symmetry

MDPI

Article

Predicting Daily Air Pollution Index Based on Fuzzy Time Series Markov Chain Model

Yousif Alyousifi [1,*], Mahmod Othman [1], Rajalingam Sokkalingam [1], Ibrahima Faye [2] and Petronio C. L. Silva [3]

1 Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia; mahmod.othman@utp.edu.my (M.O.); raja.sokkalingam@utp.edu.my (R.S.)
2 Center for Intelligent Signal and Imaging Research & Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia; ibrahima_faye@utp.edu.my
3 Instituto Federal do Norte de Minas – IFNMG, Januária 39400-149, Brazil; petronio.candido@gmail.com
* Correspondence: yousif_17007310@utp.edu.my

Received: 15 January 2020; Accepted: 11 February 2020; Published: 17 February 2020

Abstract: Air pollution is a worldwide problem faced by most countries across the world. Prediction of air pollution is crucial in air quality research since it is related to public health effects. The symmetry concept of fuzzy data transformation from a single point (crisp) to a fuzzy number is essential for the forecasting model. Fuzzy time series (FTS) is applied for predicting air pollution; however, it has a limitation caused by utilizing an arbitrary number of intervals. This study involves predicting the daily air pollution index using the FTS Markov chain (FTSMC) model based on a grid method with an optimal number of partitions, which can greatly develop the model accuracy for air pollution. The air pollution index (API) data, which was collected from Klang, Malaysia, is considered in the analysis. The model has been validated using three statistical criteria, which are the root mean (RMSE), the mean absolute percentage error (MAPE), and the Thiels' U statistic. Also, the model's validation has been investigated by comparison with some of the famous statistical models. The results of the proposed model demonstrated outperformed the other models. Thus, the proposed model could be a better option in air pollution forecasting that can be useful for managing air quality.

Keywords: air pollution index; fuzzy time series; grid partition method; Markov chain; Akaike information criterion

1. Introduction

Air pollution is a matter of concern among the public, particularly for those who live in mega-urban and industrial cities, which may have serious effects on humans and the natural environment in the future [1]. Air pollution forecasting is a high-priority in air quality research since it is related to public health effects and the natural environment [2–4]. The most widely important classical methods of time series are the autoregressive integrated moving average (ARIMA) models [5], the artificial neuron network (ANN) models [6–8], and the fuzzy time series (FTS) [9–17].

The FTS model is first introduced by Song and Chissom [18,19] based on a fuzzy set theory proposed by Zadeh [20]. Chen [21] developed the FTS model of Song and Chissom based on fuzzy logic group relations tables for reducing the computational complexity in the model. Huarng [22,23] developed Chen's [21] model by determining the effective length of intervals. Yolcu et al. [24] developed the ratio-based method based on a constrained optimization to select the length of intervals. Yu [25] improved a predicting model based on weighted fuzzy relations, which produced better forecasting results than the Chen [21] model. Cheng et al. [26] introduced the trend weighted FTS model for TAIEX forecasting by assign proper weights to individual fuzzy relationships. Effindy et al. [27]

modified a weighted FTS model for enrollment forecasting. They adopted the weighted model by adding the difference between the observed dataset across a midpoint of intervals. Tsaur [28] proposed the FTS model based on Markov chain, which is used for obtaining the largest probability using the transition probability matrix. He also used a random length of interval for the universe of discourse, which leads to a negative effect by abnormal observations and outliers. Sadaei et al. [29] developed a refined exponentially weighted FTS for forecasting the load data, which is developed prediction preciseness. More specifically, the effective interval length has been investigated by several studies based on different methods. For example, Huarng [22,23] proposed two methods for determining the effective length of intervals, which are based on averages and distribution. Yolcu et al. [24] developed Huarng's model [22] based on constrained optimization for determining the effective length of intervals. Eğrioğlu et al. [30] proposed a new method of fuzzy time series using a single variable constrained optimization to determine the best length of interval for the best forecasting accuracy. Chen et al. [31] proposed a new FTS forecasting model integrated with the granular computing approach and entropy method for stock price data. Talarposhti et al. [32] proposed a hybrid approach using optimization techniques and intelligence algorithm to determine the proper length of intervals for predicting the stock market. Cheng et al. [33] employed a rough set and utilized an adaptive expectation model to propose a new fuzzy time series to forecast the closing price. Rahim et al. [34] developed a type 2 FTS model using the sliding window technique for determining the appropriate length of intervals. Bose et al. [35] proposed a new partitioning method with the rough-fuzzy method for developing the fuzzy time series model.

Apart from that, Zuo et al. [36] developed a combining topological optimization technique in order to figure out the optimization problem in the product manufacturing process. Ning et al. [37] proposed a new method based on a chip formation model and an iterative gradient search method using Kalman filter algorithm. This optimization method has been used to inversely identify the Johnson-Cook model constants of ultra-fine-grained titanium. Ning and Liang [38] introduced a developed inverse identification technique for Johnson-Cook model constants based on the use of temperature and force data for predicting machining forces. The development of the model has been done by using an iterative gradient search method based on the Kalman filter algorithm. These types of optimization techniques can be adopted for improving the forecasting models.

More specifically, The FTS models have been utilized for forecasting environmental problems such as air quality 9-17], which are considered for predicting air pollution since the time series of air pollution may include uncertainty data and may not verify some of the statistical assumptions. Nevertheless, utilizing the FTS models in the field of air pollution is still very rare. For example, Cheng et al. [17] introduced a trend weighted FTS model to predict daily O3 concentrations. Dincer and Akkuş [12] predicted the SO2 concentrations based on a robust FTS model, which has provided good forecasting results. Koo et al. [39] made a comparison study using FTS and other statistical models for predicting air pollution events. They concluded that the proposed model outperformed the other models. Wang et al. [40] proposed a hybrid FTS method with data re-processing approaches for forecasting the main air pollutants. Yang et al. [41] proposed a forecasting system based on a combination of the fuzzy theory and advanced optimization algorithm for air pollution forecasting.

As previously mentioned, the FTS models have been utilized to solve various domain forecasting problems. However, several FTS have some issues, such as using an arbitrary length of intervals for the universe of discourse, repeated fuzzy relationships, or considering the weights of fuzzy logical relationships. According to the literature review above, some researchers proposed a partition method with complexity computations, and some have not evaluated their models, by comparison with the other recent models. Particularly, the FTS-based Markov chain model has a deficiency in determining the effective length of the interval, which was negatively affected by abnormal observations and outliers. Therefore, determining the optimal length of intervals and assigning the proper weights to present is an interesting issue that needs to be addressed. This motivated us to investigate the optimal partition number of the universe of discourse. This study proposes the FTS Markov chain

(FTSMC) model based on the grid method with the optimal number of the partitions of the universe of discourse to provide significantly improved performance in the model accuracy for air pollution forecasting. The major contribution of this study is to propose an improved model with an appropriate partition number and to implement the model for forecasting APIs as a new forecasting model in air quality research.

2. Preliminary

Fuzzy Time-Series Definitions

The fundamental steps for designing fuzzy time series models are defined universe of discourse U, divide U into an equal number of intervals, fuzzification, define fuzzy logic relation, determined forecasted values, and defuzzification. The main time series definitions of developed are listed below:

Definition 1. *Let $X(t)(t = 0, 1, 2, \ldots)$, a subset of real numbers, be the universe of discourse in which fuzzy sets $f_j(t)$ are defined. Let $F(t)$ be a collection of $f_1(t), f_2(t), \ldots$, then $F(t)$ is called a fuzzy time series, defined on $X(t)$* [17–19].

Definition 2. *Let $R(t, t-1)$ be the fuzzy logic relationship (FLR) between $F(t-1)$ and $F(t)$, which can be denoted as $F(t-1) \rightarrow F(t)$. For any t value, if $R(t, t-1)$ is independent of t, then*

$$R(t, t-1) = R(t-1, t-2) \tag{1}$$

In this case, $F(t)$ is called the time-invariant fuzzy time series, while otherwise called a time-variant fuzzy time series [18,19].

Definition 3. *Suppose that $F(t-1) = A_i$ and the $F(t) = A_j$. The relationship between two consecutive observations $(t-1)$ and $F(t)$, denoted to as the FLR can be defined as $A_i \rightarrow A_j$, where A_i and A_j are the left-hand side and right-hand side of the FLR respectively [17–19].*

3. Methodology

3.1. Study Area and Dataset

The air pollution index (API) data is classified based on the highest index value of five main air pollutants, namely, ozone (O_3), sulphur dioxide (SO_2), particulate matter (PM_{10}), carbon monoxide (CO_2), and nitrogen dioxide (NO_2), as shown in Figure 1 [42–45]. The API values are determined by the average indices for these five pollutant variables, and then the maximum value from these sub-indices is selected as the API value [3]. In Malaysia, the air pollution index (API) has been adopted as a measure of air pollution conditions. The API is a simple number that ranges from 0 to ∞ to reflect the air quality levels that are related to the health effects [3,45].

In this study, the daily API maxima values, which were gathered from an air monitoring station located in Klang, Malaysia, are considered in the analysis. The city of Klang is located nearly 32 km to the west side of Kuala Lumpur and covers a land area of about 573 km^2, as shown in Figure 2. The API dataset is divided into a training dataset, which is from the 1 January 2012 to 31 December 2013 and testing dataset, which is from the1 January 2014 to 31 December 2014. The values of API recorded at the selected monitoring station are provided by the Department of Environment of Malaysia. The total number of observations in this study is 1096. The value of API of less than 100 denotes a good air quality, while a of API greater than 100 indicates a higher degree of air pollution. The classification of states is made based on the breakpoints for API of 50, 100, 200, 300, and 300+, corresponding to good, moderate, unhealthy, very unhealthy, and hazardous states, respectively, as shown in Table 1 [3,45].

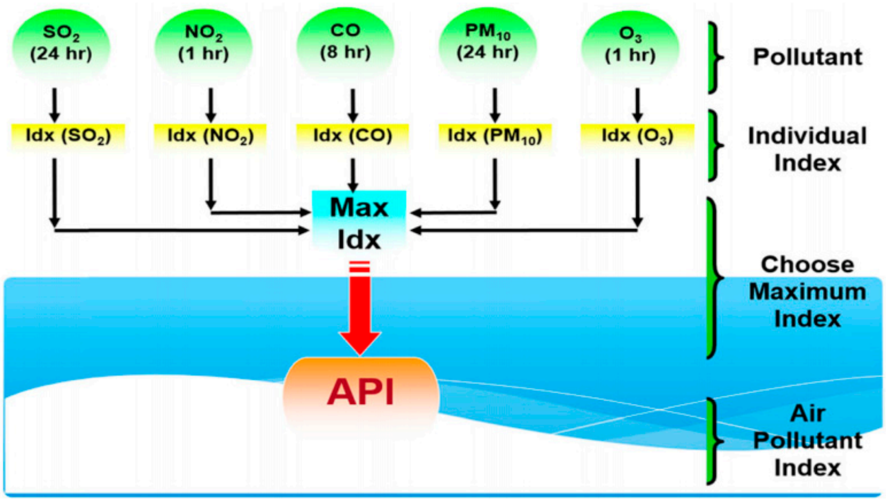

Figure 1. A method of determination for the APIs.

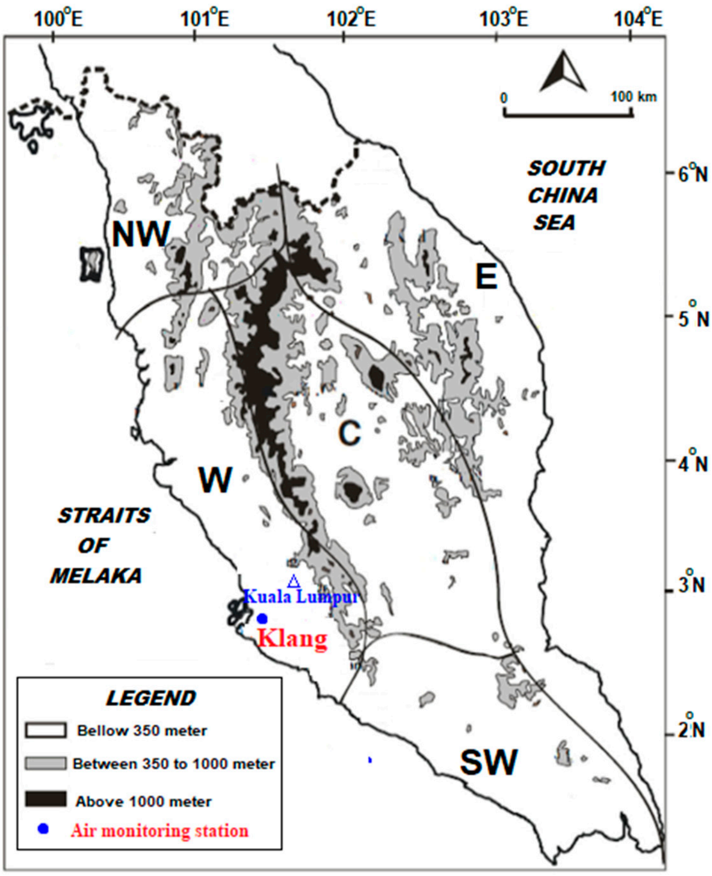

Figure 2. The air monitoring station of Klang.

Table 1. Classification of the APIs and health consequences by the Department of the environment of Malaysia [42,44].

State	Range of APIs	Air Quality Status	Health Consequences
1	[0, 50]	Good	Low pollution without any bad effect on health
2	(50, 100]	Moderate	Moderate pollution that does not pose any bad effect on health
3	(100, 200]	Unhealthy	Worsens the health condition of high-risk people that have heart and lung complications
4	(200, 300]	Very Unhealthy	Affects public health. Worsens the health condition and low tolerance of physical exercises for people with heart and lung complications
5	(300, ∞)	Hazardous	Hazardous to high-risk people and public health

3.2. Proposed Model

In this section, the simplified arithmetic operations proposed by Chen [21] and Tsaur [28] are used in the proposed algorithm (see Figure 3). The steps of the proposed model can be described as follows:

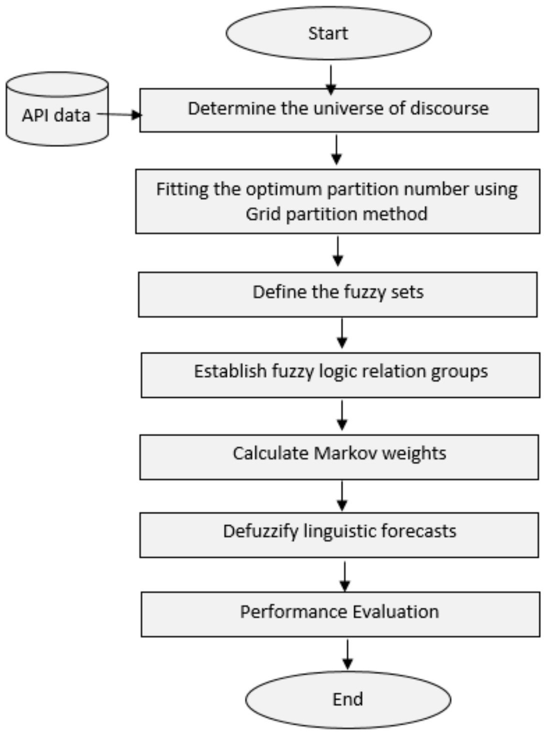

Figure 3. Flowchart of the method.

Step 1. Define the universe of discourse (U) from the available time-series data, by using the formula $U = [D_{min} - D_1, D_{max} + D_2]$, where D_{min} and D_{max} denote the minimum and the maximum value in the universe of discourse U respectively, D_1 and D_2 represent positive values.

Step 2. Partition U for the observed data using the grid partition method [21,29] based on a different number of partitions, which are 5, 6, 7, 8, ... , 50. But to avoid the redundancy, we present only 5, 10, 15, 20, 25, 30, 35, 40, and 45 numbers of partitions to determine the optimal partition number of partitions of the universe that can improve the model accuracy.

Step 3. Define the fuzzy sets A_i on U using the following equation

$$A_i = \frac{f_{A_i}(u_1)}{u_1} + \frac{f_{A_i}(u_2)}{u_2} + \ldots + \frac{f_{A_i}(u_n)}{u_n} \tag{2}$$

where f_{A_i} is the membership function of fuzzy set A_i; $f_{A_i} : U \to [0,1]$. $f_{A_i}(u_r) \in [0,1]$ and $1 \le r \le n$.

Step 4. Fuzzify the observations into fuzzy numbers based on the maximum membership value.

Step 5. Construct the fuzzy logical relationships (FLRs) and establish fuzzy logical relation groups (FLRGs) to build frequencies (count) matrix of fuzzy relation between observations.

Step 6. Generate the Markov weights (transition probability matrix) based on the frequencies of the established (FLRGs) in Step 5. The total number of states is n according to the total number of fuzzy sets. Thus, the matrix P is $P_{n \times n}$. State transition probability P_{ij}, from state A_i to state A_j. In other words, P_{ij} is the probability of observing y_{t+1} given y_t, i.e., $P_{ij} = Pr(y_{t+1} = j \big| \, y_t = i)$, which can be calculated as follows

$$P_{ij} = \frac{N_{ij}}{Ni.}, i, j = 1, 2, \ldots, n \tag{3}$$

where N_{ij} is the number of transitions from state A_i to state A_j, and $N_{i.}$ is the total number of transitions in state A_i. The transition probability matrix P is given as

$$P = \begin{bmatrix} p_{11} & p_{12} & \ldots & p_{1n} \\ p_{21} & p_{22} & & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{4}$$

where $P_{ij} \ge 0$ and $\sum_{j=1}^{n} P_{ij} = 1$.

Step 7. Calculate the forecasted values. The following rules are considered in calculating the forecasts.

Rule 1. In the case of the fuzzy logical relationship group of A_i is one-to-one, in which there only one transition for A_i (i.e., $A_i \to A_k$, with $P_{ik} = 1$ and $P_{ij} = 0, j \neq k$), then the forecasting of $F(t)$ is m_k, the midpoint of u_k, $k = 1, 2, n$, which can be calculated according to Equation (5) below

$$F\,(t+1) = m_k\, P_{ik} = m_k \tag{5}$$

Rule 2. In the case of the fuzzy logical relationship group of Ai is one-to-many, in which there are more than one transition for Ai (i.e., $Ai \to A1,\ A2,\ \ldots An, i = 1,\ 2,\ \ldots,\ n$). Thus, if the state is A_i for the actual value $Y(t)$ at time t, the forecasted value $F(t+1)$ can be determined by using Equation (6) below

$$F\,(t+1) = m_1 p_{i1} + m_1 p_{12} + \ldots + m_{i-1} p_{i(i-1)} + Y(t)\, p_{ii} + m_{i+1} p_{i(i+1)} + \ldots + m_n p_{in} \tag{6}$$

where $m_1,\ m_2,\ \ldots, m_n$ are the midpoint of $u_1,\ u_2,\ \ldots,\ u_n$ and m_i is replaced by $Y(t)$ for having information further from the state A_i at time t.

Step 8. Adjust the forecasted values by adding the differences of actual values $Y(t)$, which can adjust the forecasted values to reduce the estimated error. The adjusted forecasted values can be written by

$$\hat{F}\,(t+1) = F\,(t+1) + \text{diff}\,(Y(t)) \tag{7}$$

Step 9. Validate the model.

3.3. Model Validation

The statistical criteria used to evaluate models are MAPE, RMSE, and Thiels' U statistic, which are defined in Equations (8)–(10), respectively, where Y_i means the real data, F_i the forecasted values, and N is the total number of observations. The universe of discourse U is partitioned based on the grid method. The model is trained and tested for 5, 10, 15, 20, 25, 30, 35, and 40 number of partitions, and the results are shown in the next section.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - F_i}{Y_i}\right| \times 100 \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - F_i)^2}{N}} \tag{9}$$

$$\text{Theil's U} = \frac{\sqrt{\sum_{i=1}^{N}(Y_i - F_i)^2}}{\sqrt{\sum_{i=1}^{N} {Y_i}^2} + \sqrt{\sum_{i=1}^{N} {F_i}^2}} \tag{10}$$

4. The Implementation of the Algorithm

In this section, we will provide a result of the proposed model using the daily API data, whose plots for training data and testing data are given in Figure 4, respectively.

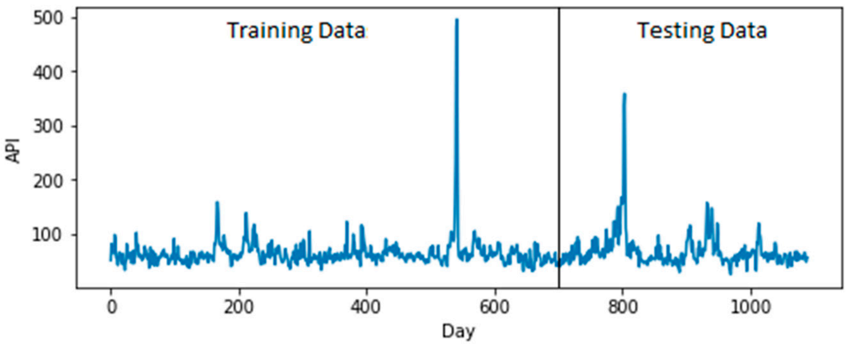

Figure 4. Time series plots of the API values of training data (2012–2013) and testing data (2014).

The implementation of the proposed model's algorithm can be done based on pyFTS [46] as follows:

Step 1. Define U for the APIs values. $U = [D_{min} - D_1,\ D_{max} + D_2]$

$$U = [25 - 5,\ 495 + 5]U = [20,\ 500]$$

Step 2. Partitioning U based on different numbers of partitions from 1 to 50. However, to prevent the redundancy where it will be too long, we have only mentioned numbers 5, 10, 15, . . . , 30 to present the partitioning as shown in Figure 5.

Step 3. Fuzzy sets are defined. Fuzzy sets A_k are determined based on the intervals u_k that already have formed using the grid method in the previous step with the function membership. Then, the fuzzy sets A_k can be written as follows using Equation (2). Table 1 reveals the fuzzy sets A_i, $(i = 1,\ 2,\ ,\ n)$. The greater the value of i indicates that the fuzzy set of API values will move from the lowest to the highest fuzzy set of API values.

Step 4. Transform APIs values into fizzy numbers and find the fuzzy logic relationships (FLRs), as shown in Table 2.

Table 3 reveals the actual API values that have been transferred to the FTS values. Then, the FLRs of these values are determined. Since u_1 has the maximum membership degree in fuzzy set A_0, observation 51 is transferred to a fuzzy set A_0. Similarly, the API values have been fuzzified.

Table 2. Values of FTS.

No	FTS Values Ai
1	$A0 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \ldots + \frac{0}{u_{27}} + \frac{0}{u_{28}} + \frac{0}{u_{29}}$
2	$A1 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \ldots + \frac{0}{u_{27}} + \frac{0}{u_{28}} + \frac{0}{u_{29}}$
3	$A2 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \ldots + \frac{0}{u_{27}} + \frac{0}{u_{28}} + \frac{0}{u_{29}}$
	.
	.
	.
29	$A28 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \ldots \frac{0.5}{u_3} + \frac{1}{u_{27}} + \frac{0.5}{u_{28}} + \frac{0}{u_{29}}$
30	$A29 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \ldots + \frac{0}{u_{27}} + \frac{0.5}{u_{28}} + \frac{1}{u_{29}}$

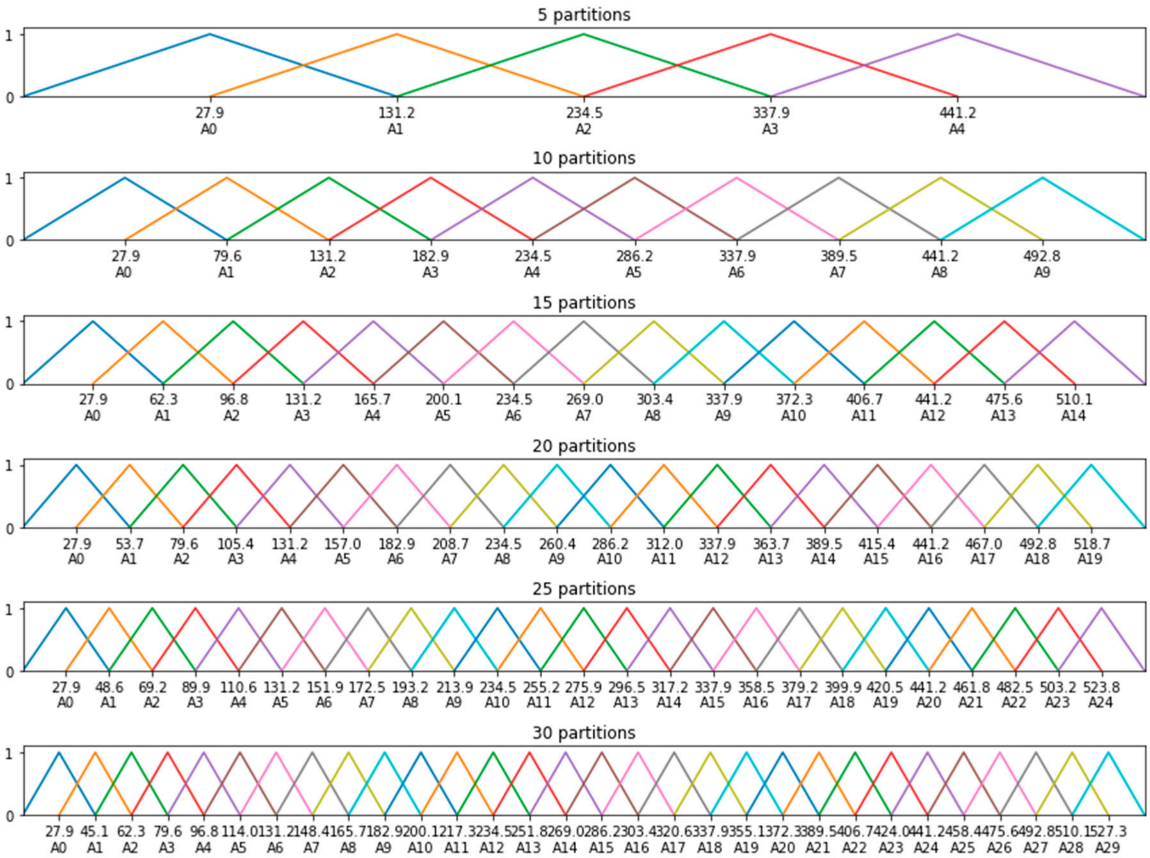

Figure 5. Partition the air pollution index data into several lengths of intervals.

Table 3. APIs as the fuzzy numbers.

N	Date	API	Fuzzy Number	Fuzzy Set Relationships
1	**2012/1/1**	51	A0	-
2	**2012/1/2**	81	A2	A0 → A2
3	**2012/1/3**	65	A1	A2 → A1
4	**2012/1/4**	70	A1	A1 → A1
5	**2012/1/5**	66	A1	A1 → A1
6	**2012/1/6**	65	A1	A1 → A1
7	**2012/1/7**	98	A3	A1 → A3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
732	**2013/12/29**	50	A0	A0 → A1
733	**2013/12/30**	61	A1	A1 → A1
734	**2013/12/31**	70	A1	A1 → A1

Step 5. Fuzzy logical relationships (FLRs) are determined, and frequencies (count) matrix of fuzzy relation between observations are determined. This step shows the FLRGs can be grouped into the fuzzy logic relationship groups (FLRGs).

It can be seen from Table 4 that thirteen groups of the FTS values are presented, which is found with several FLRs. From Table 4, transition frequency matrix or frequencies (count) matrix of fuzzy relation between observations can be determined, which could be a matrix $\boldsymbol{N}_{30\times30}$.

Table 4. FLRGs for the grid method with 30 number of partitions.

Group	Fuzzy Logical Relationships (FLRs)
G1	**A0** → (4) A0, (4) A1, (1) A3
G2	**A1** → (3) A0, (125) A1, (65) A2, (10) A3, (1) A4
G3	**A2** → (2) A0, (70) A1, (248) A2, (36) A3, (3) A4, (4) A5
G4	**A3** → A1, A2, A3, A4, A5
G5	**A4** → (2) A1, (2) A2, (11) A3, (10) A4, (1) A6, (1) A7
G6	**A5** → (2) A2, (2) A3, (3) A4, A5, A8
G7	**A6** → (1) A4, (1) A5, (1) A6
G8	**A7** → (1) A12
G9	**A8** → (2) A6, (2) A8
G10	**A12** → (1) A14
G11	**A14** → (1) A3
G12	**A26** → (1) A27
G13	**A27** → (1) A14

Step 6. Assign the Markov weights based on the matrix of frequencies from Step 5 by using Equation (4), as shown in Table 4. Then, transition process diagram could be established using the weights to visualize the Markov weighted Matrix.

Table 5 demonstrates the number of transitions of the FTS and Markov weight elements for each group. The obtained Markov weights using the grid partition method can be used for establishing the transition probability matrix $\boldsymbol{P}_{30\times30}$, which can be used for calculating the forecasting values in the next step. For instance, in the case of FLRG, it is $A_8 \rightarrow A_6, A8$. Then, value $y_{8\ 6} = 2$ and $y_{8\ 8} = 2$. Thus, $p_{8\ 6} = 1/2$ and $p_{8\ 8} = 1/2$, otherwise $p_{8\ j} = 0$.

Table 5. Markov weighted FTS based on the grid method using 30 number of partitions.

Markov Weights Elements for Each Group
A0 → A0(4/9), A1(4/9), A3(1/9)
A1 → A0(1/68), A1(125/204), A2(65/204), A3(5/102), A4(1/204)
A2 → A0(2/363), A1(70/363), 2(248/363), A3(12/121), A4(1/121), A5(4/363)
A3 → A1(2/107), A2(47/107), 3(47/107), A4(10/107), A5(10/107)
A4 → A1(2/27), A2(2/27), A3(11/27), A4(10/27), A6(1/27), (1/27)
A5 → A2 (2/9), A3 (2/9), A4 (1/9), A5(1/3), A8 (1/9)
A6 → A4 (1/3), A5 (1/3), A6 (1/3)
A7 → A12 (1)
A8 → A6 (1/2), A8 (1/2)
A12 → A14 (1)
A14 → A3 (1)
A26 → A27 (1)
A27 → A14(1)

Step 7. Calculate the forecasted values by using Equation (5) or (6) based on Markov weights. For example, the forecast value for the day (2012/1/2) is calculated by using Equation (6).

Step 8. The forecasted values are adjusted by using Equation (7). For example, in Step 7, we have found the forecast value is 56.66.

5. Model Evaluation

In this section, fitting the optimal number of partitions of the universe of discourse has been presented. In addition, to validate the proposed model, a comparison of the proposed with some existing models is provided.

5.1. Fitting the Optimal partItion Number of the Universe of Discourse

In this section, investigating the appropriate number of partitions has been done using numbers from 5 to 50 (see Table A2 in Appendix A). However, to avoid the redundancy, we present only numbers 5, 10, 15, . . . , 45. It can be seen from Table 6 and Figure 6 that the best number of partitions of API data is 30 intervals, which indicates that the proposed model produced the smallest value of MAPE, RMSE, and Theils U. This implies that the proposed model provides the best forecasting accuracy using this number of partitions as compared to the other number of partitions of APIs in terms of training and testing dataset.

Table 6. Statistical criteria for fitting the best partition number of the FTSMC model using the training dataset.

N. Partitions	RMSE	MAPE	Theils U
5	31.41	40.69	1.63
10	17.08	20.80	0.89
15	13.25	15.80	0.69
20	13.83	14.19	0.72
25	12.41	14.32	0.64
30	**11.44**	**13.15**	**0.59**
35	12.30	13.22	0.64
40	11.89	13.21	0.62
45	11.80	13.21	0.61

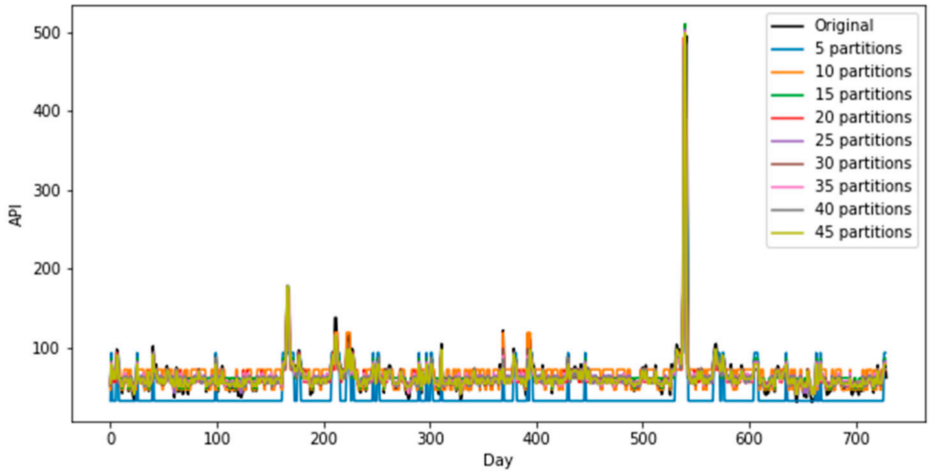

Figure 6. The FTSMC model using the grid method with different numbers of partitions using a training dataset.

More specifically, Figure 7 shows that the proposed model using the best number of partitions provides greatly improved performance in air pollution index prediction accuracy compared with the other petition numbers. This indicates that the proposed model produces accurate predicting results of the air pollution index.

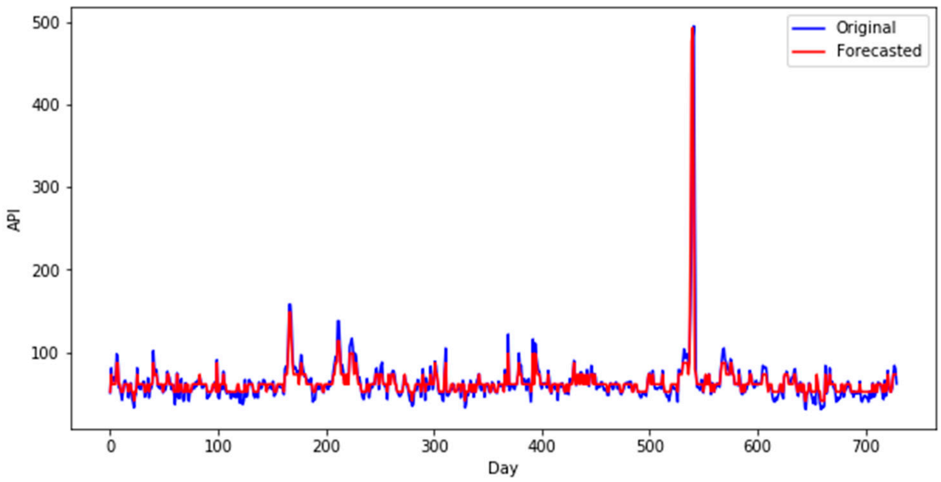

Figure 7. The FTSMC model based on the grid method with the best partition number using a training dataset.

5.2. Model's Validation

For validating the proposed model, the testing and training dataset of the APIs are used to evaluate the model performance and compare it with some of the famous existing models. Particularly, we introduce a comparison between the FTSMC model based on the optimal number of partitions and conventional FTS models that were proposed by Song & Chissom [18], Chen [21], Cheng [47], and Severiano et al. [48], which are FTS [18], CFTS [21], TWFTS [47], and HOFTS [48], respectively, to examine the performance of the proposed model. It can be seen from Table 7 and Figure 8 that the performance of the proposed model using the training dataset is very good. It has been performed with the smallest values of RMSE, MAPE, and U statistic as compared to other forecasting models. Thus, the proposed model outperformed the other forecasting models. This indicates that the proposed model is a powerful model for predicting air pollution occurrences. In addition, it could be seen from Table 8 and Figure 9 that the proposed model, using the testing dataset, outperformed the other FTS models. This implies that the model can produce a better forecasting accuracy of air pollution, which indicates that the proposed model can be modeled very well using any sort of time series.

Table 7. Statistical criteria of the proposed model and some FTS models using the training data.

Model	RMSE	MAPE	Theils U
FTS [18]	27.94	19.45	1.45
CFTS [21]	15.08	22.85	0.78
TWFTS [47]	12.84	14.28	0.62
HOFTS [48]	28.65	32.96	1.31
FTSMC The proposed model	**11.44**	**13.15**	**0.59**

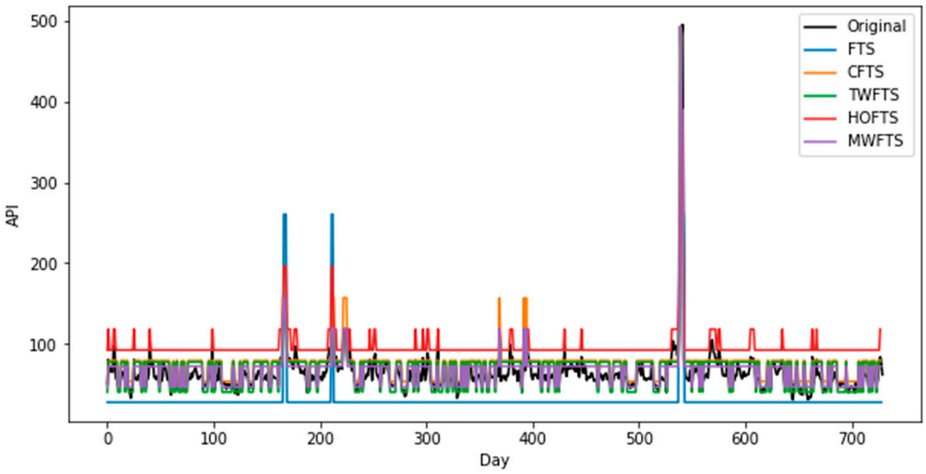

Figure 8. Comparison of the proposed model using training dataset with some FTS models proposed by Song and Chissom [18], Chen [21], Cheng [47], and Severiano et al. [48].

Table 8. Statistical criterions of the proposed model and some FTS models using the testing data.

Model	RMSE	MAPE	Theils U
FTS [18]	46.80	61.24	2.07
CFTS [21]	24.27	36.63	1.26
TWFTS [47]	18.06	18.88	0.89
HOFTS [48]	28.65	42.96	1.49
FTSMC The proposed model	**17.01**	**17.32**	**0.80**

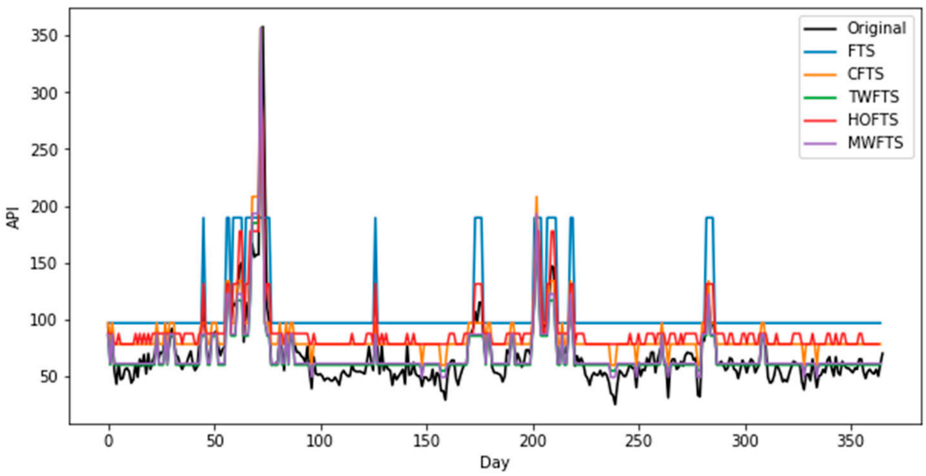

Figure 9. Comparison of the proposed model using the testing dataset with some FTS models proposed by Song and Chissom [18], Chen [21], Cheng [47], and Severiano et al. [48].

In addition, a comparison of the proposed model and some of the famous existing time series models are presented for further validation of our model. The time series models are ARMA [1], ARIMA [5,49], exponential smoothing [49], SARIMA [50], autoregressive conditional heteroskedasticity (ARCH) [51], GARCH [51], Markov chain [3], and fuzzy-ARIMA [52]. The evaluation of the model has been done based on the Akaike information criterion (AIC) [53] and Bayesian information criteria [54] using Equations (11) and (12), respectively, which are the common goodness of fit criteria for selecting the best time series models.

$$AIC = 2k - r\ \ln(L) \tag{11}$$

$$BIC = k\ \ln(r) - r\ \ln(L) \tag{12}$$

where r is the number of observations, and k is the number of parameters used in models, and $L = L(\hat{\theta})$ is the maximum value of the likelihood function of the model, which can stand for mean square error (MSE). It could be seen from Table 9 that the proposed model produced the smallest values of AIC and BIC compare to the other models. This indicates that the proposed model outperformed the other models; thus, it is an adequate model, and it could provide an accurate forecast of air pollution.

Table 9. AIC and BIC criteria of the proposed model and some of the existing models.

Prediction Model	AIC	BIC	Ranking
ARMA	9389.56	9425.39	6
ARIMA	9380.53	9415.82	4
Markov chain	9381.23	9418.71	3
ARCH	12,213.47	12,249.33	7
GARCH	12,225.42	12,261.05	8
SARIMA	9385.91	9421.28	5
Fuzzy-ARIMA	9379.94	9313.98	2
Exponential smoothing	12,942.58	12,977.24	9
FTSMC The proposed model	**9368.14**	**9406.46**	**1**

Furthermore, air pollution forecasting is based on daily API concentrations [3,39,45]. According to the time series lag test, as shown in Figure 10, we can effectively develop the performance of the fuzzy time-series Markov chain forecasting model. Based on different testing periods, the time lags of the API time series are not the same.

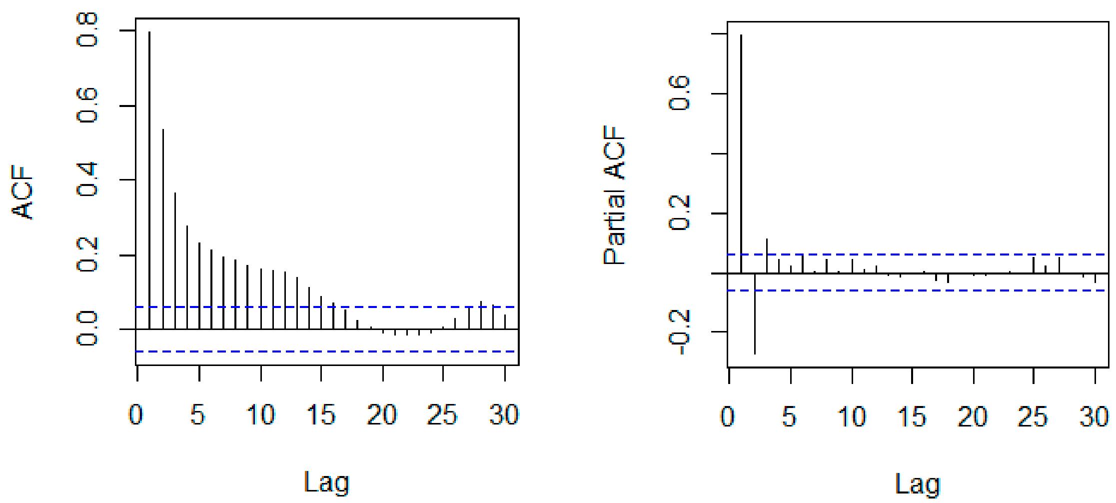

Figure 10. Training data lag test.

6. Conclusions

This study proposed the FTSMC model based on the optimal partition number for forecasting the air pollution in Malaysia using daily API data gathered from Klang for a period of three years. In this study, the Markov weights of the fuzzy logical relationships (FLRs) in the FLRG have been calculated based on the Markov transition probability. The grid partition method has been used to determine the optimal partition number of U. Then, the evaluation of the proposed model has been performed using a different number of partitions, which is chosen in order to avoid the arbitrary choosing of intervals. This is considered the first study that has ever properly defined the number of partitions in the FTSMC model. Although, the optimal number of partitions could be developing the model performance. In the proposed forecasting method, fitting the optimal number of partitions provided an improvement in the forecasting accuracy. In forecasting the daily API data, it shows that the proposed model has produced a higher prediction accuracy as compared to some FTS models. This indicated that the model could be used for forecasting air pollution data, in addition to various time-series data. For future studies, the proposed model could be performed to provide accurate results of air pollution for the sub-index variables such as PM2.5, PM10, O3, SO2, NO2, and CO, including the weather factors, such as wind speed and temperature, to provide a comprehensive exanimation of the air pollution problem. In addition, the proposed model can be developed by utilizing optimization methods such as Kalman filter, topology method, and Bayesian method, which are recommended to be employed in future works to provide an accurate forecasting model to predict air pollution. Additionally, it can be developed by combining the model with clustering and machine learning techniques in order to improve the model forecasting accuracy.

Author Contributions: Project administration, Y.A.; supervision, M.O., R.S., and I.F.; coding Y.A. and P.C.L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Grant Fundamental Research Grant Scheme (FRGS), Universiti Teknologi PETRONAS, cost center: 015MA0-021.

Acknowledgments: The authors are grateful to University Teknologi PETRONAS for providing financial support and good facilities. Additionally, they are thankful to the Department of Environment Malaysia for providing the air pollution data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. List of symbols and abbreviations in the study.

Symbol/Abbreviation	Description
$\boldsymbol{A_i}$	Fuzzy set
$\boldsymbol{U}$	Universe of discourse
$\boldsymbol{D_{min}}$	The minimum value in the universe of discourse U
$\boldsymbol{D_{max}}$	The maximum value in the universe of discourse U
$\boldsymbol{D_1}$	Positive value
$\boldsymbol{D_2}$	Positive value
$\boldsymbol{f_{A_i}}$	Membership function of fuzzy set
$\boldsymbol{u_i}$	Linguistic intervals
$\boldsymbol{F(t)}$	Fuzzy time series at time t
FLR	Fuzzy logical relationships
FLRGs	Fuzzy logical relationship groups
$\boldsymbol{m_k}$	Midpoints of the linguistic intervals u_i
$\boldsymbol{P_{ij}}$	Transition probability
$\boldsymbol{N_{ij}}$	Number of transitions
$\boldsymbol{N_{i.}}$	Total number of transitions
$\boldsymbol{P}$	Transition probability matrix
$\boldsymbol{Y(t)}$	Actual value

Table A1. *Cont.*

Symbol/Abbreviation	Description
diff $(\boldsymbol{Y}(\boldsymbol{t}))$	The difference in actual values
FTS	Fuzzy time series
FTSMC	Fuzzy time series Markov chain
ARIMA	Autoregressive integrated moving average
ANN	Artificial neuron network
SO2	Sulphur dioxide
O3	Ozone
PM_{10}	Particulate matter
CO_2	Carbon monoxide
NO_2	Nitrogen dioxide
API	Air pollution index
Thiels' U	Thiels' U statistic.
RMSE	Root mean square error
MAPE	Mean absolute percentage error
FTS	Fuzzy time series model proposed by Song
CFTS	Fuzzy time series model proposed by Chen
HOFTS	High order fuzzy time series model proposed by Severiano et al.
TWFTS	Trend weighted fuzzy time series model proposed by Cheng
AIC	Akaike information criteria
BIC	Bayesian information criteria
SARIMA	Seasonal autoregressive integrated moving average
ARMA	Autoregressive moving average
GARCH	General autoregressive conditional heteroskedasticity
ARCH	Autoregressive conditional heteroskedasticity
MSE	Mean square error
L	The maximum value of the likelihood function
ACF	Autocorrelation function
PACF	Partial autocorrelation function

Table A2. Statistical criteria for fitting the best partition number of the FTSMC model using the training dataset.

Partitions	RMSE	MAPE	Theils U
5	31.41	40.69	1.63
6	26.38	32.10	1.37
7	23.55	27.38	1.22
8	17.38	20.41	0.90
9	19.97	20.20	1.04
10	17.08	20.80	0.89
11	16.10	19.81	0.84
12	18.93	19.37	0.98
13	13.85	17.44	0.72
14	17.35	16.77	0.90
15	13.25	15.80	0.69
16	14.54	15.71	0.76
17	14.43	14.97	0.75
18	14.24	14.54	0.74
19	13.95	14.42	0.72
20	13.83	14.19	0.72
21	12.79	14.13	0.66
22	12.68	14.04	0.66
23	12.55	14.02	0.65
24	12.26	13.91	0.64
25	12.41	14.32	0.64

Table A2. *Cont.*

Partitions	RMSE	MAPE	Theils U
26	12.65	14.46	0.66
27	14.03	14.43	0.73
28	13.63	14.13	0.71
29	12.13	13.69	0.63
30	**11.44**	**13.15**	**0.59**
31	12.25	13.40	0.64
32	11.91	13.06	0.62
33	11.98	13.10	0.62
34	11.77	13.20	0.61
35	12.30	13.22	0.64
36	11.91	13.33	0.62
37	11.99	13.50	0.62
38	11.68	13.26	0.61
39	11.63	13.21	0.60
40	11.89	13.21	0.62
41	11.80	13.35	0.61
42	11.70	13.37	0.61
43	11.50	13.37	0.60
44	11.50	13.20	0.60
45	11.80	13.21	0.61
46	12.31	13.02	0.61
47	11.51	13.21	0.60
48	11.56	13.14	0.60
49	11.62	12.99	0.60

References

1. Wang, L.; Wang, J.; Tan, X.; Fang, C. Analysis of NOx Pollution Characteristics in the Atmospheric Environment in Changchun City. *Atmosphere* **2020**, *11*, 30. [CrossRef]
2. Kumar, A.; Goyal, P. Forecasting of Daily Air Quality Index in Delhi. *Sci. Total Environ.* **2001**, *409*, 5517–5523. [CrossRef] [PubMed]
3. Alyousifi, Y.; Masseran, N.; Ibrahim, K. Modeling the stochastic dependence of air pollution index data. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1603–1611. [CrossRef]
4. Rahman, N.H.A.; Lee, M.H.; Suhartono, M.T.L. Evaluation performance of time series approach for forecasting air pollution index in Johor, Malaysia. *Sains Malays.* **2016**, *45*, 1625–1633.
5. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, 1st ed.; Holden-Day: San Francesco, CA, USA, 1976.
6. David, G.S.; Rizol, P.M.S.R.; Nascimento, L.F.C. Fuzzy computational models to evaluate the effects of air pollution on children. *Rev. Paul. De Pediatr.* **2018**, *36*, 10–16. [CrossRef] [PubMed]
7. Elangasinghe, M.A.; Singhal, N.; Dirks, K.N.; Salmond, J.A. Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmos. Pollut. Res.* **2014**, *5*, 696–708. [CrossRef]
8. Rahman, N.H.A.; Lee, M.H.; Latif, M.T. Artificial neural networks and fuzzy time series forecasting: An application to air quality. *Qual. Quant.* **2015**, *49*, 2633–2647. [CrossRef]
9. Bernard, F. Fuzzy environmental Decision-making: Applications to Air Pollution. *Atmos. Environ.* **2003**, *37*, 1865–1877.
10. Heo, J.-S.; Kim, D.-S. A New Method of Ozone Forecasting Using Fuzzy Expert and Neural Network Systems. *Sci. Total Environ.* **2004**, *325*, 221–237. [CrossRef]
11. Morabito, F.C.; Versaci, M. Fuzzy Neural Identification and Forecasting Techniques to Process Experimental Urban Air Pollution Data. *Neural Netw.* **2003**, *16*, 493–506. [CrossRef]
12. Dincer, N.G.; Akkuş, Ö. A new fuzzy time series model based on robust clustering for forecasting of air pollution. *Ecol. Inform.* **2018**, *43*, 157–164. [CrossRef]

13. Aripin, A.; Suryono, S.; Bayu, S. Web based prediction of pollutant PM10 concentration using Ruey Chyn Tsaur fuzzy time series model. In Proceedings of the 2016 Conference on Fundamental and Applied Science for Advanced Technology (Confast 2016), Yogyakarta, Indonesia, 25–26 January 2016; pp. 20–46.
14. Hong, W.A.; Man, J.I.; Yili, T.A. Air Quality Index Forecast Based on Fuzzy Time Series Models. *J. Residuals Sci. Technol.* **2016**, 13.
15. Mishra, D.; Goyal, P. Neuro-fuzzy approach to forecast NO_2 pollutants addressed to air quality dispersion model over Delhi, India. *Aerosol Air Qual. Res.* **2016**, *16*, 166–174. [CrossRef]
16. Darmawan, D.; Irawan, M.I.; Syafei, A.D. Data Driven Analysis using Fuzzy Time Series for Air Quality Management in Surabaya. *Sustinere J. Environ. Sustain.* **2017**, *1*, 57–73. [CrossRef]
17. Cheng, C.H.; Huang, S.F.; Teoh, H.J. Predicting daily ozone concentration maxima using fuzzy time series based on a two-stage linguistic partition method. *Comput. Math. Appl.* **2011**, *62*, 2016–2028. [CrossRef]
18. Song, Q.; Chissom, B.S. Forecasting enrollments with fuzzy time series-Part I. *Fuzzy Sets Syst.* **1993**, *54*, 1–10. [CrossRef]
19. Song, Q.; Chissom, B.S. Forecasting enrollments with fuzzy time series-Part II. *Fuzzy Sets Syst.* **1994**, *54*, 1–10. [CrossRef]
20. Zadeh, L.A. Fuzzy sets. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]
21. Chen, S.M. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets Syst.* **1996**, *81*, 311–319. [CrossRef]
22. Huarng, K. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets Syst.* **2011**, *123*, 387–394. [CrossRef]
23. Huarng, K.; Yu, T.H.-K. Ratio-based lengths of intervals to improve fuzyy time series forecasting. *Ieee Trans. Syst. Man Cybern. Part B Cybern.* **2006**, *36*, 328–340. [CrossRef] [PubMed]
24. Yolcu, U.A. new approach based on optimization of ratio for seasonal fuzzy time series. *Iranian J. Fuzzy Syst.* **2016**, *13*, 19–36.
25. Yu, H.-K. Weighted fuzzy time series models for TAIEX forecasting. *Physica A: Stat. Mech. Appl.* **2005**, *349*, 609–624. [CrossRef]
26. Cheng, C.H.; Chen, T.L.; Teoh, H.J.; Chiang, C.H. Fuzzy time series based on adaptive expectation model for TAIEX forecasting. *Expert Syst. Appl.* **2008**, *34*, 1126–1132. [CrossRef]
27. Efendi, R.; Ismail, Z.; Deris, M.M. Improved weight Fuzzy Time Series as used in the exchange rates forecasting of US Dollar to Ringgit Malaysia. *Int. J. Comput. Intell. Appl.* **2013**, *12*, 13–29. [CrossRef]
28. Tsaur, R.C. A fuzzy time series-Markov chain model with an application to forecast the exchange rate between the Taiwan and US dolar. *Int. J. Innov. Comput. Inf. Control.* **2012**, *8*, 1349–4198.
29. Sadaei, H.J.; Enayatifar, R.; Abdullah, A.H.; Gani, A. Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search. *Inte. J. Elec. P. & Ene. Syst.* **2014**, *62*, 118–129.
30. Egrioglu, E.; Aladag, C.H.; Başaran, M.A.; Uslu, V.R.; Yolcu, U. A new approach based on the optimization of the length of intervals in fuzzy time series. *J. Intell. Fuzzy Syst.* **2011**, *22*, 15–19. [CrossRef]
31. Chen, M.Y.; Chen, B.T. A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Info.Sci.* **2018**, *294*, 227–241. [CrossRef]
32. Talarposhti, F.M.; Sadaei, H.J.; Enayatifar, R.; Guimarães, F.G.; Mahmud, M.; Eslami, T. Stock market forecasting by using a hybrid model of exponential fuzzy time series. *Inter. J. Appro. Reas.* **2019**, *70*, 79–98. [CrossRef]
33. Cheng, C.H.; Yang, J.H. Fuzzy time-series model based on rough set rule induction for forecasting stock price. *Neurocomputing* **2018**, *302*, 33–45. [CrossRef]
34. Rahim, N.F.; Othman, M.; Sokkalingam, R.; Abdul Kadir, E. Type 2 Fuzzy Inference-Based Time Series Model. *Symmetry* **2019**, *11*, 1340. [CrossRef]
35. Bose, M.; Mali, K. A novel data partitioning and rule selection technique for modeling high-order fuzzy time series. *Applied Soft Computing* **2018**, *63*, 87–96. [CrossRef]
36. Zuo, K.T.; Chen, L.P.; Zhang, Y.Q.; Yang, J. Manufacturing-and machining-based topology optimization. *Inter. J. adv. Manu. Tech.* **2006**, *27*, 531–536. [CrossRef]
37. Ning, J.; Nguyen, V.; Huang, Y.; Hartwig, K.T.; Liang, S.Y. Inverse determination of Johnson–Cook model constants of ultra-fine-grained titanium based on chip formation model and iterative gradient search. *Inter. J. Adv. Manu. Tech.* **2018**, *99*, 1131–1140. [CrossRef]

38. Ning, J.; Liang, S.Y. Inverse identification of Johnson-Cook material constants based on modified chip formation model and iterative gradient search using temperature and force measurements. *Inter. J. Adv. Manu. Tech.* **2019**, *102*, 2865–2876. [CrossRef]
39. Koo, J.W.; Wong, S.W.; Selvachandran, G.; Long, H.V. Prediction of Air Pollution Index in Kuala Lumpur using fuzzy time series and statistical models. *Air Quality, Atmosphere & Health.* **2020**, *75*, 107–111.
40. Wang, J.; Li, H.; Lu, H. Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China. *Applied Soft Computing.* **2018**, *71*, 783–799. [CrossRef]
41. Yang, H.; Zhu, Z.; Li, C.; Li, R. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Applied Soft Computing* **2019**, *87*, 105972. [CrossRef]
42. DOE Air Quality. Available online: https://www.doe.gov.my/portalv1/en/info-umum/kuality-udara/114 (accessed on 10 April 2019).
43. DOE Air Pollution Index of Malaysia. Available online: http://apims.doe.gov.my (accessed on 7 January 2020).
44. DOE Air Quality Standards. Available online: https://www.doe.gov.my/portalv1/en/info-umum/english-airquality-trend/108 (accessed on 31 January 2020).
45. Alyousifi, Y.; Ibrahim, K.; Kang, W.; Zin, W.Z.W. Markov chain modeling for air pollution index based on maximum a posteriori method. *Air Quality, Atmosphere & Health* **2019**, 1–11.
46. Silva, P.C.d.L.; Lucas, P.O.; Sadaei, H.J.; Guimarães, F.G. pyFTS: Fuzzy Time Series for Python. **2018**. [CrossRef]
47. Cheng, C.H.; Chen, T.L.; Chiang, C.H. Trend-Weighted Fuzzy Time-Series Model for TAIEX Forecasting Neural Information Processing. In *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 42, pp. 469–477.
48. Severiano, C.A.; Silva, P.C.; Sadaei, H.J.; Guimarães, F.G. Very short-term solar forecasting using fuzzy time series. In Proceedings of the 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE), Naples, Italy, 9–12 July 2017; pp. 1–6.
49. Syafei, A.D. Applying exponential state space smoothing model to short term prediction of NO2. *Jurnal Teknologi.* **2015**, 9–75. [CrossRef]
50. Lee, M.H.; Rahman, N.; Suhartono, S.; Latif, M.T.; Nor, M.; Kamisan, N. Seasonal ARIMA for forecasting air pollution index: A case study. *Am. J. Appl. Sci.* **2012**, *9*, 570–578.
51. Pahlavani, M.; Roshan, R. The comparison among ARIMA and hybrid ARIMA-GARCH models in forecasting the exchange rate of Iran. *Inter. J. Busi. Dev. Stu.* **2015**, *7*, 31–50.
52. Tseng, F.M.; Tzeng, G.H.; Yu, H.C.; Yuan, B.J. Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets Syst.* **2001**, *118*, 9–19. [CrossRef]
53. Akaike, H. A new look at the statistical model identification. *Autom Control IEEE Trans.* **1974**, *19*, 716–723. [CrossRef]
54. Konishi, S.; Kitagawa, G. Bayesian information criteria. In *Information criteria and statistical modeling*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; pp. 211–237.