Article

An Extended SEIR Model with Vaccination for Forecasting the COVID-19 Pandemic in Saudi Arabia Using an Ensemble Kalman Filter

Rabih Ghostine [1,*], **Mohamad Gharamti [2]**, **Sally Hassrouny [3]** and **Ibrahim Hoteit [4]**

1   Department of Mathematics, Kuwait College of Science and Technology, Doha 35001, Kuwait
2   National Center for Atmospheric Research, Boulder, CO 80305, USA; gharamti@ucar.edu
3   Department of Science, Kuwait College of Science and Technology, Doha 35001, Kuwait; s.hassrouny@kcst.edu.kw
4   Applied Mathematics and Computational Science, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia; ibrahim.hoteit@kaust.edu.sa
*   Correspondence: r.ghostine@kcst.edu.kw; Tel.: +965-2497-2843

Abstract: In this paper, an extended SEIR model with a vaccination compartment is proposed to simulate the novel coronavirus disease (COVID-19) spread in Saudi Arabia. The model considers seven stages of infection: susceptible (S), exposed (E), infectious (I), quarantine (Q), recovered (R), deaths (D), and vaccinated (V). Initially, a mathematical analysis is carried out to illustrate the non-negativity, boundedness, epidemic equilibrium, existence, and uniqueness of the endemic equilibrium, and the basic reproduction number of the proposed model. Such numerical models can be, however, subject to various sources of uncertainties, due to an imperfect description of the biological processes governing the disease spread, which may strongly limit their forecasting skills. A data assimilation method, mainly, the ensemble Kalman filter (EnKF), is then used to constrain the model outputs and its parameters with available data. We conduct joint state-parameters estimation experiments assimilating daily data into the proposed model using the EnKF in order to enhance the model's forecasting skills. Starting from the estimated set of model parameters, we then conduct short-term predictions in order to assess the predicability range of the model. We apply the proposed assimilation system on real data sets from Saudi Arabia. The numerical results demonstrate the capability of the proposed model in achieving accurate prediction of the epidemic development up to two-week time scales. Finally, we investigate the effect of vaccination on the spread of the pandemic.

Keywords: COVID-19 pandemic; SEIR model; mathematical modeling; ensemble Kalman filter; joint state-parameters estimation

1. Introduction

The novel coronavirus disease (COVID-19) first appeared in Wuhan city, and due to its high transmission rate, the virus has been rapidly spreading all over the world. On 8 March 2020, the WHO announced COVID-19 as a global pandemic [1]. As of 1 January 2021, there are over 100 million reported cases, and a death toll exceeding 2 million.

In the absence of a ready-to-use vaccine, and besides medical and biological research, mathematical models can play an important role in understanding and predicting disease transmission. Moreover, it helps to implement appropriate measures and efficient strategies to control the pandemic's spread and mitigate its impact.

Mathematical epidemic models based on the SIR model [2] are widely used to study the spread of a disease in a population [3–7]. Recently, several mathematical models have been developed to study the transmission dynamics of COVID-19 [8–11]. However, these models suffer from various sources of uncertainties, due to the incomplete description of the biological processes governing the disease spread, and also due to some involved parameters being poorly known. One way to mitigate these uncertainties is to constrain

epidemic forecasting models with available data. These data can be combined with the model output to improve its prediction and reduce uncertainties [12–14]. This process, known as data assimilation, sequentially adjusts the model state and parameters once data become available to keep the model as close as possible to the real track [15].

The ensemble Kalman filter (EnKF) is a popular data assimilation technique introduced by [16] to tackle large-scale, sequential, and non-linear problems. The EnKF is a Monte Carlo implementation of the traditional Kalman filter [17], operating in cycles of two steps: (i) A forecast step in which a set of state samples, or ensemble members, are integrated forward in time with the model, and (ii) an analysis step to update the forecasted members with available data based on the Bayes rule in a Gaussian framework [15,16]. Because of its affordable computational cost, robust performance, nonintrusive formulation, and ease of implementation, the EnKF have been successfully implemented in many geophysical applications, including weather forecasting [18–20], hydrology and reservoir applications [21–26], ocean modeling [27–31], and flood modeling [32–34].

In epidemiology, previous studies have used data assimilation in the context of both variational and Kalman filter methods [35–38]. The Reference [39] used the EnKF to estimate the parameters of a stochastic SEIR model. The Reference [40] applied a deterministic variant of the EnKF, the Ensemble Adjustment Kalman Filter (EAKF), in combination with a network of SEIR models, simulating different connected cities. The Reference [41] used an ensemble smoother with a multiple data assimilation (ESMDA) approach for parameter estimation in a SEIR model. The Reference [42] also implemented the ESMDA to effectively estimate the parameters of a more sophisticated SEIR model accounting for age-classes and compartments of the sick, hospitalized, and dead. The Reference [43] used an EnKF approach for parameter estimation and shot-term forecasts with a simple SIR model. The Reference [44] proposed an EnKF for estimating unmeasurable state variables and unknown parameters of a SIRV model, which takes into account the circulation of free coronavirus in the environment. In the context of variational data assimilation methods, the Reference [45] performed parameter estimation and predictions using a SIR model, while [46] proposed a modified SEIR model that distinguishes between symptomatic and asymptomatic, and conducted observation sensitivity experiments to identify suitable observing strategies. However, these studies did not consider the impact of vaccination on the pandemic spread.

This work therefore aims to study the impact of vaccination on the COVID-19 spread. An extended SEIR model comprising of seven compartments—susceptible, exposed, infectious, quarantined, recovered, deaths, and vaccinated—is then proposed. First, we conduct a mathematical analysis to illustrate the non-negativity, boundedness, epidemic equilibrium, existence, and uniqueness of the endemic equilibrium, and the basic reproduction number of the model. A Joint-EnKF is implemented for state-parameters estimation by assimilating COVID-19 data into the model. By considering uncertainties in all model states and poorly known parameters, our ensemble data assimilation approach computes updates of both state and parameter ensembles. The estimated set of model state and parameters is then used to assess the model prediction skill. The proposed approach is validated with the cases reported from the most-infected Gulf country, Saudi Arabia, showing a great ability to provide accurate and consistent forecasts up to two weeks. We finally investigate the effect of vaccination on the spread of the pandemic, showing a notable decrease in the number of cases.

The remainder of this paper is organized as follows. Section 2 presents the extended SEIR model and briefly describes the Joint-EnKF algorithm. Section 3 is devoted to the mathematical analysis of the model. The numerical results are presented and discussed in Section 4. Section 5 concludes the study, with a summary of the main results.

2. Model and Assimilation Scheme

2.1. Extended SEIR Model

We extend the SEIR model to seven compartments to simulate the epidemic of COVID-19. Seven state variables are considered within a population, that is, $S(t)$, $E(t)$, $I(t)$, $Q(t)$, $R(t)$, $D(t)$, and $V(t)$, denoting the number of susceptible, exposed (infected, but not yet infectious), infectious (not yet quarantined), quarantined (confirmed and infected), recovered, dead, and vaccinated cases, respectively. The disease transmission flow of the proposed model is sketched in Figure 1. The model is then governed by the following set of nonlinear ordinary differential equations:

$$\begin{aligned}
\frac{dS(t)}{dt} &= \Lambda - \beta S(t)I(t) - \alpha S(t) - \mu S(t), \\
\frac{dE(t)}{dt} &= \beta S(t)I(t) - \gamma E(t) + \sigma \beta V(t)I(t) - \mu E(t), \\
\frac{dI(t)}{dt} &= \gamma E(t) - \delta I(t) - \mu I(t), \\
\frac{dQ(t)}{dt} &= \delta I(t) - (1-\kappa)\lambda Q(t) - \kappa \rho Q(t) - \mu Q(t), \\
\frac{dR(t)}{dt} &= (1-\kappa)\lambda Q(t) - \mu R(t), \\
\frac{dD(t)}{dt} &= \kappa \rho Q(t), \\
\frac{dV(t)}{dt} &= \alpha S(t) - \sigma \beta V(t)I(t) - \mu V(t),
\end{aligned} \tag{1}$$

with non-negative initial conditions $S(0) = S_0$, $E(0) = E_0$, $I(0) = I_0$, $Q(0) = Q_0$, $R(0) = R_0$, $D(0) = D_0$, and $V(0) = V_0$. The coefficients Λ, β, α, μ, γ, σ, δ, κ, λ, and ρ represent the new births and new residents per unit of time, transmission rate divided by N, vaccination rate (rate of people who are vaccinated), natural death rate, average latent time, average quarantine time, mortality rate, average days until recovery, and average days until death, respectively. σ is the vaccine inefficacy ($0 \le \sigma \le 1$). So $1 - \sigma$ represents the vaccine efficacy. If $\sigma = 0$, the vaccine offers 100% protection against the disease. $N = S + E + I + Q + R + D + V$ is the total population size. Finally, system (1) is solved using a fourth-order Runge–Kutta method.
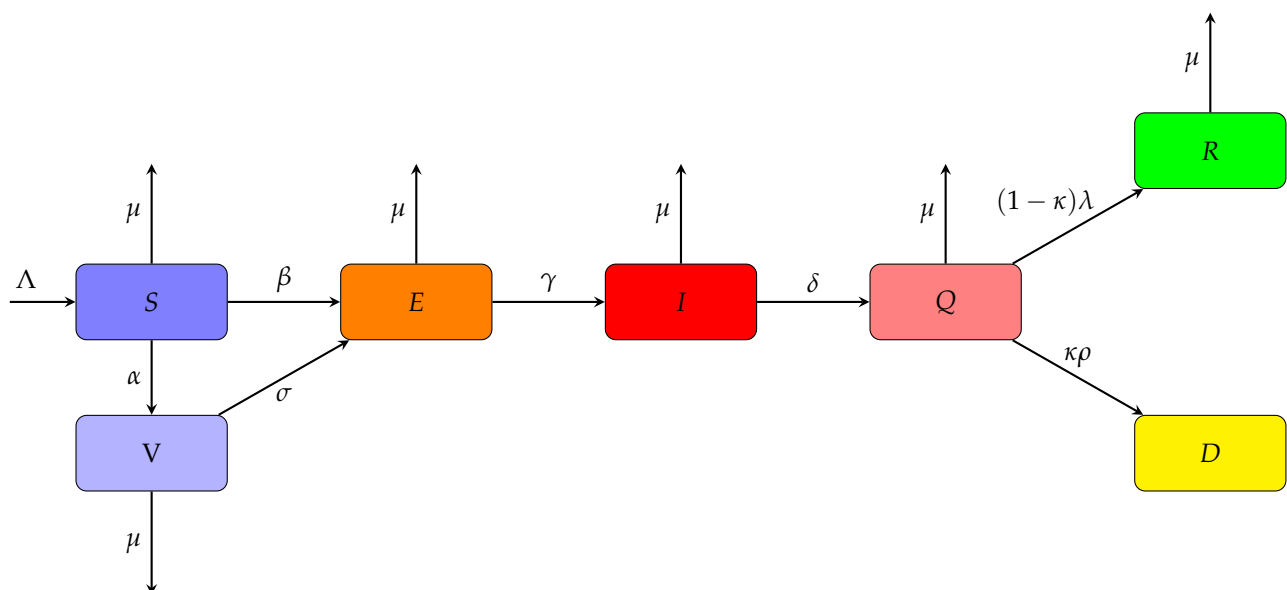

Figure 1. Disease transmission flow of the proposed model.

2.2. Ensemble Kalman Filter (EnKF)

Consider the following nonlinear data assimilation problem:

$$
\begin{aligned}
\mathbf{x}_{k+1} &= \mathcal{M}(\mathbf{x}_k, \theta) + \eta_{k+1}, \\
\mathbf{y}_{k+1} &= \mathcal{H}(\mathbf{x}_{k+1}) + \epsilon_{k+1},
\end{aligned}
\tag{2}
$$

where $\mathbf{x}_k$ is the state vector of the model at time t_k, $\mathcal{M}$ is the dynamical operator (system (1)) that integrates the model state from time t_k to time t_{k+1}, θ is the parameters vector, $\mathbf{y}_{k+1}$ is the observation vector (active, recovered, and death cases) at time t_{k+1}, $\mathcal{H}$ is the observation operator that projects the state space onto the observation space, and η_{k+1} and ϵ_{k+1} are model and observation noises assumed to follow Gaussian distributions with zero mean and covariances $\mathbf{Q}_{k+1}$ and $\mathbf{R}_{k+1}$, respectively.

At every forecast step, we integrate the ensemble members with the dynamical model (1). The predicted state mean and the associated error covariance matrix are then estimated

$$
\hat{\mathbf{x}}^f_{k+1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}^{f,i}_{k+1},
\tag{3}
$$

$$
\hat{\mathbf{P}}^f_{k+1} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \left(\mathbf{x}^{f,i}_{k+1} - \hat{\mathbf{x}}^f_{k+1} \right) \left(\mathbf{x}^{f,i}_{k+1} - \hat{\mathbf{x}}^f_{k+1} \right)^T = \mathbf{X}^f_{k+1} \left(\mathbf{X}^f_{k+1} \right)^T,
\tag{4}
$$

where N_e is the number of ensemble members.

The Joint-EnKF consists of estimating the augmented state-parameters vector $\mathbf{z}_{k+1} = \left[\mathbf{x}^T_{k+1}, \theta^T \right]$ at each time t_{k+1} given the observation $\mathbf{y}_{k+1}$. The prediction step of the Joint-EnKF becomes:

$$
\mathbf{z}^f_{k+1} = \tilde{\mathcal{M}}(\mathbf{z}_k, \theta) + \eta_{k+1} = \begin{bmatrix} \mathcal{M}(\mathbf{x}_k, \theta) + \eta_{k+1} \\ \theta \end{bmatrix},
\tag{5}
$$

where $\tilde{\mathcal{M}}$ is the augmented nonlinear dynamic operator for the joint state-parameters. With similar computations as in (4), the joint state-parameters error covariance matrix for the augmented ensemble vectors $\mathbf{z}^{f,i}_{k+1}$ is given by

$$
\hat{\mathbf{P}}^f_{k+1} = \begin{bmatrix} \hat{\mathbf{P}}^f_{\mathbf{x}_{k+1}\mathbf{x}_{k+1}} & \hat{\mathbf{P}}^f_{\mathbf{x}_{k+1}\theta_{k+1}} \\ \hat{\mathbf{P}}^f_{\theta_{k+1}\mathbf{x}_{k+1}} & \hat{\mathbf{P}}^f_{\theta_{k+1}\theta_{k+1}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^f_{\mathbf{x}_{k+1}} \left(\mathbf{X}^f_{\mathbf{x}_{k+1}} \right)^T & \mathbf{X}^f_{\mathbf{x}_{k+1}} \left(X^f_{\theta_{k+1}} \right)^T \\ \mathbf{X}^f_{\theta_{k+1}} \left(\mathbf{X}^f_{\mathbf{x}_{k+1}} \right)^T & \mathbf{X}^f_{\theta_{k+1}} \left(\mathbf{X}^f_{\theta_{k+1}} \right)^T \end{bmatrix},
\tag{6}
$$

where $\mathbf{X}_\theta$ denotes the parameter ensemble perturbations. The updated joint state and parameters are then computed using the current observation ensemble $\mathbf{y}^i_{k+1}$:

$$
\mathbf{z}^{a,i}_{k+1} = \mathbf{z}^{f,i}_{k+1} + \hat{\mathbf{P}}^f_{k+1} \tilde{\mathcal{H}}^T \left(\tilde{\mathcal{H}} \hat{\mathbf{P}}^f_{k+1} \tilde{\mathcal{H}}^T + \mathbf{R}_{k+1} \right)^{-1} \left[\mathbf{y}^i_{k+1} - \tilde{\mathcal{H}} \mathbf{z}^{f,i}_{k+1} \right],
\tag{7}
$$

where $\tilde{\mathcal{H}} = [\mathcal{H} \quad \mathbf{0}]$ is the augmented observational operator with the zero matrix.

3. Mathematical Analysis of the COVID-19 Model

In this section, we discuss the boundedness, non-negativity, epidemic equilibrium, endemic equilibrium, and basic reproduction number of the COVID-19 model (1).

3.1. Non-Negativity of the Model

Theorem 1. *If $S_0 \geq 0$, $E_0 \geq 0$, $I_0 \geq 0$, $Q_0 \geq 0$, $R_0 \geq 0$, $D_0 \geq 0$, and $V_0 \geq 0$, then the solutions of system (1) remain non-negative for all $t > 0$.*

Proof of Theorem 1. From the first equation of system (1), we have:

$$\frac{dS}{dt} \geq -\mu S. \tag{8}$$

Integrating Equation (8), we get

$$S(t) \geq S_0 e^{-\mu t} \geq 0. \tag{9}$$

Hence, $S(t)$ remains non-negative for all $t > 0$. Similarly, using the remaining equations of system (1), we obtain $E(t) \geq 0$, $I(t) \geq 0$, $Q(t) \geq 0$, $R(t) \geq 0$, $D(t) \geq 0$, and $V(t) \geq 0$. Therefore, Theorem 1 is proved. □

3.2. Boundedness of the Model

Theorem 2. *All solutions of the proposed model with non-negative initial conditions are bounded and $N(t) \leq \dfrac{\Lambda}{\mu}$ for all $t > 0$.*

Proof of Theorem 2. Adding all the equations in the system (1), the growth rate of the total population can be written as

$$\frac{dN}{dt} = \frac{dS}{dt} + \frac{dE}{dt} + \frac{dI}{dt} + \frac{dQ}{dt} + \frac{dR}{dt} + \frac{dD}{dt} + \frac{dV}{dt}. \tag{10}$$

From Equation (10), we have the following equation:

$$\frac{dN}{dt} = \Lambda - \mu N. \tag{11}$$

Integrating both sides of Equation (11) gives the solution for $N(t)$ as follows

$$N(t) = \frac{\Lambda}{\mu} + \left(N_0 - \frac{\Lambda}{\mu} \right) e^{-\mu t}. \tag{12}$$

Thus, when $t \to \infty$, we find that

$$N(t) \leq \frac{\Lambda}{\mu}. \tag{13}$$

Finally, Theorem 1 and Equation (13) give

$$0 \leq N(t) \leq \frac{\Lambda}{\mu}. \tag{14}$$

Thus, $S(t)$, $E(t)$, $I(t)$, $Q(t)$, $R(t)$, $D(t)$, and $V(t)$ are bounded and Theorem 2 is proved. □

3.3. Epidemic Equilibrium and Basic Reproduction Number of the Model

The epidemic equilibrium X^0 of system (1) is obtained by setting all the derivatives to zero with $I = 0$, that yields to:

$$X^0 = \left(\frac{\Lambda}{\mu + \alpha}, 0, 0, 0, 0, 0, \frac{\alpha \Lambda}{\mu(\mu + \alpha)} \right). \tag{15}$$

R_0 is computed using the next-generation matrix concept [47]. Let $X = (E, I)^T$. System (1) can be written as

$$X' = \mathbb{F}(X) - \mathbb{W}(X), \tag{16}$$

where $\mathbb{F}(X) = (\beta SI + \sigma \beta IV, 0)^T$ and $\mathbb{W}(X) = ((\mu + \delta)E, -\gamma E + (\mu + \delta))^T$. Their corresponding Jacobian matrices at the disease-free equilibrium are

$$F = \begin{pmatrix} 0 & \beta S^0 + \sigma \beta V^0 \\ 0 & 0 \end{pmatrix}, \qquad W = \begin{pmatrix} \mu + \gamma & 0 \\ -\gamma & \mu + \delta \end{pmatrix}. \tag{17}$$

According to [47], the basic reproduction number R_0 is the spectral radius of the next-generation matrix FW^{-1}. That gives

$$R_0 = \frac{\beta \gamma}{(\mu + \gamma)(\mu + \delta)} \left(S^0 + \sigma V^0 \right). \tag{18}$$

Substituting S^0 and V^0 from Equation (15) into Equation (20) yields the following formula for the basic reproduction number of system (1):

$$R_0 = \frac{\beta \gamma \Lambda (\mu + \alpha \sigma)}{\mu (\mu + \gamma)(\mu + \delta)(\mu + \alpha)}, \tag{19}$$

used to measure the transmission potential of a disease. It is the number of secondary cases or new infections produced by a single infectious individual in a completely susceptible population [47,48]. An epidemic is expected to increase exponentially if $R_0 > 1$, and to end if $R_0 < 1$ [47].

Theorem 3. *The disease-free equilibrium X^0 is locally asymptotically stable if $R_0 < 1$, and unstable if $R_0 > 1$.*

Proof of Theorem 3. System (1) can be written in a simplified form by eliminating the sixth equation for the death compartment (D). This is due to the fact that the state D only appears in the corresponding differential equation and has no significance on the overall system. The number of deaths can be determined from $D = N - S - E - I - Q - R - V$. The Jacobian matrix of the reduced model (1) at the disease-free equilibrium is given by

$$J\left(X^0\right) = \begin{pmatrix} -\epsilon_1 & 0 & -\beta S^0 & 0 & 0 & 0 \\ 0 & -\epsilon_2 & \beta\left(S^0 + \sigma V^0\right) & 0 & 0 & 0 \\ 0 & \gamma & -\epsilon_3 & 0 & 0 & 0 \\ 0 & 0 & \delta & -\epsilon_4 & 0 & 0 \\ 0 & 0 & 0 & \lambda(1 - \kappa) & -\mu & 0 \\ \alpha & 0 & -\sigma \beta V^0 & 0 & 0 & -\mu \end{pmatrix}, \tag{20}$$

where $\epsilon_1 = -(\alpha + \mu)$, $\epsilon_2 = -(\gamma + \mu)$, $\epsilon_3 = -(\delta + \mu)$, and $\epsilon_4 = -(\lambda(1 - \kappa) + \kappa\rho + \mu)$. $J\left(X^0\right)$ has $\lambda_1 = -\epsilon_1 < 0$, $\lambda_2 = -\epsilon_4 < 0$, $\lambda_3 = \lambda_4 = -\mu < 0$. The remaining two eigenvalues are

$$\lambda_5 = -\frac{1}{2}\left(\epsilon_2 + \epsilon_3 + \sqrt{(\epsilon_2 - \epsilon_3)^2 + 4\epsilon_2\epsilon_3 R_0}\right),$$

$$\lambda_6 = -\frac{1}{2}\left(\epsilon_2 + \epsilon_3 - \sqrt{(\epsilon_2 - \epsilon_3)^2 + 4\epsilon_2\epsilon_3 R_0}\right). \tag{21}$$

when $R_0 < 1$, we have $\lambda_5 < 0$ and $\lambda_6 < 0$. Since all the eigenvalues are negative, the disease-free equilibrium X^0 is locally asymptotically stable. If $R_0 > 1$, we have $\lambda_5 < 0$ and $\lambda_6 > 0$, and the disease-free equilibrium X^0 is locally asymptotically unstable [49]. $\quad\square$

3.4. Existence and Uniqueness of the Endemic Equilibrium

In this section, we investigate the existence and uniqueness of the endemic equilibrium when $R_0 > 1$. Let $X^* = (S^*, E^*, I^*, Q^*, D^*, V^*)$ be the endemic equilibrium of model (1). The endemic equilibrium is obtained by setting all the derivatives in model (1) to zero, giving

$$
\begin{aligned}
\Lambda - \beta S^* I^* - \alpha S^* - \mu S^* &= 0, \\
\beta S^* I^* - \gamma E^* + \sigma \beta V^* I^* - \mu E^* &= 0, \\
\gamma E^* - \delta I^* - \mu I^* &= 0, \\
\delta I^* - (1 - \kappa)\lambda Q^* - \kappa \rho Q^* - \mu Q^* &= 0, \\
(1 - \kappa)\lambda Q^* - \mu R^* &= 0, \\
\alpha S^* - \sigma \beta V^* I^* - \mu V^* &= 0.
\end{aligned}
\tag{22}
$$

Solving the first, third, and fifth equations of (22) in terms of I^* gives

$$
E^* = \frac{\epsilon_3}{\gamma} I^*, \quad Q^* = \frac{\delta}{\epsilon_4} I^*, \quad S^* = \frac{\Lambda}{\beta I^* + \epsilon_1},
\tag{23}
$$

where $\epsilon_1 = \mu + \alpha$, $\epsilon_3 = \mu + \delta$, and $\epsilon_4 = \mu + \lambda(1 - \kappa) + \kappa \rho$. Adding the second and sixth equation in (22) gives

$$
V^* = \frac{\Lambda \beta \gamma I + \Lambda \alpha \gamma - \epsilon_2 \epsilon_3 (\beta I + \epsilon_1) I}{\mu \gamma (\beta I + \epsilon_1)}.
\tag{24}
$$

Now substituting the expressions of S^*, E^*, and V^* in the second equation of (22), we get the following quadratic equation for I^*:

$$
a_2 I^{*2} + a_1 I^* + a_0 = 0,,
\tag{25}
$$

where

$$
\begin{aligned}
a_0 &= \mu \epsilon_1 \epsilon_2 \epsilon_3 (1 - R_0), \\
a_1 &= \epsilon_1 \epsilon_2 \epsilon_3 + \mu \beta \epsilon_2 \epsilon_3 - \Lambda \beta^2 \gamma \sigma, \\
a_2 &= \beta \epsilon_2 \epsilon_3.
\end{aligned}
\tag{26}
$$

It is clear that a_2 is always positive and a_0 is negative when $R_0 > 1$. Therefore, there is a unique positive value for I^*, and consequently, a unique endemic equilibrium X^* when $R_0 > 1$.

4. Results and Discussion

The proposed modeling system was applied to forecast the COVID-19 pandemic in Saudi Arabia. Data are available from the Saudi Center for Diseases Prevention and Control [50]. These include the number of deaths, as well as recovered and confirmed cases. The total number of deaths, active cases, and recovered cases were assimilated daily, and model parameters were also updated using the Joint-EnKF.

4.1. Assimilation Settings

In this section, we cover the period before the vaccine becomes available. Therefore, state and parameters related to vaccination are excluded from the model at this stage. The state vector x to be estimated is then composed of six elements

$$
x = \begin{bmatrix} S & E & I & Q & R & D \end{bmatrix},
\tag{27}
$$

and the parameters to be estimated become

$$
\theta = \begin{bmatrix} \Lambda & \beta & \mu & \gamma & \delta & \kappa & \lambda & \rho \end{bmatrix}.
\tag{28}
$$

The augmented state vector to be estimated is then:

$$z = \begin{bmatrix} x^T & \log\left(\theta^T\right) \end{bmatrix}, \tag{29}$$

where the parameters are in a logarithmic sense in order to avoid negative values in the model.

We set the initial numbers of infected and exposed people to $I(0) = E(0) = 3$, three times the number of accumulated positive tests ($Q_0 = 1$) provided by March 2nd [50]. The initial numbers of recovered, and death cases are $R(0) = D(0) = 0$. The total population and the total number of susceptible cases are $N(0) = 34{,}218{,}000$ and $S(0) = N(0) - E(0) - I(0) - Q(0) - R(0) - D(0)$, respectively. We consider the following two periods: the first period concerns the period of the beginning of the pandemic when the intervention had not been applied using a basic reproduction number greater than one, $R_0 = 2.5$, which gives $\beta_1 = 8.58 \times 10^{-9}$, while the second period corresponds to the period when the intervention was applied by the government until June 30th. During this period, the numbers of new cases are relatively constant. Therefore, we use a basic reproduction number $R_0 = 1$, which gives $\beta_2 = 3.43 \times 10^{-9}$. To reflect the uncertainty in the transmission rate, we use a standard deviation of 25%. The initial values of the remaining parameters are listed in Table 1, with a standard deviation of 10%. In addition, we assume the observation-error standard deviations to be 5% of the data value.

Table 1. Initial model parameters and biological interpretations.

Parameter	Initial Value	Description	Reference
Λ	2300 persons/day	New births and new residents	[51]
β_1	8.58×10^{-9} day^{-1}	Transmission rate before intervention	Assumed
β_2	3.43×10^{-9} day^{-1}	Transmission rate during and after intervention	Assumed
α	3.5×10^{-4} day^{-1}	Vaccination rate	[52]
μ	3×10^{-5} persons/day	Natural death rate	[51]
γ^{-1}	5.5 days	Incubation period	[42]
σ	0.05	Vaccine inefficacy	[53]
δ^{-1}	3.8 days	Infection time	[42]
κ	0.014	Case fatality rate	[50]
λ^{-1}	10 days	Recovery time	[42]
ρ^{-1}	15 days	Time until death	[42]

In order to initialize the assimilation system, we perturb the initial state using a Gaussian noise as follows:

$$x_0^i = \max[x_0(1 + \eta \cdot \sigma), 0] \qquad i = 1, 2, \cdots, N_e, \tag{30}$$

where $\sigma \sim \mathcal{N}(0, 1)$ and η is set to 10%. Similarly, the parameters' ensemble is generated by perturbing the initial values, given in Table 1, using a log-normal distribution.

To evaluate the accuracy of the assimilation system solution, we compute the Relative Mean Absolute Error (RMAE), defined as follows:

$$\text{RMAE} = \frac{1}{n} \sum_{j=1}^{n} \frac{|y(j) - x(j)|}{|y(j)|}, \tag{31}$$

where $y(j)$ represents the observed value, $x(j)$ is the model simulation, and n is the sample size of the observed data.

4.2. Sensitivity to Ensemble Size

We first study the sensitivity of the filter, to the ensemble size, N_e. Computing accurate state and parameter estimates with small ensembles is desirable to avoid an important increase in computational time. We run our assimilation system using seven ensemble sizes: $N_e = 10, 20, 50, 100, 200, 500$, and 1000. Figure 2 presents the performance of the EnKF in terms of RMAE for the different ensemble sizes. As expected, increasing the ensemble size reduces the discrepancy between the estimates and the data. A clear improvement over the case with 20 ensemble members is observed. However, increasing the ensemble size beyond 200 members does not notably enhance the performance of the filter, although it dramatically increases the computational cost.
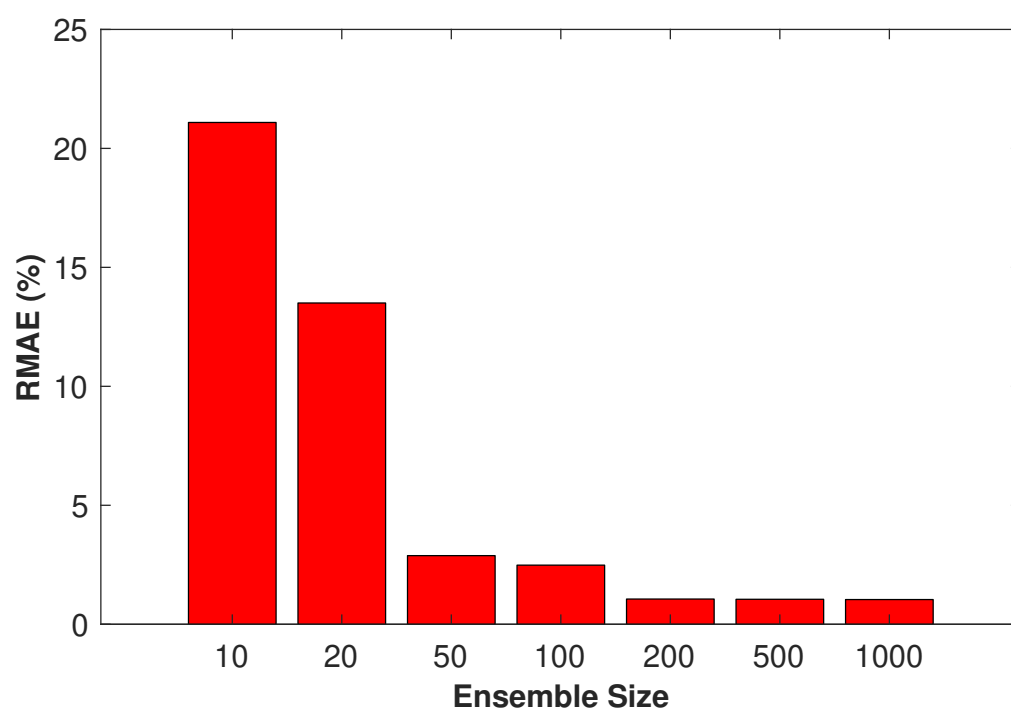
**Figure 2.** RMAE of the number of deaths as a function of ensemble size.

Figure 3 summarizes the assimilation results for the active, recovered, and confirmed cases obtained by the EnKF using 200 ensemble numbers. All results show a good fit between the predicted and observed values, with generally minor differences. The values of the model parameters estimated by the EnKF during the assimilation process are sketched in Figure 4. The parameters do not appear to be dramatically different from the initial values, except for the reproduction number and the recovery time. We observe a sharp increase in the reproduction number between March 1st and 15th. This number remains relatively high, almost $R_0 = 4$, during the pre-intervention period. During and after the intervention, the value of R_0 remains close to 1. That suggests that the intervention was very fruitful in controlling the pandemic. The recovery time increases significantly during the first two months, peaking at 16.1, and remains constant after that. We also see a slight decrease in the time until death to 13.75. The other parameters show slight deviations from the proposed initial values.
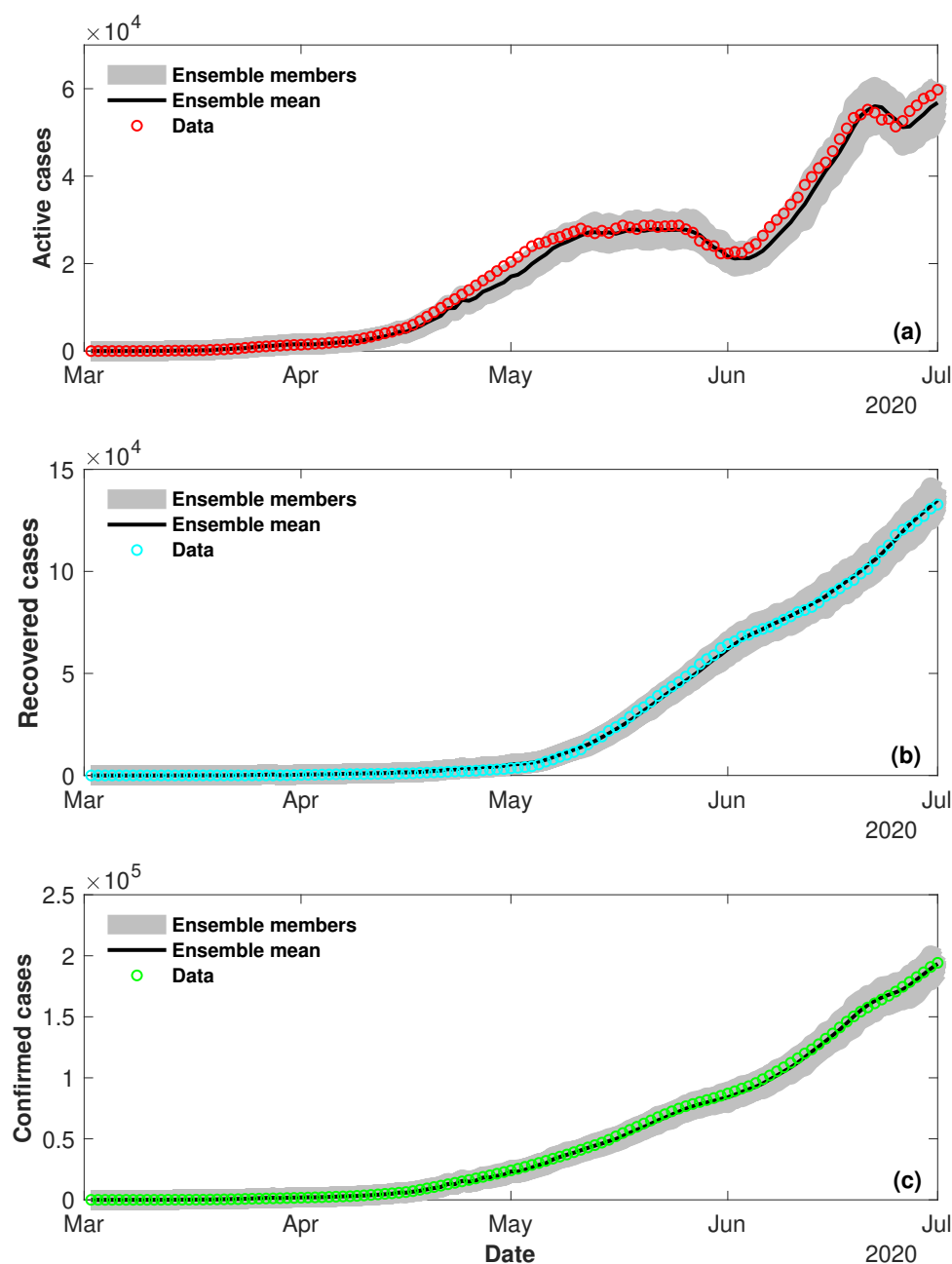
Figure 3. Assimilated numbers of: (**a**) Active cases. (**b**) Recovered cases. (**c**) Confirmed cases.

4.3. Predictability Experiment

To assess the system relevance for making short-term predictions, we daily performed two-week forecasts, starting from 1st July until 17th December (the day where the vaccination campaign started in Saudi Arabia). The estimated state and parameters during the assimilation process were used to conduct these forecasts. The RMAE of the deaths, and recovered and confirmed cases were then computed and plotted in Figure 5. The model performed very well. The values of the RMAE for the forecasted recovered and confirmed cases are always less than 5%, within the margin of the observation errors. Slightly higher RMAE are observed during the month of July for the death cases, reaching 12%. However, the model achieves a lower RMAE in the next few months, with an average of 3%.
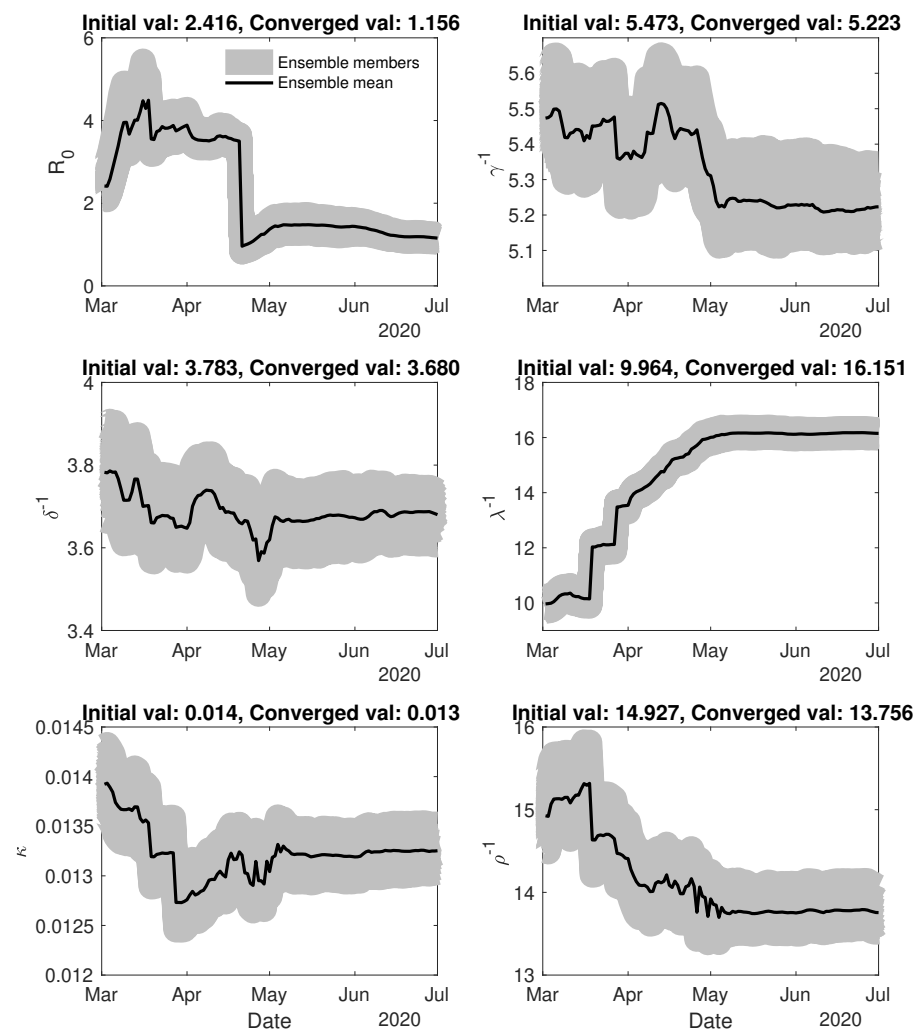
Figure 4. Time series of the estimated model parameters using the Joint-EnKF with 200 ensemble members. Initial and converged values for each parameter are also displayed.
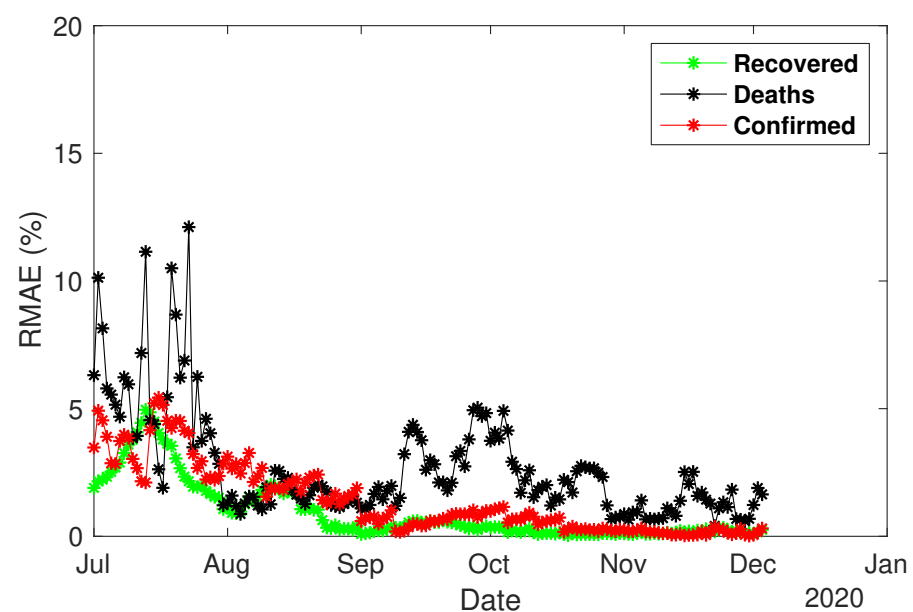
Figure 5. RMAE of the daily two-week forecasts, starting from July 1st until December 3rd.

Figure 6 shows the final predictions performed on 3rd December for the first two weeks of December. The predicted numbers of deaths, and recovered and confirmed cases are all in very good agreement with the data. While they perfectly match the data during the first 10 days, they tend to slightly underestimate or overestimate the observed values at the end of the second week. Overall, our numerical results demonstrate the capability of the proposed system in achieving an excellent two-week prediction skill.
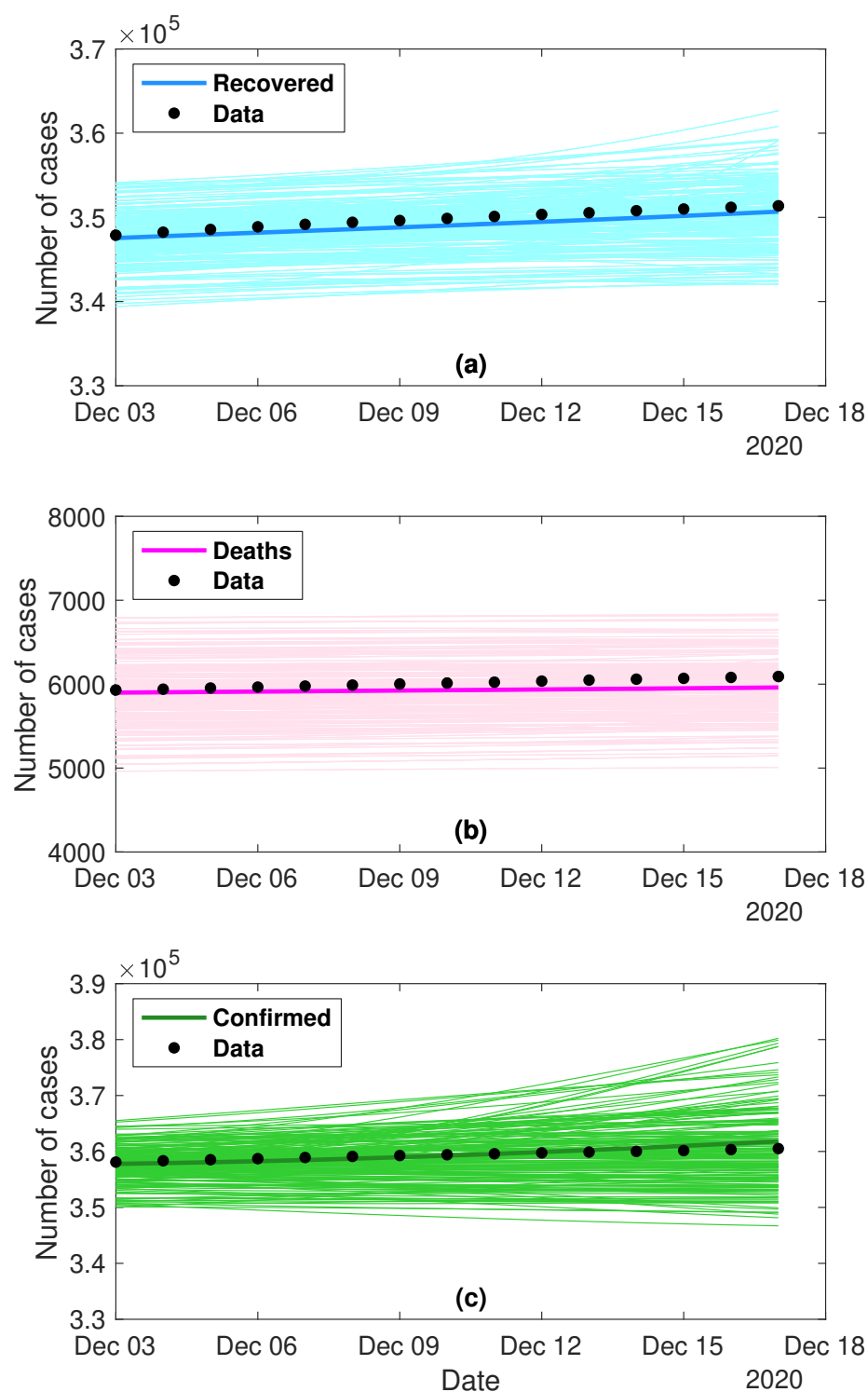

Figure 6. Predicted numbers of: (**a**) Recovered cases. (**b**) Death cases. (**c**) Confirmed cases for the first two weeks of December.

4.4. Impact of Vaccination

On 17th December, Saudi Arabia started a three-phase COVID-19 vaccination program, thus becoming the first Arab country to roll out the Pfizer-BioNTech vaccine. The first phase targets people aged 65 and above, as well as people most exposed to the disease. The second phase will include people aged over 50, as well as those with specific chronic diseases, such as asthma and diabetes. The rest of the population will get the vaccine at the third stage. However, the Ministry has identified those who should not get the vaccine. They are women who are pregnant or breastfeeding, women who plan on getting pregnant in the next two months, people who suffer from severe allergic reactions, and those who have been infected by the virus in the past 90 days.

As of 17 January 2021, nearly 295,000 people have been vaccinated against COVID-19, according to the Saudi Health Council [52]. That gives $\alpha = 3 \times 10^{-4}$. The Pfizer-BioNtech vaccine offers up to 95% protection against Covid-19 after a second dose [53], which gives $\sigma = 0.05$. We start our numerical simulation on December 17th for a period of six months with and without vaccination. Numerical results are presented in Figure 7, where $\alpha = 0$ corresponds to the model output without vaccination. We notice that the current pace of vaccination does not show a significant decrease in the number of cases. We observe only a reduction of 9.02% in the total confirmed cases, and 7.47% in the total deaths. However, the vaccination rate is expected to increase in the upcoming weeks. Therefore, we consider three more different values of α $\left(6 \times 10^{-4}, 1.2 \times 10^{-3}, \text{ and } 2.4 \times 10^{-3}\right)$. As we can see in Figure 7, intensifying the vaccination campaign starts to significantly decrease the number of cases. For $\alpha = 0.0006$, we observe a reduction up to 16.08% and 13.47% for the total confirmed cases and deaths, respectively. Finally, a notable reduction is observed for $\alpha = 0.0024$. Results show a decline by 38.67% and 33.65%, respectively.
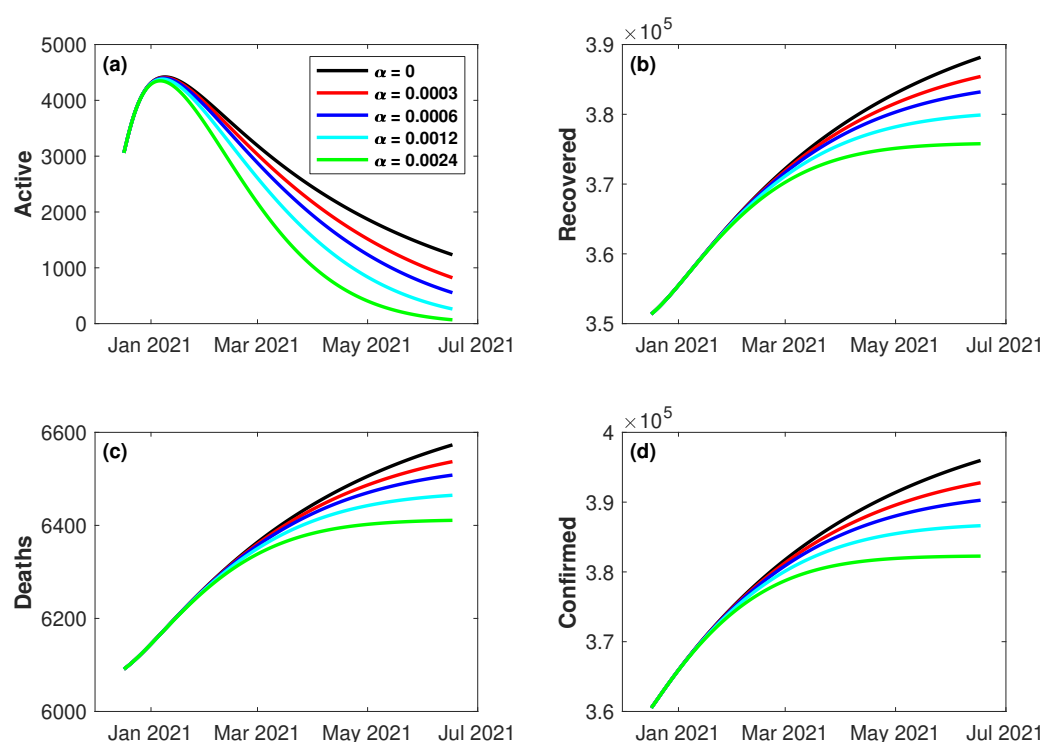
Figure 7. Impact of vaccination rate on the total number of: (**a**) Active cases. (**b**) Recovered cases. (**c**) Death cases. (**d**) Confirmed cases. $\alpha = 0$ corresponds to the model output without vaccination.

5. Conclusions

This paper presented an extended SEIR model with seven stages of infection, including vaccination, to simulate and predict the development of the novel COVID-19 outbreak. Initially, a mathematical analysis was carried out to illustrate the non-negativity, boundedness,

epidemic equilibrium, existence, and uniqueness of the endemic equilibrium, and basic reproduction number of the proposed model. To mitigate the imperfection of SEIR models and the poorly known parameters, we first applied a data assimilation framework based on a Joint-EnKF for state-parameters estimation in order to enhance the model forecast skill. We assimilated real data of the number of deaths, the number of recovered cases, and the number of active cases. In the second step, the updated state and parameters were used to perform predictions two weeks ahead in order to assess the relevance of the system of making short-term predictions of the COVID-19 pandemic. Finally, we investigated the impact of vaccination on the spread of the disease.

The proposed methodology was validated with cases reported from Saudi Arabia. In the assimilation process, the estimated numbers of active, recovered, deaths, and confirmed cases were accurately predicted by the model, with generally minor differences. In the predictability experiments, the model provided encouraging results for short-term predictions, with a relatively small ensemble size. Moreover, model predictions can be updated weekly or every two weeks to accommodate short- and medium-term predictions. Results also show that intensifying the vaccination campaign significantly decreases the number of confirmed cases and deaths. The proposed model can serve as a tool for health authorities to plan, prepare, and take appropriate measures and decisions to control the pandemic.

References

1. World Health Organization (WHO). Coronavirus Disease (COVID-19) Outbreak Situation. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 17 December 2020).
2. Kermack, W.O.; McKendrick, A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* **1927**, *115*, 700–721.
3. Giraldo, J.O.; Palacio, D.H. Deterministic SIR (Susceptible–Infected–Removed) models applied to varicella outbreaks. *Epidemiol. Infect.* **2008**, *136*, 679–687. [CrossRef] [PubMed]
4. Fung, I.C.H. Cholera transmission dynamic models for public health practitioners. *Emerg. Themes Epidemiol.* **2014**, *11*, 1–11. [CrossRef]
5. Trawicki, M.B. Deterministic SEIRs epidemic model for modeling vital dynamics, vaccinations, and temporary immunity. *Mathematics* **2017**, *5*, 7. [CrossRef]
6. Bjørnstad, O.N.; Shea, K.; Krzywinski, M.; Altman, N.S. The SEIRS model for infectious disease dynamics. *Nat. Methods* **2020**, *17*, 557–558. [CrossRef]
7. Kumar, S.; Ahmadian, A.; Kumar, R.; Kumar, D.; Singh, J.; Baleanu, D.; Salimi, M. An efficient numerical method for fractional SIR epidemic model of infectious disease by using Bernstein wavelets. *Mathematics* **2020**, *8*, 558. [CrossRef]
8. Cooper, I.; Mondal, A.; Antonopoulos, C.G. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* **2020**, *139*, 110057. [CrossRef]
9. Bärwolff, G. Mathematical modeling and simulation of the COVID-19 pandemic. *Systems* **2020**, *8*, 24. [CrossRef]
10. Ndaïrou, F.; Area, I.; Nieto, J.J.; Torres, D.F. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* **2020**, *135*, 109846. [CrossRef]
11. Sarkar, K.; Khajanchi, S.; Nieto, J.J. Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals* **2020**, *139*, 110049. [CrossRef] [PubMed]
12. Ghil, M. Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dynam. Atmos. Ocean* **1989**, *13*, 171–218. [CrossRef]
13. Holland, W.R.; Malanotte-Rizzoli, P. Assimilation of altimeter data into an ocean circulation model: Space versus time resolution studies. *J. Phys. Oceanogr.* **1989**, *19*, 1507–1534. [CrossRef]

14. Edwards, C.A.; Moore, A.M.; Hoteit, I.; Cornuelle, B.D. Regional ocean data assimilation. *Ann. Rev. Mar. Sci.* **2015**, *7*, 21–42. [CrossRef] [PubMed]
15. Hoteit, I.; Pham, D.T.; Blum, J. A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *J. Mar. Syst.* **2002**, *36*, 101–127. [CrossRef]
16. Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **2003**, *53*, 343–367. [CrossRef]
17. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Mar. Syst.* **1960**, *82*, 35–45. [CrossRef]
18. Altaf, M.; Butler, T.; Mayo, T.; Luo, X.; Dawson, C.; Heemink, A.; Hoteit, I. A comparison of ensemble Kalman filters for storm surge assimilation. *Mon. Weather Rev.* **2014**, *142*, 2899–2914. [CrossRef]
19. Buehner, M.; McTaggart-Cowan, R.; Heilliette, S. An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: VarEnKF. *Mon. Weather Rev.* **2017**, *145*, 617–635. [CrossRef]
20. Raboudi, N.F.; Ait-El-Fquih, B.; Dawson, C.; Hoteit, I. Combining Hybrid and One-Step-Ahead Smoothing for Efficient Short-Range Storm Surge Forecasting with an Ensemble Kalman Filter. *Mon. Weather Rev.* **2019**, *147*, 3283–3300. [CrossRef]
21. Reichle, R.H.; McLaughlin, D.B.; Entekhabi, D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.* **2002**, *130*, 103–114. [CrossRef]
22. Gharamti, M.; Kadoura, A.; Valstar, J.; Sun, S.; Hoteit, I. Constraining a compositional flow model with flow-chemical data using an ensemble-based Kalman filter. *Water Resour. Res.* **2014**, *50*, 2444–2467. [CrossRef]
23. Gharamti, M.; Ait-El-Fquih, B.; Hoteit, I. An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. *J. Hydrol.* **2015**, *527*, 442–457. [CrossRef]
24. Ait-El-Fquih, B.; Gharamti, M.E.; Hoteit, I. A Bayesian consistent dual ensemble Kalman filter for state-parameter estimation in subsurface hydrology. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3289–3307. [CrossRef]
25. Khaki, M.; Ait-El-Fquih, B.; Hoteit, I.; Forootan, E.; Awange, J.; Kuhn, M. A two-update ensemble Kalman filter for land hydrological data assimilation with an uncertain constraint. *J. Hydrol.* **2017**, *555*, 447–462. [CrossRef]
26. Khaki, M.; Ait-El-Fquih, B.; Hoteit, I. Calibrating land hydrological models and enhancing their forecasting skills using an ensemble Kalman filter with one-step-ahead smoothing. *J. Hydrol.* **2020**, *584*, 124708. [CrossRef]
27. Hoteit, I.; Hoar, T.; Gopalakrishnan, G.; Collins, N.; Anderson, J.; Cornuelle, B.; Köhl, A.; Heimbach, P. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dynam. Atmos. Ocean* **2013**, *63*, 1–23. [CrossRef]
28. Triantafyllou, G.; Hoteit, I.; Luo, X.; Tsiaras, K.; Petihakis, G. Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea. *J. Mar. Syst.* **2013**, *125*, 90–100. [CrossRef]
29. Gharamti, M.; Tjiputra, J.; Bethke, I.; Samuelsen, A.; Skjelvan, I.; Bentsen, M.; Bertino, L. Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Model.* **2017**, *112*, 65–89. [CrossRef]
30. Toye, H.; Kortas, S.; Zhan, P.; Hoteit, I. A fault-tolerant HPC scheduler extension for large and operational ensemble data assimilation: Application to the Red Sea. *J. Comput. Sci.* **2018**, *27*, 46–56. [CrossRef]
31. Yu, L.; Fennel, K.; Bertino, L.; El Gharamti, M.; Thompson, K.R. Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling. *Ocean Model.* **2018**, *126*, 13–28. [CrossRef]
32. Neal, J.C.; Atkinson, P.M.; Hutton, C.W. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. *J. Hydrol.* **2007**, *336*, 401–415. [CrossRef]
33. Kimura, N.; Hsu, M.H.; Tsai, M.Y.; Tsao, M.C.; Yu, S.L.; Tai, A. A river flash flood forecasting model coupled with ensemble Kalman filter. *J. Flood Risk Manag.* **2016**, *9*, 178–192. [CrossRef]
34. Ziliani, M.G.; Ghostine, R.; Ait-El-Fquih, B.; McCabe, M.F.; Hoteit, I. Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. *J. Hydrol.* **2019**, *577*, 123924. [CrossRef]
35. Bettencourt, L.M.; Ribeiro, R.M.; Chowell, G.; Lant, T.; Castillo-Chavez, C. Towards real time epidemiology: Data assimilation, modeling and anomaly detection of health surveillance data streams. In *NSF Workshop on Intelligence and Security Informatics*; Springer: Berlin, Germany 2007; pp. 79–90.
36. Rhodes, C.; Hollingsworth, T.D. Variational data assimilation with epidemic models. *J. Theor. Biol.* **2009**, *258*, 591–602. [CrossRef] [PubMed]
37. Gupta, H.; Verma, K.K.; Sharma, P. Using data assimilation technique and epidemic model to predict tb epidemic. *Int. J. Comput. Appl.* **2015**, *128*, 5. [CrossRef]
38. Eksin, C.; Paarporn, K.; Weitz, J.S. Systematic biases in disease forecasting – the role of behavior change. *Epidemics* **2019**, *27*, 96–105. [CrossRef] [PubMed]
39. Engbert, R.; Rabe, M.M.; Kliegl, R.; Reich, S. Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics. *Bull. Math. Biol.* **2021**, *83*, 1–16. [CrossRef]
40. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **2020**, *368*, 489–493. [CrossRef] [PubMed]
41. Van Wees, J.D.; Osinga, S.; Van Der Kuip, M.; Tanck, M.W.; Tutu-van Furth, A. Forecasting hospitalization and ICU rates of the COVID-19 outbreak: An efficient SEIR model. *Bull World Health Organ* **2020**. [CrossRef]
42. Evensen, G.; Amezcua, J.; Bocquet, M.; Carrassi, A.; Farchi, A.; Fowler, A.; Houtekamer, P.; Jones, C.K.; de Moraes, R.; Pulido, M.; et al. An international assessment of the COVID-19 pandemic using ensemble data assimilation. *medRxiv* **2020**. [CrossRef]

43. Yang, Q.; Yi, C.; Vajdi, A.; Cohnstaedt, L.W.; Wu, H.; Guo, X.; Scoglio, C.M. Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China. *Infect. Dis. Model.* **2020**, *5*, 563–574.

44. Nkwayep, C.H.; Bowong, S.; Tewa, J.; Kurths, J. Short-term forecasts of the COVID-19 pandemic: Study case of Cameroon. *Chaos Solitons Fractals* **2020**, *140*, 110106. [CrossRef] [PubMed]

45. Sesterhenn, J.L. Adjoint-based data assimilation of an epidemiology model for the covid-19 pandemic in 2020. *arXiv* **2020**, arXiv:2003.13071.

46. Armstrong, E.; Runge, M.; Gerardin, J. Identifying the measurements required to estimate rates of COVID-19 transmission, infection, and detection, using variational data assimilation. *Infect. Dis. Model.* **2020**, *6*, 133–147. [CrossRef]

47. Van den Driessche, P.; Watmough, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **2002**, *180*, 29–48. [CrossRef]

48. Diekmann, O.; Heesterbeek, J.A.P.; Metz, J.A. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **1990**, *28*, 365–382. [CrossRef]

49. Martcheva, M. *An Introduction to Mathematical Epidemiology*; Springer: Berlin, Germany, 2015; Volume 61.

50. Saudi Center for Diseases Prevention and Control. Available online: https://covid19.cdc.gov.sa/daily-updates/ (accessed on 17 December 2020).

51. Saudi Ministry of Health. Available online: https://www.moh.gov.sa/en/Ministry/Statistics/Indicator/Pages/Indicator-1440.aspx (accessed on 17 December 2020).

52. Saudi Health Council. Available online: https://coronamap.sa (accessed on 19 January 2021).

53. Polack, F.P.; Thomas, S.J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J.L.; Pérez Marc, G.; Moreira, E.D.; Zerbini, C.; et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603–2615. [CrossRef]