



## Research article

## Data-driven modeling and forecasting of COVID-19 outbreak for public policy making

A. Hasan<sup>a,\*</sup>, E.R.M. Putri<sup>b</sup>, H. Susanto<sup>c,d</sup>, N. Nuraini<sup>e</sup><sup>a</sup> Mærsk McKinney Møller Institute, University of Southern Denmark, Denmark<sup>b</sup> Department of Mathematics, Institut Teknologi Sepuluh Nopember, Indonesia<sup>c</sup> Department of Mathematics, Khalifa University, United Arab Emirates<sup>d</sup> Department of Mathematical Sciences, University of Essex, United Kingdom<sup>e</sup> Department of Mathematics, Institut Teknologi Bandung, Indonesia

## ARTICLE INFO

## Article history:

Received 31 July 2020

Received in revised form 11 January 2021

Accepted 13 January 2021

Available online 20 January 2021

## Keywords:

COVID-19

Modeling

Forecasting

Public policy

## ABSTRACT

This paper presents a data-driven approach for COVID-19 modeling and forecasting, which can be used by public policy and decision makers to control the outbreak through Non-Pharmaceutical Interventions (NPI). First, we apply an extended Kalman filter (EKF) to a discrete-time stochastic augmented compartmental model to estimate the time-varying effective reproduction number ( $\mathcal{R}_t$ ). We use daily confirmed cases, active cases, recovered cases, deceased cases, Case-Fatality-Rate (CFR), and infectious time as inputs for the model. Furthermore, we define a Transmission Index (TI) as a ratio between the instantaneous and the maximum value of the effective reproduction number. The value of TI indicates the “effectiveness” of the disease transmission from a contact between a susceptible and an infectious individual in the presence of current measures, such as physical distancing and lock-down, relative to a normal condition. Based on the value of TI, we forecast different scenarios to see the effect of relaxing and tightening public measures. Case studies in three countries are provided to show the practicability of our approach.

© 2021 ISA. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

The spread of the new coronavirus disease 2019 (COVID-19), originating from Wuhan China, has been worldwide and has caused a severe outbreak. The virus has infected more than 60M people with more than 1.4M confirmed deaths by the end of November 2020. The outbreak has not only triggered health crisis, but has also economic and social ones. It has become a multidimensional problem that needs to be minimized through measurable public policies.

## 1.1. Motivation

Intervention measures have been introduced to contain the outbreak and to prevent it from continuously growing and being transmitted, such as through physical distancing and lock-down measures [1,2]. To this extent, a thorough evaluation to implement the available options is urgently needed. A quantitative as well as a qualitative evaluation involving key characteristics of COVID-19 outbreak can be conducted based on epidemiological parameters [3]. As the incidence is growing, quantitative

studies on minimum physical distancing policies in Australia [4], China [5], and Italy [6] have been reported.

A control measure for the disease transmission, known as the time-varying effective reproduction number ( $\mathcal{R}_t$ ), reflects the disease extended transmission in the presence of interventions. Therefore, estimation of the time-varying effective reproduction number  $\mathcal{R}_t$  can be used to evaluate the implementation of a public policy [7]. The estimation of  $\mathcal{R}_t$  based on an epidemiological model, has an important role for an evidence-based policy making, that is also recognized by the World Health Organization (WHO) [8].

## 1.2. Literature review

A deterministic Susceptible–Infected–Recovered (SIR) model for  $\mathcal{R}_t$  estimation, assumed that the used data significantly representing the actual outbreak, was presented in [9]. Different sets of data representing levels of quarantine measures were used in [5] to describe the case growth as well as the effective reproduction number  $\mathcal{R}_t$  for each measure's levels. To accommodate uncertainties in incidence data, noise was added to the model in [7]. In the paper, the authors used inputs from daily new cases, active cases, recovered cases, and deceased cases, to estimate the spread

\* Corresponding author.

E-mail address: [agha@mmmi.sdu.dk](mailto:agha@mmmi.sdu.dk) (A. Hasan).

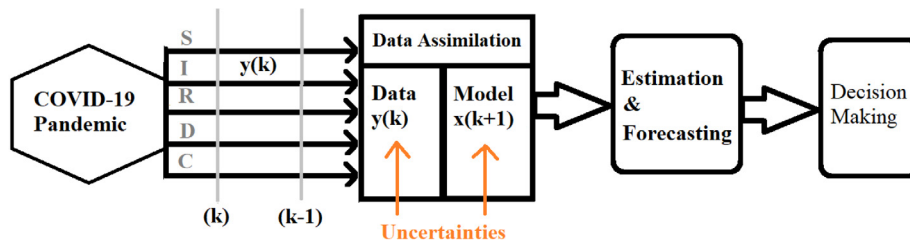


Fig. 1. Data-driven framework of COVID-19 pandemic.

Table 1

Typical time-series data reported during the pandemic.

Date	Month	Daily Confirmed (C)	Active Case (I)	Recovered (R)	Deceased (D)
⋮	⋮	⋮	⋮	⋮	⋮
28	4	25498	830745	148926	59418
29	4	28525	851065	154737	61812
30	4	30912	874215	160293	64018
1	5	36090	898527	170171	65918
2	5	29816	914246	182570	67616
3	5	27394	934901	188155	68770
⋮	⋮	⋮	⋮	⋮	⋮

of the disease. Based on the stochastic model, the study was then extended to  $\mathcal{R}_t$  estimation.

Using  $\mathcal{R}_t$ , several authors have proposed methods to forecast the evolution of the outbreak. Data-based analysis, modeling, and forecasting based on a Susceptible-Infectious-Recovered-Deceased (SIRD) model was presented in [10]. The authors fit the reported data with the SIRD model to estimate the epidemiological parameters. The main drawback when fitting the model with the data using the proposed method is that the estimated parameters can be unrealistic. The work of [11] attempted to use phenomenological models that have been validated during previous outbreaks. The model is used to generate and assess short-term forecasts of the cumulative number of confirmed reported cases. However, since COVID-19 is a new disease, the model is not reliable and the forecast can only be used for a very short term. Other authors used simple day-lag maps to investigate universality in the epidemic spreading [12]. Their results suggested that simple mean-field models can be used to gather a quantitative picture of the epidemic spreading. The main drawback is that the reproduction number was assumed to follow a heuristic continuous model, which may not described the actual transmission.

### 1.3. Contribution of this paper

In this paper, we propose a data-driven approach for COVID-19 modeling and forecasting, which can be used by public policy and decision makers to control the outbreak through Non-Pharmaceutical Interventions (NPI). Considering drawbacks in existing methods, we present two contributions: (i) estimation of the time-varying effective reproduction number  $\mathcal{R}_t$  based on real-time data fitting using an extended Kalman filter (EKF), and (ii) short to medium terms forecasting based on different public policies. As the effective reproduction number  $\mathcal{R}_t$  shows simply the extent of transmission due to population immunity or intervention in the form of public policy making [7], we propose a new measure called a Transmission Index (TI) (see Section 2.5), which describes the disease transmission relative to a normal condition. As well as  $\mathcal{R}_t$ , the value of TI can be used to measure the effectiveness of public health measures. Furthermore, TI can be used to forecast different public policy scenarios when a current measure is going to be loosen or even tightened.

### 1.4. Organization of this paper

Briefly, this paper is organized as follows. In Section 2, we discuss the methods used in evaluating the spread of the disease, including data availability and reliability, data driven framework, modeling, estimation, and forecasting. Then, we discuss the method and its applications to estimate  $\mathcal{R}_t$  and to forecast COVID-19 cases in United Arab Emirate (UAE), Australia, and Denmark in Section 3. Lastly, we present the conclusion in Section 4.

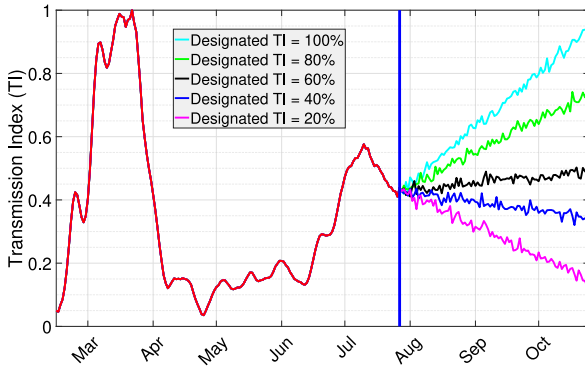
## 2. Methods

In this section, we describe the proposed data-driven modeling and forecasting approach that can be used by public policy and decision makers to control the COVID-19 pandemic through NPI. We acknowledge that no country has the exact total number of people infected with COVID-19, partially due to the lack of testing and undetected asymptomatic cases. Thus, the presented approach can only be used for countries/regions/areas that have performed mass testing with laboratory confirmation. Up to this point, we assumed that the difference between the actual and reported cases is minimum.

### 2.1. Data availability and reliability

There is a large number of generated data on COVID-19. Typically, government officials report daily confirmed cases, active cases, recovered cases, and deceased cases (see Table 1). These data are available for almost all countries and regions and can be accessed through online websites. Several websites, such as <https://www.worldometers.info> and <https://ourworldindata.org/>, also provide information regarding the number of test per capita. We will use these data in our analysis herein.

The testing positivity rate of a country, i.e., the ratio between the number of tests returning positive for COVID-19 relative to the total conducted tests, is a good metric to know whether or not the country has performed an adequate mass testing for their citizen to be able to properly monitor and control the spread of the virus. WHO advised that the positivity rate should remain at 5% or lower for at least 14 days. Only a few countries have successfully achieved this target, including UAE, Australia, and Denmark.



**Fig. 2.** Forecasting different scenarios through different designated values of TI.

**Table 2**

Interpretation of different value of designated TIs. The public measures are taken from the New Zealand COVID-19 alert system.

Level	TI	Public measures
1 (Do nothing)	100%	No public measures.
2 (Prevent)	80%	Border entry measures. Intensive testing for COVID-19. Rapid contact tracing of any positive case. Self-isolation and quarantine required. Schools and workplaces open.
3 (Reduce)	60%	People can connect with friends and family. No more than 100 people at gatherings. Keep physical distancing of 2 meters. Businesses can open to the public. Sport and recreation activities are allowed.
4 (Restrict)	40%	People must work from home. Children should learn at home if possible. Public venues are closed. No more than 10 people at gatherings. Healthcare services use virtual.
5 (Lock-down)	20%	People instructed to stay at home. Travel is severely limited. All gatherings canceled. Businesses closed except for essential services. Educational facilities closed.

**Table 3**

Parameter for the simulations.

Country	Parameters		
	N	CFR	$T_i$
UAE	9,890,402	0.15%	$9 \pm 3$
Australia	25,499,884	0.27%	$9 \pm 3$
Denmark	5,792,202	1.13%	$9 \pm 3$

**Table 4**

RRMSE between reported and estimated cases using EKF.

RRMSE <sub>x</sub>					
Country	I	R	D	C	Total
UAE	2.5e-04	1.8e-06	1.9e-06	2.4e-01	2.5e-01
Australia	2.1e-06	5.1e-06	2.1e-06	6.4e-03	6.5e-03
Denmark	4.1e-06	5.2e-02	2.2e-06	9.9e-03	6.2e-02

## 2.2. Data-driven framework

The reported cases from the pandemic will be used for two purposes: (i) to estimate the time-varying effective reproduction number  $\mathcal{R}_t$ , and (ii) to forecast the number of active, recovered, deceased, and total cases, which are important to prepare for the healthcare system.

We depicted in Fig. 1 a sketch of the data-driven framework proposed in our current study.  $S$  denotes susceptible case data. Assuming constant population,  $S$  can be obtained by subtracting

the number of population with the number of active, recovered, and deceased cases. The active case data  $I$  is the number of people who are currently infected.  $R$  and  $D$  denote the cumulative number of recovered and deceased, respectively, while  $C$  denotes the number of daily confirmed cases. To model uncertainty in the reported cases, we added a white Gaussian noise. To obtain an estimate of  $\mathcal{R}_t$ , the data are assimilated into the compartmental epidemic model (see Section 2.3 below) using EKF. Based on this estimation, we will perform forecasting for the next 90 days with different scenarios.

## 2.3. Mathematical modeling

To model the transmission of the coronavirus, we use the following discrete-time stochastic augmented compartmental model [7]:

$$S(k+1) = \left(1 - \frac{\Delta t}{N} \beta(k) I(k)\right) S(k) + w_1(k), \quad (1)$$

$$I(k+1) = (1 - (\gamma + \kappa) \Delta t) I(k) + \frac{\Delta t}{N} \beta(k) I(k) S(k) + w_2(k), \quad (2)$$

$$R(k+1) = R(k) + \gamma \Delta t I(k) + w_3(k), \quad (3)$$

$$D(k+1) = D(k) + \kappa \Delta t I(k) + w_4(k), \quad (4)$$

$$C(k+1) = (\gamma + \kappa) \Delta t I(k) + (1 - \Delta t) C(k) + w_5(k), \quad (5)$$

$$\beta(k+1) = \beta(k) + w_6(k). \quad (6)$$

Eqs. (1)–(4) can be obtained from the standard SIRD model. We chose to use a discrete-time model instead of a continuous-time model since the data is available in discrete-time. Thus, it will be easier to implement the model and the estimation algorithm. The time-varying effective reproduction number is then given by

$$\mathcal{R}_t(k) = \frac{S(k)}{S(0)} \left( \frac{\beta(k)}{\gamma + \kappa} \right) \quad (7)$$

Here, the term  $\frac{S(k)}{S(0)}$  is used to compensate the decline in the number of susceptible population, while the term  $\frac{\beta(k)}{\gamma + \kappa}$  is the expression of the basic reproduction number from the SIRD model [13]. We augment the standard SIRD model with Eqs. (5) and (6) that take into account the number of daily confirmed cases  $C(k)$  and the infectious rate  $\beta(k)$ . The infectious rate  $\beta(k)$  is assumed to be a piece-wise constant function with a jump at every one day time interval. The noise  $w_1(k)$ ,  $w_2(k)$ ,  $w_3(k)$ ,  $w_4(k)$ ,  $w_5(k)$ , and  $w_6(k)$  are added to model the uncertainty.

The system (1)–(6) have three constant parameters: the number of population  $N$ , the recovery rate  $\gamma$ , and the death rate  $\kappa$ . The recovery and death rates depend on the infectious time  $T_i$  and Case-Fatality-Rate (CFR) and are given by

$$\gamma = \frac{1 - \text{CFR}}{T_i}, \quad \kappa = \frac{\text{CFR}}{T_i}. \quad (8)$$

The infectious time is obtained from clinical data, which on average lasts for 9 days with a standard deviation of 3 days for COVID-19 [14]. The CFR is unknown and need to be estimated. However, to simplify the calculation, in this paper we assumed that it is equal to the last data of the number of deceased case divided by the total infected case. To account for under-reported case, this estimation can be divided by a correction factor  $c$ , e.g.,

$$\text{CFR} = \frac{1}{c} \frac{D(k_f)}{I(k_f)}, \quad (9)$$

where  $k_f$  denotes the index of the latest data. In our example in Section 3, we assumed the under-reported case is 3 times larger that the reported case. Thus, we take  $c = 4$ .

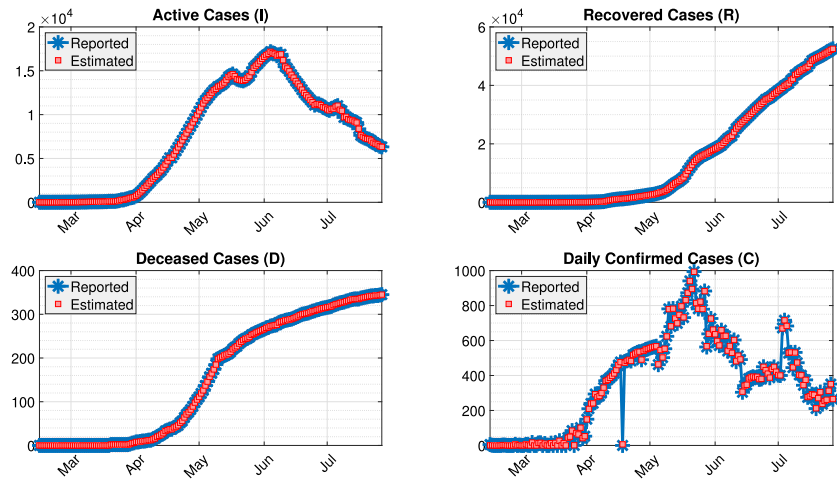


Fig. 3. Data fitting using EKF for UAE.

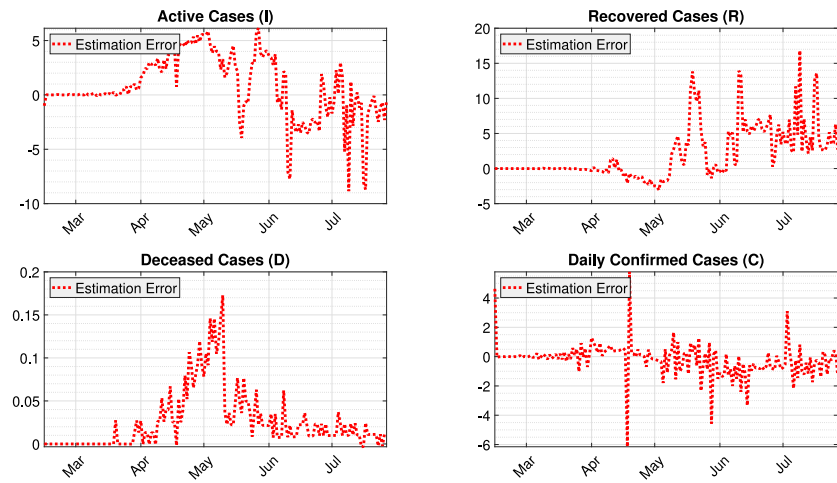


Fig. 4. Error between reported and estimated cases in UAE.

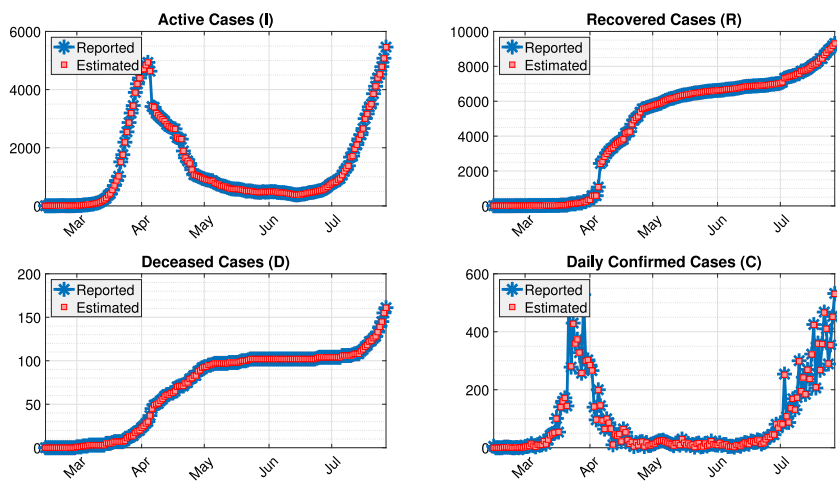


Fig. 5. Data fitting using EKF for Australia.

#### 2.4. Estimation of the time-varying effective reproduction number

The time-varying effective reproduction number  $\mathcal{R}_t(k)$  is estimated by applying EKF to the discrete-time stochastic augmented

compartmental model (1)–(6). Let us define

$$\mathbf{x}(k) = (S(k) \ I(k) \ R(k) \ D(k) \ C(k) \ \beta(k))^T. \quad (10)$$

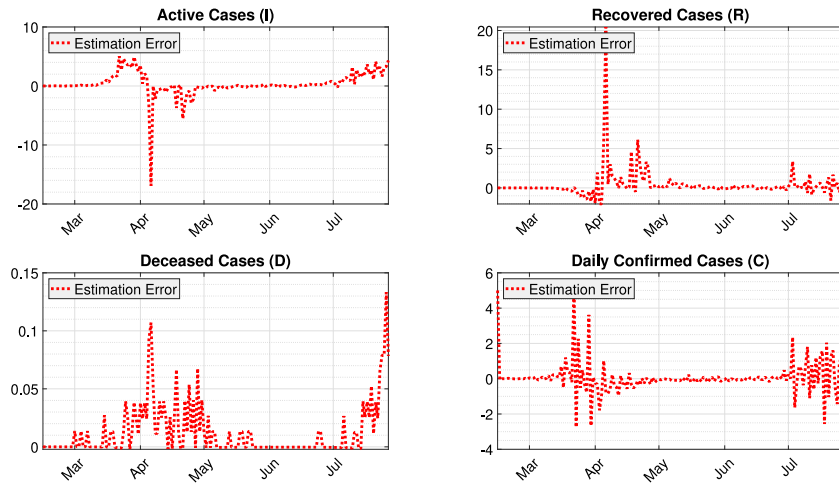


Fig. 6. Error between reported and estimated cases in Australia.

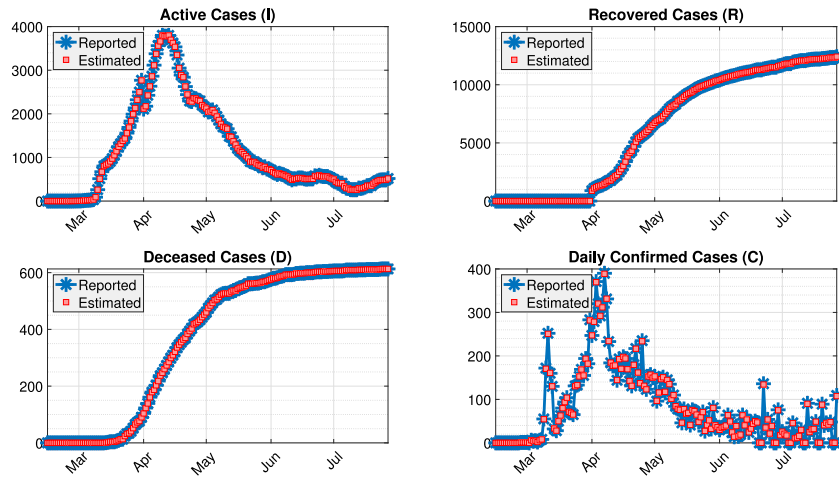


Fig. 7. Data fitting using EKF for Denmark.

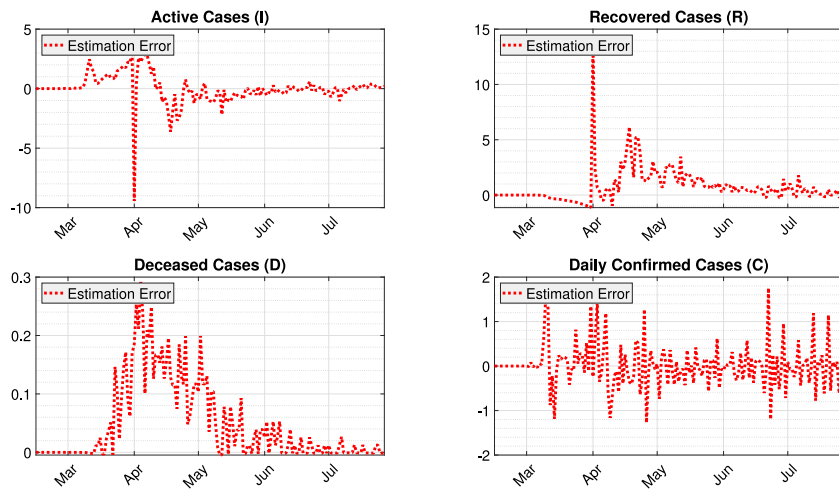


Fig. 8. Error between reported and estimated cases in Denmark.

The discrete-time augmented SIRD model (1)–(6) can be written as

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{w}(k), \quad (11)$$

where  $\mathbf{f}$  is the right hand side of (1)–(6). Let us denote  $\hat{\mathbf{x}}(k)$  as the estimate of  $\mathbf{x}(k)$  from the EKF. The EKF algorithm requires the



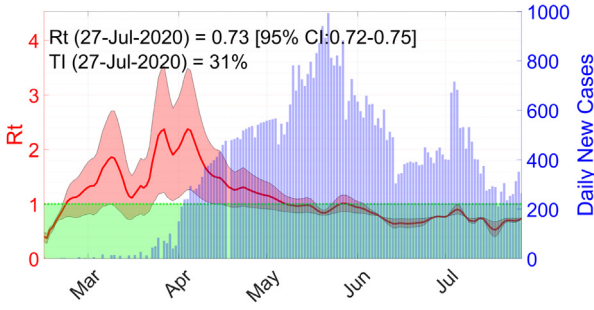


Fig. 9. The time-varying effective reproduction number  $\mathcal{R}_t$  in UAE.

Jacobian of  $\mathbf{f}$  at the estimate  $\hat{\mathbf{x}}(k)$ , that is given as

$$\mathbf{J}_f(\hat{\mathbf{x}}(k)) = \begin{pmatrix} J_{11}(\hat{\mathbf{x}}(k)) & J_{12}(\hat{\mathbf{x}}(k)) & 0 & 0 & 0 & J_{16}(\hat{\mathbf{x}}(k)) \\ J_{21}(\hat{\mathbf{x}}(k)) & J_{22}(\hat{\mathbf{x}}(k)) & 0 & 0 & 0 & J_{26}(\hat{\mathbf{x}}(k)) \\ 0 & \gamma \Delta t & 1 & 0 & 0 & 0 \\ 0 & \kappa \Delta t & 0 & 1 & 0 & 0 \\ 0 & (\gamma + \kappa) \Delta t & 0 & 0 & 1 - \Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (12)$$

where

$$J_{11}(\hat{\mathbf{x}}(k)) = 1 - \frac{\Delta t}{N} \hat{\beta}(k) \hat{I}(k), \quad (13)$$

$$J_{12}(\hat{\mathbf{x}}(k)) = -\frac{\Delta t}{N} \hat{\beta}(k) \hat{S}(k), \quad (14)$$

$$J_{16}(\hat{\mathbf{x}}(k)) = -\frac{\Delta t}{N} \hat{I}(k) \hat{S}(k), \quad (15)$$

$$J_{21}(\hat{\mathbf{x}}(k)) = \frac{\Delta t}{N} \hat{\beta}(k) \hat{I}(k), \quad (16)$$

$$J_{22}(\hat{\mathbf{x}}(k)) = 1 - (\gamma + \kappa) \Delta t + \frac{\Delta t}{N} \hat{\beta}(k) \hat{S}(k), \quad (17)$$

$$J_{26}(\hat{\mathbf{x}}(k)) = \frac{\Delta t}{N} \hat{I}(k) \hat{S}(k). \quad (18)$$

Detailed numerical implementation of the model can be found in [7]. Our algorithm has two tuning parameters: the covariance of the process noise  $\mathbf{Q}$  and the covariance of the observation noise  $\mathbf{R}$ , which can be chosen such that the Relative Root Mean Square Error (RRMSE) between the reported and estimated data

is minimized. The RRMSE for each variable  $X$  is defined as

$$\text{RRMSE}_X = \frac{1}{N_d} \sum_{i=1}^{N_d} \frac{\|X_i - \hat{X}_i\|_2^2}{\|X_i\|_2^2}, \quad (19)$$

where  $N_d$  is the number of observed days. Here,  $X_i = \{I(i), R(i), D(i), C(i)\}$  and  $\hat{X}_i = \{\hat{I}(i), \hat{R}(i), \hat{D}(i), \hat{C}(i)\}$  denote the reported and estimated data, respectively. The EKF serves as a real-time data fitting and will estimate any reported new data. Once this estimation process works, the EKF will also produce an estimate of  $\mathcal{R}_t$  from (6).

## 2.5. Forecasting

Forecasting is done for different public measure scenarios. Up to this point, we define TI as

$$\text{TI}(k) = \frac{\mathcal{R}_t(k)}{\mathcal{R}_t^{\max}(k)}. \quad (20)$$

Here  $\mathcal{R}_t^{\max}(k)$  can be interpreted as the basic reproduction number ( $\mathcal{R}_0$ ). The outcomes of different public measure scenarios are obtained through assigning different designated values of TI at the end of the prediction horizon, as illustrated in Fig. 2.

Here, the prediction horizon is 90 days. In this case, we draw a straight line perturbed by a white Gaussian noise between the current TI and the designated TI. The noise is added to simulate fluctuation in the reported cases. In our case, the designated TIs are ranging from 20% to 100%. Table 2 shows a hypothetical but rational relationship between different public measure scenarios and designated TIs. Relaxing public measures will correspond to a smaller value of designated TI and vice versa. Based on the different values, we use the discrete-time compartmental model (1)–(5) to forecast the outcome of different public measure scenarios.

## 3. Result and discussion

In this section, we run simulations for three countries: UAE, Australia, and Denmark. All data sets and MATLAB codes are available on GitHub through this link: <https://github.com/agusisma/covid19datadriven>. Parameters for the simulations are presented in Table 3. The recovery rate  $\gamma$  and the death rate  $\kappa$  are calculated using (8).

Figs. 3–8 show data fitting results using EKF and errors between reported and estimated cases in UAE, Australia, and Denmark. These figures show that our EKF algorithm estimate the

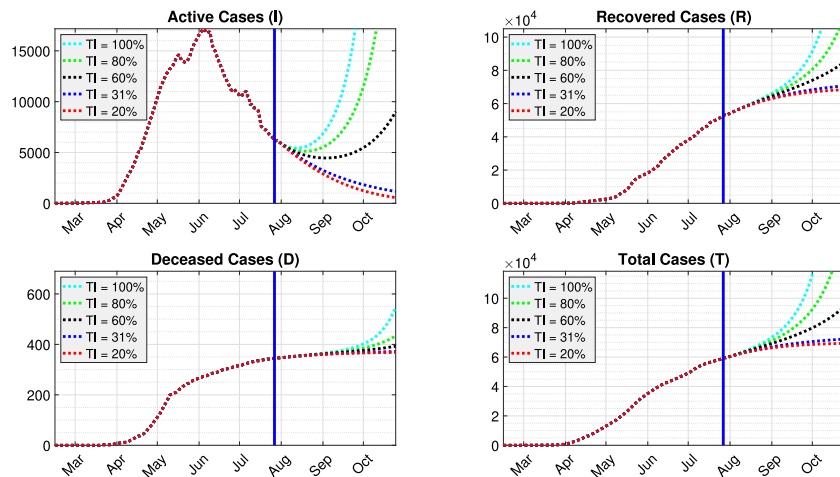


Fig. 10. Forecasting for the next 90 days in UAE.

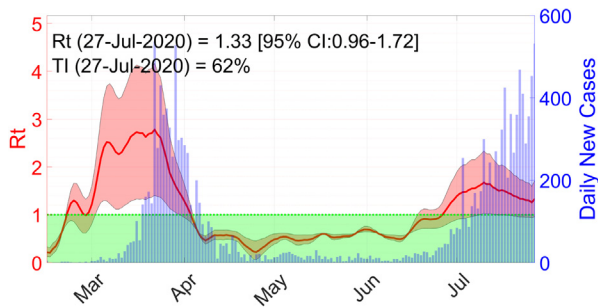


Fig. 11. The time-varying effective reproduction number  $\mathcal{R}_t$  in Australia.

actual reported cases accurately (see also Table 4). When estimating these cases, the EKF also estimates the time-varying effective reproduction number  $\mathcal{R}_t$ .

### 3.1. United Arab Emirates

The UAE has conducted more than 4.9 million tests since the outbreak or 502.14 total tests per thousand population. This brought UAE as one of the countries with the highest number of tests. The study of [15] estimated that the percentage of symptomatic COVID-19 cases reported in UAE using CFR estimates is at 98% (86%–100% of 95% credible interval).

The first confirmed cases were reported on 29 January, from an infected family of four who came to the country on holiday from Wuhan. As the number of positive cases steadily increased, the government took immediate public measures, such as the closure of schools and universities across the country until the end of the academic year in June (announced on 3 March), the suspension of prayers at mosques and all other places of worship from 16 March including the whole month of Ramadan, as well as night curfews for disinfection on 26 March for an extended period of time that limited movements within the country.

The extreme measures together with the government's wider National Screening Program, which seeks to test as many people as possible with the aims to identify, isolate and treat patients as quickly as possible, yielded positive results in the decrease of the reproduction number almost immediately afterwards, see Fig. 9. On 18 May the government announced the first day where the number of recoveries surpasses the number of new cases found. In the third week of May, the daily new cases reached its peak

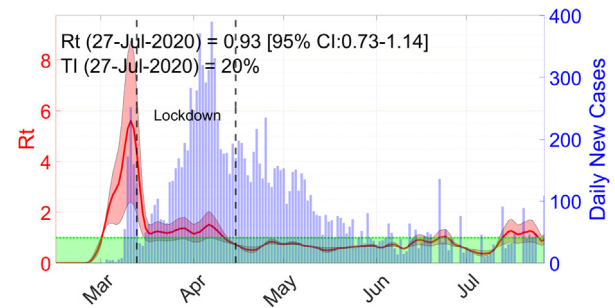


Fig. 13. The time-varying effective reproduction number  $\mathcal{R}_t$  in Denmark.

and the reproduction number was already below the threshold value  $\mathcal{R}_t = 1$ .

With the current TI at 31%, our forecast shows that the daily cases will be steadily decreasing, see Fig. 10. As restrictions in the country ease, it is important to maintain safety and preventive measures. Public negligence can increase the TI, which can result in a second wave of infection in the country.

### 3.2. Australia

The first confirmed case of COVID-19 in Australia was found in January 2020 when a traveler went back to Victoria from Wuhan, China. The number of incidence passed 1000 in March 2020 and doubled after three days. The growth of incidence during March and April is considered as the first wave of pandemic with  $\mathcal{R}_t = 3$  (see Fig. 11). The effects of pandemic in health sectors started to impact other sectors such as trade, travel, economic and finance and intensive interventions to prevent the pandemic from growing have been done [4].

The Australian Government closed the borders to all non-residents and non-citizens on 20 March 2020 and applied a 14-day self isolation for all arrivals. Quarantine/lock-down related policy such as physical distancing or self-isolation policy were applied in the form of school and workplace closure, mass gathering cancellation, contact tracing, etc. Also all non-essential services were stopped to maximize the physical distancing. The policy was extended for the next three months [16].

The effect of the interventions was indicated by the value of  $\mathcal{R}_t < 1$  in April, where the number of new cases dropped from 350 to 20 cases per day, as shown by Fig. 11.

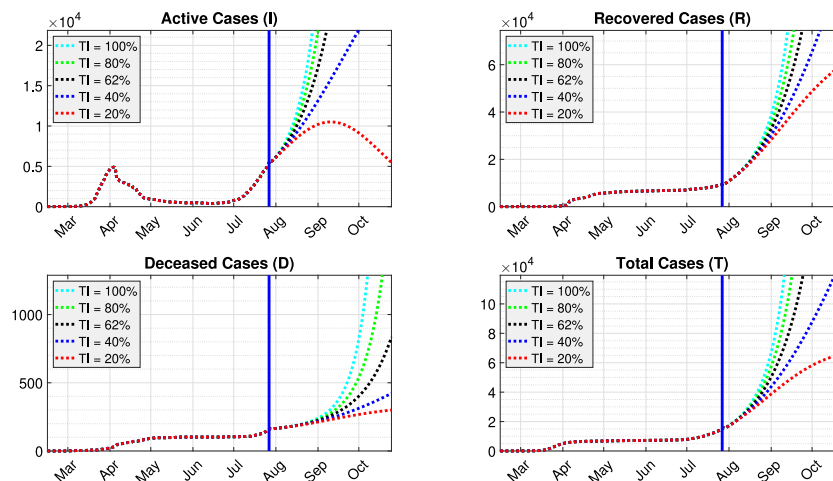


Fig. 12. Forecasting for the next 90 days in Australia.

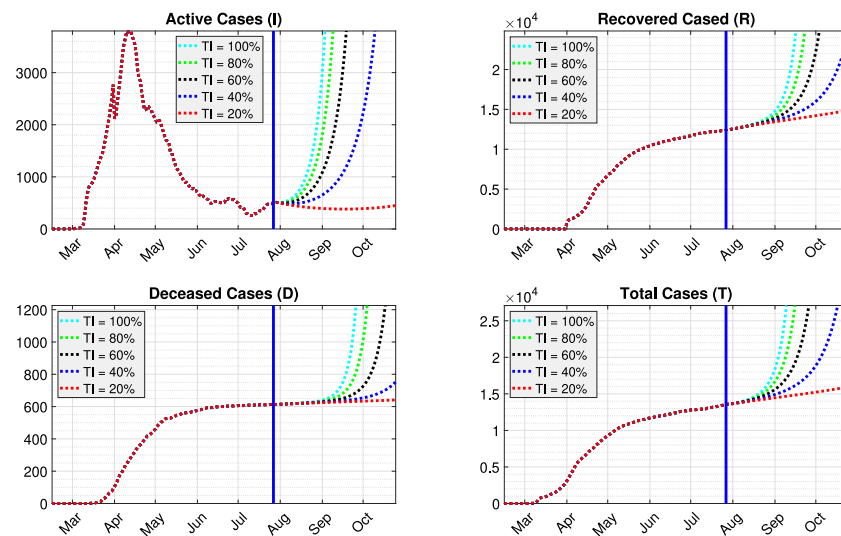


Fig. 14. Forecasting for the next 90 days in Denmark.

As the quarantine related policy was lifted in the beginning of June after a slow rate of infection ( $\mathcal{R}_t < 1$ ) for three months, there has been escalation in the number of positive cases. This urged the Australian Government applied the policy again in order to prevent not only higher cases in the second wave but also a long-term impact to all sectors. Australia TI by 27 July was 62% and its short-term projection showed that the active cases will increase sharply if there is no further intervention. The number of recovered and deceased cases will increase in the next 90 days. The short-term projection is shown in Fig. 12.

### 3.3. Denmark

Denmark confirmed its first case on 27 February, when a man who had been skiing in Lombardy in Italy returned to Denmark. The country introduced lock-down on 13 March, by ordering people working in non-essential functions in the public sector to stay at home for two weeks. Furthermore, kindergarten, primary and secondary schools, universities, libraries, indoor cultural institutions and similar places were closed. Assembly of more than ten people in public were made illegal.

The effect of the lock-down can be observed after three weeks (see Fig. 13), when  $\mathcal{R}_t < 1$ . A very slow and gradual reopening was initiated on 15 of April, by opening nurseries, kindergartens, and primary schools. However, the government will re-enforce lock-down if there are indications that the number of infections increases quickly. Universities are open only for employees, while all courses will be given online. At the end of July, there was an increase in the number of infected people, possibly due to the summer holiday. However, the government did not enforce any restriction.

Denmark's TI by 27 July was 20%. A short-term projection showed the number of active cases will be steady under the current measures. The number of recovered individuals will increase, while the number of death will decrease significantly, as can be seen from Fig. 14.

## 4. Conclusion

We have presented a data-driven approach for modeling and forecasting of COVID-19 outbreak for public policy making. The method we proposed is relied on the quality of the data. Thus, the estimated  $\mathcal{R}_t$  and the forecast results need to be carefully interpreted when the number of testing is not sufficient. By defining TI

of a country as the ratio between its current reproduction number and the highest one ever reached by the country and estimating the index, our approach can be used to produce a short-to-medium term forecast that may predict the course of COVID-19 in the near future, including the probability of an upcoming second wave. Simulation results using data from three countries showed that our approach gives a reasonable forecast and insights.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. *Lancet Infect Dis* 2020;20(6):631–3. [http://dx.doi.org/10.1016/S1473-3099\(20\)30190-0](http://dx.doi.org/10.1016/S1473-3099(20)30190-0).
- [2] Jia J, Ding J, Liu S, Liao G, Li J, Duan B, Wang G, Zhang R. Modeling the control of COVID-19: Impact of policy interventions and meteorological factors. *Electron J Differential Equations* 2020;23:1–24.
- [3] Khoshnaw SH, Shahzad M, Ali M, Sultan F. A quantitative and qualitative analysis of the COVID-19 pandemic model. *Chaos Solitons Fractals* 2020;138:1–10. <http://dx.doi.org/10.1016/j.chaos.2020.109932>.
- [4] Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Commun* 2020;11(5710):1–13.
- [5] Pang L, Liu S, Zhang X, Tian T, Zhao Z. Transmission dynamics and control strategies of COVID-19 in Wuhan, China. *J Biol Systems* 2020;28(3):1–18.
- [6] Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in lombardy, Italy: Early experience and forecast during an emergency response. *JAMA* 2020;323:1545–6.
- [7] Hasan A, Susanto H, Tjahjono VR, Kusdiantara R, Putri ERM, Hadisoemarto P, Nuraini N. A new estimation method for COVID-19 time-varying reproduction number using active cases. 2020. [arXiv:2006.03766](https://arxiv.org/abs/2006.03766).
- [8] Egger M, Johnson L, Althaus C, Schöni A, Salanti G, Low N, Norris SL. Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. *F1000Research* 2017;6(1584):1–24. <http://dx.doi.org/10.12688/f1000research.12367.2>.
- [9] Susanto H, Tjahjono V, Hasan A, Kasim M, Nuraini N, Putri E, Kusdiantara R, Kurniawan H. How many can you infect? Simple (and naive) methods of estimating the reproduction number. *Commun Biomath Sci* 2020;3(1):1–9.
- [10] Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One* 2020;15(3):1–21.
- [11] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect Dis Modell* 2020;5:256–63.
- [12] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* 2020;134:1–5.



- [13] Brauer F, Castillo-Chavez C, Feng Z. Mathematical models in epidemiology. Texts in applied mathematics, 1st ed.. Springer-Verlag; 2019.
- [14] Kai-Wang To K, Tak-Yin Tsang O, Leung W-S, Raymond Tam A, Wu T-C, Christopher Lung D, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis* 2020;20:565–74.
- [15] Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CAB, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med* 2020;18(332):1–9.
- [16] Moloney K, Moloney S. Australian quarantine policy: From centralization to coordination with mid-pandemic COVID-19 shifts. *Public Adm Rev* 2020;80(4):671–82.