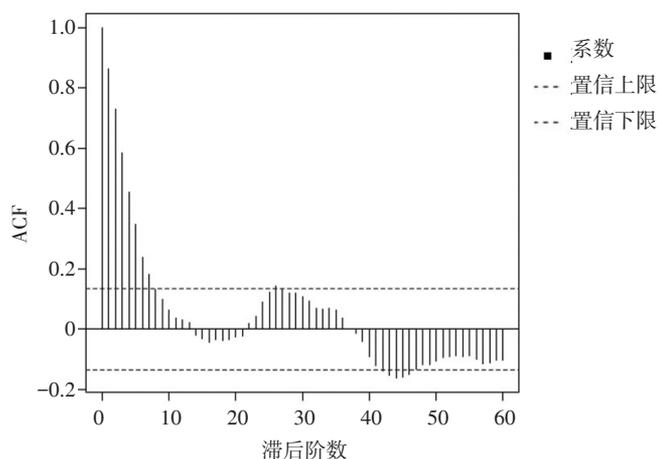


图 3 ILI% 原始时序 ACF 图

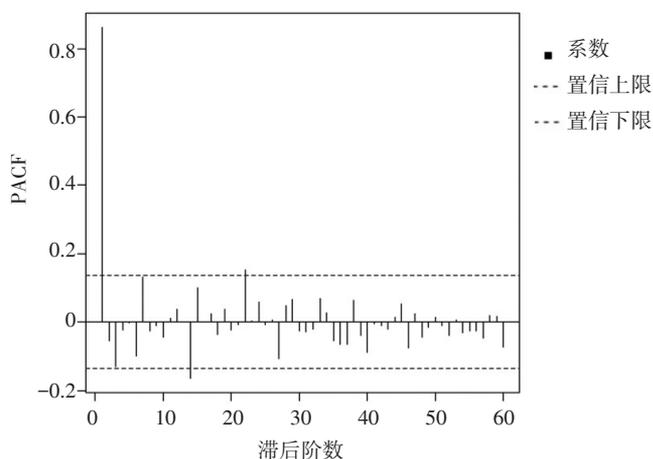


图 4 ILI% 原始时序 PACF 图

2.4 ARIMAX 模型预测评价 用 ARIMAX (1, 0, 0) (1, 0, 0)₁₂ 预测余杭区 2018 年第 1—26 周 ILI%, 预测 MAPE 值为 43.75%。第 1—6 周的实际值超出预测值 95%CI 的上限, 第 7—26 周实际值均落在预测值的 95%CI 内。见表 1。

3 讨论

本研究利用余杭区 2014 年第 1 周—2017 年第 52 周 ILI% 资料建立 ARIMA 乘积季节模型, 纳入周平均绝对湿度和周平均气温建立 ARIMAX 模型, 回代拟合 MAPE 值为 15.77%, 高于傅伟杰等 [14] 建立

表 1 2018 年第 1—26 周余杭区 ILI%预测结果分析 (%)

时间	实际值	预测值	95%CI	绝对误差	MAPE
第1周	2.62	1.74	1.20~2.28	0.88	33.63
第2周	2.70	1.58	0.86~2.29	1.12	41.62
第3周	3.93	1.67	0.84~2.50	2.26	57.53
第4周	3.70	1.40	0.50~2.30	2.30	62.22
第5周	2.91	1.52	0.57~2.48	1.39	47.63
第6周	2.65	1.55	0.55~2.54	1.10	41.64
第7周	2.25	1.50	0.47~2.52	0.75	33.48
第8周	1.17	1.44	0.40~2.49	0.27	23.37
第9周	0.97	1.46	0.40~2.52	0.49	50.89
第10周	0.81	1.37	0.30~2.44	0.56	69.28
第11周	0.83	1.42	0.34~2.50	0.59	70.26
第12周	0.92	1.34	0.25~2.43	0.41	44.84
第13周	1.05	1.39	0.30~2.49	0.34	32.44
第14周	1.14	1.31	0.21~2.40	0.17	14.77
第15周	0.76	1.34	0.24~2.44	0.58	75.85
第16周	0.92	1.31	0.21~2.42	0.40	43.31
第17周	0.53	1.31	0.20~2.41	0.78	146.42
第18周	0.77	1.31	0.21~2.42	0.54	70.40
第19周	0.57	1.28	0.17~2.39	0.71	124.61
第20周	1.28	1.33	0.22~2.44	0.05	3.95
第21周	1.19	1.25	0.14~2.35	0.05	4.26
第22周	1.46	1.28	0.17~2.39	0.18	12.11
第23周	1.28	1.28	0.17~2.39	0	0.29
第24周	1.57	1.29	0.18~2.39	0.28	18.03
第25周	1.30	1.25	0.15~2.36	0.05	3.57
第26周	1.19	1.29	0.18~2.40	0.10	8.49

ARIMA 模型的 8.124%，与赵棋锋等^[15]建立 ARIMA 季节乘积模型的 14.04% 接近。预测余杭区 2018 年第 1—26 周 ILI%，结果显示预测 MAPE 值为 43.75%。提示本研究建立的 ARIMAX 模型预测精度仍需提高，可能与余杭区 2018 年第 1—12 周 ILI% 波动较大有关。预测 MAPE 值 > 60% 的周次中 ILI% 均 < 0.9%，考虑实际值过小也会影响模型的预测精度^[16]。有研究发现日照时数是影响 ILI 的主要气象因素^[9]，本研究未纳入该气象因素，可能在一定程度上影响预测精度。

刘欣等^[8]的研究显示绝对湿度和气温与 ILI% 呈非线性关系，且流感风险在低气温和低湿度的环境下较高，与本研究中 ILI% 与平均绝对湿度呈正相关，与平均气温呈负相关的结论存在差异，可能与 Lasso 回归模型仅对气象因素进行筛选，不能对气象因素与 ILI% 的关系进行更深入的分析有关。

流感病毒的流行具有周期性特征。在流感病毒流

行期，气象因素可能会削弱 ARIMAX 模型的准确性，提示在病例阳性率较高、ILI% 实际值大于预测值时，可能存在流感高发。将预测值 95%CI 作为预警线，结合病原学监测结果，可提前预警流感高峰。

综上所述，本研究建立的 ARIMAX 模型的流感预测效果有待提高，可在今后的研究中纳入 ILI 检测标本阳性率和日照时数等变量，重新拟合模型，提高模型预测精度。

参考文献

- [1] 高福. 流感百年: 推进流感防控和研究全球合作, 中国在行动 [J]. 中华实用儿科临床杂志, 2019, 34 (2): 81-82.
- [2] 中华人民共和国国家免疫规划技术工作组流感疫苗工作组. 中国流感疫苗预防接种技术指南 (2020—2021) [J]. 中华流行病学杂志, 2020, 41 (10): 1555-1576.
- [3] 余昭, 孙琬琬, 刘社兰, 等. 呼吸道传染病监测体系运行状况分析 [J]. 预防医学, 2021, 33 (1): 101-103.
- [4] 沈冰, 沈磊, 倪晓芬, 等. 上海市原静安区成人流感样病例就诊百分比预测的自回归求和滑动平均模型构建与应用 [J]. 上海预防医学, 2017, 29 (5): 346-350.
- [5] 黄智峰, 刘晓剑, 杨连朋, 等. 流行性感胃预警方法及其应用 [J]. 疾病监测, 2016, 31 (12): 989-994.
- [6] 陈会杰, 陈叶. 2015—2017 沈阳市流感样病例与气象因素相关性分析 [J]. 预防医学论坛, 2019, 25 (9): 676-677, 682.
- [7] 丁彦红, 范俊杰, 聂清, 等. 2017—2018 年潍坊市流感样病例与气象因素相关性分析 [J]. 实用预防医学, 2020, 27 (5): 554-557.
- [8] 刘欣, 康敏, 马文军, 等. 广州市气象因素与流感样病例关系的时间序列研究 [J]. 环境卫生学杂志, 2018, 8 (5): 374-380.
- [9] 龚风云, 王凯, 樊旭成, 等. 乌鲁木齐市流感样病例与气象因素的 ARIMAX 模型预测分析 [J]. 公共卫生与预防医学, 2020, 31 (2): 4-8.
- [10] 中华人民共和国国家卫生健康委员会. 全国流感监测方案 (2017 年版) [EB/OL]. [2021-06-08]. <http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=ed1498d9e64144738cc7f8db61a39506>.
- [11] 翟红楠. 深圳市流感与大气环境的关系研究及其预测模型的建立 [D]. 武汉: 中国地质大学, 2009.
- [12] 王燕. 时间序列分析: 基于 R [M]. 北京: 中国人民大学出版社, 2015.
- [13] 王晨, 郭倩, 周罗晶. 基于 R 语言的 ARIMA 模型对流感样病例发病趋势的预测 [J]. 中华疾病控制杂志, 2018, 22 (9): 957-960.
- [14] 傅伟杰, 谢响, 曾志笠, 等. 三种模型在江西省流感样病例预测中的应用与比较 [J]. 中华疾病控制杂志, 2019, 23 (1): 101-105.
- [15] 赵棋锋, 马珊珊, 王吉玲, 等. 指数平滑法与 ARIMA 模型对流感样病例流行趋势的预测效果比较 [J]. 预防医学, 2020, 32 (4): 381-383, 387.
- [16] 沈钰钢. 嵊州市流感样病例监测结果及 ARIMA 模型预测 [D]. 杭州: 浙江大学, 2017.

收稿日期: 2021-02-04 修回日期: 2021-06-08 本文编辑: 吉兆洋