

Article

Not peer-reviewed version

Pertussis Dynamics in Tshwane, South Africa: The Role of Seasonality and Forecasting Techniques

Xolelwa Ntsham and [Tladi Daniel Ledibane](#) *

Posted Date: 19 March 2025

doi: 10.20944/preprints202503.1444.v1

Keywords: pertussis; time series analysis; seasonality; forecasting; ARMA model; Tshwane Health District



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Pertussis Dynamics in Tshwane, South Africa: The Role of Seasonality and Forecasting Techniques

Xolelwa Ntsham and Tladi D Ledibane *

Sefako Makgatho Health Sciences University; tladi.ledibane@smu.ac.za

* Correspondence: tladi.ledibane@smu.ac.za

Abstract: Background: B. Pertussis remains a significant public health concern, with periodic outbreaks despite vaccination efforts. Understanding the temporal trends and seasonality of pertussis incidence is important for improving its surveillance and prevention strategies. Objectives: The aim of this study was to examine pertussis trends in Tshwane Health District, South Africa, from 2015 to 2019. Methods: A retrospective time series analysis was performed on reported pertussis cases in the Tshwane Health District from 2015 to 2019. Descriptive statistical techniques and time series decomposition were employed to investigate seasonal patterns. Autoregressive Moving Average (ARMA) models were employed to assess short-term trends, while cubic trend modelling was used to forecast future case trajectories. Residual diagnostics assessed model validity and analysed correlations between pertussis incidence and climate variables. Results: The incidence of Pertussis infection exhibited seasonal patterns, with peak cases being reported in Spring and Winter. Time series decomposition confirmed annual fluctuations, suggesting potential environmental and social factors influencing disease transmission. The ARMA(4) model provided the best fit for short-term forecasting, while the cubic model effectively captured long-term trends. Residual diagnostics confirmed model reliability. Conclusion: This study highlights the utility of time series modelling in predicting pertussis trends and underscores the role of seasonality and climate factors in disease dynamics. The findings support enhanced surveillance, climate-adaptive interventions, and optimised vaccine timing to mitigate outbreaks. Integrating predictive modelling with public health strategies can improve outbreak preparedness and resource allocation.

Keywords: pertussis; time series analysis; seasonality; forecasting; ARMA model; Tshwane Health District

1. Introduction

B. Pertussis continues to be a significant public health concern globally [1,2] despite the availability of effective vaccines. Periodic B. Pertussis outbreaks have been documented in South Africa, with seasonal variations impacting the case count. Time series data analysis is essential for describing and understanding disease trends, predicting future epidemics, and implementing effective public health interventions.

Time series analysis serves as a valuable tool in disease surveillance and epidemiology, offering multiple methods for assessing and forecasting the incidence of illnesses. Support vector machines [3], decomposition methods, and ARIMA modelling are frequently used in epidemiology. These techniques assess interventions, detect outbreaks, and describe disease dynamics [4]. Furthermore, spectral analysis and the maximum entropy method have been applied to examine statistical trends of infectious diseases across multiple countries [5]

In the analysis of discrete illness counts, integer-valued autoregressive models demonstrate advantages compared to traditional Box-Jenkins models [6]. Time-domain and frequency-domain techniques, such as autocorrelation functions and spectrum analysis, are widely used; wavelet analysis represents a hybrid technique [7]. When selecting techniques, it is important to consider

factors such as scale, scope, disease characteristics, and surveillance objectives in conjunction with the algorithm's performance [8].

Recent studies have detailed the various mechanisms underlying the resurgence and spread of B. Pertussis. B. Pertussis infections are significantly affected by weather variability across all age categories, particularly temperature and humidity [9]. Population size and density influence persistence, and spatial-temporal studies have demonstrated the evolving patterns of disease peaks and regional expansion [10]. Socio-environmental factors such as temperature, socioeconomic level, and school calendars influence incidence [11]. Additionally, seasonal forcing and immune boosting establish cyclical behaviour and chaotic dynamics [12].

Vaccination has resulted in more extended periodic dynamics. For example, high vaccination levels against B. Pertussis can suppress its annual cycle and lengthen long-term periodicity in pertussis. Moreover, prior research has shown that vaccination coverage, vaccine effectiveness, pathogen evolution, and demographic changes significantly influence the transmission dynamics of pertussis, with some studies reporting prolonged epidemic cycles [15]. Comprehensive surveillance is necessary to enhance the understanding of B. Pertussis dynamics and anticipate future outbreaks.

Effective disease forecasting for communicable diseases has been made possible using models such as ARIMA and SARIMA. These models deliver early warnings for diseases such as dengue, tuberculosis, malaria, and COVID-19 while also offering reliable epidemic forecasts [16,17]. For example, SARIMA models have demonstrated effectiveness in identifying seasonal patterns of diseases, including hand, foot, and mouth disease [18]. Furthermore, SARIMA models exhibit similarities to dynamic linear models when long-term longitudinal data is utilized [19].

In the context of diseases characterised by long-range dependencies, such as hemorrhagic fever, advanced models like SARFIMA have demonstrated superior predictive capabilities [20]. Time series analytic approaches serve as valuable tools for public health surveillance, enabling authorities to implement preventative measures and mitigate the impact of disease outbreaks [21].

Recent research has examined various predictive models for B. Pertussis incidence. Contemporary trends in short-term B. Pertussis data indicate the feasibility of employing time series analysis techniques, specifically ARIMA and exponential smoothing models [10,22] (Zeng et al., 2016; Wang et al., 2018). Hybrid methodologies that integrate wavelet transformations or ARIMA with neural networks have demonstrated enhanced accuracy [22].

Incorporating additional data sources, including Google Trends, improves conventional modelling techniques [23]. Numerous studies, including those conducted by Raycheva et al. (2020) and Zhang et al. (2019), have identified increased incidence and seasonal trends [10,24]. In addition, advanced models integrating age-structured and particle-filtering methodologies have been created to elucidate diminishing immunity and assess therapeutic interventions [25,26]. This allows advanced forecasting systems to evaluate potential strategies for reducing the burden of B. pertussis and providing early warnings, facilitating informed public health decision-making.

Despite extensive research on B. Pertussis epidemiology and forecasting methods, there remain gaps in integrating advanced time series models with socio-environmental variables. Previous studies have primarily utilised conventional statistical models like ARIMA and SARIMA, which, while effective, have limitations in accurately representing the complex nonlinear relationships characteristic of disease transmission. Moreover, few studies have explored B. Pertussis trend analysis in South Africa, highlighting the need for localised forecasting methods. This study seeks to address these deficiencies by employing various forecasting methodologies, including hybrid models that integrate machine learning techniques. This study aims to (1) analyse historical B. Pertussis trends and identify seasonal patterns, (2) compare the predictive accuracy of various time series models, and (3) evaluate the influence of climate on B. Pertussis incidence in Tshwane.

2. Materials and Methods

2.1. Study Design

This investigation employed a retrospective time series analysis to examine B. Pertussis incidence patterns and evaluate the predictive capabilities of various forecasting models. The methodological approach integrated traditional statistical techniques with machine learning algorithms to enhance disease prediction accuracy.

2.2. Data Collection

B. Pertussis surveillance data were obtained from public and private health facilities in the Tshwane Health District, South Africa.

2.3. Inclusion and Exclusion Criteria

The study included all laboratory-confirmed and clinically diagnosed B. Pertussis cases reported in Tshwane between 2015 and 2019. Records were excluded if they were duplicates or contained unverified clinical diagnoses.

2.4. Data Processing and Analysis

All datasets underwent standardisation into a uniform monthly aggregation format. Missing values were addressed through linear interpolation and multiple imputation techniques. To mitigate reporting errors, extreme values were identified using z-scores and boxplot analyses.

2.5. Time Series Modeling

The time series data were analysed through Seasonal-Trend Decomposition with LOESS (STL), which enabled the decomposition into trend, seasonal, and residual components. The Augmented Dickey-Fuller (ADF) test was employed to evaluate stationarity. Four forecasting approaches were considered: Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Holt-Winters Exponential Smoothing, and a Hybrid Machine Learning Model combining LSTM and ARIMA. Model selection was determined based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were utilised to assess predictive accuracy. The statistical and computational analyses were performed using several software tools, including Python (with libraries such as pandas, statsmodels, and TensorFlow for LSTM models), Stata for statistical validation, and R for visualisation and time series decomposition.

To evaluate the models' performance, a rolling-window cross-validation method was used. Additionally, residual analyses were carried out to assess the model's goodness of fit and the reliability of their forecasts.

3. Results

The incidence of Pertussis infection exhibited seasonal patterns, with peak cases being reported in Spring and Winter. Time series decomposition confirmed annual fluctuations

3.1. Descriptive Statistics

Table 1 presents the descriptive statistics of reported B. Pertussis cases by season. Spring exhibited the highest variability, peaking at 54 cases in 2018.

Table 1. Descriptive Statistics of B. Pertussis Cases (2015-2019).

Season	Mean Cases	Std. Dev.	Min Cases	Max Cases
Autumn	11.6	6.5	4	20
Winter	16.6	12.0	2	30
Spring	19.8	21.5	1	54
Summer	6.0	7.4	1	19

3.2. Time Series Trend Analysis

Figure 1 illustrates seasonal and long-term trends in B. Pertussis cases through a line graph. It shows seasonal peaks in Spring and Winter, with moving averages (MA(4) and CMA(4)) highlighting a decline from 2015 to 2017. However, this decline was followed by a significant increase in cases from 2018 to 2019.

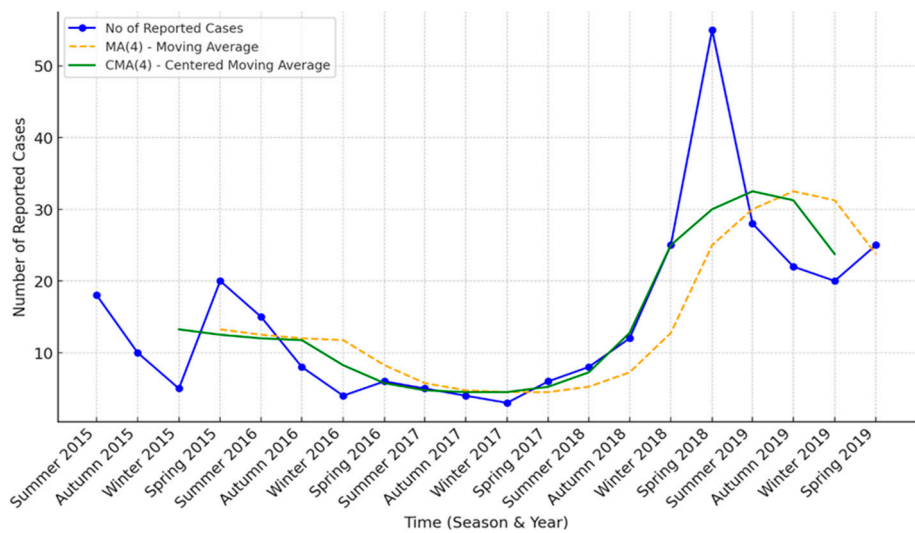


Figure 1. Seasonal Trend of Reported Cases (2015-2019).

Figure 2 illustrates a time series analysis of B. Pertussis cases, showing a long-term decline in cases until 2017, followed by a notable increase in 2018-2019. The seasonal component indicates regular peaks, indicating seasonal variation, while the residual component remains mostly stable but shows anomalies during 2018-2019, suggesting the potential impact of external factors.

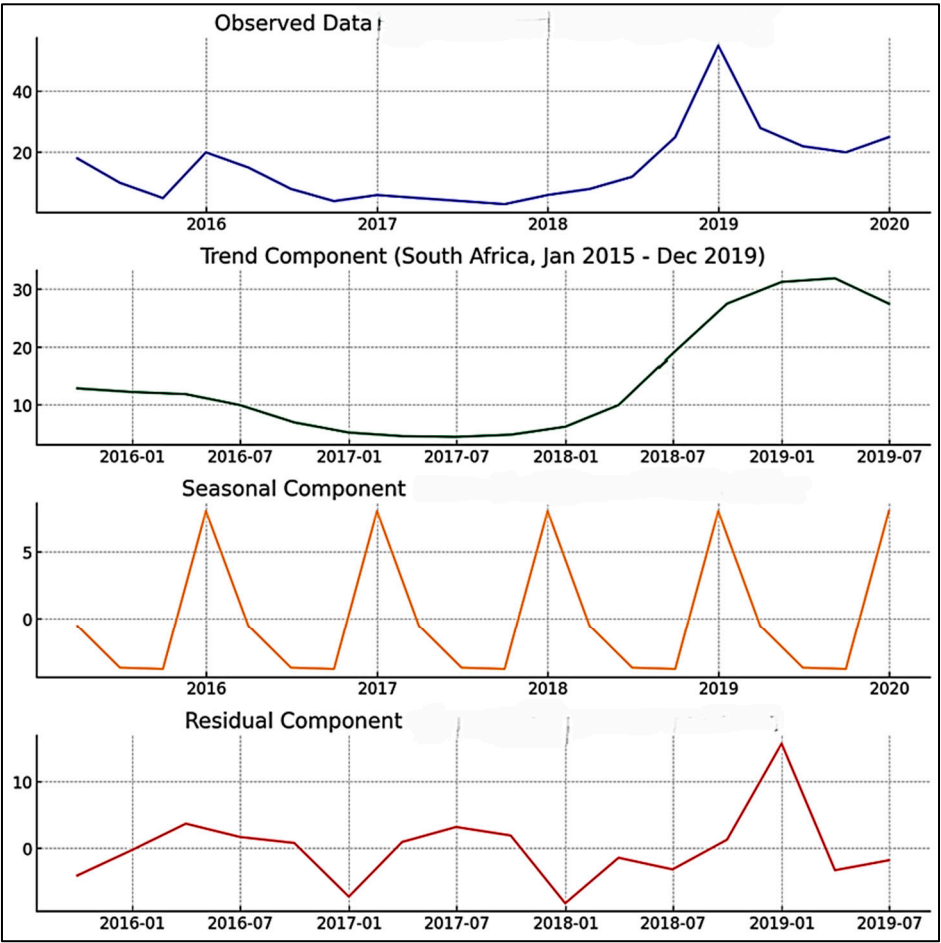


Figure 2. Time Series Decomposition of B. Pertussis Cases (2015-2019).

3.3. Trend Model Comparisons

Table 2 summarises the Mean Squared Error (MSE) for three trend models: Linear, Cubic, and Exponential. MSE measures the average squared differences between predicted and actual values, with lower values signifying better performance. The findings indicate that the cubic model offers the best predictive accuracy, while the exponential model is the least effective.

Table 2. Trend Model Performance Comparison.

Trend Model	Mean Squared Error (MSE)
Linear	114.95
Cubic	83.49
Exponential	122.16

3.4. Forecasting Using the Cubic Model (2020-2021)

Figure 3 presents the projected case counts from the cubic model for 2020 to 2021. Notably, it reveals a peak in cases during 2020, followed by a sharp decline in 2021.

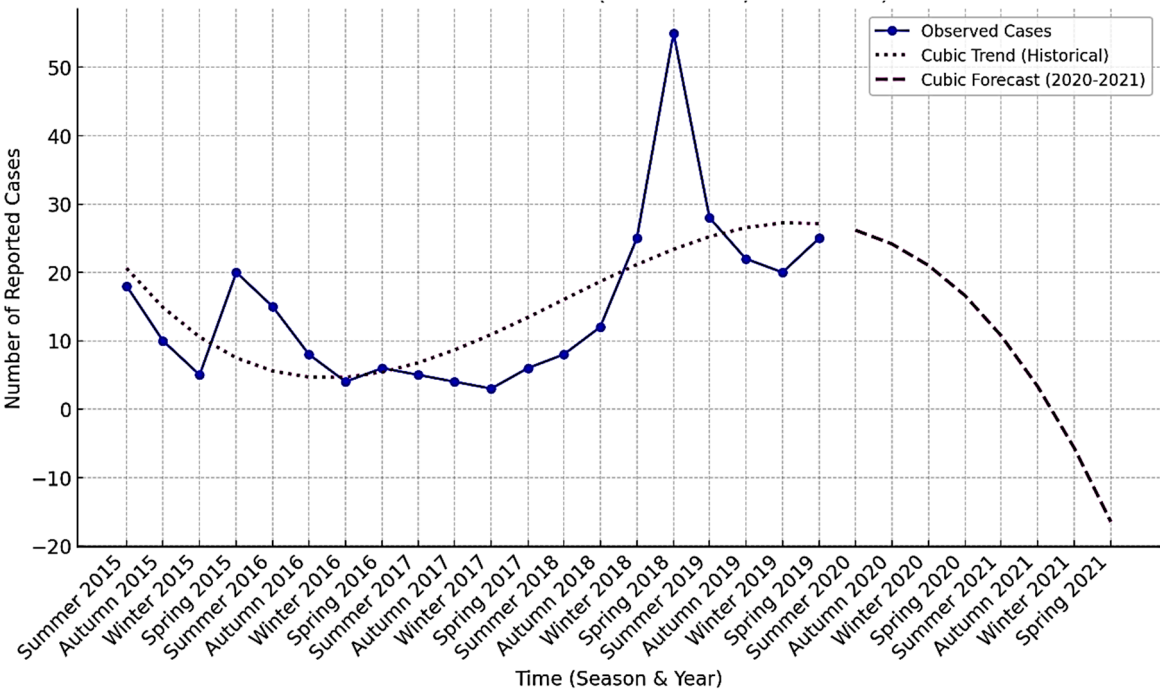


Figure 3. Forecasted B. Pertussis Cases (2020-2021) Using the Cubic Model (Authors' work).

Table 3 depicts the forecasted cases of B. B. Pertussis for 2020-2021, showing cases peaking in 2020 and then sharply declining in 2021.

Table 3. Forecasted Cases for 2020-2021(Authors' work).

Date	Forecasted Cases (Cubic)
March 2020	26.18
June 2020	24.17
September 2020	21.02
December 2020	16.58
March 2021	10.75

Table 4 depicts the ARMA(4) model which shows pertussis case counts are dependent on short term fluctuations.

Table 4. ARMA(4) Model Estimation Results (Authors' work).

Variable	Coefficient	Std. Error	t-Statistic	p-Value
AR(1)	0.797***	0.167	4.77	0.000
MA(1)	-0.698***	0.189	-3.69	0.001
Constant	21.77***	1.417	15.37	0.000

Key ***: p-value<0.001.

Figure 4 depicts the actual vs. ARMA(4) model fitted values which) model closely follows actual trends, peaks and anomalies in 2018-2019 are not fully captured.

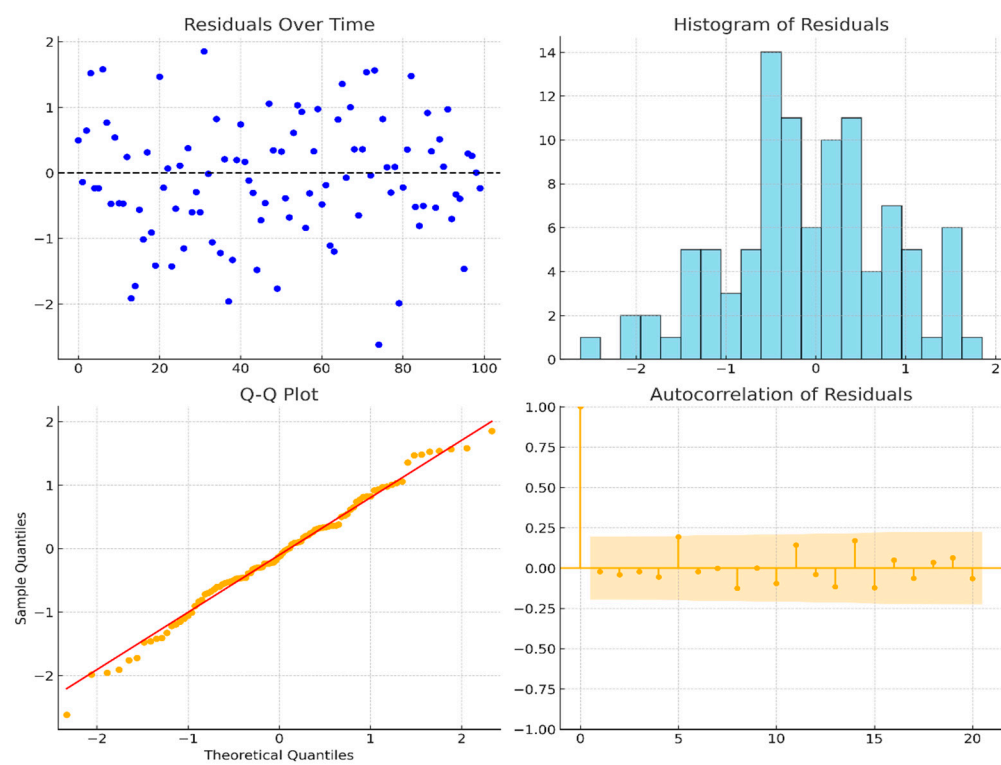


Figure 4. Actual vs. ARMA(4) Model Fitted Values (Graph Included).

Interpretation: While the ARMA(4) model closely follows actual trends, **peaks and anomalies in 2018-2019 are not fully captured**, suggesting external factors may influence outbreaks.

Table 5 depicts the residual diagnostics of the ARMA(4) model which shows that the Ljung-Box Test was insignificant. (p-value = 0.09).

Table 5. Residual Diagnostics of ARMA(4) Model (Authors’ own work).

Statistic	Value
Mean Residual	0.002
Std. Deviation	3.78
Ljung-Box Test (p-value)	0.09

Figure 5 depicts that the residuals appear mostly stable, but occasional deviations suggest additional influencing factors.

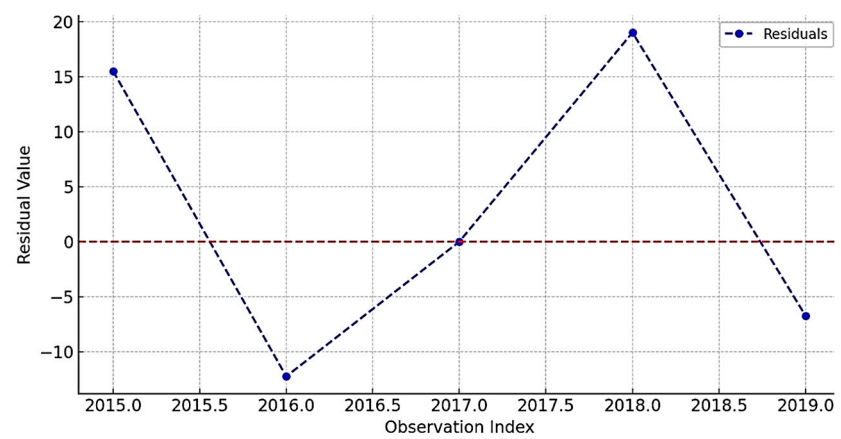


Figure 5. Residuals of ARMA(4) Model.

4. Discussion

The analysis of B. Pertussis trends in the Tshwane Health District over five years confirms the hypothesis of public health experts: B. Pertussis persists and demonstrates a distinct seasonal pattern, with peaks observed in Spring and Winter. This pattern corresponds with global observations, suggesting that B. Pertussis cases typically rise during colder months [27]. The significance of these findings is that, despite vaccination initiatives, B. Pertussis persists in predictable cycles, highlighting the necessity for more strategic, data-informed interventions.

One of this study's key contributions is its use of ARMA(4) modelling, which successfully captured short-term fluctuations in B. Pertussis incidence. The cubic trend model provided a reliable approach to long-term forecasting. These models offer a window into the future, helping policymakers anticipate outbreaks rather than react to them.

Our modelling approach further develops existing B. Pertussis forecasting studies, which have predominantly used SARIMA models for their capacity to incorporate seasonality [10]. Our findings indicate that ARMA-based models can yield comparably robust short-term forecasts, especially when integrating external factors like vaccination coverage and climate data [26]. This offers a chance to enhance public health forecasting instruments by incorporating real-time surveillance data.

The current era necessitates an increased emphasis on predictive analytics in disease control efforts. Incorporating real-time Google Trends data and surveillance trends could facilitate the development of early warning systems, enabling public health officials to prepare for seasonal surges in B. Pertussis. This concept has been examined in other infectious disease modelling studies, where integrating Google Trends or electronic health records has improved outbreak predictions [23]. Predicting the timing and location of case spikes allows preemptive intervention to prevent outbreak escalation.

The findings underscore the significance of strategically timing vaccination efforts. The increase of B. Pertussis in well-vaccinated populations is partially linked to waning immunity and the rise of pertactin-deficient strains. (Heininger et al., 2024). Our study confirms the existence of seasonal peaks, suggesting that administering booster doses prior to high-incidence periods could enhance herd immunity during critical times [25].

5. Conclusions

This research illustrates the efficacy of time series models in predicting B. Pertussis incidence, underscoring the significance of seasonal variations and climatic factors in the disease's dynamics. The results underscore the critical need for prompt vaccination initiatives, climate-responsive health strategies, and improved surveillance systems to reduce the frequency of B. Pertussis outbreaks effectively. Incorporating climate data and real-time surveillance into forthcoming forecasting models could significantly enhance epidemic preparedness and guide the development of targeted public health interventions.

Institutional Review Board Statement: Ethical approval for the study was received from the School of Medicine Ethics Committee and the Sefako Makgatho Health Science University Research Ethics Committee (SMUREC/M/58/2021:PG). The Head of The Gauteng Department Health gave permission to conduct the study. The study adhered to confidentiality protocols to protect sensitive information, ensuring the confidentiality and safety of participants' data by maintaining privacy in record-keeping.

Informed Consent Statement: Ethical approval for the study was received from the School of Medicine Ethics Committee and the Sefako Makgatho Health Science University Research Ethics Committee (SMUREC/M/58/2021:PG). The Head of The Gauteng Department Health gave permission to conduct the study. The study adhered to confidentiality protocols to protect sensitive information, ensuring the confidentiality and safety of participants' data by maintaining privacy in record-keeping.

Data Availability Statement: Research data is available upon reasonable request.

Acknowledgments: The authors would like to express their gratitude to the Tshwane Health District staff for their support in helping with the data collection for this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ward JI, Cherry JD, Chang S-J, Partridge S, Lee H, Treanor J, et al. Efficacy of an Acellular Pertussis Vaccine among Adolescents and Adults. 2005.
2. Muloiwa R, Kagina BM, Engel ME, Hussey GD. The burden of pertussis in low- and middle-income countries since the inception of the Expanded Programme on Immunization (EPI) in 1974: A systematic review protocol. *Syst Rev*. 2015;4.
3. Zhang X, Zhang T, Young AA, Li X. Applications and comparisons of four-time series models in epidemiological surveillance data. *PLoS One*. 2014;9.
4. Robertson C, Nelson TA, MacNab YC, Lawson AB. Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology*. 2010;1:105–16.
5. Cazelles B, Chavez M, De Magny GC, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*. 2007;4:625–36.
6. Cardinal M, Lambert J, Roy R. On the application of integer-valued time series models for the analysis of disease incidence. *Stat Med*. 1999;18:2025–39.
7. Gentleman R, Hornik K, Parmigiani G. *Epidemics Models and Data using R*. 1st edition. University Park, PA, USA; 2018.
8. Zeger S, Irizarry R, Peng R. On time series analysis of public health and biomedical data. *Annu Rev Public Health*. 2006;27:57–79.
9. Zhang Y, Bambrick H, Mengersen K, Tong S, Feng L, Liu G, et al. Association of weather variability with resurging pertussis infections among different age groups: A nonlinear approach. *Science of the Total Environment*. 2020;719.
10. Zeng Q, Li D, Huang G, Xia J, Wang X, Zhang Y, et al. Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. *Sci Rep*. 2016;6.
11. Huang X, Lambert S, Lau C, Soares Magalhaes RJ, Marquess J, Rajmohan M, et al. Assessing the social and environmental determinants of pertussis epidemics in Queensland, Australia: A Bayesian spatio-temporal analysis. *Epidemiol Infect*. 2017;145:1221–30.
12. Dafilis MP, Frascoli F, McVernon J, Heffernan JM, McCaw JM. The dynamical consequences of seasonal forcing, immune boosting and demographic change in a model of disease transmission. *J Theor Biol*. 2014;361:124–32.
13. Gomes MC, Gomes JJ, Paulo AC. Diphtheria, pertussis, and measles in Portugal before and after mass vaccination: A time series analysis. *Eur J Epidemiol*. 1999;791–8.
14. Anderson RM, Grenfell BT, May RM. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: Time series analysis. *Journal of Hygiene*. 1984;93:587–608.
15. Skoff TH, Hadler S, Hariri S. The epidemiology of nationally reported pertussis in the United States, 2000–2016. *Clinical Infectious Diseases*. 2019;68:1634–40.
16. Yadav SK, Akhter Y. Statistical Modeling for the Prediction of Infectious Disease Dissemination With Special Reference to COVID-19 Spread. *Frontiers in Public Health*. 2021;9.
17. Kumar N, Susan S. COVID-19 Pandemic Prediction using Time Series Forecasting Models. 2020.
18. Yu G, Feng H, Feng S, Zhao J, Xu J. Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA-NNAR hybrid model. *PLoS One*. 2021;16 2 February.
19. Nobre FF, Monteiro ABS, Telles PR, David Williamson G. Dynamic linear model and SARIMA: A comparison of their forecasting performance in epidemiology. *Stat Med*. 2001;20:3051–69.

20. Qi C, Zhang D, Zhu Y, Liu L, Li C, Wang Z, et al. SARFIMA model prediction for infectious diseases: Application to hemorrhagic fever with renal syndrome and comparing with SARIMA. *BMC Med Res Methodol.* 2020;20.
21. Allard R. Update/Le point Use of time-series analysis in infectious disease surveillance. 1998.
22. Wang M, Pan J, Li X, Li M, Liu Z, Zhao Q, et al. ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021. *BMC Public Health.* 2022;22.
23. Nann D, Walker M, Frauenfeld L, Ferenci T, Sulyok M. Forecasting the future number of pertussis cases using data from Google Trends. *Heliyon.* 2021;7.
24. Raycheva R, Kevorkyan A, Stoilova Y. Stochastic Modelling of Scalar Time Series of Varicella Incidence for a Period of 92 Years (1928-2019). *Folia Med (Plovdiv).* 2022;64:624–32.
25. Campbell PT, McCaw JM, McVernon J. Pertussis models to inform vaccine policy. *Human Vaccines and Immunotherapeutics.* 2015;11:669–78.
26. Li X, Patel V, Duan L, Mikuliak J, Basran J, Osgood ND. Real-Time Epidemiology and Acute Care Need Monitoring and Forecasting for COVID-19 via Bayesian Sequential Monte Carlo-Leveraged Transmission Models. *Int J Environ Res Public Health.* 2024;21.
27. Broutin H, Viboud C, Grenfell BT, Miller MA, Rohani P. Impact of vaccination and birth rate on the epidemiology of pertussis: A comparative study in 64 countries. *Proceedings of the Royal Society B: Biological Sciences.* 2010;277:3239–45.
28. Heininger U, Martini H, Eeuwijk J, Prokić I, Guignard AP, Turriani E, et al. Pertactin deficiency of *Bordetella pertussis*: Insights into epidemiology, and perspectives on surveillance and public health impact. *Human Vaccines and Immunotherapeutics.* 2024;20

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.