Article:

Gomez-Corral, A., Lopez Garcia, M., Palacios-Rodriguez, F. et al. (Accepted: 2026) Hitting probabilities and hitting times in time-inhomogeneous level-dependent quasi-birth-death processes. Methodology and Computing in Applied Probability. ISSN: 1387-5841 (In Press)

Hitting probabilities and hitting times in time-inhomogeneous level-dependent quasi-birth-death processes

A. Gómez-Corral[a,*], M. López-García[b], F. Palacios-Rodríguez[c], D. Taipe[d]

[a]*Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain*
[b]*Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom*
[c]*Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, Calle Tarfia s/n, 41012 Seville, Spain*
[d]*Departamento de Estadística y Ciencia de los Datos, Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, Avenida Puerta de Hierro s/n, 28040 Madrid, Spain*

Abstract

For a level-dependent quasi-birth-death process $\mathcal{X}$ with time-varying transition rates, we propose a computational approach to compute the probability law of first-passage times to higher levels, as well as related hitting probabilities, at a fixed horizon $T < \infty$. The approach involves approximating the first-passage time distributions of $\mathcal{X}$ at time T by their counterparts in a suitably defined process with piecewise-constant transition rates at an independent, Erlang-distributed horizon with S stages and mean T. The solution is exemplified by numerical experiments in the context of epidemics and queueing models.

Keywords: hitting times, hitting probabilities, quasi-birth-death process, time-varying

2010 MSC: 60J28, 92B05

*Corresponding author

Email addresses: agcorral@ucm.es (A. Gómez-Corral), m.lopezgarcia@leeds.ac.uk (M. López-García), fpalacios2@us.es (F. Palacios-Rodríguez), dtaipe@ucm.es (D. Taipe)

1. Introduction

This article focuses on quasi-birth-death (QBD) processes with time-varying transition rates. These processes arise in the performance evaluation of Markovian systems that feature non-stationary changes in their parameters, such as service systems with cyclic demands (Green et al. [23]; Margolius [36, 39, 41]), call centers with retrials (Grier et al. [24]; Mandelbaum et al. [33]), air and road traffic systems (Bookbinder [4]; Green and Kolesar [22]), computer and communication networks (Kuraya et al. [32]) and epidemics with term-time forcing (Britton and Lindholm [7]), among others. An excellent survey on queueing models with time-varying parameters is the article of Schwarz et al. [48]; for a related work, see Tripathi and Duda [52]. In the setting of epidemic models, a review of results on the influence of time heterogeneities, including seasonality, on the transmission pattern of infectious diseases can be found in Ref. [8].

Some of the aforementioned systems are based on $M_t/M_t/1$ and $M_t/PH_t/1$ queues, and $M_t/M_t/c$ and $PH_t/M_t/c$ queues with a fixed finite number of servers (Margolius [34]; Ong and Taaffe [43]; Taaffe and Ong [51]; Zhang and Coyle [60]), whence they can be readily encompassed within the class of level-independent QBD processes with time-varying transition rates. For these systems, a number of results on the transient and periodic families of the underlying asymptotic distributions can be then derived from the more general solution of Margolius [37, 38, 40]. On the contrary, $M_t/M_t/c(t)$ and $PH_t/M_t/c(t)$ queues with a time-varying number of servers (Green and Kolesar [23]; Margolius [35, 36]), $M_t/PH/\infty$ queues (Dong and Whitt [14]; Green and Kolesar [22]), time-dependent Erlang loss models and multi-server queues with retrials (Grier et al. [24]; Mandelbaum et al. [33]), and Markovian epidemic models with seasonal fluctuations (Britton and Lindholm [7]; Buonomo et al. [8]), among others, require a level-dependent (LD) structure in their QBD formulation, for which a general-purpose solution has not been formally derived.

Analytical results for the transition function of systems with an underlying time-inhomogeneous LD-QBD process can be found only under very particular distributional assumptions, as is the case with the linear rate birth and death process (Kendall [30]), the Weiss' carrier-borne process with time-varying transition rates (Dietz [12, Section 5]), certain finite birth processes arising from epidemic models (Giorno and Nobile [20]), a special class of birth and death processes with time-varying transition rates (Giorno and Nobile [18, Section 4]), and the non-homogeneous Prendiville process (Zheng [61]). We also refer to the articles by Satin et al. [46], and Zeifman et al. [59], and the references cited therein, for truncation-based approximations and related bounds in time-inhomogeneous birth and death processes. For classical results in the time-homogeneous setting, see Refs. [5, 6].

The aim of this article is to approximate the first-passage time distributions of a LD-QBD process with time-varying transition rates at a fixed finite horizon T by the corresponding ones of a suitably defined process at an independent random time Y with Erlang distribution of S stages and mean T. The Erlangization method, also referred to as Canadization (Asmussen [3]), is a well-known approach for modeling and analyzing stochastic descriptors in insurance risk models (Asmussen et al. [2]), American option values of Black-Scholes models (Carr [10]), fluid queues applied to a forestry problem (Stanford et al. [50]), Markov modulated fluid flow models (Ramaswami et al. [44]) and stochastic control problems (Yoshioka and Tsujimura [57]; Yoshioka et al. [58]). In He et al. [25], and Wu and He [56], continuous phase-type (PH) distributions are constructed as Erlangian approximations for a finite discrete probability distribution and a discrete PH distribution. The method exploits the property that $Y \Rightarrow T$ in distribution as $S \to \infty$; and, significantly, it concerns the Widder's Laplace transform inversion formula (Jagerman [28, Theorem 1]; Ramaswami et al. [44, Section 1]) for a continuous and bounded non-negative function $f(\cdot)$ on $[0, \infty)$. To be concrete, this inversion formula has the form

$$f(T) \quad = \quad \lim_{S\to\infty} \int_0^{\infty} g_{S,T}(u) f(u) du,$$

where $g_{S,T}(u)$ is the density function of Y and the convergence is uniform on compact intervals along which $f(T)$ is continuous; for further properties, see Sections III and IV in Jagerman [28], and Ref. [29]. For any time-dependent functional $f(\cdot)$ of a process $\mathcal{X}$, the integral $\int_0^\infty g_{S,T}(u)f(u)du$ can then be used as an approximation to $f(T)$. Since expressions for the integral $\int_0^\infty g_{S,T}(u)f(u)du$ are almost never known in the time-inhomogeneous framework, we shall replace here the dynamics of $\mathcal{X}$ until the expiration of the independent, Erlang-distributed time Y by those of a more tractable time-inhomogeneous process that, for a sufficiently large S, retains essential average characteristics of the original process $\mathcal{X}$. In the univariate setting of a time-inhomogeneous birth and death process, the problem is addressed by Giorno and Nobile [19] in terms of Volterra integral equations of second kind and the underlying transition function.

The remainder of the article is organized as follows. Section 2 introduces the class of LD-QBD processes with time-varying transition rates, provides background information on numerical evaluation of their transition function, and constructs, using Erlangization, time-inhomogeneous processes that preserve average dynamics of the original process, with the key feature of being analytically solvable under time-homogeneous environmental conditions. Sections 3 and 4 derive algorithmic procedures for first-passage times to higher levels in the resulting Erlangian processes. In Section 5, the results of Sections 2-4 are applied to two examples in the context of queueing and epidemic models with sinusoidal rates. Finally, Section 6 presents some concluding remarks on the methodological and computational aspects of the article.

2. Statement of the problem

Let $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$ be a regular time-inhomogeneous LD-QBD process —with $I(t)$ and $J(t)$ as the level and phase variables—, which takes values in a countable set $\mathcal{S} = \cup_{i=0}^{I^*}\mathcal{L}(i)$ with $I^* \in \mathbb{N} \cup \{\infty\}$, where the subsets $\mathcal{L}(i)$ (also termed *levels*) are specified by $\{(i,j) : j \in \{0, ..., M_i\}\}$ with $M_i \in \mathbb{N}_0$ and $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$. The infinitesimal dynamics of process $\mathcal{X}$ are governed by the

q-matrix

$$Q(t) \quad = \quad \begin{pmatrix} Q_{0,0}(t) & Q_{0,1}(t) & & & \\ Q_{1,0}(t) & Q_{1,1}(t) & Q_{1,2}(t) & & \\ & Q_{2,1}(t) & Q_{2,2}(t) & Q_{2,3}(t) & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1}$$

where $Q_{i,i'}(t)$ is a matrix of dimension $(1 + M_i) \times (1 + M_{i'})$ with elements $q_{(i,j),(i',j')}(t)$, for states (i,j) with phases $j \in \{0, ..., M_i\}$, and (i', j') with $i' \in \{\max\{0, i-1\}, i, \min\{i+1, I^*\}\}$ and $j' \in \{0, ..., M_{i'}\}$. Specifically, time-varying transition rates $q_{(i,j),(i',j')}(t)$ correspond to jumps of $\mathcal{X}$ from state (i,j) into $(i',j') \neq (i,j)$ at time t, and are right continuous in t; in the case $(i',j') = (i,j)$, it holds that

$$q_{(i,j),(i,j)}(t) \quad = \quad - \sum_{(i',j') \in \mathcal{S} \setminus \{(i,j)\}} q_{(i,j),(i',j')}(t),$$

for any state $(i,j) \in \mathcal{S}$ and time $t \geq 0$, corresponding to a conservative q-matrix.

Let $\eta_K = \inf\{t > 0 : I(t) = K\}$ be the first-passage time to $\mathcal{L}(K)$, for a predetermined finite integer $K \in \mathbb{N}$ with $K \leq I^*$ in the case $I^* < \infty$. For initial states $(i,j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$, the probability law of η_K at a fixed finite horizon T may be determined from the conditional probabilities

$$F(T|i,j;K,j') \quad = \quad P\big(\eta_K \leq T, J(\eta_K) = j' \mid (I(0), J(0)) = (i,j)\big), \tag{2}$$

for $j' \in \{0, ..., M_K\}$. They are related to the finite-state QBD process $\mathcal{X}(K) = \{(I(t;K), J(t;K)) : t \geq 0\}$ with state space

$$\mathcal{S}_{\mathcal{X}(K)} \quad = \quad \bigcup_{k=0}^{K-1} \mathcal{L}(k) \cup \mathcal{L}(K)$$

and q-matrix

$$Q_{\mathcal{X}(K)}(t) \quad = \quad \begin{pmatrix} Q(t;K-1) & Q^*(t;K-1) \\ 0_{(1+M_K) \times \sum_{k=0}^{K-1}(1+M_k)} & 0_{(1+M_K)\times(1+M_K)} \end{pmatrix},$$

where $0_{r \times c}$ is the null matrix of dimension $r \times c$, and $Q(t;K-1)$ and $Q^*(t;K-1)$ are sub-matrices that record time-varying transition rates for jumps of $\mathcal{X}$

from states in $\cup_{k=0}^{K-1}\mathcal{L}(k)$ to subsets $\cup_{k=0}^{K-1}\mathcal{L}(k)$ and $\mathcal{L}(K)$, respectively. In our arguments, states in subset $\cup_{k=K+1}^{\infty}\mathcal{L}(k)$ are discarded, since the interest is in the first passage of $\mathcal{X}$ to $\mathcal{L}(K)$; and states in $\mathcal{L}(K)$ are considered to be absorbing in process $\mathcal{X}(K)$, which allows one to make explicit the phase of $\mathcal{X}$ at its first visit to $\mathcal{L}(K)$. The partitioning of $\mathcal{S}_{\mathcal{X}(K)}$ into subsets $\cup_{k=0}^{K-1}\mathcal{L}(k)$ and $\mathcal{L}(K)$ yields the following structured form for the transition function $P_{\mathcal{X}(K)}(s,t)$ of process $\mathcal{X}(K)$:

$$P_{\mathcal{X}(K)}(s,t) = \begin{pmatrix} P(s,t;K-1) & P^*(s,t;K-1) \\ 0_{(1+M_K)\times\sum_{k=0}^{K-1}(1+M_k)} & I_{1+M_K} \end{pmatrix},$$

for $0 \le s < t$, with $P_{\mathcal{X}(K)}(t,t) = I_{\sum_{k=0}^{K}(1+M_k)}$, where I_d is the identity matrix of order d. Then, for initial states $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$, and phases $j' \in \{0, ..., M_K\}$, the value of $F(T|i,j;K,j')$ in (2) is given by the $\Big((1-\delta_{i,0})\sum_{k=0}^{i-1}(1+M_k)+1 +j, 1+j'\Big)$-th element of $P^*(s,t;K-1)$ at pair $(s,t)=(0,T)$, where $\delta_{a,b}$ denotes the Kronecker's delta.

It should be pointed out that, since $\mathcal{S}_{\mathcal{X}(K)}$ is finite, the transition function of $\mathcal{X}(K)$ satisfies the Kolmogorov forward equation (see, e.g., Heidergott et al. [26, Section 2]) and, consequently, can be written as the Péano series

$$P_{\mathcal{X}(K)}(0,T) = I_{\sum_{k=0}^{K}(1+M_k)} + \sum_{n=1}^{\infty}\int_{0\le t_1\le\cdots\le t_n\le T} Q_{\mathcal{X}(K)}(t_1)\cdots Q_{\mathcal{X}(K)}(t_n)dt_1\cdots dt_n. \quad (3)$$

The *product-integral* of the q-matrix $Q_{\mathcal{X}(K)}(\cdot)$ in (3) does not, in general, yield any analytically usable expression and its numerical treatment results in highly demanding computational costs, with the exception of special distributional assumptions; for instance, finite-state processes with a q-matrix that makes discrete changes (Rindos et al. [45]). Traditionally, the most efficient numerical procedure for evaluating (3) is the uniformization technique of van Dijk [53] (see also van Dijk et al. [54, Section 5]) and its variants (Arns et al. [1]; Burak and Korytkowski [9]; Inoue [27]; van Moorsel and Wolter [55]). For a survey on product-integration and related literature, we refer to the book of Slavík [49] and the article of Gill and Johansen [17].

Below we propose a different approach that, via Erlangization (Asmussen [3]), uses instead a more analytically tractable transition function of a certain process $\mathcal{X}'(K)$. Throughout the rest of this section, we assume that level $\mathcal{L}(K)$ is accessible from any state $(i,j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$.

In defining process $\mathcal{X}'(K)$, we first replace the fixed horizon T by an independent, Erlang-distributed horizon Y with S stages and mean T, which can be decomposed into S successive stages, each taking an independent, exponential-distributed amount of time with mean $\lambda^{-1} = S^{-1}T$. This enables us to write

$$Y = \sum_{s=1}^{S} Y_s,$$

for S independent random variables with $Y_s \sim Exp(\lambda)$, for $s \in \{1, ..., S\}$. As an important consequence of this, it holds that $Y \Rightarrow T$ in distribution (also in probability and quadratic mean) as $S \to \infty$. We associate with Y a Markov chain $\mathcal{Y} = \{Y'(t) : t \geq 0\}$ with state space $\{0\} \cup \{1, ..., S\}$ and initial probability vector $e_{1+S}^T(2)$, where $e_c(b)$ represents a column vector of order c with b-th element equal to 1, and 0 otherwise, and ^{T} denotes transposition. The q-matrix of $\mathcal{Y}$ consists of elements

$$q_{s,s'} = \begin{cases} \lambda, & \text{if } s \in \{1, ..., S-1\} \text{ and } s' = s+1, \\ \lambda, & \text{if } s = S \text{ and } s' = 0, \\ -\lambda, & \text{if } s \in \{1, ..., S\}, \\ 0, & \text{otherwise,} \end{cases}$$

whence the absorption time into state 0 in $\mathcal{Y}$ is identically distributed to Y, with $Y \sim Erlang(S, \lambda)$.

For initial states $(i,j) \in \cup_{i=0}^{K-1} \mathcal{L}(k)$, we then approximate probabilities $F(T|i,j;K,j')$ in (2) corresponding to first-passage times to states (K,j'), for $j' \in \{0, ..., M_K\}$, in process $\mathcal{X}(K)$ (equivalently, $\mathcal{X}$) by their counterparts in an *augmented* time-homogeneous process $(\mathcal{Y}, \mathcal{X}'(K)) = \{(Y'(t), I'(t;K), J'(t;K)) : t \geq 0\}$ with state space

$$\mathcal{S}_{(\mathcal{Y},\mathcal{X}'(K))} = \{0\} \cup \bigcup_{s=1}^{S} \bigcup_{k=0}^{K-1} (\{s\} \times \mathcal{L}(k)) \cup \bigcup_{s=1}^{S} (\{s\} \times \mathcal{L}(K)),$$

and q-matrix in structured form

$$Q_{(\mathcal{Y},\mathcal{X}'(K))} = \begin{pmatrix} 0 & 0^T_{c_{S,K-1}} & 0^T_{S(1+M_K)} \\ w'(K-1) & \overline{Q}'(K-1) & W'(K-1) \\ 0_{S(1+M_K)} & 0_{S(1+M_K)\times c_{S,K-1}} & 0_{S(1+M_K)\times S(1+M_K)} \end{pmatrix},$$

where

$$w'(K-1) = \begin{pmatrix} 0_{c_{S-1,K-1}} \\ \lambda 1_{c_{1,K-1}} \end{pmatrix},$$

$$\overline{Q}'(K-1) = \begin{pmatrix} Q^{[1]}(K-1) - \lambda I_{c_{1,K-1}} & \lambda I_{c_{1,K-1}} & & \\ & \ddots & \ddots & \\ & & Q^{[S-1]}(K-1) - \lambda I_{c_{1,K-1}} & \lambda I_{c_{1,K-1}} \\ & & & Q^{[S]}(K-1) - \lambda I_{c_{1,K-1}} \end{pmatrix},$$

$$W'(K-1) = \text{diag}\left(Q^{*[1]}(K-1), ..., Q^{*[S]}(K-1)\right),$$

the column vector 0_c (respectively, 1_c) has c elements, all equal to 0 (respectively, 1), and $c_{a,K-1}$ is the cardinality of subset $\cup_{s=1}^{a} \cup_{k=0}^{K-1} (\{s\} \times \mathcal{L}(k))$; i.e., $c_{a,K-1} = a \sum_{k=0}^{K-1}(1 + M_k)$. In these expressions, sub-matrices $Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$, for each stage $s \in \{1, ..., S\}$, are related to a process $\mathcal{X}'(K)$ that, under appropriate conditions for the time-varying transition rates $q_{(i,j),(i',j')}(t)$ to be locally integrable, aims to preserve the average dynamics of $\mathcal{X}(K)$ within the independent, Erlang-distributed horizon Y through the transition rates from (i, j) to (i', j')

$$q^{[s]}_{(i,j),(i',j')} = \lambda \int_{\lambda^{-1}(s-1)}^{\lambda^{-1}s} q_{(i,j),(i',j')}(t)dt, \quad (4)$$

for states $(i, j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$ and $(i', j') \neq (i, j)$ with integers $i' \in \{\max\{0, i-1\}, i, \min\{i+1, K\}\}$ and phases $j' \in \{0, ..., M_{i'}\}$. This results in sub-matrices

$Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$ with the structured form

$$
\begin{aligned}
Q^{[s]}(K-1) &= \begin{pmatrix}
Q^{[s]}_{0,0} & Q^{[s]}_{0,1} & & & & \\
Q^{[s]}_{1,0} & Q^{[s]}_{1,1} & Q^{[s]}_{1,2} & & & \\
& \ddots & \ddots & \ddots & & \\
& & Q^{[s]}_{K-2,K-3} & Q^{[s]}_{K-2,K-2} & Q^{[s]}_{K-2,K-1} \\
& & & Q^{[s]}_{K-1,K-2} & Q^{[s]}_{K-1,K-1}
\end{pmatrix}, \\
Q^{*[s]}(K-1) &= \begin{pmatrix}
0_{c_{1,K-2}\times(1+M_K)} \\
Q^{[s]}_{K-1,K}
\end{pmatrix},
\end{aligned}
$$

for $s \in \{1, ..., S\}$, and the property that, in a similar manner to $\mathcal{X}(K)$, level $\mathcal{L}(K)$ for process $\mathcal{X}'(K)$ is also accessible from any state $(i,j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$.

Remark 1. *Since $\lambda^{-1} = S^{-1}T$, the transition rate $q^{[s]}_{(i,j),(i',j')}$ in (4) can be interpreted as the mean value of $q_{(i,j),(i',j')}(t)$ in the interval $\left[S^{-1}(s-1)T, S^{-1}sT\right)$, for each stage $s \in \{1, ..., S\}$.*

Remark 2. *In the spirit of the piecewise-constant approximation in Ref. [47], another choice for sub-matrices $Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$ is specified by $Q^{[s]}(K-1) = Q(\lambda^{-1}(s-1); K-1)$ and $Q^{*[s]}(K-1) = Q^{*}(\lambda^{-1}(s-1); K-1)$, for $s \in \{1, ..., S\}$, which only requires the more general assumption that time-varying transition rates $q_{(i,j),(i',j')}(t)$ are right continuous in t, for $t \in [0, T)$.*

With this, the probabilities $F(T|i,j;K,j')$ for the first-passage time η_K for each $(K,j') \in \mathcal{L}(K)$ in process $\mathcal{X}(K)$ are approximated, for a sufficiently large S, by values

$$F_S(T|i,j;K,j') \quad = \quad \sum_{s=1}^{S} f_S(T|1,i,j;s,K,j'), \tag{5}$$

where $f_S(T|1,i,j;s,K,j')$ is defined as the hitting probability

$$P((Y'(\eta'_K), J'(\eta'_K;K)) = (s,j') \mid (Y'(0), I'(0;K), J'(0;K)) = (1,i,j)),$$

for initial states $(i,j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$ and phases $j' \in \{0, ..., M_K\}$, and η'_K is the first-passage time of $\mathcal{X}'(K)$ to $\mathcal{L}(K)$. Values $F_S(T|i,j;K,j')$ in (5) are therefore based on the replacement of $\mathcal{X}(K)$ by $\mathcal{X}'(K)$, and the Widder's inversion formula

$$F(T|i,j;K,j') \quad = \quad \lim_{S\to\infty}\int_0^\infty g_{S,T}(u)F(u|i,j;K,j')du,$$

from which $F(T|i,j;K,j')$ is approximated by the value

$$\sum_{s=1}^{S} P((Y'(\eta'_K), J'(\eta'_K;K)) = (s,j') \mid (Y'(0), I'(0;K), J'(0;K)) = (1,i,j)),$$

for a sufficiently large S.

Remark 3. *Since probabilities $f_S(T|1,i,j;s,K,j')$ are related to the absorption of $(\mathcal{Y},\mathcal{X}'(K))$ into states $(s,K,j') \in \cup_{s'=1}^{S}(\{s'\}\times\mathcal{L}(K))$, the Erlangian approximations $\{F_S(T|i,j;K,j') : T \geq 0\}$ characterize the probability law of the first-passage time η'_K to $\mathcal{L}(K)$ in process $\mathcal{X}'(K)$ before the independent, Erlang-distributed horizon Y expires.*

3. First-passage times to higher levels in process $(\mathcal{Y},\mathcal{X}'(K))$

In the construction of $(\mathcal{Y},\mathcal{X}'(K))$, process $\mathcal{Y}$ controls the environmental changes in states of $\mathcal{X}'(K)$ in such a way that, for any time $0 \leq t < Y$ with $Y'(t) = s$ and $(I'(t;K), J'(t;K)) = (i,j)$, for stages $s \in \{1,...,S\}$ and states $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$, the dynamics of $\mathcal{X}'(K)$ are linked to sub-matrices $Q_{i,i'}^{[s]}$, for $i' \in \{\max\{0,i-1\}, i, i+1\}$, instead of $Q_{i,i'}(t)$. This makes the augmented process $(\mathcal{Y},\mathcal{X}'(K))$ a versatile modelling tool for our purposes, as we will show below.

In the time-homogeneous framework of $(\mathcal{Y},\mathcal{X}'(K))$, the probability law of the first-passage time η'_K can be characterized from *restricted* Laplace-Stieltjes transforms $\varphi_{(s,i,j)}(\theta;s',K,j')$, defined as

$$E\left[e^{-\theta\eta'_K}\bar{1}_{\{(Y'(\eta'_K),J'(\eta'_K;K))=(s',j')\}} \;\middle|\; (Y'(0), I'(0;K), J'(0;K)) = (s,i,j)\right],$$

for $\Re(\theta) \geq 0$, initial states (s, i, j) with $s \in \{1, ..., S\}$ and $(i, j) \in \cup_{k=0}^{K-1} \mathcal{L}(k)$, stages $s' \in \{s, ..., S\}$ and phases $j' \in \{0, ..., M_K\}$, where $\bar{1}_A$ specifies the indicator function of subset A; note that, as a result, $f_S(T|1, i, j; s, K, j')$ in (5) is the value of $\varphi_{(1,i,j)}(\theta; s, K, j')$ at point $\theta = 0$. To be concrete, by introducing column vectors $\psi_{s,i}(\theta; s', K, j')$ —for which the $(1 + j)$-th element is given by $\varphi_{(s,i,j)}(\theta; s', K, j')$, for $j \in \{0, ..., M_i\}$—, and applying first-step analysis to $(\mathcal{Y}, \mathcal{X}'(K))$, transforms $\varphi_{(s,i,j)}(\theta; s', K, j')$ are found to be the unique solution of

$$\begin{aligned} &\Big((\theta + \lambda) I_{c_{1,K-1}} - Q^{[s]}(K-1)\Big) \Psi_s(\theta; s', K, j') \\ &\quad = (1 - \delta_{s',s}) \lambda \Psi_{s+1}(\theta; s', K, j') + \delta_{s',s} Q^{*[s]}(K-1) e_{1+M_K}(1 + j'), \quad (6) \end{aligned}$$

for stages $s \in \{1, ..., S\}$ and $s' \in \{s, ..., S\}$, with

$$\Psi_s(\theta; s', K, j') = \begin{pmatrix} \psi_{s,0}(\theta; s', K, j') \\ \psi_{s,1}(\theta; s', K, j') \\ \vdots \\ \psi_{s,K-1}(\theta; s', K, j') \end{pmatrix}.$$

By adapting arguments of Di Crescenzo et al. [11, Section 2.2], an iterative scheme for computing $\psi_{s,i}(\theta; s', K, j')$ can then be derived from (6) by exploiting the structured form of sub-matrices $Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$. This is summarized in the following result.

Theorem 1 *For any stage $s \in \{1, ..., S\}$ and phase $j' \in \{0, ..., M_K\}$, vectors $\psi_{s,i}(\theta; s, K, j')$ satisfy*

$$\psi_{s,i}(\theta; s, K, j') = \begin{cases} H_{s,i}(\theta; K) Q^{[s]}_{i,i+1} \psi_{s,i+1}(\theta; s, K, j'), & if\ i \in \{0, ..., K-2\}, \\ H_{s,K-1}(\theta; K) Q^{[s]}_{K-1,K} e_{1+M_K}(1 + j'), & if\ i = K-1, \end{cases} \quad (7)$$

with $H_{s,i}(\theta; K) = \Big((\theta + \lambda) I_{1+M_i} - Q^{[s]}_{i,i} - (1 - \delta_{i,0}) Q^{[s]}_{i,i-1} H_{s,i-1}(\theta; K) Q^{[s]}_{i-1,i}\Big)^{-1}$, and vectors $\psi_{s,i}(\theta; s', K, j')$, for $s' \in \{s + 1, ..., S\}$, satisfy

$$\psi_{s,i}(\theta; s', K, j') = \begin{cases} H_{s,i}(\theta; K) Q^{[s]}_{i,i+1} \psi_{s,i+1}(\theta; s', K, j') + h_{s,i}(\theta; s', K, j'), & if\ i \in \{0, ..., K-2\}, \\ h_{s,K-1}(\theta; s', K, j'), & if\ i = K-1, \end{cases} \quad (8)$$

where

$$h_{s,i}(\theta;s',K,j') = H_{s,i}(\theta;K) \times \left(\lambda\psi_{s+1,i}(\theta;s',K,j') + (1-\delta_{i,0})Q^{[s]}_{i,i-1}h_{s,i-1}(\theta;s',K,j')\right).$$

From Theorem 1, the probability law of η'_K on the sample paths $\{J'(\eta'_K;K) = j'\}$ of process $(\mathcal{Y},\mathcal{X}'(K))$, provided that $(1,i,j)$ is its initial state, can then be determined from the transform

$$\sum_{s'=1}^{S}\varphi_{(1,i,j)}(\theta;s',K,j') = e^T_{1+M_i}(1+j)\sum_{s'=1}^{S}\psi_{1,i}(\theta;s',K,j'),$$

for $\Re(\theta)\geq 0$, $(i,j)\in\cup_{k=0}^{K-1}\mathcal{L}(k)$ and $j'\in\{0,...,M_K\}$; and related hitting probabilities $P(J'(\eta'_K;K) = j' \mid (Y'(0),I'(0;K),J'(0;K)) = (1,i,j))$ (i.e., values $F_S(T|i,j;K,j')$ in (5)) can be evaluated as

$$e^T_{1+M_i}(1+j)\sum_{s'=1}^{S}\psi_{s,i}(0;s',K,j'),$$

for $(i,j)\in\cup_{k=0}^{K-1}\mathcal{L}(k)$ and $j'\in\{0,...,M_K\}$. It is also clear that

$$e^T_{1+M_i}(1+j)\sum_{s'=1}^{S}\sum_{j'=0}^{M_K}\psi_{1,i}(0;s',K,j') = P\big(\eta'_K < Y \mid (I'(0),J'(0)) = (i,j)\big),$$

for any $(i,j)\in\cup_{k=0}^{K-1}\mathcal{L}(k)$.

In a similar way to Theorem 1, the column vectors

$$m^{(n)}_{s,i}(s',K,j') = (-1)^n \left.\frac{d^n\psi_{s,i}(\theta;s',K,j')}{d\theta^n}\right|_{\theta=0},$$

for $n\in\mathbb{N}$, integers $i\in\{0,...,K-1\}$, stages $s\in\{1,...,S\}$ and $s'\in\{s,...,S\}$, and phases $j'\in\{0,...,M_K\}$, can be iteratively evaluated, starting with $m^{(0)}_{s,i}(s',K,j') = \psi_{s,i}(0;s',K,j')$, from

$$m^{(n)}_{s,i}(s',K,j') = \begin{cases} H_{s,i}(K)Q^{[s]}_{i,i+1}m^{(n)}_{s,i+1}(s',K,j') + h^{(n)}_{s,i}(s',K,j'), & \text{if } i\in\{0,...,K-2\},\\ h^{(n)}_{s,K-1}(s',K,j'), & \text{if } i = K-1,\end{cases} \tag{9}$$

where $H_{s,i}(K) = H_{s,i}(0;K)$ and

$$h^{(n)}_{s,i}(s',K,j') = H_{s,i}(K)\Big(nm^{(n-1)}_{s,i}(s',K,j') + (1-\delta_{i,0})Q^{[s]}_{i,i-1}h^{(n)}_{s,i-1}(s',K,j') + \delta_{s',s}\delta_{i,K-1}Q^{[s]}_{K-1,K}e_{1+M_K}(1+j') + (1-\delta_{s',s})\lambda m^{(n)}_{s+1,i}(s',K,j')\Big).$$

As a result, the $(1+j)$-th element of $\sum_{s'=1}^{S} m_{1,i}^{(n)}(s', K, j')$ corresponds to

$$E\left[(\eta'_K)^n 1_{\{J'(\eta'_K;K)=j'\}} \mid (Y'(0), I'(0;K), J'(0;K)) = (1,i,j)\right],$$

for $n \in \mathbb{N}$, initial states $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$ and phases $j' \in \{0, ..., M_K\}$.

The solution in Eqs. (7)-(8) and (9) leads to Algorithms 1 (Appendix A) and 2 (Appendix B), respectively. Specifically, vectors $m_{s,i}^{(0)}(s', K, j')$ and matrices $H_{s,i}(K)$ in Algorithm 2 must be previously evaluated from Algorithm 1 as $\psi_{s,i}(\theta; s', K, j')$ and $H_{s,i}(\theta; K)$ at point $\theta = 0$, respectively.

4. First-passage times to higher levels under taboo of level $\mathcal{L}(0)$

In Section 2, we considered hitting probabilities and hitting times for sample paths of process $\mathcal{X}$ that move from any initial state $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$ to a higher level $\mathcal{L}(K)$. Here, we focus on the subset of those sample paths that, in particular, do not visit level $\mathcal{L}(0)$ before the first passage of $\mathcal{X}$ to $\mathcal{L}(K)$; i.e., we consider the conditional probabilities

$$G(T|i,j;K,j') = P\big(\eta_K \leq T, J(\eta_K) = j', \eta_K < \eta_0 \mid (I(0), J(0)) = (i,j)\big), \tag{10}$$

for initial states $(i,j) \in \cup_{k=1}^{K-1}\mathcal{L}(k)$ and phases $j' \in \{0, ..., M_K\}$, where η_0 represents the first-passage time of $\mathcal{X}$ to $\mathcal{L}(0)$.

Observe that $G(T|i,j;K,j')$ in Eq. (10) is the probability that, under the taboo of $\mathcal{L}(0)$, process $\mathcal{X}$ reaches level $\mathcal{L}(K)$ before time T and do so in phase $j' \in \{0, ..., M_K\}$, when (i,j) is its initial state. Therefore, we may follow closely the approach of Section 2 in the sense that, for any initial state $(i,j) \in \cup_{k=1}^{K-1}\mathcal{L}(k)$, values of $G(T|i,j;K,j')$ are linked to the visit to $\mathcal{L}(K)$ in a process $\hat{\mathcal{X}}(K) = \{(\hat{I}(t;K), \hat{J}(t;K)) : t \geq 0\}$, which is defined to be a finite-state QBD process with state space $\mathcal{S}_{\hat{\mathcal{X}}(K)} = \mathcal{S}_{\mathcal{X}(K)}$ and q-matrix

$$Q_{\hat{\mathcal{X}}(K)}(t) = \begin{pmatrix} 0_{(1+M_0)\times(1+M_0)} & 0_{(1+M_0)\times\sum_{k=1}^{K-1}(1+M_k)} & 0_{(1+M_0)\times(1+M_K)} \\ \hat{Q}^{**}(t;K-1) & \hat{Q}(t;K-1) & \hat{Q}^{*}(t,K-1) \\ 0_{(1+M_K)\times(1+M_0)} & 0_{(1+M_K)\times\sum_{k=1}^{K-1}(1+M_k)} & 0_{(1+M_K)\times(1+M_K)} \end{pmatrix},$$

where $\hat{Q}^{**}(t;K-1)$, $\hat{Q}(t;K-1)$ and $\hat{Q}^{*}(t,K-1)$ record the time-varying transition rates for jumps of $\mathcal{X}$ from states in $\cup_{k=1}^{K-1}\mathcal{L}(k)$ to subsets $\mathcal{L}(0)$, $\cup_{k=1}^{K-1}\mathcal{L}(k)$ and $\mathcal{L}(K)$, respectively.

The augmented process $(\mathcal{Y},\mathcal{X}'(K))$, used in Section 2 to approximate probabilities $F(T|i,j;K,j')$ through $F_S(T|i,j;K,j')$, is now replaced by an augmented time-homogeneous process $(\hat{\mathcal{Y}},\hat{\mathcal{X}}'(K)) = \{(\hat{Y}'(t),\hat{I}'(t;K),\hat{J}'(t;K)) : t\geq 0\}$ that takes values on

$$\mathcal{S}_{(\hat{\mathcal{Y}},\hat{\mathcal{X}}'(K))} \quad = \quad \{0\}\cup\bigcup_{s=1}^{S}\bigcup_{k=1}^{K-1}\left(\{s\}\times\mathcal{L}(k)\right)\cup\bigcup_{s=1}^{S}\left(\{s\}\times\mathcal{L}(K)\right),$$

in such a way that state 0 represents either the absorption in the Markov chain $\mathcal{Y}$ or a visit to *taboo* states in $\mathcal{X}$; i.e., states in $\mathcal{L}(0)$. The q-matrix of $(\hat{\mathcal{Y}},\hat{\mathcal{X}}'(K))$ has the form of $Q_{(\mathcal{Y},\mathcal{X}'(K))}$ in Section 2 with the value $\hat{c}_{S,K-1}$ instead of $c_{S,K-1}$, sub-vector $w'(K-1) = -\overline{Q}'(K-1)1_{\hat{c}_{S,K-1}} - W'(K-1)1_{1+M_K}$, and sub-matrices

$$\overline{Q}'(K-1) \quad = \quad \begin{pmatrix} \hat{Q}^{[1]}(K-1)-\lambda I_{\hat{c}_{1,K-1}} & \lambda I_{\hat{c}_{1,K-1}} & & \\ & \ddots & \ddots & \\ & & \hat{Q}^{[S-1]}(K-1)-\lambda I_{\hat{c}_{1,K-1}} & \lambda I_{\hat{c}_{1,K-1}} \\ & & & \hat{Q}^{[S]}(K-1)-\lambda I_{\hat{c}_{1,K-1}} \end{pmatrix},$$

$$W'(K-1) \quad = \quad \mathrm{diag}\left(\hat{Q}^{*[1]}(K-1),...,\hat{Q}^{*[S]}(K-1)\right),$$

where $\hat{c}_{a,K-1} = a\sum_{k=1}^{K-1}(1+M_k)$ is the cardinality of $\cup_{s=1}^{a}\cup_{k=1}^{K-1}\left(\{s\}\times\mathcal{L}(k)\right)$, and $\hat{Q}^{[s]}(K-1)$ and $\hat{Q}^{*[s]}(K-1)$ are obtained from $Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$ by removing rows and columns of taboo states.

Then, similarly to Eq. (5), probabilities $G(T|i,j;K,j')$ in process $\hat{\mathcal{X}}(K)$ can be approximated, for a sufficiently large S, by their counterparts $G_S(T|i,j;K,j')$ in $\hat{\mathcal{X}}'(K)$; i.e., for initial states $(i,j)\in\cup_{k=1}^{K-1}\mathcal{L}(k)$ and phases $j'\in\{0,...,M_K\}$,

$$G_S(T|i,j;K,j') \quad = \quad \sum_{s=1}^{S} g_S(T|1,i,j;s,K,j'), \tag{11}$$

where $g_S(T|1,i,j;s,K,j') = P((\hat{Y}'(\hat{\eta}_K'),\hat{J}'(\hat{\eta}_K';K)) = (s,j') \mid (\hat{Y}'(0),\hat{I}'(0;K),\hat{J}'(0;K)) = (1,i,j))$, and $\hat{\eta}_K'$ is the first-passage time of $\hat{\mathcal{X}}'(K)$ to $\mathcal{L}(K)$.

By adapting the approach of Section 3 to process $(\hat{\mathcal{Y}},\hat{\mathcal{X}}'(K))$, values of $g_S(T|1,i,j;s,K,j')$ can accordingly be evaluated as hitting probabilities on states in $\cup_{s=1}^{S}(\{s\} \times \mathcal{L}(K))$. To be concrete, the probability law of the first-passage time $\hat{\eta}'_K$ to subset $\cup_{s=1}^{S}(\{s\} \times \mathcal{L}(K))$ in $(\hat{\mathcal{Y}}, \hat{\mathcal{X}}'(K))$ can be determined from restricted Laplace-Stieltjes transforms $\hat{\varphi}_{(s,i,j)}(\theta; s', K, j')$, defined as

$$E\left[e^{-\theta\hat{\eta}'_K}\bar{1}_{\{(\hat{Y}'(\hat{\eta}'_K),\hat{J}'(\hat{\eta}'_K;K))=(s',j')\}} \mid (\hat{Y}'(0), \hat{I}'(0;K), \hat{J}'(0;K)) = (s,i,j)\right],$$

for $\Re(\theta) \geq 0$, initial states (s,i,j) with $s \in \{1,...,S\}$ and $(i,j) \in \cup_{k=1}^{K-1}\mathcal{L}(k)$, stages $s' \in \{s,...,S\}$ and phases $j' \in \{0,...,M_K\}$. We do this by introducing vectors $\hat{\psi}_{s,i}(\theta; s', K, j')$ —with $(1+j)$-th elements $\hat{\varphi}_{(s,i,j)}(\theta; s', K, j')$, for $j \in \{0,...,M_i\}$—, which are found to satisfy Eqs. (7) and (8) for $s \in \{1,...,S\}$, $i \in \{1,...,K-1\}$ (instead of $i \in \{0,...,K-1\}$), stages $s' \in \{s,...,S\}$ and phases $j' \in \{0,...,M_K\}$, with matrices $H_{s,i}(\theta; K)$ and vectors $h_{s,i}(\theta; s', K, j')$ replaced by

$$\begin{aligned}\hat{H}_{s,i}(\theta;K) &= \Big((\theta+\lambda)I_{1+M_i} - \hat{Q}^{[s]}_{i,i} - (1-\delta_{i,1})\hat{Q}^{[s]}_{i,i-1}\hat{H}_{s,i-1}(\theta;K)\hat{Q}^{[s]}_{i-1,i}\Big)^{-1},\\ \hat{h}_{s,i}(\theta;s',K,j') &= \hat{H}_{s,i}(\theta;K)\\ &\quad\times\Big(\lambda\hat{\psi}_{s+1,i}(\theta;s',K,j') + (1-\delta_{i,1})\hat{Q}^{[s]}_{i,i-1}\hat{h}_{s,i-1}(\theta;s',K,j')\Big),\end{aligned}$$

respectively. This means that the probability $G_S(T|i,j;K,j')$ in Eq. (11) is the value of

$$\sum_{s'=1}^{S}\hat{\varphi}_{(1,i,j)}(\theta;s',K,j') = e^T_{1+M_i}(1+j)\sum_{s'=1}^{S}\hat{\psi}_{1,i}(\theta;s',K,j')$$

at point $\theta = 0$, and the hitting probabilities

$$P\big(\eta_K \leq T, J(\eta_K) = j', \eta_K < \eta_0 \mid (I(0), J(0)) = (i,j)\big)$$

in the original process $\mathcal{X}$ can be approximated, for a sufficiently large S, by

$$\begin{aligned}P\big(\hat{J}'(\hat{\eta}'_K;K) = j' \mid (\hat{Y}'(0), \hat{I}'(0;K), \hat{J}'(0;K)) = (1,i,j)\big)\\ = e^T_{1+M_i}(1+j)\sum_{s'=1}^{S}\hat{\psi}_{1,i}(0;s',K,j'),\end{aligned}$$

for initial states $(i,j) \in \cup_{k=1}^{K-1}\mathcal{L}(k)$ and phases $j' \in \{0,...,M_K\}$. As a result, it holds that

$$P\big(\eta_K \leq T, \eta_K < \eta_0 \mid (I(0), J(0)) = (i,j)\big) \approx e_{1+M_i}^T (1+j) \sum_{j'=0}^{M_K} \sum_{s'=1}^{S} \hat{\psi}_{1,i}(0; s', K, j'),$$

since the taboo probability $P(\eta_K \leq T, \eta_K < \eta_0 \mid (I(0), J(0)) = (i,j))$ can be expressed as the value $\lim_{S\to\infty} \sum_{j'=0}^{M_K} \sum_{s'=1}^{S} P((Y'(\eta_K), \hat{J}(\eta_K; K)) = (s', j') \mid (Y'(0), \hat{I}(0; K), \hat{J}(0; K)) = (1, i, j))$ and $e_{1+M_i}^T (1+j) \sum_{j'=0}^{M_K} \sum_{s'=1}^{S} \hat{\psi}_{1,i}(0; s', K, j') = P(\hat{\eta}'_K < Y \mid (\hat{Y}'(0), \hat{I}'(0; K), \hat{J}'(0; K)) = (1, i, j))$.

For the sake of completeness, Algorithm 3 (Appendix C) evaluates vectors $\hat{\psi}_{s,i}(0; s', K, j')$, for $i \in \{1, ..., K-1\}$, $s \in \{1, ..., S\}$, $s' \in \{s, ..., S\}$ and $j' \in \{0, ..., M_K\}$, from the above results, and Algorithm 4 (Appendix C) summarizes the iterative scheme to calculate n-th moments

$$E\left[(\hat{\eta}'_K)^n 1_{\{\hat{J}'(\hat{\eta}'_K; K) = j'\}} \mid (\hat{Y}'(0), \hat{I}'(0; K), \hat{J}'(0; K)) = (1, i, j)\right],$$

for $n \in \mathbb{N}$, initial states $(i, j) \in \cup_{k=1}^{K-1} \mathcal{L}(k)$ and phases $j' \in \{0, ..., M_K\}$.

5. Numerical work

In this section, the objective is to propose how to choose the number of stages S in our Erlangian solution and exemplify the approach with numerical experiments for $M_t/G/\infty$ queues with Erlang service times and sinusoidal arrivals, and SIR epidemic models with sinusoidal forcing of transmission. To preserve the average dynamics of the underlying processes $\mathcal{X}(K)$ and $\hat{\mathcal{X}}(K)$ in these models, the corresponding augmented processes $(\mathcal{Y}, \mathcal{X}'(K))$ and $(\hat{\mathcal{Y}}, \hat{\mathcal{X}}'(K))$ are defined from the transition rates $q_{(i,j),(i',j')}^{[s]}$ in Eq. (4).

5.1. On the selection of S

In the preceding discussions S is assumed to be selected as a sufficiently large integer, which should not be done only from the asymptotic property that $Y \Rightarrow T$ in distribution as $S \to \infty$. In the following, we adopt the usual method; i.e., we consider a matrix norm of the difference between the mass functions $\{F(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$ and their Erlangian approximations $\{F_S(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$, for initial states $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$ and fixed values T and K, and we stop the procedure when this becomes less than a predetermined tolerance. Our arguments below can be adapted, with minor modifications, to the mass function $\{G(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$ and its related approximation $\{G_S(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$, and are thus omitted here.

Consider the matrix norm $||\cdot||_\infty$[1] and measure the *overall* estimation error in terms of $||V(T|K) - V'_S(T|K)||_\infty$, where the $\Big((1-\delta_{i,0})\sum_{k=0}^{i-1}(1+M_k)+1+j, 1+j'\Big)$-th elements of $V(T|K)$ and $V'_S(T|K)$ are given by $F(T|i,j;K,j')$ and $F_S(T|i,j;K,j')$, respectively, for $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$ and $j' \in \{0,...,M_K\}$. In selecting S as the smallest integer that leads to

$$||V(T|K) - V'_S(T|K)||_\infty \quad < \quad \varepsilon, \tag{12}$$

for the prespecified tolerance $\varepsilon > 0$, the main difficulty concerns the fact that $V(T|K)$ is a sub-matrix of the transition matrix $P_{\mathcal{X}(K)}(0,T)$ in Eq. (3), while matrix $V'_S(T|K)$ is iteratively evaluated from (7)-(8). To overcome this drawback, we may introduce a suitably defined time-inhomogeneous process $\mathcal{Z}(K)$ for which an explicit expression for its transition matrix $P_{\mathcal{Z}(K)}(0,T)$ can be written. Specifically, this auxiliary process $\mathcal{Z}(K)$ takes values on the set $\cup_{k=0}^{K}\mathcal{L}(k)$ (i.e., $\mathcal{S}_{\mathcal{Z}(K)} = \mathcal{S}_{\mathcal{X}(K)}$) and its q-matrix is defined to have discrete changes as follows:

$$Q_{\mathcal{Z}(K)}(t) \quad = \quad Q^{[s]}_{\mathcal{Z}(K)}, \quad \text{if } t \in \left[\frac{(s-1)T}{S}, \frac{sT}{S}\right), \text{ for } s \in \{1,...,S\},$$

---

[1]The matrix norm $||\cdot||_\infty$ is here defined on the $\mathbb{R}$-vector space of matrices with $\sum_{k=0}^{K-1}(1+M_k)$ rows and $1+M_K$ columns, and takes the form

$$||A||_\infty \quad = \quad \max_{m\in\{1,...,\sum_{k=0}^{K-1}(1+M_k)\}} \left\{\sum_{m'=1}^{1+M_K} |a_{m,m'}|\right\},$$

for any matrix A with $\mathbb{R}$-valued elements $a_{m,m'}$.

where $Q^{[s]}_{\mathcal{Z}(K)}$ has the structured form

$$Q^{[s]}_{\mathcal{Z}(K)} = \begin{pmatrix} Q^{[s]}(K-1) & Q^{*[s]}(K-1) \\ 0_{(1+M_K)\times\sum_{k=0}^{K-1}(1+M_k)} & 0_{(1+M_K)\times(1+M_K)} \end{pmatrix},$$

and sub-matrices $Q^{[s]}(K-1)$ and $Q^{*[s]}(K-1)$, for stages $s \in \{1,...,S\}$, are those used in Section 3 in the Erlangian approximation. With this, process $\mathcal{Z}(K)$ is defined at time instants $t \in [0,T)$ —for any $t \in [T,\infty)$, $Q_{\mathcal{Z}(K)}(t)$ can be arbitrarily defined— in such a way that the product-integral form characterizing its transition function becomes

$$P_{\mathcal{Z}(K)}(0,T) = \prod_{s=1}^{S} e^{Q^{[s]}_{\mathcal{Z}(K)} S^{-1} T}, \tag{13}$$

where the ordered product $\prod_{s=1}^{S} e^{Q^{[s]}_{\mathcal{Z}(K)} S^{-1} T}$ must be evaluated from the smallest stage ($s = 1$) to the largest one ($s = S$); see, e.g., Rindos et al. [45, Equation (8.9)].

From the structured form

$$P_{\mathcal{Z}(K)}(0,T) = \begin{pmatrix} \tilde{P}(0,T;K-1) & \tilde{P}^*(0,T;K-1) \\ 0_{(1+M_K)\times\sum_{k=0}^{K-1}(1+M_k)} & I_{1+M_K} \end{pmatrix},$$

the product of matrices in (13) enables sub-matrix $\tilde{P}^*(0,T;K-1)$ to be evaluated and, therefore, matrix $V_S(T|K) = \tilde{P}^*(0,T;K-1)$ to be introduced in our arguments as a *piecewise-constant* approximation of $V(T|K)$. Then, for processes $\mathcal{X}(K)$ and $\mathcal{Z}(K)$ with $\sup_{t\in[0,T]} \left|\left|Q_{\mathcal{X}(K)}(t)\right|\right|_\infty \leq M(T;K-1)$ and $\sup_{t\in[0,T]} \left|\left|Q_{\mathcal{Z}(K)}(t)\right|\right|_\infty \leq M(T;K-1)$, for a finite constant $M(T;K-1)$[2], the triangle inequality yields

$$\begin{aligned} ||V(T|K) - V'_S(T|K)||_\infty \leq\ & \epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) \frac{e^{M(T;K-1)T} - 1}{M(T;K-1)} \\ & + ||V_S(T|K) - V'_S(T|K)||_\infty, \end{aligned} \tag{14}$$

where $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) = \sup_{t\in[0,T]} \left|\left|Q_{\mathcal{X}(K)}(t) - Q_{\mathcal{Z}(K)}(t)\right|\right|_\infty$. In particular, the first term on the right-hand side of (14) is derived from the relation

[2]For transition rates $q^{[s]}_{(i,j),(i',j')}$ defined from (4) (also from Remark 2), it is seen that $\sup_{t\in[0,T]} \left|\left|Q_{\mathcal{Z}(K)}(t)\right|\right|_\infty \leq \sup_{t\in[0,T]} \left|\left|Q_{\mathcal{X}(K)}(t)\right|\right|_\infty$, for any number S of stages.

$||V(T|K) - V_S(T|K)||_\infty \leq \left|\left|P_{\mathcal{X}(K)}(0,T) - P_{\mathcal{Z}(K)}(0,T)\right|\right|_\infty$[3] and Theorem 5 of Esquível et al. [16].

By using (14), the number S of stages can therefore be selected as the smallest integer such that

$$\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) \frac{e^{M(T;K-1)T} - 1}{M(T;K-1)} + ||V_S(T|K) - V'_S(T|K)||_\infty \quad < \quad \varepsilon,$$

for a predetermined value ε that typically ranges between 0.05 and 0.005, once the term $||V_S(T|K) - V'_S(T|K)||_\infty$ is evaluated at each S from (7)-(8) and (13). Unfortunately, it is frequently the case that there is a risk of floating-point overflow in the computation of the term $e^{M(T;K-1)T}$ when $M(T;K-1)T$ is quite large; furthermore, it may also happen that, when the cardinality of $\mathcal{Z}(K)$ becomes large, the computation of (13) is unstable. Instead, we may use either an approach based on successive iterations

$$||V_S(T|K) - V_{S-1}(T|K)||_\infty,$$

or another one based on the property

$$\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) \quad \to \quad 0,$$

as $S \to \infty$. The former is usually seen to be satisfactory when the procedure converges relatively rapidly; however, the Erlangian approximation $V_S(T|K)$ may be far from the solution $V(T|K)$ when the procedure is converging slowly. The latter is investigated in our numerical experiments in Subsections 5.2 and 5.3 by using, for a better probabilistic interpretation, the Hellinger distance of two mass functions instead of the matrix norm $||\cdot||_\infty$.

5.2. $M_t/G/\infty$ queues with Erlang service times and sinusoidal arrivals

For our first example, we consider the $M_t/G/\infty$ queue used by Green and Kolesar [22] to predict the maximum congestion from the customer arrivals

[3] The matrix norm $||\cdot||_\infty$ on the right-hand side of this inequality is defined on the $\mathbb{R}$-vector space of square matrices of order $\sum_{k=0}^{K}(1+M_k)$.

peak at queues with cyclic arrivals. Specifically, in Ref. [22] customers arrive according to a non-stationary Poisson stream with arrival rate

$$\beta(t) \quad = \quad \beta_0 \left(1 + \frac{1}{2} \sin \frac{2\pi t}{T}\right),$$

where $\beta_0 > 0$ is the overall mean arrival rate, with $T = 24$ *hours* and the peak of customer arrivals at $t = 6$ *hours*. Service times are independent, Erlang distributed with 2 stages and mean $\alpha^{-1}2$, for $\alpha > 0$, and are also mutually independent of the inter-arrival times.

In the framework of Section 2, the level and phase variables are defined here as $I(t) = J_1(t) + J_2(t)$ and $J(t) = J_2(t)$, where $J_1(t)$ and $J_2(t)$ represent the number of servers operating at stages 1 and 2, respectively, at time t. This results in a time-inhomogeneous LD-QBD process $\mathcal{X}$ that takes values on the set $\mathcal{S} = \cup_{i=0}^{\infty}\mathcal{L}(i)$, where levels $\mathcal{L}(i)$ are defined as $\{(i,j) : j \in \{0,...,i\}\}$; for a concrete specification of sub-matrices $Q_{i,i'}(t)$, for $i' \in \{\max\{0, i-1\}, i, i+1\}$, see Appendix D. Since the level variable $I(t)$ is related to congestion, the event $\{\eta_K \leq 24, J(\eta_K) = j'\}$ specifies that the system reaches a congestion level with K customers before 24 *hours* have elapsed with, in particular, $K - j'$ and j' servers operating in stages 1 and 2, respectively, at the first time when this congestion level is reached; and $F(24|i,j;K,j')$ is its conditional probability, provided that $i - j$ and j servers are initially operating in stages 1 and 2, respectively.

As discussed in the previous section, we shall begin by showing how the fact that $\lim_{S\to\infty} \epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) = 0$ can be used to select S in our approach. To do this, the solution $\mathcal{P}(K) = \{F(24|i,j;K,j') : j' \in \{0,...,K\}\}$ is first evaluated from a simulation study of $\mathcal{X}$, for congestion levels K ranging between 5 and 50, and the closeness of $\mathcal{P}(K)$ and its Erlangian approximation $\mathcal{P}_S(K) = \{F_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ is measured by using the Hellinger distance $H(\mathcal{P}(K), \mathcal{P}_S(K))$[4], for a number of stages $S = 2^m$ with $m \in \mathbb{N}_0$. Then, a numerical evidence-based criterion is proposed to select S by analyzing how

[4]For the discrete distributions $\mathcal{P}(K)$ and $\mathcal{P}_S(K)$, the Hellinger distance $H(\mathcal{P}(K), \mathcal{P}_S(K))$

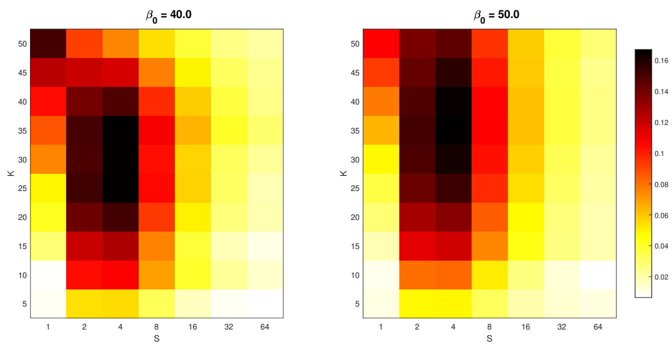

Figure 1: The Hellinger distance $H(\mathcal{P}(K), \mathcal{P}_S(K))$ as a function of S, in scenarios with $\beta_0 = 40.0$ (left) and 50.0 (right) and $\alpha = 2.0$, for congestion levels K ranging between 5 and 50.

sufficiently small values of $H(\mathcal{P}(K), \mathcal{P}_S(K))$ and $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ are related to each other in our numerical experiments.

In the simulation study, the occurrence of the event $\{\eta_K \leq 24\}$ and the stages of those servers operating at time η_K are estimated from 10^5 simulated paths of process $\mathcal{X}$ starting at the beginning of a busy period —that is, with initial state $(i, j) = (1, 0)$—, in scenarios with values $\beta_0 \in \{40.0, 50.0\}$ and $\alpha = 2.0$. From this study, Figure 1 confirms that, in terms of $H(\mathcal{P}(K), \mathcal{P}_S(K))$, the Erlangian approximation $\mathcal{P}_S(K)$ converges to $\mathcal{P}(K)$ as $S \to \infty$. However, it is also noticed that $H(\mathcal{P}(K), \mathcal{P}_S(K))$ is not necessarily a monotonically decreasing function of S, except when $K = 50$ and $\beta_0 = 40.0$ in our numerical results. In its more usual behavior, $H(\mathcal{P}(K), \mathcal{P}_S(K))$ is seen to start increasing to its maximum value and then decreases to 0 with increasing values of S, which is not easy to explain from the system parameters. For illustrative purposes, Figure 2

can be evaluated as

$$H(\mathcal{P}(K), \mathcal{P}_S(K)) = \frac{1}{\sqrt{2}} \sqrt{\sum_{j'=0}^{K} \left(\sqrt{F(24|1,0;K,j')} - \sqrt{F_S(24|1,0;K,j')} \right)^2},$$

which is equivalent to the Euclidean norm $(\sqrt{2})^{-1} || \sqrt{\mathcal{P}(K)} - \sqrt{\mathcal{P}_S(K)} ||_2$, with $0 \leq H(\mathcal{P}(K), \mathcal{P}_S(K)) \leq 1$.

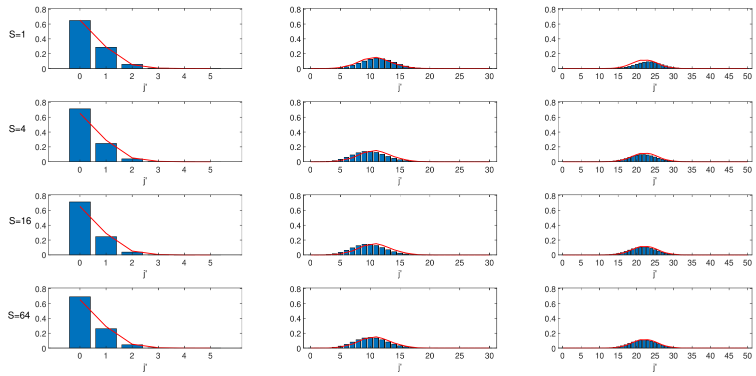

Figure 2: The solution $\mathcal{P}(K)$ (solid line in red) versus its Erlangian approximation $\mathcal{P}_S(K)$ (histogram in blue) in a scenario with $\beta_0 = 40.0$ and $\alpha = 2.0$, for congestion levels $K = 5$, 30 and 50 (from left to right).

graphically shows the closeness of $\mathcal{P}(K)$ and $\mathcal{P}_S(K)$ for different choices of K, in a scenario with $\beta_0 = 40.0$. In contrast to the behavior of $H(\mathcal{P}(K), \mathcal{P}_S(K))$, it is seen that $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ always behaves as a decreasing function in S, which is shown in Table 1 for a number of stages $S = 2^m$ with $m \in \{1, ..., 9\}$ and values $\beta_0 = 40.0$ and $\alpha = 2.0$. In particular, Tables 1 and 2, and numerical results we do not report here, show that the number $S = 128$ of stages, which is related to the first value of $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ less than 1, can be selected to obtain values of $H(\mathcal{P}(K), \mathcal{P}_S(K))$ always less than 0.025 for any K and values β_0 and α.

The above comments provide a criterion for choosing S —as an integer verifying $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) < 1$— that results in Erlangian approximations to $\mathcal{P}(K)$ sufficiently accurate in terms of $H(\mathcal{P}(K), \mathcal{P}_S(K))$ for any congestion level and scenario in our experiments, although a value of $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ close to 1 is far from its limiting value 0 as $S \to \infty$.

Following this criterion, numerical experiments are carried out to analyze, as a function of the expected service time, how plausible it is that a busy period ends before reaching a certain level K of congestion. It is clear that long service

S	$\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$
2	25.4648
8	14.9169
32	3.9144
64	1.9619
128	0.9816
256	0.4908
512	0.2454

Table 1: Values of $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ as a function of S in a scenario with $\beta_0 = 40.0$ and $\alpha = 2.0$. Note that, since the service rate is constant, the values of $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ do not depend on the congestion level K.

K	$S = 64$	$S = 128$
50	0.0191	0.0179
45	0.0219	0.0195
40	0.0231	0.0192
35	0.0270	0.0235
30	0.0210	0.0189
25	0.0172	0.0132
20	0.0176	0.0160
15	0.0110	0.0108
10	0.0140	0.0108
5	0.0068	0.0067

Table 2: Values of $H(\mathcal{P}(K), \mathcal{P}_S(K))$ as a function of K, for stages $S = 64$ and 128, in a scenario with $\beta_0 = 40.0$ and $\alpha = 2.0$.

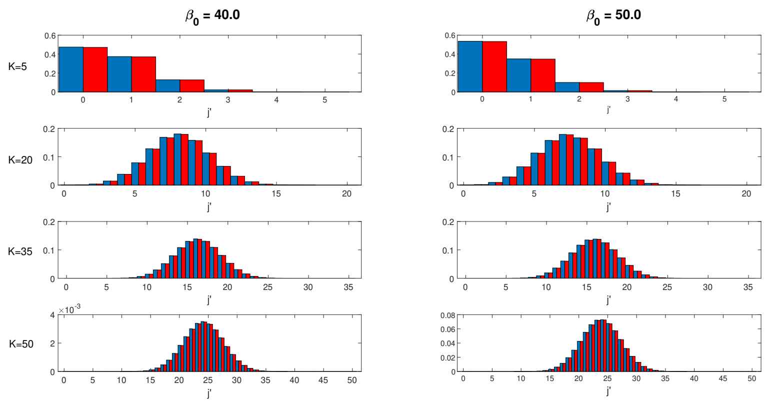

Figure 3: Erlangian approximations $\{F_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in blue) and $\{G_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in red) in a scenario with $\beta_0 = 40.0$ and $\alpha = 4.0$, for $K \in \{5, 20, 35, 50\}$.

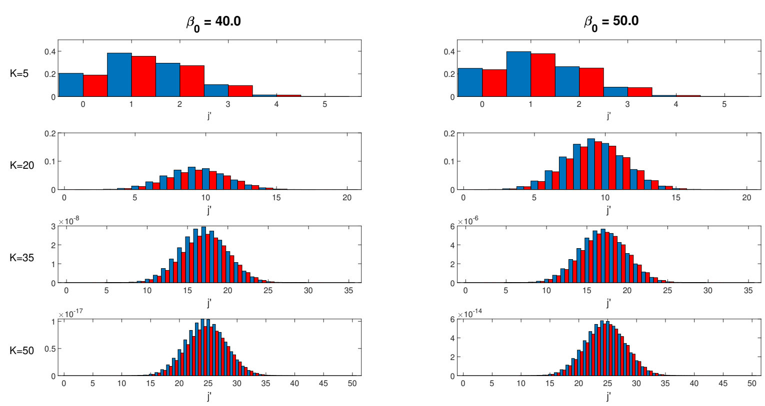

Figure 4: Erlangian approximations $\{F_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in blue) and $\{G_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in red) in a scenario with $\beta_0 = 40.0$ and $\alpha = 12.0$, for $K \in \{5, 20, 35, 50\}$.

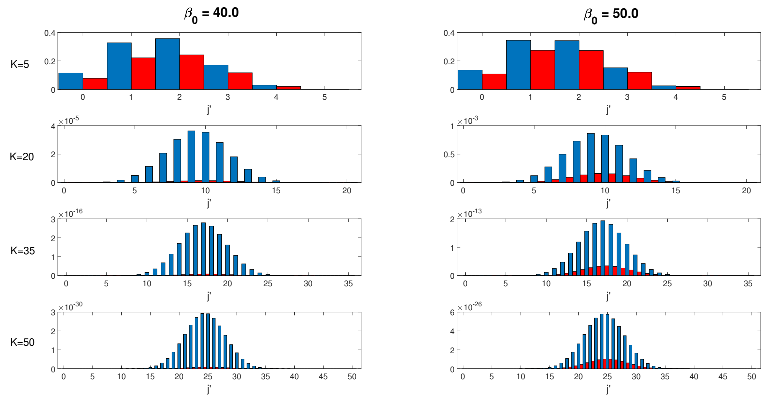

Figure 5: Erlangian approximations $\{F_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in blue) and $\{G_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ (histogram in red) in a scenario with $\beta_0 = 40.0$ and $\alpha = 24.0$, for $K \in \{5, 20, 35, 50\}$.

times result in highly congested systems, so it is not likely that a busy period will end before the system reaches small levels of congestion. This is related to the similarity between $\{F_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ and $\{G_S(24|1,0;K,j') : j' \in \{0,...,K\}\}$ shown in Figure 3 for congestion levels $K \in \{5, 20, 35, 50\}$ and service times with 30 *minutes* of mean duration (that is, $\alpha = 4.0$). In contrast, higher congestion levels are less likely to be reached before a busy period ends when service times have shorter durations, as illustrated in Figures 4 and 5 for service times with 10 *minutes* ($\alpha = 12.0$) and 5 *minutes* ($\alpha = 24.0$) of mean duration, respectively. An important observation is that, at the epoch of congestion, approximately 50% of the servers will operate in stage 1 (customers with shorter waiting times) when K is sufficiently large, while that proportion decreases when the level of congestion is smaller.

5.3. SIR epidemic models with sinusoidal forcing of transmission

In this section, we focus on the SIR (*susceptible-infective-recovered*) epidemic model, which is reasonably predictive for infectious diseases whose recovery confers lifelong immunity (Britton and Lindholm [7]; Kermack and McKendrick

[31]; Neuts and Li [42]). In particular, the following description of a Markov chain model for an SIR-like epidemic is used to reflect seasonal oscillations of influenza incidence (Dushoff et al. [15]) and requires that the contact rate be time dependent.

Consider a closed population of $N < \infty$ individuals, which are homogeneously well mixed, and assume that, from each infected individual, the disease is transmitted at random points of a non-stationary Poisson stream with rate $\beta(t)$ in such a way that the newly infected individual is selected, under independent and uniform assumptions, from the sub-population of susceptible individuals. Infectious periods are assumed to be governed by independent, exponential-distributed random variables with mean length γ^{-1}, and are also mutually independent of the infectious contact processes.

These distributional assumptions lead to a finite-state QBD process $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$, where the level and phase variables denote the number of infected and susceptible individuals —so that $R(t) = N - I(t) - J(t)$ is the number of recovered individuals— at time t. For a population consisting initially of N_I infected and N_J susceptible individuals —that is, $I(0) = N_I$ and $J(0) = N_J$ with $N_I + N_J = N$—, $\mathcal{X}$ takes values in $\mathcal{S} = \cup_{i=0}^{N} \mathcal{L}(i)$ with $\mathcal{L}(i) = \{(i, j) : i \in \{0, ..., N_I + N_J - j\}, j \in \{0, ..., N_J\}\}$, for integers $i \in \{0, ..., N\}$. Level $\mathcal{L}(0)$ consists of disease-free states, which are absorbing, whereas states in $\cup_{i=1}^{N} \mathcal{L}(i)$ are transient and correspond to an outbreak in progress. Elements of $Q_{i,i'}(t)$, for $i' \in \{\max\{0, i-1\}, i, \min\{i+1, N\}\}$, are easily derived, as shown in Appendix E.

In our numerical experiments, the seasonal transmission rate has the form

$$\beta(t) \quad = \quad \beta_0 N \left(1 + \sigma \cos\left(2\pi \left(t - \frac{1}{2}\right)\right)\right),$$

where the parameter $\sigma \in [0, 1]$ amounts to the strength of seasonal forcing and $\beta_0 > 0$ is the person-to-person contact rate in the standard SIR model (i.e., with $\sigma = 0.0$). This choice for the contact rate $\beta(t)$ is usually referred to as Dietz's model [13]; for related work, see Buonomo et al. [8], and Grassly and Fraser [21], among others. In Figure 6, we plot the time-varying reproduction

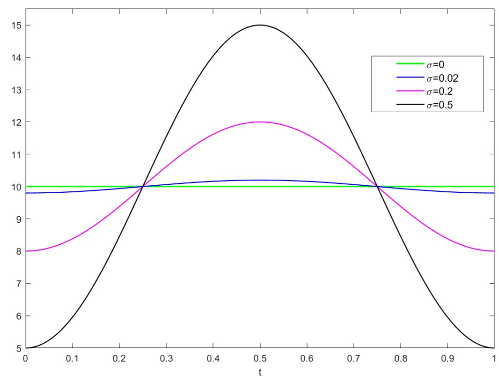

Figure 6: The time-varying reproduction number $R_0(t)$ as a function of t in the standard SIR model and SIR models with time-varying contact rates, for a scenario with $N = 100$.

number $R_0(t) = \gamma^{-1}\beta(t)$ as a function of t, for a population consisting of $N = 100$ individuals, infectious periods with mean duration $\gamma^{-1} = 7.3$ *days*, and parameters $\beta_0 = 5.0$ and $\sigma \in \{0.0, 0.02, 0.2, 0.5\}$. It is seen that, for strictly positive values of σ, $R_0(t)$ evolves periodically over an interval of length $T = 1$ *year* and reaches its maximum value at $t = 6$ *months*, while $R_0(t)$ becomes the basic reproduction number $R_0 = \gamma^{-1}\beta_0 N$ in the standard SIR model. Using the expression for $\beta(t)$, it holds that

$$R_0(t) = \left(1 + \sigma \cos\left(2\pi\left(t - \frac{1}{2}\right)\right)\right) R_0,$$

which reflects how the average transmission potential governed by R_0 can fluctuate over time due to the seasonal conditions introduced by σ. For example, when $R_0 = 10.0$ and $\sigma > 0$, the curve of $R_0(t)$ exhibits seasonal behavior in influenza transmission —from mild to more pronounced patterns— in which transmission tends to increase during winter months and decrease during warmer periods, as a result of climatic conditions, changes in social behavior, school calendars and indoor activities, among other factors; see, e.g., Dushoff et al. [15]. This temporal variability can significantly influence the spread of the disease, depending on the time of year at which the infection is introduced into the population.

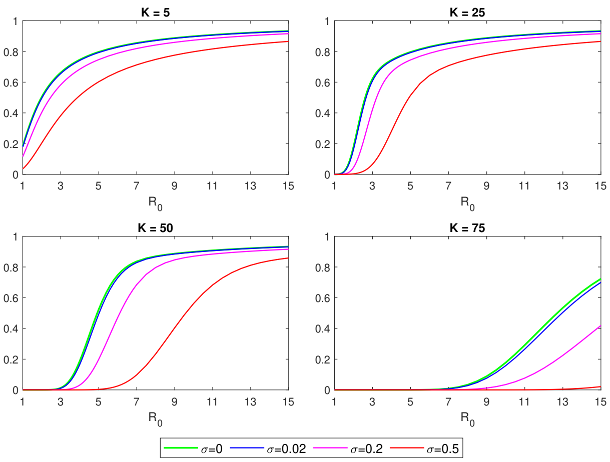

Figure 7: The Erlangian approximation of probability $P(\eta_K \leq T \mid (I(0), J(0)) = (1, 99))$ as a function of R_0 in the standard SIR model and SIR models with time-varying contact rates, for $T = 1$, $K \in \{5, 25, 50, 75\}$ and a scenario with $N = 100$.

To examine this effect, we consider an invasion time —that is, $(i, j) = (1, 99)$ is the initial state of $\mathcal{X}$— and focus on the probability $P(\eta_K \leq T \mid (I(0), J(0)) = (1, 99))$ of reaching K individuals who are simultaneously infected with influenza before an interval of length $T = 1$ *year* expires, for integers $K \in \{5, 25, 50, 75\}$. In our experiments, it is assumed that the baseline value R_0 varies between 1.0 and 15.0 (Dushoff et al. [15]), while the strength of seasonal forcing is set to $\sigma \in \{0.0, 0.02, 0.2, 0.5\}$. To approximate $P(\eta_K \leq 1 \mid (I(0), J(0)) = (1, 99))$ via $\sum_{j'=0}^{100-K} F_S(1|1, 99; K, j')$, the mass function $\{F(1|1, 99; K, j') : j' \in \{0, ..., 100 - K\}\}$ is first estimated from a simulation study similar to that of Subsection 5.2, and the number S of stages is then selected as the smallest integer such that $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ is less than one, which always results in values of the Hellinger distance of $\{F(1|1, 99; K, j') : j' \in \{0, ..., 100 - K\}\}$ and $\{F_S(1|1, 99; K, j') : j' \in \{0, ..., 100 - K\}\}$ in our SIR models less than 0.025. From Figure 7, it is observed that, regardless of R_0, the peak of infection will be more significant as seasonal variability decreases. Specifically, increasing values of the strength of seasonal forcing make it less likely to accumulate K simultaneously infected individuals before time $T = 1$ *year*, the standard SIR model being the one with the highest likelihood of such an event. This behavior occurs because seasonality introduces both periods of *low* and *high* transmission, but with different implications; in particular, low transmission will reduce the chance of influenza disease spreading sufficiently, even at high values of R_0, more significantly than high transmission will increase spread. This effect highlights how the interaction between R_0 and σ shapes the temporal pattern of $R_0(t)$, limiting the appearance of influenza outbreaks during certain time periods of the year and making it less likely to observe a significant number of simultaneously infected individuals. As was expected, for fixed K, the Erlangian approximation of $P(\eta_K \leq 1 \mid (I(0), J(0)) = (1, 99))$ increases with increasing values of R_0, regardless of σ.

6. Conclusions

In this article we construct Erlangian approximations to hitting probabilities at a fixed time $T < \infty$ for a time-inhomogeneous LD-QBD process $\mathcal{X}$. The construction replaces time T by an independent, Erlang-distributed random time Y with S stages and mean T, and uses the Widder's Laplace transform inversion formula to approximate the solution $\{F(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$ (respectively, $\{G(T|i,j;K,j') : j' \in \{0,...,M_K\}\}$), for a specified integer K and initial states $(i,j) \in \cup_{k=0}^{K-1}\mathcal{L}(k)$ (respectively, $(i,j) \in \cup_{k=1}^{K-1}\mathcal{L}(k)$), from the dynamics of a suitably defined time-inhomogeneous process $\mathcal{X}'(K)$ (respectively, $\hat{\mathcal{X}}'(K)$) before time Y expires.

An important feature of our approach is related to the definition of $(\mathcal{Y}, \mathcal{X}'(K))$ (respectively, $(\hat{\mathcal{Y}}, \hat{\mathcal{X}}'(K))$) and its relation to the process $\mathcal{Z}(K)$. Unlike process $\mathcal{Z}(K)$ —whose q-matrix makes discrete changes over S sub-intervals obtained by partition of $[0, T)$, each of length $S^{-1}T$—, the q-matrix of the augmented process $(\mathcal{Y}, \mathcal{X}'(K))$ (respectively, $(\hat{\mathcal{Y}}, \hat{\mathcal{X}}'(K))$) makes discrete changes over sub-intervals of random length, which are defined from the S independent, exponential-distributed contributions $Y_1, ..., Y_S$ to the random (Erlang-distributed) length of interval $[0, Y)$. In other words, in this article the sequence of sub-intervals $[S^{-1}(s-1)T, S^{-1}sT)$, for $s \in \{1, ..., S\}$, in the well-known piecewise-constant approximation is replaced by sub-intervals of random length Y_s, for $s \in \{1, ..., S\}$, each with a fixed mean value $S^{-1}T$. This randomization procedure allows us to avoid approximating the solution from the transition matrix $P_{\mathcal{Z}(K)}(0, T)$, especially when the underlying matrix exponentials in (13) behave numerically unstable. Instead, the Erlangian approximation is evaluated by solving systems of linear equations characterizing first-passage time distributions of process $(\mathcal{Y}, \mathcal{X}'(K))$ (respectively, $(\hat{\mathcal{Y}}, \hat{\mathcal{X}}'(K))$); see, for example, the iterative scheme for the computation of $\{F_S(T|i, j; K, j') : j' \in \{0, ..., M_K\}\}$ in Eqs. (7)-(8).

The selection of the number S of stages from (14) is likely to prove highly satisfactory when $M(T; K-1)T$ is not large. However, in certain applications, such as our examples in Subsections 5.2 and 5.3, it may become necessary to examine how the similarity of the q-matrices $Q_{\mathcal{X}(K)}(t)$ (respectively, $Q_{\hat{\mathcal{X}}(K)}(t)$) and $Q_{\mathcal{Z}(K)}(t)$, for $t \in [0, T)$ —measured in terms of $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$ (respectively, $\epsilon_{S,T}(Q_{\hat{\mathcal{X}}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot))$)— can be used to predict the closeness of the solution and its Erlangian approximation, for example, in terms of the underlying Hellinger distance of $\mathcal{P}(K)$ and $\mathcal{P}_S(K)$. To this end, a preliminary simulation study may allow one to estimate the solution and explore how sufficiently small values of $H(\mathcal{P}(K), \mathcal{P}_S(K))$ are obtained by selecting the number of stages S according to the behavior $\lim_{S\to\infty} \epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) = 0$ (respectively, $\lim_{S\to\infty} \epsilon_{S,T}(Q_{\hat{\mathcal{X}}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) = 0$). More concretely, our examples in Subsections 5.2 and 5.3 show that choosing S such that $\epsilon_{S,T}(Q_{\mathcal{X}(K)}(\cdot), Q_{\mathcal{Z}(K)}(\cdot)) < 1$ results in a sufficiently accurate approximation in terms of values of the Hellinger distance $H(\mathcal{P}(K), \mathcal{P}_S(K))$ less than 0.025.

Finally, we note that the aforementioned numerical evidence may be inherently linked to the stochastic descriptor under consideration in this article and, in particular, the sinusoidal assumption for the time-varying transition rates. Therefore, we acknowledge that our conclusions on the selection of S may not be directly applicable to other contexts. Future work could focus, for example, on the case of the transition function of a finite-state Markov chain with time-varying transition rates.

Appendix A. Algorithmic solution of Eqs. (7) and (8)

For fixed values $\theta \in \mathbb{C}$ with $\Re(\theta) \geq 0$ and phases $j' \in \{0, ..., M_K\}$, Algorithm 1 computes the vectors $\psi_{s,i}(\theta; s', K, j')$ with $i \in \{0, ..., K-1\}$, $s \in \{1, ..., S\}$ and $s' \in \{s, ..., S\}$.

Algorithm 1. Iterative evaluation of $\psi_{s,i}(\theta; s', K, j')$ with $i \in \{0, ..., K-1\}$, $s \in \{1, ..., S\}$ and $s' \in \{s, ..., S\}$.

$s := S + 1;$

$i := 0;$

while $s > 1$, do

$s := s - 1;$

$s' := s;$

$H_{s,i}(\theta; K) := \left((\theta + \lambda) I_{1+M_i} - Q^{[s]}_{i,i}\right)^{-1};$

while $i < K - 1$, do

$i := i + 1;$

$H_{s,i}(\theta; K) := \left((\theta + \lambda) I_{1+M_i} - Q^{[s]}_{i,i} - Q^{[s]}_{i,i-1} H_{s,i-1}(\theta; K) Q^{[s]}_{i-1,i}\right)^{-1};$

endwhile;

$\psi_{s,i}(\theta; s, K, j') := H_{s,i}(\theta; K) Q^{[s]}_{i,i+1} e_{1+M_{i+1}}(1 + j');$

while $i > 0$, do

$i := i - 1;$

$\psi_{s,i}(\theta; s, K, j') := H_{s,i}(\theta; K) Q^{[s]}_{i,i+1} \psi_{s,i+1}(\theta; s, K, j');$

endwhile;

while $s' < S$, do

$s' := s' + 1;$

$h_{s,i}(\theta; s', K, j') := \lambda H_{s,i}(\theta; K) \psi_{s+1,i}(\theta; s', K, j');$

while $i < K - 1$, do

$i := i + 1;$

$h_{s,i}(\theta; s', K, j') := H_{s,i}(\theta; K) \left(\lambda \psi_{s+1,i}(\theta; s', K, j') + Q^{[s]}_{i,i-1} h_{s,i-1}(\theta; s', K, j')\right);$

endwhile;

$\psi_{s,i}(\theta; s', K, j') := h_{s,i}(\theta; s', K, j');$

while $i > 0$, do

$i := i - 1;$

$\psi_{s,i}(\theta; s', K, j') := H_{s,i}(\theta; K) Q^{[s]}_{i,i+1} \psi_{s,i+1}(\theta; s', K, j') + h_{s,i}(\theta; s', K, j');$

endwhile;

endwhile;

endwhile.

Appendix B. Algorithmic solution of Eqs. (9)

Starting with $m_{s,i}^{(0)}(s', K, j') = \psi_{s,i}(0; s', K, j')$, Algorithm 2 evaluates the vectors $m_{s,i}^{(n)}(s', K, j')$, for a fixed $n \in \mathbb{N}$, from previously computed vectors $m_{s,i}^{(n-1)}(s', K, j')$.

Algorithm 2. Iterative evaluation of $m_{s,i}^{(n)}(s', K, j')$ with $i \in \{0, ..., K-1\}$, $s \in \{1, ..., S\}$ and $s' \in \{s, ..., S\}$, for a fixed phase $j' \in \{0, ..., M_K\}$.

$s := S + 1;$
$i := 0;$
while $s > 1$, do
    $s := s - 1;$
    $s' := s;$
    $h_{s,i}^{(n)}(s', K, j') := nH_{s,i}(K) m_{s,i}^{(n-1)}(s', K, j');$
    while $i < K - 1$, do
        $i := i + 1;$
        $h_{s,i}^{(n)}(s', K, j') := H_{s,i}(K) \left(n m_{s,i}^{(n-1)}(s', K, j') + Q_{i,i-1}^{[s]} h_{s,i-1}^{(n)}(s', K, j') \right);$
    endwhile;
    $h_{s,i}^{(n)}(s', K, j') := h_{s,i}^{(n)}(s', K, j') + H_{s,i}(K) Q_{K-1,K}^{[s]} e_{1+M_K}(1 + j');$
    $m_{s,i}^{(n)}(s', K, j') := h_{s,i}^{(n)}(s', K, j');$
    while $i > 0$, do
        $i := i - 1;$
        $m_{s,i}^{(n)}(s', K, j') := H_{s,i}(K) Q_{i,i+1}^{[s]} m_{s,i+1}^{(n)}(s', K, j') + h_{s,i}^{(n)}(s', K, j');$
    endwhile;
    while $s' < L$, do
        $s' := s' + 1;$
        $h_{s,i}^{(n)}(s', K, j') := H_{s,i}(K) \left(n m_{s,i}^{(n-1)}(s', K, j') + \lambda m_{s+1,i}^{(n)}(s', K, j') \right);$
        while $i < K - 1$, do
            $i := i + 1;$
            $h_{s,i}^{(n)}(s', K, j') := H_{s,i}(K) \left(n m_{s,i}^{(n-1)}(s', K, j') + \lambda m_{s+1,i}^{(n)}(s', K, j') + Q_{i,i-1}^{[s]} h_{s,i-1}^{(n)}(s', K, j') \right);$
        endwhile;
        $m_{s,i}^{(n)}(s', K, j') := h_{s,i}^{(n)}(s', K, j');$
        while $i > 0$, do
            $i := i - 1;$
            $m_{s,i}^{(n)}(s', K, j') := H_{s,i}(K) Q_{i,i+1}^{[s]} m_{s,i+1}^{(n)}(s', K, j') + h_{s,i}^{(n)}(s', K, j');$
        endwhile;
    endwhile;
endwhile.

Appendix C. Algorithmic solution under the taboo of level $\mathcal{L}(0)$

For fixed values $\theta \in \mathbb{C}$ with $\Re(\theta) \geq 0$ and phases $j' \in \{0, ..., M_K\}$, Algorithm 3 computes the vectors $\hat{\psi}_{s,i}(\theta; s', K, j')$ with $i \in \{1, ..., K-1\}$, $s \in \{1, ..., S\}$ and $s' \in \{l, ..., S\}$, from which the probability $g_S(T|1, i, j; s, K, j')$ in (11) can be evaluated as the Laplace-Stieltjes transform $\hat{\psi}_{(1,i,j)}(\theta; s, K, j') = e^T_{1+M_i}(1+j)\hat{\psi}_{1,i}(\theta; s, K, j')$ at point $\theta = 0$. Algorithm 3 also provides the vectors $\hat{m}^{(0)}_{s,i}(s', K, j') = \hat{\psi}_{s,i}(0; s', K, j')$ and matrices $\hat{H}_{s,i}(K) = \hat{H}_{s,i}(\theta; K)$ used in Algorithm 4.

Algorithm 3. Iterative evaluation of $\hat{\psi}_{s,i}(\theta; s', K, j')$ with $i \in \{1, ..., K-1\}$, $s \in \{1, ..., S\}$ and $s' \in \{l, ..., S\}$.

$s := S + 1;$
$i := 1;$
while $s > 1$, do
    $s := s - 1;$
    $s' := s;$
    $\hat{H}_{s,i}(\theta; K) := \left((\theta + \lambda)I_{1+M_i} - \hat{Q}^{[s]}_{i,i}\right)^{-1};$
    while $i < K - 1$, do
        $i := i + 1;$
        $\hat{H}_{s,i}(\theta; K) := \left((\theta + \lambda)I_{1+M_i} - \hat{Q}^{[s]}_{i,i} - \hat{Q}^{[s]}_{i,i-1}\hat{H}_{s,i-1}(\theta; K)\hat{Q}^{[s]}_{i-1,i}\right)^{-1};$
    endwhile;
    $\hat{\psi}_{s,i}(\theta; s, K, j') := \hat{H}_{s,i}(\theta; K)\hat{Q}^{[s]}_{i,i+1}e_{1+M_{i+1}}(1 + j');$
    while $i > 1$, do
        $i := i - 1;$
        $\hat{\psi}_{s,i}(\theta; s, K, j') := \hat{H}_{s,i}(\theta; K)\hat{Q}^{[s]}_{i,i+1}\hat{\psi}_{s,i+1}(\theta; s, K, j');$
    endwhile;
    while $s' < S$, do
        $s' := s' + 1;$
        $\hat{h}_{s,i}(\theta; s', K, j') := \lambda\hat{H}_{s,i}(\theta; K)\hat{\psi}_{s+1,i}(\theta; s', K, j');$
        while $i < K - 1$, do
            $i := i + 1;$
            $\hat{h}_{s,i}(\theta; s', K, j') := \hat{H}_{s,i}(\theta; K)\left(\lambda\hat{\psi}_{s+1,i}(\theta; s', K, j') + \hat{Q}^{[s]}_{i,i-1}\hat{h}_{s,i-1}(\theta; s', K, j')\right);$
        endwhile;
        $\hat{\psi}_{s,i}(\theta; s', K, j') := \hat{h}_{s,i}(\theta; s', K, j');$
        while $i > 1$, do
            $i := i - 1;$
            $\hat{\psi}_{s,i}(\theta; s', K, j') := \hat{H}_{s,i}(\theta; K)\hat{Q}^{[s]}_{i,i+1}\hat{\psi}_{s,i+1}(\theta; s', K, j') + \hat{h}_{s,i}(\theta; s', K, j');$
        endwhile;
    endwhile;
endwhile.

Algorithm 4. Iterative evaluation of $\hat{m}^{(n)}_{s,i}(s',K,j')$ with $i \in \{1,...,K-1\}$, $s \in \{1,...,S\}$ and $s' \in \{l,...,S\}$, for a fixed phase $j' \in \{0,...,M_K\}$.

$s := S+1;$

$i := 1;$

while $s > 1$, do

  $s := s-1;$

  $s' := s;$

  $\hat{h}^{(n)}_{s,i}(s',K,j') := n\hat{H}_{s,i}(K)\hat{m}^{(n-1)}_{s,i}(s',K,j');$

  while $i < K-1$, do

    $i := i+1;$

    $\hat{h}^{(n)}_{s,i}(s',K,j') := \hat{H}_{s,i}(K)\left(n\hat{m}^{(n-1)}_{s,i}(s',K,j') + \hat{Q}^{[s]}_{i,i-1}\hat{h}^{(n)}_{s,i-1}(s',K,j')\right);$

  endwhile;

  $\hat{h}^{(n)}_{s,i}(s',K,j') := \hat{h}^{(n)}_{s,i}(s',K,j') + \hat{H}_{s,i}(K)\hat{Q}^{[s]}_{K-1,K}e_{1+M_K}(1+j');$

  $\hat{m}^{(n)}_{s,i}(s',K,j') := \hat{h}^{(n)}_{s,i}(s',K,j');$

  while $i > 1$, do

    $i := i-1;$

    $\hat{m}^{(n)}_{s,i}(s',K,j') := \hat{H}_{s,i}(K)\hat{Q}^{[s]}_{i,i+1}\hat{m}^{(n)}_{s,i+1}(s',K,j') + \hat{h}^{(n)}_{s,i}(s',K,j');$

  endwhile;

  while $s' < S$, do

    $s' := s'+1;$

    $\hat{h}^{(n)}_{s,i}(s',K,j') := \hat{H}_{s,i}(K)\left(n\hat{m}^{(n-1)}_{s,i}(s',K,j') + \lambda\hat{m}^{(n)}_{s+1,i}(s',K,j')\right);$

    while $i < K-1$, do

      $i := i+1;$

      $\hat{h}^{(n)}_{s,i}(s',K,j') := \hat{H}_{s,i}(K)\left(n\hat{m}^{(n-1)}_{s,i}(s',K,j') + \lambda\hat{m}^{(n)}_{s+1,i}(s',K,j') + \hat{Q}^{[s]}_{i,i-1}\hat{h}^{(n)}_{s,i-1}(s',K,j')\right);$

    endwhile;

    $\hat{m}^{(n)}_{s,i}(s',K,j') := \hat{h}^{(n)}_{s,i}(s',K,j');$

    while $i > 1$, do

      $i := i-1;$

      $\hat{m}^{(n)}_{s,i}(s',K,j') := \hat{H}_{s,i}(K)\hat{Q}^{[s]}_{i,i+1}\hat{m}^{(n)}_{s,i+1}(s',K,j') + \hat{h}^{(n)}_{s,i}(s',K,j');$

    endwhile;

  endwhile;

endwhile.

Appendix D. Sub-matrices $Q_{i,i'}(t)$ in the $M_t/G/\infty$ queue with Erlang service times and sinusoidal arrivals

For integers $i' \in \{\max\{0, i-1\}, i, i+1\}$ and $i \in \mathbb{N}_0$, sub-matrices $Q_{i,i'}(t)$ of process $\mathcal{X}$ in Subsection 5.2 are given by

$$Q_{i,i-1}(t) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \alpha & 0 & \cdots & 0 & 0 \\ 0 & 2\alpha & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & (i-1)\alpha & 0 \\ 0 & 0 & \cdots & 0 & i\alpha \end{pmatrix},$$

$$Q_{i,i}(t) = \begin{pmatrix} -(\beta(t)+i\alpha) & i\alpha & 0 & \cdots & 0 & 0 \\ 0 & -(\beta(t)+i\alpha) & (i-1)\alpha & \cdots & 0 & 0 \\ 0 & 0 & -(\beta(t)+i\alpha) & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -(\beta(t)+i\alpha) & \alpha \\ 0 & 0 & 0 & \cdots & 0 & -(\beta(t)+i\alpha) \end{pmatrix},$$

$$Q_{i,i+1}(t) = \begin{pmatrix} \beta(t) & 0 & 0 & \cdots & 0 & 0 \\ 0 & \beta(t) & 0 & \cdots & 0 & 0 \\ 0 & 0 & \beta(t) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta(t) & 0 \end{pmatrix}.$$

Appendix E. Sub-matrices $Q_{i,i'}(t)$ in the SIR epidemic model with sinusoidal forcing of transmission

Under the assumption that $I(0) = N_I$ and $J(0) = N_J$ with $N_I + N_J = N$, the q-matrix of process $\mathcal{X}$ in Subsection 5.3 is specified from the following sub-matrices $Q_{i,i'}(t)$, for integers $i' \in \{\max\{0, i-1\}, i, \min\{i+1, N\}\}$:

$$
\begin{aligned}
Q_{i,i-1}(t) &= \begin{cases} \gamma i I_{N_J+1}, & \text{if } i \in \{1, ..., N_I\}, \\ (\gamma i I_{N-i+1}, 0_{N-i+1}), & \text{if } i \in \{N_I+1, ..., N\}, \end{cases} \\
Q_{i,i}(t) &= \operatorname{diag}(-q_{(i,0)}(t), -q_{(i,1)}(t), ..., -q_{(i,\min\{N_J, N-i\})}(t)), \\
Q_{i,i+1}(t) &= \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \frac{\beta(t)}{N} i & 0 & \cdots & 0 & 0 \\ 0 & \frac{\beta(t)}{N} 2i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \frac{\beta(t)}{N} N_J i & 0 \end{pmatrix}, \quad i \in \{1, ..., N_I - 1\}, \\
Q_{i,i+1}(t) &= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \frac{\beta(t)}{N} i & 0 & \cdots & 0 \\ 0 & \frac{\beta(t)}{N} 2i & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{\beta(t)}{N} (N-i) i \end{pmatrix}, \quad i \in \{N_I, ..., N - 1\},
\end{aligned}
$$

where $q_{(i,j)}(t) = (N^{-1}\beta(t)j + \gamma)i$.

Declarations

Funding

This research was supported by grant PID2021-125871NB-I00 funded by MICIU/AEI/10.13039/501100011033/ and by ERDF, EU.

Conflict of interest

The authors declare no conflict of interest.

Competing interests

A Gómez-Corral is a member of the editorial board of this journal, Methodology and Computing in Applied Probability.

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors consent for publication.

Data availability

This study did not generate or make use of any data set.

Materials availability

Further details on the numerical work in Section 5 can be provided by the corresponding author upon reasonable request.

Code availability

Appendix A—Appendix C include codes for the algorithms used to derive numerical results in Section 5.

Author contribution

AGC and MLG made substantial contributions to the conception and design of the work. AGC, MLG, FPR and DT drafted this work and critically reviewed it for important intellectual content. FPR and DT made substantial contributions to the creation of new software. AGC and MLG wrote the main part of the original draft of the manuscript. There were multiple rounds of revision and editing by AGC, MLG, FPR, and DT.

References

[1] Arns M, Buchholz P, Panchenko A. On the numerical analysis of inhomogeneous continuous-time Markov chains. *INFORMS J Comput.* 2010;22:416–432. doi:10.1287/ijoc.1090.0357

[2] Asmussen S, Avram F, Usabel M. Erlangian approximations for finite-horizon ruin probabilities. *ASTIN Bull.* 2002;32:267–281. doi:10.2143/AST.32.2.1029

[3] Asmussen S. Erlangization/Canadization of phase-type jump diffusions, with applications to barrier options. *J Indian Soc Probab Stat.* 2024;25:575–596. doi:10.1007/s41096-024-00209-5

[4] Bookbinder J. Multiple queues of aircraft under time-dependent conditions. *INFOR.* 1986;24:280–288. doi:10.1080/03155986.1986.11732012

[5] Bright L, Taylor PG. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Commun Stat Stoch Models.* 1995;11:497–525. doi:10.1080/15326349508807357

[6] Bright LW, Taylor PG. Equilibrium distributions for level-dependent quasi-birth-and-death processes. In: Chakravarthy SR, Alfa AS, eds. *Matrix Analytic Methods in Stochastic Models: Proceedings of the 1st International Conference*. New Jersey: Marcel Dekker; 1997:359–375.

[7] Britton T, Lindholm M. The early stage behaviour of a stochastic SIR epidemic with term-time forcing. *J Appl Probab.* 2009;46:975–992. doi:10.1239/jap/1261670683

[8] Buonomo B, Chitnis N, d'Onofrio A. Seasonality in epidemic models: A literature review. *Ricerche Mat.* 2018;67:7–25. doi:10.1007/s11587-017-0348-6

[9] Burak MR, Korytkowski P. Inhomogeneous CTMC birth-and-death models solved by uniformization with steady-state detection. *ACM T Model Comp S.* 2020;30:18. doi:10.1145/3373758

[10] Carr P. Randomization and the American put. *Rev Fin Stud.* 1998;11:597–626. doi:10.2139/ssrn.715

[11] Di Crescenzo A, Gómez-Corral A, Taipe D. A computational approach to extreme values and related hitting probabilities in level-dependent quasi-birth-death processes. *Math Comput Simulat.* 2025;228:211–224. doi:10.1016/j.matcom.2024.08.019

[12] Dietz K. Epidemics and rumours: A survey. *J R Stat Soc Ser A-G.* 1967;130:505–528. doi:10.2307/2982521

[13] Dietz K. The incidence of infectious diseases under the influence of seasonal fluctuations. In: Berger J, Bühler WJ, Repges R, Tautu P, eds., *Mathematical Models in Medicine. Lecture Notes in Biomathematics*, vol. 11. Berlin, Heidelberg: Springer; 1976:1–15. doi:10.1007/978-3-642-93048-5_1

[14] Dong J, Whitt W. Using a birth-and-death process to estimate the steady-state distribution of a periodic queue. *Nav Res Log.* 2015;62:664–685. doi:10.1002/nav.21672

[15] Dushoff J, Plotkin SE, Levin SA, Earn DJD. Dynamical resonance can account for seasonality of influenza epidemics. *Proc Natl Acad Sci U.S.A.* 2004;48:16915–16916. doi:10.1073/pnas.0407293101

[16] Esquível ML, Krasii NP, Guerreiro GR. Estimation–calibration of continuous-time non-homogeneous Markov chains with finite state space. *Mathematics.* 2024;12:668. doi:10.3390/math12050668

[17] Gill RD, Johansen S. A survey of product-integration with a view toward application in survival analysis. *Ann Stat.* 1990;18:1501–1555. doi:10.1214/aos/1176347865

[18] Giorno V, Nobile AG. On a class of birth-death processes with time-varying intensity functions. *Appl Math Comput.* 2020;379:125255. doi:10.1016/j.amc.2020.125255

[19] Giorno V, Nobile AG. On some integral equations for the evaluation of first-passage-time densities of time-inhomogeneous birth-death processes. *Appl Math Comput.* 2022;422:126993. doi:10.1016/j.amc.2022.126993

[20] Giorno V, Nobile AG. Time-inhomogeneous finite birth processes with applications in epidemic models. *Mathematics.* 2023;11:4521. doi:10.3390/math11214521

[21] Grassly NC, Fraser C. Seasonal infectious disease epidemiology. *Proc R Soc. B* 2006;273:2541–2550. doi:10.1098/rspb.2006.3604

[22] Green LV, Kolesar PJ. A note on approximating peak congestion in $M_t/G/\infty$ queues with sinusoidal arrivals. *Manage Sci.* 1998;44:S137–S144. doi:10.1287/mnsc.44.11.S137

[23] Green LV, Kolesar PJ, Soares J. Improving the SIPP approach for staffing service systems that have cyclic demands. *Oper Res.* 2001;49:549–564. doi:10.1287/opre.49.4.549.11228

[24] Grier N, Massey WA, McKoy T, Whitt W. The time-dependent Erlang loss model with retrials. *Telecommun Syst.* 1997;7:253–265. doi:10.1023/A:1019176413237

[25] He QM, Liu B, Wu H. Continuous approximations of discrete phase-type distributions and their applications to reliability models. *Perform Evaluation.* 2022;154:102284. doi:10.1016/j.peva.2022.102284

[26] Heidergott B, Leahu H, Löpker A, Pflug G. Perturbation analysis of inhomogeneous finite Markov chains. *Ad Appl Prob.* 2016;48:255–273. doi:10.1017/apr.2015.16

[27] Inoue Y. A new approach to computing the transient-state probabilities in time-inhomogeneous Markov chains. *J Oper Res Soc Jpn.* 2022;65:48–66. doi:10.15807/jorsj.65.48

[28] Jagerman DL. An inversion technique for the Laplace transform with application to approximation. *The Bell System Technical Journal.* 1978;57:669–710. doi:10.1002/j.1538-7305.1978.tb00601.x

[29] Jagerman DL. An inversion technique for the Laplace transform. *The Bell System Technical Journal.* 1982;61:1995–2002. doi:10.1002/j.1538-7305.1982.tb03096.x

[30] Kendall DG. On the generalized birth-and-death process. *Ann Math Stat.* 1948;19:1–15. doi:10.1214/aoms/1177730285

[31] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond. A* 1927;115:700–721. doi:10.1098/rspa.1927.0118

[32] Kuraya H, Masuyama H, Kasahara S. Load distribution performance of super-node based peer-to-peer communication networks: A nonstationary Markov chain approach. *Numer Algebr Control Opt.* 2011;1:593–610. doi:10.3934/naco.2011.1.593

[33] Mandelbaum A, Massey WA, Reiman M, Stolyar A. Queue lengths and waiting times for multiserver queues with abandonments and retrials. *Telecommun Syst.* 2002;21:149–171. doi:10.1023/A:1020921829517

[34] Margolius BH. A sample path analysis of the $M_t/M_t/c$ queue. *Queueing Syst.* 1999;31:59–93. doi:10.1023/A:101914592789

[35] Margolius BH. Transient solution to the time-dependent multiserver Poisson queue. *J Appl Prob.* 2005;42:766–777. doi:10.1239/jap/1127322026

[36] Margolius BH. Periodic solution to the time-inhomogeneous multi-server Poisson queue. *Oper Res Lett.* 2007;35:125–138. doi:10.1016/j.orl.2005.12.008

[37] Margolius BH. Transient and periodic solution to the time-inhomogeneous quasi-birth death process. *Queueing Syst.* 2007;56:183–194. doi:10.1007/s11134-007-9027-8

[38] Margolius BH. The matrices R and G of matrix analytic methods and the time-inhomogeneous periodic quasi-birth-and-death process. *Queueing Syst.* 2008;60:131–151. doi:10.1007/s11134-008-9090-9

[39] Margolius B. Asymptotic estimates for queueing systems with time-varying periodic transition rates. In: Andrews GE, Krattenthaler C, Krinik A, eds.,

Lattice Path Combinatorics and Applications. Berlin: Springer; 2019:307–326. doi:10.1007/978-3-030-11102-1_14

[40] Margolius B. Eulerian polynomials and quasi-birth-death processes with time-varying-periodic rates. *Contemp Math.* 2021;774:175–193. doi:10.1090/conm/774/15572

[41] Margolius BH. The periodic steady-state solution for queues with Erlang arrivals and service and time-varying periodic transition rates. *Queueing Syst.* 2023;103:45–94. doi:10.1007/s11134-022-09851-x

[42] Neuts MF, Li JM. An algorithmic study of S-I-R stochastic epidemic models. In: Heyde CC, Prohorov YV, Pyke R, Rachev ST, eds. *Athens Conference on Applied Probability and Time Series Analysis*, Lecture Notes in Statistics, vol. 114. New York: Springer; 1996:295–306. doi:10.1007/978-1-4612-0749-8_21

[43] Ong KL, Taaffe MR. Approximating nonstationary $PH(t)/PH(t)/1/c$ queueing systems. *Math Comput Simul.* 1988;30:441–452. doi:10.1016/0378-4754(88)90057-2

[44] Ramaswami V, Woolford DG, Stanford DA. The erlangization method for Markovian fluid flows. *Ann Oper Res.* 2008;160:215–225. doi:10.1007/s10479-008-0309-2

[45] Rindos A, Woolet S, Viniotis I, Trivedi K. Exact methods for the transient analysis of nonhomogeneous continuous time Markov chains. In: Stewart WJ, ed, *Computations with Markov Chains*. Boston: Kluwer Academic Publishers; 1995:121–133. doi:10.1007/978-1-4615-2241-6_8

[46] Satin Y, Korotysheva A, Kiseleva K, Shilova G, Fokicheva E, Zeifman A, Korolev V. Two-sided truncations of inhomogeneous birth-death processes. In: Claus T, Herrmann F, Manitz M, Rose O, eds. *Proceedings of the 30th European Conference on Modelling and Simulation*. ECMS Digital Library; 2016:663-668. doi:10.7148/2016-0663

[47] Satin Y, Razumchik R, Usov H, Zeifman A. Numerical computation of distributions in finite-state inhomogeneous continuous time Markov chains, based on ergodocity bounds and piecewise constant approximation. *Mathematics.* 2023;11:4265. doi:10.3390/math11204265

[48] Schwarz JA, Selinka G, Stolletz R. Performance analysis of time-dependent queueing systems: Survey and classification. *Omega.* 2016;63:170–189. doi:10.1016/j.omega.2015.10.013

[49] Slavík A. Product Integration. Its History and Applications. Matfyzpress, Praha, 2007. eudml.org/doc/202418

[50] Stanford DA, Latouche G, Woolford DG, Boychuk D, Hunchak A. Erlangized fluid queues with application to uncontrolled fire perimeter. *Stoch Models.* 2005;21:631–642. doi:10.1081/STM-200056242

[51] Taaffe MR, Ong KL. Approximating nonstationary $PH(t)/M(t)/s/c$ queueing systems. *Ann Oper Res.* 1987;8:103–116. doi:10.1007/BF02187085

[52] Tripathi S, Duda A. Time-dependent analysis of queueing systems. *INFOR.* 1986;24:199–220. doi:10.1080/03155986.1986.11732006

[53] van Dijk NM. Uniformization for nonhomogeneous Markov chains. *Oper Res Lett.* 1992;12:283–291. doi:10.1016/0167-6377(92)90086-I

[54] van Dijk NM, van Brummelen SPJ, Boucherie RJ. Uniformization: Basics, extensions and applications. *Perform Evaluation.* 2018;118:8–32. doi:10.1016/j.peva.2017.09.008

[55] van Moorsel APA, Wolter K. Numerical solution of non-homogeneous Markov processes through uniformization. In: Zobel RN, Moeller DPFM, eds., Proceedings of the 12th European Simulation MultiConference. Simulation – Past, Present and Future. Manchester: SCS Europe; 1998:710–717.

[56] Wu H, He QM. Optimizing Erlangization-based approximations for finite discrete distributions and discrete phase-type distributions. *Stoch Models.* 2023;1–29. doi:10.1080/15326349.2023.2222463

[57] Yoshioka H, Tsujimura M. Hamilton–Jacobi–Bellman–Isaacs equation for rational inattention in the long-run management of river environments under uncertainty. *Comput Math Appl.* 2022;112:23–54. doi:10.1016/j.camwa.2022.02.013

[58] Yoshioka H, Tanaka T, Aranishi F. Limit equations of adaptive Erlangization and their application to environmental management. *Comput Math Appl.* 2023;146:271–293. doi:10.1016/j.camwa.2023.07.003

[59] Zeifman A, Leorato S, Orsingher E, Satin Ya, Shilova G. Some universal limits for nonhomogeneous birth and death processes. *Queueing Syst.* 2006;52:139–151. doi:10.1007/s11134-006-4353-9

[60] Zhang J, Coyle EJ. The transient solution of time-dependent $M/M/1$ queues. *IEEE Trans Inform Theory.* 1991;37:1690–1696. doi:10.1109/18.104335

[61] Zheng Q. Note on the non-homogeneous Prendiville process. *Math Biosci.* 1998;148:1-5. doi:10.1016/S0025-5564(97)10009-8