

A Compartmental Epidemic Model Incorporating Probable Cases to Model COVID-19 Outbreak in Regions with Limited Testing Capacity

A. Hasan^{*a}, Y. Nasution^b

^a*Mærsk McKinney Møller Institute, University of Southern Denmark, Denmark*
^b*Department of Mathematics, University of Mulawarman, Indonesia*

Abstract

We propose a new compartmental epidemic model taking into account people who has symptoms with no confirmatory laboratory testing (probable cases). We prove well-posedness of the model and provide an explicit expression for the basic reproduction number (\mathcal{R}_0). We use the model together with an extended Kalman filter (EKF) to estimate the time-varying effective reproduction number (\mathcal{R}_t) of COVID-19 in West Java province, Indonesia, where laboratory testing capacity is limited. Based on our estimation, the value of \mathcal{R}_t is higher when the probable cases are taken into account. This correction can be used by decision and policy makers when considering re-opening policy and evaluation of measures.

Keywords: COVID-19, probable case, extended Kalman filter, reproduction number.

1. Introduction

Since declared as pandemic by WHO in early of March 2020, the coronavirus disease (COVID-19) has been spread to 216 countries by July 2020. At the end of July 2020, the virus has infected more than 17 million people (confirmed) and

^{*}Corresponding author. Email Address: agha@mmmi.sdu.dk (A. Hasan).

5 caused more than 650 thousand deaths [1]. Particularly in Indonesia, the number of confirmed positive infection is over 100 thousand cases with over 5,000 deaths [2]. In order to prevent and control the outbreak, in early of April 2020, the Indonesian government began to conduct a Large-Scale Social Restriction (LSSR) at several regions with severe outbreak cases, such as at Jakarta and
10 West Java Province [3]. LSSR includes measures such as closing public places, restricting public transport, and limiting travel to and from restricted regions. This regulation is being controlled by each of local government. To assess the duration of LSSR, local government relies on the estimation value of the time-varying effective reproduction number (\mathcal{R}_t) at each region [4]. The government
15 officials use sequential Bayesian method based on the discrete SIR model presented in [5] and [6]. The method requires daily confirmed infection data as inputs to estimate \mathcal{R}_t .

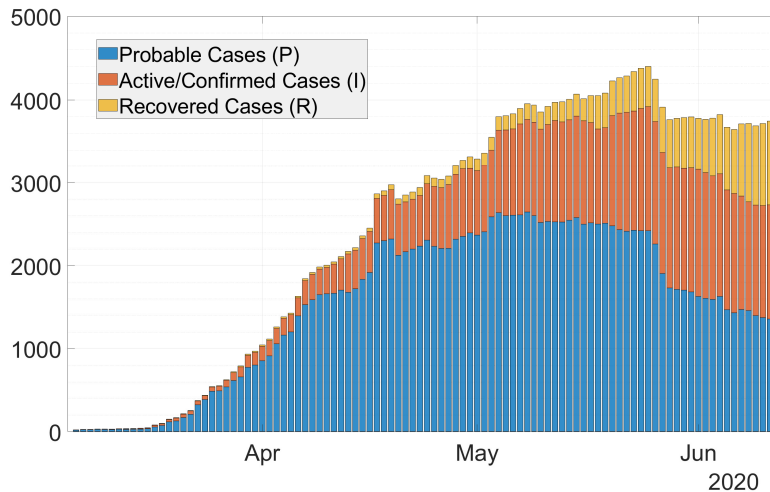


Figure 1: Stacked-bar of COVID-19 cases in West Java, Indonesia.

Based on CDC case definitions, however, there are two classifications of COVID-19 case, namely probable cases and confirmed cases [7]. A person can
20 be classified into probable case if he or she

- *meets clinical criteria AND epidemiologic evidence with no confirmatory laboratory testing performed for COVID-19,*
- *meets presumptive laboratory evidence AND either clinical criteria OR epidemiologic evidence,*
- 25 • *meets vital records criteria with no confirmatory laboratory testing performed for COVID-19.*

Details regarding clinical criteria, epidemiologic evidence, and vital records criteria can be found in [7]. Probable case is quite similar to Patient under Surveillance (*Pasien Dalam Pengawasan or PDP*) based on the Indonesian ministry of health [8]. In many provinces in Indonesia, the number of probable cases far exceed the number of confirmed cases, especially in the beginning of the pandemic. For example, Figure 1 shows a much larger number of probable cases in West Java province in Indonesia compared to active/confirmed cases. Limited laboratory testing and the existence of lag time between testing and result can be the cause of this situation [9]. Our aim is to estimate the time-varying effective reproduction number \mathcal{R}_t by utilizing daily probable case data together with daily confirmed case data. To this end, we propose a Susceptible-Probable-Infectious-Recovered (SPIR) model with four compartments as shown in Figure 2.

40 **2. Model description**

2.1. Susceptible-Probable-Infectious-Recovered (SPIR) model

The SPIR model consists of four ordinary differential equations (ODEs), describing the evolution of the population in each stage over time, and is given by:

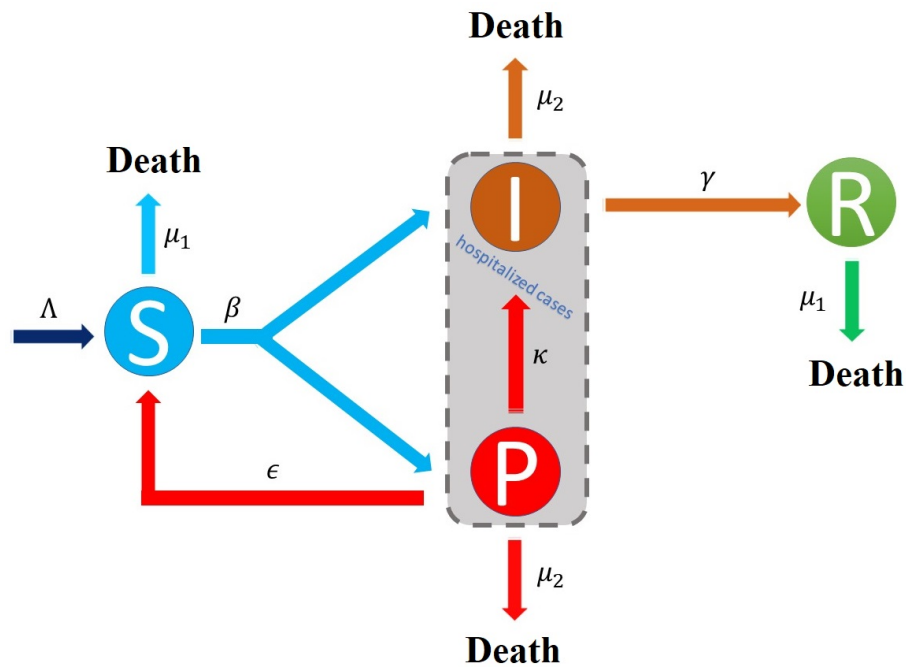


Figure 2: SPIR model is proposed to model probable cases of COVID-19 transmission process.

$$\frac{dS(t)}{dt} = \Lambda(t) - \frac{\beta S(t)(I(t) + P(t))}{N} + \epsilon P(t) - \mu_1 S(t), \quad (1)$$

$$\frac{dP(t)}{dt} = \frac{\beta S(t)P(t)}{N} - (\kappa + \epsilon + \mu_2)P(t), \quad (2)$$

$$\frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} + \kappa P(t) - (\gamma + \mu_2)I(t), \quad (3)$$

$$\frac{dR(t)}{dt} = \gamma I(t) - \mu_1 R(t). \quad (4)$$

45 Here, S , P , I , R denote susceptible case, probable case, active/confirmed case, and recovered case, respectively. The system satisfies

$$S(t) + P(t) + I(t) + R(t) = N, \quad (5)$$

expressing in mathematical terms the constancy of population N . We assume the natality $\Lambda(t)$ compensate for the death, i.e.,

$$\Lambda(t) = \mu_1 S(t) + \mu_2 P(t) + \mu_2 I(t) + \mu_1 R(t), \quad (6)$$

where μ_1 and μ_2 denote the death rates. In our model, we assume $\mu_2 > \mu_1$ since infected individuals die at faster rate. If $\mu_1 = \mu_2 = \mu$, then the natality is constant, i.e., $\Lambda = \mu N$. The remaining parameters include the infection rate β , the negative testing rate ϵ , the positive testing rate κ , and the recovery rate γ . Among these parameters, β could be considered as the most important parameter, which quantifies the transmission of the virus. β could also be defined as the average number of contacts per person per time multiplied by the probability of disease transmission. Thus, in practice this parameter is time-varying due to intervention. Remark that in order to simplify the analysis, the infection rate β in our model goes equally to I and P . However, if a Polymerase Chain Reaction (PCR) testing shows negative result, a person from compartment P can go back to compartment S . This is modelled by the negative testing rate ϵ .

2.2. Well-posedness of the SPIR model

We establish Theorem 1 and Theorem 2 to ensure the mathematical and biological well-posedness of the SPIR model (1)-(4).

Theorem 1. For $S(0), P(0), I(0), R(0) \in \mathbb{R}$, there exists $t_0 > 0$ and continuously differentiable functions $S(t), P(t), I(t), R(t) : [0, t_0) \rightarrow \mathbb{R}$, such that $(S(t), P(t), I(t), R(t))$ satisfies (1)-(4).

Proof. The Jacobian of (1)-(4) is linear with respect to $S(t), P(t), I(t)$, and $R(t)$. Thus, the system is locally Lipschitz. From Picard-Lindelöf theorem [10], there exists a unique solution for the system (1)-(4), hence completes the proof. \square

Theorem 2. *Let us assume that $S(0), P(0), I(0), R(0) \geq 0$. If the unique solution of (1)-(4) exists on the interval $[0, t_0]$ for $t_0 > 0$, then the functions $(S(t), P(t), I(t), R(t))$ will remain positive and bounded $\forall t \in [0, t_0]$, i.e., the system (1)-(4) is a dynamical system on the compact set*

$$\mathcal{K} = \{(S, P, I, R) \in \mathbb{R}_+^4 \mid \exists c > 0 \ni S + P + I + R < c\}. \quad (7)$$

⁷⁵ *Proof.* Solution of (2) is given by

$$P(t) = P(0) \exp\left(\int_0^t \left(\frac{\beta S(\tau)}{N} - (\kappa + \epsilon + \mu_2)\right) d\tau\right). \quad (8)$$

Thus, for $P(0) \geq 0$ we have $P(t) \geq 0$. Furthermore, solution for (1) is given by

$$\begin{aligned} S(t) = & S(0) \exp\left(\int_0^t \left(\frac{\beta(I(\tau) + P(\tau))}{N} + \mu_1\right) d\tau\right) \\ & + \exp\left(\int_0^t \left(\frac{\beta(I(\tau) + P(\tau))}{N} + \mu_1\right) d\tau\right) \times \\ & \int_0^t \left(\Lambda(\tau) + \epsilon P(\tau)\right) \exp\left(-\int_0^s \left(\frac{\beta(I(\tau) + P(\tau))}{N} + \mu_1\right) d\tau\right) ds. \end{aligned} \quad (9)$$

Since $\Lambda(t) \geq 0$ and $P(t) \geq 0$, then $S(t) \geq 0$. Using similar approach, we can prove $I(t), R(t) \geq 0$. Boundedness follows from the fact that $S(t) + P(t) + I(t) + R(t) = N$. \square

⁸⁰ *2.3. Basic reproduction number*

The SPIR model has a Disease-Free Equilibrium (DFE) state, which is given by $(S^*, P^*, I^*, R^*) = (N, 0, 0, 0)$. Following Lemma 1 in [11], we define:

$$F = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \text{ and } V = \begin{pmatrix} \kappa + \epsilon + \mu_2 & 0 \\ -\kappa & \gamma + \mu_2 \end{pmatrix}. \quad (10)$$

According to [12], the next generation matrix is defined as FV^{-1} and the basic reproduction number is defined as the spectral radius of the next generation matrix, i.e.,

$$\mathcal{R}_0 = \max\{\Omega(FV^{-1})\} = \max\left\{\frac{\beta}{\kappa + \epsilon + \mu_2}, \frac{\beta}{\gamma + \mu_2}\right\}, \quad (11)$$

where Ω denotes the eigenvalue of the next generation matrix FV^{-1} . Remark that, in practice $\beta = \beta(t)$ due to intervention. Thus, considering the number of susceptible individual decline over time [13], the time-varying effective reproduction number is given by:

$$\mathcal{R}_t(t) = \frac{S(t)}{S(0)} \max\left\{\frac{\beta(t)}{\kappa + \epsilon + \mu_2}, \frac{\beta(t)}{\gamma + \mu_2}\right\}. \quad (12)$$

In the next section, we use an extended Kalman (EKF) filter to estimate the infection rate $\beta(t)$, and thus estimate the time-varying effective reproduction number $\mathcal{R}_t(t)$.

3. Estimation of the time-varying effective reproduction number \mathcal{R}_t

Since $\Lambda(t) = \mu_1 S(t) + \mu_2 P(t) + \mu_2 I(t) + \mu_1 R(t)$, the SPIR model (1)-(4) can be written as follow

$$\frac{dS(t)}{dt} = -\frac{\beta(t)S(t)(I(t) + P(t))}{N} + (\epsilon + \mu_2)P(t) + \mu_2 I(t) + \mu_1 R(t), \quad (13)$$

$$\frac{dP(t)}{dt} = \frac{\beta(t)S(t)P(t)}{N} - (\kappa + \epsilon + \mu_2)P(t), \quad (14)$$

$$\frac{dI(t)}{dt} = \frac{\beta(t)S(t)I(t)}{N} + \kappa P(t) - (\gamma + \mu_2)I(t), \quad (15)$$

$$\frac{dR(t)}{dt} = \gamma I(t) - \mu_1 R(t). \quad (16)$$

3.1. Discrete-time stochastic augmented SPIR model

Discretizing the SPIR model (13)-(16) using forward Euler method and augmenting the infection rate as a sixth state variable, we obtain the following discrete-time stochastic augmented SPIR model:

$$S(k+1) = S(k) - \frac{\beta(k)S(k)(I(k) + P(k))\Delta t}{N} + (\epsilon + \mu_2)\Delta t P(k) + \mu_2\Delta t I(k) + \mu_1\Delta t R(k) + w_1(k), \quad (17)$$

$$P(k+1) = \frac{\beta(k)S(k)P(k)\Delta t}{N} + (1 - (\kappa + \epsilon + \mu_2))\Delta t P(k) + w_2(k), \quad (18)$$

$$I(k+1) = \frac{\beta(k)S(k)I(k)\Delta t}{N} + \kappa\Delta t P(k) + (1 - (\gamma + \mu_2))\Delta t I(k) + w_3(k), \quad (19)$$

$$R(k+1) = \gamma\Delta t I(k) + (1 - \mu_1\Delta t)R(k) + w_4(k), \quad (20)$$

$$\beta(k+1) = \beta(k) + w_5(k). \quad (21)$$

100 Here, we add noise $\mathbf{w}(k) = (w_1(k) \ w_2(k) \ w_3(k) \ w_4(k) \ w_5(k))^\top$ as uncertainty in the model and is assumed to be white noise Gaussian and uncorrelated. Furthermore, we assume the infection rate $\beta(k)$ is a piece-wise continuous function with jump every one day. Thus, in one day the value of $\beta(k)$ is constant and (21) is satisfied. Augmenting a parameter as a new state variable is a common
105 technique when estimating a parameter using EKF (see page 422 in [14]).

3.2. Extended Kalman filter (EKF)

In this section, we use EKF to estimate the state variables in (17)-(21) dynamically. To simplify the presentation, we define an augmented state vector

$$\mathbf{x}(k+1) = \left(S(k+1) \ P(k+1) \ I(k+1) \ R(k+1) \ \beta(k+1) \right)^\top, \quad (22)$$

and as such, the discrete-time stochastic augmented SPIR model (17)-(21) can
 110 be written as follows

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{w}(k). \quad (23)$$

Let us denote $\hat{\mathbf{x}}(k)$ as an estimated vector state from the EKF. Applying first-order Taylor series expansion to \mathbf{f} at $\hat{\mathbf{x}}(k)$, we obtain $\mathbf{f}(\mathbf{x}(k)) = \mathbf{f}(\hat{\mathbf{x}}(k)) + \mathbf{J}_{\mathbf{f}}(\hat{\mathbf{x}}(k))(\mathbf{x}(k) - \hat{\mathbf{x}}(k))$, where $\mathbf{J}_{\mathbf{f}}(\hat{\mathbf{x}}(k))$ is the Jacobian matrix of \mathbf{f} , given by

$$\mathbf{J}_{\mathbf{f}}(\hat{\mathbf{x}}(k)) = \begin{pmatrix} J_{11}(\hat{\mathbf{x}}(k)) & J_{12}(\hat{\mathbf{x}}(k)) & J_{13}(\hat{\mathbf{x}}(k)) & \mu_1 \Delta t & J_{15}(\hat{\mathbf{x}}(k)) \\ J_{21}(\hat{\mathbf{x}}(k)) & J_{22}(\hat{\mathbf{x}}(k)) & 0 & 0 & J_{25}(\hat{\mathbf{x}}(k)) \\ J_{31}(\hat{\mathbf{x}}(k)) & \kappa \Delta t & J_{33}(\hat{\mathbf{x}}(k)) & 0 & J_{35}(\hat{\mathbf{x}}(k)) \\ 0 & 0 & \gamma \Delta t & 1 - \mu_1 \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (24)$$

where

$$J_{11}(\hat{\mathbf{x}}(k)) = 1 - \frac{\beta(k)(I(k) + P(k))\Delta t}{N}, \quad (25)$$

$$J_{12}(\hat{\mathbf{x}}(k)) = -\frac{\beta(k)S(k)\Delta t}{N} + (\epsilon + \mu_2)\Delta t, \quad (26)$$

$$J_{13}(\hat{\mathbf{x}}(k)) = -\frac{\beta(k)S(k)\Delta t}{N} + \mu_2\Delta t, \quad (27)$$

$$J_{15}(\hat{\mathbf{x}}(k)) = -\frac{S(k)(I(k) + P(k))\Delta t}{N}, \quad (28)$$

$$J_{21}(\hat{\mathbf{x}}(k)) = \frac{\beta(k)P(k)\Delta t}{N}, \quad (29)$$

$$J_{22}(\hat{\mathbf{x}}(k)) = \frac{\beta(k)S(k)\Delta t}{N} + 1 - (\kappa + \epsilon + \mu_2)\Delta t, \quad (30)$$

$$J_{25}(\hat{\mathbf{x}}(k)) = \frac{S(k)P(k)\Delta t}{N}, \quad (31)$$

$$J_{31}(\hat{\mathbf{x}}(k)) = \frac{\beta(k)I(k)\Delta t}{N}, \quad (32)$$

$$J_{33}(\hat{\mathbf{x}}(k)) = \frac{\beta(k)S(k)\Delta t}{N} + 1 - (\gamma + \mu_2)\Delta t, \quad (33)$$

$$J_{35}(\hat{\mathbf{x}}(k)) = \frac{S(k)I(k)\Delta t}{N}. \quad (34)$$

115 The EKF has two main tuning parameters: the process covariance matrix \mathbf{Q}_F and the observation covariance matrix \mathbf{R}_F . Detail procedures of implementing this method can be found in [15]. Note that, the main purpose of EKF is used as real-time data fitting. Thus, the tuning parameters are chosen such that the Relative Root Mean Square Error (RRMSE) between the data and the
120 estimate is sufficiently small. The RRMSE for each variable X_i is defined as

$$\text{RRMSE}_{X_i} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{X_i - \hat{X}_i}{X_i} \right)^2, \quad (35)$$

where N_d is the number of observed days. Here, $X_i \in \{S(i), P(i), I(i), R(i)\}$ and $\hat{X}_i \in \{\hat{S}(i), \hat{P}(i), \hat{I}(i), \hat{R}(i)\}$ denote the data and the estimate, respectively.

CFR	infectious time T_i (days)	life expectancy T_l (days)
0.0425	12 ± 3	72×360

Table 1: Data used to calculate model parameters.

4. Case study: West Java province in Indonesia

We use the discrete-time stochastic augmented SPIR model together with
 125 the EKF to estimate the time-varying effective reproduction number \mathcal{R}_t in West
 Java province (population 48 million) in Indonesia. All data sets and MATLAB
 code are available on GitHub (<https://github.com/agusisma/covidPDP>). To
 simplify the model, the parameters are assumed to be constant and are obtained
 from clinical information, such as case-fatality-rate (CFR), infectious time T_i ,
 130 and life expectancy T_l , and are given by:

$$\epsilon = \kappa = \gamma = \frac{(1 - \text{CFR})}{T_i}, \quad \mu_1 = \frac{\text{CFR}}{T_l}, \quad \mu_2 = \frac{\text{CFR}}{T_i}. \quad (36)$$

The negative testing rate ϵ , the positive testing rate κ , and the recovery rate
 γ are assumed to be equal after confirmed through PCR testing. The data are
 provided in Table 1. The tuning parameters for the Kalman filter are chosen
 as $\mathbf{Q}_F = \text{diag}(10 \ 10 \ 10 \ 5 \ 0.2)$ and $\mathbf{R}_F = \text{diag}(100 \ 10 \ 5 \ 1)$, respectively. Figure
 135 3 shows results from dynamic data fitting using EKF. As can be seen from this
 figure, the discrete-time stochastic augmented SPIR model together with the
 EKF are able to model the transmissions of COVID-19 accurately.

The RRMSE_{X_i} using the chosen tuning parameters \mathbf{Q}_F and \mathbf{R}_F are given
 in Table 2, where the error is reasonably small.

RRMSE $_{X_i}$				
S	P	I	R	Total
4.1789e-16	6.6468e-04	1.3062e-06	2.2014e-06	6.6819e-04

Table 2: RRMSE between reported data and estimated data using EKF.

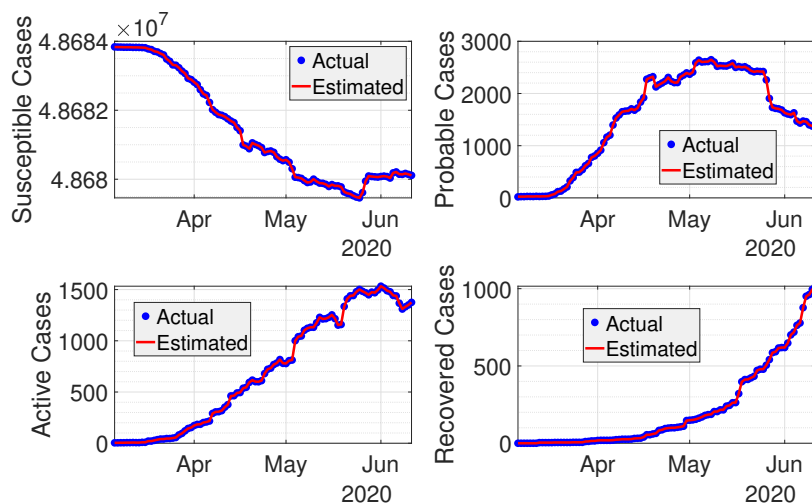


Figure 3: Real-time fitting using extended Kalman filter.

140 The estimated \mathcal{R}_t can be seen from Figure 4. Here, the value of \mathcal{R}_t is higher when the probable cases are taken into account. The difference between the estimated \mathcal{R}_t with and without probable cases becomes smaller once the rate of the probable cases decreases.

On May 6, the government of West Java province enact a Large-Scale Social
 145 Restriction (LSSR). The decision was taken based on the evaluation of \mathcal{R}_t , which according to WHO need to be done if $\mathcal{R}_t > 1$. The restriction can be relaxed if $\mathcal{R}_t < 1$ for fourteen consecutive days. Since the government officials estimated \mathcal{R}_t without considering probable cases, the restriction was relaxed on May 20 when $\mathcal{R}_t < 1$ for two weeks. Taking probable cases into account, our estimation
 150 shows that during the period of LSSR, the estimated value of \mathcal{R}_t is still above 1. Thus, the LSSR need to be extended rather than relaxed. The effect of LSSR, where $\mathcal{R}_t < 1$, can be seen after two weeks (May 24) since there are delays in infection confirmation.

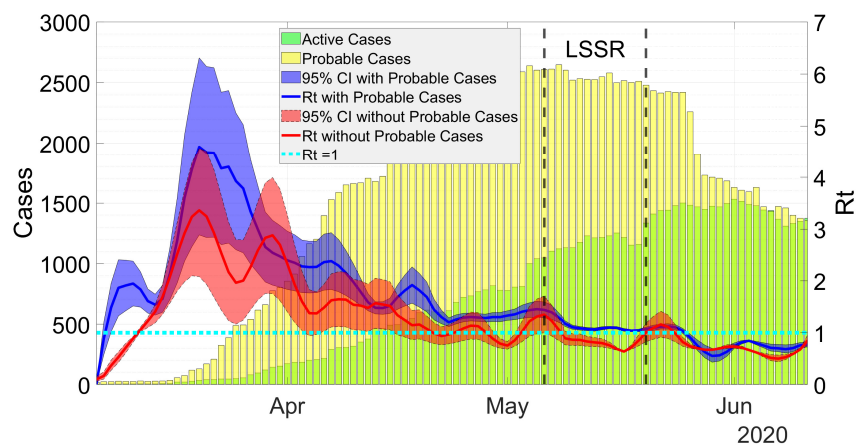


Figure 4: Estimation of \mathcal{R}_t with and without probable cases.

5. Conclusion

155 A new compartmental epidemic model taking probable cases into account has been presented in this paper. The model, called the SPIR model, consists of four compartments: Susceptible (S), Probable (P), Active (I), and Recovered (R). The model is used to estimate the time-varying effective reproduction number \mathcal{R}_t in West Java province in Indonesia. To this end, we apply an EKF to a
160 discrete-time stochastic augmented SPIR model. Numerical simulations show that when probable cases are taking into account, the value of \mathcal{R}_t is significantly higher. This results can be used to inform policy makers when deciding to loosen or tighten the measures. In general, our model and approach can be used in regions with limited testing capacity.

165 Acknowledgement

We would like to thank Dr. Rangga Ganzar Noegraha who proofreads the paper and Muhammad Fakhruddin who provides Figure 2.

References

- [1] World Health Organization, Coronavirus disease 2019 (covid-19) situation report, 2020, <https://covid19.who.int/>.
170
- [2] World Health Organization, Coronavirus disease 2019 (covid-19) situation report, 2020, country: Indonesia, <https://covid19.who.int/region/searo/country/id>.
- [3] B. Sutrisno, 50 days of indonesia's partial lockdown. is it enough for the 'new normal'?, The Jakarta Post, 28 May 2020.
175
- [4] A. Nurbaiti, West java extends psbb to june 26, The Jakarta Post, 15 June 2020.
- [5] L. Bettencourt, R. Ribeiro, Real time bayesian estimation of the epidemic potential of emerging infectious diseases, PLoS ONE 3. doi:10.1371/journal.pone.0002185.
180
- [6] A. Cori, N. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics, American Journal of Epidemiology 178. doi:10.1093/aje/kwt133.
- [7] Center for Disease Control and Prevention, Coronavirus disease 2019 (covid-19) 2020 interim case definition, <https://www.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/>.
185
- [8] A. Nurbaiti, Health ministry scraps odp, pdp terminology in new covid-19 guidelines, The Jakarta Post, 14 July 2020.
- [9] R. Ariyana, Mapping of laboratories urgent for mass covid-19 testing, The Jakarta Post, 29 April 2020.
190

- [10] R. Bartle, D. Sherbert, Introduction to Real Analysis, 4th Edition, Wiley, 2011.
- [11] P. van den Driessche, J. Watmough, Reproduction numbers and sub-
195 threshold endemic equilibria for compartmental models of disease trans-
mission, *Mathematical Biosciences* 180 (2002) 29–48. doi:10.1016/
S0025-5564(02)00108-6.
- [12] O. Diekmann, J. Heesterbeek, J. Metz, On the definition and the com-
putation of the basic reproduction ratio r_0 in models for infectious dis-
200 eases in heterogeneous populations, *J. Math. Biol.* 28 (1990) 365–382.
doi:10.1007/BF00178324.
- [13] F. Brauer, C. Castillo-Chavez, Z. Feng, *Mathematical Models in Epidemi-
ology*, 1st Edition, Springer-Verlag, 2019.
- [14] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear
205 Approaches*, 1st Edition, Wiley-Interscience, 2006.
- [15] A. Hasan, H. Susanto, V. Tjahjono, R. Kusdiantara, E. Putri, P. Hadis-
oemarto, N. Nuraini, A new estimation method for covid-19 time-varying
reproduction number using active cases, *MedRxiv (preprint)* doi:10.1101/
2020.06.28.20142158.