Acta Biotheoretica (2022) 70:25
https://doi.org/10.1007/s10441-022-09449-z

REGULAR ARTICLE

Stochastic Modeling and Forecasting of Covid-19 Deaths: Analysis for the Fifty States in the United States

Olusegun Michael Otunuga[1] · **Oluwaseun Otunuga**[2]

Received: 16 September 2021 / Accepted: 5 September 2022

Abstract

In this work, we study and analyze the aggregate death counts of COVID-19 reported by the United States Centers for Disease Control and Prevention (CDC) for the fifty states in the United States. To do this, we derive a stochastic model describing the cumulative number of deaths reported daily by CDC from the first time Covid-19 death is recorded to June 20, 2021 in the United States, and provide a forecast for the death cases. The stochastic model derived in this work performs better than existing deterministic logistic models because it is able to capture irregularities in the sample path of the aggregate death counts. The probability distribution of the aggregate death counts is derived, analyzed, and used to estimate the count's per capita initial growth rate, carrying capacity, and the expected value for each given day as at the time this research is conducted. Using this distribution, we estimate the expected first passage time when the aggregate death count is slowing down. Our result shows that the expected aggregate death count is slowing down in all states as at the time this analysis is conducted (June 2021). A formula for predicting the end of Covid-19 deaths is derived. The daily expected death count for each states is plotted as a function of time. The probability density function for the current day, together with the forecast and its confidence interval for the next four days, and the root mean square error for our simulation results are estimated.

Keywords Covid-19 · Stochastic differential equation · Probability density function · Forecast · Aggregate death · First passage time

✉ Olusegun Michael Otunuga
ootunuga@augusta.edu

Oluwaseun Otunuga
otunuga1@mail.usf.edu

[1] Department of Mathematics, Augusta University, 1120 15th Street, GE-2018, Augusta, GA 30912, USA

[2] Department of Mathematics & Statistics, University of South Florida, 4202 E Fowler Ave, Tampa, FL, USA

Published online: 16 September 2022

1 Introduction

Several mathematical models (Wu et al. 2020; Stutt et al. 2020; Linka et al. 2020; Okuonghae and Omame 2020; Ndairou et al. 2020; Ladde et al. 2020; Otunuga 2020; Mummert and Otunuga 2019; Otunuga 2018; Santosh 2020) have been developed to study the transmission of the COVID-19 virus caused by the virus species "severe acute respiratory syndrome-related corona virus", named SARS-CoV-2. The airborne transmission occurs by inhaling droplets loaded with SARS-CoV-2 particles that are expelled by infectious people. According to Wu et al. (2020), the "severe acute respiratory syndrome coronavirus" (SARS-CoV) and the "Middle East respiratory syndrome coronavirus" (MERS-CoV) are two other novel coronaviruses that emerged as major global health threats since 2002. Several public health interventions have been put in place to eradicate or reduce the spread of the disease. According to CDC, the first U.S. laboratory-confirmed case[1] of COVID-19 in the U.S. was recorded on January 20, 2020 from the samples taken 2 days earlier in Washington state. The first COVID-19 death in the United States was first reported in the same state by CDC on February 29, 2020. As of June 24, 2021, the total number of Covid-19 cases in the United States was reported by CDC to be 33, 437, 643, resulting in about 601, 221 deaths[2]. On December 11 and December 18, 2020, the United States Food and Drug Administration (FDA)[3] issued an Emergency Use Authorization (EUA) for the Pfizer-BioNTech and the Moderna COVID-19 vaccine, respectively, in the United States. An EUA for the third vaccine, the Johnson and Johnson (J &J) vaccine, was first issued in the United States on February 27, 2021[4]. A pause on the usage of the J &J vaccine was recommended by CDC and the FDA on April 13, 2021 due to the serious blood clots (a condition called thrombosis) in six women between the ages of 18 and 49 years with thrombocytopenia syndrome (TTS)[5] following the usage of the vaccine. As of June 7, 2021, about 51.5% of the total population of the United States have received at least one dose of the vaccination, with 41.9% fully vaccinated. The study done in this work is to check the effects of these interventions. That is, with these vaccines, we check if the number of deaths resulting from the Covid-19 disease slows down as time proceeds and as the number of those who are vaccinated increases.

The trajectory of the aggregate death counts in most states in the United States follows the same dynamics. At first, it follows a somewhat exponential trajectory, with its growth slowing down at some point and speeding up at other points. With public health interventions like vaccination mitigating the growth of the virus, this pattern is expected to continue until a certain steady-state is reached. This dynamic

[1] https://www.cdc.gov/museum/timeline/covid19.html, accessed 06.04.2022

[2] https://covid.cdc.gov/covid-data-tracker, accessed 06.24.2021 at 9:09PM.

[3] https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-frequently-asked-questions, accessed 03.09.2021.

[4] https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/janssen-covid-19-vaccine, accessed 06.08.2021.

[5] https://www.healthline.com/health/vaccinations/johnson-and-johnson-vaccine, accessed 06.07.2021.

follows roughly the well known Verhulst logistic equation and its generalization (Beddington and May 1977; Li and Wang 2010; Prajneshu 1980; Pelinovsky et al. 2020; Pella and Tomlinson 1969; Wang et al. 2020). The model was first derived by Verhulst (1838) to study population growth. Some other methods (Baud et al. 2020; Bhapkar et al. 2020; Satpathy et al. 2021; Kaciroti et al. 2021) have been developed to estimate mortality following the Covid-19 infection. In this work, we consider the logistic model

$$dN = \frac{\bar{\boldsymbol{\mu}}}{\mathcal{K}}(\mathcal{K} - N)N\, dt, \quad N(t_0) = N_0, \tag{1.1}$$

where $t_0 \geq 0$, $N_0 > 0$, $N(t)$ denotes the total number of Covid-19 death counts at time t, $\mathcal{K}$ is the maximum number of Covid-19 deaths, and $\bar{\boldsymbol{\mu}}$ is the per capita initial growth rate, to interpret the aggregate number of COVID-19 death trajectories in the United States. We see from (1.1) that $\frac{dN}{dt} > 0$ on $(0, \mathcal{K})$ and $\frac{d^2N}{dt^2} = \frac{\bar{\boldsymbol{\mu}}^2}{\mathcal{K}^2}(\mathcal{K} - N)(\mathcal{K} - 2N)N$. From this, we have $\frac{d^2N}{dt^2} > 0$ on the interval $(0, \mathcal{K}/2)$, and $\frac{d^2N}{dt^2} < 0$ on $(\mathcal{K}/2, \mathcal{K})$. That is, the trajectory of the aggregate number speeds up in the interval $(0, \mathcal{K}/2)$ and slows down in the interval $(\mathcal{K}/2, \mathcal{K})$. Also, since $N(t)$ represents the aggregate number of deaths at a given time t, it follows that the daily number of deaths is at the maximum at the time when $N(t) = \mathcal{K}/2$.

From this analysis, if the current day aggregate death counts in time series is more than $\mathcal{K}/2$, then we know that its growth is slowing down and the virus spread has been controlled. Otherwise, the virus is spreading, with speeding growth. The problem with using this model to analyze the aggregate number of cases is that it fails to account for the fluctuations or perturbations in the data resulting from fluctuations in the rates of infection or death. These noises/fluctuations can be caused by many factors like the rates at which Covid-19 testing is done, vaccination rates, mask use per capital, social behavior, public health intervention (Linka et al. 2020), and so on. In addition, CDC[6] reported on their websites that counting exact aggregate confirmed and probable COVID-19 cases and deaths is not possible due to delays in reporting from different voluntary jurisdictions. The number of death cases reported on CDC's website might not be complete because it takes several weeks for death records to be processed, coded, submitted, and tabulated on the National Center for Health Statistics (NCHS). As a result, these cause the counts to fluctuate substantially, with a possibility of a negative number of probable cases reported on a given day if more probable cases were disproven than were initially reported on that day.

Several authors (Lv et al. 2019; Li and Wang 2010; Lungu and Øksendal 1997; Yang et al. 2019; Gardiner 1985) have worked on model (1.1) and its extension to a stochastic model. In this work, we derive a stochastic model governing the aggregate number of Covid-19 death counts in the United States by extending model (1.1) to a stochastic case. The proposed model is better in the sense that it captures

[6] https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36/data.

fluctuations in the aggregate counts better than the widely used logistic model (1.1) with lesser root mean square error. Our aim in this study is to determine whether the virus infection and death counts resulting from the infection are still growing sharply or slowing down. Upon analyzing the COVID-19 data collected from CDC, we see that the path of the aggregate number of Covid-19 death counts over time follows a logistic model with irregular trajectories. We assume this fluctuation is caused by many factors such as listed above, causing the per capita growth rate to fluctuate over time. We account for the fluctuations in this rate by extending the deterministic logistic model (1.1) to a stochastic model. This is done by assuming the parameter $\bar{\boldsymbol{\mu}}$ is not constant over time, but fluctuates about a particular mean value. In order to estimate the epidemiological parameters $\mathcal{K}$ (the carrying capacity of Covid-19 death counts), $\bar{\boldsymbol{\mu}}$ (the per capita initial growth rate), and the death rate noise intensity, we derive the transition probability density function for the aggregate number of death counts and apply a Maximum Likelihood Estimate (MLE) scheme. The distribution is also used to calculate the expected aggregate count at a particular period in time. Using the parameter estimates for the fitted data, we forecast the cumulative number of Covid-19 deaths and provide a 95% confidence interval for the forecast.

The organization of the work done is as follows: In Sect. 2, we derive a stochastic model describing the cumulative number of deaths by assuming $\bar{\boldsymbol{\mu}}$ is not constant, but changes with time and fluctuates around a mean value. We show that the model is well defined, with a unique closed-form solution. In Sect. 3, the transition probability density function for the aggregate number of death counts is derived. Using the MLE scheme, we estimate the epidemiological parameters in the model. Since the aggregate count $N(t)$ is random, we calculate the expected number of Covid-19 aggregate death counts for each states in the United States. We show that with probability one the aggregate count remains in the interval $(0, \mathcal{K})$ if it starts from there. In Sect. 4, the expected first hitting time when the aggregate count $N(t)$ reaches $\mathcal{K} - \epsilon$ is calculated for some small positive constant $\epsilon > 0$. We also estimate the expected first passage time when the aggregate death counts started slowing down. Numerical simulations, forecast, and analysis of the aggregate death counts for the fifty states in the United States are carried out in Sect. 5. The summary of the work done is given in Sect. 6.

2 Methodology

2.1 Data Sources

The Covid-19 aggregate death counts in the United States is collected from the United States Centers for Disease Control and Prevention (CDC) website and provided by the CDC Case Task Force[7]. The data was collected for the period ranging from January 22, 2020 to June 24, 2021, and it includes the date of counts, state/

[7] https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36, accessed 06.25.2021.

jurisdiction, total/aggregate number of death cases (including total confirmed and probable deaths), number of new and new probable deaths with the date and time the records were created. The definition of each of these can be found on the CDC's website[7].

2.2 Modeling the Covid-19 Aggregate Death Counts

In this section, we describe the dynamics of the aggregate number of deaths in the United States by extending the well-known deterministic logistic model (1.1) to a stochastic differential equation. Analysis of the data (see Figs. 1, 2) shows that its growth rate fluctuates with time. Some works (Lagarto and Braumann 2014; Mazzuco et al. 2018; Zocchetti and Consonni 1994) have been done in analyzing the distribution of mortality rate. The dynamics of male and female crude death rates of the Portuguese population over the period 1940–2009 were modeled in the work of Lagarto and Braumann (2014) using a bi-dimensional stochastic Gompertz model with correlated Wiener processes. Zocchetti and Consonni (1994) showed in their work that when the number of deaths is sufficiently elevated, the Gauss distribution (also referred to as the normal distribution) can be used as a good approximation distribution for the variability in the mortality rate. In their work, Mazzuco et al. (2018) derived a new model for mortality rate based on the mixture of a half-normal distribution with a generalization of the skew-normal distribution. The Wiener process (Khasminskii 2012; Kloeden and Platen 1995; Mao 2007; Øksendal 2003), often called Brownian motion, is a real-valued continuous-time stochastic process $\{W(t) : t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with stationary independent Gaussian increments such that $W(0) = 0$ with probability one, $W(t + \Delta t) - W(t)$ is normally distributed with mean 0 and variance Δt, and $W(t) - W(s)$ is independent of the past random variable $W(u)$, $0 \leq u \leq s$. In addition, the random variables $\{W(t_j) - W(s_j),\ j = 1, 2, \cdots, n\}$ are jointly independent, for $0 \leq s_1 < t_1 \leq s_2 < t_2 \cdots \leq s_n < t_n < \infty$. The independence and stationarity of the increment, together with the continuous (almost everywhere) sample path of the process $W(t)$ lead to its great tractability, making it one of the most important stochastic process in continuous time used in biological systems to model perturbed epidemiological parameters. Following a similar assumption made in the work of Prajneshu (1980), Yang et al. (2019), and Otunuga (2021b), where a logistic stochastic population model with the population's transition probability density function were derived to study the distribution of a population subjected to a continuous spectrum of disturbances, with fluctuations in the intrinsic growth rate, we assume the dynamics of $\bar{\boldsymbol{\mu}}$ shouldn't be constant over time, but instead be driven by fluctuations that can be modeled to follow a process of the form

$$\bar{\boldsymbol{\mu}} dt = \mu dt + \sigma \circ dW(t), \tag{2.1}$$

where $W(t)$ is a standard Wiener process, μ is the average death counts per capita initial growth rate, σ is the noise intensity, and $\circ$ is the Stratonovich integral symbol (Arnold 1974). We use the Stratonovich calculus instead of the Itô calculus to describe this dynamic simply because it obeys the traditional rule of chain rule and

allows white noise to be treated as a regular derivative of a Brownian or Wiener process (West et al. 1979; Wong and Zakai 1965). For more reading on the Itô and Stratonovich calculus, we direct readers to the work of Kloeden and Platen (1995) and Øksendal (2003). Substituting into (1.1), the proposed stochastic differential equation (SDE) for the aggregate number of deaths is given by

$$dN = \frac{\mu}{\mathcal{K}}(\mathcal{K} - N)N\,dt + \frac{\sigma}{\mathcal{K}}(\mathcal{K} - N)N \circ dW(t), \quad N(t_0) = N_0, \tag{2.2}$$

where $N_0 > 0$, $\mathcal{K}$ is the carrying death capacity, μ and σ are as described in (2.1). We note here that the stochasticity added in model (2.2) leads to a non-monotonic cumulative death sample path, a property that makes it able to capture irregularities in the sample path better than the deterministic counterpart, with a smaller root mean square error. The interpretation of the stochastic differential equation as a Stratonovich differential equation follows the work of Otunuga (2019, 2021a). We shall later compare model (2.2) with its deterministic equivalent (1.1) and show that model (2.2) performs better in capturing the trajectory of (and the noise in) the aggregate death counts. We convert (2.2) into a Itô stochastic differential equation as

$$dN = \left(\frac{\mu}{\mathcal{K}}(\mathcal{K} - N)N + \frac{\sigma^2}{2\mathcal{K}^2}(\mathcal{K} - N)(\mathcal{K} - 2N)N\right)dt + \frac{\sigma}{\mathcal{K}}(\mathcal{K} - N)N dW(t), \quad N(t_0) = N_0. \tag{2.3}$$

In order to show that model (2.3) is biologically feasible, we show in Theorem 2 (using Corollary 3.1 of Khasminskii (2012)) that the solution $N(t)$ of (2.3) exists, and it remains in $(0, \mathcal{K})$ with probability one whenever it starts from there. A statement of the corollary, together with definition of some terminologies in the corollary are given below.

Definition 1 *L*-operator.

Given a one-dimensional stochastic differential equation

$$dx = f(t,x)dt + g(t,x)dW(t), \;\; x(t_0) = x_0, \tag{2.4}$$

we define the *L*-operator (Mao 2007) associated with (2.4) as

$$L = \frac{\partial}{\partial t} + f(t,x)\frac{\partial}{\partial x} + \frac{1}{2}g^2(t,x)\frac{\partial^2}{\partial x^2}. \tag{2.5}$$

If L acts on a nonnegative function $V(t, x)$ which is continuously differentiable with respect to t and twice continuously differentiable with respect to x, then

$$LV(t,x) = \frac{\partial V(t,x)}{\partial t} + f(t,x)\frac{\partial V(t,x)}{\partial x} + \frac{1}{2}g^2(t,x)\frac{\partial^2 V(t,x)}{\partial x^2}.$$

The usefulness of the expression for $LV(t, x)$ is seen in the Itô Lemma (a Lemma which gives the formula for the stochastic analogue of the chain rule or change of variable rule in calculus), which simply states that the stochastic differential of $V(t, x)$ is given by

$$dV(t,x) = LV(t,x)dt + g(t,x)\frac{\partial V(t,x)}{\partial x}dW(t). \tag{2.6}$$

We see here that $LV(t, x)$ is the drift part of the differential $dV(t, x)$.

Theorem 1 (*From Corollary* 3.1 *of Khasminskii* (2012))

Let D_n be an increasing sequence of open sets whose closure are contained in an open set D such that $\bigcup D_n = D$. Suppose that the drift and diffusion coefficients $f(t, x)$ and $g(t, x)$, respectively, in (2.4) satisfy the Lipschitz and linear growth conditions in $(0, \infty) \times D_n$ and there exists a function $V(t, x)$, twice continuously differentiable in x and continuously differentiable in t in the domain $(0, \infty) \times D$, which satisfies

$$LV \leq cV,$$
$$\inf_{t>0,\ x\in D\setminus D_n} V(t,x) \to \infty \ \ as \ \ n \to \infty,$$

for some positive constant c. Then for every random variable $x(t_0)$ independent of $W(t) - W(t_0)$, there exists a solution $x(t)$ of (2.4) which is an almost surely continuous stochastic process and is unique up to equivalence[8] *provided that $\mathbb{P}(x(t_0) \in D) = 1$. Moreover the solution satisfies the relation*

$$\mathbb{P}(x(t) \in D) = 1 \quad \text{for all} \ \ t \geq t_0.$$

Since the drift and diffusion coefficients of (2.3) are non-linear, the classical existence and uniqueness theorem of SDE (Kloeden and Platen 1995; Khasminskii 2012; Mao 2007; Øksendal 2003) does not apply. We use Theorem 1 to prove the existence and uniqueness of the solution of (2.3) in the interval $(0, \mathcal{K})$ in Theorem 2.

Theorem 2 *Let the stochastic differential equation (2.3) be given for any $t_0 \geq 0$ and initial value $N_0 \in (0, \mathcal{K})$ independent of $W(t) - W(t_0)$. Then there exists a unique global positive solution $N : [t_0, \infty) \to \mathfrak{R}^+$ such that with probability one $N(t) \in (0, \mathcal{K})$. That is,*

$$\mathbb{P}\{N(t) \in (0, \mathcal{K}),\ t \geq t_0\} = 1.$$

Proof Define the sequence $\{D_r\}$ by

$$D_r = \left(\frac{1}{r}, \mathcal{K} - \frac{1}{r}\right), r = 1, 2, \cdots.$$

Clearly, $\{D_r\}$ is an increasing sequence of open sets whose closures are contained in $(0, \mathcal{K})$. The drift and diffusion coefficients $b(N) = \frac{\mu}{\mathcal{K}}(\mathcal{K} - N)N + \frac{\sigma^2}{2\mathcal{K}^2}(\mathcal{K} - N)(\mathcal{K} - 2N)N$ and $g(N) = \frac{\sigma}{\mathcal{K}}(\mathcal{K} - N)N$, respectively,

[8] Two solutions $x_1 : [t_0, T] \to \mathfrak{R}^+$ and $x_2 : [t_0, T] \to \mathfrak{R}^+$ are said to be equivalent if $\mathbb{P}\big(x_1(t) = x_2(t) \ \text{for all} \ t \in [t_0, T]\big) = 1$.

of (2.3) satisfy the Lipschitz and linear growth conditions locally in D_r. Define a function $V : (0, \mathcal{K}) \to \mathfrak{R}^+$ by

$$V(N) = \frac{\mathcal{K}}{N(\mathcal{K} - N)}.$$

Applying the L-operator ((2.5)) associated with (2.3) on V, we have

$$\begin{aligned} LV =& \frac{\partial V}{\partial t} + b(N)\frac{\partial V}{\partial N} + \frac{g^2(N)}{2}\frac{\partial^2 V}{\partial N^2} \\ =& \frac{\mu}{\mathcal{K}}(\mathcal{K} - N)N\left(-\frac{1}{N^2} + \frac{1}{(\mathcal{K} - N)^2}\right) + \frac{\sigma^2}{\mathcal{K}^2}(\mathcal{K} - N)^2N^2\left(\frac{1}{N^3} + \frac{1}{(\mathcal{K} - N)^3}\right) \\ &+ \frac{\sigma^2}{2\mathcal{K}^2}(\mathcal{K} - N)(\mathcal{K} - 2N)N\left(-\frac{1}{N^2} + \frac{1}{(\mathcal{K} - N)^2}\right) \\ \leq& \frac{\mu}{\mathcal{K}}\frac{\mathcal{K}}{\mathcal{K} - N} + \frac{\sigma^2}{\mathcal{K}^2}\left(\frac{\mathcal{K}^3}{(\mathcal{K} - N)N} - 3\mathcal{K}\right) + \frac{\sigma^2}{2\mathcal{K}^2}\left(\frac{\mathcal{K}(\mathcal{K} - N)N}{(\mathcal{K} - N)^2} + \frac{2(\mathcal{K} - N)N^2}{N^2}\right) \\ \leq& \mathcal{C}V, \end{aligned}$$

where $\mathcal{C} = \mu + 3\sigma^2/2$. For any $x \in D\backslash D_r = \left(0, \frac{1}{r}\right] \bigcup \left[\mathcal{K} - \frac{1}{r}, \mathcal{K}\right)$, we have

$$\inf_{t>0,\ x\in D\backslash D_r} V(x) = r + \frac{1}{\mathcal{K}} \to \infty \ \ as \ \ r \to \infty.$$

The result follows from Theorem 1. □

We give the exact solution of (2.3) in the following theorem. This solution will be used in the simulation process to plot the sample path of the death counts.

Theorem 3 *For any given initial value $N_0 \in (0, \mathcal{K})$, the exact solution of the SDE (2.3) is obtained as*

$$N(t) = \frac{\mathcal{K}}{1 + \frac{\mathcal{K} - N_0}{N_0}\Phi^{-1}(t)}, \tag{2.7}$$

where

$$\Phi(t) = e^{\mu(t-t_0)+\sigma(W(t)-W(t_0))}.$$

Proof As shown in Theorem (2), there is a unique global positive solution $N(t) \in (0, \mathcal{K})$ for all $t \geq t_0$ with probability one. Define

$$u = \ln\left(\frac{N}{\mathcal{K} - N}\right), \quad N \in (0, \mathcal{K}). \tag{2.8}$$

It follows from (2.3) and (2.6) that

$$du = \mu dt + \sigma dW(t), \ \ u(t_0) = u_0, \tag{2.9}$$

with solution

$$u(t) = u_0 + \mu(t - t_0) + \sigma(W(t) - W(t_0)). \tag{2.10}$$

The result follows by substituting (2.10) into (2.8) and solving for $N(t)$. □

Remark 1 Theorem 2 shows that the aggregate number of deaths cannot grow past a particular value, $\mathcal{K}$, with probability one if its starting point is in $(0, \mathcal{K})$. If $0 < N_0 < \mathcal{K}$, then it follows from Theorems 2 and 3 that the feasible epidemiological region of interest for the solution $N(t)$ is the set

$$\mathcal{T} = \left\{n \in \mathfrak{R}^+ \mid 0 < n < \mathcal{K}\right\}. \tag{2.11}$$

The following theorem shows that if $\mu > \sigma^2/2$ and the process $N(t)$ starts from $(0, \mathcal{K})$, then it converges almost surely to a random variable N_∞ with finite expectation.

Theorem 4 *For any given initial condition $N_0 \in (0, \mathcal{K})$, the process $N(t)$ is a submartingale if $\mu > \sigma^2/2$. That is,*

$$\mathbb{E}[N(t)|N(s),\ s \le t] \ge N(s).$$

Furthermore, there exists a random variable N_∞ such that $\mathbb{E}\left[\left(N_\infty\right)\right] < \infty$ and

$$\lim_{t\to\infty} N(t) = N_\infty. \tag{2.12}$$

Proof It follows from (2.2) that for $s \le t$, we have

$$N(t) = N(s) + \int_s^t \left(\frac{\mu}{\mathcal{K}}(\mathcal{K} - N(u))N(u) + \frac{\sigma^2}{2\mathcal{K}^2}(\mathcal{K} - N(u))(\mathcal{K} - 2N(u))N(u)\right) du + \int_s^t \frac{\sigma}{\mathcal{K}}(\mathcal{K} - N(u))N(u)dW(u).$$

If $\mu > \sigma^2/2$, we obtain

$$\mathbb{E}[N(t)|N(s)] = N(s) + \int_s^t \left(\frac{\mu}{\mathcal{K}}(\mathcal{K} - N(u))N(u) + \frac{\sigma^2}{2\mathcal{K}^2}(\mathcal{K} - N(u))(\mathcal{K} - 2N(u))N(u)\right) du \ge N(s).$$

Clearly, we see from (2.7) that $0 < N(t) < \mathcal{K}$ for all $t \ge 0$. Hence, $\sup_{t\ge 0} \mathbb{E}\left[\left(N(t)^+\right)\right] < \infty$. By the Martingale Convergence Theorem, we have $\mathbb{E}\left(N_\infty\right) < \infty$ and equation (2.12) is satisfied. □

Remark 2 The Martingale Convergence Theorem is a stochastic analogue of the Monotone Convergence Theorem. Here, we also see that a submartingale property is a stochastic analogue of a non-decreasing sequence. Since aggregate death count is expected to be increasing with time and the process $N(t)$ is stochastic, we need a condition that will guarantee a stochastic analogue of an increasing function.

Theorem 4 gives such condition. It shows that the expected aggregate death counts is an increasing function provided $N_0 \in (0, \mathcal{K})$ and $\mu > \sigma^2/2$. It has been shown in several works (Mendez et al. 2012; Otunuga 2018, 2020) that the presence of environmental perturbations can affect the dynamic nature of biological systems if the noise intensity grows beyond a certain value. The latter condition shows that the noise intensity, σ, of the environmental perturbation must not be allowed to grow beyond a certain function of the average death count's growth rate μ if the submartingale property is to be maintained. Following Theorem 4, we assume for the rest of this work that the noise intensity $\sigma < \sqrt{2\mu}$.

We discuss and analyze the distribution of the random aggregate number of death counts process $N(t)$ by first deriving its transition probability density function. The distribution will be used to estimate the epidemiological parameters $\mathcal{K}$, μ, and σ, and to calculate the expected aggregate number of death counts at a particular time t in the United States. These estimates will later be used in simulating and forecasting the total death counts.

3 Probability Distribution of the Aggregate Number of Deaths Following (2.3)

Let $p_N(n|t, N_0)$ represents the transition probability density function (PDF) for the aggregate death counts $N(t)$ given t and N_0. Following the results in (2.8) and (2.10), the transition probability density function $p_N(n|t, N_0)$ is obtained as

$$p_N(n|t,N_0) = \frac{\mathcal{K}}{n(\mathcal{K}-n)\sigma\sqrt{2\pi(t-t_0)}} \exp\left(-\frac{\left(\ln\left(\frac{n}{(\mathcal{K}-n)}\frac{(\mathcal{K}-N_0)}{N_0}\right) - \mu(t-t_0)\right)^2}{2(t-t_0)\sigma^2}\right), \quad 0 < n < \mathcal{K}. \tag{3.1}$$

The purpose of deriving this PDF is to be able to estimate the epidemiological parameters in model (2.3) using the MLE scheme.

3.1 Parameter Estimates

Let T be a number corresponding to the current date the Covid-19 aggregate data is collected, and $t_0 < t_1 < \cdots < t_m = T$ be a partition P of the interval $[t_0, T]$. Denote $N(t_j)$ by N_j and let $N_0, N_1, N_2, \cdots, N_m$ be samples satisfying (2.3) at a given time. Let $\Delta t_j = t_j - t_{j-1}$, $j = 1, 2, \cdots, m$. The likelihood and log-likelihood functions $L(\Theta|N)$ and $\mathcal{L}(\Theta|N)$, respectively, of the samples are obtained using (3.1), the transformation (2.8), and the distribution of u in (2.10), as

$$
\begin{aligned}
L(\Theta|N) &= \prod_{j=1}^{m} p_N(N_j|t_j, N_{j-1}) \\
&= \prod_{j=1}^{m} \frac{\mathcal{K}}{N_j(\mathcal{K}-N_j)\sigma\sqrt{2\pi\Delta t_j}} \exp\left(-\frac{\left(\ln\left(\frac{N_j}{(\mathcal{K}-N_j)}\frac{(\mathcal{K}-N_{j-1})}{N_{j-1}}\right) - \mu\Delta t_j\right)^2}{2\Delta t_j\sigma^2} \right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{L}(\Theta|N) = {} & m\ln\mathcal{K} - \sum_{j=1}^{m} \ln\left(N_j(\mathcal{K}-N_j)\right) - \frac{m}{2}\ln\sigma^2 - \frac{1}{2\sigma^2} \\
& \sum_{j=1}^{m} \frac{1}{\Delta t_j}\left(\ln\left(\frac{N_j}{(\mathcal{K}-N_j)}\frac{(\mathcal{K}-N_{j-1})}{N_{j-1}}\right) - \mu\Delta t_j\right)^2,
\end{aligned} \tag{3.2}
$$

where $\Theta = \{\mathcal{K}, \mu, \sigma\}$ represents the parameter set to be estimated. The maximum likelihood estimates $\hat{\mathcal{K}}$, $\hat{\mu}$, $\hat{\sigma}^2$ of $\mathcal{K}$, μ, σ^2 are estimated from (3.2) as

$$
\begin{aligned}
\hat{\mu} &= \frac{\sum_{j=1}^{m} \ln\left(\frac{N_j}{(\hat{\mathcal{K}}-N_j)}\frac{(\hat{\mathcal{K}}-N_{j-1})}{N_{j-1}}\right)}{\sum_{j=1}^{m}\Delta t_j} = \frac{\ln\left(\frac{N(T)}{(\hat{\mathcal{K}}-N(T))}\frac{(\hat{\mathcal{K}}-N_0)}{N_0}\right)}{T-t_0}, \\
\hat{\sigma}^2 &= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{\Delta t_j}\left(\ln\left(\frac{N_j}{(\hat{\mathcal{K}}-N_j)}\frac{(\hat{\mathcal{K}}-N_{j-1})}{N_{j-1}}\right) - \hat{\mu}\Delta t_j\right)^2,
\end{aligned} \tag{3.3}
$$

where $\hat{\mathcal{K}}$ satisfies

$$
\begin{aligned}
& \frac{m}{\hat{\mathcal{K}}} - \sum_{j=1}^{m}\frac{1}{\hat{\mathcal{K}}-N_j} - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{m}\frac{1}{\Delta t_j}\left(\ln\left(\frac{N_j}{(\hat{\mathcal{K}}-N_j)}\frac{(\hat{\mathcal{K}}-N_{j-1})}{N_{j-1}}\right) - \hat{\mu}\Delta t_j\right) \\
& \left(\frac{N_j - N_{j-1}}{(\hat{\mathcal{K}}-N_{j-1})(\hat{\mathcal{K}}-N_j)}\right) = 0.
\end{aligned} \tag{3.4}
$$

Remark 3 The initial point N_0 can also be estimated for better simulation result. In this case, the estimates $\hat{\mathcal{K}}$, $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{N}_0$ of $\mathcal{K}$, μ, σ^2, and N_0 reduce to

$$
\begin{aligned}
\hat{\mu} &= \frac{\ln\left(\frac{N(T)}{(\hat{\mathcal{K}}-N(T))}\frac{(\hat{\mathcal{K}}-N_1)}{N_1}\right)}{T-t_0}, \\
\hat{N}_0 &= \frac{\hat{\mathcal{K}}}{1+\frac{\hat{\mathcal{K}}-N_1}{N_1}e^{\hat{\mu}\Delta t}}, \\
\hat{\sigma}^2 &= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{\Delta t_j}\left(\ln\left(\frac{N_j}{(\hat{\mathcal{K}}-N_j)}\frac{(\hat{\mathcal{K}}-N_{j-1})}{N_{j-1}}\right) - \hat{\mu}\Delta t_j\right)^2\Big|_{N_0=\hat{N}_0},
\end{aligned} \tag{3.5}
$$

where $\hat{\mathcal{K}}$ satisfies

$$\frac{m}{\hat{\mathcal{K}}} - \sum_{j=1}^{m} \frac{1}{\hat{\mathcal{K}} - N_j} - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^{m} \frac{1}{\Delta t_j} \left(\ln \left(\frac{N_j}{(\hat{\mathcal{K}} - N_j)} \frac{(\hat{\mathcal{K}} - N_{j-1})}{N_{j-1}} \right) - \hat{\mu} \Delta t_j \right) \left(\frac{N_j - N_{j-1}}{(\hat{\mathcal{K}} - N_{j-1})(\hat{\mathcal{K}} - N_j)} \right) \Big|_{N_0 = \hat{N}_0} = 0. \tag{3.6}$$

3.2 Expected and Simulated Number of Deaths

Since the aggregate death counts $N(t)$ is a random process, it is important to calculate the expected number of total deaths at each given time. Given the initial value N_0, the expected aggregate number of deaths, denoted $\mathbb{E}\left[N(t)|N_0\right]$, at time t is calculated from (3.1) as

$$\mathbb{E}\left[N(t)|N_0 = n_0\right] = \int_0^{\mathcal{K}} n\, p_N(n|t, N_0)\, dn = \int_{-\infty}^{\infty} \frac{\mathcal{K}}{\sigma\sqrt{2\pi t}\left(1 + \frac{\mathcal{K}-N_0}{N_0} e^{-u}\right)} \exp\left[-\frac{(u-\mu t)^2}{2\sigma^2 t}\right] du. \tag{3.7}$$

We show in the next theorem that for each time t, the expected death counts falls in $(0, \mathcal{K})$ if the death counts starts there.

Theorem 5 *If $N_0 \in (0, \mathcal{K})$, then $0 < \mathbb{E}[N(t)|N_0] < \mathcal{K}$.*

Proof If $N_0 \in (0, \mathcal{K})$, then it follows from (3.7) that

$$\begin{aligned} \mathbb{E}[N(t)|N_0] &= \int_{-\infty}^{\infty} \frac{\mathcal{K}}{\sigma\sqrt{2\pi t}\left(1 + \frac{\mathcal{K}-N_0}{N_0} e^{-u}\right)} \exp\left[-\frac{(u-\mu t)^2}{2\sigma^2 t}\right] du \\ &< \int_{-\infty}^{\infty} \frac{\mathcal{K}}{\sigma\sqrt{2\pi t}} \exp\left[-\frac{(u-\mu t)^2}{2\sigma^2 t}\right] du = \mathcal{K}. \end{aligned}$$

□

Theorem 5 shows that the expected aggregate number of deaths will always be in the feasible region $\mathcal{T}$ if the initial point N_0 starts from there. The following theorem gives the total deaths expected on the long run and shows condition under which this number converges to the point $N = \mathcal{K}$ using Theorem 5.3 of Khasminskii (2012).

Theorem 6 *If $\mu > \sigma^2/2$ and $N_0 \in (0, \mathcal{K})$, then the total deaths expected on the long run is $\mathcal{K}$.*

Proof Consider the random process $z(t) = \mathcal{K} - N(t)$. It follows from (3.1) that the probability density function $p_z(z|t, z_0)$ of the random variable z given the initial point z_0 is obtained as

$$p_z(z|t,z_0)=\frac{\mathcal{K}}{(\mathcal{K}-z)z\sigma\sqrt{2\pi(t-t_0)}}\exp\left(-\frac{\left(\ln\left(\frac{(\mathcal{K}-z)}{z}\frac{z_0}{(\mathcal{K}-z_0)}\right)-\mu(t-t_0)\right)^2}{2(t-t_0)\sigma^2}\right),\quad 0<z<\mathcal{K}.$$

For any $\delta>0$,

$$0\le\sup_{|z_0|>\delta,\ t\ge t_0}\mathbb{E}\left[z(t)|z_0\right]\le\int_{-\infty}^{\infty}\frac{\mathcal{K}}{\sigma\sqrt{2\pi(t-t_0)}\left(1+\frac{\mathcal{K}-\delta}{\delta}e^u\right)}$$
$$\exp\left[-\frac{\left(u-\mu(t-t_0)\right)^2}{2\sigma^2(t-t_0)}\right]du\to 0\quad\text{as}\quad\delta\to 0$$

and

$$0<\mathbb{E}\left[z(t)|z_0\right]<\int_{-\infty}^{\infty}\frac{\mathcal{K}}{\sigma\sqrt{2\pi(t-t_0)}\left(\frac{\mathcal{K}-z_0}{z_0}e^u\right)}\exp\left[-\frac{\left(u-\mu(t-t_0)\right)^2}{2\sigma^2(t-t_0)}\right]du$$
$$=\frac{z_0}{\mathcal{K}-z_0}\exp\left(-\left(\mu-\frac{\sigma^2}{2}\right)(t-t_0)\right).$$

We deduce from the Squeeze Theorem that

$$\lim_{t\to\infty}\mathbb{E}[z(t)|z_0]=0. \tag{3.8}$$

Hence, the result follows from Khasminskii (2012). □

4 Predicting the End of Covid-19 Death

Since $N(t)$ denotes the aggregate death counts at a particular time t, it follows that dN/dt will describe the daily death counts. Following Theorems 2 and 6, we know that the daily death counts converges to zero as $N(t)$ converges asymptotically to $\mathcal{K}$. So, in order to calculate an approximate time that people will stop dying of Covid-19, we need to calculate, in an $\epsilon>0$ neighborhood of $\mathcal{K}$, a time when the aggregate number $\mathcal{K}-\epsilon$ is reached. For some positive small constant ϵ, define the open interval

$$\mathcal{A}_\epsilon=(\mathcal{K}-\epsilon,\mathcal{K}). \tag{4.1}$$

Following Theorem 6, we plan to calculate the first-hitting-time τ_ϵ until the process $N(t)$ enters $\mathcal{A}_\epsilon$.

Definition 2 We define the first passage time τ_ϵ as

$$\tau_\epsilon = \inf\left\{t > t_0 \ : \ N(t) = \mathcal{K} - \epsilon\right\}. \tag{4.2}$$

Let $g(t) = \mathbb{P}\left(\tau_\epsilon \le t\right)$ and $f_{\tau_\epsilon}(t) = dg(t)/dt$ be the First Passage Time Density (FPTD), the probability that the aggregate number of death counts $N(t)$ has first reached a point $\mathcal{K}-\epsilon$ at exactly time t. We derive $f_{\tau_\epsilon}(t)$ and the expected first hitting time in the theorem below for small ϵ.

Theorem 7 *If $N_0 \in (0, \mathcal{K})$, then the probability $f_{\tau_\epsilon}(t)$ that the aggregate number of death counts $N(t)$ has first reached a point $\mathcal{K} - \epsilon$ at exactly time t is obtained as*

$$f_{\tau_\epsilon}(t) = \frac{1}{\sigma\sqrt{2\pi(t-t_0)^3}} \ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right) \exp\left(-\frac{\left(\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right) - \mu(t-t_0)\right)^2}{2\sigma^2(t-t_0)}\right), \quad t > t_0.$$

Furthermore, for t_0=0, the expected first hitting time $\mathbb{E}\left(\tau_\epsilon\right)$ is obtained as

$$\mathbb{E}\left(\tau_\epsilon\right) = \frac{1}{\mu}\ln\left(\frac{(\mathcal{K}-\epsilon)\left(\mathcal{K}-N_0\right)}{\epsilon N_0}\right). \tag{4.3}$$

$$\mathbb{P}\left(\tau_\epsilon \le t\right) = \frac{1}{\sqrt{\pi}} \int_{\frac{\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right)-\mu(t-t_0)}{\sigma\sqrt{2(t-t_0)}}}^{\infty} e^{-y^2}\, dy + \frac{e^{\frac{2\mu}{\sigma^2}\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right)}}{\sqrt{\pi}} \int_{\frac{\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right)+\mu(t-t_0)}{\sigma\sqrt{2(t-t_0)}}}^{\infty} e^{-y^2}\, dy.$$

Proof

From this, and the Fundamental Theorem of Calculus, we have

$$\begin{aligned} f_{\tau_\epsilon}(t) &= \frac{d}{dt}\mathbb{P}\left(\tau_\epsilon \le t\right) \\ &= \frac{1}{\sigma\sqrt{2\pi(t-t_0)^3}} \ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right) \exp\left(-\frac{\left(\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right) - \mu(t-t_0)\right)^2}{2\sigma^2(t-t_0)}\right), \quad t > t_0, \end{aligned}$$

and for t_0=0,

$$\begin{aligned} \mathbb{E}\left(\tau_\epsilon\right) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \left(t^{-1/2}\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right)\right) \exp\left(-\frac{\left(\ln\left(\frac{\mathcal{K}-\epsilon}{\epsilon}\frac{\mathcal{K}-N_0}{N_0}\right) - \mu t\right)^2}{2\sigma^2 t}\right) dt \\ &= \frac{1}{\mu}\ln\left(\frac{(\mathcal{K}-\epsilon)\left(\mathcal{K}-N_0\right)}{\epsilon N_0}\right), \quad 0 < \epsilon < \mathcal{K}. \end{aligned}$$

□

Equation (4.3) and Theorem 7 can be used to calculate an approximate expected time when people in the United States will stop dying of the Covid-19 by making ϵ to be as small as possible.

Remark 4 For the solution $N_{\sigma=0}(t) = \frac{\mathcal{K}}{1+\frac{\mathcal{K}-N_0}{N_0}e^{-\mu t}}$ of (1.1), the time $T_\epsilon = \frac{1}{\mu}\ln\left(\frac{(\mathcal{K}-\epsilon)(\mathcal{K}-N_0)}{\epsilon N_0}\right)$ at which $N_{\sigma=0}(t) = \mathcal{K}-\epsilon$ satisfies $T_\epsilon = \mathbb{E}(\tau_\epsilon)$.

As shown in Sect. 1 with respect to model (1.1), the aggregate death count's growth speeds up in the interval $(0, \mathcal{K})$ and slows down in the interval $(\mathcal{K}/2, \mathcal{K})$. This shows that the maximum number of daily deaths can be calculated as $\max\left(\frac{dN}{dt}\right) = \frac{\mu}{4}\mathcal{K}$, occurring at time $T_{\mathcal{K}/2} = \frac{1}{\mu}\ln\left(\frac{\mathcal{K}-N_0}{N_0}\right) = \mathbb{E}(\tau_{\mathcal{K}/2})$. In the next theorem, we calculate the expected first time when the aggregate death counts reaches the size $\mathcal{K}/2$ for the stochastic case.

Corollary 8 *If $N_0 \in (0, \mathcal{K})$, we have $\lim_{\epsilon\to 0^+}\mathbb{E}(\tau_\epsilon) = +\infty$. That is, the process $N(t)$ never reaches the point $\mathcal{K}$. Also, $\mathbb{E}(\tau_{\mathcal{K}/2}) = \frac{1}{\mu}\ln\left(\frac{(\mathcal{K}-N_0)}{N_0}\right)$.*

Proof The proof follows from (4.3) and Theorem 7. □

The expected first time passage $\mathbb{E}(\tau_{\mathcal{K}/2})$ is analogous to the deterministic time $T_{\mathcal{K}/2}$ when the process $N(t)$ starts slowing down. That is, the time where the maximum number of daily deaths occurs.

As discussed in Sect. 2, we showed, with respect to the deterministic model (1.1), the current day's aggregate death count's growth is slowing down starting from the moment when the current day's aggregate count is more than $\mathcal{K}/2$. Otherwise, the virus is spreading, with speeding growth. Numerical results in Figs. 7 and 8 also show, for the stochastic case, that the expected aggregate count slows down starting from the time $\mathbb{E}(\tau_{\mathcal{K}/2})$.

5 Numerical Simulation and Forecast for the Aggregate Death Counts in the United States

As discussed earlier, $W(t)$ is a Wiener process that depends continuously on $t \in [0, T]$ with independent increment property such that $W(t_0 = 0) = 0$, $W(t) - W(s) \sim \sqrt{t-s}\,N(0,1)$ for $0 \le s < t \le T$, where $N(0, 1)$ is the standard normal distribution. Let $\Delta W_j = W_j - W_{j-1}$, $j = 1, 2, \cdots, m$, where $W_{j+1} = W(t_{j+1})$, $t_j = t_0 + (j-1)\Delta t_j$. We discretize the Wiener processes ΔW_j with time step Δt_j, and $W(t_j)$ as $\Delta W_j \sim \sqrt{\Delta t_j}\,N(0,1)$ and $W(t_j) = W(t_{j-1}) + \Delta W_j$, $j = 2, 3, \cdots, m$ with $W(t_1) = \Delta W_1$.

Let $\hat{N}_j$ be the discretized aggregate death counts satisfying the solution (2.7) at time t_j, $j = 1, 2, \cdots, m$. We estimate $\hat{N}_j$ using (2.7) as

$$\hat{N}_j = \frac{\hat{\mathcal{K}}}{1 + \frac{\hat{\mathcal{K}}-N_0}{N_0}\exp\left(-\hat{\mu}\,t_j - \hat{\sigma}\,W(t_j)\right)}, \quad j = 1, 2, \cdots, m, \tag{5.1}$$

Table 1 Parameter estimates using stochastic model (2.3) for the first day Covid-19 deaths is recorded to June 20, 2021

State	$\widehat{\mathcal{K}}$	$\hat{\mu}$	$\hat{\sigma}$	N_0	RMSE	RAMSE	$\mathbb{E}(\tau_{\mathcal{K}/2})$
AK	366	0.0235	0.0190	1.0172	11.3997	11.2450	250.6866 (23-Nov-2020)
AL	12269	0.0142	0.0111	237.8943	448.4912	434.3204	276.0435 (29-Nov-2020)
AR	6020.2	0.0208	0.0168	28.0941	136.1947	117.4587	257.9212 (07-Dec-2020)
AZ	19306	0.0137	0.0328	419.7449	996.1320	843.8170	280.1291 (25-Dec-2020)
CA	71273	0.0139	0.0212	648.3989	3806.2	3686.1	338.2706 (05-Jan-2021)
CO	6941.6	0.0128	0.0442	265.8343	438.0126	323.4614	258.6176 (29-Nov-2020)
CT	8824.2	0.0080	0.0434	1638.1	731.9427	539.3944	200.3560 (04-Oct-2020)
DE	2018.8	0.0085	0.0274	166.8309	97.1994	77.3889	288.3975 (08-Jan-2021)
FL	40794	0.0126	0.0000	1.2765	1191.6	1191.6	273.3731 (07-Dec-2020)
GA	25150	0.0113	0.0040	817.7736	645.5268	639.6664	300.7165 (07-Jan-2021)
HI	517.11	0.0178	0.0001	6.0115	16.7624	16.7624	249.3262 (05-Dec-2020)
IA	6308.2	0.0180	0.0334	58.1454	243.9217	184.7225	261.9713 (12-Dec-2020)
ID	2122.0	0.0210	0.0243	10.7009	49.0611	33.2558	251.9702 (03-Dec-2020)
IL	26890	0.0119	0.0358	1.3726	1407.4	1013.6	249.2160 (21-Nov-2020)
IN	14358	0.0153	0.0422	298.5343	803.2925	603.8770	255.4122 (27-Nov-2020)
KS	5259.6	0.0233	0.0270	6.1572	155.4604	142.3990	290.0863 (27-Dec-2020)
KY	8567.7	0.0130	0.0180	97.4734	202.7383	177.6680	344.4590 (25-Feb-2021)
LA	11488	0.0102	0.0166	1.1347	436.0073	375.6429	218.4941 (19-Oct-2020)
MA	19462	0.0084	0.0337	3.0196	1289.9	983.2224	210.8422 (15-Oct-2020)
MD	10656	0.0091	0.0335	995.9400	568.9196	397.9094	255.0706 (28-Nov-2020)
ME	868.01	0.0165	0.0563	7.0728	53.0840	40.3454	296.6184 (18-Jan-2021)
MI	22940	0.0085	0.0323	1862.2	1221.2	931.3025	292.8191 (20-Dec-2020)
MN	7759.4	0.0152	0.0372	170.5692	376.5030	277.8504	251.9335 (28-Nov-2020)
MO	9325.2	0.0196	0.0268	80.8314	302.6209	236.9409	243.3070 (16-Nov-2020)
MS	7649.2	0.0144	0.0230	264.3824	270.1491	214.1997	233.0122 (24-Oct-2020)
MT	1651	0.0237	0.0001	2.8173	26.8629	28.8074	269.1468 (20-Dec-2020)
NC	14141	0.0155	0.0210	207.0690	487.1274	427.1619	271.8087 (11-Dec-2020)
ND	1523	0.0338	0.0156	0.6061	37.0740	37.0678	231.9706 (14-Nov-2020)
NE	2294.8	0.0216	0.0402	11.5529	101.2707	77.5800	246.9092 (29-Nov-2020)
NH	1497.1	0.0106	0.0424	79.2622	88.0697	60.7974	280.1738 (28-Dec-2020)
NJ	26374	0.0091	0.0516	4.8141	2381.3	1602.9	180.5213 (07-Sep-2020)
NM	4380.8	0.0209	0.0535	22.4488	232.2403	160.4579	255.3504 (03-Dec-2020)
NV	6068.2	0.0146	0.0242	101.9782	225.0800	191.6449	280.5323 (22-Dec-2020)
NY	23267	0.0070	0.0359	3360.5	1663.7	1313.5	267.5434 (08-Dec-2020)
OH	23400	0.0123	0.0271	496.0223	1173	1068.3	314.31 (28-Jan-2021)
OK	7550.7	0.0219	0.0352	21.9981	249.1718	188.7447	267.6021 (12-Dec-2020)
OR	2785.0	0.0164	0.0287	25.2546	83.2097	60.4606	288.1383 (28-Dec-2020)
PA	2996.6	0.0113	0.0419	1240.5	1790.9	1334.7	283.8442 (27-Dec-2020)
RI	2986.6	0.0102	0.0375	252.4832	190.8369	143.3343	240.9393 (15-Nov-2020)
SC	10576	0.0141	0.0132	251.1180	391.9012	373.3641	264.1748 (05-Dec-2020)
SD	2027	0.0286	0.0380	0.9668	51.6401	43.1584	268.6586 (05-Dec-2020)

Table 1 (continued)

State	$\widehat{\mathcal{K}}$	$\hat{\mu}$	$\hat{\sigma}$	N_0	RMSE	RAMSE	$\mathbb{E}(\tau_{\mathcal{K}/2})$
TN	12757	0.0211	0.0325	51.1127	378.6108	261.3577	263.1342 (08-Dec-2020)
TX	53563	0.0154	0.0215	959.9323	1975.8	1734.3	260.4598 (30-Nov-2020)
UT	2424.9	0.0164	0.0270	24.7834	61.0702	37.3714	279.9737 (27-Dec-2020)
VA	14144	0.0102	0.0229	458.0540	542.5815	471.6913	336.5766 (15-Feb-2021)
VT	301.44	0.0100	0.0436	11.7290	19.7914	16.1092	330.1658 (12-Feb-2021)
WA	6749.6	0.0095	0.0179	377.1470	224.7571	191.6103	298.7084 (24-Dec-2020)
WI	8047	0.0200	0.0294	41.1695	285.9822	232.0415	264.1555 (09-Dec-2020)
WV	2890.9	0.0231	0.0314	4.2413	67.4599	47.9361	283.0646 (06-Jan-2021)
WY	734	0.0340	0.0180	0.1363	14.1754	14.4533	252.5651 (22-Dec-2020)

where $\widehat{\mathcal{K}}, \hat{\mu}, \hat{\sigma}$ are calculated in Sect. 3. In order to generate pseudo samples for each point $\hat{N}_j$, we define $\Delta W_j^l = W_j^l - W_{j-1}^l$, $j = 1, 2, \cdots, m$, $l = 1, 2, \cdots, L$, for sample size m and L number of simulations. Using Milstein scheme (Gaines and Lyons 1994), the l-th discretized solution $N_j^l \equiv N(t_j)^l$ of (2.3) satisfies

$$\begin{aligned} N_j^l &= N_{j-1}^l + \left(\frac{\hat{\mu}}{\hat{\mathcal{K}}}\left(\mathcal{K} - N_{j-1}^l\right)N_{j-1}^l + \frac{\hat{\sigma}^2}{2\hat{\mathcal{K}}^2}\left(\hat{\mathcal{K}} - N_{j-1}^l\right)\left(\hat{\mathcal{K}} - 2N_{j-1}^l\right)N_{j-1}^l\right)\Delta t_j \\ &\quad + \frac{\hat{\sigma}}{\hat{\mathcal{K}}}\left(\mathcal{K} - N_{j-1}^l\right)N_{j-1}^l \Delta W_j^l \sqrt{\Delta t_j} \\ &\quad + \frac{\hat{\sigma}^2}{2\hat{\mathcal{K}}^2} N_{j-1}^l \left(\hat{\mathcal{K}} - N_{j-1}^l\right)\left(\hat{\mathcal{K}} - 2N_{j-1}^l\right)\left(\left(\Delta W_j^l\right)^2 - 1\right)\Delta t_j, \end{aligned} \tag{5.2}$$

for $j = 1, 2, \cdots, m$, $l = 1, 2, \cdots, L$. We use the estimate (5.1), together with the estimated parameters $\hat{\mathcal{K}}$, $\hat{\mu}$ and $\hat{\sigma}$ in (3.5)–(3.6) to fit the aggregate COVID-19 death counts in the fifty states in the United States from the period when the first death count is recorded to June 20, 2021, and also to forecast from June 21, 2021 to June 24, 2021. Model (5.2) and (3.1) are used to generate the probability density function for the aggregate death counts for each time t_j. Let $\hat{N}_{j,\sigma=0}$ denote the deterministic equivalent of (5.1), which is the discretization of solution of (1.1) with $\sigma = 0$. The parameters in $\hat{N}_{j,\sigma=0}$ are also estimated using the Non-Linear Least Square estimate scheme (Coleman and Li 1996; May 1963) for model comparison purposes. Denote the root mean square error for the deterministic and stochastic discretization scheme by RMSE and RAMSE, respectively. We define

$$\begin{aligned} \text{RMSE} &= \left[\frac{1}{m}\sum_{j=1}^{m}\left(\hat{N}_{j,\sigma=0} - N(t_j)\right)^2\right]^{\frac{1}{2}}, \\ \text{RAMSE} &= \left[\frac{1}{m}\sum_{j=1}^{m}\left(\hat{N}_j - N(t_j)\right)^2\right]^{\frac{1}{2}}, \end{aligned} \tag{5.3}$$

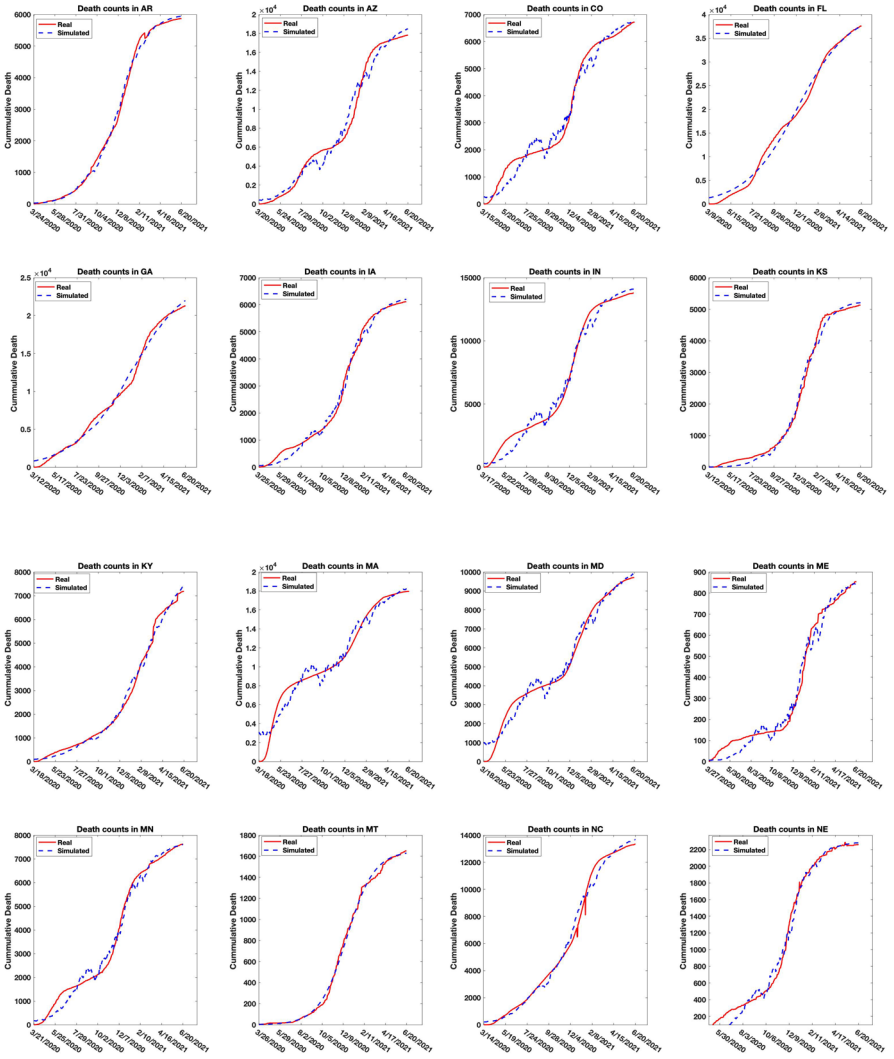

Fig. 1 Real and simulated aggregate counts of COVID-19 death counts for the states: AR, AZ, CO, FL, GA, IA, IN, KS, KY, MA, MD, ME, MN, MT, NC, NE in the United States

where $\{N(t_j)\}_{j=1}^{m}$ is the real aggregate death counts data. In order to show the superiority of model (2.3) over (1.1), we compare the root mean square errors RMSE and RAMSE and show that model (2.3) has a smaller root mean square error.

Table 1 shows the parameter estimates for the stochastic model (2.3) together with the root mean square errors RMSE and RAMSE in (5.3) for the deterministic and stochastic cases, respectively, using the Covid-19 aggregate death counts in the United States for the period when the first death case is reported to June 20, 2021. Here, N_0 denotes the estimate of the starting value when the first death case is reported. The expected first time the aggregate death counts is more than half its

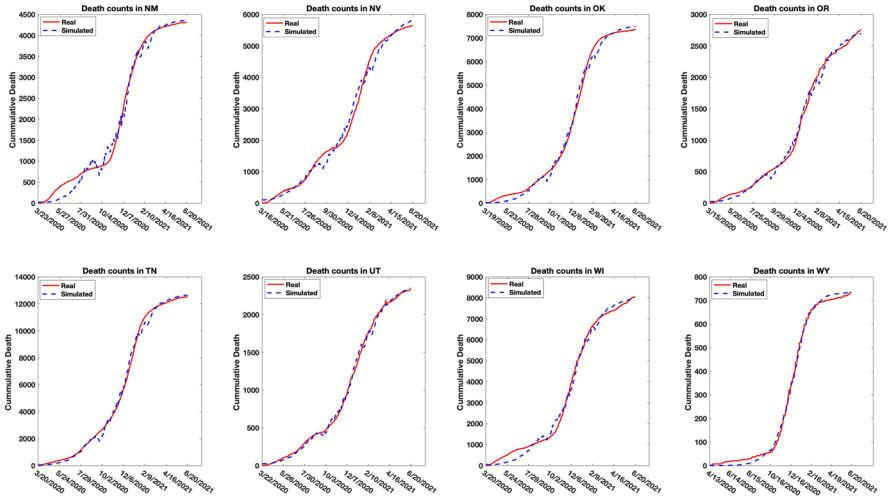

Fig. 2 Real and simulated aggregate counts of COVID-19 death counts for the states: NM, NV, OK, OR, TN, UT, WI, WY in the United States

carrying capacity is calculated. The root mean square error RMSE was calculated by first estimating the parameters in the deterministic model (1.1) using the Non-Linear Least Square estimate scheme (May 1963; Coleman and Li 1996). The estimated parameters for the deterministic model are not reported in this work. Within the analysis period, we see that the expected aggregate death counts started slowing down around mid December when the first vaccine was administered for most states. A quick comparison of RMSE and RAMSE in Table 1 shows that the stochastic model (2.3) performs better than the deterministic model (1.1) in describing the trajectory of the aggregate count of Covid-19 in the United States. In order to minimize space, we only show the real and simulated death counts for 24 out of 50 states in the United States in Figs. 1 and 2.

Figures 1 and 2 show the real and simulated aggregate death counts for some of the fifty states in the United States. In order to forecast the aggregate death counts from June 21, 2021 to June 24, 2021, we analyze the data starting from June 4, 2021 to June 20, 2021. The parameter estimates are shown in Table 2.

Table 2 contains parameter estimates derived using data set from June 4, 2021 to June 20, 2021. Here, N_0 denotes the estimate of the starting value for June 4, 2021. These estimates are used in the forecast for the aggregate death counts from June 21, 2021 to June 24, 2021. In Table 2, $N_{06/21/2021}$ denotes the forecast aggregate death counts estimate for June 21, 2021. The 95% confidence interval for the forecast estimate is also calculated and presented in Figs. 3 and 4.

Figures 3 and 4 show simulated and forecast estimate for the aggregate death counts of Covid-19 in the United States. The parameter estimates used for the simulation are given in Table 2 using the data set for June 4, 2021 to June 20, 2021. These parameters are used to forecast the aggregate death counts for June 21, 2021 to June 24, 2021.

Table 2 Parameter estimates using stochastic model (2.3) from June 4, 2021 to June 20, 2021

State	$\widehat{\mathcal{K}}$	$\hat{\mu}$	$\hat{\sigma}$	N_0	$N_{06/21/2021}$	$N_{06/22/2021}$	$N_{06/23/2021}$	$N_{06/24/2021}$
AK	368.4	0.0938	0.0179	360.41	366.75	366.87	367.01	367.11
AL	11441	0.0057	0.0001	11299	11312	11313	11314	11315
AR	6013.6	0.0131	0.0018	5842.1	5876.2	5877.7	5879.5	5881.0
AZ	28387	0.0015	31527	17665	17847	17857	17867	17877
CA	93187	0.00047	0.0001	62447	62620	62630	62639	62649
CO	6925.7	0.0337	0.0001	6593.5	6740.5	6746.4	6752.2	6757.8
CT	8279.6	0.0800	0.0001	8241.6	8270.6	8271.2	8271.9	8272.5
DE	3021.4	0.00057	0.0001	1675.0	1682.7	1683.1	1683.5	1683.9
FL	38744	0.0300	0.0039	36925	37643	37671	37703	37729
GA	28840	0.0042	0.0001	20917	21341	21364	21387	21410
HI	4824.5	0.0011	0.0001	498.64	507.73	508.24	508.75	509.26
IA	6156.3	0.0520	0.0001	6060.2	6118.2	6120.1	6121.9	6123.6
ID	2160.4	0.0295	0.0001	2101.3	2125.3	2126.3	2127.3	2128.2
IL	26192	0.0249	0.0001	25282	25603	25618	25631	25645
IN	18348	0.0014	0.0001	13702	13789	13794	13799	13804
KS	5320.7	0.0176	0.0064	5081.1	5138.8	5141.0	5144.3	5146.2
KY	7213.6	0.0634	0.0043	7166.4	7198.1	7199.0	7199.9	7200.7
LA	17722	0.0012	0.0001	10602	10696	10701	10706	10711
MA	18433	0.0088	0.00022	17898	17974	17978	17982	17985
MD	10000	0.0151	0.0001	9635.8	9720.3	9724.4	9728.4	9732.4
ME	1068.6	0.0075	0.0036	836.96	857.57	858.35	859.73	860.41
MI	21561	0.0263	0.0015	20541	20909	20925	20941	20956
MN	8094.2	0.0129	0.00039	7535.8	7644.0	7649.3	7654.8	7659.9
MO	9424.0	0.0266	0.00069	9141.8	9246.6	9251.0	9255.5	9259.8
MS	7601.6	0.0128	0.00069	7326.0	7380.2	7382.8	7385.5	7388.1
MT	2749.7	0.0024	0.0006	1628.7	1654.9	1656.1	1657.8	1658.9
NC	18353	0.0015	0.0001	13245	13342	13348	13353	13358
ND	1530.0	0.0605	0.0194	.514.2	1524.1	1524.4	1524.7	1524.9
NE	2259	0.2955	0.0978	2243.4	2258.9	2258.9	2258.9	2258.9
NH	1811.7	0.0027	0.0005	1353.2	1368.8	1369.5	1370.5	1371.2
NJ	26778	0.0158	0.0003	26259	26385	26391	26397	26403
NM	4526.7	0.0052	0.0012	4315.9	4332.9	4333.7	4334.7	4335.4
NV	10653	0.0013	0.0001	5592.0	5652.4	5655.7	5659.1	5662.4
NY	20320	0.0128	0.0003	19810	19912	19917	19922	19927
OH	20761	0.0190	0.0001	19948	20176	20187	20197	20208
OK	8005.6	0.0060	0.0026	7315.4	7373.4	7375.6	7379.4	7381.4
OR	2877.7	0.0304	0.0001	2680.4	2760.1	2763.5	2766.8	2770.0
PA	28203	0.0222	0.0008	27305	27592	27605	27618	27630
RI	2796.8	0.0064	0.0001	2717.1	2725.5	2726.0	2726.4	2726.9
SC	10210	0.0057	0.0004	9771.0	9811.3	9813.3	9815.5	9817.5
SD	2028.9	0.0982	0.0001	2020.4	2027.4	2027.6	2027.7	2027.8
TN	13509	0.0027	0.0001	12474	12520	12523	12525	12527
TX	51977	0.0236	0.0001	50572	51050	51071	51092	51112
UT	2395.8	0.0159	0.0001	2309.1	2330.1	2331.1	2332.1	2333.1
VA	11616	0.0282	0.0031	11218	11369	11375	11382	11388

Table 2 (continued)

State	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\sigma}$	N_0	$N_{06/21/2021}$	$N_{06/22/2021}$	$N_{06/23/2021}$	$N_{06/24/2021}$
VT	256.4	0.00005	0.0001	256.00	256.00	256.00	256.00	256.0
WA	12305	0.000	0.0009	5824.1	5809.7	5807.3	5807.9	5805.0
WI	8267.0	0.0263	0.0001	7942.5	8061.8	8067.0	8072.1	8077.0
WV	2882.0	0.0995	0.0002	2800.1	2868.0	2869.3	2870.5	2871.6
WY	3320.9	0.0021	0.0011	718.2092	737.1472	737.8835	739.2320	739.87

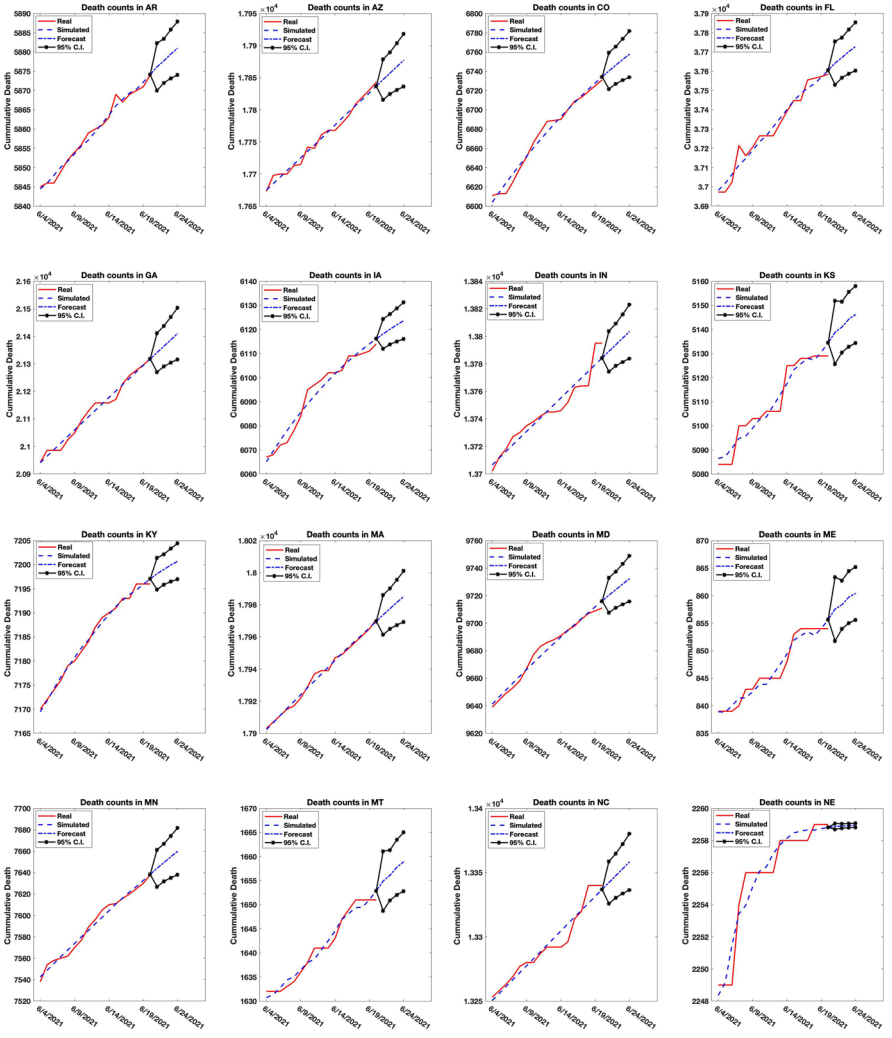

Fig. 3 Forecast of aggregate counts of COVID-19 death counts for the states: AR, AZ, CO, FL, GA, IA, IN, KS, KY, MA, MD, ME, MN, MT, NC, NE in the United States

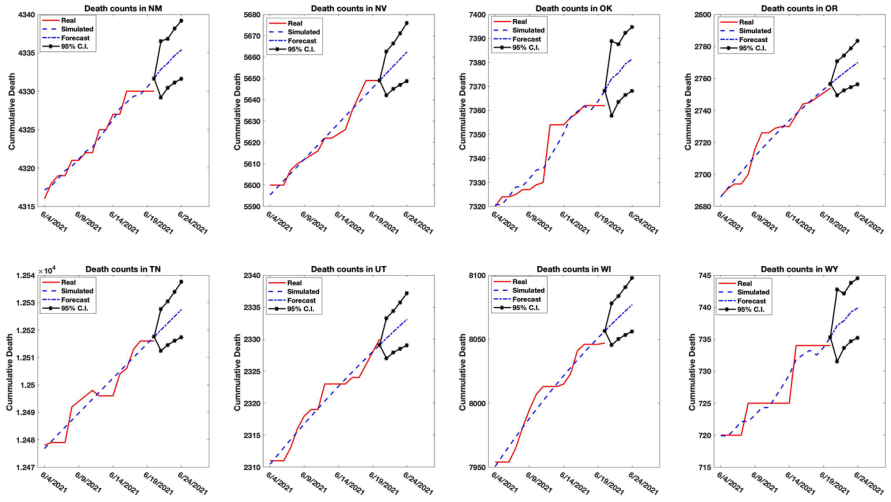

Fig. 4 Forecast of aggregate counts of COVID-19 death counts for the states: NM, NV, OK, OR, TN, UT, WI, WY in the United States

In order to verify the validity of the obtained probability density function (3.1), we show in the following graphs the comparison of the probability density function for the random variable N_T^l given in (5.2) for $l = 10,000$ simulations with the probability density function in (3.1) by setting $t = t_m = T$. The time $t = T$ corresponds to the day: June 24, 2021.

By generating the histogram of $\{N_T^l\}_{l=1}^{10,000}$ in (5.2), we show the comparison of the graphs of the probability density function for the random variable N_T^l and $p_N(N|T, N_0)$ in Figs. 5 and 6. The graphs show that the probability density function concentrates on a particular value. To know what this value is, we calculate the expected value of the aggregate death counts obtained in (3.7) for each of the fifty states in the United States and noticed the probability density function concentrates on the expected aggregate count $\mathbb{E}(N(T))$ on June 24, 2021, with $t = t_m = T$ denoting June 24, 2021. We also noticed that this value is close to the equilibrium point $\mathcal{K}$, which is the maximum aggregate death counts as at the time this research is conducted. Define

$$N_{T,\max} = \arg\max_{N\in(0,\mathcal{K})} p_N(N|T, N_0) \tag{5.4}$$

The comparison of the value $N_{T,\max}$ where the probability density function $p_N(N|T, N_0)$ concentrates on, with the expected aggregate count on June 24, 2021 is shown in Table 3. The plot of the expected aggregate death counts is plotted in Figs. 7 and 8 as a function of time.

Using the result obtained in (3.7), we plot the graph of the expected value of the aggregate death counts $\mathbb{E}\big(N(t)|N_0\big)$ with time for each states in the United States in Figs. 7 and 8. The graph shows that the expected value of the aggregate

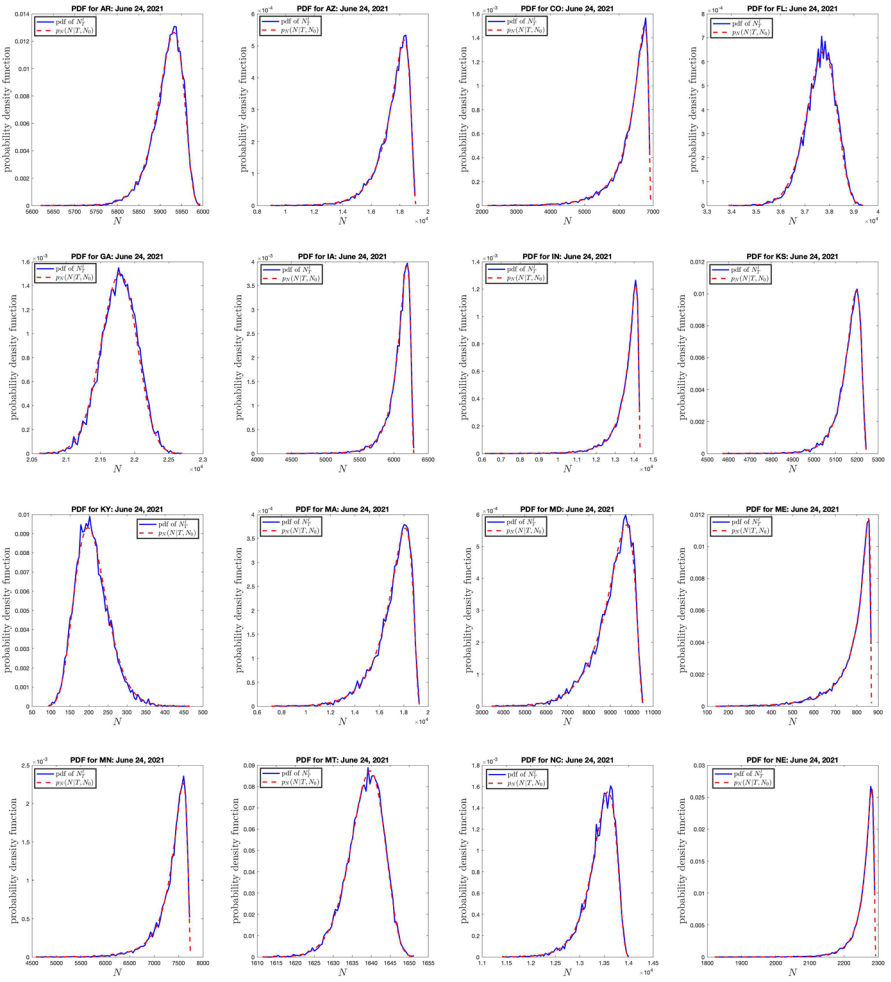

Fig. 5 Probability density function for COVID-19 death counts for the states: AR, AZ, CO, FL, GA, IA, IN, KS, KY, MA, MD, ME, MN, MT, NC, NE on June 24, 2021

death counts is increasing for $t \geq 0$. We also see from the graphs that the expected aggregate death counts started slowing down sometimes around the month of December 2020 for most states. This number is still slowing down as at the time this analysis is carried out (June 2021).

6 Summary and Discussion

In this work, we study and analyze the aggregate death counts $N(t)$ resulting from the Covid-19 virus infection in the United States. Recent studies use the deterministic logistic model to analyze this counts by assuming the death count's per capita

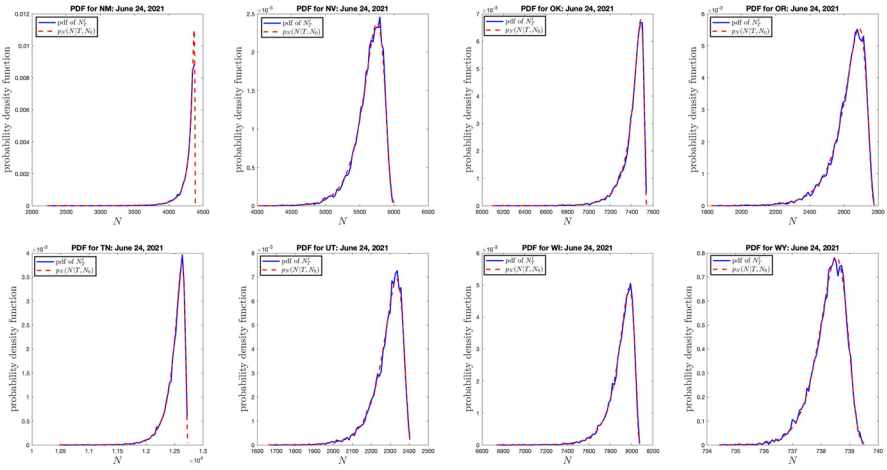

Fig. 6 Probability density function for COVID-19 death counts for the states: NM, NV, OK, OR, TN, UT, WI, WY on June 24, 2021

growth rate of the Covid-19 virus is constant over time. Our studies show that this is not the case. We assume, based on some analysis, that this growth rate can be affected by external perturbations causing fluctuations that can be modeled as a white noise described using a Wiener process. This assumption is used to modify and extend the existing logistic model to a stochastic differential equation. We analyze this model by first showing that it has a unique solution and its solution is bounded, with probability one. Analogue to an increasing function, we show that the process $N(t)$ is a submartingale that converges almost surely to a random variable on the long run if μ is greater than $\sigma^2/2$. By calculating the first hitting time when the aggregate death counts reach $\mathcal{K} - \epsilon$ for small $\epsilon > 0$, we calculate an approximate expected time when the death counts slow down for each state in the United States. To do this, we first calculate the probability of the first passage time τ_ϵ described in (4.2), and later calculate the expected first passage time. Our result shows that the aggregate death counts is now slowing down as at June 2021 when this analysis is conducted. By comparing the estimate $\mathcal{K}/2$ with the estimate $N(T)$ and $\mathbb{E}\left[N(T)|N_0\right]$ for the current day, June 24, 2021, our analysis shows that the Covid-19 death crisis slows down in the month of June in most states in the United States.

By deriving the transition probability density function for the process $N(t)$, we show that the expected aggregate death counts is bounded, and approaches $\mathcal{K}$ asymptotically. We also studied the distribution of the counts for the time T when this analysis is conducted and our result shows that the distribution concentrates on the expected aggregate count for that day (June 24, 2021).

Using the Maximum Likelihood Estimate scheme, we estimate the epidemiological parameters $\mathcal{K}$, μ, and σ using (3.3)–(3.4). These are used together with the Milstein scheme (5.2) to simulate and forecast the aggregate death counts for each states in the United States. A 95% confidence interval is provided for the forecast result. To show that our model (2.3) performs better than existing model (1.1), we compare

Table 3 Table containing $N_{T,\max}$ with the expected aggregate count $\mathbb{E}[N(T)]$ on June 24, 2021

State	$N_{T,\max}$	$\mathbb{E}\left[N(T)\vert N_0\right]$
AK	365	364
AL	11, 651	11, 603
AR	5932	5915
AZ	18, 316	17, 528
CA	65, 907	64, 308
CO	6752	6344
CT	8391	7680
DE	1700	1600
FL	37, 763	37, 592
GA	21, 786	21, 757
HI	504	503
IA	6188	6072
ID	2116	2106
IL	25, 766	24, 581
IN	14, 094	13, 583
KS	5198	5161
KY	7269	7153
LA	10, 711	10, 584
MA	18, 129	17, 085
MD	9747	9111
ME	854	780
MI	20, 256	18, 718
MN	7595	7367
MO	9234	9179
MS	7474	7401
MT	1640	1638
NC	13, 581	13, 412
ND	1526	1526
NE	2281	2257
NH	1399	1264
NJ	25, 936	24, 031
NM	4369	4279
NV	5754	5630
NY	20, 259	18, 310
OH	20, 735	19, 714
OK	7485	7408
OR	2689	2624
PA	28, 225	25, 783
RI	2822	2647
SC	10, 004	9943
SD	2026	2020
TN	12, 630	12, 507
TX	51, 732	51, 136

Table 3 (continued)

State	$N_{T,\max}$	$\mathbb{E}[N(T)\|N_0]$
UT	2334	2285
VA	11, 499	10, 979
VT	267	230
WA	5828	5687
WI	7977	7899
WV	2854	2822
WY	739	738

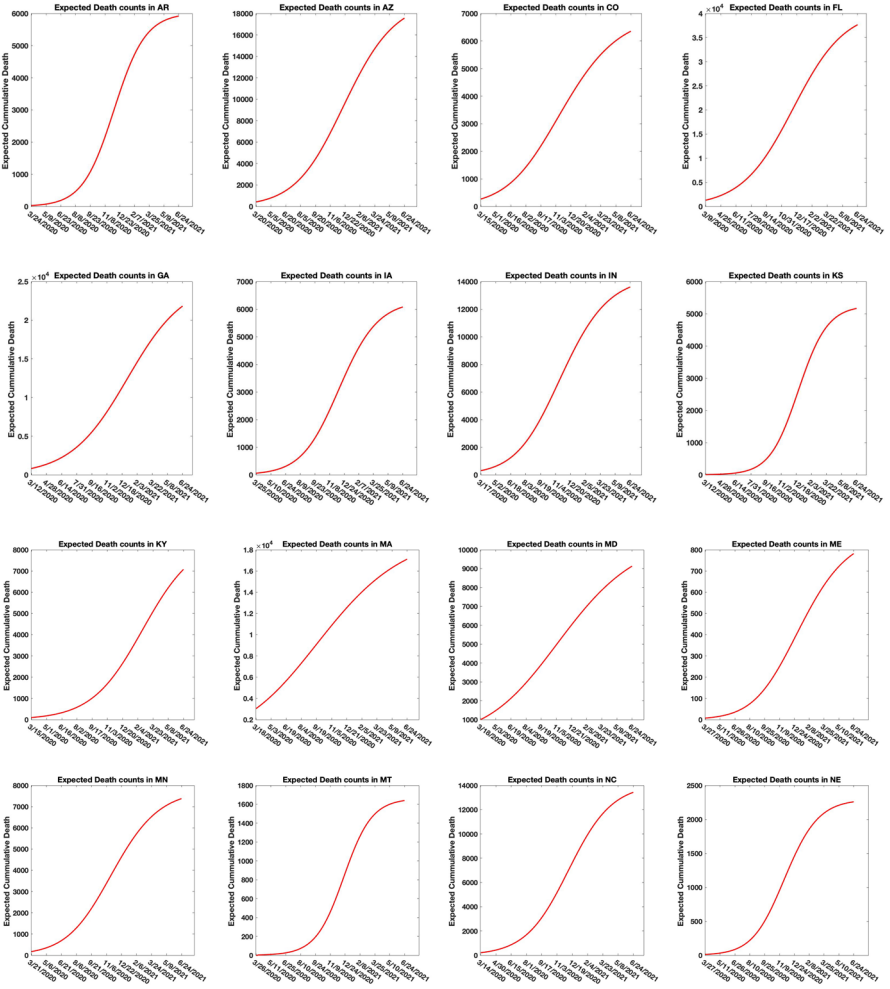

Fig. 7 Expected aggregate counts for the states: AR, AZ, CO, FL, GA, IA, IN, KS, KY, MA, MD, ME, MN, MT, NC, NE in the United States

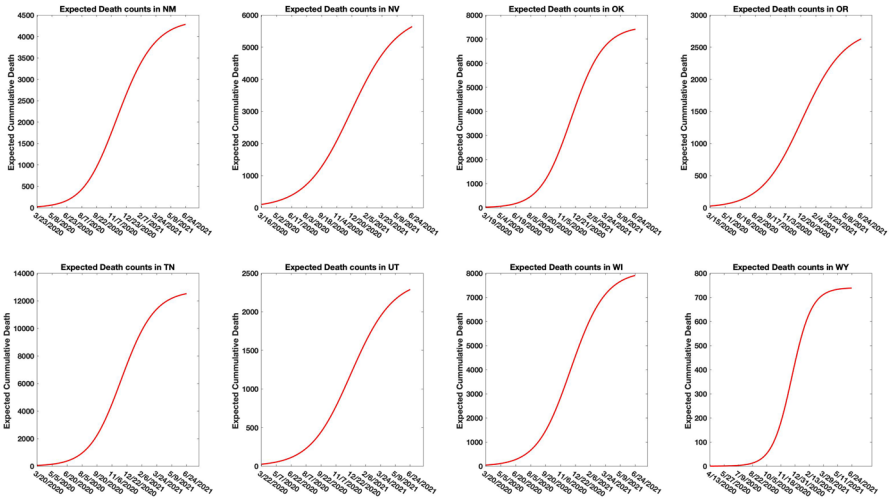

Fig. 8 Expected aggregate counts for the states: NM, NV, OK, OR, TN, UT, WI, WY in the United States

the root mean square errors RAMSE and RMSE for the stochastic and deterministic models, respectively, and show that RAMSE $<$ RMSE. This research is still ongoing and an update will be provided when available.

References

Arnold L (1974) Stochastic differential equations: theory and applications. Wiley, New York

Beddington JR, May RM (1977) Harvesting natural populations in a randomly fluctuating environment. Science 197:463–465

Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar Léo, Favre Guillaume (2020) Real estimates of mortality following COVID-19 infection. Lancet Infect Dis 20(7):773. https://doi.org/10.1016/S1473-3099(20)30195-X

Bhapkar HR, Mahalle PN, Dey N, Santosh KC (2020) Revisited COVID-19 mortality and recovery rates: are we missing recovery time period? J Med Syst 44(12):202

Coleman TF, Li Y (1996) An interior, trust region approach for nonlinear minimization subject to bounds. SIAM J Optim 6:418–45

Gaines JG, Lyons TJ (1994) Random generation of stochastic area integrals. SIAM J Appl Math 54(4):1132–1146

Gardiner CW (1985) Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer-Verlag, New York

Lv J, Liu H, Zou X (2019) Stationary distribution and persistence of a stochastic predator-prey model with a functional response. J Appl Anal Comput 9(1):1–11

Kaciroti NA, Lumeng C, Parekh V, Boulton ML (2021) A bayesian mixture model for predicting the COVID-19 related mortality in the United States. Am J Trop Med Hyg 104(4):1484–1492

Khasminskii R (2012) Stochastic stability of differential equations, 2nd edn. Springer-Verlag, Berlin Heidelberg, p 66

Kloeden PE, Platen E (1995) Numerical solution of stochastic differential equations. Springer-Verlag, New York

Lagarto S, Braumann CA (2014) Modeling human population death rates: ABi-dimensional stochastic Gompertz model with correlated wiener processes. In: Pacheco A, Santos R, Oliveira M, Paulino

C (eds) New advances in statistical modeling and applications. Studies in theoretical and applied statistics. Springer, Cham
Li W, Wang K (2010) Optimal harvesting policy for general stochastic logistic population model. J Math Anal Appl 368:420–428
Linka K, Peirlinck M, Kuhl E (2020) The reproduction number of COVID-19 and its correlation with public health interventions. Comput Mech 66(4):1035–1050. https://doi.org/10.1007/s00466-020-01880-8
Lungu EM, Øksendal B (1997) Optimal harvesting from a population model in a Stochastic Crowded Environment. Math Biosci 145:47–75
May RM (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math 11(2):431–41
Mao X (2007) Stochastic differential equations and applications, 2nd edn. Horwood, Chichester
Mazzuco S, Scarpa B, Zanotto L (2018) A mortality model based on a mixture distribution function. Popul Stud 72(3):1–10
Mendez V, Campos D, Horsthemke W (2012) Stochastic fluctuations of the transmission rate in the susceptible-infected-susceptible epidemic model. Phys Rev E 86:011919
Ndairou F, Area I, Nieto JJ, Torres DFM (2020) Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solitons Fractals 135:109846
Okuonghae D, Omame A (2020) Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria. Chaos Solitons Fractals 139:110032
Øksendal B (2003) Stochastic differential equations, An introduction with applications. Springer-Verlag, Berlin Heidelberg, New York
Ladde GS, Otunuga OM, Ladde NS (2020) Local lagged adapted generalized method of moments Dynamic Process. U.S. Patent Number: 10719578
Otunuga OM (2021) Time-dependent probability distribution for the number of infection in a stochastic SIS model: case study COVID-19. Chaos Solitons Fractals 147:110983
Otunuga OM (2021) Time-dependent probability density function for general stochastic logistic population model with harvesting effort. Phys A 573:1–33
Otunuga OM (2020) Qualitative analysis of a stochastic SEITR epidemic model with multiple stages of infection and treatment. Infect Dis Modell 5:61–90
Otunuga OM (2019) Closed-form probability distribution of number of infections at a given time in a stochastic SIS epidemic model. Heliyon 5:1–12
Mummert A, Otunuga OM (2019) Parameter identification for a stochastic SEIRS epidemic model: case study influenza. J Math Biol 79(2):705–729. https://doi.org/10.1007/s00285-019-01374-z
Otunuga OM (2018) Global stability for a 2n + 1 dimensional HIV/AIDS epidemic model with treatments. Math Biosci 5:138–52
Pelinovsky E, Kurkin A, Kurkina O, Kokoulina M, Epifanova A (2020) Logistic equation and COVID-19. Chaos Solitons Fractals 140:110241
Pella JS, Tomlinson PK (1969) A generalised stock-production model. Bull Int Am Trop Tuna Commun 13:421–496
Prajneshu (1980) Time dependent solution of the logistic model for population growth in random environment. J Appl Prob 17:1083–1086
Santosh KC (2020) COVID-19 prediction models and unexploited data. J Med Syst 44:170
Satpathy S, Mangla M, Sharma N, Deshmukh H, Mohanty S (2021) Predicting mortality rate and associated risks in COVID-19 patients. Spat Inf Res 29(4):455–464
Stutt Rojh, Retkute R, Bradley M, Gilligan CA, Colvin J (2020) A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic. Proc R Soc A 476:20200376
Wang P, Zheng X, Li J, Zhu B (2020) Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos Solitons Fractals 139:110058
Verhulst Pierre-Francois (1838) Notice sur la loi que la population poursuit dans son accroissement. Correspondance Mathématique et Physique. 10:113-121. Retrieved 3 Dec 2014
West BJ, Bulsara AR, Lindenberg K, Seshadri V, Shuler KE (1979) Stochastic processes with non-additive fluctuations: I. Itô and Stratonovich calculus and the effects of correlations. Phys A 97(2):211–233

Wong E, Zakai M (1965) On the convergence of ordinary integrals to stochastic integrals. Ann Math Stat 36(5):1560–1564

Wu JT, Leung K, Leung GM (2020) ' 'Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 35(10225):689–697

Yang B, Cai Y, Wang K, Wang W (2019) Optimal harvesting policy of logistic population model in a randomly fluctuating environment. Phys A 526:120817

Zocchetti C, Consonni D (1994) Mortality rate and its statistical properties. Med Lav 85(4):327–43