2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)
April 09-11. 2021, ISTTS Surabaya, Indonesia

Analysis of the Spread of COVID-19 in Local Areas in Indonesia

2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) | 978-1-6654-0514-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/EIConCIT50028.2021.9431906

1st Alvian Burhanuddin
Magister Informatika
Universitas Islam Negeri Maulana Malik Ibrahim
Malang, Indonesia
19841009@student.uin-malang.ac.id

2nd Fachrul Kurniawan
Magister Informatika
Universitas Islam Negeri Maulana Malik Ibrahim
Malang, Indonesia
fachrulk@ti.uin-malang.ac.id

***Abstract*—COVID-19 has spread in various countries, one of which is Indonesia. Several provinces in Indonesia have reported individuals who have tested positive for the virus. The spread that has not been completed makes prediction difficult, especially in the heuristic model. Therefore, this paper tries to approach the SIR and SIRD models to determine developments and estimate their distribution in Indonesia. We compared the two models to predict the peak spread of the virus in the local area. The SIRD model predicts that there will be a peak on day 178, while SIR is on day 177. This difference is influenced by the Alpha value, which is obtained from virus induction.**

***Index Terms*—Data mining, time series, SIR Model**

I. Introduction

COVID-19 was first reported in Wuhan's city at the end of 2019 and has spread rapidly to various other countries. Within three months, WHO declared COVID-19 a global pandemic . This virus is a variant of the coronavirus; in several countries, research has been carried out on how it develops in the region, such as France [1], Japan [2], and India [3]. The spread of the COVID-19 virus also exists in Indonesia. Lots of sectors in Indonesia were paralyzed when this virus arrived. On the other hand, several solutions in various fields have also emerged, for example in the field of education [4] and economics [5].

The latest data from the official website of the Indonesian Ministry of Health records that nearly 12,000 people spread across various provinces have been infected with this virus . The development of data on the site is increasing every day, especially for positive cases. considering that when this paper was written, there was no vaccine to control the virus. This paper wants to conduct research on the development of the spread in Indonesia using a mathematical epidemic approach using the SIR and SIRD models. With this research, it is hoped that the level of distribution and forecasts of Indonesia's peak season can be achieved.

II. Related Research

Nuraini [6] conducted research on the development of COVID-19 in Indonesia using Richard's curve. In this study, predictions and development information were carried out using Korea's parameters to be compared with data in Indonesia. In this paper, we try to cover the deficiency of these parameters by estimating parameters according to Indonesia's existing data conditions. Fanelli [7] conducted an analysis to compare the spread of COVID-19 in Italy, China, and France, using the SIRD model. The study found that this virus has a different spread rate and mortality ratio in each country. Italy has a death rate of 4-8, while China has a lower mortality rate of 1-3. Meanwhile, based on observations using this model, the recovery rate is not much different in the observed countries.

In a different paper, Wangping et al [8] perform the same comparison using the SIR method. The study found that the reproductive value of COVID-19 in Italy was 4.1 (95 % CI: 2.156.77), and China had a reproductive value of 3.15 (95 % CI: 1.715.21). This different discovery resulted from the faster Chinese state in carrying out the lockdown process in several cities. A value of 1.5 times greater (approximately in the range between 4.2 - 4.5) was found if only at that time the late lockdown was imposed in several cities in China. Rajesh, Aditya also studied the effects of the lockdown [3] in India. The researcher compared the spread of the COVID-19 virus in the period before the lockdown and after the lockdown was enforced within 100 days. As in previous studies [8], the spread α value was reduced by 10% from the previous one without lockdown.

III. Methodology

A. Parameter Estimation

This paper complements the parameter estimates for α and β, using Kermack and McKendrick's [9] equations about biological epidemics. Meanwhile, to determine the total recovery of an individual, the resulting inequality is reduced to.

$$\int_{-\infty}^{\infty} tf(dt) = \frac{1}{\alpha} \quad (1)$$

When textually explained, the mean value of the total time an infected person could infect is . From here, the α value can be determined. To determine the β value, if it is known that the number of S not affected by the virus, it should be found in another equation from Kermack and McKendrick's model. However, in this paper, we use the Least-Squares non-linear method to estimate the parameters. This is because the Least-Squares non-linear method is quite good at determining estimates, for example, in Xi [10] research on predicting short-term airplane passengers. We use Differential Evolution (DE) [11] to perform a calibration on the results of the non-linear Least-Squares equation. We take DE use because our data uses time series, which based on research [12] this method is better and quite convergent with other data.

B. SIR Model

This section is a general discussion of models. There have been several books for a detailed SIR modeling conversation, for example, Mogilner's book [13] and Martcheva's book [14]. So it is good to shorten it. We only write a few points

that we consider essential and relevant in this paper.In general, SIR has a scheme:

$$N = S(t) + I(t) + R(t) \quad (2)$$

Where N is the total population, S is the notation of the suspect population, I is the infected population, and R is the population that has recovered from the infection. Whereas (t) is the notation of the variable taken from the time-series. From the schematic, write the model:

$$\frac{dS}{dt} = -\beta SI \quad (3)$$

$$\frac{dI}{dt} = \beta SI - \alpha I \quad (4)$$

$$\frac{dR}{dt} = \alpha I \quad (5)$$

In the equation 3-5, the value of β is the transmission rate, and the value of α is the value of the removal / recovered rate

C. SIRD Model

The model of SIRD has similarities to SIR, wherein in this model, the value of mortality, which is denoted by δ, is added. In general, the equation of this SIR derivative is written:

$$\frac{dI}{dt} = \beta SI - (1-\delta)\alpha I - \delta\rho I \quad (6)$$

$$\frac{dI}{dt} = (1-\delta)\alpha I \quad (7)$$

$$\frac{dI}{dt} = \delta\rho I \quad (8)$$

The equation 3 does not change, wherein the equation 4 and 5 there is a change with the ratio of deaths caused by the virus. Other studies using a similar method are [15], [16]. Rajendrakumar conducted research using this method in the Indian state, while Caccavo used it for Italy. To find the value of the $\delta(t)$ value, you need the delta's initial value δ_0. As in determining the β model, we also used the least-squares method to find the appropriate value in this study in the δ can be seen in Fig. 1.

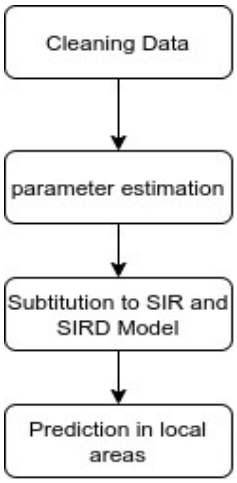

Fig. 1. An overview analysis of spread covid-19 in local areas

IV. Discussion

A. Data

We collect datasets that are open to the public on the Ristek Dikti page. From this site, we retrieve data belonging to KawalCovid. For this study, we took data from only one province to reduce complexity. The data we take is the number of positive patients, the number of patients who died, and the number of patients recovered. We take this data in a timeseries, starting on 18 March 2020 - 08 May 2020. For population data, we refer directly to the Central Statistics Agency (BPS). Where the basis for taking this amount, we refer to the publications published in 2020 [17].

B. SIR Model

In conducting our tests, we looked for the value of α first. We found several references regarding the duration of the incubation period for the COVID-19 virus, namely with a range of 5 days [18], between 7 days to 10 days [19]. From these two studies, we will compare the values of α = 0.2 and α = 0.125. Then perform a regression to find the value of beta in the SIR method.

We were fitting model our data based on the two α and β values we studied previously. Next, we compared the results obtained from alpha and beta, respectively, against real data. This difference in alpha value occurs because of different assumptions about the induction period of the virus. In Fig. 2. we consider the induction period to be five days, while in Fig. 3. the induction period is eight days. The higher the assumption of the virus's induction value, the more it will affect the suitability of the model with real data. After knowing that the model matches the data set, we make predictions based on that data. As we mentioned in the collect data section, we collected a 51-day dataset, which we will use to determine when the peak season of this pandemic is. For each peak season, we calculate it based on the existing values in the previous model. Details can be seen in the Table I.

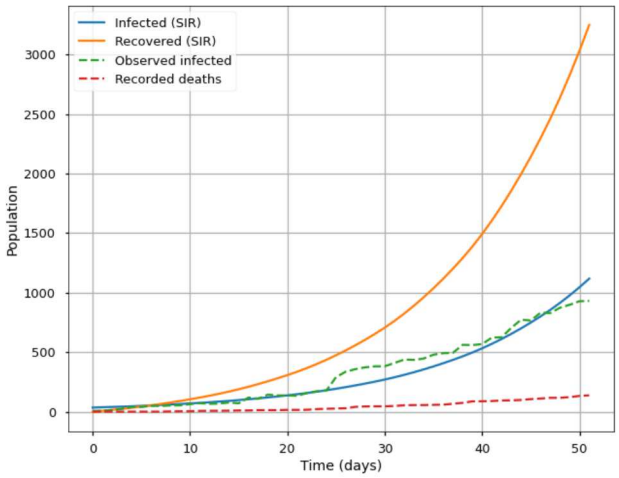

Fig. 2. fitting with α 0.2

TABLE I. *A* SIR Model

Method	α	β
SIR	0.2	0.266699
SIR	0.125	0.191698

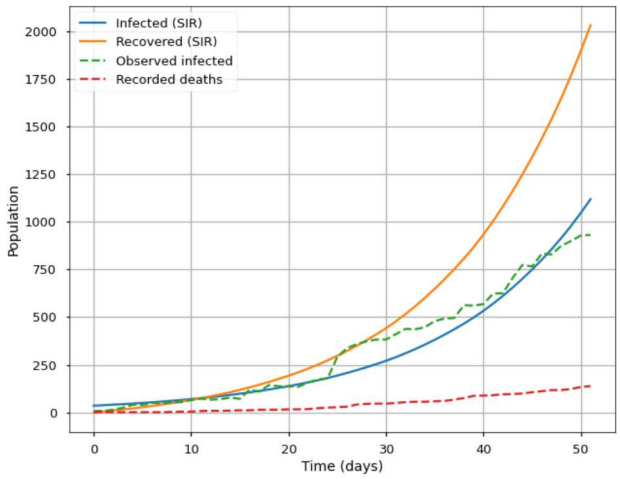

Fig. 3. fitting with α 0.12

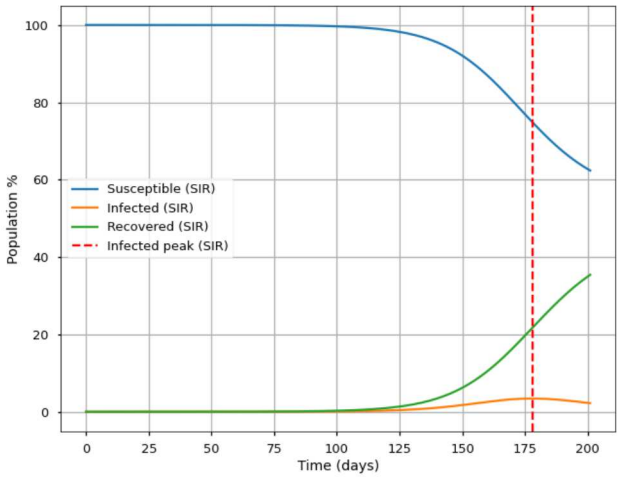

Fig. 4. predict with α 0.2

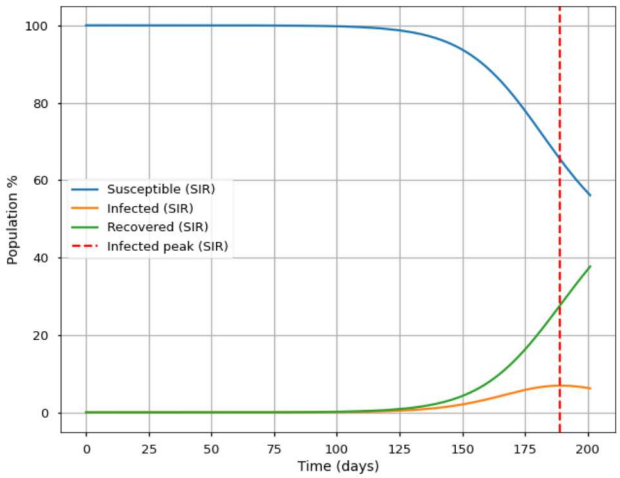

Fig. 5. predict with α 0.12

The perpendicular line represents the prediction peak season based on each α. Based on the 51-day dataset and α with a value of 0.2 (Fig. 4), we find that the peak season will occur on day 178, with the total infected users reaching 3.43 % of the total population. As for the prediction with a value of α 0.12 (Fig. 5), we find that the peak season will occur on day 189, with the total infected population reaching 6.91 % of the total population. On peak season, Susceptible drops to 40 %. Meanwhile, the population considered cured almost reached 40%. In this paper, we do not write a prediction of when the pandemic will end because more variables affect it.

C. *SIRD Model*

Testing on the SIRD model also refers to the SIR modelling in the sir section. α value has the same value, namely 0,2 and 0,12 can be seen in Fig. 6 and 7. However, the difference is in the δ value, which, at this stage, we determine based on the known β value in the SIR model. We initiate the value of δ_0 is 0, based on when the infected event occurred the first time there was no death. From the regression calculation, we get the results for the values α, β, and δ shown in the Table II. Unlike the sir section, which explains one by one the fitting results of the α value, in this section, we will directly compare the SIR and SIRD models.

TABLE II. REGRESSION RESULTS FOR SIRD VALUES

Method	alpha	beta	delta
SIRD	**0.2**	**0.266699**	**0.00984287**
SIRD	0.125	0.191698	0.00983505

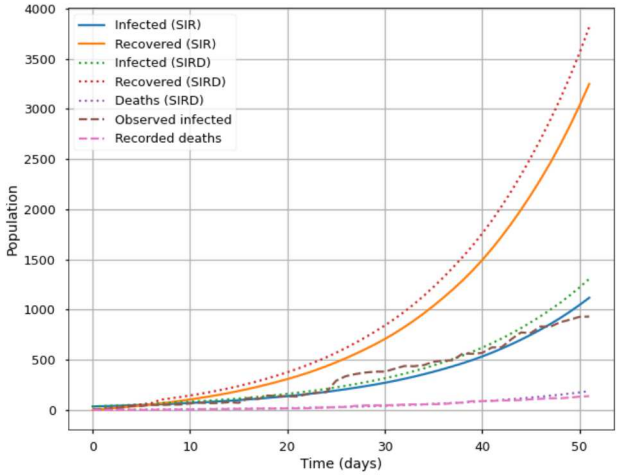

Fig. 6. fitting SIRD with α 0.2

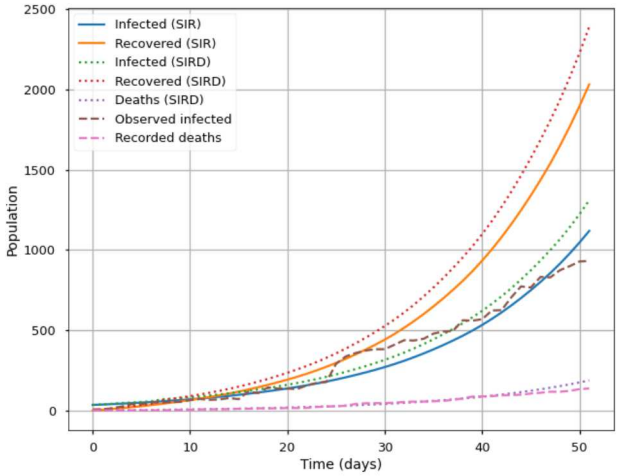

Fig. 7. fitting SIRD with α 0.12

If we look at the Fig. 6 and 7, the modelling result from SIRD has a slightly larger value than the value generated by the SIR model. This value is increased because it is influenced by the different models to get the value of I(t), where this difference is reflected in the 4 and 7 models. In the 7 models, the value of the _ coefficient in the equation is added. Furthermore, based on the SIRD model, we also predict the peak season based on existing datasets. Where to make predictions, we use the values in the Table II.

From our experiment, we get the result that there is a slightly different peak season in the SIR method. At an α value of 0.2, the peak season occurs on the 177th day. SIRD is different from the SIR Model prediction on day 178, a difference of 1 day. Meanwhile, if the α value is 0.12, the peak season occurs when the 188th day. It also experiences a 1-day difference from the SIR model. In the SIRD model, the variable number of deaths of the total infected individuals is also introduced. Where in the image indicated by a green line. When using α 0.2, the mortality rate at peak season is 1.5% of the population. Meanwhile, when the α value was 0.12, the mortality rate was 2.6% can be seen in Fig. 8 and 9.

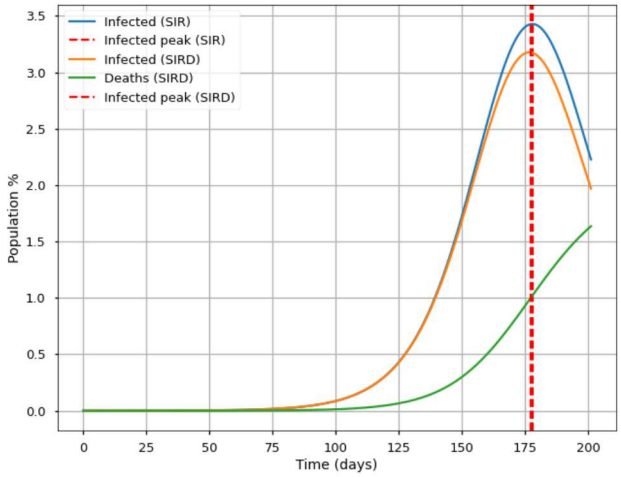

Fig. 8. SIRD Prediction using α 0.2

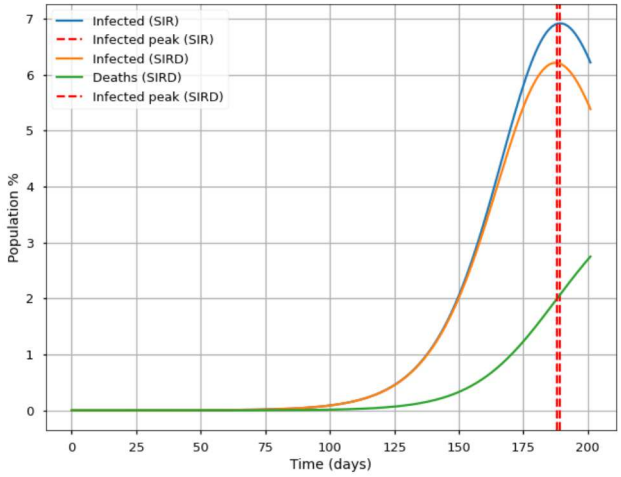

Fig. 9. SIRD Prediction using α 0.12

CONCLUSSION

In this paper, we perform data modeling based on a mathematical model for epidemiology, in which we involve the SIR and SIRD models. For the dataset, we collect data within the local scope of provinces in Indonesia. From SIR and SIRD modeling, it can be seen that the range of individual infections in the peak season ranges from 3 - 7% of the total population. Meanwhile, the peak season occurs between 175-190 days from when the individual was infected, where this value arises from the assumption that the incubation value is five days and eight days. During the incubation period, we performed regression modeling to find the coefficient of β. So that if further research is carried out, it is necessary to check the beta value through other methods, for example, the social network model. Besides, it is also necessary to check the lockdown effect because, in this paper, we have an assumption that there is no quarantine in the dataset we get.

REFERENCES

[1] S. B. Stoecklin *et al.*, "First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020," *Eurosurveillance*, vol. 25, no. 6, p. 2000094, 2020.

[2] K. Kakimoto, H. Kamiya, T. Yamagishi, T. Matsui, M. Suzuki, and T. Wakita, "Initial investigation of transmission of COVID-19 among crew members during quarantine of a cruise ship—Yokohama, Japan, February 2020," *Morb. Mortal. Wkly. Rep.*, vol. 69, no. 11, pp. 312–313, 2020.

[3] A. Rajesh, H. Pai, V. Roy, S. Samanta, and S. Ghosh, "CoVID-19 prediction for India from the existing data and SIR (D) model study," *medRxiv*. Cold Spring Harbor Laboratory Press, 2020.

[4] F. Kurniawan, F. A. Dwiyanto, G. Budiprasetyo, and M. I. Esha, "Sustainable development for smart city education based on features and parameters assessment modelling slope one algorithm," *J. Green Eng.*, vol. 10, no. 10, pp. 9801–9811, 2020.

[5] S. Susilawati, R. Falefi, and A. Purwoko, "Impact of COVID-19's Pandemic on the Economy of Indonesia," *Budapest Int. Res. Critics Inst. Humanit. Soc. Sci.*, vol. 3, no. 2, pp. 1147–1156, 2020.

[6] N. Nuraini, K. Khairudin, and M. Apri, "Modeling simulation of COVID-19 in Indonesia based on early endemic data," *Commun. Biomath. Sci.*, vol. 3, no. 1, pp. 1–8, 2020.

[7] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons & Fractals*, vol. 134, p. 109761, 2020.

[8] J. Wangping *et al.*, "Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China," *Front. Med.*, vol. 7, p. 169, 2020.

[9] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics. II.—The problem of endemicity," in *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 1932, vol. 138, no. 834, pp. 55–83.

[10] G. Xie, S. Wang, and K. K. Lai, "Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches," *J. Air Transp. Manag.*, vol. 37, pp. 20–26, 2014.

[11] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[12] R. K. Ursem and P. Vadstrup, "Parameter identification of induction motors using differential evolution," in *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, 2003, vol. 2, pp. 790–796.

[13] A. Mogilner, "Mathematics of cell motility: have we got its number?," *J. Math. Biol.*, vol. 58, no. 1, pp. 105–134, 2009.

[14] M. Martcheva, *An introduction to mathematical epidemiology*, vol. 61. Springer, 2015.

[15] A. L. Rajendrakumar *et al.*, "Epidemic Landscape and Forecasting of SARS-CoV-2 in India," *J Epidemiol Glob Heal.*, vol. 11, no. 1, pp. 55–59, 2020, doi: 10.2991/jegh.k.200823.001.

[16] D. Caccavo, “Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model,” *medRxiv*. Cold Spring Harbor Laboratory Press, 2020.

[17] Badan Pusat Statistik, “BPS-Statistics of Jawa Timur Province, e-catalog.” 2020.

[18] S. A. Lauer *et al.*, “The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application,” *Ann. Intern. Med.*, vol. 172, no. 9, pp. 577–582, 2020.

[19] S. Lei *et al.*, “Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection,” *EClinicalMedicine*, vol. 21, p. 100331, 2020.