

# 2005—2023 年南昌市丙型肝炎流行特征及基于贝叶斯结构时间序列模型的发病趋势预测

莫佳丽<sup>1,2</sup>, 涂志斌<sup>3</sup>, 肖勇<sup>4</sup>, 路亮<sup>3</sup>, 刘星雨<sup>1,2</sup>, 况杰<sup>1,2</sup>

1. 南昌大学公共卫生学院流行病学教研室, 江西 南昌 330006; 2. 疾病预防与公共卫生江西省重点实验室;  
3. 南昌市疾病预防控制中心; 4. 南昌市青云谱区疾病预防控制中心

**摘要:**目的 分析 2005—2023 年南昌市丙型肝炎流行特征, 并探讨贝叶斯结构时间序列 (BSTS) 模型预测丙肝发病趋势的应用价值, 为该市丙肝防控提供科学依据。方法 收集 2005 年 1 月至 2023 年 12 月南昌市丙肝发病数据, 采用时间序列分解法解析月发病数据的趋势和季节组分。运用 R 软件构建 BSTS 模型, 其中 2005 年 1 月至 2022 年 12 月的数据作为训练集拟合 BSTS 模型, 2023 年 1—12 月数据作为测试集评估模型的预测效果, 并将其预测准确性与自回归整合移动平均 (ARIMA) 模型进行比较, 采用平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE)、均方根误差 (RMSE) 和均方根百分比误差 (RMSPE) 评估预测准确性。结果 南昌市丙肝发病总体呈上升趋势, 2017 年后发病数的增加有所减缓。发病数显示出明显的周期性和季节性变化, 发病高发于春季 3—5 月, 1—2 月为低谷。BSTS 模型预测性能指标 MAE = 9.67、MAPE = 17.61%、RMSE = 11.99 和 RMSPE = 17.84 均小于 ARIMA 模型预测性能指标 MAE = 12.12、MAPE = 29.03%、RMSE = 15.30、RMSPE = 33.62。基于 BSTS 模型预测的 2024 年 1 月至 2024 年 12 月南昌市丙肝发病总数为 308 (95% CI: 68 ~ 583) 例, 月均发病数为 26 (95% CI: 6 ~ 48) 例。结论 南昌市丙肝发病存在周期性和季节性波动; BSTS 模型预测性能更优, BSTS 模型可为丙肝精准防控提供技术支撑。

**关键词:** 丙型肝炎; 发病; 预测; 贝叶斯结构时间序列模型; 自回归整合移动平均模型

中图分类号: R183.9 文献标志码: A 文章编号: 1003-8507(2025)04-583-07

DOI: 10.20043/j.cnki.MPM.202410209

## Epidemiological characteristics of hepatitis C and incidence trend prediction based on the BSTS model, Nanchang, 2005 – 2023

MO Jia-li\*, TU Zhi-bin, XIAO Yong, LU Liang, LIU Xing-yu, KUANG Jie

\* Department of Epidemiology, School of Public Health, Nanchang University, Nanchang, Jiangxi 330006, China

**Abstract: Objective** To analyze the epidemiological characteristics of hepatitis C in Nanchang from 2005 to 2023 and evaluate the application value of the Bayesian structural time series (BSTS) model in predicting hepatitis C incidence trends, providing a scientific basis for the prevention and control of hepatitis C in the city. **Methods** Data on hepatitis C incidence in Nanchang from January 2005 to December 2023 were collected. The trend and seasonal components of the monthly incidence data were analyzed using time series decomposition. The BSTS model was constructed using R software, with data from January 2005 to December 2022 used as the training set to fit the model, and data from January to December 2023 used as the test set to evaluate the model's predictive performance. The prediction accuracy of the BSTS model was compared with that of the autoregressive integrated moving average (ARIMA) model using mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and root mean square percentage error (RMSPE) as evaluation metrics. **Results** The overall incidence of hepatitis C in Nanchang has been on an upward trend, though the rate of increase has slowed since 2017. The data reveals significant cyclical and seasonal fluctuations, with a peak incidence in the spring months from March to May and a trough from January to February. The prediction performance indicators of the BSTS model (MAE = 9.67, MAPE = 17.61%, RMSE = 11.99 and RMSPE = 17.84) were all lower than those of the ARIMA model (MAE = 12.12, MAPE = 29.03%, RMSE = 15.30, RMSPE = 33.62). Based on the BSTS model, the total predicted number of hepatitis C cases in Nanchang from January 2024 to December 2024 is 308 (95% CI: 68 – 583), with an average monthly incidence of 26 cases

基金项目: 国家自然科学基金(82160645, 82360667); 江西省自然科学基金(20212BAB206091); 江西省卫生健康委科技计划(202410613)

作者简介: 莫佳丽(2001—), 女, 硕士在读, 研究方向: 流行病与卫生统计学

通信作者: 况杰, E-mail: kuangjie@ncu.edu.cn

(95% CI: 6 - 48). **Conclusion** Hepatitis C incidence in Nanchang exhibits periodic and seasonal fluctuations. The BSTS model outperforms the ARIMA model in prediction performance and can provide technical support for the precise prevention and control of hepatitis C.

**Keywords:** Hepatitis C; Incidence; Prediction; Bayesian structural time series model; Autoregressive integrated moving average model

丙型肝炎病毒(Hepatitis C Virus, HCV)可引起肝失代偿、肝硬化、肝癌等严重肝脏相关疾病<sup>[1]</sup>,是全球人口死亡的重要原因<sup>[2]</sup>。2019 年全球 HCV 感染人数达 5 800 万,其中 54 万人死亡,占全球死亡总数的 0.96%<sup>[3]</sup>。当前我国 HCV 及 HCV 感染相关肝脏疾病的发病数呈上升趋势,疾病防控形势仍旧严峻<sup>[4]</sup>。2016 年,WHO 提出到 2030 年消除病毒性肝炎作为重大公共卫生威胁的宏伟目标,强调新发感染下降 90%<sup>[5]</sup>。在此背景下,加强 HCV 早期防控,构建精细化模型及时预测疾病流行趋势,预警疾病暴发隐患并减少新发病例显得尤为重要。自回归滑动平均混合 (autoregressive integrated moving average, ARIMA) 模型因其简单性、快速适用性被广泛应用于时间序列预测,是评估传染病流行趋势最常用的模型<sup>[6-7]</sup>。但 ARIMA 模型适用平稳数据,在建模过程中需假定趋势、季节性、随机波动等三个时间序列主要成分在观察时间内保持恒定,这导致模型未能准确反映疾病传播的实际动态<sup>[8]</sup>。相比之下,贝叶斯结构时间序列 (Bayesian structural time series, BSTS) 模型允许模型参数随时间波动来正确反映随机行为,以弥补传统模型的不足<sup>[9]</sup>。其能够整合先验知识、处理复杂数据结构以及提供全面的不确定性评估,具有更高的灵活性<sup>[8]</sup>。目前基于贝叶斯结构时间序列模型预测丙肝流行趋势的研究相对较少,进一步探讨有助于构建精细化模型,优化疾病流行趋势预测效能,促进丙肝防控策略的精确制定。因此,本研究基于南昌市丙型肝炎发病数据,分析丙型肝炎流行特征,利用 BSTS 探索预测南昌市丙型肝炎流行趋势的优化模型,为该市疾病监测预警与防控提供技术支撑。

## 1 研究资料和方法

**1.1 研究资料** 依据《丙型肝炎诊断:(WS213 - 2018)》,从南昌市传染病疫情监测系统获取 2005 年 1 月至 2023 年 12 月南昌市丙肝发病数据,共计 228 个数据点。训练集纳入 216 个数据点 (2005.1—2022.12),用于构建 ARIMA 和 BSTS 模型并进行拟合。为验证模型预测的准确性,将 12 个数据点 (2023.1—2023.12) 纳入测试集。采用普雷斯科特过滤器<sup>[10]</sup> (Hodrick - Prescott, HP) 从原始时间序列数据中分解出趋势和周期成分,揭示数据的季节性波动。采用时间序列分解法<sup>[11]</sup> 获得季节指数 (Seasonal

Index, SI), 季节指数表示特定时期内,发病率偏离平均水平的程度。SI > 1 表示高风险季节, < 1 为低风险季节。

**1.2 ARIMA 模型** ARIMA 模型的标准形式为 ARIMA(p, d, q)(P, D, Q)<sub>s</sub>, 其主要用于分析具有周期性和季节性特征的时间序列数据。模型中“p”代表非季节性部分的自回归项数,“d”是非季节性差分的阶数,“q”表示非季节性移动平均的项数。对应地,“P”代表季节性部分的自回归项数,“D”是季节性差分的阶数,“Q”表示季节性移动平均的项数,而“S”则指定模型的季节周期长度<sup>[12]</sup>。建模<sup>[13]</sup> 包括以下步骤:首先检验时间序列数据的平稳性,这里运用单位根检验 (ADF 检验);其次,数据平稳后运用自相关函数 (ACF) 和偏自相关函数 (PACF) 识别模型阶数,参数的选定在 R 软件中可通过贝叶斯信息准则 (BIC) 标准自动完成,选取 BIC 值最小的模型以确保纳入最优参数;而后应用白噪声检验 (Ljung - Box Q), 判断残差是否为随机序列;拟合质量合格后则可进行模型预测。构建 ARIMA 模型时应谨慎处理原始数据的平稳化,合理选择模型参数,以提高预测的精确性和可靠性。

**1.3 BSTS 模型** BSTS 模型整合了卡尔曼滤波器、Spike 和 slab 回归,以及贝叶斯模型平均这三个关键组件<sup>[14]</sup>。卡尔曼滤波器用于加强模型中的局部回归分析、线性趋势估计以及季节性因素的整合;Spike 和 slab 回归允许加入新的回归变量以考虑其他潜在的影响因素;模型的最终预测结果则通过贝叶斯模型平均获得。构建 BSTS 模型时须配置适宜的趋势组件和季节组件,趋势组件包括局部水平、局部线性趋势、半局部线性趋势和共享局部水平等趋势等分量<sup>[8,14]</sup>。本研究中采用局部线性趋势,因为该分量基于趋势均值和斜率随机游走假设,适用短期预测且可降低误差<sup>[15]</sup>。研究以丙肝的月发病数为例,季节参数固定为 12,以模拟月度变化。BSTS 模型在分析中有效地利用了先验知识,并允许模型参数适应时间的变化,从而能够精确处理包含众多协变量的复杂数据集,防止发生过拟合。此外,该模型采取基于数据的分析策略,能够应对不同规模的时间序列数据,不受限于预设的假设条件 (例如,构建 ARIMA 模型,数据须满足平稳性、线性等条件)。最终,通过优化后验分布的最大化, BSTS 模型增强对数据中不确定性的管理,提升

预测结果的精确度和可信度。这些特性使得 BSTS 模型在时间序列分析领域表现出较高的适应性和精确性<sup>[8-9]</sup>。

**1.4 统计学分析** 从《江西统计年鉴》获得南昌市人口数据,南昌市丙型肝炎发病率由月度报告的发病数除以平均年中人口所得。发病数据的趋势分量和季节组分量由 R 语言过滤器和时间序列分解法进行分析。ARIMA 模型的构建运用 R 软件的“tseries”和“forecast”包,而 BSTS 模型则通过“bsts”包实现。在本研究中,为比较两种模型的预测效果,我们计算了以下指标:平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方根误差(RMSE)和均方根百分比误差(RMSPE)。检验水准均为 0.05。BSTS 模型和 ARIMA 模型的预测性能差异通过校正的迪博德-马里亚诺检验<sup>[16]</sup>(Modified Diebold - Mariano, MDM)来验证。

**2 结果**

**2.1 南昌市丙肝流行特征** 南昌市 2005 年 1 月到 2023 年 12 月共报告丙型肝炎病例 8 813 例,2019 年报告的病例数最多(594 例,10.61/10 万),2005 年报告的病例数最少(266 例,5.89/10 万)。总体年均发病率为 8.69/10 万,月均发病率为 0.72/10 万,具体见图 1。丙肝发病数据的趋势和周期序列由 HP 过滤器绘制,结果显示南昌市丙肝流行总体呈上升趋势(图 2);1 至 12 月发病数据的季节指数由时间序列分解法获得,分别为 0.70、0.84、1.16、1.14、1.10、1.03、1.01、1.08、1.03、1.00、0.95、0.96;以上表明南昌市丙肝发病数显示出明显的周期性和季节性变化,每年初(1 月至 2 月)报告病例数较低,而春季(3 月至 5 月)病例数量明显增多,其余月份保持相对稳定。

**2.2 基于 ARIMA 模型的丙肝发病预测** 丙肝发病数据序列呈非平稳状态,显示季节性波动。为消除季

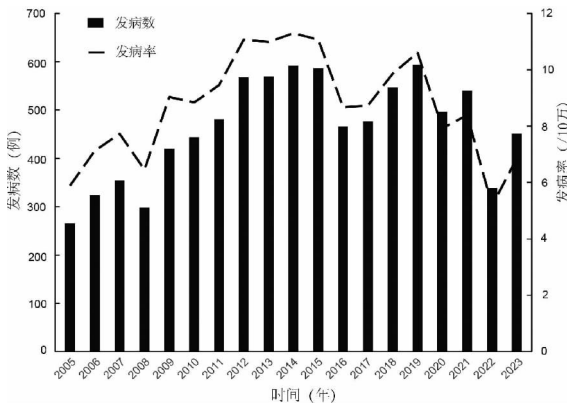


图 1 南昌市 2005—2023 年丙肝年发病数和发病率

Fig. 1 Annual incidence and morbidity of hepatitis C in Nanchang from 2005 to 2023

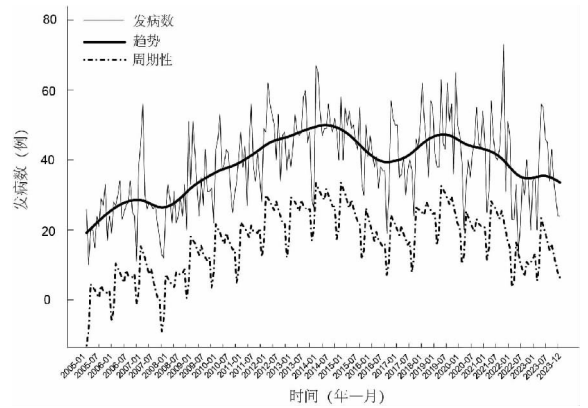


图 2 南昌市 2005 年 1 月至 2023 年 12 月丙肝逐月发病数时序图

Fig. 2 Time series of monthly incidence of hepatitis C in Nanchang from January 2005 to December 2023

节效应,将数据进行季节性差分,差分后的 ACF 和 PACF 图(图 3)提示序列仍存在结构依赖,如趋势性或长期相关性,序列还需进行非季节性差分处理。R 中使用 ndiffs() 函数,得出一阶拆分后变为平稳序列(ADF = -9.036, P = 0.01),据此 d、D 均为 1。且 ACF 和 PACF 图(图 3)在 1 月和 12 月的滞后点显示出局部峰值,据此判断 p、q、P 和 Q 的阶数不超过 1。结合以上条件,根据贝叶斯信息准则(BIC)获得最佳模型 ARIMA(1,1,1)(0,1,1)<sub>12</sub>。对模型系数进行进一步检验,AR1 = -0.52 (t = -6.88, P = 0.04), MA1 = -0.70 (t = -14.76, P < 0.01), SMA1 = -0.35 (t = -13.56, P < 0.01)。残差的 ACF 和 PACF 图示结果表明所有相关系数均在置信区间内(图 4A)。Ljung - Box Q 检验结果表明残差为白噪声(表 1)。以上分析表明,以南昌市丙肝发病数据建立的最佳 ARIMA 模型为 ARIMA(1,1,1)(0,1,1)<sub>12</sub>,经诊断该模型可用于适用于对测试集数据进行预测,结果见表 2。

表 1 ARIMA 和 BSTS 模型残差的 Ljung - Box Q 检验

Table 1 Ljung - Box Q test for the residuals from ARIMA and BSTS models

滞后期 (月)	ARIMA 模型		BSTS 模型	
	$\chi^2$ 值	P 值	$\chi^2$ 值	P 值
1	0.017	0.895	0.277	0.599
2	0.375	0.829	0.809	0.667
3	0.421	0.936	4.283	0.232
4	0.636	0.959	7.284	0.122
5	0.775	0.979	7.887	0.163
6	0.776	0.993	8.585	0.198
7	0.966	0.995	10.351	0.170
8	4.020	0.855	14.699	0.065
9	4.521	0.874	14.939	0.093
10	4.603	0.917	15.110	0.128

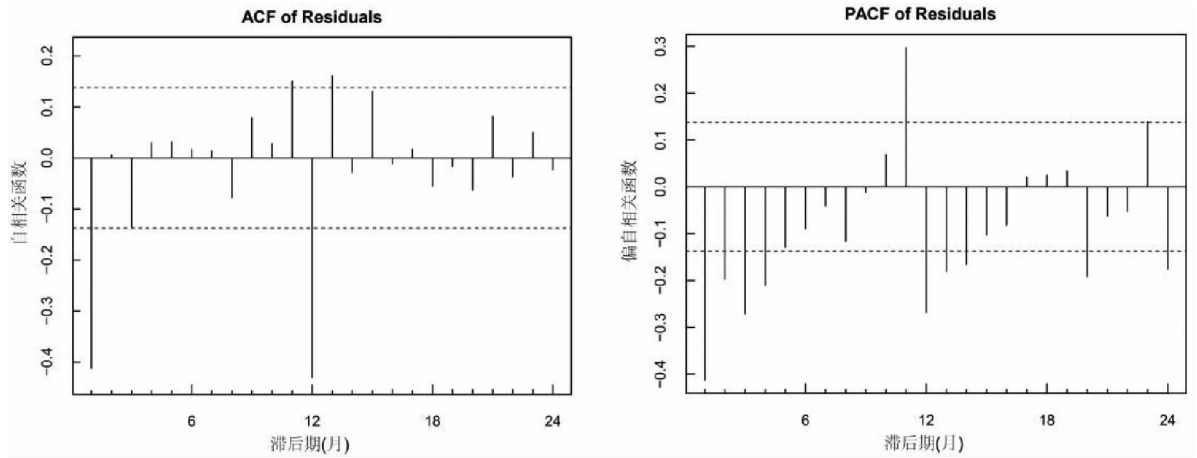


图3 丙肝发病序列季节差分的自相关和偏自相关函数

Fig.3 ACF and PACF plots of seasonally differenced hepatitis C incidence series

**2.3 基于 BSTS 模型的丙肝发病预测** BSTS 模型的模拟过程中,在 BSTS 模型的状态结构中纳入局部水平模型,500 次 MCMC 模拟后抽样处于平稳状态。所得残差标准差参数的后验均值为 7.51,训练数据集向前预测模拟的残差标准差为 9.55,残差方差较原始序列方差增加了 63.45%,拟合优度统计量为 0.32。进

一步对模型进行检验,残差的 ACF 和 PACF 图中大部分相关系数位于可信区间内(图 4B),Ljung - Box Q 检验结果表明 BSTS 模型的残差为白噪声(表 1)。因此,可用基于局部水平的 BSTS 模型对测试集进行预测,具体结果见表 2。

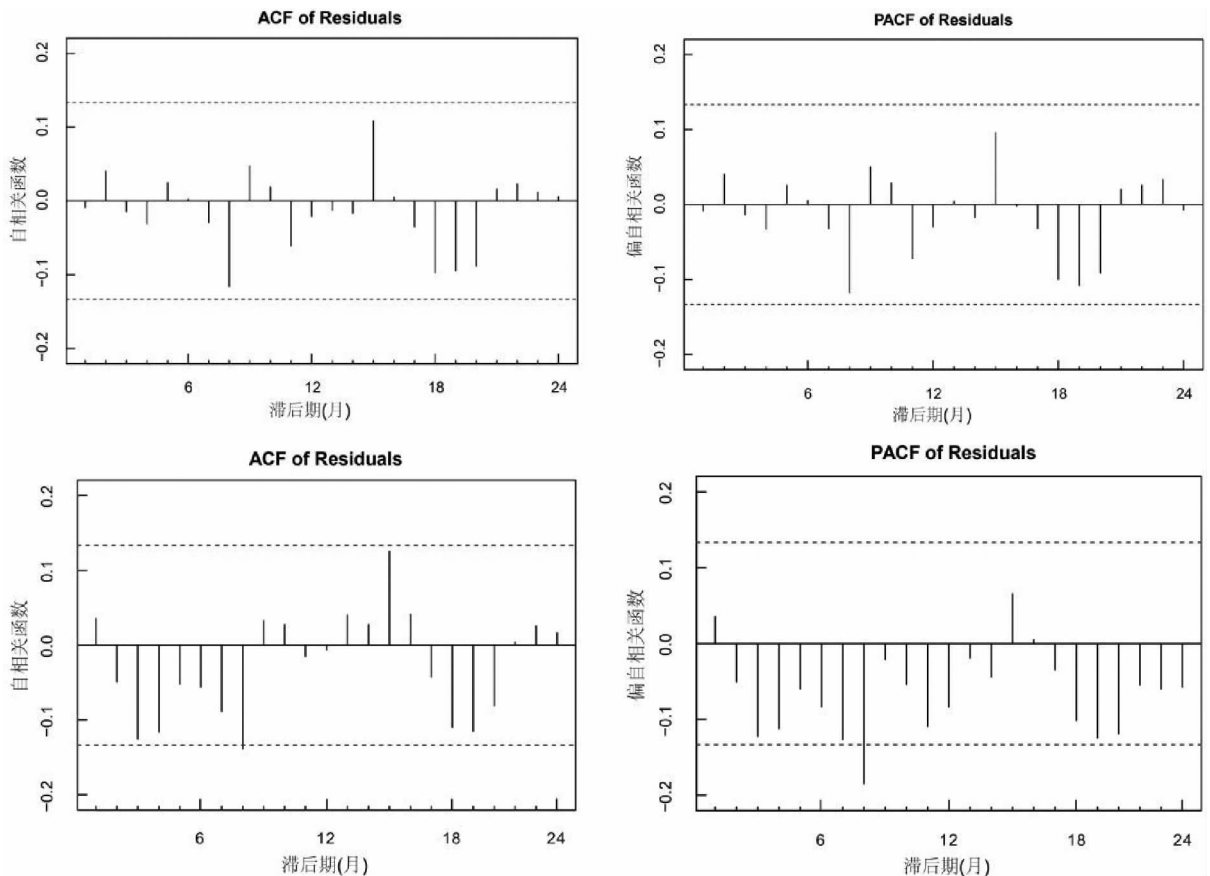


图4 ARIMA (A)和 BSTS(B)模型残差序列的 ACF 和 PACF 图

Fig.4 Autocorrelation and Partial Autocorrelation Plots of Residual Series for ARIMA (A) and BSTS (B) Models

**表 2** ARIMA(1,1,1)(0,1,1)<sub>12</sub>和 BSTS 模型预测的 2023 年 1 月至 12 月南昌市丙肝发数**Table 2** Predicted hepatitis C cases in Nanchang from January to December 2023 using ARIMA (1, 1, 1) (0, 1, 1)<sub>12</sub> and BSTS models

月份	实际发病数(例)	ARIMA 模型预测发病数(95% CI)(例)	BSTST 模型预测发病数(95% CI)(例)
1	20	18(0~36)	18(0~35)
2	41	16(3~35)	19(0~37)
3	56	34(14~53)	35(14~56)
4	55	5(13~45)	29(10~49)
5	46	26(6~46)	28(9~51)
6	44	32(11~52)	29(8~51)
7	34	29(8~49)	26(5~50)
8	43	34(14~55)	28(2~53)
9	34	32(11~53)	26(7~51)
10	29	25(4~47)	22(4~47)
11	24	32(10~53)	25(3~53)
12	24	30(8~51)	23(6~50)

**2.4 BSTS 模型和 ARIMA 模型性能比较** 两种模型预测准确性的比较结果见表 3。结果显示 ARIMA 模型预测性能指标(MAE、MAPE、RMSE 和 RMSPE)均大于 BSTS 模型,且 MDM 检验结果显示  $P$  值均  $< 0.05$ ,表明 BSTS 模型预测性能优于 ARIMA 模型。此外,基于 BSTS 模型重新模拟 2005 年 1 月至 2023 年 12 月数据,并对南昌市 2024 年 1 月至 2024 年 12 月丙肝流行趋势进行预测,预测表明南昌市 2024 年 1 月至 2024 年 12 月丙肝发病总数为 308(95% CI:68~583)例,月均发病数为 26(95% CI:6~48)例。具体结果见表 4。

**表 3** ARIMA(1,1,1)(0,1,1)<sub>12</sub>和 BSTS 模型的预测性能比较  
**Table 3** Comparison of Forecasting Performance between ARIMA (1,1,1)(0,1,1)<sub>12</sub> and BSTS Models

指标	ARIMA 模型	BSTST 模型	MDM	$P$ 值
MAE	12.12	9.67	2.164	$< 0.001$
MAPE%	29.03	17.61	3.572	$< 0.001$
RMSE	15.30	11.99	3.187	$< 0.001$
RMSPE	33.62	17.84	3.915	$< 0.001$

### 3 讨论

丙肝的疾病负担沉重,是危害我国人民健康的重大公共卫生问题<sup>[4]</sup>。增强 HCV 早期监测预警能力,及时预测疾病流行趋势具有重要意义。BSTS 时间序列预测模型是近年来传染病发病预测的新型工具,本研究将其预测性能同 ARIMA 模型进行比较。研究发现,BSTST 模型预测性能优于 ARIMA 模型,其模型性能指标(MAE、MAPE、RMSE 和 RMSPE)均小于 ARIMA 模型。

**表 4** 基于 BSTS 模型预测的南昌市 2024 年 1 月至 2024 年 12 月丙肝发病数**Table 4** Predicted hepatitis C cases in Nanchang from January 2023 to December 2024 based on BSTS model

时间(年.月)	预测发病数(95% CI)(例)
2024.01	19(1~38)
2024.02	23(3~45)
2024.03	39(19~59)
2024.04	32(10~54)
2024.05	31(9~53)
2024.06	33(9~54)
2024.07	28(4~52)
2024.08	32(5~58)
2024.09	29(3~55)
2024.10	25(4~51)
2024.11	28(0~55)
2024.12	25(5~55)

本次研究发现,南昌市丙肝发病总体呈上升趋势,与全国丙型肝炎发病情况一致<sup>[7]</sup>。根据发病趋势图可知,2017 年后南昌市丙肝发病数的增加有所减缓,这可能是由于丙肝直接抗病毒药物(direct-acting antiviral agents, DAAs)开始纳入使用<sup>[17]</sup>。DAAs 疗程短,治愈率高达 90% 以上,高治愈率促使患病群体缩小,帮助减少疾病传播。同时,为响应到 2030 年消除丙肝病毒为重大威胁的目标,中国采取了一系列积极措施来促进丙肝病毒的预防和治疗<sup>[18]</sup>,包括通过了《病毒性肝炎预防和控制计划(2017-2020 年)》和《消除丙型肝炎作为公共卫生威胁的行动工作计划(2021-2030 年)》,以上均促进了丙肝疫情的积极防控。时间序列结果显示,南昌市丙肝发病具有周期性和季节性,每年 3-5 月为高峰,1-2 月为低谷。这一结果与相关报道相似<sup>[7,19]</sup>。1-2 月为低谷可能与“春节效应”有关<sup>[15]</sup>,节假日相关部门值勤率低容易造成传染病漏报情况。而每年 3-5 月发病数增加可能与温度有关。先前研究指出,HCV 病毒稳定性与温度相关,在室温 25℃ 存活最宜<sup>[20]</sup>。3-5 月南昌市气温回升,增强 HCV 病毒活性,与寒冷的 1-2 月相比更有利于传播感染。

MAPE 是衡量模型准确性最常用的指标之一,其值越小代表预测模型准确性越好<sup>[16]</sup>。Pao 等学者<sup>[21]</sup>提出的模型评价标准认为 MAPE 值小于 20% 代表模型准确性良好;介于 20%~50% 的代表模型是合理的;而 MAPE 值大于 50% 则代表预测模型不准确。本研究中 BSTS 时间序列模型预测性能的 MAPE 小于 20%,这说明 BSTS 模型具有较高的预测性能,能够揭示南昌市丙肝发病数据的潜在趋势,实现丙肝流行趋势动态的精准评估。这一结果与河南省丙肝发病趋势模型性能研究相符,该研究分别比较 SARIMA 模型

与 BSTS 模型向前 15 步、向前 12 步预测准确性,分析发现在预测丙肝发病趋势动态时,BSTS 模型表现更佳<sup>[19]</sup>。BSTS 模型在其他传染病的流行趋势预测中也表现出极高的预测性能,Xu 等人<sup>[1]</sup>利用 BSTS 模型预测中国急性出血性结膜炎(AHC)的流行趋势,结果表明与 ARIMA 模型相比,该模型在预测 AHC 的流行趋势方面具有更高的应用价值。除传染病发病趋势预测外,BSTS 模型在疾病负担预测也表现出色<sup>[23]</sup>,以上表明 BSTS 时间序列模型可成为疾病监测和预防策略的宝贵工具。

本次研究拟合的最佳 ARIMA 模型为 ARIMA(1, 1, 1)(0, 1, 1)<sub>12</sub>,与河南省丙肝发病趋势模型 SARIMA(0, 1, 1)(2, 1, 0)<sub>12</sub><sup>[19]</sup>、中国丙型肝炎流行预测研究的 SARIMA(3, 0, 0)(0, 1, 2)<sub>12</sub>模型均不同<sup>[7]</sup>,这种差异可能源于 ARIMA 模型建立在线性假设的基础上,无法识别数据中的所有信息。因此改变分析的地区或时段时,需要重新构建预测模型<sup>[15]</sup>。相比之下,BSTS 模型作为一种随机状态空间模型,能够独立地分析趋势、季节性和回归成分。它采取基于数据的分析方法,对特定理论假设的依赖性较小,通过整合先验知识以处理复杂数据结构,从而使得模型在估计中更具灵活性,提供全面的不确定性评估。这些特性使得 BSTS 模型能够有效应对不同规模的时间序列数据<sup>[8-9, 24]</sup>。本研究中 ARIMA 模型预测性能的 MAPE 小于 50%,这说明 ARIMA 模型在丙肝疾病流行趋势预测研究仍具有重要价值。疾病监测工作者应选择多种模型拟合及预测各类传染病,结合 ARIMA 传统模型与 BSTS 时间序列模型,综合评估疾病流行趋势,促进防控策略的精确制定。此外,在应用 BSTS 时需要注意以下几点:(1)确保足够的时间序列数据集以便模型拟合,通常需 8 个以上的季节周期时间序列<sup>[15]</sup>。(2)及时更新数据以便修正参数,避免由于预测步数的增加导致模型性能下降。(3)BSTS 模型允许包括趋势、季节性和回归组件等多种形式的时间序列动态;根据具体的应用场景选择合适的模型组件是关键。(4)部分传染病如丙肝具有隐匿性强的特点,难以完全避免漏报发生;疾病监测以被动监测为主,使得模型预测与真实情况存在偏倚。

综上,本研究采用 BSTS 模型与 ARIMA 模型对南昌市丙肝发病趋势进行预测,结果显示 BSTS 模型在预测准确性上明显优于 ARIMA 模型。BSTS 模型凭借其在处理时间序列数据上的灵活性和对复杂数据结构的提取能力,可作为传染病流行趋势分析的有力工具。

**利益冲突声明** 本研究不存在任何利益冲突

## 参考文献

- [1] Martinello M, Solomon SS, Terrault NA, et al. Hepatitis C [J]. *Lancet*, 2023, 402(10407): 1085-1096.
- [2] Younossi ZM, Wong G, Anstee QM, et al. The global burden of liver disease [J]. *Clinical Gastroenterology and Hepatology: the Official Clinical Practice Journal of the American Gastroenterological Association*, 2023, 21(8): 1978-1991.
- [3] GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019 [J]. *Lancet*, 2020, 396(10258): 1204-1222.
- [4] Zhao ZY, Chu MJ, Guo YC, et al. Feasibility of hepatitis C elimination in China: from epidemiology, natural history, and intervention perspectives [J]. *Frontiers in Microbiology*, 2022, 13: 884598.
- [5] Cui FQ, Blach S, Manzengo mingiedi C, et al. Global reporting of progress towards elimination of hepatitis B and hepatitis C [J]. *The Lancet. Gastroenterology & Hepatology*, 2023, 8(4): 332-342.
- [6] Wang M, Jiang Z, You M, et al. An autoregressive integrated moving average model for predicting varicella outbreaks - China, 2019 [J]. *China CDC Weekly*, 2023, 5(31): 698-702.
- [7] Wang YB, Qing SY, Liang ZY, et al. Time series analysis - based seasonal autoregressive fractionally integrated moving average to estimate hepatitis B and C epidemics in China [J]. *World Journal of Gastroenterology*, 2023, 29(42): 5716-5727.
- [8] Feroze N. Forecasting the patterns of COVID - 19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models [J]. *Chaos Solitons & Fractals*, 2020, 140: 110196.
- [9] Ding WH, Li YY, Bai YC, et al. Estimating the effects of the COVID - 19 outbreak on the reductions in tuberculosis cases and the epidemiological trends in China: a causal impact analysis [J]. *Infection and Drug Resistance*, 2021, 14: 4641-4655.
- [10] Itikhar H, Daniyal M, Qureshi M, et al. A hybrid forecasting technique for infection and death from the mpox virus [J]. *Digit Health*, 2023, 9: 20552076231204748.
- [11] Hao J, Liu F. Improving long - term multivariate time series forecasting with a seasonal - trend decomposition - based 2 - dimensional temporal convolution dense network [J]. *Scientific Reports*, 2024, 14(1): 1689.
- [12] Schaffer AL, Dobbins TA, Pearson SA. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large - scale health interventions [J]. *BMC Medical Research Methodology*, 2021, 21(1): 58.
- [13] 萧楚瑶,黎婷婷,付若楠,等.基于 ARIMA 与 LSTM 模型的乌鲁木齐百日咳发病预测研究 [J]. *现代预防医学*, 2024, 51(21): 3877-3882.
- [14] Xiao CY, Li TT, Fu RN, et al. Research on pertussis incidence prediction in Urumqi based on ARIMA and LSTM models [J]. *Modern Preventive Medicine*, 2024, 51(21): 3877-3882. (In Chinese)
- [15] Katarina B, Gunardi. 2023. Optimization of bayesian structural time series (BSTS) applications in forecasting stock prices through state components selection [C] // Proceedings of the 8th International Conference on the Applications of Science and Mathematics, Singapore: Springer Nature Singapore: 229-248.
- [16] 张珍,刘星言,李言言,等.基于贝叶斯结构时间序列模型评估我国梅毒流行趋势 [J]. *疾病监测*, 2023, 38(9): 1087-1093.
- [17] Zhang Z, Liu XY, Li YY, et al. Estimated prevalence of syphilis

- in China based on Bayesian structural time series model [J]. *Disease Surveillance*, 2023, 38(9): 1087–1093. (In Chinese)
- [16] Zhou J, Li HQ, Zhong WL. A modified Diebold – Mariano test for equal forecast accuracy with clustered dependence[J]. *Economics Letters*, 2021, 207: 110029.
- [17] Du XY, Mi JR, Cheng HC, et al. Uptake of hepatitis C direct – acting antiviral treatment in China: a retrospective study from 2017 to 2021[J]. *Infectious Diseases of Poverty*, 2023, 12(1): 28.
- [18] Song YY, Li YC, Cheng HC, et al. Eliminate hepatitis C as a public health threat: a narrative review of strategies, gaps, and opportunities for China [J]. *Infectious Diseases and Therapy*, 2022, 11(4): 1427–1442.
- [19] 张珍,刘星言,邢莹莹,等. BSTS 和 SARIMA 模型在河南省丙肝发病趋势预测中的性能比较研究[J]. *江苏预防医学*, 2023, 34(6): 723–726.
- Zhang Z, Liu XY, Xing YY, et al. Comparative performance of BSTS and SARIMA models in the prediction of hepatitis C incidence trend in Henan Province [J]. *Jiangsu Journal of Preventive Medicine*, 2023, 34(6): 723–726. (In Chinese)
- [20] Song HS, Li J, Shi S, et al. Thermal stability and inactivation of hepatitis C virus grown in cell culture[J]. *Virology Journal*, 2010, 7(1): 40.
- [21] Pao HT. Forecasting energy consumption in Taiwan using hybrid nonlinear models[J]. *Energy*, 2009, 34(10): 1438–1446.
- [22] Xu G, Fan T, Zhao Y, et al. Predicting the epidemiological trend of acute hemorrhagic conjunctivitis in China using Bayesian structural time – series model [J]. *Scientific Reports*, 2024, 14(1): 17364.
- [23] Bai Z, Wang H, Shen C, et al. The global, regional, and national patterns of change in the burden of non – malignant upper gastrointestinal diseases from 1990 to 2019 and the forecast for the next decade[J]. *International Journal of Surgery*, 2024.
- [24] Navas Thorakkattle M, Farhin S, Khan AA. Forecasting the trends of covid – 19 and causal impact of vaccines using Bayesian structural time series and ARIMA [J]. *Annals of Data Science*, 2022, 9(5): 1025–1047.

收稿日期:2024-10-16

## (上接第 582 页)

- [17] Li C, Gu L, Shi FY, et al. Serum liver enzymes and risk of stroke: Systematic review with meta – analyses and Mendelian randomization studies[J]. *European Journal of Neurology*, 2024, 31(12): e16506.
- [18] Davies NM, Holmes MV, Davey smith G. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians[J]. *BMJ*, 2018, 362: k601.
- [19] Xu L, Jiang CQ, Lam TH, et al. Mendelian randomization estimates of alanine aminotransferase with cardiovascular disease: Guangzhou Biobank Cohort study[J]. *Human Molecular Genetics*, 2017, 26(2): 430–437.
- [20] Au yeung SL, Borges MC, Wong THT, et al. Evaluating the role of non – alcoholic fatty liver disease in cardiovascular diseases and type 2 diabetes: a Mendelian randomization study in Europeans and East Asians[J]. *International Journal of Epidemiology*, 2023, 52(3): 921–931.
- [21] Guo YL, Gao J, Liu Y, et al. An examination of causal associations and shared risk factors for diabetes and cardiovascular diseases in the East Asian population: A Mendelian randomization study[J]. *Frontiers in Endocrinology*, 2023, 14: 1132298.
- [22] Sakaue S, Kanai M, Tanigawa Y, et al. A cross – population Atlas of genetic associations for 220 human phenotypes [J]. *Nature Genetics*, 2021, 53(10): 1415–1424.
- [23] Kim YJ, Moon S, Hwang MY, et al. The contribution of common and rare genetic variants to variation in metabolic traits in 288,137 East Asians[J]. *Nature Communications*, 2022, 13(1): 6642.
- [24] Ishigaki K, Akiyama M, Kanai M, et al. Large – scale genome – wide association study in a Japanese population identifies novel susceptibility loci across different diseases [J]. *Nature Genetics*, 2020, 52(7): 669–679.
- [25] Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries [J]. *Nature*, 2022, 611(7934): 115–123.
- [26] Burgess S, Thompson SG. Multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects [J]. *American Journal of Epidemiology*, 2015, 181(4): 251–260.
- [27] Burgess S, Scott RA, Timpson NJ, et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors [J]. *European Journal of Epidemiology*, 2015, 30(7): 543–552.
- [28] Bowden J, Davey smith G, Haycock PC, et al. Consistent estimation in mendelian randomization with Some invalid instruments using a weighted median estimator [J]. *Genetic Epidemiology*, 2016, 40(4): 304–314.
- [29] Bowden J, Davey smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression [J]. *International Journal of Epidemiology*, 2015, 44(2): 512–525.
- [30] Verbanck M, Chen CY, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases [J]. *Nature Genetics*, 2018, 50(5): 693–698.
- [31] Pazoki R, Vujkovic M, Elliott J, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes [J]. *Nature Communications*, 2021, 12(1): 2579.
- [32] Sarin SK, Kumar M, Eslam M, et al. Liver diseases in the Asia – Pacific region: a Lancet Gastroenterology & Hepatology Commission [J]. *The Lancet. Gastroenterology & Hepatology*, 2020, 5(2): 167–228.
- [33] Lee Y, Seo JH. Potential causal association between elevated Gamma – Glutamyl transferase level and stroke: atwo – sample Mendelian randomization study [J]. *Biomolecules*, 2023, 13(11): 1592.
- [34] Franzini M, Corti A, Martinelli B, et al. Gamma – glutamyltransferase activity in human atherosclerotic plaques— biochemical similarities with the circulating enzyme [J]. *Atherosclerosis*, 2009, 202(1): 119–127.
- [35] Karmen A, Wroblewski F, Ladue JS. Transaminase activity in human blood [J]. *Journal of Clinical Investigation*, 1955, 34(1): 126–131.
- [36] Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two – sample Mendelian randomization [J]. *Genetic Epidemiology*, 2016, 40(7): 597–608.

收稿日期:2024-11-04