J. Ris. & Ap. Mat. Vol. 07 No. 01 (2023) pp. 1-12
Jurnal Riset dan Aplikasi Matematika
e-ISSN: 2581-0154
URL: journal.unesa.ac.id/index.php/jram

PREDICTING TUBERCULOSIS MORBIDITY RATE IN INDONESIA USING WEIGHTED MARKOV CHAIN MODEL

RAHMAT AL KAFI[1*], ANGGIA ABYGAIL SIHOMBING[2], DIAN LESTARI[3]

[1,2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

*rahmat.alkafi@sci.ui.ac.id

ABSTRACT

In this work, the Weighted Markov Chain (WMC) model for time series data forecasting is examined. The Markov Chain model has been generalized in this model. In order to forecast the morbidity rate in 2021, the WMC model was used to data on tuberculosis (TB) morbidity rates in Indonesia from 2000 to 2020. The WMC model's output takes the form of a state that is represented by the interval that contains the expected morbidity. In the first stage, the simulation results of the WMC model are analyzed, with an emphasis on the number of states and the biggest step in the Markov chain. In this research, the maximum step and the number of states were combined in 10 different ways. The analysis's study revealed that the maximum step and the number of states had no impact on the predictive value of the morbidity rate. The WMC model's projections for the morbidity rate in 2021 are presented in the second stage. These forecasts are then verified by the predictions from the Simple Exponential Smoothing (SES) approach, and it is concluded that these predictions are fairly consistent.

Keywords: Chi-Square, Premium, Simple Exponential Smoothing

1 Introduction

One of the most contagious respiratory illnesses in the world, tuberculosis (TB) is brought on by the bacteria Mycobacterium tuberculosis [1]. Because of their rod-like structure and acid resistance, these bacteria are frequently referred to as Acid-Fast Bacteria. The majority of TB bacteria are commonly detected to contaminate pulmonary TB by infecting the lung parenchyma, however these bacteria can also cause extrapulmonary TB by infecting other organs like the pleura, lymph nodes, bones, and other Extrapulmonary organs [2]. Globally, there were 9.9 million TB cases by 2020 [3]. While some people in Indonesia require insurance to cover the expense of the treatment, treatment for tuberculosis patients has a relatively high cost. The amount of medical expense coverage is based on the customer's premium payment. The morbidity rate of customers who have developed the disease is one of the factors taken into account when determining premiums for customers who suffer from the condition [4].

The estimate of the likelihood that a disease would spread throughout a community is known as the morbidity rate. Illness, sickness, injury, or disability are all examples of morbidity. The morbidity rate, which is determined by the number of complaints about certain diseases, reflects the general state of public health and can tell us how widespread a disease is in a given community [5]. Insurance firms can create a health insurance policy specific to TB disease and

---

2010 *Mathematics Subject Classification*: 60G07, 60J20

Tanggal Masuk: 05-08-22; direvisi: 04-10-22; diterima: 28-04-23

set the amount of premium based on the perceived risk by knowing the TB morbidity rate in the future.

By making predictions or performing forecasting, which is the process of determining what will happen in the future using historical morbidity rate data, it is possible to determine the morbidity rate in the future. Several researchers have conducted research on forecasting morbidity rates, as in [6], [7], [8], [9], [10], [11], and [12]. Some of them used machine learning and the other used statistical approaches.

This study presents the analysis of the WMC model in predicting the TB morbidity rate in Indonesia. Many researchers have used the WMC method to predict a quantity, including Yang et al. for the probability interval of wind power in China [13], Gui and Choi for precipitation in Yangzhou [14], Yan et al. for mobile phone user mobility in China [15], and Siregar et al. for the consumer price index in Medan [16]. All previous studies show a satisfactory result. However, in this study, the Simple Exponential Smoothing (SES) method is considered to compare the actual results with the predicted outcomes of the morbidity rate in order to assess how well the WMC model predicted the morbidity rate.

2 Literature Review

2.1 Weighted Markov Chain

The extended Markov Chain model, known as the Weighted Markov Chain (WMC), is used to analyze the state of the dynamic process and its transition rule. The weight of each state is the primary distinction between the WMC model and the Markov Chain model. It is not necessary for the state weight in the WMC to be 1 or 0, but it can be estimated to be one of these values using the Pearson correlation coefficient as provided in this paper.

The steps of using WMC to predict time series data are given below [17].

1. Calculate the mean ($\bar{y}$) and the standard deviation (s) of the set of observations, respectively, which are given by Equation (1) and Equation (2).

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t \tag{1}$$

$$s = \sqrt{\frac{\sum_{t=1}^{n}(y_t-\bar{y})^2}{n-1}} \tag{2}$$

where y_t denotes the observation time t and n denotes the number of observations.

2. Classify the observations into m states using the mean and the standard deviation of the observations.
3. Construct the frequency matrix, $F = (f_{ij})$, as the following:

$$F = \begin{pmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mm} \end{pmatrix}$$

where f_{ij} denotes the number of the stochastic process $\{X_i\}$ transitioning one-step from state i to state j.

4. Construct the one-step transition probabilities matrix, $P = (p_{ij})$, and the marginal matrix $Q = (q_i)$, whose entries are given as follows

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^{m} f_{ij}} \tag{3}$$

$$q_i = \frac{\sum_{j=1}^{m} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}} \tag{4}$$

with p_{ij} denotes the probability that a process in state i will be in state j after one-step transition, q_i denotes the probability that a process makes a transition from state i.

5. Conduct the Chi-Square test to check the Markov property of the given stochastic process. The statistical hypothesis in this test is as follows:
H_0: the stochastic process has no Markov property
H_1: the stochastic process has a Markov property
The test statistic is
$$\chi^2 = 2\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}\left|\ln\frac{p_{ij}}{q_i}\right| \tag{5}$$
The critical region is:
$$\text{Reject } H_0 \text{ if } \chi^2 > \chi^2_{\alpha;(m-1)^2}$$
with α denotes the significance level and $(m-1)^2$ denotes the degree of freedom with m denotes the number of states.
6. Predict the next observation using WMC that depends on the weight of Markov Chain w_k which is given by Equation (6):
$$w_k = \frac{|r_k|}{\sum_{k=1}^{L}|r_k|} \tag{6}$$
with
$$r_k = \frac{\sum_{t=1}^{n-k}(y_t-\bar{y})(y_{t+k}-\bar{y})}{\sum_{t=1}^{n}(y_t-\bar{y})^2} \tag{7}$$
is the Pearson correlation, with $k \in \{1,2,\dots,L\}$. L denotes the maximum step used for prediction, which means a process will be in a state in the future time depending on the states at L previous times. y_t denotes the observation at time t, $\bar{y}$ denotes the average of the observed values of y_t, and n denotes the number of observations. Therefore, the WMC formula is given as follows
$$\hat{p}_{ij} = \sum_{k=1}^{L} w_k p_{ij}^{(k)} \tag{8}$$
for every $j \in \{1,2,\dots,m\}$. $p_{ij}^{(k)}$ can be obtained from k-step transition probabilities matrix P^k where $p_{ij}^{(k)}$ denotes the probability that a process in state i will be in state j after k-step transition. $\hat{p}_{ij}$ denotes the probability that the observation x_t will be in state j in the future time. The prediction result of WMC is in the form of state, denoted by j and obtained as the following:
$$\underset{j\in\{1,2,\dots,m\}}{\arg\max}\,\hat{p}_{ij}$$

2.2 Simple Exponential Smoothing

Brown first proposed the Simple Exponential Smoothing (SES) approach for forecasting in 1956. The forecast equation for the SES method is as follows [18]:
$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1}) \tag{9}$$
with α represents the smoothing parameter, $\alpha \in [0,1]$. A_t and F_t denote the actual and the fitted value of observation at time t, consecutively. However, this method can only predict a quantity for one period ahead.

3 Results and Discussion

The Table 1 below consists of Tuberculosis Morbidity Rate in Indonesia from 2000 to 2020 [19]. It shows that the values are declining over time.

Table 1: Tuberculosis morbidity rate (%) in Indonesia

Year	Morbidity Rate
2000	0.370
2001	0.369
2002	0.367
2003	0.366
2004	0.363
2005	0.360
2006	0.357
2007	0.353
2008	0.349
2009	0.345
2010	0.342
2011	0.338
2012	0.335
2013	0.332
2014	0.329
2015	0.325
2016	0.322
2017	0.319
2018	0.316
2019	0.312
2020	0.301

The simulation results of WMC model in forecasting the TB morbidity rate in Indonesia are analyzed. The morbidity rates from 2000 to 2019 are used as training data, and the morbidity rate in 2020 is used as testing data. Additionally, the WMC model's prediction result is contrasted with the SES method's prediction result. The outcome of each calculation step in the WMC algorithm is as follows:

(1) Mean $(\bar{y})$ and the standard deviation (s) of the observations from 2000 to 2019, according to Equation (1) and Equation (2), consecutively are $\bar{y} = 0.343$ and $s = 0.019152$.

(2) Divide the morbidity rates into m states (classes) based on $\bar{y} = 0.343$ and $s = 0.019152$. This study considered the case of $m = 4$ and $m = 6$. Both classes, respectively, are presented in Table 2 and Table 3.

Table 2: Morbidity rate classification for $m = 4$

State	Standard for Grading	Interval
1	$y < \bar{y} - s$	$y < 0.3243$
2	$\bar{y} - s \leq y < \bar{y}$	$0.3243 \leq y < 0.3435$
3	$\bar{y} \leq y < \bar{y} + s$	$0.3435 \leq y < 0.3626$
4	$y \geq \bar{y} + s$	$y \geq 0.3626$

Table 3: Morbidity rate classification for $m = 6$

State	Standard for Grading	Interval
1	$y < \bar{y} - s$	$y < 0.3243$
2	$\bar{y} - s \leq y < \bar{y} - 0.5s$	$0.3243 \leq y < 0.3339$
3	$\bar{y} - 0.5s \leq y < \bar{y}$	$0.3339 \leq y < 0.3435$
4	$\bar{y} \leq y < \bar{y} + 0.5s$	$0.3435 \leq y < 0.3530$
5	$\bar{y} + 0.5s \leq y < \bar{y} + s$	$0.3530 \leq y < 0.3626$
6	$y \geq \bar{y} + s$	$y \geq 0.3626$

The Tuberculosis morbidity rate in Indonesia is shown in Tables 4 and 5, which are divided into four states and six states, respectively.

Table 4: Tuberculosis morbidity rate and state transitions for $m = 4$

Year	Morbidity Rate	State	State Transition
2000	0.370	4	
2001	0.369	4	44
2002	0.367	4	44
2003	0.366	4	44
2004	0.363	4	44
2005	0.360	3	43
2006	0.357	3	33
2007	0.353	3	33
2008	0.349	3	33
2009	0.345	3	33
2010	0.342	2	32
2011	0.338	2	22
2012	0.335	2	22
2013	0.332	2	22
2014	0.329	2	22
2015	0.325	2	22
2016	0.322	1	21
2017	0.319	1	11
2018	0.316	1	11
2019	0.312	1	11

Table 5: Tuberculosis morbidity rate and state transitions for $m = 6$

Year	Morbidity Rate	State	State Transition
2000	0.370	6	
2001	0.369	6	66
2002	0.367	6	66
2003	0.366	6	66
2004	0.363	6	66
2005	0.360	5	65
2006	0.357	5	55
2007	0.353	5	55
2008	0.349	4	54
2009	0.345	4	44
2010	0.342	3	43
2011	0.338	3	33
2012	0.335	3	33
2013	0.332	2	32
2014	0.329	2	22
2015	0.325	2	22
2016	0.322	1	21
2017	0.319	1	11
2018	0.316	1	11
2019	0.312	1	11

(3) Hence by Table 4 and Table 5, the frequency matrix for $m = 4$ and $m = 6$ is given by Equation (10) and Equation (11), consecutively.

$$F = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 4 \end{pmatrix} \tag{10}$$

$$F = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \tag{11}$$

(4) According to Equation (3) and Equation (4), and Equation (10), the one-step transition probability matrix P and the marginal matrix Q for $m = 4$, respectively, are as follows:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.16667 & 0.83333 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.16 \\ 0.315789474 \\ 0.263157895 \\ 0.263157895 \end{pmatrix}$$

According to Equation (3) and Equation (4), and Equation (11), the one-step transition probability matrix P and the marginal matrix Q for $m = 6$, respectively, are as follows:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.33333 & 0.66667 & 0 & 0 & 0 & 0 \\ 0 & 0.33333 & 0.66667 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.33333 & 0.66667 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.16 \\ 0.157894737 \\ 0.157894737 \\ 0.105263158 \\ 0.157894737 \\ 0.3 \end{pmatrix}$$

(5) The Markov property is then conducting for $m = 4$ and $m = 6$, with the statistic value of χ^2 is calculated by using Equation (5). This study considered the significance level of $\alpha = 0.05$. The results of the Chi-Square test for both cases are presented by Table 6.

Table 6: The Chi-Square test for $m = 4$ and $m = 6$

m	χ^2	$\chi^2_{0.05;(m-1)^2}$	**Result**
4	40.94416722	3.325	Reject H_0
6	48.51889783	37.652	Reject H_0

Therefore by Table 6, the stochastic process of Tuberculosis morbidity rate in Indonesia for $m = 4$ and $m = 6$ follow Markov property.

(6) The next step is calculating Pearson autocorrelation coefficient r_k using Equation (7), followed by calculating the weight of Markov Chain w_k using Equation (6). This study involved one to five maximum steps ($L = 1,2,3,4,5$) in predicting future morbidity rate. The resulting r_k and w_k are presented in Table 7 and Table 8, respectively.

Table 7: Autocorrelation Coefficient of each k

k	**1**	**2**	**3**	**4**	**5**
r_k	0.864685	0.725833	0.583602	0.435330	0.289927

Table 8: The weight of the Markov Chain of each k for every $L = 1,2,3,4,5$

k	w_k				
	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
1	1	0.54365	0.39772	0.33137	0.29823
2		0.45635	0.33385	0.27816	0.25034
3			0.26843	0.22365	0.20129
4				0.16683	0.15015
5					0.09999

A prediction of the TB morbidity rate in year 2020 is found first using Equation (8) in order to warrant the accuracy of the WMC, before projecting the TB morbidity rate in year 2021. Equation (8) is constructed based on the number of states m and the maximum step L.

1. **Case of $m = 4$**

- $L = 1$

Estimating the morbidity rate for 2020 involves the state in the previous year and the one-step transition probability matrix P. The result is presented in Table 9.

Table 9: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 1$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2019	1	1	1	1	0	0	0
			$\hat{p}_{ij}$	1	0	0	0

From Table 9, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 2$

Estimating the morbidity rate for 2020 involves the state in the past two years, and the two-step transition probability matrix $P^2 = PP$. The result is presented in Table 10.

Table 10: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 2$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2019	1	1	0.54365	0.54365	0	0	0
2018	1	2	0.45635	0.45635	0	0	0
			$\hat{p}_{ij}$	1	0	0	0

From Table 10, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 3$

Estimating the morbidity rate for 2020 involves the state in the past three years, and the three-step transition probability matrix $P^3 = PPP$. The result is presented in Table 11.

Table 11: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 3$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2019	1	1	0.39772	0.39772	0	0	0
2018	1	2	0.33385	0.33385	0	0	0
2017	1	3	0.26843	0.26843	0	0	0
			$\hat{p}_{ij}$	1	0	0	0

From Table 11, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 4$

Estimating the morbidity rate for 2020 involves the state in the past four years, and the four-step transition probability matrix $P^4 = PPPP$. The result is presented in Table 12.

Table 12: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 4$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2019	1	1	0.331367	0.331367	0	0	0
2018	1	2	0.278156	0.278156	0	0	0
2017	1	3	0.223649	0.223649	0	0	0
2016	1	4	0.166828	0.166828	0	0	0
			$\hat{p}_{ij}$	1	0	0	0

From Table 12, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 5$

Estimating the morbidity rate for 2020 involves the state in the past five years, and the five-step transition probability matrix $P^5 = PPPPP$. The result is presented in Table 13.

Table 13: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 5$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2019	1	1	0.298231	0.298231	0	0	0
2018	1	2	0.250341	0.250341	0	0	0
2017	1	3	0.201285	0.201285	0	0	0
2016	1	4	0.150146	0.150146	0	0	0
2015	2	5	0.099996	0.059810	0.040186	0	0
			$\hat{p}_{ij}$	0.959814	0.040186	0	0

From Table 13, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 0.959814.

In the case of $m = 4$ it can be concluded that for every $L \in \{1,2,3,4,5\}$, the TB morbidity rate in 2020 will be on state 1 ($y < 0.3243$). This predicted state is in accordance with the actual TB morbidity rate in 2020, which is 0.301 which belongs to the interval of state 1.

2. **Case of $m = 6$**
   - $L = 1$

   Estimating the morbidity rate for 2020 involves the state in the previous year and the one-step transition probability matrix $P = (p_{ij})$. The result is presented in Table 14.

Table 14: The predicted TB morbidity rate in 2020 with $m = 6$ and $L = 1$

Year	i	k	w_k	$w_k p_{ij}$					
				$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
2019	1	1	1	1	0	0	0	0	0
			$\hat{p}_{ij}$	1	0	0	0	0	0

From Table 14, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 2$

Estimating the morbidity rate for 2020 involves the state in the past two years, and the two-step transition probability matrix $P^2 = PP$. The result is presented in Table 15.

Table 15: The predicted TB morbidity rate in 2020 with $m = 6$ and $L = 2$

Year	i	k	w_k	$w_k p_{ij}$					
				$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
2019	1	1	0.54365	0.54365	0	0	0	0	0
2018	1	2	0.45635	0.45635	0	0	0	0	0
			$\hat{p}_{ij}$	1	0	0	0	0	0

From Table 15, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 3$

Estimating the morbidity rate for 2020 involves the state in the past three years, and the three-step transition probability matrix $P^3 = PPP$. The result is presented in Table 16.

Table 16: The predicted TB morbidity rate in 2020 with $m = 6$ and $L = 3$

Year	i	k	w_k	$w_k p_{ij}$					
				$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
2019	1	1	0.39772	0.39772	0	0	0	0	0
2018	1	2	0.33385	0.33385	0	0	0	0	0
2017	1	3	0.26843	0.26843	0	0	0	0	0
			$\hat{p}_{ij}$	1	0	0	0	0	0

From Table 16, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 4$

In this case, estimating the morbidity rate for 2020 involves the state in the past four years, and the four-step transition probability matrix $P^4 = PPPP$. The result is presented in Table 17.

Table 17: The predicted TB morbidity rate in 2020 with $m = 6$ and $L = 4$

Year	i	k	w_k	$w_k p_{ij}$					
				$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
2019	1	1	0.331367	0.331367	0	0	0	0	0
2018	1	2	0.278156	0.278156	0	0	0	0	0
2017	1	3	0.223649	0.223649	0	0	0	0	0
2016	1	4	0.166828	0.166828	0	0	0	0	0
			$\hat{p}_{ij}$	1	0	0	0	0	0

From Table 17, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 1.

- $L = 5$

In this case, estimating the morbidity rate for 2020 involves the state in the past five years, and the five-step transition probability matrix $P^5 = PPPPP$. The result is presented in Table 18.

Table 18: The predicted TB morbidity rate in 2020 with $m = 6$ and $L = 5$

Year	i	k	w_k	$w_k p_{ij}$					
				$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
2019	1	1	0.298231	0.298231	0	0	0	0	0
2018	1	2	0.250341	0.250341	0	0	0	0	0
2017	1	3	0.201285	0.201285	0	0	0	0	0
2016	1	4	0.150146	0.150146	0	0	0	0	0
2015	2	5	0.099996	0.086828	0.013168	0	0	0	0
			$\hat{p}_{ij}$	0.986832	0.013168	0	0	0	0

From Table 18, the TB morbidity rate in 2020 belongs to state 1 ($y < 0.3243$) with the probability of 0.986832.

In the case of $m = 6$ it can be concluded that for every $L \in \{1,2,3,4,5\}$, the TB morbidity rate in 2020 will be on state 1 ($y < 0.3243$). This predicted state is in accordance with the actual TB morbidity rate in 2020, which is 0.301 which belongs to the interval of state 1.

According to the simulation results that came before, the predicted outcomes for the TB morbidity rate in 2020 are unaffected by the difference in the number of states m and the value of L. Afterwards, the case of $m = 4$ and $L = 5$ is considered to run WMC model in predicting TB morbidity rate for year of 2021. The predicting result is presented in Table 19.

Table 19: The predicted TB morbidity rate in 2020 with $m = 4$ and $L = 5$

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2020	1	1	0.298231	0.298231	0	0	0
2019	1	2	0.250341	0.250341	0	0	0
2018	1	3	0.201285	0.201285	0	0	0

Year	i	k	w_k	$w_k p_{ij}$			
				$j = 1$	$j = 2$	$j = 3$	$j = 4$
2017	1	4	0.150146	0.150146	0	0	0
2016	1	5	0.099996	0.099996	0	0	0
			$\hat{p}_{ij}$	1	0	0	0

Table 19 indicates that

$$\arg\max_{j\in\{1,2,3,4\}}\{1,0,0,0\} = 1$$

It means the TB morbidity rate in 2021 belongs to state 1 ($y < 0.3243$). The prediction of TB morbidity rate in 2021 by the WMC model is then compared with the result obtained from the SES method. This step aims to check the accuracy of state prediction generated by the WMC model.

The α of 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 and 1 are considered and in this study the parameter α will not be estimated since it almost gives the same result with the α of 0.9, 0.95 and 1. The best value of α is based on the smallest absolute error in 2020 and the results show that the α of 0.9, 0.95 and 1 are the best values that can be used in the forecasting equation of the SES method, with the absolute error of 0.011 and the predicted TB morbidity rate in 2021 is 0.301. It is obvious that the morbidity rate of 0.301 belongs to the interval of state 1 ($y < 0.3243$) in the WMC model. Therefore, the results of the predicted state generated by the WMC model can be considered as one of indicators in making policy that related to TB disease prevention. Moreover, the insurance company that covers TB disease can justify the insurance premium in the future given the predicted morbidity rate in 2021.

4 Conclusion

By the WMC model, the morbidity rate in 2021 is predicted to be less than 0.3243. Both the number of states and the maximum step of prediction have no impact on the predicting results. In this case study, the WMC model are quite credible for prediction since it is in congruence with the predicting result of the SES method which is commonly used to predict the data with trend.

References

[1] P. K. Nathella and S. Babu, ‘Influence of diabetes mellitus on the immunity to human tuberculosis’, *Immunology*, vol. 152, pp. 13-24, 2017.

[2] Kementerian Kesehatan Republik Indonesia, *Pedoman nasional pelayanan kedokteran tata laksana Tuberkulosis*. Jakarta: Gerdunas TB, 2020.

[3] World Health Organization, *Global Tuberculosis Report*. Geneva: WHO, 2020.

[4] J. C. Yue, HC. Wang, YY. Leong, and WP. Su, ‘Using Taiwan National health insurance database to model cancer incidence and mortality rates’, *Insur.: Math. Econ.,* vol. 78, pp. 316-324, 2018.

[5] J. M. Last, R. A. Spasoff, S. S. Harris, and M. C. Thuriaux, *A dictionary of epidemiology 4th edition*. New York: Oxford University Press, 2000.

[6] K. Kesorn, P. Ongruk, J. Chompoosri, A. Phumee, U. Thavara, A. Tawatsin, and P. Siriyasatien, ‘Morbidity rate prediction of Dengue Hemorrhagic Fever (DHF) using the Support Vector Machine and the Aedes aegypti infection rate in similar climates and

geographical areas', *PLOS,* vol. 10, pp. 1-16, 2015.

[7] W. Bekou, M. Adjobimey, O. Adjibode, G. Ade, A. D. Harries, and S. Anagonou, 'Tuberculosis case finding in Benin, 2000-2014 and beyond: A retrospective cohort and time series study', *Tuberc. Res. Treat.,* vol. 2016, pp. 1-9, 2016.

[8] Q. Song, MR. Zhao, XH. Zhou, Y. Xue, and YJ. Zeng, 'Predicting gastrointestinal infection morbidity based on environmental pollutant', *Ecol. Indic.,* vol. 82, pp. 76-81, 2017.

[9] G. J. Costa, M. J. G. Mello, C. G. Ferreira, A. Bergmann, and L. C. S. Thuler, 'Increased incidence, morbidity and mortality rates for lung cancer in women in Brazil between 2000 and 2014: An analysis of three types of sources of secondary data', *Lung Cancer,* vol. 125, pp. 77-85, 2018.

[10] R. C. M. Ribeiro, T. A. Quadros, J. J. S. Ausique, O. A. Chase, P. S. S. Campos, P. C. S. Junior, J. F. S. Almeida, and G. T. Marques, 'Forecasting incidence of tuberculosis cases in Brazil based on various univariate time-series models', *Int. J. Innov. Educ. Res.,* vol. 7, pp. 894-909, 2019.

[11] X. Fang, W. Liu, J. Ai, M. He, Y. Wu, Y. Shi, W. Shen, and C. Bao, 'Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China', *BMC Infect. Dis.,* vol. 20, pp. 1-8, 2020.

[12] D. Kapusta, S. Krivtsov, and D. Chumachenko, 'Holt's Linear model of COVID-19 morbidity forecasting in Ukraine', *CEUR Workshop Proceedings,* vol. 2917, pp. 16-25, 2021.

[13] X. Yang, X. Ma, N. Kang, and M. Maihemuti, 'Probability interval prediction of wind power based on KDE method with rough sets and Weighted Markov Chain', *IEEE Access,* vol. 6, pp. 51556-51565, 2018.

[14] Y. Gui and H. Choi, 'Prediction of precipitation based on Sliding Grey theory and Weighted Markov Chain', *2018 4th International Conference on Computational Intelligence & Communication Technology (CICT),* pp. 1-5, 2018.

[15] M. Yan, S. Li, C. A. Chan, Y. Shen, and Y. Yu, 'Mobility prediction using a Weighted Markov model based on mobile user classification', *Sensors,* vol. 21, 2021.

[16] I. P. S. Siregar, R. Widyasari, and N. H. Prasetya, 'A weighted Markov chain application to predict consumer price index when facing pandemic covid-19', *Quadratic: J. Inn. Tech. Math. & Math. Educ.,* vol. 1, pp. 41-46, 2021.

[17] R. A. Kafi, Y. R. Safitri, Y. Widyaningsih, and B. D. Handari, 'Comparison of weighted Markov chain and fuzzy time series Markov chain in forecasting stock closing price of company X', *AIP Conf. Proc.,* vol. 2168, 2019.

[18] F. Sidqi and I. D. Sumitra, 'Forecasting product selling using single exponential smoothing and double exponential smoothing methods', *IOP Conf. Ser.: Mater. Sci. Eng.,* vol. 662, 2019.

[19] The World Bank IBRD-IDA, *Incidence of Tuberculosis (per 100,000 people)*. 2021. Accessed: March 2022