

# 基于SARIMA时间序列模型的台风频次预测

王 依, 黄培煌

闽江学院数学与数据科学学院(软件学院), 福建 福州

收稿日期: 2023年11月26日; 录用日期: 2023年12月21日; 发布日期: 2023年12月29日

## 摘 要

台风是一种破坏力极大的灾害性天气, 因此预测和预报台风, 历来是气象工作的一项重要任务。对台风进行精准预测, 并制定相应的预防和应急措施, 是减轻台风造成灾难的重要手段。本文基于季节性差分自回归滑动平均模型(Seasonal Autoregressive Integrated Moving Average, SARIMA)研究台风频次的预测方法。该模型通过考虑时间序列的季节性和趋势性变化研究台风频次, 旨在对未来台风频次提供准确预测方法。通过预处理台风发生频次相关数据, 对时间序列进行平稳性检验以及白噪声测试, 采用AIC遍历对模型定阶, 计算模型的均方根误差(RMSE)和平均绝对误差(MAE)并绘制模型的残差分布和自相关图, 分析比较后认为该模型的拟合效果较好。最后对2024年1月至2024年12月的台风频次进行预测, 为提高自然灾害应对和相关政策制定提供了有力支持。

## 关键词

SARIMA模型, 台风频次, 季节性影响, 求和自回归移动平均模型

# Typhoon Frequency Prediction Based on SARIMA Time Series Model

Yi Wang, Peihuang Huang

School of Mathematics and Data Science (School of Software), Minjiang University, Fuzhou Fujian

Received: Nov. 26<sup>th</sup>, 2023; accepted: Dec. 21<sup>st</sup>, 2023; published: Dec. 29<sup>th</sup>, 2023

## Abstract

Typhoon is a highly destructive and catastrophic weather, so predicting and forecasting typhoons has always been an important task in meteorological work. Accurately predicting typhoons and developing corresponding prevention and emergency measures is an important means of mitigating disasters caused by typhoons. This article is based on the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to study the prediction method of typhoon frequency.

This model studies typhoon frequency by considering the seasonal and trend changes of time series, aiming to provide accurate prediction methods for future typhoon frequencies. By preprocessing the frequency related data of typhoons, conducting stationarity tests and white noise tests on the time series, using AIC traversal to determine the order of the model, calculating the root mean square error (RMSE) and mean absolute error (MAE) of the model, and drawing the residual distribution and autocorrelation diagram of the model, it is believed that the fitting effect of the model is good after analysis and comparison. Finally, the prediction of typhoon frequency from January 2024 to December 2024 provides strong support for improving natural disaster response and related policy formulation.

## Keywords

SARIMA Model, Typhoon Frequency, Seasonal Impact, Sum Autoregressive Moving Average Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 研究背景与意义

自然灾害,尤其是台风活动,对社会和环境造成了巨大的影响。这种天气现象不仅对人们的生命和财产构成威胁,还可能导致长期的社会和经济影响。随着全球气候变化的不断加剧,对台风活动进行准确预测变得尤为关键,以更好地理解 and 应对其潜在威胁。在过去的研究中,时间序列分析已经被广泛应用于气象和气候相关的领域。

时间序列是指将某种现象在不同时间的观测值按时间先后顺序排列而成的数列[1]。通常,一组时间序列的时间间隔为固定值,因此可以将时间序列作为离散时间数据进行分析和处理。时间序列分析广泛应用于金融经济、农业、气象水文、信号处理、工程等领域[2]。其中,SARIMA (Seasonal Autoregressive Integrated Moving Average)模型以其对时间序列数据中季节性和趋势性变化的有效捕捉而备受关注。SARIMA 模型在考虑历史数据的基础上,结合了自回归、差分和移动平均等成分,使其成为一种强大的工具,可用于对台风频次进行预测。本文探索了 SARIMA 模型在台风频次预测中的潜在应用,旨在通过考虑季节性和趋势性变化,对未来台风频次进行预测。

### 1.2. 模型选取

过去几十年来,时间序列分析已经成为研究气象学、经济学和其他领域问题的重要方法。相关文献强调了时间序列分析在捕捉随时间变化的数据模式和趋势中的优越性。自回归(AR)、差分(I)、移动平均(MA)等组成部分的组合在解释时间序列中的自相关性和趋势性方面取得了显著进展[3] [4]。

SARIMA 模型作为时间序列分析的扩展,通过引入季节性成分,弥补了传统 ARIMA 模型对季节性数据分析的不足。相关文献表明,在捕捉时间序列中的季节性和趋势性变化方面,SARIMA 模型具有显著的优越性[5]。其综合考虑了自回归、差分、移动平均和季节性成分,使其成为适用于研究多变量时间序列的强大工具。在过去的研究中,SARIMA 模型在气象学、经济学和流行病学等领域的成功应用引起了广泛关注。对这些应用的深入分析表明,SARIMA 模型在处理季节性数据和趋势性变化方面具有出色的性能。

对于气象灾害的预测, 台风频次的研究一直备受关注。先前的文献突出了台风预测的重要性, 尤其是在气候变化日益引起关注的今天。过去的研究努力提高了对台风路径和强度的预测准确性, 但频次预测的可靠性同样至关重要, 特别是在风险管理和紧急响应方面。由于台风生成频次具有明显的季节性, 因此本文采用具有季节性差分的 SARIMA 模型对台风频次进行建模和预测。SARIMA 模型在时间序列分析和台风预测中的重要性, 为本研究提供了坚实的理论基础。在此基础上, 本文将进一步探讨 SARIMA 模型在预测台风频次方面的潜在优势, 并通过实证研究验证其在实际应用中的效果。

## 2. SARIMA 季节性差分自回归滑动平均模型

### 2.1. 模型介绍

时间序列分析方法由 Box-Jenkin 于 1976 年提出, 它适用于各种领域的时间序列分析。ARIMA 模型是一种用来实现时间序列预测分析的模型, ARIMA 预测模型通常可以分解为长期趋势、季节性变化和随机波动三方面。

由于台风发生频次具有明显的季节性, 且季节效应、长期趋势效应和随机波动之间可能存在复杂的交互影响关系, 因此使用 ARIMA 模型的预测效果并不好, 所以本文采用季节性差分的 SARIMA 模型[6][7]对台风发生频次进行预测, 其基本模型为  $ARIMA(p,d,q) \times (P,D,Q)[s]$ , 其中  $d$  和  $D$  分别为使时间序列平稳化所需要进行的逐期差分和季节差分的阶数,  $p$  和  $q$  分别为自回归和移动平均的阶数,  $P$  和  $Q$  分别为季节自回归和季节移动平均的阶数[8],  $s$  为季节差分步长。

上述 SARIMA 模型可记为:

$$\varphi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_q(B^s)u_t \quad (1)$$

其中,  $\varphi_p(B)$  为非季节自回归多项式;  $\Phi_p(B^s)$  为季节  $P$  阶自回归算子多项式;  $\nabla^d$  为差分算子;  $\nabla_s^D$  为季节差分算子;  $y_t$  为原始时间序列;  $\theta_q(B)$  为非季节移动平均多项式;  $\Theta_q(B^s)$  为季节  $Q$  阶移动平均算子多项式;  $u_t$  为白噪声序列[9][10][11]。

### 2.2. SARIMA 模型实现基本步骤

#### (1) 数据平稳性检验

在分析时间序列数据之前, 首先要求时间序列是平稳的, 因为非平稳时间序列会出现伪回归, 导致结果不确定。因此, 需要先对时间序列绘图, 观察数据是否平稳, 并进行 ADF 单位根检验, 进一步判断是否为平稳时间序列。若数据是平稳的, 则可以直接使用 SARIMA 模型进行建模分析, 否则需要进行  $d$  阶差分 and  $D$  阶差分处理[12], 直到数据通过平稳性检验。

#### (2) 白噪声检验

白噪声检验, 即判断经过处理的稳定时间序列数据是否是随机序列, 因为随机序列不具有分析意义。

#### (3) 确定参数

实现前两个步骤后, 所需要进行分析的时间序列已成为满足建立 SARIMA 模型条件的平稳的、非随机的时间序列数据[13]。因此, 接下来需要对 SARIMA( $p,d,q$ )( $P,D,Q$ )[ $s$ ]模型进行参数的确定。可以通过分析时间序列的自相关系数图(ACF 图)和偏自相关系数图(PACF 图), 判断时间序列是否为拖尾或截尾, 由此确定趋势自回归阶数  $p$ 、趋势移动平均阶数  $q$ 、季节性自回归阶数  $P$  以及季节性移动平均阶数  $Q$  [14], 但是该方法具有一定的主观性, 因此还可借助 AIC (最小化信息量准则)和 BIC (贝叶斯信息准则)方法, 通过遍历参数寻找 AIC 或 BIC 值最小的模型, 从而确定模型的最佳参数[15][16][17]。

#### (4) 判断检验

建立 SARIMA( $p,d,q$ )( $P,D,Q$ )[ $s$ ]模型后, 需要确定模型的残差序列是否为白噪声, 即是否属于随机序列, 若是, 则说明原始时间序列中的信息已经被提取, 不用再进一步分析了, 否则需要重新进行参数的调整和确定[18]。此外, 还要观察残差图是否符合正态分布。

#### (5) 模型评估及预测

根据所选的最佳 SARIMA( $p,d,q$ )( $P,D,Q$ )[ $s$ ]模型, 对测试集进行预测, 将预测值与观测值进行比较。可通过计算均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)等标准来测试模型的拟合效果[19]。这三个判断标准的计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (4)$$

若误差较小, 则说明该模型的拟合效果较好, 具有良好的预测性能, 若模型误差较大, 则说明模型预测的性能较差。若模型拟合效果较好, 则可以使用该模型对未来数据进行预测[20]。

本文基于 SARIMA 时间序列模型对台风频次进行预测, 主要包括数据平稳化处理、白噪声检验、确定模型参数、判断检验、模型评估及预测五个步骤, 其框架如图 1 所示。

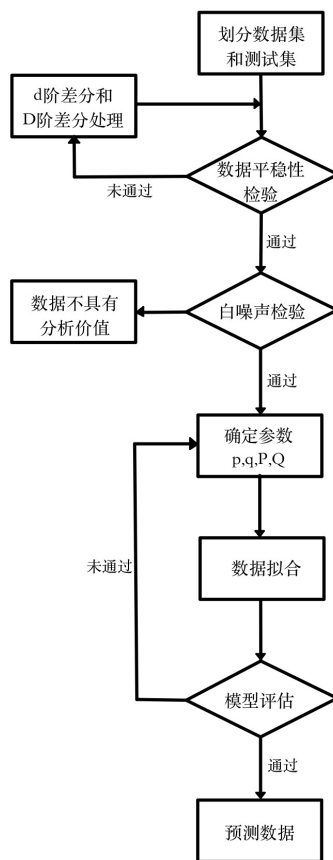


Figure 1. SARIMA time series model establishment process

图 1. SARIMA 时间序列模型建立流程

### 3. 实验研究

#### 3.1. 划分数据集和测试集

本文中所用数据为 2000 年 1 月~2022 年 12 月的台风生成频次, 所采用的模型为 SARIMA 模型。

将 2000 年 1 月~2022 年 12 月的台风频次数据集进行划分, 2000 年 1 月~2019 年 12 月的前 240 个数据作为训练集, 2020 年 1 月~2022 年 12 月的 36 个数据作为测试集。

#### 3.2. 时间序列平稳性检验

绘制 2000 年 1 月~2022 年 12 月台风频次的时间序列分解图, 如图 2 所示。原始时间序列经 ADF 检验, 得到  $ADF = -3.61$ , 而置信水平为 1%、5% 和 10% 的临界值分别为  $-3.46$ 、 $-2.87$  和  $-2.57$ 。由此可以看出, ADF 统计量小于这三组临界值, 同时  $p = 0.0056$ , 小于 0.05, 因此可以拒绝原假设, 接受备择假设, 即认为该时间序列是平稳的[21], 因此可以直接使用 SARIMA 模型进行建模分析。

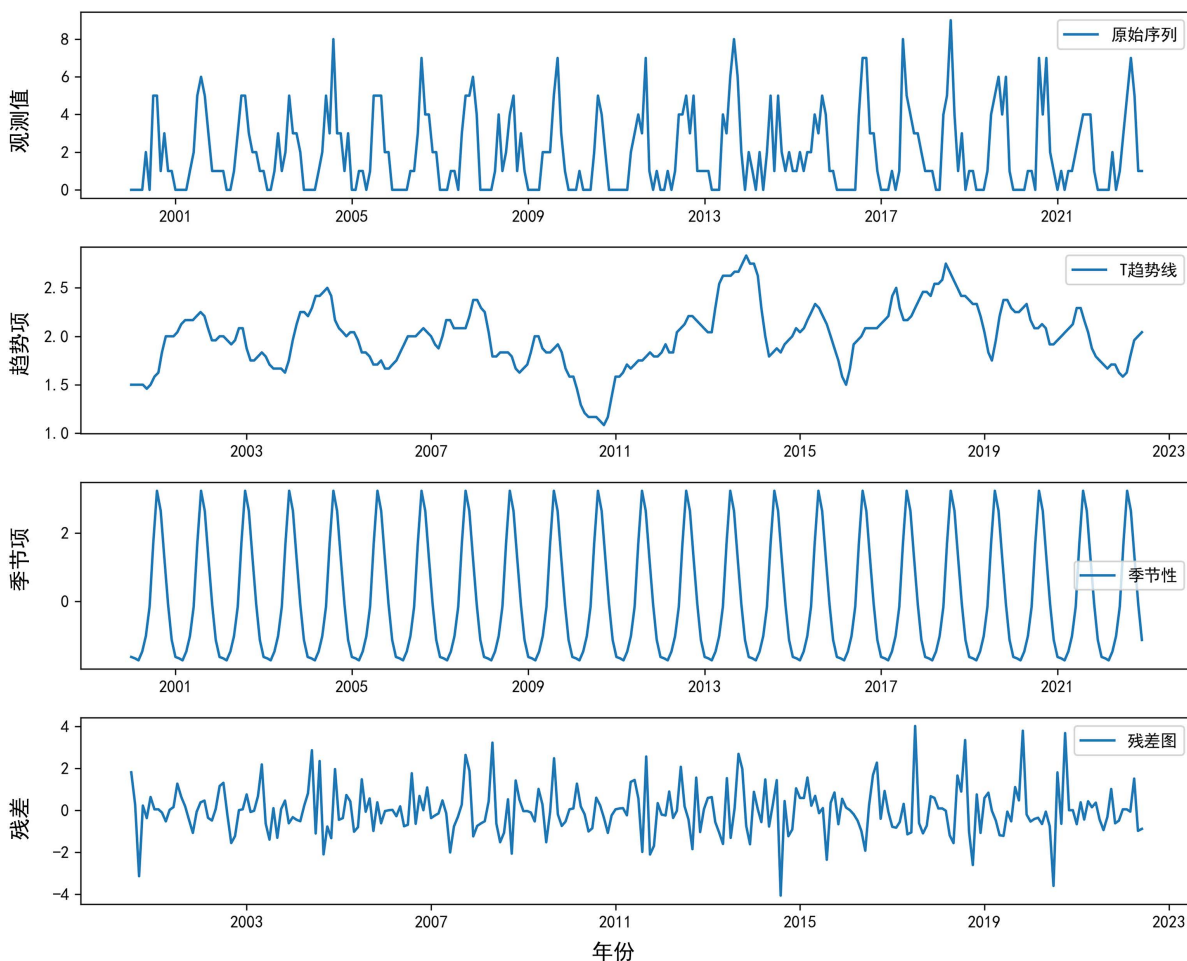


Figure 2. Time series decomposition of typhoon frequency from January 2000 to December 2022

图 2. 2000 年 1 月~2022 年 12 月台风频次的时间序列分解图

#### 3.3. 白噪声测试

并非所有的平稳时间序列都具有相关性, 这种纯随机的、没有记忆的时间序列被称为“白噪声”序

列。“白噪声”序列虽然是平稳序列,但其过去的行为对将来的发展没有任何影响,不具有分析价值和待挖掘的相关性[22]。因此需要在建模前进行白噪声测试,以确保该平稳时间序列具有相关性,存在继续分析的价值。

Ljung-Box 检验法可用于检验  $m$  阶滞后范围内时间序列的自相关性是否显著,来判断时间序列是否为白噪声序列。本文运用 Ljung-Box 检验法,借助 python 编程工具 statsmodels 库中的 `acorr_ljungbox` 函数对该平稳时间序列进行白噪声测试。原始平稳数据时间序列经检验,  $p$  值小于 0.001,说明该时间序列为非白噪声时间序列,即非随机序列,可以继续分析。

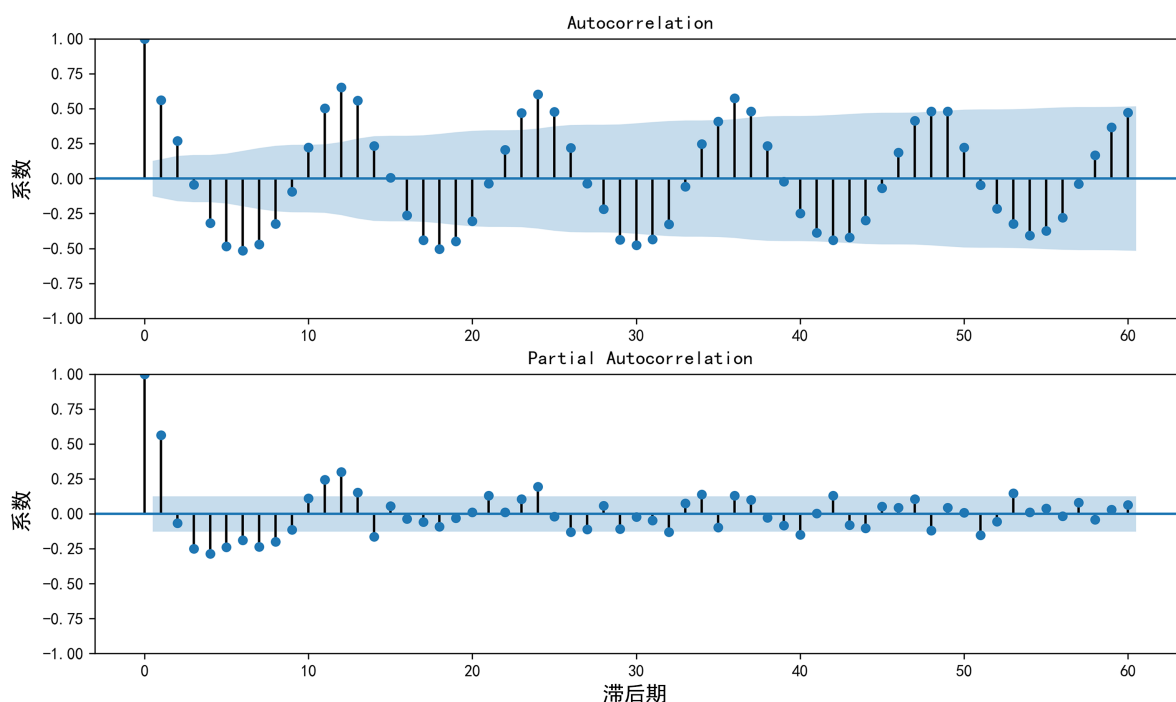
### 3.4. 模型定阶

SARIMA( $p,d,q$ )( $P,D,Q$ )[ $s$ ]总共 7 个参数,可以分成 2 类,3 个非季节参数( $p,d,q$ ),和 4 个季节参数( $P,D,Q,s$ )。

自回归阶数  $p$ 、趋势移动平均阶数  $q$ 、季节性自回归阶数  $P$ 、季节性移动平均阶数  $Q$  可以根据 ACF 图和 PACF 图来确定,也可以通过遍历 AIC 和 BIC 最小参数组合来确定[23]。本文通过绘制 ACF 图和 PACF 图观察以及借助 `pmdarima` 库中的 `auto.arima()` 函数确定模型的最佳参数为  $p=0, d=0, q=0, P=1, D=0, Q=1$ 。

原始数据的自相关图和偏自相关图,如图 3 所示。`auto.arima()` 函数可以通过遍历自动尝试不同的阶数组合并挑选出可能的最优模型,它选择最优模型的判断标准默认是 AIC,信息准则越小,说明参数的选择越好。由图 2 的季节性图可以看出,原序列有明显的周期性,而且是以 12 个月为一个周期,由此确定  $s=12$ 。

综上,确定模型为 SARIMA(0,0,0)(1,0,1)[12]。经过白噪声测试,模型的残差为随机序列,说明信息已经被提取。



**Figure 3.** Typhoon frequency autocorrelation and partial autocorrelation maps of raw data from January 2000 to December 2019  
**图 3.** 2000 年 1 月~2019 年 12 月原始数据台风频次自相关图和偏自相关图

### 3.5. 模型评估和残差检验

要对时间序列模型进行评估, 可以使用各种指标来衡量模型的性能和预测的准确性。本文通过计算均方根误差(RMSE)和平均绝对误差(MAE)对模型进行评估。

RMSE 是预测值与观测值之间差异的平方和的平均值的平方根, 它衡量了模型的平均预测误差大小, 数值越小越好。MAE 是预测值与观测值之间差异的绝对值的平均值它衡量了模型的平均预测误差的大小, 数值越小越好[24] [25]。本文计算 SARIMA(0,0,0)(1,0,1)[12]模型的均方根误差  $RMSE = 1.3932$ , 平均绝对误差  $MAE = 1.0389$ 。运用该模型预测测试集 2020 年 1 月~2022 年 12 月的台风频次, 使用 Matplotlib 库绘制了预测结果与观测值的对比图, 如图 4 所示, 可以直观地观察预测结果与实际观测值之间的差异。其中, 黄色实线为预测结果, 蓝色实线为观测结果。

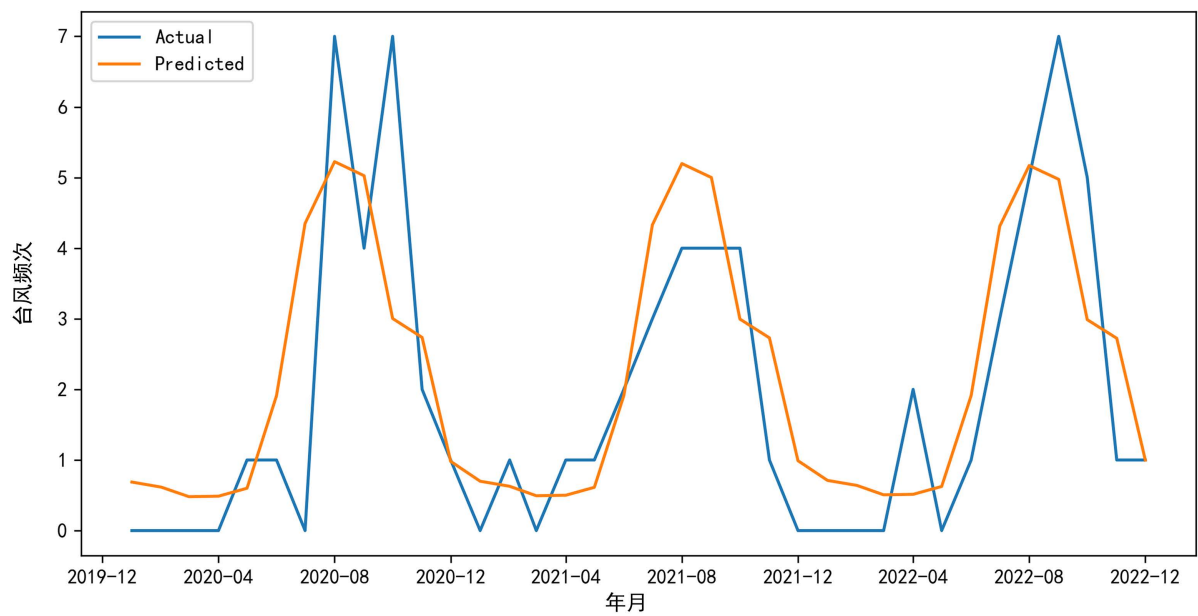
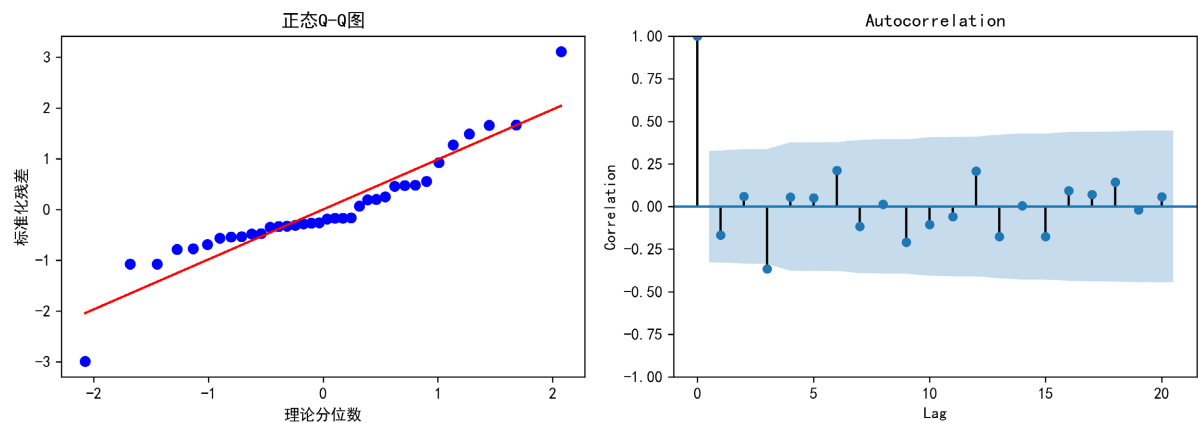
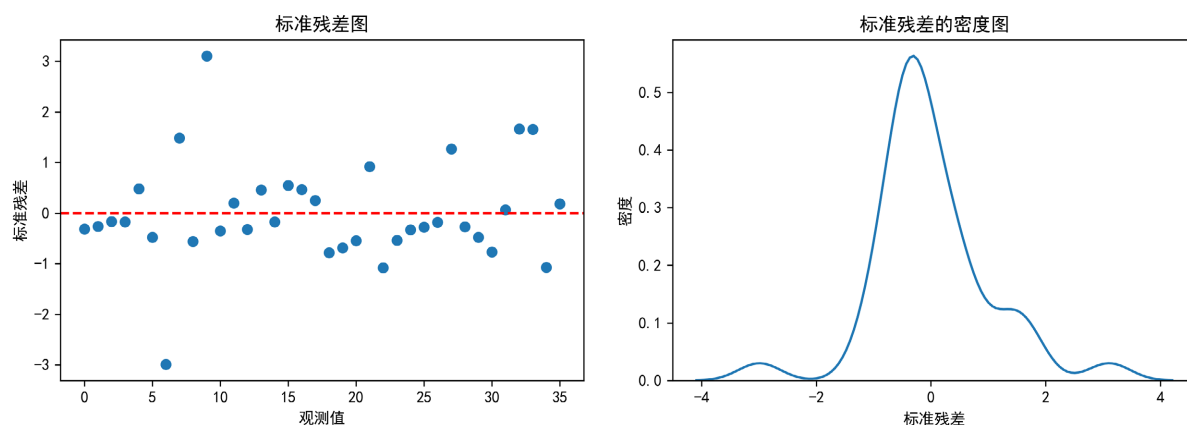


Figure 4. Comparison of predicted and observed typhoon frequency in 2024  
图 4. 2024 年台风发生频次预测值与观测值对比

最后使用 Matplotlib 库绘制 SARIMA(0,0,0)(1,0,1)[12]的残差分布和自相关图, 如图 5 所示。





**Figure 5.** Typhoon frequency SARIMA (0,0,0) × Residual distribution and autocorrelation diagram of the (1,0,1) [12] model  
**图 5.** 台风频次 SARIMA(0,0,0)×(1,0,1)[12]模型残差分布和自相关图

通过观察,可以发现整体数据均分布在标准红线两侧,偏差相对较小。自相关图(ACF 图)中第二个点之后的所有点都在蓝色置信区间内,说明模型残差的自相关为理想结果。由模型检验结果图可知,标准残差图中的点基本落在一条水平带附近,且模型残差密度图接近正态分布[26]。上述检验可以证明,残差接近正态分布且互相独立,证明建模的合理性。

### 3.6. 预测数据

根据拟合的 SARIMA(order=(0,0,0), seasonal\_order=(1,0,1,12))模型对 2024 年台风发生频次做为期 12 个月的预测,结果表 1 所示,并绘制折线图,如图 6 所示:

**Table 1.** Predict the frequency of typhoons in each month of 2024

**表 1.** 预测 2024 年各月的台风频次

未来月份	预测值
1	0.73
2	0.66
3	0.52
4	0.53
5	0.65
6	1.92
7	4.27
8	5.12
9	4.93
10	2.97
11	2.71
12	1.02



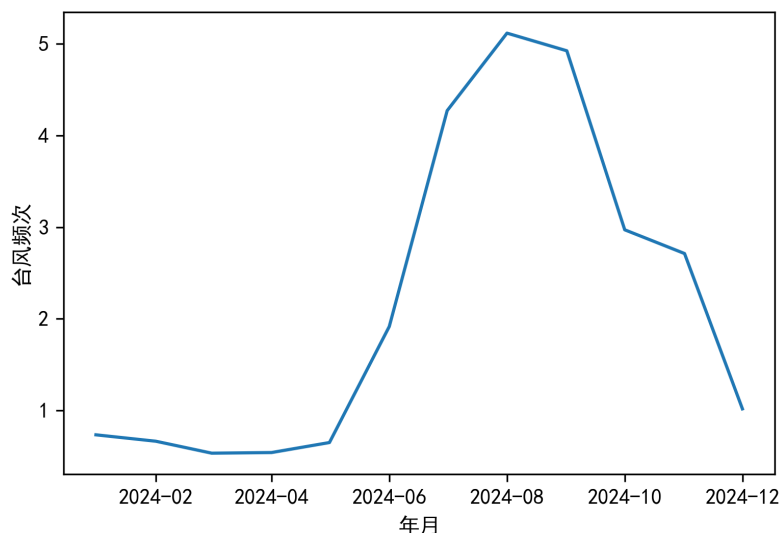


Figure 6. Monthly typhoon frequency prediction results for 2024  
图 6. 2024 年逐月台风频次预测结果

#### 4. 结论

本文基于时间序列分析中的 SARIMA 模型对 2024 年的逐月台风频次进行了预测。图 6 给出了模型预测的 2024 年各月的台风生成频次。通过时间序列平稳化, 数据白噪声测试, 模型定阶, 模型评估以及残差检验等步骤对 2000 年 1 月至 2019 年 12 月逐月的台风生成频次序列建立了最佳的 SARIMA 模型, 并用该模型对 2020 年 1 月至 2022 年 12 月逐月的台风生成频次进行了预测。结果表明: 预测结果与观测值较为吻合, 台风在 7 至 10 月生成频次较多, 较为活跃。

虽然 SARIMA 模型在对时间序列进行分析时, 只考虑了其自身历史时间序列的变化而不直接考虑其他因素对时间序列的影响, 但这并不等于可以完全忽视这些因素的影响, 因为时间序列本身正是在各种相关因素的相互作用下形成的[27]。对历史数据变动规律的概括, 也正是对各种复杂因素影响时间序列规律的概括。

从模型预测的台风生成频次的结果来看, SARIMA 模型对台风发生频次具有一定的预测能力, 但仍存在漏报、错报的情况, 其精确度需要在接下来的工作中进行进一步的提高, 以实现更好的防灾减灾作用。

#### 参考文献

- [1] 张利. 基于时间序列 ARIMA 模型的分析预测算法研究及系统实现[D]: [硕士学位论文]. 镇江: 江苏大学, 2008.
- [2] 陈秋怡, 崔妮, 汤景泰. 基于时间序列模型的传播预测方法研究[J]. 中国传媒大学学报(自然科学版), 2022, 29(3): 57-65.
- [3] 刘珊, 陈幸荣, 蔡怡. 基于 ARIMA 模型的台风频次预测[J]. 海洋预报, 2016, 33(1): 53-58.
- [4] 孙义, 周隄隄. 基于 Python 的金融时间序列 ARIMA 模型教学[J]. 现代信息技术, 2021, 5(10): 192-195.
- [5] Xu, S.J., Chan, H., et al. (2019) Forecasting the Demand of the Aviation Industry Using Hybrid Time Series SARIMA-SVR Approach. *Transportation Research Part E-Logistics and Transportation Review*, **122**, 169-180. <https://doi.org/10.1016/j.tre.2018.12.005>
- [6] 潘兴强, 马瑞, 杨天池, 等. 应用 Python 编程语言构建宁波市水痘发病率预测的季节性 ARIMA 模型[J]. 中国疫苗和免疫, 2022, 28(1): 83-87+104.
- [7] 肖良. 基于季节性 ARIMA 模型的居民消费水平预测[J]. 统计与决策, 2016(8): 83-86.

- [8] 安致远, 何恩球. 基于 SARIMA 时间序列模型的区域快递需求预测——以江苏省为例[J]. 物流科技, 2022, 45(20): 63-66+70.
- [9] 管亚平. 基于 SARIMA 模型的城市热岛季节性时序预测研究[J]. 科学技术创新, 2024(7): 111-114.
- [10] 李志超, 刘升. 基于 ARIMA 模型、灰色模型和回归模型的预测比较[J]. 统计与决策, 2019, 35(23): 38-41.
- [11] Ren, F.T., Zhang, C.L., *et al.* (2022) A Hybrid Method for Power Demand Prediction of Electric Vehicles Based on SARIMA and Deep Learning with Integration of Periodic Features. *Energy*, **250**, Article ID: 123738.
- [12] 邓方进, 王绪本, 李德伟. ARIMA 模型在 LMT 数据处理中的应用[J]. 物探化探计算技术, 2017, 39(5): 612-619.
- [13] 周浩, 李虹, 张岳琴, 等. SARIMA 模型在山西省手足口病发病预测中的应用[J]. 中国预防医学杂志, 2024, 24(2): 117-121.
- [14] 段然, 庞建华, 张良钧. 基于 SARIMA 模型的铁路站点客流量预测研究[J]. 数学的实践与认识, 2019, 49(9): 1-10.
- [15] 陈璇, 郑崇伟, 左常鹏, 等. 台风活跃季月活动频次指数的构建及其应用[J]. 厦门大学学报(自然科学版), 2020, 59(3): 394-400.
- [16] 汪路, 卢莹, 赵海坤. 台风灾害时空特征分析与评估模型构建[J]. 灾害学, 2023, 38(4): 187-194.
- [17] 郑浩然, 潘雨青, 李世伟, 等. 基于季节性 ARIMA 模型的小区供水预测[J]. 计算机应用与软件, 2018, 35(1): 118-122+294.
- [18] Moeeni, H.B., *et al.* (2017) Integrated SARIMA with Neuro-Fuzzy Systems and Neural Networks for Monthly Inflow Prediction. *Water Resources Management*, **31**, 2141-2156. <https://doi.org/10.1007/s11269-017-1632-7>
- [19] 王承杰, 孟彦菊. 基于季节性 ARIMA 模型的民航客运量预测[J]. 统计与管理, 2022(5): 88-93.
- [20] 李颖若, 韩婷婷, 汪君霞, 等. ARIMA 时间序列分析模型在臭氧浓度中长期预报中的应用[J]. 环境科学, 2021, 42(7): 3118-3126.
- [21] 陈丽敏, 刘成伟, 梁新民, 等. 基于 Python 语言的 ARIMA 模型在江西省食源性疾病发病率预测中的应用[J]. 中国食品卫生杂志, 2024, 35(3): 458-463.
- [22] Diop, M.-D., Kamdem, J.S., *et al.* (2024) Multiscale Agricultural Commodities Forecasting Using Wavelet-SARIMA Process. *Journal of Quantitative Economics*, **21**, 1-40. <https://doi.org/10.1007/s40953-022-00329-4>
- [23] 杨振昊, 张俊波, 杨晨星, 等. 基于 SARIMA 模型的我国水产品消费价格指数预测[J]. 海洋湖沼通报, 2021, 43(2): 131-138.
- [24] 陈刚, 郭晓梅. 基于时间序列模型的非平衡数据的过采样算法[J]. 信息与控制, 2021, 50(5): 522-530.
- [25] 翁渊瀚, 李南. 基于时间序列模型的文本数据压缩存储算法[J]. 吉林大学学报: 工学版, 2024, 53(7): 2109-2114.
- [26] 孙嘉琪, 王晓晔, 杨鹏, 等. 基于时间序列模型和情感分析的情感趋势预测[J]. 计算机工程与设计, 2021, 42(10): 2938-2945.
- [27] Shad, M.S., Singh, Y.D., *et al.* (2022) Forecasting of Monthly Relative Humidity in Delhi, India, Using SARIMA and ANN Models. *Modeling Earth Systems and Environment*, **8**, 4843-4851. <https://doi.org/10.1007/s40808-022-01385-8>