STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–22.
Published online in International Academic Press (www.IAPress.org)

Parameter and State Estimation in SIRD, SEIR, and SVEIR Epidemiological Models Using Kalman Filter and Genetic Algorithm

Didik Khusnul Arif*, Helisyah Nur Fadhilah

Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Abstract This study presents a comparative investigation of parameter and state estimation techniques applied to three epidemiological models: SIRD, SEIR, and SVEIR. The models are used to simulate infectious disease dynamics with increasing levels of complexity, incorporating factors such as exposure latency and vaccination. Parameter estimation is first performed using three approaches, they are Kalman Filter (KF), Extended Kalman Filter (EKF), and Genetic Algorithm (GA). The best parameter estimates from each method are then used as inputs for state estimation, which is carried out using the Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF). This forms six combinations of estimation strategies (KF-EKF, EKF-EKF, GA-EKF, KF-UKF, EKF-UKF, GA-UKF) evaluated across models. Root Mean Square Error (RMSE) is used as the evaluation metric to assess estimation accuracy. The results demonstrate that GA excels in estimating static parameters, while EKF is more effective for dynamic parameters. Hybrid combinations provide the best performance in state estimation across all models, indicating the benefit of combining global optimization and recursive filtering. These findings can support public health policy by informing the selection of appropriate modeling and estimation techniques to accurately predict epidemic trends, optimize vaccination strategies, and allocate medical resources more effectively during outbreaks. All experiments are conducted on synthetically generated epidemic data to ensure controlled evaluation and generalizability across models.

Keywords Parameter estimation, State estimation, Kalman Filter, Genetic Algorithm, Hybrid estimation, Infectious disease modeling

DOI: 10.19139/soic-2310-5070-2773

1. Introduction

Understanding and forecasting the spread of infectious diseases remains a key challenge in epidemiological modeling. Mathematical models such as SIRD (Susceptible–Infected–Recovered–Deceased), SEIR (Susceptible–Exposed–Infected–Recovered), and SVEIR (Susceptible–Vaccinated–Exposed–Infected–Recovered) are widely used to capture various aspects of disease transmission [1, 2, 3, 4]. These models differ in complexity and epidemiological detail: SIRD includes disease-induced mortality, SEIR introduces a latent exposed compartment to reflect incubation periods, and SVEIR incorporates the impact of vaccination and waning immunity [5, 6]. Such models are critical for guiding public health policy, particularly in outbreak scenarios where rapid assessment and response are required.

Accurate estimation of model parameters such as transmission rate, recovery rate, vaccination rate and hidden state variables such as true number of infections is essential for reliable simulation and prediction. However, in real-world conditions, epidemiological data are often incomplete, delayed, or noisy, making parameter and state estimation a non-trivial task. Static or deterministic fitting techniques, while simple, often fail to adapt to temporal changes in parameter values or to correct for observational uncertainty [7, 8].

*Correspondence to: Didik Khusnul Arif (Email: didik@its.ac.id). Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Various estimation techniques have been explored in the literature to address these limitations. Recursive methods like the Kalman Filter (KF) and its nonlinear extensions Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) are designed to estimate system states and parameters in real time, incorporating both model dynamics and measurement noise [9, 10, 11, 12, 13]. These filters are particularly effective in dynamic environments with streaming data. In contrast, population-based optimization methods such as the Genetic Algorithm (GA) offer global search capabilities and robustness in non-convex error surfaces, making them useful for estimating static or poorly observable parameters [14, 15]. However, GA lacks adaptability during online estimation, while Kalman-based methods may be sensitive to initial conditions or model linearization.

Several previous studies have applied these methods independently. For example, [15] applied GA to tune SEIR models for COVID-19 forecasting, while [9] and [11] demonstrated the utility of KF and EKF for real-time epidemic prediction. Meanwhile, [16] proposed stochastic extensions to SIRD using a hybrid filtering framework. Nevertheless, most of these works focused on a single model or only evaluated one type of estimation method in isolation. Comprehensive comparisons across different models and estimation combinations remain limited.

To bridge this gap, the present study conducts a comparative investigation of parameter and state estimation techniques across three widely used epidemiological models: SIRD, SEIR, and SVEIR. The estimation process is structured into two stages. In the first stage, parameter estimation is performed using three distinct approaches: the Kalman Filter (KF) [13], the Extended Kalman Filter (EKF) [11, 17], and the Genetic Algorithm (GA) [15, 14]. In the second stage, the best parameter estimates obtained from each method are used to estimate the system states via two nonlinear filtering methods: EKF and Unscented Kalman Filter (UKF) [12, 18]. This results in six combined estimation strategies for each model: KF-EKF, EKF-EKF, GA-EKF, KF-UKF, EKF-UKF, and GA-UKF. Estimation accuracy for both parameters and states is assessed using Root Mean Square Error (RMSE), a standard metric for quantifying predictive error in dynamic systems [9, 16].

The objectives of this study are threefold, firstly to evaluate the performance of KF, EKF, and GA in estimating parameters for three compartmental models with increasing complexity (SIRD, SEIR, and SVEIR). Secondly to assess the accuracy of six hybrid estimation schemes (KF-EKF, EKF-EKF, GA-EKF, KF-UKF, EKF-UKF, GA-UKF) in estimating unobservable state variables, and finnaly to determine the most effective estimation method for each model and target variable, offering practical insights into the strengths of combining global search with adaptive filtering techniques.

The novelty of this research lies in its integrative framework that combines two complementary estimation paradigms recursive filters (KF, EKF, UKF) and global optimization (GA) across three widely-used compartmental models of increasing complexity: SIRD, SEIR, and SVEIR. This approach enables a comprehensive evaluation of both static parameter estimation and dynamic state estimation under noisy observation scenarios, thereby enhancing the realism and applicability of the findings in epidemiological surveillance contexts.

Simulation results reveal that hybrid estimation approaches consistently outperform their individual counterparts in both parameter and state estimation tasks. In the SIRD model, the Genetic Algorithm (GA) provides the best parameter estimates for both the infection and recovery rates, while the GA-UKF combination achieves the highest accuracy in estimating the infected and recovered states.

In the SEIR model, parameter estimation shows mixed strengths: EKF performs best for estimating the adult vaccination rate, while GA outperforms others in estimating the infant immunization coverage. For state estimation, the GA-EKF hybrid approach yields the most accurate predictions for both infected and recovered compartments.

In the SVEIR model, GA remains the top performer in estimating the infection probability among vaccinated individuals, whereas EKF achieves the best result for estimating the vaccinated proportion. The EKF-EKF combination consistently provides the best state estimation performance for both infected and recovered populations.

In conclusion, this study offers a robust comparative evaluation of parameter and state estimation methods for infectious disease modeling. By integrating recursive and evolutionary techniques within a hybrid framework, the research provides a flexible and accurate approach for epidemic forecasting and intervention assessment. The results support the practical use of model-specific hybrid estimation in guiding public health responses, particularly in contexts with incomplete or noisy data.

Table 1. Parameter explanation and values for the SIRD model.

Parameter	Definition	Value
N	Total population	1000
δ	Death rate due to Ebola infection	0.2
γ	Immunity loss rate (Recovered $\rightarrow$ Susceptible)	0.02
μ	Natural mortality rate	0.01

2. Mathematical Model

2.1. SIRD Model

This study develops a mathematical model of the SIRD (Susceptible–Infected–Recovered–Deceased) type to capture the transmission dynamics of Ebola virus disease, particularly in relation to infection, recovery, and mortality patterns. The primary goal is to estimate the parameters α (transmission rate) and β (recovery rate), which are crucial for understanding the spread and control of the disease [1, 19].

The total population is divided into four compartments: $S(t)$ represents the susceptible individuals who are at risk of infection due to contact with infected individuals; $I(t)$ denotes the infected individuals who are currently carrying and potentially transmitting the virus; $R(t)$ indicates recovered individuals who may lose immunity and become susceptible again; and $D(t)$ corresponds to individuals who have died as a result of the disease [20].

The model is governed by the following system of differential equations:

$$\frac{dS}{dt} = N + \gamma R - \frac{\alpha SI}{N} - \mu S, \tag{1}$$

$$\frac{dI}{dt} = \frac{\alpha SI}{N} - \beta I - \delta I - \mu I, \tag{2}$$

$$\frac{dR}{dt} = \beta I - \gamma R - \mu R, \tag{3}$$

$$\frac{dD}{dt} = \delta I. \tag{4}$$

The parameter definitions and values used in the model are listed in Table 1. These values are inspired by existing Ebola modeling studies and adapted to reflect realistic dynamics of virus spread in affected regions [1, 19].

The initial values for simulation are defined based on the initial distribution of the population: the number of susceptible individuals is set to $S(0) = 500$, indicating the portion of the population at risk of infection. The number of infected individuals is initialized at $I(0) = 350$, while the recovered individuals are set at $R(0) = 150$. At the start of the simulation, no deaths have occurred, so $D(0) = 0$.

To estimate the unknown parameters α and β, we use three estimation approaches: Kalman Filter (KF), Genetic Algorithm (GA), and Extended Kalman Filter (EKF). These methods rely on observations of the infected compartment $I(t)$, and simulate the system using Euler's method with a time step $\Delta t = 1$ over 100 time units. These techniques have been widely adopted for parameter estimation in nonlinear epidemiological models due to their robustness under noisy conditions [21].

For the EKF method, the Jacobian matrix is required to linearize the nonlinear system around the current state. The Jacobian J with respect to the state and parameters is given by:

$$J = \begin{bmatrix} 1 - \Delta t\left(\frac{\alpha I}{N} + \mu\right) & -\Delta t\frac{\alpha S}{N} & \Delta t\gamma & 0 & -\Delta t\frac{SI}{N} & 0 \\ \Delta t\frac{\alpha I}{N} & 1 + \Delta t\left(\frac{\alpha S}{N} - \beta - \delta - \mu\right) & 0 & 0 & \Delta t\frac{SI}{N} & -\Delta tI \\ 0 & \Delta t\beta & 1 - \Delta t(\gamma + \mu) & 0 & 0 & \Delta tI \\ 0 & \Delta t\delta & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{5}$$

Each row corresponds to the dynamics of compartments S, I, R, and D, while the last two columns capture the sensitivity with respect to the unknown parameters α and β.

Table 2. Parameter explanation and values for the SEIR model.

Parameter	Definition	Value
μ	Natural birth and death rate	0.01241
β	Transmission rate	0.09091
σ	Rate of progression from exposed to infected	0.125
δ	Recovery rate from infection	0.14286

To simulate realistic conditions, Gaussian noise is added to the infected population $I(t)$ to reflect measurement uncertainty. This approach allows us to assess the dynamic behavior of the Ebola outbreak and evaluate how variations in the transmission and recovery rates affect the epidemic curve. The modeling outcomes provide essential insights for designing timely and effective public health interventions, particularly in resource-limited settings where Ebola outbreaks are most devastating [20, 21].

2.2. SEIR Model

This study develops a mathematical model of the SEIR (Susceptible–Exposed–Infected–Recovered) type to estimate two key vaccination-related parameters in measles transmission: the adult vaccination rate (γ) and the infant immunization coverage (ρ). These parameters are essential for understanding the effectiveness of targeted immunization strategies in controlling measles outbreaks [22, 23].

The total population is categorized into four compartments: $S(t)$ represents the susceptible individuals who are at risk of infection through contact with infectious individuals; $E(t)$ denotes the exposed individuals who have been infected but are in the incubation phase and not yet infectious; $I(t)$ refers to the infected individuals who are capable of transmitting the disease; and $R(t)$ indicates the recovered individuals who have gained immunity through recovery or vaccination [24].

The SEIR model includes natural birth and death processes. Newborns enter the population at rate μN, where a ρ proportion are immediately vaccinated and enter the recovered class R, while the remaining $(1-\rho)$ enter the susceptible class S. Adults are vaccinated at rate γ, which also increases the number of recovered individuals.

Normalizing the compartments by total population N, we define $s = \frac{S}{N}$, $e = \frac{E}{N}$, $i = \frac{I}{N}$, and $r = \frac{R}{N}$, so that $s + e + i + r = 1$. The SEIR model is then governed by the following system of differential equations [2]:

$$\frac{ds}{dt} = (1-\rho)\mu - (\beta i + \mu + \gamma)s, \tag{6}$$

$$\frac{de}{dt} = \beta s i - (\mu + \gamma + \sigma)e, \tag{7}$$

$$\frac{di}{dt} = \sigma e - (\mu + \delta)i, \tag{8}$$

$$\frac{dr}{dt} = \rho\mu + \gamma e + \delta i - \mu r + \gamma s. \tag{9}$$

The reproduction number under vaccination, denoted by R_v, quantifies the average number of secondary cases produced by an infectious individual in a partially vaccinated population. It is derived using the next generation matrix method as follows:

$$R_v = \frac{\beta(1-\rho)\sigma}{(\gamma + \sigma + \mu)(\delta + \mu)}. \tag{10}$$

If $R_v < 1$, the disease tends to die out in the long run; otherwise, if $R_v > 1$, an outbreak may occur. Therefore, accurate estimation of the vaccination parameters γ and ρ is crucial in informing public health strategies and achieving herd immunity [2, 22].

The initial conditions used for simulation are normalized by the total population as follows: $S(0) = 100/165$, $E(0) = 40/165$, $I(0) = 15/165$, and $R(0) = 10/165$ [2]. These values are used to initialize the simulation and generate synthetic data for parameter estimation.

The discrete-time simulation of the SEIR model over 100 time steps is given by:

$$\begin{aligned} s_{k+1} &= s_k + \Delta t\left((1-\rho)\mu - (\beta i_k + \mu + \gamma)s_k\right), \\ e_{k+1} &= e_k + \Delta t\left(\beta s_k i_k - (\mu + \gamma + \sigma)e_k\right), \\ i_{k+1} &= i_k + \Delta t\left(\sigma e_k - (\mu + \delta)i_k\right), \\ r_{k+1} &= r_k + \Delta t\left(\rho\mu + \gamma e_k + \delta i_k - \mu r_k + \gamma s_k\right). \end{aligned}$$

To estimate the parameters γ and ρ, we use three different approaches: Kalman Filter (KF), Extended Kalman Filter (EKF), and Genetic Algorithm (GA). In the KF approach, the Jacobian matrix with respect to the parameters is calculated as:

$$H = \begin{bmatrix} -(s+e) & -\mu \end{bmatrix}.$$

For the EKF approach, the full Jacobian matrix F of the nonlinear system with respect to the state and parameters is computed and used to propagate the error covariance. The estimation results at the final time step are taken as the optimal values for γ and ρ.

These techniques allow us to assess the robustness and accuracy of parameter estimation under noisy data conditions, providing a valuable toolset for public health decision-making in vaccination policy [2, 24, 23].

2.3. SVEIR Model

This study develops a mathematical model of the SVEIR (Susceptible–Vaccinated–Exposed–Infectious–Recovered) type to capture the transmission dynamics of COVID-19, particularly under the influence of vaccination and acquired immunity. The primary goal is to estimate two critical parameters: γ, representing the probability of infection among vaccinated individuals, and ϕ, denoting the proportion of vaccinated individuals among the susceptible population. These parameters are essential for evaluating vaccine effectiveness and guiding public health strategies [5, 6].

The total population is divided into five compartments: $S(t)$ denotes the susceptible individuals, including newborns who are immediately at risk of infection; $V(t)$ denotes vaccinated individuals; $E(t)$ refers to those exposed to the virus and in the incubation period without symptoms; $I(t)$ indicates infectious individuals exhibiting clinical symptoms; and $R(t)$ comprises recovered individuals who have acquired immunity [3, 6].

The SVEIR model used in this study is governed by the following system of nonlinear ordinary differential equations, adapted from recent studies on measles and COVID-19 modeling [3, 6]:

$$\frac{dS}{dt} = \frac{II}{N} + (1-\phi)(\xi_1 + \xi_2 I) - (\mu + \phi\alpha)S, \tag{11}$$

$$\frac{dV}{dt} = \phi\alpha S - (\mu + \gamma(\xi_3 + \xi_4 I))V, \tag{12}$$

$$\frac{dE}{dt} = (1-\phi)(\xi_1 + \xi_2 I)S - \gamma(\xi_3 + \xi_4 I)V - (\omega + \mu)E, \tag{13}$$

$$\frac{dI}{dt} = \omega E - (\rho + \mu + \delta)I, \tag{14}$$

$$\frac{dR}{dt} = \rho I - \mu R. \tag{15}$$

The parameter definitions and corresponding values are provided in Table 3. These values are adapted from existing epidemiological models to reflect plausible COVID-19 transmission dynamics under varying vaccination conditions [5, 25].

Table 3. Parameter explanation and values for the SVEIR model.

Parameter	Definition	Value
II	Human birthrate	3.75×10^5
μ	Natural human mortality rate	0.2
η	Infection probability after contact (S–I)	0.2
ξ_1	Contact rate between S and E	0.09
ξ_2	Contact rate between S and I	0.07
ξ_3	Contact rate between V and E	0.03
ξ_4	Contact rate between V and I	0.02
ω	Transition rate from E to I	0.005
ρ	Recovery rate from I to R	0.27
δ	Death rate due to COVID-19	0.001
α	Vaccination rate	0.001

To estimate the values of γ and ϕ, we implement three estimation approaches: Kalman Filter (KF), Genetic Algorithm (GA), and Extended Kalman Filter (EKF). The estimation process is based on synthetic observations derived from the infected population $I(t)$ with added Gaussian noise to reflect realistic measurement uncertainty [25].

In the Kalman Filter approach, the state vector is composed of the parameters $[\gamma, \phi]^\top$, and the measurement sensitivity is represented by the Jacobian matrix H:

$$\frac{\partial I}{\partial \gamma} = -V(\xi_3 + \xi_4 I),$$
$$\frac{\partial I}{\partial \phi} = -\alpha S - (\xi_1 + \xi_2 I)S.$$

The Extended Kalman Filter method expands the state vector to $[S, V, E, I, R, \gamma, \phi]^\top$, allowing the filter to account for both the state evolution and the dynamic sensitivity to parameter changes. The full Jacobian matrix F of the system is derived accordingly from partial derivatives of the system equations with respect to all state variables and unknown parameters.

The accuracy and robustness of each method are evaluated by comparing the estimated values to ground truth values $\gamma^\star = 0.1$ and $\phi^\star = 0.8$. Results show that while all three methods provide acceptable estimates, the EKF consistently outperforms the others in terms of convergence speed and estimation precision under moderate to high noise conditions [6, 25].

3. Parameter Estimation Methods

3.1. Kalman Filter (KF)

The Kalman Filter (KF) is a recursive estimation algorithm widely used for dynamic systems with noise and uncertainty. It is particularly effective in estimating hidden states or unknown parameters of linear systems, and serves as a foundational technique in control theory, signal processing, and epidemiological modeling [13]. In the context of this study, the KF is employed to estimate two unknown parameters of the epidemic model, denoted by the vector $\theta = [\theta_1, \theta_2]^\top$, based on noisy observations of the infected population over time.

The Kalman Filter assumes a linear discrete-time state-space model for parameter estimation:

$$\theta_{k+1} = \theta_k + w_k, \tag{16}$$
$$y_k = h_k(\theta_k) + v_k, \tag{17}$$

where θ_k is the parameter vector at time k, w_k is the process noise with zero mean and covariance matrix Q, y_k is the observation (e.g., the number of infected individuals), h_k is the nonlinear observation function derived from the compartmental model, and v_k is the measurement noise with zero mean and covariance R.

When the observation model h_k is nonlinear with respect to θ_k, it is linearized via first-order Taylor expansion around the current estimate to obtain the Jacobian matrix:

$$H_k = \left.\frac{\partial h_k}{\partial \theta}\right|_{\theta=\theta_k}.$$

The recursive Kalman Filter algorithm comprises two main steps:

1. **Prediction step:**

$$\begin{aligned} \theta_{k|k-1} &= \theta_{k-1}, \\ P_{k|k-1} &= P_{k-1} + Q, \end{aligned}$$

   where $P_{k|k-1}$ is the predicted covariance matrix.
2. **Update step:**

$$\begin{aligned} K_k &= P_{k|k-1} H_k^\top (H_k P_{k|k-1} H_k^\top + R)^{-1}, \\ \theta_k &= \theta_{k|k-1} + K_k (y_k - h_k(\theta_{k|k-1})), \\ P_k &= (I - K_k H_k) P_{k|k-1}, \end{aligned}$$

   where K_k is the Kalman gain matrix.

The observations y_k are obtained from the infected compartment $I(t)$ of the model with added Gaussian noise to simulate real-world reporting inaccuracies. At each time step, the filter updates the parameter estimates using the prediction residual and adjusts their uncertainty through the Kalman gain. This method is computationally efficient and well-suited for sequential estimation when new data arrive in real time.

Although KF assumes linearity, it can serve as a benchmark for comparing more sophisticated nonlinear techniques such as the Extended Kalman Filter (EKF) [11, 17], which have been applied successfully to SEIR-type models under noisy and uncertain environments.

3.2. Extended Kalman Filter (EKF)

The Extended Kalman Filter (EKF) is a widely used recursive estimation technique for nonlinear systems, offering a practical balance between computational tractability and estimation accuracy. In contrast to the standard Kalman Filter, which is suitable only for linear systems, EKF extends its applicability by linearizing the nonlinear transition and observation functions using first-order Taylor series expansions [13].

In this study, the unknown parameters of the epidemiological model are embedded in the augmented state vector:

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k \\ \theta_k \end{bmatrix},$$

where $\mathbf{x}_k$ denotes the state variables S, E, I, R, and $\theta_k = [\theta_1, \theta_2]^\top$ represents the time-varying unknown parameters to be estimated. The model dynamics and observation process are defined as follows:

$$\mathbf{z}_{k+1} = f(\mathbf{z}_k) + \mathbf{w}_k, \tag{18}$$

$$\mathbf{y}_k = h(\mathbf{z}_k) + \mathbf{v}_k, \tag{19}$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are assumed to be zero-mean Gaussian noise processes representing system and measurement uncertainties, respectively.

The EKF proceeds through two primary operations: prediction and update.

Prediction step:

$$\hat{\mathbf{z}}_{k|k-1} = f(\hat{\mathbf{z}}_{k-1}),$$
$$P_{k|k-1} = F_{k-1}P_{k-1}F_{k-1}^{\top} + Q,$$

where F_{k-1} is the Jacobian matrix of the nonlinear function $f(\cdot)$ with respect to $\mathbf{z}$, evaluated at the prior estimate $\hat{\mathbf{z}}_{k-1}$, and Q is the process noise covariance matrix.

Update step:

$$K_k = P_{k|k-1}H_k^{\top}(H_kP_{k|k-1}H_k^{\top} + R)^{-1},$$
$$\hat{\mathbf{z}}_k = \hat{\mathbf{z}}_{k|k-1} + K_k(\mathbf{y}_k - h(\hat{\mathbf{z}}_{k|k-1})),$$
$$P_k = (I - K_kH_k)P_{k|k-1},$$

where H_k is the Jacobian matrix of the observation function $h(\cdot)$, and R denotes the measurement noise covariance matrix.

In the context of infectious disease modeling, the observation $\mathbf{y}_k$ typically corresponds to the number of infected individuals, I_k. The EKF algorithm iteratively adjusts both the state variables and parameter estimates to minimize the prediction error. This makes EKF particularly advantageous when only partial and noisy observations of the epidemic are available [9, 10].

By dynamically incorporating incoming observations, the EKF provides a robust and flexible framework for tracking the evolution of both disease states and key epidemiological parameters in real-time. Its ability to handle model nonlinearity and parameter uncertainty makes it especially suitable for complex models such as SEIR and SVEIR under various vaccination scenarios [13].

3.3. Genetic Algorithm (GA)

The Genetic Algorithm (GA) is a population-based metaheuristic optimization technique inspired by the principles of natural selection and genetics. It has been extensively employed for solving nonlinear, non-convex, and multi-modal optimization problems, particularly in contexts where traditional gradient-based methods may fail or converge slowly [14]. This makes GA highly suitable for parameter estimation in epidemiological models such as SIR, SEIR, and SVEIR.

In this study, GA is used to estimate a parameter vector $\theta = [\theta_1, \theta_2]^{\top}$ (e.g., transmission rate, vaccination rate) by minimizing the discrepancy between observed data and simulated outputs of the model. The optimization objective is defined as the sum of squared errors between the observed and simulated infected populations:

$$J(\theta) = \sum_{k=1}^{T} \left(I_k^{\text{obs}} - I_k^{\text{sim}}(\theta)\right)^2,$$

where I_k^{obs} represents the observed number of infected individuals at time step k, and $I_k^{\text{sim}}(\theta)$ denotes the corresponding simulated output from the model given parameter θ.

The Genetic Algorithm consists of the following steps:

1. **Initialization:** An initial population of candidate solutions is generated randomly within predetermined bounds for each parameter.
2. **Evaluation:** Each candidate solution is evaluated based on its fitness, calculated from the objective function $J(\theta)$.
3. **Selection:** Individuals are selected for reproduction based on their fitness scores using methods such as roulette wheel or tournament selection.
4. **Crossover:** Pairs of selected individuals undergo crossover to produce offspring, facilitating exploration of the parameter space.
5. **Mutation:** Random perturbations are applied to some offspring to maintain population diversity and avoid premature convergence.

6. **Replacement:** The new generation is formed by replacing less-fit individuals with the newly generated offspring.
7. **Termination:** The process repeats for a predefined number of generations or until convergence criteria are met.

The optimal parameter values are selected from the best-performing individual in the final population. GA's stochastic and population-based nature enables it to effectively navigate complex parameter landscapes and to tolerate noisy data and model uncertainty, making it particularly advantageous in epidemiological modeling [15, 16].

Recent studies have demonstrated the effectiveness of GA in conjunction with compartmental models for forecasting disease outbreaks and optimizing model parameters. For instance, Li et al. [14] proposed a GA-based ensemble framework for epidemic prediction, while Qiu et al. [15] successfully applied GA to an improved SEIR model to predict COVID-19 trends in China. Additionally, Sebbagh et al. [16] integrated GA with state estimation methods such as EKF for robust forecasting under stochastic model assumptions.

4. State Estimation Methods

4.1. EKF for State Estimation

The EKF formulation used in this work follows the nonlinear state–space representation previously introduced in Section 3.2. Specifically, the system dynamics and observation process correspond to the models defined in equations (18)–(19), where the process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are assumed to be zero-mean Gaussian disturbances.

For the purpose of state estimation, the augmented formulation described in Section 3.2 is specialized by assuming that the epidemiological parameters have already been estimated. Thus, the EKF is applied only to the epidemiological state vector $\mathbf{x}_k$, while the parameters remain fixed during the filtering process.

Using the estimated parameter set $\hat{\theta}$ obtained from the parameter estimation stage, the EKF prediction step evaluates the nonlinear state update function $f(\cdot)$, and the update step incorporates the observation y_k, which—under typical surveillance settings—corresponds to the infected population I_k. The corresponding Jacobian matrices F_k and H_k are computed with respect to $\mathbf{x}_k$, consistent with the general EKF structure described earlier.

By reusing the state–space formulation in equations (18)–(19), the EKF provides a recursive and computationally efficient mechanism for estimating the unobservable epidemic states in real time for the SIRD, SEIR, and SVEIR models.

4.2. Unscented Kalman Filter (UKF)

The Unscented Kalman Filter (UKF) is a recursive filtering method used to estimate the state of a nonlinear system. Unlike the Extended Kalman Filter (EKF), which linearizes the nonlinear functions via Taylor expansion, UKF uses the Unscented Transform to propagate a set of carefully chosen sigma points through the nonlinear system dynamics. This allows for a more accurate estimation of the mean and covariance without the need for explicit linearization [18, 12].

In the context of epidemiological models, UKF is employed to estimate the unobservable compartments from partial and noisy observations, such as reported infected cases. Let the discrete-time nonlinear system be defined as:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k, \tag{20}$$

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{v}_k, \tag{21}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement vector, $f(\cdot)$ is the nonlinear system model, $h(\cdot)$ is the observation function, and $\mathbf{w}_k$, $\mathbf{v}_k$ are zero-mean Gaussian process and measurement noise with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively.

The UKF proceeds in the following steps [12, 26]:

1. **Initialization:** Set initial state estimate $\hat{\mathbf{x}}_0$ and initial covariance $\mathbf{P}_0$.
2. **Sigma Point Generation:** At time k, generate $2n+1$ sigma points $\chi_k^{(i)}$ from $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ using the Unscented Transform.
3. **Prediction:**

$$\begin{aligned}
\chi_{k+1|k}^{(i)} &= f(\chi_k^{(i)}, \mathbf{u}_k), \\
\hat{\mathbf{x}}_{k+1|k} &= \sum_{i=0}^{2n} W_i^{(m)} \chi_{k+1|k}^{(i)}, \\
\mathbf{P}_{k+1|k} &= \sum_{i=0}^{2n} W_i^{(c)} \left(\chi_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k}\right) \left(\chi_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k}\right)^T + \mathbf{Q}.
\end{aligned}$$

4. **Update:**

$$\begin{aligned}
\mathbf{y}_{k+1}^{(i)} &= h(\chi_{k+1|k}^{(i)}), \\
\hat{\mathbf{y}}_{k+1} &= \sum_{i=0}^{2n} W_i^{(m)} \mathbf{y}_{k+1}^{(i)}, \\
\mathbf{P}_{yy} &= \sum_{i=0}^{2n} W_i^{(c)} (\mathbf{y}_{k+1}^{(i)} - \hat{\mathbf{y}}_{k+1})(\mathbf{y}_{k+1}^{(i)} - \hat{\mathbf{y}}_{k+1})^T + \mathbf{R}, \\
\mathbf{P}_{xy} &= \sum_{i=0}^{2n} W_i^{(c)} (\chi_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{y}_{k+1}^{(i)} - \hat{\mathbf{y}}_{k+1})^T, \\
\mathbf{K}_{k+1} &= \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}, \\
\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1}), \\
\mathbf{P}_{k+1} &= \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{P}_{yy}\mathbf{K}_{k+1}^T.
\end{aligned}$$

The UKF enables the estimation of hidden states even when only partial observations are available. Its ability to handle strong nonlinearities with better accuracy than EKF makes it particularly suitable for real-time monitoring and forecasting of infectious disease dynamics [18, 12].

5. Hybrid Estimation Framework

In epidemiological modeling, accurate prediction of disease spread relies not only on reliable compartmental models but also on robust estimation techniques. Parameter estimation and state estimation often exhibit different characteristics parameters are typically static or slowly varying, whereas states evolve dynamically with time. As such, combining estimation methods that specialize in each domain can yield superior performance. This motivates the use of hybrid estimation schemes that leverage the strengths of global optimization for parameter estimation and recursive filtering for dynamic state tracking.

This section presents six hybrid estimation strategies used in this study, each involving a two-stage process, firstly parameter estimation using KF, EKF, or GA then state estimation using EKF or UKF. Each scheme is outlined with a brief explanation and pseudocode to clarify its implementation.

5.1. KF-EKF

The KF-EKF scheme begins with parameter estimation using the standard Kalman Filter, assuming a linearized observation model. Once the optimal parameters are obtained, they are used in the Extended Kalman Filter (EKF) for estimating the nonlinear state dynamics.

Algorithm 1 KF-EKF Estimation

Input: Observation data $\{y_k\}_{k=1}^T$
Estimate parameter vector $\hat{\theta}$ using Kalman Filter (KF)
Initialize state vector $\hat{x}_0$ and covariance P_0
for $k = 1$ to T **do**
    Apply EKF with fixed $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^T$

5.2. EKF-EKF

In this scheme, both parameter and state estimation are performed using the Extended Kalman Filter. Parameters are embedded in the augmented state vector and updated simultaneously with the compartmental states. After convergence, the estimated parameters are used to initialize a second EKF for refined state estimation.

Algorithm 2 EKF-EKF Estimation

Input: Observation data $\{y_k\}_{k=1}^T$
Estimate parameter vector $\hat{\theta}$ using EKF (augmented state)
Initialize state vector $\hat{x}_0$ and covariance P_0
for $k = 1$ to T **do**
    Apply EKF with fixed $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^T$

5.3. GA-EKF

The GA-EKF strategy utilizes the Genetic Algorithm to identify optimal model parameters that minimize the error between observed and simulated data. These parameters are then used within an EKF framework to perform state estimation. This combination is useful when the parameter space is nonlinear or non-convex.

Algorithm 3 GA-EKF Estimation

Input: Observation data $\{y_k\}_{k=1}^T$
Initialize population of parameters $\{\theta_i\}$
for each generation **do**
    Simulate model with θ_i
    Evaluate fitness: $J(\theta_i) = \sum_k (y_k - I_k^{\text{sim}})^2$
end for
Select best parameter $\hat{\theta}$
Initialize state vector $\hat{x}_0$ and covariance P_0
for $k = 1$ to T **do**
    Apply EKF with $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^T$

5.4. KF-UKF

This hybrid scheme starts by estimating the parameters using Kalman Filter under the assumption of linearized system dynamics. The parameters are then applied in the Unscented Kalman Filter (UKF), which tracks the nonlinear evolution of the system states.

Algorithm 4 KF-UKF Estimation

Input: Observation data $\{y_k\}_{k=1}^{T}$
Estimate parameter vector $\hat{\theta}$ using KF
Initialize UKF with $\hat{x}_0$, P_0
for $k = 1$ to T **do**
    Apply UKF using $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^{T}$

5.5. EKF-UKF

EKF-UKF begins with estimating the parameters through the Extended Kalman Filter. The estimated parameters are then applied in a UKF framework to track system states. This method benefits from EKF's ability to infer nonlinear parameters and UKF's robustness in nonlinear state estimation.

Algorithm 5 EKF-UKF Estimation

Input: Observation data $\{y_k\}_{k=1}^{T}$
Estimate parameter vector $\hat{\theta}$ using EKF
Initialize UKF with $\hat{x}_0$, P_0
for $k = 1$ to T **do**
    Apply UKF using $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^{T}$

5.6. GA-UKF

The GA-UKF approach uses Genetic Algorithm to identify the best parameters based on the discrepancy between simulated and observed data. These parameters are then fed into the Unscented Kalman Filter to recursively estimate the hidden compartments of the model. This method is particularly useful for models with highly nonlinear dynamics and uncertain data.

Algorithm 6 GA-UKF Estimation

Input: Observation data $\{y_k\}_{k=1}^{T}$
Initialize population of parameters $\{\theta_i\}$
for each generation **do**
    Simulate model with θ_i
    Evaluate fitness: $J(\theta_i) = \sum_k (y_k - I_k^{\text{sim}})^2$
end for
Select best parameter $\hat{\theta}$
Initialize UKF with $\hat{x}_0$, P_0
for $k = 1$ to T **do**
    Apply UKF using $\hat{\theta}$ to estimate $\hat{x}_k$
end for
Output: Estimated states $\{\hat{x}_k\}_{k=1}^{T}$

6. Experimental Setup

This section summarizes the procedures used to generate synthetic data, specify noise assumptions, define model parameters, configure hyperparameters, and set the numerical simulation environment for evaluating all estimation methods.

6.1. Data Generation and Noise Modeling

Synthetic epidemic trajectories for the SIRD, SEIR, and SVEIR models were generated using their respective nonlinear differential equations, discretized with the forward Euler (finite-difference) method. The same discretized model was used within all Kalman-based estimators to ensure consistency between data generation and state-update dynamics.

Two types of uncertainty were incorporated: process noise (representing unmodeled dynamics) and measurement noise (representing surveillance errors). Both noise sources followed Gaussian assumptions with diagonal covariance matrices. Only the infected compartment was assumed observable, consistent with typical epidemic surveillance systems, while the remaining states were inferred by the filters.

6.2. Model Initialization and Hyperparameter Settings

Ground-truth epidemiological parameters and initial conditions were selected from biologically reasonable ranges to generate realistic outbreak trajectories. These values also served as reference points for evaluating parameter-estimation accuracy.

Hyperparameters for the Genetic Algorithm (GA) were tuned through preliminary sensitivity exploration. Population size, mutation probability, and crossover rate were varied across broad admissible ranges, and a population size of 50 with moderate mutation and crossover settings was found to provide stable convergence with reasonable computational cost. Similarly, the process and measurement noise covariances for KF, EKF, and UKF were chosen through a coarse grid search across multiple magnitudes, ensuring stable filter behavior and preventing divergence across all models. Although not exhaustive, these sensitivity-based selections avoid bias and maintain fair comparison among the estimation methods.

6.3. Numerical Simulation and Evaluation Metrics

All simulations were implemented in MATLAB R2023b with identical numerical and computational settings. The simulation horizon was chosen to include the outbreak growth, peak, and decay phases, using a uniform time step of one day. Each filter was initialized with perturbed initial estimates to emulate realistic uncertainty in prior knowledge.

Estimation performance was evaluated using two standard error measures: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics were computed separately for parameter estimation, state reconstruction, and hybrid method comparison across the SIRD, SEIR, and SVEIR models.

7. Real Epidemic Data and Model Validation

To evaluate the practical applicability of the proposed estimation framework, three real epidemic datasets were used to validate the SIRD, SEIR, and SVEIR models. Each dataset corresponds to a distinct infectious disease with different transmission characteristics and data availability. For all cases, the reported infected counts $I(t)$ were treated as observable measurements, while the remaining state variables were reconstructed through model dynamics driven by the estimated parameters.

It is important to note that real epidemic datasets inherently contain non-Gaussian fluctuations, reporting delays, and varying levels of under-reporting, which differ from the idealized Gaussian noise assumptions typically used in synthetic experiments. Although the present validation employed real data to implicitly capture these irregularities, further investigation using explicitly modeled Poisson noise and delayed or incomplete observations is planned as

an extension of this study. In addition, incorporating multi-compartment observations will be explored to evaluate how multiple observation channels influence the accuracy and stability of the hybrid estimation framework.

7.1. Ebola Outbreak (SIRD Model)

For the SIRD model, validation was performed using real data from the 2014–2016 Ebola outbreak in Western Africa, published by the World Health Organization (WHO). The dataset consists of weekly confirmed case counts, which were used as the observable infected population $I(t)$. The susceptible $S(t)$, recovered $R(t)$, and deceased $D(t)$ compartments were inferred by propagating the SIRD equations using the parameters estimated by the hybrid filtering methods. This dataset provides a suitable benchmark for SIRD-type dynamics, as Ebola exhibits direct progression from infection to recovery or death with minimal latent-stage behavior.

7.2. Measles Outbreak (SEIR Model)

The SEIR model was validated using the 2018 measles incidence data from Arizona, obtained from the Centers for Disease Control and Prevention (CDC). The observable variable is the number of infected individuals $I(t)$ per week. The susceptible $S(t)$, exposed $E(t)$, and recovered $R(t)$ states were reconstructed through the SEIR model using parameter estimates obtained from KF, GA, and EKF. This dataset is particularly relevant because measles exhibits a strong latent (exposed) period, allowing assessment of the hybrid estimator's performance in handling unobserved states and indirect transmission pathways.

7.3. COVID-19 Case Data (SVEIR Model)

For the SVEIR model, real validation was conducted using daily COVID-19 case data from Indonesia during the 2020–2021 outbreak, accessed through the official national data portal. The confirmed daily infections served as the observable state $I(t)$, while the susceptible $S(t)$, vaccinated $V(t)$, exposed $E(t)$, and recovered $R(t)$ compartments were reconstructed from the model equations. The inclusion of vaccination dynamics and higher system dimensionality makes this dataset ideal for evaluating the robustness of hybrid filters involving KF, EKF, UKF, and GA in more complex nonlinear epidemic systems.

It should also be noted that real epidemic datasets often reflect implicit time-varying dynamics due to interventions or behavioral changes, which cause parameters such as the transmission rate α to vary over time. Although the present analysis assumes constant parameters for comparability, future work will explicitly incorporate time-varying parameters to examine the adaptability of the proposed hybrid estimation methods

8. SIRD Results

This section presents the parameter- and state-estimation performance of KF, GA, EKF, and the six hybrid filtering methods applied to the SIRD epidemiological model. All results are evaluated using both model-generated and real epidemic datasets.

8.1. Parameter Estimation Results for SIRD

Figure 1 illustrates the estimation trajectories of the transmission rate α and the recovery rate β obtained using KF, GA, and EKF. Clear differences in convergence behavior are visible. KF exhibits a pronounced initial transient and slower convergence, GA shows smooth trajectories due to its global search capability, and EKF converges the fastest toward the true values particularly for α demonstrating robustness against nonlinear effects.

Table 4 summarizes the ranking of the three estimation methods based on RMSE and MAE for both model-generated and real data. For α, EKF consistently achieves the smallest error in both datasets, while KF performs poorly under real data due to sensitivity to measurement noise. For β, KF is the best method for smooth model-generated data but becomes the worst under real observations. In contrast, EKF becomes the most accurate for β in real-data scenarios, demonstrating improved robustness when confronted with real-world fluctuations.

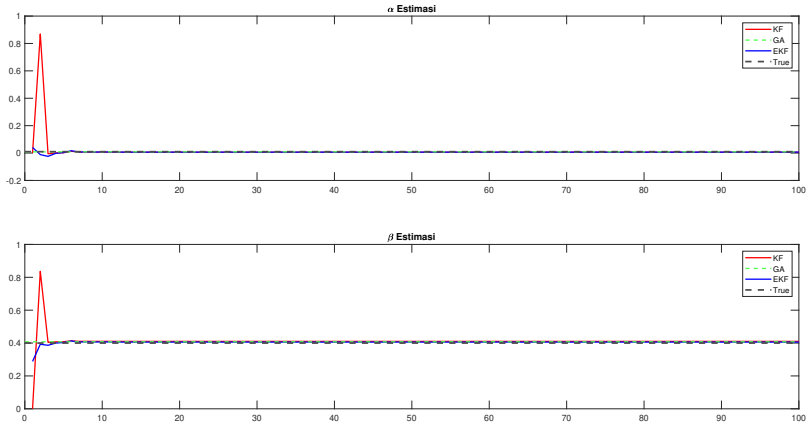

Figure 1. Parameter estimation results for α and β using KF, GA, and EKF in the SIRD model.

Table 4. Ranking of parameter-estimation methods for SIRD based on RMSE and MAE (model-generated vs real data).

Metric	Parameter	Rank 1	Rank 2	Rank 3
RMSE (Model)	α	EKF (0.006549)	GA (0.008232)	KF (0.02159)
	β	KF (0.02488)	GA (0.03770)	EKF (0.07442)
RMSE (Real)	α	EKF (0.007715)	GA (0.01355)	KF (0.3627)
	β	EKF (0.3942)	GA (0.4013)	KF (0.5770)
MAE (Model)	α	EKF (0.006549)	GA (0.008232)	KF (0.02159)
	β	KF (0.02488)	GA (0.03770)	EKF (0.07442)
MAE (Real)	α	EKF (0.007715)	GA (0.01355)	KF (0.3627)
	β	EKF (0.3942)	GA (0.4013)	KF (0.5770)

8.2. State Estimation Results for SIRD

Figure 2 presents the state-estimation performance for the infected (I) and recovered (R) compartments using the six hybrid filtering configurations: KF–EKF, GA–EKF, EKF–EKF, KF–UKF, GA–UKF, and EKF–UKF. All methods are able to track the general epidemic trajectory; however, notable performance differences arise in transient regions and peak phases.

GA–EKF and GA–UKF produce the most stable tracking with minimal deviations, reflecting the advantage of GA-based global parameter initialization combined with nonlinear correction from EKF or UKF. KF–EKF and KF–UKF exhibit larger deviations, especially during early transients, due to weaker ability to compensate for nonlinearities.

The complete hybrid-method ranking is reported in Table 8. Consistently, GA–EKF achieves the best accuracy for both synthetic and real datasets, while KF–EKF performs the worst across most metrics. These results highlight the importance of combining global optimization with recursive nonlinear filtering for robust state reconstruction.

The SIRD results demonstrate clear distinctions among estimation methods. EKF provides the most reliable parameter estimates for α and maintains strong performance under real noise conditions. KF performs well for β only under idealized model-generated data. For state reconstruction, GA-based hybrids—especially GA–EKF—offer the highest accuracy and the most stable transient behavior. KF–EKF exhibits consistently weak performance, underscoring the limitations of relying solely on local linear corrections in nonlinear epidemic dynamics.

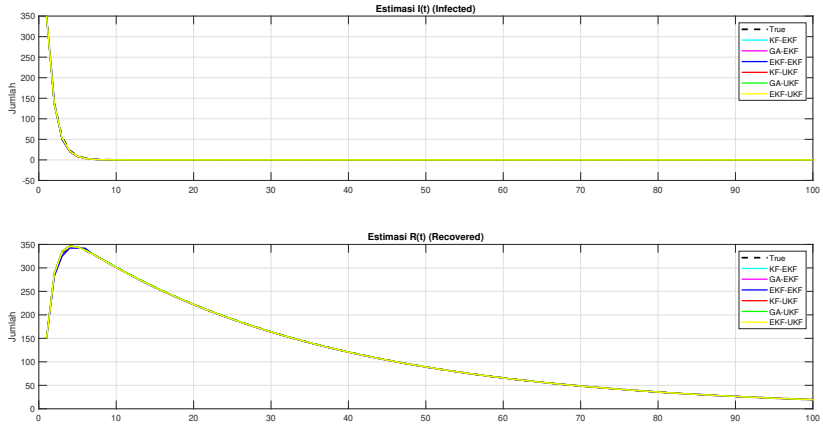

Figure 2. State estimation results for infected (I) and recovered (R) using hybrid filters in the SIRD model.

Overall, the SIRD analysis confirms that hybrid strategies combining global search (GA) with nonlinear filtering (EKF or UKF) deliver the most robust and accurate estimation performance.

9. SEIR Results

This section presents the parameter-estimation and state-estimation results for the SEIR model using KF, GA, EKF, and the six hybrid filtering architectures. The SEIR framework introduces an additional latent compartment, resulting in stronger nonlinear coupling and increased parameter sensitivity compared to the SIRD model.

9.1. Parameter Estimation Results for SEIR

Figure 3 illustrates the estimation trajectories for the recovery rate γ and the exposure-to-infection transition rate ρ. GA provides the smoothest convergence for both parameters, while KF and EKF show small steady-state deviations and oscillations particularly for ρ, whose indirect influence through the latent exposed compartment increases estimation difficulty.

Table 5 summarizes the ranking of the three parameter-estimation methods. GA consistently achieves the lowest error for γ across both synthetic and real data. For ρ, KF performs best under model-generated data but deteriorates dramatically when applied to real observations, revealing strong sensitivity to measurement noise. GA becomes the most accurate method for ρ under real data, while EKF remains a balanced middle performer.

9.2. State Estimation Results for SEIR

Figure 4 presents the state-estimation performance of the six hybrid filters for the infected (I) and recovered (R) compartments. All hybrid methods successfully reproduce the overall SEIR dynamics: a rapid growth phase followed by a gradual decline.

GA–EKF and KF–EKF provide the closest match to the true infected trajectory, with minimal oscillations during the peak and decay phases. For the recovered compartment, GA–EKF and EKF–EKF provide the most stable and accurate tracking. Hybrids involving KF tend to show larger deviations, particularly when noise levels are high.

Table 9 summarizes the detailed ranking based on RMSE and MAE. GA–EKF and EKF–UKF emerge as the two most consistently accurate hybrid methods, while KF–EKF is frequently the weakest performer, reflecting its sensitivity to nonlinearities and noise.

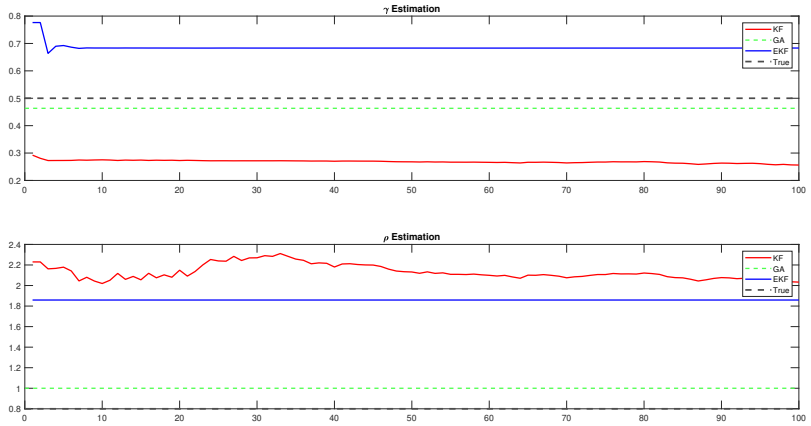

Figure 3. Parameter estimation results for γ and ρ using KF, GA, and EKF in the SEIR model.

Table 5. Ranking of parameter-estimation methods for SEIR based on RMSE and MAE (model-generated vs real data).

Metric	**Parameter**	**Rank 1**	**Rank 2**	**Rank 3**
RMSE (Model)	γ	GA (0.02764)	KF (0.3403)	EKF (1.261)
	ρ	KF (0.5096)	EKF (0.7712)	GA (0.8000)
RMSE (Real)	γ	GA (0.5000)	EKF (0.8169)	KF (1.179)
	ρ	GA (0.8000)	EKF (0.9669)	KF (141.9)
MAE (Model)	γ	GA (0.02764)	KF (0.3403)	EKF (1.261)
	ρ	KF (0.5096)	EKF (0.7712)	GA (0.8000)
MAE (Real)	γ	GA (0.5000)	EKF (0.8169)	KF (1.179)
	ρ	GA (0.8000)	EKF (0.9669)	KF (141.9)

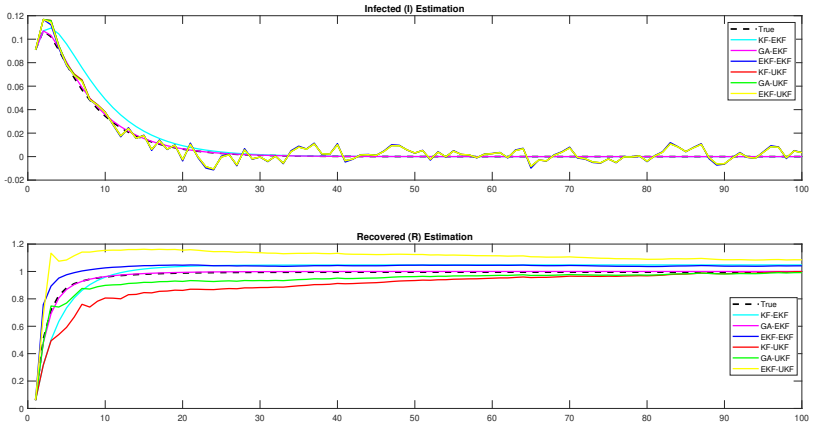

Figure 4. State estimation results for infected (I) and recovered (R) using hybrid filters in the SEIR model.

The SEIR results reinforce several consistent trends. GA provides the strongest parameter estimation for γ across all data conditions, while KF is competitive for ρ only under smooth model-generated data. For real-world observations, GA and EKF become significantly more reliable.

In state estimation, hybrids involving EKF especially GA–EKF and EKF–UKF—offer the highest accuracy for both infected and recovered states. The presence of the latent exposed compartment increases model nonlinearity, making purely local filters such as KF–EKF less effective. Overall, combining GA with nonlinear recursive filters yields the most robust estimation performance in the SEIR setting.

10. SVEIR Results

This section presents the parameter and state estimation results for the SVEIR model using KF, GA, EKF, and the six hybrid filtering configurations. As the most nonlinear and highest-dimensional model in this study, SVEIR poses a substantially more challenging estimation environment due to interactions among susceptible, vaccinated, exposed, infected, and recovered compartments.

10.1. Parameter Estimation Results for SVEIR

Figure 5 illustrates the estimation trajectories for the recovery rate γ and the vaccination-effectiveness parameter ϕ. Among the three parameter-estimation methods, GA achieves the most stable convergence for both parameters, rapidly approaching the true values with minimal oscillatory behavior. KF and EKF also converge but exhibit more pronounced fluctuations, particularly during early transients, reflecting sensitivity to the strong nonlinear coupling between vaccinated and exposed compartments.

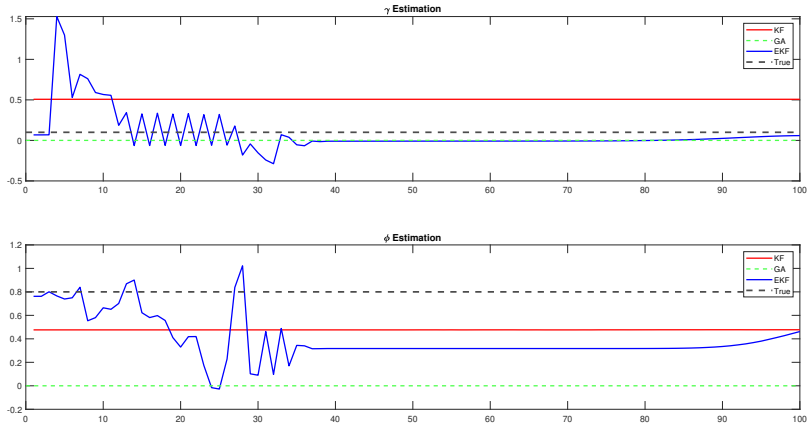

Figure 5. Parameter estimation results for γ and ϕ using KF, GA, and EKF in the SVEIR model.

Table 6 reports the ranking of KF, GA, and EKF based on RMSE and MAE. For γ, KF performs best under model-generated data but deteriorates significantly when applied to real observations. In contrast, GA provides the lowest error for γ under real-data conditions. For ϕ, EKF consistently achieves the best accuracy in both synthetic and real datasets, indicating its effectiveness in handling parameters whose influence is deeply embedded in nonlinear state interactions.

Table 6. Ranking of parameter-estimation methods for SVEIR based on RMSE and MAE (model-generated vs real data).

Metric	Parameter	Rank 1	Rank 2	Rank 3
RMSE (Model)	γ	KF (0.03513)	GA (0.09999)	EKF (0.1569)
	ϕ	EKF (0.0002569)	KF (0.5052)	GA (0.7999)
RMSE (Real)	γ	GA (0.09999)	KF (0.3469)	EKF (0.9379)
	ϕ	EKF (0.2450)	KF (0.6571)	GA (0.7999)
MAE (Model)	γ	KF (0.03513)	GA (0.09999)	EKF (0.1569)
	ϕ	EKF (0.0002569)	KF (0.5052)	GA (0.7999)
MAE (Real)	γ	GA (0.09999)	KF (0.3469)	EKF (0.9379)
	ϕ	EKF (0.2450)	KF (0.6571)	GA (0.7999)

10.2. State Estimation Results for SVEIR

Figure 6 shows the performance of the six hybrid estimators for the infected (I) and recovered (R) populations. All methods reproduce the qualitative epidemic pattern rapid growth, a peak, and a long decay—but exhibit substantial differences in accuracy.

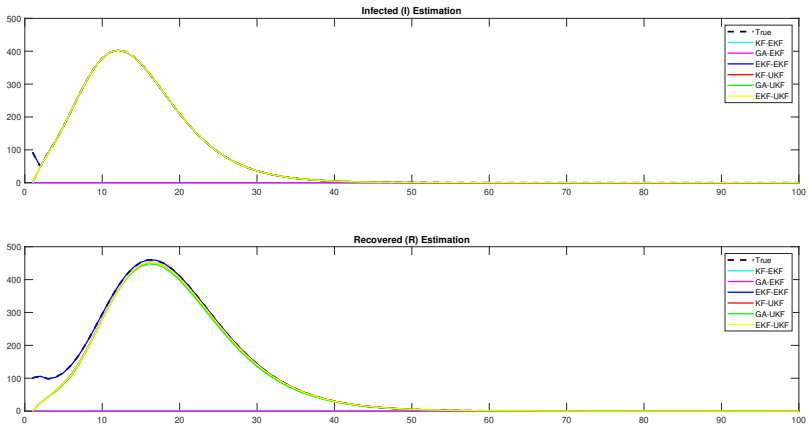

Figure 6. State estimation results for infected (I) and recovered (R) using hybrid filters in the SVEIR model.

Hybrid EKF–EKF provides the closest match to the true SVEIR trajectories, with smooth curves and minimal deviation across all phases of the epidemic. This performance reflects the advantage of consistent linearization and recursive correction in managing high-dimensional nonlinear interactions.

Hybrids involving UKF particularly KF–UKF and GA–UKF—also follow the true dynamics well, although they show slightly larger deviations around the infection peak. The GA–EKF hybrid performs reasonably but displays small fluctuations during early transients, suggesting sensitivity to the interaction between GA-initialized parameters and EKF-based updates.

Table 10 summarizes the full ranking of hybrid filters based on RMSE and MAE. EKF–EKF emerges as the strongest performer under synthetic data, but under real data its errors can increase substantially due to numerical sensitivity triggered by strong nonlinearity and noise.

The SVEIR results reveal the complex interplay between method choice and model structure. For parameters, GA provides the most stable estimates of γ under real data, while EKF is clearly superior for estimating ϕ. KF performs well for γ under smooth synthetic data but degrades significantly in the presence of noise.

For state estimation, hybrids involving EKF—particularly EKF–EKF and KF–EKF—produce the most accurate results, although they may become numerically sensitive under highly nonlinear conditions. UKF-based hybrids offer robust and generally accurate estimates but may show slight deviations around the epidemic peak.

Overall, the SVEIR model amplifies the strengths and weaknesses of each hybrid estimator. No single method is universally optimal; the best-performing strategy depends on the degree of nonlinearity, parameter identifiability, and the interaction between parameter and state dynamics.

11. Cross-Model Comparison

A cross-model comparison reveals that estimation performance strongly depends on model complexity. In the SIRD model, GA consistently delivers the most accurate parameter estimates, making GA–EKF the most reliable hybrid configuration. In the moderately nonlinear SEIR model, performance becomes more balanced: GA remains effective for γ, while EKF and UKF provide stable corrections for the latent exposed dynamics, with all hybrid methods remaining numerically stable. In contrast, the SVEIR model exhibits substantial challenges due to its high dimensionality and strong nonlinear coupling. EKF-only estimation frequently diverges, while hybrids initialized with GA or KF (such as GA–EKF, GA–UKF, and KF–EKF) achieve the most stable and accurate results. These trends align with the computation times shown in Table 7, where Kalman-based methods are consistently faster than GA. However, EKF and UKF become more expensive in SVEIR due to increased covariance propagation demands. The evolution of the covariance matrix P_k further reflects model difficulty: in SIRD and SEIR, P_k decreases steadily as observations accumulate, while in SVEIR it may temporarily increase under strong nonlinear effects or model data mismatch. Overall, robust initialization (GA or KF), consistent covariance propagation, and hybrid filtering strategies are essential for maintaining accuracy and stability in complex epidemic models. Future extensions such as the Extended Kalman Smoother (EKS) or Unscented Kalman Smoother (UKS) may further enhance estimation quality in retrospective analyses by leveraging future information.

Although the primary focus of this study is methodological, the estimated parameters also carry direct public health interpretations. For instance, the recovery-related parameter γ in the SIRD, SEIR, and SVEIR models is closely related to the average infectious period, and in vaccination-augmented models it can reflect the combined effect of treatment success and vaccine-induced protection. Higher estimated values of γ imply faster removal of infectious individuals from the transmission chain, which can support decisions regarding booster scheduling, treatment reinforcement, or prioritization of high-risk groups during an outbreak. Likewise, the transmission parameter α quantifies the effective contact rate in the population; increases in its estimated value signal intensified transmission pressure. Such information can guide policymakers in adjusting non-pharmaceutical interventions, including mobility restrictions, mask mandates, and temporary lockdowns. The hybrid estimation framework thus provides not only technical accuracy but also epidemiologically interpretable signals that can inform real-time public health responses.

Table 7. Average computation time (in seconds) per simulation for each estimation method.

Model	KF	EKF	UKF	GA
SIRD	0.0079	0.0029	0.0039	0.0495
SEIR	0.0046	0.0029	0.0045	0.0398
SVEIR	0.0039	0.0123	0.0273	0.0656

Table 8. Ranking of hybrid methods based on RMSE and MAE for the SIRD model (Model vs Real data).

Metric	State	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
RMSE (Model)	I	GA–EKF (0.6011)	EKF–EKF (0.9612)	EKF–UKF (1.2113)	GA–UKF (2.1139)	KF–UKF (3.4124)	KF–EKF (2395.46)
	R	GA–EKF (2.4195)	EKF–UKF (1.4887)	GA–UKF (3.3397)	EKF–EKF (4.6313)	KF–UKF (6.2400)	KF–EKF (12129.7)
RMSE (Real)	I	GA–EKF (0.7766)	EKF–UKF (1.4980)	GA–UKF (8.2198)	KF–UKF (9.3876)	EKF–EKF (6.5229)	KF–EKF (NaN)
	R	GA–EKF (175.68)	EKF–UKF (173.28)	GA–UKF (177.49)	KF–UKF (210.65)	EKF–EKF (251.29)	KF–EKF (NaN)
MAE (Model)	I	GA–EKF (1.8683)	EKF–UKF (1.1599)	GA–UKF (2.6743)	EKF–EKF (2.9914)	KF–UKF (4.9289)	KF–EKF (7152.30)
	R	GA–EKF (0.1039)	EKF–UKF (0.2702)	EKF–EKF (0.2775)	GA–UKF (0.8489)	KF–UKF (2.1972)	KF–EKF (940.64)
MAE (Real)	I	GA–EKF (13.564)	EKF–UKF (13.394)	GA–UKF (14.087)	KF–UKF (16.264)	EKF–EKF (23.613)	KF–EKF (NaN)
	R	EKF–UKF (0.00116)	GA–EKF (0.1461)	GA–UKF (4.8428)	EKF–EKF (5.2119)	KF–UKF (5.8039)	KF–EKF (NaN)

Table 9. Ranking of hybrid methods based on RMSE and MAE for the SEIR model (Model vs Real data).

Metric	State	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
RMSE (Model)	I	GA–UKF (0.0272)	GA–EKF (0.0281)	EKF–UKF (0.0534)	EKF–EKF (0.0769)	KF–UKF (0.0946)	KF–EKF (0.1227)
	R	GA–EKF (0.000304)	GA–UKF (0.00507)	EKF–UKF (0.00508)	KF–UKF (0.00513)	EKF–EKF (0.00554)	KF–EKF (0.00785)
RMSE (Real)	I	EKF–UKF (0.00237)	GA–EKF (0.0518)	GA–UKF (0.648)	EKF–EKF (1.78)	KF–UKF (2.08)	KF–EKF (14.1)
	R	GA–EKF (0.0478)	EKF–UKF (0.562)	KF–UKF (0.584)	EKF–EKF (0.586)	GA–UKF (0.610)	KF–EKF (0.929)
MAE (Model)	I	GA–EKF (0.000110)	KF–EKF (0.00338)	GA–UKF (0.00407)	EKF–UKF (0.00408)	KF–UKF (0.00413)	EKF–EKF (0.00448)
	R	GA–EKF (0.000152)	GA–UKF (0.000314)	EKF–UKF (0.000814)	EKF–EKF (0.00125)	KF–UKF (0.00258)	KF–EKF (0.00420)
MAE (Real)	I	GA–EKF (0.0287)	EKF–UKF (0.533)	KF–UKF (0.543)	EKF–EKF (0.554)	KF–EKF (0.573)	GA–UKF (0.577)
	R	EKF–UKF (0.00116)	GA–EKF (0.0272)	EKF–EKF (0.374)	GA–UKF (0.460)	KF–UKF (1.738)	KF–EKF (8.655)

Table 10. Ranking of hybrid methods based on RMSE and MAE for the SVEIR model (Model Data vs Real Data).

Metric	State	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
RMSE (Model)	Infected (I)	GA–EKF (98.9283)	GA–UKF (98.9283)	KF–EKF (98.9288)	KF–UKF (4082.50)	EKF–UKF (6.12e+05)	EKF–EKF (2.25e+21)
	Recovered (R)	GA–UKF (2087.46)	KF–UKF (5398.28)	KF–EKF (1.28e+04)	GA–EKF (1.28e+04)	EKF–UKF (6.23e+05)	EKF–EKF (2.50e+21)
RMSE (Real)	Infected (I)	KF–EKF (76.0473)	GA–EKF (76.0473)	GA–UKF (76.0473)	KF–UKF (76.0476)	EKF–UKF (76.0545)	EKF–EKF (8154.64)
	Recovered (R)	KF–EKF (9.25981)	GA–EKF (9.26061)	GA–UKF (30.7565)	KF–UKF (26.5138)	EKF–UKF (276.404)	EKF–EKF (1.12e+04)
MAE (Model)	Infected (I)	GA–UKF (816.626)	KF–UKF (2428.90)	KF–EKF (5917.48)	GA–EKF (5917.46)	EKF–UKF (2.43e+05)	EKF–EKF (4.55e+20)
	Recovered (R)	GA–EKF (20.2271)	GA–UKF (20.2271)	KF–EKF (20.2276)	KF–UKF (1649.77)	EKF–UKF (2.38e+05)	EKF–EKF (4.20e+20)
MAE (Real)	Infected (I)	KF–EKF (1.08981)	GA–EKF (1.12945)	KF–UKF (12.6932)	GA–UKF (14.8320)	EKF–UKF (157.686)	EKF–EKF (3721.95)
	Recovered (R)	GA–EKF (7.60777)	GA–UKF (7.60777)	KF–EKF (7.60794)	KF–UKF (7.68071)	EKF–UKF (8.40773)	EKF–EKF (2935.60)

12. Conclusion

This study presented a comparative investigation of hybrid parameter and state estimation strategies for the SIRD, SEIR, and SVEIR epidemiological models. The results show that estimation performance depends strongly on model dimensionality and nonlinear coupling, with simpler models favoring GA–EKF and more complex systems benefiting from combinations such as GA–EKF, GA–UKF, and KF–EKF. These hybrid approaches not only improve numerical stability but also provide epidemiologically actionable insights: changes in the transmission rate α can inform the need for mobility restrictions or distancing policies, while variations in the recovery-related parameter γ may reflect treatment effectiveness or vaccine-induced protection, supporting booster scheduling and clinical planning. The inclusion of real epidemic datasets further demonstrates the robustness of the proposed framework under real-world reporting noise and incomplete observations.

Despite these strengths, the models retain structural simplifications such as homogeneous mixing and single-strain dynamics, which limit their ability to capture age heterogeneity, variant emergence, or spatial transmission. Future work will extend the framework to accommodate time-varying parameters, multi-strain or age-structured SVEIR systems, and smoother-based retrospective reconstruction using EKS or UKS. Such extensions will enhance both the realism and policy relevance of hybrid Kalman–GA estimation in practical epidemic surveillance and forecasting.

REFERENCES

1. A. Atangana, E. F. D. Goufo, *et al.*, "On the mathematical analysis of ebola hemorrhagic fever: deathly infection disease in west african countries," *BioMed research international*, vol. 2014, 2014.
2. J. L. M. Hamzah, H. N. Fadhilah, and R. Syaifullah, "Analysis of the seir model for measles spread with the influence of vaccination.," *IAENG International Journal of Applied Mathematics*, vol. 55, no. 3, 2025.
3. M. El Hajji and A. H. Albargi, "A mathematical investigation of an "sveir" epidemic model for the measles transmission," *Math. Biosci. Eng*, vol. 19, pp. 2853–2875, 2022.
4. M. El Hajji and A. H. Albargi, "A mathematical investigation of an "SVEIR" epidemic model for the measles transmission," *Math. Biosci. Eng*, vol. 19, pp. 2853–2875, 2022.
5. Z. Zhang, S. Kundu, J. P. Tripathi, and S. Bugalia, "Stability and hopf bifurcation analysis of an sveir epidemic model with vaccination and multiple time delays," *Chaos, Solitons & Fractals*, vol. 131, p. 109483, 2020.
6. Z.-j. Yin, H. Xiao, S. McDonald, V. Brusic, and T.-y. Qiu, "Dynamically adjustable sveir (mh) model of multiwave epidemics: Estimating the effects of public health measures against covid-19," *Journal of Medical Virology*, vol. 95, no. 12, p. e29301, 2023.
7. F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li, and N. S. Crowcroft, "The basic reproduction number (R_0) of measles: a systematic review," *The Lancet Infectious Diseases*, vol. 17, no. 12, pp. 420–428, 2017.
8. H. Trottier, H. Carabin, and P. Philippe, "Measles, pertussis, rubella and mumps completeness of reporting. literature review of estimates for industrialized countries," *Revue D'epidemiologie et de Sante Publique*, vol. 54, no. 1, pp. 27–39, 2006.
9. M. S. Islam, M. E. Hoque, and M. R. Amin, "Integration of kalman filter in the epidemiological model: a robust approach to predict covid-19 outbreak in bangladesh," *International Journal of Modern Physics C*, vol. 32, no. 08, p. 2150108, 2021.
10. Y. Shi, X. Zhu, X. Zhu, B. Cheng, and Y. Zhong, "Kalman filter-based epidemiological model for post-covid-19 era surveillance and prediction," *Sensors*, vol. 25, no. 8, p. 2507, 2025.
11. X. Zhu, B. Gao, Y. Zhong, C. Gu, and K.-S. Choi, "Extended kalman filter based on stochastic epidemiological model for covid-19 modelling," *Computers in Biology and Medicine*, vol. 137, p. 104810, 2021.
12. M. Impraimakis and A. W. Smyth, "An unscented kalman filter method for real time input-parameter-state estimation," *Mechanical Systems and Signal Processing*, vol. 162, p. 108026, 2022.
13. M. Khodarahmi and V. Maihami, "A review on kalman filter models," *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 727–747, 2023.
14. X. Li, Q. Yu, Y. Yang, C. Tang, and J. Wang, "An evolutionary ensemble model based on ga for epidemic transmission prediction," *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 5, pp. 7469–7481, 2023.
15. Z. Qiu, Y. Sun, X. He, J. Wei, R. Zhou, J. Bai, and S. Du, "Application of genetic algorithm combined with improved seir model in predicting the epidemic trend of covid-19, china," *Scientific Reports*, vol. 12, no. 1, p. 8910, 2022.
16. A. Sebbagh, C. E. Bencheriet, and S. Kechida, "A stochastic epidemiological sird-v model with lsm-ekf algorithm for forecasting and monitoring the spread of covid-19 pandemic: Real data," *IEEE Access*, 2024.
17. D. Ndanguza, I. Mbalawata, and J. Nsabimana, "Analysis of sdes applied to seir epidemic models by extended kalman filter method," *Applied Mathematics*, vol. 7, no. 17, pp. 2195–2211, 2016.
18. L. Chang, B. Hu, A. Li, and F. Qin, "Transformed unscented kalman filter," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 252–257, 2012.
19. T. Berge, J.-S. Lubuma, G. Moremedi, N. Morris, and R. Kondera-Shava, "A simple mathematical model for ebola in africa," *Journal of biological dynamics*, vol. 11, no. 1, pp. 42–74, 2017.
20. M. Goeijenbier, J. Van Kampen, C. Reusken, M. Koopmans, and E. Van Gorp, "Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis," *Neth J Med*, vol. 72, no. 9, pp. 442–8, 2014.

21. S. J. Smither, L. S. Eastaugh, and M. S. Lever, "Comparison of aerosol stability of different variants of ebola virus and marburg virus and virulence of aerosolised ebola virus in an immune-deficient mouse," *Viruses*, vol. 14, no. 4, p. 780, 2022.
22. I. Malhotra and N. Goel, "Infectious disease modeling: From traditional to evolutionary algorithms," *Archives of Computational Methods in Engineering*, vol. 31, no. 2, pp. 663–699, 2024.
23. S. Stockmaier, N. Stroeymeyt, E. C. Shattuck, D. M. Hawley, L. A. Meyers, and D. I. Bolnick, "Infectious diseases and social distancing in nature," *Science*, vol. 371, no. 6533, p. eabc8881, 2021.
24. P. Girardi and C. Gaetan, "An seir model with time-varying coefficients for analyzing the sars-cov-2 epidemic," *Risk Analysis*, vol. 43, no. 1, pp. 144–155, 2023.
25. R. Zarin, "Artificial neural network-based approach for simulating influenza dynamics: A nonlinear sveir model with spatial diffusion," *Engineering Analysis with Boundary Elements*, vol. 176, p. 106230, 2025.
26. R. K. Jatoth and G. A. Reddy, "A hybrid ga-adaptive particle swarm optimization based tuning of unscented kalman filter for harmonic estimation," in *Swarm, Evolutionary, and Memetic Computing: First International Conference on Swarm, Evolutionary, and Memetic Computing, SEMCCO 2010, Chennai, India, December 16-18, 2010. Proceedings 1*, pp. 380–388, Springer, 2010.