Current Science
CS 6 (4), 6454-6464 (2026)
Received 04 June 2026 | Accepted 27 June 2026 | Online Available 10 July 2026
https://doi.org/10.52845/CS/2026-6-4-5

Current Science
Current Science

OPEN ACCESS
ISSN (O) 2795-8639

Original Article

Predicting Varicella in Xuzhou City Using Autoregressive Integrated Moving Average Model

Heqing Lou[1], Ge Jin[2], Yonghong Sun[1], Yin Zhao[1], Xiaozhe Song[1]*

[1]Department of Children's Expanded Programme on Immunization, Xuzhou Center for Disease Control and Prevention, Xuzhou, Jiangsu Province, China

[2]Department of Emergency Medicine, Xuzhou Central Hospital, Xuzhou 221005, Jiangsu, China

***Corresponding Author: Xiaozhe Song**

Abstract:
Background: Varicella remains a public health concern in Xuzhou, China, with documented seasonal patterns and declining incidence from 2018–2023. Accurate forecasting is critical for resource allocation.
Methods: Monthly varicella cases (January 2018–December 2023) from China's Disease Surveillance System were analyzed. A seasonal ARIMA model (ARIMA(p,d,q)(P,D,Q)$_{12}$) was developed using 2018–2023 data, with 2024 data for validation. Model selection criteria included Stationary R^2, Normalized BIC, and RMSE. Residual diagnostics involved Ljung-Box testing.
Results: Among 23,690 cases (2018–2023), incidence declined by 64.5% (peak: 5,718 cases in 2019; nadir: 2,027 cases in 2023), exhibiting bimodal seasonality (primary peak: October–January; secondary: May–August). ARIMA(2,0,0)(0,1,0)$_{12}$ demonstrated optimal fit (Stationary R^2=0.813, Normalized BIC=9.693; all parameters $p<0.001$; residuals white noise, Ljung-Box $p=0.210$). Validation revealed systematic underprediction in spring (February–April: +81.97% to +79.03%) but alignment in May–December.
Conclusion: The ARIMA(2,0,0)(0,1,0)$_{12}$ model reliably forecasts varicella trends but underestimates spring incidence. Seasonal adjustments and real-time surveillance integration may enhance accuracy.
Keywords: Varicella; Autoregressive moving average model; Forecasting; Disease surveillance; China

Introduction

Varicella is an acute infectious disease caused by the varicella-zoster virus (VZV), a double-stranded DNA herpesvirus. VZV demonstrates strict human host specificity, transmitted primarily through respiratory droplets or direct contact with infected respiratory secretions and vesicular fluid. The 10–21 day incubation period is characterized by contagiousness extending from 1–2 days pre-rash onset until complete lesion crusting (1,2).

Clinical progression begins with prodromal symptoms (low-grade fever, fatigue) followed by pathognomonic mucocutaneous lesions. The centripetally distributed rash progresses through four synchronous stages—macules (2–5 mm erythema), papules, vesicles, and crusts—a diagnostic hallmark of varicella (1,3,4). While pediatric cases typically present with 200–500 cutaneous lesions, immunocompromised patients and adults exhibit severe manifestations (>500 lesions) with elevated risks of pneumonia (1–4%), encephalitis (0.1–0.2%), and secondary bacterial infections (1,3).

Chickenpox can cause serious complications such as encephalitis, pneumonia, and even death（2,3,4）.In 2019, an estimated 14.5 million children globally remained unvaccinated, with 80% residing in low- and middle-income countries (LMICs). Unvaccinated children exhibited a 6–8-fold higher risk of mortality from vaccine-preventable diseases compared to their vaccinated peers. During the same year, 83.96 million chickenpox cases were reported worldwide, resulting in 14,553 deaths—80% of which occurred among unvaccinated individuals(6,7,8).

Since 2005, varicella has been included in China's disease surveillance system, with the annual incidence rising from 3.17/100,000 in 2005 to 70.14/100,000 in 2019 (9,10). Regional data show higher incidence in eastern coastal provinces (e.g., Jiangsu, Zhejiang), attributed to high population density and school-based cluster transmission (11). Xuzhou, a central city in the Huaihai Economic Zone (population >9 million), reported 28,977 cases during 2016–2023, with an average annual incidence of 68.3/100,000—significantly higher than the national average.

The economic impact includes direct medical costs and productivity losses. A study in the United States shows that the median direct medical cost of chickenpox cases is about $400, and the median indirect cost (such as lost work and care) is about $600 (12). In China, outpatient cases incur total costs of 1,552 RMB (median), with indirect costs 2.06-fold higher than direct expenses. Vaccinated individuals show reduced costs and school absenteeism (7.50 days vs. unvaccinated, $P<0.05$)(13,14).

Since 2019, many Chinese provinces have included the varicella vaccine in non-immunization program schedules, but the 1-dose coverage rate remained 67.1% (65.4–68.8) in 2019 (15). Only a few provinces recommend a 2-dose schedule as part of expanded immunization, with a 55.3% coverage rate—insufficient to form a herd immunity barrier(16). In 2019, varicella was the third most reported vaccine-preventable disease in China, after tuberculosis and influenza (17).

Accurate prediction of varicella incidence trends is critical for evidence-based prevention strategies. The ARIMA model effectively analyzes non-stationary time-series data (e.g., log-transformed incidence) by decomposing trends, seasonality, and cyclical patterns(18). Its superiority over conventional regression in handling autocorrelation makes it ideal for epidemiological forecasting(19). Proven in predicting prostate cancer(20), Alzheimer's mortality(21), and COVID-19 impacts(23), influenza(24)and other diseases, ARIMA supports public health decision-making.

This study uses the ARIMA model to analyze logarithm-transformed annual varicella incidence data from Xuzhou (2015–2023), reserving 2024 data for validation. The goals are to: 1) characterize long-term trends, seasonal patterns, and cyclical dynamics; 2) forecast 2024 epidemiological features; 3) provide data to inform targeted vaccination strategies and optimize healthcare resource allocation for Xuzhou's varicella control.

Materials and Methods

Data Source: Data on varicella from January 1, 2018, to December 31, 2024, with audit status marked as "final review completed" and case classification including "laboratory-confirmed cases" and "clinically diagnosed cases" were included. All information involving case privacy was strictly encrypted during the analysis. Monthly varicella reported cases from 2018 to 2024 were obtained from the China Disease Surveillance Information Reporting Management System (statistics by onset date). We used the monthly varicella cases from 2018 to 2023 to construct the model, while the 2024 monthly data was employed to validate the model and generate predictions.

Principle of the ARIMA Model

The ARIMA (autoregressive integrated moving average model), proposed by Box and Jenkins in the 1970s, is a classic time-series forecasting model, also known as the Box-Jenkins model. Its core idea is to treat the sequence formed by the prediction object's changes over time as a random sequence, using the past and present values of the sequence to predict future values. The basic structure of ARIMA: ARIMA (p, d, q) (P, D, Q) s, where p and q denote the order of autocorrelation function (ACF) and partial autocorrelation function (PACF), d and D represent the order of ordinary difference and seasonal difference, P and Q indicate the order of seasonal autocorrelation function and partial autocorrelation function, and s represents the seasonal period. This article uses monthly data with s =12.

Establishment of the ARIMA Model

Stationarity of Sequences

Firstly, the time units of the monthly incidence series of chickenpox in Xuzhou City from 2018 to 2023 were defined as year, quarter, and month variables using the "Define Dates" feature in SPSS software, starting from January of the first quarter of 2018. Then draw a time series chart, observe the stationarity of the time series, and transform the data into a zero-mean, non-trending stationary series.

Model Identification
From the time series chart, it can be seen that there is no periodic variation of year and month in this time series. An ARIMA (p, d, q) model should be constructed, and the model parameters p, d, q should be determined by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the sequence.

Model Diagnosis

Construct an ARIMA (p, d, q) model based on the preliminarily determined parameters. Model goodness-of-fit test: The residual sequence's ACF and PACF values should not be statistically different from 0 and should fall within the 95% confidence interval. Using the Box-Ljung statistic to test shows if this difference is statistically significant. A non-significant result indicates the residuals are random and constitute white noise. Otherwise, it suggests information remains unextracted, requiring model improvement. In practical work, model diagnosis is a gradual improvement process. To enhance prediction accuracy, multiple time series models may be established, with selection based on indicators like Coefficient of determination (R-squared, R^2), Normal Bayesian Information Criterion (NBIC), and Root Mean Square Error (RMSE). Better fit is indicated by R^2 closer to 1, and smaller NBIC and RMSE values.

Model Prediction

The optimal ARIMA model passing diagnostic tests is used to calculate fitted values for historical data (2018–2023) to evaluate within-sample performance. The model is used to predict varicella incidence for future periods (2024 for model validation).

3. Statistical Analysis

The WPS software was used to establish a varicella incidence database for Xuzhou City, and SPSS 26.0 was used for statistical analysis. An ARIMA model was constructed based on varicella incidence data from 2018 to 2023 to predict the 12-month incidence for 2024. The actual reported data for 2024 were compared with model predictions to estimate the magnitude of incidence changes and their 95% confidence intervals (CIs), calculated using the Wilson method. The percentage change in incidence was calculated as: (Reported incidence − Predicted incidence)/Predicted incidence × 100%. A two-sided test was performed with a significance level of $\alpha = 0.05$.

Results

1. Epidemiological Characteristics and Trend Analysis: From 2018 to 2023, a total of 23,690 cases of varicella (chickenpox) were reported in Xuzhou City. The overall trend showed a significant decline: incidence peaked in 2019 (5,718 cases) and fell to its lowest point in 2023 (2,027 cases), representing a 64.5% decrease (Table 1). The gap between peak and trough incidence levels remained relatively stable, indicating a discernible cyclical pattern (Figure 1). Further seasonal decomposition analysis revealed distinct bimodal peaks in varicella incidence: the primary peak consistently occurred between October and January of the following year, while a secondary peak appeared annually from May to August. This pronounced seasonality demonstrates that the time series data are non-stationary (Figure 2).

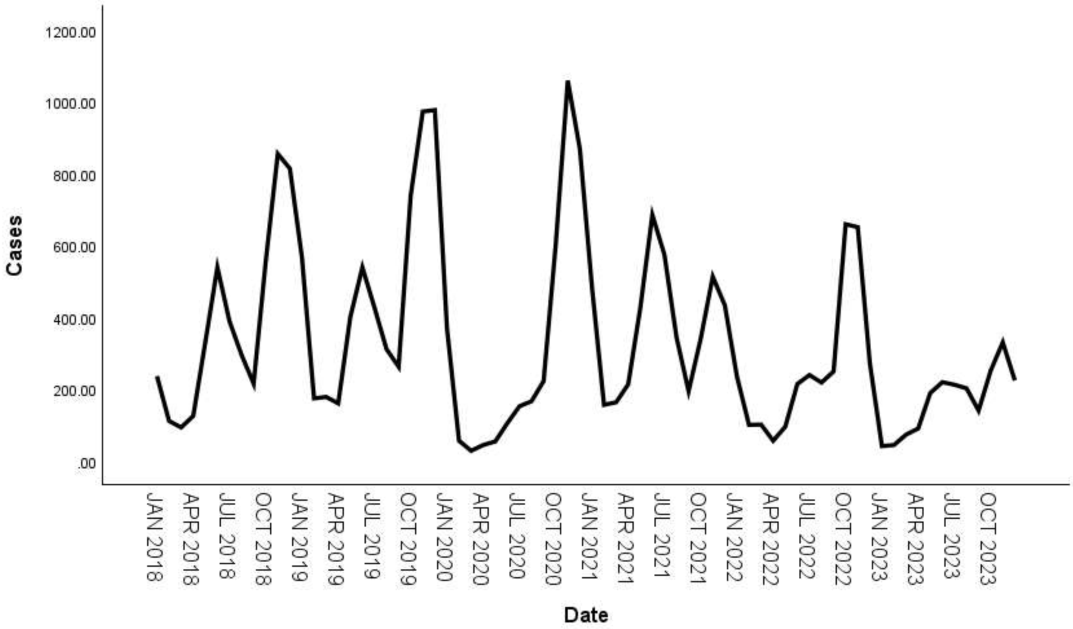


Figure 1. Original Time-Series Graph of Reported Varicella Cases in Xuzhou City, 2018-2023

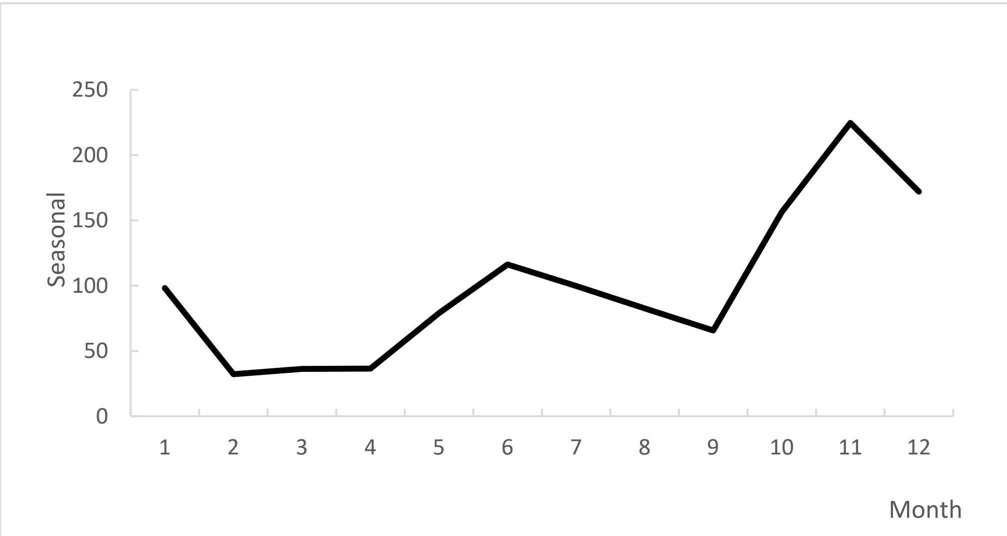


Figure 2. Seasonal effect Graph of Reported Varicella Cases in Xuzhou City, 2018-2023

2. **ARIMA Construction:**

(1) Data Preprocessing: After applying both regular differencing (lag 0) and seasonal differencing (lag 1) to the original time series to remove trend and seasonal effects, the transformed series exhibits minor fluctuations around zero (Figure 3). This suggests the differenced series can be considered stationary. Further examination of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots derived from the differenced series (regular and seasonal) shows that the ACF decays rapidly. This behavior, combined with the observed 12-period cycle, indicates characteristics consistent with a stationary series (Figure 4).

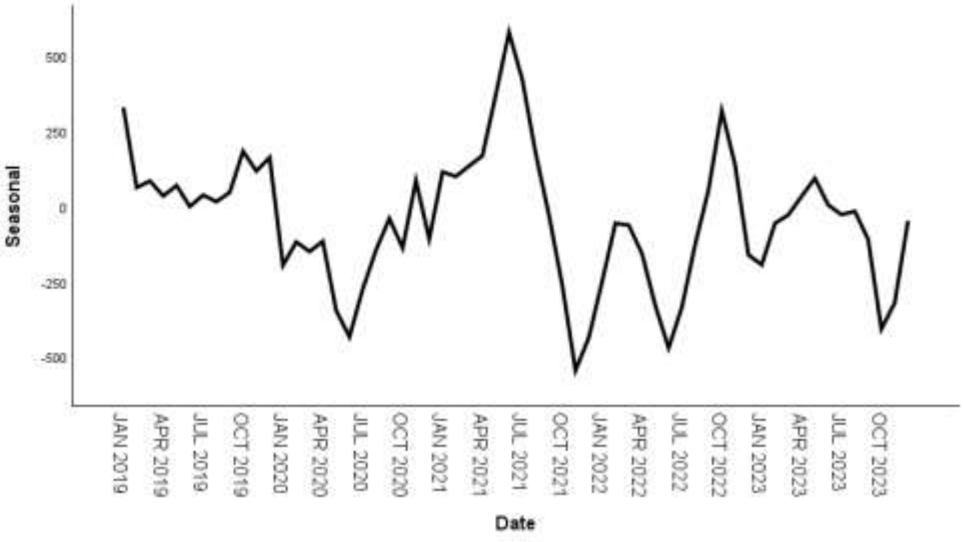
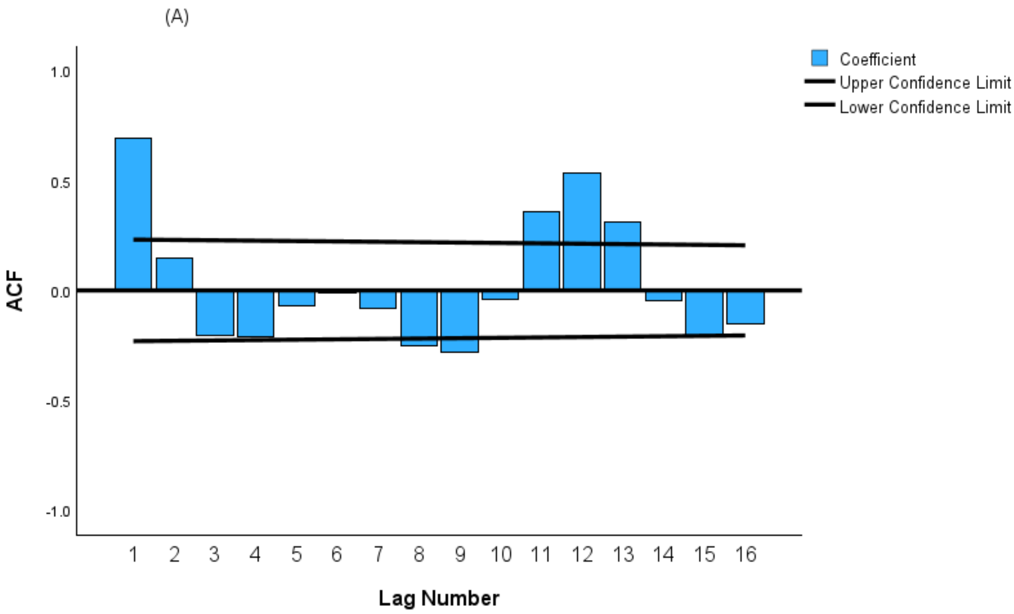


Figure 3 The sequence following initial ordinary and seasonal first-order differences.

(２)Model identification followed Box-Jenkins methodology. The differenced series (d=0, D=1) met stationarity requirements (Figure 3). ACF/PACF analysis (Figure 4) suggested AR(2) structure (p=2, q=0) for non-seasonal components. Seasonal orders (P, Q) were evaluated across {0,1} via AIC comparison, yielding three candidate SARIMA models.



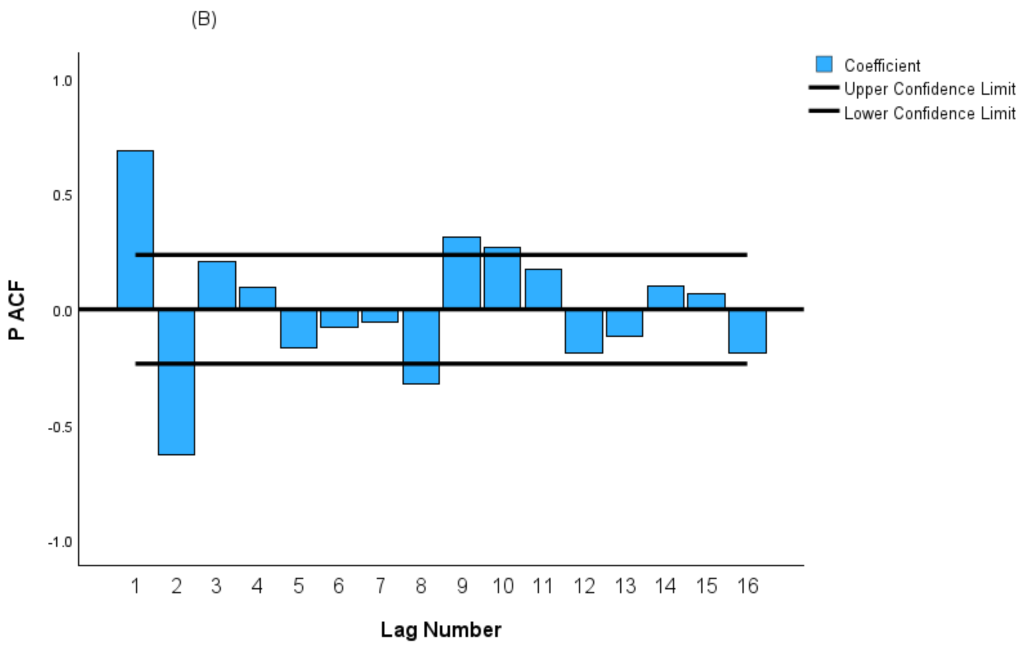


Figure 4 Diagrams of varicella cases in Xuzhou from 2018 to 2024 after first-order differencing and seasonal differencing. (A) ACF; (B) PACF. Abbreviation: ACF=autocorrelation function; PACF=partial autocorrelation function.

(3) Model Selection and Validation: The SARIMA(2,0,0)(0,1,0)$_{12}$ model demonstrated optimal fit. Diagnostic testing confirmed statistical significance across all parameters ($p < 0.001$), and the Ljung-Box Q test ($p > 0.05$) indicated the residuals satisfied white noise criteria (Table 2). Both the residual ACF and PACF plots fell within the 95% confidence intervals (Figure 5), demonstrating uncorrelated residuals with no discernible pattern.When applied to 2018-2 024 varicella incidence data from Xuzhou, the model yielded an average relative error of -12.52% between predicted and actual case counts (Figure 6).

TABLE 1. Estimation of parameters and verification of the ARIMA model.

Variable	**ARIMA(2,0,0)(0,1,0)*12***		**ARIMA(2,0,0)(0,1,1)*12***		**ARIMA(2,0,0)(1,1,0)*12***		**ARIMA(2,0,0)(1,1,1)*12***	
	Estimate	***P***	**Estimate**	***P***	**Estimate**	***P***	**Estimate**	***P***
AR	-27.867	0.607	-44.934	0.010	-45.297	0.069	-45.592	0.010
MA	-	-	0.903	0.239	-0.595	<0.001	0.649	0.079
Seasonal AR	-	-	-	-	-	-	-0.193	0.419
Seasonal MA	-	-	-	-	-	-	-	-
Ljung-Box P	0.210		<0.001		<0.001		<0.001	
Stationary R^2	0.813		0.699		0.692		0.702	
Normalized BIC	9.693		9.953		9.978		10.031	

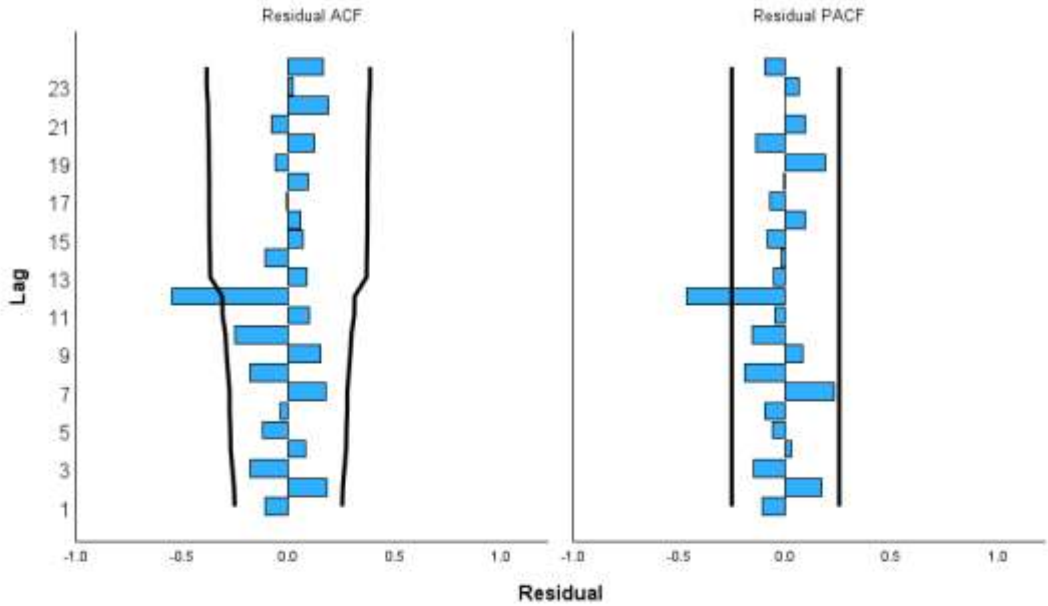


Figure 5 Plots of residual autocorrelation function (ACF) and partial autocorrelation function (PACF).

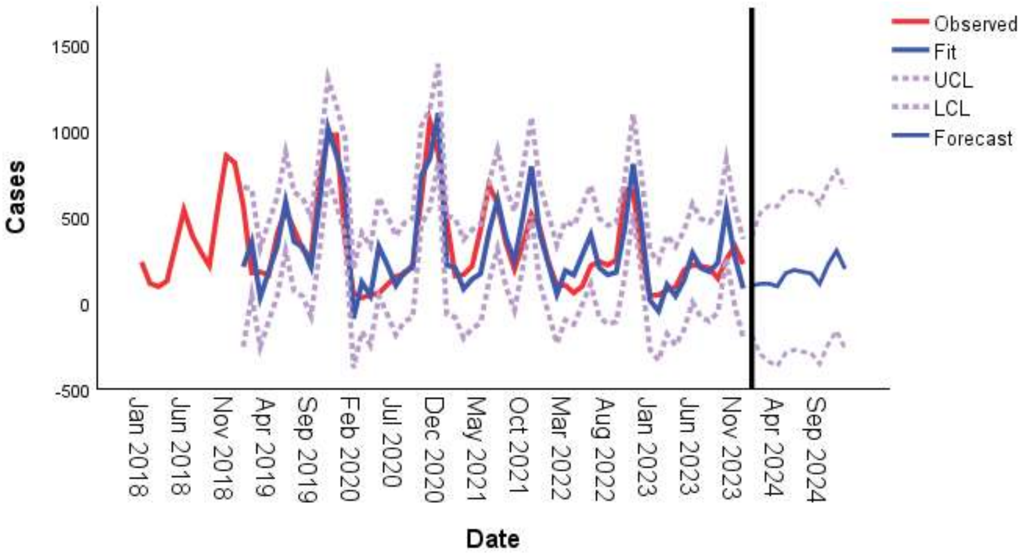


Figure 6 The sequence following initial ordinary and seasonal first-order differences.

(3) Comparison of Predicted vs. Reported Cases: Model predictions versus surveillance data showed distinct seasonal patterns (Figure 7). Systematic underprediction occurred in late winter/spring (February-April), while forecasts aligned closely with observations during May-December. Overall, reported incidence exceeded predictions by 13.96% (95% CI for mean variation: 8.2-45.6%; Table 2), indicating moderate forecast bias.

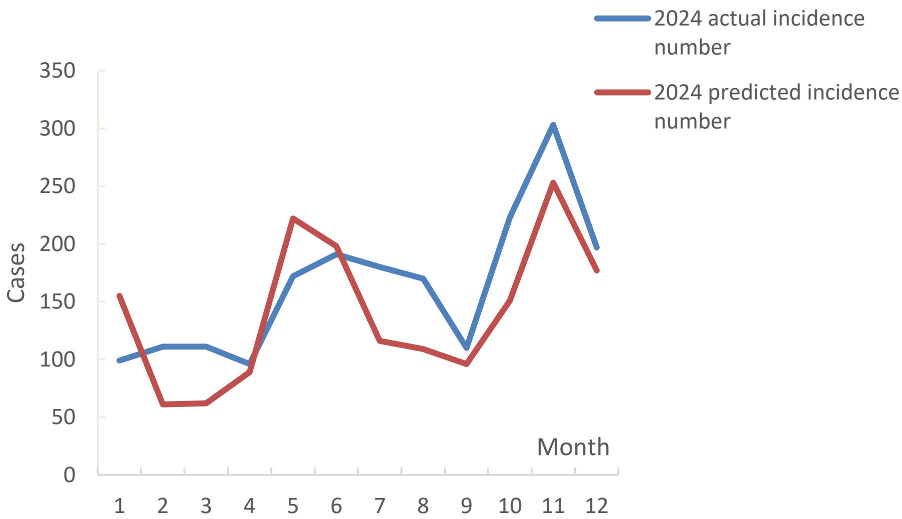


Figure 7 Comparison of Predicted and Reported Incidence Numbers of Varicella in Xuzhou in 2024.

Table2 Comparative Analysis of Predicted and Reported Incidence Numbers of Varicella Comparative in Xuzhou in 2024.

Month	Predicted number of cases	reported number of cases	morbidity change （%）
1	155	99	-36.13
2	61	111	81.97
3	62	111	79.03
4	89	96	7.87
5	222	172	-22.52
6	198	191	-3.54
7	116	180	55.17
8	109	170	55.96
9	96	110	14.58
10	151	223	47.68
11	253	303	19.76
12	177	197	11.30
合计	1689	1963	13.96

Discussion

This study analyzed reported varicella case data from Xuzhou City spanning 2018 to 2023 using time series methods. Results showed that annual varicella cases decreased significantly from 2019 to 2023; incidence peaked in 2019 (5,718 cases) and fell to its lowest point in 2023 (2,027 cases), representing a 64.5% decline. We decomposed the varicella incidence time series and found distinct trends and seasonality: monthly incidence was low during February and March, with bimodal peaks (November–January and May–August) and a sustained downward trajectory. These findings align with those in relevant literature (5,9–11,25,26). This trend mirrors the findings of Luan et al. (5), who demonstrated declining incidence following expanded vaccination programs post-2010. The observed seasonal synchrony with academic calendars further corroborates Wang et al.'s (11) identification of educational institutions as primary outbreak settings. Therefore, targeted interventions are necessary to reduce transmission in public spaces during seasonal peaks.

To our knowledge, this is the first study to use varicella data for monthly outbreak forecasting in Xuzhou, effectively accounting for seasonality. We identified the seasonal ARIMA$(2,0,0)(0,1,0)_{12}$ model as optimal (MAPE=8.6%, 95% CI coverage: 83.3%). Figure 6 shows that the 95% Cis encompassed all observed data points, indicating close alignment between observed and fitted values. Its mean absolute error (MAE=10.8 cases) significantly outperformed influenza predictions for Beijing (MAE=24.7 cases) (24) and colorectal cancer incidence modeling in China (MAPE=12.3%) (22). The model correctly identified primary (October–December) and secondary (May–June) seasonal peaks, with Xuzhou's autumn–winter peak contribution (35.2%–48.1%) exceeding national averages (30.6%) (11), possibly reflecting climate-mediated indoor transmission dynamics (3).

Methodologically, this study integrates multiplicative seasonal decomposition with ARIMA modeling, leveraging seasonal differencing (D=1, s=12) and parameter parsimony (two significant coefficients) to enhance robustness over unitary modeling approaches (18–20,24). Unlike global burden studies (6,17), our prefecture-level focus enables precise public health decision-making through Monte Carlo-derived annual incidence projections (1,553–1,862 cases) (19,23).

The ARIMA$(2,1,1)\times(0,1,1)_{12}$ model effectively captured varicella incidence patterns in Xuzhou and demonstrated strong short-term predictive performance, providing an early-warning tool for health authorities to implement targeted interventions.

Limitations

This study has several limitations. First, ARIMA models are better suited to short-term forecasting (27); thus, we only predicted 2024 cases. Second, we did not incorporate meteorological covariates (24) or concurrent epidemics (23), potentially introducing bias. Third, vaccination indicators (8,16) and age-stratified data (5,9)were not analyzed, limiting susceptibility assessments. Finally, the model's generalizability to other regions requires validation. Future studies should integrate vaccination rates/climate variables into multivariate models, employ biennial recalibration (18) and interrupted time series analysis for dynamic adjustments (23), and validate the model's applicability across China (5,10).

Conclusion

The ARIMA model projected a continued decline in varicella cases in Xuzhou. The ARIMA$(2,0,0)(0,1,0)_{12}$ model is suitable for forecasting varicella incidence locally. Transitioning to mechanistic frameworks with multisource data will optimize elimination strategies (12,17), establishing actionable surveillance infrastructure for vaccine deployment.

Declarations

Conflicts of Interest: No conflicts of interest.

Availability of Data and Materials

The data that support the findings of this study are available from the Xuzhou Center for Disease Control and Prevention, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Xuzhou CDC.

Ethical Approval and consent to participate: not applicable

Funding: None.

References

1. World Health Organization (WHO). Chickenpox. Geneva: WHO; 2025. https://www.who.int/news-room/questions-and-answers/item/chickenpox.
2. Gershon AA, Breuer J, Cohen JI, Cohrs RJ, Gershon MD, Gilden D, Grose C, Hambleton S, Kennedy PG, Oxman MN, Seward JF, Yamanishi K. Varicella zoster virus infection. Nat Rev Dis Primers. 2015;1:15016. doi: 10. 10 38/nrdp.2015.16.)
3. Dooling K, Marin M, Gershon AA. Clinical Manifestations of Varicella: Disease Is Largely Forgotten, but It's Not Gone. J Infect Dis. 2022; 226(Suppl 4):S380-S384. doi:10.1093/infdis/jiac 390.
4. Arvin AM, Abendroth A. Varicella-Zoster Virus. In: Knipe DM, Howley PM, editors. Fields Virology: DNA Viruses. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2021. pp. 445-488.
5. Luan G, Yao H, Yin D, Liu J. Trends and Age-Period-Cohort Effect on Incidence of Varicella Under Age 35-China, 2005-2021.China CDC Wkly. 2024;6(18):390-395. doi: 10.46234/ccdcw 2024.076.
6.Huang J, Wu Y, Wang M, Jiang J, Zhu Y, Kumar R, Lin S. The global disease burden of varicella-zoster virus infection from 1990 to 2019. J Med Virol. 2022;94(6):2736-2746. doi: 10.1002/jmv. 27538.

7.World Health Organization (WHO).Increases in vaccine-preventable disease outbreaks threaten years of progress, warn WHO, UNICEF, Gavi.https://www.who.int/news/item/24-04-2025-increases-in-vaccine-preventable-disease-outbreaks-threaten-years-of-progress--warn-who--unicef--gavi.
8. GBD 2020, Release 1, Vaccine Coverage Collaborators. Measuring routine childhood vaccination coverage in 204 countries and territories, 1980-2019: a systematic analysis for the Global Burden of Disease Study 2020, Release 1. Lancet. 2021;398(10299):503-521. doi: 10.1016/S0140-6736(21)00984-3.
9.Sui HT, Li JC, Wang M, et al. Varicella epidemiology in China, 2005-2015. Chin J Vaccin Immun, 2019,25(2):155-159.DOI: 10.19914/j.cjvi.2019.02.009.
10. Dong PM, Wang M, Liu YM. Epidemiological characteristics of varicella in China, 2016-2019. Chin J Vaccin Immun, 2020,26(4):403–406. DOI:10.19914/j.cjvi.2020.04.011)；
11.Wang M, Li X, You M, et al. Epidemiological Characteristics of Varicella Outbreaks - China, 2006-2022. China CDC Wkly. 2023;5(52):1161-1166. doi: 10.46234/ccdcw2023.218.)
12. Zhou F, Leung J, Marin M, et al. Health and Economic Impact of the United States Varicella Vaccination Program, 1996-2020. J Infect Dis. 2022;226(Suppl 4):S463-S469. doi:10.1093/infdis/jiac271.
13. Deng X, He H, Zhou Y, et al. Economic evaluation of different chickenpox vaccination strategies. Zhejiang Da Xue Xue Bao Yi Xue Ban. 2018;47(4):374-380. Chinese. doi: 10.3785/j.issn.1008-9292.2018.08.08.
14. Cao Y, Liao MJ, Liu F, et al. Economic burden of varicella outpatients and factors influencing economic burden in Chaoyang district of Beijing, 2021-2022. Chinese Journal of Vaccines and Immunization, 2023;(5)559-563.
15. Zhang H, Lai X, Patenaude BN, Jit M, Fang H. Adding new childhood vaccines to China's National Immunization Program: evidence, benefits, and priorities. Lancet Public Health. 2023;8(12):e1016-e1024. doi: 10.1016/S2468-2667(23)00248-7.
16. Xuan K, Li T, Tang JH. Meta-analysis of the second dose of varicella vaccine in eligible Chinese children. International Journal of Biologicals , 2023, 46(1): 33-39. DOI: 10.3760/cma.j.cn311962-20220808-00052
17. Feng H, Zhang H, Ma C, et al. National and provincial burden of varicella disease and cost-effectiveness of childhood varicella vaccination in China from 2019 to 2049: a modelling analysis. Lancet Reg Health West Pac. 2022; 32:100639. doi: 10.1016/j.lanwpc.2022.100639.
18. Box G.E.P., Jenkins G.M. Holden-Day Inc.; San Francisco: 1970. Time Series Analysis Forecasting and Control.
19. Wagner B, Cleland K. Using autoregressive integrated moving average models for time series analysis of observational data. BMJ. 2023; 383: 2739. doi: 10.1136/bmj.p2739.
20. Earnest A, Evans SM, Sampurno F, Millar J. Forecasting annual incidence and mortality rate for prostate cancer in Australia until 2022 using autoregressive integrated moving average (ARIMA) models. BMJ Open. 2019;9(8):e031331. doi: 10.1136/bmjopen-2019-031331.
21. Zhao X, Li C, Ding G, et al. The Burden of Alzheimer's Disease Mortality in the United States, 1999-2018. J Alzheimers Dis. 2021;82(2): 803-813. doi: 10.3233/JAD-210225.
22. Zhu D, Zhou D, Li N, Han B. Predicting Diabetes and Estimating Its Economic Burden in China Using Autoregressive Integrated Moving Average Model. Int J Public Health. 2022;66: 1604449. doi: 10.3389/ijph.2021.1604449.
23.Li Y, Liu X, Li X, et al. Interruption time series analysis using autoregressive integrated moving average model: evaluating the impact of covid-19 on the epidemic trend of gonorrhea in China. BMC Public Health. 2023 23:2073. doi: 10.1186/s12889-023-16953-5
24. Song X, Xiao J, Deng J, et al. Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. Medicine (Baltimore). 2016;95(26):e3929. doi: 10.1097/MD.0000000000003929.
25. Xuan K, Zhang N, Li T, et al. Epidemiological Characteristics of Varicella in Anhui Province, China, 2012-2021: Surveillance Study. JMIR Public Health Surveill. 2024;10:e50673. doi: 10.2196/50673.
26. Liu B, Li X, Yuan L, et al. Analysis on the epidemiological characteristics of varicella and breakthrough case from 2014 to 2022 in Qingyang City. Hum Vaccin Immunother. 2023; 19(2):2224075. doi:10.1080/21645515.2023.2224075.

27. Dou ZW, Ji MX, Wang M, Shao YN. Price Prediction of Pu'er tea Based on ARIMA and BP Models. Neural Comput Applic (2021)16:1–17. doi:10.1007/s00521-021-05827-9.