Article

Mathematical Modeling of Immune Responses against SARS-CoV-2 Using an Ensemble Kalman Filter

Rabih Ghostine ^{1,*} , Mohamad Gharamti ², Sally Hassrouny ³ and Ibrahim Hoteit ⁴¹ Department of Mathematics, Kuwait College of Science and Technology, Doha 35001, Kuwait² National Center for Atmospheric Research, Boulder, CO 80305, USA; gharamti@ucar.edu³ School of Engineering, American International University, Doha 35001, Kuwait; s.hassrouny@aiu.edu.kw⁴ 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, a mathematical model was developed to simulate SARS-CoV-2 dynamics in infected patients. The model considers both the innate and adaptive immune responses and consists of healthy cells, infected cells, viral load, cytokines, natural killer cells, cytotoxic T-lymphocytes, B-lymphocytes, plasma cells, and antibody levels. First, a mathematical analysis was performed to discuss the model's equilibrium points and compute the basic reproduction number. The accuracy of such mathematical models may be affected by many sources of uncertainties due to the incomplete representation of the biological process and poorly known parameters. This may strongly limit their performance and prediction skills. A state-of-the-art data assimilation technique, the ensemble Kalman filter (EnKF), was then used to enhance the model's behavior by incorporating available data to determine the best possible estimate of the model's state and parameters. The proposed assimilation system was applied on the real viral load datasets of six COVID-19 patients. The results demonstrate the efficiency of the proposed assimilation system in improving the model predictions by up to 40%.

Keywords: mathematical modeling; SARS-CoV-2; immune response; ensemble Kalman filter; joint state–parameters estimation

Citation: Ghostine, R.; Gharamti, M.; Hassrouny, S.; Hoteit, I. Mathematical Modeling of Immune Responses against SARS-CoV-2 Using an Ensemble Kalman Filter. *Mathematics* **2021**, *9*, 2427. <https://doi.org/10.3390/math9192427>

Academic Editor: Dumitru Baleanu

Received: 7 August 2021

Accepted: 2 September 2021

Published: 30 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since being first identified in the late fall of 2019, the outbreak of coronavirus disease (COVID-19) has received increasing attention among the scientists all over the world in order to understand the epidemiology of the disease and curb the spread of the virus. COVID-19 is a contagious disease caused by the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Infected people will experience mild to moderate symptoms such as a cough, headache, fever, and in some cases, severe pneumonia [1,2]. The development of a vaccine to end the disease has undergone significant progress during the last two years. However, the virus is still going on, as the mortality rate and daily confirmed cases remain high in many countries. As efforts to eradicate the disease continue, in addition to biological and medical research, mathematical models have been suggested to play a vital role in describing and forecasting disease transmission, as well as help to implement efficient measures and strategies to control the spread of the pandemic and reduce its impact [3].

Several models have recently been proposed and implemented in order to understand the dynamics of COVID-19. Most of these models are based on the traditional epidemiological model, called the susceptible–exposed–infectious–recovered (SEIR) model, which was introduced to study the pandemic transmission within a human community [4–14]. To date, very few attempts have been devoted to describe the interaction between the immune system and SARS-CoV-2 [15–18]. An important limitation of these models is the

presence of several sources of uncertainties, which may strongly degrade their prediction performance and lead to large inaccuracies in their outputs. Aside from the incomplete description of the biological processes, the poor characterization of the model parameters is also considered a major source of uncertainty. Therefore, it is crucial to quantify and reduce the model's uncertainties before using them for decision making.

Data assimilation techniques provide efficient tools for quantifying and reducing the uncertainties of such models [19,20]. They constrain prior information about the state and parameters of the underlying dynamical system with available data to compute posterior updated estimates of the parameters and state variables [21], based upon which improved predictions can be obtained with the model. This process, which finds its roots in the atmospheric and oceanic weather communities, sequentially nudges model predictions towards incoming observations, so the model trajectory remains as close as possible to the true trajectory [22]. The ensemble Kalman filter (EnKF) is an efficient data assimilation technique that has been successfully implemented in many fields including atmospheric and oceanic weather forecasting [23–27], hydrology and reservoir applications [28–31], and flood modeling [32,33]. EnKF is a Monte Carlo implementation of the classical linear Kalman filter (KF) [34], operating with cycles of forecast and update steps. In the forecast step, a set of members sampled from the state probability distribution were integrated forward in time with the model. In the update step, a Gaussian Kalman update is applied, once a new datum becomes available [35], from which the next assimilation cycle is initiated [22,35]. Unlike the KF, the EnKF can efficiently handle model nonlinearities without the need to compute tangent linear approximations of the model as for the extended Kalman filter.

In this work, we propose a mathematical model for the innate and adaptive immune responses against SARS-CoV-2 infection. First, we conducted a mathematical analysis to derive the basic reproduction number and discuss the equilibrium states of the model. To improve the model prediction skill, an efficient state–parameters estimation framework based on a joint-EnKF assimilation scheme was developed. By considering uncertainties in the model state variables and poorly known parameters, the proposed ensemble data assimilation system updates both state and parameters ensembles every time a new observation become available. This methodology was validated with real SARS-CoV-2 viral load data obtained from a hospital in Munich. We assessed the sensitivity of the system to the ensemble size and particularly focused on the relevance of the joint-EnKF for improving the prediction capability of the immune response mathematical model.

The rest of the study is organized as follows. Section 2 presents the mathematical model and briefly outlines the joint-EnKF algorithm. Section 3 is dedicated to the mathematical analysis of the model. The numerical results are presented and discussed in Section 4. Key findings and conclusions are summarized in Section 5.

2. Model and Assimilation Scheme

2.1. Mathematical Model

The mathematical model describes the interaction dynamics of both the human innate and adaptive immune system with SARS-CoV-2 as follows. Healthy cells $H(t)$ become infected at a rate β . The virus $V(t)$ is produced by infected cells at a rate k_1 . The innate immune system instantly responds after the viral infection. Cytokines $C(t)$ are a crucial element of the innate system and play a major role in inhibiting the viral replication at a rate p_2 . Specific cytokines activate natural killer cells $N(t)$ at a rate r , which play an essential role in eliminating infected cells at a rate p_5 . At the same time, cytokines activate the adaptive immune system, particularly the cytotoxic T-lymphocytes $T(t)$ at a rate λ_1 . T-cells kill virus infected cells at a rate p_1 and afterwards activate B-lymphocytes $B(t)$ at a rate λ_2 . A ratio k of B-cells differentiate into plasma cells in an antigen-dependent manner to produce antibodies at a rate η . They specifically recognize and neutralize the virus at a rate p_3 . A schematic diagram characterizing the proposed model is presented in

Figure 1. Finally, the system describing these interactions is expressed as a set of ordinary differential equations:

$$\begin{aligned}
 \frac{dH(t)}{dt} &= \Pi - \beta H(t)V(t) - \mu_1 H(t), \\
 \frac{dI(t)}{dt} &= \beta H(t)V(t) - p_1 T(t)I(t) - p_5 N(t)I(t) - \mu_2 I(t), \\
 \frac{dV(t)}{dt} &= k_1 I(t) - p_2 C(t)V(t) - p_3 A(t)V(t) - \mu_3 V(t), \\
 \frac{dC(t)}{dt} &= k_2 I(t) - \mu_4 C(t), \\
 \frac{dN(t)}{dt} &= rC(t) - \mu_5 N(t), \\
 \frac{dT(t)}{dt} &= \lambda_1 C(t)T(t) - \mu_6 T(t), \\
 \frac{dB(t)}{dt} &= \lambda_2 T(t)B(t) - \mu_7 B(t), \\
 \frac{dP(t)}{dt} &= k B(t)V(t) - \mu_8 P(t), \\
 \frac{dA(t)}{dt} &= G(t - \tau)\eta P(t) - p_4 A(t)V(t) - \mu_9 A(t),
 \end{aligned} \tag{1}$$

where G is the Heaviside step function defined as follows:

$$G(t - \tau) = \begin{cases} 0 & \text{if } t < \tau \\ 1 & \text{if } t > \tau \end{cases}, \tag{2}$$

where the time delay τ is the time required to produce antibodies from plasma cells. Detailed descriptions and definitions of all the model state and parameters are given in Table 1.

Figure 1. Schematic diagram of the proposed mathematical model. The cyan arrows indicate production, black arrows indicate activation, and orange arrows indicate attack.

Table 1. Initial model variables and parameters and their biological interpretations.

Variable	Initial Value	Description	Reference
H	4×10^5 cells per mL	Healthy cells	[36]
I	3×10^{-4} cells per mL	Infected cells	[36]
V	357 RNA per mL	Viral load	[36]
C	0 cells per mL	Cytokines	[16]
N	100 cells per mL	Natural killer cells	[16]
T	500 cells per mL	T-cells	[16]
B	100 cells per mL	B-cells	[16]
P	0 cells per mL	Plasma cells	[36]
A	0 molecules per ml	Antibodies level	[36]
Parameter	Initial value	Description	Reference
Π	4×10^3 cells mL ⁻¹ day ⁻¹	Production rate of healthy cells	[37]
β	1.5×10^{-7} mL (RNA copies) ⁻¹ day ⁻¹	Rate at which healthy cells are infected	[17]
p_1	0.001 mL cells ⁻¹ day ⁻¹	Rate at which T-cells destroy infected cells	[38]
p_2	(0–1) mL cells ⁻¹ day ⁻¹	Rate at which viral particles are neutralized by cytokines	[16]
p_3	(0–1) mL molecules ⁻¹ day ⁻¹	Rate at which viral particles are neutralized by antibodies	[16]
p_4	3×10^{-7} mL (RNA copies) ⁻¹ day ⁻¹	Rate at which virus neutralizes antibodies	[36]
p_5	5.74×10^{-4} mL cells ⁻¹ day ⁻¹	Rate at which infected cells are diminished by NK cells	[39]
k	10^{-5} ml (RNA copies) ⁻¹ day ⁻¹	Plasma cells recruitment	[36]
k_1	(8.2–525) day ⁻¹	Production rate of virus from infected cells	[17]
k_2	(0–10) day ⁻¹	Production rate of cytokines	[16]
r	0.52 day ⁻¹	Activation rate of NK cells	[39]
λ_1	0.1 mL cells ⁻¹ day ⁻¹	Activation rate of T cells	[40]
λ_2	0.01 mL cells ⁻¹ day ⁻¹	Activation rate of B cells	[40]
η	(0–1) day ⁻¹	Antibodies production rate	[41]
μ_1	0.14 day ⁻¹	Natural death rate of healthy cells and protected cells	[40]
μ_2	(0–1) day ⁻¹	Natural death rate of infected cells	[16]
μ_3	(0–1) day ⁻¹	Clearance rate of virus	[16]
μ_4	0.7 day ⁻¹	Natural death rate of cytokines	[16]
μ_5	0.07 day ⁻¹	Natural death rate of NK cells	[39]
μ_6	1 day ⁻¹	Natural death rate of T cells	[40]
μ_7	0.2 day ⁻¹	Natural death rate of B cells	[36]
μ_8	4 day ⁻¹	Natural death rate of plasma cells	[18]
μ_9	0.07 day ⁻¹	Natural death rate of antibodies	[40]
τ	7–14 days	Time delay for antibody production	[16]

2.2. Ensemble Kalman Filter (EnKF)

Consider the following space–time dynamical system:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathcal{M}(\mathbf{x}_k, \theta) + \eta_{k+1}, \\ \mathbf{y}_{k+1} &= \mathcal{H}(\mathbf{x}_{k+1}) + \epsilon_{k+1}, \end{aligned} \quad (3)$$

where the model's prognostic state is denoted by $\mathbf{x}_k$, $\mathcal{M}$ is the nonlinear predictive model and θ denotes a set of modeling parameters. The observations at time t_{k+1} are given by $\mathbf{y}_{k+1}$. The state variables are mapped into the observation space using the $\mathcal{H}$ operator. Model errors η_{k+1} are assumed Gaussian, centered at zero with a prescribed covariance $\mathbf{Q}_{k+1}$. Observation errors ϵ_{k+1} , on the other hand, are assumed to be uncorrelated with the model errors and further follow $\mathcal{N}(0, \mathbf{R}_{k+1})$.

During the forecast step, the model is used to integrate an ensemble of states forward in time. The prior sample mean and covariance are then approximated as

$$\hat{\mathbf{x}}_{k+1}^f = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_{k+1}^{f,i}, \quad (4)$$

$$\hat{\mathbf{P}}_{k+1}^f = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{x}_{k+1}^{f,i} - \hat{\mathbf{x}}_{k+1}^f) (\mathbf{x}_{k+1}^{f,i} - \hat{\mathbf{x}}_{k+1}^f)^T = \mathbf{X}_{k+1}^f (\mathbf{X}_{k+1}^f)^T, \quad (5)$$

where N_e is the ensemble size.

By augmenting the state and parameters— $\mathbf{z}_{k+1} = [\mathbf{x}_{k+1}^T, \boldsymbol{\theta}^T]^T$ —the joint-EnKF uses y_{k+1} to update the prior joint state as follows:

$$\mathbf{z}_{k+1}^f = \tilde{\mathcal{M}}(\mathbf{z}_k, \boldsymbol{\theta}) + \eta_{k+1} = \begin{bmatrix} \mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta}) + \eta_{k+1} \\ \boldsymbol{\theta} \end{bmatrix}, \quad (6)$$

such that $\tilde{\mathcal{M}}$ is a modified nonlinear function operating on the augmented state–parameters vector. Using the ensemble members of both the state and parameters, the joint prior covariance can be formed as

$$\hat{\mathbf{P}}_{k+1}^f = \begin{bmatrix} \hat{\mathbf{P}}_{\mathbf{x}_{k+1}\mathbf{x}_{k+1}}^f & \hat{\mathbf{P}}_{\mathbf{x}_{k+1}\boldsymbol{\theta}_{k+1}}^f \\ \hat{\mathbf{P}}_{\boldsymbol{\theta}_{k+1}\mathbf{x}_{k+1}}^f & \hat{\mathbf{P}}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{k+1}}^f \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{\mathbf{x}_{k+1}}^f (\mathbf{X}_{\mathbf{x}_{k+1}}^f)^T & \mathbf{X}_{\mathbf{x}_{k+1}}^f (\mathbf{X}_{\boldsymbol{\theta}_{k+1}}^f)^T \\ \mathbf{X}_{\boldsymbol{\theta}_{k+1}}^f (\mathbf{X}_{\mathbf{x}_{k+1}}^f)^T & \mathbf{X}_{\boldsymbol{\theta}_{k+1}}^f (\mathbf{X}_{\boldsymbol{\theta}_{k+1}}^f)^T \end{bmatrix}, \quad (7)$$

where $\mathbf{X}_{\boldsymbol{\theta}}$ represents the departure of the parameter ensemble from the mean, i.e., ensemble anomalies. A linear Kalman update is applied to compute the analysis state and parameter ensembles:

$$\begin{pmatrix} \mathbf{x}_{k+1}^{a,i} \\ \boldsymbol{\theta}_{k+1}^{a,i} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{k+1}^{f,i} \\ \boldsymbol{\theta}_{k+1}^{f,i} \end{pmatrix} + \begin{pmatrix} \mathbf{X}_{\mathbf{x}_{k+1}}^f \\ \mathbf{X}_{\boldsymbol{\theta}_{k+1}}^f \end{pmatrix} \cdot \mathbf{w}_i, \quad (8)$$

$$\text{where } \mathbf{w}_i = \left(\mathcal{H} \mathbf{X}_{k+1}^f \right)^T \left[\left(\mathcal{H} \mathbf{X}_{k+1}^f \right) \left(\mathcal{H} \mathbf{X}_{k+1}^f \right)^T + R \right]^{-1} \cdot \left(y_{k+1}^i - \mathcal{H} \mathbf{x}_{k+1}^{f,i} \right).$$

3. Mathematical Analysis of the Immune Response Model

This section discusses the equilibrium states, the stability of the disease-free equilibrium, and the basic reproduction number of the immune response model (1).

3.1. Disease-Free Equilibrium and Basic Reproduction Number of the Model

Model (1) has a unique disease-free equilibrium X_0 obtained by setting all the right-hand side to zero with $V = 0$ as follows:

$$X_0 = \left(\frac{\Pi}{\mu_1}, 0, 0, 0, 0, 0, 0, 0 \right). \quad (9)$$

The basic reproduction number R_0 represents the average number of secondary infected cells produced by a single infected cell in the infection-free environment. We calculated it using the next-generation method [42]. Let F and W denote the matrix of the new infection and the transmission matrix, respectively. At the disease-free equilibrium X_0 , F and W are given by

$$F = \begin{pmatrix} 0 & \beta \Pi / \mu_1 \\ 0 & 0 \end{pmatrix}, \quad W = \begin{pmatrix} \mu_2 & 0 \\ -k_1 & \mu_3 \end{pmatrix}. \quad (10)$$

Following [42], the basic reproduction number R_0 is the dominant eigenvalue of the next-generation matrix FW^{-1} , so that:

$$R_0 = \frac{\Pi \beta k_1}{\mu_1 \mu_2 \mu_3}. \quad (11)$$

3.2. Analysis of the Steady States

Model (1) has four steady states:

1. The disease-free equilibrium X_0 .
2. The virus persistence equilibrium in the absence of immune responses given by

$$X_1 = (H_1, I_1, V_1, 0, 0, 0, 0, 0, 0), \quad (12)$$

where:

$$H_1 = \frac{\Pi}{\mu_1 R_0}, \quad I_1 = \frac{\mu_1 \mu_3 (R_0 - 1)}{\beta k_1}, \quad V_1 = \frac{\mu_1 (R_0 - 1)}{\beta}. \quad (13)$$

We can clearly see that the equilibrium X_1 exists when $R_0 > 1$.

3. The virus persistence equilibrium in the absence of the adaptive immune response given by

$$X_2 = (H_2, I_2, V_2, C_2, N_2, 0, 0, 0, 0), \quad (14)$$

where:

$$H_2 = \frac{\Pi}{\beta V_2 + \mu_1}, \quad I_2 = \frac{\mu_4 C_2}{k_2}, \quad V_2 = \frac{k_1 \mu_4 C_2}{k_2 (p_2 C_2 + \mu_3)}, \quad N_2 = \frac{r C_2}{\mu_5}, \quad (15)$$

and C_2 is the root of the quadratic equation:

$$a_2 C_2^2 + a_1 C_2 + a_0 = 0, \quad (16)$$

with:

$$a_0 = \frac{\mu_5 k_2 (1 - R_0)}{\mu_1 \mu_2 \mu_3}, \quad a_2 = r p_5 (\beta \mu_4 k_1 + \mu_1 k_2 p_2). \quad (17)$$

It is clear that a_2 is always positive, therefore, Equation (16) has a unique positive solution when $a_0 < 0$. Thus, $R_0 > 1$ ensures the existence of X_2 .

4. The all cells and virus co-existence equilibrium is given by

$$X_3 = (H_3, I_3, V_3, C_3, N_3, T_3, B_3, P_3, A_3), \quad (18)$$

where:

$$T_3 = \frac{\mu_7}{\lambda_2}, \quad C_3 = \frac{\mu_6}{\lambda_1}, \quad N_3 = \frac{r \mu_6}{\lambda_1 \mu_5}, \quad I_3 = \frac{\mu_4 \mu_6}{k_2 \lambda_1}. \quad (19)$$

From the first equation of system (1), one has:

$$H_3 = \frac{\Pi}{\beta V_3 + \mu_1}, \quad (20)$$

and from the third one:

$$V_3 = \frac{k_1 I_3}{p_2 C_3 + p_3 A_3 + \mu_3}. \quad (21)$$

Putting the expressions of H_3 and V_3 into the second equation of system (1), we obtain the following expression for A_3 :

$$A_3 = \frac{\mu_1 \mu_2 \mu_3 (R_0 - R_1 - 1)}{\mu_1 p_3 p_5 N_3 + \mu_1 p_1 p_3 T_3 + \mu_1 \mu_2 p_3}, \quad (22)$$

where:

$$R_1 = \frac{(p_5 N_3 + p_1 T_3 + \mu_2)(\beta k_1 I_3 + \mu_1 p_2 C_3) + \mu_1 \mu_3 (p_5 N_3 + p_1 T_3)}{\mu_1 \mu_2 \mu_3}. \quad (23)$$

Finally, the last two equations of system (1) give:

$$P_3 = \frac{p_4 A_3 V_3 + \mu_9 A_3}{G(t - \tau) \eta}, \quad B_3 = \frac{\mu_8 P_3}{\lambda_3 V_3}. \quad (24)$$

Since X_3 should be positive, one can see that the equilibrium X_3 exists if $R_0 > R_1 + 1$.

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

Proof of Theorem 1. The Jacobian matrix of the model (1) at the disease-free equilibrium is defined as

$$J(X^0) = \begin{pmatrix} -\mu_1 & 0 & -\beta\Pi/\mu_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\mu_2 & \beta\Pi/\mu_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_1 & -\mu_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_2 & 0 & -\mu_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r & -\mu_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\mu_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\mu_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\mu_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & G(t-\tau)\eta & -\mu_9 \end{pmatrix}. \quad (25)$$

Clearly, $J(X^0)$ has the eigenvalues $\lambda_1 = -\mu_1$, $\lambda_2 = -\mu_4$, $\lambda_3 = -\mu_5$, $\lambda_4 = -\mu_6$, $\lambda_5 = -\mu_7$, $\lambda_6 = -\mu_8$, and $\lambda_7 = -\mu_9$, which are all negative. The other two eigenvalues λ_8 and λ_9 are solutions of the quadratic equation $\lambda^2 + a_1\lambda + a_0 = 0$, where:

$$a_0 = \mu_2\mu_3(1 - R_0), \quad a_1 = \mu_2 + \mu_3. \quad (26)$$

a_1 is always positive and for $R_0 < 1$, a_0 becomes also positive. By the Routh–Hurwitz condition, the last two eigenvalues λ_8 and λ_9 are negative. Hence, the disease-free equilibrium X^0 is locally asymptotically stable when $R_0 < 1$. For the case $R_0 > 1$, a_0 becomes negative and one of the two eigenvalues λ_8 and λ_9 is positive. Therefore, whenever $R_0 > 1$, the disease-free equilibrium X^0 is unstable [43]. $\square$

4. Results and Discussion

The aforementioned model is applied to forecast the SARS-CoV-2 viral load. Real data including daily viral load measurements from six patients in a hospital in Munich (Germany) are obtained from [44]. The assimilation impact is assessed by comparing the data assimilation output obtained using the joint-EnKF to those of the model free-run without assimilation.

4.1. Assimilation Settings

The state vector ($\mathbf{x}$) to be assimilated includes nine variables: the number of healthy cells H , infected cells I , viral load V , cytokines C , natural killer cells N , cytotoxic T-lymphocytes T , B-lymphocytes B , plasma cells P and antibody levels A :

$$\mathbf{x} = [H \ I \ V \ C \ N \ T \ B \ P \ A], \quad (27)$$

and the parameters vector ($\boldsymbol{\theta}$) contains all model parameters listed in Table 1, meaning that:

$$\boldsymbol{\theta} = [\Pi \ \beta \ p_1 \ p_2 \ p_3 \ p_4 \ p_5 \ k \ k_1 \ k_2 \ r \ \lambda_1 \ \lambda_2 \ \eta \ \mu_1 \ \mu_2 \ \mu_3 \ \mu_4 \ \mu_5 \ \mu_6 \ \mu_7 \ \mu_8 \ \mu_9 \ \tau]. \quad (28)$$

In the joint-EnKF, the state and parameters are simultaneously estimated. Thus, the augmented state vector becomes:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}^T & \log(\boldsymbol{\theta}^T) \end{bmatrix}. \quad (29)$$

To prevent negative input, all parameters are taken in a logarithmic form during the estimation procedure. The total number of estimated variables by the joint state is 33.

The initial state and parameter values are outlined in Table 1 with $p_2 = 0.001$, $p_3 = 0.05$, $k_1 = 379$, $k_2 = 5$, $\eta = 0.05$, $\mu_2 = 0.65$, $\mu_3 = 0.9$, and $\tau = 7$. To initial-

ize the simulation, the state ensemble at time zero is drawn from a truncated Gaussian distribution using the initial state x_0 as:

$$\mathbf{x}_0^i = \max[x_0 + (\kappa x_0)\sigma, 0], \quad i = 1, 2, \dots, N_e. \quad (30)$$

A preprocessing step is performed to ensure all initial state variables are non-negative. The standard deviation of the initial ensemble distribution is chosen as κx_0 , where $\kappa = 10\%$ and σ is a standard normal distribution. Likewise, the parameter ensemble is initialized starting from the nominal parameters of Table 1 in which a log-normal distribution, with a 10% standard deviation, is used to draw the parameter samples. A larger initial uncertainty range is assigned to parameter β where the standard deviation is set to 40%. To prescribe the uncertainty in the data, the observation error covariance at any time t_k is assumed diagonal (i.e., observations error correlations in space vanish) with variances set to 5% of the data values.

The prediction skill of the data assimilation framework is evaluated using the following relative absolute error (RAE):

$$\text{RAE} = \frac{\sum_{k=1}^K |y^o(k) - y^f(k)|}{\sum_{k=1}^K |y^o(k)|}, \quad (31)$$

where $y^o(k)$ denotes the k^{th} value of the observed data, $y^f(k)$ is the associated forecast mean and K is the sample size of the observed data at that time.

4.2. Sensitivity to Ensemble Size

Initially, we investigated the EnKF sensitivity to the ensemble size N_e , which is a key feature of the EnKF. Running the filter using large ensemble sizes is undoubtedly more costly. Therefore, computing reliable state and parameter estimation using small ensembles is often desired. We run the assimilation system using six ensembles with different sizes varying from $N_e = 25$ to 1000, and report the average RAE of the forecasted viral load in Figure 2. As expected, increasing the ensemble size improves the performance of the filter. As shown in Figure 2, the use of 100 ensemble members leads to a nearly 35% decrease in the RAE compared to using 25 members. However, increasing the ensemble size beyond 100 only has a marginal effect on the accuracy (roughly 6%) relative to the increment in the runtime cost. Therefore, we run the rest of our simulations with 100 members.

Figure 2. RAE of the viral load for Patient 1 for different ensemble sizes.

4.3. Assimilation Results

The time series of the viral load from the data, free-run (without data assimilation) and joint-EnKF run for all patients are shown in Figure 3. Compared to the model free-run, the joint-EnKF suggests significant improvements and yields more accurate viral load estimates. Although a few data points are not particularly well estimated by the joint-EnKF, the overall behavior of the viral load is well reproduced by the filter. To allow a quantifiable validation, Table 2 outlines the bias values comparing the measured and predicted viral load of the model free-run and the joint-EnKF for the six patients. The joint-EnKF results are also presented in terms of percentage improvement with respect to the model free-run prediction. The significantly improved filter estimates demonstrate the relevance of the data assimilation of the viral load information into the immune response model.

Figure 3. Time series of the viral load for (a) Patient 1; (b) Patient 2; (c) Patient 3; (d) Patient 4; (e) Patient 5; and (f) Patient 6—relative to the model free-run (without assimilation), joint-EnKF, and the real data [44].

Table 2. Bias of the predicted viral load using the model free-run and the joint-EnKF and improvement (%) with respect to the model.

Patient	Model	Joint-EnKF	Improvement %
1	0.1496	0.0792	47.06
2	0.1646	0.1110	32.56
3	0.3566	0.1297	63.63
4	0.3131	0.2498	20.22
5	0.1853	0.1078	41.82
6	0.3255	0.1953	40.00

Closer examination of the results suggests that the viral load measurements for Patient 1 WERE very well predicted by the joint-EnKF. In particular, the viral load peak and the duration of high viral load WERE well captured. The joint-EnKF provides lower RAE compared to the model free-run, with an improvement reaching up to 50%. The model underestimates the peak level and does not accurately predict the rest of the hydrograph. For patients 2, 4 and 5, the viral load was satisfactory predicted by both the joint-EnKF and the model free-run, although the later fails to reproduce the peak level at the end of the Patient 5 hydrograph. The joint-EnKF further improves the prediction of the viral load for those patients by approximately 33%, 20% and 42%, respectively. Both the joint-EnKF and the model free-run overestimate the peak level for Patient 3. While the joint-EnKF accurately predicts the rest of the hydrograph, the model free-run still overestimates the duration of the high viral load. The joint-EnKF is able to improve the viral load prediction by almost 63%. The same conclusion can be drawn for Patient 6, for whom the duration of the high viral load was overpredicted by the model free-run while the rest of the hydrograph is underestimated.

Overall, the performance of the model with and without data assimilation can be summarized as follows: (1) the peak level is well predicted almost everywhere by the filter, while it is occasionally overestimated or underestimated by the model free-run; (2) the model free-run solution is smooth compared to the joint- EnKF, which delineates the viral load dynamic much more accurately; and (3) the joint-EnKF is able to enhance the viral load estimation for all patients, improving the model solution by up to 40% on average.

Figure 4 displays the time evolution of selected model parameters as estimated by the joint-EnKF for Patient 1. As seen, all parameters smoothly converge to a biologically acceptable range of values. The initial ensemble shrinks at each iteration showing a reduction of 95% in the standard deviation at the end of the assimilation process. A similar assessment was made for the other parameters. Table 3 presents a summary of the selected parameters estimated by the joint-EnKF for all patients. The time delay for antibody production τ of patients 1, 2 and 5 was approximately 7 days and significantly lower than that for patients 3, 4 and 6 with values of 9.52, 11.96 and 10.17 days, respectively. These results are quite consistent with the viral load hydrograph in Figure 3. As we can see, the viral load of patients 1, 2 and 5 start to decrease after almost 7 days. However, the decrease in the viral load for the other patients takes more time.

Table 3. Selected parameters and their converged values estimated by the joint-EnKF.

Patient	Π	β	p_2	k_1	r	τ
1	3824.63	2.60×10^{-7}	0.2889	395.17	0.5318	6.83
2	3948.25	3.26×10^{-7}	0.2943	376.06	0.5291	7.22
3	4036.26	3.11×10^{-7}	0.2746	383.67	0.5157	9.52
4	3941.78	1.43×10^{-7}	0.2840	369.57	0.5180	11.96
5	3916.61	2.07×10^{-7}	0.2789	383.86	0.5190	7.43
6	3990.69	1.54×10^{-7}	0.2835	372.31	0.5127	10.17

Figure 4. Time evolution of selected parameters estimated by the joint-EnKF for Patient 1.

5. Conclusions

This paper formulated a nine-compartment mathematical model for simulating SARS-CoV-2 dynamics and the immune system response in infected patients. The model consists of healthy cells, infected cells, viral load, cytokines, natural killer cells, cytotoxic T-lymphocytes, B-lymphocytes, plasma cells and antibody levels. We computed the basic reproduction number and discussed the equilibrium states of the model. In order to reduce uncertainties in the model and improve its prediction skill, a joint state–parameter estimation scheme based on the EnKF was implemented. The joint-EnKF uses current viral load observations and applies a Kalman-based update to the dynamic state variables and parameters. Finally, we demonstrated the efficiency of the proposed methodology with real SARS-CoV-2 viral load data for six patients, while comparing its performance with the model free-run (i.e., without assimilation).

Sensitivity experiments using different ensemble sizes with the EnKF suggested that 100 members are sufficient to obtain reliable estimates with reasonable computational

cost. During assimilation, the parameters' ensembles were constrained by the data where the spread was reduced and eventually converged to a biologically acceptable range of values. In addition, the proposed joint-EnKF was able to reproduce the overall evolution of the dynamic process of the viral load evolution, the peak values, and the duration of the high viral load with significantly better accuracy than the model free-run. We also performed a quantitative analysis using the RAE comparing the measured and predicted viral load, and reported the percentage of improvements achieved by the joint-EnKF. The joint-EnKF consistently achieved better results for all patients, leading to an approximately 40% improvement over the model free-run. This study demonstrated the relevance of the proposed assimilation system for enhancing the estimates of the immune response mathematical model.

Author Contributions: Conceptualization, R.G.; methodology, R.G. and M.G.; software, R.G. and M.G.; validation, R.G. and M.G.; formal analysis, R.G. and S.H.; writing—original draft preparation, R.G. and S.H.; writing—review and editing, I.H.; supervision, I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.; et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [\[CrossRef\]](#)
2. Sohrabi, C.; Alsafi, Z.; O'Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int. J. Surg.* **2020**, *76*, 71–76. [\[CrossRef\]](#)
3. Brauer, F.; Castillo-Chavez, C.; Feng, Z. *Mathematical Models in Epidemiology*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 32.
4. Choi, W.; Shim, E. Optimal strategies for vaccination and social distancing in a game-theoretic epidemiologic model. *J. Theor. Biol.* **2020**, *505*, 110422. [\[CrossRef\]](#) [\[PubMed\]](#)
5. 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. [\[CrossRef\]](#)
6. Mahajan, A.; Sivadas, N.A.; Solanki, R. An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India. *Chaos Soliton Fractals* **2020**, *140*, 110156. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Ndairou, F.; Area, I.; Nieto, J.J.; Torres, D.F. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Soliton Fractals* **2020**, *135*, 109846. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Nkwayep, C.H.; Bowong, S.; Tewa, J.; Kurths, J. Short-term forecasts of the COVID-19 pandemic: Study case of Cameroon. *Chaos Soliton Fractals* **2020**, *140*, 110106. [\[CrossRef\]](#)
9. Saha, S.; Samanta, G. Modelling the role of optimal social distancing on disease prevalence of COVID-19 epidemic. *Int. J. Dyn. Control* **2020**, *9*, 1053–1077. [\[CrossRef\]](#)
10. Sarkar, K.; Khajanchi, S.; Nieto, J.J. Modeling and forecasting the COVID-19 pandemic in India. *Chaos Soliton Fractals* **2020**, *139*, 110049. [\[CrossRef\]](#)
11. 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**. [\[CrossRef\]](#)
12. 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.
13. Choi, W.; Shim, E. Optimal strategies for social distancing and testing to control COVID-19. *J. Theor. Biol.* **2021**, *512*, 110568. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ghostine, R.; Gharamti, M.; Hassrouny, S.; Hoteit, I. An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble Kalman filter. *Mathematics* **2021**, *9*, 636. [\[CrossRef\]](#)
15. Du, S.Q.; Yuan, W. Mathematical modeling of interaction between innate and adaptive immune responses in COVID-19 and implications for viral pathogenesis. *J. Med. Virol.* **2020**, *92*, 1615–1628. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Ghosh, I. Within host dynamics of SARS-CoV-2 in humans: Modeling immune responses and antiviral treatments. *arXiv* **2020**, arXiv:2006.02936.

17. Hernandez-Vargas, E.A.; Velasco-Hernandez, J.X. In-host mathematical modelling of COVID-19 in humans. *Annu. Rev. Control* **2020**, *50*, 448–456. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Chimal-Eguia, J.C. Mathematical Model of Antiviral Immune Response against the COVID-19 Virus. *Mathematics* **2021**, *9*, 1356. [\[CrossRef\]](#)
19. Edwards, C.A.; Moore, A.M.; Hoteit, I.; Cornuelle, B.D. Regional ocean data assimilation. *Ann. Rev. Mar. Sci.* **2015**, *7*, 21–42. [\[CrossRef\]](#)
20. Hoteit, I.; Luo, X.; Bocquet, M.; Kohl, A.; Ait-El-Fquih, B. Data assimilation in oceanography: Current status and new directions. *New Front. Oper. Oceanogr.* **2018**, *2018*, 465–512.
21. Hoteit, I.; Luo, X.; Pham, D.T. Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. *Mon. Weather Rev.* **2012**, *140*, 528–542. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. 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. [\[CrossRef\]](#)
25. 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. [\[CrossRef\]](#)
26. 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. [\[CrossRef\]](#)
27. 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. [\[CrossRef\]](#)
28. 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. [\[CrossRef\]](#)
29. 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. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. 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. [\[CrossRef\]](#)
32. 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. [\[CrossRef\]](#)
33. 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. [\[CrossRef\]](#)
34. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Mar. Syst.* **1960**, *82*, 35–45. [\[CrossRef\]](#)
35. Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **2003**, *53*, 343–367. [\[CrossRef\]](#)
36. Nikin-Beers, R.; Ciupe, S.M. The role of antibody in enhancing dengue virus infection. *Math. Biosci.* **2015**, *263*, 83–92. [\[CrossRef\]](#)
37. Duffin, R.P.; Tullis, R.H. Mathematical models of the complete course of HIV infection and AIDS. *J. Theor. Biol.* **2002**, *4*, 215–221. [\[CrossRef\]](#)
38. Clapham, H.E.; Tricou, V.; Van Vinh Chau, N.; Simmons, C.P.; Ferguson, N.M. Within-host viral dynamics of dengue serotype 1 infection. *J. R. Soc. Interface* **2014**, *11*, 20140094. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Ben-Shachar, R.; Koelle, K. Minimal within-host dengue models highlight the specific roles of the immune response in primary and secondary dengue infections. *J. R. Soc. Interface* **2015**, *12*, 20140886. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Sasmal, S.K.; Dong, Y.; Takeuchi, Y. Mathematical modeling on T-cell mediated adaptive immunity in primary dengue infections. *J. Theor. Biol.* **2017**, *429*, 229–240. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Gujarati, T.P.; Ambika, G. Virus antibody dynamics in primary and secondary dengue infections. *J. Math. Biol.* **2014**, *69*, 1773–1800. [\[CrossRef\]](#) [\[PubMed\]](#)
42. 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. [\[CrossRef\]](#)
43. Martcheva, M. *An Introduction to Mathematical Epidemiology*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 61.
44. Wölfel, R.; Corman, V.M.; Guggemos, W.; Seilmaier, M.; Zange, S.; Müller, M.A.; Niemeyer, D.; Jones, T.C.; Vollmar, P.; Rothe, C.; et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* **2020**, *581*, 465–469. [\[CrossRef\]](#) [\[PubMed\]](#)