·论著·

SARIMA 和 SARFIMA 模型在我国结核病流行趋势预测中的性能比较研究*

河南医药大学公共卫生学院流行病与卫生统计学教研室（453000）
卿思雨　梁姿悦　席　越　张冰洁　薛晨璐　梁艺芳　王永斌△

【提　要】 目的　比较季节性自回归差分移动平均模型（seasonal autoregressive integrated moving average，SARIMA）与季节性自回归分数差分移动平均模型（seasonal autoregressive fractionally integrated moving average，SARFIMA）对我国结核病（tuberculosis，TB）流行趋势的预测性能，并基于最优模型预测 2023—2030 年 TB 发病趋势。**方法**　收集 2004 年 1 月—2023 年 10 月全国 TB 月发病数据，采用 Joinpoint 回归以平均年度百分比变化（average annual percentage change，AAPC）分析长期趋势，并结合时间序列分解法探讨季节性特征。以 2004 年 1 月—2022 年 10 月为训练集、2022 年 11 月—2023 年 10 月为测试集，基于赤池信息准则（Akaike information criterion，AIC）和贝叶斯信息准则（Bayesian information criterion，BIC）优选模型参数，通过平均绝对误差（mean absolute error，MAE）、平均绝对百分比误差（mean absolute percentage error，MAPE）、均方根误差（root mean square error，RMSE）等 5 项指标评估模型性能。**结果**　TB 发病率呈显著下降趋势（$AAPC_{2004—2023}$=−3.48，95%*CI*：−4.39~−2.77），每年 3 月达高峰（季节性指数 1.16），2 月为低谷（0.84）。SARFIMA（0，0.49，1）（2，0.44，2）$_{12}$ 的预测性能优于 SARIMA（0，1，1）（2，1，2）$_{12}$，其 MAE（0.53 vs.0.92）、MAPE（0.16 vs.0.28）、RMSE（0.90 vs.1.12）等指标显著降低，残差检验显示 SARFIMA 残差为白噪声（Ljung−Box *Q*=0.31，*P*>0.05）。预测的 2023—2030 年 TB 年均发病率为 4.48/10 万（95% *CI*：1.95~7.01）。**结论**　SARFIMA 模型通过分数差分技术有效捕捉 TB 发病率的长记忆性特征，预测精度显著优于 SARIMA，建议将其纳入 TB 监测系统以动态优化防控策略，助力实现 2030 年消除 TB 目标。

【关键词】 结核病　SARFIMA 模型　SARIMA 模型　预测　时间序列分析
【中图分类号】 R18　　**【文献标志码】** A　　**DOI** 10.11783/j.issn.1002−3674.2026.03.001

A Comparative Study on the Performance of SARIMA and SARFIMA Models in Predicting the Epidemic Trend of Tuberculosis in China

Qing Siyu, Liang Ziyue, Xi Yue, et al (*Department of Epidemiology and Health Statistics, School of Public Health, Henan Medical University, Xinxiang* 453000)

【Abstract】 Objective　To systematically compare the predictive performance of the seasonal autoregressive fractionally integrated moving average (SARFIMA) model and the seasonal autoregressive integrated moving average (SARIMA) model for tuberculosis (TB) epidemic trends in China, and to forecast TB incidence from 2023 to 2030 based on the optimal model. **Methods**　Monthly TB incidence data from January 2004 to October 2023 were collected. Joinpoint regression with average annual percentage change (AAPC) was used to analyze long-term trends and seasonal patterns. The dataset from January 2004 to October 2022 was designated as the training set, and November 2022 to October 2023 as the test set. Model parameters were optimized using the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Performance was evaluated using five metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and others. **Results**　TB incidence showed a significant declining trend ($AAPC_{2004—2023}$=−3.48, 95%*CI*: −4.39 to −2.77), with seasonal peaks in March (seasonal index: 1.16) and troughs in February (0.84). The SARFIMA (0, 0.49, 1) (2, 0.44, 2)$_{12}$ model outperformed SARIMA (0, 1, 1) (2, 1, 2)$_{12}$ across all metrics, with lower MAE (0.53 vs. 0.92), MAPE (0.16 vs. 0.28), and RMSE (0.90 vs. 1.12). Residual tests confirmed white noise for SARFIMA (Ljung-Box *Q*=0.31, *P* > 0.05). The projected annual TB incidence rate for 2023 to 2030 is 4.48 per 100000 (95%*CI*: 1.95 to 7.01). **Conclusion**　By leveraging fractional differencing to capture long-memory characteristics, the SARFIMA model achieved superior accuracy over SARIMA in predicting TB trends. Integrating SARFIMA into TB surveillance systems is recommended to dynamically optimize prevention strategies and advance progress toward the 2030 TB elimination target.

【Key words】 Tuberculosis; SARFIMA model; SARIMA model; Forecast; Time series analysis

结核病（tuberculosis，TB）是由结核分枝杆菌引起的全球重大公共卫生威胁，空气传播特性导致以肺结核为主的复杂流行模式，全球约 1/4 人口曾感染。尽管已有防控措施，但 TB 仍是仅次于 COVID−19 的第二大单一传染病致死病因[1]。我国作为 TB 高负担国家，疫情防控依赖精准的流行趋势预测。

传统时间序列模型中，季节性自回归差分移动平均模型（seasonal autoregressive integrated moving average，SARIMA）因能处理趋势与季节效应，在传染病预测中广泛应用[2−10]，但其整数阶差分仅适用于短记忆性数据，对长记忆性序列可能导致过度差分。相比之下，季节性自回归分数差分移动平均模型（seasonal autoregressive fractionally integrated moving average，

*基金项目：河南省自然科学基金项目（222300420265）；河南省高等学校重点科研项目（21A330004）
△通信作者：王永斌，E-mail：wybwho@163.com

SARFIMA）通过分数阶差分（0<d<1）兼顾长记忆性与复杂季节性，在肾综合征出血热、病毒性肝炎等疾病预测中表现更优[11-13]。然而，SARFIMA尚未应用于TB流行趋势预测。

本研究基于全国TB月发病数据（2004—2023年），构建SARFIMA模型，结合COVID-19干扰控制变量，系统评估其预测效能，旨在为TB防控提供更精准的模型工具，并预测2023—2030年流行趋势。

材料与方法

1. 数据收集

2004年1月至2023年10月全国TB月发病数据源自中国疾病预防控制中心（https://www. chinacdc. cn/），人口数据来自《国家统计年鉴》。所有病例均依据国家卫生健康委员会发布的TB诊断标准（http://www. nhc. gov. cn/wjw/s9491/wsbz. shtml）确诊。研究将2004年1月至2022年10月的数据作为训练集，2022年11月至2023年10月的数据作为测试集，用于模型拟合与验证。为控制COVID-19大流行的影响，设置虚拟变量：2020年1月—2022年2月赋值为1，其余时段为0。

2. 研究方法

（1）SARIMA模型原理

SARIMA通过差分运算将非平稳序列转化为平稳序列，并融合季节性与非季节性自回归（autoregressive，AR）、移动平均（moving average，MA）成分，其结构表示为：SARIMA$(p,d,q)(P,D,Q)_s$，其中：(p,d,q)分别为非季节性自回归阶数、差分次数和移动平均阶数；$(P,D,Q)_s$分别为季节性自回归阶数、差分次数和移动平均阶数；s为季节周期（s=12）[14-15]。

（2）SARIMA模型建模步骤

①序列平稳化

白噪声检验：通过Ljung-Box Q检验（P<0. 05）确认序列非随机性[16]。

平稳性检验：联合增广迪基-福勒检验（augmented Dickey-Fuller test，ADF）与科瓦特科夫斯基-菲利普斯-施密特-辛检验（Kwiatkowski-Phillips-Schmidt-Shin test，KPSS）判定序列平稳性。非平稳序列需通过差分处理（1阶季节差分D=1与1阶非季节差分d=1）[15-18]。

②参数识别与优化

基于自相关函数（autocorrelation function，ACF）与偏自相关函数（partial autocorrelation function，PACF）的拖尾/截尾特征初选$(p,q)(P,Q)$组合。通过赤池信息准则（Akaike information criterion，AIC）、贝叶斯信息准则（Bayesian information criterion，BIC）筛选最优模型，优先选择AIC/BIC值最小、对数似然（log-likelihood，LL）值最大的组合。

③残差诊断

采用Ljung-Box Q检验验证残差是否为白噪声（P>0. 05）。通过ACF/PACF图与正态性检验图（quantile-quantile plot，Q-Q图）检验残差自相关性及正态性[15,19]。

④预测验证

基于测试集（2022. 11—2023. 10）计算预测误差指标，包括平均绝对误差（mean absolute error，MAE）、平均绝对百分比误差（mean absolute percentage error，MAPE）、均方根误差（root mean square error，RMSE）等。

（3）SARFIMA模型原理

ARFIMA模型通过分数阶差分（0<d<1）处理具有长记忆性的时间序列，解决了SARIMA模型因整数差分导致的过度差分问题[20]。其核心优势在于捕捉ACF的双曲线衰减特征（长记忆性）[13]，而SARIMA模型仅适用于ACF呈指数衰减的短记忆性序列。

（4）SARFIMA模型建模步骤

①长记忆性检验

采用重标极差法（R/S分析）计算赫斯特指数（Hurst index，H）以量化时间序列的长记忆性，其中分数阶差分参数与H的关系为$d=H-0.5$[13]。赫斯特指数的判断标准为：0≤H<0. 5时序列呈反持久性布朗运动（“均值回复”）；H=0. 5时为标准随机游走，两者均无长记忆性，适用传统SARIMA模型；0. 5<H<1时序列具有长记忆性，自相关函数呈双曲线衰减，适用SARFIMA模型；H=1时序列完全线性相关（实际罕见）[10]。

②分数阶差分与模型结构

ARFIMA模型通过分数阶差分处理长记忆性序列，避免传统整数差分导致的信息丢失[20]。SARFIMA模型作为其季节性扩展，允许序列进行分数差分，将SARIMA的整数差分阶数扩展为分数值，适用于同时具有长记忆性和季节性的序列，其模型结构如下：

（a）简单分数差分模型：$(1-B)^d x_t=\varepsilon_t$（$d\in(-0.5,0.5)$，无季节性）；

（b）季节性分数差分模型：$(1-B^s)^d x_t=\varepsilon_t$（结合季节性与分数差分）；

（c）SARFIMA模型：$(1-B^s)^d\Omega(B)x_t=\Theta(B)\varepsilon_t$，包含季节自回归[$\Omega(B)$]和移动平均[$\Theta(B)$]成分，实现长记忆性与季节性的联合建模[13]。模型参数表示为SARIMA$(p,d^1,q)(P,D^1,Q)_s$，其中：$d^1=d+d_f$（d：整数差分阶数，d_f：分数差分参数，$d_f=H-0.5$）；$D^1=D+D_f$（D：季节整数差分，D_f：季节分数差分）[13,20]。

③建模步骤与软件实现

步骤与SARIMA一致（数据检验→定阶→诊断→

预测），通过 R 语言 forecast 和 arfima 包（版本 4. 3. 2）实现，以 LL 值最大、AIC/BIC 值最小为最优模型筛选标准。纳入 COVID-19 影响变量（2020. 1—2022. 2 为 1，其余为 0），控制疫情对 TB 就诊与传播的干扰。

（5）两种模型的比较

采用 2022 年 11 月—2023 年 10 月数据作为测试集，通过以下指标量化比较模型预测精度：

①MAE $=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t|$；

②MAPE $=\frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right|$；

③RMSE $=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$；

④平均误差率（mean error rate，MER）：MER $=\frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\hat{y}_t|}{y_t}$；

⑤均方根预测误差（root mean squared prediction error，RMSPE）：RMSPE $=\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t-\hat{y}_t}{y_t}\right)^2}$；

其中，MAE、MAPE 和 RMSE 反映预测值与实际值的绝对偏差；MER 和 RMSPE 衡量相对误差与预测稳定性。指标值越小，模型预测性能越优[13,20]。

结　果

1. 数据描述

2004 年 1 月至 2023 年 10 月，中国累计报告 TB 病例 3827797 例，年均发病率为 88. 35/10 万（月均 7. 36/10 万）。Joinpoint 回归分析显示（图 1），TB 发病率总体呈显著下降趋势（平均年度百分比变化，average annual percentage change，$AAPC_{2004—2023}$=−3. 48%，95% *CI*：−4. 39%~−2. 77%），但存在两个显著性拐点，形成三阶段特征：2004—2006 年呈快速上升（年度百分比变化，annual percentage change，APC=14. 75%，95% *CI*：5. 43%~22. 84%），2006—2018 年转为缓慢下降（APC=−3. 57%，95% *CI*：−4. 41%~−2. 70%），2018 年后加速下降（APC=−9. 73%，95% *CI*：−16. 62%~−7. 03%）。季节特征分析表明（图 2），发病呈规律性波动：1—2 月为低谷期（季节性指数 0. 85、0. 84），3—4 月达高峰（指数 1. 16、1. 15），其余月份趋近年均水平（指数 1. 00±0. 05），提示春季为 TB 防控关键窗口期。

2. SARIMA 模型构建与验证

基于 2004 年 1 月至 2023 年 10 月全国 TB 月发病率数据（图 3 ~ 4），原始序列呈现显著趋势性与季节性特征，经单位根检验确认为非平稳白噪声序列。ADF 检验结果显示统计量为−0. 91（*P*=0. 33），未能拒绝原假设（存在单位根）；而 KPSS 检验统计量为 3. 72（*P*<0. 001），拒绝趋势平稳假设，表明需通过差分处理实现平稳化。采用 1 阶非季节差分（*d*=1）联合 1 阶季节差分（*D*=1，*s*=12）后，ADF 检验统计量显著改善至 −17. 30（*P*<0. 001），确认序列达到平稳状态（图 5）。

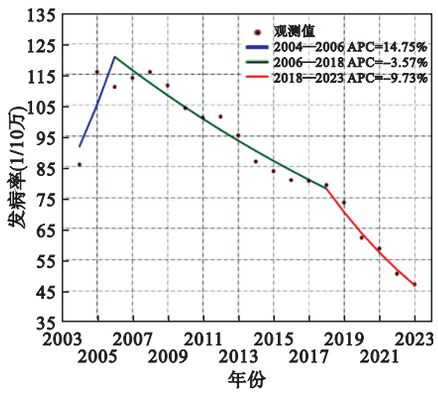


图 1　2004—2023 年中国 TB 发病率趋势分解图

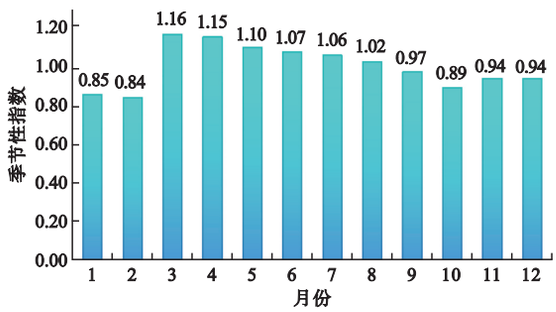


图 2　TB 季节性指数图

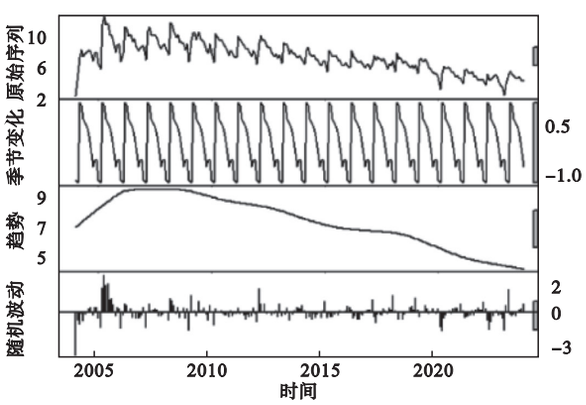
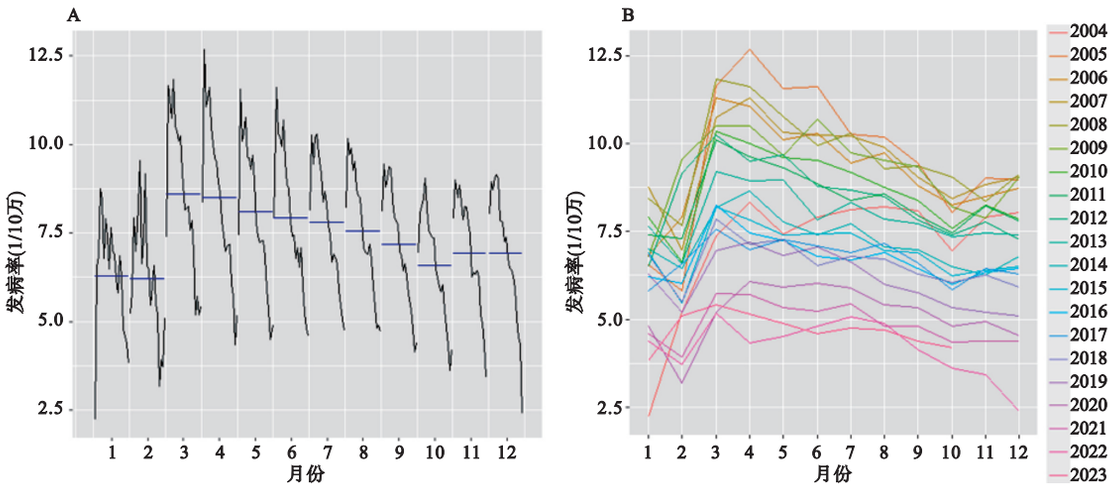


图 3　2004 年 1 月—2023 年 10 月 TB 月发病率原始序列与分解图

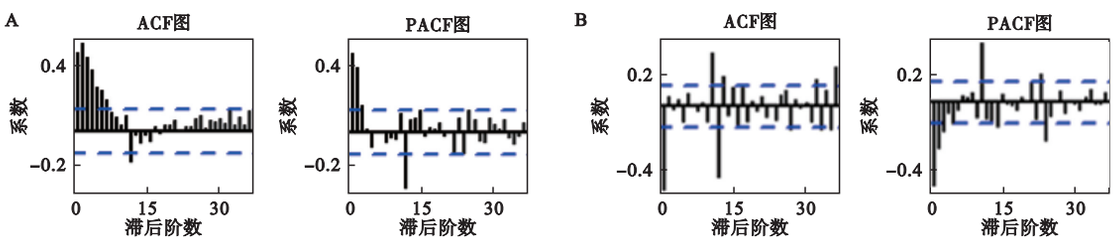
利用 R 语言 forecast 包的 auto. arima 函数自动优化模型，依据 AIC=448. 02 与 BIC=468. 52，确定最优模型为 SARIMA（0，1，1）（2，1，2）$_{12}$。模型参数估计结果如表 1 所示：非季节性移动平均项（ma1=−0. 614，*P*<0. 001）、季节性自回归项（sar1=−0. 924，*P*<0. 001；sar2=−0. 303，*P*<0. 05）及移动平均项（sma1=0. 276，*P*<0. 05；sma2=−0. 425，*P*<0. 001）均通过显著性检验，表明模型拟合效果良好。

进一步对模型残差进行诊断（图 6），ACF 图（图 6A）显示，除滞后 11 阶与 32 阶外，残差自相关系数均落入 95% 置信区间内，可认为残差相互独立；Ljung-Box *Q* 检验统计量为 38. 69（*P*>0. 05），支持残差序列为白噪声的假设。此外，残差序列的正态性分布图（图 6



注：A. 原始序列月份分布：展示2004—2023年TB月发病率原始数据连接情况，“—”表示平均值；B. 原始序列年份趋势（按月展示）：呈现2004—2023年各年份TB月发病率趋势

图4 2004—2023年TB月发病率原始序列特征



注：A. 1阶季节性差分；B. 1阶季节性差分+非季节性差分

图5 SARIMA模型差分后的ACF图和PACF图

B）和Q-Q图（图6C）显示残差近似服从正态分布。上述结果表明，SARIMA$(0,1,1)(2,1,2)_{12}$模型能够有效捕捉TB发病率的动态特征，可用于后续预测分析。

表1 SARIMA$(0,1,1)(2,1,2)_{12}$参数估计与显著性检验

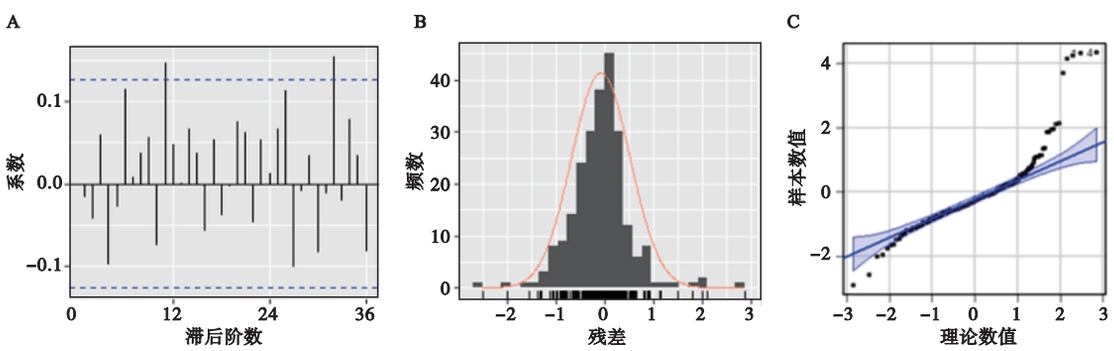
参数	估计值	标准误	z值	P值
ma1	−0.613941	0.052532	−11.6870	<0.001
sar1	−0.923873	0.124901	−7.3968	<0.001
sar2	−0.302937	0.110031	−2.7532	<0.05
sma1	0.276017	0.129492	2.1315	<0.05
sma2	−0.425119	0.115764	−3.6723	<0.001

3. SARFIMA模型

基于R/S分析计算得出H=0.82，表明TB发病序列具有显著长记忆性（0.5<H<1），适用于构建SARFIMA模型。分数差分阶数由公式$d_f=H-0.5$=0.32确定，结合LL值最大、AIC/BIC值最小（LL=95.77；AIC=−173.55；BIC=−142.76）筛选最优模型，最终选定SARFIMA$(0,0.49,1)(2,0.44,2)_{12}$。模型残差诊断显示（图7）：Ljung-Box Q检验统计量Q=0.31（P>0.05），ACF与PACF图均无显著滞后项，表明残差为白噪声序列，满足模型有效性要求。

4. 模型评价与应用

如表2所示，SARFIMA$(0,0.49,1)(2,0.44,2)_{12}$模型在测试集（2022年11月—2023年10月）的相对误差（relative error，RE）显著低于SARIMA$(0,1,1)(2,1,2)_{12}$模型。具体而言，SARFIMA的MAE为0.53（SARIMA：0.92，降幅42.4%），MAPE为0.16%（SARIMA：0.28%，降幅42.9%），且RMSE（0.90



注：A. ACF图；B. 残差正态性分布图；C. Q-Q图

图6 SARIMA模型残差序列诊断图

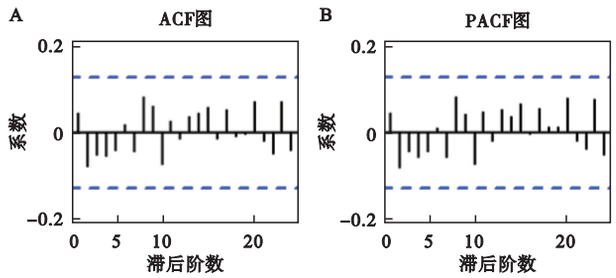


图 7　SARFIMA 模型残差序列 ACF 图和 PACF 图

vs. 1. 12)等指标均显示其预测效能更优(表 3)。进一步应用 SARFIMA 模型对 2030 年前 TB 发病趋势进行预测(图 8),结果显示发病率将于 2023 年后进入平台期,年均值为 4. 48/10 万(95% *CI*: 1. 958/10 万~7. 01/10 万),其中 2029 年可能出现小幅回升至 4. 63/10 万(表 4)。值得注意的是,远期预测的置信区间范围逐年扩大(如 2030 年:1. 43/10 万~7. 75/10 万),提示需结合动态监测数据优化防控策略。

表 2　SARIMA 与 SARFIMA 候选模型预测值及相对误差比较

时间(年-月)	实际值	SARIMA $(0,1,1)(2,1,2)_{12}$		SARFIMA$(0,0.49,1)(2,0.44,2)_{12}$	
		预测值	*RE*(%)	预测值	*RE*(%)
2022-11	3. 42	3. 82	11. 70	3. 82	11. 68
2022-12	2. 40	3. 57	48. 75	3. 87	61. 29
2023-01	3. 81	3. 49	8. 40	3. 93	3. 04
2023-02	5. 09	2. 29	55. 01	2. 51	50. 68
2023-03	5. 41	4. 22	22. 00	4. 97	8. 14
2023-04	5. 16	4. 07	21. 12	4. 50	12. 71
2023-05	4. 89	4. 09	16. 36	4. 66	4. 69
2023-06	4. 59	3. 99	13. 07	4. 48	2. 48
2023-07	4. 74	4. 32	8. 86	4. 83	1. 97
2023-08	4. 71	4. 07	13. 59	4. 70	0. 25
2023-09	4. 38	3. 68	15. 98	4. 33	1. 13
2023-10	4. 20	3. 25	22. 62	3. 95	6. 01

表 3　SARIMA 与 SARFIMA 候选模型评估指标对比

模型	AIC	BIC	MAE	MAPE	RMSE	MER	RMSPE
SARIMA$(0,1,1)(2,1,2)_{12}$	448. 02	468. 52	0. 92	0. 28	1. 12	0. 21	0. 41
SARFIMA$(0,0.49,1)(2,0.44,2)_{12}$	−173. 55	−142. 76	0. 53	0. 16	0. 90	0. 12	0. 32

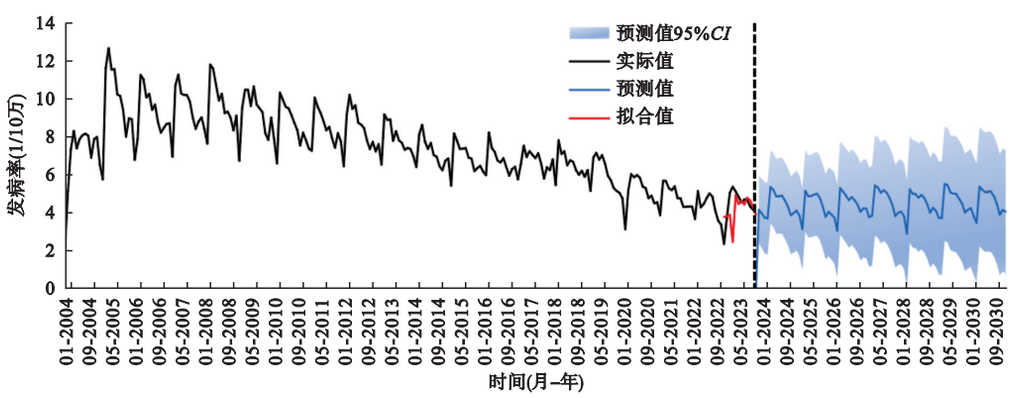


图 8　SARFIMA$(0,0.49,1)(2,0.44,2)_{12}$对全国 2023—2030 年 TB 月发病率拟合及预测图

表4　2023—2030年全国TB发病率SARFIMA模型预测值及95%*CI*

时间(年)	预测值(1/10万)	95%*CI*(1/10万)
2023	4. 11	2. 31~5. 91
2024	4. 51	2. 52~6. 49
2025	4. 44	2. 23~6. 64
2026	4. 47	1. 97~6. 97
2027	4. 60	1. 93~7. 28
2028	4. 48	1. 63~7. 33
2029	4. 63	1. 60~7. 67
2030	4. 59	1. 43~7. 75
均值	4. 48	1. 95~7. 01

讨　论

传染病预测模型是公共卫生早期预警的关键工具。传统统计方法如神经网络、灰色预测等虽应用广泛,但难以捕捉疾病流行的长记忆性与复杂动态特征[21-23]。时间序列分析通过整合时间变量替代多元影响因素,在慢性传染病短期预测中具有独特优势[24]。世界卫生组织《2023 年全球结核病报告》显示,TB 仍是全球第二大单一传染病死因,死亡人数约为艾滋病的两倍,精准预测其流行趋势对防控至关重要。

1. 模型创新与预测效能

本研究首次将 SARFIMA 应用于 TB 发病率预测。SARFIMA 通过分数差分克服了传统 SARIMA 模型整数差分导致的长记忆性信息丢失问题,能有效捕捉 TB 流行的双曲线衰减特征(长记忆性)和不规则季节性波动[9-10]。从预测性能指标看,SARFIMA 的 MAE(降幅 42. 4%)、MAPE(降幅 42. 9%)等核心指标显著优于 SARIMA,且残差白噪声检验(Q=0. 31, P>0. 05)与拟合趋势图进一步验证了其可靠性。这一优势与 SARFIMA 在金融市场、交通流和空气质量预测等其他领域的表现一致[25-31],也证实了其在传染病长记忆性动态建模中的适用性,将其纳入 TB 监测体系可提升预测精度,为防控资源调配提供科学依据。

2. 流行病学启示与防控策略

研究预测显示，TB发病率在2023年后将进入平台期（年均4.48/10万，95% *CI*：1.95/10万~7.01/10万），且远期预测不确定性逐年增加（2030年95% *CI*：1.43/10万~7.75/10万）。这一趋势可能与COVID-19大流行导致的TB诊断延误积压、防控资源挤兑及公众就医行为改变密切相关[32-33]。世界卫生组织指出，2022年全球TB病例的“复苏性增长”反映了既往漏诊患者的累积效应[1]，提示需加强主动筛查、优化分级诊疗体系以应对平台期风险。此外，TB发病存在显著季节性波动（3—4月高峰，1—2月低谷），与春运人口流动增加暴露风险、冬季维生素D缺乏抑制免疫力等有关[34-35]，因此冬春季节需强化高危人群监测与健康宣教。

3. 区域异质性与模型普适性

2004—2023年我国TB发病率呈现“先升后降、近年加速”的阶段性特征，与河南省研究一致[36]，但与山西、山东等地差异显著[24,37]，反映出区域经济水平、卫生资源配置等因素导致的流行异质性[38]。这种异质性提示需构建区域特异性预测模型以适配本地化防控需求。尽管SARFIMA在本研究中表现出优势，但其在跨区域、跨病种应用中的普适性仍需验证，后续研究可结合机器学习方法进一步捕捉非线性时空交互效应。

本研究存在以下局限性：一是数据依赖被动监测系统，可能存在漏报偏倚，影响模型精度；二是SARFIMA对样本量要求较高（建议$n \geqslant 100$），在短期数据中应用受限；三是未整合社会经济、气候等外部协变量，可能低估复杂驱动因素的影响。未来需通过多源数据融合（如电子健康档案、气象数据）提升模型精度，并在不同流行病学场景中验证SARFIMA的迁移能力，结合空间流行病学方法构建多维度预测模型，提升防控策略的区域针对性。

参 考 文 献

[1] Bagcchi S. WHO's global tuberculosis report 2022[J]. The Lancet Microbe,2023,4(1): e20.

[2] Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France[J]. Science of the Total Environment,2020,729: 138817.

[3] 黎婷婷,尹钰,邹莹,等. 基于ARIMA模型的乌鲁木齐市水痘发病趋势预测与分析[J]. 实用预防医学,2024,31(8): 1008-1011.

[4] 王小亭,李明春,周宪君. 2005—2021年佳木斯市流行性腮腺炎流行病学特征与预测模型研究[J]. 中国初级卫生保健,2024,38(1): 91-95.

[5] 钱晓涵,郑敏,姚永明,等. 贵州省2010—2022年梅毒流行特征分析及ARIMA模型预测[J]. 应用预防医学,2023,29(6): 383-387.

[6] 马一鸣,丁勇. 我国2012—2021年4种肝炎流行趋势的时间序列分析和预测[J]. 南京医科大学学报（自然科学版）,2024,44(1): 72-79.

[7] 朱思谨,孙聪,初天慈,等. 2012—2022年烟台市肺结核流行病学特征及趋势预测[J]. 中国医院统计,2024,31(1): 35-40,46.

[8] 滕冲,王玉兰,刘柳,等. 2013—2022年北京市东城区肺结核报告发病流行特征及发病预测分析[J]. 中国防痨杂志,2024,46(4): 397-402.

[9] 刘凯,张晞. 基于分数阶差分ARIMA模型的煤炭消费预测[J]. 矿业科学学报,2017,2(5): 489-496.

[10] 王晓瑞,刘元,王彤. 基于分数整合自回归移动平均模型的山西省手足口病预测研究[J]. 疾病监测,2023,38(7): 865-871.

[11] 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.

[12] Wang Y,Liang Z,Qing S,et al. Application of an ARFIMA model to estimate hepatitis C epidemics in Henan,China[J]. American Journal of Tropical Medicine and Hygiene,2024,110(2): 404-411.

[13] Qi C,Zhang D,Zhu Y,et al. SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with SARIMA[J]. BMC Medical Research Methodology,2020,20(1): 243.

[14] 阎涵. 辽宁省719例艾滋病病人/感染者生存状况及影响因素分析[D]. 沈阳: 中国医科大学,2018.

[15] 郭玉秀. 2004—2016年我国艾滋病流行特征分析及月发病率预测研究[D]. 合肥: 安徽医科大学,2019.

[16] 刘娟. 基于时间序列理论方法的流感病毒DNA序列特征分析[D]. 无锡: 江南大学,2011.

[17] 胡学锋,吴霜,吴媛媛,等. 1997—2020年中国大陆登革热病例数量变化及时间序列拟合研究[J]. 中国国境卫生检疫杂志,2022,45(3): 216-219.

[18] 李阳,陈坚,廖建英,等. 基于ARIMA时间序列的广西百色市艾滋病流行趋势分析[J]. 右江民族医学院学报,2023,45(1): 134-138.

[19] 李丽,刘亦新,杨云,等. 学校结核病时间序列分析及预测模型研究[J]. 公共卫生与预防医学,2022,33(5): 106-110.

[20] Chang F,Huang H,Chan AHS,et al. Capturing long-memory properties in road fatality rate series by an autoregressive fractionally integrated moving average model with generalized autoregressive conditional heteroscedasticity: a case study of Florida, the United States, 1975-2018[J]. Journal of Safety Research,2022,81: 216-224.

[21] Gan R,Chen N,Huang D. Comparisons of forecasting for hepatitis in Guangxi Province,China by using three neural networks models[J]. PeerJ,2016,4: e2684.

[22] Zheng Y,Zhang L,Zhu X,et al. A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China[J]. PLoS One,2020,15(6): e0234660.

[23] 袁莺楹,董建成. 基于数学模型的疾病预测方法比较研究[J]. 软件导刊,2009,8(5): 108-110.

[24] 梁建国. 山西省2010—2020年结核病流行特征分析及发病预测研究[D]. 太原: 山西医科大学,2023.

[25] 叶静,赵凯,王传稳,等. 基于ARFIMA-GARCH模型的黄金价格分析及预测[J]. 青岛大学学报（自然科学版）,2014,27(4): 10-14.

[26] 陈鹏;李星野. 基于ARFIMA-GARCH模型族的黄金价格预测分析[J]. 电子商务,2018(5): 42-44.

[27] 汪妍. 基于ARFIMA-GARCH-LSTM混合模型的股价预测研究[D]. 上海: 上海师范大学,2022.

（下转第332页）

态FBG与NSCLC患者OS关联估计的准确性。

参 考 文 献

[1] 张雪燕,马晓天,侯子茜,等. 胸腔镜非小细胞肺癌术后患者肺部并发症风险预测模型的建立[J]. 中国卫生统计,2024,41(1): 102-105.

[2] 李艳华,魏素菊,王俊艳,等. 非小细胞肺癌患者凝血功能异常的相关因素与预后分析[J]. 中国肺癌杂志,2014,17(11): 789.

[3] Hoffman PC,Mauer AM,Vokes EE. Lung cancer [J]. Lancet,2000,355(9202): 479-485.

[4] PDQ Adult Treatment Editorial Board. Non-Small Cell Lung Cancer Treatment (PDQ®): Health Professional Version [EB/OL]. (2025-05-07)[2025-11-23]. https://www. ncbi. nlm. nih. gov/books/NBK65865/.

[5] Vasudevan S, Krishna V, Mehta A. Lung cancer in non-smokers: clinicopathological and survival differences from smokers [J]. Cureus,2022,14(12): e32417.

[6] Sheikh M,Mukeriya A,Shangina O,et al. Postdiagnosis smoking cessation and reduced risk for lung cancer progression and mortality: a prospective cohort study [J]. Annals of Internal Medicine,2021,174(9): 1232-1239.

[7] Torres-Roman JS, Valcarcel B, Martinez-Herrera JF, et al. Mortality Trends for Lung Cancer and Smoking Prevalence In Peru [J]. Asian Pacific Journal of Cancer Prevention: APJCP,2022,23(2): 435.

[8] Jones CC, Mercaldo SF, Blume JD, et al. Racial disparities in lung cancer survival: the contribution of stage,treatment,and ancestry [J]. Journal of Thoracic Oncology,2018,13(10): 1464-1473.

[9] Wei S, Tian J, Song X, et al. Causes of death and competing risk analysis of the associated factors for non-small cell lung cancer using the Surveillance, Epidemiology, and End Results database[J]. Journal of Cancer Research and Clinical Oncology, 2018, 144(1): 145-155.

[10] Wang W, Dou S, Dong W, et al. Impact of COPD on prognosis of lung cancer: from a perspective on disease heterogeneity [J]. International Journal of Chronic Obstructive Pulmonary Disease, 2018, 13: 3767-3776.

[11] Qi Y, Fu J. Research on the coagulation function changes in non small cell lung cancer patients and analysis of their correlation with metastasis and survival [J]. J Buon,2017,22(2): 462-467.

[12] 燕启田. D-二聚体、纤维蛋白原水平与肝癌根治术后患者预后的相关性[J]. 中国民康医学,2023,35(17): 127-129.

[13] 庞杏峰. 回顾性分析凝血指标PT、FIB在原发性肝癌中的变化及其预后价值[D]. 南宁:广西医科大学,2022.

[14] Zhang K,Xu Y,Tan S,et al. The association between plasma fibrinogen levels and lung cancer: a meta-analysis [J]. Journal of Thoracic Disease,2019,11(11): 4492.

[15] Zhang Y,Cao J,Deng Y,et al. Pretreatment plasma fibrinogen level as a prognostic biomarker for patients with lung cancer [J]. Clinics,2020,75: e993.

[16] Kim KH, Park TY, Lee JY, et al. Prognostic significance of initial platelet counts and fibrinogen level in advanced non-small cell lung cancer [J]. Journal of Korean Medical Science,2014,29(4): 507.

[17] Jiang HG,Li J,Shi SB,et al. Value of fibrinogen and D-dimer in predicting recurrence and metastasis after radical surgery for non-small cell lung cancer [J]. Medical Oncology,2014,31: 1-9.

[18] Mitsui S, Tanaka Y, Doi T, et al. Prognostic value of preoperative plasma fibrinogen levels in resected stage I non-small cell lung cancer [J]. Thoracic Cancer,2022,13(10): 1490-1495.

[19] Fay MP,Graubard BI,Freedman LS,et al. Conditional logistic regression with sandwich estimators: application to a meta-analysis [J]. Biometrics,1998,54(1):195-208.

[20] Harrell FE. Regression modeling strategies: with applications to linear models,logistic regression,and survival analysis [M]. New York: Springer,2001.

[21] Palumbo JS, Talmage KE, Massari JV, et al. Platelets and fibrin (ogen) increase metastatic potential by impeding natural killer cell‐mediated elimination of tumor cells [J]. Blood, 2005, 105(1): 178-185.

[22] Xiao Y,Xie H,Xie Z,et al. Kinetics of postdiagnosis platelet count with overall survival of pancreatic cancer: a counting process approach [J]. Cancer Medicine,2016,5(5): 881-887.

[23] 肖媛媛,许传志,赵耐青. 常用生存分析模型及其对时依性协变量效应的估计方法[J]. 中国卫生统计,2016,33(3): 543-547.

(责任编辑:邓　妍)

(上接第327页)

[28] 陈健红. 基于WRV-ARFIMA-T模型股市波动率的长记忆性及预测研究[D]. 武汉: 武汉理工大学,2019.

[29] 陈雪,申建红,徐文慧,等. 基于ARFIMA模型的钢材价格预测研究[J]. 河北工程大学学报(自然科学版),2020,37(3): 64-68.

[30] 靳小飞. 基于ARFIMA混合模型的短时交通流预测研究[D]. 武汉: 华中科技大学,2020.

[31] 马跃. 基于ARFIMA-GARCH-F模型对重庆市空气质量指数动态VaR极值的研究[D]. 重庆: 西南大学,2017.

[32] 鲁进,孙文艳,肖作汉,等. 新冠肺炎疫情对世界结核病防控的影响[C]//中国防痨协会2023年第34届全国学术大会暨结核病诊防治新技术推广应用论坛论文汇编. 2023: 104.

[33] 乐博昕,刘效峰,王娜,等. 基于ARIMA的新冠肺炎疫情前后结核病流行趋势预测与分析[J]. 实用预防医学,2022,29(11): 1299-1302.

[34] 龚浩;郭在金;周罗晶. SARIMA模型和BPNN模型在江苏省肺结核发病预测中的应用[J]. 中国预防医学杂志,2024,25(2): 239-244.

[35] Kohei Y,Sumi A,Kobayashi N. Time-series analysis of monthly age-specific numbers of newly registered cases of active tuberculosis in Japan from 1998 to 2013[J]. Epidemiology and Infection,2016,144(11): 2401-2414.

[36] 蒋建国,孙定勇,张艳秋,等. 应用ARIMA模型对河南省肺结核流行趋势预测分析[J]. 现代疾病预防控制,2023,34(7): 495-499.

[37] 田庆,刘永鹏,张晶晶,等. ARIMA乘积季节模型在山东省肺结核发病预测中的应用[J]. 山东大学学报(医学版),2021,59(7): 112-118.

[38] 张曼晖. 我国肺结核控制目标实现及影响因素研究[D]. 北京:中国疾病预防控制中心,2019.

(责任编辑:邓　妍)