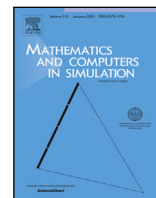




Contents lists available at ScienceDirect

# Mathematics and Computers in Simulation

journal homepage: [www.elsevier.com/locate/matcom](http://www.elsevier.com/locate/matcom)



## Original articles

# A computational approach to extreme values and related hitting probabilities in level-dependent quasi-birth–death processes

A. Di Crescenzo <sup>a</sup>, A. Gómez-Corral <sup>b,\*</sup>, D. Taïpe <sup>c</sup>

<sup>a</sup> Dipartimento di Matematica, Università degli Studi di Salerno, Via Giovanni Paolo II, 132, Fisciano, I-84084, Italy

<sup>b</sup> Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain

<sup>c</sup> Departamento de Estadística y Ciencia de Datos, Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

## ARTICLE INFO

MSC:  
60J28  
92B05

**Keywords:**

Epidemic model  
First-passage time  
Hitting probability  
Quasi-birth–death process

## ABSTRACT

This paper analyzes the dynamics of a level-dependent quasi-birth–death process  $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$ , i.e., a bi-variate Markov chain defined on the countable state space  $\cup_{i=0}^{\infty} l(i)$  with  $l(i) = \{(i, j) : j \in \{0, \dots, M_i\}\}$ , for integers  $M_i \in \mathbb{N}_0$  and  $i \in \mathbb{N}_0$ , which has the special property that its  $q$ -matrix has a block-tridiagonal form. Under the assumption that the first passage to the subset  $l(0)$  occurs in a finite time with certainty, we characterize the probability law of  $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$ , where  $I_{\max}$  is the running maximum level attained by process  $\mathcal{X}$  before its first visit to states in  $l(0)$ ,  $\tau_{\max}$  is the first time that the level process  $\{I(t) : t \geq 0\}$  reaches the running maximum  $I_{\max}$ , and  $J(\tau_{\max})$  is the phase at time  $\tau_{\max}$ . Our methods rely on the use of restricted Laplace–Stieltjes transforms of  $\tau_{\max}$  on the set of sample paths  $\{I_{\max} = i, J(\tau_{\max}) = j\}$ , and related processes under taboo of certain subsets of states. The utility of the resulting computational algorithms is demonstrated in two epidemic models: the SIS model for horizontally and vertically transmitted diseases; and the SIR model with constant population size.

## 1. Introduction

In this work, we are interested in continuous-time Markov chains  $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$  defined on a countable state space  $S$ , which we partition as  $\cup_{i=0}^{\infty} l(i)$  with subsets  $l(i) = \{(i, j) : j \in \{0, \dots, M_i\}\}$ , for integers  $M_i \in \mathbb{N}_0$  and  $i \in \mathbb{N}_0$ . The  $q$ -matrix of  $\mathcal{X}$  is assumed to be conservative and irreducible, and exhibits the structured form

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & & & & & \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & Q_{i,i-1} & Q_{i,i} & Q_{i,i+1} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & & & \ddots \end{pmatrix}, \tag{1}$$

where the sub-matrices  $Q_{i,i'}$  have entries  $q_{(i,j),(i',j')}$ , for pairs  $(i, j)$  with  $i \in \mathbb{N}_0$  and  $j \in \{0, \dots, M_i\}$ , and  $(i', j')$  with  $i' \in \{\max\{0, i - 1\}, i, i + 1\}$  and  $j' \in \{0, \dots, M_{i'}\}$ . This means that process  $\mathcal{X}$  can move in one step from state  $(i, j)$ , for  $i \in \mathbb{N}_0$  and  $j \in \{0, \dots, M_i\}$ ,

\* Corresponding author.

E-mail addresses: [adicrescenzo@unisa.it](mailto:adicrescenzo@unisa.it) (A. Di Crescenzo), [agcorral@ucm.es](mailto:agcorral@ucm.es) (A. Gómez-Corral), [dtaipe@ucm.es](mailto:dtaipe@ucm.es) (D. Taïpe).

URLs: <https://docenti.unisa.it/005305/home> (A. Di Crescenzo), <https://blogs.mat.ucm.es/agomez-corral> (A. Gómez-Corral).

<https://doi.org/10.1016/j.matcom.2024.08.019>

Received 27 March 2024; Received in revised form 3 July 2024; Accepted 16 August 2024

Available online 24 August 2024

0378-4754/© 2024 The Author(s). Published by Elsevier B.V. on behalf of International Association for Mathematics and Computers in Simulation (IMACS). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

to state  $(i', j')$  only if  $|i' - i| \leq 1$  and  $j' \in \{0, \dots, M_{i'}\}$ , and, consequently, the values of  $I(t)$  can only change by  $-1, 0$  and  $1$  units at jump times of  $\mathcal{X}$ . Process  $\mathcal{X}$  is commonly referred to as a level-dependent quasi-birth–death (LD-QBD) process, with the first variable  $I(t)$  termed the *level*, and the second one  $J(t)$  the *phase*; for convenience, we also refer to the subset  $l(i)$  as the  $i$ th *level*. Markov chains with LD-QBD structure can be thought of as the natural generalization of the uni-variate birth–death process (see, e.g., Anderson [1, Chapter 8]), competition processes (Reuter [2]) and related models, including prey–predator processes (Hitchcock [3]; Ridle-Rowe [4]), two-species competition processes (Gómez-Corral and López García [5]; Ridler-Rowe [6]), and birth/birth–death processes with biological applications (Ho et al. [7]). In the multi-variate setting, the multi-variate competition process of Iglehart [8], including a variety of birth–death processes on the  $m$ -orthant (Anderson [1]; Bharucha-Reid [9]; Jonckheere and Shneer [10]), may be easily formulated as a bi-variate LD-QBD process by defining the sum of all components of a point as the level variable. This means that points  $(i_1, \dots, i_m) \in \mathbb{N}_0^m$  with  $i_1 + \dots + i_m = i$  define level  $l(i)$ , and jumps of process  $\mathcal{X}$  amount to either the *birth* or the *death* of an individual affecting a single component, or the *mutation* of an individual moving from one coordinate to another one.

LD-QBD processes have been extensively investigated by several researchers, among whom are Baumann and Sandmann [11,12], Bright and Taylor [13], Phung-Duc et al. [14], and Takine [15]. Readers are also referred to Chapter 12 of Latouche and Ramaswami [16], the overviews of Bright and Taylor [17], and Kharoufeh [18], and references therein. The focus in these references is mainly on the stationary vector of a regular process  $\mathcal{X}$  and its expression in terms of *rate matrices*  $\{R_i : i \in \mathbb{N}_0\}$ , where  $R_i$  consists of expected rates of visit to state  $(i + 1, j')$  in level  $l(i + 1)$  per unit of the local time of  $l(i)$  spent in state  $(i, j)$ . To be concrete, the solution of Bright and Taylor [13] is developed as an extension of previous matrix-geometric methods of Neuts [19] and the logarithmic reduction solution in Ref. [20] – which are both related to a level-independent QBD process, where  $Q_{i,i-1} = Q_2$ , for  $i \in \{2, 3, \dots\}$ , and  $Q_{i,i} = Q_1$  and  $Q_{i,i+1} = Q_0$ , for  $i \in \mathbb{N}$  – for the level-dependent case. From the probabilistic interpretation of  $R_i$  and a suitable continued fraction representation, Phung-Duc et al. [14] provide a novel algorithmic procedure to evaluate the underlying rate matrices, which is found to be more efficient than that in Ref. [13]; for a related work, see Baumann and Sandmann [11]. In Ref. [12], Baumann and Sandmann present a matrix-analytic solution for numerically computing stationary expectations for long-run averages of additive functionals, such as long-run average costs and reward rates, without at first computing the stationary vector of  $\mathcal{X}$ . To the best of our knowledge, a complete explicit description of the rate matrices  $\{R_i : i \in \mathbb{N}_0\}$  has appeared in the literature only for level-independent QBD processes where  $Q_0$  or  $Q_2$  are of rank 1 (Ramaswami and Latouche [21]); for a specific class of level-independent QBD processes, see also the closed-form solution of van Leeuwen and Ninands [22], which reduces to counting lattice paths in the transition diagram from the time the process leaves a certain level until its first return to this level. The approach of Baumann [23] (see also Baumann and Sandmann [12]) for LD-QBD processes relies on truncating the state space  $S$  at a suitably selected level  $l(N)$ , which results in inevitable errors for which lower and upper bounds on stationary expectations are efficiently computed; see Ref. [23, Section 4]. Rastpour et al. [24] propose a novel method to calculate lower and upper bounds of the rate matrices of a LD-QBD process. The approach in Ref. [24] allows the truncation level  $N$  to be endogenously selected to ensure that, for performance measures that can be expressed as stationary probabilities, the calculated values are within the desired error tolerance relative to the actual values. QBD processes on a finite state space have been extensively studied in the literature, specifically in Refs. [25–29], among others. In the setting of LD-QBD processes with an *explosive* state space (i.e., as the cardinality  $1 + M_i$  of level  $l(i)$  increases exponentially with  $i$ ), Takine [15] presents approximation techniques for computing performance measures. For the analysis of quasi-stationary regime, see the work of Bean et al. [30], and Ramaswami and Taylor [31].

In this paper, the focus is on the extreme values of the LD-QBD process  $\mathcal{X}$  before the first entrance to states in  $l(0)$ , and hitting probabilities of the underlying phase process  $\{J(t) : t \geq 0\}$  as the maximum level of  $\mathcal{X}$  is visited for the first time. A first motivation to study extreme values and related hitting probabilities in a LD-QBD process comes from applications. For instance, the structured form of the  $q$ -matrix in (1) is often obtained directly in the case of retrial queues (Artalejo and Gómez-Corral [32]), in such a way that the length of a busy period is equivalent to the time to reach level  $l(0)$  in a suitably defined LD-QBD process; for a variety of retrial queues, see the papers by Dayar and Orhan [33], Jeganathan et al. [34], Liu and Zhao [35], Phung-Duc and Kawanishi [36], Phung-Duc [37], Sanga and Charan [38], and Saravanan et al. [39], among others. We shall demonstrate here that our approach (Section 2) does not need to exploit further properties of the retrial queue under consideration, except for the specific expressions of sub-matrices  $Q_{i,i'}$ , for  $i \in \mathbb{N}_0$  and  $i' \in \{\max\{0, i - 1\}, i, i + 1\}$ , in Eq. (1) defining the model. This is also applicable to queueing models with impatient customers and flexible matching mechanism (Chai et al. [40]). In epidemics (Baumann and Sandmann [41]; Lefèvre and Simon [42]), the first visit of  $\mathcal{X}$  to level  $l(0)$  usually amounts to the duration of an outbreak, and the maximum level of  $\mathcal{X}$  and the state of  $\{J(t) : t \geq 0\}$  at the first access to  $l(0)$  are related to the maximum number of simultaneously infected hosts during the outbreak and the state of the population at the peak of infection, respectively. As in the case of retrial queues, our approach (Section 2) should be seen to be a unified procedure in the analysis of epidemic models at the first time of maximum incidence of the pathogen, including multi-strain models (Chalub et al. [43]), stochastic chemical kinetics (Dayar et al. [44]), gene family evolution (Diao et al. [45]), resource-limited environments (Gómez-Corral et al. [46]), and recurrence in tumors (Santana et al. [47]).

As a second motivation, let us recall that extreme values of the LD-QBD process  $\mathcal{X}$  in (1) are analyzed by Mandjes and Taylor [48] (see also Javier and Fralix [49]) in terms of the *running maximum level*  $\bar{I}(T) = \sup\{I(t) : t \in [0, T]\}$  attained by  $\mathcal{X}$  at an independent exponentially distributed time  $T$  and, via Erlangization, at a predetermined time  $T$ ; for a related work in retrial queues, see Gómez-Corral and López García [50]. We shall complement here the work of Mandjes and Taylor [48] by characterizing the running maximum level of a LD-QBD process at time  $T$  that amounts to the first time process  $\mathcal{X}$  makes a transition to states contained in  $l(0)$ , and a detailed description of the phase process  $\{J(t) : t \geq 0\}$  as the running maximum is first attained. Interested readers in seeing a simpler description in epidemic models are referred to Amador et al. [51].

This paper proceeds as follows. In Section 2, we first introduce the basic notation for first-passage times of process  $\mathcal{X}$ , and give the formal definition of extreme value and phase as the maximum level is first attained. The distributional properties of the first time to reach the maximum level are analyzed in terms of taboo Laplace–Stieltjes transforms of certain suitably defined first-passage times (Section 2.2), and the probability law of the maximum level visited by process  $\mathcal{X}$  (Section 2.3). In Section 3, the computational approach is exemplified with two epidemic models: the first one is used to illustrate the solution in the case of a LD-QBD process with infinitely many states (Section 3.1), and the second one to do so in the case of a finite QBD process (Section 3.2). Section 4 concludes the paper with a brief discussion.

## 2. Extreme values of a LD-QBD process and related hitting probabilities

Throughout this section, the LD-QBD process  $\mathcal{X}$  is assumed to be regular and we study the dynamics of  $\mathcal{X}$  before the first visit to level  $l(0)$ , under the assumption that the first passage from any initial state  $(i_0, j_0) \in S \setminus l(0)$  to level  $l(0)$  occurs in a finite time with certainty. This assumption amounts to the fact that the null column vector is the largest non-negative solution bounded above by 1 to  $Q^*v = 0$ , where  $Q^*$  is the matrix obtained by removing from  $Q$  in (1) the rows and columns corresponding to states in  $l(0)$ ; see Kulkarni [52, Theorem 6.13].

For ease of presentation, we adopt the notation and terminology used in our recent work [28,29,53] and related papers [13,48,49]. More particularly, we define the first-passage time to level  $l(i)$  as  $\tau_{l(i)} = \inf\{t \geq 0 : I(t) = i\}$ , for integers  $i \in \mathbb{N}_0$ , and we let  $I_{\max}$  denote the running maximum level attained by  $\mathcal{X}$  before the first entrance into any state in level  $l(0)$ , and  $\tau_{\max}$  be the first time to reach the value  $I_{\max}$ . As a result,  $J(\tau_{\max})$  is the phase variable at time  $\tau_{\max}$ , and  $I_{\max} = I(\tau_{\max})$ .

**Remark 1.** Readers should note that Javier and Fralix [49] analyze the joint probability law of both the state  $(I(t), J(t))$  of process  $\mathcal{X}$  and its associated running maximum level  $\bar{I}(t)$  at any arbitrary time  $t$ , whereas the state  $(I_{\max}, J(\tau_{\max}))$  is inherently linked to the random time  $\tau_{\max}$ , provided that the entrance of  $\mathcal{X}$  to any state in level  $l(0)$  is certain.

Therefore, the analytical treatment of  $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$  contributes to a better understanding of the dynamics of LD-QBD processes in terms of their running maximum level. Furthermore, it contributes in a differentiated way to the general theory on structured Markov chains, recently addressed by the authors in the special case of QBD processes with finitely many states (Gómez-Corral and López-García [28]; Gómez-Corral et al. [29]) and infinitely many states (Gómez-Corral et al. [53]).

For later use, we introduce some notation as follows:  $0_a$  and  $1_a$  are the column vectors of order  $a$  such that all their entries are equal to 0 and 1, respectively;  $e_a(b)$  denotes the column vector of order  $a$  with its  $b$ th entry equals 1, and 0 elsewhere;  $0_{a \times b}$  and  $I_a$  are the null matrix of dimension  $a \times b$  and the identity matrix of order  $a$ , respectively;  $T$  denotes transposition; and  $\delta_{a,b}$  denotes the Kronecker’s delta.

### 2.1. Statement of the problem

Our objective is to characterize the conditional distribution of the random vector  $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$ , provided that  $(I(0), J(0)) = (i_0, j_0)$ , for a predetermined initial state  $(i_0, j_0) \in S \setminus l(0)$ . This distribution can be denoted by  $P(\tau_{\max}, I_{\max}, J(\tau_{\max}) | (i_0, j_0))$  in a concise manner, and it is related to the probabilities

$$P\left((\tau_{\max}, I_{\max}, J(\tau_{\max})) \in A \times \{(i, j)\} \mid (I(0), J(0)) = (i_0, j_0)\right),$$

for subsets  $A \in \beta[0, \infty)$  and states  $(i, j) \in \cup_{k=i_0}^{\infty} l(k)$ , where  $\beta[0, \infty)$  is the Borel  $\sigma$ -algebra on  $[0, \infty)$ . The problem is equivalent to determine the probability law of  $\tau_{\max}$  on those sample paths of process  $\mathcal{X}$  satisfying  $\{I_{\max} = i, J(\tau_{\max}) = j\}$ , for any state  $(i, j) \in \cup_{k=i_0}^{\infty} l(k)$ .

To start with, we notice that  $\tau_{\max}$  is a mixed random variable with a point mass at zero on the subset  $\{I_{\max} = i_0, J(\tau_{\max}) = j_0\}$ ; i.e., we have that

$$P(\tau_{\max} = 0, I_{\max} = i_0, J(\tau_{\max}) = j_0 \mid (I(0), J(0)) = (i_0, j_0)) = \delta_{(i_0, j_0), (i_0, j_0)} P_{(i_0, j_0)}(i_0),$$

where  $P_{(i_0, j_0)}(i) = P(I_{\max} = i \mid (I(0), J(0)) = (i_0, j_0))$ . On the sample paths of  $\mathcal{X}$  satisfying  $\{I_{\max} = i, J(\tau_{\max}) = j\}$ , for states  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ ,  $\tau_{\max}$  is strictly positive and behaves like a continuous random variable, for which we let  $f(t, i, j \mid i_0, j_0)$  denote its conditional density for  $t > 0$ , given that  $(I(0), J(0)) = (i_0, j_0)$ . For states  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ , we can then characterize the distribution of  $\tau_{\max}$  from the restricted Laplace–Stieltjes transform

$$\psi_{(i_0, j_0)}(\theta; i, j) = \int_0^{\infty} e^{-\theta t} f(t, i, j \mid i_0, j_0) dt,$$

for  $\Re(\theta) \geq 0$ , which we may also express as

$$\psi_{(i_0, j_0)}(\theta; i, j) = E\left[e^{-\theta \tau_{\max}} 1_{\{I_{\max}=i, J(\tau_{\max})=j\}} \mid (I(0), J(0)) = (i_0, j_0)\right],$$

where  $1_B$  is the indicator function of the subset  $B$ ; i.e.,  $1_B(\omega) = 1$  if  $\omega \in B$ , and 0 otherwise. As a result, the hitting probabilities  $P(\tau_{\max} > 0, I_{\max} = i, J(\tau_{\max}) = j \mid (I(0), J(0)) = (i_0, j_0))$ , for states  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ , may be directly evaluated as  $\psi_{(i_0, j_0)}(\theta; i, j) |_{\theta=0}$ .

It is worth noting that  $\psi_{(i_0, j_0)}(\theta; i, j)$  is inherently linked to the random interval  $[0, \tau_{l(0)})$  in such a way that the running maximum level  $I_{\max} = i$  is first attained by process  $\mathcal{X}$  at time  $\tau_{\max}$ , with  $\tau_{\max} < \tau_{l(0)}$ , and the subset of states  $\cup_{k=i+1}^{\infty} l(k)$  is then avoided during

the interval  $[\tau_{\max}, \tau_{l(0)})$ . More concretely, the random time  $\tau_{\max}$  on  $\{J(\tau_{\max}) = j\}$  is identically distributed to the first-passage time  $\tau_{l(i)}$  on  $\{J(\tau_{l(i)}) = j\}$  under the taboo of states in  $l(0)$ , which yields the expression

$$f(t, i, j | i_0, j_0) = g(t, i, j | i_0, j_0) P_{(i,j)}(i), \tag{2}$$

where  $g(t, i, j | i_0, j_0)$  denotes the density of  $\tau_{l(i)}$  on  $\{J(\tau_{l(i)}) = j\}$  under the taboo of states in  $l(0)$ , for  $t > 0$ . The term  $P_{(i,j)}(i)$  in (2) guarantees that, provided that  $(I(\tau_{\max}), J(\tau_{\max})) = (i, j)$ , the maximum level that  $\mathcal{X}$  visits during the residual interval  $[\tau_{\max}, \tau_{l(0)})$  is not greater than the level variable at time  $\tau_{\max}$ . Then, the expression for  $f(t, i, j | i_0, j_0)$  in (2) follows by applying the strong Markov property to the stopping time  $\tau_{(i,j)}$  – defined as the time of visit to state  $(i, j)$  – and the natural filtration.

This has the following immediate consequence.

**Lemma 1.** For any fixed state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ , the restricted Laplace–Stieltjes transform of  $\tau_{\max}$  on  $\{\tau_{\max} > 0, I_{\max} = i, J(\tau_{\max}) = j\}$  is determined by

$$\psi_{(i_0, j_0)}(\theta; i, j) = \varphi_{(i_0, j_0)}(\theta; i, j) P_{(i,j)}(i), \tag{3}$$

for  $\Re(\theta) \geq 0$ , where  $\varphi_{(i_0, j_0)}(\theta; i, j)$  is defined by  $\int_0^{\infty} e^{-\theta t} g(t, i, j | i_0, j_0) dt$ .

### 2.2. Taboo Laplace–Stieltjes transforms of $\tau_{l(i)}$

Lemma 1 tells us how to compute the restricted Laplace–Stieltjes transform  $\psi_{(i_0, j_0)}(\theta; i, j)$ , for states  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ , from the taboo Laplace–Stieltjes transform  $\varphi_{(i_0, j_0)}(\theta; i, j)$  and the marginal distribution of the running maximum level  $I_{\max}$ . Here, we show how one may compute  $\varphi_{(i_0, j_0)}(\theta; i, j)$  itself. To that end, we first introduce taboo Laplace–Stieltjes transforms  $\varphi_{(i', j')}(\theta; i, j)$  in a more general setting by replacing the initial state  $(i_0, j_0) \in S \setminus l(0)$  by the current state  $(i', j')$  of process  $\mathcal{X}$  at any arbitrary time, for  $(i', j') \in \cup_{k=1}^{i-1} l(k)$ .

**Theorem 1.** For any fixed state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ , the column vectors  $\Phi_{i'}(\theta; i, j) = (\varphi_{(i', j')}(\theta; i, j) : j' \in \{0, \dots, M_{i'}\})$ , for integers  $i' \in \{1, \dots, i-1\}$ , are uniquely characterized as the solution of the system of linear equations

$$\begin{cases} (\theta I_{1+M_{i'}} - Q_{i', i'}) \Phi_{i'}(\theta; i, j) &= (1 - \delta_{i', 1}) Q_{i', i'-1} \Phi_{i'-1}(\theta; i, j) + Q_{i', i'+1} \Phi_{i'+1}(\theta; i, j), \quad i' \in \{1, \dots, i-2\}, \\ (\theta I_{1+M_{i-1}} - Q_{i-1, i-1}) \Phi_{i-1}(\theta; i, j) &= Q_{i-1, i-2} \Phi_{i-2}(\theta; i, j) + Q_{i-1, i} e_{1+M_i} (1 + j). \end{cases} \tag{4}$$

Moreover, the column vectors  $m_{i'}^{(n)}(i, j)$  consisting of the  $n$ th moments of  $\tau_{l(i)}$  on  $\{\tau_{l(i)} < \tau_{l(0)}, J(\tau_{l(i)}) = j\}$ , for  $n \in \mathbb{N}$  and integers  $i' \in \{1, \dots, i-1\}$ , satisfy

$$\begin{cases} -Q_{i', i'} m_{i'}^{(n)}(i, j) &= nm_{i'}^{(n-1)}(i, j) + (1 - \delta_{i', 1}) Q_{i', i'-1} m_{i'-1}^{(n)}(i, j) + Q_{i', i'+1} m_{i'+1}^{(n)}(i, j), \quad i' \in \{1, \dots, i-2\}, \\ -Q_{i-1, i-1} m_{i-1}^{(n)}(i, j) &= nm_{i-1}^{(n-1)}(i, j) + Q_{i-1, i-2} m_{i-2}^{(n)}(i, j), \end{cases} \tag{5}$$

with  $m_{i'}^{(0)}(i, j) = \Phi_{i'}(\theta; i, j)|_{\theta=0}$ .

**Proof.** Let us fix the state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$  and assume that process  $\mathcal{X}$  jumps from its current state  $(i', j') \in \cup_{k=1}^{i-1} l(k)$  to state  $(i'', j'')$ , which occurs with probability  $q_{(i', j'), (i'', j'')} q_{(i', j')}$ , where  $q_{(i', j')} = -q_{(i', j'), (i', j')}$ . Then, the conditional distribution of  $\tau_{l(i)}$  on the set  $\{\tau_{l(i)} < \tau_{l(0)}, J(\tau_{l(i)}) = j\}$  is identical to the distribution of  $\eta_{(i', j')} + \tau_{(i'', j''), (i, j)}$ , where  $\eta_{(i', j')}$  is the sojourn time in  $(i', j')$  – which is exponentially distributed with mean  $q_{(i', j')}^{-1}$  – and  $\tau_{(i'', j''), (i, j)}$  is the first-passage time to state  $(i, j)$  starting from  $(i'', j'')$  under the taboo of states in the levels  $l(0)$  and  $l(i) \setminus \{(i, j)\}$ . It is clear that  $\tau_{(i'', j''), (i, j)} = 0$  almost surely if  $i' = i-1$  and  $(i'', j'') = (i, j)$ , and is strictly positive with density  $g(t, i, j | i'', j'')$  for  $t > 0$  otherwise. Thus, we have that

(i) For states  $(i', j') \in \cup_{k=1}^{i-2} l(k)$ ,

$$\begin{aligned} \varphi_{(i', j')}(\theta; i, j) &= (1 - \delta_{i', 1}) \sum_{j''=0}^{M_{i'-1}} \frac{q_{(i', j'), (i'-1, j'')}}{\theta + q_{(i', j')}} \varphi_{(i'-1, j'')}(\theta; i, j) + \sum_{j''=0, j'' \neq j'}^{M_{i'}} \frac{q_{(i', j'), (i', j'')}}{\theta + q_{(i', j')}} \varphi_{(i', j'')}(\theta; i, j) \\ &+ \sum_{j''=0}^{M_{i'+1}} \frac{q_{(i', j'), (i'+1, j'')}}{\theta + q_{(i', j')}} \varphi_{(i'+1, j'')}(\theta; i, j). \end{aligned} \tag{6}$$

(ii) For states  $(i', j') \in l(i-1)$ ,

$$\varphi_{(i-1, j')}(\theta; i, j) = \sum_{j''=0}^{M_{i-2}} \frac{q_{(i-1, j'), (i-2, j'')}}{\theta + q_{(i-1, j')}} \varphi_{(i-2, j'')}(\theta; i, j) + \sum_{j''=0, j'' \neq j'}^{M_{i-1}} \frac{q_{(i-1, j'), (i-1, j'')}}{\theta + q_{(i-1, j')}} \varphi_{(i-1, j'')}(\theta; i, j) + \frac{q_{(i-1, j'), (i, j)}}{\theta + q_{(i-1, j')}}. \tag{7}$$

Straightforward algebra allows us to verify that, by multiplying (6)–(7) by  $\theta + q_{(i', j')}$  and  $\theta + q_{(i-1, j')}$ , respectively, the matrix versions of the resulting equations are (4). Then, Eqs. (5) are derived by taking derivatives in (4) at  $\theta = 0$ , and observing that the column vector  $m_{i'}^{(n)}(i, j) = (-1)^n d^n \Phi_{i'}(\theta; i, j) / d\theta^n |_{\theta=0}$  contains the taboo expectations

$$E \left[ \tau_{l(i)}^n \mathbf{1}_{\{\tau_{l(i)} < \tau_{l(0)}, J(\tau_{l(i)}) = j\}} \mid (I(0), J(0)) = (i', j') \right],$$

for any current state  $(i', j') \in \cup_{k=1}^{i'-1} I(k)$ . This completes the proof.  $\square$

An important observation is that the equations in (4) (respectively, (5)) become a structured system of linear equations in the unknowns  $\Phi_{i'}(\theta; i, j)$  (respectively,  $m_{i'}^{(n)}(i, j)$ ) with  $i' \in \{1, \dots, i - 1\}$ , whence its solution can be evaluated by applying any general-purpose procedure, such as block Gaussian elimination; see, e.g., Stewart [54]. The solution of (4) and (5) may be then expressed as

$$\begin{aligned} \Phi_{i'}(\theta; i, j) &= \begin{cases} H_{i'}(\theta; i, j)\Phi_{i'+1}(\theta; i, j), & \text{if } i' \in \{1, \dots, i - 2\}, \\ H_{i-1}(\theta; i, j)e_{1+M_{i'}}(1 + j), & \text{if } i' = i - 1, \end{cases} \\ m_{i'}^{(n)}(i, j) &= \begin{cases} H_{i'}(i, j)m_{i'+1}^{(n)}(i, j) + h_{i'}(i, j), & \text{if } i' \in \{1, \dots, i - 2\}, \\ h_{i-1}(i, j), & \text{if } i' = i - 1, \end{cases} \end{aligned}$$

where

$$\begin{aligned} H_{i'}(\theta; i, j) &= \left( \theta I_{1+M_{i'}} - Q_{i',i'} - (1 - \delta_{i',1})Q_{i',i'-1}H_{i'-1}(\theta; i, j) \right)^{-1} Q_{i',i'+1}, \\ h_{i'}(i, j) &= \left( -Q_{i',i'} - (1 - \delta_{i',1})Q_{i',i'-1}H_{i'-1}(i, j) \right)^{-1} \left( (1 - \delta_{i',1})Q_{i',i'-1}h_{i'-1}(i, j) + nm_{i'}^{(n-1)}(i, j) \right), \end{aligned}$$

for  $i' \in \{1, \dots, i - 1\}$ . In these expressions, the column vectors  $m_{i'}^{(0)}(i, j)$  with  $i' \in \{1, \dots, i - 1\}$  record the values  $\varphi_{(i',j')}(\theta; i, j)|_{\theta=0}$ , for integers  $j' \in \{0, \dots, M_{i'}\}$ , and matrices  $H_{i'}(i, j)$  with  $i' \in \{1, \dots, i - 1\}$  are defined by  $H_{i'}(\theta; i, j)|_{\theta=0}$ .

This is summarized in Algorithms A–B, where vectors  $m_{i'}^{(0)}(i, j)$  and matrices  $H_{i'}(i, j)$  in Algorithm B are derived from Algorithm A by selecting  $\theta = 0$ . As in the well-known forward-elimination–backward-substitution method for solving a system of linear equations, the iterative scheme proceeds in two steps. During the first step, the procedure progressively reduces the subset of states  $\cup_{i'=1}^{i-1} I(i')$  by removing one level and evaluating the auxiliary matrices  $H_{i'}(\theta; i, j)$  and vectors  $h_{i'}(i, j)$  at each iteration; once these matrices and vectors are evaluated, the procedure iteratively computes the unknown vectors  $\Phi_{i'}(\theta; i, j)$  and  $m_{i'}^{(n)}(i, j)$  in the second step, by adding back one level at each iteration.

**Algorithm A.** Computation of the column vectors  $\Phi_{i'}(\theta; i, j)$  with  $i' \in \{1, \dots, i - 1\}$ , for  $\Re(\theta) \geq 0$  and a fixed state  $(i, j) \in \cup_{k=i_0+1}^{\infty} I(k)$ .

Step 1:  $i' := 1$ ;

$$H_{i'}(\theta; i, j) := (\theta I_{1+M_{i'}} - Q_{i',i'})^{-1} Q_{i',i'+1};$$

while  $i' < i - 1$ , repeat

$$i' := i' + 1;$$

$$H_{i'}(\theta; i, j) := (\theta I_{1+M_{i'}} - Q_{i',i'} - Q_{i',i'-1}H_{i'-1}(\theta; i, j))^{-1} Q_{i',i'+1};$$

endwhile.

Step 2:  $\Phi_{i'}(\theta; i, j) := H_{i'}(\theta; i, j)e_{1+M_{i'+1}}(1 + j)$ ;

while  $i' > 1$ , repeat

$$i' := i' - 1;$$

$$\Phi_{i'}(\theta; i, j) := H_{i'}(\theta; i, j)\Phi_{i'+1}(\theta; i, j);$$

endwhile.

**Algorithm B.** Computation of the column vectors  $m_{i'}^{(n)}(i, j)$  with  $i' \in \{1, \dots, i - 1\}$ , for  $n \in \mathbb{N}$  and a fixed state  $(i, j) \in \cup_{k=i_0+1}^{\infty} I(k)$ , in terms of  $m_{i'}^{(n-1)}(i, j)$  with  $i' \in \{1, \dots, i - 1\}$ .

Step 1:  $i' := 1$ ;

$$h_{i'}(i, j) := -nQ_{i',i'}^{-1}m_{i'}^{(n-1)}(i, j);$$

while  $i' < i - 1$ , repeat

$$i' := i' + 1;$$

$$h_{i'}(i, j) := (-Q_{i',i'} - Q_{i',i'-1}H_{i'-1}(i, j))^{-1};$$

$$h_{i'}(i, j) := h_{i'}(i, j)(Q_{i',i'-1}h_{i'-1}(i, j) + nm_{i'}^{(n-1)}(i, j));$$

endwhile.

Step 2:  $m_{i'}^{(n)}(i, j) := h_{i'}(i, j)$ ;

while  $i' > 1$ , repeat

$$i' := i' - 1;$$

$$m_{i'}^{(n)}(i, j) := H_{i'}(i, j)m_{i'+1}^{(n)}(i, j) + h_{i'}(i, j);$$

endwhile.

From Algorithms A–B, the taboo Laplace–Stieltjes transform of  $\tau_{I(i)}$  on  $\{\tau_{I(i)} < \tau_{I(0)}, J(\tau_{I(i)}) = j\}$  is easily obtained as

$$\varphi_{(i_0, j_0)}(\theta; i, j) = e_{1+M_{i_0}}^T (1 + j_0)\Phi_{i_0}(\theta; i, j),$$

and the expectation  $E \left[ (\tau_{\max})^n 1_{\{I_{\max}=i, J(\tau_{\max})=j\}} \right] (I(0), J(0) = (i_0, j_0))$  is evaluated as

$$e_{1+M_{i_0}}^T (1 + j_0)m_{i_0}^{(n)}(i, j)P_{(i, j)}(i),$$

for any state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ . Applying Euler and Post-Widder algorithms (Abate and Whitt [55]; Cohen [56]), we may invert numerically the transform  $\varphi_{(i_0, j_0)}(\theta; i, j)$  from Algorithm A, for any state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$  and the initial state  $(i_0, j_0) \in S \setminus l(0)$ . This yields the density  $g(t, i, j | i_0, j_0)$  of the first-passage time  $\tau_{l(i)}$  on the sample paths  $\{\tau_{l(i)} < \tau_{l(0)}, J(\tau_{l(i)}) = j\}$  of process  $\mathcal{X}$ , from which we have that  $f(t, i, j | i_0, j_0) = g(t, i, j | i_0, j_0)P_{(i, j)}(i)$  by (2).

### 2.3. Marginal distribution of $I_{\max}$

Using processes under a taboo (see, e.g., Section 2.1 in Gómez-Corral and López García [5]), the mass function  $\{P_{(i_0, j_0)}(i) : i \in \mathbb{N}_0\}$  of  $I_{\max}$  is found to be  $P_{(i_0, j_0)}(i) = 0$  if  $i \in \{0, \dots, i_0 - 1\}$ ; otherwise,  $P_{(i_0, j_0)}(i) = F_{\max}(i; i_0, j_0) - (1 - \delta_{i, i_0})F_{\max}(i - 1; i_0, j_0)$ , where the probability distribution function  $F_{\max}(i; i_0, j_0) = P(I_{\max} \leq i | (I(0), J(0)) = (i_0, j_0))$  has the form

$$F_{\max}(i; i_0, j_0) = e_{c(i)}^T ((1 - \delta_{i_0, 1})c(i_0 - 1) + 1 + j_0) (-T^{-1}(i)t_0(i)), \tag{8}$$

for  $i \in \{i_0, i_0 + 1, \dots\}$ . In this expression,  $c(i) = \sum_{i'=1}^i (1 + M_{i'})$  is the cardinality of the subset of states  $\cup_{k=1}^i l(k)$ , the matrix  $T(i)$  consists of infinitesimal rates  $q_{(i, j), (i', j')}$ , for states  $(i, j), (i', j') \in \cup_{k=1}^i l(k)$ , and  $t_0(i)$  is the column vector

$$t_0(i) = \begin{pmatrix} Q_{1,0} 1_{1+M_0} \\ 0_{c(i)-(1+M_1)} \end{pmatrix}.$$

It is worth noting that  $-T^{-1}(i)$  in (8) is a matrix of expected total times spent in states of  $\cup_{k=1}^i l(k)$ , whence  $-T^{-1}(i)$  is a non-negative matrix and  $-T^{-1}(i)t_0(i)$  contains probabilities that the first visit of  $\mathcal{X}$  to  $l(0)$  occurs in a finite time, provided that process  $\mathcal{X}$  does not visit states in the subset  $\cup_{k=i+1}^{\infty} l(k)$ , from any initial state in  $\cup_{k=1}^i l(k)$ .

Algorithm C derives the mass function of  $I_{\max}$  in a recursive manner. Its proof mostly repeats arguments of Ref. [5, Section 2.1], whence we omit the details.

**Algorithm C.** Computation of the non-null probabilities  $P_{(i_0, j_0)}(i)$ , for  $i \in \{i_0, i_0 + 1, \dots\}$  and a predetermined initial state  $(i_0, j_0) \in S \setminus l(0)$ .

```

Step 1:  $i := 1$ ;
 $-T^{-1}(i) := -Q_{i,i}^{-1}$ ;
if  $i = i_0$ , then
     $F_{\max}(i; i_0, j_0) := e_{c(i)}^T (1 + j_0) (-T^{-1}(i)t_0(i))$ ;
     $P_{(i_0, j_0)}(i) := F_{\max}(i; i_0, j_0)$ ;
endif.
Step 2: While  $F_{\max}(i; i_0, j_0) < 1 - \varepsilon$ , repeat
     $i := i + 1$ ;
     $A_{1,2} := \begin{pmatrix} 0_{c(i-2) \times (1+M_i)} \\ Q_{i-1,i} \end{pmatrix}$ ;
     $A_{2,1} := \begin{pmatrix} 0_{(1+M_i) \times c(i-2)}, Q_{i,i-1} \end{pmatrix}$ ;
     $B_{2,2} := (-Q_{i,i} - A_{2,1}(-T^{-1}(i-1))A_{1,2})^{-1}$ ;
     $B_{2,1} := B_{2,2}A_{2,1}(-T^{-1}(i-1))$ ;
     $B_{1,2} := -T^{-1}(i-1)A_{1,2}B_{2,2}$ ;
     $B_{1,1} := -T^{-1}(i-1)(I_{c(i-1)} + A_{1,2}B_{2,1})$ ;
     $-T^{-1}(i) := \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}$ ;
    if  $i \geq i_0$ , then
         $F_{\max}(i; i_0, j_0) := e_{c(i)}^T (c(i_0 - 1) + 1 + j_0) (-T^{-1}(i)t_0(i))$ ;
         $P_{(i_0, j_0)}(i) := F_{\max}(i; i_0, j_0) - F_{\max}(i - 1; i_0, j_0)$ ;
    endif;
endwhile.
    
```

In evaluating  $-T^{-1}(i)$  from the previously determined matrix  $-T^{-1}(i - 1)$ , Algorithm C (Step 2) computes the inverse matrix  $B_{2,2}$  of dimension  $1 + M_i$ . Since the number  $1 + M_i$  of phases does not necessarily increase with increasing values of  $i$ , it is not possible to give, in principle, general recommendations on how to address potential numerical precision problems in Algorithm C. These problems may occur due to either a large number  $1 + M_i$  of phases – for some level –, or to the additional operations needed to iteratively evaluate the matrices  $-T^{-1}(i)$  until at least  $1 - \varepsilon$  of mass from  $F_{\max}(i; i_0, j_0)$  is accumulated.

Note that, from an appropriate use of Algorithms A and C, it is possible to evaluate the following expressions for the hitting probabilities at time  $\tau_{\max}$  and the  $n$ th moment of  $J(\tau_{\max})$  on  $\{I_{\max} = i\}$ :

(i) For states  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$ ,

$$P(I_{\max} = i, J(\tau_{\max}) = j | (I(0), J(0)) = (i_0, j_0)) = \varphi_{(i_0, j_0)}(\theta; i, j) \Big|_{\theta=0} P_{(i, j)}(i).$$

**Table 1**  
Transitions from state  $(i, s)$  of process  $\mathcal{Y}$  and related events in the SIS model with vertical and horizontal transmission.

Event type	Event function	Event condition	Rate
New person-to-person infection	$(i + 1, s - 1)$	$i \neq 0, s \neq 0$	$\beta is$
Birth of an infectious individual	$(i + 1, s)$	$i \neq 0$	$(1 - p)\beta_I i$
Birth of a susceptible individual	$(i, s + 1)$	$(i, s) \neq (0, 0)$	$p\beta_I i + \beta_S s$
Death of an infectious individual	$(i - 1, s)$	$i \neq 0$	$\delta_I i$
Death of a susceptible individual	$(i, s - 1)$	$s \neq 0$	$\delta_S s$
Recovery of an infectious individual	$(i - 1, s + 1)$	$i \neq 0$	$\gamma i$

(ii) For  $n \in \mathbb{N}_0$  and integers  $i \in \{i_0 + 1, i_0 + 2, \dots\}$ ,

$$E \left[ (J(\tau_{\max}))^n 1_{\{I_{\max}=i\}} \mid (I(0), J(0)) = (i_0, j_0) \right] = \sum_{j=0}^{M_I} j^n \varphi_{(i_0, j_0)}(\theta; i, j) \Big|_{\theta=0} P_{(i, j)}(t).$$

To conclude the section, we briefly present the solution for a finite LD-QBD process.

**Remark 2.** For a process  $\mathcal{X}$  on the finite state space  $S = \cup_{i=0}^N I(i)$ , Algorithms A–B should be executed progressively for any state  $(i, j) \in \cup_{k=i_0+1}^N I(k)$  and a fixed initial state  $(i_0, j_0) \in \cup_{k=1}^{N-1} I(k)$ . In Algorithm C, the probabilities  $P_{(i_0, j_0)}(i)$  can be evaluated for integers  $i \in \{i_0, i_0 + 1, \dots, N\}$  by modifying Step 2 with the condition  $i < N - 1$  and noting that  $F_{\max}(N; i_0, i_0) = 1$ .

### 3. Epidemic models

In this section, the analytical solution in Section 2 is exemplified with two epidemic models: the SIS model for a pathogen transmitted both vertically and horizontally; and the SIR model with constant population size. For convenience, the number  $I(t)$  is related to infectious individuals in Sections 3.1–3.2, and the level of process  $\mathcal{X}$  is suitably defined according to the extinction of the population (Section 3.1) and the duration of an outbreak (Section 3.2).

#### 3.1. The SIS epidemic model for vertically and horizontally transmitted pathogens

In the terminology of Reuter [2], the SIS model with vertical and horizontal transmission is defined in Ref. [57] as a competition process  $\mathcal{Y} = \{(I(t), S(t)) : t \geq 0\}$ , where  $I(t)$  and  $S(t)$  are related to the number of infectious and susceptible individuals at time  $t$ , respectively. For states  $(i, s), (i', s') \in \mathbb{N}_0 \times \mathbb{N}_0$  with  $(i', s') \neq (i, s)$ , the dynamics of  $\mathcal{Y}$  are specified by the infinitesimal rates

$$q_{(i,s),(i',s')} = \begin{cases} \beta is, & \text{if } i' = i + 1 \text{ and } s' = s - 1, \\ (1 - p)\beta_I i, & \text{if } i' = i + 1 \text{ and } s' = s, \\ \delta_S s, & \text{if } i' = i \text{ and } s' = s - 1, \\ p\beta_I i + \beta_S s, & \text{if } i' = i \text{ and } s' = s + 1, \\ \delta_I i, & \text{if } i' = i - 1 \text{ and } s' = s, \\ \gamma i, & \text{if } i' = i - 1 \text{ and } s' = s + 1, \end{cases} \tag{9}$$

and  $q_{(i,s),(i,s)} = -(\beta is + (\delta_I + \beta_I + \gamma)i + (\delta_S + \beta_S)s)$ , for  $(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0$ , where  $\beta, \beta_S, \beta_I, \delta_S, \delta_I$  and  $\gamma$  are strictly positive, and  $p \in (0, 1)$ . In particular, the value  $\beta$  in (9) is the contact rate; parameters  $\beta_S$  and  $\delta_S$  (respectively,  $\beta_I$  and  $\delta_I$ ) are the birth and death rates of susceptible (respectively, infectious) individuals;  $\gamma$  is the recovery rate of an infectious individual; and  $p$  represents the proportion of the offspring of infectious parents that is born as susceptible. For a detailed description of events, see Table 1.

Epidemic models with vertical and horizontal transmission (Busenberg and Cooke [58]) are motivated by TORCH infections. The acronym TORCH refers to the congenital transmission of Toxoplasmosis, Other infections (including hepatitis B, coxsackievirus, syphilis, varicella-herpes zoster, and parvovirus B19, among others), Rubella, Cytomegalovirus, and Herpes simplex. The choice  $p \in (0, 1)$  in process  $\mathcal{X}$  results in the natural generalization of the SIS model with demography (Allen [59, Chapter 7]), where the prevalence probability  $1 - p$  of congenitally acquired disease is closely linked to the pathogen and can be very significant or negligible, as is the case of the human immunodeficiency virus (with estimated values varying between 31 – 36%, measured in terms of cases per 100 live births) and the parasitic protozoan *Toxoplasma gondii* (between 0.01 – 0.033%), respectively; see e.g. Neu et al. [60].

In Figs. 1–3, we focus on the life cycle of a population initially consisting of a single infectious individual and twenty susceptible ones, i.e.,  $I(0) = 1$  and  $S(0) = 20$ . Thus, the analytical results in Section 2 have to be applied here to process  $\mathcal{X} = \{(N(t), I(t)) : t \geq 0\}$  with  $N(t) = I(t) + S(t)$ , instead of  $\mathcal{Y}$ . Process  $\mathcal{X}$  can be routinely formulated as a LD-QBD process<sup>1</sup> defined on  $S = \cup_{n=0}^{\infty} I(n)$  with levels  $I(n) = \{(n, i) : i \in \{0, \dots, n\}\}$ , for  $n \in \mathbb{N}_0$ , in such a way that the time taken to access the absorbing state  $(0, 0)$  (i.e., level  $I(0)$ ) corresponds to the extinction time. This allows quantifying the incidence of the disease at the time of maximum population size through the number  $I(\tau_{\max})$  of infectious individuals, which acts as the phase of  $\mathcal{X}$  at time  $\tau_{\max}$ . The numerical examples in Figs. 1–3 are then derived from Algorithms A–C (i.e., Eqs. (4), (5) and (8)) and the corresponding expressions of submatrices  $Q_{n,n'}$ , for  $n \in \mathbb{N}$  and  $n' \in \{n - 1, n, n + 1\}$ , in process  $\mathcal{X}$  (Appendix A).

<sup>1</sup> Specifically, we let  $N(t)$  and  $I(t)$  be the level and the phase variables of  $\mathcal{X}$ , respectively. Hence,  $N_{\max} = N(\tau_{\max})$  is the maximum population size before extinction.

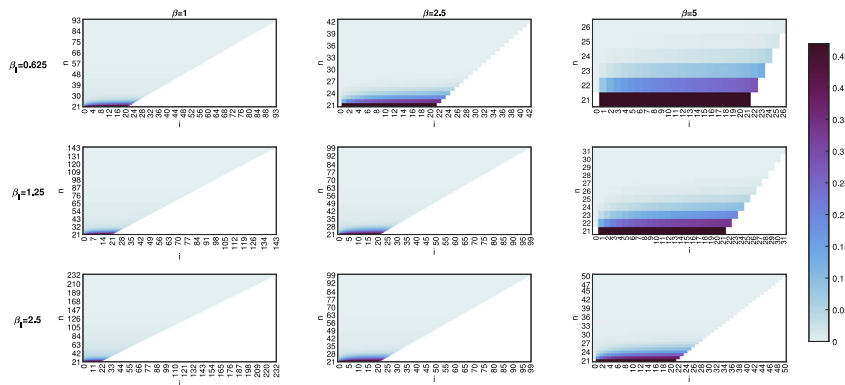


Fig. 1. Conditional probabilities  $P(N_{\max} = n, I(\tau_{\max}) \leq i | (N(0), I(0)) = (21, 1))$ , for integers  $n \in \{21, \dots, n_{0.95}\}$  and  $i \in \{\delta_{n,21}, \dots, n\}$ , in scenarios with rates  $\beta \in \{1.0, 2.5, 5.0\}$  and  $\beta_I \in \{0.625, 1.25, 2.5\}$ .

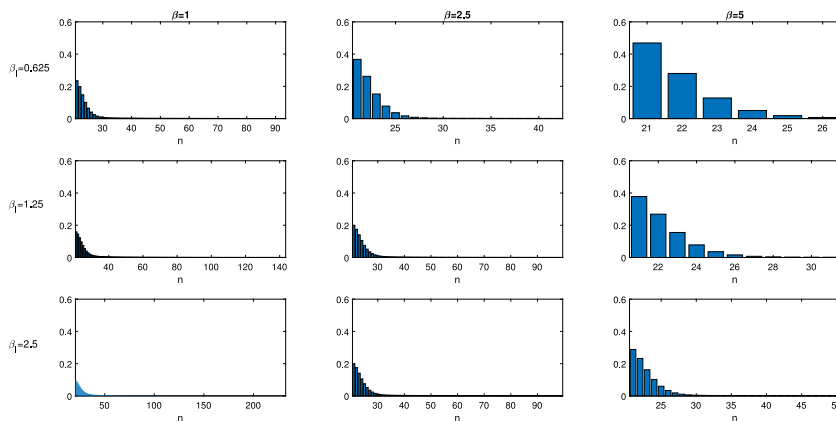


Fig. 2. The mass function  $\{P(N_{\max} = n | (N(0), I(0)) = (21, 1)) : n \in \{21, \dots, n_{0.95}\}\}$  in scenarios with rates  $\beta \in \{1.0, 2.5, 5.0\}$  and  $\beta_I \in \{0.625, 1.25, 2.5\}$ .

Similarly to Ref. [57, Section 4], birth and death rates in our experiments are assumed to be related to each other through the equalities  $\delta_S = \beta_S$ ,  $\beta_S = 2\beta_I$  and  $\delta_I = 2\delta_S$ . Hence, the compartment of susceptible individuals is expected to be demographically stable, offspring from infectious parents are expected to be born less frequently than those from susceptible parents, and susceptible individuals have a longer life expectancy than infectious ones. With the selections  $p = 0.2$  and  $\gamma^{-1} = 1.0$ , it is supposed that 20% of newborn offspring of infectious parents will be susceptible and that an infectious individual will be recovered, on average, in 1 unit of time. Contact rates  $\beta \in \{1.0, 2.5, 5.0\}$  and birth rates  $\beta_I \in \{0.625, 1.25, 2.5\}$  yield nine scenarios in Figs. 1–3, for which the extinction of the population is certain and occurs in a finite mean time.<sup>2</sup> For convenience, intervals on the axes  $ox$  and  $oy$  in Fig. 1, and on the  $ox$  axis in Figs. 2–3 are chosen to concentrate at least 0.95 of the probability, whence the integer  $n_{0.95}$  in these figures satisfies  $F_{\max}(n_{0.95} - 1; 21, 1) < 0.95 \leq F_{\max}(n_{0.95}; 21, 1)$ .

It is seen that, in our experiments, the transmission dynamics of the pathogen are limited and the duration of the outbreak is expected to be short. This is mainly related to the values of  $P(N_{\max} = n, I(\tau_{\max}) \leq i | (N(0), I(0)) = (21, 1))$  in Fig. 1, especially because the most significant ones correspond to the integers  $n = 21$  and  $i = 1$  (i.e., the initial population size and the initially infectious individual), regardless of the choice of  $(\beta, \beta_I)$ . In Fig. 2, it is seen that the mass function of  $N_{\max}$  is unimodal and is mainly concentrated on the initial population size and the closest values to it. Note that, for a fixed value of  $\beta_I$ , shorter life cycles are expected to occur when isolation measures between compartments are relaxed (i.e., increasing values of  $\beta$ ), which is explained by seeing that infectious individuals die at a higher rate than susceptible ones. Although a higher birth rate from infectious parents leads to an increased number of infectious newborns, it will also generate a greater number of susceptible newborns, who are born in a lower proportion than infectious ones but with longer life spans and a higher birth rate. Therefore, heavier-tailed distributions of  $N_{\max}$  – which would predict longer durations of a life cycle – are observed when isolation measures between compartments become more significant and, in particular, the birth rate from infectious parents increases.

<sup>2</sup> Since process  $\mathcal{X}$  is regular, the inequality  $\min\{\beta_S, \beta_I\} < \delta_S + \delta_I$  can be seen as a sufficient condition for  $\mathcal{X}$  to be absorbed almost surely, provided that  $\min\{\delta_S, \delta_I\} = \min\{\beta_S, \beta_I\}$ ; see Ref. [53, Theorems 1 and 3]. Indeed, the absorption of  $\mathcal{X}$  occurs in a finite mean time from any initial non-absorbing state.

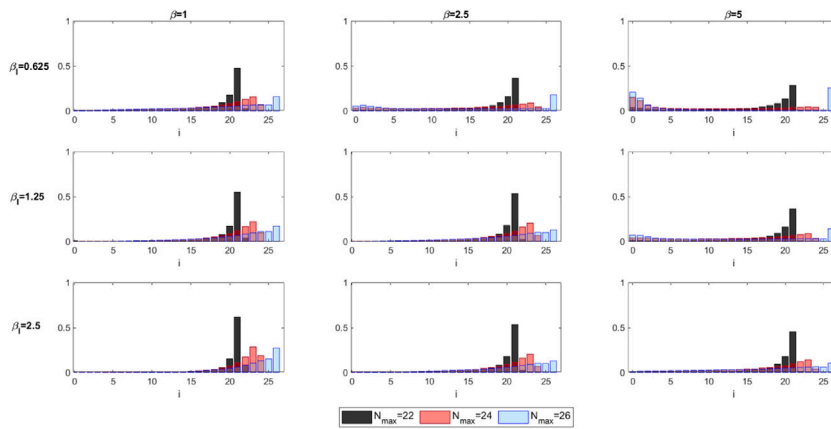


Fig. 3. Conditional probabilities  $\{P(I(\tau_{\max}) = i | (N(0), I(0)) = (21, 1), N_{\max} = n) : i \in \{0, \dots, n\}\}$ , for selected numbers  $n \in \{22, 24, 26\}$ , in scenarios with rates  $\beta \in \{1.0, 2.5, 5.0\}$  and  $\beta_j \in \{0.625, 1.25, 2.5\}$ .

Table 2  
Transitions from state  $(i, s)$  of process  $\mathcal{X}$  and related events in the SIR model.

Event type	Event function	Event condition	Rate
New infection	$(i + 1, s - 1)$	$i \neq 0, s \neq 0$	$N^{-1} \beta i s$
Recovery of an infectious individual	$(i - 1, s)$	$i \neq 0$	$\gamma i$

A detailed description of the compartment of infectious individuals at the time of maximum population size is displayed in Fig. 3 in terms of the conditional probabilities  $P(I(\tau_{\max}) = i | (N(0), I(0)) = (21, 1), N_{\max} = n)$ , for integers  $i \in \{0, \dots, n\}$  and selections  $n \in \{22, 24, 26\}$ . The unimodal conditional distribution of  $I(\tau_{\max})$ , provided that  $N_{\max} = n$ , reflects how a more accurate estimation of the maximum population size  $N_{\max}$  will allow a better prediction of the incidence of the disease in terms of  $I(\tau_{\max})$ , as observed, for example, in the case  $n = 22$  and any of our scenarios. For some values of  $(\beta, \beta_j)$ , the selection  $n = 26$  leads to a bimodal conditional distribution of  $I(\tau_{\max})$ , with a first peak at a small integer  $i$  and a second one at  $i = n$ . This behavior is linked to the inherent stochasticity of the Markov chain model, for which sample paths of process  $\mathcal{X}$  on  $\{N_{\max} = n\}$  in the case  $n = 26$  and some pairs  $(\beta, \beta_j)$  may make it significantly plausible that the outbreak has a short duration in time and a minimal incidence in terms of the number  $I(\tau_{\max})$  of infectious individuals, and at the same time a major outbreak with values of  $I(\tau_{\max})$  comparable to the maximum population size can also be observed.

### 3.2. The SIR epidemic model

In the SIR model (Kermack and McKendrick [61]; Neuts and Li [62]), the process  $\mathcal{X} = \{(I(t), S(t)) : t \geq 0\}$  records the number  $I(t)$  of infectious individuals and the number  $S(t)$  of susceptible individuals, whereas the number  $R(t)$  of removed or recovered individuals is given by  $R(t) = N - I(t) - S(t)$ , where  $N$  is the (constant) population size. The process  $\mathcal{X}$  is then defined on the state space

$$S = \{(i, s) : i \in \{0, \dots, N_I + N_S - s\}, s \in \{0, \dots, N_S\}\},$$

for initial numbers of  $N_I$  infectious individuals and of  $N_S$  susceptible individuals – consequently,  $N = N_I + N_S$  with  $R(0) = 0$  –, and its  $q$ -matrix  $Q$  has entries

$$q_{(i,s),(i',s')} = \begin{cases} N^{-1} \beta i s, & \text{if } i' = i + 1 \text{ and } s' = s - 1, \\ -(N^{-1} \beta i s + \gamma i), & \text{if } i' = i \text{ and } s' = s, \\ \gamma i, & \text{if } i' = i - 1 \text{ and } s' = s, \\ 0, & \text{otherwise,} \end{cases}$$

for states  $(i, s), (i', s') \in S$ , where the parameters  $\beta$  and  $\gamma$  are the contact and recovery rates, respectively, and are assumed to be strictly positive; see Table 2 and Appendix B.

The SIR model is possibly the simplest compartmental model and is commonly taken as the basic way to define more general ones (Yang et al. [63]), both in the deterministic (Krögen and Schlickeiser [64]) and stochastic (Allen [59]) frameworks. In the above formulation, it is seen as a reasonably predictive model for pathogens that are transmitted by person-to-person contact in closed populations, and whose recovery confers lifelong resistance, such as measles and mumps.

In Figs. 4–5 and Table 3, we consider a small community consisting initially of one infectious and twenty-four susceptible individuals, i.e.,  $I(0) = 1$  and  $S(0) = 24$ . An infectious individual is assumed to recover, on average, in  $\gamma^{-1} = 1$  unit of time,

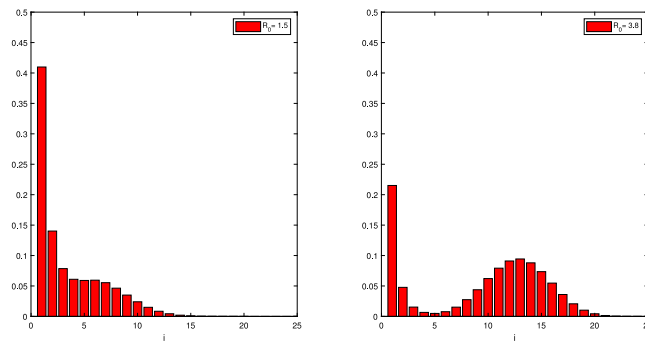


Fig. 4. The mass function  $\{P(I_{\max} = i | (I(0), S(0)) = (1, 24)) : i \in \{1, \dots, 25\}\}$  for populations with value  $\mathcal{R}_0 = 1.5$  (left) and 3.8 (right).

Table 3

Expected values  $E[\tau_{\max} | (I(0), S(0)) = (1, 24), I_{\max} = i]$  for selected numbers  $i \in \{3, 8, 13, 18, 23\}$ , and populations with  $\mathcal{R}_0 = 1.5$  and 3.8.

$\mathcal{R}_0$	$i$	$E[\tau_{\max}   (I(0), S(0)) = (1, 24), I_{\max} = i]$
1.5	3	0.84726
	8	2.44037
	13	2.27654
	18	2.07524
	23	1.91487
3.8	3	0.41006
	8	1.73395
	13	1.45888
	18	1.26555
	23	1.13595

and the contact rate  $\beta$  is determined from the basic reproductive number  $\mathcal{R}_0 = \gamma^{-1}\beta$ . The choice of  $\mathcal{R}_0$  values corresponds to the estimated mean values in Ref. [65, Table 1] for the 1918 Spanish influenza. Fig. 4 illustrates the transmission dynamics in terms of the mass function<sup>3</sup>  $\{P(I_{\max} = i | (I(0), S(0)) = (1, 24)) : i \in \{1, \dots, 25\}\}$ , which is plotted for values  $\mathcal{R}_0 = 1.5$  (spring wave) and 3.8 (fall wave), and Table 3 gives the expected values  $E[\tau_{\max} | (I(0), S(0)) = (1, 24), I_{\max} = i]$ , for selected numbers  $i \in \{3, 8, 13, 18, 23\}$ . We notice that, not surprisingly, the less interaction between individuals, the less spread of the pathogen will be observed. In particular, the maximum number  $I_{\max}$  of infectious individuals in the spring wave has a unimodal distribution with a clear peak at point  $i = 1$ , meaning that the event  $\{I_{\max} = 1\}$  occurs more frequently than the others. As a result, the disease is more likely to be eradicated as soon as the initially infectious individual recovers than to spread to the subpopulation of susceptible individuals. However, the latter may also occur with significant probability. A more relevant interaction between individuals in the fall wave yields a bimodal distribution of  $I_{\max}$ , with a first peak at point  $i = 1$  and a second one at  $i = 13$ . This illustrates how increasing values of  $\mathcal{R}_0$  may cause major outbreaks – i.e., outbreaks with a significant number of infectious cases – at the same time that the disease could also disappear at an early stage of the epidemic. This fact highlights the need to analyze the maximum number  $I_{\max}$  of infectious individuals as a random variable (i.e., in terms of its mass function) rather than its expected value or its deterministic counterpart.

To further investigate how the values of  $I_{\max}$  affect the subpopulation of recovered individuals at the time  $\tau_{\max}$  of greatest incidence of the pathogen, a detailed description of  $R(\tau_{\max})$  is displayed in Fig. 5, in terms of the mass function  $\{P(R(\tau_{\max}) = r | (I(0), S(0)) = (1, 24), I_{\max} = i) : r \in \{0, \dots, 25 - i\}\}$ , for integers  $i \in \{3, 8, 13, 18, 23\}$ . It is important to note that the most likely event  $\{I_{\max} = i\}$  in Fig. 4 is linked to the choice  $i = 1$ , irrespectively of  $\mathcal{R}_0$ , which leads to  $R(\tau_{\max}) = 0$  almost surely (since  $\tau_{\max} = 0$ ), provided that  $(I(0), S(0)) = (1, 24)$  and  $I_{\max} = 1$ . The particular selection  $i^* = 13$  is related to the second peak of the bimodal distribution of  $I_{\max}$  in the fall wave. Therefore, the conditional probabilities  $\{P(R(\tau_{\max}) = r | (I(0), S(0)) = (1, 24), I_{\max} = i^*) : r \in \{0, \dots, 25 - i^*\}\}$  must be seen as an accurate description of the variability of  $R(\tau_{\max})$  for a major outbreak in the case  $\mathcal{R}_0 = 3.8$ , whereas they should not be usually appropriate when  $\mathcal{R}_0 = 1.5$  as  $i^* = 13$  leads to an event  $\{I_{\max} = i^*\}$  that it will hardly ever occur. Our experiments always found that the conditional distribution of  $R(\tau_{\max})$  is unimodal, regardless of the selected integer  $i$ .

#### 4. Conclusions

For a regular LD-QBD process  $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$  with countably many states, the probability law of the random vector  $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$  has been characterized under the assumption that the first passage from any initial state  $(i_0, j_0) \in S \setminus I(0)$  to level

<sup>3</sup> Since  $I(t)$  and  $S(t)$  specify the level and the phase variables of  $\mathcal{X}$ , respectively,  $I_{\max} = I(\tau_{\max})$  represents the maximum size of the compartment of infectious individuals during an outbreak.

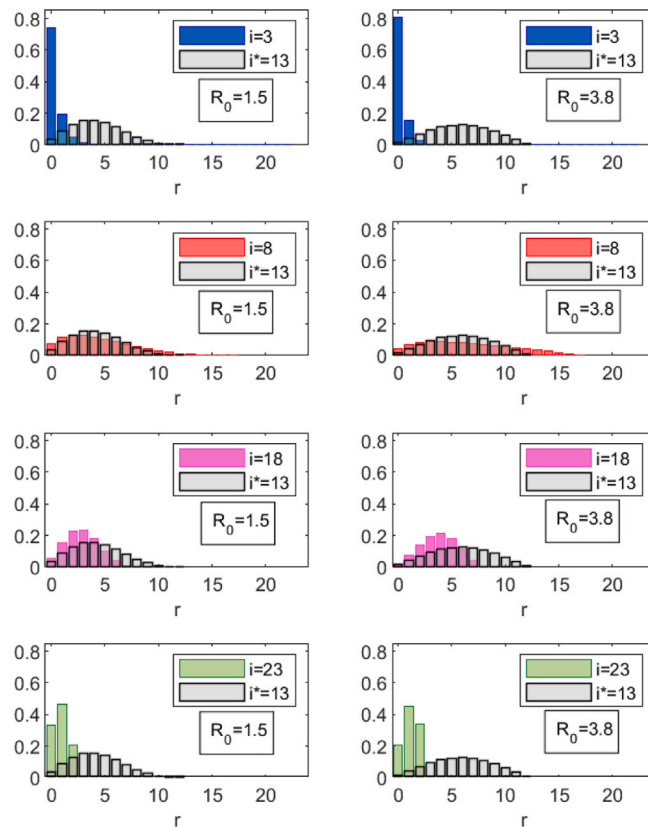


Fig. 5. Conditional probabilities  $\{P(R(\tau_{\max}) = r | I(0), S(0) = (1, 24), I_{\max} = i) : r \in \{0, \dots, 25 - i\}\}$ , for selected numbers  $i \in \{3, 8, 13, 18, 23\}$ , and populations with value  $R_0 = 1.5$  (left) and 3.8 (right).

$l(0)$  is certain. For this purpose, efficient algorithms for computing the marginal distribution of  $I_{\max}$ , and restricted Laplace–Stieltjes transforms of  $\tau_{\max}$  on the sample paths of  $\mathcal{X}$  such that  $\{I_{\max} = i, J(\tau_{\max}) = j\}$ , have been developed as an extension of the well-known block-Gaussian elimination technique, whence the algorithmic complexity of the solution in Sections 2.2 and 2.3 is similar to that in Refs. [27, Algorithm A] and [28, Algorithms 1A–3A] in the finite case. To be concrete, the most intensive effort in Algorithms A–B lies in the inversion of matrices with dimension  $1 + M_{i'}$ , which implies that the computational complexities of Algorithms A and B are  $O\left(\sum_{i'=1}^{i-1} (1 + M_{i'})^3\right)$  and  $O\left(\sum_{i'=1}^i (1 + M_{i'})^3\right)$ , respectively, for a fixed state  $(i, j) \in \cup_{k=i_0+1}^{\infty} l(k)$  and the initial state  $(i_0, j_0)$ . In Algorithm C, the effort needed for each integer  $i$  is  $O((1 + M_i)^3)$  and, consequently, the computational complexity of Algorithm C is  $O\left(\sum_{i=1}^{i^*} (1 + M_i)^3\right)$ , where  $i^* \in \mathbb{N}$  is the smallest integer satisfying  $F_{\max}(i^*; i_0, j_0) \geq 1 - \epsilon$ ; in the finite case, we may select  $i^* = N - 1$ .

Algorithms A–C involve matrix inversion, for which the key assumption is that process  $\mathcal{X}$  has a finite number of phases at each level, regardless of the number (finite or denumerable) of levels. In the framework of LD-QBD processes with infinitely many possible values of the phase variable, the study of  $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$  should require an analytical treatment and related numerical procedures that will be substantially different from those used in this paper; in particular, the underlying analysis could benefit from the approach described by Motyer and Taylor [66] for the level-independent case.

The computational approach can be thought of as an alternative one to that of Mandjes and Taylor [48], and Javier and Fralix [49] for the running maximum  $\bar{I}(T)$  attained by process  $\mathcal{X}$  during a time interval  $[0, T]$  when  $T$  is either independent and exponentially distributed time, or a fixed time. Here, time  $T$  is described as the first time that process  $\mathcal{X}$  visits states in level  $l(0)$ , so  $T$  is closely linked to the dynamics of  $\mathcal{X}$  and, for practical use, amounts to the duration of a life cycle in a population, of an outbreak in epidemics, and of a busy period in a queueing model. In the context of epidemics, an aspect that deserves further exploration is how phase  $J(\tau_{\max})$  and its probability law – when level  $l(t)$  in process  $\mathcal{X}$  is defined as the number of infectious individuals – could be used to characterize a stochastic version of the herd immunity threshold (see, e.g., Britton et al. [67]), for which it is crucial to notice that no more endemic transmissions are expected to occur after time  $\tau_{\max}$ .

**Data availability**

No data was used for the research described in the article. Details on numerical experiments in Section 3 will be made available from the corresponding author upon reasonable request.





- [37] T. Phung-Duc, Asymptotic analysis for Markovian queues with two types of nonpersistent retrial customers, *Appl. Math. Comput.* 265 (2015) 768–784, <http://dx.doi.org/10.1016/j.amc.2015.05.133>.
- [38] S.S. Sanga, G.S. Charan, Fuzzy modeling and cost optimization for machine repair problem with retrial under admission control F-policy and feedback, *Math. Comput. Simulation* 211 (2023) 214–240, <http://dx.doi.org/10.1016/j.matcom.2023.03.036>.
- [39] V. Saravanan, V. Poongothai, P. Godhandaraman, Performance analysis of a multi server retrial queueing system with unreliable server, discouragement and vacation model, *Math. Comput. Simulation* 214 (2023) 204–226, <http://dx.doi.org/10.1016/j.matcom.2023.07.008>.
- [40] X. Chai, T. Jiang, L. Li, W. Xu, L. Lin, On a many-to-many matched queueing system with flexible matching mechanism and impatient customers, *J. Comput. Appl. Math.* 416 (2022) 114573, <http://dx.doi.org/10.1016/j.cam.2022.114573>.
- [41] H. Baumann, W. Sandmann, Structured modeling and analysis of stochastic epidemics with immigration and demographic effects, *PLoS ONE* 11 (3) (2016) e0152144, <http://dx.doi.org/10.1371/journal.pone.0152144>.
- [42] C. Lefèvre, M. Simon, SIR-type epidemic models as block-structured Markov processes, *Methodol. Comput. Appl. Probab.* 22 (2020) 433–453, <http://dx.doi.org/10.1007/s11009-019-09710-y>.
- [43] F.A.C.C. Chalub, A. Gómez-Corral, M. López-García, F. Palacios-Rodríguez, A Markov chain model to investigate the spread of antibiotic-resistant bacteria in hospitals, *Stud. Appl. Math.* 151 (2023) 1498–1524, <http://dx.doi.org/10.1111/sapm.12637>.
- [44] T. Dayar, W. Sandmann, D. Spieler, V. Wolf, Infinite level-dependent QBD processes and matrix-analytic solutions for stochastic chemical kinetics, *Adv. in Appl. Probab.* 43 (2011) 1005–1026, <http://dx.doi.org/10.1239/aap/1324045696>.
- [45] J. Diao, T.L. Stark, D.A. Liberles, M.M. O'Reilly, B.R. Holland, Level-dependent QBD models for the evolution of a family of gene duplicates, *Stoch Models* 36 (2) (2020) 285–311, <http://dx.doi.org/10.1080/15326349.2019.1680296>.
- [46] A. Gómez-Corral, M.J. Lopez-Herrero, D. Taïpe, A Markovian epidemic model in a resource-limited environment, *Appl. Math. Comput.* 458 (2023) 128252, <http://dx.doi.org/10.1016/j.amc.2023.128252>.
- [47] L.M. Santana, S. Ganesan, G. Bhanot, A quasi birth-and-death model for tumor recurrence, *J. Theoret. Biol.* 480 (2019) 175–191, <http://dx.doi.org/10.1016/j.jtbi.2019.07.017>.
- [48] M. Mandjes, P. Taylor, The running maximum of a level-dependent quasi-birth-death process, *Probab. Engrg. Inform. Sci.* 30 (2) (2016) 212–223, <http://dx.doi.org/10.1017/S026996481500039X>.
- [49] K. Javier, B. Fralix, On the study of the running maximum and minimum level of level-dependent quasi-birth-death processes and related models, *J. Appl. Probab.* 60 (1) (2023) 14–29, <http://dx.doi.org/10.1017/jpr.2022.22>.
- [50] A. Gómez-Corral, M. López García, Maximum queue lengths during a fixed time interval in the  $M/M/c$  retrial queue, *Appl. Math. Comput.* 235 (2014) 124–136, <http://dx.doi.org/10.1016/j.amc.2014.02.074>.
- [51] J. Amador, D. Armesto, A. Gómez-Corral, Extreme values in SIR epidemic models with two strains and cross-immunity, *Math. Biosci. Eng.* 16 (2019) 1992–2022, <http://dx.doi.org/10.3934/mbe.2019098>.
- [52] V.G. Kulkarni, *Modeling and Analysis of Stochastic Systems*, third ed., Chapman & Hall/CRC Press, Boca Raton, 2017.
- [53] A. Gómez-Corral, J. Langwade, M. López-García, C. Molina-París, Sufficient conditions for regularity, positive recurrence and absorption in level-dependent QBD processes and related block-structured markov chains, *Math Meth Appl Sci.* 46 (6) (2023) 6756–6766, <http://dx.doi.org/10.1002/mma.8938>.
- [54] W.J. Stewart, *Introduction To the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, 1994.
- [55] J. Abate, W. Whitt, Numerical inversion of Laplace transforms of probability distributions, *ORSA J. Comput.* 7 (1995) 36–43, <http://dx.doi.org/10.1287/ijoc.7.1.36>.
- [56] A.M. Cohen, *Numerical Methods for Laplace Transform Inversion*, Springer, New York, 2007.
- [57] A. Gómez-Corral, F. Palacios-Rodríguez, M.T. Rodríguez-Bernal, On the exact reproduction number in SIS epidemic models with vertical transmission, *Comput. Appl. Math.* 42 (2023) 291, <http://dx.doi.org/10.1007/s40314-023-02424-5>.
- [58] S. Busenberg, K. Cooke, *Vertically Transmitted Diseases*, Springer, Berlin, 1992.
- [59] L.J.S. Allen, *An Introduction To Stochastic Processes with Applications To Biology*, second ed., Chapman and Hall/CRC, New York, 2010.
- [60] N. Neu, J. Duchon, P. Zachariah, TORCH infections, *Clin Perinatol* 42 (2015) 77–103, <http://dx.doi.org/10.1016/j.clp.2014.11.001>.
- [61] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 115 (1927) 700–721, <http://dx.doi.org/10.1098/rspa.1927.0118>.
- [62] M.F. Neuts, J.M. Li, An algorithmic study of S-I-R stochastic epidemic models, in: C.C. Heyde, Y.V. Prohorov, R. Pyke, S.T. Rachev (Eds.), *Athens Conference on Applied Probability and Time Series Analysis*, in: *Lecture Notes in Statistics*, vol. 114, Springer, New York, 1996, pp. 295–306, [http://dx.doi.org/10.1007/978-1-4612-0749-8\\_21](http://dx.doi.org/10.1007/978-1-4612-0749-8_21).
- [63] W. Yang, D. Zhang, L. Peng, C. Zhuge, L. Hong, Rational evaluation of various epidemic models based on the COVID-19 data of China, *Epidemics* 37 (2021) 100501, <http://dx.doi.org/10.1016/j.epidem.2021.100501>.
- [64] M. Kröger, R. Schlickeiser, Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor, *J. Phys. A* 53 (2020) 505601, <http://dx.doi.org/10.1088/1751-8121/abc65d>.
- [65] P. van den Driessche, Reproduction numbers of infectious disease models, *Infect. Dis. Model.* 2 (3) (2017) 288–303, <http://dx.doi.org/10.1016/j.idm.2017.06.002>.
- [66] A.J. Motyer, P.G. Taylor, Decay rates for quasi-birth-and-death processes with countably many phases and tridiagonal block generators, *Adv. in Appl. Probab.* 38 (2006) 522–544, <http://dx.doi.org/10.1239/aap/1151337083>.
- [67] T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2, *Science* 369 (2020) 846–849, <http://dx.doi.org/10.1126/science.abc6810>.