PREDICTION OF JAKARTA CITY AIR QUALITY INDEX:
MODIFIED DOUBLE EXPONENTIAL SMOOTHING APPROACHES

SEN HANSUN*, ARYA WICAKSANA AND MARCEL BONAR KRISTANDA
Informatics Department
Universitas Multimedia Nusantara
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten 15811, Indonesia
*Corresponding author: seng.hansun@lecturer.umn.ac.id
arya.wicaksana@umn.ac.id; marcel.bonar@lecturer.umn.ac.id
Received January 2021; revised May 2021

ABSTRACT. Air is a necessary element for all living beings. The quality of air could affect the well-being of the people, and therefore, it is important to make sure that the air quality could be measured properly. One standard measurement of air quality is the Air Quality Index (AQI). In this study, the Jakarta City AQI, as one of the cities that suffer from bad air quality due to several factors, will be predicted using two modified hybrid prediction methods, namely the B-WEMA and the H-WEMA methods. The comparative results in two collected datasets, i.e., the Central Jakarta and the South Jakarta AQIs, showed that B-WEMA excels H-WEMA in predicting the air quality.

Keywords: AQI, B-WEMA, Double exponential smoothing, H-WEMA, Jakarta City

1. Introduction. Air Quality Index (AQI) is a measurement index that shows the quality of air in a region, to be precise, to measure how much the air in a region is free from pollution [1]. There are six groups of AQI levels, starting from the ‘Good’ level with an AQI range of 0-50, ‘Moderate’ level with an AQI range of 51-100, ‘Unhealthy for sensitive groups’ level with an AQI range of 101-150, ‘Unhealthy’ level with an AQI range of 151-200, ‘Very Unhealthy’ level with an AQI range of 201-300, and lastly, ‘Hazardous’ level with an AQI score more than 300 [2].

There are some major pollutants in the air, such as ground-level ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, and particle pollution [3]. Those pollutants can be used as key indicators in measuring the AQI of a region. One of the most commonly used in the literature is PM2.5, i.e., the particulate matter with a diameter less than 2.5μm [4-7].

As Indonesia’s capital city, Jakarta was ranked 11th as a city with the worst air quality in the world [8]. It has an average AQI value of 110, which is categorized as Unhealthy for its citizen. Therefore, as one of the cities with the most population in the world [9], there is a need to have a proper measurement to predict the AQI for Jakarta City.

There are several previous pieces of research that have put their main focus on this domain of field. Cassano et al. [10] have used two different Recurrent Neural Network (RNN) models to predict the Apulia region’s pollutants level. By applying this approach to the region, they could get the pollution levels alerts up to five days ahead. Xu and Pei [11] proposed an AQI prediction model based on error Back Propagation (BP), which was optimized by using Particle Swarm Optimization (PSO). They found that the PSO-BP model could reduce the iteration time and improve prediction results. Jiao et al. [12] used the Long Short-Term Memory (LSTM) in predicting AQI through several attributes, such as temperature, PM2.5, PM10, and SO2. The experimental results on the data provided

DOI: 10.24507/ijicic.17.04.1363
by the environmental protection department of Shanghai, China, show that LSTM could predict the AQI well.

Moreover, Ma et al. [13] also implemented a spatially transferred bi-directional LSTM network to predict the air quality at new stations with a data shortage problem. Based on the case study in Anhui, China, they found that the proposed method could achieve 35.21% lower RMSE on average in new stations. Schürholz et al. [14] have developed an artificial intelligence context-aware AQI prediction tool, called MyAQI, which was further implemented and evaluated in Melbourne Urban Area, Australia. They found that high precision levels could be reached (90-96%) in four air quality monitoring stations. Lastly, Jyothi et al. [15] studied and analyzed the air pollution in three cities (six major sites) in Kerala, India, by using the AQI. One of their findings confirmed that Particulate Matter (PM) concentration plays a significant role as a pollutant in those areas. Based on the previous researches and people’s growing awareness to have a better quality of life and environment, the prediction of AQI has become a popular subject in this domain.

As explained above, many prediction techniques can be found in the literature ranging from statistical to machine learning methods, but one of the most basic and popular in the statistical methods domain is the Moving Average (MA), which is famous for its usefulness, easiness, objectiveness, and reliability [16]. It also has been studied by many researchers, who develop many other variants of MAs, such as the B-WEMA and the H-WEMA. Both of them are relatively new methods introduced in 2016 as improved versions of the double exponential smoothing method [17,18]. It was concluded that each proposed method excels its building-block methods in terms of the prediction accuracy level.

The exponential smoothing method was chosen in this study due to its simplicity and its ability to incorporate trend and seasonality in data [19]. Specifically, there are some published works that used the exponential smoothing family methods, such as the single exponential smoothing, the double exponential smoothing, and the Holt-Winters model to predict the air quality index in a region [20-22].

In brief, the goal of this study is to predict the Jakarta City AQI level by using B-WEMA and H-WEMA methods. The prediction will be based on the regions’ AQI historical data by using the PM2.5 factor. Moreover, the prediction results of both methods will be compared to each other by using MSE and MAPE criteria. The incremental contribution of this work is the modification we made in finding the associated smoothing factor constants for each method by using a brute-force approach. After a brief description of the dataset being used in this study in Section 2, both prediction methods and the error criteria will be further described in the subsequent Sections 3 to 5. The prediction results and comparative analysis will be explained in Section 6. Lastly, some concluding remarks will end the organization of this paper.

2. The Datasets. This research was started by conducting the data collection phase. The datasets being used in this study were collected from the AirNow website. AirNow (https://airnow.gov) is the legitimate website under the US Environmental Protection Agency (EPA). It collects real-time air quality measurements from over 2000 monitoring sites maintained by the local air quality agencies [23]. The data provided by AirNow had been used in many pieces of research, as can be found in the work of Chai et al. [24] and Buonocore et al. [25], to name a few. In particular, it holds the data of hourly recorded AQI values for both Central Jakarta and South Jakarta, two districts in Jakarta City. We used both datasets, which were stored on the site and could be accessed freely [26,27]. Table 1 shows the information contained in both datasets; meanwhile, Table 2 shows some records saved in the datasets, namely the Central Jakarta and the South Jakarta.
Table 1. Information on datasets

<table>
<thead>
<tr>
<th>District Representation</th>
<th>Principal Parameter</th>
<th>Number of Attributes</th>
<th>Data Records</th>
<th>Missed Records</th>
<th>Number of Used Records</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>JakartaCentral_PM2.5_2020_YTD</strong></td>
<td>Central Jakarta</td>
<td>PM2.5</td>
<td>14</td>
<td>01/01/2020 01:00 01/03/2020 00:00</td>
<td>2&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>JakartaSouth_PM2.5_2020_YTD</strong></td>
<td>South Jakarta</td>
<td>PM2.5</td>
<td>14</td>
<td>01/01/2020 01:00 01/03/2020 00:00</td>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Missed records are for 25/02/2020 16:00 and 25/02/2020 17:00

<sup>b</sup> Missed records are for 06/02/2020 16:00 and 06/02/2020 17:00

Table 2. Sample records in both datasets

<table>
<thead>
<tr>
<th>Central Jakarta</th>
<th>South Jakarta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date time</td>
<td>Date time</td>
</tr>
<tr>
<td>01/01/2020 01:00</td>
<td>01/01/2020 01:00</td>
</tr>
<tr>
<td>01/01/2020 02:00</td>
<td>01/01/2020 02:00</td>
</tr>
<tr>
<td>01/01/2020 03:00</td>
<td>01/01/2020 03:00</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>01/03/2020 00:00</td>
<td>01/03/2020 00:00</td>
</tr>
<tr>
<td>AQI</td>
<td>AQI</td>
</tr>
<tr>
<td>119</td>
<td>156</td>
</tr>
<tr>
<td>79</td>
<td>136</td>
</tr>
<tr>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>69</td>
<td>86</td>
</tr>
</tbody>
</table>

Following this section, two hybrid moving average methods will be discussed, followed by the explanation of two error measurement criteria, i.e., MSE and MAPE.

3. The Brown’s Weighted Exponential Moving Average Method. The first hybrid algorithm being discussed here is the B-WEMA method. It was first introduced in 2016 in a publication by Hansun [17]. Since then, B-WEMA has been accepted and used in different kinds of applications, such as Foreign Exchange forecasting [28], foreign tourist arrivals prediction [29], and the rate of return of a stock composite index data [30].

Basically, this method combines the weighting factor formula used in the Weighted Moving Average (WMA) with the calculation procedure of Brown’s DES. The detailed modified procedure can be written as three general steps as follows.

Step 1) Calculate the base value, \( B_t \), using Equation (1).

\[
B_t = \frac{n P_m + (n-1)P_{m-1} + \cdots + 2P_{(m-n+2)} + P_{(m-n+1)}}{n + (n-1) + \cdots + 2 + 1}
\]  

Step 2) By using the B-DES procedure, find the forecasting value with four iterative steps as follows.

2a) Set the initial values for \( S'_t \) (single-smoothed series) and \( S''_t \) (double-smoothed series) as formulized in Equation (2).

\[
S'_{t-1} = S''_{t-1} = B_t
\]  

2b) Find the next values for both smoothed series by using \( \alpha \), as shown in Equation (3) and Equation (4) below. In this study, to find the smoothing factor constant, \( \alpha \), we employ a brute-force approach, which iteratively runs for 100 times ranging from 0 to 1.

\[
S'_t = \alpha Y_t + (1-\alpha)S'_{t-1}
\]  

\[
S''_t = \alpha S'_t + (1-\alpha)S''_{t-1}
\]  

2c) Find the estimated level (\( L_t \)) and trend (\( T_t \)) at time \( t \) using Equation (5) and Equation (6).

\[
L_t = 2S'_t - S''_{t-1}
\]
\[ T_t = \frac{\alpha}{1-\alpha} (S_t' - S_{t-1}''') \]  
(6)

2d) Find the forecasted value for \( Y_{t+k} \) for any \( k \geq 1 \), as shown in Equation (7).

\[ F_{t+k} = L_t + kT_t \]  
(7)

Here, \( Y_t \) denotes the actual observation value at time \( t \) and \( F_t \) denotes the forecasted value at time \( t \).

Step 3) Repeat Steps 1) and 2) until each data point has been gone through.

4. **The Holt’s Weighted Exponential Moving Average Method.** H-WEMA is another type of double exponential smoothing method. It is an improved version from the combination of WMA and Holt’s DES methods as described in [18]. It has been applied to predicting the large capital stock [31], the domestic tourist arrivals [32], the number of natural disaster risk predictions [33], and the case notification rate of tuberculosis disease [34].

Similar to B-WEMA, H-WEMA consists of three modified general steps as follows.

Step 1) Calculate the base value, \( B_t \), using Equation (1) as in B-WEMA.

Step 2) By using the H-DES procedure, find the forecasting value with three iterative steps as follows.

2a) Set the initial values for \( L_t \) and \( T_t \) as shown in Equation (8) and Equation (9).

\[ L_{t-1} = B_{t-1} \]  
(8)

\[ T_{t-1} = B_t - B_{t-1} \]  
(9)

2b) Find the following values for both smoothed series using Equation (10) and Equation (11). Similar to the modification in B-WEMA, to find the smoothing factor constants, \( \alpha \) and \( \beta \), we employ a brute-force approach, which iteratively runs for 100 times ranging from 0 to 1.

\[ L_t = \alpha A_t + (1-\alpha)(L_{t-1} + T_{t-1}) \]  
(10)

\[ T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \]  
(11)

2c) Find the forecasted value by using Equation (7) as can be found in B-WEMA.

Step 3) Repeat Steps 1) and 2) until each data point has been gone through.

5. **The Error Criteria.** There are two popular forecast error criteria that are used in this study, namely the MSE (Mean Square Error) and the MAPE (Mean Absolute Percentage Error). If MSE’s advantage can be drawn from its simplicity, MAPE has the advantage that it can be used across a different scale of unit measurements being used in a study.

As described in Mukhlashin and Nugraha [30], MSE can be found by using Equation (12) as follows.

\[ MSE = \frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2 \]  
(12)

where \( n \) refers to the total number of data, \( Y_t \) is the real value of data, and \( F_t \) is the forecasted value of data.

Meanwhile, MAPE can be calculated by using Equation (13) as follows [35].

\[ MAPE = \text{mean}(|100e_t/Y_t|) \]  
(13)

where \( e_t = Y_t - F_t \) and MAPE is in percentage value.
6. **Main Results.** To compare the prediction results of the Air Quality Index (AQI) in Jakarta City, specifically for Central Jakarta and South Jakarta, we used both B-WEMA and H-WEMA methods. Since the data were recorded on an hourly basis, we set the parameters for initial data as 24 with four different span values, i.e., 3, 6, 9, and 12. Then, we calculate the MSE and the MAPE values, as shown in Table 3 and Table 4, for Central Jakarta and South Jakarta, respectively.

<table>
<thead>
<tr>
<th>Span</th>
<th>Method</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-WEMA</td>
<td>H-WEMA</td>
<td>B-WEMA</td>
</tr>
<tr>
<td>3</td>
<td>129.65414</td>
<td>149.71983</td>
<td>138.13473</td>
</tr>
<tr>
<td>6</td>
<td>10.40268</td>
<td>11.62010</td>
<td>10.90634</td>
</tr>
<tr>
<td>9</td>
<td>143.01918</td>
<td>146.36821</td>
<td>145.13224</td>
</tr>
<tr>
<td>12</td>
<td>11.18721</td>
<td>11.41547</td>
<td>11.33927</td>
</tr>
</tbody>
</table>

**Table 3. Central Jakarta forecast error**

<table>
<thead>
<tr>
<th>Span</th>
<th>Method</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-WEMA</td>
<td>H-WEMA</td>
<td>B-WEMA</td>
</tr>
<tr>
<td>3</td>
<td>130.18710</td>
<td>149.85343</td>
<td>139.71281</td>
</tr>
<tr>
<td>6</td>
<td>7.12362</td>
<td>7.93280</td>
<td>7.50828</td>
</tr>
<tr>
<td>9</td>
<td>143.66295</td>
<td>147.05133</td>
<td>145.24085</td>
</tr>
<tr>
<td>12</td>
<td>7.67918</td>
<td>7.79858</td>
<td>7.72816</td>
</tr>
</tbody>
</table>

**Table 4. South Jakarta forecast error**

From those tables, it can be inferred that the best span number parameter to be used for the B-WEMA method is 3 and for the H-WEMA method are 9 and 12, since they gave the lowest MSE and MAPE forecast error values in Central Jakarta’s AQI prediction. Similarly, the best span number parameter to be applied for the B-WEMA method is 3 and for the H-WEMA method is 12 since they gave the lowest MSE and MAPE forecast error values in South Jakarta’s AQI prediction.

Moreover, by using the same parameters, we could find the best $\alpha$ value for B-WEMA and the best $\alpha$ and $\beta$ values for H-WEMA, as shown in Table 5 and Table 6. These smoothing factor constants were found by using a brute-force approach introduced in this study.

From the forecast error results, we know that B-WEMA gives better prediction results for AQI both in Central Jakarta and South Jakarta. It has lower MSE and MAPE values.

<table>
<thead>
<tr>
<th>Span</th>
<th>B-WEMA</th>
<th>H-WEMA</th>
<th>H-WEMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>3</td>
<td>0.48</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>0.42</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>0.40</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>0.39</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 5. Central Jakarta’s best parameters**
Table 6. South Jakarta’s best parameters

<table>
<thead>
<tr>
<th>Span</th>
<th>B-WEMA</th>
<th>H-WEMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.48</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>0.42</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>0.39</td>
<td>1.00</td>
</tr>
</tbody>
</table>

than the prediction results from H-WEMA. The best $\alpha$ value for B-WEMA is 0.48 for both datasets; meanwhile, the best $\alpha$ and $\beta$ values for H-WEMA are 1.00 and 0.00 for Central Jakarta and 1.00 and 0.01 for South Jakarta. The prediction graphs for each dataset by using the best parameters of both methods are shown in Figure 1 until Figure 4.

7. Conclusions. Two hybrid MA methods, namely, B-WEMA and H-WEMA, have been successfully applied to predicting the Air Quality Index (AQI) in Central Jakarta and South Jakarta regions. Moreover, based on the comparison of forecast error values by using MSE and MAPE criteria, it was found that B-WEMA could give better prediction results.
compared to H-WEMA for both datasets being used in this study. In the near future, the research’s results can be compared with other prediction methods, ranging from the moving average family, such as the SES and Holt-Winter TES, to more sophisticated and complex machine learning methods, such as the Bayes regression [36] and the recurrent neural networks [37].

Acknowledgment. This work is partially supported by Kemenristekdikti (Contract No. 1581/HD/LPPM-UMN/IV/2019) and LPPM UMN. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation of the paper.

REFERENCES


