Stochastic SEIR Epidemic Model in a Homogeneous Community By Kalman Filter Parameter Estimation

Ricca D. Callis

EN 625.641.82.SU20 Introduction to Stochastic Differential Equations

Johns Hopkins University

12/15/2020

Introduction

Epidemiology is the study of the distribution (frequency, pattern) and determinants (causes, risk factors) of health-related states and events (not just diseases) in specified populations (e.g., Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy, et al., 2014). It is the foundation of public health and provides the architecture for policy decisions and best practices by identifying risk factors for disease and targets for preventative healthcare. In spite of advances in epidemiology over the last century, infectious disease continues to plague human populations, reducing life expectancy, increasing mortality, burdening quality of life, and inflicting wide-ranging socioeconomic disruptions. Traditionally, epidemiology utilized time-series and regression analyses to identify transmission parameters and create predictions regarding the distribution and burden of the disease (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014). These techniques, however, tend to scale poorly when modeling complex, nonlinear systems with reciprocal causality, feedback loops, tipping points, or other threshold effects common in communicable diseases (Kreuger & Osgood, 2015).

To incorporate randomness into these mathematical models, the terms are formulated as probabilities at which an event occurs, rather than rates of occurrences. To study and predict disease outbreak, the stochastic SEIR (Susceptible-Exposed-Infectious-Recovered) model, for example, outlines the probabilities with which individuals become susceptible to infection, exposed to an infection, infected with a disease, and recovered from a disease (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014). With known parameters, the SEIR model is useful for prediction purposes, but it is much more limited for emerging communicable diseases. Even when applied to influenza, a common and well-

researched endemic infection, with established estimates of parameter values and a surplus of historical data, model predictions created early during an outbreak differ greatly from final observed rates.

Complex, nonlinear systems are unpredictable and involve many factors which change rapidly and are, therefore, difficult to quantify (e.g., risk perception impacting contact patterns, shifting vaccine attitudes, compliance fatigue, diagnosing patterns, or weather). The divergence is made more likely due to the fact that so many of these factors, including changes in human contact patterns, play a substantial role in disease transmission and may not be captured by the model (e.g., Kennedy et al, 2014; Kreuger & Osgood, 2015). Dynamic models which employ statistical filtering and estimation methods such as Sequential Monte Carlo (SMC) and Markov Chain Monte Carlo (MCMC) methods create robust present-moment predictions and the opportunity to update the model based on continued empirical observations. Among estimation algorithms, Kalman filtering is often included to create estimates based on consensus of empirical data and provide model predictions using Maximum Likelihood Estimation (MLE).

Consequently, a general stochastic SEIR model will be produced with an associated Kalman filter to simulate outbreak prediction estimates. The model's predecessors, SIR (Susceptible-Infectious-Recovered) model, will be provided for context. Afterward, the SEIR epidemic model will be formulated as a system of deterministic nonlinear differential equations and then changed to a system of stochastic nonlinear differential equations (SDEs). The numerical simulation of the resulting SDEs will be solved by the Euler-Maruyama approach, and parameters will be estimated by adaptive Markov Chain Monte Carlo and extended Kalman filter methods.

Epidemiology Modeling

Mathematical models are used to identify and better understand a system, simulate a system's behavior, predict future behavior, and ultimately provide system control (Sameni, 2020). Thus, mathematical modeling of infectious diseases is used to study the mechanisms by which diseases spread, to predict the future course of an outbreak, and to evaluate strategies to control an epidemic (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014; Kratzer, 2016). To do this, epidemic models relate population-level disease dynamics to basic host and pathogen properties to explore and predict the infection process. In order to model the progress of an epidemic in a large population, population diversity must be reduced to a few key characteristics which are relevant to the infection under consideration. For example, models studying common childhood diseases that confer long-lasting immunity typically divide the population into three categories (referred to as compartments): those who are susceptible to the disease, those who are infected, and those who have recovered and are immune (known as the SIR model; Kermack & McKendrick, 1927). This idea has been extended to SEIR epidemic model where the population can be partitioned into four compartments: susceptible (S), latent or exposed (E), infectious (I), and recovered (R; Ndanguza et al).

Compartmental Models

Compartmental models represent differential equations of dynamic systems. A compartment is an abstract entity representing the quantity of interest (volume, number, density, etc.;Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Sameni, 2020). Each variable of interest is equivalent to a system state and is abstractly represented by a single compartment and visually conceptualized as a box. Each

compartment is assumed to be internally homogeneous, which implies that all entities assumed inside the compartment are indistinguishable. The compartments interact with one another through a set of rate equations, visually represented by arrows between the compartments. Therefore, compartmental models can be converted to a set of first order linear or nonlinear equations (and vice versa), by writing the net flow into a compartment (Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Sameni, 2020). Specifically, the edge weights are multiplied by the state variable of their start node and added to (subtracted from) the rate change equation of the end node (start node). External inputs can be considered to be originated from an external node with value 1. The system of equations represent a state-space model and after setting initial conditions, the system of equations can be solved either analytically or numerically (Costa, 2017; Sameni, 2020).

From this perspective, for an n compartment system, the compartment variables can be considered as state variables denoted in vector form as $\boldsymbol{x}(t) = [x_1(t), \dots, x_n(t)]^T$. The compartmental model provides a graphical representation of the state-space model:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{w}(t);\ \boldsymbol{\theta}(t), t)$$

$$\boldsymbol{y}(t) = \boldsymbol{g}(\boldsymbol{x}(t);\ \boldsymbol{\theta}(t), t) + \boldsymbol{v}(t) \qquad (1)$$

where,

$\boldsymbol{f}(\cdot)$ is the state dynamics function corresponding to the compartmental model graph (which can be time-variant and nonlinear)

$w(t) = [w_1(t), \dots, w_1(t)]^T$ represents deterministic or stochastic external system inputs

$\boldsymbol{y}(t) = [y_1(t), \dots, y_m(t)]^T$ is the vector of observable model invariables considered as outputs (the measures)

$\boldsymbol{g}(\cdot)$ is the function that maps the state variables to the observations (measurements)

$v(t) = [v_1(t), \dots, v_m(t)]^T$ is the vector of measurement inaccuracies, considered as additive noise

$\theta(t) = \left[\theta_1(t), \dots, \theta_p(t)\right]^T$ is a vector of model parameters to be set or identified

The state-space form of (1) implies that one may eventually be able to estimate and predict the compartment variables from noisy measurements, using state-space estimation techniques, such as the Kalman or extended Kalman filter.

A compartmental model is linear, when its rate flow factors are independent of the state variables and nonlinear when its rate flow factors are dependent on the state variables (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Sameni, 2020). A compartmental model is time-invariant, when its rate flow factors are independent of time and time-variant when its rate flow factors are dependent of time. Compartmental models may be open or closed. In closed systems, the quantities are only passed between the compartments, while in open systems the quantities may flow into or out of the whole system. In a closed compartmental model, the sum of all the differential equations of the system is zero (for all *t*).

A key inference problem for these systems is that the epidemic process operates in continuous time but available data are almost always discrete snapshots of the system (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, McKinley et al., 2014; Sameni, 2020). As a result, these systems are likely to form (partial) time-series counts of individuals in some, but not necessarily all, of the epidemic states. This makes evaluation of the likelihood difficult unless we introduce latent variables to account for the unobserved (continuous) events or using a discrete-time model where we will assume that

we wish to fit a continuous-time model to discrete sampled data. The latter approach will be expanded upon below.

In epidemiology, mathematical models assume the population can be divided into a set of distinct compartments (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Sameni, 2020). These compartments are defined with respect to disease status. To study disease transmission, epidemiologists consider the rates at which individuals progress through different epidemiological states over time. A functional form is then chosen to describe the movements of individuals between states, governed by a set of epidemiological parameters.

Three Compartment Model: Susceptible-Infected-Recovered (SIR)

The simplest epidemiological compartmental model was described by Kermack and McKendrick in 1927, and included three compartments: susceptible (S), infected (I), recovered (R; see Figure 1). This model examined two flow processes. First, a susceptible host (S) can become infected by an infected host (I) at some rate (parameter *b*). This leads to the susceptible host leaving the S compartment and entering the I compartment. Second, an infected host either dies or recovers which moves the host into compartment R. This is described by parameter *g* in the model.

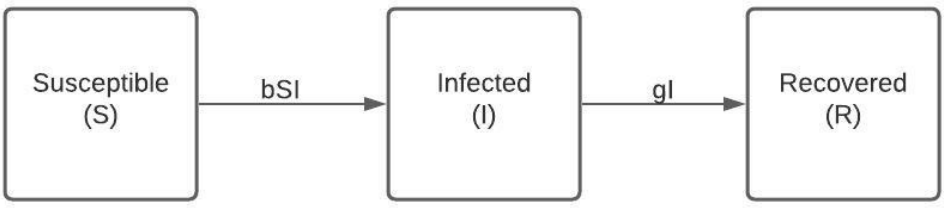

FIGURE 1. Basic Susceptible-Infected-Recovered epidemiological model without immunity.

In this model, the total population of *N* individuals exposed to an epidemic disease at each time instant *t* is divided into three groups (each represented by a compartment): the susceptible group fraction denoted by S(*t*), the infected group fraction denoted by I(*t*), and the recovered group fraction denoted by R(*t*). Accordingly, the system is closed and we have:

$$S(t) + I(t) + R(t) = 1$$

Each compartment represents a differential equation. The susceptible compartment holds hosts susceptible to disease infection at time t. The rate at which new infections occur is bSI, for some positive constant b.

$$\frac{dS}{dt} = -bS(t)I(t) \quad (2)$$

From this general equation, we can observe that this is a first-order differential equation because its highest order is a first derivative.

The Infected/Infectious compartment holds hosts infected with the disease at time t. The rate at which individuals can enter the removed/recovered class occurs at the rate gI, for some positive constant g. The inverse of the recovery rate indicates the length of the symptomatic infectious period.

$$\frac{dI}{dt} = bS(t)I(t) - gI(t) \quad (3)$$

From this general equation, we can observe that this is a first-order differential equation because its highest order is a first derivative.

The recovered compartment holds hosts who have recovered from the disease at time t. This basic model assumes lifetime immunity, such that once recovered, hosts do not return to the susceptible compartment.

$$\frac{dR}{dt} = gI(t) \quad (4)$$

From this general equation, we can observe that this is a first-order differential equation because its highest order is a first derivative.

For the SIR model, the total population size (N) is a constant because:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = -bSI + (bIS - gI) + gI = 0 \quad (5)$$

Note, however, that for purposes of this model, total population was given a numerical value higher than 0. If, for example, I = 0, then there are no invectives and the right-hand side of all three equations is 0, so nothing will change. While I(0) is normally small relative to N, we must have I(0) > 0 for an epidemic to develop. Equation (4) says that if I = 0 at time 0 (or any time), then dI/dt = 0 as well, and there can never be any increase from the 0-level of infection.

SIR Solution: Euler Method

A specific model solution modeling the number of hosts in each compartment, given initial conditions at time *t* would depend on already acquired data regarding parameters *b* and *g* and would also assume, as stated earlier, that the population is constant. It should be noted, however, that public health officials usually use the SIR model to estimate parameters *b* and *g* because they already have established host data for each compartment. For example, public health officials may utilize the Euler method to estimate the *b* (the average number of transmissions from an infected person in a time period).

$$b = \frac{S_n - S_{n+1}}{S_n I_n} \quad (6)$$

Public health officials also often study solutions at or near equilibria by ignoring R(*t*) and instead focus on (S, I) at (*b*, 0).

The Euler Method allows for a step-wise solution to the SIR model, given established parameters b and g, initial values of S, I, and R at time 0, and change in time (Δt). Since each host is always in one of the three compartments, exact values of S, I, and R can be calculated at discrete-time steps (t+1). To complete the final step, a series of discrete-time models must be calculated to derive the continuous-time model of disease transmission. The general premise of the Euler Method allows one to generate a sequence of y-values given a first derivative ("slope"; $\frac{dy}{dt}$) at any point (t,y) see (Smith & Moore, 2004).

Within the Euler method, S_{n+1}, I_{n+1}, and R_{n+1} are the number of susceptible, infected/infectious, and removed hosts at time (n+1). Δt is a small change in time. For every small increase in time (Δt), the new number of hosts in the susceptible compartment depends on the previous number of hosts in the susceptible compartment and the rate at which new infections occur (bSI), for some positive constant b.

$$s_{n+1} = s_n - bS_nI_n\Delta t$$

For every small increase in time (Δt), the new number of hosts in the infected/infectious compartment depends on the previous number of hosts in the infected compartment and the rate at which susceptible hosts become infected (bSI) and the rate at which infected hosts can enter the removed compartment (gI).

$$I_{n+1} = I_n(bS_n - g)\Delta t$$

For every small increase in time (Δt), the new number of hosts in the removed compartment depends on the previous number of hosts in the removed compartment plus the rate at which individuals recover from infection, gI, for some positive constant g.

$$R_{n+1} = R_n + gI_n\Delta t$$

The ODE for flow counting processes become:

$$\frac{dN_{SI}}{dt} = \mu_{SI}(t)S(t), \qquad \frac{dN_{IR}}{dt} = \mu_{IR}I(t)$$

SIR Solution: Reduced to One SDE

The methods presented above use a system of three differential equations with three unknown functions. To reduce this problem to a single differential equation, we can consider I (3) as a function of S (2). According to the chain rule:

$$\frac{dI}{dS} = \frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{bIS - gI}{-bIS} = \frac{g}{bS} - 1 \qquad (7)$$

Multiply both sides of Equation *(7)* by dS:

$$dI = \left(\frac{g}{bS}\right) - 1dS \qquad (8)$$

Integrate one side with respect to I and the other with respect to S:

$$\int dI = \int \left(\frac{g}{bS} - 1\right) dS \qquad (9)$$

$$I = \frac{g}{b}\ln(S) - S + C \qquad (10)$$

In Equation (10), C is an arbitrary constant which can be found by using the initial condition I(0):

$$I(0) - \frac{g}{b}\ln(S0) + S(0) = C \qquad (11)$$

Given initial conditions, S(0), I(0), g, and b, these values can be substituted back into Equation (11) to solve for C. Finally, the value of C can be substituted back into Equation (10) to solve for I. The value of C acts to maintain the relationship between S and I in a manner that holds as the epidemic (started from some set of initial conditions) proceeds. Different initial conditions will yield different values of C (assuming parameters *b* and *g* remain stable) and will trace a family of curves in the (S,I) plane. Equation (10) is constant along each curve, a

concept referred to as conserved quantity (e.g., Llyod, 2017). The conserved quantity is used to find the size of an epidemic, equating the value at the start of the epidemic to its value at the end (i.e., as *t* approaches infinity). The full explanation of this calculation is, however, beyond the scope of this paper.

SIR Model Behavior

From these general equations, we can observe how the compartmentalized groups will act as $t \to \infty$ (also see Figure 2). Note that since R(t) = N – S(t) – I(t), the system is generally reduced to a system of two ODEs: Equation (2) and Equation (3). Equation (2) shows that the susceptible group will decrease over time and approach zero. Equation (3) shows that recovered group will increase and approach population size (N) over time. To fully understand the infected and infectious group [Equation (3)], we have to take the integral of Equation (3) from 0 to t, which yields:

$$g \int_0^t I(s)ds = R(t) \quad (12)$$

Solve for R(t):

$$R(t) = N - S(t) - I(t) \quad (13)$$

Combine:

$$g \int_0^t I(s)ds = N - S(t) - I(t) \quad (14)$$

Thus, as t goes to infinity, $I(t) \to 0$:

$$I(t) \to N - S(\infty) - g \int_0^t I(s)ds \quad (15)$$

Unlike the susceptible group, the infected/infectious group is not always negative or zero; nor is it always positive or zero, as it is with the removed group. As observed by examining Equation (2), the rate of change for the infected/infectious group depends on b, g, and S(t). When bS(t) is less than g, the rate of change for the infected/infectious group is negative. However, when bS(t) is greater than g, the rate of change for the infected/infectious group is positive. Finally, when bS(t) is equal to g, then the rate of change for the infected/infectious group is zero.

Thus, what we see is that the disease always dies out. Equation (2) shows that for all initial conditions, $I(\infty) = 0$. While equation (3) shows that for a sufficiently large enough t, $\frac{dR}{dt} > \frac{gI(\infty)}{2} > 0$ which implies that $R(\infty) = \infty$.

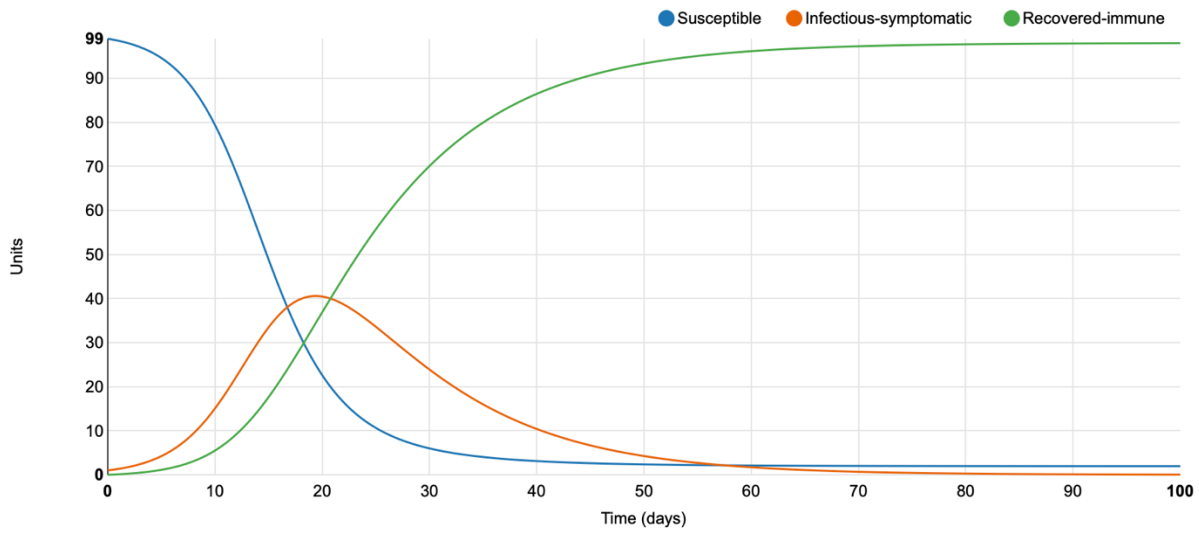

FIGURE 2. SIR model behavior where $\Delta = 0.1$. Y-axis indicates population count and x axis indicates time in days. Created with Epidemix (see Muellner et al., 2018).

SIR Model Assumptions

It should be noted that the SIR model will not accurately model all vector-borne diseases. For simplicity, this model combines both infected hosts and those infected but not infectious into one compartment. As will be expanded upon below, more complex models separate the two categories (known as the SEIR model). Further, this model assigns lifelong immunity to those recovered (i.e., removed) from the disease. More complex models may need to separate whether hosts within this compartment live, die, fully recover and receive life-long immunity, or reintegrate back into the susceptible compartment (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Sameni, 2020).

The SIR model also makes the following assumptions:

1. The population is fixed and closed.
2. The disease is contagious and transfers via contact
3. The only way a person can leave a group is to change infection status. To leave the susceptible group, a host must become infected/infectious. Once recovered, the individual is removed from the model (either through death or by receiving lifelong immunity).
4. Age, sex, social status, race, access to healthcare, geographical location, and seasonality, do not affect the probability of being infected.
5. There is no inherited immunity.
6. There is no disease control mechanism or vaccine.
7. The members of the population mix homogeneously (i.e., have the same number of interactions with each other and are distributed homogeneously in space).

Reproductive Ratio

For any real outbreak, the number of initially infected hosts (I_0) is usually not known, nor is the exact time the outbreak starts. Thus, the infection's growth rate may need to be estimated by obtaining two values of I at two time points during the initial growth rate (i.e., I_1 at t_1 and I_2 at t_2) and the solve for I_0.

To determine the number of new infections that would result from the introduction of a single infectious individual into an entirely susceptible population, public health officials use the Reproductive Ratio, r_0. Note that standard notation for Reproductive Ratio and initial Removed compartment both utilize R_0. For clarity, r_0 will be used henceforth to refer to Reproductive Ratio.

The Reproductive Ratio helps illustrate whether a particular population is at risk of a disease epidemic (e.g., Heffernan et al, 2005; Roberts, 2007). More specifically, the reproductive number is defined as the average number of new infected (and infectious) individuals caused by one infectious individual. The basic reproductive number is the reproductive number in a scenario where everyone is susceptible. For the SIR model shown above, this quantity is:

$$r_0 = \frac{bS_0}{g} \qquad (16)$$

where S_0 is the initial number of susceptibles.

Equation (16) is clearly affected by both infection (b) and recovery (g) rates. When. $r_0 > 1$, the occurrence of the disease will increase. Here, I(t) increases, reaches a maximum, and then decreases to zero as $t \to \infty$. When $r_0 < 1$, the occurrence of the disease will decrease and eventually die out. Here, I(t) decreases monotonically to zero as $t \to \infty$. However, when $r_0 = 1$, the disease occurrence will remain constant. We can see that values of r_0 close to 1 produce very

slow-growing epidemics, while values of r_0 much greater than one produce fast, explosive epidemics.

From Deterministic to Stochastic Modeling

The mathematical SIR modeling of disease outbreak is largely done deterministically, by means of systems of deterministic ordinary differential equations (ODE), partial differential equation (PDE), or delay differential equation (DDE; e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al., 2014, Ndanguza et al., 2016). Deterministically, the flow between compartments is represented as rates. This makes the ODEs simplistic to analyze and capable of explaining the change or evolution of the system, identifying points where change occurs, and illustrating the effect of the starting point to the initial solution. However, such modeling does not take into consideration the naturally stochastic processes governing the demography nor the inherently stochastic nature of disease outbreaks.

The outbreak of a contagious disease in a large population is a stochastic event (e.g., Costa, 2017; Ndanguza et al., 2016; Sameni, 2020). Starting from a single infected individual, the infection is transmitted to others in a stochastic manner, either by direct contact, proximity, or environmental traces (infected objects left over in the environment). Then, the newly infected hosts probabilistically transmit the infection to encountered healthy individuals. During the primary stages of an epidemic outbreak, the encounters between health and infected individuals are statistically independent. Thus, the likelihood of multiple infected people meeting a single healthy individual is probabilistically low. As a result, assuming that each infected individual contaminates r_0 new people on average, if $r_0 > 1$, the disease spreads exponentially from one time-step to another (for example on a daily basis). However, in a finite population, exponential growth has a limit. Depending on the population size and contact patterns, the probability of

infected people encountering independent healthy individuals decreases. Therefore, after the initial outbreak that exponentially spreads among the population, the infected population tend to encounter each other and repeated healthy ones (the healthy individuals already contacted by another infected person). Hence, the stochastic model of infection propagation, somehow saturates.

To ensure that the outputs from the model can be interpreted robustly, it is vital to account for parameter uncertainty, as well as stochasticity arising from the model dynamics (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al, 2014, Ndanguza et al., 2016; Sameni, 2020). To account for randomness or uncertainty in the model, stochastic processes formulate terms as probabilities at which an event occurs, over the index space of time, for a state space consisting of positive integers from 1 to the total population size N. Index and state spaces can be continuous or discrete. Stochastic processes thus define the probability of disease transmission as state transitions governed by sets of probability equations. Hence, multiple realizations of the system will result in a distribution of outcomes, even for a fixed set of parameter values (i.e. with no parameter uncertainty).

Stochastic models converge to deterministic models when the population size becomes large (e.g., Ndanguza et al., 2016). However, some phenomena do not satisfy the law of large numbers and, therefore, are genuinely stochastic. This approach more robustly models the beginning phase of an outbreak, because the number of infectious individuals is small and includes factors for uncertainty in the estimation of parameters and states (as they are equipped with standard error). It also more robustly models the end of an outbreak (i.e., the extinction of an outbreak), which occurs when the epidemic process deviates from the expected level.

The probabilistic models used for modeling such disease outbreaks regularly employ Poisson distributions for the probability of contact between infectious and healthy subjects. Let x(*t*) denote the number of infected individuals of a population at time *t*. Assume that the likelihood of infection increases with the number of infected hosts. Then, variations in the population of the number of infected between time *t* and $t + \Delta$ (over relatively small intervals) is proportional to the number of infected individuals, i.e.,

$$\frac{dx(t)}{dt} \approx \frac{x(t+\Delta) - x(t)}{\Delta} = \phi(t)x(t) \qquad (17)$$

where $\phi(t)$ is the reproduction function, which models how the infected population evolves over time and accounts various probabilistic factors including the rate of infection transmission, population density, and contact patterns. Denoting the k^{th} generation of the infection spread by $x_k \triangleq x(k\Delta)$, (17) can be discretized as follows:

$$x_{k+1} = [1 + \Delta\phi(k\Delta)]x_k$$

If we define the reproduction number as $r_k \triangleq [1 + \Delta\phi(k\Delta)]$, it is evident that the population at the discretized time index *k* can be recursively found from the initial condition x_0:

$$x_k = (r_{k-1}r_{k-2}\cdots r_0)x_0$$

We can see that, if for all *k*, $r_k < 1$ (or equivalently $\phi(t) < 0$), the infection would decay to zero. Otherwise, if $r_k > 1$ or equivalently $\phi(t) > 0$), the infection spreads. If the reproduction function is constant ($\phi(t) = \lambda$), we have a constant reproduction number ($r_0 = 1 + \lambda\Delta$), resulting in an exponential growth or decay:

$$x_k = x_0 r_0^k$$

More generally, the reproduction function $\phi(t)$ (or r_0 in the discrete case) can be a time-varying function of factors such as the total susceptible population, the population of the exposed individuals (carriers of the disease but without symptoms), contact patterns, and countermeasures

such as social distancing and lockdowns. The notion of reproduction function (number) and its impact on epidemic outbreak generalizes to eigen-analysis of vector valued dynamic epidemic models (when the population is divided into compartments), enabling the stability analysis of such models.

Compartmental Models Via SDEs

Continuous-time Markov chains are the basic tool for building discrete population epidemic models. The Markov property, which states that the future evolution of the process depends only on the current state, lets us specify a model by giving the transition probabilities on small intervals together with initial conditions. For the SIR model in a closed population, we have:

$$P[N_{SI}(t+\delta) = N_{SI}(t) + 1] = b_{SI}(t)S(t)\delta + o(\delta)$$

$$P[N_{SI}(t+\delta) = N_{SI}(t)] = 1 - b_{SI}(t)S(t) + o(\delta)$$

$$P[N_{IR}(t+\delta) = N_{IR}(t) + 1] = g_{SI}(t)I(t) + o(\delta)$$

$$P[N_{IR}(t+\delta) = N_{IR}(t)] = 1 - g_{SI}(t)I(t) + o(\delta)$$

Deterministically, the Euler method approach $\left(\frac{dx}{dt} = h(x,t)\right)$ initializes the numerical solution at the known starting value $\tilde{x}(0) = x(0)$. Then, for $k = 1, 2, \dots, n$, it is assumed that the gradient $\frac{dx}{dt}$ is approximately constant over the small time interval $k\delta \le t \le (k+1)\delta$. Therefore, $\tilde{x}\big((k+1)\delta\big) = \tilde{x}(k\delta) + \delta h(\tilde{x}(k\delta), k\delta)$. This defines $\tilde{x}(t)$ for those *t* that are multiples of δ. But, we can suppose instead that $\tilde{x}(t)$ is constant between these discrete times. We look for a numerical solution with state variables $\tilde{S}(k,\delta), \tilde{I}(k\delta), \tilde{R}(k\delta)$. The counting process for the flows between compartments are $\widetilde{N}_{SI}(t)$ and $\widetilde{N}_{IR}(t)$. The counting processes are

related to the number of individuals in the compartments by the same flow equations we had before:

$$\Delta\tilde{S} = -\Delta\widetilde{N}_{SI}$$

$$\Delta\tilde{I} = \Delta\widetilde{N}_{SI} - \Delta\widetilde{N}_{IR}$$

$$\Delta\tilde{R} = \Delta\widetilde{N}_{IR}$$

The Euler method extends naturally to SDEs by:

$$\frac{dX}{dt} = h(X) + \sigma\frac{dB}{dt}$$

where $B(t)$ is Brownian motion and so $\frac{dB}{dt}$ is Gaussian white noise. The Euler-Maruyama approximation is given by:

$$\tilde{X}\big((k+1)\delta\big) = \tilde{X}(k\delta) + \delta h\left(\tilde{X}(k\delta)\right) + \sigma\sqrt{\delta}Z_k$$

where $Z_1, Z_2, \ldots$ is a sequence of independent standard normal random variables, i.e., $Z_k \sim Normal\ (0,1)$.

Deterministic SEIR Model

Although the SIR model has been studied extensively and applied to a wide range of diseases, it assumes that after the transmission of the disease to a susceptible individual, the individual becomes immediately infective. However, in most cases, initial transfer of bacterial cells or viruses are typically very small. Thus, for a period of time, the pathogen abundance is too low in the tagged individual for active transmission to other susceptible individuals. This is known as a latent period (or incubation). During the latent period, an individual is infected (i.e., exposed to the virus) but not yet infective. The Susceptible-Exposed-Infective-Removed (SEIR) model more accurately represents disease transmission than the corresponding SIR model due to the inclusion of the latent period (i.e., the exposed compartment; e.g., Costa, 2017; Ndanguza et

al., 2016). The SEIR model has a slower growth rate, since after the pathogen invasion, susceptible individuals need to pass through an incubation period (i.e., the exposed compartment) before continuing through the transmission processes.

Similar to the SIR model, when N is large, the population can be represented by a deterministic SEIR model of ODEs, where the independent variable is time and the dependent variable is the number of individuals in the various compartments S, E, I, and R. More accurately, this deterministic model is an appropriate limit of the corresponding stochastic version expanded upon below.

Model Overview

Notation

$S = S(t) = \#\ of\ people\ in\ population\ susceptible\ to\ the\ disease\ at\ time\ t$

$E = E(t) = \#\ of\ people\ in\ population\ exposed/infected\ by\ the\ disease\ at\ time\ t$

$I = I(t) = \#\ of\ people\ in\ population\ who\ are\ infectious\ \ at\ time\ t$

$R = R(t) = \#\ of\ people\ in\ population\ \ recovered\ from\ the\ disease\ at\ time\ t$

There are a number of parameters which will need to be either modeled or estimated from the data. It is assumed that these parameters are time invariant though more sophisticated efforts and information could produce time– varying models. A description of these parameters is listed below.

Parameters

$P(I) = Probability\ of\ disease\ transmission$

$\beta = Transmission\ rate\ per\ person$

$\beta \frac{I(t)}{N} = Conversion\ rate\ for\ susceptible\ to\ exposed\ (aka\ Force\ of\ Infection)$

$k = Number\ of\ days\ during\ the\ incubation\ period$

$\gamma = Recovery/death\ rate$

Disease Dynamics

Consider a closed population, where birth and death rates are ignored. Suppose that the total effective population size N is divided into four compartments, such that at each time $t \geq 0$, every individual in the population is either susceptible, exposed, infected, or recovered/removed (see Figure 3). The epidemic starts at $t = 0$ in a specified state, often the state with one infectious individual, called the index case and thought of as being externally infected, and the rest being susceptible (e.g.,Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Costa et al., 2005; Kennedy et al., 2014; Ndanguza et al., 2016; Sameni, 2020).

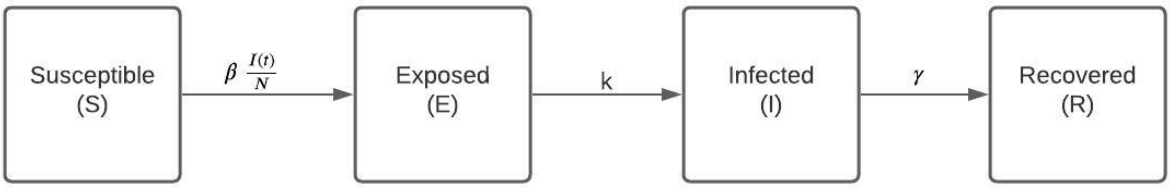

FIGURE 3. SEIR Model in homogeneous, closed population.

In the static models, the rate of infection (i.e., transition probability) is fixed; whereas in dynamic models, the rate of infection depends on the number of people already infected. Here, susceptible individuals in compartment S migrate into compartment E after exposure to an infected individual at the exposure rate. In epidemiology, the exposure rate is known as the exposure rate is the force of infection, which is equivalent to a transition probability, is not a constant as it is in decision trees or other Markov models. Instead, it is defined as $\beta \frac{I(t)}{N}$, where β is the transmission rate per person per day, and $\frac{I(t)}{N}$ is the number of infected individuals in the total population. This is a realistic way of modeling the chances a person is infected. The larger

the proportion infected $\frac{I(t)}{N}$, the more likely it is people can become infected. β, the transmission rate, is a function of the rate of contact and the probability of transmission given contact. During an outbreak, social policies such as lockdowns, "social distancing" (which should be referred to as physical distancing), and mask-wearing are attempts to lower either the rate of contact or the probability of transmission (therefore, lowering β). In this simple model, we capture all these features with the parameter β, but we could define $\beta = rc \cdot p(T)$

Exposed individuals undergo an average incubation period of k days before progressing to the infectious compartment (I). k indicates the incubation rate, or the rate at which exposed people become infectious. This latency or incubation period delays the initial spread of the disease Infectious individuals remain infectious for a specified amount of time, governed by an infection's period distribution after which they stop being infectious and recover. The infectious periods are defined to be independent and identically distributed (also independent of the contact processes).

Infectious individuals move to the recovered compartment (R) at the per-capita rate γ. Recovery, or removal, is carried out either through isolation from the rest of the population, immunization against infection, recovery from the disease with full immunity against reinfection, or death caused by the disease. These characterizations of recovered members are different from an epidemiological perspective but are often equivalent from a modeling point of view which takes into account only the state of an individual with respect to the disease (e.g., Chowell, et al., 2009).

Since we assume that the time-scale of the epidemic is much faster than characteristic times for demographic processes (natural birth and death), background demographic processes are not included in the model.

The Model

To examine disease transition over time, we model the transition probabilities, understanding that the change in the number of infected (I) per unit of time depends on the susceptible becoming exposed and infected and the infected recovering. The transition process in modeled by the following system of nonlinear ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N}$$

$$\frac{dE(t)}{dt} = \beta \frac{S(t)I(t)}{N} - kE(t)$$

$$\frac{dI(t)}{dt} = kE(t) - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \qquad (18)$$

where $S(t), E(t), I(t),$ and $R(t)$ denote the number of susceptible, exposed, infected, and removed at time *t*, respectively. The population size, $N(t)$ is constant, given by:

$$S(t) + I(t) + R(t) = N \Rightarrow \frac{dN}{dt} = 0 \qquad (19)$$

As we are dealing with a closed population, there is no need of recording all compartment's occupancy level. We can also see that from *(19)*, $R = N - S - E - I$ because $N = S + E + 1 + R$. Therefore, the number of recovered individuals is not reported and the deterministic SEIR model (18) becomes:

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N}$$

$$\frac{dE(t)}{dt} = \beta \frac{S(t)I(t)}{N} - kE(t)$$

$$\frac{dI(t)}{dt} = kE(t) - \gamma I(t) \quad (20)$$

Since the model is Markovian all these events happen at rates depending only on the current state. When an infection occurs, the number of susceptibles decreases by 1 and the number of exposed increases by 1; when a latency period ends, the number of exposed decreases by 1 and the number of infectives increases by 1; and finally when there is a recovery, the number of infectives decreases by 1 and the number of recovered increases by 1. Thus, if we look at "proportions", the corresponding changes are $-1 = N$ and $+1 = N$, justifying the use of a deterministic model.

Due to the fact that the exposed and infected compartments are random, however, we can approximate a Markovian SEIR model.

Assumptions

The full deterministic SEIR model described above operate under the following simplifying assumptions:

1. The population is fixed, homogenous, and closed.
2. Sufficiently short time-scale: $t \in [t_0, t_0 + \Delta t]$ where Δt is 3-4 weeks.
3. At $t = 0$, there are no individuals in the recovered compartment, i.e., $R(t_0) = 0$
4. For $t \in [t_0, t_0 + \Delta t], S(t) = S(t_0) = S_0$
5. The only way a person can leave a group is to change infection status.
6. Age, sex, social status, race, access to healthcare, geographical location, and seasonality, do not affect the probability of being infected.

7. There is no inherited immunity and there is insufficient time for R (recovered individuals) to return to the population of susceptible individuals. Thus, an individual reaching compartment R will never come back into the system.
8. There is no disease control mechanism or vaccine.
9. The members of the population mix homogeneously (i.e., have the same number of interactions with each other and are distributed homogeneously in space).
10. Parameters $\beta, k,$ and γ do not vary with respect to time.

As mentioned above, noises have crucial effects in epidemic models. Therefore, it's essential to change this system of ODEs to a system of SDEs.

Model Behavior

As seen in Figure 4, the variable S is monotonically decreasing, whereas E and I are increasing and decreasing after a period of time. This is due to recruitment and migrations from those compartments. The compartment R is monotonically increasing because all individuals reaching this state remain within it. The infectious period is exponentially distributed with parameter γ. The system of ODE equations **Error! Reference source not found.** will be used for all further studies and analysis.

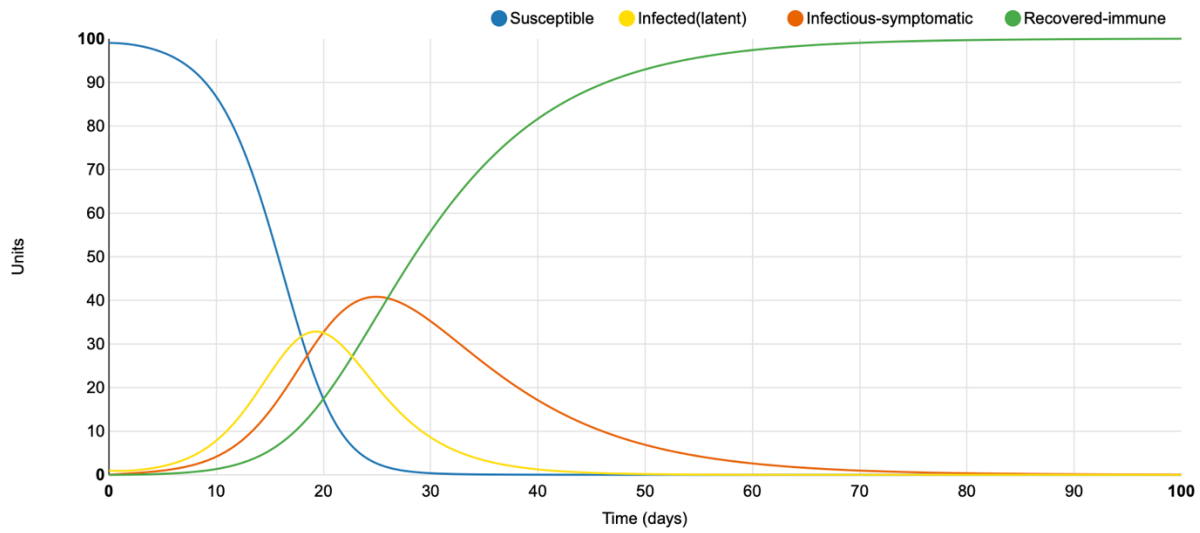

FIGURE 4. SEIR epidemic model. The susceptible variable is decreasing since some of its candidates are immigrating to E. By this time, E and I are increasing and decreasing after a given period. R is increasing exponentially. Created with Epidemix (see Muellner et al., 2018).

Reproductive Number & Final Size

The deterministic SEIR and SIR share the two most important properties in that they have the same basic reproduction number r_0 and give the same final size (assuming the initial number of infectives/exposed are positive but negligible in both cases; See Figure 5 below). In both SEIR and SIR models, $r_0 = \beta/\gamma$. Or, $r_0 = \beta \cdot \frac{1}{\gamma}$. Here, we can interpret β has the rate at which a person is infected (transmission rate) and $\frac{1}{\gamma}$ as the length of time a person is in the infected state.

The basic reproductive number, r_0 is defined as the "average number of secondary cases arising from a typical primary case in an entirely susceptible population" (Blackwood & Childs, 2018). Generally speaking, it is a model number used to understand how contagious a disease is. One might think of r_0 as the average (or expected) number of cases generated by one case, or the average number of individuals an infected person infects. The r_0 depends both on the force of transmission and the recovery rate. Intuitively, the higher the recovery rate, the less time a person

stays infected transmitted the disease. The higher the force of infection, the higher the r_0. Interestingly, r_0 is not a fixed number. The number of people (on average) an infected person infects depends on the features of the disease and the behavior of people. Specifically, the rate of transmission depends on the probability of transmission after contact (a feature of the disease) and the contact rate (behavior driven): $\beta = cr \cdot p(T)$. Clearly, the relationship between how quickly a person is infected (contact rate, probability of transmission given contact) and infection duration are key to determine the dynamics of the model. If $r_0 < 1$, the disease stops. If $r_0 = 1$, the disease becomes endemic (infections change steadily). If $r_0 > 1$, the disease will continue to spread unless something structurally changes to lower r_0.

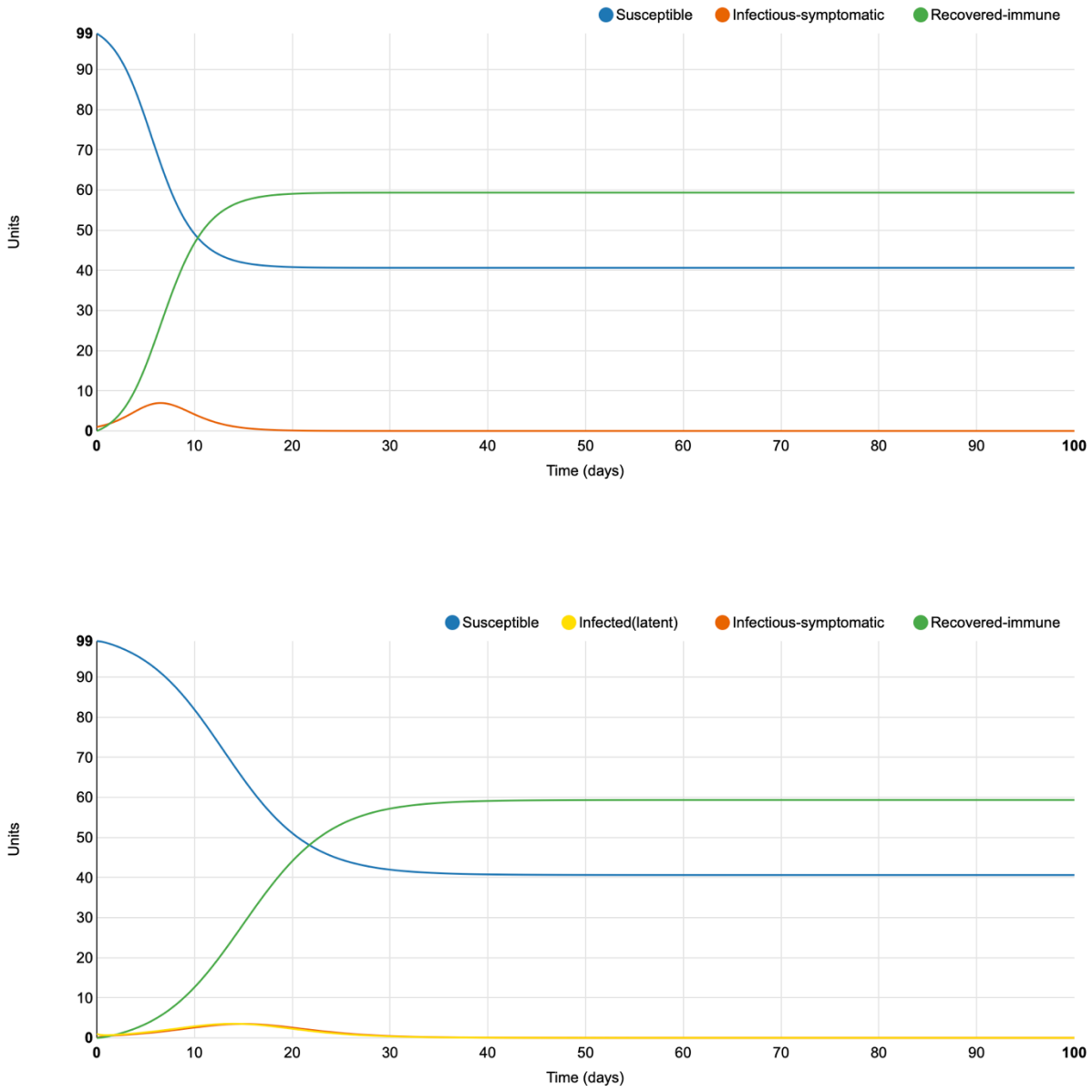

FIGURE 5. Comparison between SIR (top) and SEIR (bottom) with similar parameters, $\beta = 1.5$, $\gamma = 1$, and SEIR k = 1. In both models we see the susceptible variable decreasing as its candidates immigrate to E (or I). As this occurs, E and I are increasing for a period. In both models, R increases exponentially. The largest difference between the two is that the SEIR model (bottom) has a longer delay (due to the incubation period). Created with Epidemix (see Muellner et al., 2018).

Deterministic SEIR models are often used in the case of sufficiently large populations (i.e., in the thermodynamic limit). Larger populations are much more likely to meet the

assumption of homogeneity. Therefore, they can be viewed as limit models of a stochastic model class. However, they still lack the ability to account for uncertainty within the mode.

Stochastic SEIR Model

Compartmental Model Overview

The stochastic SEIR model adopts the same general model dynamics as the deterministic SEIR model *(19)*. We consider a closed population, where birth and death rates are ignored. Suppose that the total effective population size N is divided into four compartments, such that at each time $t \geq 0$, every individual in the population is either susceptible, exposed, infected, or removed. Susceptible individuals in compartment S migrate into compartment E after exposure to an infected individual. The rate at which this occurs is proportional to the number of contacts *c* with the infectious population I(*t*) times the probability of disease transmission per contact β times the proportion of the population which is infectious: $P(I) = \beta I(t)$. Contacts are mutually independent. We assume homogeneous mixing (i.e. random mixing) between individuals and, therefore, the fraction I(*t*)/N is the probability of a random contact with an infectious individual in a population of size N.

A stochastic process is said to be a Markov process or have the Markov property if the conditional probability of the future state conditional on both present and past states depends only on the present state of the process. In stochastic Markov models, parameters are also given a probability distribution. The incubation period for the exposed individuals are assumed independent and identically distributed according to an exponential distribution *k*. Exposure to the disease occurs when contacts with any given infected individual at the time points of a time homogenous Poisson process with intensity $\beta I(t)/N$. If a contacted individual is still susceptible, then (s)he becomes exposed. Exposed individuals are unable to infect until after the

incubation period is completed, at which time the individual becomes infective. After the completion of the infectious period, the individual is recovered (or, removed) and plays no further role in the epidemic spread. Recovery has an exponential distribution with parameter γ. All Poisson processes, incubation, and infectious periods are assumed independent. As a result, the stochastic SEIR model is a continuous-time 3-dimensional Markov chain $X = \{S(t), E(t), I(t): t \geq 0\}$, that records the number of susceptible, exposed, and infective individuals at any time point. The number of recovered individuals can be obtained indirectly by $R(t) = N - S(t) - E(t) - I(t)$.

The epidemic starts at $t = 0$ in a specified state, often the state with one infectious individual, called the index case and thought of as being externally infected, and the rest being susceptible $\big(S(0), E(0), I(0), R(0)\big) = (N, 0, 1, 0)$ (e.g., Artalejo, et al., 2015; Britton & Pardoux, 2019). Then, we can easily see that the state space of X is $S = S_1 - S_2$, where $S_1 = \{(s, e, i): 0 \leq s, e, i \leq N, S + e + i \leq N\}$ and $S_2 = \{(s, e, 0): 0 \leq s, e \leq N, s + e = N\}$. The infection ends when $E(t) = I(t) = 0$. Thus, the set of the absorbing states is $S_A = \{(s, 0,0): 0 \leq s \leq N - 1\}$. Since the set of transient states $S_T = S - S_A$ is a finite set, absorption occurs in a finite time with probability 1.

Reproductive Number

The reproductive number maintains its same general definition as described in the deterministic SEIR model. Specifically, $r_0 = \beta/\gamma$. Or, $r_0 = \beta \cdot \frac{1}{\gamma}$. Here, we can interpret β has the rate at which a person is infected (transmission rate) and $\frac{1}{\gamma}$ as the length of time a person is in the infected state. Sometimes, the rate β of having infectious contacts is replaced by an overall rate of contact c multiplied by the probability p of contact leading to infection. So, $\beta = cp$ and $r_0 = cpi$ (e.g., Britton & Pardoux, 2019).

Research may also focus on the probability for a given susceptible to escape infection. The instantaneous infectious force from the infective to a susceptible is β/N and the random duration of the infectious period is I. Conditional upon $I = x$, the escape probability is then $e^{-(\beta/N)x}$. The unconditional escape probability is then

$$\mathbb{P}(escape) = \mathbb{E}\left(e^{-\beta I/N}\right) = \psi_1(-\beta/N)$$

where $\psi_1(b) = \mathbb{E}(e^{bI})$ is the moment generating function of the infectious period. So, $\psi(-b)$ is the Lapace transform.

Mathematical Modeling

An Ito SDE, where Lipschitz and growth conditions are satisfied to ensure the existence and uniqueness of strong solutions, can be written as:

$$dx(t) = \boldsymbol{f}(x(t), t, \boldsymbol{\theta})dt + \boldsymbol{L}(\boldsymbol{x}(t), t, \boldsymbol{\theta})d\boldsymbol{B}(t) \quad (21)$$

The drift vector function $\theta \in \Phi \subseteq \mathbb{R}^d$ is the vector of parameters to be estimated, $\boldsymbol{L}$: $\mathbb{R}^n \cdot [0, \infty] \cdot \boldsymbol{\Phi} \to \mathbb{R}^{n \cdot m}$ is a diffusion function, and is m-dimension Brownian motion with diffusion matrix $\boldsymbol{Q}_c \in \mathbb{R}^{m \cdot m}$. If $\boldsymbol{B}(t) = 0$, the SDE becomes the normal ordinary differential equation.

The stochastic SEIR model presented will use stochastic processes with continuous index space and discrete state space. The Euler-Maruyama discretization scheme will be used to approximate (21) as:

$$x(t) = x(t_{k-1}) + f(x(t_{k-1}), t_{k-1}, \theta)\Delta t + L(x(t_{k-1}), t_{k-1}, \theta)\Delta B(t_{k-1}) \quad (22)$$

where $\Delta t = t_k - t_{k-1}$.

Ito SDEs are derived from a system of ordinary differential equations *(19)*. Let $x_i, i = 1, 2, 3$ denote the random variables for SEI, respectively. For a small time interval Δt, there are 3 possibilities in changes in the vector $X = [x_1, x_2, x_3]^T$. The transition probabilities are:

$$P(s+k, i+j, l+m) = f(x) = \begin{cases} \frac{\beta SI}{N}\Delta t, & (k,j,m) = (-1,1,0) \\ kE\Delta t, & (k,j,m) = (0,-1,1) \\ \gamma I\Delta t, & (k,j,m) = (0,0,-1) \\ 1 - \left(\frac{\beta SI}{N} + kE + \gamma I\right)\Delta t, & (k,j,m) = (-1,1,0) \\ 0, & otherwise \end{cases}$$

Compartmental changes in a small time period Δt are calculated as follows:

The transition $(\Delta \boldsymbol{x})_1 = [-1,1,0]^T$ yields the probability $p_1 = \left(\frac{\beta x_1 x_3}{N}\right)\Delta t$

The transition $(\Delta \boldsymbol{x})_2 = [0,-1,1]^T$ yields the probability $p_1 = kx_2\Delta t$

The transition $(\Delta \boldsymbol{x})_2 = [0,0,-1]^T$ yields the probability $p_1 = \gamma x_3\Delta t$

The ODE system (4) is converted into the SDE in the form: (needs #)

$$d\boldsymbol{x}(t) = F\big(t, \boldsymbol{x}(t)\big)dt + G\big(t, \boldsymbol{x}(t)\big)dW(t)$$

where W_i are white noises and Winner processes, $B_t dt = dW_t$ is a stochastic processes, and B_t is Brownian motion.

Then,

$$E(\Delta \boldsymbol{x}) = \sum_{i=1}^{3} p_i(\Delta \boldsymbol{x}_1) + p_2(\Delta \boldsymbol{x}_2) + p_3(\Delta \boldsymbol{x}_3)$$

$$= p_1\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + p_2\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + p_3\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-\beta x_1 x_3}{N} \\ \frac{-\beta x_1 x_3}{N} \\ 0 \end{pmatrix}\Delta t + \begin{pmatrix} 0 \\ -kx_2 \\ kx_2 \end{pmatrix}\Delta t + \begin{pmatrix} 0 \\ 0 \\ -\gamma x_3 \end{pmatrix}\Delta t$$

$$= \begin{pmatrix} \frac{-\beta x_1 x_3}{N} \\ \frac{-\beta x_1 x_3}{N} - kx_2 \\ kx_{2-\gamma x_3} \end{pmatrix}\Delta t = F(x_1, x_2, x_3)\Delta t$$

The covariance matrix becomes:

$$E\big((\Delta\boldsymbol{x})(\Delta\boldsymbol{x})\big) = \sum_{i=1}^{3} p_i(\Delta x)_i(\Delta x)_i^T = p_1(\Delta\boldsymbol{x})_1(\Delta\boldsymbol{x})_1^T + p_2(\Delta x)_2(\Delta\boldsymbol{x})_2^T + p_3(\Delta x)_3(\Delta\boldsymbol{x})_3^T$$

$$= p_1\begin{pmatrix}-1\\1\\0\end{pmatrix}(-1\ \ 1\ 0) + p_2\begin{pmatrix}0\\-1\\1\end{pmatrix}(0\ -1\ 1) + p_3\begin{pmatrix}0\\0\\-1\end{pmatrix}(0\ 0\ -1)$$

$$= p_1\begin{pmatrix}1 & -1 & 0\\-1 & 1 & 0\\0 & 0 & 0\end{pmatrix} + p_2\begin{pmatrix}0 & 0 & 0\\0 & 1 & -1\\0 & -1 & 1\end{pmatrix} + p_3\begin{pmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{pmatrix}$$

$$= \begin{pmatrix}p_1 & -p_1 & 0\\-p_1 & p_1+p_2 & -p_2\\0 & -p_2 & p_2+p_3\end{pmatrix} = \begin{pmatrix}\frac{\beta x_1 x_3}{N} & -\frac{\beta x_1 x_3}{N} & 0\\-\frac{\beta x_1 x_3}{N} & -\frac{\beta x_1 x_3}{N} + kx_2 & -kx_2\\0 & -kx_2 & kx_2 + \gamma x_3\end{pmatrix}\Delta t$$

$$= V(x_1, x_2, x_3)\Delta t$$

Thus, the SDE model has the form:

$$d\boldsymbol{X}(t) = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})dt + \big(\boldsymbol{V}(X, \boldsymbol{\theta})\big)^{1/2} d\boldsymbol{W}(t) \qquad (23)$$

An optimal estimate of the model predicted/measurement corrected state of a disease outbreak can be obtained via the Kalman filter. The discussion is, hereafter, framed in the context of Kalman filters.

Parameter Estimation

As mentioned above, infectious diseases are notoriously difficult to predict and manage. While anticipating future behavior of an outbreak and understanding how interventions will affect the spread of the disease are paramount for policy makers, disease dynamics are largely unknown for emerging pathogens such as COVID-19 or SARS. Dynamic models can project possible epidemic evolutions and assess tradeoffs between interventions. These models are traditionally parameterized and calibrated when they are constructed, even though underlying

parameters are dependent on unpredictable factors. Statistical filtering and estimation methods for dynamic models, such as sequential Monte Carlo (SMC) method and Markov chain Monte Carlo (MCMC) methods, can be used to estimate model parameters as information becomes available. The Kalman Filter can be used to adapt a model to better approximate an epidemic outbreak even when the parameters in question are not empirically or logically observable (e.g., Chowell, 2009).

State Space Models & Kalman Filtering

To understand Kalman filters, we must first review a few basics regarding state space models. State space models attempt to describe phenomena with two characteristics (e.g., Brauer, et al., 2019; Britton & Pardoux, 2019; Chowell, et al., 2009; Costa, 2017; Kennedy et al., 2014; Peng, 2020). First, there's an underlying system that has time-varying dynamic relationships. Here, the state of the system at time *t* is related to the state of the system at time *t*-1. Second, we cannot observe the true underlying state of the system but rather we observe a noisy version of the system. If we suppose there's an initial state $x_0 \sim N(x_0^0, P_0^0)$, for $t = 1, 2, \ldots$, we want to estimate the subsequent states $x_1, x_2, \ldots, x_n$. At each time point, we observe some data y_t in order to incorporate it into our estimation of x_t. Then, we have an observation equation: $y_t = Ax_t + v_t$, where $v_t \sim N(0, \sigma^2)$. And an state equation: $x_t = \theta x_{t-1} + w_t$, where $w_t \sim N(0, \tau^2)$. The parameters θ, σ, τ are assumed to be known (like tuning parameters) and the goal is to produce an estimate of x_t for all *t* of interest.

Although this will be expanded upon below, the basic one-dimensional Kalman filter will compute estimates for x_1 and P_1 given the current state as:

$$x_1^0 = \theta x_0^0$$

$$P_1^0 = \theta^2 P_0^0 + \tau^2$$

With a new observation y_1, the estimate can be updated:

$$x_1^1 = x_1^0 + K_1(y_1 - x_1^0)$$

$$P_1^1 = (1 - K_1)P_1^0$$

Where $K_1 = P_1^0/(P_1^0 + \sigma^2)$. Then, the new estimate x_t combined with the current state x_{t-1}^{t-1} and variance P_{t-1}^{t-1} allows for a prediction:

$$x_t^{t-1} = \theta x_{t-1}^{t-1}$$

$$P_t^{t-1} = \theta^2 P_{t-1}^{t-1} + \tau^r$$

Given new information y_t, the estimate is updated:

$$x_t^t = x_t^{t-1} + K_t(y_t - x_t^{t-1})$$

$$P_t^t = (1 - K_t)P_t^{t-1}$$

Where

$$K_t = \frac{P_t^{t-1}}{P_t^{t-1} + \sigma^2}$$

Is the Kalman gain coefficient. Clearly, when noise is high, σ^2 is large, K_t approaches 0, and y_t has a small influence. On the other hand, if σ^2 is small, then the filtered value x_t^t will adjust more in the direction of y_t. An overview of the process can be found in Figure 6 below.

FIGURE 6. Kalman Filter process. Image obtained from Teow (2018).

Implementing a Kalman Filter

The Kalman Filter is a method that computes the posterior states based on the Maximum Likelihood of a linear Gaussian dynamical system (e.g., Costa et al., 2005; Evensen, 2009; Labbe, 2020; Ndanguza et al., 2016; Rhudy, et al., 2017; Welch & Bishop, 2006). The ordinary Kalman filter originally proposed by Kalman in 1960, obtains an estimator of a conditional mean of a state vector, given a series of observations, and is obtained as the best linear minimum mean square error estimate. The ordinary Kalman filter has been used by many to describe linear models, such as dynamic regression and nonhomogenous autoregressive processes (e.g., Soyer, 2016). Once the observation and state error distributions are specified as Gaussian, the Kalman filter equations can be obtained by using standard Bayesian prior to posterior updates. Not only is it applicable for parameter estimation, but it can also be used for forecasting distributions.

Here, however, we have unknown parameters β, k, γ in nonlinear SDE. Therefore, an extended Kalman filter (eKF) must be utilized in a manner similar to that explored by Mbalawata et al. (2012). Thus, we must consider a more general nonlinear set of eKF equations:

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}) + \boldsymbol{w}_{k-1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{v}_k$$

Where $\boldsymbol{f}$ is the vector-valued nonlinear state transition function and $\boldsymbol{h}$ is the vector-valued, nonlinear observation or output function. To handle the nonlinear functions, Jacobian matrices are calculated in order to linearize the equations at each time step about the previous state:

$$\boldsymbol{F}_k = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right|_{\widehat{\boldsymbol{X}}_{k-1}}$$

$$\boldsymbol{H}_k = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\widehat{\boldsymbol{X}}_{k|k-1}}$$

Then, the eKF becomes:

$$\widehat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}(\widehat{\boldsymbol{x}}_{k-1}, \boldsymbol{u}_{k-1})$$

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1}\boldsymbol{F}_{k-1}^T + \boldsymbol{Q}_{k-1}$$

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

$$\widehat{\boldsymbol{x}}_k = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_x\left[\boldsymbol{z}_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k-1}\right)\right]$$

$$\boldsymbol{P}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k|k-1}$$

In (19), however, measurement noises are additive. Therefore, a nonlinear continuous-discrete eKF should be employed as a Taylor series expansion based approximation of the general Bayesian continuous-discrete filter (e.g., Nganguza et al., 2016):

$$d\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t), t, \boldsymbol{\theta})dt + \boldsymbol{L}(\boldsymbol{x}(t), t, \boldsymbol{\theta})d\boldsymbol{B}(t)$$

$$\boldsymbol{y}_t = \boldsymbol{h}\left(\boldsymbol{x}(t_k)\right) + \boldsymbol{r}_k$$

where $\boldsymbol{x}(t_k)$ is the state at time $t_k, \theta \in \boldsymbol{\Phi} \subseteq \mathbb{R}^d$ is the vector of parameters to be estimated, $\boldsymbol{f}: \mathbb{R}^n \cdot [0,\infty] \cdot \boldsymbol{\Phi} \to \mathbb{R}^n$ is the dynamic model function, $\boldsymbol{L}: \mathbb{R}^n \cdot [0,\infty] \cdot \boldsymbol{\Phi} \to \mathbb{R}^{n \cdot 5}$ is the matrix valued function, $t \mapsto \boldsymbol{B}(t)$ is s-dimension Brownian motion with diffusion matrix, $\boldsymbol{Q}_c \in R^{m \cdot m}$, $\boldsymbol{y}_k \in R^m$is the measurement at time t_k, $\boldsymbol{h}: \mathbb{R}^n \mapsto \mathbb{R}^m$ is the measurement model function, $r_k \sim \mathbb{N}(0, \boldsymbol{R}_k)$ is the Gaussian measurement noise with $\boldsymbol{R}_k \in \mathbb{R}^{m \cdot m}$ being the covariance matrix of the measurement error at t_k. At time t_0, the state is assumed to have the prior distribution $p\big(\boldsymbol{x}(t_0)\big) = \mathbb{N}(\boldsymbol{x}(t_0)|\boldsymbol{m}_0, \boldsymbol{P}_0)$, where $\boldsymbol{m}_0$ is the predictive initial mean and $\boldsymbol{P}_0$ is the predictive initial covariance.

Prediction Step

$$\frac{d\boldsymbol{m}_k^-(t)}{dt} = \boldsymbol{f}(\boldsymbol{m}_k^-(t), t, \boldsymbol{\theta})$$

$$\frac{d\boldsymbol{P}_k^-(t)}{dt} = \boldsymbol{F}_x(\boldsymbol{m}_k^-(t), t, \boldsymbol{\theta})\boldsymbol{P}_k^-(t) + \boldsymbol{P}_k^-(t)\boldsymbol{F}_x^T(\boldsymbol{m}_k^-(t), t, \boldsymbol{\theta}) + \sum(\boldsymbol{m}_k^-(t), t, \boldsymbol{\theta})$$

Where $\sum(\boldsymbol{m}_k^-(t), t, \boldsymbol{\theta}) = \boldsymbol{L}(\boldsymbol{m}(t), t, \boldsymbol{\theta})\boldsymbol{Q}_c\boldsymbol{L}^T(\boldsymbol{m}, t, \boldsymbol{\theta})$ and $\boldsymbol{F}_x(\boldsymbol{x}, t, \boldsymbol{\theta})$ is the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}, t)$ with respect to **x**. The initial conditions are $\boldsymbol{m}_k^-(t_{k-1}) = m_{k-1}, \boldsymbol{P}_k^-(t_{k-1}) = P_{k-1}$, and the prediction result is given as $\boldsymbol{m}_k^- \triangleq \boldsymbol{m}_k^-(t_k), \boldsymbol{P}_k^- \triangleq \boldsymbol{P}_k^-(t_k)$

Update Step

$$\mu_k = \boldsymbol{h}(\boldsymbol{m}_k^-, t)$$

$$S_k = H_k(\boldsymbol{m}_m^-, t)\boldsymbol{P}_k^- H_x^T(\boldsymbol{m}_k^-, t) + \boldsymbol{R}_k$$

$$K_k = \boldsymbol{P}_k^-\boldsymbol{H}_x^T(\boldsymbol{m}_k^-, t)\boldsymbol{S}_k^{-1}$$

$$\boldsymbol{m}_k = \boldsymbol{m}_k^- + \boldsymbol{K}_k(y_k - \mu_k)$$

$$\boldsymbol{P}_k = P_k^- - \boldsymbol{K}_k\boldsymbol{S}_k\boldsymbol{K}_K^T$$

Where $H_x(\boldsymbol{x}, t)$ is the Jacobian matrix of $\boldsymbol{h}(\boldsymbol{x}, t)$.

Posterior Function Approximation

We can use the Gaussian approximation of the posterior distribution from the continuous-discrete eKF:

$$\varphi(\boldsymbol{\theta}) = \sum_{k=1}^{N} \frac{1}{2} \ln|2\Pi S_k| + \frac{1}{2} \sum_{k=1}^{N} (y_k - \mu_k)^T S_k^{-1} (y_k - \mu_k) - \ln p(\theta)$$

Markov Chain Monte Carlo

Here, we draw samples $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(3)}$ from the posterior distribution $p(\theta|y_1, \ldots, y_M)$ and then approximate the expectation as the sample average:

$$E[g(\theta|y_1, \ldots, y_M)] \approx \frac{1}{N} \sum_{i=1}^{N} g\left(\theta^{(i)}\right)$$

Using a Metropolis-Hastings Algorithm:

Draw a candidate point $\theta^{(\theta)}$, from an initial distribution $0_o(\boldsymbol{\theta})$

For n = 0, 1, 2, …

Sample a candidate point θ^* from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^n)$

Accept the candidate point and set $\boldsymbol{\theta}^{(n+1)}\boldsymbol{\theta}^*$ with the probability

$$A\left(\theta^{(n)}, \theta^*\right) = min\left\{1, \frac{q(\theta^{(n)}, \theta^*)}{A(\theta^*, \theta^{(n)})} exp\left(\varphi\left(\theta^{(n)}\right) - \varphi(\theta^*)\right)\right\}$$

Generate $u \sim U(1,0)$ from Uniform distribution

Accept $\boldsymbol{\theta}^* \; if \; u \leq A\left(\boldsymbol{\theta}^{(n)}, \boldsymbol{\theta}^*\right)$

Previous Applications

There has been a wide range of Kalman filter applications within epidemiology. Cazelles and Chau (1995) used a Kalman filter to analyze HIV/AIDs in Paris, France. Gourieroux and Jasiak (2020) used Kalman filter to analyze COVID-19 virus propagation using the SIRs model. Hu et al. (2020) used Kalman filter integrated with geographically weighted regression to

estimate incidence of Hand, Foot, and Mouth disease. Jang et al. (2018) used an ensemble Kalman filter to explore Hepatitis B Viral infection during antiviral therapy. Kelemen et al. (2006) used a Kalman filter to remove white noise on microarray-based molecular diagnosis. Krishnamurthy et al. (2014) used ensemble Kalman filter and optimal statistical interpolation to retrospectively study the "Black Death" plague of 14th century Europe. Reis and Shaman (2016) used a Kalman filter on SIRs model of Respiratory Syncytial Virus. Sameni (2020) used multiple SIRs models to examine COVID-19. Similar research was conducted by Nkwayep et al. (2020). Yang et al (2015) used a temporal inference modified network model in conjunction with an ensemble adjustment Kalman filter to study the spatial-temporal progression of Ebola in Sierra Leone between 2014 and 2015. Yang et al. (2015) used an ensemble adjustment Kalman filter and a modified particle filter with a SIR model of Influenza in Hong Kong between 1998 to 2013. Similar research was conducted by Zhang (2016). However, very little research has applied the Kalman filter to a SEIRs model.

Conclusion

This paper has discussed the use of an extended Kalman filter (eKF) in a nonlinear, closed SEIR model in order to better approximate an epidemic outbreak even when the parameters in question are not empirically or logically observable. The eKF consists of two processes for updating state variables and process covariance estimates (which specify the degree of accuracy of model estimates): continuous time updates and discrete-time update. The continuous time updates govern the behavior of the aggregate SIER model and process covariance estimates between measurement points whereas the discrete-time updates modify system state and process covariance estimates with a consensus estimate derived from both measurement and model estimates.

It should be noted, however, that the Kalman filter does have a disadvantages when applied to epidemiological models. While the ordinary Kalman filter relies on linear models, the eKF relies on linearized versions of nonlinear models and on the maximum likelihood estimate. This may be problematic for some communicable diseases. Kalman filters also assume Gaussian distributed errors for both measurement error and process noise.

References

Antunes, A. C. L., Jensen, V. F., & Jensen, D. (2019). Unweaving tangled mortality and antibiotic consumption data to detect disease outbreaks – Peaks, growths, and foresight in swine production. *PLOS One, 14(10),* e0223250.

Blackwood, J. C., & Childs, L.M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical Models in Epidemiology*. Springer. New York, NY.

Britton, T., & Pardoux, E. (2019). *Stochastic Epidemic Models with Inference*. Springer. Cham, Switzerland.

Cazelles, B., & Chau, N. P. (1997). Using the Kalman Filter and Dynamic Models to Assess the Changing HIV/AIDS Epidemic. *Mathematical Biosciences, 140,* 131-154.

Chowell, G., Hyman, J. M., Bettencourt, L. M. A., Castillo-Chavez, C. (2009). *Mathematical and Statistical Approaches in Epidemiology*. Springer. New York, NY.

Costa, P. J. (2017). *Applied Mathematics for the Analysis of Biomedical Data*. Wiley. Hoboken, NJ.

Costa, P. J., Dunyak, J. P., & Mohtashemi, M. (2005). Models, Prediction, and Estimation of Outbreaks of Infectious Disease, *Proceedings, IEEE Southeast Conference,* Ft. Lauderdale, FL, USA, 174-178.

Evensen, G. (2009). Data Assimilation: The Ensemble Kalman Filter (2nd Ed.). Springer. New York, NY.

Eksin, C., Paarporn, K., & Weitz, J. S. (2019). Systematic Biases in Disease Forecasting – The Role of Behavior Change. *Epidemics, 27,* 96-106.

Elliott, M. N., et al. (2009). Improving Disparity Estimates for Rare Racical/Ethnic Groups with Trend Estimation and Kalman Filtering: An Application to the National Health Interview, *Health Services Research, 44(5),* 1622-1639.

Giamberardino, P. D., & Iacoviello, D. (2017). Optimal Control of SIR Epidemic Model with State Dependent Switching Cost Index. *Biomedical Signal Processing and Control, 31,* 377-380.

Gourieroux, C., & Jasiak, J. (2020). Analysis of Virus Propagation: A Transition Model Representation of Stochastic Epidemiological Models. *arXiv:2006.10265 [q-bio.PE].*

Heffernan, J.M., Smith, R.D., & Wahl, L.M. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society, 2(4),* 281-293.

Hethcote, H.W. (2000). The mathematics of infectious disease. *SIAM Review, 42(4),* 599-653.

Hethcote, H.W. (2009). The basic epidemiology models: Models, expressions for r0, parameter estimation, and applications. *Mathematical understanding of infectious disease dynamics, 16,* 1-16.

Hu, B., Qiu, W., Xu, C., & Wang, J. (2020). Integration of a Kalman Filter in the Geographically Weighted Regression for Modeling the Transmission of Hand, Foot, and Mouth Disease. *BioMed Central Public Health, 20,* 479.

Huang, W., & Provan, G. (2016). An Improved State Filter Algorithm for SIR Epidemic Forecasting, (from proceedings of the Twenty-second European Conference on Artificial Intelligence (ECAI 2016),The Hague, The Netherlands, 29th August - 2nd September), *Frontiers in Artificial Intelligence and Applications, 285,* 524-532.

Jang, J., Jang, K., Kwon, H., & Lee, J. (2018). Feedback Control of an HBV Model Based on Ensemble Kalman Filter and Differential Evolution. *Mathematical Biosciences and Engineering, 15(3)*, 667-691.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME, 82*, 35-45.

Keleman, J. Z., Kertesz-Farkas, A., Kocsor, A., & Puskas, L. G. (2006). Kalman Filtering for Disease-State Estimation from Microarray Data. *Bioinformatics, 22(24)*, 3047-3053.

Kennedy, W. G., Agarwal, N., & Yang, S. J. (2014). Social Computing, Behavioral-Cultural Modeling, and Prediction. 6th International Conference, SBP. Springer. Washington, DC.

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society London,* 115, 700-721.

Kratzer, G. (2016). Simulation Based Inference in Epidemic Models. Masters Thesis in Bio-Statistics of Zurich University.

Kreuger, K., & Osgood, N. (2015). Particle Filtering Using Agent-Based Transmission Models. *Proceedings of the 2015 Winter Simulation Conference*.

Krishnamurthy, A., Cobb, L., Mandel, J., & Beezley, J. (2014). Bayesian Tracking of Emerging Epidemics Using Ensemble Optimal Statistical Interpolation. *Spatial and Spatio-Temporal Epidemiology, 10,* 39-48.

Labbe, R.R. (2020). Kalman and Bayesian Filters in Python. Obtained from https://drive.google.com/file/d/0By_SW19c1BfhSVFzNHc0SjduNzg/view on November 2 on November 1, 2020.

Mbalawata, I. S., Sarkka, S., & Haario, H. (2012). Parameter Estimation in Stochastic Differential Equations with Markov Chain Monte Carlo and Nonlinear Kalman Filtering. *Computational Statistics, 28,* 1195-1223.

McKinley, T. J., Ross, J. V., Deardon, R., & Cook, A. R. (2014). Simulation-based Bayesian Inference for Epidemic Models. *Computational Statistics and Data Analysis, 71*, 434-447.

Mohammadi, B. (2015). Ensemble Kalman Filters and Geometric Characterization of Sensitivity Spaces for Uncertainty Quantification in Optimization. *Computer Methods in Applied Mechanics and Engineering, 290,* 228-249.

Muellner, U. et al. (2018). Epidemix – An Interactive Multi-model Application for Teaching and Visualizing Infectious Disease Transmission. *Epidemics, 23,* 49-54.

Ndanguza, D., Mbalawata, I.S., & Nsabimana, J. P. (2016). Analysis of SDEs Applied to SEIR Epidemic Models by Extended Kalman Filter Method. *Applied Mathematics, 7,* 2195-2211.

Nkwayep, C. H., et al. (2020). Short-term Forecasts of the COVID-19 Pandemic: A Study Case of Cameroon. *Chaos, Solitons, and Fractcals, 140,* 110106.

O'Neill, P. D., & Wen, C. H. (2012). Modelling and Inference for Epidemic Models Featuring Non-linear Infection Pressure. *Mathematical Biosciences, 238,* 38-48.

Peng, R. D. (2020). A Very Short Course on Time Series Analysis. Obtained from https://bookdown.org/rdpeng/timeseriesbook/

Reis, J., & Shaman, J. (2016). Retrospective Parameter Estimation and Forecast of Respiratory Syncytial Virus in the United States. *PLOS Computational Biology, 12(10)*, e1005133.

Roberts, M.G. (2007). The pluses and minuses of Ro. Modeling Infectious Diseases in Humans and Animals. Princeton University Press.

Rhudy, M. B., Salguero, R. A., & Holappa, K. (2017). A Kalman Filtering Tutorial for Undergraduate Students. *International Journal of Computer Science & Engineering Survey, 8*(1), 1-18.

Safarishahrbijari, A., et al. (2017). Predictive accuracy of particle filtering in dynamic models supporting outbreak projections. *BioMed Central Infectious Diseases, 17,* 648-660.

Sameni, R. (2020). Mathematical Modeling of Epidemic Diseases; A Case Study of the COVID-19 Coronavirus, *arXiv:2003.11371 [q-bio.PE].*

Smith, D. & Moore, L. (2004). *The SIR model for spread of disease – Euler's method for systems.* Retrieved August 5, 2018, from *https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-introduction*

Soyer, R. (2016). Kalman Filtering and Sequential Bayesian Analysis. *WIREs Computational Statistics, 10,* e1438.

Teow, J. (2018). *Understanding Kalman Filters with Python*. Obtained from https://medium.com/@jaems33/understanding-kalman-filters-with-python-2310e87b8f48

Welch, G., & Bishop, G. *An Introduction to Kalman Filter*. Obtained from https://cis.temple.edu/~latecki/Courses/CIS750-03/Papers/KalmanFilterSIGGRAPH2001.pdf

Xia, C., et al. (2013). Effects of Delayed Recovery and Nonuniform Transmission on the Spreading of Diseases in Complex Networks. *Physica A, 392,* 1577-1585.

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLOS Computational Biology, 10(4)*, e1003583.

Yang, W., et al. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology, 11(7),* e1004383.

Yang, W., et al. (2015). Transmission Network of the 2014-2015 Ebola Epidemic in Sierra Leone. *Journal of the Royal Society Interface, 12,* 20150536.

Yang, J., & Zhang, Y. (2020). Epidemic Spreading of Evolving Community Structure. *Chaos, Solitons and Fractals, 140*, 110101.

Young, P. C., Pedregal, D. J., & Tych, W. (1999). Dynamic Harmonic Regression. *Journal of Forecasting, 18*, 369-394.

Zeng, X., & Ghanem, R. (2020). Dynamics Identification and Forecasting of COVID-19 by Switching Kalman Filters. *Computational Mechanics, 66,* 1179-1193.

Zhang, Y. (2016). Influenza Trend Prediction Using Kalman Filter and Particle Filter. Obtained from https://www.researchgate.net/publication/324069016_Influenza_trend_prediction_using_Kalman_filter_and_particle_filter on November 1, 2020.

Zimmer, C., Yaesoubi, R., & Cohen, T. (2017). A Likelihood Approach for Real-Time Calibration of Stochastic Compartmental Epidemic Models. *PLOS Computational Biology, 13(1)*, e1005257.