Journal of Mathematical Biology (2024) 88:31
https://doi.org/10.1007/s00285-024-02049-0

Mathematical Biology

On cognitive epidemic models: spatial segregation versus nonpharmaceutical interventions

Guodong Liu[1,2] · Hao Wang[2] · Xiaoyan Zhang[1]

Received: 16 May 2023 / Revised: 19 November 2023 / Accepted: 19 January 2024

Abstract
Fick's law and the Fokker–Planck law of diffusion are applied to manifest the cognitive dispersal of individuals in two reaction-diffusion SEIR epidemic models, where the disease transmission is illustrated by nonlocal infection mechanisms in heterogeneous environments. Building upon the well-posedness of solutions, threshold dynamics are discussed in terms of the basic reproduction numbers for the two cognitive epidemic models. The numerical investigation reveals that the Fokker–Planck law can better describe the diffusion of individuals by taking different dispersal strategies of exposed individuals in our cognitive epidemic models, and provides some insights on spatial segregation and nonpharmaceutical interventions: (i) spatial segregation occurs in the random diffusion model when the nonlocal infection radius is small, while it appears in the symmetric diffusion model when the radius is large; (ii) nonpharmaceutical interventions on restricting the dispersal of exposed and infected individuals do not contribute to reducing the infection proportion, but rather eliminate the disease in a region, which expands as the nonlocal infection radius increases. We additionally find that the final infection size in the random diffusion model is significantly smaller than that in the symmetric diffusion model and decreases as the nonlocal infection radius increases.

Keywords SEIR epidemic model · Nonlocal infection · Cognitive diffusion · Spatial segregation · Nonpharmaceutical interventions

Mathematics Subject Classification 92D30 · 92D50 · 35K57 · 37N25

✉ Hao Wang
hao8@ualberta.ca

1 School of Mathematics, Shandong University, Jinan 250100, China

2 Interdisciplinary Lab for Mathematical Ecology and Epidemiology, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton T6G 2G1, Canada

Published online: 26 February 2024

1 Introduction

1.1 Laws of diffusion

The movement of a physical quantity or biological organism in space is frequently modeled by diffusion equations. The homogeneous diffusion with a constant coefficient is commonly employed to illustrate the Brownian motion of microscopic particles characterized by irregular and non-directional movement. It has been applied to describe the dispersal or migration of population in spatial ecology by assuming the diffusivity is spatially homogeneous (Skellam 1951; Okubo and Levin 2001; Contrell and Cosner 2003). The diffusion equation with such homogeneous diffusion is read as

$$\partial_t u(x,t) = d\Delta u(x,t), \tag{1.1}$$

where $u(x,t)$ is the concentration of particles, or the density of individuals at position x and time t. The positive constant d is the spatially homogeneous coefficient of diffusion, and Δ is the Laplace operator.

However, spatial heterogeneity is ubiquitous in nature, accordingly, diffusivity tends to vary spatially. There are two laws to model diffusion in heterogeneous media: Fick's law (Fick 1855) and the Fokker–Planck law, also known as Chapman's law (Chapman 1928) of diffusion. In correspondence with these two different laws, one finds two possible diffusion equations which are expressed as

$$\partial_t u(x,t) = \nabla \cdot (d(x)\nabla u(x,t)), \tag{1.2}$$

and

$$\partial_t u(x,t) = \Delta(d(x)u(x,t)), \tag{1.3}$$

respectively. Here, the positive smooth function $d(x)$ is the spatially varying coefficient of diffusion. $\nabla\cdot$ and ∇ are the divergence and gradient operators, respectively.

These two laws of diffusion can be obtained from the dispersal of individuals among discrete patches (Okubo and Levin 2001; Wang et al. 2022). For Fick's law, it is assumed that the dispersal rates between two patches are equal. For the Fokker–Planck law, it is assumed that the dispersal rates from one patch to the two adjacent patches are equal. From the perspective of the patch model, the diffusion following Fick's law is usually called symmetric diffusion (Kim et al. 2019) while that following the Fokker–Planck law is known as random diffusion (Wang et al. 2022). From another viewpoint, both laws of diffusion can be seen as special cases of the more general Fokker–Planck equation (Bengfort et al. 2016). Fick's law assumes that the velocity of each step of movement depends on the conditions at the end point of that step, while the Fokker–Planck law assumes that the velocity depends on the conditions at the starting point of each step. In Andreucci et al. (2019), Andreucci et al. considered the diffusion laws by a density field $u(x,t)$ in terms of the flux vectorial field $J(x,t)$, in the diffusion equation of the form

$$\partial_t u(x,t) = -\nabla \cdot J(x,t),$$

where $J(x,t) = -d(x)\nabla u(x,t)$ for Fick's law and $J(x,t) = -\nabla(d(x)u(x,t))$ for the Fokker–Planck law.

We view the two laws of diffusion from a mathematical perspective. Given a bounded domain $\Omega \subset \mathbb{R}^n$ ($n \geq 1$) and an unknown function $u : \bar{\Omega} \to \mathbb{R}$, consider the second-order uniform elliptic operator $\mathfrak{L}$ having the divergence form

$$\mathfrak{L}u := \sum_{i,j=1}^{n} (a^{ij}(x)u_{x_i})_{x_j} + \sum_{i=1}^{n} b^i(x)u_{x_i} + c(x)u$$

for given coefficient functions a^{ij}, b^i, c ($i, j = 1, \ldots, n$), where $(a^{ij})_{n\times n}$ is positive definite and symmetric. In addition, there exist positive constants θ and Θ such that

$$\theta|y|^2 \leq \sum_{i,j=1}^{n} a^{ij}(x)y_i y_j \leq \Theta|y|^2$$

for $x \in \Omega$ and $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, which is known as uniform elliptic condition. The three types of diffusion mentioned above satisfy the uniform elliptic operator in the divergence form:

- Homogeneous diffusion term $d\Delta u$, where $a^{ij}(x) = d\delta_{ij}$ and $b^i = c \equiv 0$, here δ_{ij} is the Kronecker symbol;
- Symmetric diffusion term $\nabla \cdot (d(x)\nabla u)$, where $a^{ij}(x) = d(x)$ and $b^i = c \equiv 0$;
- Random diffusion term $\Delta(d(x)u)$, where $a^{ij}(x) = d(x)$, $b^i(x) = d_{x_i}(x)$ and $c(x) = \Delta d(x)$.

Note that the random diffusion is the divergence of the gradient of $d(x)u$, i.e.,

$$\Delta(d(x)u) = \nabla \cdot \nabla(d(x)u) = \nabla \cdot (d(x)\nabla u) + \nabla \cdot (u\nabla d(x)).$$

As a consequence, the difference between the two types of diffusion is the advection term $\nabla \cdot (u\nabla d(x))$. More precisely,

$$\nabla \cdot (d(x)\nabla u) = \nabla d(x) \cdot \nabla u + \Delta d(x)u.$$

The last term in the above equation is the growth adaptation term and $\Delta d(x)$ is the curvature of the diffusivity (Bengfort et al. 2016). If u represents the density of the population, then the curvature $\Delta d(x)$ represents the growth rate of the population. When $\Delta d(x) > 0$, there is extra net growth of the population. When $\Delta d(x) < 0$, there is additional net decay of the population. When $\Delta d(x) = 0$, the only distinction between Fick's law and the Fokker–Planck law is the term $\nabla d(x) \cdot \nabla u$. It is obvious that if $d(x) \equiv d =$ constant, then the symmetric and random diffusion becomes the homogeneous diffusion.

Consider two types of homogeneous boundary conditions $\nabla u \cdot \mathbf{n} = 0$ and $\nabla(d(x)u) \cdot \mathbf{n} = 0$. Here, $\mathbf{n}$ is the unit outer normal vector at the boundary. The first one is known as the Neumann boundary condition. Note that

$$\nabla(d(x)u) \cdot \mathbf{n} = d(x)\nabla u \cdot \mathbf{n} + \nabla d(x) \cdot \mathbf{n}u,$$

which is a combination of the Neumann and Dirichlet boundary conditions. If the diffusion equations (1.1) or (1.2) are associated with the homogeneous Neumann boundary condition, the total population number of (1.1) or (1.2) is conserved. This is also true for (1.3) if it takes the second boundary condition. Hence, these two types of boundary conditions are zero-flux boundary conditions.

In applied sciences, various situations involve the assumption of two distinct laws of diffusion. In Schnitzer (1993), Schnitzer mentioned that different diffusion laws may be followed in different cases. The author gave two systems and remarked that one system follows Fick's law and the other one follows the Fokker–Planck law. In Bringuier (2011), three experimental examples of diffusion in heterogeneous media are examined to verify the law of diffusion, and the observed law can be of Fick, Fokker–Planck, or hybrid type. The question of whether diffusion in heterogeneous media should follow either Fick's law or the Fokker–Planck law has long been debated in the literature. The theory of diffusion and dispersion is not well understood yet and the choice of the diffusion law is nontrivial (Andreucci et al. 2019).

1.2 Model formulation

The practical importance of understanding the transmission mechanism of infectious disease is steadily increasing. Epidemic models are effective tools for exploring the evolutionary dynamics of epidemics to provide strategies for disease regulation. It should be mentioned that in 1927, Kermack and McKendrick proposed a simple deterministic susceptible-infected-recovered (SIR) epidemic model in Kermack and McKendrik (1927) with the bilinear incidence mechanism to represent the direct disease transmission between susceptibles and infectives. The Kermack-McKendrick model indeed plays a pivotal role in subsequent developments in the study of the dynamics of infectious diseases. Typically, epidemic models are characterized by nonlinearity to describe the transmission of infectious diseases. In Kendall (1965), Kendall generalized the classical Kermack-McKendrick model to a space-dependent integro-differential equation, which is written as

$$\begin{cases} \dfrac{\partial S(x,t)}{\partial t} = -\beta S(x,t) \displaystyle\int_{-\infty}^{\infty} K(x-y)I(y,t)dy, \ t > 0, \\ \dfrac{\partial I(x,t)}{\partial t} = \beta S(x,t) \displaystyle\int_{-\infty}^{\infty} K(x-y)I(y,t)dy - \gamma I(x,t), \ t > 0, \\ \dfrac{\partial R(x,t)}{\partial t} = \gamma I(x,t), \ t > 0. \end{cases}$$

Here, $S(x,t)$ and $I(x,t)$ are the density of susceptibles and infectives at position x and time t, respectively. β and γ are positive constants, accounting for the infection rate and recovery rate, respectively. The nonlinear interaction term of infectives with susceptibles in an unbounded space is given by $S(x,t)\int_{-\infty}^{\infty} K(x-y)I(y,t)dy$, where the kernel $K(x-y)$ measures the contributions of infected individuals at position y to the infection of susceptible individuals at position x.

Inspired by this infection mechanism, we consider the nonlocal infection mechanism on a bounded domain $\Omega \subset \mathbb{R}^n$ $(n \geq 1)$ with smooth boundary $\partial\Omega$. Define $\mathcal{K}: L^1(\Omega) \to L^1(\Omega)$:

$$\mathcal{K}(I)(x) := \int_{\Omega} K(x-y)I(y)dy,$$

where $K: \Omega \to \mathbb{R}_+$ is a smooth nonnegative function with compact support, and not identically zero. We also assume

$$\int_{\Omega} K(x)dx = 1$$

as used in Kendall (1965). This nonlinear interaction between susceptibles and infectives explains that some infectious diseases can spread over long distances via droplet and airborne transmission. For example, an infected individual with influenza can infect others in the same room or even in the same neighbourhood without close contact between the infected patient and susceptibles. This gives rise to the nonlocal infection mechanism in a bounded domain $S(x)\int_{\Omega} K(x-y)I(y)dy$ between susceptibles at position x and infectives at position y around x. In fact, such an infection mechanism allows both local and nonlocal interactions between susceptibles and infectives (Capasso 1978). Usually, the kernel $K(x-y)$ depending on x and y is selected in such a way

$$K(x-y) = 0, \text{ if } y \notin B(x,a), \ \forall x, y \in \Omega,$$

where $B(x,a)$ is an open ball with the center x and radius a in Ω, and local and nonlocal infection can be distinguished according to a in following senses:

- If a is small enough compared with geometrical size $|\Omega|$ of the habitat Ω, i.e., $a \ll |\Omega|$, such infection is local.
- If a is of the same order as $|\Omega|$, such infection is nonlocal.

In fact, for the first case, $K(x-y) \to \delta_0(x)$ as $a \to 0^+$ for $x, y \in \Omega$, where δ_0 is the Dirac function at point 0. The infection kernel $K(x)$ can be regarded as a standard mollifier, which can be found in Appendix C of Evans (2022).

In the theory of epidemiology, it has been recognized that environmental heterogeneity and individual mobility are significant factors that should be included in the spread of infectious diseases. Allen et al. (2008) investigated a reaction-diffusion susceptible-infected-susceptible (SIS) epidemic model where susceptible and infected individuals enter the system at constant diffusion rates. The impact of spatial heterogeneity and individual movement on the persistence and extinction of disease is

studied. Li et al. (2017, 2018) found that the birth-death rate can influence the dynamics of epidemics subtly. The effects of advection and cross-diffusion in epidemic models have been considered (Cui et al. 2017, 2021; Cui and Lou 2016; Li et al. 2020). Reaction-diffusion epidemic models in time-periodic environments have also been studied extensively (Peng and Zhao 2012; Liu and Lou 2022; Liu et al. 2019; Liang et al. 2019). For instance, Liu et al. (2019) considered the effects of seasonality and nonlocal infection in a reaction-diffusion SIR epidemic model.

However, the models in the above-mentioned literature did not include the compartment of exposed (latently infected) individuals and ignored the influence of their movement. For some infectious diseases, infected individuals can experience incubation before showing symptoms, such as HIV/AIDS, rabies, and malaria. Song et al. (2019) formulated a reaction-diffusion epidemic model in a heterogeneous environment with constant diffusion rates, and extensively discussed the asymptotic properties of the basic reproduction number. It is commonly believed that exposed individuals, even if they show no symptoms, can infect susceptible individuals and spread the disease geographically. Motivated by this, we establish the following susceptible-exposed-infected-recovered (SEIR) epidemic model frame with two nonlocal infection processes $S(x,t)\int_\Omega K(x-y)E(y,t)dy$ and $S(x,t)\int_\Omega K(x-y)I(y,t)dy$ in a heterogeneous environment:

$$\begin{cases} \partial_t S - \text{Diffusion of } S = \Lambda(x) - \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad - \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - \mu(x)S, & x\in\Omega,\ t>0, \\ \partial_t E - \text{Diffusion of } E = \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad + \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - (\sigma(x)+\mu(x))E, & x\in\Omega,\ t>0, \\ \partial_t I - \text{Diffusion of } I = \sigma(x)E - (\gamma(x)+\mu(x)+\alpha(x))I, & x\in\Omega,\ t>0, \\ \partial_t R - \text{Diffusion of } R = \gamma(x)I - \mu(x)R, & x\in\Omega,\ t>0, \end{cases} \tag{1.4}$$

where Λ is the recruitment rate of susceptibles; β_1 and β_2 measure the infection rate of susceptibles infected by the exposed and the infectives, respectively; μ and α are the natural death rate and the disease-induced rate of the population, respectively. The exposed individuals become infected with a rate σ, and γ means the recovery rate of infectives.

Cognitive movement is of vital importance in studying the spatial spread of an infectious disease. Animals or humans have the perception or knowledge to move or disperse, which can be modeled mechanistically based on assumptions (Wang and Salmaniw 2023). Symmetric and random diffusions in a heterogeneous environment provide two fundamental movement mechanisms of humans or animals by constructing the dispersal function $d(x)$ in (1.2) and (1.3). Recovered individuals move without direction, thus we take homogeneous diffusion with a positive constant diffusion rate d_R in the R equation. However, it is not reasonable to treat the diffusion rates for S, E, and I as constants. Susceptibles escape from sites or regions with high transmis-

sion rates, such as indoor playgrounds and supermarkets. In these areas, the diffusion rate of susceptibles is high, therefore it can be regarded as an increasing function of transmission rates, i.e.,

$$\text{the diffusion rate of } S = f(x) := f\left(\sum_{i=1}^{2} a_i \beta_i(x)\right),$$

where a_i is the coefficient for the linear combination of β_i for $i = 1, 2$. Infected individuals stay in areas where the recovery rate is high, such as hospitals or regions with sufficient medical resources. In these areas, the diffusion rate of the infected is low, hence it can be seen as a decreasing function of the recovery rate, i.e.,

$$\text{the diffusion rate of } I = h(x) := h(\gamma^{-1}(x)).$$

For the exposed group, there are two cases for its diffusion rate. In one situation, they know they have contacted directly or indirectly with some infectives, and they may have been infected. On this condition, they behave the same as infectives, so the diffusion rate can be chosen as a function of the inverse of the recovery rate. In the other situation, they are not aware of infection without any symptoms, thus they behave like susceptibles, in which scenario the diffusion rate can be chosen as a function of transmission rates. Hence,

$$\text{the diffusion rate of } E = g(x) := g\left(\sum_{i=1}^{2} b_i \beta_i(x)\right) \text{ or } g(\gamma^{-1}(x)),$$

where b_i is the coefficient for the linear combination of β_i for $i = 1, 2$. It is obvious that g is an increasing function with respect to β_i or γ^{-1}.

Having constructed the diffusion rates for different individuals with human cognition and perception, we now apply Fick's law and the Fokker–Planck law of diffusion in the SEIR epidemic models. When we take symmetric diffusion following Fick's law, system (1.4) is expressed as

$$\begin{cases} \partial_t S - \nabla \cdot (f(x)\nabla S) = \Lambda(x) - \beta_1(x) S \displaystyle\int_\Omega K(x-y) E(y,t)dy \\ \quad - \beta_2(x) S \displaystyle\int_\Omega K(x-y) I(y,t)dy - \mu(x) S, & x \in \Omega,\ t > 0, \\ \partial_t E - \nabla \cdot (g(x)\nabla E) = \beta_1(x) S \displaystyle\int_\Omega K(x-y) E(y,t)dy \\ \quad + \beta_2(x) S \displaystyle\int_\Omega K(x-y) I(y,t)dy - (\sigma(x) + \mu(x)) E, & x \in \Omega,\ t > 0, \\ \partial_t I - \nabla \cdot (h(x)\nabla I) = \sigma(x) E - (\gamma(x) + \mu(x) + \alpha(x)) I, & x \in \Omega,\ t > 0, \\ \partial_t R - d_R \Delta R = \gamma(x) I - \mu(x) R, & x \in \Omega,\ t > 0, \\ \nabla S \cdot \mathbf{n} = \nabla E \cdot \mathbf{n} = \nabla I \cdot \mathbf{n} = \nabla R \cdot \mathbf{n} = 0, & x \in \partial\Omega,\ t > 0, \\ S(x,0) = S_0(x),\ E(x,0) = E_0(x),\ I(x,0) = I_0(x),\ R(x,0) = R_0(x), & x \in \Omega. \end{cases} \tag{1.5}$$

Here, we utilize the homogeneous Neumann boundary condition. On the other hand, when we take random diffusion following the Fokker–Planck law, then system (1.4) becomes

$$\begin{cases} \partial_t S - \Delta(f(x)S) = \Lambda(x) - \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad - \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - \mu(x)S, & x \in \Omega,\ t > 0, \\ \partial_t E - \Delta(g(x)E) = \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad + \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - (\sigma(x)+\mu(x))E, & x \in \Omega,\ t > 0, \\ \partial_t I - \Delta(h(x)I) = \sigma(x)E - (\gamma(x)+\mu(x)+\alpha(x))I, & x \in \Omega,\ t > 0, \\ \partial_t R - d_R\Delta R = \gamma(x)I - \mu(x)R, & x \in \Omega,\ t > 0, \\ \nabla(f(x)S)\cdot\mathbf{n} = \nabla(g(x)E)\cdot\mathbf{n} = \nabla(h(x)I)\cdot\mathbf{n} = \nabla R\cdot\mathbf{n} = 0, & x \in \partial\Omega,\ t > 0, \\ S(x,0) = S_0(x),\ E(x,0) = E_0(x),\ I(x,0) = I_0(x),\ R(x,0) = R_0(x), & x \in \Omega. \end{cases} \tag{1.6}$$

Both in systems (1.5) and (1.6), homogeneous boundary conditions are used to represent zero flux across the boundary. Furthermore, we make the following assumptions for the parameters and initial data:

(**H1**) The rates $\Lambda, \mu, \sigma, \gamma, \alpha$, and $\beta_i (i = 1, 2)$ are spatially heterogeneous, positive and Hölder continuous functions on $\bar{\Omega}$.

(**H2**) The diffusion rates $f, g, h \in C^2(\bar{\Omega})$. There exist two positive constants m_0 and M_0 such that

$$m_0 \le f(x), g(x), h(x) \le M_0,\ \forall x \in \bar{\Omega}.$$

(**H3**) The initial data $S_0(x)$, $E_0(x)$, $I_0(x)$, and $R_0(x)$ are nonnegative and continuous functions, and the total numbers of the exposed and the infected are strictly positive, i.e.,

$$\int_\Omega [E_0(x) + I_0(x)]dx > 0.$$

(H2) is assumed to satisfy the uniform elliptic condition. We call systems (1.5) and (1.6) cognitive epidemic models as these models include the perception of humans in the dispersal. By taking the cognitive dispersal rates into account, susceptible individuals have the perception to avoid being infected by infective individuals, and infected individuals have the awareness to seek medical resources. It seems that symmetric diffusion and random diffusion incorporate chemotaxis or advection effects in describing the directed movements of individuals in epidemic modeling.

Based on the two cognitive epidemic models, it is natural to consider the joint effects of cognitive dispersal and nonlocal infection on spatial segregation. Spatial segregation is a spontaneous phenomenon, where two similar species separate from each other in their habitat (Schelling 1969), holding significance in disease regulation. On the other hand, nonpharmaceutical interventions, especially restriction of individual movement,

have been significant in controlling infectious diseases. It is imperative to compare the effectiveness of spatial segregation and nonpharmaceutical interventions in reducing infection fraction and eliminating the disease.

1.3 Outline of contents

Throughout the paper, let $\mathcal{F}^* = \max_{x\in\bar{\Omega}} \mathcal{F}$ and $\mathcal{F}_* = \min_{x\in\bar{\Omega}} \mathcal{F}$ for positive function $\mathcal{F}$, where $\mathcal{F}$ can be taken as Λ, μ, σ, γ, α and $\beta_i (i = 1, 2)$. If we are concerned with the corresponding elliptic problems of models (1.5) and (1.6), we call the nonnegative steady state solution as the disease-free equilibrium (DFE) if $I(x) = 0$ for all $x \in \Omega$, and the endemic equilibrium (EE) if $I(x) > 0$ for some $x \in \Omega$.

The paper addresses four main aspects. From a theoretical standpoint, we investigate the solvability and threshold dynamics of our two cognitive epidemic models (1.5) and (1.6) with nonlocal infection. As an extension work of Wang et al. (2022), we examine and contrast the impacts of various diffusion strategies on spatial segregation within SEIR epidemic models. On the other hand, we are devoted to making clear which law of diffusion is better to describe the dispersal of individuals in epidemic models. Meanwhile, we conduct a comparative analysis of the efficacy of spatial segregation and nonpharmaceutical interventions in disease control.

The rest of the work is organized as follows. In Sects. 2 and 3, we establish the wellposedness of solutions and basic reproduction numbers for models (1.5) and (1.6), and study the threshold dynamics in terms of the basic reproduction numbers. In Sect. 4, global stability for the DFE and EE is investigated for a homogeneous case with local infection. Furthermore, we perform sensitivity analysis of the basic reproduction number and the exposed group in the steady state with respect to the infection rates. We explore the effects of spatial segregation and nonpharmaceutical inventions by numerical methods in Sect. 5. We summarize and discuss our findings in the last section.

2 The epidemic model with symmetric diffusion

Let $\mathbf{X} := C(\bar{\Omega}, \mathbb{R}^4)$ be a Banach space equipped with the supreme norm $\|\cdot\|$. Its positive cone is denoted by $\mathbf{X}_+ := C(\bar{\Omega}, \mathbb{R}^4_+)$ with the interior $\text{int}(\mathbf{X}_+) \neq \emptyset$. Hence, $(\mathbf{X}, \mathbf{X}_+)$ is an ordered Banach space. For each $i = 1, 2, 3, 4$, let $T_i(t) : C(\bar{\Omega}, \mathbb{R}) \to C(\bar{\Omega}, \mathbb{R})$ be the C_0 semigroup generated by the operator A_i subject to homogeneous Neumann boundary condition in system (1.5), where $A_1 = \nabla \cdot (f\nabla) - \mu$, $A_2 = \nabla \cdot (g\nabla) - (\sigma + \mu)$, $A_3 = \nabla \cdot (h\nabla) - (\gamma + \mu + \alpha)$ and $A_4 = d_R\Delta - \mu$. Note that $T(t) := (T_1(t), T_2(t), T_3(t), T_4(t))$ is a C_0 semigroup on $\mathbf{X}$ with an infinitesimal generator $A := (A_1, A_2, A_3, A_4)$. It then follows from Smith (1995) that $T_i(t)$ is compact and strongly positive for each $t > 0$ and $i = 1, 2, 3, 4$. Let $-\mu_i < 0$ $(i = 1, 2, 3, 4)$ be the principal eigenvalue of A_i subject to homogeneous Neumann boundary condition. Thus we have $\|T_i(t)\| \leq m_i e^{-\mu_i t}$ for each $t > 0$, where m_i $(i = 1, 2, 3, 4)$ are positive constants.

Define $\bar{F} := (F_1, F_2, F_3, F_4) : \mathbf{X}_+ \to \mathbf{X}$ by

$$\begin{aligned}
F_1(\psi)(x) &= \Lambda(x) - \beta_1(x)\psi_1(x)\int_\Omega K(x-y)\psi_2(y)dy \\
&\quad - \beta_2(x)\psi_1(x)\int_\Omega K(x-y)\psi_3(y)dy, \\
F_2(\psi)(x) &= \beta_1(x)\psi_1(x)\int_\Omega K(x-y)\psi_2(y)dy \\
&\quad + \beta_2(x)\psi_1(x)\int_\Omega K(x-y)\psi_3(y)dy, \\
F_3(\psi)(x) &= \sigma(x)\psi_2(x), \\
F_4(\psi)(x) &= \gamma(x)\psi_3(x),
\end{aligned}$$

for any $\psi = (\psi_1, \psi_2, \psi_3, \psi_4) \in \mathbf{X}_+$. Given a vector-valued function $u := (u_1, u_2, u_3, u_4) \in \mathbf{X}$, then system (1.5) can be rewritten as the following abstract differential equation:

$$\begin{cases} \dfrac{du(t)}{dt} = Au(t) + \bar{F}(u(t)), \ t > 0, \\ u(0) = \psi, \end{cases}$$

or the following integral equation:

$$u(t) = T(t)\psi + \int_0^t T(t-s)\bar{F}(u(s))ds. \tag{2.1}$$

2.1 Well-posedness

Before we establish the well-posedness of the solution to model (1.5), we first demonstrate two preliminary lemmas. According to Lou and Zhao (2011), the following statement is well-known.

Lemma 2.1 *Assume (H1)–(H2) hold. The reaction-diffusion*

$$\begin{cases} \partial_t w = \nabla \cdot (f(x)\nabla w) + \Lambda(x) - \mu(x)w, & x \in \Omega, \ t > 0, \\ \nabla w \cdot \mathbf{n} = 0, & x \in \partial\Omega \end{cases}$$

admits a unique positive steady state $w^0(x)$ which is globally asymptotic stable in $C(\bar{\Omega}, \mathbb{R})$, where $f(x)$, $\Lambda(x)$ and $\mu(x)$ are the same as those in the model (1.5). Furthermore, $w^0 = \frac{\Lambda}{\mu}$ if Λ and μ are positive constants.

For any $\psi \in \mathbf{X}_+$ and $\delta > 0$, it follows that

$$\begin{aligned}&\psi(x) + \delta \bar{F}(\psi)(x)\\ &\geq \left(\psi_1(1-\delta\beta_1(x)\int_\Omega K(x-y)\psi_2(y)dy - \delta\beta_2(x)\int_\Omega K(x-y)\psi_3(y)dy), \psi_2, \psi_3, \psi_4\right)^T.\end{aligned}$$

By choosing δ small enough, we have $\psi(x) + \delta \bar{F}(\psi)(x) \in \mathbf{X}_+$. Hence, the following subtangential condition holds

$$\lim_{\delta \to 0^+} \delta^{-1} \mathrm{dist}(\mathbf{X}_+, \psi(x) + \delta \bar{F}(\psi)(x)) = 0.$$

By Corollary 4 in Martin and Smith (1990), we have the following lemma on the existence and uniqueness of the solution to the model (1.5).

Lemma 2.2 *Assume (H1)–(H3) hold. For the initial data* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$, *model* (1.5) *admits a unique solution* (S, E, I, R) *on the maximal interval of existence* $[0, T_{\max})$, *where* $T_{\max} \leq \infty$. *Moreover, if* $T_{\max} < \infty$, *then*

$$\limsup_{t \to T_{\max}} (\|S(\cdot,t)\| + \|E(\cdot,t)\| + \|I(\cdot,t)\| + \|R(\cdot,t)\|) = \infty.$$

Let $\Phi(t) : \mathbf{X}_+ \to \mathbf{X}_+$ be the solution semiflow of the model (1.5), i.e.,

$$\Phi(t)\psi = (S(\cdot,t), E(\cdot,t), I(\cdot,t), R(\cdot,t)), \ \forall\, t \geq 0.$$

We then give the uniform boundedness of the solution to the model (1.5) based on Lemma 2.2.

Theorem 2.1 *Assume (H1)–(H3) hold. For the initial data* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$, *system* (1.5) *admits a unique solution* (S, E, I, R) *on* $\bar{\Omega} \times [0, \infty)$. *Moreover, there exists a positive constant* M *depending on the initial data such that*

$$\|S(\cdot,t)\|, \|E(\cdot,t)\|, \|I(\cdot,t)\|, \|R(\cdot,t)\| \leq M, \ \forall\, t \geq 0,$$

and there exist positive constants N *independent of the initial data such that*

$$\limsup_{t\to\infty} \|S(\cdot,t)\|, \limsup_{t\to\infty} \|E(\cdot,t)\|, \limsup_{t\to\infty} \|I(\cdot,t)\|, \limsup_{t\to\infty} \|R(\cdot,t)\| \leq N.$$

Furthermore, the solution semiflow $\Phi(t) : \mathbf{X}_+ \to \mathbf{X}_+$ *admits a global attractor, which is a nonempty, compact, and invariant set.*

Proof The existence and uniqueness of the solution on $[0, T_{\max})$ can be obtained by Lemma 2.2. It suffices to verify $T_{\max} = \infty$.

It follows from the S-equation of (1.5) that

$$\partial_t S \leq \nabla \cdot (f(x)\nabla S) + \Lambda^* - \mu_* S, \ x \in \Omega, \ t \in [0, T_{\max}).$$

By the standard comparison principle for parabolic equations and Lemma 2.1, one has

$$S(x,t) \le \frac{\Lambda^*}{\mu_*} =: M_1 = N_1,\ x \in \bar{\Omega},\ t \in [0, T_{\max}). \tag{2.2}$$

Adding the four equations in (1.5) and integrating over Ω by parts yield

$$\begin{aligned}\partial_t \int_\Omega (S+E+I+R)dx &= \int_\Omega \Lambda dx - \int_\Omega \mu(S+E+I+R)dx - \int_\Omega \alpha I dx \\ &\le \Lambda^*|\Omega| - \mu_* \int_\Omega (S+E+I+R)dx,\ t \in [0, T_{\max}).\end{aligned}$$

By Gronwall's inequality, we have

$$\begin{aligned}\int_\Omega (S+E+I+R)dx \le e^{-\mu_* t} \int_\Omega (S_0+E_0+I_0+R_0)dx \\ +(1-e^{-\mu^* t})M_1|\Omega|,\ t \in [0, T_{\max}).\end{aligned} \tag{2.3}$$

Application of (2.1) to E-equation in (1.5), one has

$$\begin{aligned}E(\cdot,t) = T_2(t)E_0 + \int_0^t T_2(t-s)\Big[\beta_1(\cdot)S(\cdot,s)\int_\Omega K(\cdot-y)E(y,s)dy \\ +\beta_2(\cdot)S(\cdot,s)\int_\Omega K(\cdot-y)I(y,s)dy\Big]ds,\ t \in [0, T_{\max}).\end{aligned}$$

Hence, by $\|T_2(t)\| \le m_2 e^{-\mu_2 t}$ and (2.2), we obtain

$$\begin{aligned}\|E(\cdot,t)\| &\le \|T_2(t)E_0\| \\ &\quad + \int_0^t \Big\| T_2(t-s)\Big[\beta_1(\cdot)S(\cdot,s)\int_\Omega K(\cdot-y)E(y,s)dy \\ &\quad +\beta_2(\cdot)S(\cdot,s)\int_\Omega K(\cdot-y)I(y,s)dy\Big]\Big\| ds \\ &\le m_2 e^{-\mu_2 t}\|E_0\| + C_1 \int_0^t e^{-\mu_2(t-s)} \int_\Omega (E(y,s)+I(y,s))dyds,\ t \in [0, T_{\max}),\end{aligned}$$

where $C_1 = m_2 \sup_{x,y\in\Omega} K(x-y)\max\{\beta_1^*, \beta_2^*\}M_1|\Omega|$. Combining this inequality with (2.3), it produces

$$\begin{aligned}\|E(\cdot,t)\| &\le m_2 e^{-\mu_2 t}\|E_0\| + C_1\int_0^t e^{-\mu_2(t-s)}\left[e^{-\mu_* s}\int_\Omega (S_0+E_0+I_0+R_0)dx + M_1|\Omega|\right]ds\\ &\le m_2 e^{-\mu_2 t}\|E_0\| + C_1 e^{-\mu_2 t}\int_0^t\left[e^{(\mu_2-\mu_*)s}\int_\Omega (S_0+E_0+I_0+R_0)dx + e^{\mu_2 s}M_1|\Omega|\right]ds\\ &\le m_2 e^{-\mu_2 t}\|E_0\| + \frac{C_1 e^{-\mu_2 t}|\Omega|}{\mu_2-\mu_*}(1-e^{-(\mu_2-\mu_*)t})\|S_0+E_0+I_0+R_0\|\\ &\quad +(1-e^{-\mu_2 t})\frac{C_1 M_1|\Omega|}{\mu_2},\ t\in[0,T_{\max}).\end{aligned} \tag{2.4}$$

Replacing μ_* by a smaller positive number such that $\mu_2-\mu_*>0$, it can be derived that

$$\begin{aligned}\|E(\cdot,t)\| &\le m_2\|E_0\| + \frac{C_1|\Omega|}{\mu_2-\mu_*}\|S_0+E_0+I_0+R_0\|\\ &\quad +\frac{C_1 M_1|\Omega|}{\mu_2} =: M_2,\ t\in[0,T_{\max}).\end{aligned} \tag{2.5}$$

It follows from the I-equation in (1.5) and (2.1) that

$$I(\cdot,t) = T_3(t)I_0 + \int_0^t T_3(t-s)\sigma E(\cdot,s)ds,\ t\in[0,T_{\max}).$$

Thus, by virtue of $\|T_3(t)\|\le m_3 e^{-\mu_3 t}$, we can get

$$\begin{aligned}\|I(\cdot,t)\| &\le \|T_3(t)I_0\| + \sigma^*\int_0^t \|T_3(t-s)E(\cdot,s)\|ds\\ &\le m_3 e^{-\mu_3 t}\|I_0\| + \frac{m_3\sigma^*\|E\|}{\mu_3}(1-e^{-\mu_3 t}),\ t\in[0,T_{\max}).\end{aligned} \tag{2.6}$$

Therefore, according to (2.6), we have

$$\|I(\cdot,t)\| \le m_3\|I_0\| + \frac{m_3\sigma^*}{\mu_3}M_2 =: M_3,\ t\in[0,T_{\max}). \tag{2.7}$$

Similar to (2.7), we can obtain

$$\|R(\cdot,t)\| \le M_4,\ t\in[0,T_{\max}). \tag{2.8}$$

It should be pointed out that the positive constants M_2, M_3, M_4 depend on the initial data. In view of (2.2), (2.5), (2.7) and (2.8), we know $T_{\max}=\infty$ by Lemma 2.2. That is to say, the above estimates hold for all $t\ge 0$. It can be observed that $M = \max\{M_1, M_2, M_3, M_4\}$.

We now prove the existence of N. It follows from (2.4) that

$$\limsup_{t\to\infty} \|E(\cdot,t)\| \le \frac{C_1 N_1 |\Omega|}{\mu_2} =: N_2. \tag{2.9}$$

It can be derived from (2.6) that

$$\limsup_{t\to\infty} \|I(\cdot,t)\| \le \frac{m_3\sigma^*}{\mu_3} N_2 =: N_3. \tag{2.10}$$

It can be found that the positive constants N_2 and N_3 are independent of the initial data. Similarly, there also exists a positive constant N_4 independent of the initial data such that

$$\limsup_{t\to\infty} \|R(\cdot,t)\| \le N_4. \tag{2.11}$$

Consequently, $N = \max\{N_1, N_2, N_3, N_4\}$. In view of (2.2), (2.9), (2.10) and (2.11), we know the solution is ultimately bounded, i.e., the solution semiflow $\Phi(t)$ is pointwisely dissipative. On the other hand, it follows from Wu (1996) that $\Phi(t)$ is continuous and compact for $t > 0$. As a result, $\Phi(t)$ admits a global attractor by Hale (1988). □

2.2 Basic reproduction number

According to Lemma 2.1, there exits a unique DFE $\mathcal{E}^0 = (S^0, 0, 0, 0)$ of model (1.5). Linearizing the second and third equations of model (1.5) at $\mathcal{E}^0$ gives

$$\begin{cases} \partial_t \xi - \nabla\cdot(g\nabla\xi) = \beta_1 S^0 \displaystyle\int_\Omega K(x-y)\xi(y,t)dy \\ \quad + \beta_2 S^0 \displaystyle\int_\Omega K(x-y)\zeta(y,t)dy - (\sigma+\mu)\xi, & x\in\Omega,\ t>0, \\ \partial_t \zeta - \nabla\cdot(h\nabla\zeta) = \sigma\xi - (\gamma+\mu+\alpha)\zeta, & x\in\Omega,\ t>0, \\ \nabla\xi\cdot\mathbf{n} = \nabla\zeta\cdot\mathbf{n} = 0, & x\in\partial\Omega,\ t>0. \end{cases} \tag{2.12}$$

Setting $(\xi,\zeta) = e^{\lambda t}(\psi_E, \psi_I)$ in the linearized system (2.12), we obtain the following eigenvalue problem of model (1.5):

$$\begin{cases} \nabla\cdot(g\nabla\psi_E) + \beta_1 S^0 \displaystyle\int_\Omega K(x-y)\psi_E(y)dy + \beta_2 S^0 \int_\Omega K(x-y)\psi_I(y)dy \\ \quad - (\sigma+\mu)\psi_E = \lambda\psi_E, & x\in\Omega, \\ \nabla\cdot(h\nabla\psi_I) + \sigma\varphi_E - (\gamma+\mu+\alpha)\psi_I = \lambda\psi_I, & x\in\Omega, \\ \nabla\psi_E\cdot\mathbf{n} = \nabla\psi_I\cdot\mathbf{n} = 0, & x\in\partial\Omega. \end{cases} \tag{2.13}$$

It then follows from the Krein-Rutman theorem (Hess 1991) that the eigenvalue problem (2.13) admits a unique principal eigenvalue λ^0 associated with a positive eigenfunction (ψ_E^0, ψ_I^0).

We adopt the next-generation operator approach developed in Wang and Zhao (2012) to derive the basic reproduction of model (1.5) with symmetric diffusion, then L, $F(x)$ and $V(x)$ in Wang and Zhao (2012) can be respectively defined as $L = \mathrm{diag}(\nabla \cdot (g\nabla), \nabla \cdot (h\nabla))$,

$$F(x) = \begin{pmatrix} \beta_1 S^0 \frac{\partial \int_\Omega K(x-y)\xi dy}{\partial \xi}|_{\xi=0} & \beta_2 S^0 \frac{\partial \int_\Omega K(x-y)\zeta dy}{\partial \zeta}|_{\zeta=0} \\ 0 & 0 \end{pmatrix},$$

$$V(x) = \begin{pmatrix} \sigma+\mu & 0 \\ -\sigma & \gamma+\mu+\alpha \end{pmatrix}.$$

Note that both E and I are infected compartments, and $F(x)$ represents the infection process. $L - V$ means the synthetical influences of L and V, where L signifies the mobility of individuals, and $-V$ measures the mortality and transfer of individuals in infected compartments. Let $\Psi(t) : C(\bar{\Omega}, \mathbb{R}^2) \to C(\bar{\Omega}, \mathbb{R}^2)$ be the solution semigroup generated by $L - V$ with zero-flux boundary condition. Suppose the distribution of initial infection described by $\bar{\psi}(x) = (\psi_E, \psi_I)$ and then the distribution of those infective members becomes $\Psi(t)\bar{\psi}(x)$ as time evolves. Hence, the distribution of new infection at time t is $F(x)\Psi(t)\bar{\psi}(x)$. The distribution of total new infections is

$$\int_0^\infty F(x)\Psi(t)\bar{\psi}(x)dt = F(x)\int_0^\infty \Psi(t)\bar{\psi}(x)dt =: \mathcal{L}(\bar{\psi})(x).$$

Then the basic reproduction number of the model (1.5) can be defined by the spectral radius of $\mathcal{L}$ as

$$\mathcal{R}_0^{(1.5)} = r(\mathcal{L}).$$

According to Wang and Zhao (2012), we have the following property between the basic reproduction number of the model (1.5) and the principal eigenvalue of the problem (2.13).

Lemma 2.3 *The following assertion of $\mathcal{R}_0^{(1.5)}$ and λ^0 is valid:*

$$\mathrm{sign}(\mathcal{R}_0^{(1.5)} - 1) = \mathrm{sign}(\lambda^0).$$

2.3 Threshold dynamics

In terms of the basic reproduction number $\mathcal{R}_0^{(1.5)}$, we establish the threshold dynamics including disease extinction and persistence for model (1.5).

Theorem 2.2 *Assume (H1)–(H3) hold. The following statements hold:*

(i) *If $\mathcal{R}_0^{(1.5)} < 1$, then $\mathcal{E}^0$ is globally asymptotically stable;*

(ii) *If* $\mathcal{R}_0^{(1.5)} > 1$, *model* (1.5) *is uniformly persistent, i.e., there exists a constant* $\eta > 0$ *such that for any* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$ *with* $E_0 \not\equiv 0$ *or* $I_0 \not\equiv 0$, *the solution* (S, E, I, R) *satisfies*

$$\liminf_{t\to\infty} S(x,t), \liminf_{t\to\infty} E(x,t), \liminf_{t\to\infty} I(x,t), \liminf_{t\to\infty} R(x,t) \geq \eta, \ \forall x \in \Omega.$$

Moreover, model (1.5) *admits at least one EE.*

Proof We first prove (i). The local asymptotically stability of $\mathcal{E}^0$ follows from Theorem 3.1 of Wang and Zhao (2012). We only verify the global attractivity of $\mathcal{E}^0$. Fix $\epsilon_1 > 0$. There exists $T_1 > 0$ such that

$$0 \leq S(\cdot, t) \leq S^0(\cdot) + \epsilon_1, \ \forall\, t \geq T_1.$$

Let $(\hat{E}(x,t), \hat{I}(x,t))$ be the solution to the following problem

$$\begin{cases} \partial_t \hat{E} - \nabla \cdot (g \nabla \hat{E}) = \beta_1 (S^0 + \epsilon_1) \displaystyle\int_\Omega K(x-y)\hat{E}(y,t)dy \\ \quad + \beta_2 (S^0 + \epsilon_1) \displaystyle\int_\Omega K(x-y)\hat{I}(y,t)dy - (\sigma + \mu)\hat{E}, & x \in \Omega,\ t > T_1, \\ \partial_t \hat{I} - \nabla \cdot (h \nabla \hat{I}) = \sigma \hat{E} - (\gamma + \mu + \alpha)\hat{I}, & x \in \Omega,\ t > T_1, \\ \nabla \hat{E} \cdot \mathbf{n} = \nabla \hat{I} \cdot \mathbf{n} = 0, & x \in \partial\Omega,\ t > T_1. \end{cases}$$

Suppose there exists a constant $c_1 > 0$ such that $c_1(\hat{\psi}_E, \hat{\psi}_I) \geq (E(x, T_1), I(x, T_1))$ for $x \in \Omega$, where $(\hat{\psi}_E, \hat{\psi}_I)$ is the positive eigenfunction corresponding to the principal eigenvalue $\lambda^0(S^0 + \epsilon_1)$ of the elliptic eigenvalue problem (2.13) with S^0 replaced by $S^0 + \epsilon_1$. It follows from Lemma 2.3 that $\lambda^0(S^0 + \epsilon_1) < 0$ since $\mathcal{R}_0^{(1.5)} < 1$. Then by the standard comparison principle for parabolic equations, one has

$$(E(\cdot,t), I(\cdot,t)) \leq (\hat{E}(\cdot,t), \hat{I}(\cdot,t)) = c_1(\hat{\psi}_E, \hat{\psi}_I) e^{\lambda^0(S^0+\epsilon_1)(t-T_1)}, \ \forall t \geq T_1.$$

Hence,

$$\lim_{t\to\infty} (E(\cdot,t), I(\cdot,t)) = (0,0).$$

Moreover, it follows from the R-equation that $\lim_{t\to\infty} R(x,t) = 0$ for $x \in \Omega$. Then by Lemma 2.1, one has $\lim_{t\to\infty} S(x,t) = S^0(x)$ for $x \in \Omega$. This completes the proof of assertion (i).

We then verify (ii). Denote

$$U_0 := \{\varphi = (\varphi_1, \varphi_2, \varphi_3, \varphi_4) \in \mathbf{X}_+ | \varphi_2 \neq 0 \text{ and } \varphi_3 \neq 0\},$$
$$\partial U_0 := \{\varphi \in \mathbf{X}_+ | \varphi_2 = 0 \text{ or } \varphi_3 = 0\}.$$

Note that $\mathbf{X}_+ = U_0 \cup \partial U_0$. Moreover, U_0 and ∂U_0 are relatively open and closed subsets of $\mathbf{X}_+$, respectively. The solution semiflow $\Phi(t)$ of the model (1.5) admits a global attractor indicated by Theorem 2.1. Set $M_\partial = \{\varphi \in \partial U_0 : \Phi(t)\varphi \in \partial U_0,\ \forall t > 0\}$.

Claim 1. $\Phi(t)U_0 \subset U_0,\ \forall t > 0$.

It follows directly from the strong maximum principle and Hopf boundary lemma for parabolic equations.

Claim 2. $\bigcup_{\varphi \in M_\partial} \omega(\varphi) = \{\mathcal{E}^0\}$, where $\omega(\varphi)$ is the omega limit set of the forward orbit $\gamma^+(\varphi) = \{\Phi(t)\varphi : t \geq 0\}$.

Since $\varphi \in M_\partial$, we know $\varphi_2 = \varphi_3 = 0$, then $E(x,t) = I(x,t) = R(x,t) = 0$ for all $x \in \bar{\Omega}$, $t \geq 0$. As a result, $S(\cdot,t) \to S^0$ uniformly as $t \to \infty$, implying the claim holds, i.e., $\{\mathcal{E}^0\}$ is a compact and isolated invariant set for $\Phi(t)$ restricted in M_∂.

Claim 3. There exists some positive constant ϵ_2 independent of initial data such that

$$\limsup_{t\to\infty} \|\Phi(t)\varphi - \mathcal{E}^0\| > \epsilon_2.$$

Argue indirectly. Suppose there exists $\bar{\varphi} \in U_0$ such that $\limsup_{t\to\infty} \|\Phi(t)\bar{\varphi} - \mathcal{E}^0\| < \epsilon_2$. Then there exists $T_2 > 0$ such that

$$S(\cdot,t) > S^0(\cdot) - \epsilon_2,\ \forall t \geq T_2.$$

Let $(\check{E}(x,t), \check{I}(x,t))$ be the solution to the following parabolic equations:

$$\begin{cases} \partial_t \check{E} - \nabla\cdot(g\nabla\check{E}) = \beta_1(S^0-\epsilon_2)\displaystyle\int_\Omega K(x-y)\check{E}(y,t)dy \\ \quad + \beta_2(S^0-\epsilon_2)\displaystyle\int_\Omega K(x-y)\check{I}(y,t)dy - (\sigma+\mu)\check{E}, & x\in\Omega,\ t > T_2, \\ \partial_t \check{I} - \nabla\cdot(h\nabla\check{I}) = \sigma\check{E} - (\gamma+\mu+\alpha)\check{I}, & x\in\Omega,\ t > T_2, \\ \nabla\check{E}\cdot\mathbf{n} = \nabla\check{I}\cdot\mathbf{n} = 0, & x\in\partial\Omega,\ t > T_2. \end{cases}$$

Assume there exists a constant $c_2 > 0$ such that $c_2(\check{\psi}_E, \check{\psi}_I) \geq (E(x,T_2), I(x,T_2))$ for $x \in \Omega$, where $(\check{\psi}_E, \check{\psi}_I)$ is the positive eigenfunction corresponding to the principal eigenvalue $\lambda^0(S^0-\epsilon_2)$ of the elliptic eigenvalue problem (2.13) with S^0 replaced by $S^0-\epsilon_2$. Then Lemma 2.3 gives $\lambda^0(S^0+\epsilon_0) > 0$ since $\mathcal{R}_0^{(1.5)} > 1$. By the comparison principle, one has

$$(E(x,t), I(x,t)) \geq (\hat{E}(x,t), \hat{I}(x,t)) = c_2(\check{\psi}_E, \check{\psi}_I)e^{\lambda^0(S^0-\epsilon_2)(t-T_2)},\ \forall t \geq T_2.$$

Hence,

$$\lim_{t\to\infty}(E(\cdot,t), I(\cdot,t)) = (\infty,\infty),$$

a contradiction to Theorem 2.1. This verifies Claim 3.

According to the above claims, the uniform persistence can be derived by Zhao (2017). It then follows from Magal and Zhao (2005) that there exists at least one positive steady state. □

3 The epidemic model with random diffusion

Let $fS =: \tilde{S}$, $gE =: \tilde{E}$ and $hI =: \tilde{I}$. We then obtain the following equivalent system of the system (1.6):

$$\begin{cases} \partial_t \tilde{S} - f\Delta\tilde{S} = f\Lambda(x) - \beta_1(x)\tilde{S}\int_\Omega K(x-y)\tilde{E}(y,t)g^{-1}dy \\ \quad - \beta_2(x)\tilde{S}\int_\Omega K(x-y)\tilde{I}(y,t)h^{-1}dy - \mu(x)\tilde{S}, & x\in\Omega,\ t>0, \\ \partial_t \tilde{E} - g\Delta\tilde{E} = \beta_1(x)gf^{-1}\tilde{S}\int_\Omega K(x-y)\tilde{E}(y,t)g^{-1}dy \\ \quad + \beta_2(x)gf^{-1}\tilde{S}\int_\Omega K(x-y)\tilde{I}(y,t)h^{-1}dy - (\sigma(x)+\mu(x))\tilde{E}, & x\in\Omega,\ t>0, \\ \partial_t \tilde{I} - h\Delta\tilde{I} = \sigma(x)hg^{-1}\tilde{E} - (\gamma(x)+\mu(x)+\alpha(x))\tilde{I}, & x\in\Omega,\ t>0, \\ \partial_t R - d_R\Delta R = \gamma(x)h^{-1}\tilde{I} - \mu(x)R, & x\in\Omega,\ t>0, \\ \nabla\tilde{S}\cdot\mathbf{n} = \nabla\tilde{E}\cdot\mathbf{n} = \nabla\tilde{I}\cdot\mathbf{n} = \nabla R\cdot\mathbf{n} = 0, & x\in\partial\Omega,\ t>0, \\ \tilde{S}(x,0) = fS_0(x),\ \tilde{E}(x,0) = gE_0(x), \tilde{I}(x,0) = hI_0(x),\ R(x,0) = R_0(x), & x\in\Omega. \end{cases} \tag{3.1}$$

For each $i = 1,2,3,4$, let $J_i(t) : C(\bar{\Omega},\mathbb{R}) \to C(\bar{\Omega},\mathbb{R})$ be the C_0 semigroups generated by the operator B_i subject to homogeneous Neumann boundary condition in (3.1), where $B_1 = f\Delta - \mu$, $B_2 = g\Delta - (\sigma+\mu)$, $B_3 = h\Delta - (\gamma+\mu+\alpha)$ and $B_4 = d_R\Delta - \mu$. Note that $B_4 = A_4$ and $J_4 = T_4$. Hence, $J(t) := (J_1(t), J_2(t), J_3(t), J_4(t))$ is a C_0 semigroup on $\mathbf{X}$ with an infinitesimal generator $B := (B_1, B_2, B_3, B_4)$. With reference to Smith (1995), $J_i(t)$ is compact and strongly positive for each $t > 0$ and $i = 1,2,3,4$. Let $-\nu_i < 0$ $(i = 1,2,3,4)$ be the principal eigenvalue of B_i subject to homogeneous Neumann boundary condition. Thus we have $\|J_i(t)\| \le \tilde{m}_i e^{-\nu_i t}$ for each $t > 0$, where $\tilde{m}_i$ $(i = 1,\dots,4)$ are positive constants. It is obvious that $\nu_4 = \mu_4$ and $\tilde{m}_4 = m_4$. Set $\tilde{F} = (\tilde{F}_1, \tilde{F}_2, \tilde{F}_3, \tilde{F}_4) : \mathbf{X}_+ \to \mathbf{X}$, where

$$\begin{aligned} \tilde{F}_1(\phi)(x) &= f\Lambda(x) - \beta_1(x)\phi_1(x)\int_\Omega K(x-y)\phi_2(y)g^{-1}dy - \beta_2(x)\phi_1(x) \\ &\quad \int_\Omega K(x-y)\psi_3(y)h^{-1}dy, \\ \tilde{F}_2(\phi)(x) &= \beta_1(x)gf^{-1}\phi_1(x)\int_\Omega K(x-y)\phi_2(y)g^{-1}dy + \beta_2(x)gf^{-1}\phi_1(x) \\ &\quad \int_\Omega K(x-y)\phi_3(y)h^{-1}dy, \\ \tilde{F}_3(\phi)(x) &= \sigma(x)hg^{-1}\phi_2(x), \\ \tilde{F}_4(\phi)(x) &= \gamma(x)h^{-1}\phi_3(x), \end{aligned}$$

for any $\phi = (\phi_1, \phi_2, \phi_3, \phi_4) \in \mathbf{X}_+$.

Similar to the system (1.5), given a vector-valued function $\omega := (\omega_1, \omega_2, \omega_3, \omega_4) \in \mathbf{X}$, system (1.6) can be rewritten as the following abstract differential equation:

$$\begin{cases} \dfrac{d\omega(t)}{dt} = B\omega(t) + \tilde{F}(\omega(t)), \ t > 0, \\ \omega(0) = \phi, \end{cases}$$

or the following integral equation:

$$\omega(t) = J(t)\phi + \int_0^t J(t-s)\tilde{F}(\omega(s))ds. \tag{3.2}$$

3.1 Well-posedness

Let $\tilde{\Phi}(t) : \mathbf{X}_+ \to \mathbf{X}_+$ be the solution semiflow of the model (1.6). Similar to Theorem 2.1, we then give the well-posedness of the model (1.6).

Theorem 3.1 *Assume (H1)–(H3) hold. For the initial data* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$, *model* (1.6) *admits a unique solution* (S, E, I, R) *on* $\bar{\Omega} \times [0, \infty)$. *Moreover, there exists a positive constant* $\tilde{M}$ *depending on the initial data such that*

$$\|S(\cdot, t)\|, \|E(\cdot, t)\|, \|I(\cdot, t)\|, \|R(\cdot, t)\| \le \tilde{M}, \ \forall\, t \ge 0, \tag{3.3}$$

and there exists positive constants $\tilde{N}$ *independent of the initial data such that*

$$\limsup_{t\to\infty} \|S(\cdot, t)\|, \limsup_{t\to\infty} \|E(\cdot, t)\|, \limsup_{t\to\infty} \|I(\cdot, t)\|, \limsup_{t\to\infty} \|R(\cdot, t)\| \le \tilde{N}. \tag{3.4}$$

Furthermore, the solution semiflow $\tilde{\Phi}(t)$ *admits a global attractor, which is a nonempty, compact, and invariant set.*

Proof It follows from the variation of constant formula (3.2) that

$$\tilde{S}(t) \le J_1(t)(fS_0) + \int_0^t J_1(s)(f\Lambda(x))ds, \ t \in [0, \tilde{T}_{\max}),$$

where $[0, \tilde{T}_{\max})$ is the maximal existence interval for the solution to (3.1). Hence, by assumption (H2),

$$\begin{aligned} \|\tilde{S}(t)\| &\le \tilde{m}_1 M_0 e^{-\nu_1 t}\|S_0\| + \frac{\tilde{m}_1 M_0 \Lambda^*}{\nu_1}(1 - e^{-\nu_1 t}) \\ &\le \tilde{m}_1 M_0 \|S_0\| + \frac{\tilde{m}_1 M_0 \Lambda^*}{\nu_1} =: \tilde{M}_1, \ t \in [0, \tilde{T}_{\max}). \end{aligned} \tag{3.5}$$

Here, $\tilde{M}_1$ depends on the initial data. In order to obtain the L^1-estimate of $\tilde{S}+\tilde{E}+\tilde{I}+R$, we add the four equations in (1.6), and integrate the result by parts over Ω. Then we can also get (2.3) due to the zero-flux boundary conditions. Moreover,

$$\int_\Omega (\tilde{S}+\tilde{E}+\tilde{I}+R)dx \le \max\{M_0,1\}\int_\Omega (S+E+I+R)dx.$$

Combining this inequality with (2.3), we obtain

$$\begin{aligned}\int_\Omega (\tilde{S}+\tilde{E}+\tilde{I}+R)dx \le{}& \max\{M_0,1\}e^{-\mu_* t}\int_\Omega (S_0+E_0+I_0+R_0)dx\\ &+(1-e^{-\mu^* t})\max\{M_0,1\}\frac{\Lambda^*}{\mu_*}|\Omega|,\ t\in[0,\tilde{T}_{\max}).\end{aligned}$$

Analogous to the proof of Theorem 2.1, we can then get similar estimates as those in the proof of Theorem 2.1 for the system (3.1). It then follows from Pao (2012) that $\tilde{T}_{\max}=\infty$. This gives the global existence and boundedness of the system (3.1), that is, there exists a positive constant M' depending on the initial data such that

$$\|\tilde{S}(\cdot,t)\|, \|\tilde{E}(\cdot,t)\|, \|\tilde{I}(\cdot,t)\|, \|\tilde{R}(\cdot,t)\| \le M',\ \forall\, t\ge 0.$$

Since $\tilde{T}_{\max}=\infty$, it follows from (3.5) that

$$\limsup_{t\to\infty}\|\tilde{S}(\cdot,t)\| \le \frac{\tilde{m}_1 M_0\Lambda^*}{\nu_1} =: \tilde{N}_1$$

Here, $\tilde{N}_1$ is independent of the initial data. In the same manner, we can get some positive constants independent of the initial data such that $\tilde{E}$, $\tilde{I}$ and $\tilde{R}$ are also ultimately bounded. Hence, there exists a positive constant N' independent of the initial data such that

$$\limsup_{t\to\infty}\|\tilde{S}(\cdot,t)\|, \limsup_{t\to\infty}\|\tilde{E}(\cdot,t)\|, \limsup_{t\to\infty}\|\tilde{I}(\cdot,t)\|, \limsup_{t\to\infty}\|R(\cdot,t)\| \le N'.$$

According to $S=f^{-1}\tilde{S}$, $E=g^{-1}\tilde{E}$, and $I=h^{-1}\tilde{I}$, one has (3.3) and (3.4) also hold according to assumption (H2). There also exists a global attractor with reference to Hale (1988). □

3.2 Basic reproduction number

In view of Remark 2.2 in Wang and Zhao (2012), we can also define the basic reproduction number for model (1.6) using the next-generation operator approach. First, we provide the following lemma, where the symmetric diffusion in Lemma 2.1 is replaced by the random diffusion. The proof follows from a similar procedure of Claim 1 in the proof Theorem 2.2 in Freedman and Zhao (1997).

Lemma 3.1 *Assume (H1)–(H2) hold. The reaction-diffusion*

$$\begin{cases} \partial_t v = \Delta(f(x)v) + \Lambda(x) - \mu(x)v, & x \in \Omega,\ t > 0, \\ \nabla(fv)\cdot \mathbf{n} = 0, & x \in \partial\Omega, \\ v(x,0) = \psi, & x \in \Omega. \end{cases} \tag{3.6}$$

admits a unique positive steady state $v^0(x)$ which is globally asymptotic stable in $C(\bar{\Omega}, \mathbb{R})$, where $f(x)$, $\Lambda(x)$ and $\mu(x)$ are the same as those in the model (1.6).

Proof By virtue of standard theory for parabolic equations, one has (3.6) admits a unique solution $v(t,\psi)$ on $[0,\infty)$ with $\psi \in C(\bar{\Omega}, \mathbb{R}_+)$. Let $Q(t)$ be the solution semiflow of $\Delta(f\cdot) - \mu$ with zero-flux boundary condition, i.e., $Q(t)\psi = v(t,\psi)$.

Let $W := f(x)v$. Then (3.6) becomes

$$\begin{cases} \partial_t W = f\Delta W + f\Lambda(x) - \mu(x)W, & x \in \Omega,\ t > 0, \\ \nabla W \cdot \mathbf{n} = 0, & x \in \partial\Omega, \\ W(x,0) = f\psi, & x \in \Omega. \end{cases}$$

By standard comparison arguments and assumption (H2), one has

$$\frac{m_0\Lambda_*}{\mu^*} \le W \le \frac{M_0\Lambda^*}{\mu_*}.$$

Then for any $\psi \in C(\bar{\Omega}, \mathbb{R})$, the omega limit set $\omega(\psi)$ satisfies

$$\omega(\psi) \subseteq \left\{ \varphi : \frac{m_0\Lambda_*}{M_0\mu^*} \le \varphi \le \frac{M_0\Lambda^*}{m_0\mu_*} \right\}.$$

Hence, for any $\varphi > \psi$, we can infer $Q(t)\varphi \gg Q(t)\psi$ for all $t > 0$.

Let $h(x,v) := \Lambda(x) - \mu(x)v$. It is clear that $h(x,v)$ is strictly subhomogeneous in the sense that $h(x,\tau v) > \tau h(x,v)$ for any $\tau \in (0,1)$ and $v \gg 0$.

For any $\psi \in C(\bar{\Omega}, \mathbb{R}_+)$ and $\varsigma \in (0,1)$, let $V(t) = v(t,\varsigma\psi) - \varsigma v(t,\psi)$. It follows from (3.6) that $V(0) = 0$ and $V(t)$ satisfies

$$\begin{cases} \partial_t V = \Delta(fV) + H(t,x)V + h(x,\varsigma v(t,\psi)) - \varsigma h(x, v(t,\psi)), & x \in \Omega,\ t > 0, \\ \nabla(fV)\cdot \mathbf{n} = 0, & x \in \partial\Omega, \end{cases}$$

where

$$H(t,x) = \int_0^1 \frac{\partial h(x, sv(t,\varsigma\psi) + (1-s)v(t,\psi))}{\partial v} ds.$$

Let $U(t,s)$, $t \geq s \geq 0$, be the evolution operator of the nonautonomous linear parabolic system:

$$\begin{cases} \partial_t V = \Delta(fV) + H(t,x)V, & x \in \Omega,\ t > 0, \\ \nabla(fV)\cdot \mathbf{n} = 0, & x \in \partial\Omega. \end{cases}$$

It follows from Theorem 7.4.1 in Smith (1995) that $U(t,s)$ is strongly positive by a transformation, i.e., $U(t,s)\phi \gg 0$ for any $\phi > 0$. According to the formula of variants of constants,

$$V(t) = \int_0^t U(t,s)\,[h(\cdot, \varsigma v(s,\psi)) - \varsigma h(\cdot, v(s,\psi))]\,ds,\ t \geq 0.$$

Note that $v(s,\phi) \gg 0$ for $s \gg 0$ and $h(x,v)$ is strictly subhomogeneous in v, one has

$$h(\cdot, \varsigma v(s,\psi)) - \varsigma h(\cdot, v(s,\psi))ds > 0 \text{ in } C(\bar{\Omega}, \mathbb{R}),$$

As a result, $V(t) \gg 0$ for all $t > 0$, that is, $Q(t)(\varsigma\psi) > \varsigma Q(t)\psi$ for $\varsigma \in (0,1)$ and $\psi \gg 0$.

Hence, by Zhao (2017), one has $Q(t)$ has a positive equilibrium $v^*(x)$ such that $\omega(\psi) = v^* \in C(\bar{\Omega}, \mathbb{R})$ for any $\psi \in C(\bar{\Omega}, \mathbb{R})$. □

It then follows from Lemma 3.1 that there exits a unique DFE $\bar{\mathcal{E}}^0 = (\bar{S}^0, 0, 0, 0)$ for model (1.6). Linearizing the E- and I-equations in model (1.6) gives

$$\begin{cases} \partial_t\theta - \Delta(g\theta) = \beta_1\bar{S}^0\displaystyle\int_\Omega K(x-y)\theta(y,t)dy \\ \quad + \beta_2\bar{S}^0\displaystyle\int_\Omega K(x-y)\vartheta(y,t)dy - (\sigma+\mu)\theta, & x \in \Omega,\ t > 0, \\ \partial_t\vartheta - \Delta(h\vartheta) = \sigma\theta - (\gamma+\mu+\alpha)\vartheta, & x \in \Omega,\ t > 0, \\ \nabla(g\theta)\cdot\mathbf{n} = \nabla(h\vartheta)\cdot\mathbf{n} = 0, & x \in \partial\Omega,\ t > 0. \end{cases} \tag{3.7}$$

Setting $(\theta, \vartheta) = e^{\lambda t}(\phi_E, \phi_I)$, the linearized system (3.7) becomes

$$\begin{cases} \Delta(g\phi_E) + \beta_1\bar{S}^0\displaystyle\int_\Omega K(x-y)\phi_E(y)dy + \beta_2\bar{S}^0\int_\Omega K(x-y)\phi_I(y)dy \\ \quad - (\sigma+\mu)\phi_E = \lambda\phi_E, & x \in \Omega, \\ \Delta(h\phi_I) + \sigma\phi_E - (\gamma+\mu+\alpha)\phi_I = \lambda\phi_I, & x \in \Omega, \\ \nabla(g\phi_E)\cdot\mathbf{n} = \nabla(h\phi_I)\cdot\mathbf{n} = 0, & x \in \partial\Omega. \end{cases} \tag{3.8}$$

It then follows from the Krein-Rutman theorem that (3.8) admits a unique principal eigenvalue $\bar{\lambda}^0$ equipped with a positive eigenfunction (ϕ_E^0, ϕ_I^0).

For the random diffusion model (1.6), L, $F(x)$ in Wang and Zhao (2012) can be respectively defined as $L = \mathrm{diag}(\Delta(g\cdot), \Delta(h\cdot))$,

$$F(x) = \begin{pmatrix} \beta_1 \bar{S}^0 \frac{\partial \int_\Omega K(x-y)\theta dy}{\partial \theta}|_{\theta=0} & \beta_2 \bar{S}^0 \frac{\partial \int_\Omega K(x-y)\vartheta dy}{\partial \vartheta}|_{\vartheta=0} \\ 0 & 0 \end{pmatrix},$$

and $V(x)$ is the same as that in Sect. 2.

Let $\bar{\Psi}(t) : C(\bar{\Omega}, \mathbb{R}^2) \to C(\bar{\Omega}, \mathbb{R}^2)$ be the solution semigroup generated by $L - V$ associated with zero-flux boundary condition, and $\bar{\phi}(x) = (\phi_E, \phi_I)$ the distribution of initial infection. Then the basic reproduction number for model (1.6) is defined by the spectral radius of $\bar{\mathcal{L}}$ as

$$\mathcal{R}_0^{(1.6)} = r(\bar{\mathcal{L}}),$$

where

$$\bar{\mathcal{L}}(\bar{\phi})(x) := \int_0^\infty F(x)\bar{\Psi}(t)\bar{\phi}(x)dt = F(x)\int_0^\infty \bar{\Psi}(t)\bar{\phi}(x)dt. \tag{3.9}$$

Lemma 3.2 *The following assertion of $\mathcal{R}_0^{(1.6)}$ and $\bar{\lambda}^0$ is valid:*

$$\mathrm{sign}(\mathcal{R}_0^{(1.6)} - 1) = \mathrm{sign}(\bar{\lambda}^0).$$

Proof Let $B^0 := L - V$. Denote by $\mathbf{X}^1 := C(\bar{\Omega}, \mathbb{R}^2)$ and $\mathbf{X}^1_+ := C(\bar{\Omega}, \mathbb{R}^2_+)$. It follows from the maximum principle that $\bar{\Psi}(t)$ is strongly positive, i.e., $\bar{\Psi}(t)\mathbf{X}^1_+ \subseteq \mathbf{X}^1_+$ for all $t \geq 0$. According to Theorem 3.12 in Thieme (2009), B^0 is resolvent-positive, and

$$(\lambda I - B^0)^{-1}\bar{\phi} = \int_0^\infty e^{-\lambda t}\bar{\Psi}(t)\bar{\phi}dt, \ \forall \lambda > \bar{\lambda}^0, \ \bar{\phi} \in \mathbf{X}^1. \tag{3.10}$$

Note $\bar{\lambda}^0 < 0$. Let $\lambda = 0$ in (3.10), one has

$$-(B^0)^{-1}\bar{\phi} = \int_0^\infty \bar{\Psi}(t)\bar{\phi}dt, \ \forall \bar{\phi} \in \mathbf{X}^1.$$

By (3.9), one has $\bar{\mathcal{L}} = -F(B^0)^{-1}$. Define $D := B^0 + F$, which generates a positive C_0 semigroup. By Theorem 3.12 in Thieme (2009), D is resolvent-positive. Then by Theorem 3.5 in Thieme (2009), $\bar{\lambda}^0 = s(D)$ has the same sign as $r(\bar{\mathcal{L}}) - 1 = \mathcal{R}_0^{(1.6)} - 1$, where $s(D)$ is the spectral bound of D defined by

$$s(D) = \sup\{\mathrm{Re}\lambda : \lambda \in \sigma(D), \sigma(D) \text{ is the spectrum of } D\}.$$

□

3.3 Threshold dynamics

Similar to Theorem 2.2, we can obtain the following threshold dynamics according to $\mathcal{R}_0^{(1.6)}$.

Theorem 3.2 *Assume (H1)–(H3) hold. The following assertions hold:*

(i) *If* $\mathcal{R}_0^{(1.6)} < 1$, *then* $\bar{\mathcal{E}}^0$ *is globally asymptotically stable;*

(ii) *If* $\mathcal{R}_0^{(1.6)} > 1$, *then model* (1.6) *is uniformly persistent, i.e., there exists* $\eta_0 > 0$ *such that for any* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$ *with* $E_0 \not\equiv 0$ *or* $I_0 \not\equiv 0$, *the solution* (S, E, I, R) *satisfies*

$$\liminf_{t\to 0} S(x,t), \liminf_{t\to 0} E(x,t), \liminf_{t\to 0} I(x,t), \liminf_{t\to 0} R(x,t) \geq \eta_0, \ \forall x \in \Omega.$$

Moreover, model (1.6) *admits at least one EE.*

Proof The proof for global attractivity of the DFE $\bar{\mathcal{E}}^0$ is similar to that of $\mathcal{E}^0$ in Theorem 2.2(i) in view of Lemma 3.1. The proof of the uniform persistence for the system (1.6) is similar to that of the system (1.5) in Theorem 2.2(ii). Hence we omit it here. We only give the locally asymptotic stability of $\bar{\mathcal{E}}^0$ of the system (1.6).

Indeed, let $P(t)$ be the solution semigroup generated by D subject to the zero-flux boundary conditions, and $\omega(P)$ be the exponential growth bound of $P(t)$. Since $\mathcal{R}_0^{(1.6)} < 1$, one has $s(D) < 0$ by the proof of Lemma 3.2. By Theorem 3.14 in Thieme (2009), $\omega(P) = s(D) < 0$. Since the R-equation in the model (1.6) can be decoupled by the other three equations, we have the linearized S-equation of the model (1.6) at the DFE $\bar{\mathcal{E}}^0$:

$$\begin{cases} \partial_t U - \Delta(fU) = -\mu U - \beta_1 \bar{S}^0 \displaystyle\int_\Omega K(x-y)\theta(y,t)dy \\ \quad - \beta_2 \bar{S}^0 \displaystyle\int_\Omega K(x-y)\vartheta(y,t)dy, & x \in \Omega,\ t > 0, \\ \nabla(fU)\cdot \mathbf{n} = 0, & x \in \partial\Omega,\ t > 0, \end{cases}$$

where θ and ϑ are given in (3.7). It is obvious that $s(\Delta(f\cdot) - \mu) < 0$ satisfies the condition (A5) in Wang and Zhao (2012). Let $T(t)$ be the solution semigroup of the first three equations in the model (1.6). Denote by $\omega(T)$ the exponential growth bound of $T(t)$. Then $\omega(T) < 0$. It then follows from Theorem 2.1 in Desch and Schappacher (2006) that $\bar{\mathcal{E}}^0$ is locally asymptotically stable. □

4 The local epidemic model in a homogeneous case

Let $K_a(x) := \frac{1}{a^k} K(\frac{x}{a})$ for $x \in \Omega$ and $a > 0$. In fact, the function $K_a(x)$ is smooth and satisfies $\int_{\mathbb{R}^n} K_a(x)dx = 1$ with compact support in the open ball $B(0,a)$. The readers can refer to Appendix C of Evans (2022) for more properties for the function $K_a(x)$. It is clear that the support size of K_a decreases as a decreases, indicating the

effective infection area decreases. As $a \to 0^+$, K_a converges to the Dirac function δ_0. If we consider $K(\cdot)$ in models (1.5) and (1.6) replaced by K_a, then the nonlocal infection terms become the local infection terms, i.e.,

$$\beta_1 S \int_\Omega K_a(x-y)E(y,t)dy + \beta_2 S \int_\Omega K_a(x-y)I(y,t)dy \to \beta_1 SE + \beta_2 SI, \text{ as } a \to 0^+.$$

This implies the connection between nonlocal infection and local infection.

In this section, we consider systems (1.5) and (1.6) with local infection in a homogeneous environment, where Λ, μ, σ, γ, α and $\beta_i (i = 1, 2)$ are positive constants. Then $f(\sum_{i=1}^2 a_i\beta_i)$, $g(\sum_{i=1}^2 b_i\beta_i)$ or $g(\gamma^{-1})$, and $h(\gamma^{-1})$ are also positive constants. As a result, systems (1.5) and (1.6) degenerate to

$$\begin{cases} \partial_t S - f\Delta S = \Lambda - \beta_1 SE - \beta_2 SI - \mu S, & x \in \Omega,\ t > 0, \\ \partial_t E - g\Delta E = \beta_1 SE + \beta_2 SI - (\sigma+\mu)E, & x \in \Omega,\ t > 0, \\ \partial_t I - h\Delta I = \sigma E - (\gamma+\mu+\alpha)I, & x \in \Omega,\ t > 0, \\ \partial_t R - d_R\Delta R = \gamma I - \mu R, & x \in \Omega,\ t > 0, \\ \nabla S \cdot \mathbf{n} = \nabla E \cdot \mathbf{n} = \nabla I \cdot \mathbf{n} = \nabla R \cdot \mathbf{n} = 0, & x \in \partial\Omega,\ t > 0, \\ S(x,0) = S_0(x),\ E_0(x) = E_0(x),\ I(x,0) = I_0(x),\ R(x,0) = R_0(x), & x \in \Omega. \end{cases} \tag{4.1}$$

With a similar procedure in the proof of Theorem 2.1, we also can give the well-posedness of the model (4.1).

Proposition 4.1 *For any initial date* $(S_0, E_0, I_0, R_0) \in \mathbf{X}_+$, *model* (4.1) *admits a unique bounded solution* $(S, E, I, R) \in \mathbf{X}_+$ *for all* $t \geq 0$.

It follows from Theorem 3.4 in Wang and Zhao (2012) that the basic reproduction number of the model (4.1) can be defined by

$$\mathcal{R}_0 = \frac{\beta_1 \Lambda/\mu}{\sigma+\mu} + \frac{\beta_2 \sigma \Lambda/\mu}{(\sigma+\mu)(\gamma+\mu+\alpha)},$$

which is the same as that of the ODE compartment model. The two parts of the basic reproduction number $\mathcal{R}_0$ can be regarded as the influence of the exposed E and the infected I, respectively.

Proposition 4.2 *There exist two equilibria for model* (4.1)*: the DFE* $\tilde{\mathcal{E}}^0 = (\frac{\Lambda}{\mu}, 0, 0, 0)$*; and the EE* $\tilde{\mathcal{E}}^1 = (S_1, E_1, I_1, R_1)$ *for* $\mathcal{R}_0 > 1$, *where*

$$S_1 = \frac{(\sigma+\mu)(\gamma+\mu+\alpha)}{\beta_1(\gamma+\mu+\alpha)+\beta_2\sigma},\ E_1 = \frac{\Lambda-\mu S_1}{\sigma+\mu},\ I_1 = \frac{\sigma E_1}{\gamma+\mu+\alpha},\ R_1 = \frac{\gamma I_1}{\mu}.$$

It is easy to check that $\Lambda - \mu S_1 > 0$ as $\mathcal{R}_0 > 1$. Employing the Lyapunov functional method, we can give the globally asymptotic stability for the DEF and EE of the model (4.1) in terms of the $\mathcal{R}_0$.

4.1 Global stability

Theorem 4.1 *The following assertions are valid.*

(i) *If* $\mathcal{R}_0 < 1$, *then* $\tilde{\mathcal{E}}^0$ *is globally asymptotically stable in* $\mathbf{X}_+$;
(ii) *If* $\mathcal{R}_0 > 1$, *then* $\tilde{\mathcal{E}}^1$ *is globally asymptotically stable in* $\mathbf{X}_+$.

Proof Since the R-equation can be decoupled by the first three equations of (4.1), we only consider the system composed by the first three equations.

We first prove (i). Set $\breve{S} := \frac{\Lambda}{\mu}$. Define $W_1(t) : C(\bar{\Omega}, \mathbb{R}_+) \to \mathbb{R}$ as

$$W_1(t) = \int_\Omega \left[S - \breve{S} - \breve{S} \ln \frac{S}{\breve{S}} + E + \frac{\beta_2 \breve{S}}{\gamma + \mu + \alpha} I \right] dx, \ \forall t > 0.$$

It follows from direct calculations that

$$\begin{aligned}
\frac{dW_1(t)}{dt} &= -f \int_\Omega \frac{\breve{S}}{S^2} |\nabla S|^2 dx + \int_\Omega \left[\left(1 - \frac{\breve{S}}{S} \right) (\Lambda - \beta_1 SE - \beta_2 SI - \mu S) \right. \\
&\quad \left. + (\beta_1 SE + \beta_2 SI - (\sigma + \mu)E) + \frac{\beta_2 \breve{S}}{\gamma + \mu + \alpha} (\sigma E - (\gamma + \mu + \alpha) I) \right] dx \\
&= -f \int_\Omega \frac{\breve{S}}{S^2} |\nabla S|^2 dx + \Lambda \int_\Omega \left(2 - \frac{\breve{S}}{S} - \frac{S}{\breve{S}} \right) dx \\
&\quad + \int_\Omega \left(\beta_1 \breve{S} + \frac{\beta_2 \sigma \breve{S}}{\gamma + \mu + \alpha} - (\sigma + \mu) \right) dx \\
&\leq 0,
\end{aligned}$$

where we used the fact that $\Lambda = \mu \breve{S}$ in the second equality and $\mathcal{R}_0 < 1$ in the last inequality. It is obvious that $\frac{dW_1(t)}{dt} = 0$ if and only if $(S, E, I) = (\breve{S}, 0, 0)$. Hence, $W_1(t)$ is a Lyapunov functional. By some standard arguments, it is evident that

$$(S, E, I) \to (\breve{S}, 0, 0) \text{ in } (L^2(\Omega))^3, \ \text{as } t \to \infty.$$

From the uniform boundedness in Proposition 4.1, the parabolic L^p-theory, Sobolev embedding theorem, and a standard compactness argument (see, e.g., Theorem A2 in Brown et al. (1981)) guarantee that there exist positive constants C and T_0 such that

$$\|S(\cdot, t)\|_{C^2(\bar{\Omega})} + \|E(\cdot, t)\|_{C^2(\bar{\Omega})} + \|I(\cdot, t)\|_{C^2(\bar{\Omega})} \leq C, \ \forall t \geq T_0.$$

Therefore, the Sobolev embedding theorem allows one to assert

$$(S, E, I) \to (\breve{S}, 0, 0) \text{ in } (L^\infty(\Omega))^3, \ \text{as } t \to \infty.$$

This implies that the DFE $\tilde{\mathcal{E}}^0$ attracts all solutions of (4.1).

We then prove (ii). $\mathcal{R}_0 > 1$ guarantees the existence of EE. Define $W_2(t) : C(\bar{\Omega}, \mathbb{R}_+) \to \mathbb{R}$ as

$$W_2(t) = \int_\Omega \Big[S - S_1 - S_1 \ln \frac{S}{S_1} + E - E_1 - E_1 \ln \frac{E}{E_1} + \frac{\beta_2 S_1}{\gamma + \mu + \alpha} \left(I - I_1 - I_1 \ln \frac{I}{I_1} \right) \Big] dx, \ \forall t > 0.$$

Thus, a straightforward computation shows

$$\begin{aligned} \frac{dW_2(t)}{dt} &= -f \int_\Omega \frac{S_1}{S^2} |\nabla S|^2 dx + \int_\Omega \left(1 - \frac{S_1}{S}\right) (\Lambda - \beta_1 SE - \beta_2 SI - \mu S) dx \\ &\quad -g \int_\Omega \frac{E_1}{E^2} |\nabla E|^2 dx + \int_\Omega \left(1 - \frac{E_1}{E}\right) (\beta_1 SE + \beta_2 SI - (\sigma + \mu) E) dx \\ &\quad - \frac{\beta_2 S_1 h}{\gamma + \mu + \alpha} \int_\Omega \frac{I_1}{I^2} |\nabla I|^2 dx + \frac{\beta_2 S_1}{\gamma + \mu + \alpha} \int_\Omega \left(1 - \frac{I_1}{I}\right) (\sigma E - (\gamma + \mu + \alpha) I) dx \\ &= -f \int_\Omega \frac{S_1}{S^2} |\nabla S|^2 dx - g \int_\Omega \frac{E_1}{E^2} |\nabla E|^2 dx - h \int_\Omega \frac{I_1}{I^2} |\nabla I|^2 dx \\ &\quad + \beta_1 S_1 E_1 \int_\Omega \left(2 - \frac{S_1}{S} - \frac{S}{S_1}\right) dx + \beta_2 S_1 I_1 \int_\Omega \left(3 - \frac{S_1}{S} - \frac{SIE_1}{S_1 I_1 E} - \frac{EI_1}{E_1 I}\right) dx \\ &\le 0, \end{aligned}$$

where we use the fact

$$\begin{aligned} \Lambda &= \beta_1 S_1 E_1 + \beta_2 S_1 I_1 + \mu S_1, \ \beta_1 S_1 E_1 + \beta_2 S_1 I_1 \\ &= (\sigma + \mu) E_1, \ \sigma E_1 = (\gamma + \mu + \alpha) I_1, \end{aligned}$$

in the second equality. We can verify $\frac{dW_2(t)}{dt} = 0$ if and only if $(S, E, I) = (S_1, E_1, I_1)$, implying that $W_2(t)$ is a Lyapunov functional. Then by a similar argument as in the proof of (i), one has

$$(S, E, I) \to (S_1, E_1, I_1) \text{ in } (L^\infty(\Omega))^3, \ \text{as } t \to \infty.$$

□

For $\mathcal{R}_0 > 1$, Theorem 4.1(ii) indicates the uniqueness of EE for model (4.1) with local infection in a spatially homogeneous environment. Compared with Theorems 2.2(ii) and 3.2(ii), the uniform persistence only guarantees the existence of EE for the models in spatially heterogeneous environments, but the uniqueness is not derived.

4.2 Sensitivity analysis

In this subsection, we explore the sensitivity of $\mathcal{R}_0$, E_1, and I_1 to the infection rate β_1 of exposed individuals and the infection rate β_2 of infected individuals. We apply the formula in Chitnis et al. (2008) to analyze the sensitivity:

$$\Upsilon_p^u := \frac{\partial u}{\partial p} \cdot \frac{p}{u}. \tag{4.2}$$

It means the sensitivity of u to parameter p.

We first compare the sensitivity of $\mathcal{R}_0$ to the parameters β_1 and β_2. The sensitivity index of $\mathcal{R}_0$ to β_1 and β_2 can be directly calculated by (4.2) in the following:

$$\Upsilon_{\beta_1}^{\mathcal{R}_0} = \frac{\beta_1(\gamma+\mu+\alpha)}{\beta_1(\gamma+\mu+\alpha)+\beta_2\sigma}$$

and

$$\Upsilon_{\beta_2}^{\mathcal{R}_0} = \frac{\beta_2\sigma}{\beta_1(\gamma+\mu+\alpha)+\beta_2\sigma}.$$

It is easy to observe that the sensitivity index of I_1 to the parameters β_1 and β_2 is equal to the sensitivity index of E_1 to the parameters. So it suffices to give the sensitivity of E_1 to β_1 and β_2, which is

$$\Upsilon_{\beta_1}^{E_1} = \frac{\beta_1\mu(\sigma+\mu)(\gamma+\mu+\alpha)^2}{(\beta_1(\gamma+\mu+\alpha)+\beta_2\sigma)^2(\Lambda-\mu S_1)}$$

and

$$\Upsilon_{\beta_2}^{E_1} = \frac{\beta_2\mu(\sigma+\mu)(\gamma+\mu+\alpha)\sigma}{(\beta_1(\gamma+\mu+\alpha)+\beta_2\sigma)^2(\Lambda-\mu S_1)}.$$

We take $\Lambda = 150$, $\mu = 0.18$, $\sigma = 0.5$, $\alpha = 0.05$, and $\gamma = 0.5$ to give illustrations. There is no doubt that the infection probability by infected individuals is greater than that by exposed individuals, that is, $\beta_2 \geq \beta_1$. Therefore, we choose $\beta_1 \in [0.01, 0.05]$ and $\beta_2 \in [0.05, 0.1]$.

The sensitivity of $\mathcal{R}_0$ and E_1 to β_1 and β_2 is illustrated in Fig. 1a, b, respectively. It is apparent that $\Upsilon_{\beta_i}^{\mathcal{R}_0} > \Upsilon_{\beta_i}^{E_1}$. We then compare $\Upsilon_{\beta_1}^{\mathcal{R}_0}$ with $\Upsilon_{\beta_2}^{\mathcal{R}_0}$. Letting $\Upsilon_{\beta_1}^{\mathcal{R}_0} = \Upsilon_{\beta_2}^{\mathcal{R}_0}$, we can find $\beta_2 = \frac{\gamma+\mu+\alpha}{\sigma}\beta_1$. This is a straight line which is the intersection of the two surfaces $\Upsilon_{\beta_1}^{\mathcal{R}_0}$ and $\Upsilon_{\beta_2}^{\mathcal{R}_0}$. The sensitivity of $\mathcal{R}_0$ to β_2 is higher than that to β_1 if $\beta_2 > \frac{\gamma+\mu+\alpha}{\sigma}\beta_1$, that is, on the left-hand side of the intersection of the surfaces in Fig. 1a. On the contrary, $\Upsilon_{\beta_1}^{\mathcal{R}_0} > \Upsilon_{\beta_2}^{\mathcal{R}_0}$ if $\beta_2 < \frac{\gamma+\mu+\alpha}{\sigma}\beta_1$, that is, on the right-hand side of the intersection of the surfaces. It is easy to verify that the intersection line in Fig. 1b is also $\beta_2 = \frac{\gamma+\mu+\alpha}{\sigma}\beta_1$. The sensitivity of $\Upsilon_{\beta_1}^{E_1}$ and $\Upsilon_{\beta_2}^{E_1}$ in Fig. 1b can be analyzed in the same way as in Fig. 1a.

5 Spatial segregation versus nonpharmaceutical interventions

In this section, we perform a series of numerical examinations with the aim of exploring the effects of spatial segregation and nonpharmaceutical interventions on the control of

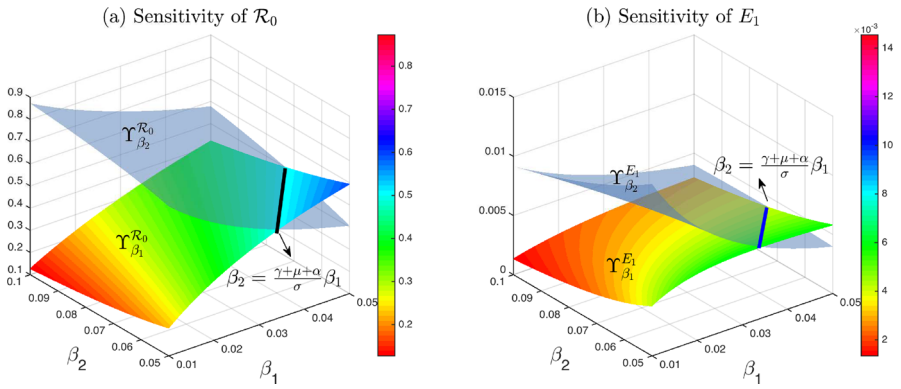

Fig. 1 Sensitivity analysis of $\mathcal{R}_0$ and E_1 to β_1 and β_2

infectious diseases. In order to compare with the dynamics of the symmetric diffusion model (1.5) and the random diffusion model (1.6), we also consider a homogeneous diffusion model taking the form

$$\begin{cases} \partial_t S - \bar{f}\Delta S = \Lambda(x) - \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad - \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - \mu(x)S, & x\in\Omega,\ t>0, \\ \partial_t E - \bar{g}\Delta E = \beta_1(x)S\displaystyle\int_\Omega K(x-y)E(y,t)dy \\ \quad + \beta_2(x)S\displaystyle\int_\Omega K(x-y)I(y,t)dy - (\sigma(x)+\mu(x))E, & x\in\Omega,\ t>0, \\ \partial_t I - \bar{h}\Delta I = \sigma(x)E - (\gamma(x)+\mu(x)+\alpha(x))I, & x\in\Omega,\ t>0, \\ \partial_t R - d_R\Delta R = \gamma(x)I - \mu(x)R, & x\in\Omega,\ t>0, \\ \nabla S\cdot\mathbf{n} = \nabla E\cdot\mathbf{n} = \nabla I\cdot\mathbf{n} = \nabla R\cdot\mathbf{n} = 0, & x\in\partial\Omega,\ t>0, \\ S(x,0)=S_0(x),\ E(x,0)=E_0(x),\ I(x,0)=I_0(x),\ R(x,0)=R_0(x), & x\in\Omega, \end{cases} \tag{5.1}$$

where $\bar{f}$, $\bar{g}$ and $\bar{h}$ are positive constants.

The directed migration or dispersal of individuals contributes to the spatial pattern formation of segregation, playing a vital role in population regulation. Furthermore, it is beneficial for the waning and eventual elimination of the disease. We adopt the segregation indices χ and κ in Wang et al. (2022) to measure the degree of segregation of two groups u_1 and u_2 of the population in Ω, where

$$\chi(u_1,u_2) = \max_{x\in\bar{\Omega}}\{u_1-u_2\}\cdot\min_{x\in\bar{\Omega}}\{u_1-u_2\},$$

and

$$\kappa(u_1,u_2) = \frac{\|u_1-u_2\|_{L^1(\Omega)}}{\|u_1\|_{L^1(\Omega)}+\|u_2\|_{L^1(\Omega)}}.$$

In fact, χ is used to determine whether there is a segregation phenomenon between u_1 and u_2, while $0 \leq \kappa \leq 1$ can describe the degree of segregation. If $\chi \geq 0$, then there is no segregation phenomenon. If $\chi < 0$ and κ is close to 0, then the segregation is weak, and such phenomenon disappears as κ becomes 0. On the contrary, if $\chi < 0$ and κ is close to 1, the segregation is considered strong or perfect. We designate susceptible group S as u_1, and infected groups E and I as u_2 for the reason that both exposed and infected individuals are contagious. We employ ρ to represent the weight of I among the infectives concerning disease transmission. Accordingly, the weight of E is represented by $1-\rho$. That is to say, $u_2 := (1-\rho)E+\rho I$, which suits the above given segregation indices χ and κ. Generally, we have $\rho > \frac{1}{2}$ because infected individuals (symptomatic) are normally more infectious than exposed individuals (asymptomatic). Besides transmission probabilities (or rates), other factors such as weather conditions, population density, epidemic severity, and disease-induced fatality can also regulate the value of ρ. In the following simulations, the weight $\rho = 0.8$.

To measure the final infection size, namely, the total number of exposed and infected individuals, we introduce the following infection fraction (I.F.) index:

$$\text{I.F.} = \frac{\|E + I\|_{L^1(\Omega)}}{\|S + E + I + R\|_{L^1(\Omega)}},$$

where S, E, I and R are at the steady state.

We carry out the numerical examples in a one-dimensional interval $\Omega = (0, 2\pi)$. The infection and recovery rates are taken as

$$\begin{aligned} &\beta_1(x) = 0.002(\cos x + 1.3),\ \beta_2(x) = 0.05(\cos x + 1.2), \\ &\gamma(x) = 0.3(\sin x + 1.1),\ x \in (0, 2\pi). \end{aligned}$$

The diffusion rates of susceptible individuals and infected individuals are taken as

$$f\left(\sum_{i=1}^{2} a_i \beta_i(x)\right) = 100(\beta_1(x) + \beta_2(x)),\ h(\gamma^{-1}(x)) = \frac{1}{\gamma(x)},\ x \in (0, 2\pi).$$

respectively, where $a_1 = a_2 = 1$. Furthermore, the diffusion rate of exposed individuals can be taken as either $g\left(\sum_{i=1}^{2} b_i \beta_i(x)\right) = f\left(\sum_{i=1}^{2} a_i \beta_i(x)\right)$ or $g(\gamma^{-1}(x)) = h(\gamma^{-1}(x))$, depending on the perception of exposed individuals. The infection-recovery rates and diffusion rates of susceptibles and infectives are illustrated in Fig. 2a, b, respectively. It can be seen from Fig. 2a that $\beta_1(x) < \beta_2(x)$ for all $x \in (0, 2\pi)$. In addition, the infection rate is the lowest at $x = \pi$, while the recovery rate is the highest at $x = \frac{\pi}{2}$ and the lowest at $x = \frac{3\pi}{2}$. Figure 2b shows that susceptible individuals at $x = \pi$ and infected individuals at $x = \frac{\pi}{2}$ have the lowest diffusion rates because they tend to stay at these sites to avoid infection or to get medical treatment. On the other hand, infected individuals have a high diffusion rate at $x = \frac{3\pi}{2}$ because the recovery rate is the lowest here. They will escape from this site to seek medical resources. The diffusion rate of the recovered d_R is 10 in simulations.

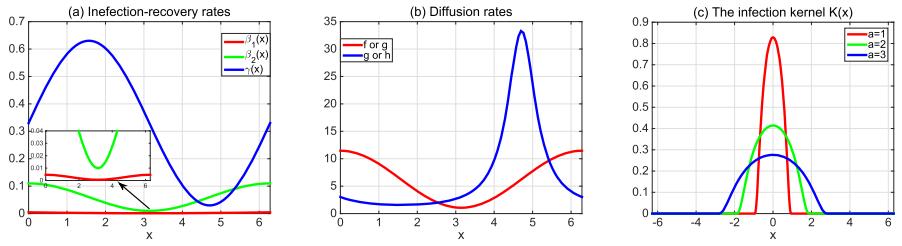

Fig. 2 The infection-recovery rates, diffusion rates, and nonlocal infection radii

For the homogeneous diffusion model (5.1), we take the spatially constant diffusion rates of susceptible individuals and infected individuals as

$$\bar{f} = \frac{1}{2\pi}\int_0^{2\pi} f\left(\sum_{i=1}^{2} a_i\beta_i(x)\right)dx,\ \bar{h} = \frac{1}{2\pi}\int_0^{2\pi} h(\gamma^{-1}(x))dx,$$

respectively, and the spatially constant diffusion rate $\bar{g}$ of exposed individuals is equal to $\bar{f}$ or $\bar{h}$. It can be calculated that $\bar{f} = 6.26$ and $\bar{h} = 7.2376$.

We adopt the nonlocal infection kernel in the same format as described in Liu et al. (2019):

$$K(x) = \begin{cases} \dfrac{2.2523}{a} e^{\frac{a^2}{|x|^2-a^2}}, & |x| < a, \\ 0, & |x| \geq a, \end{cases}$$

where a is the nonlocal infection radius. The unit of a can be interpreted as meters, representing the maximum distance over which the infection can spread. The infection kernel $K(x)$ with different infection radii is shown in Fig. 2c. The infection area increases while the infection strength decreases with the increase of the infection radius.

The initial values are taken as $S_0 = 500$ and $E_0 = I_0 = R_0 = 100$. The other parameter values are chosen the same as those in Fig. 1.

5.1 Segregation by cognitive diffusion

In this subsection, we consider spatial segregation induced by cognitive diffusion with the influence of nonlocal infection. Three examples with infection radii $a = 0, 1, 3$ are performed, respectively.

Example 1 Segregation with local infection.

Here we provide the distribution of the steady state for the homogeneous diffusion model (5.1), symmetric diffusion model (1.5) and random diffusion model (1.6) with local infection, that is, the infection radius $a = 0$. In addition, we pay attention to the impacts of different dispersal strategies adopted by exposed individuals. In Fig. 3a–c, the distribution of the steady state is shown when the dispersal rate of exposed

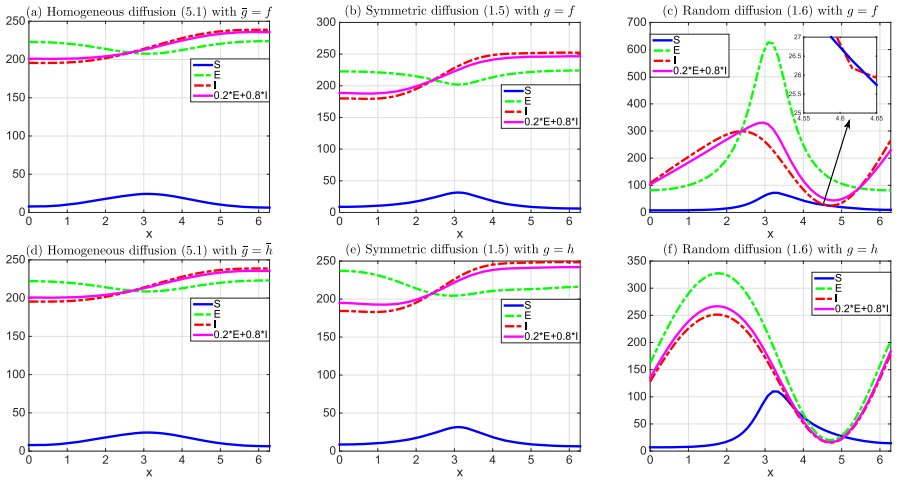

Fig. 3 The distribution of the steady state of epidemic models with homogeneous, symmetric, and random diffusion if the infection is local

individuals is equal to that of susceptible individuals, and in Fig. 3d–f, the distribution of the steady state is shown when the dispersal rate of exposed individuals is equal to that of infected individuals.

Intuitively, from Fig. 3a, b, d, e, the symmetric diffusion model and the homogeneous diffusion model have a similar distribution of the steady state regardless of whether the movement mechanism of exposed individuals equals to that of susceptible or infected individuals. However, the distribution of the steady state in the random diffusion model behaves rather differently from that in the symmetric or homogeneous diffusion models. Moreover, we can find that the distribution of exposed individuals is similar to that of susceptible individuals when the diffusion rate of exposed individuals equals that of susceptible individuals from Fig. 3c and similar to that of infected individuals when it equals to that of infected individuals from Fig. 3f. In contrast, the symmetric diffusion model and homogeneous diffusion model do not display such characteristics. Additionally, we obtain two distinct cases: near the position $x = \frac{3\pi}{2}$ with the low recovery rate and the high diffusion rate of infected individuals, the density of infectives is high in the homogeneous and symmetric diffusion model, while it is low in the random diffusion model.

According to Table 1, we can find there is no segregation phenomenon between susceptibles and infectives including the exposed and the infected in the homogeneous and symmetric diffusion models because of the positive χ, and the infection fraction is very close. For the random diffusion model, there is a segregation phenomenon when $g = h$ due to the negative χ. In fact, there is also a segregation phenomenon between S and I in the random diffusion model when $g = f$ because $\chi(S, I) = -55.3406$ and $\kappa(S, I) = 0.6954$. However, there is no segregation between S and E. Therefore, the segregation when $g = f$ is weaker than that when $g = h$. The segregation in the random diffusion model when $g = h$ may contribute to the smallest infection.

Table 1 Segregation indices and infection fraction for $a = 0$

Diffusion of E	Epidemic models	χ	κ	I.F.
$\bar{g} = \bar{f}$	Homogeneous diffusion (5.1)	4.2595×10^4	0.8256	0.5347
$g = f$	Symmetric diffusion (1.5)	4.2640×10^4	0.8284	0.5438
$g = f$	Random diffusion (1.6)	5.8516×10^3	0.6799	0.4607
$\bar{g} = \bar{h}$	Homogeneous diffusion (5.1)	4.2596×10^4	0.8257	0.5347
$g = h$	Symmetric diffusion (1.5)	4.2560×10^4	0.8253	0.5396
$g = h$	Random diffusion (1.6)	-5.1608×10^3	0.5096	0.3941

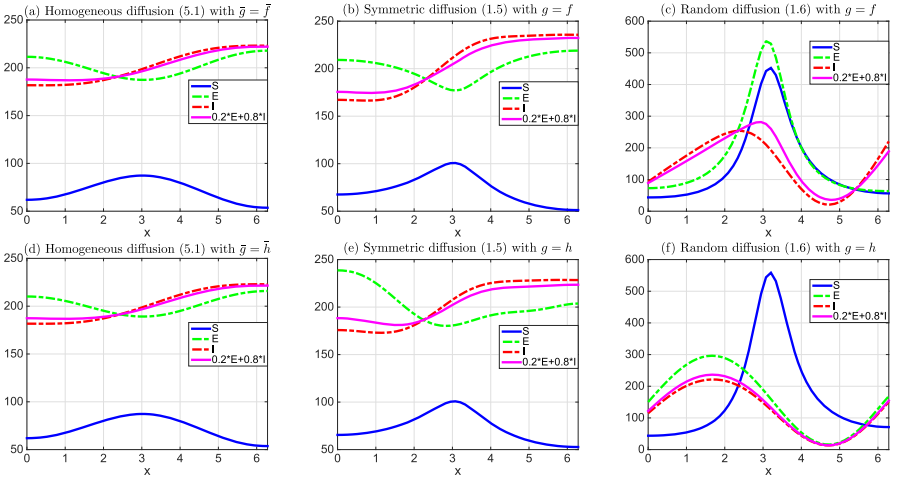

Fig. 4 The distribution of the steady state of epidemic models with homogeneous, symmetric, and random diffusion when nonlocal infection radius $a = 1$

Example 2 Segregation with $a = 1$.

The distribution of the steady state of the three diffusion models with the nonlocal infection radius $a = 1$ is given in Fig. 4. Compared Fig. 4 with Fig. 3, when taking nonlocal infection into consideration, the total number of susceptibles is increasing while that of infectives including exposed and infected individuals is decreasing. Similar to Fig. 3, the homogeneous and symmetric diffusion models also have a similar distribution of the steady state, while the random diffusion model still behaves rather differently. The different distribution in different diffusion models at $x = \frac{3\pi}{2}$ is also like Fig. 3.

It can be observed from Table 2 that there is no segregation phenomenon in the homogeneous and symmetric diffusion models because χ is positive in these two models, and there is a segregation phenomenon in the random diffusion model for negative χ. Compared the value of κ in the random diffusion model in Table 2 with that in Table 1, we can find κ gets smaller when there is a nonlocal infection in the random diffusion model, indicating the segregation is getting weaker. However, although the segregation driven by random diffusion is poor, the infection fraction decreases.

Table 2 Segregation indices and infection fraction for $a = 1$

Diffusion of E	Epidemic models	χ	κ	I.F.
$\bar{g} = \bar{f}$	Homogeneous diffusion (5.1)	1.7941×10^4	0.4364	0.4989
$g = f$	Symmetric diffusion (1.5)	1.8131×10^4	0.4485	0.5088
$g = f$	Random diffusion (1.6)	-2.5755×10^4	0.2678	0.3828
$\bar{g} = \bar{h}$	Homogeneous diffusion (5.1)	1.7935×10^4	0.4362	0.4988
$g = h$	Symmetric diffusion (1.5)	1.6876×10^4	0.4371	0.4998
$g = h$	Random diffusion (1.6)	-7.0273×10^4	0.4171	0.3046

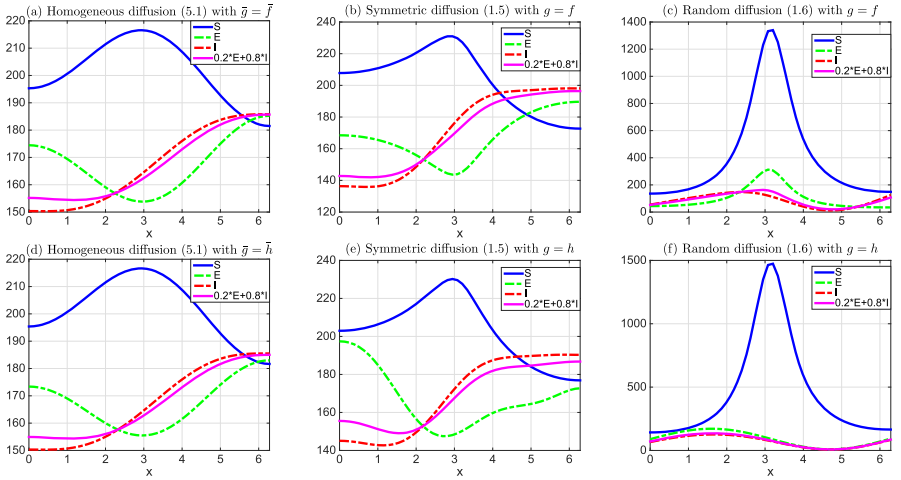

Fig. 5 The distribution of the steady state of epidemic models with homogeneous, symmetric, and random diffusion when nonlocal infection radius $a = 3$

Meanwhile, the infection proportion is still the lowest in the random diffusion model with $g = h$.

Example 3 Segregation with $a = 3$.

Figure 5 shows the solution distribution of the three diffusion models with the infection radius $a = 3$. The symmetric diffusion model remains a similar distribution of the steady state to the homogeneous diffusion model. With the infection radius increasing from 1 to 3, the total number of susceptible individuals is increasing abruptly in the random diffusion model. There is also a slight increase of susceptibles in the homogeneous and symmetric diffusion models. In addition, susceptible individuals gather at $x = \pi$ where the infection rate is the lowest. On the contrary, the total number of infectives is decreasing in the three diffusion models.

By observing Table 3, a rather different phenomenon occurs with the increase of infection radius: there is no segregation phenomenon in the random diffusion model while the segregation appears in the homogeneous and symmetric diffusion models.

Table 3 Segregation indices and infection fraction for $a = 3$

Diffusion of E	Epidemic models	χ	κ	I.F.
$\bar{g} = \bar{f}$	Homogeneous diffusion (5.1)	-238.6669	0.0743	0.4151
$g = f$	Symmetric diffusion (1.5)	-1.6778×10^3	0.0868	0.4280
$g = f$	Random diffusion (1.6)	4.9305×10^4	0.6609	0.2173
$\bar{g} = \bar{h}$	Homogeneous diffusion (5.1)	-189.4131	0.0741	0.4181
$g = h$	Symmetric diffusion (1.5)	-669.2330	0.0752	0.4166
$g = h$	Random diffusion (1.6)	6.5154×10^4	0.7441	0.1568

However, the segregation is rather weak in the homogeneous and symmetric diffusion models because the value of κ is close to zero.

In view of the above three examples, it seems that the random diffusion model can characterize individual movement more accurately based on the fact that the distribution of the exposed group is similar to the susceptible group when $g = f$ and similar to the infected group when $g = h$. While the homogeneous and symmetric diffusion models never display such characteristics. The Fokker–Planck law better captures the cognitive diffusion of individuals. This may account for why the final infection size of infectives in the random diffusion model is always smaller than that in the homogeneous and symmetric diffusion models. Moreover, in the random diffusion model, if exposed individuals take the same dispersal strategy as infected individuals, the final size is then smaller than that in the case where exposed individuals take the same dispersal strategy as susceptible individuals. Another interesting numerical result derived from the above three examples is that the segregation phenomenon disappears in the random diffusion model while it appears in the homogeneous or symmetric models, as the nonlocal infection radius increases.

Comparing the distribution of infectives in the symmetric diffusion model and random diffusion model, different diffusion laws lead to distinguished distributions of infectives. At $x = \frac{3\pi}{2}$, the recovery rate is low but the diffusion rate of infectives is high. We can find in the symmetric diffusion model, the number of infectives around this position is very high, but in the random diffusion model, the number of infectives is rather low. This gives rise to two distinct cases, which should be examined further from observed reality.

5.2 Nonpharmaceutical interventions

In this subsection, we explore the effects of nonpharmaceutical interventions by considering the asymptotic behaviors of the EE of model (5.1) as $\bar{g} \to 0$ or $\bar{g}, \bar{h} \to 0$. In the literature (Allen et al. 2008; Li et al. 2017, 2018; Cui et al. 2017, 2021; Li et al. 2020), some systematic methods were established to investigate the asymptotic properties of the EE on the basis of compactness arguments, singular perturbation arguments, elliptic regularity theory, etc., for some reaction-diffusion epidemic models with frequency-dependent or bilinear incidence mechanisms. Whereas, we give the asymptotic results of the EE numerically due to the difficulties caused by the nonlocal

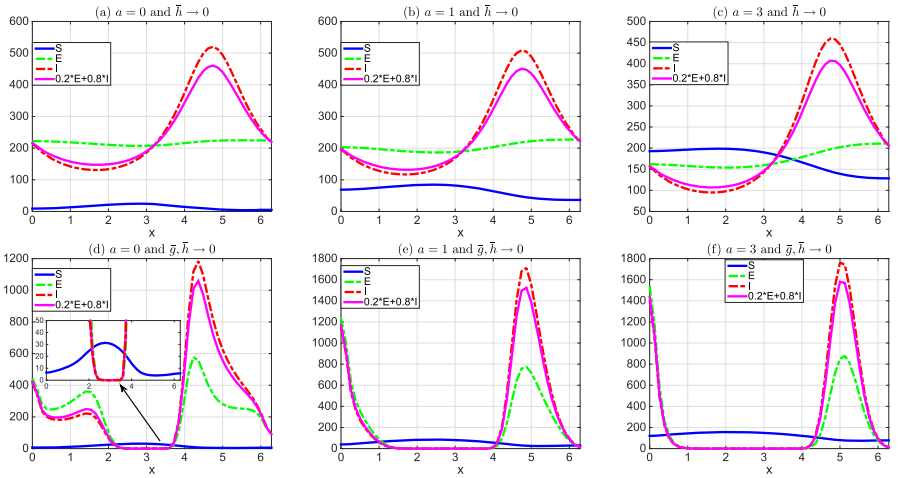

Fig. 6 Nonpharmaceutical interventions on restricting the dispersal of infected individuals or both exposed and infected individuals on the homogeneous diffusion model (5.1) with different infection radii

infection. Here, we take $\bar{h} = 10^{-6}$ or $\bar{g} = \bar{h} = 10^{-6}$ to represent the movement restriction of infected individuals or both exposed and infected individuals, respectively.

If only the movement of infected individuals is restricted, the distribution of susceptible, exposed, and infected individuals in the steady state is illustrated in Fig. 6a–c, where the infection radius is taken as 0, 1, and 3, respectively. If the movement of both exposed and infected individuals is restricted, the distribution is displayed in Fig. 6d–f, where three infection radii are also taken.

It follows from Fig. 6a–c that the number of susceptibles is expanding with the increase of infection radius, by only restricting the movement of infected individuals. Meanwhile, there is a slight decrease in the total number of exposed and infected groups. By observing Fig. 6d–f, we find that the infectious disease can be eliminated in a region by restricting the movement of both exposed and infected individuals. Moreover, the region expands with the increase of the infection radius. If the diffusivity of both exposed and infected individuals is limited, though there is a disease-free region, infectives can aggregate at the position where the recovery rate is the lowest. This aggregation phenomenon may neglect the fact that the diffusion rates of infectives at this position are high in reality because the infectives will escape from such areas to get sufficient medical treatment. In addition, such a position moves to the right gradually with the increase of nonlocal infection radius.

It is evident that restricting the movement of infectives does not help decrease the infection proportion by comparing the values of I.F. with those in Tables 1-4. When $\bar{h} \to 0$, the value of I.F. decreases with the increase of infection radius. When $\bar{g}, \bar{h} \to 0$, there is segregation in the three cases. It has stronger segregation when $a = 0$ than $a = 1$ by comparing the values of κ. Consequently, the infection fraction in the model without nonlocal infection is lower than that in the model with infection

Table 4 Segregation indices and infection fraction for nonpharmaceutical interventions

Interventions on E and I	Infection radius a	χ	κ	I.F.
	$a = 0$	5.7904×10^4	0.8681	0.6079
$\bar{h} \to 0$	$a = 1$	1.9814×10^4	0.5788	0.5853
	$a = 3$	-2.4250×10^4	0.2996	0.5226
	$a = 0$	-3.2819×10^4	0.9000	0.6567
$\bar{g}, \bar{h} \to 0$	$a = 1$	-1.2652×10^5	0.7898	0.6843
	$a = 3$	-2.3472×10^5	0.7482	0.6525

radius $a = 1$. The values of κ are close when $a = 1$ and $a = 3$, but the model with a bigger infection radius has a smaller infection fraction.

6 Discussion

In the present work, we applied Fick's law and the Fokker–Planck law of diffusion to model population movement in heterogeneous environments. By taking the perception concerning the dispersal of individuals into consideration, two SEIR epidemic models with cognitive diffusion were established and studied. Particularly, the diffusive behaviors of exposed individuals have been examined according to different dispersal strategies. If exposed individuals have the perception that they may be infected, they adopt the same movement mechanism as infected individuals. If they are not aware that they have been exposed in some high-risk regions, their dispersal strategy goes with susceptible individuals. In addition, the two models include nonlocal infections between susceptibles and infectives, which can describe more realistic disease transmission.

Theoretically, we obtained the well-posedness of solutions for the two SEIR epidemic models with symmetric diffusion and random diffusion, which were given in Theorems 2.1 and 3.1, respectively. Then utilizing the next-generation operator approach, the basic reproduction numbers were established to study the threshold dynamics of the disease. It is found that if the basic reproduction number is less than one, then the disease can be extinct. The disease will prevail if the basic reproduction number is greater than one. The threshold dynamics for the two models were established in Theorems 2.2 and 3.2, respectively. Furthermore, a limit case was also considered in a homogeneous environment, in which the nonlocal infection mechanism degenerates to the local infection mechanism (or bilinear infection mechanism) by letting the infection radius tend to 0^+. Global stability of the DFE and EE was given in Theorem 4.1 by the Lyapunov functional method, which indicates the uniqueness of the EE. However, in Theorem 2.2(ii) and 3.2(ii), we only recognized the existence of the EE, failing to derive the uniqueness using the dynamical system theory approach. It is still a challenge to prove the uniqueness of the EE for our reaction-diffusion models with environmental heterogeneity. Moreover, we can only grant the implicit definition of basic reproduction numbers for models (1.5) and (1.6) by the spectral

radius of the next-generation operator in heterogeneous environments. However, the explicit formula of basic reproduction was established for the locally infective model (4.1) in a homogeneous environment.

Wang et al. (2022) has investigated the effects of dispersal rates and favorableness of the habitat on spatial segregation of susceptible and infected individuals in SIS epidemic models. In our work, we further explored the effect of nonlocal infection on segregation in SEIR epidemic models. Additionally, we conducted a comparative analysis of the efficacy of spatial segregation and nonpharmaceutical interventions on epidemic dynamics. The numerical findings are summarized below:

- The nonlocal infection has different impacts on spatial segregation in different diffusion models. The segregation arises in the random diffusion model with a small nonlocal infection radius and in the symmetric diffusion model with a large radius.
- The final infection size in the random diffusion model is significantly smaller than that in the symmetric diffusion model. The infection fraction decreases as the nonlocal infection radius increases in the homogeneous, symmetric, and random diffusion models. This finding is consistent with the observation in Liu et al. (2019), which can be attributed to the diminishing infection strength as the infection radius increases.
- In the homogeneous and symmetric diffusion models, we observe a high density of infected individuals around positions characterized by a low recovery rate and a high diffusion rate. Conversely, the random diffusion model exhibits a lower density of infectives in the same region.
- The distribution of the steady state of the model with homogeneous diffusion is always similar to that with symmetric diffusion, whereas the model with random diffusion behaves in a contrasting manner. The numerical illustrations in Wang et al. (2022) also displayed these properties. It suggests that the homogeneous and symmetric diffusion share more similar underlying mechanisms.
- Nonpharmaceutical interventions on restricting the dispersal of exposed and infected individuals do not help to reduce the infection proportion but can effectively eliminate the disease in a region. Additionally, such a disease-free region expands as the nonlocal infection radius increases.

The numerical examples performed in Sect. 5 provided some insights not only on disease control but also on epidemic modeling. The applicability of Fick's law versus the Fokker–Planck law in modeling diffusion in heterogeneous environments has been a topic of controversy in the literature. We obtained two distinct cases by the three numerical examples in Sect. 5.1: in the random diffusion model, the distribution of the steady state of exposed individuals closely resembles that of susceptible individuals if the dispersal function of exposed individuals is identical to that of susceptible individuals, while in the case where the dispersal functions of exposed and infected individuals are identical, both exposed and infected individuals also exhibit a similar distribution at the steady state. Nevertheless, the homogeneous and symmetric diffusion models never manifest this characteristic. The Fokker–Planck law appears to better capture human diffusion patterns in our epidemic models from this perspective. Notably, the infective density at a location characterized by a low recovery rate and

high diffusion rate in the symmetric diffusion model contradicts that observed in the random diffusion model. In the symmetric diffusion model, the infective density is high at such a site, reflecting the low recovery rate, whereas, in the random diffusion model, the density is low, corresponding to the high diffusion rate. Deciding the most appropriate diffusion law becomes challenging when considering these inconsistencies. It is an intriguing outcome deserving further investigation to shed light on its underlying mechanisms. In contrasting Fick's law with the Fokker–Planck type law of diffusion, it is significant to note that the Fokker–Planck type includes an advection term. Employing data fitting with real-world data is a viable approach to assess the significance of advection in a specific infectious disease. This method can aid in selecting a more appropriate diffusion type.

Plenty of infectious diseases such as influenza, brucellosis, measles, etc., exhibit discernible seasonal patterns, characterized by high transmission rates during specific seasons of the year. For instance, the influenza virus tends to be more prevalent during winter or spring months. Liu et al. (2019) found that the presence of seasonality has complicated implications on the dynamics of the disease. Therefore, the consideration of the seasonal variations in epidemic models is essential for ensuring accurate predictions of disease outbreaks and the formulation of effective disease control strategies. However, the mathematical complexity increases when incorporating periodic parameters like infection and recovery rates into these models. Investigating the cognitive diffusion mechanisms in epidemic models that incorporate seasonality deserves further exploration.

Acknowledgements This work was completed during the period of the first author's joint Ph.D. study in the Interdisciplinary Lab for Mathematical Ecology & Epidemiology, directed by the corresponding author, at the University of Alberta supported by China Scholarship Council. The research of the second and corresponding author was partially supported by Natural Sciences and Engineering Research Council of Canada (Individual Discovery Grant RGPIN-2020-03911 and Discovery Accelerator Supplement Award PGPAS-2020-00090) and the Canada Research Chairs Program (Tier 1 Canada Research Chair Award). The research of the third author was partially supported by National Natural Science Foundation of China (No. 11571200) and Natural Science Foundation of Shandong Province (No. ZR2021MA062).

Data availability The manuscript has no associated data.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen LJS, Bolker BM, Lou Y, Nevai AL (2008) Asymptotic profiles of the steady states for an SIS epidemic reaction–diffusion model. Discrete Contin Dyn Syst 21(1):1–20

Andreucci D, Cirillo ENM, Colangeli M, Gabrieli D (2019) Fick and Fokker–Planck diffusion law in inhomogeneous media. J Stat Phys 174:469–493

Bengfort M, Malchow H, Hilker FM (2016) The Fokker–Planck law of diffusion and pattern formation in heterogeneous environments. J Math Biol 73:683–704

Bringuier E (2011) Particle diffusion in an inhomogeneous medium. Eur J Phys 32(4):975–992

Brown KJ, Dunne PC, Gardner RA (1981) A semilinear parabolic system arising in the theory of superconductivity. J Differ Equ 40(2):232–252

Capasso V (1978) Global solution for a diffusive nonlinear deterministic epidemic model. SIAM J Appl Math 35(2):274–284
Chapman S (1928) On the Brownian displacements and thermal diffusion of grains suspended in a non-uniform fluid. Proc R Soc Lon Ser A 119(781):34–54
Chitnis N, Hyman JM, Cushing JM (2008) Determining important parameters in the spread of malaira through the sensitivity analysis of a mathematical model. Bull Math Biol 70:1272–1296
Contrell RS, Cosner C (2003) Spatial ecology via reaction–diffusion equations. Wiley, New York
Cui R, Lou Y (2016) A spatial SIS model in advective heterogeneous environments. J Differ Equ 261(6):3305–3343
Cui R, Lam K-Y, Lou Y (2017) Dynamics and asymptotic profiles of steady states of an epidemic model in advective environments. J Differ Equ 263(4):2343–2373
Cui R, Li H, Peng R, Zhou M (2021) Concentration behavior of endemic equilibrium for a reaction–diffusion–advection SIS epidemic model with mass action infection mechanism. Calc Var Partial Differ Equ 60(5):184
Desch W, Schappacher W (2006) Linearized stability for nonlinear semigroups. In: Differential equations in banach spaces: proceedings of a conference held in Bologna, July 2–5, 1985. Springer, Berlin, pp 61–73
Evans LC (2022) Partial differential equations. American Mathematical Society, Providence
Fick A (1855) Über diffusion. Ann Phys 170(1):59–86
Freedman HI, Zhao X-Q (1997) Global asymptotics in some quasimonotone reaction–diffusion systems with delays. J Differ Equ 137(2):340–362
Hale JK (1988) Asymptotic behavior of dissipative systems, mathematical surveys and monographs. American Mathematical Society, Providence
Hess P (1991) Periodic-parabolic boundary value problems and positivity. Longman, London
Kendall DG (1965) Mathematical models of the spread of injection. In: Mathematics and computer science in biology and medicine, pp 213–225
Kermack WO, McKendrik AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc A 115(772):700–721
Kim Y-J, Seo H, Yoon C (2019) Asymmetric dispersal and evolutional selection in two-patch system. Discrete Contin Dyn Syst 40(6):3571–3593
Li H, Peng R, Wang FB (2017) Varying total population enhances disease persistence: qualitative analysis on a diffusive epidemic model. J Differ Equ 262(2):885–913
Li H, Peng R, Wang Z (2018) On a diffusive susceptible-infected-susceptible epidemic model with mass action mechanism and birth-death effect: analysis, simulations, and comparison with other mechanisms. SIAM J Appl Math 78(4):2129–2153
Li H, Peng R, Xiang T (2020) Dynamics and asymptotic profiles of endemic equilibrium for two frequency-dependent SIS epidemic models with cross-diffusion. Eur J Appl Math 31(1):26–56
Liang X, Zhang L, Zhao X-Q (2019) Basic reproduction ratios for periodic abstract functional differential equations (with application to a spatial model for lyme disease). J Dyn Differ Equ 31:1247–1278
Liu S, Lou Y (2022) Classifying the level set of principal eigenvalue for time-periodic parabolic operators and applications. J Funct Anal 282(4):109338
Liu Z, Shen Z, Wang H, Jin Z (2019) Analysis of a local diffusive SIR model with seasonality and nonlocal incidence of infection. SIAM J Appl Math 79(6):2218–2241
Lou Y, Zhao X-Q (2011) A reaction-diffusion malaria model with incubation period in the vector population. J Math Biol 62(4):543–568
Magal P, Zhao X-Q (2005) Global attractors and steady states for uniformly persistent dynamical systems. SIAM J Math Anal 37(1):251–275
Martin RH, Smith HL (1990) Abstract functional-differential equations and reaction-diffusion systems. Trans Am Math Soc 321(1):1–44
Okubo A, Levin SA (2001) Diffusion and ecological problems: modern perspectives. Springer, Berlin
Pao CV (2012) Nonlinear parabolic and elliptic equations. Springer, Berlin
Peng R, Zhao X-Q (2012) A reaction-diffusion SIS epidemic model in a time-periodic environment. Nonlinearity 25(5):1451
Schelling TC (1969) Models of segregation. Am Econ Rev 59(2):488–493
Schnitzer MJ (1993) Theory of continuum random walks and application to chemotaxis. Phys Rev E 48(4):2553–2568
Skellam JG (1951) Random dispersal in theoretical populations. Biometrika 38(1/2):196–218

Smith HL (1995) Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems. American Mathematical Society, Providence
Song P, Lou Y, Xiao Y (2019) A spatial SEIRS reaction–diffusion model in heterogeneous environment. J Differ Equ 267(9):5084–5114
Thieme HR (2009) Spectral bound and reproduction number for infinite-dimensional population structure and time heterogeneity. SIAM J Appl Math 70(1):188–211
Wang H, Salmaniw Y (2023) Open problems in PDE models for knowledge-based animal movement via nonlocal perception and cognitive mapping. J Math Biol 86(5):71
Wang W, Zhao X-Q (2012) Basic reproduction numbers for reaction–diffusion epidemic models. SIAM J Appl Dyn Syst 11(4):699–717
Wang H, Wang K, Kim Y-J (2022) Spatial segregation in reaction–diffusion epidemic models. SIAM J Appl Math 82(5):1680–1709
Wu J (1996) Theory and applications of partial functional-differential equations. Springer, Berlin
Zhao X-Q (2017) Dynamical systems in population biology, 2nd edn. Springer, New York