

Estimation of Transmission Rate and Recovery Rate of SIR Pandemic Model Using Kalman Filter

Wahyu Sukestyastama Putra*, Afrig Aminuddin, Ibnu Hadi Purwanto, Rakhma Shafrida Kurnia, Ika Asti Astuti

Faculty of Computer Science, Universitas Amikom, Yogyakarta, Sleman, Indonesia

Abstract—COVID-19 is a global pandemic that significantly impacts all aspects. The number of victims who died makes this disease so terrible. Various policies continue to be pursued to reduce the spread and impact of COVID-19. The spread of a disease can be modeled in differential equation modeling. This differential equation modeling is known as the SIR Model. A differential equation can be expressed in a state-space model. The state-space model is a model that is widely used to design a modern control system. This research carried out the transmission rate and recovery rate estimates in the SIR pandemic model. Estimation of the transmission rate and recovery rate in this study poses a challenge to the value of the number of people confirmed as infected. The experimental result shows that the transmission and recovery rates can be estimated using the data for the infected and recovered persons. Estimates of infected and recovered people were conducted using the Kalman Filter.

Keywords—Kalman filter; pandemic; SIR model

I. INTRODUCTION

COVID-19 is a global pandemic disease that is currently challenging for various countries worldwide. The current effort is to prevent an explosion of infected cases, considering the limited medical facilities available. This condition will be more difficult if not controlled because the growth in cases of infection is exponential [1]. The spread of a disease can be modeled in differential equation modeling. This differential equation modeling is known as the SIR Model. A differential equation can be expressed in a state-space model. The state-space model is a model that is widely used to design a modern control system. Modeling the distribution of COVID-19 to the SIR state-space model is interesting because this model can control the spread of measured diseases. The advantage of controlling using state-space modeling is that the model supports MIMO (Multiple Input and Multiple Output) cases. Two essential parameters must be known in the SIR model: transmission and recovery rate.

Transmission and recovery rates are the points to observe a process of controlling a disease outbreak in this model. The system model can be divided into input, output, and process. In this study, the input in the model is a suspect person (a person who is likely to be infected). Meanwhile, state-space modeling is carried out in the process stage with the state of the infected person and recovered person. At the same time, the output is an active case of an infected person. The output becomes a dynamic system because it adds a migration factor to the control system model.

According to previous research, the SIR model is used to predict and control a pandemic [2]–[9]. However, the pandemic SIR model has a problem because the transmission rate and recovery rate in the SIR model are constant [10]–[16]. Using these constants in predicting the number of infected and recovered is unsuitable for a condition that does not restrict population movement [17]–[22]. In this study, the authors are interested in identifying the transmission and recovery rates of the SIR model in real time. The updated results of the transmission rate and recovery rate values are then used to predict the number of states in the SIR model. State estimation needs to be done because detection bias of infected people can occur in a pandemic control.

Kalman filter is an algorithm that can be used to estimate the state of the SIR model. The Kalman filter is a two-step process that involves prediction and measurement updates. In the prediction step, the Kalman filter uses the current state of the system and the system's dynamics to predict the state at the next time step. In the measurement update step, the Kalman filter uses the predicted state and the new measurement to compute a more accurate estimate of the system's state.

The rest of this paper is organized as follows. Section II presents the related works of the SIR model implementation in various pandemics. Section III describes the proposed methods used in this study. Section IV presents the result and discussion. Finally, Section V concludes this study.

II. RELATED WORKS

Research on SIR models was conducted in 2010, and an H1N1 disease distribution model was designed using the SIR Model [2]. G. Yang [3] also carried out the control design using the SIR model. Based on these two studies, it can be concluded that the pandemic can be modeled into a control system. Research related to control systems using the SIR Model continues to grow. H. Weiss [4] formulated several policy examples to prevent the spread of disease based on the SIR model. W. Chen [5] modeled the Ebola distribution model using the SIR Model. Using the SIR model, W. Huang and G. Provan [6] designed several forecasting filters. The use of the SIR model to form state space was carried out in [7] and [8] forecasted the spread of influenza using the SIR state-space model approach.

Meanwhile, related to the COVID-19 pandemic, Chen et al. [10] added an undetectable infected person factor to the SIR model for the COVID-19 disease. The model uses two types of infected people: detachable infected persons, people with symptoms, and Asymptomatic infected persons. The SIR

*Corresponding Author.

equation turns into a matrix with the order of 3×3 . The dataset was taken from the National Health Commission of the People's Republic of China (NHC) daily Outbreak Notification. The experimental result showed that the method could predict infected and recovered persons with an error of 3%.

G. Fabricius and A. Maltz [11] have identified the threshold for disease spread in the SIR model with local and global contacts. The study modeled a SIR model into the Stochastic and deterministic models. The research built space detection parameters in a region to avoid the exponential spread of disease. In comparison, [12] and [13] carried out mathematical modeling of the transmission and control of COVID-19. The method used in [12] is the stochastic transmission method, while in [13], the logistic equation is used to describe and interpret the SIR epidemic model. This study indicates that research [12] made early predictions of locations with the potential to transmit disease based on data models that had occurred in Wuhan. Whereas [13] shows that identifying the transmission rate in the SIR model can be done using the logistics equation model.

C. Tsay et al. [14] use the SEAIR model. The SEAIR model is a modification of the SEIR (Susceptible-Exposed-Infected-Recovered) model. The research built an optimal control for the model by using the parameters of social distancing, quarantining, rate of testing, recovery rate, death rate, and initial exposure. In comparison, [18] used a modified generalized Lotka-Volterra (gLV) model. It developed a control for the model using immigration, infection, recovery, dead, and control parameters. At the same time, the form of the augmented model is to add a birth factor. While the state in the model used is estimated using the Extended Kalman Filter method. In comparison, [17] uses the SEIQR (Susceptible-Exposed-Infected- Quarantine-Recovered) model. It used a fixed control parameter on the migration parameter (M). The main finding of this study was that researchers were able to develop a mathematical model that could be used to observe the dynamics of COVID-19.

A. Abuhasel et al. [19] use the SIR model to predict cases of the spread of COVID-19 in the Kingdom of Saudi Arabia (KSA). At the same time, the ARIMA model is used to predict prevalence cases. The data used in this research is daily case data in Saudi Arabia (KSA). The results show that the SIR model can predict the development of infection cases and shows that the policies taken by the government are appropriate. At the same time, the ARIMA model shows that this model is an estimation model with current and past data with a high correlation and showed a small error. In comparison, [20] used the SIR model to predict the distribution of disease in a community. This prediction helps determine what anticipation needs to be done to control the spread of COVID-19. The data used in this study are data from various countries such as China, South Korea, India, Australia, the USA, and Italy.

III. PROPOSED METHOD

The SIR Pandemic model is simple. The model consists of S (susceptible person), I (Infected person), and R (recovered person). A susceptible person is a person who has the potential to be infected with the disease. Meanwhile, infected people are

people who have been infected with the disease. The rate of change from susceptible persons to infected persons is known as the transmission rate. Meanwhile, the rate of change from an infected person to a recovered person is called the recovery rate. In this study, an estimate of the transmission rate and recovery rate of the COVID-19 epidemic was carried out. The dataset used in this study is the confirmation data for COVID-19 based on the public database.

To estimate the transmission rate and recovery rate, the SIR epidemic model is first carried out in the form of a differential equation which is written as follows,

$$\frac{dS(t)}{dt} = \frac{-\beta(t)S(t)X(t)}{n}, \quad (1)$$

$$\frac{dX(t)}{dt} = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t) \quad (2)$$

$$\frac{dR(t)}{dt} = \gamma(t)X(t) \quad (3)$$

where $\frac{dS(t)}{dt}$ is the rate of change of susceptible persons, $\frac{dX(t)}{dt}$ is the rate of change of confirmed infected persons, and $\frac{dR(t)}{dt}$ is the rate of change of people recovering from infection. A susceptible person is a person who has the potential to be infected.

The differential equation (1) to (3) is then changed to the form of the differential equation into:

$$S(t + 1) - S(t) = \frac{-\beta(t)S(t)X(t)}{n} \quad (4)$$

$$X(t + 1) - X(t) = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t) \quad (5)$$

$$R(t + 1) - R(t) = \gamma(t)X(t). \quad (6)$$

where n is the known population in an area. The differential equation (4) to (6) is then converted into the state-space model equation with the assumption that $n = S(t)$ becomes:

$$\begin{bmatrix} S(t + 1) \\ X(t + 1) \\ R(t + 1) \end{bmatrix} = \begin{bmatrix} 1 & -\beta & 0 \\ 0 & 1 + \beta - \gamma & 0 \\ 0 & \gamma & 1 \end{bmatrix} \begin{bmatrix} S(t) \\ X(t) \\ R(t) \end{bmatrix} \quad (7)$$

with:

$S(t)$ = Susceptible person.

$X(t)$ = Infected person.

$R(t)$ = Recovered person.

$\beta(t)$ = Transmission rate

$\gamma(t)$ = Recovery rate

In (7), the infected persons and the recovered person do not depend on the state of susceptible person. Thus (7) can be converted into:

$$\begin{bmatrix} X(t + 1) \\ R(t + 1) \end{bmatrix} = \begin{bmatrix} 1 + \beta - \gamma & 0 \\ \gamma & 1 \end{bmatrix} \begin{bmatrix} X(t) \\ R(t) \end{bmatrix} \quad (8)$$

with the output equation:

$$y = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X(t) \\ R(t) \end{bmatrix} \quad (9)$$

Estimates of $X(t)$ (infected person) and $R(t)$ (recovered person) were performed using the Kalman Filter method. The steps of the Kalman filter algorithm are presented in Fig. 1.

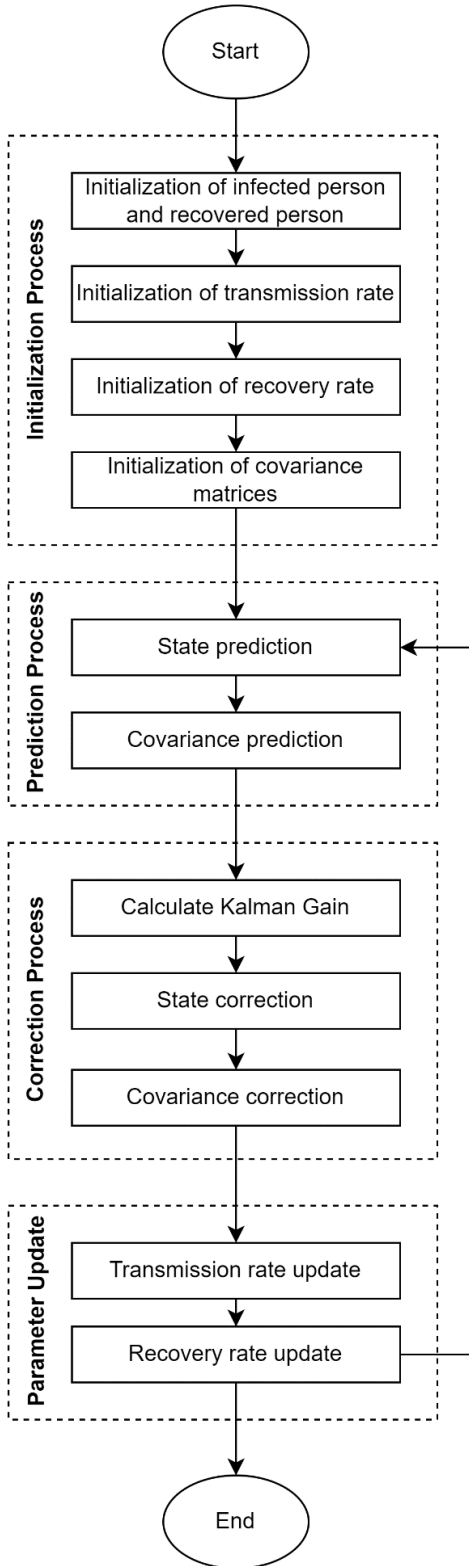


Fig. 1. The flow diagram of the proposed method.

Step 1: Initialization Process

- 1) Initialization of $X_{k-1}(t)$ and $R_{k-1}(t)$
- 2) Initialization of $\beta(t)$ using:

$$\beta(t) = \frac{(X(t)-X(t-1))+(R(t)-R(t-1))}{X(t-1)} \quad (10)$$

- 3) Initialization of $\gamma(t)$ using:

$$\gamma(t) = \frac{R(t)-R(t-1)}{X(t-1)} \quad (11)$$

- 4) Initialization of covariance matrices

Step 2: Prediction Process

- 1) State prediction using:

$$\begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix} = \begin{bmatrix} 1 + \beta(t) - \gamma(t) & 0 \\ \gamma(t) & 1 \end{bmatrix} \begin{bmatrix} X_{k-1}(t) \\ R_{k-1}(t) \end{bmatrix} \quad (12)$$

- 2) Covariance prediction using:

$$\hat{P}_k(t) = AP_{k-1}(t)A^T + Q, \quad (13)$$

with Q is a covariance noise and A is transition matrices written as:

$$A = \begin{bmatrix} 1 + \beta(t) - \gamma(t) & 0 \\ \gamma(t) & 1 \end{bmatrix} \quad (14)$$

Step 3: Correction Process

- 1) Calculate Kalman Gain using:

$$K_k = \frac{\hat{P}_k(t)H^T}{H\hat{P}_k(t)H^T + R} \quad (15)$$

which R is measurement noise, and H represents measurement matrices written as:

$$H = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \quad (16)$$

- 2) State Correction using:

$$\begin{bmatrix} X_k(t) \\ R_k(t) \end{bmatrix} = \begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix} + K_k(z_k - \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{X}_k(t) \\ \hat{R}_k(t) \end{bmatrix}) \quad (17)$$

which z_k is the measurement data

- 3) Covariance correction using:

$$P_k = (I - K_k \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix})\hat{P}_k(t). \quad (18)$$

Step 4: Parameter Update

- 1) Transmission rate $\beta(t)$ update using:

$$\beta(t) = \frac{(X_k(t)-X_k(t-1))+(R_k(t)-R_k(t-1))}{X_k(t-1)} \quad (19)$$

- 2) Recovery rate $\gamma(t)$ update using:

$$\gamma(t) = \frac{R_k(t)-R_k(t-1)}{X_k(t-1)} \quad (20)$$

Step 5: Go to Step 2

IV. EXPERIMENTAL RESULTS

Estimating the number of infected and recovered people is a topic to do because early detection of COVID-19 infection is less massive. This detection is increasingly difficult with the community's attitude, who still think that being infected with COVID-19 is a bad thing that has the impact of being excluded by the community. Awareness of the dangers of COVID-19 has not been good because it was found that people had not implemented health protocols in an orderly manner. With both conditions, matters are complicated because the movement of people who ignore health protocols increases the potential for infection and becomes an infection agent for others. This condition is amplified by the lack of mass testing for COVID-19 and undetectable infected person in the early pandemic.

In this study, the transmission rate and recovery rate are estimated using the SIR pandemic model. The estimation of the transmission rate and recovery rate in this study poses a challenge to the value of the number of people confirmed as infected. The number of people confirmed as infected is essential in determining the value of the transmission and recovery rate. An estimate of the number of infected people and the number of people recovered was carried out to overcome this problem. Estimation is done using the Kalman Filter.

The results of estimating the infected and recovered person using the Kalman Filter are depicted in Fig. 2 and Fig. 3. The estimation results show that the estimates and the confirmed data have the same trend. An error covariance matrix tuning in the estimation algorithm is performed to obtain the slightest difference between the estimated recovered person and the confirmed recovered person. The difference between the estimate and the confirmed recovered person is used as a reference.

The difference between the estimates and the confirmed recovered person is then used to determine the estimation error of the proposed algorithm. The recovered person estimation error is depicted in Fig. 4. Fig. 4 shows that a fairly large estimation error occurs in the first 15 days of the estimation. This error occurs because of the low number of cases of confirmed recovered, so the percentage of estimation error is large. After the 20th day, the estimated error of the recovered person is decreased to less than 5%.

The error of the recovered person is then used as the basis for determining the estimated error of the infected person. Given the estimated error value, an upper and lower limit of the estimate can be made. This limit is then used as the basis for the validation of the data for the confirmed infected person. If the data for the confirmed infected person are within the estimated range, then the data has good validity. However, the data has dubious validity if it is outside the estimated range. The estimated range of the infected person is depicted in Fig. 5. In Fig. 5, the lower limit of the estimate is depicted using the yellow line, and the blue line represents the upper limit of the estimate. Meanwhile, the confirmed infected person is depicted using the orange line. From the comparative data between the confirmed infected person and the estimated range of the infected person in Fig. 5, it can be seen that the position of the confirmed infected person is not always within the estimation

range. Thus there are several data of confirmed infected persons who are in doubt.

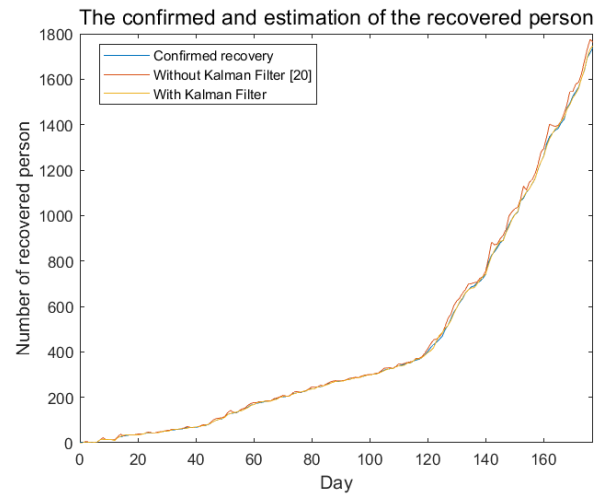


Fig. 2. The confirmed and estimation of the recovered person.

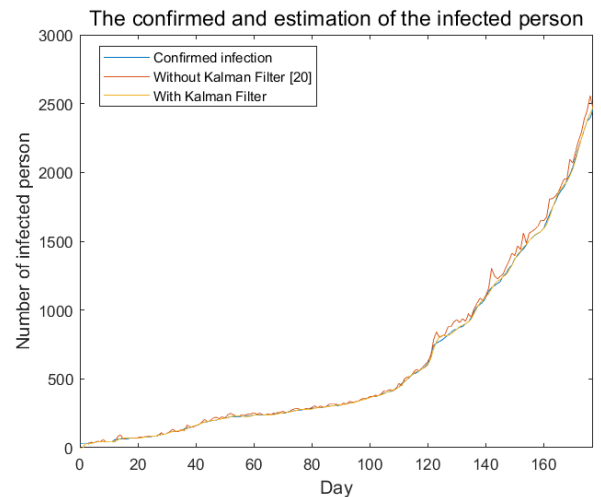


Fig. 3. The confirmed and estimation of the infected person.

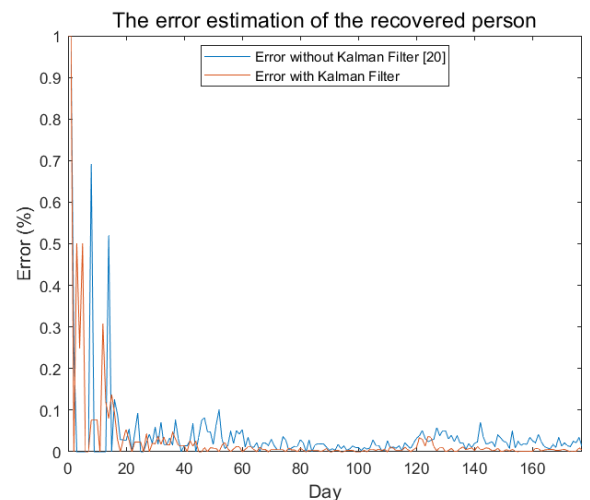


Fig. 4. The error estimation of the recovered person.

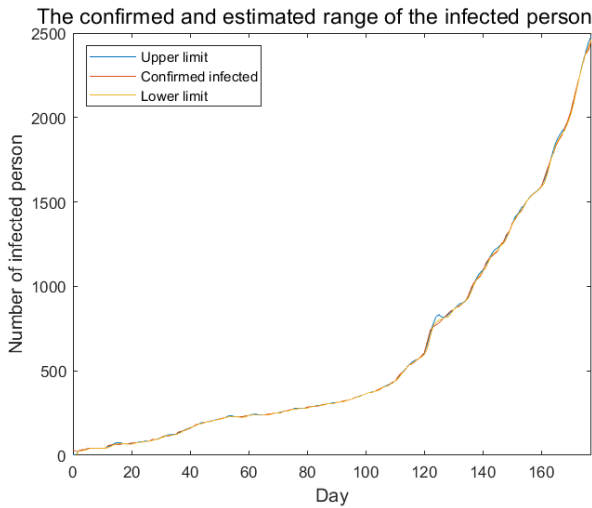


Fig. 5. The confirmed and estimated range of the infected person.

Fig. 6 shows data for confirmed infected persons outside the estimated limits. In Fig. 6, it can be seen that for the estimated day 0 to day 100, the confirmed infected person is still within the estimated limit, and if it exceeds the estimated limit, it only has a slight difference, namely a maximum of 2. However, after the 100th day, the confirmed infected person is almost always outside the estimated range. So the number of confirmed infected persons has questionable validity. On the other hand, the estimated recovered person has a reasonably good estimation error. Thus the transmission rate and recovery rate are predicted using estimated data.

Estimated infected people and estimated recovered people are used to estimate the transition rate and recovery rate. The results of the transmission rate estimation are depicted in Fig. 7. Fig. 7 shows that the transmission rate values from day 0 to day 40 experience convergence characterized by decreasing oscillations. After convergence, the transmission rate oscillates with a value above zero. This value indicates that the transmission process is still occurring. Meanwhile, the transmission rate is said to stop if the transmission rate is zero.

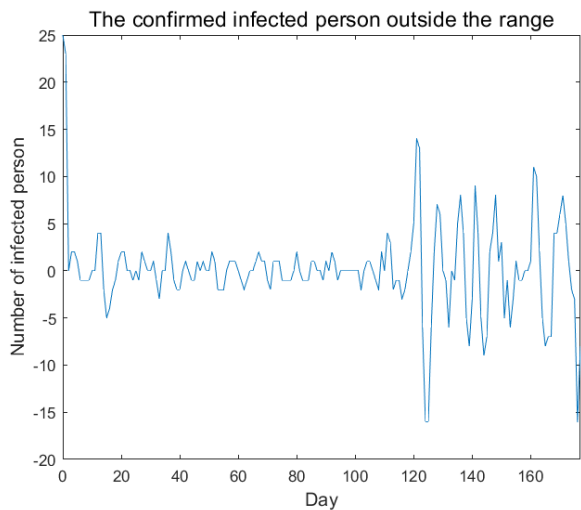


Fig. 6. The confirmed infected person outside the range.

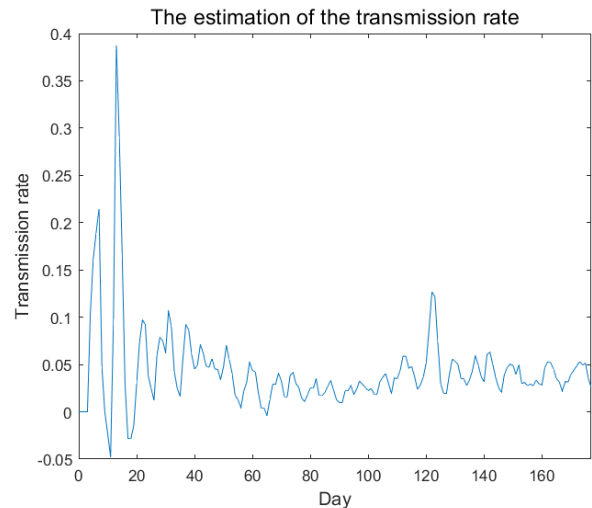


Fig. 7. The estimation of the transmission rate.

The results of the recovery rate estimation are depicted in Fig. 8. Fig. 8 shows that the recovery rate values from day 0 to day 40 experience a convergence like what happened to the transmission rate marked by reduced estimation oscillations. After convergence, the recovery rate also experiences an oscillation like what happens in the transmission rate. However, the oscillation amplitude at the recovery rate is smaller than at the transmission rate. The recovery rate of the oscillation value is also above zero. It means that the recovery process of an infected person is also in progress.

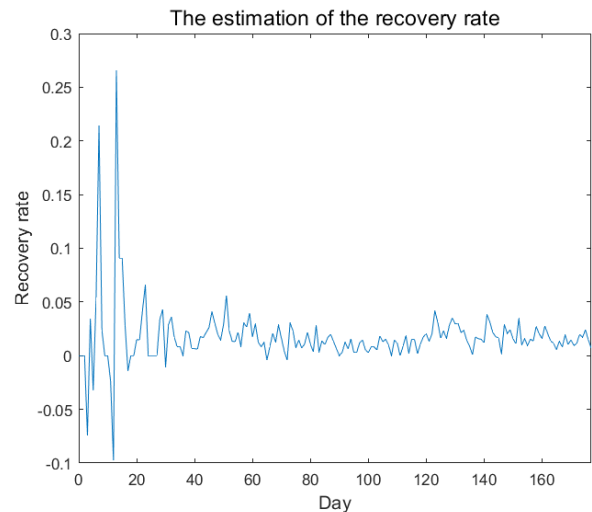


Fig. 8. The estimation of the recovery rate.

V. CONCLUSIONS

The research shows that the transmission rate and recovery rate can be estimated using the estimated data for the infected and recovered persons. Estimates of infected and recovered people are carried out using the Kalman Filter. Estimates of the infected and recovered people are carried out to address the data of the confirmed infected persons whose validity is doubtful. In further research, developing a more comprehensive pandemic model is necessary. In pandemic

modeling, SIR has not included factors for the asymptomatic infected person. Thus, research on developing pandemic models needs to be carried out.

ACKNOWLEDGMENT

This research funded by Research Department of Universitas Amikom Yogyakarta – Indonesia with contract Number 063/KONTRAK-LEMLIT/AMIKOM/ VII/2020, July, 6th 2020.

REFERENCES

- [1] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, p. 2022, 2022, doi: 10.14569/IJACSA.2022.0130665.
- [2] Y. Liu, "Investigation of Prediction and Establishment of SIR Model for H1N1 Epidemic Disease," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, 2010, pp. 1–4, doi: 10.1109/ICBBE.2010.5517654.
- [3] G. Yang, "Optimal control for an SIR epidemic model," in 2011 Chinese Control and Decision Conference (CCDC), 2011, pp. 515–518, doi: 10.1109/CCDC.2011.5968235.
- [4] H. Weiss Sir Ronald Ross, "The SIR model and the Foundations of Public Health," *Mater. MATemàtics Vol.*, vol. 17, no. 3, pp. 1887–1097, 2013.
- [5] W. Chen, "A Mathematical Model of Ebola Virus Based on SIR Model," in 2015 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration, 2015, pp. 213–216, doi: 10.1109/ICIICII.2015.135.
- [6] W. Huang and G. Provan, "An improved state filter algorithm for SIR epidemic forecasting," in Proceedings of the Twenty-second European Conference on Artificial Intelligence, 2016, pp. 524–532.
- [7] E. Sebastian, P. Victor, and T. Nadu, "a State Space Approach for Sir Epidemic Model," *Int. J. Differ. Equations*, vol. 12, no. 1, pp. 79–87, 2017.
- [8] D. Osthus, K. S. Hickmann, P. C. Caragea, D. Higdon, and S. Y. Del Valle, "Forecasting seasonal influenza with a state-space SIR model.," *Ann. Appl. Stat.*, vol. 11, no. 1, pp. 202–224, Mar. 2017, doi: 10.1214/16-AOAS1000.
- [9] F. F. Abdulloh, M. Rahardi, A. Aminuddin, S. D. Anggita, and A. Y. A. Nugraha, "Observation of Imbalance Tracer Study Data for Graduates Employability Prediction in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, pp. 169–174, Oct. 2022, doi: 10.14569/IJACSA.2022.0130820.
- [10] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons," pp. 1–10, 2020.
- [11] G. Fabricius and A. Maltz, "Exploring the threshold of epidemic spreading for a stochastic SIR model with local and global contacts," *Phys. A Stat. Mech. its Appl.*, vol. 540, p. 123208, 2020, doi: <https://doi.org/10.1016/j.physa.2019.123208>.
- [12] J. Kucharski et al., "Early dynamics of transmission and control of COVID-19: a mathematical modelling study.," *Lancet. Infect. Dis.*, Mar. 2020, doi: 10.1016/S1473-3099(20)30144-4.
- [13] M. De La Sen and A. Ibeas, "On a Sir Epidemic Model for the COVID-19 Pandemic and the Logistic Equation," *Discret. Dyn. Nat. Soc.*, vol. 2020, 2020, doi: 10.1155/2020/1382870.
- [14] Tsay, F. Lejarza, M. A. Stadtherr, and M. Baldea, "Modeling, state estimation, and optimal control for the US COVID-19 outbreak," *Sci. Rep.*, vol. 10, no. 1, p. 10711, 2020, doi: 10.1038/s41598-020-67459-8.
- [15] Aminuddin, "Android Assets Protection Using RSA and AES Cryptography to Prevent App Piracy," 2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020, pp. 461–465, Nov. 2020, doi: 10.1109/ICOIACT50329.2020.9331988.
- [16] Q. Waseem, W. Isni Sofiah Wan Din, A. Aminuddin, M. Hussain Mohammed, and R. F. Alfa Aziza, "Software-Defined Networking (SDN): A Review," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 30–35, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9972067.
- [17] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. K. Kar, "A model based study on the dynamics of COVID-19: Prediction and control," *Chaos, Solitons & Fractals*, vol. 136, p. 109889, 2020, doi: <https://doi.org/10.1016/j.chaos.2020.109889>.
- [18] B. Younes and Z. Hasan, "COVID-19: Modeling, prediction, and control," *Appl. Sci.*, vol. 10, no. 11, pp. 1–14, 2020, doi: 10.3390/app10113666.
- [19] K. A. Abuhasel, M. Khadr, and M. M. Alquraish, "Analyzing and forecasting COVID-19 pandemic in the Kingdom of Saudi Arabia using ARIMA and SIR models," *Comput. Intell.*, no. September, pp. 1–14, 2020, doi: 10.1111/coin.12407.
- [20] Cooper, A. Mondal, and C. G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," *Chaos, Solitons and Fractals*, vol. 139, p. 110057, 2020, doi: 10.1016/j.chaos.2020.110057.
- [21] Cucus, L. B. Aji, A.-F. Bin Mubarak Ali, A. Aminuddin, and L. D. Farida, "Selection of Prospective Workers Using Profile Matching Algorithm on Crowdsourcing Platform," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 122–126, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9972155.
- [22] Geetha Mohan, A.-F. Mubarak Ali, M. Ariff Ameen, B. Lathakumary Vijayan, A. Aminuddin, and W. Widayani, "Predictive Models Using Supervised Neural Network for Pollutant Removal Efficiency in Petrochemical Wastewater Treatment," 2022 5th Int. Conf. Inf. Commun. Technol., pp. 116–121, Aug. 2022, doi: 10.1109/ICOIACT55506.2022.9971929.