Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:98

https://doi.org/10.28919/cmbn/8802

ISSN: 2052-2541

ENHANCING INFECTIOUS DISEASE MODELLING THROUGH A KALMAN-BASED (HCRD-R) EPIDEMIOLOGICAL APPROACH. APPLICATION TO COVID-19 DATA DURING THE VACCINATION PERIOD

PARTHENA CHRISTODOULIDOU, VASILEIOS E. PAPAGEORGIOU*, GEORGE TSAKLIDIS

Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

Abstract. The COVID-19 pandemic represented one of the most pressing global health crises of the 21st century, imposing substantial challenges upon national healthcare systems. Efforts to model this disease have become paramount in devising informed strategies for its containment. However, prevailing mathematical methodologies predominantly rely on deterministic models, which often oversimplify the complex dynamics of epidemic spread as they cannot quantify the uncertainty that accompanies this physical phenomenon. Notably, there is a dearth of models that specifically address the nuanced dynamics of hospitalized and intensive care unit (ICU) admitted cases, crucial for healthcare system planning. This paper, introduces a novel HCRD-R model, tailored to focus exclusively on hospitalized, ICU admitted, recovered and deceased cases in combination with a Kalman filter for predictive analysis within hospital settings. This focused approach aims to optimize resources while ensuring reliability in modeling and efficiently predicting critical epidemiological scenarios. The reliability of the model is examined on daily recording of COVID-19 in Italy, over a time interval of 106 days which takes place after the onset of the vaccination period. Several comparisons with deterministic and stochastic alternatives are also explored, validating the efficiency of the proposed epidemiological model. Finally, through our analysis we underscore the enhanced efficacy of stochastic epidemiological models over their deterministic counterparts, rendering them suitable for assessing the severity of

*Corresponding author

E-mail address: vpapageor@math.auth.gr

Received August 01, 2024

existing and new epidemic outbreaks.

Keywords: Kalman filters; state-space models; stochastic modelling; differential equations; epidemiology; COVID-19.

2020 AMS Subject Classification: 62L20, 62P10, 60G35.

1. INTRODUCTION

During December 2019, the world witnessed the emergence of the novel coronavirus, SARS-CoV-2, known as COVID-19. Originating in Wuhan, China, the virus quickly traversed international borders, leaving an indelible mark on global health. From January 2020 till 2023, the World Health Organization declared this highly infectious disease a pandemic and since then, the recorded count indicates that over 700 million people have contracted received the infection, with a reported mortality exceeding 7 million [1]. The pervasive nature of COVID-19 is underscored by its capacity for transmission from asymptomatic carriers, posing an insidious challenge to containment efforts. As nations implemented a spectrum of containment and detection strategies, such as vaccines, quarantines and testing, the relentless increase in diagnosed cases persisted. This surge placed a considerable toll on both economic and healthcare systems [2].

The present pandemic has been modeled in several cases, as most contagious diseases, using compartmental models. The most widely recognized are the SIR (Susceptible, Infected, Recovered) [3], SEIR (Susceptible, Exposed, Infected, Recovered) [4], [5], [6] or the SEIRD (Susceptible, Exposed, Infected, Recovered, Deceased) [7], [8] models. However, through the course of the pandemic, more notable fluctuations in disease parameters were introduced. The emergence of variants such as alpha, delta or omicron [9] is linked to increased contagiousness, amplified risk of hospitalization, ICU admission and mortality. However, prevailing studies on COVID-19 modeling are predominantly centered around deterministic epidemiological models, describing the evolution of the pandemic through systems of ordinary differential equations.

These systems lack the capacity to incorporate stochastic behaviors inherent in the COVID-19 pandemic, originating from errors in records and time-varying transmission behavior. Therefore,

a transition from deterministic to stochastic approaches is a practical consideration, aligning with the evolving landscape of disease modeling [10], [11], [12]. The incorporation of the Kalman filter in epidemiological modeling is motivated by its capabilities in handling noisy data, providing a methodological advantage in the context of infectious diseases, where fluctuations and uncertainties are inherent [13], [14], [15]. Other notable analysis which incorporate Kalman filtering techniques for the evaluation of the spread of infectious disease can be found in [16], [17], [18], [19], [20], [21]. These attempts mainly emphasize on the utilization of the extended Kalman filter (EKF) due to the non-linearity which is associated with the infection of susceptible individuals. Moreover, there are several attempts that aim to avoid the statistical assumptions that accompany Kalman filters, employing computational techniques derived from particle filtering [22], [23], [24] [25], [26]. However, none of the aforementioned papers consider the existence of severe infections that may lead to hospitalizations or ICU admissions.

Another class of stochastic methodologies commonly referenced in literature -distinct from the Kalman filters- involves the application of continuous-time Markov chains. They are frequently utilized to model epidemiological systems of moderate complexity, typically containing 2 to 4 states [27], [28], [29], [30], [31], [32], [33], [34], [35]. In contrast to our analysis, these models focus on specific epidemiological features, such as the extinction time of a disease, the final size (total number) of infections or deaths, the optimal revaccination time, the maximum cardinality of infections, etc. Once again, we underline that these schemes and respective stochastic epidemiological descriptors do not consider the existence of hospitalizations, ICU admissions or open population dynamics [2], [36], [37], [38].

Aiming to address the aforementioned limitations, in this paper, we propose a novel HCRD (Hospitalized, ICU admitted, Recovered, Deceased) epidemiological model which is combined with the integration of the Kalman filter methodology with parameter estimation. This model is focused solely on the population within hospitals, which enriches the existing literature, where specialized predictions considering the subpopulation of severely infected patients which

transition inside or outside hospital facilities are highlighted. Additionally, the incorporation of the cardinality of individuals who recovered outside hospitals or ICUs provides a significant enhancement of the generated predictions.

We test the reliability of our model on reported data from Italy, covering a period of 106 days, providing daily estimates for hospitalized, ICU admitted, recovered and deceased cases. In the results section, a comprehensive comparison of the developed models was conducted to assess the utility of Kalman-based models. The key attribute underlying the efficacy of our approach lies in the integration of the Kalman filter, which enables the model to assimilate information from the observed data, thereby enhancing its predictive capacity amidst the intricate dynamics of the phenomenon.

Given the challenges posed by the COVID-19 pandemic, there is a significant need to investigate the complex dynamics of hospitalizations, as they represent a quite informative indicator regarding the severity of the pandemic's infection wave. Employing mathematical models more focused on the effects of the pandemic within healthcare systems is essential in order to gain understanding of its impact. By concentrating on the specific dynamics of hospital admissions, these models would enable precise predictions of the strain on healthcare facilities, guiding proactive resource allocation and capacity planning. This targeted approach facilitates informed decision-making for public health interventions, such as implementing timely measures and optimizing vaccination strategies to mitigate the burden on hospitals.

2. PRELIMINARIES

2.1 Proposed HCRD and HCRD-R models

In this paragraph, we introduce the HCRD compartmental scheme, through which we can examine the impact of coronavirus on Italy's healthcare system based on real-time records. As previously mentioned, our model specifically concentrates on four distinct states/compartments relevant to patients that are admitted to hospitals: hospitalized, ICU-admitted, recovered and

deceased cases. Figure 1 illustrates the transitions between the abovementioned states in the proposed epidemiological model, providing a visual representation of its dynamics.

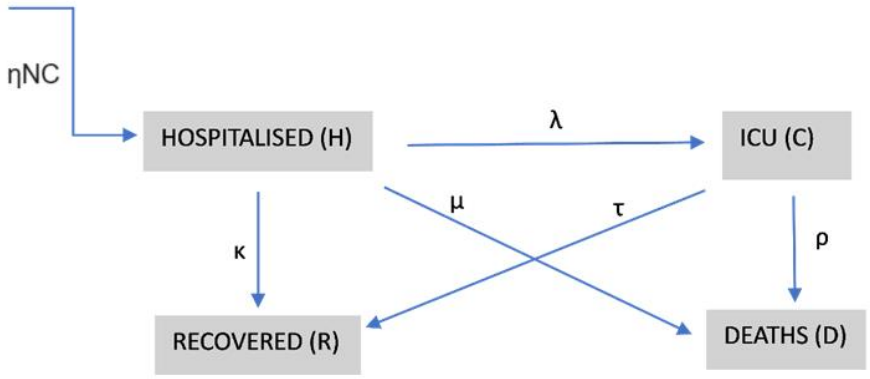

FIGURE 1. Diagram of the HCRD epidemiological model

$$\begin{aligned} \frac{dH}{dt} &= \eta NC(t) - \lambda H(t) - \kappa H(t) - \mu H(t) \\ \frac{dC}{dt} &= \lambda H(t) - \tau C(t) - \rho C(t) \\ \frac{dR}{dt} &= \kappa H(t) + \tau C(t) \\ \frac{dD}{dt} &= \mu H(t) + \rho C(t) \end{aligned} \tag{1}$$

TABLE 1. Parameter and state definitions concerning the introduced HCRD model

Symbol	Definition of Parameter/State
$NC(t)$	New cases
$H(t)$	Hospitalized
$C(t)$	ICU admitted
$R(t)$	Recovered
$D(t)$	Deceased
η	Hospitalization rate
λ	Transition rate from hospital to ICU
κ	Recovery rate of hospitalized cases
τ	Recovery rate of ICU admitted cases
μ	Death rate of hospitalized cases
ρ	Death rate of ICU admitted cases

The HCRD model is characterized by the aforementioned continuous-time deterministic system. In this system, $H(t), C(t), R(t)$ and $D(t)$ represent the cardinality of hospitalized, ICU admitted, recovered and deceased compartments during the examined time interval. The model's parametric space includes the transition rates $\lambda, \kappa, \mu, \tau, \rho$, where λ signifies the admission rate in ICU, κ, τ denote the recovery rates from a hospital or ICU admission and μ, ρ represent the mortality rates of hospitalized and the ICU-admitted patients, respectively. Parameter η signifies the daily hospitalization rate, which can be estimated from daily records. The product of this factor with the number of new cases ($NC(t)$), signify the daily number of new hospital entries. We note that all parameters are non-negative constant rates, thus the system's parametric space is $(\eta, \lambda, \kappa, \tau, \mu, \rho) \in \mathbb{R}_{+}^{6}$. Table 1 summarizes the definitions of the notation employed for the utilized deterministic system of differential equations.

The second model, HCRD-R incorporates an additional transition to account for individuals who have recovered outside the hospital setting, as signified by the second "R" in its abbreviation. Additionally, it refines the previous epidemiological scheme by incorporating a temporal adaptation, selecting hospital entries from the preceding t^* days. In each case, the researcher can choose t^* to align with the characteristics of the data. In our analysis, we select $t^* = 7\ days$. However, it is essential to note that this duration may vary in different epidemics, contingent upon the specific characteristics of the phenomenon. The product of the parameter η with the number of new cases ($NC(t - t^*)$), signify the daily number of new hospital entries. According to the above, we lead to

$$\begin{aligned} \frac{dH}{dt} &= \eta NC(t - t^*) - \lambda H(t) - \kappa H(t) - \mu H(t) \\ \frac{dC}{dt} &= \lambda H(t) - \tau C(t) - \rho C(t) \\ \frac{dR}{dt} &= \kappa H(t) + \tau C(t) + (1 - \eta)\, NC(t - t^*) \\ \frac{dD}{dt} &= \mu H(t) + \rho C(t) \end{aligned} \quad . \tag{2}$$

The attributes of the continuous-time formulation need to be derived from the discrete time

observations of real-time data, making the discretization of the system a necessary step in this analysis. Commencing with the initial system, it can be discretized using finite differences, specifically employing the expression $\frac{du}{dt} = \frac{u(t+\Delta t)-u(t)}{\Delta t}$. This results in the transformation of the first ODE's system into a state-space representation. As a result, the discrete-time system takes the form

$$\begin{aligned} H_{t+1} &= \eta N C_t + (1-\lambda-\kappa-\mu)H_t \\ C_{t+1} &= \lambda H_t + (1-\tau-\rho)C_t \\ R_{t+1} &= R_t + \kappa H_t + \tau C_t \\ D_{t+1} &= D_t + \mu H_t + \rho C_t \end{aligned} \quad . \tag{3}$$

This discretized system is more suitable for interpreting population movements within hospitals during discrete time steps. The states at time t can be assembled in a vector $\boldsymbol{x}_t^T = [H_t, C_t, R_t, D_t]$. Similarly, the second system undergoes the discretization process, following the same rationale. The resulting discrete-time system is given by

$$\begin{aligned} H_{t+1} &= \eta N C_{t-t^*} + (1-\lambda-\kappa-\mu)H_t \\ C_{t+1} &= \lambda H_t + (1-\tau-\rho)C_t \\ R_{t+1} &= R_t + \kappa H_t + \tau C_t + (1-\eta)N C_{t-t^*} \\ D_{t+1} &= D_t + \mu\, H_t + \rho C_t \end{aligned} \tag{4}$$

2.2 Stochastic HCRD and HCRD-R models

In this section, our focus shifts on the transformation of the deterministic models HCRD and HCRD-R into stochastic equivalents which can account for the existing uncertainty that accompanies the description of physical phenomena, such as epidemic outbreaks. This conversion necessitates a meticulous reformulation of the ODEs into a state-space representation. This method is grounded in discretizing the ODE system, opting for a time step $\Delta t = 1$ to align with the daily reporting frequency of pandemic-related observations. We remind that vector $\boldsymbol{x}_t = [H_t\ C_t\ R_t\ D_t]^T$ represents the state vector characterizing the daily cardinality of hospitalized, ICU admitted, recovered and deceased cases at time step t. Therefore, in the initial model the state-space representation is expressed as follows

$$\tilde{\boldsymbol{x}}_t = f(\tilde{\boldsymbol{x}}_{t-1}) + \boldsymbol{w}_t\,, \tag{5}$$

where

$$f(\tilde{\boldsymbol{x}}_{t-1}) = \begin{bmatrix} \eta N C_{t-1} + (1-\lambda-\kappa-\mu)H_{t-1} \\ \lambda H_{t-1} + (1-\tau-\rho)C_{t-1} \\ R_{t-1} + \kappa H_{t-1} + \tau C_{t-1} \\ D_{t-1} + \mu\, H_{t-1} + \rho\, C_{t-1} \end{bmatrix}, \tag{6}$$

while vector $\boldsymbol{w}_t$ represents the introduction of Gaussian noise with covariance matrix $\boldsymbol{Q}$, signifying the transition from the deterministic to the stochastic perspective. The incorporation of white noise encapsulates the uncertainty inherent in the transitions between the model's states that the deterministic approach fails to capture. Sources of uncertainty may derive from perturbations of admission rates, asymptomatic infections and population's heterogeneity. Therefore, it is evident that the transition to a stochastic point of view seems necessary.

For the introduced (HCRD-R) model the state-space representation can be derived in a similar manner as

$$f_R(\tilde{\boldsymbol{x}}_{t-1}) = \begin{bmatrix} \eta N C_{t-t^*-1} + (1-\lambda-\kappa-\mu)H_{t-1} \\ \lambda H_{t-1} + (1-\tau-\rho)C_{t-1} \\ R_{t-1} + \kappa H_{t-1} + \tau C_{t-1} + (1-\eta)N C_{t-t^*-1} \\ D_{t-1} + \mu H_{t-1} + \rho\, C_{t-1} \end{bmatrix}. \tag{7}$$

To leverage the available observations, we introduce the following second stochastic equation, where $\boldsymbol{v_t}$ also denotes white Gaussian noise with zero mean and covariance matrix $\boldsymbol{R}$

$$\boldsymbol{y}_t = h(\tilde{\boldsymbol{x}}_t) + \boldsymbol{v}_t\,, \tag{8}$$

The inclusion of white noise in this equation is essential to account for uncertainties inherent in daily observations, which may originate from errors and delays in records. In our case, both the state-space system and the observation equation involve linear functions. Consequently, $h(.)$ can be expressed through a matrix $\boldsymbol{H}$ that takes the form of a 4-dimensional identity matrix. This reformulates equation (8) as

$$\boldsymbol{y}_t = \tilde{\boldsymbol{x}}_t + \boldsymbol{v}_t\,. \tag{9}$$

The Kalman filter is a well-suited algorithm for systems that can be effectively modeled using linear dynamics. The application of the Kalman filter involves the following sequential steps:

i) Set the initial system state and its corresponding covariance matrix

$$\widehat{\boldsymbol{x}}_0 = E[\boldsymbol{x}_0]\,, \tag{10}$$

$$\widehat{\boldsymbol{P}}_0 = E[(\boldsymbol{x_0} - \widehat{\boldsymbol{x}}_0)(\boldsymbol{x}_0 - \widehat{\boldsymbol{x}}_0)]^T\,. \tag{11}$$

ii) During the prediction step, compute the a priori state estimate

$$\widehat{\boldsymbol{x}}_{t|t-1} = f(\widehat{\boldsymbol{x}}_{t-1|t-1})\,, \tag{12}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{F}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}_t^T + \boldsymbol{Q}\,, \tag{13}$$

where $\boldsymbol{F}_t = \frac{df}{dt}|_{\widehat{\boldsymbol{x}}_{t-1|t-1}}$.

iii) Calculate the Kalman Gain

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^T\big(\boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{R}\big)^{-1}\,. \tag{14}$$

iv) During the update phase, compute the a posteriori estimate considering daily records

$$\widehat{\boldsymbol{x}}_{t|t} = \widehat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t\big(\widehat{\boldsymbol{y}}_t - \boldsymbol{H}_t\widehat{\boldsymbol{x}}_{t|t-1}\big), \tag{15}$$

$$\boldsymbol{P}_{t|t} = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H}_t)\boldsymbol{P}_{t|t-1}\,. \tag{16}$$

v) Repeat the same process for $t = 1, \dots, T$, where T is the time series' length.

2.3 Datasets

In this section, we introduce the two datasets comprising daily observations that we utilized to implement the HCRD and HCRD-R modeling schemes to offer estimations of the progression of the pandemic in hospitals. First, we utilize the dataset sourced from [39] an open-access database provided by the University of Oxford. This dataset offers daily observations of deceased individuals, hospitalized cases, total cases and ICU patients, which are subsequently extracted for analysis.

Moreover, we access data on the daily count of currently positive cases from the database available on the website [40]. This webpage is updated daily and provides official COVID-19 data for all European Union member states, sourced from national resources. By subtracting the daily counts of currently positive and deceased cases from the daily cases, we calculate the total number of recoveries. It is important to emphasize that we have verified the compatibility of these two

datasets. Both datasets consistently reflect the same numbers for deceased, hospitalizes and ICU patients. Additionally, the column representing total cases in the first dataset aligns perfectly with the column for total positive cases in the second dataset, further validating the consistency between datasets.

3. Main Results

3.1 Simulations of the deterministic HCRD and HCRD-R models

For this analysis, we assume that the model's parameters remain constant. To estimate them we conduct numerous simulations, applying the least squares method with the root mean square error (RMSE) as the criterion for parameter selection. Throughout this process, we iteratively adjust the model's parameters to minimize the discrepancy between the observed data and the model's predictions. This allows the determination of parameters that provide the best fitting performance with the real-time data of the pandemic in Italy. This simulation spans 106 days – from June 8, 2022, to September 21, 2022 – during which we observe a single wave of hospitalizations peaking after 49 days.

By assuming fixed parameters, we disregard factors such as the emergence of COVID-19 variants, which could potentially influence hospital or ICU admission rates, as well as death and recovery rates. It is important to highlight that the COVID-19 pandemic data from Italy serves solely for verification and validation purposes in this study. However, the proposed method is designed to be generic and applicable across various contexts, therefore it can also be utilized to model the spread of COVID-19 pandemic in other countries or regions. The optimal model parameters for the deterministic HCRD and HCRD-R models are presented in Table 2.

TABLE 2. The values of the deterministic models' parameters

	η	λ	κ	τ	μ	ρ
Deterministic HCRD model	0.048	0.02	0.48	0.44	0	0.01
Deterministic HCRD-R model	0.141	0.02	0.7	0.62	0.01	0.01

It is notable that in the deterministic HCRD model, the optimal parameter value for the transition from the hospitalized to the deceased state is identified as 0. This indicates that, within the framework of this model, the least RMSE value is achieved under the assumption that there are no direct transitions from the hospitalized to the deceased state. This observation may derive from the fact that during the examined period, the ratio of deaths generated directly from the "Hospitalized" compartment could be considered negligible since nursing stuff and doctors have faced several COVID-19 cases and gained valuable experience against severe infections. However, the absolute absence of deceased cases originating from this compartment seems infeasible. As a result, for the improved epidemiological form, namely the HCRD-R scheme, we encounter a correction of this estimate, obtaining $\mu = 0.01$. For both models, we encounter lower mortality rates (μ, ρ) compared to the respective recovery ones (τ, μ). This phenomenon is in accordance with the literature and daily records, as the percentage of patients which have recovered from COVID-19 has increased dramatically during the vaccination period.

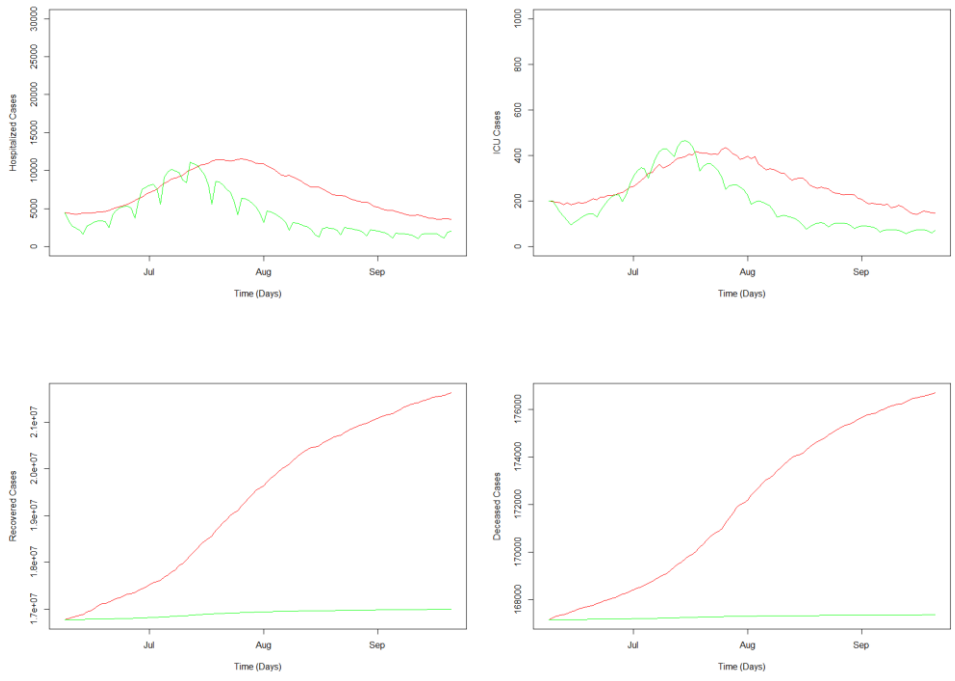

FIGURE 2. Comparison of the predictions generated from the deterministic HCRD model with daily observations (red line: observed data, green line: deterministic model's predicitons)

We proceed with the exploration of the performance of the HCRD model. This model serves as the foundation of our epidemic modeling framework, capturing the dynamics of hospitalizations, ICU admissions, recoveries and deaths. Figure 2 illustrates the comparison between the model predictions and the observed data, providing insights into the performance of the model in capturing the dynamics of the COVID-19 pandemic within hospital settings.

The diagrams indicate that the estimates for hospitalizations and ICU admissions closely align with the observed data. However, there appears to be less accuracy in estimating the recovered and deceased states. The model appears to have captured the trends in the hospitalized and ICU states reasonably well, but adopts a more linear trend for the recovered and deceased states. This likely stems from the relatively large population size in the recovered state compared to the hospital and ICU states, indicating that the recovered state is less susceptible to their influence. Additionally, the minimal changes observed in the deceased state can be attributed to the small values of parameters μ and ρ which reduce its sensitivity to changes.

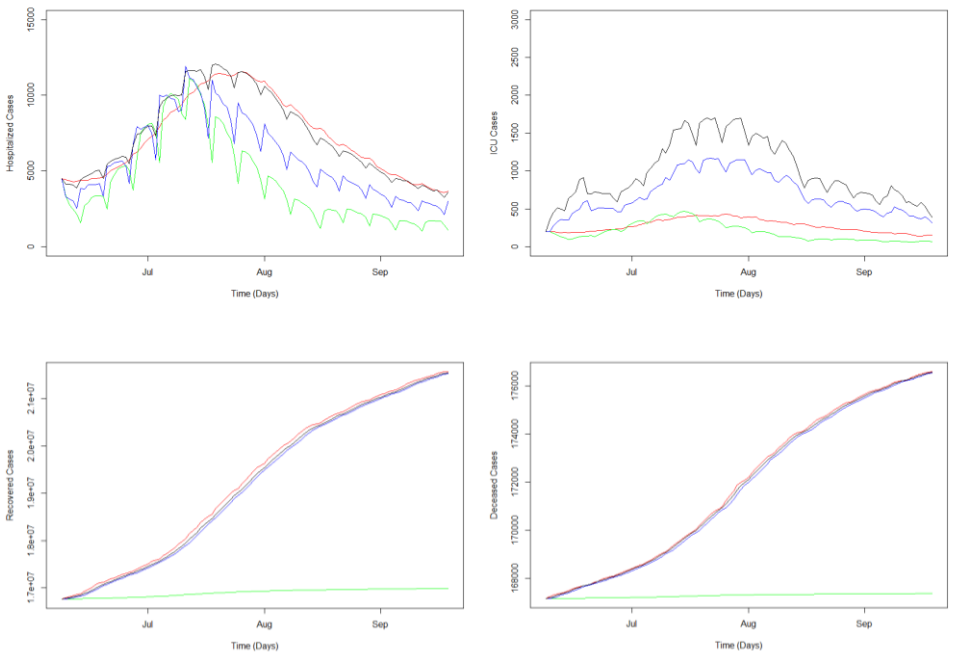

FIGURE 3. Comparison of the daily observations with the predictions of the deterministic model, the a priori and a posteriori estimates (red line: observed data, green line: deterministic model's predictions, blue: a priori estimate of the stochastic model, black: a posteriori estimate of the stochastic model)

Now, we introduce the Kalman filter into our modeling framework, incorporating white noise $\boldsymbol{w}_t \sim N(\mathbf{0}, \boldsymbol{Q})$ in the state equation and $\boldsymbol{v}_t \sim N(\mathbf{0}, \boldsymbol{R})$ in the observation equation, where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are diagonal matrices, the elements of which are estimated in the same manner as the model's parameters. Figure 3 provides a visual comparison between the observed data, the predictions of the deterministic model, the a priori and a posteriori estimates produced by the Kalman filtering procedure.

From Figure 3, it is evident that the predictive ability of the model significantly improves in the hospitalized, recovered and deceased states when utilizing the Kalman filter. The a priori estimates for these states converge more to the observed data than those from the deterministic model, suggesting that incorporating noise into the deterministic models, leading to the construction of a stochastic epidemiological equivalents, is highly beneficial. Furthermore, the a posteriori estimation, which integrates information from the observed data, further refines our predictions, demonstrating even greater accuracy. However, the ICU state presents an exception where the deterministic model appears to provide a better fit than stochastic approaches. In this case, the a posteriori estimation does not improve upon the a priori estimation, indicating that for the ICU state, the deterministic model is the most effective. This observation is further supported by the root mean square errors (RMSE) of the predictions compared to the observations, as depicted in Table 3.

TABLE 3. Root mean square errors of model's predictions compared to daily observations

HCRD model \ RMSE	H	C	R	D
Deterministic	3702	112	2883747	5651
A priori est.	2100	454	122649	166
A posteriori est.	598	778	76484	77

Moving on to the second model of the analysis, the HCRD-R model, there are notable differences compared to the HCRD model. In the HCRD-R model, we select new cases, which is used for the calculation of the new hospital entries, from t^* days ago as input, where in our case

$t^* = 7\ days$. Additionally, we modify the equation for the recovered state to include cases that have recovered outside the hospital. Upon examining the predictions from the first model, we observe fluctuations in the hospitalized and ICU admitted states that likely arise from the time series of new cases. To address this, we apply a moving average to smooth the pattern of daily observations. After experimenting with different orders, we find that an order of 8 provides the optimal balance between smoothing and retaining the information of the time series. Figure 4 illustrates the smoothing of the time series of new cases.

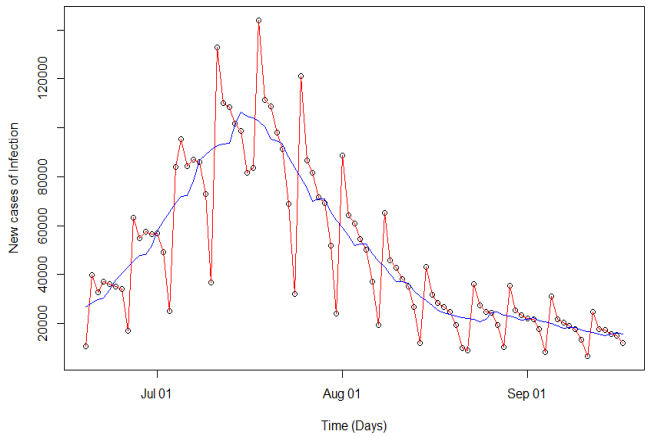

FIGURE 4. Smoothing of new cases based on the MA (8) filter (red line: original time series, blue: smoothed time series)

With these enhancements, we developed the new and improved HCRD-R model to predict the population dynamics across the different states. Figure 5 illustrates the comparison between the HCRD-R model predictions and the observed data. In the four diagrams, the improvement in fitting is evident, particularly noticeable in the hospitalized and ICU admitted states, where the estimated lines exhibit smoother trajectories compared to the original HCRD model. Moreover, we observe significant improvements in the recovered and deceased states compared to the HCRD deterministic model. These improvements are reflected in trajectories that closely follow the trend of the observed data, contrasting with the more linear predicted lines of the original model. This improvement is also apparent when examining the RMSEs of the two models, shown in Table 4.

TABLE 4. RMSEs of the two deterministic models

Model	H	C	R	D
Deterministic HCRD	3702	112	2883747	5651
Deterministic HCRD-R	3833	98	291207	835

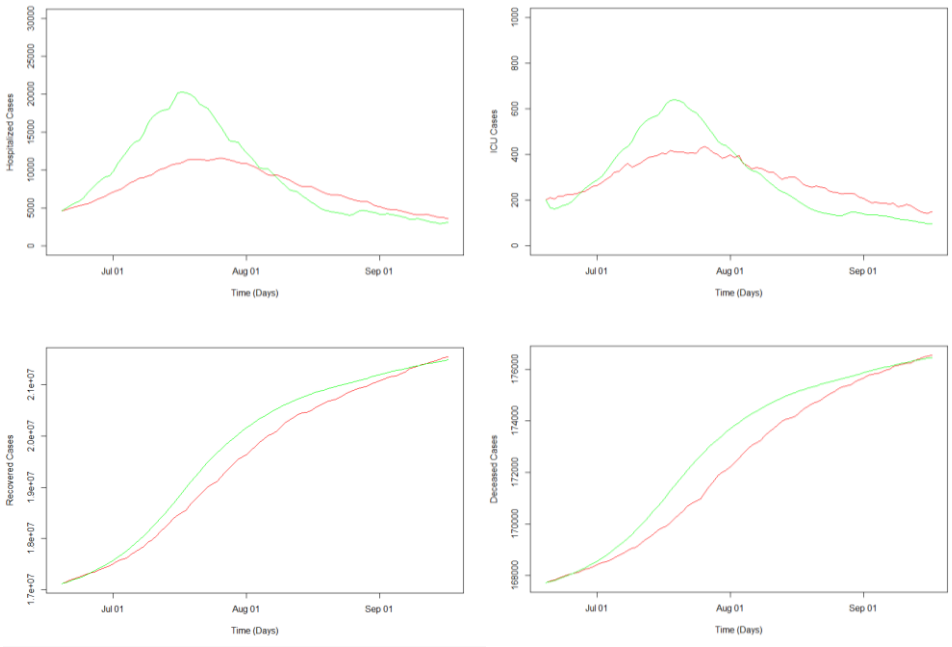

FIGURE 5. Visual comparison of the deterministic HCRD-R model with the daily observations (red line: observed data, green line: Deterministic model's predictions)

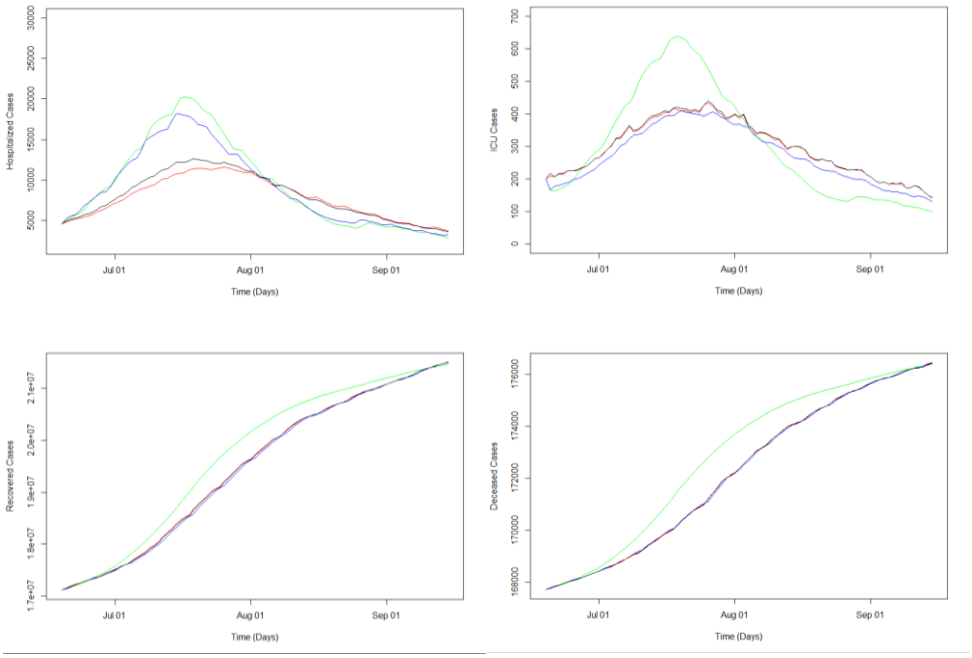

FIGURE 6. Comparison of the observations with the deterministic model, the a priori and the a posteriori estimate (red line: observed data, green line: deterministic model's predictions, blue: a priori estimate of the stochastic model, black: a posteriori estimate of the stochastic model)

Following the same methodological approach, we proceed to apply the Kalman filter to the HCRD-R deterministic model. The integration of the Kalman filter significantly enhanced the predictive capabilities of the previous model. Given this improvement, we anticipate a similar enhancement in the accuracy and reliability of the predictions for the HCRD-R model. Figure 6 provides a visual representation of the outcomes obtained from the HCRD-R with Kalman filtering across all states.

TABLE 5. RMSEs of the HCRD-R model

HCRD-R model	H	C	R	D
Deterministic	3833	98	291207	835
A priori est.	3070	24	38369	42
A posteriori est.	575	2.9	9411	10.8

TABLE 6. RMSEs of all models for the four compartments of the examined epidemiological scheme

Model	H	C	R	D
Deterministic HCRD	3702	112	2883747	5651
Stochastic HCRD	598	778	76484	77
Deterministic HCRD-R	3833	98	291207	835
Stochastic HCRD-R	**3070**	**24**	**38369**	**42**

The diagrams indicate that the Kalman filter significantly enhanced the predictions of the HCRD-R model, particularly for the hospitalized and ICU admitted states. Furthermore, the model closely aligns with the observations of the recovered and deceased states. As observed previously, the inclusion of noise in the deterministic model has led to improvements in prediction accuracy, particularly noticeable in the ICU, recovered and deceased states. Furthermore, the integration of the observed data in the a posteriori estimation has further enhanced the accuracy of estimations across all states. Table 5 presents the RMSEs of the deterministic HCRD-R and its Kalman filter application, while Table 6 adds the RMSEs of the HCRD model and its stochastic counterpart, facilitating a comprehensive comparison of all models.

4. DISCUSSION

Epidemics, particularly exemplified by the rapid global spread of the novel SARS-CoV-2 present significant challenges for national healthcare systems. The surge in cases strains hospital and ICU capacities, leading to a decline in the quality of patient care [41]. Mathematical modeling of infectious diseases typically adopts deterministic approaches [42], often extending standard models like the SIR model to include additional differential equations for states such as deceased, hospitalized and ICU admitted cases [43], [44], [45]. However, the presence of multiple states complicates the system, posing challenges for investigating the impact of restrictive measures,

such as lockdowns and masks, on healthcare resource management.

In this paper, we introduce a novel HCRD model and its enhanced version, HCRD-R, with parameter estimation through the method of least squares. Our aim is to address limitations observed in existing models by offering a specialized framework tailored to the assessment of hospital and ICU admissions and discharges. The introduced scheme is of significant importance since it facilitates the precise evaluation of conditions inside hospital facilities. Efficient prediction of patient admissions will contribute to the effective distribution of resources, avoiding situations like those encountered in Italian hospitals during the initial stages of the COVID pandemic, which led to excess mortality.

The HCRD-R model incorporates several improvements over the HCRD one. Firstly, we modify the input of new cases in the system to be chosen t^* days prior to the day of admission. Additionally, we implement a smoothing technique for the time series of new cases to enhance data accuracy. Furthermore, we extend the "Recovered" state to include individuals who have recovered outside the hospital. These models consider daily COVID-19 cases as input multiplied by an estimated parameter (η) to represent the rate at which new cases are admitted to hospital. It is evident that thorough COVID-19 testing was conducted to those admitted to hospitals and ICUs. Consequently, the daily observations of these two states serve as reliable data for validating the fitting performance of the proposed model.

The Kalman filter methodology leverages both daily observations and the model equations, integrating them while introducing white noise to adopt a stochastic approach. This enhancement significantly improves the model's fitting ability. The presented results, which involve fitting the spread of the disease in Italy over approximately 4 months during the vaccination period, affirm the reliability of our model.

We contend that employing this stochastic approach is essential. Despite the reliability of the data from hospitals and ICUs, there remains room for errors. In our model, the existence of errors associated with the daily records is justified by the addition of the white noise. Additionally,

incorporating all possible transitions to the epidemiological scheme aiming to better explain the phenomenon's dynamics would necessitate a more computationally expensive model, rendering the estimation process unprofitable. As a result, we have included only the necessary transition rates leading to 6 parameters $(\eta, \lambda, \kappa, \mu, \tau, \rho)$ that have to be estimated before the application of the Kalman methodology. According to the above, we believe that the adoption of stochastic modeling approaches is more suitable.

In epidemic modeling, deterministic models are frequently favored for their simplicity compared to the stochastic models. Despite the widespread use, their limitations are often acknowledged but overlooked [46]. While deterministic models offer the advantage of being amenable to analysis and occasionally yield explicit solutions, their main drawback lies in their oversimplification, rendering conclusions less applicable to real-world epidemics. Introducing additional complexity to these models enhances realism but also increases analytical challenges and introduces additional uncertainties through the incorporation of more parameters. Important features to enhance the realism of epidemic models is the inclusion of individual heterogeneity, or the introduction of randomness in certain features such as actual transmissions, social structures and latent periods. Stochastic models enable individuals to exhibit diverse behaviors specified by random distributions, such as the white noise employed in this analysis [47], [48], [49]

As evident in the results section, the HCRD-R model outperforms the HCRD model in fitting the observed data. This fact is observed both visually, through the figures, and quantitatively, via the RMSEs for each state. Additionally, the Kalman filter generally provides a closer approximation of reported data compared to their deterministic counterparts, except for the ICU estimation in the first model (Table 6). Notably, among the models evaluated, the stochastic HCRD-R model exhibits the closest alignment with the reported data.

In analyzing the RMSE values across different states and models, notable variations were observed. In the deterministic HCRD model, the RMSE increased by 20.58% in the hospitalized, 366.7% in the ICU admitted, 7415.8% in the recovered and 10786.5% in the deceased state.

Similarly, the deterministic HCRD-R model exhibited increase in RMSE by 24.85% in the hospitalized, 308% in the ICU admitted, 659% in the recovered and 1888% in the deceased state. In contrast, the stochastic HCRD model displayed a decrease in RMSE by 80.5% in the hospitalized state, but experienced increases by 3141.7% in the ICU admitted, 99.3% in the recovered and 83.3% in the deceased state. Consequently, the stochastic HCRD-R model demonstrates notably higher accuracy compared to the other models in depicting population dynamics within hospital settings.

Control measures like vaccination has shown notable efficiency against severe infections and hospitalizations. For instance, the authors in [50] reported a 63% decrease in anticipated total deaths due to vaccination, while in [51] they indicated reductions of 63.5% and 65.6% in non-ICU hospitalizations and ICU admissions, respectively, over a 300-day period in the United States. Additionally, researchers in [52] found that vaccination was 72% effective in preventing deaths, with a slightly lesser effect against the B.1.1.7 variant compared to non-B.1.1.7 strains (70% and 78% reduction, respectively). Finally, according to [53], mortality rates in areas with varying levels of vaccination coverage (low, medium, and high) decreased by 60%, 75%, and 81%, respectively, during the first half of 2021.

The model's effectiveness is constrained by several factors that warrant consideration. Firstly, parameter uncertainty [54] introduces challenges in accurately estimating key variables such as transition rates and intervention effects, which can impact the reliability of model's predictions. Additionally, the model's effectiveness is tempered by data limitations, too [55]. These limitations result in gaps in comprehending essential aspects of population dynamics, along with data quality issues, such as reporting errors that introduce uncertainties to the outputs impending its ability to provide accurate and reliable predictions. Finally, capturing the evolving nature of epidemics over time presents a complex challenge [56], as the dynamics of disease spread and the appearance of variants and interventions may fluctuate unpredictably. These limitations underscore the need for careful interpretation of results and ongoing refinement of the model to enhance its robustness and

applicability in diverse epidemiological contexts.

5. CONCLUSIONS

This paper presents a novel epidemiological Kalman filter model tailored specifically to track hospitalized, ICU admitted, recovered and deceased cases, demonstrating its efficacy over a vaccination period of 106 days in Italy. Notably, this model could also be used for analyzing future pandemics and build effective strategies in better healthcare resource planning. By leveraging the Kalman filter, the model integrates comprehensive information from daily records, capturing the evolving dynamics of the pandemic and its strains on healthcare systems. Successfully navigating these complex dynamics, the proposed model surpasses its deterministic counterparts, yielding notably smaller RMSE values. Consequently, we advocate for increased utilization of stochastic methods in the future. These techniques can assist public health authorities not only in gauging the severity of a novel epidemic outbreak but also in assessing the effectiveness of intervention strategies. This enables the pursuit of an optimal equilibrium between health benefits and socio-economic implications.

In future research endeavors, extending the application of our model to other regions or different time periods prior to the vaccination campaign would be valuable. Additionally, exploring longer time intervals to assess changes in hospital and ICU admissions before and after vaccination could provide insights into the impact of vaccination campaigns on disease dynamics [57]. Furthermore, incorporating vaccination and other intervention strategies as control inputs [58] in parallel with information concerning chronic diseases [59] would enable the evaluation and comparison of optimal strategies for reducing patient admissions. Moreover, employing likelihood-based parameter estimation methods [15], [60] could further enhance the accuracy of our model's predictive capabilities and overall reliability. Lastly, expanding the model to include additional states could also enhance its predictive capabilities, enabling it to forecast outcomes across a broader range of scenarios.

FUNDING

This research has been funded by the Special Account for Research of Aristotle University of Thessaloniki (ELKE AUTH).

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] World Health Organisation. Accessed: Apr. 01, 2024. https://www.who.int.

[2] Z. Xu, H. Zhang, Z. Huang, A continuous Markov-chain model for the simulation of COVID-19 epidemic dynamics, Biology. 11 (2022), 190. https://doi.org/10.3390/biology11020190.

[3] I. Cooper, A. Mondal, C.G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, Chaos Solitons Fractals. 139 (2020), 110057. https://doi.org/10.1016/j.chaos.2020.110057.

[4] O.N. Bjørnstad, K. Shea, M. Krzywinski, et al. The SEIRS model for infectious disease dynamics, Nat. Methods 17 (2020), 557–558. https://doi.org/10.1038/s41592-020-0856-2.

[5] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, Nonlinear Dyn. 101 (2020), 1667–1680. https://doi.org/10.1007/s11071-020-05743-y.

[6] P. Girardi, C. Gaetan, An SEIR model with time-varying coefficients for analyzing the SARS-CoV-2 epidemic, Risk Anal. 43 (2021), 144–155. https://doi.org/10.1111/risa.13858.

[7] K. Rajagopal, N. Hasanzadeh, F. Parastesh, et al. A fractional-order model for the novel coronavirus (COVID-19) outbreak, Nonlinear Dyn. 101 (2020), 711–718. https://doi.org/10.1007/s11071-020-05757-6.

[8] I. Korolev, Identification and estimation of the SEIRD epidemic model for COVID-19, J. Econometrics. 220 (2021), 63–85. https://doi.org/10.1016/j.jeconom.2020.07.038.

[9] V.E. Papageorgiou, G. Tsaklidis, An improved epidemiological-unscented Kalman filter (hybrid SEIHCRDV-UKF) model for the prediction of COVID-19. Application on real-time data, Chaos Solitons Fractals. 166 (2023), 112914. https://doi.org/10.1016/j.chaos.2022.112914.

[10] X. Zhu, B. Gao, Y. Zhong, et al. Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling, Computers Biol. Med. 137 (2021), 104810. https://doi.org/10.1016/j.compbiomed.2021.104810.

[11] V.E. Papageorgiou, G. Tsaklidis, A stochastic particle extended SEIRS model with repeated vaccination:

Application to real data of COVID-19 in Italy, Math. Methods Appl. Sci. 47 (2024) 6504–6538. https://doi.org/10.1002/mma.9934.

[12] V.E. Papageorgiou, Novel stochastic descriptors of a Markovian SIRD model for the assessment of the severity behind epidemic outbreaks, J. Franklin Inst. 361 (2024), 107022. https://doi.org/10.1016/j.jfranklin.2024.107022.

[13] M.S. Islam, M.E. Hoque, M.R. Amin, Integration of Kalman filter in the epidemiological model: A robust approach to predict COVID-19 outbreak in Bangladesh, Int. J. Mod. Phys. C. 32 (2021), 2150108. https://doi.org/10.1142/s0129183121501084.

[14] B. Cazelles, N.P. Chau, Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic, Math. Biosci. 140 (1997), 131–154. https://doi.org/10.1016/s0025-5564(96)00155-1.

[15] J. Song, H. Xie, B. Gao, Y. Zhong, C. Gu, K.-S. Choi, Maximum likelihood-based extended Kalman filter for COVID-19 prediction, Chaos Solitons Fractals. 146 (2021), 110922. https://doi.org/10.1016/j.chaos.2021.110922.

[16] X. Zhu, B. Gao, Y. Zhong, et al. Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling, Computers Biol. Med. 137 (2021), 104810. https://doi.org/10.1016/j.compbiomed.2021.104810.

[17] R. Ghostine, M. Gharamti, S. Hassrouny, et al. An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble Kalman filter, Mathematics. 9 (2021), 636. https://doi.org/10.3390/math9060636.

[18] D. Ndanguza, I.S. Mbalawata, J.P. Nsabimana, Analysis of SDEs applied to SEIR epidemic models by extended Kalman filter method, Appl. Math. 07 (2016), 2195–2211. https://doi.org/10.4236/am.2016.717175.

[19] A. Sebbagh, S. Kechida, EKF-SIRD model algorithm for predicting the coronavirus (COVID-19) spreading dynamics, Sci. Rep. 12 (2022), 13415. https://doi.org/10.1038/s41598-022-16496-6.

[20] R. Lal, W. Huang, Z. Li, An application of the ensemble Kalman filter in epidemiological modelling, PLoS ONE. 16 (2021), e0256227. https://doi.org/10.1371/journal.pone.0256227.

[21] K.K. Singh, S. Kumar, P. Dixit, et al. Kalman filter based short term prediction model for COVID-19 spread, Appl. Intell. 51 (2020), 2714–2726. https://doi.org/10.1007/s10489-020-01948-1.

[22] V.E. Papageorgiou, P. Kolias, A novel epidemiologically informed particle filter for assessing epidemic phenomena. Application to the monkeypox outbreak of 2022, Inverse Probl. 40 (2024). 035006. https://doi.org/10.1088/1361-6420/ad1e2f.

[23] A. Endo, E. van Leeuwen, M. Baguelin, Introduction to particle Markov-chain Monte Carlo for disease

dynamics modellers, Epidemics. 29 (2019), 100363. https://doi.org/10.1016/j.epidem.2019.100363.

[24] D. Calvetti, A. Hoover, J. Rose, et al. Bayesian dynamical estimation of the parameters of an SE(A)IR COVID-19 spread model, preprint, (2020). http://arxiv.org/abs/2005.04365.

[25] X. Yang, S. Wang, Y. Xing, et al. Bayesian data assimilation for estimating instantaneous reproduction numbers during epidemics: Applications to COVID-19, PLoS Comput. Biol. 18 (2022), e1009807. https://doi.org/10.1371/journal.pcbi.1009807.

[26] W. Yang, A. Karspeck, J. Shaman, Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics, PLoS Comput. Biol. 10 (2014), e1003583. https://doi.org/10.1371/journal.pcbi.1003583.

[27] A. Marfak, D. Achak, A. Azizi, et al. The hidden Markov chain modelling of the COVID-19 spreading using Moroccan dataset, Data Brief. 32 (2020), 106067. https://doi.org/10.1016/j.dib.2020.106067.

[28] X. Ding, Y. Ma, J. Tang, Markov Chain models in Covid-19 prediction: State-of-the-art and future perspectives, Highlights Sci. Eng. Technol. 49 (2023), 527–535. https://doi.org/10.54097/hset.v49i.8606.

[29] J.R. Artalejo, A. Economou, M.J. Lopez-Herrero, Stochastic epidemic models with random environment: quasi-stationarity, extinction and final size, J. Math. Biol. 67 (2012), 799–831. https://doi.org/10.1007/s00285-012-0570-5.

[30] A. Economou, A. Gómez-Corral, M. López-García, A stochastic SIS epidemic model with heterogeneous contacts, Physica A: Stat. Mech. Appl. 421 (2015), 78–97. https://doi.org/10.1016/j.physa.2014.10.054.

[31] J.R. Artalejo, A. Economou, M.J. Lopez-Herrero, The stochastic SEIR model before extinction: Computational approaches, Appl. Math. Comput. 265 (2015), 1026–1043. https://doi.org/10.1016/j.amc.2015.05.141.

[32] V. E. Papageorgiou and G. Tsaklidis, A stochastic SIRD model with imperfect immunity for the evaluation of epidemics, Appl Math Model 124 (2023), 768–790. https://doi.org/10.1016/j.apm.2023.08.011.

[33] M. Gamboa, M. Lopez-Herrero, Measures to assess a warning vaccination level in a stochastic SIV model with imperfect vaccine, Stud. Appl. Math. 148 (2022), 1411–1438. https://doi.org/10.1111/sapm.12479.

[34] R. Fernández-Peralta, A. Gómez-Corral, A structured Markov chain model to investigate the effects of pre-exposure vaccines in tuberculosis control, J. Theor. Biol. 509 (2021), 110490. https://doi.org/10.1016/j.jtbi.2020.110490.

[35] A. Gómez-Corral, M. López-García, M.T. Rodríguez-Bernal, On time-discretized versions of the stochastic SIS epidemic model: a comparative analysis, J. Math. Biol. 82 (2021), 46. https://doi.org/10.1007/s00285-021-01598-y.

[36] C. Twumasi, L. Asiedu, E.N.N. Nortey, Markov chain modeling of HIV, tuberculosis, and hepatitis B transmission in Ghana, Interdiscip. Perspect. Infect. Dis. 2019 (2019), 9362492. https://doi.org/10.1155/2019/9362492.

[37] R. Yaesoubi, T. Cohen, Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies, Eur. J. Oper. Res. 215 (2011), 679-687. https://doi.org/10.1016/j.ejor.2011.07.016.

[38] M. Al-Zoughool, T. Oraby, H. Vainio, et al. Using a stochastic continuous-time Markov chain model to examine alternative timing and duration of the COVID-19 lockdown in Kuwait: what can be done now?, Arch. Public Health. 80 (2022), 22. https://doi.org/10.1186/s13690-021-00778-y.

[39] Our World in Data. Accessed: Jan. 01, 2022. https://ourworldindata.org.

[40] data.europa.eu. Accessed: Jan. 01, 2022. https://data.europa.eu/en.

[41] L. Veneti, E. Seppälä, M.L. Storm, et al. Increased risk of hospitalisation and intensive care admission associated with reported cases of SARS-CoV-2 variants B.1.1.7 and B.1.351 in Norway, December 2020 – May 2021, PLoS ONE. 16 (2021), e0258513. https://doi.org/10.1371/journal.pone.0258513.

[42] G. Brown, M. Ozanne, Statistical models for infectious diseases: A useful tool for practical decision-making, Amer. J. Trop. Med. Hyg. 101 (2019), 1–2. https://doi.org/10.4269/ajtmh.19-0354.

[43] J.R.T. Garbin, F.M.C. Leite, C.S.S. Dell'Antonio, et al. Hospitalizations for coronavirus disease 2019: an analysis of the occurrence waves, Sci. Rep. 14 (2024), 5777. https://doi.org/10.1038/s41598-024-56289-7.

[44] L. Rainer, F. Bachner, K. Eglau, et al. Comorbidities and COVID-19 hospitalization, ICU admission and hospital mortality in Austria, Wien. Klin. Wochenschr. 134 (2022), 856–867. https://doi.org/10.1007/s00508-022-02036-9.

[45] V.E. Papageorgiou, G. Vasiliadis, G. Tsaklidis, Analyzing the asymptotic behavior of an extended SEIR model with vaccination for COVID-19, Mathematics. 12 (2023), 55. https://doi.org/10.3390/math12010055.

[46] M. Roberts, V. Andreasen, A. Lloyd, et al. Nine challenges for deterministic epidemic models, Epidemics. 10 (2015), 49–53. https://doi.org/10.1016/j.epidem.2014.09.006.

[47] T. Britton, D. Lindenstrand, Epidemic modelling: Aspects where stochasticity matters, Math. Biosci. 222 (2009), 109–116. https://doi.org/10.1016/j.mbs.2009.10.001.

[48] D. Olabode, J. Culp, A. Fisher, et al. Deterministic and stochastic models for the epidemic dynamics of COVID-19 in Wuhan, China, Math. Biosci. Eng. 18 (2021), 950–967. https://doi.org/10.3934/mbe.2021050.

[49] I. Nåsell, Stochastic models of some endemic infections, Math. Biosci. 179 (2002), 1–19. https://doi.org/10.1016/s0025-5564(02)00098-6.

[50] O.J. Watson, G. Barnsley, J. Toor, et al. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study, Lancet Infect. Dis. 22 (2022), 1293–1302. https://doi.org/10.1016/s1473-3099(22)00320-6.

[51] S.M. Moghadas, T.N. Vilches, K. Zhang, et al. The impact of vaccination on coronavirus disease 2019 (COVID-19) outbreaks in the United States, Clin. Infect. Dis. 73 (2021), 2257–2264. https://doi.org/10.1093/cid/ciab079.

[52] K. Jabłońska, S. Aballéa, M. Toumi, The real-life impact of vaccination on COVID-19 mortality in Europe and Israel, Public Health. 198 (2021), 230–237. https://doi.org/10.1016/j.puhe.2021.07.037.

[53] C. Dye, The benefits of large scale covid-19 vaccination, BMJ (2022), o867. https://doi.org/10.1136/bmj.o867.

[54] R. McCabe, M.D. Kont, N. Schmit, et al. Communicating uncertainty in epidemic models, Epidemics. 37 (2021), 100520. https://doi.org/10.1016/j.epidem.2021.100520.

[55] G. Fiscon, F. Salvadore, V. Guarrasi, et al. Assessing the impact of data-driven limitations on tracing and forecasting the outbreak dynamics of COVID-19, Computers Biol. Med. 135 (2021), 104657. https://doi.org/10.1016/j.compbiomed.2021.104657.

[56] S.N. Park, H.-H. Kim, K.B. Lee, Inherently high uncertainty in predicting the time evolution of epidemics, Epidemiol. Health. 43 (2021), e2021014. https://doi.org/10.4178/epih.e2021014.

[57] G. Webb, A COVID-19 epidemic model predicting the effectiveness of vaccination in the US, Infect. Dis. Rep. 13 (2021), 654–667. https://doi.org/10.3390/idr13030062.

[58] A. Olivares, E. Staffetti, Optimal control-based vaccination and testing strategies for COVID-19, Computer Methods Progr. Biomed. 211 (2021), 106411. https://doi.org/10.1016/j.cmpb.2021.106411.

[59] V. Papageorgiou, G. Tsaklidis, Modeling of premature mortality rates from chronic diseases in Europe, investigation of correlations, clustering and granger causality, Commun. Math. Biol. Neurosci. 2021 (2021), 67. https://doi.org/10.28919/cmbn/5926.

[60] S. Cauchemez, N.M. Ferguson, Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London, J. R. Soc. Interface. 5 (2008), 885–897. https://doi.org/10.1098/rsif.2007.1292.