APPLICATION OF THE HOLT-WINTERS EXPONENTIAL SMOOTHING METHOD ON THE AIR POLLUTION STANDARD INDEX IN SURABAYA

Silmi Muna¹, Kuntoro²

¹,² Faculty of Public Health, Universitas Airlangga, East Java, Indonesia
Corresponding Author: Silmi Muna
E-mail: silmimuna2@gmail.com

ABSTRACT

The Air Pollution Standards Index (APSI) is an indicator that shows how clean or polluted the air is in a city. It also portrays the health impacts towards the people who breathe it in. Based on the Indonesian Ministry of Environment monitoring through the Air Quality Monitoring Station (AQMS), the city of Surabaya only had 22 up to 62 days of air categorized as good in a year. The purpose of this study was to forecast APSI as a scientific-based reference for making decisions and policies that were appropriate in tackling the effects of air pollution on health. This study was non-obstructive or non-reactive research. The research method used was time series to identify the time relationship. The data used were secondary data taken from the APSI documents from 2014 to 2019 at the Surabaya City Environment Agency. The results of this study obtained the best model through α (0.8), γ (0.5), and δ (0.6) with the values of MAPE (0.104355), MAD (0.00842), and MSD (0.00105) calculated with the Holt-Winters exponential smoothing method. The highest produced forecast value of APSI was in September 2020, and the smallest was in January 2020. This study suggests the government of Surabaya to create policies and programs to suppress the number within APSI.

Keywords: Air Pollution Standards Index, exponential smoothing, holt-winters

INTRODUCTION

Air is a necessity for all living things to breathe. The Air Pollution Standard Index (APSI) is a report on the air quality of a city, providing information on air cleanliness and pollution, as well as the health impacts of inhaling air.

Surabaya, one of Indonesia’s big cities, in general has been progressing quite fast. One of the sectors progressing quickly is transportation. Currently, for industrial progress, it produces various types of vehicles and transportation for the community as transportation nowadays is a necessity. The advancement of transportation industry has...
caused a continuous increase in the number of cars each year, leading to traffic jams and the increase of air pollutants. In Indonesia’s big cities, one problem often faced is pollution (Prahards, 2018).

The results of the Air Quality Monitoring Station (AQMS) report from the Indonesian Ministry of Environment showed that in the six major cities of Pekanbaru, Medan, Jambi, Bandung, Surabaya, and Jakarta, there were only 22 to 62 days of good air per year (Darmawan, 2018).

Forecasting is a science to predict, estimate or describe a current or future event, using previous or past data. The purpose of forecasting is to provide a guide for the future. The future itself could not be ascertained, and thus forecasting methods are needed to assist in planning (Encik, 2016). Time series is a method of forecasting in statistics using time-series data or previous quantitative data. The time series method could also develop an event, an event and its relationship with or influence on other occasions.

The Holt-Winter method is a completion of the Holt-Brown method. Holt’s two-parameter model, also known as linear exponential smoothing, could be used for data of trend patterns. However, data of both trend and seasonal patterns cannot use Holt’s linear exponential smoothing because the method cannot detect seasonal patterns. The Holt-Winters exponential smoothing method is a double extension of the exponential smoothing method designed for seasonal trends, trend changes, and seasonal correlations (Gelper, 2010).

All activities carried out do not necessarily require forecasting for future decisions or planning. Nevertheless, forecasting is an essential method for planning and estimating the efficiency and effectiveness of the future.

According to the Regulation of the Indonesia Ministry of Environment Number 41 of 199 concerning Air Pollution Control in Chapter 2 Part 6 Article 12, the Air Pollution Standard Index (APSI) must deal with the health of humans, plants, animals, aesthetic values, and buildings. Therefore, Surabaya needs a forecast calculation by APSI to plan decisions and sort strategies based on the results. The data for APSI forecasting in Surabaya was taken from the Surabaya City Air Pollution Standard Index periodically for the past five years.

The goal of the current research was to predict the Surabaya APSI figure in 2020 based on APSI data from January 2014 to December 2019 using the Holt-Winter exponential smoothing forecasting model.

METHODS

The research was conducted using secondary data from the APSI archives and reporting documents in Surabaya. These documents were then collected from 4 APSI equipment stations spread across the city. The time series method was employed to observe the relationship between several times. This study was conducted at the Surabaya Environmental Service by collecting secondary data of APSI within the city from January 2014 to December 2019.

The data were then divided into two in-sample and out-sample data. In-sample data were from the 1st to the 60th period (January 2014-December 2018). Out-sample data were from 61st to 72nd (January 2019-December 2019). The data were used to validate the results of forecasting. The best models found used the trial and error methods α, β, and γ between the value range of 0 to 1.

Data Preparation

Data prepared for pattern testing were the Air Pollution Standard Index numbers from January 2014 to December 2019. Historical data were recorded and reported at the Surabaya City Environmental Service. The data were then plotted in accordance of time to determine whether the mean and variance were stationary or not. If the data variances were stationary, the transformation process would be carried out. On the other hand, if the data means were not stationary, differencing would be carried out.

The data contained a trend that had a significant Autocorrelation Function (ACF) where the initial lags first, then gradually and progressively decrease towards zero. The autocorrelation coefficient for time lag 1 was often extensive (direction 1) (Makridakis 1999).

Holt-Winters Exponential Smoothing Model Development

The numerical data in APSI are influenced by the seasons. Previous research by
Ahmad (2016) said that the fluctuation of high and low concentration was greatly influenced by the seasons. Data from Surabaya’s APSI could be considered seasonal as they were in line with the study. After all, concentration is influenced by seasonality.

After the data were deemed appropriate and forecasting was done, the next step would be choosing the proper model to predict data. The Holt-Winter Exponential Smoothing method is the appropriate method for seasonal data. This method requires three parameters, which are $\alpha$ as a constant for overall smoothing, $\beta$ as a constant smoothing trend, and $\gamma$ as a seasonal smoothing constant. Parameters were determined through trial and error until the smallest error was obtained.

**Implementation of the Holt-Winters Exponential Smoothing Model**

The model implementation begun by determining the seasonal length of the data. Next, it was determined that the length of the forecast period was in units of years. Determining the exponential smoothing value, estimating trend value, and estimating the seasonal value were the three parameters of this model.

**Calculating Forecast Error**

Not all forecasts are 100% accurate as they all allow for errors in projections. Good forecasting could minimize result errors; therefore, the forecasting method would be measured by the error rate or errors that occurs. The smaller the result errors, the better and more precise the forecasting. The measures of accuracy used to calculate errors in forecasting were Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Deviation (MSD).

The MAPE, MAD, and MSD formulas for in-sample and out-sample are:

MAPE = \( \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{Y_t - F_t}{Y_t} \right| \)

MAD = \( \frac{1}{n} \sum_{t=1}^{n} \left| Y_t - F_t \right| \)

MSD = \( \frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2 \)

Information:

- $n$ = amount of data
- $Y_t$ = actual data at time $t$
- $F_t$ = forecast data at time $t$

**Forecasting Results of the Smallest Error**

After determining the value of MAPE, MAD, and MSD, the next step was to determine a mathematical model with the lowest error rate for forecasting to predict the Surabaya’s APSI number in the next few years.

**RESULTS**

This study predicated a forecast to find the APSI number and thus provided a solution for the forecasting results. The APSI trends of Surabaya in 2014-2019 were presented in Figure 1.

![Figure 1. Surabaya City APSI Figures Graph 2014-2019](source: Surabaya City APSI Figures monthly report 2014-2019)

Time series are an estimate of the future based on the previous value of a variable and/or previous errors to find patterns in historical data series that can be used as a basis for future forecasting considerations (Irawan, 2017). The data plot formed determines the method used. The city’s APSI data plot was seasonal as seen in Figure 2.

![Figure 2. The Data Plot on the Air Pollution Standard Index for Surabaya in 2014-2019](source: Laporan bulan Januari-Desember 2014-2019)

Figure 2 shows that the APSIs had increased at a certain pattern point. They were
included as seasonal data for having repeated fluctuations even though the amplitude was not the same every year. This illustrated that the data plot had seasonal patterns.

**Data Plot Identification**

Identification of data patterns determines the appropriate forecasting method. Figure 1 shows that the city’s APSIs formed a seasonal pattern. The existence of a seasonal component in the data indicated that the data were not stationary. Data are stationary when the fluctuation is around a constant average value.

![Autocorrelation Function for C4](image)

**Figure 3.** Surabaya City Air Pollution Standard Index Autocorrelation Graph

The autocorrelation graph in Figure 3 shows that the APSIs in Surabaya were not stationary. There was one line on the ACF graph with a value that exceeded the significant regional boundary and slowly decreased to zero.

The range of values indicated that the two autocorrelation values for the APSI number data were outside the limit of their significance at 0.23. It was thus concluded that the data were not stationary. Data with a trend pattern had decreased or increased over time (Son, 2015).

![Box-Cox Plot of Angka ISP4 Kota Surabaya](image)

**Figure 4.** Box-Cox Plot Data of Surabaya City Air Pollution Standard Index

The results of good lambda values of Lower Confidence Limits (CL) and Upper CL were at a value of -1.13 to 1.48, while the best lambda value was 0.25. Rounded up, the value became 0.00.

Figure 4 shows that the numerical data of APSI were not stationary invariants. This result could be seen from the rounded value, which is 0. The data were stationary if the rounded value (lambda) was 1 (Son, 2015).

**Determination of Smoothing Constants by Trial and Error**

Data used within the in-sample were the 1st to 60th period, while the the out-sample were collected from the 61st to 72nd period. This was then used to validate the forecasting results of the best models found. The results of the smoothing constant identification by trial and error were presented in Table 1.

The determination of constants using trial and error can be seen in Table 1. Through trial and error calculations, it was calculated MAPE (0.216869), MAD (0.141082), and MSD (0.338572) were the best values with $\alpha$ (0.8), $\lambda$ (0.5), and $\delta$ (0.6). In this study, the MAPE was at 0.21%.

**Table 1.** Values of MAPE, MAD, MSD from In-sample and Out-sample Data

|       | In-sample                  |       | Out-sample                      |
|-------|----------------------------|-------|---------------------------------|
| Model 1 | $\alpha = 0.9$ | MAPE = 0.2183 | Model 1 | $\alpha = 0.9$ | MAPE = 1.4325 |
|       | $\lambda = 0.9$ | MAD = 0.13866 | $\lambda = 0.9$ | MAD = 0.73719 |
|       | $\delta = 0.6$ | MSD = 0.42184 | $\delta = 0.6$ | MSD = 2.34751 |
| Model 2 | $\alpha = 0.9$ | MAPE = 0.2368 | Model 2 | $\alpha = 0.9$ | MAPE = 1.8314 |
|       | $\lambda = 0.13$ | MAD = 0.1577 | $\lambda = 0.9$ | MAD = 0.93518 |
|       | $\delta = 0.5$ | MSD = 0.2877 | $\delta = 0.5$ | MSD = 1.59895 |
| Model 3 | $\alpha = 0.8$ | MAPE = 0.2163 | Model 3 | $\alpha = 0.8$ | MAPE = 1.5268 |
|       | $\lambda = 0.5$ | MAD = 0.14108 | $\lambda = 0.5$ | MAD = 0.76734 |
|       | $\delta = 0.6$ | MSD = 0.33857 | $\delta = 0.6$ | MSD = 1.77000 |

*Model 3 In-sample was the selected model with the smallest value of MAPE.*
Transformation

In-sample time series data, which variances were not stationary, were then transformed using the Box-Cox transformation method. The purpose of this was to reduce data variation then the error value of the forecast. Wei (2006) in his book theorized that the transformation was carried out with the formula of the natural logarithm (ln) or LOGTEN (number) because the rounded value is 0.

The transformation through trial and error calculations obtained the best values of MAPE (0.104355), MAD (0.00842), and MSD (0.001232) with \( \alpha \) (0.8), \( \gamma \) (0.5), and \( \delta \) (0.6). These results have proven that transforming data could reduce the error value. The MAPE value in the initial data before transformation ranged from 0.21% to 0.10%.

The data that had been transformed were then forecasted on the in-sample data to see the error value. The calculation of the smoothing constant by trial and error can be seen in Table 1. Table 1 shows that the error value became small after the transformation shown in Table 2 was carried out.

**Table 2.** Values of MAPE, MAD, MSD

| In-sample | Model 1 | \( \alpha = 0.9 \) | MAPE = 0.1053 |
|-----------|---------|------------------|---------------|
|           | \( \lambda = 0.9 \) | MAD = 0.00839 |
|           | \( \delta = 0.6 \) | MSD = 0.00155 |

| Model 2 | \( \alpha = 0.9 \) | MAPE = 0.1141 |
|         | \( \lambda = 0.09930 \) | MAD = 0.00105 |
|         | \( \delta = 0.5 \) |

| Model 3 | \( \alpha = 0.8 \) | MAPE = 0.1043 |
|         | \( \lambda = 0.5 \) | MAD = 0.00842 |
|         | \( \delta = 0.6 \) | MSD = 0.00123 |

| Out-sample | Model 1 | \( \alpha = 0.9 \) | MAPE = 1.4325 |
|            | \( \lambda = 0.9 \) | MAD = 0.73719 |
|            | \( \delta = 0.6 \) | MSD = 2.34751 |

| Model 2 | \( \alpha = 0.9 \) | MAPE = 1.8314 |
|         | \( \lambda = 0.93518 \) | MAD = 1.59895 |
|         | \( \delta = 0.5 \) |

| Model 3 | \( \alpha = 0.8 \) | MAPE = 1.5268 |
|         | \( \lambda = 0.8 \) | MAD = 0.76734 |
|         | \( \delta = 0.6 \) | MSD = 1.77000 |

*Model 3 In-sample is the selected model with the smallest value of MAPE

**Determination of the Best Equation Model**

The best model for Holt-Winters Exponential Smoothing was the smallest MAPE, MAD, or MSD values. Between in-sample and out-sample, the best model was obtained at a MAPE (0.104355), MAD (0.00842), and MSD (0.001050) with \( \alpha \) (0.8), \( \gamma \) (0.5), and \( \delta \) (0.6).

**Forecasting Results**

The model equation obtained from trial and error was used to predict the APSI figure for January 2020 to December 2020. The calculation results of the PSI numbers were presented in the figure.

**Figure 5.** Graph of the forecast data of Air Pollution Standard Index numbers in 2020

Figure 5 shows the fluctuating forecast results of the 2020 APSI scores from January to December. In 2020 the lowest APSI score was in January (48.2354), which was categorized as good. Meanwhile, the highest APSI score was in September (76.1988), which was categorized as moderate.

**DISCUSSION**

**Data Plot Identification**

The results showed that the APSI data in Surabaya formed a seasonal pattern. Moreover, a seasonal component and drastic changes in the data indicated that it was not stationary. A study conducted by Sinay (2017) explained, that stationary was a state where there was no change invariance, and the mean (mean) over time or both are not constant (no increase or decrease) at any time. It could be said that if the time series was stationary, when the mean and variance of the series were not...
affected by changes in the observed time, then it is a statistical balance.

Data from the Environmental Service, namely the monthly reports of Surabaya City APSI of 2014-2019, were obtained and showed as many as 72 periods (observations or points). According to Makridakis (1999), stationarity is the absence of drastic changes in data. In contrast, when viewed from the data plot, there were powerful or extreme data changes in a specific month.

Jatmiko (2017) said that data patterns and data stationarity could be seen from the Autocorrelation Function (ACF) chart. If the chart results dropped rapidly towards zero during the second or third break, the results did not move. However, the research results in Figure 3 illustrated the opposite. The ACF chart did not start from zero but decreased down to zero at the last lag. The graph in Figure 3 showed a repetitive movement from left to right. It illustrated the data's seasonal pattern.

If the time series data were not stationary, there were two ways to make it so, through difference or transformation. Figure 3 shows that the numerical data of APSI were not stationary in the variants. If the variant data are not stationary, then a transformation is carried out to eliminate the non-stationary aspect (Makridakis, 1999). In line with Makridakis’ research, this study performed a modification of the in-sample data to make it stationary.

**Determination of Smoothing Constants with Trial and Error**

According to Makridakis (1999), determining smoothing constants by trial and error takes a long time because all the constant values are entered. Then the most appropriate value is chosen to produce the smallest MAPE, MAD, and MSD. The disadvantage of calculations by trial and error is time-taking. However, it could produce the maximum constant value to get the smallest value from MAPE, MAD, and MSD.

This study took the ideal in-sample and out-sample models, with the smallest MAPE, MAD, and MSD values obtained with the same α, γ, and δ. If the smallest value is not received from the same α, γ, and δ, a choice of either in-sample or out-sample must be made. Decision-making resulted in out-sample because it provided a validation of smoothing constants determination to reduce uncertainty in the forecasting results (Makridakis, 1999).

Determination of constants using trial and error calculations is illustrated in Table 1. Through trial and error calculations, the results obtained were MAPE (0.216869), MAD (0.141082), and MSD (0.338572) with α (0.8), λ (0.5), and δ (0.6). This study resulted in MAPE equals to 0.21%. Therefore, the requirement of less than 10% had been met. This result indicated that the forecasting of 2020 APSI numbers from January to December was considered good. According to Makridakis’ (2002) research, the best forecasting model was that with a MAPE value of less than 10%.

**Transformation**

Time series data of APSI, both in-sample and out-sample, were not stationary in variants. This could be determined from the rounded value of the Box-Cox Plot graph, which is 0.

The Box-Cox transformation was developed by Cox and Box to normalize data and homogenize variance. The Box-Cox transformation is a powerful transformation with a single parameter of λ (lambda). The value of λ is usually between -5 to 5. The value of λ determines the relative strength of a transformation.

Wei (2006) stated that if the value of λ equals to 0, the correct transformation is done by transforming the natural logarithm (ln). This transformation occurs when the data are related to the mean following the geometric mean and time. If the average treatment is more significant, the variant also gets bigger. The homogeneity of the variants or the variety between treatments is not appropriate. For example, the APSI data related to time were reported monthly.

Following the purpose of the transformation to reduce data variation and the forecast's error value, the determination of constants used trial and error or its calculations (See Table 1). Trial and error calculations found the values of MAPE (0.104355), MAD (0.00842), and MSD (0.001232) with α (0.8), γ (0.5), and δ (0.6). From these results, it could be proven that reducing the error value by transforming data was possible. The value of MAPE in
the data before the transformation was at 0.21%. Meanwhile, after being transformed, the MAPE value was at 0.10%. These results rely on the MAPE category (Son, 2018).

In Makridakis’ (2002) research, the excellent forecasting category was when the MAPE value was within the range of <10%. The smallest value of MAPE showed that the forecasting results were close to the actual value. The MAPE value in this study, after being transformed, was 0.10%, which was considered very good.

**Determining the Best Equation Model**

This study is in line with research on the best model. The smallest value of MAPE was chosen with a value of $\alpha$ (0.8), $\gamma$ (0.5) and $\delta$ (0.6), resulting in MAPE (0.10%), MAD (0.008%), and MSD (0.001%). Previously, the smallest value of MAPE showed that the model was getting better. The best model was determined from the smallest value of MAPE. Thus, the smaller the MAPE value, the better the model. The best model of the Holt-Winters exponential smoothing was the third model among the three models that had been tested (see Table 2).

**Forecasting Results**

The results of the time series plots for the APSIs had fluctuated from year to year; therefore, it was suitable to use the multiplicative the Holt-Winter method. The multiplicative Holt-Winters method used had $\alpha$, $\gamma$, and $\delta$ parameters determined by trial and error to obtain the slightest error (MAPE, MAD, and MSD) among the three models tested. After receiving the smallest error value, the best model was determined and then used to calculate the forecast value for the next 12 months.

According to Utami (2017), the exponential smoothing forecasting method, done with new data, is given greater weight. This means that forecasting for the next period will depend on the previous one. For example, the forecast results of January 2020 depended on the data of December 2019; therefore, the forecasts of February 2020 depended on the forecast of January 2020.

The 2020 APSI forecasts in Surabaya from January to December could be seen in Figure 5. The projection results on the original data obtained the forecasting results of as many as 72 histories and points from January 2014 to December 2019.

**CONCLUSIONS AND SUGGESTIONS**

**Conclusions**

This study concludes that the time series figures of Surabaya City APSI’s in 2014-2019 proved that using the Holt-Winters method for seasonal data patterns resulted in the best model after transformation. The values of MAPE (0.104355), MAD (0.00842), MSD (0.001050), and the trial and error method produced a forecast with the smallest error value. The best model with the Holt-Winters method was generated through $\alpha$ (0.8), $\gamma$ (0.5), and $\delta$ (0.6). The lowest forecasting result in the 2020 APSI figure was of January (48.2354), while the highest was of September (76.1988).

**Suggestion**

The forecasting results for the next 12 months showed an increasing fluctuation pattern. The suggestion of this study are to reduce the number on Air Pollution Standard Index (APSI) by increasing public transportation facilities that allow less operation of private vehicles. Furthermore, policies or regulations that support people in making wiser decisions to choose transportation should also be created, in addition to making programs or activities related to plants in the city of Surabaya.

To make forecasting more accurate with a smaller degree of deviation, it is mandatory to adjust the data patterns with the forecasting method. The author also suggests using methods other than the Holt-Winters’ exponential refinement, such as predicting data that contain seasonal and trend elements to obtain a newer and more accurate model.

**REFERENCES**

Ahmad, EF and Santoso, M. 2016. Analysis of Characterization of Concentration and Air Particulate Composition (Case Study: Surabaya). Journal of Chemistry VALENSI, 2 (2), pp. 97–103

Darmawan, R., 2018. Environmental Health Risk Assessment of NO2 Ambient Level and Toll Collectors Officer’S
Health Complaints. Journal of Environmental Health, 10 (1), pp. 116.

Encik Rosalina, Sigit Sugiarto, MDHG, 2016. Holt-Winter Forecasting Method to Predict the Number of Visitors to the Library of Riau University. FMIPA Repository, 7 (1), pp. 1–8.

Gelper, S., Fried, R. and Croux, C., 2010. Robust forecasting with exponential and holt-winters smoothing. Journal of Forecasting, 29 (3), pp. 285–300.

Irawan, A. and Sutomo, AH, 2017. The incident of ARI in Pekanbaru. Journal Of Community Medicine and Public Health, 33 (1), pp. 15-32.

Jatmiko, YA, Rahayu, RL and Darmawan, G., 2017. Comparison of the accuracy of the Holt-Winters Method of Shallot Production Forecasting Results with Singular Spectrum Analysis (SSA). Mathematical Journal, 3 (1), pp. 13.

Makridakis, S., Wheelwright, S; and Victor, 1999. Methods and Applications of Forecasting. 2nd edition ed. Jakarta: Erlangga.

Makridakis, S., Wheelwright, SC; and McGee, M ... 2002. Forecasting Methods and Application. Canada: John Wiley and Sons Inc.

Government Regulation Number 41 of 199 concerning Air Pollution Control. Jakarta: President of the Republic of Indonesia.

Prahardis, R., Syauqi, D. and Akbar, SR, 2018. Implementation of Air Pollution Monitoring System Based on Air Pollution Standard Index with Finite State Machine Modeling. Journal of Information Technology Development and Computer Science, 2 (9), pp. 3128-3137

Putra, DAWS, Hartomo, KD and Tanone, R., 2018. Drought Prediction Model Using the Holt-Winters Method (Case Study: Boyolali District). Indonesian Journal of Computing and Modeling, 1 (1), pp. 36–41.

Putra, RE and Indriyani, T., 2015. Application of Association Rules with Apriori Algorithm for Air Pollutant Analysis in Surabaya. National Seminar on Indonesian Information Systems, 2 (3), pp.2–3.

Sinay, LJ, Pentury, T. and Anakotta, D., 2017. Rainfall Forecasting in Ambon City Using the Holt-Winters Exponential Smoothing Method Rainfall Forecasting in Ambon City Using the Holt- Winter Exponential Smoothing Method. Journal of Mathematics and Applied Science, 11 (2), pp. 101–108.

Utami, R. and Atmojo, S., 2017. Comparison of the Holt Exponential Smoothing and Winter Exponential Smoothing Methods for forecasting souvenir sales. Asian Scientific Journal of Information Technology, 11 (2), pp. 123.

Wei, W., 2006. Time Series Analysis: Univariate and Multivariate Methods. 2, illustrate ed. California: University of California.