J. Math. Biol. (2013) 67:799–831
DOI 10.1007/s00285-012-0570-5

Mathematical Biology

Stochastic epidemic models with random environment: quasi-stationarity, extinction and final size

J. R. Artalejo · A. Economou · M. J. Lopez-Herrero

Received: 20 December 2011 / Revised: 9 July 2012 / Published online: 15 August 2012
© Springer-Verlag 2012

Abstract We investigate stochastic SIS and SIR epidemic models, when there is a random environment that influences the spread of the infectious disease. The inclusion of an external environment into the epidemic model is done by replacing the constant transmission rates with dynamic rates governed by an environmental Markov chain. We put emphasis on the algorithmic evaluation of the influence of the environmental factors on the performance behavior of the epidemic model.

Keywords Stochastic epidemic · Random environment · Quasi-stationarity · Extinction time

Mathematics Subject Classification (2000) 92D30 · 60J22

1 Introduction

Many epidemiological models employ time-dependent parameters (e.g. periodic contact rates) for a better understanding of the infectious disease dynamics. In particular, seasonal patterns have been largely observed in a range of diseases including influenza,

J. R. Artalejo (✉)
Department of Statistics and Operations Research, Faculty of Mathematics,
Complutense University of Madrid, 28040 Madrid, Spain
e-mail: jesus_artalejo@mat.ucm.es

A. Economou
Department of Mathematics, University of Athens, Panepistemiopolis, 15784 Athens, Greece
e-mail: aeconom@math.uoa.gr

M. J. Lopez-Herrero
School of Statistics, Complutense University of Madrid, 28040 Madrid, Spain
e-mail: lherrero@estad.ucm.es

childhood infections (e.g. measles) and vector-borne diseases (e.g. malaria). For a classification of infectious diseases by mechanism of seasonality and timing of outbreak, we refer to the survey paper by Altizer et al. (2006). The existing literature shows that seasonal changes are due to a variety of causes such as changing contact patterns of social behavior, influence of climatic variables and changing due to the control measures taken during an outbreak of the disease. Detailed discussion including an identification of the mechanisms that explain the potential causes of seasonality can be found in Altizer et al. (2006), Grassly and Fraser (2006) and Keeling and Rohani (2007). Despite the efforts already made, the study of potential causes and consequences of seasonal patterns is still a challenging topic for epidemiologists. A detailed review of the role of the seasonality in the dynamics of infectious diseases is not our objective here, but a few general bibliographical comments are presented in what follows.

It should be pointed out that seasonality patterns are often related to changes in the environmental factors (e.g. temperature, humidity and other climate characteristics). In this paper, an *environment* is understood as any external factor that influences the parameters of the epidemic model. Our revision of the literature reveals that most of the existing work has been done for models where the changes in the environment are deterministic. For example, it is commonly assumed that the disease is transmitted at time t according to a sinusoidal function of the form $\beta(t) = \beta_0(1+\omega\cos 2\pi t)$; that is, the contact rate varies along the time but not randomly. In contrast, as far as the authors know, models with stochastic changes of the environment have received little attention.

Mathematical models with deterministic environmental rates are formulated following both deterministic approaches based on differential equations and difference equations (e.g. Bacaër and Gomes 2009; Stone et al. 2007; Wesley and Allen 2009) and stochastic approaches based on Markov chains and branching and diffusion processes (e.g. Prajneshu et al. 1986; Trapman et al. 2004). Mixed formulations integrating deterministic and stochastic techniques have also been considered in a number of papers (e.g. Kuske et al. 2007; Parham and Michael 2011). Another classification of this literature can be done on the basis of discrete-time models (e.g. Franke and Yakubu 2006; Metcalf et al. 2009) versus continuous-time models (e.g. most of the previous cited works).

On the other hand, among those biological models where the environmental factors vary randomly, we mention a recent paper by Britton and Lindholm (2009) where the random environment of an SIR stochastic epidemic is modeled in terms of a cyclically alternating renewal process with a finite number of environmental states. A related example is the work by Steinsaltz et al. (2011) for stage-structured populations driven by a random environment which forms a discrete-time Markov chain. Also, another related reference is the paper by Gray et al. (2012) where the effect of a Markovian environment on the deterministic SIS model is discussed.

Realistic mathematical models for infectious diseases are often analytically intractable due to non-linearity inherent to the transmission rates. In this sense, the inclusion of random environmental mechanisms makes the mathematical models even more complicated. For example, it is well-known (Bacaër 2007; Wesley and Allen 2009) that the interpretation of the transmission factor is strongly affected by the seasonal conditions (i.e., the environmental state in progress) at the time when the

infection is introduced. Thus, a compromise between realistic assumptions and mathematical tractability is needed.

The objective of this paper is to provide further insight of infection disease transmission in stochastic epidemic models driven by random environments governed by a continuous-time Markov chain (CTMC). More concretely, here we concentrate on the SIS and SIR epidemic models operating in random varying environment conditions. Since the classic SIS and SIR models form the basis of many more complex epidemic models, we think that our study is relevant in its own right, but emphasize that the approach can be extended to more sophisticated but realistic models.

We next identify some specific objectives addressed in our study.

(a) We employ a Markov modulated environmental mechanism rather than a deterministic environmental formulation. This random approach drives environmental changes that determine all the transmission rates. In this sense, we extend most of the previous studies where typically the environmental factors only influence the contact rates.
(b) For epidemic models subject to environmental influence, it is important to be able to compute a meaningful transmission factor. We provide a definition of the transmission factor which is a natural extension of the analog quantity for the epidemic models with constant transmission rates (see Appendix 6.1). Its computational calculation is very simple and stable.
(c) We perform an exhaustive analysis of the main measures of the spread and persistence of the infectious disease. To this end, our investigation covers the detailed analysis of the quasi-stationary distribution, the extinction time and the final size of the epidemic.
(d) We develop two average approximations, which are proved to be efficient when the rates of the CTMC driven the environment evolve either essentially faster, or slower, than the epidemic transmission rates.
(e) Our attention is focused on developing efficient methods for the algorithmic analysis of stochastic epidemic models with random environment. We show how algebraic computations of the most relevant epidemic indicators can be reduced to deal only with non-negative matrices of small order.

The organization of the rest of the paper is as follows. In Sect. 2, we first introduce the mathematical formulation of the SIS epidemic model with random environment. Then, we concentrate on the quasi-stationary distribution and the extinction time. We develop algorithmic schemes for their analysis, making emphasis on the computation of the quasi-stationary vector as well as the Laplace transform and the moments of the extinction time. Section 3 presents a parallel analysis for the SIR epidemic model subject to random environmental changes. Since the quasi-stationary distribution does not assign positive mass to all transient states, we supplement the analysis by investigating the alternative provided by the ratio of expectations distribution (RE-distribution). Some numerical experiments that illustrate the results and shed light on the influence of the environmental changes are presented in Sect. 4. The implications of our results are discussed in Sect. 5. Several auxiliary results, including our definition for the environmental transmission factor, are presented in Appendix.

2 The SIS epidemic model with random environment

2.1 Model description

The SIS model with random environment (abbreviated as SIS-re) is a CTMC $X = \{(I(t), E(t)); t \geq 0\}$ on the state space $S = \{0, \ldots, N\} \times S_E$, where the environmental mechanism is governed by an irreducible CTMC $X_E = \{E(t); t \geq 0\}$, with state space $S_E = \{1, \ldots, E\}$ and infinitesimal generator $\mathbf{Q}_E$ with rates $q_E(e, e')$, for $1 \leq e, e' \leq E$. The sojourn time at state $e \in E$ is exponentially distributed with rate $q_E(e) = -q_E(e, e) = \sum_{e' \neq e} q_E(e, e')$. In the SIS-re stochastic model a closed homogeneous population of size N is considered. At time t, the population consists of $I(t)$ infected individuals and $S(t) = N - I(t)$ susceptible individuals. When the environment is in state $e \in S_E$, the infection is transmitted at rate $\lambda_i(e)$ as a result of an effective contact between one of the current i infective individuals and a susceptible individual. Accordingly, the recovery rate is $\mu_i(e)$ as long as $(I(t), E(t)) = (i, e)$. As usual in the SIS paradigm, the SIS-re model assumes that a recovered individual does not acquire immunity.

The infection ends when $I(t) = 0$. Thus, the state space S can be decomposed into the absorbing macro-state $S_A = \{(0, e); 1 \leq e \leq E\}$ and the set of transient states $S_T = S - S_A$ with cardinality $g = NE$. Since S_T is a finite irreducible set, absorption occurs in a finite time with probability 1. The infinitesimal transition rates of the Markov chain X are as follows

$$q_{(i,e)(i',e')} = \begin{cases} \lambda_i(e), & \text{if } (i', e') = (i + 1, e), \\ \mu_i(e), & \text{if } (i', e') = (i - 1, e), \\ q_E(e, e'), & \text{if } (i', e') = (i, e'), e' \neq e, \\ -q_{(i,e)}, & \text{if } (i', e') = (i, e), \\ 0, & \text{otherwise,} \end{cases}$$

where $q_{(i,e)} = \lambda_i(e) + \mu_i(e) + q_E(e)$, for $(i, e) \in S$, and $\lambda_0(e) = \lambda_N(e) = \mu_0(e) = 0$, for $e \in S_E$.

The state space S can be partitioned as $S = \cup_{i=0}^{N} l(i)$, where $l(i)$ is the i-th level defined by $l(i) = \{(i, e); e \in S_E\}$, for $0 \leq i \leq N$. Then, we may express the infinitesimal generator $\mathbf{Q}$ of X in the block tridiagonal form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{00} & \mathbf{Q}_{01} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{10} & \mathbf{Q}_{11} & \mathbf{Q}_{12} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Q}_{N-1,N-2} & \mathbf{Q}_{N-1,N-1} & \mathbf{Q}_{N-1,N} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_{N,N-1} & \mathbf{Q}_{NN} \end{pmatrix},$$

where the blocks $\mathbf{Q}_{ii'}$ are of dimension $E \times E$ and $\mathbf{0}$ is a matrix of zeroes.

The specification of the non-zero blocks $\mathbf{Q}_{ii'}$ describing the motion from level $l(i)$ to level $l(i')$ is as follows

$$\begin{aligned}
\mathbf{Q}_{i,i+1} &= \mathbf{D}_i^{\lambda} = diag(\lambda_i(1), \ldots, \lambda_i(E)), \quad 0 \le i \le N-1,\\
\mathbf{Q}_{i,i-1} &= \mathbf{D}_i^{\mu} = diag(\mu_i(1), \ldots, \mu_i(E)), \quad 1 \le i \le N,\\
\mathbf{Q}_{ii} &= \mathbf{Q}_E - (1-\delta_{i0})\mathbf{D}_i^{\mu} - (1-\delta_{iN})\mathbf{D}_i^{\lambda}, \quad 0 \le i \le N,
\end{aligned}$$

where δ_{ab} stands for the Kronecker's function defined by 1, when $a = b$, and it equals 0, otherwise.

For each state $(i, e) \in S$, in analogy with the rates of the classical SIS model without random environment, we may consider the following infection and recovery rates

$$\begin{aligned}
\lambda_i(e) &= \frac{\beta_e}{N} i(N-i),\\
\mu_i(e) &= \gamma_e i.
\end{aligned}$$

where β_e is the effective contact rate and γ_e is the individual recovery rate, given that $E(t) = e$. Then, the blocks $\mathbf{Q}_{ii'}$ become

$$\begin{aligned}
\mathbf{Q}_{i,i+1} &= \frac{i(N-i)}{N}\mathbf{D}^{\beta} = \frac{i(N-i)}{N} diag(\beta_1, \ldots, \beta_E), \quad 0 \le i \le N-1,\\
\mathbf{Q}_{i,i-1} &= i\mathbf{D}^{\gamma} = i\, diag(\gamma_1, \ldots, \gamma_E), \quad 1 \le i \le N,\\
\mathbf{Q}_{ii} &= \mathbf{Q}_E - i\mathbf{D}^{\gamma} - \frac{i(N-i)}{N}\mathbf{D}^{\beta}, \quad 0 \le i \le N.
\end{aligned}$$

2.2 The quasi-stationary regime

The stationary distribution of the SIS-re stochastic model is degenerate in the sense that only the absorbing macro-state $S_A = l(0)$ has positive probability mass. The quasi-stationary distributions have been widely used to measure the distribution of a Markov chain given that the absorption has not yet taken place. In the SIS-re model, the quasi-stationary probabilities, u_{ie}, for $1 \le i \le N$ and $e \in E$, are defined as

$$u_{ie} = \lim_{t\to\infty} P\{(I(t), E(t)) = (i, e) \,|T > t\}, \quad (i, e) \in S_T,$$

where $T = \inf\{t \ge 0 \,|X(t) \in S_A\}$ denotes the absorption time.

We notice that the set of transient states S_T is finite and irreducible. Then, it is well known (Darroch and Seneta 1967) that the quasi-stationary distribution is the unique normalized left eigenvector corresponding to the eigenvalue with maximal real part $-\alpha$ of the sub-generator $\mathbf{Q}_T$ associated with the transient states. We remark that α is real and positive, so α can also be viewed as the eigenvalue with minimal absolute value. This result provides a natural option for numerical computations.

Let $\mathbf{u}$ be the row vector containing the quasi-stationary probabilities partitioned according to the levels $l(i)$; that is, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iE})$, for $1 \le i \le N$. In what follows, we present an algorithm for the numerical computation of the quasi-stationary vector $\mathbf{u}$. Our algorithm is a level reduction algorithm that exploits the block tridiagonal structure of the generator $\mathbf{Q}$ in combination with the

inverse power method. A distinguishing feature of the inverse power method is that the underlying algebraic technique is primarily concerned with the determination of a single eigenvector. This is our case because we are only interested in finding $\mathbf{u}$ and its associated eigenvalue $-\alpha$, rather that all the eigenvectors of the matrix $\mathbf{Q}_T$. In this context, the inverse power method, also called inverse iteration, is recommended in the specialized literature (Wilkinson 1988; Ciarlet 1989) as the most useful method.

The general idea of the inverse power method is to compute the desired eigenvector $\mathbf{u}$ by using the recursive scheme

$$\begin{aligned} \widehat{\mathbf{u}}^{(n)}\mathbf{Q}_T &= \mathbf{u}^{(n-1)}, \\ \mathbf{u}^{(n)} &= \widehat{\mathbf{u}}^{(n)}/\max(\widehat{\mathbf{u}}^{(n)}), \end{aligned}$$

for $n \geq 1$, until some stopping criterion.

Algorithm 1

Step 1. Choose an initial non-zero guess $\mathbf{u}^{(0)} = (\mathbf{u}_1^{(0)}, ..., \mathbf{u}_N^{(0)})$.

Step 2. Set $\mathbf{Q}_N^* = \mathbf{Q}_{NN}$.

Step 3. For $i = N-1, N-2, ..., 1$ do

$$\mathbf{Q}_i^* = \mathbf{Q}_{ii} + \mathbf{Q}_{i,i+1}\left(-\mathbf{Q}_{i+1}^*\right)^{-1}\mathbf{Q}_{i+1,i}.$$

Step 4. For $i = 1, 2, ..., N-1$ do

$$\mathbf{R}_i = \mathbf{Q}_{i,i+1}\left(-\mathbf{Q}_{i+1}^*\right)^{-1}, \ \mathbf{G}_i = \left(-\mathbf{Q}_{i+1}^*\right)^{-1}\mathbf{Q}_{i+1,i}.$$

Step 5. For $n = 1, 2, ...$ till convergence

5.1. Set $\mathbf{v}_N^{(n)} = \mathbf{u}_N^{(n-1)}$.

5.2. For $i = N-1, N-2, ..., 1$ do

$$\mathbf{v}_i^{(n)} = \mathbf{u}_i^{(n-1)} + \mathbf{v}_{i+1}^{(n)}\mathbf{G}_i.$$

5.3. Set $\widehat{\mathbf{u}}_1^{(n)} = \mathbf{v}_1^{(n)}\left(\mathbf{Q}_1^*\right)^{-1}$.

5.4. For $i = 1, 2, ..., N-1$ do

$$\widehat{\mathbf{u}}_{i+1}^{(n)} = \mathbf{v}_{i+1}^{(n)}\left(\mathbf{Q}_{i+1}^*\right)^{-1} + \widehat{\mathbf{u}}_i^{(n)}\mathbf{R}_i.$$

5.5. Scale $\widehat{\mathbf{u}}^{(n)}$ by setting $\mathbf{u}^{(n)} = \widehat{\mathbf{u}}^{(n)}/\max(\widehat{\mathbf{u}}^{(n)})$.

Step 6. If the stopping criterion is satisfied, then compute

$$\begin{aligned} \mathbf{u} &\simeq \mathbf{u}^{(n)}/\mathbf{u}^{(n)}\mathbf{e}_g, \\ -\alpha &\simeq (\max(\widehat{\mathbf{u}}^{(n)}))^{-1}. \end{aligned}$$

The notation $\mathbf{e}_g$ stands for the column vector of order g of 1s. We use the notation $\max(\widehat{\mathbf{u}}^{(n)})$ to denote the element of maximum modulus of the vector $\widehat{\mathbf{u}}^{(n)}$.

A detailed proof of the algorithm is omitted here, but we point out that the block tridiagonal structure of the generator $\mathbf{Q}$ of the SIS-re model has been exploited to facilitate the use of Gaussian elimination. A typical assumption to apply the power method is that the matrix under consideration is diagonalizable. However, it is not a

necessary assumption. In the case of the matrix $\mathbf{Q}_T$, the eigenvalue $-\alpha$ has algebraic multiplicity 1 and it suffices to use the power method.

For general discussion and bibliographical coverage on other methods for the calculation of eigenvalues and eigenvectors (e.g. the Jacobi method, the Givens-Householder method, the LR and QR algorithms), we refer again to the books by Wilkinson (1988) and Ciarlet (1989). In particular, the later reference could be useful for those readers interested in technical details related to the proof of Algorithm 1.

2.3 Extinction time

The extinction time quantifies the spread of the epidemic on the whole population and describes the time until the end of the epidemic process. The extinction time T is indeed the absorption time of X, which was already defined in Sect. 2.2. In Appendix 6.2, we summarize some results for the extinction time which are valid in the general framework of a CTMC on a finite state space. Those results provide closed-form solutions for the distribution of T and its moments, but they require to deal with inverses and powers of matrices of dimension $g = NE$ having positive and negative entries, which is a source of numerical instability. Thus, our objective in this section is to exploit the block tridiagonal structure of the SIS-re stochastic model in order to obtain simple recursive schemes that involve small matrices of dimension E with non-negative entries.

For every initial state $(i, e) \in S_T$, we define the conditional extinction time T_{ie}, with Laplace transform $\varphi_{ie}(s) = E\left[e^{-sT_{ie}}\right]$ and k-th moment $m_{ie}^k = E\left[T_{ie}^k\right]$, for $k \geq 0$. By using a first-step argument, we find that

$$(s + \lambda_i(e) + \mu_i(e) + q_E(e))\,\varphi_{ie}(s) = \mu_i(e)\varphi_{i-1,e}(s) + (1 - \delta_{iN})\lambda_i(e)\varphi_{i+1,e}(s) + \sum_{\substack{e'=1 \\ e' \neq e}}^{E} q_E(e, e')\varphi_{ie'}(s),\ (i, e) \in S_T.$$

Moreover, $\varphi_{0e}(s) = 1$, for $e \in S_E$. In matrix form, the system yields

$$\left(s\mathbf{I}_E + \mathbf{D}_i^{\lambda} + \mathbf{D}_i^{\mu} - \mathbf{Q}_E\right)\boldsymbol{\varphi}_i(s) = \mathbf{D}_i^{\mu}\boldsymbol{\varphi}_{i-1}(s) + (1 - \delta_{iN})\mathbf{D}_i^{\lambda}\boldsymbol{\varphi}_{i+1}(s), \quad 1 \leq i \leq N,$$

where the vectors $\boldsymbol{\varphi}_i(s) = (\varphi_{i1}(s), \ldots, \varphi_{iE}(s))'$, for $0 \leq i \leq N$, comprise the Laplace transforms $\varphi_{ie}(s)$ partitioned according to the orbit levels $l(i)$. The notation $\mathbf{I}_E$ refers to the identity matrix of order E.

The starting value of the density function $f_{T_{ie}}(x)$ follows from the Tauberian result $f_{T_{ie}}(0) = \lim_{s\to\infty} s\varphi_{ie}(s) = \delta_{i1}\mu_i(e)$.

By differentiating k times the Laplace transforms with regard to s and setting $s = 0$, we obtain the following block tridiagonal system for the moments:

$$\left(\mathbf{D}_i^{\lambda} + \mathbf{D}_i^{\mu} - \mathbf{Q}_E\right)\mathbf{m}_i^k = \mathbf{D}_i^{\mu}\mathbf{m}_{i-1}^k + (1 - \delta_{iN})\mathbf{D}_i^{\lambda}\mathbf{m}_{i+1}^k + k\mathbf{m}_i^{k-1}, \quad 1 \leq i \leq N, k \geq 1,$$

where $\mathbf{m}_i^k = \left(m_{i1}^k, \ldots, m_{iE}^k\right)'$, for $0 \leq i \leq N$, and $\mathbf{m}_0^k = \mathbf{0}_E$, for $k \geq 1$, with $\mathbf{m}_i^0 = \mathbf{e}_E$, for $0 \leq i \leq N$. We notice that $\mathbf{0}_E$ denotes the column vector of order E of 0s.

The solution of this system may be obtained by using a forward elimination backward substitution block algorithm (Ciarlet 1989; Varga 2000) and reduction of some matrices. In this way, we obtain the following algorithm for the recursive computation of the moments m_{ie}^k in terms of the moments m_{ie}^{k-1} of one order less.

Algorithm 2

Step 1. Set $\mathbf{G}_1 = \mathbf{D}_1^{\lambda} + \mathbf{D}_1^{\mu} - \mathbf{Q}_E$.

Step 2. For $i = 2, ..., N$ do

$$\mathbf{G}_i = \mathbf{D}_i^{\lambda} + \mathbf{D}_i^{\mu} - \mathbf{Q}_E - \mathbf{D}_i^{\mu}\mathbf{G}_{i-1}^{-1}\mathbf{D}_{i-1}^{\lambda}.$$

Step 3. Set $\mathbf{w}_1(k) = \mathbf{G}_1^{-1} k \mathbf{m}_1^{k-1}$.

Step 4. For $i = 2, ..., N$ do

$$\mathbf{w}_i(k) = \mathbf{G}_i^{-1}\left(k\mathbf{m}_i^{k-1} + \mathbf{D}_i^{\mu}\mathbf{w}_{i-1}(k)\right).$$

Step 5. Set $\mathbf{m}_N^k = \mathbf{w}_N(k)$.

Step 6. For $i = N-1, ..., 1$ do

$$\mathbf{m}_i^k = \mathbf{w}_i(k) + \mathbf{G}_i^{-1}\mathbf{D}_i^{\lambda}\mathbf{m}_{i+1}^k.$$

It is possible to prove that the matrices $\mathbf{G}_i$, for $1 \leq i \leq N$, are M-matrices. In fact, $\mathbf{G}_i$ has non-positive off-diagonal elements, positive row sums and its inverse $\mathbf{G}_i^{-1}$ exists and has positive entries. As a result, Algorithm 2 provides a simple recursive scheme where only small non-negative matrices are involved in the algebraic calculations. This property is numerically desirable to avoid instability inherent to the simultaneous presence of positive and negative terms. The key to obtain this advantage is to introduce the reductions $\widetilde{\mathbf{G}}_i = \widehat{\mathbf{G}}_i^{-1}$ and $\mathbf{G}_i = -\mathbf{D}_i^{\lambda}\widetilde{\mathbf{G}}_i$, for $1 \leq i \leq N-1$, where $\widehat{\mathbf{G}}_1 = (\mathbf{D}_1^{\lambda} + \mathbf{D}_1^{\mu} - \mathbf{Q}_E)^{-1}(-\mathbf{D}_1^{\lambda})$ and $\widehat{\mathbf{G}}_i = (\mathbf{D}_i^{\lambda} + \mathbf{D}_i^{\mu} - \mathbf{Q}_E + \mathbf{D}_i^{\mu}\widehat{\mathbf{G}}_{i-1})^{-1}(-\mathbf{D}_i^{\lambda})$, for $2 \leq i \leq N-1$, are the matrices associated with the standard formulation of the forward elimination backward substitution block algorithm.

It is not our objective here to establish a comparison among different methods of solution for block systems of equations. However, we would like to mention the existence of iterative methods like the successive overrelaxation block method. Varga (2000) comments that the advantage of the forward elimination backward substitution method is that it is a direct method but in order to improve the matrix storage on the computer one could use the successive overrelaxation method. In such a case, the choice of an appropriate overrelaxation parameter is a key point in order to guarantee a sufficiently fast convergence. Other methods for solving block tridiagonal systems take advantage of symmetry but it is not the case in this paper.

The system of equations for the Laplace transforms may be solved in a similar manner. The following minor modifications are needed: (i) change the notation from $\mathbf{G}_i$, $\mathbf{w}_i(k)$ and $\mathbf{m}_i^k$ by $\mathbf{G}_i(s)$, $\mathbf{w}_i(s)$ and $\boldsymbol{\varphi}_i(s)$, respectively, (ii) add $s\mathbf{I}_E$ to the right-

hand side of the expressions given in Steps 1 and 2, and (iii) replace $k\mathbf{m}_1^{k-1}$ by $\mathbf{D}_1^{\mu}\mathbf{e}_E$, in the Step 3, and $k\mathbf{m}_i^{k-1} + \mathbf{D}_i^{\mu}\mathbf{w}_{i-1}(k)$ by $\mathbf{D}_i^{\mu}\mathbf{w}_{i-1}(s)$, in the Step 4. Our interest in the computation of $\varphi_{ie}(s)$ is that it is required for getting the density function of T_{ie} by numerical inversion (Cohen 2007).

3 The SIR epidemic model with random environment

3.1 Model description

We now describe the SIR-re stochastic model; that is, we consider how the classical SIR assumptions are modified in the presence of random environment factors. The assumption that the total population is closed implies that, at any time t, $S(t) + I(t) + R(t) = N$, where $S(t)$, $I(t)$ and $R(t)$ denote the number of susceptibles, infectives and removals, respectively, at time t. Thus, the system state of the SIR-re model can be described by a CTMC $X = \{(I(t), S(t), E(t)); t \geq 0\}$, with state space $S = \{(i, j); 0 \leq i \leq m + n, 0 \leq j \leq \min(n, m + n - i)\} \times S_E$ and initial state $(I(0), S(0)) = (m, n)$, with $m \geq 1$ and $m + n = N$. The environmental process $\{E(t); t \geq 0\}$ is defined as in Sect. 2.1. If the environmental state is $E(t) = e \in S_E = \{1, \ldots, E\}$, then transitions from a state (i, j), with $i \geq 1$, can be either to state $(i+1, j-1)$ at a rate $\lambda_{ij}(e)$, due to an infection, or to state $(i-1, j)$ at a rate $\mu_i(e)$, due to a removal. We consider that the epidemic ends as soon as $I(t) = 0$, so the set of absorbing states S_A consists of the macro-states $\{(0, j, e); e \in S_E\}$, for $0 \leq j \leq n$, and $\lambda_{0j}(e) = 0$, for $e \in S_E$. In addition, we have some trivial null rates, that is $\lambda_{i0}(e) = \mu_0(e) = 0$, for $e \in S_E$.

The elements of the infinitesimal generator are now given by

$$q_{(i,j,e)(i',j',e')} = \begin{cases} \lambda_{ij}(e), & \text{if } (i', j', e') = (i+1, j-1, e), \\ \mu_i(e), & \text{if } (i', j', e') = (i-1, j, e), \\ q_E(e, e'), & \text{if } (i', j', e') = (i, j, e'), e' \neq e, \\ -q_{(i,j,e)}, & \text{if } (i', j', e') = (i, j, e), \\ 0, & \text{otherwise,} \end{cases}$$

where $q_{(i,j,e)} = \lambda_{ij}(e) + \mu_i(e) + q_E(e)$, for $(i, j, e) \in S$.

We consider that $S(t)$ is the level of process X. Then, we notice that $S = \cup_{j=0}^{n} l(j)$, where the j-th level is given by $l(j) = \{(i, j, e); 0 \leq i \leq m + n - j, e \in S_E\}$, for $0 \leq j \leq n$. The cardinality of $l(j)$ is $|l(j)| = (m + n - j + 1)E$, for $0 \leq j \leq n$. This partition of the state space leads to the following block triangular representation of the generator $\mathbf{Q}$:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{00} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{10} & \mathbf{Q}_{11} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Q}_{n-1,n-2} & \mathbf{Q}_{n-1,n-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}_{n,n-1} & \mathbf{Q}_{nn} \end{pmatrix},$$

where the blocks $\mathbf{Q}_{jj'}$ are of dimension $|l(j)| \times |l(j')|$.

The blocks $\mathbf{Q}_{j,j-1}$, for $1 \leq j \leq n$, correspond to transitions due to infections. By introducing the sub-levels $l(i,j) = \{(i,j,e); e \in S_E\}$, for $0 \leq i \leq m+n-j$, the block $\mathbf{Q}_{j,j-1}$ can be partitioned as $\mathbf{Q}_{j,j-1} = \left(\mathbf{Q}_{(i,j)(i',j-1)}\right)$, where the sub-blocks $\mathbf{Q}_{(i,j)(i',j-1)}$, for $0 \leq i \leq m+n-j$, are matrices of dimension $E \times E$ given by

$$\begin{aligned}\mathbf{Q}_{(i,j)(i+1,j-1)} &= \mathbf{D}^{\lambda}_{ij} = diag(\lambda_{ij}(1), \ldots, \lambda_{ij}(E)),\\ \mathbf{Q}_{(i,j)(i',j-1)} &= \mathbf{0}, \quad 0 \leq i' \leq m+n-j+1, i' \neq i+1.\end{aligned}$$

On the other hand, the blocks $\mathbf{Q}_{jj}$, for $0 \leq j \leq n$, are associated with transitions due to recoveries, as well as the environmental transition rates. The blocks $\mathbf{Q}_{jj}$ are structured in sub-block triangular form. Thus, we have $\mathbf{Q}_{jj} = \left(\mathbf{Q}_{(i,j)(i',j)}\right)$, where $\mathbf{Q}_{(i,j)(i',j)}$, for $0 \leq i \leq m+n-j$, are sub-blocks of dimension $E \times E$ defined as follows:

$$\begin{aligned}\mathbf{Q}_{(i,j)(i-1,j)} &= \mathbf{D}^{\mu}_{i} = diag(\mu_i(1), \ldots, \mu_i(E)),\\ \mathbf{Q}_{(i,j)(i,j)} &= \mathbf{Q}_E - \mathbf{D}^{\lambda}_{ij} - \mathbf{D}^{\mu}_{i},\\ \mathbf{Q}_{(i,j)(i',j)} &= \mathbf{0}, \quad 0 \leq i' \leq m+n-j+1, i' \notin \{i-1, i\}.\end{aligned}$$

An appeal to the classical SIR model motivates the consideration in what follows of the following specific choice of the diagonal matrices $\mathbf{D}^{\lambda}_{ij}$ and $\mathbf{D}^{\mu}_{i}$:

$$\begin{aligned}\mathbf{D}^{\lambda}_{ij} &= \frac{ij}{N}\mathbf{D}^{\beta} = \frac{ij}{N} diag(\beta_1, \ldots, \beta_E), \quad 0 \leq j \leq n, 0 \leq i \leq m+n-j,\\ \mathbf{D}^{\mu}_{i} &= i\mathbf{D}^{\gamma} = i\, diag(\gamma_1, \ldots, \gamma_E), \quad 0 \leq i \leq m+n,\end{aligned}$$

where β_e and γ_e respectively denote the contact rate and the recovery rate, given that $E(t) = e$.

3.2 Quasi-stationary and ratio of expectations distributions

We start this section by recalling that the classical SIR model without random environment has a finite reducible set of transient states. This fact influences the quasi-stationary distribution which concentrates all the probability mass in the state $(i,j) = (1,0)$, when $\lambda_{ij} = ij\beta/N$ and $\mu_i = i\gamma$ (Artalejo and Lopez-Herrero 2010; van Doorn and Pollett 2008). Our first objective in this section is to elucidate if the existence of random environment factors modify, or not, this assignment of the quasi-stationary probability mass.

In the SIR-re model each sub-level $l(i,j)$ constitutes a class of communicating states. The set of transient states $S_T = \cup_{j=0}^{n} \cup_{i=1}^{m+n-j} l(i,j)$ is obtained by removing from each level $l(j)$ the first E states associated to the sub-level $l(0,j)$. Thus, the blocks $\mathbf{Q}^{T}_{jj'}$ of the sub-generator $\mathbf{Q}_T$ are obtained by removing the first E rows and the first E columns of $\mathbf{Q}_{jj'}$. Obviously, the resulting matrix $\mathbf{Q}_T$ preserves the block

triangular structure. As a result, it follows that the set of eigenvalues of $\mathbf{Q}_T$ is given by

$$sp\left(\mathbf{Q}_T\right) = \bigcup_{j=0}^{n} sp\left(\mathbf{Q}_{jj}^T\right) = \bigcup_{j=0}^{n} \bigcup_{i=1}^{m+n-j} sp\left(\mathbf{Q}_E - \mathbf{D}_{ij}^{\lambda} - \mathbf{D}_i^{\mu}\right),$$

and $-\alpha = \max\left\{-\alpha_{ij}; 0 \leq j \leq n, 1 \leq i \leq m+n-j\right\}$, where $-\alpha_{ij}$ is the eigenvalue of $\mathbf{Q}_{(i,j)(i,j)} = \mathbf{Q}_E - \mathbf{D}_{ij}^{\lambda} - \mathbf{D}_i^{\mu}$ with maximal real part.

The key point is to determine the sub-level where the maximum is reached. To this end, we define the matrices $\mathbf{A} = -\mathbf{Q}_{(1,0)(1,0)} = \mathbf{D}_1^{\mu} - \mathbf{Q}_E$ and $\mathbf{B}_{ij} = -\mathbf{Q}_{(i,j)(i,j)} = \mathbf{D}_{ij}^{\lambda} + \mathbf{D}_i^{\mu} - \mathbf{Q}_E$, for $0 \leq j \leq n$ and $1 \leq i \leq m+n-j$. For any pair $(i, j) \neq (1, 0)$, we notice that $\mathbf{B}_{ij} = \mathbf{A} + \mathbf{D}_{ij}$, where $\mathbf{D}_{ij} = \mathbf{D}_{ij}^{\lambda} + \mathbf{D}_{i-1}^{\mu}$.

Since $\mathbf{D}_1^{\mu}$ and $\mathbf{D}_{ij}$, with $(i, j) \neq (1, 0)$, are diagonal matrices with positive diagonal entries, the matrices $\mathbf{A}$ and $\mathbf{B}_{ij}$ are strict diagonally dominant matrices. Thus, they are indeed M-matrices and their inverse matrices $\mathbf{A}^{-1}$ and $\mathbf{B}_{ij}^{-1}$ are positive matrices. Moreover, we have that $\mathbf{A} \leq \mathbf{B}_{ij}$ and $\mathbf{A} \neq \mathbf{B}_{ij}$. Therefore, the Perron-Frobenius theorem for non-negative matrices can be applied to obtain that $0 < \rho(\mathbf{B}_{ij}^{-1}) < \rho(\mathbf{A}^{-1})$ (Ciarlet 1989), where $\rho(\mathbf{M})$ denotes the spectral radius of matrix $\mathbf{M}$ (i.e., the largest absolute value attained by any eigenvalue of $\mathbf{M}$).

We now recall that a number $\lambda \in sp(\mathbf{M}) \Longleftrightarrow -\lambda \in sp(-\mathbf{M})$ and $\lambda \in sp(\mathbf{M}) \Longleftrightarrow \lambda^{-1} \in sp(\mathbf{M}^{-1})$. Thus, we notice that the eigenvalue of matrix $-\mathbf{A}$ with maximal real part (respectively $-\mathbf{B}_{ij}$) is $-\rho^{-1}(\mathbf{A}^{-1})$ (respectively $-\rho^{-1}(\mathbf{B}_{ij}^{-1})$). Consequently, $-\alpha_{10} = -\rho^{-1}(\mathbf{A}^{-1}) > -\rho^{-1}(\mathbf{B}_{ij}^{-1}) = -\alpha_{ij}$, for $0 \leq j \leq n$ and $1 \leq i \leq m+n-j$, with $(i, j) \neq (1, 0)$, and we conclude that $\alpha = \alpha_{10}$.

From the general results given in van Doorn and Pollett (2008) for the case where S_T is finite and reducible, it is now easy to conclude that the quasi-stationary probabilities $\mathbf{u} = \left\{u_{ije}; (i, j, e) \in S_T\right\}$ of the SIR-re model only assign positive mass to the states $(1, 0, e)$, for $e \in S_E$. This result provides a natural generalization of the scalar analog for the classical SIR stochastic model.

The sub-vector $\mathbf{u}_{10}$ containing the positive probabilities satisfies

$$\mathbf{u}_{10}\left(\mathbf{Q}_E - \mathbf{D}_1^{\mu}\right) = -\alpha_{10}\mathbf{u}_{10} = -\left(\sum_{e=1}^{E} u_{10e}\gamma_e\right)\mathbf{u}_{10},$$

subject to the normalizing condition $\mathbf{u}_{10}\mathbf{e}_E = 1$. It should be pointed out that the normalized quasi-stationary vector $\mathbf{u}_{10}$ and the environmental stationary vector $\boldsymbol{\pi}_E$ of the generator $\mathbf{Q}_E$ are different (i.e., $\mathbf{u}_{10} \neq \boldsymbol{\pi}_E$).

We now turn our attention to the RE-distribution as an alternative approach for measuring the behavior of the epidemic before the absorption (Artalejo and Lopez-Herrero 2010). The RE-distribution is indeed a variant of the ratio of means studied in Darroch and Seneta (1967). Given the initial state $(I(0), S(0), E(0)) = (m, n, i_0) \in S_T$, we define the ratio

$$P_{(m,n,e_0)}(i, j, e) = \frac{E_{(m,n,e_0)}[T_{ije}^*]}{E_{(m,n,e_0)}[T]}, \quad (i, j, e) \in S_T,$$

where T^*_{ije} is the time that the Markov chain X spends in state $(i, j, e) \in S_T$ before absorption. The denominator represents the expected time to absorption given the initial state (m, n, i_0).

For a fixed state $(\bar{i}, \bar{j}, \bar{e}) \in S_T$, we may derive the equations governing the dynamics of the expectations $E_{(i,j,e)}[T^*_{\bar{i}\bar{j}\bar{e}}]$

$$\begin{aligned}
&\left(\lambda_{ij}(e) + \mu_i(e) + q_E(e)\right) E_{(i,j,e)}[T^*_{\bar{i}\bar{j}\bar{e}}] \\
&= \delta_{(i,j,e)(\bar{i},\bar{j},\bar{e})} + \mu_i(e) E_{(i-1,j,e)}[T^*_{\bar{i}\bar{j}\bar{e}}] + \lambda_{ij}(e) E_{(i+1,j-1,e)}[T^*_{\bar{i}\bar{j}\bar{e}}] \\
&+ \sum_{\substack{e'=1 \\ e' \neq e}}^{E} q_E(e, e') E_{(i,j,e')}[T^*_{\bar{i}\bar{j}\bar{e}}], \ (i, j, e) \in S_T,
\end{aligned}$$

where $E_{(0,j,e)}[T^*_{\bar{i}\bar{j}\bar{e}}] = 0$, for $0 \leq j \leq n$ and $e \in S_E$.

In matrix-form, this system of equations can be written as

$$\begin{aligned}
\left(\mathbf{D}^{\lambda}_{ij} + \mathbf{D}^{\mu}_{i} - \mathbf{Q}_E\right) \mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e}) &= \mathbf{D}^{\mu}_{i} \mathbf{m}_{i-1,j}(\bar{i}, \bar{j}, \bar{e}) \\
&+ \mathbf{D}^{\lambda}_{ij} \mathbf{m}_{i+1,j-1}(\bar{i}, \bar{j}, \bar{e}) + \delta_{(i,j)(\bar{i},\bar{j})} \mathbf{e}_E(\bar{e}),
\end{aligned}$$

where $\mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e}) = \left(E_{(i,j,1)}[T^*_{\bar{i}\bar{j}\bar{e}}], \ldots, E_{(i,j,E)}[T^*_{\bar{i}\bar{j}\bar{e}}]\right)'$, for $0 \leq j \leq n$ and $1 \leq i \leq m + n - j$, and $\mathbf{e}_E(\bar{e})$ is a column vector of order E with zero entries, except for the $\bar{e}$-th one which is equal to 1.

At this point, we observe that the block triangular form of the above system can be used to develop a recursive scheme for the computation of the expectations $\mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e})$. Calculations are performed in the order $0 \leq j \leq n$ and $0 \leq i \leq m+n-j$. For $i = 0$, we use the boundary conditions $\mathbf{m}_{0j}(\bar{i}, \bar{j}, \bar{e}) = \mathbf{0}_E$. Then, for fixed values of j and i, we solve a system of the form $(\mathbf{D} - \mathbf{Q}_E)\mathbf{x} = \mathbf{c}$, where $\mathbf{D} = \mathbf{D}^{\lambda}_{ij} + \mathbf{D}^{\mu}_{i}$, $\mathbf{x} = \mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e})$ and $\mathbf{c} = \mathbf{D}^{\mu}_{i} \mathbf{m}_{i-1,j}(\bar{i}, \bar{j}, \bar{e}) + \mathbf{D}^{\lambda}_{ij} \mathbf{m}_{i+1,j-1}(\bar{i}, \bar{j}, \bar{e}) + \delta_{(i,j)(\bar{i},\bar{j})} \mathbf{e}_E(\bar{e})$. In Appendix 6.3, it is showed that the solution of such a system is stable as only multiplication of small positive matrices are involved.

The recursive method for computing $\mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e})$ is summarized in the following algorithm.

Algorithm 3

Step 1. Set $\mathbf{m}_{00}(\bar{i}, \bar{j}, \bar{e}) = \mathbf{0}_E$.

Step 2. For $i = 1, ..., m + n$ do

$$\mathbf{m}_{i0}(\bar{i}, \bar{j}, \bar{e}) = (\mathbf{D}^{\mu}_{i} - \mathbf{Q}_E)^{-1} \left(\mathbf{D}^{\mu}_{i} \mathbf{m}_{i-1,0}(\bar{i}, \bar{j}, \bar{e}) + \delta_{(i,0)(\bar{i},\bar{j})} \mathbf{e}_E(\bar{e})\right).$$

Step 3. For $j = 1, ..., n$ do

3.1. Set $\mathbf{m}_{0j}(\bar{i}, \bar{j}, \bar{e}) = \mathbf{0}_E$.

3.2. For $i = 1, ..., m + n - j$ do

$$\mathbf{m}_{ij}(\bar{i},\bar{j},\bar{e}) = \left(\mathbf{D}^{\lambda}_{ij} + \mathbf{D}^{\mu}_{i} - \mathbf{Q}_E\right)^{-1} \left(\mathbf{D}^{\mu}_{i}\mathbf{m}_{i-1,j}(\bar{i},\bar{j},\bar{e}) + \mathbf{D}^{\lambda}_{ij}\mathbf{m}_{i+1,j-1}(\bar{i},\bar{j},\bar{e}) + \delta_{(i,j)(\bar{i},\bar{j})}\mathbf{e}_E(\bar{e})\right).$$

For particular cases, including the cyclic environment and the star environment (see Sect. 4), we may obtain explicit formulas for $\mathbf{m}_{ij}(\bar{i},\bar{j},\bar{e})$ from Appendix 6.3 by taking $d_e = \lambda_{ij}(e)+\mu_i(e)$ and $c_e = \mu_i(e)E_{(i-1,j,e)}[T^*_{\bar{i}\bar{j}\bar{e}}]+\lambda_{ij}(e)E_{(i+1,j-1,e)}[T^*_{\bar{i}\bar{j}\bar{e}}]+\delta_{(i,j,e)(\bar{i},\bar{j},\bar{e})}$, for $e \in S_E$.

Finally, summing $E_{(m,n,e_0)}[T^*_{ije}]$ over $(i, j, e) \in S_T$, we obtain $E_{(m,n,e_0)}[T]$. Once the probabilities $P_{(m,n,e_0)}(i, j, e)$ have been computed, we may calculate the unconditional probabilities

$$P_{(m,n)}(i,j) = \sum_{e_0=1}^{E}\left(\sum_{e=1}^{E} P_{(m,n,e_0)}(i,j,e)\right)\pi_{e_0}.$$

At this point, we remark that the RE-distribution gives positive probability to all transient states as opposed to the quasi-stationary mass which is concentrated on the sub-level $l(1, 0)$. In this sense, the quasi-stationary distribution is not appropriate for quantifying an SIR model with acute infections. In general, when an epidemic ends in a short period, the quasi-stationary regime is not reached and the system is quantified better by giving the RE-distribution, which can be employed despite how long the absorption time is. If the absorption time is large (e.g. an SIS model with transmission factor sufficiently large), then the effect of the initial state vanishes as far as the time increases. Then, the RE-distribution and the quasi-stationary distribution are close.

For the SIS model without random environment, Artalejo and Lopez-Herrero (2010) showed that the RE-distributions are stochastically monotone with respect to the initial state. As a result, the RE-distribution can be used to construct approximations of the quasi-stationary distribution. Also, Barbour and Pollett (2010) have investigated how the quasi-stationary distribution can be approximated in total variation by the stationary distribution of a return process, which has the same behavior of the epidemic process while in S_T but, on reaching the absorption, it is instantaneously returned to S_T according to a certain probability distribution (see also Bartlett 1960). We notice that the RE-distribution corresponds to the case where the return process restarts always at a fixed initial transient state.

To conclude this section, we refer the reader to Appendix 6.4 where, for the sake of completeness, we have summarized the computation of the RE-distribution for the SIS-re model.

3.3 Extinction time

We first introduce the random variable T_{ije} representing the extinction time of the SIR-re epidemic, given the current state $(i, j, e) \in S_T$. We also define the corresponding Laplace transforms, $\varphi_{ije}(s)$, and moments, m^k_{ije}, for $k \geq 0$.

By using first-step analysis, we get that $\varphi_{ije}(s)$ satisfy

$$\big(s+\lambda_{ij}(e)+\mu_i(e)+q_E(e)\big)\,\varphi_{ije}(s)=\mu_i(e)\varphi_{i-1,je}(s)+\lambda_{ij}(e)\varphi_{i+1,j-1,e}(s)+\sum_{\substack{e'=1\\ e'\neq e}}^{E} q_E(e,e')\varphi_{ije'}(s),\ (i,j,e)\in S_T.$$

For the absorbing states, we have $\varphi_{0je}(s)=1$, for $0\le j\le n$ and $e\in S_E$. Thus, in matrix-form the system assumes the block triangular form

$$\Big(s\mathbf{I}_E+\mathbf{D}_{ij}^{\lambda}+\mathbf{D}_{i}^{\mu}-\mathbf{Q}_E\Big)\,\boldsymbol{\varphi}_{ij}(s)=\mathbf{D}_{i}^{\mu}\boldsymbol{\varphi}_{i-1,j}(s)+\mathbf{D}_{ij}^{\lambda}\boldsymbol{\varphi}_{i+1,j-1}(s),$$

for $0\le j\le n$ and $1\le i\le m+n-j$, where $\boldsymbol{\varphi}_{ij}(s)=\big(\varphi_{ij1}(s),\ldots,\varphi_{ijE}(s)\big)'$.

The initial value of the density is given by $f_{T_{ije}}(0)=\delta_{i1}\mu_i(e)$.

For determining the moments m_{ije}^k, we first introduce the vector notation $\mathbf{m}_{ij}^k=(m_{ij1}^k,\ldots,m_{ijE}^k)'$ and notice that $\mathbf{m}_{0j}^k=\mathbf{0}_E$, for $k\ge 1$, and $\mathbf{m}_{ij}^0=\mathbf{e}_E$, for $0\le j\le n$ and $0\le i\le m+n-j$. Moreover, after appropriate differentiation of the Laplace transforms, we find that

$$\Big(\mathbf{D}_{ij}^{\lambda}+\mathbf{D}_{i}^{\mu}-\mathbf{Q}_E\Big)\,\mathbf{m}_{ij}^k=\mathbf{D}_{i}^{\mu}\mathbf{m}_{i-1,j}^k+\mathbf{D}_{ij}^{\lambda}\mathbf{m}_{i+1,j-1}^k+k\mathbf{m}_{ij}^{k-1},$$

for $0\le j\le n$, $1\le i\le m+n-j$ and $k\ge 1$.

The recursive scheme for the computation of the k-th moments $\mathbf{m}_{ij}^k$ is similar to that given in Algorithm 3 for $\mathbf{m}_{ij}(\bar{i},\bar{j},\bar{e})$. In fact, the linear systems governing $\mathbf{m}_{ij}^k$ and $\mathbf{m}_{ij}(\bar{i},\bar{j},\bar{e})$ are identical except for the independent term. Calculations are now done in the order $k\ge 0$, $0\le j\le n$ and $0\le i\le m+n-j$. The modifications in the algorithm are as follows: (i) replace $\mathbf{m}_{ij}(\bar{i},\bar{j},\bar{e})$ by $\mathbf{m}_{ij}^k$, and (ii) replace $\delta_{(i,0)(\bar{i},\bar{j})}\mathbf{e}_E(\bar{e})$ by $k\mathbf{m}_{i0}^{k-1}$ in Step 2 and $\delta_{(i,j)(\bar{i},\bar{j})}\mathbf{e}_E(\bar{e})$ by $k\mathbf{m}_{ij}^{k-1}$ in Step 3.2.

Obviously, another similar algorithmic result for the computation of $\boldsymbol{\varphi}_{ij}(s)$ can be readily derived, but its presentation is omitted here.

3.4 Final size of the epidemic

The final size measures the severity of an epidemic by counting the total number of infected individuals including the initial number m. If there are j susceptible individuals when $I(t)$ reaches the state 0, then the final size of the epidemic is $m+n-j$. As a result, the probabilities $p_f(j)$, for $m\le j\le m+n$, associated with the final size of the epidemic can be determined by computing the absorption probabilities of the SIR-re Markov chain.

Let f_{ije} be the absorption probability by the state $(0,j^*,e^*)$, for $0\le j^*\le n$ and $e^*\in S_E$, given that the initial state is $(i,j,e)\in S$. The first-step analysis for the

absorption probabilities yields

$$\begin{aligned}\left(\lambda_{ij}(e)+\mu_i(e)+q_E(e)\right) f_{ije} &= \mu_i(e) f_{i-1,je} + \lambda_{ij}(e) f_{i+1,j-1,e} \\ &+ \sum_{\substack{e'=1 \\ e'\neq e}}^{E} q_E(e,e') f_{ije'},\ (i,j,e)\in S_T,\end{aligned}$$

where the boundary conditions are $f_{0je} = \delta_{(j,e)(j^*,e^*)}$. The knowledge of the environmental state in progress when the absorption is reached is irrelevant with respect to the final size of the epidemic, so we may employ the alternative boundary condition $f_{0je} = \delta_{jj^*}$, for $e \in S_E$, in order to calculate directly the absorption probability associated with the sub-level $l(0, j^*)$.

By defining $\mathbf{f}_{ij} = (f_{ij1}, \ldots, f_{ijE})'$, we may rewrite the system for the absorbing probabilities in matrix-form

$$\left(\mathbf{D}_{ij}^{\lambda} + \mathbf{D}_{i}^{\mu} - \mathbf{Q}_E\right)\mathbf{f}_{ij} = \mathbf{D}_{i}^{\mu}\mathbf{f}_{i-1,j} + \mathbf{D}_{ij}^{\lambda}\mathbf{f}_{i+1,j-1},$$

for $0 \le j \le n$ and $1 \le i \le m+n-j$, where $\mathbf{f}_{0j} = \delta_{jj^*}\mathbf{e}_E$.

Once more, the algorithm for the recursive computation of $\mathbf{f}_{ij}$ is a minor variant of Algorithm 3 in Sect. 3.2. The necessary modifications are: (i) replace $\mathbf{m}_{ij}(\bar{i}, \bar{j}, \bar{e})$ by $\mathbf{f}_{ij}$, (ii) replace the initial conditions in Step 1 and Step 3.1 by $\delta_{0j^*}\mathbf{e}_E$ and $\delta_{jj^*}\mathbf{e}_E$, respectively, and (iii) delete the term $\delta_{(i,j)(\bar{i},\bar{j})}\mathbf{e}_E(\bar{e})$ in Step 2 and Step 3.2.

Since $(I(0), S(0)) = (m, n)$, we may be interested in computing f_{mne}, for $e \in S_E$, and the unconditional absorption probability $f_{mn} = \mathbf{f}_{mn}\boldsymbol{\pi}_E$. The computation of the final size distribution is completed once one solves the system of equations $n+1$ times, one for each choice of the sub-level $(0, j^*)$. Then, we have $p_f(m+n-j^*) = f_{mn}(j^*)$, for $0 \le j^* \le n$, where the notation on the right hand side refers to the unconditional absorption by $l(0, j^*)$.

4 Numerical experiments

In this section, we present a set of numerical results that provide insight on the influence of the random environment on the epidemic descriptors. We structure our results into four sets of experiments. Firstly, we introduce two average approximations of the stochastic model with random environment. These approximations are constructed on the basis of simple constant rate epidemic models. A second set of numerical examples is focused on the quasi-stationary distribution of the SIS-re model. In Sect. 4.3, we turn our attention to the computation of the distribution of the extinction time and its moments. Finally, in Sect. 4.4, we focus on the transient behavior of the number of infectives in an SIS-re model with alternating environment.

An important feature of the models under study is that they are characterized by a large number of system parameters. For example, in the SIS-re we need to specify the population size, N, the environmental transmission rates $\lambda_i(e)$ and $\mu_i(e)$, for $(i, e) \in S$, as well as the choice of the generator $\mathbf{Q}_E$. We next describe two specifications for the

environmental generator. On the other hand, concrete choices of the environmental transmission rates will be specified along the numerical examples.

1. Cyclic environment. This particular choice represents the case where the environment changes cyclically between E different states. The case $E = 2$ corresponds to the alternating environment largely used in stochastic modeling. As a related possibility, we mention the epidemic models with term-time forcing where the contact rate is subject to discontinuous but cyclic changes. In comparison with the term-time forcing model studied by Britton and Lindholm (2009), we reduce to the case where the sojourn time in the environment $e \in S_E$ is exponentially distributed with rate δ_e. However, here we allow the environmental forces to also affect the recovery rates.

The infinitesimal generator $\mathbf{Q}_E$ of the cyclic environment has the following structure:

$$\mathbf{Q}_E = \begin{pmatrix} -\delta_1 & \delta_1 & 0 & \cdots & 0 & 0 \\ 0 & -\delta_2 & \delta_2 & \cdots & 0 & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & -\delta_{E-1} & \delta_{E-1} \\ \delta_E & 0 & 0 & \cdots & 0 & -\delta_E \end{pmatrix}.$$

It can be easily proved that the stationary probabilities $\boldsymbol{\pi}_E = (\pi_1, \ldots, \pi_E)$ of the cyclic environment generator are given by

$$\pi_e = \frac{\frac{1}{\delta_e}}{\sum_{e=1}^{E} \frac{1}{\delta_e}}, \quad e \in S_E.$$

2. Star environment. In the case of a star environment, we suppose that the environment resides typically in an environmental state $e = 1$, but once in a while it deviates to other states as well. After a deviation to one of the non-typical states $e \in S_E - \{1\}$, the environment is regulated and returns to its natural state 1. Hence, the infinitesimal generator $\mathbf{Q}_E$ is given by

$$\mathbf{Q}_E = \begin{pmatrix} -\delta_1 & \delta_1 p_2 & \cdots & \delta_1 p_{E-1} & \delta_1 p_E \\ \delta_2 & -\delta_2 & \cdots & 0 & 0 \\ & \ddots & \ddots & \ddots & \\ \delta_E & 0 & \cdots & 0 & -\delta_E \end{pmatrix},$$

where $q_E(e) = \delta_e$ is the sojourn rate in the environmental state $e \in S_E$, and p_e are the routing environmental probabilities, with $\sum_{e=2}^{E} p_e = 1$.

The stationary probabilities $\boldsymbol{\pi}_E = (\pi_1, \ldots, \pi_E)$ of the generator $\mathbf{Q}_E$ are now as follows:

$$\pi_e = \frac{\frac{p_e}{\delta_e}}{\frac{1}{\delta_1} + \sum_{e=2}^{E} \frac{p_e}{\delta_e}}, \quad e \in S_E, (p_1 = 1).$$

4.1 The two average approximations

The price to be paid by the consideration of a random mechanism governing the environmental changes is to deal with a more complex epidemic model both from the analytical and the computational point of view. Here, we introduce two average approximations based on simple models with constant transmission rates.

1. The average descriptor approximation (ADA). Let us consider E marginal epidemic models of SIS or SIR-type. The e-th marginal model is defined as the epidemic model whose constant transmission rates are equal to the rates associated with the e-th environmental state in the random environment model. The ADA of any epidemic descriptor is then obtained as a weighted average of the corresponding descriptor of the marginal epidemic models, where the weights are determined by the stationary distribution $\boldsymbol{\pi}_E$. For example, given a population with one initial infective, we may construct $E[T_1^{ADA}] = \sum_{e=1}^{E} E[T_1(e)]\pi_e$, where $E[T_1(e)]$ is the expected time to extinction in an SIS epidemic model with constant contact and recovery rates $\lambda_i(e)$ and $\mu_i(e)$, for $0 \leq i \leq N$.

2. The average rate approximation (ARA). New model rates can be defined as an average of the corresponding rates of the marginal epidemic models. For example, an ARA individual recovery rate can be defined as $\gamma_{ARA} = \sum_{e=1}^{E} \gamma_e \pi_e$. The epidemic model with random environment can be approximated by the analog model with constant ARA rates.

We now show how the ADA and ARA are helpful to gain understanding of the stochastic model with random environment. Let us assume that the SIS-re model is under study. The key point here is to compare the rates $q_E(e, e')$ of $\mathbf{Q}_E$ versus the environmental transmission rates. If the rates $q_E(e, e')$ vary slowly in comparison to $\lambda_i(e)$ and $\mu_i(e)$, then it is expected that the e-th marginal model currently under operation may have enough time to approach its stationary regime before the sojourn time at state $e \in S_E$ expires. Thus, intuition suggests that the ADA is reliable when the rates of $\mathbf{Q}_E$ are small enough. On the other hand, the ARA is expected to provide an accurate approximation of the epidemic model with random environment when the rates of $\mathbf{Q}_E$ vary essentially faster than the environmental transmission rates. Then, in average, the random environment model operates with the ARA rates. These intuitive reasonings are confirmed in the following numerical illustrations; see, for example, Fig. 1.

In Table 1, we deal with an SIS-re model with $N = 100$ and $E = 3$. Both the cyclic and the star environment are considered. In the case of the cyclic environment we assume that $\delta_e = \delta(E + 1 - e)^2$, with $\delta = 0.5$, for $e \in S_E$, while the sojourn rates of the star environment have the form $\delta_e = \delta e$, with $\delta = 1.0$, for $e \in S_E$, and $p_2 = 0.75$. The following three scenarios for the environmental effective contact rates and the individual recovery rates are considered:

A) $(\beta_1, \beta_2, \beta_3) = (\beta, \beta + \varepsilon, \beta - \varepsilon) = (1.0, 1.6, 0.4)$ (i.e., $\beta = 1.0$ and $\varepsilon = 0.6$) and $\gamma_e = 0.8$, for $e \in S_E$.
B) $(\beta_1, \beta_2, \beta_3) = (\beta, \beta\varepsilon_2, \beta\varepsilon_3) = (1.0, 0.5, 0.1)$ and $\gamma_e = e$, for $e \in S_E$.

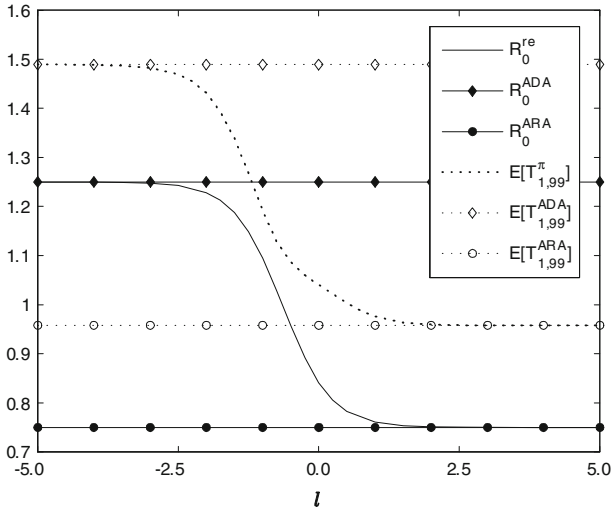

Fig. 1 R_0^{re} and $E[T_1^{\pi}]$ as a function of l

C) $\beta_e = e$ and $\gamma_e = E + 1 - e$, for $e \in S_E$.

The above scenarios cover a number of additive and multiplicative choices for β_e and γ_e. In particular, the deviations ε_2 and ε_3 in scenario B) can be interpreted as the proportion of susceptibles that can be contacted when $e \in S_E$. For each choice of $\mathbf{Q}_E$ and a scenario, we display from top to bottom the following epidemic indicators: the transmission factor R_0, the mean value, $E[T_1]$, and the standard deviation, $\sigma(T_1)$, of the extinction time and the mean value, $E[\mathbf{u}]$, and the standard deviation, $\sigma(\mathbf{u})$, of the quasi-stationary distribution. Notation R_0 should be interpreted as R_0^{re} for the SIS-re model (see Appendix 6.1), as well as $R_0^{ADA} = \sum_{e=1}^{E} R_0(e)\pi_e$, with $R_0(e) = \beta_e/\gamma_e$, and $R_0^{ARA} = \beta^{ARA}/\gamma^{ARA} = \left(\sum_{e=1}^{E} \beta_e\pi_e / \sum_{e=1}^{E} \gamma_e\pi_e\right)$. In the same manner, $E[T_1]$ and $\sigma(T_1)$ refer to the unconditional mean value $E[T_1^{\boldsymbol{\pi}}] = \sum_{e=1}^{E} E[T_{1e}]\pi_e$ and to $\sigma(T_1^{\boldsymbol{\pi}}) = \left(\sum_{e=1}^{E} E[T_{1e}^2]\pi_e - E^2[T_1^{\boldsymbol{\pi}}]\right)^{1/2}$ for the SIS-re model, but $E[T_1]$ and $\sigma(T_1)$ also denote the ADA and ARA, as they were described above. We also employ in the table the common notation $E[\mathbf{u}]$ and $\sigma(\mathbf{u})$ to refer both to the quasi-stationary moments of the SIS-re model as well as to their ADA and ARA.

An examination of the results in the table reveals the following interesting observations:

- In the scenario A, the transmission factor is the same for the SIS-re model and the two average approximations. We notice that, in this case, $\gamma_e = 0.8$, for $e \in S_E$. In fact, when the individual recovery rate is a constant (i.e., $\gamma_e = \gamma$, for $e \in S_E$), it is easy to prove that $R_0^{re} = R_0^{ADA} = R_0^{ARA} = \gamma^{-1} \sum_{e=1}^{E} \beta_e\pi_e$.
- For $R_0 < 1$, it is observed that the ADA and ARA characteristics provide upper and lower bounds of the SIS-re epidemic descriptor. However, in the case $R_0 > 1$, many entries show that the SIS- re descriptor takes values out of the ADA-ARA bounds. This behavior will be analyzed later on (see Fig. 6).

Table 1 The ADA and ARA approximations

	$Cyclic\,\mathbf{Q}_E$			$Star\,\mathbf{Q}_E$		
	ADA	$SIS\text{-}re$	ARA	ADA	$SIS\text{-}re$	ARA
Scenario A						
R_0	0.83673	0.83673	0.83673	1.40000	1.40000	1.40000
$E[T_1]$	14390458.5	2.70093	2.56222	20146650.2	70.74177	143.4922
$\sigma(T_1)$	66342664.9	4.31283	3.84495	77755519.1	177.4597	365.4481
$E[\mathbf{u}]$	11.78318	6.28859	4.16672	24.27685	25.12884	25.50697
$\sigma(\mathbf{u})$	18.54468	6.42979	3.29550	16.93759	10.90144	9.07557
Scenario B						
R_0	0.15204	0.10931	0.09307	0.75190	0.65001	0.59791
$E[T_1]$	0.59768	0.41379	0.39542	2.20976	1.21751	1.09825
$\sigma(T_1)$	1.77571	0.43989	0.40510	4.71958	1.62469	1.37111
$E[\mathbf{u}]$	1.57704	1.25432	1.10115	5.22130	3.45789	2.32181
$\sigma(\mathbf{u})$	2.23346	0.64951	0.33337	5.09645	3.08514	1.70269
Scenario C						
R_0	2.41496	2.15495	1.96969	0.65714	0.56103	0.52173
$E[T_1]$	5.33×10^{17}	79.23375	22523070.7	4.14×10^{16}	0.57791	0.53362
$\sigma(T_1)$	9.41×10^{17}	142.3715	40011759.2	2.98×10^{17}	0.75532	0.63946
$E[\mathbf{u}]$	50.01302	46.02433	48.12469	6.60436	3.91364	2.00513
$\sigma(\mathbf{u})$	27.44199	20.15906	7.29586	15.16916	5.30271	1.39254

- In the scenario C, the magnitude of the ADA value of $E[T_1]$ and $\sigma(T_1)$ is of order 10^{17}. This is observed for both the cyclic and the star environments. This extremely large value is caused by the individual contribution of the 3rd marginal SIS model, where $R_0(3) = \beta_3/\gamma_3 = 3.0$. Then, $E[T_1(3)] = 7.2\times10^{17}$. Moreover, $\pi_3 = 36/49$ (cyclic case) so the contribution of this marginal environment to $E[T_1^{ADA}]$ is very significant.
- The ARA values of $E[T_1]$ and $\sigma(T_1)$ in scenario C for the cyclic environment are also quite large. Now the explanation is related to R_0^{ARA}, which is equal to 1.96969.

The next Fig. 1 supplements Table 1. We now consider an $SIR\text{-}re$ model with initial state $(m, n) = (1, 99)$ operating under the alternating environment

$$\mathbf{Q}_E = \begin{pmatrix} -4\times10^{l} & 4\times10^{l} \\ 10^{l} & -10^{l} \end{pmatrix}, \quad l \in (-5, 5)\,,$$

where the parameter l can be interpreted as the speed of the environmental changes in a logarithmic scale.

The alternating environment can be viewed as the stochastic analog of the case where the deterministic contact rate $\beta(t)$ is approximated as β^+ in the high season and as β^- in the low season (see, e.g. Stone et al. 2007). In the deterministic model the seasons change sequentially leading to periodicity dynamics. On the other hand, the alternating stochastic process $X_E = \{E(t); t \geq 0\}$ has piecewise constant paths

in agreement with the switching between the two environmental states. However, the periodicity is no longer preserved when we deal with the transient probabilities $p_{ij}(t) = P\{E(t) = j \,|E(0) = i\}, 1 \leq i, j \leq 2$ (Kulkarni 1995). The conclusion is that the possibility to capture periodicity dynamics is an important difference between the deterministic and the stochastic alternating models.

The environmental transmission rates are chosen as $(\beta_1, \beta_2) = (5.75, 0.25)$ and $(\gamma_1, \gamma_2) = (1.0, 2.0)$. Then, we observe that $\boldsymbol{\pi}_E = (0.2, 0.8)$, $R_0^{ADA} = 1.25$ and $R_0^{ARA} = 0.75$, while the two average approximations of $E[T_{1,99}]$ are given by $E[T_{1,99}^{ADA}] = 1.48903$ and $E[T_{1,99}^{ARA}] = 0.95787$. R_0^{re} and $E[T_{1,99}^{\boldsymbol{\pi}}]$ have been represented in the figure as a function of the rate speed $l \in (-5, 5)$. In agreement with our previous comments, it is observed that the ADA is very accurate if $l < -3$; that is, the SIR-re epidemic indicators converge to the ADA values when the environmental rates in $\mathbf{Q}_E$ are slow. The convergence to the ARA values is a bit faster. For $l > 2$, we may conclude that the SIR-re descriptors are close enough to the ARA values.

4.2 The alternating SIS model: quasi-stationary characteristics

In this example we deal with the SIS-re model with alternating environment

$$\mathbf{Q}_E = \begin{pmatrix} -\delta_1 & \delta_1 \\ \delta_2 & -\delta_2 \end{pmatrix},$$

where $\delta_2 = 1.0$ and $\delta_1 = 0.1, 0.25, 0.5, 1.0, 2.0, 4.0$ and 10.0, so $\boldsymbol{\pi}_E = \left(\frac{1}{\delta_1+1}, \frac{\delta_1}{\delta_1+1}\right)$, with $\pi_1 = 0.\widehat{90}, 0.80, 0.\widehat{6}, 0.5, 0.\widehat{3}, 0.2$ and $0.\widehat{09}$. In addition, we assume that the population size is $N = 100$ and choose the environmental transmission rates as $\beta_e = \beta e$ and $\gamma_e = 3 - e$, for $1 \leq e \leq 2$, with $\beta = 0.1, 0.25, 0.5, 1.0, 2.0, 4.0$ and 10.0.

In Table 2, we display the main characteristics of the quasi-stationary distribution. For each pair (β, δ_1), we give from top to bottom R_0^{re}, $E[\mathbf{u}]$, $\sigma(\mathbf{u})$ and the modes of $\mathbf{u}$. The transmission factor is indeed explicit and it is given by $R_0^{re} = \frac{\beta}{\delta_1+4}\left(2 + \frac{\delta_1(2\delta_1+5)}{\delta_1+1}\right)$.

A summary of the observations inferred from the table is as follows:

- The expected value $E[\mathbf{u}]$ increases for increasing values of β and δ_1. In particular, π_2 increases with δ_1, which implies to increase the time that the process spends in $E(t) = 2$. Since $R_0(2) = 2\beta$ quadruples the value of $R_0(1) = \beta/2$, the increasing behavior of $E[\mathbf{u}]$ is explained.
- The intuition says that the quasi-stationary mass tends to be concentrated around 1 or N, as far as $\beta \to 0$ or $\beta \to \infty$, respectively. As a result, the standard deviation $\sigma(\mathbf{u})$ has a maximum as a function of β. As a function of δ_1, the behavior of $\sigma(\mathbf{u})$ follows a variety of different patterns.
- For $R_0^{re} < 1$, entries in the table show that the mode of the quasi-stationary distribution is $m_1 = 1$. On the other hand, for $R_0^{re} \geq 1$, we observe a more interesting behavior including bimodal distributions. This fact is analyzed in Figs. 2 and 3.

Table 2 Alternating SIS model: R_0^{re} and quasi-stationary characteristics

δ_1	β						
	0.1	0.25	0.5	1	2	4	10
0.1	0.06031	0.15077	0.30155	0.60310	1.20620	2.41241	6.03104
	1.11444	1.39254	2.54690	6.82234	14.94212	52.99791	81.20279
	0.37062	0.83064	2.70782	10.02041	16.21953	12.38348	6.09348
	1	1	1	1	1, 73	50, 87	80, 95
0.25	0.07294	0.18235	0.36470	0.72941	1.458 82	2.91764	7.29411
	1.13099	1.43984	2.68585	8.87048	22.87064	57.81034	82.94914
	0.39670	0.87245	2.70564	11.26271	20.72745	15.31173	7.19352
	1	1	1	1	5, 72	51, 87	80, 95
0.5	0.08888	0.22222	0.44444	0.88888	1.77777	3.55555	8.88888
	1.14515	1.48407	2.85181	11.23683	32.32164	63.56716	85.07768
	0.41734	0.90933	2.74960	12.35854	23.00425	16.68684	7.84014
	1	1	1	1	13, 72	52, 87	81, 95
1	0.11000	0.27500	0.55000	1.10000	2.20000	4.40000	11.00000
	1.16053	1.53675	3.10048	15.00478	44.44030	70.50620	87.72335
	0.43864	0.95299	2.86346	13.63320	22.42759	16.10367	7.83000
	1	1	1	1	25, 72	55, 87	81, 95
2	0.13333	0.33333	0.66666	1.33333	2.66666	5.33333	13.33333
	1.17700	1.59943	3.47559	21.32503	56.30874	77.03065	90.33572
	0.46075	1.00639	3.08052	14.67732	18.32548	13.36929	6.94779
	1	1	1	1	73	63, 87	82, 95
4	0.15500	0.38750	0.77500	1.55000	3.10000	6.20000	15.50000
	1.19402	1.67159	4.02336	30.60085	64.75987	81.78430	92.36805
	0.48328	1.07028	3.42354	13.88896	12.98414	9.69063	5.49030
	1	1	1	35	73	87	95
10	0.17662	0.44155	0.88311	1.76623	3.53246	7.06493	17.66233
	1.21476	1.76960	4.98178	41.06915	70.65337	85.16557	93.92275
	0.51057	1.16047	4.01841	10.02224	7.86886	5.85166	3.65005
	1	1	1	43	73	87	95

In Fig. 2, we plot the marginal quasi-stationary probabilities of the number of infected individuals (i.e., $u_i = u_{i1} + u_{i2}$, for $1 \leq i \leq 100$) in the case $\delta_1 = 1.0$ and $\beta = 0.5, 1.0, 2.0$ and 10.0. For small values of β the transmission factor is also small, then one might expect a small epidemic, so the quasi-stationary distribution is unimodal and the mode is at the origin; that is, $m_1 = 1$. As far as β increases, R_0^{re} also increases and the epidemic becomes larger. Then, the two observed modes $m_1 < m_2$ are explained by the alternating switching between the transmission factors $R_0(1) < R_0(2)$. An increment in the value of β implies a shift to the right of the modes. For $\delta_1 = 1.0$, we notice that $\pi_1 = \pi_2 = 0.5$ and $R_0(1) < R_0(2)$, so the peak of mode m_2 is higher than the peak of m_1; in other words, $u_{m_1} < u_{m_2}$.

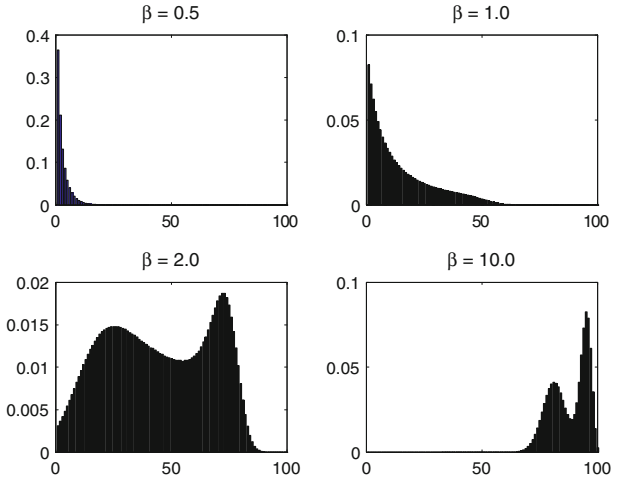

Fig. 2 Quasi-stationary distribution of infectives versus β

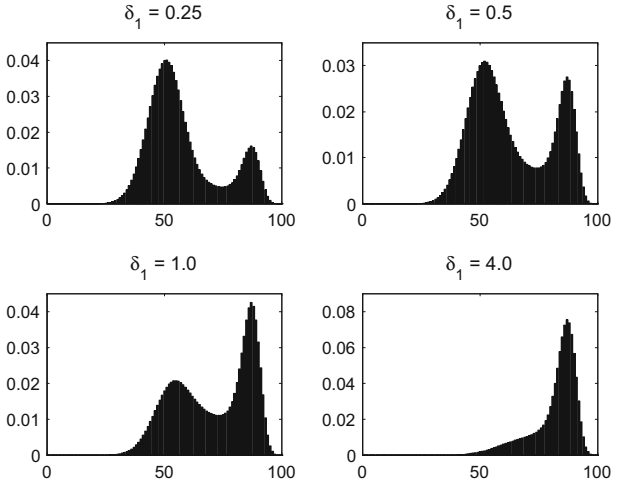

Fig. 3 Quasi-stationary distribution of infectives versus δ_1

Figure 3 is dual to Fig. 2. Now we fix $\beta = 4.0$ and plot the quasi-stationary probabilities for $\delta_1 = 0.25, 0.5, 1.0$ and 4.0. For $\delta_1 = 0.25$, we have $\pi_1 = 0.8$ and $\pi_2 = 0.2$. Since the state $e = 1$ is more frequently visited, we observe that the highest peak corresponds to the mode m_1. However, π_2 increases as a function of δ_1. Thus, the peaks are firstly inverted (case $\delta_1 = 1.0$) and finally only one mode exists (case $\delta_1 = 4.0$).

4.3 Time to extinction analysis

In this section, we present numerical illustrations regarding the computation of the density function and the moments of the extinction time in SIS-re and SIR-re mod-

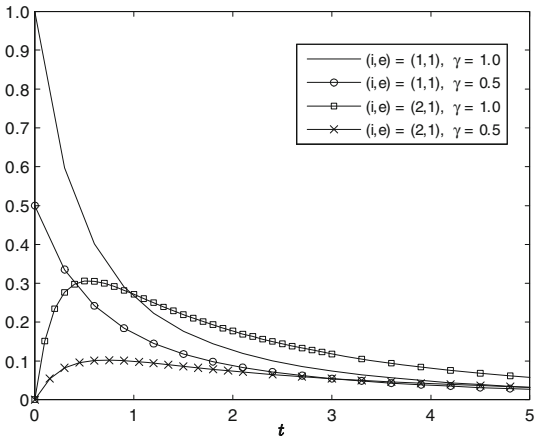

Fig. 4 Densities of T_{i1} in an SIS-re model

els. We have implemented the computation of the density function following two approaches. A first possibility is to use a scaling and squaring method to compute a Padé approximation of the underlying matrix exponential (see Appendix 6.2 and Moler and van Loan 2003). The use of Fourier series methods for the numerical inversion of the Laplace transforms gives another possibility (Cohen 2007). Our initial motivation to employ two different approaches was to perform a parallel computation in order to check correctness and accuracy by agreement of the two methods.

The two approaches were implemented in double precision Fortran 90 routines in a personal computer of 2.40 GHz and 2GB RAM. The CPU time was shorter when we use numerical inversion of Laplace transforms. This is probably due to the fact that our implementation of the Padé method was based on a more cumbersome multiple link of Fortran subroutines. For further discussion, we refer the reader to our previous work in Artalejo et al. (2012) and to the survey paper by Moler and van Loan (2003).

We first focus on an SIS-re model with star environment, where $N = 100$, $E = 3$, $(\delta_1, \delta_2, \delta_3) = (0.5, 0.25, 0.25)$ and $p_2 = 0.1$. The stationary environmental distribution is $\boldsymbol{\pi}_E = \left(\frac{1}{3}, \frac{1}{15}, \frac{3}{5}\right)$. In addition, we choose the contact rates as $(\beta_1, \beta_2, \beta_3) = (1.0, 1.5, 0.5)$. We also consider two constant choices for the recovery rates: $A)\gamma = 1.0$ (scenario A), and $B)\gamma = 0.5$ (scenario B). For each scenario, in Fig. 4 we plot the density functions $f_{T_{ie}}(x)$ of the extinction time when the initial states are $(i, e) = (1, 1)$ and $(i, e) = (2, 1)$.

We observe that $f_{T_{1e}}(0) = \gamma$, while $f_{T_{2e}}(0) = 0$; that is, the numerical inversion confirms the Tauberian result $f_{T_{ie}}(0) = \delta_{i1}\mu_i(e)$ given in Sect. 2.3. For a given initial state, we see that the density with the heaviest tail is associated with the recovery rate $\gamma = 0.5$. This fact is in agreement with the expected value of the extinction times, which are given by $E[T_{11}] = 2.13903$ and $E[T_{21}] = 3.42768$, in the scenario A, and by $E[T_{11}] = 62.15860$ and $E[T_{21}] = 99.58247$, in the scenario B.

It is known that the density of the extinction time in the classical SIR epidemic model may exhibit a bimodal shape (Artalejo et al. 2012). In Fig. 5, the analysis of the bimodal nature of the extinction time is extended to the context of an SIR-re model with a cyclic environment consisting of $E = 4$ environments and $\delta_e = \delta e$, for $e \in S_E$.

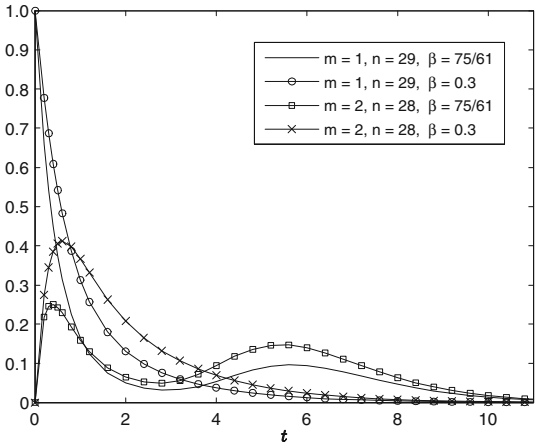

Fig. 5 Densities of T_{mn1} in an SIR-re model

We assume that $E(0) = 1$, $(m, n) = (1, 29)$ and $(2, 28)$, $\beta_e = \beta e$ and $\gamma_e = 1.0$, for $e \in S_E$. For this choice of the model parameters, we have $R_0^{re} = \frac{48}{25}\beta$. For each initial state (m, n), we consider two possibilities: $A)\beta = 75/61$ (scenario A), and $B)$ $\beta = 0.3$ (scenario B).

The densities plotted in the figure show that the case $\beta = 75/61$ leads to bimodal densities, while curves associated with $\beta = 0.3$ have only one mode close to the origin. Moreover, once more, we observe that the value of $f_{T_{mn1}}(0) = \delta_{m1}$ agrees with the Tauberian result. The shape of the densities can be interpreted as follows. When $\beta = 0.3$ ($R_0^{re} = 0.57600$) one might expect a minor epidemic, so $f_{T_{mn1}}(x)$ is unimodal. However, for $\beta = 75/61$ ($R_0^{re} = 2.36065$) and small values of m, either a small or a major epidemic can occur. This explains the observed bimodal shape.

In what follows, we turn our attention to the moments of the extinction time. To this end, in the next example we consider an SIS-re model with $N = 100$, $E = 4$ and a cyclic environment with $\delta_e = \delta e$, for $e \in S_E$. The stationary distribution of such an environment is $\boldsymbol{\pi}_E = \left(\frac{12}{25}, \frac{6}{25}, \frac{4}{25}, \frac{3}{25}\right)$. We assume that $\beta_e = \beta e$ and $\gamma_e = E + 1 - e$, for $e \in S_E$. Then, in Table 3, we report in each cell from top to down the values of R_0^{re}, $E[T_1^{\boldsymbol{\pi}}]$ and $\sigma(T_1^{\boldsymbol{\pi}})$, when δ and β take the values 0.1, 0.25, 0.5, 1.0, 2.0, 4.0 and 10.0.

In fact, the transmission factor follows the explicit formula $R_0^{re} = \beta \frac{480\delta^3+960\delta^2+648\delta+120}{770\delta^3+1345\delta^2+770\delta+120}$, so R_0^{re} is increasing with β, but it is easy to check that it is a decreasing function of δ. Moreover, we point out that $\lim_{\delta\to 0} R_0^{re} = R_0^{ADA} = \beta$ and $\lim_{\delta\to\infty} R_0^{re} = R_0^{ARA} = \frac{48}{77}\beta$.

With respect to $E[T_1^{\boldsymbol{\pi}}]$ and $\sigma(T_1^{\boldsymbol{\pi}})$, we observe the following features:

- For each fixed δ, we notice that $E[T_1^{\boldsymbol{\pi}}]$ and $\sigma(T_1^{\boldsymbol{\pi}})$ are increasing functions of the contact rate β.
- For $\beta \leq 1.0$, both descriptors decrease as far as the sojourn rate δ increases. However, for $\beta > 1.0$, we observe that $E[T_1^{\boldsymbol{\pi}}]$ and $\sigma(T_1^{\boldsymbol{\pi}})$ have a minimum, as they are considered as functions of δ. This behavior is analyzed in more detail in Fig. 6.

Table 3 Cyclic SIS-re model: R_0^{re} and moments of the extinction time

δ	β						
	0.1	0.25	0.5	1.0	2.0	4.0	10.0
0.1	0.09226	0.23065	0.46131	0.92263	1.84527	3.69055	9.22639
	0.41289	0.47685	0.64833	1.16349	2.75265	7.88163	6.99×10^{12}
	0.54078	0.75750	1.29616	2.49023	5.45834	12.75698	8.97×10^{12}
0.25	0.08553	0.21384	0.42768	0.85537	1.71074	3.42149	8.55372
	0.39187	0.43122	0.52484	0.83766	2.01056	7.72499	8.41×10^{12}
	0.46345	0.55823	0.79310	1.49843	3.71801	12.48517	1.07×10^{13}
0.5	0.07936	0.19840	0.39680	0.79360	1.58720	3.17440	7.93600
	0.37509	0.40418	0.46941	0.70430	1.85115	10.57500	1.19×10^{13}
	0.41393	0.47147	0.61139	1.14721	3.41437	17.51156	1.52×10^{13}
1.0	0.07347	0.18369	0.36738	0.73477	1.46955	2.93910	7.34775
	0.36048	0.38385	0.43412	0.61629	1.91889	22.66557	2.88×10^{13}
	0.37983	0.42100	0.51698	0.92186	3.71457	38.35188	3.66×10^{13}
2.0	0.06890	0.17227	0.34454	0.68909	1.37818	2.75636	6.89090
	0.34978	0.37016	0.41285	0.56205	2.16764	94.94838	7.39×10^{13}
	0.36025	0.39413	0.47015	0.78318	4.52200	159.5708	9.27×10^{13}
4.0	0.06593	0.16483	0.32967	0.65935	1.31870	2.63740	6.59351
	0.34305	0.36194	0.40090	0.53229	2.55398	943.1931	1.60×10^{14}
	0.35024	0.38072	0.44742	0.70903	5.74018	1541.796	1.98×10^{14}
10.0	0.06385	0.15964	0.31929	0.63859	1.27718	2.55436	6.38591
	0.33845	0.35647	0.39326	0.51400	3.16099	79704.21	2.11×10^{14}
	0.34434	0.37284	0.43418	0.66482	7.67206	126308.1	2.55×10^{14}

Entries associated with the case $\beta = 2.0$ in Table 3 show that $E[T_1^{\boldsymbol{\pi}}]$ and $\sigma(T_1^{\boldsymbol{\pi}})$ may not be an increasing function of the transmission factor R_0^{re}. Roughly speaking this occurs because these environmental descriptors are defined as weighted sums of the corresponding descriptor given that $E(0) = e$. For example, to get understanding about $E[T_1^{\boldsymbol{\pi}}] = \sum_{e=1}^{4} E[T_{1e}]\pi_e$, in Fig. 6 we plot the unconditional mean value $E[T_1^{\boldsymbol{\pi}}]$, as well as its four contributions $E[T_{1e}]$ versus δ. In this case, we have $E[T_1^{ADA}] = 8.75 \times 10^{49}$ and $E[T_1^{ARA}] = 4.34624$.

Despite that the magnitude of $E[T_1^{ADA}]$ is too large to be represented in the figure, we observe that all the curves drop below the ARA value $E[T_1^{ARA}]$. In fact, the numerical results show the following limiting results:

$$\lim_{\delta\to 0} E[T_{1e}] = E[T_1(e)], \quad 1 \le e \le 4,$$
$$\lim_{\delta\to \infty} E[T_{1e}] = E[T_1^{ARA}],$$

where the mean extinction times of the e-th marginal SIS model are given by $E[T_1(1)] = 0.34414$, $E[T_1(2)] = 13.26926$, $E[T_1(3)] = 3.63\times10^{17}$ and $E[T_1(4)] =$

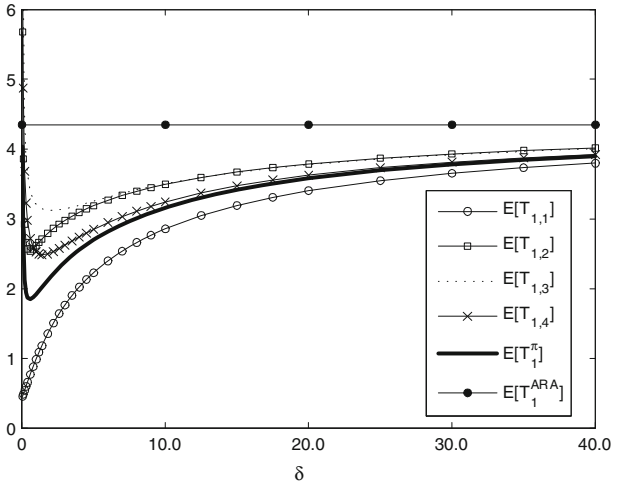

Fig. 6 The extinction times versus δ

7.29×10^{50}. Now the minimum exhibited by $E[T_1^{\boldsymbol{\pi}}]$ in Table 3 is fully explained, because $E[T_{1e}]$ has the same shape for $2 \leq e \leq 4$; that is, while $R_0(e) > 1.0$.

To conclude this section, we next consider an SIR-re model with initial state $(m, n) = (1, 29)$, which operates under a star environment with $E = 3$, $(\delta_1, \delta_2, \delta_3) = (0.5, 0.25, 0.25)$ and $p_2 = p \in (0, 1)$. In addition, we assume that $E(0) = 1$, $(\beta_1, \beta_2, \beta_3) = (\beta, \beta + \epsilon, \beta - \epsilon)$ and $\gamma_e = \gamma$, for $e \in S_E$. The extreme case $\epsilon = 0$ corresponds to the classical SIR model with transmission rates (β, γ). We notice that $\boldsymbol{\pi}_E = \left(\frac{1}{3}, \frac{2p}{3}, \frac{2(1-p)}{3}\right)$ and $R_0^{re} = \frac{1}{\gamma}\left(\beta + \frac{2}{3}(2p-1)\epsilon\right)$.

From the explicit formula for R_0^{re}, we notice that the transmission factor is an increasing function of p and β, but it decreases as far as γ increases. On the other hand, $p = 0.5$ acts as a threshold of the behavior of R_0^{re} with respect to the deviation parameter ϵ; that is, R_0^{re} is decreasing in ϵ, for $p < 0.5$, $R_0^{re} = \frac{\beta}{\gamma}$ (constant), for $p = 0.5$, and finally R_0^{re} is increasing as a function of ϵ, for $p > 0.5$. This behavior can be explained by noticing that the marginal transmission factors are given by $(R_0(1), R_0(2), R_0(3)) = \left(\frac{\beta}{\gamma}, \frac{\beta+\epsilon}{\gamma}, \frac{\beta+\epsilon}{\gamma}\right)$, so an increment in p means also an increment of the probability π_2 associated with the highest marginal transmission factor $R_0(2)$.

The analysis of the unconditional expected time to extinction $E[T_{1,29}^{\boldsymbol{\pi}}]$ is more interesting. In Fig. 7, we display R_0^{re} and $E[T_{1,29}^{\boldsymbol{\pi}}]$ as a function of ϵ, for $\beta = 1.0$ and $\gamma = 0.5$. Like in the previous SIS-re example, we observe that increasing values of R_0^{re} does not necessarily imply a reduction in the value of $E[T_{1,29}^{\boldsymbol{\pi}}]$; see in the figure the cases $p = 0.75$ and $p = 0.9$, where peaks around $\epsilon = 0.4$ and $\epsilon = 0.85$ are respectively observed. A possible explanation is as follows. The balanced case $p = 0.5$ implies that $\pi_e = 1/3$, for $e \in S_E$, and $R_0^{re} = 2.0$, irrespective of the value of ϵ. For this case, we notice that $E[T_{1,29}^{\boldsymbol{\pi}}]$ decreases with increasing values of ϵ. In

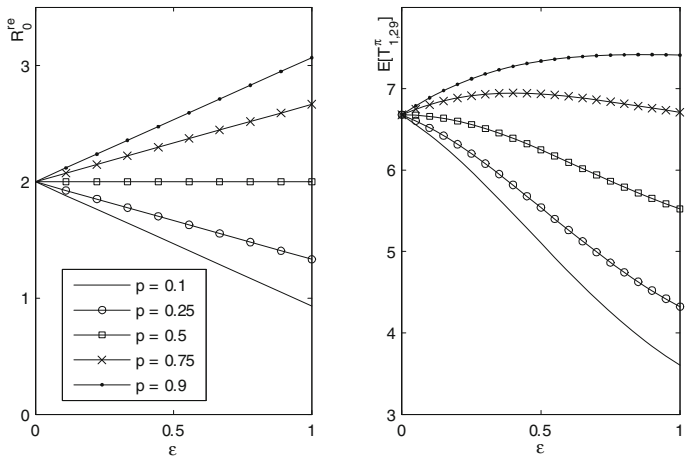

Fig. 7 R_0^{re} and $E[T_{1,29}^{\boldsymbol{\pi}}]$ versus ϵ

particular, we notice that $\lim_{\epsilon \to 1} R_0(3) = 0$. To choose $p < 0.5$ implies to increase π_3 and, consequently, to give support to the lowest marginal transmission factor $R_0(3)$. Thus, $E[T_{1,29}^{\boldsymbol{\pi}}]$ decreases with ϵ. On the other hand, if $p > 0.5$ then $R_0(2)$ is positively supported. As a result, $E[T_{1,29}^{\boldsymbol{\pi}}]$ is able to invert the decreasing trend and its has a mode in the cases $p = 0.75$ and $p = 0.9$. Finally, it is increasing in ϵ, for $p = 0.95$.

4.4 An illustration of transient behavior

This example illustrates the transient behavior in an SIS-re model with alternating environment. Once more, we assume that $N = 100$, $E = 2$, $(\beta_1, \beta_2) = (5.75, 0.25)$ and $\gamma_e = e$, for $1 \leq e \leq 2$, so the marginal transmission factor switches between $R_0(1) = 5.75$ and $R_0(2) = 0.125$. We consider the following two scenarios for $\mathbf{Q}_E$: $A)(\delta_1, \delta_2) = (0.04, 0.01)$, and $B)(\delta_1, \delta_2) = (4.0, 1.0)$. For both scenarios, we find that $\boldsymbol{\pi}_E = (0.2, 0.8)$.

We are interested in the behavior of the number of infected individuals $I(t)$, given that the epidemic is longer than t. To this end, in Fig. 8 we plot the transient expectation $M(t) = \sum_{(i,e) \in S_T} E[I(t) \,|T_{ie} > t, X(0) = (i, e)\,]a_{ie}$. Our choices for the initial state at time $t = 0$ are $a_{ie} = \delta_{(i,e)(1,1)}$ (one initial infective) and $a_{ie} = \delta_{(i,e)(100,1)}$ (all the population is infected). The computation of $M(t)$ is easily derived from the matrix exponential $\exp\{\mathbf{Q}_T t\}$ (see Appendix 6.2).

The curves in the figure show that, irrespective of the initial condition, the transient expectation $M(t)$ converges to the mean value of the quasi-stationary distribution of infected individuals; that is, we observe that $\lim_{t \to \infty} M(t) = E[\mathbf{u}_i] = \sum_{i=1}^{100} i u_i (E[u])$, where $E[\mathbf{u}_i] = 75.58020$, in the scenario A, and $E[\mathbf{u}_i] = 12.92589$, in the scenario B. Comparing R_0^{re} in both scenarios, we observe that $R_0^{re} = 1.22846$, for scenario A, while $R_0^{re} = 0.84090$, for scenario B. This fact explains why $E[\mathbf{u}_i]$ is higher for scenario A.

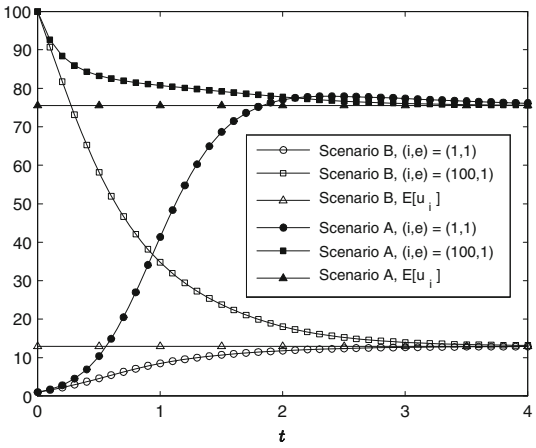

Fig. 8 The expected number of infectives $M(t)$

5 Concluding remarks

Our aim in this paper is to provide a detailed analytical and algorithmic study of the infectious disease dynamics when there is a random environment subject to stochastic changes. This is a key distinguished point in this work, as much of the existing related literature is mainly focused on mathematical models with deterministic environmental rates. It should be pointed out that the proposed numerical methods require little implementation effort because they rely on very simple and stable recursive formulas.

We consider a continuous-time approach for epidemic models of SIS and SIR -type driven by random environments which form a Markov chain. Although the possibility of considering that environmental factors may affect several, or even all, transition rates has been considered earlier (e.g. Bacaër and Gomes 2009; Steinsaltz et al. 2011; Wesley and Allen 2009), most of the previous works assume that the environmental changes only influence the contact transmission rates. In the framework assumed here, the effect of environmental variability in other transition rates (e.g. recovery rates, demographic rates) may also be incorporated.

Two tools play a significant role for understanding stochastic disease transmission under a random environment framework. First of all, we refer to the transmission factor. It is well-known that the transmission factor R_0 in a deterministic model is frequently interpreted as a threshold value, so a disease will die out, when $R_0 < 1$, but it may become endemic, if $R_0 > 1$. This interpretation is not possible for stochastic models with a finite subset of transient states, since the epidemic extinction is reached in a finite time with probability one. The generalized transmission factor R_0^{re}, under consideration in this paper, provides a natural extension of the classical definition in models with constant transmission rates. Despite that the length of the epidemic may not be an increasing function of R_0^{re}, it gives an accurate measure of the infection spread in the population. Roughly speaking, as far as the magnitude of R_0^{re} increases, it is harder to control the epidemic. Most of the numerical experiments in Section

4 were successfully interpreted by combining the role played by R_0^{re}, including the related marginal and initial transmission factors $R_0(e)$ and R_{0e}, and the ADA and ARA introduced in Sect. 4.1. These two average approximations are the second fundamental tool for getting understanding from the epidemic indicators, as far as they are interpreted in terms of a comparison between the environmental sojourn rates and the epidemic transmission rates.

The previous comments provide a summary of the novelty and the strengths of the random environment approach assumed in this paper. In what follows, we also comment on the limitations of our study. Despite that many assumptions are idealized representations of the underlying real infectious disease, most stochastic epidemic models are complex from an analytical point of view. The models with stochastic random environment are not an exception. In fact, the consideration of the environmental generator $\mathbf{Q}_E$ and the subsequent transmission rates $\lambda_i(e)$ and $\mu_i(e)$ imply a notable increment of the number of system parameters. As a result, the main hurdles here are probably related to statistical and prediction issues.

The study can be continued in several directions. For example, it would be interesting to relax the exponentiality of the sojourn times as well as the inherent assumptions of individual homogeneity and randomly-mixing population. The present study is oriented to epidemic models but a forthcoming work might explore how stochastic random environments influence the dynamics of other biological models including competition and predation. We also notice that the methodology can be easily extended to the discrete-time case. Generalization to more sophisticated models is a challenging topic but a compromise between mathematical tractability and realistic assumptions is always a premise.

Acknowledgments The authors would like to thank the referees for their useful comments and suggestions. This work was supported by the Government of Spain (Department of Science and Innovation) and the European Commission through project MTM 2011-23864.

6 Appendix

6.1 The transmission factor

The classical biological interpretation of the transmission factor presents R_0 as the expected number of contacts produced by one individual during the period of infectivity, when the individual is introduced into a completely susceptible population. With the help of R_0, the parameter region where the time to extinction is short can be identified by small values of R_0. In contrast, if R_0 is large, then the epidemic tends to persist for a very long time, so a state of quasi-stationary equilibrium may be reached before a random fluctuation leads to the extinction of the epidemic.

We next consider a natural extension that takes into account the environmental effects. To this end, we define the environmental transmission factor R_0^{re} as follows:

$$R_0^{re} = \sum_{e=1}^{E} R_{0e}\pi_e,$$

where R_{0e} represents the transmission factor given that the infectious individual is introduced at any time t in which $E(t) = e$. These conditional transmission factors R_{0e}, for $e \in S_E$, are governed by the system of equations

$$R_{0e} = \frac{\gamma_e}{\gamma_e + q_E(e)} \frac{\beta_e}{\gamma_e + q_E(e)} + \frac{q_E(e)}{\gamma_e + q_E(e)} \left(\frac{\beta_e}{\gamma_e + q_E(e)} + \sum_{\substack{e'=1 \\ e' \neq e}}^{E} \frac{q_E(e, e')}{q_E(e)} R_{0e'} \right), \quad e \in S_E.$$

When the environmental state is $e \in S_E$, the two first summands in the right hand side of the above formula count the expected number of infections that occur until the next environmental change or the end of the infectivity period, whatever happens first. Given that the environment turns to state e', the last term corresponds to the expected number of infections that occur thereafter until the end of the infectivity period.

By introducing the column vector $\mathbf{R}_0^{re} = (R_{01}, \ldots, R_{0E})$, the system of equations becomes

$$\mathbf{R}_0^{re} = \left(\mathbf{D}^{\gamma} - \mathbf{Q}_E\right)^{-1} \mathbf{D}^{\beta} \mathbf{e}_E.$$

In particular, explicit formulas for $\mathbf{R}_0^{re}$ can be obtained for the cyclic environment and the star environment by considering the appropriate choice of the generator $\mathbf{Q}_E$ (see Sect. 4) and setting $d_e = \gamma_e$ and $c_e = \beta_e$, for $e \in S_E$ (see Appendix 6.3).

6.2 General results for the unconditional extinction time of a CTMC with finite state

Let X be an absorbing CTMC on a finite state space S; that is, the state space S consists of a set S_T of transient states where the process evolves until it escapes to a set of absorbing states S_A, so $S = S_A \cup S_T$. Suppose the initial distribution $\mathbf{a} = \{a_i; i \in S_T\}$, where $a_i = P\{X(0) = i\}$. Then, the unconditional distribution function of the extinction time T (i.e., the time to reach the absorption) is given by

$$P\{T \leq x\} = 1 - \mathbf{a} \exp\{\mathbf{Q}_T x\}\mathbf{e}, \quad x \geq 0.$$

In fact, the (i, j)-th element of the matrix exponential $\exp\{\mathbf{Q}_T x\}$ amounts to the probability $P\{X(x) = j, T > x \,|X(0) = i\}, i \in S_T$.

Since $\mathbf{Q}_T$ is the sub-generator associated with the transient states, it is an invertible matrix. The following statements are readily verified (Li 2010):

$$\begin{aligned} f_T(x) &= -\mathbf{a} \exp\{\mathbf{Q}_T x\}\mathbf{Q}_T \mathbf{e}, \quad x \geq 0, \\ \varphi_T(s) &= -\mathbf{a} \left(s\mathbf{I} - \mathbf{Q}_T\right)^{-1} \mathbf{Q}_T \mathbf{e}, \quad \mathrm{Re}(s) \geq 0, \\ E[T^k] &= k! \mathbf{a} \left(-\mathbf{Q}_T\right)^{-k} \mathbf{e}, \quad k \geq 1, \end{aligned}$$

where $f_T(x), \varphi_T(s)$ and $E[T^k]$ denote the density function, the Laplace transform and the k-th moment of the extinction time T, respectively. We refer to Sect 4.3 for some discussion on the numerical computation of the density $f_T(x)$.

In epidemiology, it is often known that a certain epidemic has been evolving for a long time and that it has not reached the extinction yet. This situation provides the necessary motivation to study the distribution of the time to extinction from quasi-stationarity $T_{\mathbf{u}}$ (i.e., the distribution of T given that the initial distribution is $\mathbf{a} = \mathbf{u}$). The distribution of $T_{\mathbf{u}}$ has been studied in a number of stochastic epidemic models including the SIS epidemic model and the Verhulst logistic model (Nåsell 2001), SIR epidemic models with gamma distributions (Andersson and Britton 2000) and endemic infections with demography (Andersson and Lindenstrand 2011; Nåsell 1999). All these papers employ a similar methodology to prove that $T_{\mathbf{u}}$ follows an exponential distribution with rate $\sum_{i\in S_{TA}} u_i q_{iS_A}$, where the set S_{TA} is defined as $S_{TA} = \{i \in S_T \,|\exists j \in S_A \text{ such that } q_{ij} > 0\}$ and $q_{iS_A} = \sum_{j\in S_A} q_{ij}$. For a general unified treatment, including both continuous and discrete-time Markov chains, we refer the reader to the recent paper by Artalejo (2012).

6.3 Solving the system $(\mathbf{D} - \mathbf{Q}_E)\mathbf{x} = \mathbf{c}$

A system of linear equations of the form $(\mathbf{D} - \mathbf{Q}_E)\mathbf{x} = \mathbf{c}$, where $\mathbf{D} = diag(d_1, \ldots, d_E), \mathbf{x} = (x_1, \ldots, x_E)'$ and $\mathbf{c} = (c_1, \ldots, c_E)'$, arises several times along this paper (e.g. Sect. 3, Appendix 6.1). The numerical computation of $\mathbf{x}$ amounts to obtain the inverse matrix $(\mathbf{D} - \mathbf{Q}_E)^{-1}$. Since $\mathbf{D} - \mathbf{Q}_E$ is a diagonally dominant M-matrix, its inverse matrix is always positive.

We next give a general explicit solution for the particular cases of the cyclic environment and the star environment described in Sect. 4. We notice that the scalar algebraic calculations are reduced to only deal with positive numbers, so the computation of $\mathbf{x}$ is stable.

In the case of a cyclic environment, the system $(\mathbf{D} - \mathbf{Q}_E)\mathbf{x} = \mathbf{c}$ reduces to the equations

$$\begin{aligned}(d_e + \delta_e)x_e - \delta_e x_{e+1} &= c_e, \quad 1 \le e \le E-1,\\ (d_E + \delta_E)x_E - \delta_E x_1 &= c_E.\end{aligned}$$

The recursive iteration of the first equation yields

$$x_e = \sum_{i=e}^{E} \frac{c_i}{\delta_i} \prod_{j=e}^{i} \frac{\delta_j}{d_j + \delta_j} + x_1 \prod_{i=e}^{E} \frac{\delta_i}{d_i + \delta_i}, \quad 2 \le e \le E.$$

Then, x_1 is determined from the remaining equation. It is as follows

$$x_1 = \frac{\sum_{i=1}^{E} \frac{c_i}{\delta_i} \prod_{j=1}^{i} \frac{\delta_j}{d_j+\delta_j}}{1 - \prod_{i=1}^{E} \frac{\delta_i}{d_i+\delta_i}}.$$

On the other hand, in the case of a star environment, the system $(\mathbf{D}-\mathbf{Q}_E)\mathbf{x}=\mathbf{c}$ is given by

$$\begin{aligned}(d_1+\delta_1)x_1-\delta_1\sum_{e=2}^{E}p_e x_e &= c_1,\\ -\delta_e x_1+(d_e+\delta_e)x_e &= c_e,\quad 2\le e\le E.\end{aligned}$$

From the second equation, we have

$$x_e=\frac{c_e+\delta_e x_1}{d_e+\delta_e},\quad 2\le e\le E.$$

Now we use the first equation and the fact that $\sum_{e=2}^{E}p_e=1$ to get

$$x_1=\frac{c_1+\delta_1\sum_{e=2}^{E}\frac{p_e c_e}{d_e+\delta_e}}{d_1+\delta_1\sum_{e=2}^{E}\frac{p_e d_e}{d_e+\delta_e}}.$$

6.4 SIS-re model: the ratio of expectations distribution

We now consider the RE-distribution for the SIS-re epidemic model. Let T_{ie}^{*} be the time that the chain $X=\{(I(t),E(t));t\ge 0\}$ spends in state $(i,e)\in S_T$ before absorption. Then, given that $(I(0),E(0))=(i_0,e_0)\in S_T$, we define

$$P_{(i_0,e_0)}(i,e)=\frac{E_{(i_0,e_0)}\left[T_{ie}^{*}\right]}{E_{(i_0,e_0)}\left[T\right]},\quad (i,e)\in S_T.$$

For any fixed state $(\bar{i},\bar{e})\in S_T$, we apply first-step analysis to obtain

$$\left(\mathbf{D}_i^{\lambda}+\mathbf{D}_i^{\mu}-\mathbf{Q}_E\right)\mathbf{m}_i(\bar{i},\bar{e})=\mathbf{D}_i^{\mu}\mathbf{m}_{i-1}(\bar{i},\bar{e})+(1-\delta_{iN})\mathbf{D}_i^{\lambda}\mathbf{m}_{i+1}(\bar{i},\bar{e})+\delta_{i\bar{i}}\mathbf{e}_E(\bar{e}),$$

where $\mathbf{m}_i(\bar{i},\bar{e})=\left(E_{(i,1)}[T_{\bar{i}\bar{e}}^{*}],\ldots,E_{(i,E)}[T_{\bar{i}\bar{e}}^{*}]\right)'$, for $0\le i\le N$, with $\mathbf{m}_0(\bar{i},\bar{e})=\mathbf{0}_E$.

The system has the same block tridiagonal form as the system for the moments of the extinction times. Hence, the recursive scheme in Algorithm 2 remains valid by changing $k\mathbf{m}_i^{k-1}$ by $\delta_{i\bar{i}}\mathbf{e}_E(\bar{e})$.

Of course, summing $E_{(i_0,e_0)}\left[T_{ie}^{*}\right]$ over $(i,e)\in S_T$, we obtain $E_{(i_0,e_0)}[T]$ so $P_{(i_0,e_0)}(i,e)$ is determined.

References

Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P (2006) Seasonality and the dynamics of infectious diseases. Ecol Lett 9:467–484

Andersson H, Britton T (2000) Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. J Math Biol 41:559–580

Andersson P, Lindenstrand D (2011) A stochastic SIS epidemic with demography: initial stages and time to extinction. J Math Biol 62:333–348
Artalejo JR (2012) A general proof of the time to extinction from quasi-stationarity. Physica A 391: 4483–4486
Artalejo JR, Lopez-Herrero MJ (2010) Quasi-stationarity and ratio of expectations distributions: a comparative study. J Theor Biol 266:264–274
Artalejo JR, Economou A, Lopez-Herrero MJ (2012) Stochastic epidemic models revisited: analysis of some continuous performance measures. J Biol Dyn 6:189–211
Bacaër N (2007) Approximation of the basic reproduction number R_0 for vector-borne diseases with a periodic vector population. Bull Math Biol 69:1067–1091
Bacaër N, Gomes MGM (2009) On the final size of epidemics with seasonality. Bull Math Biol 71: 1954–1966
Barbour AD, Pollett PK (2010) Total variation approximation for quasi-stationary distributions. J Appl Prob 47:934–946
Bartlett MS (1960) Stochastic population models in ecology and epidemiology. Methuen, London
Britton T, Lindholm M (2009) The early stage behaviour of a stochastic SIR epidemic with term-time forcing. J Appl Prob 46:975–992
Ciarlet PG (1989) Introduction to numerical linear algebra and optimization. Cambridge University Press, Cambridge
Cohen AM (2007) Numerical methods for Laplace transform inversion. Springer, New York
Darroch JN, Seneta E (1967) On quasi-stationary distributions in absorbing continuous-time finite Markov chains. J Appl Probab 4:192–196
Franke JE, Yakubu A (2006) Discrete-time SIS epidemic model in a seasonal environment. SIAM J Appl Math 66:1563–1587
Grassly NC, Fraser C (2006) Seasonal infectious disease epidemiology. Proc R Soc B 273:2541–2550
Gray A, Greenholgh D, Mao X, Pan J (2012) The SIS epidemic model with Markovian switching. J Math Anal Appl 394:496–516
Keeling MJ, Rohani P (2007) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton
Kulkarni VG (1995) Modeling and analysis of stochastic systems. Chapman & Hall, London
Kuske R, Gordillo LF, Greenwood P (2007) Sustained oscillations via coherence resonance in SIR. J Theor Biol 245:459–469
Li QL (2010) Constructive theory in stochastic models with applications: the RG-factorizations. Springer/Tsinghua University Press, Berlin/Beijing
Metcalf CJE, Bjørnstad ON, Grenfell BT, Andreasen V (2009) Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. Proc R Soc B 276:4111–4118
Moler CB, van Loan CF (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev 45:3–49
Nåsell I (1999) On the time to extinction in recurrent epidemics. JR Stat Soc B 61:309–330
Nåsell I (2001) Extinction and quasi-stationarity in the Verhulst logistic model. J Theor Biol 211:11–27
Parham PE, Michael E (2011) Outbreak properties of epidemic models: the roles of temporal forcing and stochasticity on pathogen invasion dynamics. J Theor Biol 271:1–9
Prajneshu Gupta CK, Sharma U (1986) A stochastic epidemic model with seasonal variations in infection rate. Biom J 28:889–895
Steinsaltz D, Tuljapurkar S, Horvitz C (2011) Derivatives of the stochastic growth rate. Theor Popul Biol 80:1–15
Stone L, Olinky R, Huppert A (2007) Seasonal dynamics of recurrent epidemics. Nature 446:533–536
Trapman P, Meester R, Heesterbeek H (2004) A branching model for the spread of infectious animal diseases in varying environments. J Math Biol 49:553–576
van Doorn EA, Pollett PK (2008) Survival in a quasi-death process. Linear Algebra Appl 429:776–791
Varga RS (2000) Matrix iterative analysis. Springer, Berlin
Wesley CL, Allen LJS (2009) The basic reproduction number in epidemics with periodic demographics. J Biol Dyn 3:116–129
Wilkinson JH (1988) The algebraic eigenvalue problem. Oxford University Press, Oxford