Walisongo Journal of Information Technology
Vol 7 No 2 (2026): 19-28. DOI: 10.21580/wjit.2025.7.2.16856
ISSN 2715-0143 (online) | 2714-9048 (printed)

Seasonal Periods Seasonal Time Series Models for Forecasting Pertussis Endemic with Holt-Winter's Exponential Smoothing

Ashri Shabrina Afrah[1], Nur Fitriyah Ayu Tunjung Sari[2], Shoffin Nahwa Utama[3], Tri Mukti Lestari[4]
[123] Universitas Islam Negeri Maulana Malik Ibrahim Malang
Corresponding Author : ashri.shabrina@ti.uin-malang.ac.id

Abstract: *Accurate forecasting of pertussis cases can support disease surveillance and public health planning. This study evaluates the application of Holt–Winters exponential smoothing to forecast weekly pertussis cases in the United States. The study uses weekly case data reported by the Centers for Disease Control and Prevention (CDC) and the Florida Department of Health, covering the period from 2008 to 2012, with four additional observations from the final weeks of 2007 included as initial data. Two Holt–Winters configurations were evaluated using seasonal periods of 4 and 13 weeks. The dataset was divided into training and testing sets at a ratio of 80:20, and forecasting accuracy was assessed using the Mean Absolute Percentage Error (MAPE). The reported results show that the 4-week configuration achieved a MAPE of 34.7898%, whereas the 13-week configuration achieved a MAPE of 40.7340%. Thus, the 4-week configuration produced a lower MAPE than the 13-week configuration for the data evaluated in this study. However, the results should be interpreted within the scope of the selected dataset and seasonal configurations. Further evaluation using alternative seasonal periods, clearly specified testing procedures, and additional datasets is necessary to determine the robustness and generalizability of the forecasting performance.*

Keyword : *Holt–Winters Exponential Smoothing; Pertussis; Time Series Forecasting; MAPE; Disease Surveillance*

Copyright © 2025 Walisongo Journal of Information Technology

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1. Introduction

Pertussis, commonly known as whooping cough, is a highly contagious respiratory disease caused by the bacterium Bordetella pertussis. Despite the availability of vaccination, pertussis remains a public health concern because immunity may wane over time and the disease can cause severe complications, particularly in infants and young children (Gabutti & Rota, 2012). (Havers et al., 2015) reports that pertussis occurs worldwide and that reported cases may increase during particular periods, although the disease does not exhibit a distinct seasonal pattern in every setting. These characteristics highlight the importance of monitoring changes in reported case numbers and understanding their temporal distribution. Time-series forecasting can contribute to this process by using historical observations to estimate future

case counts, thereby providing additional information for disease surveillance and public health planning (Chowell & Skums, 2024; Hand, 2009) . However, forecasts should be interpreted as estimates subject to uncertainty rather than as definitive predictions of future outbreaks.

Previous studies have investigated the temporal characteristics and forecasting of pertussis incidence using different statistical approaches. (De Greeff et al., 2009) examined long-term periodicity and seasonal patterns in pertussis incidence in the Netherlands, reporting variation in the timing of annual incidence peaks across age groups. Their findings demonstrate that temporal patterns can differ between populations and demographic groups, indicating that seasonal characteristics should be examined in the context of the data being analyzed. (Wang et al., 2018) developed a wavelet-based hybrid model combining Seasonal Autoregressive Integrated Moving Average (SARIMA) and Nonlinear Autoregressive (NAR) approaches to model pertussis incidence in China. In another study, (Nann et al., 2021) investigated the use of Google Trends data to improve SARIMA-based forecasts of weekly pertussis cases. Their findings indicated that incorporating pertussis-related search data improved predictive performance over the traditional model during the evaluated period. Together, these studies illustrate the potential of time-series modeling and supplementary information for forecasting pertussis cases. Nevertheless, differences in geographic coverage, observation periods, data sources, and model specifications mean that forecasting performance should be evaluated for the particular dataset and prediction task under investigation.

Holt–Winters exponential smoothing is an established forecasting approach that incorporates level, trend, and seasonal components. The method is associated with the development of exponential smoothing models by (Holt, 2004) and (Winters, 1960) , and it remains an important approach for forecasting time series with trend and seasonal variation (Gardner, 2006; Hyndman, R.J., & Athanasopoulos, 2015). Its practical implementation requires appropriate decisions regarding the seasonal structure, smoothing parameters, and initial component values. (Chatfield & Yar, 1988) discussed practical issues in Holt–Winters forecasting, including the selection of starting values and smoothing parameters. Similarly, (Hand, 2009) provided a statistical treatment of exponential smoothing methods and their underlying state-space formulations. The selection of a forecasting method and its configuration should therefore consider the characteristics of the observed series rather than assume that a particular method is universally superior. Furthermore, comparisons between forecasting methods do not always produce a consistent winner across different datasets, as illustrated by the comparative evaluation of seasonal exponential smoothing methods by (Bartolomei & Sweet, 1989). These considerations support the need to evaluate alternative model configurations using a clearly defined and consistent assessment procedure.

The seasonal component is particularly relevant to the present study because the analyzed observations represent weekly pertussis cases. Seasonal patterns identified in monthly or annual epidemiological data cannot automatically be assumed to correspond to shorter cycles in weekly observations. The choice of seasonal period should therefore be treated as a modeling decision that requires empirical evaluation. In this regard, the present study examines two Holt–Winters configurations with seasonal periods of 4 and 13 weeks. These configurations represent alternative cycle lengths for comparison; they are not assumed to represent established biological or epidemiological cycles of pertussis. The forecasting performance of the two configurations is assessed using the Mean Absolute Percentage Error (MAPE), which expresses forecast errors relative to observed values. However, the interpretation of forecast accuracy requires care because different accuracy measures have different properties and limitations (Hyndman & Koehler, 2006). Consequently, the reported MAPE values should be interpreted in conjunction with the evaluation procedure, the characteristics of the dataset, and the scope of the forecasting task.

Based on these considerations, this study evaluates the application of Holt–Winters exponential smoothing to weekly pertussis cases in the United States using data covering 2008–2012, with additional observations from the final weeks of 2007 reported as initial data in the manuscript. The study compares two seasonal configurations, with periods of 4 and 13 weeks, and evaluates their forecasting accuracy using MAPE. The specific objective is to determine which of the two tested configurations produces the lower MAPE under the experimental conditions used in this study. Accordingly, the contribution of this research is a comparative evaluation of alternative seasonal-period configurations for Holt–Winters forecasting of weekly pertussis cases. The study does not seek to establish a universally optimal seasonal period or to demonstrate the superiority of Holt–Winters over other forecasting methods, since such conclusions would require broader comparative experiments and additional datasets.

2. Method

Forecast of Time Series Data

Forecasting is an analytical process used to estimate future values based on historical and current data, thereby supporting informed decision-making and the selection of appropriate actions. *Forecasting* is needed to create effective and efficient planning. There are 3 types of forecasting, namely:

1. Quantitative Forecasting

   Quantitative forecasting is carried out when the required quantitative data are available. There are 2 types of quantitative forecasting, they are time series and explanatory.

2. Qualitative Forecasting

   This type of forecasting is carried out when there is no available data at all or there are only some quantitative data available, but the qualitative are all there which can be used in forecasting.

3. Unpredictable Forecasting

   This type of forecasting is done when there is little or no needed information for forecasting.

Forecasting on time series data is quantitative. Historical data usually consists of a series of observations over time. This series is known as a time series. In forecasting, an analysis of the data pattern is usually carried out to see its tendency to continue in the future. Infectious disease surveillance data, such as weekly or monthly case counts, can exhibit temporal patterns that are suitable for time-series analysis and forecasting. Previous studies have applied time-series forecasting methods to dengue cases and pertussis incidence(Fitria et al., 2017; Nann et al., 2021).

Holt-Winter's Exponential Smoothing Method

Holt-Winter's Exponential Smoothing method is commonly called as Triple Exponential Smoothing (Chatfield & Yar, 1988). This method is able to handle time-series data containing trends as well as seasons and uses 3 smoothing constants. In this method, there are 3 smoothing equations, namely level smoothing, trend smoothing, and seasonal smoothing.

The equations used in Holt-Winter's Exponential Smoothing method are shown in Equation (1), Equation (2), Equation (3), and Equation (4):
Level smoothing:

$$L = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (1)$$

Trend smoothing:

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (2)$$

Seasonal smoothing:

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s} \quad (3)$$

Forward forecasting:

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (4)$$

Where:
Yt = actual data in t period
Lt = level estimation
bt = trend estimation
St = seasonal estimation
α = smoothing constant for level estimation
β = smoothing constant for trend estimation
γ = smoothing constant for seasonal estimation
s = season length
m = number of forward forecasting
F(t+m) = forecasting value of m forward period

Model Assessment Method

Accuracy test of the forecasting is carried out to measure the suitability of the forecasting results with the historical data used. The accuracy test is carried out using statistical measures, both standard and relative statistical measures. This needs to be done in an effort to validate the model created. In general, the test is carried out based on the residual value, as seen in the Equation (5) below.

$$e_t = X_t - Y_t \quad (5)$$

Where et means an error value, Xt is the actual data, and Yt is the forcasting value.

In addition to the standard measurement, tests based on the relative error (ret) values of forecasting are also carried out by using Equation (6).

$$re_t = \frac{X_t - Y_t}{X_t} \times 100 \quad (6)$$

Where ret is the relative error, Xt is the actual data, and Yt is the forcasting value.

One of the methods for forecasting accuracy based on the forecasting relative error value is the Mean Absolute Percentage Error (MAPE), according to Equation (7) (Chang, Wang, & Liu, 2007):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|re_t| \quad (7)$$

Where ret is the relative error and n is the number of data.

The smaller the value of MAPE, the more accurate a model does the forecasting. The MAPE value guideline are shown in Table 1.

Table 1. MAPE Value Guideline

MAPE Value	Description
≤ 10	Forecasting result is highly accurate
10-20	Forecasting result is very good
20-50	Forecasting result is decent (good enough)
> 50	Forecasting result is inaccurate

Table 1 shows that the value of MAPE should be 50% or below. If the MAPE value is above 50%, the forecasting results tend to be inaccurate. If the MAPE values below 10%, we can assume that the forecasting results are highly accurate.

Data and Methodology

The data used in this study are the numbers of weekly pertussis cases in the United States in 2008-2012. This data was taken from the official website of the Centers of Disease Control and Prevention (CDC) and the Florida Department of Health. The data illustrates the numbers of weekly pertussis cases in all US States, both infecting US citizens and foreign nationals. Table 2 below displays an overview of the data used here.

Table 2. Data of Weekly Pertussis Cases in the US

Year	Week	Number of Case
2007	49	60
2007	50	86
2007	51	40
2007	52	92
2008	1	34
2008	2	62
2008	3	63
2008	4	59
2008	5	217
2008	6	101
2008	7	56
2008	8	28
…	…	…
2012	52	167

As shown in Table 2., we use the 5 years data of the numbers of weakly pertussis cases. To simplify the seasonal time-series analysis of the data, 4 rows of data on the number of pertussis cases were added in December 2007, namely in the 49th, 50th, 51st, and 52nd weeks of the year.

According to previous research (Nann, et al., 2021), the data on the number of pertussis cases in the United States has a seasonal time-series pattern. To determine whether the data obtained really has the pattern, it is necessary to do an autocorrelation test. The most commonly used autocorrelation test method is the Durbin-Watson Test.

The Equation (8) is for conducting The Durbin-Watson Test [10]:

$$DW = 2[1 - \rho(1)] \quad (8)$$

DW is the Durbin-Watson Test autocorrelation test value and ρ(1) is the 1st order autocorrelation value of the data to be analyzed. According to the study conducted by Guha and Bandyopadhyay (Guha & Bandhyopadyay, 2016), Durbin-Watson Test value between 0 until 1.5 indicates that the data has a dependency on time so that time-series model can be applied.

The next step is analyzing the data using Holt-Winter's Exponential Smoothing Method. The data is divided in to two categories, those are training data and testing data. The proportions of training and testing data is 80:20.

In other research (De Greeff, et al., 2009), a study was conducted on the pattern of data on the number of pertussis cases in the United States. From this study, it can be concluded that the data on the number of pertussis cases in the United States is seasonal time series data, with the highest number of cases occurring in August. According to this study, a 4-week (monthly) seasonal pattern can be used on the data.

Pertussis is an infectious disease caused by a viral infection. One of the conditions that support the life of viruses as disease vectors is the weather conditions of an environment. In the WHO report on climate change and infectious diseases, it is stated that the population of disease vectors is closely related to conditions of air temperature, humidity and rainfall (Surahman, 2015). In temperate to subtropical countries such as the United States, there are 4 seasons that occur in a year, namely spring (March-May), summer (June-August), autumn (September-November), and winter (December-February). The weather in each season has special characteristics, including air temperature, humidity, and rainfall. Therefore, a second model was created to analyze the data on the number of pertussis cases in the United States, namely a model with a seasonal pattern of 13 weeks (1 season).

Based on the explanation, time series models used in this study are present in Table 3. To obtain the value of smoothing level constant (α), trend level constant (β), and seasonal level constant (γ), we use Solver add ins of MS Excel.

Table 3. Time Series Model for Weekly Pertussis Case Forecast

Parameters	Model 1	Model 2
Smoothing level constant (α)	0.1	0.1
Trend level constant (β)	0.1	0.1
Seasonal level constant (γ)	0.1	0.1
Season length (s)	4	13

To obtain forecasting results with the created model, a simulation is carried out with the Python Programming Language in the Google Collaboratory application. Both models were

simulated, then tested to get the accuracy measurements using the statistical measure of Mean Absolute Percentage Error (MAPE). be presented.

3. Result and Discussion

After collecting data, an autocorrelation test was done to find out whether the data has dependency on time. Therefore, Durbin-Watson Test was carried out with the following Equation (8):

$$DW=2[1-\rho(1)]$$

$$DW=2(1-0.859506)$$

$$= 0.280988$$

Durbin-Watson Test shows the value is 0,280988, thus it can be assumed that the data are suitable for time-series model.

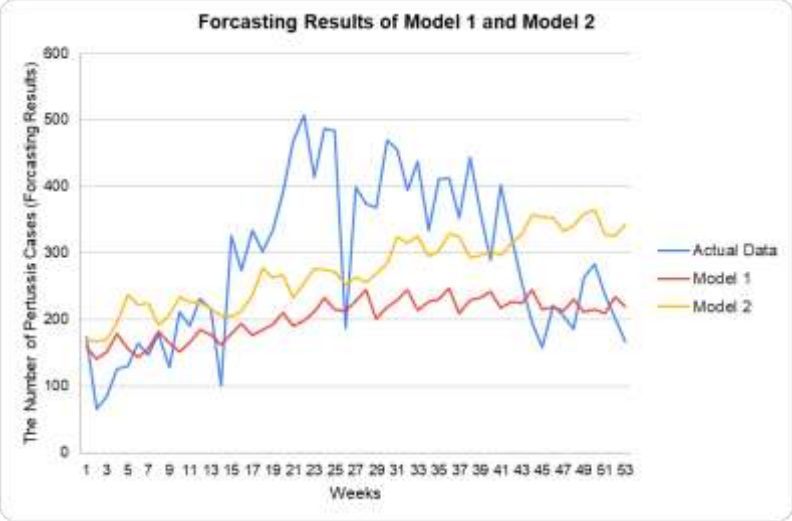
Then, simulations were done on the two models (as in Table 3). The simulation results show the predictions of the number of pertussis cases during 53 weeks. The overview of the forecasting results of the two models can be seen in Table 4. and Figure 1. below.

Table 4. Forecasting Results of Model 1 and Model 2

Week	Model 1	Model 2
1	158.3688402	169.9314398
2	141.5328118	166.6514357
3	151.3354217	170.7004505
4	178.6823462	195.6257415
5	156.0901591	237.027663
6	143.3473456	222.7535412
7	156.4695533	224.1543247
8	181.9151134	191.0992992
9	164.845218	206.6634819
10	151.5797999	233.5380708
11	166.3028143	227.0468748
12	184.5829932	225.3923524
13	177.0852941	216.8113603
14	162.5913762	205.3755029
15	179.2309933	204.8317457
16	193.6380153	212.7113697
17	176.225261	235.8846367

...	...	...
53	219.8765097	343.060059

Figure 1. Graphic of forecasting results of pertussis cases spread with Holt-Winter's Exponential Smoothing Method



Besides, accuracy tests of the forecasting were also done using statistic measure of Mean Absolute Percentage Error (MAPE). The values of MAPE can be seen on Table 5.

Table 5. MAPE Values of the Forecasting

Statistic Measure	Model 1	Model 2
MAPE	34.78980129333%	40.7339504170%

This paper aims to develop a forecasting model for pertussis endemic by using The Holt-Winter's Exponential Smoothing with an appropriate seasonal pattern.Based on the values of MAPE in Table 5, it can be found that Model 1 has a better MAPE than Model 2. This result is in accordance with the other research that explain that a 4-week seasonal pattern is suitable for the spread of pertussis endemic forecasting (De Greeff, et al., 2009). In addition, the MAPE value is also better than the model in another study using the SARIMA method, which is 58.59% (Nann, et al., 2021). This shows that time series analysis with a seasonal pattern of 4 weeks (1 month) is more suitable to be applied to data on the number of pertussis cases in the United States for the data used in this research, although further studies by using other data of pertussis endemic spread is required.

4. Conclusion

This study evaluates the performance of the Holt–Winters exponential smoothing method for forecasting weekly pertussis cases using two different seasonal periods. The results indicate that Model 1, which uses a four-week seasonal period, achieved a lower Mean Absolute

Percentage Error (MAPE) of 34.7898% than Model 2, which uses a 13-week seasonal period and achieved a MAPE of 40.7339%. These findings suggest that the four-week configuration provided better forecasting accuracy than the 13-week configuration for the dataset evaluated in this study. However, this result does not establish that a four-week seasonal period is universally more appropriate for forecasting pertussis cases. Further studies using alternative seasonal periods, clearly defined evaluation procedures, and additional datasets are needed to assess the robustness and generalizability of the findings.

5. References

Bartolomei, S. M., & Sweet, A. L. (1989). A note on a comparison of exponential smoothing methods for forecasting seasonal series. *International Journal of Forecasting*, *5*(1). https://doi.org/10.1016/0169-2070(89)90068-X

Chatfield, C., & Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *The Statistician*, *37*(2). https://doi.org/10.2307/2348687

Chowell, G., & Skums, P. (2024). Investigating and forecasting infectious disease dynamics using epidemiological and molecular surveillance data. In *Physics of Life Reviews* (Vol. 51). https://doi.org/10.1016/j.plrev.2024.10.011

De Greeff, S. C., Dekkers, A. L. M., Teunis, P., Rahamat-Langendoen, J. C., Mooi, F. R., & De Melker, H. E. (2009). Seasonal patterns in time series of pertussis. *Epidemiology and Infection*, *137*(10). https://doi.org/10.1017/S0950268809002489

Fitria, J. D., Karohmah, N., & Sunarmi. (2017). Peramalan Jumlah Pasien DBD di Rsud Dr. Soeselo Slawi dengan Metode Dekomposisi dan Triple Exponential Smoothing Winter's. *Seminar Nasional Pendidikan, Sains Dan Teknologi*.

Gabutti, G., & Rota, M. C. (2012). Pertussis: A review of disease epidemiology worldwide and in Italy. In *International Journal of Environmental Research and Public Health* (Vol. 9, Number 12). https://doi.org/10.3390/ijerph9124626

Gardner, E. S. (2006). Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting*, *22*(4). https://doi.org/10.1016/j.ijforecast.2006.03.005

Hand, D. J. (2009). Forecasting with Exponential Smoothing: The State Space Approach by Rob J. Hyndman, Anne B. Koehler, J. Keith Ord, Ralph D. Snyder. *International Statistical Review*, *77*(2). https://doi.org/10.1111/j.1751-5823.2009.00085_17.x

Havers, F. P., Moro, P. L., Hariri, S., & Skoff, T. (2015). Centers for Disease Control and Prevention. "Pertussis" from Epidemiology and Prevention of Vaccine-Preventable Diseases. *Washington D.C. Public Health Foundation, 2015.*, *13th Editi*.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1). https://doi.org/10.1016/j.ijforecast.2003.09.015

Hyndman, R.J., & Athanasopoulos, G. (2015). Forecasting: Principles and Practice (3rd ed). *OTexts: Melbourne, Australia. OTexts.Com/Fpp3*.

Nann, D., Walker, M., Frauenfeld, L., Ferenci, T., & Sulyok, M. (2021). Forecasting the future number of pertussis cases using data from Google Trends. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08386

Wang, Y., Xu, C., Wang, Z., Zhang, S., Zhu, Y., & Yuan, J. (2018). Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. *PLoS ONE*, *13*(12). https://doi.org/10.1371/journal.pone.0208404

Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, *6*(3). https://doi.org/10.1287/mnsc.6.3.324