

A MARTINGALE FORMULATION FOR STOCHASTIC COMPARTMENTAL SUSCEPTIBLE-INFECTED-RECOVERED (SIR) MODELS TO ANALYZE FINITE SIZE EFFECTS IN COVID-19 CASE STUDIES

XIA LI

Department of Mathematics
University of California, Los Angeles

CHUNTIAN WANG

UA Department of Mathematics
The University of Alabama

HAO LI

Department of Mathematics
University of California, Los Angeles

ANDREA L. BERTOZZI

Departments of Mathematics and Mechanical and Aerospace Engineering
University of California, Los Angeles

ABSTRACT. Deterministic compartmental models for infectious diseases give the mean behaviour of stochastic agent-based models. These models work well for counterfactual studies in which a fully mixed large-scale population is relevant. However, with finite size populations, chance variations may lead to significant departures from the mean. In real-life applications, *finite size effects* arise from the variance of individual realizations of an epidemic course about its fluid limit. In this article, we consider the classical stochastic Susceptible-Infected-Recovered (SIR) model, and derive a martingale formulation consisting of a deterministic and a stochastic component. The deterministic part coincides with the classical deterministic SIR model and we provide an upper bound for the stochastic part. Through analysis of the stochastic component depending on varying population size, we provide a theoretical explanation of *finite size effects*. Our theory is supported by quantitative and direct numerical simulations of theoretical infinitesimal variance. Case studies of coronavirus disease 2019 (COVID-19) transmission in smaller populations illustrate that the theory provides an envelope of possible outcomes that includes the field data.

2020 *Mathematics Subject Classification.* Primary: 60G51; Secondary: 35Q91.

Key words and phrases. Applications of continuous-time Markov processes on discrete state spaces, applications of stochastic analysis (to PDE, etc.), PDEs in connection with game theory, economics, social and behavioral sciences, martingale representation, independent Poisson clocks.

This work is supported by NSF grant DMS-2027438, ARO MURI Grant W911NF1810208, NSF grant DMS-2027277, and Simons Math + X Investigator Award number 510776.

1. Introduction. The COVID-19 pandemic provides the first instance in 100 years to study the worldwide dynamics of a novel virus outbreak. Early on in the pandemic we observed many localized outbreaks on cruise ships, skilled nursing facilities, and through other congregating mechanisms such as conferences, parties, and places of worship. Several groups developed models to inform public health, the most notable being the March 16, 2020 report by Imperial College [24] that forecast over 2 million deaths in the US and over 500,000 deaths in the UK if non-pharmaceutical interventions (NPIs) were not implemented. Within days, lockdowns were enforced in both countries. Over a year since that time we have observed a variety of patterns, most of which did not follow the trends predicted by the early reports. The most obvious reason for these differences is the reduction of the reproductive number due to NPIs. An additional important factor is the departure from simple mixture models that results from isolating populations because of NPIs or simply because of geography. One way to model such effects is to consider an agent-based model within a finite population size in which the role of stochasticity is more pronounced due to the population size. This is the focus of our work.

Mathematical modeling has an important role in studying infectious diseases on both long and short timescales - for modern viruses such as AIDS, Severe Acute Respiratory Syndrome (SARS), Zika, and the novel coronavirus disease 2019 (COVID-19). For epidemic modeling, the standard approach involves compartmental models in which the population is divided into compartments representing individuals in one of several states, e.g. the susceptible (S), infectious (I), recovered (R) and exposed (E). This mathematical framework can lead to a variety of epidemic models, such as SIR, SIS, SEIR, SIRS (see e.g. [45], [49], [43], [9]). Compartmental modeling has been applied to epidemic study of COVID-19, HIV, etc [63, 50, 11, 7, 8]. These models can have an agent-based form with the most granular being the stochastic models. Between the stochastic SIR and the continuum limit are the Hawkes and HawkesN models in which a single self-exciting point process describes virus transmission between people [11]. Recent work has shown a theoretical connection between HawkesN and continuum compartmental models [59]. Because of the granularity of the stochastic compartmental model, its fluid limit is often adopted to reduce computational complexity. By the law of large numbers, stochastic compartment models with Markov processes can be approximated by their deterministic ODE counterpart. The central limit effect arises under the diffusion scaling. And the fluid-scaled model converges to a Gaussian diffusion of stochastic differential equations about the deterministic solution. All these approximation methods require a large population size to control variances. As the population size decreases the fluid-limit approximation acquires larger stochastic variances, and *finite size effects* may arise. For example, empirical variances of realizations of stochastic models can be at the same magnitude of the inverse of the total population size ([33, 34]). Since population scale within a relatively small community (like small counties, cruise ships) is very likely to fall within the regime of considerable stochastic deviations, quantitative study and analysis of *finite size effects* is relevant to real disease statistics. To this end, we investigate stochastic variability in the stochastic SIR model driven by independent Poisson clocks, which will be referred to as the SIR-IPC Model. We adopt Poisson clocks, rather than time steps discretized with a fixed duration. A more realistic model should treat all events as occurring independently, according to their own stochastic clock, rather than occurring at regular intervals, and arriving according to the same schedule. Time steps

are turned into exponentially distributed random variables with the introduction of Poisson clocks. Moreover, independent Poisson clocks allow us to use the theory of continuous-time Markov pure jump processes, e.g. a martingale approach (see e.g. [20, 44, 46, 47, 48]). And this leads to a martingale formulation. In prior literature, only the components of the martingale formulation are derived (see e.g. [1], [2], [72]), while study of the complete formulation has been lacking. Specifically, theoretical infinitesimal variances have also been studied in e.g. [14, 49], [2], [72], [73]. However, their main purpose is to explore ways to approximate stochastic compartmental models with a large population size, such as fluid limit, diffusion limit, and linear SDE approximations. With the martingale formula applicable to all population scales, we can quantitatively analyze the *finite size effects* under small populations.

It is possible to merge independent Poisson clocks to obtain systems that are driven by only one or two Poisson clocks. And this idea is worked out in [1], [2], [72]. There the SIR Models are driven by two Poisson clocks, one for transitions from S to I compartment, the other for those from I to R compartment. Here in our model, we assume that individual events each occur at a certain Poisson rate. These events include infectious contacts of each pair of susceptible and infected individuals, and recovery of each infectious individuals. We also assume the absence of vital dynamics with a deterministic population size parameter N . With the martingale approach, we decompose the process into a deterministic component and a stochastic component, corresponding to the infinitesimal means and variances of the process, respectively. The deterministic component leads to a fluid-limit-type continuum analogue, which coincides with the deterministic SIR model (rescaled over N) as in [38]. Computer simulations show that as N decreases, the SIR-IPC Model deviates further from the deterministic SIR model and *finite size effects* arise. The stochastic component is bounded by quantities of the same order of magnitude of $1/N$. We find that the square root of the stochastic component scales as $1/\sqrt{N}$. This is not a surprise, as it is expected based on the law of large numbers that the standard deviation scales like $1/\sqrt{N}$. Furthermore, the smaller the population size is, the larger the deviation of the stochastic simulations from the deterministic ones are. We provide a theoretical estimation for this scaling, building on the analysis of continuous-time Markov processes. Output of numerical experiments of theoretical infinitesimal variances supports our theory.

Furthermore, we conduct numerical experiments with several sets of parameters that mimic real world cases, including smaller US counties and a cruise ship. As the pandemic progresses, it permeates smaller counties and rural areas [51] [61]. Since the outbreak of COVID-19 on the Diamond Princess Cruise Ship at least 25 other such vessels have confirmed COVID-19 cases and studies of the transmission of the disease on cruise ship have drawn consideration attention (e.g. [6]).

The article is organized as follows. In Section 2, the stochastic SIR Model is introduced. The martingale formulation is derived in Section 3. Based on the martingale formulation, a deterministic and continuum analogue of the stochastic SIR model is derived (Section 3.3). Simulations are run to compare the stochastic SIR model to its continuum limit, illustrating the importance of *finite size effects* (Section 3.4). In Section 4 the *finite size effects* are analyzed theoretically based upon the martingale formulation. The theory is supported by simulations and field data in Section 5.

2. Stochastic SIR-IPC model.

2.1. Overview and notations. Our setting is similar to the continuous-time-Markov-processes stochastic SIR models in [1], [59] and [72]. We assume a continuous time variable $t \in [0, \infty)$, and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a deterministic and fixed population size N . The classical stochastic SIR model without vital dynamics consists of three compartments: susceptible, infected, and recovered individuals. For $\omega \in \Omega$, we assume that at a given time t the numbers of the compartments are $\mathbf{S}_N(\omega, t)$, $\mathbf{I}_N(\omega, t)$, and $\mathbf{R}_N(\omega, t)$, so that $N = \mathbf{S}_N(\omega, t) + \mathbf{I}_N(\omega, t) + \mathbf{R}_N(\omega, t)$, respectively. For simplicity of modeling, we assume that there is no vital dynamics, that is, we view deaths as a subset of recovered individuals and humans' natural resistance to the disease does not introduce new susceptible people after recovery. Below we assume that all the parameters are independent of N , unless otherwise specified. We assume that the initial data are deterministic and denoted as

$$(\mathbf{S}_N(\omega, 0), \mathbf{I}_N(\omega, 0), \mathbf{R}_N(\omega, 0)) = (S_{0N}, I_{0N}, R_{0N}). \quad (1)$$

A Poisson clock governs the infectious contact of any pair of a susceptible and an infectious individuals. There are in total $\mathbf{S}_N(\omega, t)\mathbf{I}_N(\omega, t)$ such pairs at time t . Each individual Poisson clock advances according to a Poisson process with rate β/N . We denote this clock as the “i-clock”. Suppose that the i-clock advances at time t^- . At time t , with probability $p_1 \in [0, 1]$ the susceptible of the pair associated with the clock contracts the disease and transitions to the infectious compartment through contact.

Another Poisson clock governs the recovery of the infectious individuals. Each infectious individual is assigned a Poisson clock, denoted as “r-clock”, which advances with rate γ . Suppose that the “r-clock” advances at time t^- . At time t , the infectious individual with the clock is recovered, with probability $p_2 \in [0, 1]$.

All the Poisson clocks are exponentially distributed independent random variables and are independent with time increments. As in the classical SIR literature, βp_1 denotes the transmission rate constant, and γp_2 denotes recovery rate constant. Thus the reproduction number $\mathcal{R}_0$ for our model is

$$\mathcal{R}_0 = \frac{\beta p_1}{\gamma p_2}. \quad (2)$$

Remark 1. To facilitate the computation of the SIR-IPC model, we show the following equivalent description of the SIR-IPC model with merged Poisson clocks. All the i-clocks can be merged into one master i-clock to govern the arrivals of infectious contact. Once the master i-clock advances, with probability p_1 a randomly chosen susceptible individual contracts the disease. The master i-clock advances with rate $\beta \mathbf{I}_N(\omega, t^-) \mathbf{S}_N(\omega, t^-) / N$. Likewise all the Poisson r-clocks can be merged into one master r-clock as we do for the i-clocks. Once the master r-clock advances, one randomly chosen infectious individual is recovered, and with probability p_2 develops immunity and becomes disinfected. The master r-clock advances according to a Poisson process with rate $\gamma \mathbf{I}_N(\omega, t^-)$. In general, theory regarding the merging and splitting of Poisson processes (see e.g. [21]) implies that independent Poisson clocks can be treated as one merged Poisson process; and with probability in proportion to the rate of each Poisson clock, the Poisson clocks compete to advance first in the merged Poisson process. More precisely, suppose that $X_1(t)$, $X_2(t)$, ..., $X_N(t)$ denotes the number of arrivals corresponding to independent Poisson processes with

arriving rates $\lambda_1, \dots, \lambda_N$, respectively. Then $\sum_{n=1}^N X_n(t)$ has the same distribution as the number of arrivals corresponding to one Poisson process with arriving rate $\sum_{n=1}^N \lambda_n(t)$. Furthermore, any particular arrival of the merged process has probability $\lambda_n / \sum_{n=1}^N \lambda_n(t)$ of originating from the n -th process, for $n = 1, 2, \dots, N$, independent of all other arrivals and their origins. The variance of the merged process at time t can be derived as follows:

$$\text{Var} \left(\sum_{n=1}^N X_n(t) \right) = \sum_{n=1}^N \text{Var} (X_n(t)) = \sum_{n=1}^N \lambda_n t = t \sum_{n=1}^N \lambda_n. \quad (3)$$

When $\lambda_n \equiv \lambda$ for all n , we have

$$\text{Var} \left(\sum_{n=1}^N X_n(t) \right) = tN\lambda. \quad (4)$$

2.2. Small-time-interval probabilities for compartment fractions. We denote the fractions of each compartment as

$$(\mathbf{s}_N(\omega, t), \mathbf{i}_N(\omega, t), \mathbf{r}_N(\omega, t)) := \left(\frac{\mathbf{S}_N}{N}(\omega, t), \frac{\mathbf{I}_N}{N}(\omega, t), \frac{\mathbf{R}_N}{N}(\omega, t) \right). \quad (5)$$

with the initial condition of population fractions denoted as

$$(\mathbf{s}_N(\omega, 0), \mathbf{i}_N(\omega, 0), \mathbf{r}_N(\omega, 0)) := (\mathbf{s}_{0N}, \mathbf{i}_{0N}, \mathbf{r}_{0N}). \quad (6)$$

This together with (1) implies that

$$(\mathbf{s}_{0N}, \mathbf{i}_{0N}, \mathbf{r}_{0N}) = \left(\frac{S_{0N}}{N}, \frac{I_{0N}}{N}, \frac{R_{0N}}{N} \right). \quad (7)$$

With similar idea as in e.g. [1], we derive the small-time-interval probabilities for compartment fractions. For Δt a short time interval, we have

$$\begin{aligned} \mathbb{P}(\Delta \mathbf{s}_N(\omega, t), \Delta \mathbf{i}_N(\omega, t)) &= (k, j) | (\mathbf{s}_N(\omega, t), \mathbf{i}_N(\omega, t) = (s, i)) \\ &= \begin{cases} p_1 \beta N i s \Delta t + o_1(\Delta t), & (k, j) = (-\frac{1}{N}, \frac{1}{N}), \\ p_2 \gamma N i \Delta t + o_2(\Delta t), & (k, j) = (0, -\frac{1}{N}), \\ 1 - p_1 \beta N i s \Delta t - p_2 \gamma N i \Delta t + o_3(\Delta t), & (k, j) = (0, 0), \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (8)$$

where $\Delta \mathbf{s}_N(\omega, t) = \mathbf{s}_N(\omega, t + \Delta t) - \mathbf{s}_N(\omega, t)$, $\Delta \mathbf{i}_N(\omega, t) = \mathbf{i}_N(\omega, t + \Delta t) - \mathbf{i}_N(\omega, t)$ and $o_l(\Delta t)$ are functions of Δt that satisfy $\lim_{\Delta t \rightarrow 0} o_l(\Delta t)/\Delta t = 0$, $l = 1, 2, 3$.

Remark 2. Although the set up of the Markov pure jump process implies (8), the inverse is not true. For example, the set of small-time-interval equations do not uniquely determine the rates of the Poisson clocks. For the same reason, combining the parameters p_1 and β as one parameter leads to a different stochastic SIR model. For one thing, the variance of a Bernoulli random variable with parameter p_1 is $p_1(1 - p_1)$, which is not linear with p_1 .

3. Main results using a martingale approach. The martingale formulation of a Markov pure jump process characterizes the process as the sum of an integral part involving the infinitesimal mean and a martingale part involving the infinitesimal variance.

3.1. Martingale formulation of the stochastic SIR-IPC model. For every t , we define $\mathbf{s}_N(t) := \{\mathbf{s}_N(\omega, t) : \omega \in \Omega\}$. In a similar way we can define the stochastic processes $\mathbf{i}_N(t)$ and $\mathbf{r}_N(t)$ associated with the stochastic SIR-IPC model. As $(\mathbf{s}_N(t), \mathbf{i}_N(t), \mathbf{r}_N(t))$ is a Markov pure jump process with state space $\mathbb{R}_+^3$, a martingale formulation (see e.g. [20, 21, 40, 48, 62]) can be derived. We first introduce the notion of blow-up time (see e.g. [31]).

Definition 3.1. It is possible that for a certain realization of SIR-IPC Model, the Poisson clocks may generate time increments that add up to be finite, say time τ and $\tau < \infty$. In this case the corresponding stochastic process is defined only for $0 \leq t < \tau$. By the time $t = \tau$ the Poisson clocks will have made an infinite number of advances. We say that the stochastic process explodes, and τ is called the blow-up time. This is an uncommon occurrence and it is a different issue from the process extinguishing itself, in which the infected population goes to zero in finite time. The latter is very real and of interest for finite size epidemics.

In the previous literature (e.g. [73, 72, 1, 2, 3, 14, 34, 35]), there are works that consider the infinitesimal mean and variance of stochastic SIR models but typically only in the cases when the total population N increase to infinity. In contrast, we are interested in the dynamics of the SIR-IPC model for moderate size N in which the finite size effect matters. Thus we derive the full martingale formulation encompassing the infinitesimal mean and variance of the SIR-IPC Model below.

Theorem 3.2. *Before the possible blow-up time, $(\mathbf{s}_N(t), \mathbf{i}_N(t), \mathbf{r}_N(t))$ can be written as*

$$\begin{cases} \mathbf{s}_N(t) = \mathbf{s}_{0N} + \int_0^t \mathcal{G}_N^{(1)}(w) dw + \mathcal{M}_N^{(1)}(t), \\ \mathbf{i}_N(t) = \mathbf{i}_{0N} + \int_0^t \mathcal{G}_N^{(2)}(w) dw + \mathcal{M}_N^{(2)}(t), \\ \mathbf{r}_N(t) = \mathbf{r}_{0N} + \int_0^t \mathcal{G}_N^{(3)}(w) dw + \mathcal{M}_N^{(3)}(t), \end{cases} \quad (9)$$

where $\mathcal{M}_N^{(l)}(t) = \mathcal{M}^{(l)}(\mathbf{s}_N(t), \mathbf{i}_N(t), \mathbf{r}_N(t))$, $l = 1, 2, 3$, are martingales that start at $t = 0$ as zeros, and $\mathcal{G}_N^{(l)}(t) = \mathcal{G}^{(l)}(\mathbf{s}_N(t), \mathbf{i}_N(t), \mathbf{r}_N(t))$, $l = 1, 2, 3$, are the infinitesimal means for $\mathbf{s}_N(t)$, $\mathbf{i}_N(t)$, and $\mathbf{r}_N(t)$, respectively, and

$$\mathcal{G}_N^{(1)}(t) = -p_1 \beta \mathbf{i}_N(t) \mathbf{s}_N(t), \quad (10)$$

$$\mathcal{G}_N^{(2)}(t) = p_1 \beta \mathbf{i}_N(t) \mathbf{s}_N(t) - p_2 \gamma \mathbf{i}_N(t) \quad (11)$$

$$\mathcal{G}_N^{(3)}(t) = p_2 \gamma \mathbf{i}_N(t). \quad (12)$$

The variances of $\mathcal{M}_N^{(l)}(t)$, $l = 1, 2, 3$, can be characterized in the following way:

$$\text{Var}(\mathcal{M}_N^{(l)}(t)) = \int_0^t \mathbb{E}[\mathcal{V}_N^{(l)}(w)] dw, \quad l = 1, 2, 3, \quad (13)$$

where $\mathcal{V}_N^{(l)}(t) = \mathcal{V}_N^{(l)}((s_N(t), i_N(t), r_N(t)))$, $l = 1, 2, 3$, are the infinitesimal variances for $s_N(t)$, $i_N(t)$, and $r_N(t)$ respectively, and

$$\mathcal{V}_N^{(1)}(t) = \frac{1}{N} p_1 \beta i_N(t) s_N(t), \quad (14)$$

$$\mathcal{V}_N^{(2)}(t) = \frac{1}{N} p_1 \beta i_N(t) s_N(t) + \frac{1}{N} p_2 \gamma i_N(t), \quad (15)$$

$$\mathcal{V}_N^{(3)}(t) = \frac{1}{N} p_2 \gamma i_N(t). \quad (16)$$

Proof of Theorem 3.2. To prove Theorem 3.2, we compute the infinitesimal means and variances for the Markov pure jump process $(s_N(t), i_N(t), r_N(t))$ for a fixed N , using the methods in e.g., [5, 16, 22, 30, 37, 44, 52, 53, 56, 58].

As $\mathcal{G}_N^{(l)}(t)$, $l = 1, 2, 3$, are the infinitesimal means for $s_N(t)$, $i_N(t)$, and $r_N(t)$, respectively, from (8) we have (10) - (12). As $\mathcal{V}_N^{(l)}(t)$, $l = 1, 2, 3$, are the infinitesimal variance of $s_N(t)$, $i_N(t)$, and $r_N(t)$, respectively, we have

$$\mathcal{V}_N^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{E} \left[(\Delta s_N(t^-))^2 \middle| (s_N(t^-), i_N(t^-), r_N(t^-)) \right] \quad (17)$$

From (17) we obtain (14). In a similar way we obtain (15), and (16). With the infinitesimal means and variances obtained, we apply Theorem (1.6), [20] or Theorem 3.32, [48], to obtain (9), and apply Exercise 3.8.12 of [13] to obtain (13). The proof of Theorem 3.2 is completed. $\square$

3.2. Estimates of the martingale variances. Here we derive upper bounds for variances of martingales of the martingale formulation (9).

Theorem 3.3. *Before the possible blow-up time, we have the following estimates:*

$$\text{Var} \left(\mathcal{M}_N^{(1)}(t) \right) \leq \frac{1}{N} s_{0N}, \quad (18)$$

$$\text{Var} \left(\mathcal{M}_N^{(2)}(t) \right) \leq \frac{1}{N} [s_{0N} + p_2 \gamma (i_{0N} + s_{0N}) t], \quad (19)$$

$$\text{Var} \left(\mathcal{M}_N^{(3)}(t) \right) \leq \frac{1}{N} p_2 \gamma (i_{0N} + s_{0N}) t. \quad (20)$$

Proof of Theorem 3.3. Taking expectation on both sides of (9)₁, we have

$$\mathbb{E}[s_N(t)] = s_{0N} + \int_0^t \mathbb{E}[\mathcal{G}_N^{(1)}(w)] dw. \quad (21)$$

This together with (10), (13) and (14) implies

$$\text{Var} \left(\mathcal{M}_N^{(1)}(t) \right) = \frac{1}{N} \int_0^t \mathbb{E}[p_1 \beta i_N(w) s_N(w)] dw = \frac{1}{N} (s_{0N} - \mathbb{E}[s_N(t)]). \quad (22)$$

As $\mathbb{E}[s_N(t)] \geq 0$, we can bound the right-hand-side of (22) from above by s_{0N}/N , and obtain (18).

From (13) and (16) we infer

$$\begin{aligned} \text{Var} \left(\mathcal{M}_N^{(3)}(t) \right) &= \frac{1}{N} p_2 \gamma \int_0^t \mathbb{E}[i_N(w)] dw \\ &\leq \frac{1}{N} p_2 \gamma \int_0^t (i_{0N} + s_{0N}) dw \\ &= \frac{1}{N} p_2 \gamma (i_{0N} + s_{0N}) t. \end{aligned} \quad (23)$$

From (23) we obtain (20).

From (13) - (16) and we obtain

$$\text{Var} \left(\mathcal{M}_N^{(2)}(t) \right) = \text{Var} \left(\mathcal{M}_N^{(1)}(t) \right) + \text{Var} \left(\mathcal{M}_N^{(3)}(t) \right). \quad (24)$$

This together with (18) and (20) implies (19). $\square$

Theorem 3.3 implies that the variances of the martingales of $\mathbf{s}_N(t)$, $\mathbf{i}_N(t)$, $\mathbf{r}_N(t)$ are bounded by the initial states of each compartment, and the total population N . Moreover, they are bounded by quantities with an equal or lower order of magnitude than $1/N$.

3.3. Deterministic analogue of the stochastic SIR model. With a similar derivation of the hydrodynamic limit of Markov pure jump processes [23, 27, 40, 39, 62, 64, 68, 69], we can find a deterministic analogue of the (renormalized) stochastic SIR model when N increases to infinity. The analysis is based on the martingale formulation (9), and the conclusion of Theorem 3.3, where the martingale variances are shown to have an equal or lower order of magnitude than $1/N$.

And we see that for N large, it is reasonable to set the the infinitesimal mean vector as the generator of the deterministic flow of a deterministic analogue of the stochastic SIR model. And by (10), (11) and (12) we obtain

$$\begin{cases} \frac{\partial s(t)}{\partial t} = -p_1 \beta i(t) s(t), \\ \frac{\partial i(t)}{\partial t} = p_1 \beta i(t) s(t) - p_2 \gamma i(t), \\ \frac{\partial r(t)}{\partial t} = p_2 \gamma i(t), \\ s(0) = s_0, \quad i(0) = i_0, \quad r(0) = r_0, \end{cases} \quad (25)$$

where $s(t)$, $i(t)$, and $r(t)$, $t \in (0, \infty)$ are the deterministic versions of $\mathbf{s}_N(t)$, $\mathbf{i}_N(t)$, and $\mathbf{r}_N(t)$, respectively, and $s_0 = S_0/N$, $i_0 = I_0/N$, $r_0 = R_0/N$. We note that the ODEs given in Eqs. (25) does not explicitly include the population size parameter N , and is independent of N . Note that the deterministic analogue of the stochastic SIR model (25) is the same as the classical deterministic SIR model (without vital dynamics) as in e.g. [38] rescaled by N .

3.4. Numerical simulations. We compare the stochastic and deterministic analogue of the stochastic SIR model through simulations. For the stochastic case, we use the classical Gillespie algorithm for continuous-time Markov processes (see e.g. [25]). Roughly speaking, we first simulate the sojourn times using exponential distribution generators, and then simulate the sample paths of the embedded discrete-time Markov process as described in Section 2.1. The algorithm ensures that independent Poisson clocks advance randomly without a prescribed diagram. Namely, we do not require that the i-clocks and r-clocks advance sequentially with a pre-determined arrangement. In this way, we are able to produce a number of sample paths corresponding to random realizations of sequences of events. The predictive power of our simulation is strengthened by this effort of proper usage of the algorithm.

We focus on the cases when the reproduction number $\mathcal{R}_0$ is near the self-sustaining level of $\mathcal{R}_0 = 1$, as is typical in a scenario where public health measures trade off

against economic and educational needs of the population¹. Specifically, we consider the cases when $\mathcal{R}_0 = 0.95, 1.1, 1.2$, and 1.3 , respectively. Parameters of the simulations are recorded in figure caption.

In all the cases, we show output of the SIR-IPC model simulations with N taking on values $10^{2.5}, 10^3, 10^{3.5}$, and 10^4 , by the blue, orange, green, and red lines, respectively, in Figs. 1b, 1d, 1f, 1h. For the case when $\mathcal{R}_0 = 0.95$, we infer from (25) that the parameters and data used to create the blue, orange, green, and red lines in Fig. 1b give rise to the same set of solution $(s(t), i(t), r(t))$ to the deterministic SIR Model. Therefore, we only display the deterministic output once in Fig. 1a. This same arrangement also applies to other cases with different values of $\mathcal{R}_0$, which are displayed in Figs. 1c, 1e, and 1g, respectively. In all the figures, black, magenta, and cyan lines represent $s(t)$, $i(t)$, and $r(t)$ associated with the deterministic simulation, respectively.

The *finite size effects* are exhibited. As the population size decreases, deviations of dynamics of the stochastic SIR Model from its continuum equation increase. The same simulation output is also observed over other random paths.

4. Mathematical analysis of the *finite size effects*. In Theorem 3.3, a global upper bound of the same order of magnitude of $1/N$ is derived and the result indicates the order of the magnitude of the martingale infinitesimal variances when $N \rightarrow \infty$. To further understand the martingale variances when N is relatively smaller, in this section, we analyze the *finite size effects*, based on the martingale formulation, and simulations are run which support our theoretical conclusion.

4.1. Scaling property of the stochastic component. We analyze the deterministic and stochastic component of the martingale formulation with varying population size N . We fix the initial condition of the compartment fractions and denote them as

$$(s_{0N}, i_{0N}, r_{0N}) \equiv (s_0, i_0, r_0). \quad (26)$$

We analyze the martingale formulation of the stochastic SIR model with varying population size N . Applying (9) and (13) to $(s_N(t), i_N(t), r_N(t))$ over a small time step Δt , we obtain

$$\begin{cases} s_N(t + \Delta t) = s_N(t) + \mathcal{G}_N^{(1)}(t)\Delta t + \mathcal{M}_N^{(1)}(t + \Delta t) - \mathcal{M}_N^{(1)}(t), \\ i_N(t + \Delta t) = i_N(t) + \mathcal{G}_N^{(2)}(t)\Delta t + \mathcal{M}_N^{(2)}(t + \Delta t) - \mathcal{M}_N^{(2)}(t), \\ r_N(t + \Delta t) = r_N(t) + \mathcal{G}_N^{(3)}(t)\Delta t + \mathcal{M}_N^{(3)}(t + \Delta t) - \mathcal{M}_N^{(3)}(t). \end{cases} \quad (27)$$

By (13) and additivity of the variance in time for martingales, we have

$$\sqrt{\text{Var}(\Delta \mathcal{M}_N^{(l)}(t))} \cong \sqrt{\mathbb{E}[\mathcal{V}_N^{(l)}(t)] \Delta t}, \quad l = 1, 2, 3, \quad (28)$$

where $\Delta \mathcal{M}_N^{(l)}(t) = \mathcal{M}_N^{(l)}(t + \Delta t) - \mathcal{M}_N^{(l)}(t)$, $l = 1, 2, 3$. This together with (27) implies that the infinitesimal variances are the key to estimate the infinitesimal standard deviation of the stochastic component, and the deviation of the trajectories of the evolution of the model from its deterministic component.

¹As a specific example we note that the dynamic reproductive number has been estimated weekly in Los Angeles County using a Bayesian SEIR model applied to hospital demand data - during the period July 2020-April 2021 it remained within a 25% window of one [10].

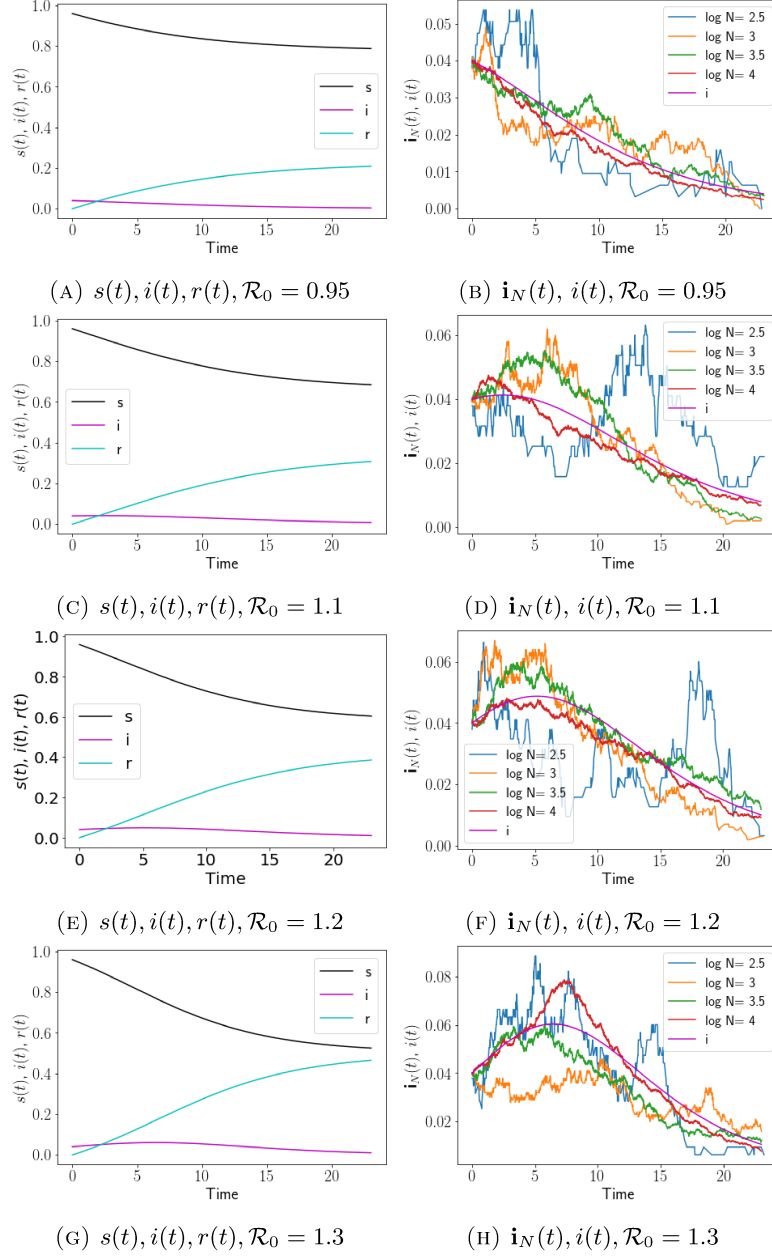


FIGURE 1. Plots of infected compartment fractions $i_N(t)$ for the stochastic SIR and $s(t), i(t)$, and $r(t)$ for deterministic SIR. For both models, $p_1 = 0.5$, $p_2 = 0.5$, $\gamma = 1$, $T = 23$, $s_0 = s_{0N} = 0.96$, $i_0 = i_{0N} = 0.04$, $r_0 = r_{0N} = 0$, and In Figs. 1b and 1a, 1d and 1c, 1f and 1e, and 1h and 1g, $\beta = 0.95, 1.1, 1.2, 1.3$, respectively. For the SIR-IPC model displayed in Figs. 1b, 1d, 1f, and 1h, $N = 10^{2.5}, 10^3, 10^{3.5}, 10^4$ in every panel.

We perform estimates at the first time step. At time zero, from (14) - (16) we infer

$$V_N^{(1)}(0) = \frac{1}{N} p_1 \beta i_0 s_0, \quad (29)$$

$$\mathcal{V}_N^{(2)}(0) = \frac{1}{N} (p_1 \beta i_0 s_0 + p_2 \gamma i_0), \quad (30)$$

$$\mathcal{V}_N^{(3)}(0) = \frac{1}{N} p_2 \gamma i_0, \quad (31)$$

which implies that the infinitesimal variances for the attractiveness are each inversely proportional to N :

$$\mathcal{V}_N^{(l)}(0) \propto \frac{1}{N}, l = 1, 2, 3. \quad (32)$$

This together with (28) implies that at the first time step we have

$$\text{Var} \left(\mathcal{M}_N^{(l)}(\Delta t) \right) \propto \frac{1}{N}, l = 1, 2, 3. \quad (33)$$

From (33) and (27) we infer that at the first time step a smaller value of N leads to a larger deviation of the trajectory of $(\mathbf{s}_N(\omega, t), \mathbf{i}_N(\omega, t), \mathbf{r}_N(\omega, t))$ from its deterministic component, and the hotspots develop temporal transience in the simulations. This explains the *finite size effects* at the first time step. This suggests that smaller value of N leading to a larger deviation remains to be true at an arbitrary later time, namely,

$$\mathcal{V}_N^{(l)}(t) > \mathcal{V}_{\tilde{N}}^{(l)}(t), \text{ for } 0 < N < \tilde{N} \text{ and } t > 0, l = 1, 2, 3, \quad (34)$$

which leads to a theory of the *finite size effects* at an arbitrary later time.

Next we estimate the right-hand-side of formulas (14) - (16), so that we can estimate $\mathcal{V}_N^{(l)}(t), l = 1, 2, 3$, at later times for $t > 0$. In the prior literature of stochastic partial differential equations (see e.g. [12, 17, 57, 67, 71]), usually Itô calculus is applied, as the infinitesimal variances are multiplied by Wiener processes, and the population or number of particles is required to be large, which does not fit with the cases on which our focus here.

We conjecture that formulas (29) - (31) can be written for an arbitrary $t > 0$ with the same leading order:

$$\mathcal{V}_N^{(1)}(t) \propto \frac{1}{N} p_1 \beta i_0 s_0, \quad (35)$$

$$\mathcal{V}_N^{(2)}(t) \propto \frac{1}{N} (p_1 \beta i_0 s_0 + p_2 \gamma i_0), \quad (36)$$

$$\mathcal{V}_N^{(3)}(t) \propto \frac{1}{N} p_2 \gamma i_0. \quad (37)$$

If our conjecture above is true, then we can fix a time period $[T_1, T_2]$, and take average of both sides of (35) - (37) over the time period $[T_1, T_2]$, and obtain

$$\overline{\mathcal{V}_N^{(1)}} \propto \frac{1}{N} p_1 \beta i_0 s_0, \quad (38)$$

$$\overline{\mathcal{V}_N^{(2)}} \propto \frac{1}{N} (p_1 \beta i_0 s_0 + p_2 \gamma i_0), \quad (39)$$

$$\overline{\mathcal{V}_N^{(3)}} \propto \frac{1}{N} p_2 \gamma i_0, \quad (40)$$

where $\overline{\mathcal{V}_N^{(l)}}$, $l = 1, 2, 3$ denotes the time average of the infinitesimal variance over this time period:

$$\overline{\mathcal{V}_N^{(l)}} := \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \mathcal{V}_N^{(l)}(t) dt, \quad l = 1, 2, 3. \quad (41)$$

Here we slightly abuse the notation and omit the dependence of the time average over T_1 and T_2 .

4.2. Numerical simulations. To check the validity of (34), we perform direct simulations of the infinitesimal standard deviations, namely

$$\sigma_N^{(l)}(t) = \sqrt{\overline{\mathcal{V}_N^{(l)}}}, \quad l = 1, 2, 3. \quad (42)$$

Example output can be found in Figure. 2.

Figs. 2a, 2c, and 2e show results of $\sigma_N^{(l)}(t)$, $l = 1, 2, 3$, in the case with $\mathcal{R}_0 = 0.95 < 1$. The blue, orange, green and red lines show results with the same realizations from simulations with the corresponding colors in Fig. 1b, respectively. Figs. 2b, 2d, and 2f show results of $\sigma_N^{(l)}(t)$, $l = 1, 2, 3$, in the case with $\mathcal{R}_0 = 1.3 > 1$. The blue, orange, green and red lines show results with the same realizations from simulations with the corresponding colors in Fig. 1h, respectively.

The output of the simulations supports the validity of (34). The same simulation results are also observed over other random paths.

We validate our conjecture of (38) - (40) through the following numerical simulation. Fig. 3 shows the log-log plot with error bars for (38) - (40). The lines show the theoretical scaling with slope as -1 and the x -intercepts as $\log(p_1\beta i_0 s_0)$, $\log(p_1\beta i_0 s_0 + p_2\gamma i_0)$, and $\log(p_2\gamma i_0)$, respectively, and the error bars show the true scaling with the x -coordinate and y -coordinate as follows:

$$(x, y) = \left(\log N, \log \left(\overline{\mathcal{V}_N^{(l)}} \right) \right), \quad l = 1, 2, 3, \quad (43)$$

for $N = 10^{2.5}, 10^3, 10^{3.5}, 10^4$, and $l = 1, 2, 3$. Here $[T_1, T_2]$ are chosen as $[0, 10]$, $[1, 11]$, $[2, 12]$, ..., and $[13, 23]$. The minimum and maximum values of $y = \log \left(\overline{\mathcal{V}_N^{(l)}} \right)$, $l = 1, 2, 3$, taken over all such intervals are set as the limits of the error bars.

Figs. 3a, 3c, and 3e show results in the cases when $\mathcal{R}_0 = 0.95$ for $l = 1, 2$, and 3 , respectively. The error bars with x -axis as $2.5, 3, 3.5$, and 4 show results with the same realization from simulations of the blue, orange, green, and red lines in Fig. 1b, respectively. Figs. 3b, 3d, and 3f show results in the cases when $\mathcal{R}_0 = 0.13$ for $l = 1, 2$, and 3 , respectively. The error bars with x -axis as $2.5, 3, 3.5$, and 4 , show results with the same realization from simulations of the blue, orange, green, and red lines in Fig. 1h, respectively.

The output shows that the error bars are short and fall mostly on the straight lines representing the theory. The same simulation results are also observed over other random paths. These results support the validity of equations (38) - (40) and our theory for the *finite size effects* based on the martingale formulation.

5. Case studies of a small county and a cruise ship. In the following experiments, we set parameters to mimic real world cases of a small county or a cruise ship. And we will display plots for the fractional of compartment and the infinitesimal standard deviations, and analyze the role of stochastic variances in each case. The infinitesimal variances represent chance variations that leads to deviations from the deterministic SIR model. Through simulations of infinitesimal variances as in (14)

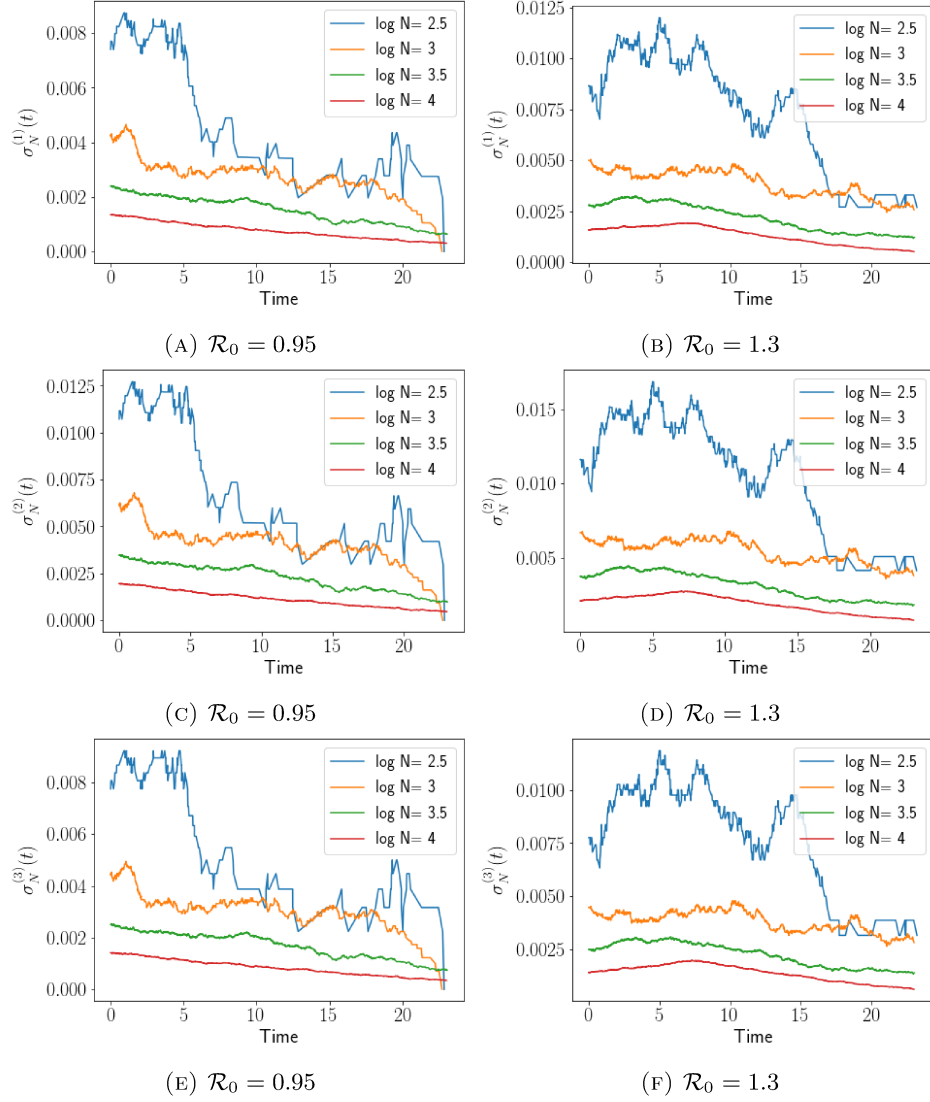


FIGURE 2. Examples of the infinitesimal standard deviation for $\sigma_N^{(l)}(t)$, $l = 1, 2, 3$. In both cases, $p_1 = 0.5$, $p_2 = 0.5$, $\gamma = 1$, $T = 23$, $s_0 = \mathbf{s}_{0N} = 0.96$, $i_0 = \mathbf{i}_{0N} = 0.04$, $r_0 = \mathbf{r}_{0N} = 0$. In the cases with $\mathcal{R}_0 = 0.95 < 1$ in Figs 2a, 2c, and 2e, $\beta = 0.95$. And the blue, orange, green, and red lines show results with the same realization from simulations with the corresponding colors in Fig 1b, respectively. In the cases with $\mathcal{R}_0 = 1.3 > 1$ in Figs 2b, 2d, and 2f, $\beta = 0.95$. And the blue, orange, green, and red lines show results with the same realization from simulations with the corresponding colors in Fig 1h, respectively.

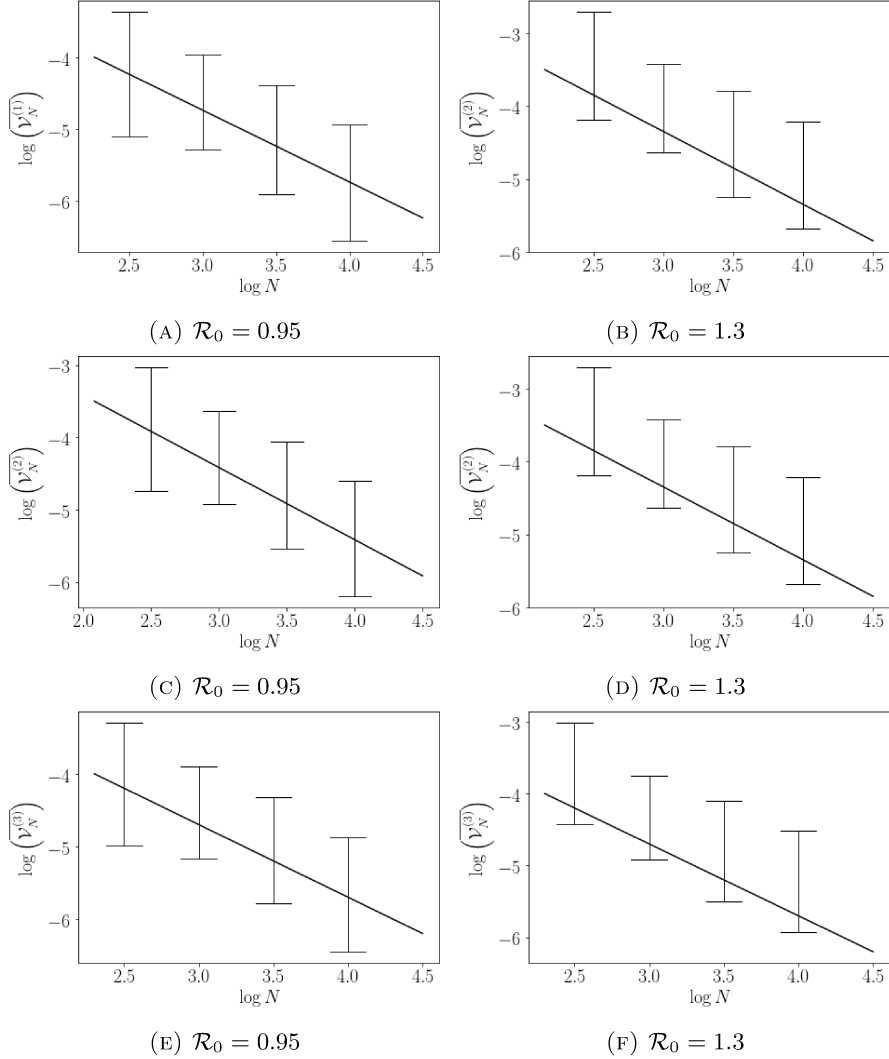


FIGURE 3. Comparisons of the log-log plot of the theoretical and empirical scaling for $\overline{\mathcal{V}}_N^{(l)}$, $l = 1, 2, 3$, as in (38) - (40). The vertical bars are the error bars of the empirical scaling of the infinitesimal variance from the same realization as in Figs. 1b, and 1h. We set $[T_1, T_2]$ as $[0, 10]$, $[1, 11]$, $[2, 12]$, $\dots$ and $[13, 23]$. The minimum and maximum values of $y = \log(\overline{\mathcal{V}}_N^{(l)})$, $l = 1, 2, 3$, taken over all such intervals are set as the limits of the error bars. The straight lines with slope -1 show the theoretical scaling with the x -intercepts as $\log(p_1\beta i_0 s_0)$, $\log(p_1\beta i_0 s_0 + p_2\gamma i_0)$, and $\log(p_2\gamma i_0)$, for $l = 1, 2, 3$, respectively. In all the figures, $p_1 = 0.5$, $p_2 =$, $\gamma = 1$, $T = 23$, $s_0 = \mathbf{s}_{0N} = 0.96$, $i_0 = \mathbf{i}_{0N} = 0.04$, and $r_0 = \mathbf{r}_{0N} = 0$. In Figs. 3a, 3c, and 3e, $\beta = 0.95$. In Figs. 3a, 3c, and 3e, $\beta = 1.3$.

- (16), we can produce envelopes that encompass a majority of the random paths. Realistically, our findings suggest that for small populations, the probabilities of rare events, e.g. the early-die-out or early-outbreak are important.

We study the outbreak of COVID-19 in Churchill County, Nevada from July, 2020 to March, 2021 and on the Diamond Princess cruise ship from February to March, 2020. Since early testing on the Diamond Princess was done by sampling from the population due to limited availability of testing, there is an under-reporting issue in the infected counts. The data of Churchill County was collected from [19] and the reported number of confirmed cases of Princess Diamond, and the quarantine process were retrieved from publicly available sources including the Princess Cruise website of the Carnival Cooperation [18] and the official website of Ministry of Health, Labor and Welfare, Japan (Ministry of Health, Labor and Welfare, Japan [54]).

We plot the daily confirmed cases with 30 realisations from the SIR-IPC models in fig. 4. For fig. 4a, the total population is set to be Churchill County's population $N = 25715$ and the reproduction number $\mathcal{R}_0 = 1.007$ using linear regression optimized by a grid search. For fig. 4c the total population is set to be the total number of passengers and crew on Diamond Princess $N = 3711$. Cruise ships carry a large number of people in confined spaces with relative homogeneous mixing, amplifying an already highly transmissible disease. Following the estimation of the dynamic reproduction number in [60], we take an average of these estimates from Feb. 5 to Feb. 26, resulting in $\mathcal{R}_0 = 2.02$. With these smaller populations, the infinitesimal standard deviations are non-negligible in the stochastic SIR model. Public health officials thus need to be aware of the variability of possible outcomes due to finite size effects.

6. Discussion. Compartmental models are a powerful tool to predict and control infectious diseases. In this article, we study theoretically and quantitatively the *finite size effects* arising in stochastic compartmental models, where individual realizations of the models deviate from their mean-field limit.

We apply compound Poisson processes to classical SIR compartmental models without vital dynamics. The result is a continuous-time Markov pure jump process, and a martingale approach can be applied to this process. The process is expressed as the sum of a deterministic and a stochastic component, which provides us with a tool to study both the statistical and stochastic features of the process. The deterministic part coincides with the classical deterministic SIR model. By providing a bound of the variances of the martingale, we show that the continuum mean-field limit is indeed the deterministic SIR model. Simulations results also align with the conclusion.

However, a small population size leads to stochastic fluctuations that deviate from the deterministic SIR model. Quantitative and theoretical analysis of these *finite size effects* have not been well-studied. We find a theoretical explanation for the *finite size effects* by observing that the stochastic component of the martingale formulation scales as the inverse of the square root of the population size. A larger variance both in the outbreak size and its temporal behavior arises as population size decreases. This scaling property is verified at time zero with equilibrium initial data. Direct numerical simulations of the theoretical infinitesimal variance support our theory as output shows that the dependence on population size remains to be true in later times. Here we simulate with fixed initial compartment fractions and

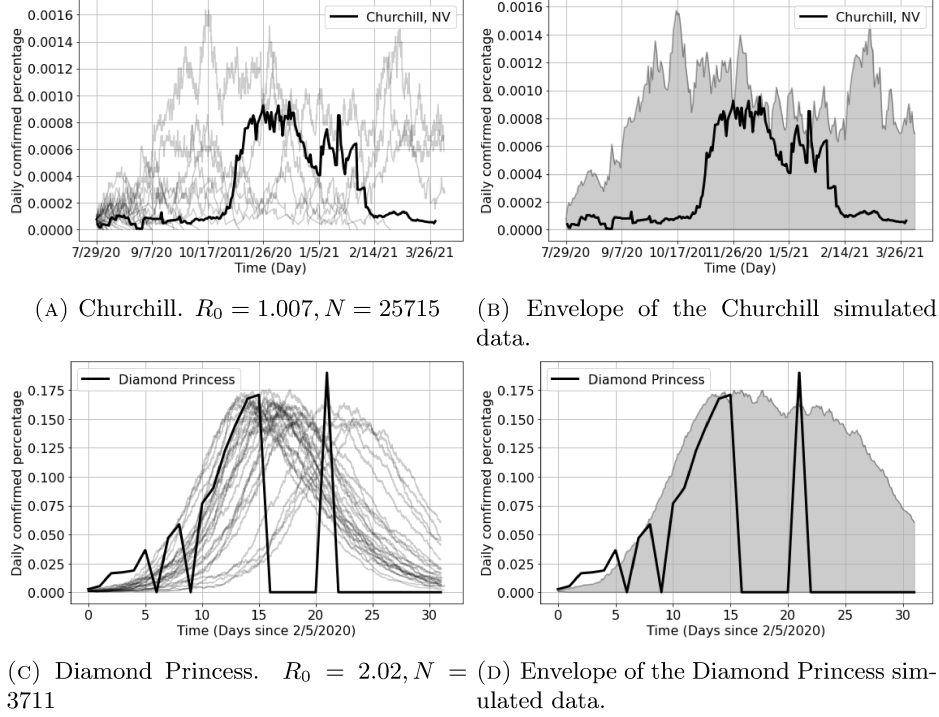
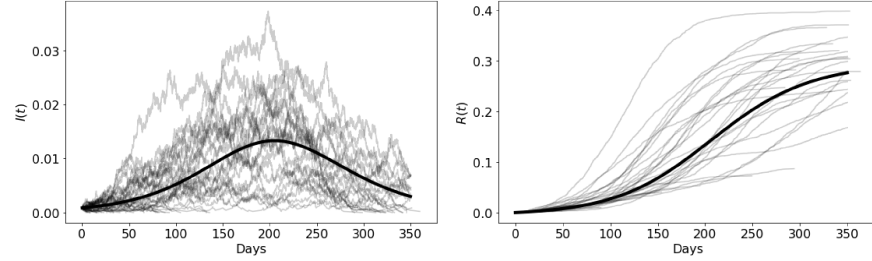


FIGURE 4. Comparison of field data (solid black line) for daily confirmed case percentages with 30 realisations of the stochastic SIR model for Churchill County, NV and the Diamond Princess Cruise Ship.

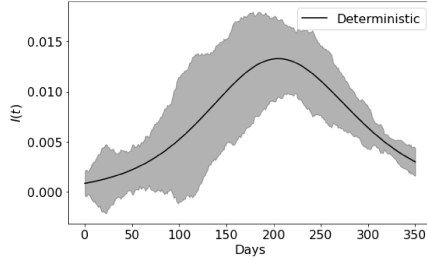
vary the total population size. To the best of our knowledge, this is the first time that simulations of theoretical infinitesimal variances for stochastic compartmental models are implemented. In previous works only empirical variations are simulated (see e.g. [59], [33]), and the focus was on the effects of statistical fluctuations on the reproduction number assuming varying initial compartment fractions [4]. We also simulate *finite size effects* with small populations and analyze with real data. All the simulations support our theory. Our results exhibit the danger of fitting data collected during an outbreak to deterministic counterparts of the stochastic compartmental models, especially for small populations.

Finite size effects are observed in Markov processes for population dynamics in many fields ([15, 26, 28, 29, 32, 36, 41, 42, 55, 65, 66, 70]). But there have been few quantitative studies to date about *finite size effects* in epidemics. On one hand, the methodology developed here may be broadly useful for quantitative social and natural sciences. On the other, this work provides a mathematical and theoretical framework that may contribute to epidemic policy for public health agencies.

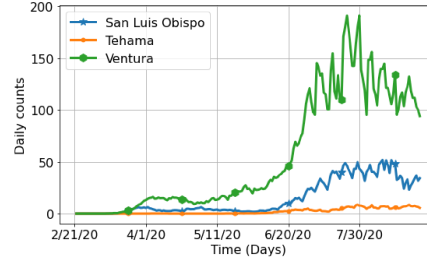
So far, the analysis is purely theoretical. It remains an open problem to integrate the real-world data to the model and glean some indicators from the data to control the disease. One example is that in order to control the spread, the government should know when and where to take certain measures to reduce β .



(A) Deterministic $i(t)$ with 50 simulations of Stochastic $i(t)$. The parameters are estimated using daily infected counts and total population of San Luis Obispo county, CA and $\beta = 0.13, \gamma = 0.11, N = 4777$



(C) Deterministic $i(t)$ with its theoretical Confidence band. The bandwidth is $6 * \sqrt{V_I(t)}$.



(D) 10-day moving average of the daily confirmed cases in 3 counties in California. Ventura's peak is around end of July, San Luis Obispo's peak is around mid August.

FIGURE 5

From Figure 5c, with a smaller population, the envelope of synthetic data can be wide. Additionally, the process with a smaller population also has more paths that die out early in the process (see fig. 2a, 2c, 2e), and 5a. Our work provides a good guide for authorities of smaller populations like cruise ships and small towns to estimate risk over time in order to prepare for the outbreak. It is important to bear in mind that the broader variations in the pandemic caused by the smaller population would lead to a wide outcome when it comes to estimating risk. In the past hundreds of years of human history, there have been several infectious diseases, including SARS, MERS and H1N1. However, only few of them, e.g. COVID-19 and 1918 influenza escalated into a pandemic. It is important to understand when, why and how a disease dies out and our next step is to quantitatively study the die-out event and its relations to finite size population and reproduction number $\mathcal{R}_0$.

Acknowledgments. We thank Professors Thomas Liggett, Jingyu Huang, Qiang Zeng, and Martina Hofmanova for helpful discussions.

REFERENCES

- [1] L. J. S. Allen, [An introduction to stochastic epidemic models](#), In *Mathematical Epidemiology*, Lecture Notes in Math., **1945** (2008), 81–130.
- [2] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, 2nd edition, CRC Press, Boca Raton, FL, 2011.
- [3] L. J. Allen and E. J. Allen, [A comparison of three different stochastic population models with regard to persistence time](#), *Theoretical Population Biology*, **64** (2003), 439–449, <https://www.sciencedirect.com/science/article/pii/S0040580903001047>.
- [4] E. Almaraz and A. Gómez-Corral, [On SIR-models with Markov-modulated events: Length of an outbreak, total size of the epidemic and number of secondary infections](#), *Discrete Contin. Dyn. Syst. Ser. B*, **23** (2018), 2153–2176.
- [5] D. Applebaum, *Lévy Processes and Stochastic Calculus*, vol. 116 of Cambridge Studies in Advanced Mathematics, 2nd edition, Cambridge Studies in Advanced Mathematics, 116. Cambridge University Press, Cambridge, 2009.
- [6] P. Azimi, Z. Keshavarz, J. G. Cedeno Laurent, B. Stephens and J. G. Allen, [Mechanistic transmission modeling of COVID-19 on the diamond princess cruise ship demonstrates the importance of aerosol transmission](#), *Proceedings of the National Academy of Sciences*, **118** (2021).
- [7] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd edition, Hafner Press [Macmillan Publishing Co., Inc.], New York, 1975.
- [8] M. S. Bartlett, [The critical community size for measles in the united states](#), *Journal of the Royal Statistical Society: Series A (General)*, **123** (1960), 37–44.
- [9] F. K. Batista, A. M. del Rey and A. Queiruga-Dios, [A review of seir-d agent-based model](#), In *Distributed Computing and Artificial Intelligence*, 16th International Conference, *Special Sessions*, **1004** (2019), 133–140.
- [10] T. Belin, A. Bertozzi, N. Chaudhary, T. Graves, J. Guterman, M. C. Jarashow, R. J. Lewis, J. Marion, F. Schoenberg, M. Shah, J. Tolles, E. Traub, K. Viele and F. Wu, Projections of Hospital-based Healthcare Demand due to COVID-19 in Los Angeles County May 24, 2021, 2021, http://file.lacounty.gov/SDSInter/dhs/1107440_{COVID}-19ProjectionPublicUpdateLewis05.24.21English.pdf.
- [11] A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short and D. Sledge, [The challenges of modeling and forecasting the spread of COVID-19](#), *Proc. Natl. Acad. Sci.*, **117** (2020), 16732–16738, <https://www.pnas.org/content/117/29/16732>.
- [12] B. M. Bibby and M. Sørensen, [Martingale estimation functions for discretely observed diffusion processes](#), *Bernoulli*, **1** (1995), 17–39.
- [13] K. Bichteler, *Stochastic Integration with Jumps*, Encyclopedia of Mathematics and its Applications, 89. Cambridge University Press, Cambridge, 2002.
- [14] E. Buckingham-Jeffery, V. Isham and T. House, [Gaussian process approximations for fast inference from infectious disease data](#), *Math. Biosci.*, **301** (2018), 111–120.
- [15] Y. Chang, J. C. Tzou, M. J. Ward and J. C. Wei, [Refined stability thresholds for localized spot patterns for the Brusselator model in \$\mathbb{R}^2\$](#) , *European J. Appl. Math.*, **30** (2019), 791–828.
- [16] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd edition, Modern Birkhäuser Classics, Birkhäuser/Springer, New York, 2014.
- [17] F. Comte, V. Genon-Catalot and A. Samson, [Nonparametric estimation for stochastic differential equations with random effects](#), *Stochastic Process. Appl.*, **123** (2013), 2522–2551.
- [18] P. Cruises, Princess cruise lines (2020) Diamond Princess updates, 2021, https://www.princess.com/news/notices_and_advisories/notices/diamond-princess-update.html.
- [19] E. Dong, H. Du and L. Gardner, [An interactive web-based dashboard to track COVID-19 in real time](#), *The Lancet Infectious Diseases*, **20** (2020), 533–534.
- [20] R. Durrett, *Stochastic Calculus*, A practical introduction. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.
- [21] R. Durrett, *Essentials of Stochastic Processes*, Springer Texts in Statistics, Springer-Verlag, New York, 1999.
- [22] R. Durrett, *Probability Models for DNA Sequence Evolution*, Probability and its Applications (New York), Springer-Verlag, New York, 2002.
- [23] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence.*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986.

- [24] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. v. Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley and A. C. Ghani, [Impact of non-pharmaceutical interventions \(NPIs\) to reduce COVID-19 mortality and healthcare demand, Report 9, Imperial College COVID-19 Response Team, Imperial College London, London, United Kingdom, 2020.](#)
- [25] D. T. Gillespie, [A general method for numerically simulating the stochastic time evolution of coupled chemical reactions](#), *J. Comput. Phys.*, **22** (1976), 403–434.
- [26] D. Gomez, M. J. Ward and J. Wei, [The linear stability of symmetric spike patterns for a bulk-membrane coupled Gierer-Meinhardt model](#), *SIAM J. Appl. Dyn. Syst.*, **18** (2019), 729–768.
- [27] M. Z. Guo, G. C. Papanicolaou and S. R. S. Varadhan, [Nonlinear diffusion limit for a system with nearest neighbor interactions](#), *Comm. Math. Phys.*, **118** (1988), 31–59, <http://projecteuclid.org/euclid.cmp/1104161907>.
- [28] F. Hamba, [Turbulent energy density in scale space for inhomogeneous turbulence](#), *J. Fluid Mech.*, **842** (2018), 532–553.
- [29] I. Hameduddin, C. Meneveau, T. A. Zaki and D. F. Gayme, [Geometric decomposition of the conformation tensor in viscoelastic turbulence](#), *J. Fluid Mech.*, **842** (2018), 395–427.
- [30] S. W. He, J. G. Wang and J. A. Yan, *Semimartingale Theory and Stochastic Calculus*, Kexue Chubanshe (Science Press), Beijing; CRC Press, Boca Raton, FL, 1992.
- [31] P. G. Hoel, S. C. Port and C. J. Stone, *Introduction to Stochastic Processes*, The Houghton Mifflin Series in Statistics, Houghton Mifflin Co., Boston, Mass., 1972.
- [32] S. J. Illingworth, J. P. Monty and I. Marusic, [Estimating large-scale structures in wall turbulence using linear models](#), *J. Fluid Mech.*, **842** (2018), 146–162.
- [33] V. Isham, [Assessing the variability of stochastic epidemics](#), *Mathematical Biosciences*, **107** (1991), 209–224.
- [34] V. Isham, Stochastic models for epidemics with special reference to AIDS, *Ann. Appl. Probab.*, **3** (1993), 1–27, [http://links.jstor.org/sici?sici=1050-5164\(199302\)3:1<1:SMFEWS>2.0.CO;2-4&origin=MSN](http://links.jstor.org/sici?sici=1050-5164(199302)3:1<1:SMFEWS>2.0.CO;2-4&origin=MSN).
- [35] V. Isham, [Stochastic models for epidemics](#), In *Celebrating Statistics*, Oxford Statist. Sci. Ser., Oxford Univ. Press, Oxford, **33** (2005), 27–54.
- [36] J. Jiménez, [Coherent structures in wall-bounded turbulence](#), *J. Fluid Mech.*, **842** (2018), P1, 100.
- [37] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.
- [38] W. O. Kermack, A. G. McKendrick and G. T. Walker, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **115** (1927), 700–721, <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118>.
- [39] C. Kipnis, S. Olla and S. R. S. Varadhan, [Hydrodynamics and large deviation for simple exclusion processes](#), *Comm. Pure Appl. Math.*, **42** (1989), 115–137.
- [40] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, Berlin, 1999.
- [41] T. Kolokolnikov, M. Ward, J. Tzou and J. Wei, [Stabilizing a homoclinic stripe](#), *Philos. Trans. Roy. Soc. A*, **376** (2018), 20180110, 13 pp.
- [42] T. Kolokolnikov and J. Wei, [Pattern formation in a reaction-diffusion system with space-dependent feed rate](#), *SIAM Rev.*, **60** (2018), 626–645.
- [43] I. Krishnarajah, A. Cook, G. Marion and G. Gibson, [Novel moment closure approximations in stochastic epidemics](#), *Bull. Math. Biol.*, **67** (2005), 855–873.
- [44] H. Kunita, *Lectures on Stochastic Flows and Applications*, Springer-Verlag, Berlin, 1986.
- [45] Z. Li, Z. Teng, D. Hong and X. Shi, [Comparison of three sis epidemic models: Deterministic, stochastic and uncertain](#), *Journal of Intelligent & Fuzzy Systems*, **35** (2018), 5785–5796.
- [46] T. M. Liggett, Interacting Markov processes, In *Biological Growth and Spread (Proc. Conf., Heidelberg, 1979)*, **38** (1980), 145–156.
- [47] T. M. Liggett, *Interacting Particle Systems*, Springer-Verlag, New York, 1985.
- [48] T. M. Liggett, *Continuous Time Markov Processes*, American Mathematical Society, Providence, RI, 2010.

- [49] X. Liu, A. Mubayi, D. Reinhold and L. Zhu, [Approximation methods for analyzing multiscale stochastic vector-borne epidemic models](#), *Math. Biosci.*, **309** (2019), 42–65.
- [50] A. L. Lloyd, [Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics](#), *Theoretical Population Biology*, **60** (2001), 59–71, <http://www.sciencedirect.com/science/article/pii/S0040580901915254>.
- [51] S. McMiNN, R. Talbot and J. ENG, America’s 200,000 COVID-19 deaths: Small cities and towns bear a growing share, <https://www.npr.org>, <https://www.npr.org/sections/health-shots/2020/09/22/914578634/americas-200-000-COVID-19-deaths-small-cities-and-towns-bear-a-growing-share>.
- [52] M. Métivier, *Semimartingales*, A course on stochastic processes. de Gruyter Studies in Mathematics, 2. Walter de Gruyter & Co., Berlin-New York, 1982.
- [53] M. Métivier and J. Pellaumail, *Stochastic Integration*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London-Toronto, Ont., 1980.
- [54] L. Ministry of Health and J. Welfare, Ministry of health, labor and welfare, Japan (2020) about coronavirus disease 2019 (COVID-19), 2020. https://www.mhlw.go.jp/stf/newpage_09276.html.
- [55] F. Paquin-Lefebvre, W. Nagata and M. J. Ward, [Pattern formation and oscillatory dynamics in a two-dimensional coupled bulk-surface reaction-diffusion system](#), *SIAM J. Appl. Dyn. Syst.*, **18** (2019), 1334–1390.
- [56] S. Peszat and J. Zabczyk, *Stochastic Partial Differential Equations with Lévy Noise*, Cambridge University Press, Cambridge, 2007.
- [57] B. L. S. Prakasa Rao, *Statistical Inference for Diffusion Type Processes*, Edward Arnold, London; Oxford University Press, New York, 1999.
- [58] P. E. Protter, *Stochastic Integration and Differential Equations*, 2nd edition, Springer-Verlag, Berlin, 2005.
- [59] M.-A. Rizoïu, S. Mishra, Q. Kong, M. Carman and L. Xie, [SIR-Hawkes: Linking epidemic models and hawkes processes to model diffusions in finite populations](#), *International World Wide Web Conferences Steering Committee*, Republic and Canton of Geneva, CHE, (2018), 419–428.
- [60] J. Rocklöv, H. Sjödin and A. Wilder-Smith, COVID-19 outbreak on the diamond princess cruise ship: Estimating the epidemic potential and effectiveness of public health countermeasures, *Journal of Travel Medicine*, **27** (2020), taaa030.
- [61] E. Schumaker and M. Nichols, An american tragedy: Inside the towns hardest hit by coronavirus, *abcNEWS*, <https://abcnews.go.com/Health/small-towns-face-COVID-19-pandemic-us-passes/story?id=74271392>.
- [62] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Classics in Mathematics, Springer-Verlag, Berlin, 2006.
- [63] W.-Y. Tan and S. Tang, [A general markov model of the hiv epidemic in populations involving both sexual contact and iv drug use](#), *Mathematical and Computer Modelling*, **19** (1994), 83–132.
- [64] B. Tóth and B. Valkó, [Onsager relations and Eulerian hydrodynamic limit for systems with several conservation laws](#), *J. Statist. Phys.*, **112** (2003), 497–521.
- [65] W. H. Tse and M. J. Ward, [Asynchronous instabilities of crime hotspots for a 1-D reaction-diffusion model of urban crime with focused police patrol](#), *SIAM J. Appl. Dyn. Syst.*, **17** (2018), 2018–2075.
- [66] J. C. Tzou, M. J. Ward and J. C. Wei, [Anomalous scaling of Hopf bifurcation thresholds for the stability of localized spot patterns for reaction-diffusion systems in two dimensions](#), *SIAM J. Appl. Dyn. Syst.*, **17** (2018), 982–1022.
- [67] M. Uchida, [Approximate martingale estimating functions for stochastic differential equations with small noises](#), *Stochastic Process. Appl.*, **118** (2008), 1706–1721.
- [68] S. R. S. Varadhan, Entropy methods in hydrodynamic scaling, In *Proceedings of the International Congress of Mathematicians*, **1** (1995), 196–208.
- [69] S. R. S. Varadhan, Lectures on hydrodynamic scaling, In *Hydrodynamic Limits and Related Topics (Toronto, ON, 1998)*, Fields Inst. Commun., Amer. Math. Soc., Providence, RI, **27** (2000), 3–40.
- [70] M. J. Ward, [Spots, traps, and patches: Asymptotic analysis of localized solutions to some linear and nonlinear diffusive systems](#), *Nonlinearity*, **31** (2018), 189–239.
- [71] G.-W. Weber, P. Taylan, Z.-K. Görgülü, H. A. Rahman and A. Bahar, [Parameter estimation in stochastic differential equations](#), In *Dynamics, Games and Science. II*, **2** (2011), 703–733.

- [72] P. Yan, [Distribution theory, stochastic processes and infectious disease modelling](#), In *Mathematical Epidemiology*, Lecture Notes in Math., **1945** (2008), 229–293.
- [73] P. Yan and G. Chowell, [Beyond the initial phase: Compartment models for disease transmission](#), *Quantitative Methods for Investigating Infectious Disease Outbreaks*, **70** (2019), 135–182.

Received July 2021; revised February 2022; early access March 2022.

E-mail address: xli51@math.ucla.edu

E-mail address: cwang27@ua.edu

E-mail address: hao.li1947@gmail.com

E-mail address: bertozzi@math.ucla.edu