ACCEPTED VERSION

Nicolas P. Rebuli, N. G. Bean, J. V. Ross
Hybrid Markov chain models of S-I-R disease dynamics
Journal of Mathematical Biology, 2017; 75(3):521-541

© Springer-Verlag Berlin Heidelberg 2016

*This is a post-peer-review, pre-copyedit version of an article published in **Journal of Mathematical Biology.** The final authenticated version is available online at:* http://dx.doi.org/10.1007/s00285-016-1085-2

PERMISSIONS

https://www.springer.com/gp/open-access/publication-policies/self-archiving-policy

Self-archiving for articles in subscription-based journals

Springer journals' policy on preprint sharing.

By signing the Copyright Transfer Statement you still retain substantial rights, such as self-archiving:

Author(s) are permitted to self-archive a pre-print and an author's accepted manuscript version of their Article.
……..

b. An Author's Accepted Manuscript (AAM) is the version accepted for publication in a journal following peer review but prior to copyediting and typesetting that can be made available under the following conditions:

(i) Author(s) retain the right to make an AAM of their Article available on their own personal, self-maintained website immediately on acceptance,

(ii) Author(s) retain the right to make an AAM of their Article available for public release on any of the following 12 months after first publication ("Embargo Period"): their employer's internal website; their institutional and/or funder repositories. AAMs may also be deposited in such repositories immediately on acceptance, provided that they are not made publicly available until after the Embargo Period.

An acknowledgement in the following form should be included, together with a link to the published version on the publisher's website: "This is a post-peer-review, pre-copyedit version of an article published in [insert journal title]. The final authenticated version is available online at: http://dx.doi.org/[insert DOI]".

When publishing an article in a subscription journal, without open access, authors sign the Copyright Transfer Statement (CTS) which also details Springer's self-archiving policy.

See Springer Nature terms of reuse for archived author accepted manuscripts (AAMs) of subscription articles.

24 June 2020

http://hdl.handle.net/2440/115564

Noname manuscript No.
(will be inserted by the editor)

Hybrid Markov chain models of S-I-R disease dynamics

Nicolas P. Rebuli · N. G. Bean · J. V. Ross

Received: date / Accepted: date

Abstract Deterministic epidemic models are attractive due to their compact nature, allowing substantial complexity with computational efficiency. This partly explains their dominance in epidemic modelling. However, the small numbers of infectious individuals at early and late stages of an epidemic, in combination with the stochastic nature of transmission and recovery events, are critically important to understanding disease dynamics. This motivates the use of a stochastic model, with continuous-time Markov chains being a popular choice. Unfortunately, even the simplest Markovian S-I-R model – the so-called general stochastic epidemic – has a state space of order N^2, where N is the number of individuals in the population, and hence computational limits are quickly reached. Here we introduce a hybrid Markov chain epidemic model, which maintains the stochastic and discrete dynamics of the Markov chain in regions of the state space where they are of most importance, and uses an approximate model – namely a deterministic or a diffusion model – in the remainder of the state space. We discuss the evaluation, efficiency and accuracy of this hybrid model when approximating the distribution of the duration of the epidemic and the distribution of the final size of the epidemic. We demonstrate that the computational complexity is $\mathcal{O}(N)$ and that under suitable conditions our approximations are highly accurate.

Keywords Markov population processes · epidemiology · fluid approximation · diffusion approximation

Mathematics Subject Classification (2000) 60J27 · 60J28 · 92D30 · 60J22 · 60J60

Nicolas P. Rebuli
E-mail: nicolas.rebuli@adelaide.edu.au

School of Mathematical Sciences, University of Adelaide, SA 5005, Australia

ARC Centre of Excellence for Mathematical and Statistical Frontiers, University of Melbourne, VIC 3010, Australia

1 Introduction

Compartmental continuous-time Markov chain models (CTMC)s are of substantial importance to mathematical epidemiology (Bartlett, 1956; Rand and Wilson, 1991; Fox, 1993; Grenfell et al, 1998; Keeling et al, 2000; Spagnolo et al, 2003; Coulson et al, 2004). They capture the stochastic individual–to–individual nature of disease transmission which is particularly important when there are small numbers of infectious individuals, such as during the early or late stages of an epidemic. However, the state space of these models is typically $\mathcal{O}(N^d)$, where d is the number of compartments in the model and N is the population size. Hence when N is large, it is often more efficient to analyse an approximation of the CTMC.

Kurtz (1970, 1971) and Barbour (1974, 1976, 1980b,a) established a deterministic and a diffusion approximation of suitably-scaled *density dependent Markov population processes* which are asymptotic in N. In practice, these approximations are highly accurate for finite N but are known to be inaccurate if the population of at least one compartment of the underlying CTMC is close to zero. The hybrid models presented in this paper combine a CTMC with its associated deterministic or diffusion approximation in such a way as to appeal to the strengths of both models while also addressing their respective weaknesses.

We consider the so-called general stochastic epidemic model (Bartlett, 1949, 1956; Bailey, 1950, 1957; Kendall, 1965; Kermack and McKendrick, 1927), otherwise known as the *Susceptible–Infectious–Removed* (SIR) CTMC, which is a common representation of the population level dynamics of many viral infections, where following recovery from the disease, an infectious individual is henceforth permanently immune. Our focus will be on approximating the distribution of the duration of the epidemic and the distribution of the final size of the epidemic. The duration of the epidemic is defined as the length of time before the final infectious individual is removed from the population, and the final size of the epidemic is defined as the total number of individuals who experience an infection (including those who were infected initially) before the final infectious individual is removed from the population. Jenkinson and Goutsias (2012) and Black and Ross (2015) presented highly efficient approaches for calculating the distribution of the duration of the epidemic and the final size of the epidemic, respectively, directly from the SIR CTMC. As the SIR CTMC is a density dependent Markov population process, these distributions have been studied in detail via the deterministic and diffusion approximations (Kermack and McKendrick, 1927; Ethier and Kurtz, 2008).

Hybrid models of the SIR CTMC have also generated substantial interest. In particular, Barbour (1975) presented an asymptotic approximation for the distribution of the duration of the epidemic similar to our own *hybrid fluid model*. The key difference being that Barbour used the dynamics of a branching process during the early and final stages of the epidemic, whereas our model uses the dynamics of the CTMC. In addition, an asymptotic approximation of the distribution of the final size of the epidemic similar to our own *hybrid*

diffusion model has been investigated by a number of authors (Andersson and Britton, 2000; Ball and Neal, 2010; Nagaev and Startsev, 1970; Scalia-Tomba, 1985; Watson, 1980, 1981; Lefèvre, 1990). According to Lefèvre (1990), all of these previous models use the dynamics of a branching process to model the sub-critical component of the final size and a Gaussian approximation to model the super-critical component of the final size. These approaches differ from our hybrid models because we use CTMC dynamics whenever the number of infectious individuals is low. For comparison, we compare the accuracy of our hybrid fluid model and hybrid diffusion model to these asymptotic approximations in calculating the distribution of the duration of the epidemic and the distribution of the final size of the epidemic.

Hybrid discrete–continuous approximations of CTMC dynamics have also been presented by Sazonov et al (2011) and Safta et al (2015). Sazonov et al (2011) presented a two-stage model for approximating the dynamics of the SIR CTMC. Their model has the dynamics of a branching process during the early stages of the epidemic and deterministic dynamics thereafter. Sazonov et al (2011) used their model to approximate the distribution of the time of the peak of the outbreak. Safta et al (2015) presented a numerical scheme for approximating the distribution of CTMC models of chemical reaction networks. Their approach uses CTMC dynamics for the compartments of the process which are less than a particular threshold and diffusion dynamics otherwise. In the context of the SIR CTMC, this means that during the early stages of the epidemic, the susceptible class has diffusion dynamics while the infectious class has CTMC dynamics.

As we shall see, the hybrid models we introduce here reduce the complexity of algorithms required to compute distributions of interest from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ in exchange for a minor reduction in accuracy. This enables us to compute an accurate approximation of these distributions in a reasonable amount of time even for population sizes of order 10^7.

This paper is organised as follows. Section 2 introduces notation for the SIR CTMC and presents its fluid limit and diffusion limit approximations. Section 3 introduces the hybrid fluid model and applies it to calculating the distribution of the duration of the epidemic and the distribution of the final size of the epidemic. Section 4 introduces the hybrid diffusion model and applies it to calculating the distribution of the final size of the epidemic. Section 5 discusses the details of calculating the solutions to the systems of equations which arise in Sections 3 and 4. Finally Section 6 discusses possible improvements and extensions.

2 The SIR epidemic model

Let $\{\mathbf{X}(t)\}_{t\geq 0}$ denote the SIR CTMC (Bailey (1950)) which takes values (S, I) from the two-dimensional lattice

$$\mathcal{X} = \left\{(S, I) \in \mathbb{Z}_+^2 \,:\, S + I \leq N\right\}. \tag{1}$$

Let $\boldsymbol{\ell}_1 = (-1, 1)$ and $\boldsymbol{\ell}_2 = (0, -1)$ denote the stochiometries (jumps) of $\mathbf{X}(t)$ and $q_X(\mathbf{x}, \mathbf{x} + \boldsymbol{\ell}_j)$ denote the transition rate from $\mathbf{x}$ to $\mathbf{x} + \boldsymbol{\ell}_j$, for $j = 1, 2$. Then for all $\mathbf{x} \in \mathcal{X}$, the positive transition rates of $\mathbf{X}(t)$ are:

$$q_X(\mathbf{x}, \mathbf{x} + \boldsymbol{\ell}_1) = \frac{\beta}{N-1} SI \qquad \text{if } \mathbf{x} + \boldsymbol{\ell}_1 \in \mathcal{X}, \tag{2}$$

$$q_X(\mathbf{x}, \mathbf{x} + \boldsymbol{\ell}_2) = \gamma I \qquad \text{if } \mathbf{x} + \boldsymbol{\ell}_2 \in \mathcal{X}, \tag{3}$$

where β is the effective transmission rate parameter and $1/\gamma$ is the average infectious period of an individual.

2.1 The fluid approximation

Consider the scaled process $\mathbf{X}(t)/N$ which takes the scaled values $(S/N, I/N)$, with (S, I) in $\mathcal{X}$. Under minor technical conditions, the fluid limit theorem (Theorem 3.1 of Kurtz (1970)) implies that as $N \to \infty$, the scaled process $\mathbf{X}(t)/N$ converges uniformly in probability over finite time intervals to the unique deterministic trajectory $\mathbf{x}(t)$, provided $\mathbf{x}(0) = \mathbf{X}(0)/N$. Given an initial state, the deterministic trajectory $\mathbf{x}(t)$ takes values (s, i) from the continuum $[0, 1]^2$, with $s + i \leq 1$, and is the unique solution to the system of ordinary differential equations (ODE)s

$$\frac{ds}{dt} = -\beta si, \tag{4}$$

$$\frac{di}{dt} = \beta si - \gamma i. \tag{5}$$

For a finite population, the deterministic trajectory $N\mathbf{x}(t)$ is a working approximation for the average dynamics of $\mathbf{X}(t)$ and is commonly referred to as its fluid approximation.

2.2 The diffusion approximation

Under minor technical conditions, the diffusion limit theorem (Theorem 3.5 of Kurtz (1971)) implies that $\sqrt{N}(\mathbf{X}(t)/N - \mathbf{x}(t))$ converges weakly over finite time intervals to a Gaussian diffusion process as $N \to \infty$, provided $\mathbf{x}(0) = \mathbf{X}(0)/N$. This Gaussian diffusion process has mean value $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}(t)$, where $\boldsymbol{\Sigma}(t)$ is the unique solution to the system of ODEs

$$\frac{d\,\boldsymbol{\Sigma}}{dt} = \left(\mathbf{B} + \mathbf{B}^T\right)\boldsymbol{\Sigma} + \mathbf{G}, \tag{6}$$

with

$$\mathbf{B} = \begin{bmatrix} -\beta i & -\beta s \\ \beta i & \beta s - \gamma \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \beta si & -\beta si \\ -\beta si & \beta si + \gamma i \end{bmatrix},$$

and initial value $\boldsymbol{\Sigma}(0) = (\boldsymbol{\Sigma}_{i,j}(0) = 0,\ i, j = 1, 2)$. For a finite population, the Gaussian diffusion process with mean $N\mathbf{x}(t)$ and variance-covariance matrix

$N\boldsymbol{\Sigma}(t)$ is a working approximation for $\mathbf{X}(t)$, and we refer to it as the diffusion approximation.

It is worth noting that the fluid and diffusion approximations of Kurtz (1970, 1971) are uniform only over finite time intervals and can not be used to directly approximate the distribution of the duration of the epidemic or the distribution of the final size of the epidemic. However, see section 11.4 of Ethier and Kurtz (2008), in particular remark 11.4.2.

3 The hybrid fluid model

In this section we introduce the hybrid fluid model whose joint dynamics in $S(t)$ and $I(t)$ are determined by either the SIR CTMC or the fluid approximation, depending on the number of infectious individuals.

3.1 Model formulation

Let $\{\mathbf{Y}(t)\}_{t\geq 0}$ denote the hybrid fluid process, which takes values (S^Y, I^Y) from the hybrid discrete-continuous state space $\mathcal{Y}$, which is defined next. For now, we fix the *threshold* $\widehat{I}$ as a constant in $\{0, 1, \ldots, N\}$ and define the state space $\mathcal{Y}$ as the union of $\mathcal{Y}^{MC}$ and $\mathcal{Y}^{DE}$, where

$$\mathcal{Y}^{MC} = \left\{ \left(S^Y, I^Y\right) \in \mathbb{Z}_+^2 \,:\, S^Y + I^Y \leq N,\, I^Y \leq \widehat{I} \right\}$$

and

$$\mathcal{Y}^{DE} = \left\{ \left(S^Y, I^Y\right) \in \mathbb{R}_+^2 \,:\, S^Y + I^Y \leq N,\, I^Y \geq \widehat{I} \right\}.$$

The hybrid fluid process switches dynamics depending on which subset of $\mathcal{Y}$ it is in. In particular, when $\mathbf{Y}(t)$ is in the subset $\mathcal{Y}^{MC}$ it has the dynamics of the SIR CTMC $\mathbf{X}(t)$, and when $\mathbf{Y}(t)$ is in the subset $\mathcal{Y}^{DE}$ it has the dynamics of the fluid approximation $N\mathbf{x}(t)$. The dynamics of $\mathbf{Y}(t)$ at the intersection of $\mathcal{Y}^{MC}$ and $\mathcal{Y}^{DE}$, denoted $\mathcal{T}^{MC}$, require careful consideration.

According to the fluid dynamics of $\mathbf{Y}(t)$ (equation (5)) the rate of change of I^Y with respect to time is positive if $S^Y > N/R_0$, where $R_0 = \beta/\gamma$ is the basic reproductive number. This means that if $\mathbf{Y}(t)$ hits the state $\mathbf{y}_1 = (S_1^Y, \widehat{I})$ in $\mathcal{T}^{MC}$, where $S_1^Y > N/R_0$, then the fluid dynamics will immediately force $\mathbf{Y}(t)$ out of $\mathcal{T}^{MC}$ and into $\mathcal{Y}^{DE}$. In contrast, if $S_1^Y \leq N/R_0$ then the fluid dynamics will force $\mathbf{Y}(t)$ to remain in its current state until a discrete event occurs. As such, we define

$$\mathcal{T}_1^{MC} = \left\{ (S^Y, \widehat{I}) \in \mathcal{Y}^{MC} \,:\, S^Y \in \left\{ \left\lfloor \frac{N}{R_0} \right\rfloor + 1, \ldots, N - \widehat{I} \right\} \right\},$$

as the set of states which force $\mathbf{Y}(t)$ to switch from CTMC dynamics to ODE dynamics and

$$\mathcal{T}_2 = \left\{ (S^Y, \widehat{I}) \in \mathcal{Y}^{DE} \,:\, S^Y \in \left[0, \frac{N}{R_0} \right] \right\}$$

as the set of states which force $\mathbf{Y}(t)$ to switch from ODE dynamics to CTMC dynamics. We denote the integer components of $\mathcal{T}_2$ as $\mathcal{T}_2^{MC}$ which is defined as the intersection of $\mathcal{Y}^{MC}$ and $\mathcal{T}_2$.

Since the fluid dynamics of $\mathbf{Y}(t)$ are deterministic, we can determine which state in $\mathcal{T}_2$ the fluid dynamics terminate in as well as the total duration of the fluid dynamics, provided the initial state in $\mathcal{T}_1^{MC}$ of the fluid dynamics is known. A phase-plane analysis of the fluid approximation reveals that if the hybrid fluid process hits the state $\mathbf{y}_1$ in $\mathcal{T}_1^{MC}$ then the process will hit the state $\mathbf{y}_2$ in $\mathcal{T}_2$ where the number of susceptible individuals is given by

$$S_2^Y(\mathbf{y}_1) = -\frac{N}{R_0}\mathcal{W}\left(-\left(\frac{S_1^Y R_0}{N}\right)\exp\left(-\frac{S_1^Y R_0}{N}\right)\right), \tag{7}$$

where $\mathcal{W}(x)$ denotes the principal branch of the Lambert-W function. In addition, as the number of susceptible individuals monotonically decreases with time, the duration of the fluid dynamics is given by the integral

$$t(\mathbf{y}_1) = \frac{N}{\beta}\int_{S_1^Y}^{S_2^Y(\mathbf{y}_1)}\left\{u\left[\frac{1}{R_0}\log\left(\frac{S_1^Y}{u}\right)+\frac{1}{N}\left(u - S_1^Y - \widehat{I}\right)\right]\right\}^{-1} du. \tag{8}$$

Similar equations to (7) and (8) are used by Barbour (1974).

When the hybrid fluid process switches from fluid dynamics to CTMC dynamics a discretisation process must occur before the CTMC dynamics can resume. As the fluid dynamics only approximate the expected value of the system, we decided to round the number of susceptible individuals $S_2^Y(\mathbf{y}_1)$ as follows:

$$\begin{aligned} &\text{round down to } \lfloor S_2^Y(\mathbf{y}_1)\rfloor \quad &&\text{with probability } 1-\left(S_2^Y(\mathbf{y}_1)-\lfloor S_2^Y(\mathbf{y}_1)\rfloor\right),\\ &\text{round up to } \lfloor S_2^Y(\mathbf{y}_1)\rfloor+1 \quad &&\text{with probability } \left(S_2^Y(\mathbf{y}_1)-\lfloor S_2^Y(\mathbf{y}_1)\rfloor\right). \end{aligned} \tag{9}$$

Once $\mathbf{Y}(t)$ enters a state in $\mathcal{T}_2^{MC}$ it resumes CTMC dynamics. However, the only events which are possible from states in $\mathcal{T}_2^{MC}$ are recovery events.

Figure 1 is a representation of the state space of the hybrid fluid model for a population of $N = 15$ individuals and a threshold of $\widehat{I} = 3$. The green points are states from the discrete set $\mathcal{Y}^{MC}$, and the continuum $\mathcal{Y}^{DE}$ is the region with $I^Y \geq \widehat{I}$, and $S^Y \leq N - I^Y$. The state space $\mathcal{Y}$ is the union of these two sets. The sets $\mathcal{T}_1^{MC}$ and $\mathcal{T}_2^{MC}$ are represented by the upward and downward pointing triangles, respectively. The trajectories of $N\mathbf{x}(t)$ are represented as the black curves emanating from the set $\mathcal{T}_1^{MC}$ which intersect with $\mathcal{T}_2$ in integer states where the number of susceptible individuals is given by equation (7). The duration that the process spends on each of these trajectories is calculated from equation (8) and the final state of the diffusion dynamics is either rounded up or down. Finally, the arrows in $\mathcal{Y}^{MC}$ represent the transitions of the CTMC dynamics.

For the remainder of this section we use the hybrid fluid process to approximate the distribution of the duration of the epidemic and the distribution of the final size of the epidemic.

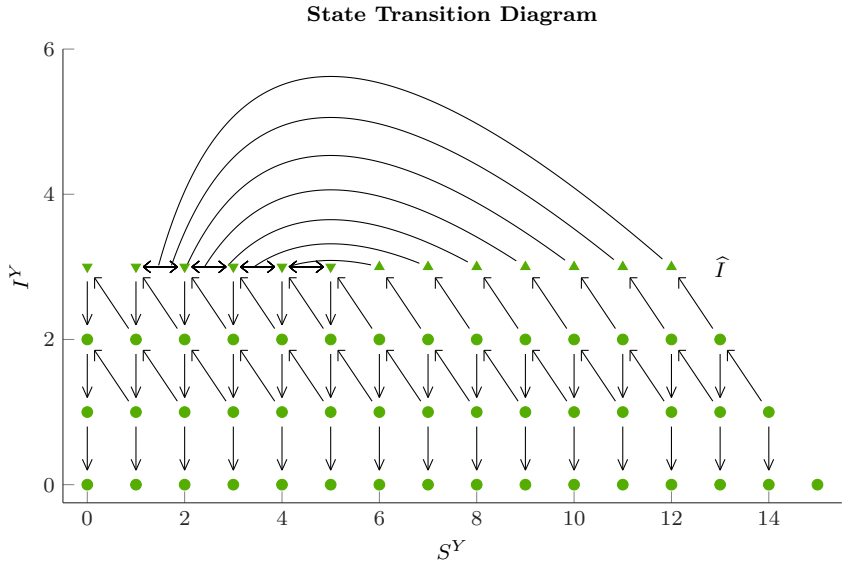

Fig. 1 The state transition diagram of the hybrid fluid process with $N = 15$ and $\widehat{I} = 3$. The green points are the discrete states from $\mathcal{Y}^{MC}$, and the continuum $\mathcal{Y}^{DE}$ is the set of states with $I^Y \geq \widehat{I}$, and $S^Y \leq N - I^Y$. The upward (downward) pointing triangles are states from which $\mathbf{Y}(t)$ switches from CTMC to fluid (fluid to CTMC) dynamics, which are contained within the set $\mathcal{T}_1^{MC}$ ($\mathcal{T}_2^{MC}$). The black curves emanating from states in $\mathcal{T}_1^{MC}$ are the deterministic trajectories of $N\mathbf{x}(t)$ through $\mathcal{Y}^{DE}$.

3.2 Duration of the epidemic

A system of delayed differential equations (DDEs) describing the flow of probability through the discrete states in $\mathcal{Y}^{MC}$ is derived by separately considering the probability flux on three disjoint subsets of $\mathcal{Y}^{MC}$. Within each of these subsets, the flux of probability between states in $\mathcal{Y}^{MC}$ must be treated differently due to the way in which probability flows between $\mathcal{Y}^{MC}$ and $\mathcal{Y}^{DE}$. In the first scenario we consider the set $\mathcal{D} = \mathcal{Y}^{MC} \setminus (\mathcal{T}_1^{MC} \cup \mathcal{T}_2^{MC})$, on which no probability flows out of $\mathcal{Y}^{MC}$. In the second and third scenarios we consider the sets $\mathcal{T}_1^{MC}$ and $\mathcal{T}_2^{MC}$ on which probability flows from $\mathcal{Y}^{MC}$ to $\mathcal{Y}^{DE}$ and from $\mathcal{Y}^{DE}$ to $\mathcal{Y}^{MC}$, respectively. The system of DDEs allow us to calculate the distribution of $\mathbf{Y}(t)$ on $\mathcal{Y}^{MC}$ for $t \geq 0$, which we utilise for calculating the distribution of the duration of the epidemic.

3.2.1 Scenario 1

We begin by defining the probability distribution of the hybrid fluid process as the vector $\phi_Y(\mathbf{y}\,;t) = \Pr(\mathbf{Y}(t) = \mathbf{y} \,|\, \mathbf{Y}(0) = \mathbf{y}_0)$ which denotes the probability mass that $\mathbf{Y}(t)$ is in the state $\mathbf{y}$ in $\mathcal{Y}^{MC}$ at the time t, for $t \geq 0$, given that the initial state of the process is $\mathbf{y}_0$ in $\mathcal{D}$. For all $\mathbf{y}$ in the set $\mathcal{D}$, the

probability flux is governed by the forward equations

$$\frac{d}{dt}\phi_Y(\mathbf{y}\,;\,t) = \sum_{\mathbf{y}'\in\mathcal{Y}^{MC}} \phi_Y(\mathbf{y}'\,;\,t)\; q_X(\mathbf{y}',\mathbf{y}). \tag{10}$$

3.2.2 Scenario 2

Now consider the set of states through which probability flows out of $\mathcal{Y}^{MC}$ and into $\mathcal{Y}^{DE}$. Since the hybrid fluid process instantly switches from CTMC dynamics to fluid dynamics when it hits a state in $\mathcal{T}_1^{MC}$, the flux of probability into any state in $\mathcal{T}_1^{MC}$ is always equal to the probability flux out. Consequently, the net flux for any state $\mathbf{y}$ in $\mathcal{T}_1^{MC}$ is zero and consequently $\phi_Y(\mathbf{y}\,;\,t) = 0$ for all $t > 0$.

3.2.3 Scenario 3

Now consider the set of states through which probability flows into $\mathcal{Y}^{MC}$ from $\mathcal{Y}^{DE}$. According to equations (8) and (9) the flux of probability into the state $\mathbf{y}_1$ in $\mathcal{T}_1$ at time t is distributed amongst two corresponding states in $\mathcal{T}_2^{MC}$ at time $t + t(\mathbf{y}_1)$. Suppose $\mathbf{y}_1$ is a state in $\mathcal{T}_1^{MC}$ and $\mathbf{y}$ is a state in $\mathcal{T}_2^{MC}$. In addition, define $\Pr(\mathbf{y}|\mathbf{y}_1)$ as the probability that the hybrid fluid process switches from fluid dynamics to CTMC dynamics through the state $\mathbf{y}$, conditioned on switching from CTMC dynamics to fluid dynamics through the state $\mathbf{y}_1$, given by equation (9). Then the flux of probability into the state $\mathbf{y}$ at time t is given by

$$\sum_{\mathbf{y}'\in\mathcal{D}} \phi_Y(\mathbf{y}'\,;\,t - t(\mathbf{y}_1))\; q_X(\mathbf{y}',\mathbf{y}_1) \sum_{\mathbf{y}_1\in\mathcal{T}_1^{MC}} \Pr(\mathbf{y}\,|\,\mathbf{y}_1),$$

with the condition that $\phi_Y(\mathbf{y}\,;\,u) = 0$ if $u < 0$.

Since the hybrid fluid process has CTMC dynamics when it is in the state $\mathbf{y}$, the probability flux when in state $\mathbf{y}$ is also influenced by other states in $\mathcal{Y}^{MC}$. Therefore, for all $\mathbf{y}$ in $\mathcal{T}_2^{MC}$, the probability flux is governed by the system of DDEs

$$\begin{aligned}\frac{d}{dt}\phi_Y(\mathbf{y}\,;\,t) = &\sum_{\mathbf{y}''\in\mathcal{Y}^{MC}} \phi_Y(\mathbf{y}''\,;\,t)\; q_X(\mathbf{y}'',\mathbf{y}) \\ &+ \sum_{\mathbf{y}'\in\mathcal{D}} \phi_Y(\mathbf{y}'\,;\,t - t(\mathbf{y}_1))\; q_X(\mathbf{y}',\mathbf{y}_1) \sum_{\mathbf{y}_1\in\mathcal{T}_1^{MC}} \Pr(\mathbf{y}\,|\,\mathbf{y}_1)\,. \end{aligned} \tag{11}$$

We solve the system of DDEs (10)—(11) on the discrete set $\mathcal{Y}^{MC}$ using a modified version of the Implicit Euler scheme of Jenkinson and Goutsias (2012). In order to deal with the delayed terms in equation (11) we store the flux of probability into each state in $\mathcal{T}_1^{MC}$ in a secondary array and after $t(\mathbf{y}_1)$ time units have elapsed the probability returns to the system through the corresponding states in $\mathcal{T}_2^{MC}$. Under the assumption that $t(\mathbf{y}_1)$ is the same

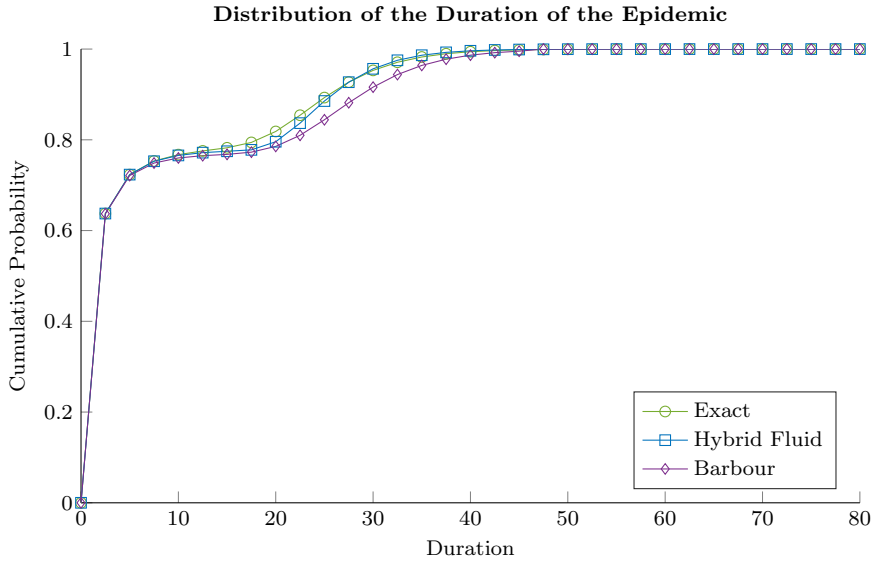

Fig. 2 The distribution of the duration of the epidemic calculated from the CTMC model, hybrid fluid model, and Barbour's model for $R_0 = 1.3$ and $N = 1000$ with one initially infectious individual.

for all $\mathbf{y}_1$ in $\mathcal{T}_1^{MC}$ the exact solution to the system of DDEs (10)—(11) is available (Yi and Ulsoy, 2006). However, the solution involves computing a large number of matrix exponentials and is less computationally efficient than the numerical scheme outlined above.

The probability that the duration of the epidemic is at most t is the sum over $\phi_Y(\mathbf{y}\,;\,t)$ for $\mathbf{y}$ in $\mathcal{A}$, where $\mathcal{A} = \{(S^Y, I^Y) \in \mathcal{Y}^{MC} : I^Y = 0\}$ denotes the set of absorbing states of $\mathcal{Y}^{MC}$. For future reference, we define $\mathcal{B} = \mathcal{Y}^{MC} \setminus \mathcal{A}$ as the set of transient states of $\mathcal{Y}^{MC}$. The system of DDEs (10)—(11) is calculated using Algorithm 1 of Section 5.

3.2.4 Numerical results.

Throughout this section we use a time grid which ranges from 0 to 80 with a time step of 0.01 because this ensures that the epidemic will be extinct with high probability before the terminal time (for $N \leq 10000$), and that the implicit Euler scheme of Jenkinson and Goutsias achieves a global L_1-error of $\mathcal{O}(10^{-2})$. We set $R_0 = 1.3$, for which our procedure for determining an appropriate threshold (inequality (17) in Section 5.2) gives $\widehat{I} = 17$.

Figure 2 shows the distribution of the duration of the epidemic calculated from the CTMC model (green with circles), hybrid fluid model (blue with squares), and Barbour's model (Barbour, 1975) (purple with diamonds) for $N = 1000$ with one initially infectious individual. The hybrid fluid model provides an accurate approximation of the distribution of the duration of the epidemic. Barbour's model approximates the sub-critical component of the

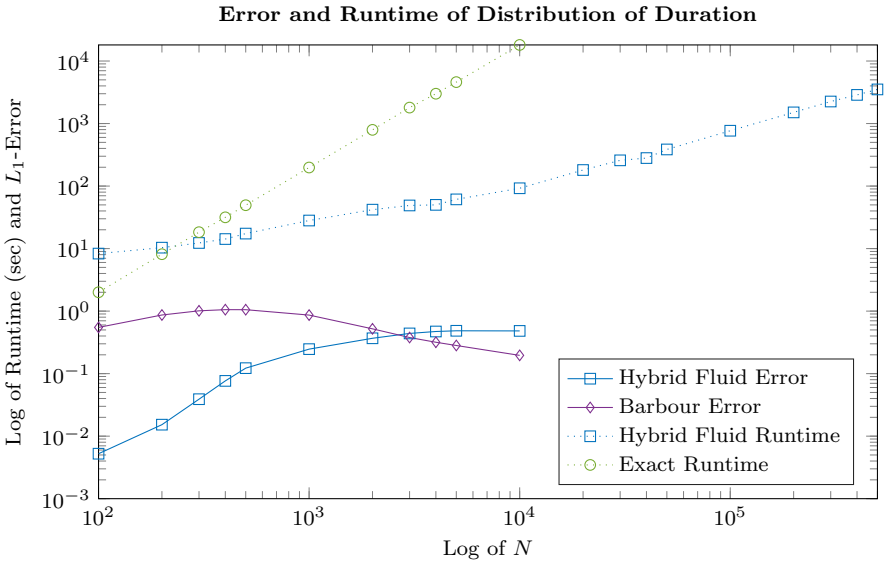

Fig. 3 The L_1-error and runtime of the distribution of the duration of the epidemic for the SIR CTMC, hybrid fluid model and Barbour's asymptotic approximation. The L_1-error of the hybrid fluid model is much better than Barbour's L_1-error for moderate N, and its runtime is $\mathcal{O}(N)$. Here we used $R_0 = 1.3$ for which $\widehat{I} = 17$ (inequality (17)), and the initial state $(N-1, 1)$.

duration of the epidemic accurately but the hybrid fluid model captures the super-critical component of the duration of the epidemic more accurately.

Figure 3 shows the runtime of the CTMC model (dotted green with circles) and the runtime of the hybrid fluid model (dotted blue with squares) across a range of N with $R_0 = 1.3$. The slope of the line from the hybrid fluid model is approximately one, which indicates that the asymptotic runtime for using Algorithm 1 on the hybrid fluid model to calculate the distribution of the duration of the epidemic is of $\mathcal{O}(N)$, corresponding to the number of states in the state space of the hybrid fluid model which is approximately $\widehat{I}N$. Irrespective of the population size, Barbour's asymptotic approximation is effectively instantaneous to compute so its runtime has not been included in Figure 3.

Figure 3 also shows the L_1-error of the hybrid fluid model (solid blue with squares) and the L_1-error of Barbour's model (solid purple with diamonds). The L_1-error of the hybrid fluid model shows a significant improvement over Barbour's model for N less than 10^3 but appears to be consistent with Barbour's model for larger N. The L_1-error of the hybrid fluid model appears to increase with N which suggests that the main source of disagreement between the CTMC model and the hybrid fluid model is the length of time over which the CTMC is approximated by the fluid model. Although the L_1-error of the hybrid fluid approximation can generally be improved by increasing the threshold $\widehat{I}$, the hybrid fluid approximation does not show a significant im-

provement over Barbour's asymptotic approximation unless $\widehat{I}$ is large enough that the probability of $\mathbf{Y}(t)$ hitting the subset $\mathcal{Y}^{DE}$ is insignificant.

3.3 Final size of the epidemic

The distribution of the final size of the epidemic can be deduced from the system of DDEs (10)—(11) by numerically integrating the system until such a time that the epidemic is almost surely over. This approach, however, is not the most computationally efficient. The distribution of the final size of the epidemic is deduced from the hitting distribution of the embedded jump chain of the hybrid fluid process on the absorbing set $\mathcal{A}$.

Let $\{\mathbf{Y}_n\}_{n\geq 0}$ denote the embedded jump process of $\mathbf{Y}(t)$ which takes values in $\mathcal{Y}^{MC}$ and denote by $p_Y(\mathbf{y}, \mathbf{y}+\boldsymbol{\ell}_j)$ the probability that the jump process transitions from the state $\mathbf{y}$ to $\mathbf{y}+\boldsymbol{\ell}_j$, for $j = 1, 2$. For all $\mathbf{y}$ in $\mathcal{Y}^{MC} \setminus \mathcal{T}_1^{MC}$, the positive jump probabilities are:

$$
\begin{aligned}
p_Y(\mathbf{y}, \mathbf{y}+\boldsymbol{\ell}_1) &= \frac{\beta S}{\beta S + \gamma(N-1)} && \text{if } \mathbf{y}+\boldsymbol{\ell}_1, \mathbf{y}+\boldsymbol{\ell}_2 \in \mathcal{Y}^{MC}, \\
p_Y(\mathbf{y}, \mathbf{y}+\boldsymbol{\ell}_2) &= \frac{\gamma(N-1)}{\beta S + \gamma(N-1)} && \text{if } \mathbf{y}+\boldsymbol{\ell}_1, \mathbf{y}+\boldsymbol{\ell}_2 \in \mathcal{Y}^{MC}, \\
p_Y(\mathbf{y}, \mathbf{y}+\boldsymbol{\ell}_2) &= 1 && \text{if } \mathbf{y}+\boldsymbol{\ell}_1 \notin \mathcal{Y}^{MC} \text{ and } \mathbf{y}+\boldsymbol{\ell}_2 \in \mathcal{Y}^{MC};
\end{aligned}
$$

and for all $\mathbf{y}$ in $\mathcal{T}_1^{MC}$ the positive jump probabilities are given by equation (9).

Fix $\mathbf{y}_0$ in $\mathcal{B}$ and let $h_Y(\mathbf{y})$ denote the probability that $\mathbf{Y}_n$ ever hits the state $\mathbf{y}$ in $\mathcal{Y}^{MC}$, given the initial state $\mathbf{y}_0$. Then the hitting probabilities $h_Y(\mathbf{y})$ for all $\mathbf{y}$ in $\mathcal{Y}^{MC}$ are the minimal non-negative solution to the system of linear equations

$$
h_Y(\mathbf{y}) = \sum_{\mathbf{y}' \in \mathcal{Y}^{MC}} h_Y(\mathbf{y}')\, p_Y(\mathbf{y}', \mathbf{y}), \tag{12}
$$

with $h_Y(\mathbf{y}_0) = 1$. The distribution of the final size of the epidemic, given the initial state $\mathbf{y}_0$, is the $(N+1)\times 1$ vector with entries $h_Y(\mathbf{y})$ for all $\mathbf{y}$ in $\mathcal{A}$. The solution to the system of linear equations (12) is calculated using Algorithm 1.

3.3.1 Numerical results.

Figure 4 shows the distribution of the final size of the epidemic calculated from the CTMC model (green with circles) and the hybrid fluid model (blue with squares) for $R_0 = 1.3$ and $N = 1000$ with one initially infectious individual. The hybrid fluid model approximates the sub-critical component of the final size accurately but fails to approximate the super-critical component of the final size distribution.

Figure 5 shows the runtime of the CTMC model (dotted green with circles) and the runtime of the hybrid fluid model (dotted blue with squares) across a range of N for $R_0 = 3$. The asymptotic slope of the curve of the runtime for the

hybrid fluid model is approximately one, which indicates that the asymptotic runtime of Algorithm 1 on the hybrid fluid model is $\mathcal{O}(N)$.

Figure 5 also shows the L_1-error of the hybrid fluid model (solid blue with squares). The L_1-error of the hybrid fluid model appears to converge to a value around 66% of the largest possible L_1-error, suggesting that $\mathbf{Y}(t)$ approximates well the $1/R_0$ proportion of sample paths which become extinct close to $S^Y = N$, but fails to approximate the $1 - 1/R_0$ proportion of sample paths which become extinct near $S^Y = 0$. This confirms our intuition that the source of disagreement between $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ propagates from the time interval over which the fluid approximation is used to approximate the underlying CTMC. Although $\widehat{I} = 17$ has been identified as a reasonable threshold (inequality (17) in Section 5.2), the asymptotic error may generally be decreased by selecting a larger threshold. However, the L_1-error is fairly insensitive to changing the threshold.

The hybrid fluid model approximates the time that the SIR CTMC spends above the threshold quite well, but the variability in the S component of the model when switching from fluid dynamics to CTMC dynamics, not captured by the fluid approximation, is important in determining the final size distribution.

4 The hybrid diffusion model

We now introduce the hybrid diffusion model which uses the diffusion approximation to capture the fluctuations of the underlying CTMC about the deterministic trajectory of the fluid approximation. The hybrid diffusion model is constructed in a similar way to the hybrid fluid model except that it uses the diffusion approximation in place of the fluid approximation. The hybrid diffusion model is used to calculate the distribution of the final size of the epidemic.

4.1 Model formulation

Let $\{\mathbf{Z}(t)\}_{t\geq 0}$ denote the hybrid diffusion process, which takes values (S^Z, I^Z) in the set $\mathcal{Y}$. As with the hybrid fluid process, the hybrid diffusion process switches dynamics depending on which subset of $\mathcal{Y}$ it is in. In particular, when $\mathbf{Z}(t)$ is in the subset $\mathcal{Y}^{MC}$ it has the dynamics of the SIR CTMC $\mathbf{X}(t)$, and when $\mathbf{Z}(t)$ is in the subset $\mathcal{Y}^{DE}$ it has the dynamics of the diffusion approximation. The dynamics of the hybrid diffusion process at the interface $\mathcal{T}$ are considered in more detail.

Since the fluid approximation provides the mean drift of the diffusion approximation, if $\mathbf{Z}(t)$ hits a state in $\mathcal{T}_1^{MC}$ there is a high probability that the diffusion dynamics will immediately force $\mathbf{Z}(t)$ out of $\mathcal{T}_1^{MC}$ and into $\mathcal{Y}^{DE}$. Due to the stochastic nature of the diffusion dynamics, if $\mathbf{Z}(t)$ is in $\mathcal{Y}^{DE}$, then there is a non-zero probability that $\mathbf{Z}(t)$ will hit any state $(S^Z, \widehat{I})$ such that

S^Z is in $\mathbb{R}$. In which case (i) if $S^Z > N/R_0$ then there is a high probability that $\mathbf{Z}(t)$ will be forced back into $\mathcal{Y}^{DE}$, and (ii) if $S^Z \leq N/R_0$ then there is a high probability that $\mathbf{Z}(t)$ will become trapped close to the boundary $\mathcal{T}_2$. It follows that we allow the hybrid diffusion process to switch between CTMC dynamics and diffusion dynamics upon hitting a state in $\mathcal{T}_1^{MC}$ and force $\mathbf{Z}(t)$ to switch from diffusion dynamics to CTMC dynamics upon hitting a state in $\mathcal{T}_2$.

As we are only using the hybrid diffusion process to approximate the distribution of the final size of the epidemic, we only need to use the diffusion dynamics of $\mathbf{Z}(t)$ to calculate the jump probabilities of the states in $\mathcal{T}_1^{MC}$. Let $\mathbf{z}_1 = (S_1^Z, \widehat{I})$ be a state in $\mathcal{T}_1^{MC}$, then it follows from Ethier and Kurtz (2008) that the next hitting distribution of $\mathbf{Z}(t)$ on the set of states with $\widehat{I}$ infectious individuals is normally distributed with mean $S_2^Y(\mathbf{z}_1)$ (equation (7)) and variance

$$\boldsymbol{\Sigma}_{1,1}(t(\mathbf{z}_1)) + \frac{\boldsymbol{\Sigma}_{2,2}(t(\mathbf{z}_1))}{\left(1 - 1/\left(R_0\, S_2^Y(\mathbf{z}_1)\right)\right)^2} + \frac{2\,\boldsymbol{\Sigma}_{1,2}(t(\mathbf{z}_1))}{1 - 1/\left(R_0\, S_2^Y(\mathbf{z}_1)\right)}, \tag{13}$$

where $\boldsymbol{\Sigma}$ is governed by equation (6) and $t(\mathbf{z}_2)$ is given by equation (8). Let $F(u\,;\,\mathbf{z}_1)$ denote the cumulative density function of this hitting distribution, given that $\mathbf{Z}(t)$ switched from CTMC dynamics to diffusion dynamics through the state $\mathbf{z}_1$ in $\mathcal{T}_1^{MC}$. Then for all $\mathbf{z}_1 = (S_1^Z, \widehat{I})$ in $\mathcal{T}_1^{MC}$ and $\mathbf{z} = (S^Z, \widehat{I})$ in $\mathcal{T}^{MC}$, the positive jump probabilities are

$$p_Z(\mathbf{z}_1, \mathbf{z}) = \begin{cases} F\left(\frac{1}{2}\,;\,\mathbf{z}_1\right) & \text{if } S^Z = 0, \\ F\left(S^Z + \frac{1}{2}\,;\,\mathbf{z}_1\right) - F\left(S^Z - \frac{1}{2}\,;\,\mathbf{z}_1\right) & \text{if } 1 \leq S^Z \leq S_1^Z - 2, \\ 1 - F\left(S_1^Z - \frac{1}{2}\,;\,\mathbf{z}_1\right) & \text{if } S^Z = S_1^Z - 1. \end{cases} \tag{14}$$

To account for the event that $\mathbf{Z}(t)$ hits a state $\mathbf{z} = (S^Z, \widehat{I})$ in $\mathcal{T}_1^{MC}$ from a state in $\mathcal{Y}^{DE}$, we define the following switching rule: (i) with probability $\gamma(N-1)/(\gamma(N-1)+\beta S^Z)$, $\mathbf{Z}(t)$ switches back to CTMC dynamics and has an instantaneous recovery event, and (ii) with probability $\beta S^Z/(\gamma(N-1)+\beta S^Z)$, $\mathbf{Z}(t)$ restarts diffusion dynamics from the state $\mathbf{z}$.

4.2 Final size of the epidemic

Let $\{\mathbf{Z}_n\}_{n\geq 0}$ denote the embedded jump process of $\mathbf{Z}(t)$ which takes values in $\mathcal{Y}^{MC}$. Then, for all $\mathbf{z} \in \mathcal{Y}^{MC} \setminus \mathcal{T}_1^{MC}$ the positive jump probabilities of $\mathbf{Z}_n$ are identical to those of $\mathbf{Y}_n$, and for all $\mathbf{z} \in \mathcal{T}_1^{MC}$ the positive jump probabilities are given by equation (14), in conjunction with the switching rule.

Fix $\mathbf{z}_0$ in $\mathcal{B}$, and let $h_Z(\mathbf{z})$ denote the probability that $\mathbf{Z}_n$ ever reaches the state $\mathbf{z}$ in $\mathcal{Y}^{MC}$, given the initial state $\mathbf{z}_0$. Then the hitting probabilities $h_Z(\mathbf{z})$ for all $\mathbf{z}$ in $\mathcal{Y}^{MC}$ are the minimal non-negative solution to the system of linear equations

$$h_Z(\mathbf{z}) = \sum_{\mathbf{z}' \in \mathcal{Y}^{MC}} h_Z(\mathbf{z}')\, p_Z(\mathbf{z}', \mathbf{z}), \tag{15}$$

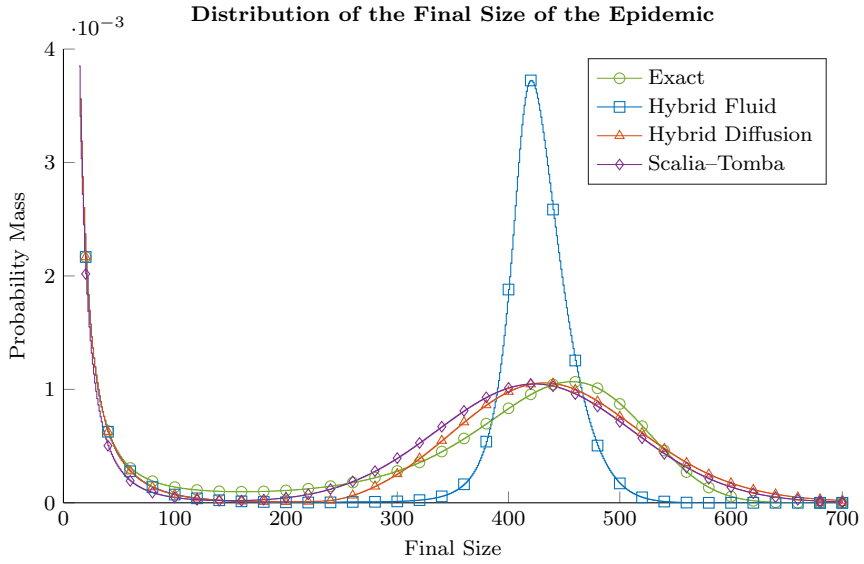

Fig. 4 The distribution of the final size of the epidemic calculated from the CTMC, hybrid fluid model, hybrid deterministic model, and Scalia–Tomba for $R_0 = 1.3$ and $N = 1000$ with one initially infectious individual.

with $h_Z(\mathbf{z}_0) = 1$. The distribution of the final size of the epidemic, given the initial state $\mathbf{z}_0$, is the $(N+1) \times 1$ vector with entries $h_Z(\mathbf{z})$, for all $\mathbf{z}$ in $\mathcal{A}$.

4.2.1 Numerical results.

Figure 4 shows the distribution of the final size of the epidemic calculated from the hybrid diffusion model (red with triangles) and Scalia-Tomba (1985) (purple with diamonds). The hybrid diffusion model and Scalia–Tomba's model approximate the sub-critical component of the final size accurately but neither model succeeds in fully describing the non-normality exhibited by the super-critical component of the epidemic.

Figure 5 shows the runtime of the hybrid diffusion model (dotted ochre with triangles). The asymptotic slope of the runtime line is approximately one, which indicates that the asymptotic runtime of Algorithm 1 for the hybrid diffusion model is $\mathcal{O}(N)$. The time difference between the runtime of the hybrid fluid model and the hybrid diffusion model corresponds to the time difference in calculating the hitting distributions of equations (9) and (14). Irrespective of N, Scalia–Tomba's approximation is effectively instantaneous to compute so its runtime has not been included in Figure 5.

Figure 5 also shows the L_1-error of the hybrid diffusion model (solid ochre with triangles) and Scalia–Tomba's model (solid purple with triangles). As N increases, the L_1-error of the hybrid diffusion approximation decreases achieving a minimum of a constant of order 10^{-2}, thereby showing a significant

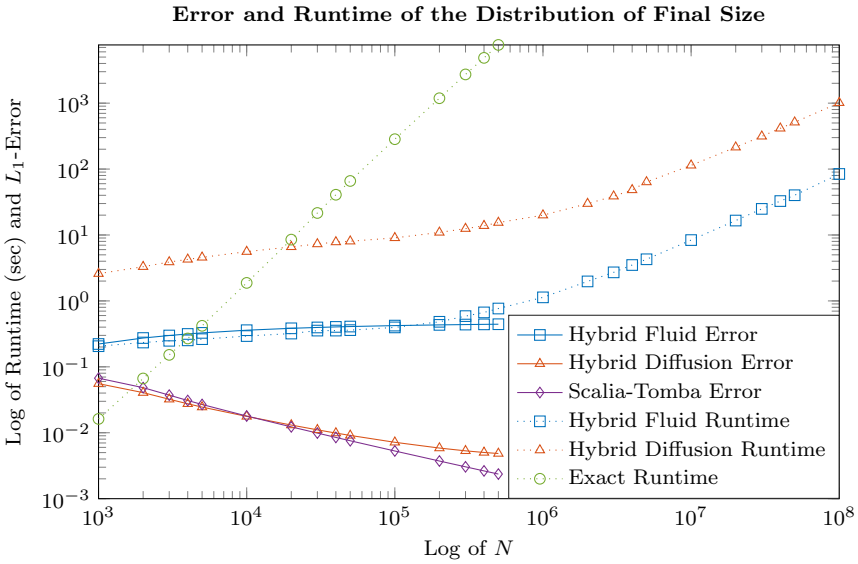

Fig. 5 The L_1-error and runtime of the distribution of the final size of the epidemic for both hybrid models and Scalia's asymptotic approximation, compared to the SIR CTMC. The error in the hybrid fluid model and the hybrid diffusion model is at-best constant of $\mathcal{O}(10^0)$ and $\mathcal{O}(10^{-3})$, respectively. The asymptotic slope of the runtime of Algorithm 1 on the hybrid models suggest that they are of computational complexity $\mathcal{O}(N)$ compared to the $\mathcal{O}(N^2)$ complexity of the Black and Ross algorithm. Here we used $R_0 = 1.3$ and $\widehat{I} = 17$ with the initial state $(N-1, 1)$.

improvement over the accuracy of the hybrid fluid model. Although the L_1-error can generally be decreased by increasing the threshold, the hybrid diffusion model does not achieve a significant improvement over Scalia–Tomba's approximation unless the probability that $\mathbf{Z}(t)$ hits a state in $\mathcal{Y}^{DE}$ is very small.

5 Numerical implementation

This section presents the algorithm we use to calculate the distribution of the duration of the epidemic and the distribution of the final size of the epidemic from the systems of equations presented in Sections 3 and 4, and describes our approach to calculating a suitable value for the threshold $\widehat{I}$.

5.1 Degree of advancement representation

The SIR CTMC is known as a population process because its state space is defined using the population numbers S and I. Alternatively, the degree-of-advancement (DA) representation (Jenkinson and Goutsias, 2012; Black and

Ross, 2015) is a counting process which tracks the number of infection and recovery events.

Let $\{\mathbf{N}(t)\}_{t\geq 0}$ denote the DA representation of either the hybrid fluid process or the hybrid diffusion process which takes values (N_I, N_R) from the state space $\mathcal{N}$. The DA numbers N_I and N_R are uniquely identified from the population numbers S and I since

$$N_I = N - S, \qquad N_R = N - S - I.$$

We define the DA sets $\mathcal{N}$, $\mathcal{N}^{MC}$ and $\mathcal{N}_1^T$ as the DA representations of the population sets $\mathcal{Y}$, $\mathcal{Y}^{MC}$ and $\mathcal{T}_1^{MC}$, respectively.

The DA representation is more amenable to numerical analysis than the population representation because the DA numbers N_I and N_R are monotonically increasing in time. Thus, in order to calculate a quantity for a particular state, one only needs to calculate the same quantity for all states leading up-to that state. For example, in order to calculate the hitting probability of the state (N_I, N_R) in $\mathcal{N}^{MC}$, one only needs to calculate the hitting probability of every state (N_I', N_R') in $\mathcal{N}^{MC}$ such that $(N_I', N_R') \neq (N_I, N_R)$ with $N_I' \leq N_I$ and $N_R' \leq N_R$.

We order the states in $\mathcal{N}^{MC}$ such that the state (N_I, N_R) precedes the state (N_I', N_R'), denoted $(N_I, N_R) \succ (N_I', N_R')$, if and only if

$$N_I - N_R < N_I' - N_R' \qquad \text{or} \qquad N_I - N_R = N_I' - N_R' \text{ and } N_I > N_I'. \quad (16)$$

We index each state $\mathbf{n}_k$ in $\mathcal{N}^{MC}$ by $k = 1, 2, \ldots, |\mathcal{N}^{MC}|$ such that $\mathbf{n}_1 \succ \mathbf{n}_2 \succ \cdots \succ \mathbf{n}_{|\mathcal{N}^{MC}|}$. Define $\delta k_1 = N - N_I + N_R$ and $\delta k_2 = N + 2 - N_I + N_R$ as the change in the index k due to an infection event or a recovery event, respectively. We now describe how we use the DA representation to calculate the distribution of the duration of the epidemic and the distribution of the final size of the epidemic.

In order to calculate the distribution of the final size of the epidemic we must calculate the solution to the systems of linear equations (12) and (15). Let $\boldsymbol{\varphi}$ be the $|\mathcal{N}^{MC}| \times 1$ vector whose kth element is the probability of ever hitting the state $\mathbf{n}_k$ in $\mathcal{N}^{MC}$, for $k = 1, 2, \ldots, |\mathcal{N}^{MC}|$. In addition, let $f(k, k')$ be the jump probability from the state $\mathbf{n}_k$ to $\mathbf{n}_{k'}$ (with $f(k, k) = 0$), for $k, k' = 1, 2, \ldots, |\mathcal{N}^{MC}|$. Then, if $\boldsymbol{\varphi}$ is initialised as the distribution of $\mathbf{N}(0)$ on $\mathcal{N}^{MC}$, the distribution of the final size of the epidemic is calculated by iteratively updating the entries of $\boldsymbol{\varphi}$ via Algorithm 1, until the algorithm terminates.

In order to calculate the distribution of the duration of the epidemic we must iteratively solve the system of linear equations (10)—(11) over a grid of time points (Jenkinson and Goutsias, 2012). This can be achieved using Algorithm 1 by simply re-defining $\boldsymbol{\varphi}$ and $f(k, k')$. Let $\boldsymbol{\varphi}$ be the $|\mathcal{N}^{MC}| \times 1$ vector whose kth element is the probability mass of the state $\mathbf{n}_k$ in $\mathcal{N}^{MC}$ at time $t + \Delta t$, for $k = 1, 2, \ldots, |\mathcal{N}^{MC}|$. In addition, let $f(k, k')$ be the transition rate from the state $\mathbf{n}_k$ to $\mathbf{n}_{k'}$, multiplied by Δt, (with $f(k, k) = \sum_{k' \neq k} f(k, k')$), for $k, k', = 1, 2, \ldots, |\mathcal{N}^{MC}|$. Then, if $\boldsymbol{\varphi}$ is initialised as the distribution of $\mathbf{N}(t)$,

the distribution of $\mathbf{N}(t+\Delta t)$ is calculated by iteratively updating the entries of $\boldsymbol{\varphi}$ via Algorithm 1, until the algorithm terminates.

When calculating the distribution of the final size of the epidemic from the hybrid diffusion model, we reduce the computational over-head of Algorithm 1 by only calculating the mean and variance of the hitting distribution (14) for a subset of states in $\mathcal{T}_1^{MC}$, and then extrapolating to the rest of the states in $\mathcal{T}_1^{MC}$ using a linear interpolant. More specifically, let $\boldsymbol{\theta}(\mathbf{y}_1) = \left(S_2^Y(\mathbf{y}_1), \boldsymbol{\Sigma}_{1,1}(t(\mathbf{y}_1)), \boldsymbol{\Sigma}_{1,2}(t(\mathbf{y}_1)), \boldsymbol{\Sigma}_{1,1}(t(\mathbf{y}_1))\right)$ for $\mathbf{y}_1$ in $\mathcal{T}_1^{MC}$, and $\mathcal{T}^* = \{(S^Y, \widehat{I}) \in \mathcal{T}_1^{MC} : S^Y = S_0^Y, S_0^Y + k, S_0^Y + 2k, \ldots, N - \widehat{I}\}$ where $S_0^Y = \lfloor N/R_0 \rfloor$ and k is a positive integer. Then we evaluate $\boldsymbol{\theta}(\mathbf{y}_1)$ for every $\mathbf{y}_1$ in $\mathcal{T}^*$ and use the output to approximate $\boldsymbol{\theta}(\mathbf{y}_1)$ for every $\mathbf{y}_1$ in $\mathcal{T}_1^{MC} \setminus \mathcal{T}^*$ using a linear interpolant. We found 30 to be a robust choice for k which provides a substantial computational advantage over the $k = 1$ case without accumulating too much error.

5.2 Choice of threshold

Our approach to determining a suitable threshold is based on using the branching process approximation of the SIR CTMC to estimate how large the subcritical epidemic will grow before it becomes extinct. Our reasoning for this approach is that we want an expression for the threshold which can be computed effectively instantaneously which will provide a threshold that is large enough to minimise the effect of switching dynamics on the subcritical component of the epidemic, without accruing unnecessary computational over-head. In order to achieve this, we condition our underlying CTMC on extinction (Waugh, 1958) and then investigate the distribution of the maximum of the corresponding branching process approximation (Ball and Donnelly, 1995).

Let $\{U(t)\}_{t\geq 0}$ denote the branching process approximation of the population of infectious individuals (conditioned on extinction) at time t, which takes values $0, 1, 2, \ldots$, and define $M = \sup_{0\leq t\leq\infty} U(t)$. Then $\widehat{I}$ is defined as the minimum m, for $m = 0, 1, 2, \ldots$, which satisfies $\Pr(M \geq m) \leq \epsilon$. In particular, using Section 5 of Ball and Donnelly (1995), the threshold $\widehat{I}$ is the minimum m which satisfies the inequality

$$m \geq U(0) + \frac{\log\left(R_0^{U(0)} + \epsilon - 1\right) - \log(\epsilon)}{\log(R_0)}. \tag{17}$$

In the event that $R_0 < 1$, R_0 is replaced by $1/R_0$ in inequality (17) as the process $U(t)$ will almost surely become extinct. However, inequality (17) can not be used if $R_0 = 1$. We determined that 5×10^{-3} is a suitable value for ϵ, in $[0, 1)$, due to the following observation. Note that, choosing a smaller ϵ leads to a larger choice of $\widehat{I}$ and hence, generally, more accurate results but larger computational runtimes.

For the distribution of the final size (duration) of the epidemic, the ochre (green) curve with triangles (circles) in Figure 6 shows the minimum threshold

Algorithm 1: Algorithm for calculating the distribution of the duration of the epidemic and the distribution of the final size of the epidemic.

Initialise the index k as $2N+1$ and initialise $\boldsymbol{\varphi}$.
for $N_R = 0, \ldots, N$ **do**
  Store the initial index $k_0 = k$, and update the current entry of $\boldsymbol{\varphi}$
  $\varphi_k = \varphi_k/(1 + f(k,k))$,
  for $N_I = N_R + 1, \ldots, \min\{N_R + \widehat{I} - 1, N - 1\}$ **do**
    Update the elements of $\boldsymbol{\varphi}$ which correspond to states in $\mathcal{N}^{MC}$ through their relationship with states in $\mathcal{N}^{MC} \setminus \mathcal{N}_1^T$.
    $\varphi_{k+\delta k_1} = \varphi_{k+\delta k_1} + \varphi_k\, f(k, k+\delta k_1)$,
    $\varphi_{k-\delta k_2} = \varphi_{k-\delta k_2} + \varphi_k\, f(k, k-\delta k_2)$.
    Update the state index $k = k + \delta k_1$.
  if $N_R < N - \widehat{I} - \lfloor N/R_0 \rfloor$ **then**
    for $j = 1, \ldots, N - N_I$ **do**
      Provided the probability of ever hitting the state $\mathbf{n}_k$ is greater than 10^{-7}, update the elements of $\boldsymbol{\varphi}$ which correspond to states in $\mathcal{N}_2^T$ through their relationship with states in $\mathcal{N}_1^T$.
      – For the final size of the epidemic
        $\varphi_{k-j} = \varphi_{k-j} + \varphi_k\, f(k, k-j)$.
      – For the duration of the epidemic, φ_k is stored in an additional array and the delayed flux $\varphi_k^{delayed}$ is used instead
        $\varphi_{k-j} = \varphi_{k-j} + \varphi_k^{delayed}\, f(k, k-j)$.
  else if $N_R < N$ **then**
    Update the elements of $\boldsymbol{\varphi}$ which correspond to states in $\mathcal{N}^{MC}$ through their relationship with states in $\mathcal{N}_2^T$.
    $\varphi_{k-\delta k_2} = \varphi_{k-\delta k_2} + \varphi_k\, f(k, k-\delta k_2)$.
  Reset the state index $k = k_0 - 1$.

required to achieve at most 0.1 (0.25) L_1-error when calculated with the hybrid diffusion (fluid) model. We chose these values because they correspond to the worst-case scenarios of Scalia-Tomba (1985) and Barbour (1974). Taking ϵ to be 5×10^{-3} produces the blue curve with squares which ensures a higher threshold than the ochre and green curves and hence ensures that the L_1-error in the distribution of the final size (duration) of the epidemic is at most 0.1 (0.25). However, this guarantee breaks down when R_0 exceeds 7.5, which we discuss next. As N increases, the minimum threshold required to achieve at most 0.1 (0.25) L_1-error in the distribution of the final size (duration) of the epidemic decreases and the threshold determined by inequality (17) stays the same. In addition, the point at which inequality (17) breaks down for the distribution of the final size of the epidemic increases.

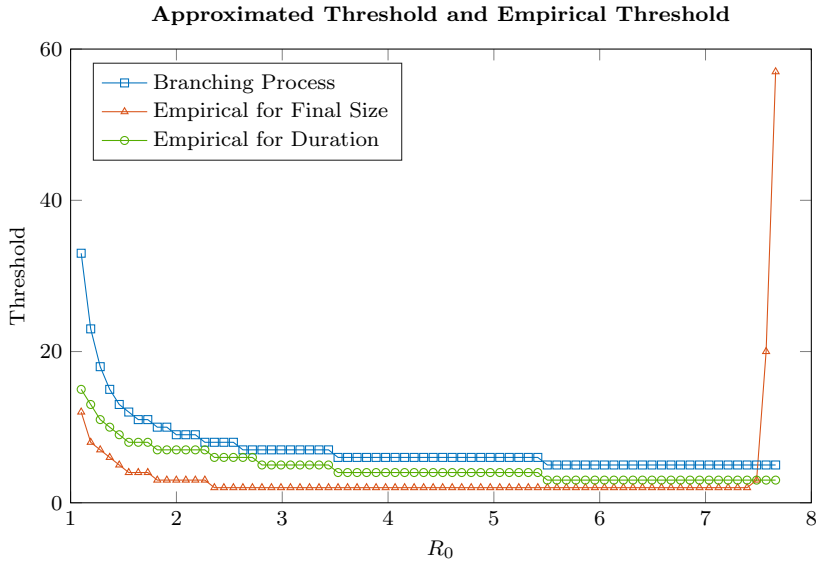

Fig. 6 For the distribution of the final size (duration) of the epidemic, the ochre (green) curve with triangles (circles) shows the minimum threshold $\widehat{I}$ which achieves an L_1-error of 0.1 (0.25). The blue curve with squares shows the threshold determined by inequality (17) using $\epsilon = 5 \times 10^{-3}$ which achieves at most 0.1 (0.25) L_1-error in the distribution of the final size (duration) of the epidemic, provided R_0 is less than approximately 7.5. Here we used $N = 10,000$ and the initial state $(N-1, 1)$.

Figure 7 shows the threshold determined by inequality (17) provides an L_1-error for the distribution of the final size of the epidemic which is at most 0.24. The divergence of the approximate distribution from the exact distribution manifests as an inaccurate approximation of the probability that the final size of the epidemic is N, $N-1$ or $N-2$. This divergence occurs when the diffusion approximation comes close to the absorbing boundary with $S = 0$ because the SIR CTMC is able to be absorbed by this set but the diffusion approximation is not. Figure 7 shows that the L_1-error decreases for $R_0 \geq 13$ which is because the probability that the final size of the epidemic is equal to $N-1$ or $N-2$ becomes negligible as R_0 becomes very large. The loss of the ability of inequality (17) to provide a reliable threshold is characterised as the region of R_0 for which the mean number of susceptible individuals during the fluid dynamics of $\mathbf{Z}(t)$ is less than eight.

6 Discussion

In this paper we have introduced two hybrid Markov chain models for approximating the distribution of the duration/final size of the SIR CTMC. These models are novel in the sense that no other hybrid models of the SIR CTMC have had the dynamics of the SIR CTMC during the early and final stages

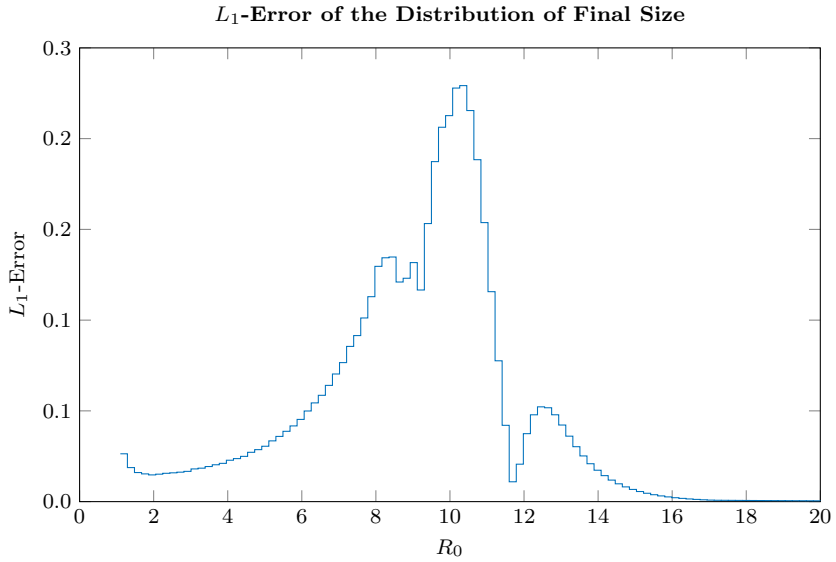

Fig. 7 The L_1-error of the distribution of the final size of the epidemic using inequality (17) to calculate the threshold. The error exceeds 0.1 on an interval of R_0 from approximately 7.5 to 11 and is at most 0.24. This issue arises when the fluid approximation of S falls below approximately eight susceptible individuals. Here we used $N = 10,000$ and the initial state $(N-1, 1)$.

of the epidemic. As a result, these models preserve the important stochastic features of the SIR CTMC which occur during these phases of the epidemic. Namely, the probability that the epidemic becomes extinct close to $S = N$, and the variability in the amount of time before the number of infectious individuals assumes an exponential-like trajectory. For the case of the general stochastic epidemic, we have been able to use these hybrid models to derive expressions for the distribution of the duration of the epidemic and the distribution of the final size of the epidemic which can be solved numerically in $\mathcal{O}(N)$ time, as opposed to the $\mathcal{O}(N^2)$ of the original CTMC model. This has enabled us to calculate the distribution of the duration of the epidemic and the final size of the epidemic for populations up to order 10^7, within a matter of hours. Our approximations of the distribution of the duration of the epidemic and the distribution of the final size of the epidemic achieve a similar level of accuracy to the existing asymptotic approximations and we believe that our methodology has the additional advantage of being straightforward, intuitive and generalisable.

6.1 Improvements

The hybrid models presented here were observed to provide inaccurate approximations of the distribution of the final size of the epidemic for a particular

region of R_0. This is because the S^Z component of the mean trajectory of the diffusion approximation comes close to the $S = 0$ absorbing boundary of the Markov chain, thereby causing the diffusion approximation to break down. This motivates modifying the diffusion hybrid model to include an additional threshold on the number of susceptible individuals which is conceptually similar to the two-stage model of Safta et al (2015).

6.2 Extensions

The general stochastic epidemic is a simple model which is often embedded within more complex models. The hybrid models presented here may be useful in scenarios where the embedded SIR model has a large population size, as would be the case when modelling disease dynamics within-hosts and between-hosts. A notable deficiency of the general stochastic epidemic is that it is not biologically plausible for an infectious period to be exponentially distributed. As such, the hybrid models presented here may be extended to incorporate phase-type infectious periods as would be the case with a SIIR model. Alternatively, the hybrid modelling methodology presented here could be extended to a wide range of epidemic models. Epidemic models like the SEIR model or the SIRS model would be natural extensions. Finally, the hybrid modelling methodology could also be used for developing a computationally efficient scheme for generating realisations from an underlying, potentially complex, CTMC.

References

Andersson H, Britton T (2000) Stochastic Epidemic Models and Their Statistical Analysis, Lecture Notes in Statistics, vol 151, 1st edn. Springer-Verlag New York, doi: 10.1007/978-1-4612-1158-7

Bailey NTJ (1950) A simple stochastic epidemic. Biometrika doi: 10.1093/biomet/37.3-4.193

Bailey NTJ (1957) The Mathematical Theory of Epidemics. Griffin, London

Ball F, Donnelly P (1995) Strong approximations for epidemic models. Stoch Proc Appl doi: 10.1016/0304-4149(94)00034-Q

Ball F, Neal P (2010) Workshop on Branching Processes and Their Applications, Lecture Notes in Statistics, vol 197, Springer Berlin Heidelberg, Germany, chap Applications of branching processes to the final size of SIR epidemics, pp 207–223. doi: 10.1007/978-3-642-11156-3˙15

Barbour A (1975) The duration of the closed stochastic epidemic. Biometrika doi: 10.1093/biomet/62.2.477

Barbour AD (1974) On a functional central limit theorem for Markov population processes. Adv Appl Prob doi: 10.2307/1426205

Barbour AD (1976) Quasi-stationary distributions in Markov population processes. Adv Appl Prob doi: 10.2307/1425906

Barbour AD (1980a) Biological Growth and Spread, Lecture Notes in Biomathematics, vol 38, Springer Berlin Heidelberg, Germany, chap Density dependent Markov population processes, pp 36–49. doi: 10.1007/978-3-642-61850-5˙4

Barbour AD (1980b) Equilibrium distributions of Markov population processes. Adv Appl Prob doi: 10.2307/1426422

Bartlett MS (1949) Some evolutionary stochastic processes. J R Statist Soc 11:211–229

Bartlett MS (1956) Deterministic and stochastic models for recurrent epidemics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, vol 4

Black AJ, Ross J (2015) Computation of epidemic final size distributions. J Theoret Biol doi: 10.1016/j.jtbi.2014.11.029

Coulson T, Rohani P, Pascual M (2004) Skeletons, noise and population growth: The end of an old debate? Trends Ecol Evol doi: 10.1016/j.tree.2004.05.008

Ethier SN, Kurtz TG (2008) Markov Processes: Characterisation and Convergence. John Wiley and Sons, Inc., New Jersey, doi: 10.1002/9780470316658

Fox GA (1993) Life history evolution and demographic stochasticity. Evol Ecol doi: 10.1007/BF01237731

Grenfell BT, Wilson K, Finkenstadt BF, Coulson TN, Murray S, Albon SD, Pemberton JM, Clutton-Brock TH, Crawley MJ (1998) Noise and determinism in synchronized sheep dynamics. Nature doi: 10.1038/29291

Jenkinson G, Goutsias J (2012) Numerical integration of the master equation in some models of stochastic epidemiology. PLoS ONE doi: 10.1371/journal.pone.0036160

Keeling MJ, Wilson HB, Pacala SW (2000) Reinterpreting space, time lags, and functional responses in ecological models. Science doi: 10.1126/science.290.5497.1758

Kendall DG (1965) Mathematical models of the spread of infection. Mathematics and Computer Science in Biology and Medicine

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond A doi: 10.1098/rspa.1927.0118

Kurtz TG (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. J Appl Prob doi: 10.2307/3212147

Kurtz TG (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. J Appl Prob doi: 10.2307/3211904

Lefèvre C (1990) Stochastic Processes in Epidemic Theory, Lecture Notes in Biomathematics, vol 86, Springer Berlin Heidelberg, Brussels, chap Stochastic Epidemic Models for SIR Infectious Diseases: a Brief Survey of the Recent General Theory, pp 1–12. doi: 10.1007/978-3-662-10067-7˙1

Nagaev AV, Startsev AN (1970) The asymptotic analysis of a stochastic model of an epidemic. Theory of Probability and Its Applications

Rand DA, Wilson HB (1991) Chaotic stochasticity: A ubiquitous source of unpredictability in epidemics. Proc R Soc Lond B doi: 10.1098/rspb.1991.0142

Safta C, Sargsyan K, Debusschere B, Najm HN (2015) Hybrid discrete/continuum algorithms for stochastic reaction networks. J Comput Phys doi: 10.1016/j.jcp.2014.10.026

Sazonov I, Kelbert M, Gravenor MB (2011) A two-stage model for the SIR outbreak: Accounting for the discrete and stochastic nature of the epidemic at the initial contamination stage. Math Biosci doi: 10.1016/j.mbs.2011.09.002

Scalia-Tomba G (1985) Asymptotic final-size distribution for some chain-binomial processes. Adv Appl Prob doi: 10.2307/1427116

Spagnolo B, Fiasconaro A, Valenti D (2003) Noise induced phenomena in Lotka-Volterra systems. Fluct Noise Lett doi: 10.1142/S0219477503001245

Watson R (1980) A useful random time-scale transformation for the standard epidemic model. J Appl Prob doi: 10.2307/3213022

Watson R (1981) An application of a martingale central limit theorem to the standard epidemic model. Stoc Proc Appl doi: 10.1016/0304-4149(81)90023-5

Waugh WAO (1958) Conditioned Markov processes. Biometrika doi: 10.1093/biomet/45.1-2.241

Yi S, Ulsoy A (2006) Solution of a system of linear delay differential equations using the matrix Lambert function. In: Proc. of the 25th American Control Conference, Minneapolis, pp 2433–2438, doi: 10.1109/ACC.2006.1656585