Computers in Biology and Medicine 137 (2021) 104810

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling

Xinhe Zhu [a,*], Bingbing Gao [b], Yongmin Zhong [a], Chengfan Gu [c], Kup-Sze Choi [c]

[a] School of Engineering, RMIT University, Victoria, Australia
[b] School of Automation, Northwestern Polytechnical University, Xi'an, China
[c] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
COVID-19 modelling
Stochastic epidemiological model
Social distancing
Re-infection
And extended kalman filter

ABSTRACT

This paper presents a new stochastic-based method for modelling and analysis of COVID-19 spread. A new deterministic Susceptible, Exposed, Infectious, Recovered (Re-infected) and Deceased-based Social Distancing model, named SEIR(R)D-SD, is proposed by introducing the re-infection rate and social distancing factor into the traditional SEIRD (Susceptible, Exposed, Infectious, Recovered and Deceased) model to account for the effects of re-infection and social distancing on COVID-19 spread. The deterministic SEIRD(R)D-SD model is further converted into the stochastic form to account for uncertainties involved in COVID-19 spread. Based on this, an extended Kalman filter (EKF) is developed based on the stochastic SEIR(R)D-SD model to simultaneously estimate both model parameters and transmission state of COVID-19 spread. Simulation results and comparison analyses demonstrate that the proposed method can effectively account for the re-infection and social distancing as well as uncertain effects on COVID-19 spread, leading to improved accuracy for prediction of COVID-19 spread.

1. Introduction

At the end of 2019, the novel coronavirus, SARS-CoV-2 (COVID-19), firstly appeared in Wuhan, the city of Hubei province in the People's Republic of China and spread rapidly around the world. The pathogenesis of the virus is characterized by respiratory tract infection, which directly leads to pneumonia showing ground glass alveolar angiography. The COVID-19 virus is contagious from the people infected even though they may not show symptoms (asymptomatic infections). Although China, United States, Italy, Australia and other countries have successively adopted various containment and detection measures, the cumulative number of diagnoses is still increasing every day. The occasional rebound has also hampered the implementation of economic recovery plans. In order to better control and monitor the epidemic, mathematical modelling of COVID-19 becomes an area of active research.

Predictive mathematical models for epidemics are essential to understand the propagative characteristics of COVID-19 and to implement the intervention and preparedness measures for controlling the disease spread. The current existing research efforts on prediction of infectious diseases are mainly dominated by the agent-based and compartmental models. The agent-based model involves a complex process to define agent behaviours together with their associated interaction mechanisms and intervention rules, leading to expensive computations. Chang et al. developed an agent-based model to study the effect of social distancing (SD) compliance on COVID-19 spread [28]. Kerr et al. proposed a methodology of COVID-19 agent-based simulator, which is used to explore different intervention scenarios [29]. However, both methods depend on a large number of samples and rules, leading to the difficulty of parameter identification. They also require an expensive sensitivity analysis to determine the prediction robustness. Therefore, these two methods can only use a relatively small number of agents for COVID-19 modelling, leading to the limited modelling accuracy.

Comparing to the agent-based model, the compartmental model is simple and correlated to observation [28]. It involves a dynamic process based on how the population is divided into different compartments to describe the transmission state [1]. The SIR (Susceptible, Infectious and Recovered) model divides the population into the susceptible, infectious and recovered compartments to describe the state of disease spread, where people who are susceptible to infection will possibly be infected, and the infected people will be recovered with a certain rate. The SEIR (Susceptible, Exposed, Infectious and Recovered) model introduces the

* Corresponding author.
E-mail addresses: xinhe.zhu@rmit.edu.au, xinhe.zhu@outlook.com (X. Zhu).

https://doi.org/10.1016/j.compbiomed.2021.104810
Received 15 June 2021; Received in revised form 24 August 2021; Accepted 24 August 2021
Available online 28 August 2021

exposed compartment into the SIR model to describe the intermediate state between the susceptible and infected people. Since the ferocity of the epidemic has claimed many lives, its lethality cannot be ignored. However, neither SIR nor SEIR considers the lethality of the disease. The deceased compartment is thus introduced in parallel to the recovered compartment to describe the possibility of disease transmission from infected people via the transmitting rate from the infected to deceased compartment [2]. The SIRD (Susceptible, Infectious, Recovered and Deceased) model introduces the deceased compartment into the SIR model to consider the fatal condition. Similarly, the SEIRD (Susceptible, Exposed, Infectious, Recovered and Deceased) model introduces the deceased compartment into the SEIR model to describe disease transmission between humans.

Since the outbreak of COVID-19 pandemic, especially in the absence of vaccines, various SD measures such as flight restriction, school closure, indoor activity restrictions and quarantine [28] have been widely adopted by governments to reduce the cross-infection possibility [3,4]. Therefore, it is necessary to account for the effect of SD compliance on COVID-19 spread in epidemiological modelling. Further, patients cannot develop the lifelong immunity after recovery and the SARS-CoV-2 virus mutates over time, causing immune evasion [5,8]. Therefore, it is also necessary to take into account the deceased compartment and the re-infection rate from the recovered to susceptible compartment into epidemiological modelling. Hagger et al. studied a social cognition model by taking into account the distance between individuals in the SEIR model [6]. However, this model does not consider the deceased people and re-infection effect. Malkov studied how the possibility of reinfection shapes the epidemiological dynamic based on the SEIR model [31]. However, the lethality of COVID-19 is not considered. Further, all containment measures are integrated into the transmission rate, unable to characterize the effects of various kinds of containment measures on COVID-19 spread.

Distinct from other infectious diseases, COVID-19 has a randomly variable incubation period. It mutates with varying infectivity and pathogenicity (e.g. the B.1.1.7 and B.1.617.2 variants have increased infectivity and shorter incubation period of about 24 h), making the incubation period of COVID-19 and its associated infection rate involve randomness [8,27]. Moreover, since the potential sources of infection are unknown, asymptomatic infections are difficult to detect, resulting in uncertainties in reported infection cases [7]. The inadequate contact tracing, lack of population-wide PCR testing and short-term policy changes also cause the uncertainties in reported data on COVID-19 [32]. However, the existing studies on COVID-19 modelling are dominated by deterministic epidemiological models for describing the epidemiological evolution deterministically via ordinary differential equations, unable to model the stochastic behaviours of the COVID-19 epidemic [9,10]. Therefore, it is also necessary to develop a stochastic epidemiological model to account for random or stochastic events involved in the COVID-19 transmission system.

In addition to an epidemiological model, dynamic modelling of COVID-19 pandemic also requires a real-time algorithm to estimate the transmission state online. The recursive least square (RLS) is a traditional method for online parameter estimation in epidemiological modelling [1,11]. It can generate optimal state estimation via minimizing the linear least-squares cost function related to system observations. As an improvement of RLS, the Kalman filter (KF) introduces the system state equation in RLS to calculate the propagation of system state via a prediction process. It can achieve optimal state estimation in the accuracy of minimum mean square error, even in the absence of observations. Kumar et al. developed a KF based on a linear forest regression model that describes the correlations between infected individuals to predict the future trend of COVID-19 [11]. Arroyo-Marioli et al. used KF to estimate the basic reproduction number of COVID-19 based on a linearized form of the SIR model [2]. Nevertheless, RLS and KF can be applied to linear systems only [12], while the existing epidemiological models for COVID-19 forecast are nonlinear. The constrained least-squares (CLS) [13] and Markov chain Monte Carlo (MCMC) [14] are the commonly used estimation method for nonlinear epidemiological models. However, since CLS and MCMC are based on maximum posteriori estimates of probability density function, the accuracy of both methods heavily depends on the sample size [13]. Further, both methods also involve expensive computations and can only be conducted in an offline manner. Accordingly, they are unsuitable for characterising random uncertainties involved in epidemiological model parameters for COVID-19 prediction.

As an improvement of KF for nonlinear systems, the extended Kalman filter (EKF) dynamically linearizes the nonlinear system model to employ the traditional KF for online state estimation. Comparing to CLS and MCMC, EKF is a simple iterative algorithm with significant computational efficiency for nonlinear epidemiological modelling [15,16]. So far, there has been very limited research on using EKF for epidemiological modelling, especially in COVID-19. Just recently, Hassan et al. developed an EKF based on the SEIR model for modelling of COVID-19 spread, but without considering the exposed patients and incubation period [17]. Younes and Hassan developed an EKF based on the Lotka-Volterra model to estimate COVID-19 spread [18]. However, the predator-prey interaction mechanism described by the Lotka-Volterra model has a very limited capacity to model the complex characteristics of the natural transmission process of COVID-19. Song et al. studied a novel maximum likelihood based EKF to estimate COVID-19 spread [9]. However, since this method is based on a deterministic epidemiological model, it is incapable of characterizing the stochastic characteristics of COVID-19 spread. Further, it does not consider the re-infection and SD effects either.

This paper presents a new stochastic-based method for estimation and prediction of COVID-19 spread. This method introduces the deceased compartment with the death rate, and the re-infection rate that characterizes the transmission from the infected back to susceptible compartment, into the SEIRD model, leading to a new deterministic Susceptible, Exposed, Infectious, Recovered (Re-infected) and Deceased-based Social Distancing model, named SEIR(R)D-SD, to account for both re-infection and the SD effects of COVID-19. Subsequently, the stochastic version of the deterministic SEIR(R)D-SD model is constructed according to the probabilities of independent random changes occurred in the system to account for the stochastic characteristics of COVID-19 spread. Based on this framework, an EKF algorithm is developed for online estimation of the spreading behaviours of COVID-19, where the system state equation and system observation equation are constructed and the model parameters of SEIR(R)D-SD are also augmented into the system state to simultaneously estimate both model parameters and system state. Simulation results are consistent with the COVID-19 epidemic in Australia, where the multiple outbreak waves are accurately captured by the proposed method [33]. They also reveal that SD restriction can postpone the COVID-19 outbreak and the re-infection rate can reflect the non-lifelong immunity characteristic of COVID-19.

2. Methodology

2.1. SEIR(R)D-SD model

The SEIRD model has an additive deceased compartment associated with the death rate [19,20]. It constitutes the following time-continuous deterministic system

$$\begin{cases} \frac{dS}{dt} = \frac{-\beta S}{N} I \\ \frac{dE}{dt} = \frac{\beta S}{N} I - \alpha E \\ \frac{dI}{dt} = \alpha E - \gamma I \\ \frac{dR}{dt} = \gamma I \\ \frac{dD}{dt} = \mu I \end{cases} \tag{1}$$

where S, E, I, R and D denote the susceptible, exposed, infectious, recovered and deceased compartments, N is the total population; and α, β, γ and μ are the model parameters, where α is the infection rate which is the inverse of the incubation period, β is the exposing rate, γ is the recovery rate, and μ is the death rate. For simplicity and consideration of the limited immigration-emigration effect on population due to border restriction policies, N is generally considered as a constant for COVID-19 modelling [30,33].

Suppose that the community of all the compartments is closed, i.e.,

$$S + E + I + R + D = N \tag{2}$$

By introducing the SD factor ρ into the deterministic SEIRD model to study the SD effect and by introducing a factor κ in S and R to represent the rate from the recovered to susceptible compartment to account for the re-infected population [20], the deterministic SEIR(R)D-SD model can be written as

$$\frac{dS}{dt} = \frac{-\rho\beta IS}{N} + \kappa R \tag{3}$$

$$\frac{dE}{dt} = \frac{\rho\beta IS}{N} - \alpha E \tag{4}$$

$$\frac{dI}{dt} = \alpha E - (\gamma + \mu) I \tag{5}$$

$$\frac{dR}{dt} = \gamma I - \kappa R \tag{6}$$

$$\frac{dD}{dt} = \mu I \tag{7}$$

where κ is the re-infection rate, and ρ is the SD factor which is dynamically changed with the compliance levels of various SD policies such as travel restriction, lockdown or semi-lockdown, self-quarantine and school closures [28].

Fig. 1 illustrates the structure of the SEIR(R)D-SD model. The deceased compartment is outside the loop of disease spread to leave the community with the death rate μ, while the rest of the compartments form a loop, where the susceptible compartment which is initially presumed as the total population size, is transferred to the exposed compartment with the exposing rate β and the SD factor ρ, and the recovered compartment will return to the susceptible compartment to cyclically transmit the disease in the community.

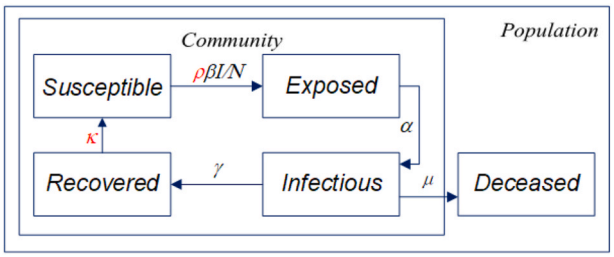

Fig. 1. Structure of the SEIR(R)D-SD model.

2.2. Stochastic SEIR(R)D-SD model

In this section, we discuss how to convert the above deterministic SEIRD(R)D-SD model into a stochastic model according to the possibilities of all the independent random changes occurred in the system.

According to (2),

$$R(t) = N - (S(t) + E(t) + I(t) + D(t)) \tag{8}$$

Substituting (8) into (3)–(7), the deterministic SEIR(R)D-SD model becomes

$$dS(t)/dt = (-\rho\beta I(t)S(t))/N + \kappa[N - (S(t) + I(t) + E(t) + D(t))] \tag{9}$$

$$dE(t)/dt = \rho\beta I(t)S(t)/N - \alpha E(t) \tag{10}$$

$$dI(t)/dt = \alpha E(t) - (\gamma + \mu) I(t) \tag{11}$$

$$dD(t)/dt = \mu I(t) \tag{12}$$

Equation (9)~(12) can be combined into the following form

$$d\boldsymbol{x}(t) = f(\boldsymbol{x}(t))dt \tag{13}$$

where $\boldsymbol{x}(t) = [S(t) \quad E(t) \quad I(t) \quad D(t)]^{\mathrm{T}}$ is the four-dimensional system state consisting of the susceptible, exposed, infected and death compartments, and $f(\cdot)$ is the nonlinear system function.

Discretizing (13) in time domain, we can have the following discrete-time SEIR(R)D-SD model

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + f(\boldsymbol{x}_k) \tag{14}$$

where $\boldsymbol{x}_k$ is the system state at time point k.

The deterministic discrete system (14) involves four random changes with each occurred to at least one of the state parameters. Define the jth random change as

$$\boldsymbol{r}_j = \begin{cases} \boldsymbol{\lambda}_j & \text{with probability } p_j \\ 0_{4\times 1} & \text{with probability } 1 - p_j \end{cases} (j = 1, 2, 3, 4) \tag{15}$$

where p_j denotes the probability of the jth change and $\boldsymbol{\lambda}_j$ denotes the transition of the system state under the jth change, both of which are obtained from (9)~(12) and given in Table 1.

By summing each random change for the system state, a simple stochastic form of (14) can be written as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \sum_{j=1}^{4} \boldsymbol{r}_j \tag{16}$$

where it is known from (15) that $\sum_{j=1}^{4} \boldsymbol{r}_j$ obeys the normal distribution with expectation $f(\boldsymbol{x}_k) = \sum_{j=1}^{4} p_j \boldsymbol{\lambda}_j$ and variance $\boldsymbol{G}(\boldsymbol{x}_k) = \sum_{j=1}^{4} p_j \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j$.

Approximating the random changes $\boldsymbol{r}_j$ using a normal random vector $\sigma_j \sim \mathcal{N}(0, 1)$ via the central limit theorem [21,22] yields

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + f(\boldsymbol{x}_k) + \boldsymbol{g}(\boldsymbol{\sigma}_k) \tag{17}$$

where $\boldsymbol{\sigma} = [\sigma_1 \quad \sigma_2 \quad \sigma_3 \quad \sigma_4]$, and $\boldsymbol{g}(\boldsymbol{\sigma}_k) = \boldsymbol{G}^{1/2}\boldsymbol{\sigma}_k$ which is subject to the Gaussian distribution.

Table 1
Random changes involved in the SEIR(R)D-SD model where Δt denotes the time step.

Transition of change	Probability
$\boldsymbol{\lambda}_1 = [-1 \quad 1 \quad 0 \quad 0]^{\mathrm{T}}$	$p_1 = \beta S_k I_k N^{-1} \Delta t$
$\boldsymbol{\lambda}_2 = [0 \quad -1 \quad 1 \quad 0]^{\mathrm{T}}$	$p_2 = \alpha E_t \Delta t$
$\boldsymbol{\lambda}_3 = [\kappa \quad 0 \quad -1 \quad 0]^{\mathrm{T}}$	$p_3 = (\gamma + \mu) I_t \Delta t$
$\boldsymbol{\lambda}_4 = [-\kappa \quad 0 \quad 0 \quad 1]^{\mathrm{T}}$	$p_4 = \mu D_t \Delta t$

2.3. System state and observation equations

Simplifying (17) yields

$$\boldsymbol{x}_{k+1} = \phi(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{18}$$

where $\phi(\boldsymbol{x}_k) = \boldsymbol{x}_k + f(\boldsymbol{x}_k)$ and $\boldsymbol{w}_k$ is the system noise.

Since the reported data are in terms of the infectious, recovered and dead compartments only, the system observation is constructed as

$$z_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{19}$$

where z_k is the system observation (i.e., the reported data); $\boldsymbol{v}_k$ is the observation noise which is assumed to be white noise with covariance $\boldsymbol{R}$ and is independent of $\boldsymbol{w}_k$; and $\boldsymbol{H}_k$ is the observation matrix which is expressed as

$$\boldsymbol{H}_k = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

The nonlinear function $\phi(\boldsymbol{x}_k)$ can be linearized as

$$\phi(\boldsymbol{x}_k) = \begin{Bmatrix} S_k - \beta\frac{S_k I_k}{N} + \kappa\left(N - \sum(S_k, E_k, I_k, D_k)\right) \\ E_k + \beta\frac{S_k I_k}{N} - \alpha E_k \\ I_k + \alpha E_k - (\gamma + \mu) I_k \\ D_k + \mu D_k \end{Bmatrix} \approx \text{Jacobian}(\phi(\boldsymbol{x}_k)) \begin{bmatrix} S_k \\ E_k \\ I_k \\ D_k \end{bmatrix} = \boldsymbol{F}_k \boldsymbol{x}_k \tag{21}$$

where $\boldsymbol{F}_k$ is the Jacobian matrix, which is expressed as

$$\boldsymbol{F}_k = \begin{bmatrix} 1 - \frac{\rho\beta I_k}{N} & 0 & -\frac{\rho\beta S_k}{N} - \kappa\left(N - \sum(S_k, E_k, I_k, D_k)\right) & 0 \\ -\frac{\rho\beta I_k}{N} & 1 - \alpha & \frac{\rho\beta S_k}{N} & 0 \\ 0 & \alpha & 1 - (\gamma + \mu) & 0 \\ 0 & 0 & \mu & 1 \end{bmatrix} \tag{22}$$

Since the epidemiological model parameters are random unknowns, they must also be estimated in the filtering process to account for their randomness. Accordingly, we augment the model parameters into the system state as

$$\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix} \tag{23}$$

where $\boldsymbol{X}_k$ is the augmented system state and $\boldsymbol{\theta}_k$. collects the model parameters including the infection rate α, exposing rate β, recovery rate γ, death rate μ , re-infection factor k and SD factor ρ.

Correspondingly, the system state equation (18) becomes

$$\boldsymbol{X}_{k+1} = \boldsymbol{\Phi}_k \boldsymbol{X}_k + \boldsymbol{W}_k. \tag{24}$$

where $\boldsymbol{W}_k$ is the process noise which is assumed as a white noise with covariance $\boldsymbol{Q}$, and $\boldsymbol{\Phi}_k$ is the augmented system function which is represented as

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{F}_k & 0 \\ 0 & \boldsymbol{I} \end{bmatrix} \tag{25}$$

where $\boldsymbol{I}$ is the 6×6 unit matrix.

The EKF procedure for estimating the model parameters and transmission state involves the following steps:

i) Set the initial system state and its associated covariance

$$\widehat{\boldsymbol{X}}_0 = E[\boldsymbol{X}_0] \tag{26}$$

$$\widehat{\boldsymbol{P}}_0 = E\left[\left(\boldsymbol{X}_0 - \widehat{\boldsymbol{X}}_0\right)\left(\boldsymbol{X}_0 - \widehat{\boldsymbol{X}}_0\right)^{\mathrm{T}}\right] \tag{27}$$

ii) Calculate the predicted state and its associated covariance

$$\widehat{\boldsymbol{X}}^{-}_{k+1|k} = \phi\left(\widehat{\boldsymbol{X}}_k\right) \tag{28}$$

$$\boldsymbol{P}^{-}_{k+1|k} = \boldsymbol{\Phi}_k \boldsymbol{P}_k \boldsymbol{\Phi}_k + \boldsymbol{Q}_k \tag{29}$$

iii) Calculate the Kalman gain

$$\boldsymbol{K}_k = \boldsymbol{P}^{-}_{k+1|k} \boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{P}^{-}_{k+1|k}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}_k\right) \tag{30}$$

iv) Update the estimated state and its associated covariance

$$\widehat{\boldsymbol{X}}_{k+1} = \widehat{\boldsymbol{X}}^{-}_{k+1|k} + \boldsymbol{K}_k\left(\widehat{z}_k - \boldsymbol{H}\widehat{\boldsymbol{X}}^{-}_{k+1|k}\right) \tag{31}$$

$$\widehat{\boldsymbol{P}}_{k+1} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}^{-}_{k+1|k}(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^{\mathrm{T}} + \boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k \tag{32}$$

v) Repeat (28)–(32) until all iterations are processed.

3. Performance evaluation

Simulations were conducted to comprehensively evaluate the performance of the proposed method for COVID-19 modelling in terms of the following aspects: (i) the effectiveness of the proposed deterministic SEIR(R)D-SD model in comparison with the classical SEIRD model; (ii) the effectiveness of the proposed EKF based on the stochastic SEIR(R)D-SD model in comparison with the numerical solution of the deterministic SEIR(R)D-SD model; and (iii) the effectiveness of the proposed EKF based on the stochastic SEIR(R)D-SD model in comparison with the numerical solution of the classical SEIRD and the proposed deterministic SEIRD(R)D-SD models based on CLS parameter identification for modelling of the pandemic in Australia from daily reported data. It should be noted that the COVID-19 pandemic in Australia is considered for verification and validation purposes only, while the proposed method is generic and independent of application cases and thus it can also be used to predict COVID-19 pandemics in any other countries/

Table 2
The values of the model parameters.

	α	β	γ	μ	κ	ρ
SEIRD model	0.14	0.6	0.12	0.01	0	0
Deterministic SEIR(R)D-SD model	0.14	0.6	0.12	0.01	0.005	0.6

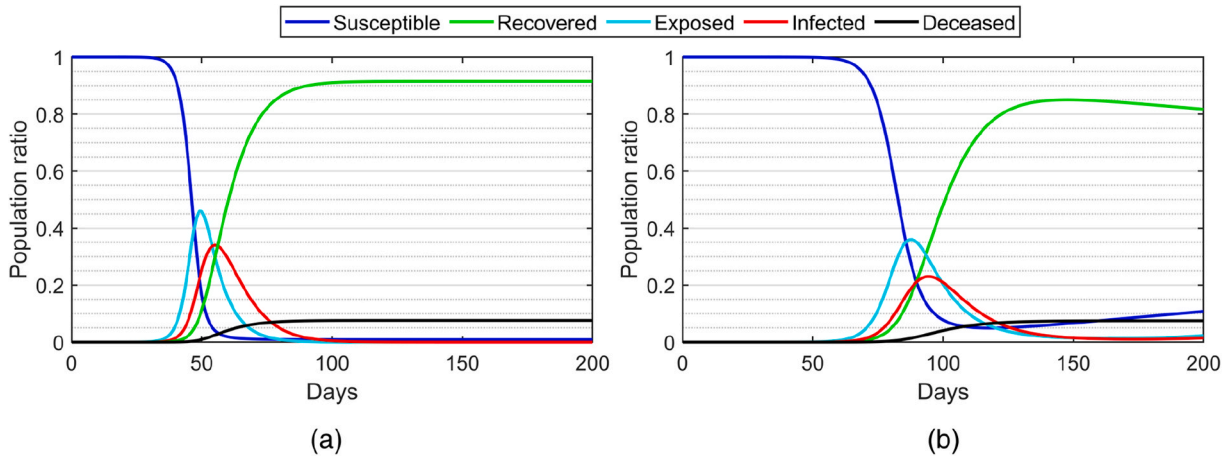

Fig. 2. The population ratios of the susceptible, exposed, infected, recovered and deceased compartments calculated by the numerical solutions of (a) the classical SEIRD model; and (b) the proposed deterministic SEIR(R)D-SD model ($\rho = 0.5$ and $\kappa = 0.005$).

regions.

The modelling accuracy is evaluated in terms of the root mean-squares error (RMSE), which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{N}\left(\widehat{\boldsymbol{X}}_k - \boldsymbol{X}_{true}\right)^2}{N}} \tag{33}$$

where $\boldsymbol{X}_{true}$ denotes the ground truth and N the number of samples.

3.1. Deterministic SEIR(R)D-SD model

Simulation tests were conducted to evaluate the performance of the deterministic SEIR(R)D-SD model in comparison with the classical SEIRD model. The simulation time was set to 200 days. The model parameters are given in Table 2.

Fig. 2(a) illustrates the numerical solution of the classical SEIRD model. The susceptible population decreases and the recovered population increases exponentially from day 40–100, whereas both exposed and infected populations quickly increase and then decrease after day 50. After day 100, the populations of all the compartments remain stable, where the susceptible, infected and exposed populations decrease to zero while the deceased population remains at a constant. For the deterministic SEIR(R)D-SD model, as shown in Fig. 2(b), due to the SD effect, the peaks of the exposed, infected and recovered populations are lower than those of the classical SEIRD model.

Despite the steep drop, the susceptible population does not drop to zero and slightly increases after day 100 due to the re-infection effect. Due to the SD effect, the peaks of the exposed and infected populations are delayed from day 50–100, comparing to those of the SEIRD model. Due to the transfer from the recovered to infected compartment via the re-infection rate, the infected population gradually increases after the steep decrease from day 50–100, while the recovered population gradually decreases after the steep increase from day 50–100. The results

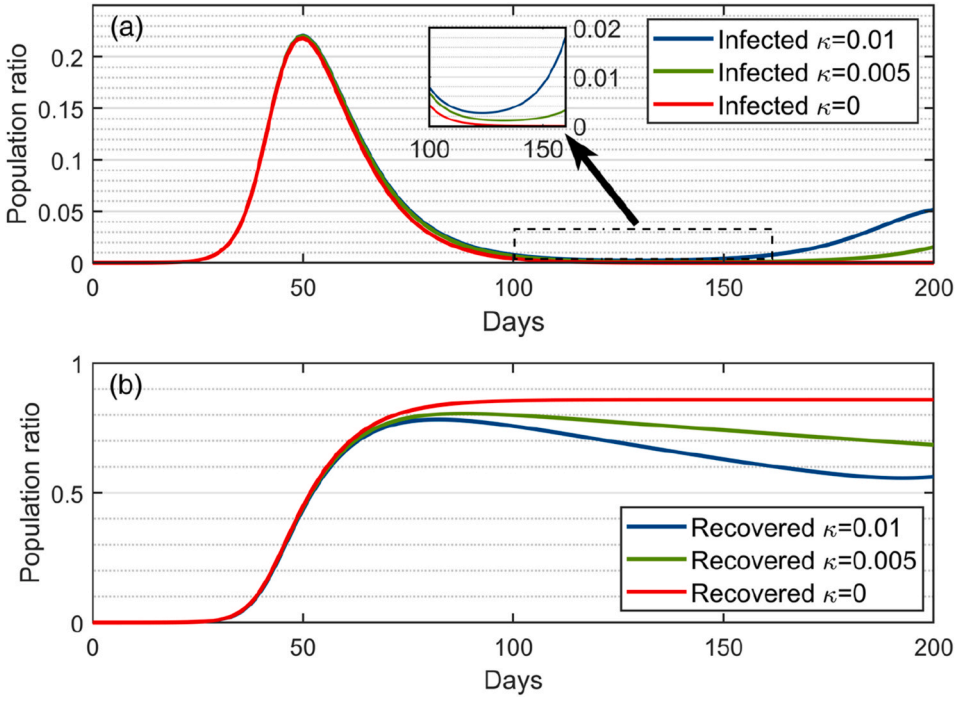

Fig. 3. The population ratios of the infected (a) and recovered (b) compartments calculated by the numerical solution of the deterministic SEIR(R)D-SD model under three different re-infection rates ($\kappa = 0$, $\kappa = 0.005$ and $\kappa = 0.01$).

show that SD restriction can postpone the disease outbreak and the re-infection rate can reflect the non-lifelong immunity characteristic of COVID-19.

Simulations were also conducted to evaluate the effects of the re-infection rate κ and SD factor ρ.

Fig. 3(a) illustrates the variations of the infected populations under three different κ values, i.e., $\kappa = 0$, $\kappa = 0.005$ and $\kappa = 0.01$, without involving the SD effect (i.e. the natural spread case of $\rho = 1$). In the case of $\kappa = 0$ (lifelong immunity), the infected population rises steeply from zero to the peak and then quickly reverts to zero and eventually remains at zero after day 120. In the case of $\kappa = 0.005$, the infected population has a similar trend as that in the case of $\kappa = 0$. before day 140, while it gradually increases after day 140 since the re-infected people cause the disease to spread in the community. In the case of $\kappa = 0.01$, the infected population has a similar trend as in the cases of $\kappa = 0$ and $\kappa = 0.005$ before day 130, whereas it gradually increases again after day 130. It can be seen that the larger the κ value is, the more the infected population will rebound, leading to the higher chance that the disease will break out again in the future.

Fig. 3(b) illustrates the trends of the recovered population ratio under the three different re-infection rates. Before day 50, the recovered populations for all the three cases have a similar trend and all rise quickly. This is because the recovered population is too small at the beginning of disease spread, leading to a similar re-infection effect for the three cases. After day 50, the recovered population becomes stable for the cases of $\kappa = 0$, while the recovered populations for both cases of $\kappa = 0.005$ and $\kappa = 0.01$gradually decrease, where the decrease in the case of $\kappa = 0.01$ is almost the twice of that in the case of $\kappa = 0.005$. The larger the re-infected rate is, the more the recovered population will decrease. The resultant increase of the infected population and the resultant decrease of the recovered population after day 130 indicate that future COVID-19 outbreaks may occur.

Simulation tests were also conducted to evaluate the SD effect. As shown in Fig. 4, both exposed and infected populations decrease significantly, and their peaks are also postponed when the SD factor ρ is changed from 1 to 0.5. The peak population ratios for both populations in the case of $\rho = 0.6$ are only about 20% and 40% smaller than those in the case without SD restriction (i.e., $\rho = 1$). Further, in the case of $\rho = 0.5$, the ratios of both populations are about 40% and 50% smaller than those in the case of natural spread (i.e., $\rho = 1$). This is because SD restriction directly reduces the transmission rate from the susceptible to exposed compartment. However, since SD restriction does not affect the total exposed population, the cases of SD restriction (i.e., $\rho = 0.6$ and $\rho = 0.5$) postpone the peaks of both exposed and infected populations for about 40 days and 60 days comparing to the case without SD restriction (i.e., $\rho = 1$). It can be seen from the results that the smaller the SD factor is, the lower the risk of disease outbreak will be. However, SD restriction can only keep the transmission rate from the susceptible to exposed compartment at a low level to delay disease outbreaks, while unable to prevent them.

3.2. EKF based on stochastic SEIR(R)-D model

To evaluate the stochastic SEIR(R)D-SD model and its associated EKF, the observation data were generated by adding a random white noise of covariance $\boldsymbol{Q} = 0.001$ in the numerical solution of the deterministic SEIR(R)D-SD as shown in Fig. 2(c) to simulate the actual reported data of COVID-19 spread that involve uncertainties.

Based on the observation data shown in Fig. 5, both model parameters and transmission state are estimated by the proposed EKF based on the stochastic SEIR(R)D-SD model. Fig. 6 illustrates the model parameters estimated by EKF with reference to their true values given in Table 2. It can be seen that the EKF estimations of the model parameters closely approximate their true values. As shown in Table 3, the estimation RMSEs (Root Mean Square Errors) are 0.0032, 0.011, 0.0018, 0.00085, 0.0036 and 0.012 for parameters α, β, γ, μ, κ and ρ, respectively, demonstrating that the proposed EKF based on the stochastic SEIR(R)D-SD model can effectively estimate the model parameters.

Fig. 7 illustrates the errors of the compartment populations estimated by the proposed EKF from the noisy observation data. The estimations of each compartment population resulted from the noisy observation data converge to their true values quickly, with the RMSE of 0.0032 for the susceptible, 0.0021 for the exposed, 0.0035 for the infected, and 0.00064 for the deceased population. This demonstrates that the proposed EKF based on the stochastic SEIR(R)D-SD model can effectively predict the transmission state from noisy observation data.

To further evaluate the stochastic SEIR(R)D-SD model, the non-noisy measurement data, i.e., the numerical solution shown in Fig. 2(b), were also used as observations to estimate the compartment populations by the proposed EKF based on the stochastic SEIR(R)D-SD model. As shown in Fig. 7, the estimations of each compartment population from the non-noisy observation data converge to their true values within a very short time period, demonstrating the stochastic SEIR(R)D-SD model is a particular case of the deterministic SEIR(R)D-SD model in the ideal condition. Table 4 lists the RMSEs of the compartment populations estimated by the proposed EKF based on the stochastic SEIR(R)D-SD model from both noisy and non-noisy observation data.

3.3. COVID-19 pandemic in Australia

In Australia, a hotspot of the COVID-19 pandemic, as of September 30, 2020, the Australian government reported 27,078 total confirmed cases and 886 deaths [23]. The pandemic started in January and reached the first peak on April 6, 2020. Its second wave started at the end of June 2020 and reached the peak on 9th August. The Australian government adopted a series of SD measures such as travel restriction policies, school

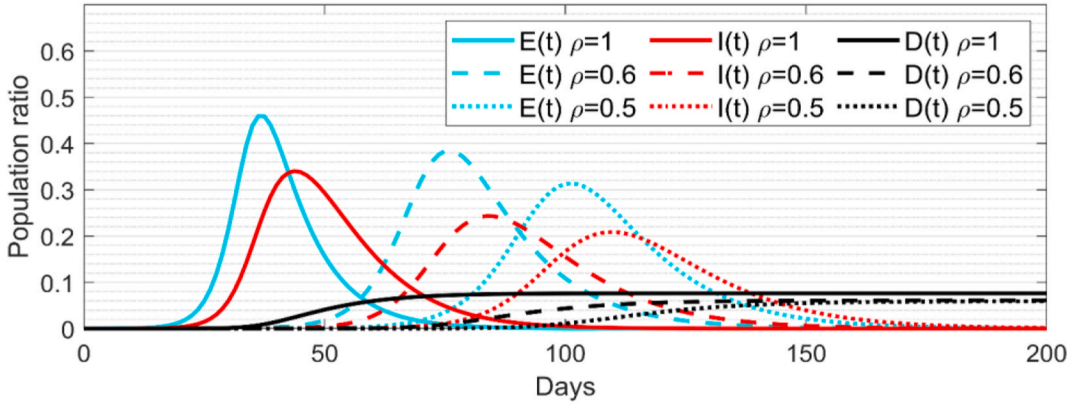

Fig. 4. The population ratios of the exposed, infected and deceased compartments calculated by the numerical solution of the deterministic SEIR(R)D-SD model.

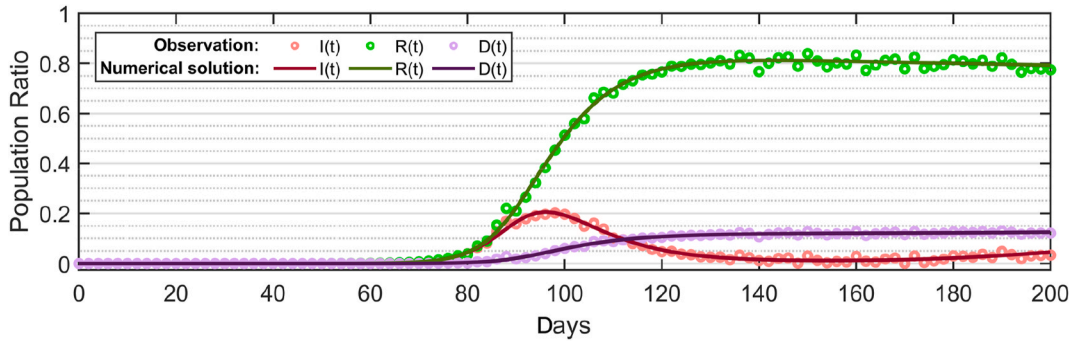

Fig. 5. Simulated observation data generated by adding a white noise in the numerical solution of the deterministic SEIR(R)D-SD model.

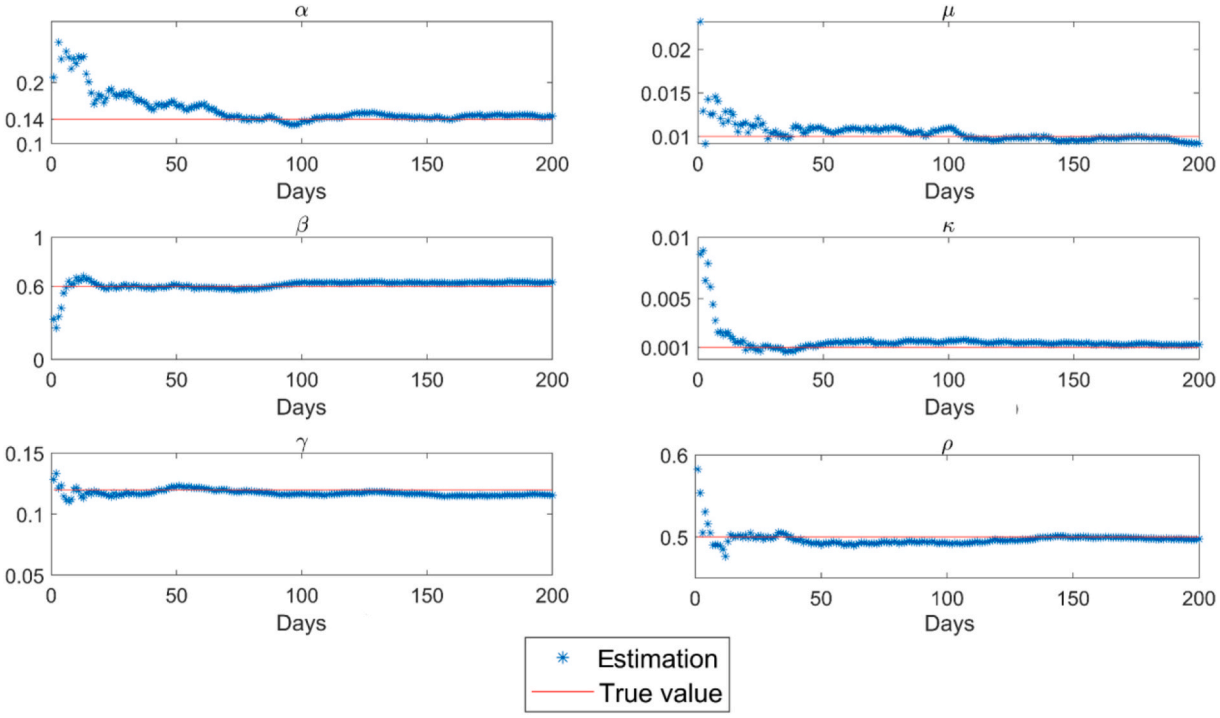

Fig. 6. Model parameters estimated by the proposed EKF based on the stochastic SEIR(R)D-SD model.

Table 3
RMSEs of the model parameters estimated by the proposed EKF based on the stochastic SEIR(R)D-SD model.

	α	β	γ	μ	κ	ρ
EKF	0.0032	0.011	0.0018	0.00085	0.0036	0.012

closure, indoor activity restrictions and quarantine to control the virus spread. The infected cases were almost vanished in the mid of June 2020. However, when the compliance level of SD restriction was relaxed at the end of May 2020 due to a good progress in controlling the first outbreak and the urgent desire for economic recovery, the second outbreak occurred at the end of June 2020.

Simulation trials were conducted by tracking and analysing the COVID-19 spread in Australia during the outbreak period of 230 days from 22nd January to September 8, 2020. According to the report of *United Nations Population Division* [24], the Australian population were about 25.5 million during the COVID-19 pandemic. We collected the reported cases within 230 days from 22nd January to September 8, 2020 from the World Health Organization (WHO) Novel Coronavirus Situation Report [25]. As shown by the reported data in Fig. 8, the COVID-19 spread in Australia has two outbreaks from 22nd January to September 8, 2020, where the first outbreak was occurred on about day 75 and the second outbreak started from day 160.

Since the true values for the actual cases are unknown, the reported data were taken as reference for calculation of estimation error. The initial state and noise covariances were set based on the observation data on the first day of the simulation analysis. The initial values of the model parameters are given in Table 5. For comparison analysis, the transmission state was also calculated from the SEIRD and deterministic SEIR(R)D-SD models based on parameter identification via the offline CLS algorithm [1,26] under the same conditions.

Fig. 8 illustrates the infected, recovered, and deceased populations calculated by the numerical solutions of the SEIRD model and deterministic SEIR(R)D-SD models based on CLS as well as those estimated by the proposed EKF based on the stochastic SEIR(R)D-SD model. As shown in Fig. 8(a), the numerical solution of the SEIRD model presents only one peak (i.e., only one outbreak). The maximum infected population calculated from the SEIRD model is about 850 more than and about 10 days earlier than that in the report data. Further, the infected population quickly drops to zero after the only peak, leading to a significant error. Different from that of the SEIRD model, the numerical solution of the

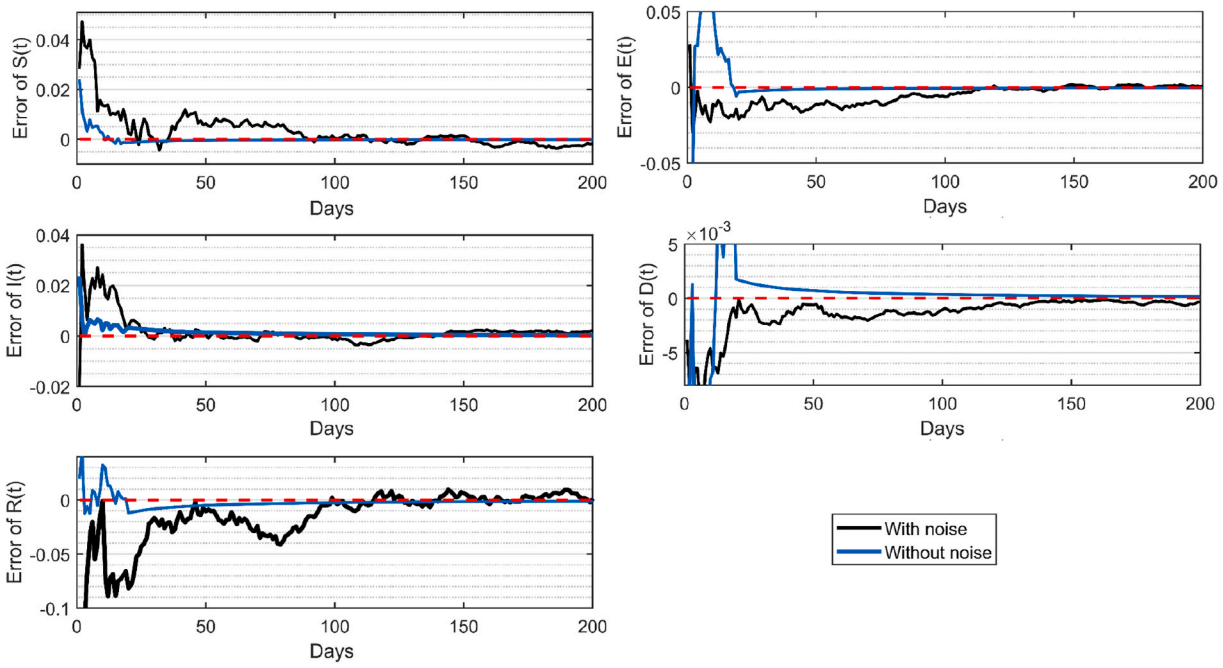

Fig. 7. Estimation error by the proposed EKF based on the stochastic SEIR(R)D-SD model.

Table 4
RMSEs of the proposed EKF based on the stochastic SEIR(R)D-SD model.

	Susceptible	Exposed	Infected	Recovered	Deceased
Without noise	3.2×10^{-3}	2.1×10^{-3}	1.2×10^{-3}	3.5×10^{-3}	6.4×10^{-4}
With noise	8.2×10^{-3}	4.4×10^{-3}	3.1×10^{-3}	5.5×10^{-3}	1.2×10^{-3}

deterministic SEIR(R)D-SD model for the infected population presents two outbreaks. Although the first outbreak is occurred closely to that in the report data, due to the inability to account for uncertainties involved in COVID-19 spread, the numerical solution of the deterministic SEIR(R)D-SD model for the infected population involves an obvious error, leading to the second outbreak with a delay of about 20 days and a deviation of 1237 infections comparing to that in the reported data. In contrast, the solution of the proposed EKF based on stochastic SEIR(R)D-SD model for the infected population is much closer to the report data than that of the deterministic SEIR(R)D-SD model, and the estimated two outbreaks are in a good agreement with those in the report data.

As shown in Fig. 8(b) and (c), the solutions of the recovered and deceased populations by the three methods have a similar trend as those of the infected population. The numerical solutions of the SEIRD model for the recovered and deceased populations remain at a constant value after day 120 and 110, respectively, both of which miss the second outbreak. Although the solutions from the deterministic and stochastic SEIR(R)D-SD models follow the report data and present the two outbreaks, the solution estimated by the proposed EKF based on the

Table 5
The initial values of the model parameters for the COVID-19 pandemic in Australia.

Parameters	β_0	α_0	γ_0	μ_0	κ_0	ρ_0
Value	0.63	0.24	0.03	0.01	0.001	0.5

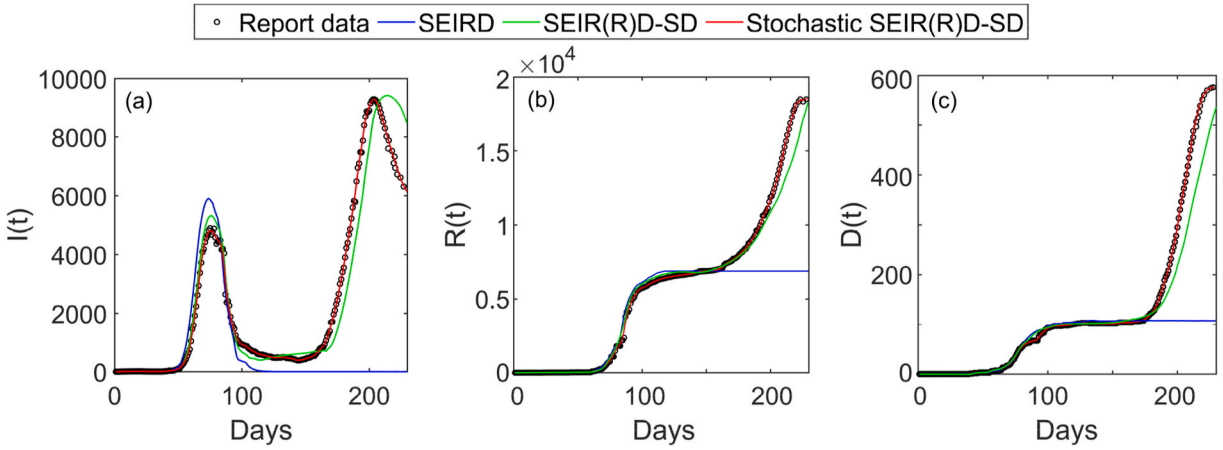

Fig. 8. The infected, recovered and deceased populations by the numerical solutions of the SEIRD and deterministic SEIR(R)D-SD models based on CLS and the estimation solution of the proposed EKF based on the stochastic SEIR(R)D-SD model for the COVID-19 pandemic in Australia: (a) the infected population; (b) the recovered population; and (c) the deceased population.

Table 6
RMSEs and prediction means of the infected, recovered and deceased populations by the numerical solutions of the SEIRD and deterministic SEIR(R)D-SD models based on CLS as well as the proposed EKF based on the stochastic SEIR(R)D-SD model for the COVID-19 pandemic in Australia.

Compartments	Numerical solution of the SEIRD model based on CLS		Numerical solution of the deterministic SEIR(R)D-SD model based on CLS		EKF based on the stochastic SEIR(R)D-SD model		Report data
	RMSE	Mean	RMSE	Mean	RMSE	Mean	Mean
Infected	5064	800	472	1484	203	1631	1649
Recovered	6614	3466	556	4028	225	4254	4203
Deceased	311	64	35	92	17	101	103

stochastic SEIR(R)D-SD model approximates the reported data more closely than the numerical solution of the deterministic SEIR(R)D-SD model for both recovered and decreased populations.

The RMSEs for the infected, recovered and deceased populations are shown in Table 6. The RMSEs of the infected, recovered and deceased populations obtained by the numerical solution of the SEIRD model are 5064, 6614 and 311. The corresponding RMSEs of the deterministic SEIR(R)D-SD model are 472, 556 and 35, which are about 10 times smaller than those of the SEIRD model, whereas the corresponding RMSEs of the proposed EKF based on the stochastic SEIR(R)D-SD model are 203, 225 and 17, which are more than twice smaller than those of the deterministic SEIR(R)D-SD model and more than 20 times smaller than those of the SEIRD model. Thus, it is evident that the proposed EKF based on the stochastic SEIR(R)D-SD model has much higher accuracy than the SEIRD and deterministic SEIR(R)D-SD models for modelling of COVID-19 spread. Table 6 also compares the prediction means of the three methods with reference to the means of the reported data, which further verifies the above conclusion.

4. Conclusion

This paper presents a new method for COVID-19 modelling. The novelties of this method are: (i) a deterministic SEIR(R)D-SD model is developed to account for the re-infection and SD effects on COVID-19 spread; (ii) a stochastic SEIR(R)D-SD model is developed from the deterministic SEIR(R)D-SD model to account for uncertainties involved in COVID-19 spread; and (iii) based on the stochastic SEIR(R)D-SD model, an EKF algorithm is developed to simultaneously estimate both model parameters and transmission state. Simulations and comparison analyses demonstrate that the proposed method can effectively account for the re-infection and SD effects as well as uncertainties on COVID-19 spread, leading to increased accuracy for COVID-19 modelling.

The future research work will focus on improvement of the proposed method to account for the vaccination effect especially due to the vaccination rollout on COVID-19 spread. A new compartment and its associated rate will be introduced into the proposed SEIR(R)D-SD model to characterize the behaviours of vaccination, leading to a new stochastic epidemiological model and its associated real-time estimation algorithm for COVID-19 modelling.

CRediT authorship contribution statement

Xinhe Zhu: Formal analysis, Writing – original draft. **Bingbing Gao:** Writing – original draft. **Yongmin Zhong:** Writing – original draft. **Chengfan Gu:** Writing – original draft. **Kup-Sze Choi:** Formal analysis.

Declaration of competing interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

[1] R. Sameni, Mathematical Modeling of Epidemic Diseases; a Case Study of the COVID-19 Coronavirus, 2020 arXiv preprint arXiv:2003.11371.
[2] F.A. Marioli, F. Bullano, C. Rondón-Moreno, Tracking R of COVID-19: a new real-time estimation using the Kalman filter, PloS One 16 (1) (2020), e0244474.
[3] S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L.Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, T. Wu, Publisher Correction: the effect of large-scale anti-contagion policies on the COVID-19 pandemic, Nature 585 (7824) (2020) E7.
[4] S.B. Ramezani, A. Amirlatifi, S. Rahimi, A novel compartmental model to capture the nonlinear trend of COVID-19, Comput. Biol. Med. 134 (2021) 104421.
[5] Y. Shi, Y. Wang, C. Shao, J. Huang, J. Gan, X. Huang, E. Bucci, M. Piacentini, G. Ippolito, G. Melino, COVID-19 infection: the perspectives on immune responses, Cell Death Differ. 27 (5) (2020) 1451–1454.
[6] M.S. Hagger, S.R. Smith, J.J. Keech, S.A. Moyers, K. Hamilton, Predicting social distancing intention and behavior during the COVID-19 Pandemic: an integrated social cognition model, Ann. Behav. Med. 54 (10) (2020) 713–727.
[7] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, J. Wang, Z. Hu, Y. Yi, H. Shen, Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China, Sci. China Life Sci. 63 (5) (2020) 706–711.
[8] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182 (4) (2020) 812–827, e819.
[9] J. Song, H. Xie, B. Gao, Y. Zhong, C. Gu, K.S. Choi, Maximum likelihood-based extended Kalman filter for COVID-19 prediction, Chaos, Solit. Fractals 146 (2021) 110922.
[10] H. Khataee, I. Scheuring, A. Czirok, Z. Neufeld, Effects of social distancing on the spreading of COVID-19 inferred from mobile phone data, Sci. Rep. 11 (1) (2021).
[11] S. Kumar, K.K. Singh, P. Dixit, M. Kumar Bajpai, Kalman filter based short term prediction model for COVID-19 spread, Appl. Intell. 51 (5) (2020) 2714–2726.
[12] B. Gao, G. Hu, Y. Zhong, X. Zhu, Cubature rule-based distributed optimal fusion with identification and prediction of kinematic model error for integrated UAV navigation, Aero. Sci. Technol. (2020) 106447.
[13] Y. Marzouk, D. Xiu, A stochastic collocation approach to bayesian inference in inverse problems, Comput. Phys. Commun. 6 (4) (2009) 826–847.
[14] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures, Proc. Natl. Acad. Sci. U.S.A. 117 (19) (2020) 10484–10491.
[15] P.S.R. Diniz, Kalman Filters. *Adaptive Filtering: Algorithms and Practical Implementation*, Springer International Publishing, Cham, 2020, pp. 431–456.
[16] X. Zhu, B. Gao, Y. Zhong, C. Gu, K.S. Choi, Extended Kalman filter for online soft tissue characterization based on Hunt-Crossley contact model, J. Mech. Behav. Biomed. Mater. 123 (2021) 104667.
[17] A. Hasan, H. Susanto, V.R. Tjahjono, R. Kusdiantara, E.R.M. Putri, P. Hadisoemarto, N. Nuraini, A New Estimation Method for COVID-19 Time-Varying Reproduction Number Using Active Cases, 2020 arXiv preprint arXiv: 2006.03766.
[18] A. Bani Younes, Z. Hasan, COVID-19: Modeling, prediction, and control, Appl. Sci. 10 (11) (2020).
[19] I. Korolev, Identification and estimation of the SEIRD epidemic model for COVID-19, J. Econom. 220 (1) (2021) 63–85.
[20] A.M. Salman, I. Ahmed, M.H. Mohd, M.S. Jamiluddin, M.A. Dheyab, Scenario analysis of COVID-19 transmission dynamics in Malaysia with the possibility of reinfection and limited medical resources scenarios, Comput. Biol. Med. 133 (2021) 104372.
[21] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, CRC press, 2013.
[22] E. Allen, Modeling with Itô Stochastic Differential Equations, vol. 22, Springer Science & Business Media, 2007.
[23] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, Lancet Infect. Dis. 20 (5) (2020) 533–534.
[24] United Nations, World Population Prospects 2019: Highlights, Department of Economic and Social Affairs, Population Division, 2019.

[25] World Health Organization, Novel Coronavirus (2019-nCoV) Situation Reports, vol. 3, 2020.
[26] E. Loli Piccolomini, F. Zama, Monitoring Italian COVID-19 spread by a forced SEIRD model, PloS One 15 (8) (2020), e0237417.
[27] S. Rao, M. Singh, An evolving public health crisis caused by the rapid spread of the SARS-CoV-2 Delta variant, DHR Proceedings 1 (S4) (2021) 6–8.
[28] S.L. Chang, N. Harding, C. Zachreson, O.M. Cliff, M. Prokopenko, Modelling transmission and control of the COVID-19 pandemic in Australia, Nat. Commun. 11 (1) (2020).
[29] C.C. Kerr, R.M. Stuart, D. Mistry, R.G. Abeysuriya, K. Rosenfeld, G.R. Hart, D. J. Klein, Covasim: an agent-based model of COVID-19 dynamics and interventions, PLoS Comput. Biol. 17 (7) (2021), e1009149.
[30] P. Nouvellet, S. Bhatia, A. Cori, K.E.C. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, N.F. Brazeau, L. Cattarino, L.V. Cooper, H. Coupland, Z.M. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, B.A. Djaafara, I. Dorigatti, O.D. Eales, S.L. van Elsland, F. F. Nascimento, C.A. Donnelly, Reduction in mobility and COVID-19 transmission, Nat. Commun. 12 (1) (2021).
[31] E. Malkov, Simulation of coronavirus disease 2019 (COVID-19) scenarios with possibility of reinfection, Chaos, Solit. Fractals 139 (2020) 110296.
[32] M.J. Keeling, T.D. Hollingsworth, J.M. Read, Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19), J. Epidemiol. Community 74 (10) (2020) 861–866.
[33] S.J. Bickley, H.F. Chan, A. Skali, D. Stadelmann, B. Torgler, How does globalization affect COVID-19 responses? Glob. Health 17 (1) (2021) 57.