TURNING POINTS AND RELAXATION OSCILLATION CYCLES IN SIMPLE EPIDEMIC MODELS

MICHAEL Y. LI[†,‖], WEISHI LIU[‡,**], CHUNHUA SHAN[†,††], AND YINGFEI YI[†,§,¶,††]

Abstract. We study the interplay between effects of disease burden on the host population and the effects of population growth on the disease incidence, in an epidemic model of SIR type with demography and disease-caused death. We revisit the classical problem of periodicity in incidences of certain autonomous diseases. Under the assumption that the host population has a small intrinsic growth rate, using singular perturbation techniques and the phenomenon of the delay of stability loss due to turning points, we prove that large amplitude relaxation oscillation cycles exist for an open set of model parameters. Simulations are provided to support our theoretical results. Our results offer new insight to the classical periodicity problem in epidemiology. Our approach relies on analysis far away from the endemic equilibrium and contrasts sharply with the method of Hopf bifurcations.

Key words. Epidemic models, periodicity in disease incidence, inter-epidemic period, turning point, delay of stability loss, relaxation oscillation cycles

AMS subject classifications. 34C26, 92D25

1. Introduction. Investigation of oscillations in disease incidence is of fundamental importance in mathematical epidemiology. Empirical data of disease incidence has shown clearly identifiable cyclic patterns in many common diseases, including diseases for which environmental influences do not appear to play an important role, such as measles, pertussis, chicken pox and mumps [2, 20]. Mechanisms for this type of "autonomous oscillation" have been extensively studied in the mathematical epidemiology literature. These include, together with papers that introduced them, time delays in the transmission process [14, 20], varying total population size with density dependent demography and transmission [1, 22], nonlinear incidence forms [16], discrete age-structures with a non-symmetric contact matrix among age groups [8], and seasonality in the transmission process in both deterministic and stochastic models [3, 4, 12, 20]. The mathematical approach for these earlier work has been bifurcation analysis (e.g. Hopf bifurcation theory), which analyzes model behaviours in a neighbourhood of an endemic equilibrium. In the case of Hopf bifurcation, certain degree of complexity needs to be introduced into the transmission process to produce instability of the endemic equilibrium, and the bifurcation may occur in parameter regimes that are not biologically realistic. For more complete reviews of related work, we refer the reader to [2, 13].

In the present paper, we apply a singular perturbation approach to this investigation. Our goal is to reveal a simple and biologically sound mechanism that can produce large-amplitude oscillations in disease incidence. Our basic assumption is

[†]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, T6G 2G1, Canada

[‡]Department of Mathematics, University of Kansas, Lawrence, KS 66045 USA

[§]School of Mathematics, Jilin University, Changchun, 130012, PRC

[¶]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

[‖]Research partially supported by Natural Science and Engineering Research Council of Canada and Canada Foundation for Innovation

[**]Research partially supported by University of Kansas GRF Award #2301055 and Jilin University Tang Ao-Qing Professorship

[††]Research partially supported by National Science Foundation and Natural Science and Engineering Research Council of Canada

that the host population has a small intrinsic growth rate $\varepsilon > 0$, the difference between the natural birth rate and the natural death rate. This slow-growth assumption is not biologically unrealistic. Demographic data has shown that annual population growth rates in many industrialized countries have been only slightly above zero, in the range 0.01-0.001 per year, for a long period of time [26]. The slow-growth assumption may also apply to animal population on live-stock farms, where, for economic reasons, population may be kept near its carrying capacity where the growth is close to zero. Using the intrinsic growth rate ε as a perturbation parameter, we show that a standard SIR epidemic model can be reformulated as a *singularly* perturbed problem. Applying techniques from geometric singular perturbation and global center manifold theory, we prove that, for an open and biologically realistic parameter regime, stable periodic oscillations exist in rather simple SIR models. Furthermore, our analysis demonstrates that the periodic solution has a large amplitude of order $O(1)$. This overcomes a common drawback of Hopf bifurcation analysis where the bifurcating periodic solutions are of small amplitude.

Our analysis reveals an important characteristic of the model, the existence of a *turning point*. This is a point on the slow manifold with the population size at the critical community size to support an epidemic [2]. Accompanying the presence of the turning point is a critical phenomenon called *delay of stability loss*, in which a solution starts with a fast motion to approach a vicinity of the slow manifold, moves slowly along the slow manifold, passes through the turning point and continues the slow motion along the slow manifold, then, up to some point, moves away from the slow manifold in a fast motion (see, e.g., [7, 17, 18, 23, 24, 25]). In our model, the slow manifold is in the disease-free region, and the time period a solution spends in the vicinity of the slow manifold corresponds to the inter-epidemic period with low disease incidence: the period between epidemics (fast dynamics) away from the slow manifold. The fast-slow oscillations characterize the global dynamics of the model and capture the qualitative nature of the oscillatory behaviors in empirical disease data. Our analysis of the simple SIR models has demonstrated that, the existence of turning points and the associated delay of stability loss due to the slow growth of the population, offers a simple and robust mechanism for sustained oscillations of disease incidence.

Mathematically, our singular perturbation analysis is carried out for a 3-dimensional system, and the presence of turning points leads to a significant challenge. At a turning point, two eigenvalues are zero. This results in the loss of normal hyperbolicity of the 1-dimensional slow manifold, and the standard geometric singular perturbation theory of Fenichel [9, 10] no longer applies. Another difficulty we encounter in the analysis is having to deal with the nonlinear dynamics in a large neighbourhood of the slow manifold. Such a difficulty does not seem to appear in the analysis of many other biological models, e.g., in the analysis of relaxation oscillation of a predator-prey model [19].

The primary objective of our paper is to establish the mathematical framework and carry out detailed mathematical analysis for the singular perturbation approach to the study of epidemic models. We have chosen a simple SIR model to keep the mathematical technicality to its minimum, and the analysis is applicable to more complex models. In a subsequent paper, we will investigate relaxation oscillations in an SEIR model and give a more in depth discussion of biological implications of the mathematical results. Singular perturbation approach and associated asymptotic analysis have been successfully applied to the analysis of relaxation oscillation phenomena in many mechanical, physical, chemical, and biological systems. We hope that our study will lead to more applications of singular perturbation analysis to the study of disease transmission processes.

2. The model and statements of main results.

2.1. The model problem. Consider the spread of an infectious disease in a host population of size N. Partition the population into susceptible, infectious, and recovered classes, and denote the sizes by S, I, and R, respectively, so that $N = S + I + R$.

In the absence of the disease, we assume that N satisfies

$$N' = \varepsilon g(N),$$

where constant $\varepsilon > 0$ is assumed to be small. A typical example of $g(N)$ is the quadratic form $N(1 - N/N^*)$, such that N has the logistic growth with carrying capacity N^* and intrinsic growth rate ε. It is natural to require the following.

(A1) *The function $g(N)$ satisfies*

$$g''(N) < 0, \; g(0) = g(N^*) = 0 \;\; \textit{for some} \;\; N^* > 0.$$

As a consequence, we have the following properties.

LEMMA 2.1. *Assume* **(A1)***. Then, N^* is unique, and $g(N) > 0$ for $N \in (0, N^*)$ and $g(N) < 0$ for $N > N^*$.*

We further assume that the per capita natural death rate is a constant $d > 0$, and newborns $b(N)$ has a density dependent form $b(N) = dN + \varepsilon g(N)$. For simplicity, we assume that all newborns are susceptible to the disease. We consider the type of diseases that spreads through direct contact of hosts, and incidence is given by $h(S, N)I$, where $h(S, N)$ is a smooth function. We will assume the following basic properties on $h(S, N)$.

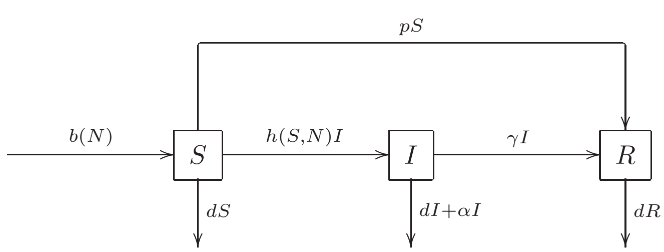
(A2) *The function $h(S, N)$ is increasing in S and $h(0, N) = 0$.*

A specific form of $h(S, N)$ that is commonly used is

$$h(S, N) = \frac{\beta(N) S^q}{K + S}, \quad q \geq 1, \; K \geq 0.$$

This incidence form $h(S, N)I$ includes the bilinear incidence βSI (with $K = 0, \; q = 2$), nonlinear incidence $\beta S^{q-1} I$ (with $K = 0, \; q > 2$), standard incidence $\frac{\lambda SI}{N}$ (with $\beta(N) = \lambda/N, K = 0, q = 2$), and the saturation incidence $\frac{\beta SI}{K+S}$.

The transmission process is demonstrated in the following diagram.



The parameter γ denotes the recovery rate, and p denotes vaccination rate for a simple vaccination strategy. We assume that the infectious individuals suffer a disease-caused death αI with a constant rate α. It is assumed that disease confers permanent immunity, and all parameters are assumed to positive. The transfer diagram leads to the following system of differential equations

$$\begin{aligned} S' &= b(N) - h(S,N)I - (d+p)S, \\ I' &= h(S,N)I - (d+\gamma+\alpha)I, \\ R' &= pS + \gamma I - dR. \end{aligned} \tag{2.1}$$

As a consequence, the total population size N satisfies

$$N' = \varepsilon g(N) - \alpha I. \tag{2.2}$$

It follows that, for $\varepsilon > 0$ and $\alpha > 0$, N varies with time, and model (2.1) is a 3-dimensional system.

Using $b(N) = dN + \varepsilon g(N)$ and replacing the R equation by (2.2), we rewrite the model (2.1) as the following equivalent system

$$\begin{aligned} S' &= dN + \varepsilon g(N) - h(S,N)I - (d+p)S, \\ I' &= h(S,N)I - aI, \\ N' &= \varepsilon g(N) - \alpha I, \end{aligned} \tag{2.3}$$

where $a = d + \alpha + \gamma$. We study system (2.3) for $\varepsilon \geq 0$ in the feasible region

$$\mathcal{D} = \{(S,I,N) \in \mathbb{R}^3 : S \geq 0,\ I \geq 0, N \geq 0 \ \text{ and } S + I \leq N \leq N^*\}.$$

From Lemma 2.1 and equation (2.2) we know that $N' < 0$ if $N > N^*$. If follows that the region $\mathcal{D}$ is positively invariant with respect to system (2.3) and globally attracts all non-negative solutions of (2.3).

Global dynamics of model (2.3) for the case $\varepsilon = 0$ were studied in [11]. It was shown that the essential dynamics consist of a local, stable, two dimensional invariant manifold and, on the invariant manifold, a line of equilibria exists and all other solutions are heteroclinic orbits each connecting a pair of equilibria. This is a highly unstable structure and small perturbations can dramatically change the nature of the global dynamic. We will study the global dynamics of (2.3) for the case $\varepsilon > 0$ and show that, under certain conditions, there exists a stable relaxation periodic cycle for small ε.

In the rest of this section, we describe the structure of the equilibria and their stability, and state our main result on relaxation oscillations.

2.2. Structure of equilibria and statement of the main result. For $\varepsilon \geq 0$, $(0,0,0)$ and $(S^*, 0, N^*)$, with N^* defined in **(A1)** and $S^* = dN^*/(d+p)$ are equilibria of system (2.3).

Proposition 2.2. *There are no other equilibria for $\varepsilon > 0$ if and only if*

$$h(S(N), N) < a = d + \gamma + \alpha$$

for all N where

$$S(N) = \frac{d}{d+p}N - \varepsilon \frac{a-\alpha}{\alpha(d+p)} g(N).$$

Furthermore, if $h(S,N) < a$ *for all* S *and* N*, then the equilibrium* $(S^*, 0, N^*)$ *attracts all solutions except* $(0,0,0)$*. The global dynamics are trivial.*

Proof. The first statement can be checked directly. Assume that $h(S,N) < a$ for all S and N. Then, for any initial condition other than $(0,0,0)$, the solution $(S(t), I(t), N(t))$ satisfies that $I(t) \to 0$ as $t \to +\infty$ and, on the plane $\{I = 0\}$, $(S(t), N(t)) \to (S^*, N^*)$ as $t \to \infty$. ∎

In this work, we will focus on the cases where nontrivial dynamics are possible. In view of the statements in Proposition 2.2, we assume the following.

(A3) *The function* $h(dN/(d+p), N)$ *is non-decreasing for* $N \in (0, N^*)$*. There is a unique* $N_0 \in (0, N^*)$ *such that* $h(S_0, N_0) = a$ *where* $S_0 = \frac{d}{d+p} N_0$*. Furthermore,* $\frac{d}{d+p} h_S(S_0, N_0) + h_N(S_0, N_0) > 0$*.*

We note that $(S_0, 0, N_0)$ in **(A3)** has both dynamical and biological significance. In the case when $h(S,N) = \beta S$, the equation $h(S_0, N_0) = a$ becomes $\beta S_0 = d + \gamma + \alpha$ and thus $S_0 = (d + \gamma + \alpha)/\beta$. In the classical SIR model with no demography ($b = d = 0$) and no disease-caused death ($\alpha = 0$), we have $S_0 = \gamma/\beta$, which is known as the critical size of susceptible population to sustain an epidemic [2, 12]. The dynamical significance of point $(S_0, 0, N_0)$ is that it is a turning point, whose existence is the foundation of the relaxation oscillation phenomenon.

Lemma 2.3. *Assume that* **(A3)** *holds. For* $\varepsilon > 0$ *small, there is a unique equilibrium* $E_\varepsilon = (S_\varepsilon, I_\varepsilon, N_\varepsilon)$ *with* $S_\varepsilon, I_\varepsilon, N_\varepsilon > 0$*, and* $E_\varepsilon \to (S_0, 0, N_0)$ *as* $\varepsilon \to 0$*.*

Proof. In addition to $(0,0,0)$ and $(S^*, 0, N^*)$, other equilibria of system (2.3) are determined by

$$h(S,N) = a,\ I = \frac{\varepsilon}{\alpha} g(N),\ S = \frac{d}{d+p} N - \varepsilon \frac{a-\alpha}{\alpha(d+p)} g(N).$$

The N coordinates are roots of

$$f(N;\varepsilon) := h\left(\frac{d}{d+p} N - \varepsilon \frac{a-\alpha}{\alpha(d+p)} g(N), N\right) - a = 0.$$

It follows from assumption **(A3)** that

$$f(N_0;0) = 0,\ f_N(N_0;0) = \frac{d}{d+p} h_S(S_0,N_0) + h_N(S_0,N_0) > 0.$$

An application of the Implicit Function Theorem gives that, for $\varepsilon > 0$ small, there is N_ε such that $f(N_\varepsilon;\varepsilon) = 0$ and $N_\varepsilon \to N_0$ as $\varepsilon \to 0$. Note that the corresponding I-coordinate is $I_\varepsilon = \frac{\varepsilon}{\alpha} g(N_\varepsilon) > 0$ for $\varepsilon > 0$ small. ∎

Stability of equilibria of system (2.3) is described in the next result, whose proof is given in Appendix I. Denote

$$\Delta_0 = \Big(\frac{a}{\alpha} - \frac{d}{d+p}\Big) h_S(S_0,N_0) g(N_0) - (d+p) g_N(N_0). \tag{2.4}$$

Theorem 2.4. *Assume that* **(A1)**, **(A2)** *and* **(A3)** *hold. Then, for* $\varepsilon > 0$ *small,*

(i) *the equilibria* $(0,0,0)$ *and* $(S^*, 0, N^*)$ *are saddles each with two negative eigenvalues and one positive eigenvalue;*

(ii) *the equilibrium* E_ε *always has a real negative eigenvalue and a pair of complex conjugate eigenvalues. If* $\Delta_0 > 0$*, then the complex eigenvalues have a negative real part and* E_ε *is locally stable; if* $\Delta_0 < 0$*, then the complex eigenvalues have a positive real part and* E_ε *is a saddle.*

A rough statement of our main result is given in the following. A more technical statement (Theorem 4.3) of this result and its proof will be given in Section 4.

THEOREM 2.5. *Assume that* **(A1)**, **(A2)** *and* **(A3)** *hold. Then, for system* (2.3) *with* $\varepsilon > 0$ *small, one of the following holds*

(i) *the equilibrium* E_ε *is a sink and it attracts all orbits except equilibria* $(0,0,0)$ *and* $(S^*, 0, N^*)$*;*

(ii) *there exists an invariant annulus-like or disk-like two dimensional region that attracts all but equilibria orbits and contains at least one stable periodic orbit.*

We note that, for fixed $\varepsilon > 0$ small, as Δ_0 varies from positive to negative, in view of statement (ii) in Theorem 2.4, it is possible that a periodic solution can be created through a supercritical Hopf bifurcation of E_ε. This has been extensively studied for many biological models in the literature. We will not pursue along this direction. Instead, we will investigate the existence of a relaxation oscillation using a global approach. More precisely, we will treat ε as a parameter, first understand the limiting global behaviours when $\varepsilon = 0$, and then examine how a relaxation oscillation is created for $\varepsilon > 0$, far from the endemic equilibrium E_ε. In particular, the example in Section 4.3 shows that a stable relaxation oscillation may exist even if E_ε is stable.

3. Global dynamics of system (2.3) **for** $\varepsilon = 0$**.** In this section, we give a complete description of the dynamics for the limiting system (2.3) at $\varepsilon = 0$. The result extends the work in [11] for a semi-local description of the dynamics. We recall that system (2.3) for $\varepsilon = 0$ is

$$\begin{aligned} S' &= dN - h(S,N)I - (d+p)S, \\ I' &= (h(S,N) - a)I, \\ N' &= -\alpha I, \end{aligned} \tag{3.1}$$

with feasible region $\mathcal{D} = \{(S,I,N) \in \mathbb{R}^3 \ : \ S \geq 0,\ I \geq 0, N \geq 0,\ S + I \leq N \leq N^*\}$, which is positively invariant for (3.1).

It can be verified that the disease-free plane $\{I = 0\}$ and the half-line

$$\mathcal{Z}_0 := \left\{ S = \frac{dN}{d+p},\ I = 0,\ N \geq 0 \right\}$$

are both invariant under system (3.1). In particular, $\mathcal{Z}_0$ consists of equilibria of (3.1).

3.1. A complete characterization of dynamics of (3.1). On the invariant plane $\{I = 0\}$, all solutions $(S(t), I(t), N(t))$ satisfy that

$$I(t) \equiv 0,\ N(t) \equiv N(0), \ \text{ and } \ S(t) \to \frac{d}{d+p} N(0) \ \text{ as } \ t \to \infty.$$

The set $\mathcal{Z}_0$ of equilibria attracts all solutions within $\{I = 0\}$.

The linearization at each point $(dN/(d+p), 0, N) \in \mathcal{Z}_0$ is

$$\begin{pmatrix} -(d+p) & -h(dN/(d+p),N) & d \\ 0 & h(dN/(d+p),N) - a & 0 \\ 0 & -\alpha & 0 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = -(d+p) < 0$, and $\lambda_3 = h(dN/(d+p), N) - a$. The eigenvectors associated with λ_1 and λ_2 span the plane $\{I = 0\}$ and that associated with λ_3 is transversal to the plane $\{I = 0\}$. The eigenvalue $\lambda_3 = h(dN/(d+p), N) - a$ changes sign across the point $(S_0, 0, N_0) \in \mathcal{Z}_0$, where S_0 and N_0 are defined in (**A3**).

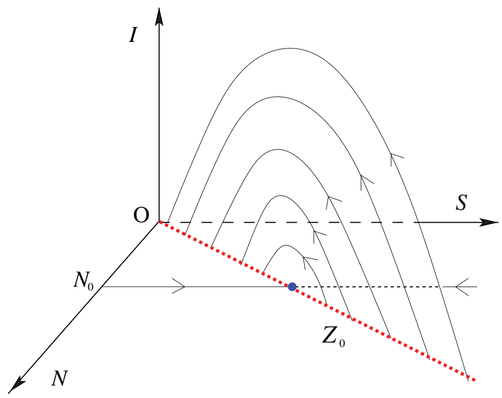


FIG. 1. *Heteroclinic structure of (2.3) with $\varepsilon = 0$*

The complete dynamics for the case $\varepsilon = 0$ are described in the following result, and depicted in Figure 1. The proof is given in Appendix I.

THEOREM 3.1. *Assume that* **(A2)** *and* **(A3)** *are satisfied. The the following statements hold.*

(i) *Every solution of system* (3.1) *is bounded for* $t \geq 0$ *and the set* $\mathcal{Z}_0$ *is the global attractor;*

(ii) *The unstable manifold of each equilibrium* $(dN/(d+p), 0, N) \in \mathcal{Z}_0$ *with* $N > N_0$ *is a heteroclinic orbit to an equilibrium* $(\bar{S}, 0, \bar{N}) \in \mathcal{Z}_0$ *with* $0 < \bar{N} < N_0$. *The relationship* $\bar{N}_1 < \bar{N}_2 < N_0$ *if* $N_1 > N_2 > N_0$ *holds. Furthermore,* $\lim_{N\to\infty} \bar{N} = N_\infty \in (0, N_0)$.

We denote by $M(\mathcal{Z}_0)$ the two dimensional invariant manifold that consists of heteroclinic orbits established in Theorem 3.1 - (ii), and define a map

$$H : (N_0, \infty) \to (0, N_0), \quad H(N) = \bar{N}$$

where $\bar{N}$ is defined by the heteroclinic orbits in Theorem 3.1 - (ii). The invariant manifold $M(\mathcal{Z}_0)$ and the map H will play important roles in our results on relaxation oscillations for model (2.3) with $\varepsilon > 0$.

3.2. Persistence of $M(\mathcal{Z}_0)$ for $\varepsilon > 0$ small. We are interested in whether or not the invariant manifold $M(\mathcal{Z}_0)$ will persist for $\varepsilon > 0$ small; that is, for $\varepsilon > 0$ small, whether or not there is an invariant manifold M_ε for system (2.3) so that $M_\varepsilon \to M(\mathcal{Z}_0)$ as $\varepsilon \to 0$.

Recall that, when $\varepsilon = 0$, for each equilibrium $w = (\frac{d}{d+p}N, 0, N) \in \mathcal{Z}_0$, the eigenvalues of the linearization at w are

$$\lambda_1 = 0, \quad \lambda_2 = -(d+p), \quad \lambda_3 = h\Big(\frac{d}{d+p}N, N\Big) - a.$$

Based on the relative size of eigenvalues, the consideration can be divided into two cases.

Case 1: $a = d + \alpha + \gamma < d + p$. It follows that $h(\frac{d}{d+p}N, N) - a > -(d+p)$ for all $N \geq 0$. At each point $w \in \mathcal{Z}_0$, we have $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_2$. Applying a center manifold theorem in [5, 6] to the invariant set $\mathcal{Z}_0$, we obtain the existence of a 2-dimensional center manifold $W^c(\mathcal{Z}_0)$. The center manifold $W^c(\mathcal{Z}_0)$ is invariant under (3.1), contains $\mathcal{Z}_0$ and all orbits bounded in the vicinity of $\mathcal{Z}_0$. At each $w \in \mathcal{Z}_0$, the tangent space $T_w W^c(\mathcal{Z}_0)$ is spanned by the eigenvectors associated with λ_1 and λ_3 (both are larger than λ_2). Most importantly, the center manifold theorem guarantees the persistence of $W^c(\mathcal{Z}_0)$ for $\varepsilon > 0$ small. In general, a center manifold may not be unique but any center manifold will contain all orbits that are bounded in the vicinity of $\mathcal{Z}_0$. Therefore, for this model problem, $W^c(\mathcal{Z}_0)$ coincides with $M(\mathcal{Z}_0)$ and is unique, and $M(\mathcal{Z}_0)$ persists for $\varepsilon > 0$.

Case 2: $a = d+\alpha+\gamma \geq d+p$**.** In this case, there exists a unique $\hat{N} \in [0, N_0)$ such that $h(\frac{d}{d+p}N, N) - a > -(d+p)$ for $N > \hat{N}$ but $h(\frac{d}{d+p}N, N) - a \leq -(d+p)$ for $N \leq \hat{N}$. The general results on center manifolds in [5, 6] cannot be applied to the whole set $\mathcal{Z}_0$ to obtain a two-dimensional center manifold. For any fixed $\delta > 0$, the results in [5, 6] can be applied to the subset $\mathcal{Z}_0^\delta := \mathcal{Z}_0 \cap \{N \geq \hat{N} + \delta\}$ but the corresponding center manifold $W^c(\mathcal{Z}_0^\delta)$ will only be a proper subset of $M(\mathcal{Z}_0)$. It turns out, for $\varepsilon > 0$, parts of relaxation oscillations could occur outside $W^c(\mathcal{Z}_0^\delta)$ for all $\delta > 0$. We take the advantage of a crucial property that the set $\{I = 0\}$ is invariant under system (2.3) for all $\varepsilon \geq 0$, and show that $M(\mathcal{Z}_0)$ persists for $\varepsilon > 0$ small even though it is *not* normally hyperbolic. This is established in Appendix II. This persistence result appears to be contradictory to Mãñé's result that an invariant manifold is persistent if and only if it is normally hyperbolic ([21]). It is not, since the persistence in Mãñé's result is with respect to all small perturbations while the perturbations in our system are special: they leave the set $\{I = 0\}$ invariant. As mentioned above, it is possible that a portion of a relaxation oscillation occurs over the region where $N < \hat{N}$. In the limit as $\varepsilon \to 0$, this portion approaches $\mathcal{Z}_0$ along the eigenvector associated with $-(p+q)$ in general.

3.3. The map H near N_0. The map $H : (N_0, \infty) \to (0, N_0)$ defined in Theorem 3.1 will be a key ingredient for our main result on relaxation oscillations. Detailed global properties of H seem to be not achievable. On the other hand, it is possible to examine properties of H near N_0 based on an approximation of $W^c(\mathcal{Z}_0)$ near $(S_0, 0, N_0)$, or simply, a center manifold $W^c(S_0, 0, N_0)$ of the equilibrium $(S_0, 0, N_0)$. Note that, the eigenvalues at $(S_0, 0, N_0)$ are $\lambda_1 = \lambda_3 = 0 > \lambda_2 = -(d+p)$. Thus, for an equilibrium $w \in \mathcal{Z}_0$ near $(S_0, 0, N_0)$, the corresponding eigenvalues satisfy $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_2$. As a consequence, $W^c(S_0, 0, N_0) \subset M(\mathcal{Z}_0)$, and hence, is unique. It should be pointed out that, in general, a center manifold may not be unique.

3.3.1. An approximation of the center manifold $W^c(S_0, 0, N_0)$. We look for an approximation of the center manifold $W^c(S_0, 0, N_0)$ in the vicinity of $(S_0, 0, N_0)$ as the graph of a function

$$S = \frac{d}{d+p}N + U(N, I)I = \frac{d}{d+p}N + a_0(N)I + a_1(N, I)I^2.$$

The form is justified by the fact that $\{I = 0\}$ is invariant and $W^c(S_0, 0, N_0) \cap \{I = 0\} \subset \mathcal{Z}_0$.

Taking the derivative of $S = \frac{d}{d+p}N + U(N, I)I$ with respect to t, we have

$$S' = \frac{d}{d+p}N' + a_0' I N' + a_{1,N} I^2 N' + a_0 I' + 2a_1 I I' + a_{1,I} I^2 I'.$$

From (3.1) we have

$$\begin{aligned}
dN-&h\Big(\frac{d}{d+p}N+a_0(N)I+a_1(N,I)I^2,N\Big)I\\
&-(d+p)\Big(\frac{d}{d+p}N+a_0(N)I+a_1(N,I)I^2\Big)\\
=&-\alpha\Big(\frac{d}{d+p}+a_0'I+a_{1,N}I^2\Big)I\\
&+(a_0+2a_1I+a_{1,I}I^2I')\Big(h\Big(\frac{d}{d+p}N+a_0(N)I+a_1(N,I)I^2,N\Big)-a\Big)I.
\end{aligned}$$

Expanding h at the point $(bN/(b+p),N)$ we get

$$\begin{aligned}
dN-&h\Big(\frac{d}{d+p}N,N\Big)I-(d+p)\Big(\frac{d}{d+p}N+a_0(N)I\Big)+O(I^2)\\
=&-\frac{\alpha d}{d+p}I+a_0\Big(h\Big(\frac{d}{d+p}N,N\Big)-a\Big)I+O(I^2).
\end{aligned}$$

Comparing coefficients of I^0 and I^1 we obtain

$$a_0(N)=\frac{\frac{\alpha d}{d+p}-h\left(\frac{d}{d+p}N,N\right)}{d+p+h\left(\frac{d}{d+p}N,N\right)-a}.$$

Note that we restrict the approximation of $W^c(S_0,0,N_0)$ near $(S_0,0,N_0)$. Thus, N is close to N_0, and hence, the denominator in the above expression is close to $d+p>0$.

Near equilibrium $(S_0,0,N_0)$, the center manifold $W^c(S_0,0,N_0)$ is given as the graph of the function

$$\begin{aligned}
S=&\frac{d}{d+p}N+a_0(N)I+O(I^2)\\
=&\frac{d}{d+p}N+\frac{\frac{\alpha d}{d+p}-h\left(\frac{d}{d+p}N,N\right)}{d+p+h\left(\frac{d}{d+p}N,N\right)-a}I+O(I^2).
\end{aligned}\tag{3.2}$$

On the center manifold $W^c(S_0,0,N_0)$ and near $(S_0,0,N_0)$, system (3.1) is reduced to a 2-dimensional system

$$\begin{aligned}
I'&=h\left(\frac{dN}{d+p}+a_0(N)I+O(I^2),N\right)I-aI,\\
N'&=-\alpha I.
\end{aligned}\tag{3.3}$$

3.3.2. Properties of the map H near N_0.

PROPOSITION 3.2. *The map H satisfies $H(N_0)=N_0$, $H'(N_0)=-1$ and*

$$H''(N_0)=-\frac{2}{\alpha}a_0(N_0)h_S\Big(S_0,N_0\Big).$$

Proof. Set $v(t) = N(t) - N_0$. In terms of (I, v), system (3.3) becomes

$$I' = h\left(S_0 + \frac{dv}{d+p} + a_0(N_0+v)I + O(I^2), N_0 + v\right) I - aI,$$
$$v' = -\alpha I.$$

Since $\{I = 0\}$ is invariant and the map H is defined through the dynamics where $I > 0$, we divide the two equations above to get

$$\frac{dI}{dv} = -\frac{1}{\alpha}\left(h\Big(S_0 + \frac{dv}{d+p} + a_0(N_0+v)I + O(I^2), N_0+v\Big) - a\right). \tag{3.4}$$

Expanding the right-hand side at $v = 0$ leads to

$$\begin{aligned} &h\Big(S_0 + \frac{dv}{d+p} + a_0(N_0+v)I + O(I^2), N_0+v\Big) - a \\ &\quad = h_S \cdot \Big(\frac{dv}{d+p} + (a_0 + a_0'v)I\Big) + h_N \cdot v + \frac{1}{2}h_{SS}\cdot\Big(\frac{dv}{d+p} + (a_0+a_0'v)I\Big)^2 \\ &\qquad + h_{SN}\cdot\Big(\frac{dv}{d+p} + (a_0+a_0'v)I\Big)v + \frac{1}{2}h_{NN}\cdot v^2 + O(I^2, v^2 I, I^3), \end{aligned}$$

where the partial derivatives of h are all evaluated at (S_0, N_0), and $a_0 = a_0(N_0)$ and $a_0' = a_0'(N_0)$. Denote

$$L = \frac{d}{d+p}h_S + h_N \quad\text{and}\quad Q = \frac{d^2}{(d+p)^2}h_{SS} + \frac{2d}{d+p}h_{SN} + h_{NN}. \tag{3.5}$$

Equation (3.4) becomes

$$\begin{aligned} \frac{dI}{dv} &= -\frac{1}{2\alpha}(2Lv + Qv^2) \\ &\quad - \frac{1}{\alpha}\left(h_S + \Big(\frac{d}{d+p}h_{SS} + h_{SN}\Big)v\right)(a_0 + a_0'v)I + O(I^2, v^2 I, v^3). \end{aligned} \tag{3.6}$$

By the existence and smoothness of solutions and smooth dependence of solutions on parameters, for v small, we look for solutions of the form $I(v) = c_0 + c_1 v + c_2 v^2 + O(v^3)$. Substituting $I(v)$ into (3.6) and comparing terms of like powers in v we get

$$\begin{aligned} c_1 &= -\frac{1}{\alpha}a_0 h_S c_0 + O(c_0^2), \\ c_2 &= -\frac{1}{2\alpha}L - \frac{1}{2\alpha}\Big(a_0' h_S - \frac{1}{\alpha}a_0^2 h_S^2 + \frac{d}{d+p}a_0 h_{SS} + a_0 h_{SN}\Big)c_0 + O(c_0^2). \end{aligned} \tag{3.7}$$

Thus, near $v = 0$, the solution of (3.6) is

$$I(v) = c_0 - \frac{a_0 h_S v}{\alpha}c_0 - \frac{L}{2\alpha}v^2 + O(c_0 v^2). \tag{3.8}$$

To define H, we need the initial condition $I(v) = 0$ at $v = N - N_0$ for $N > N_0$ and $N - N_0 \ll 1$. We can then determine the value c_0 corresponding to this initial condition. From (3.8),

$$0 = c_0 - \frac{a_0 h_S \cdot (N - N_0)}{\alpha}c_0 - \frac{L}{2\alpha}(N-N_0)^2 + O(c_0(N-N_0)^2),$$

or equivalently,

$$c_0\left(1-\frac{1}{\alpha}a_0h_S\cdot(N-N_0)+O(N-N_0)^2\right)=\frac{L}{2\alpha}(N-N_0)^2.$$

Thus,

$$c_0=\frac{L}{2\alpha}(N-N_0)^2+\frac{La_0h_S}{2\alpha^2}(N-N_0)^3+O(N-N_0)^4. \tag{3.9}$$

The value of $H(N)$ satisfies $I(H(N)-N_0)=0$. Note that

$$H(N)-N_0=H(N)-H(N_0)=H'(N_0)(N-N_0)+\frac{1}{2}H''(N_0)(N-N_0)^2+O(N-N_0)^3.$$

It then follows from $I(H(N)-N_0)=0$, (3.7), (3.8) and (3.9) that

$$\begin{aligned}0=&\frac{L}{2\alpha}(N-N_0)^2+\frac{La_0h_S}{2\alpha^2}(1-H'(N_0))(N-N_0)^3\\&-\frac{L}{2\alpha}\left(H'(N_0)(N-N_0)+\frac{1}{2}H''(N_0)(N-N_0)^2\right)^2+O(N-N_0)^4.\end{aligned}$$

Comparing $(N-N_0)^2$ terms gives that $H'(N_0)=-1$ (due to also that H is decreasing). The $(N-N_0)^3$ terms then yield

$$\frac{La_0h_S}{\alpha^2}+\frac{L}{2\alpha}H''(N_0)=0.$$

This completes the proof. □

3.4. A discussion and the link to the main result. In this section, we summarize the results for system (3.1), discuss the impact of the sign changing eigenvalue $h(S,N)-a$, and provide mathematical and biological motivations for our main result.

For $N<N_0$, $h(dN/(d+p),N)-a<0$, and it implies that, for an initial state $(S(0),I(0),N(0))$ near the region $\{I=0,\ N<N_0\}$, $I(t)$ decreases and the solution converges to an equilibrium in $\mathcal{Z}_0$ with $N<N_0$. Biologically speaking, if the total population N is below the critical community size N_0, or equivalently, the number of susceptibles S is below the critical size $S_0=dN_0/(d+p)$, then the population can not sustain an epidemic and the disease dies out.

We describe the dynamics for solutions with initial conditions near the other region $\{I=0,\ N>N_0\}$ in three stages.

Stage I: For an initial state $(S(0),I(0),N(0))$ with $N>N_0$, $h(dN/(d+p),N)-a>0$ for small $t>0$ and $I(t)$ increases initially. In biological terms, if the population size surpasses the critical community size N_0, then any initial infection will lead to a disease outbreak.

Stage II: As $I(t)$ increases away from $\{I=0\}$, the dynamics outside $\{I=0\}$ become dominant; in particular, $N(t)$ decreases. Once $N(t)<N_0$ (or equivalently $S(t)<S_0$), we know that $h(S,N)-a<0$ and $I(t)$ begins to decrease.

Stage III: As time goes on, $I(t)$ continues to decrease. Eventually the solution will enter a vicinity of the region $\{I=0,\ N<N_0\}$ and is attracted to an equilibrium in $\mathcal{Z}_0$ with $N<N_0$. The disease outbreak leads to an epidemic but the disease eventually dies out.

We see that, when $\varepsilon=0$, the model (3.1) only describes epidemics of the disease; the disease eventually dies out. There is no mechanism for the recurrence of the disease if the population growth is zero. This is parallel to the classical SIR model with no demography and disease caused death.

When $\varepsilon > 0$, solutions of system (2.3) with $N(0) > N_0$ and $I(0)$ small go through Stages I and II as described above, but Stage III will no longer be the terminal stage. In this case, the disease-free set $\{I = 0\}$ remains invariant. The half line $\mathcal{Z}_0$ also remains invariant but is no longer a set of equilibria. Instead, $\mathcal{Z}_0$ becomes an orbit for which N increases with t with speed of order $O(\varepsilon)$. For this reason, $\mathcal{Z}_0$ is called the slow manifold for small $\varepsilon > 0$.

Stage IV: When $\varepsilon > 0$, for a solution in the vicinity of $\mathcal{Z}_0$ with $N < N_0$ during Stage III, it will follow an orbit on the slow manifold $\mathcal{Z}_0$ by the continuous dependence on initial conditions. As $N(t)$ increases beyond the critical community size N_0 and the solution enters the region $\{I = 0,\ N > N_0\}$. As a consequence, $I(t)$ begins to increase and the solution repeats Stages I - III, leading to another epidemic. The period during which the solution moves along the slow manifold is the inter-epidemic period. We see that, when $\varepsilon > 0$, the fall of susceptible population during an epidemic and the recovery of the susceptible population during the inter-epidemic period produces an oscillating behavior.

In summary, for $\varepsilon > 0$ small, all orbits, except for solutions on $\{I = 0\}$ and E_ε of system (2.3), will exhibit oscillating behaviours. Three key conditions are responsible for the mechanism of oscillation:

(C0) the plane $\{I = 0\}$ is invariant for $\varepsilon \geq 0$,

(C1) the assumption on the natural growth $g(N)$ of the total population in the absence of disease, and

(C2) the sign changing assumption of the eigenvalue $\lambda_3 = h(S, N) - a$.

In the language of singular perturbation theory, condition (C2) means that the point $(S_0, 0, N_0)$ at which $h(S_0, N_0) - a = 0$ is a *turning point*. This point marks the level of N or S that separates the region of disease decline from that of disease rise. The condition (C1) implies that, on $\mathcal{Z}_0$, with the population growth and increase of susceptibles from newborns, all orbits move from region of disease decline where $N < N_0$ to region of disease rise where $N > N_0$. Conditions (C0) - (C2) imply that the turning point $(S_0, 0, N_0)$ is associated with the delay of stability-loss ([7, 17, 18, 23, 24, 25]). We emphasize that, while condition (C0) holds true naturally for the specific model we consider, it is, however, highly degenerate in general when turning points are present. Without condition (C0), presence of turning points can make $\{I = 0\}$ non-normally hyperbolic ([9, 15]) and destroy the persistence of $\{I = 0\}$ for $\varepsilon > 0$. The impact of turning points for $\varepsilon > 0$ can be extremely difficult to investigate.

While the oscillating behaviors of the SIR model when $\varepsilon > 0$ as described above are biologically intuitive and mathematically verifiable, they can be decayed oscillations. The important mathematical question with biological significance is whether or not there exists a stable periodic oscillation. Our main result in next section characterizes, for the existence of stable periodic solutions, abstract conditions in general and verifiable sufficient conditions in particular. Those periodic oscillations, if exist, will typically have a large period of order $O(\varepsilon^{-1})$.

4. Global dynamics of (2.3) for $\varepsilon > 0$ small. Recall that the two-dimensional invariant manifold $M(\mathcal{Z}_0)$ from Theorem 3.1 persists to M_ε for $\varepsilon > 0$ small. We use the properties of M_ε to establish an abstract result from geometric singular perturbations with turning point, focusing on results on relaxation oscillations. Due to the lack of explicit global representation of $M(\mathcal{Z}_0)$, not all abstract results can be transformed back to the concrete model (2.3) in the sense that the corresponding conditions are not easy to verify. For some sufficient conditions on the existence of periodic oscillations, we are able to transform the conditions back to the original model and they are verifiable.

4.1. Formulation of a singularly perturbed problem. For $\delta > 0$ small, let M be the manifold consisting of all heteroclinic orbits from $(S, 0, N)$ with $N_0 < N < N^* + \delta$, together with the point $(S_0, 0, N_0)$. Then, M persists in the sense as discussed in Case 2 of Section 3.2 and proved in Appendix II. Let M_ε be the perturbed manifold of M for $\varepsilon > 0$ small; that is, M_ε is invariant and $M_\varepsilon \to M$ as $\varepsilon \to 0$. Due to the fact that $\{I = 0\}$ is invariant for all ε and the set $\mathcal{Z}_0$ is normally hyperbolic within $\{I = 0\}$, we have $\mathcal{Z}_0$ persists for $\varepsilon > 0$ small; that is, $\mathcal{Z}_\varepsilon = M_\varepsilon \cap \{I = 0\}$ persists as a portion of the boundary of M_ε.

Let $\phi(u, v; \varepsilon)$ for $(u, v) \in \mathcal{R}$ be a parameterization of the center manifold M_ε, where $\mathcal{R}$ is a bounded domain in $\{u \geq 0, v \geq 0\}$ to be further characterized later on. We require that,

- (P1) for $\varepsilon = 0$, the heteroclinic orbits are determined by $v =$ const so that decreasing in u corresponds to increasing in time;
- (P2) for $\varepsilon \geq 0$, the set $\mathcal{Z}_\varepsilon$ corresponds to the curve $\{v = T(u)\}$ for function $T : (0, U) \to (0, V)$ with $T(U) = V$ where (U, V) corresponds to the point $(S, 0, N^* + \delta) \in \mathcal{Z}$ and hence $\{v = V\}$ corresponds to the heteroclinic orbit from $(S, 0, N^* + \delta) \in \mathcal{Z}_0$; Therefore,
$$\mathcal{R} = \{(u, v) : 0 < u < U, T(u) \leq v < V\};$$
- (P3) for $\varepsilon \geq 0$, the point $(u, v) = (u_0, T(u_0))$ corresponds to the point $(S_0, 0, N_0)$, $(u, v) = (u^0, T(u^0))$ corresponds to $(S^*, 0, N^*)$.

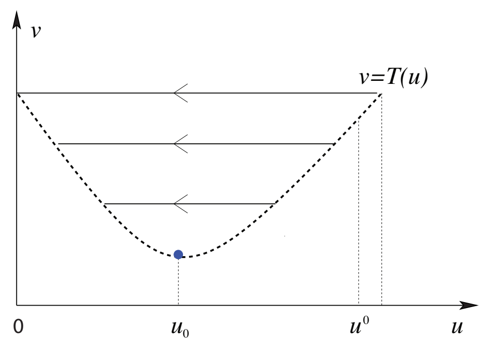


FIG. 2. *Heteroclinic structure of (4.2) with* $\varepsilon = 0$

In terms of $(u, v) \in \mathcal{R}$, suppose that system (2.3) on the center manifold can be put into the form

$$u' = F(u, v; \varepsilon), \quad v' = G(u, v; \varepsilon). \tag{4.1}$$

We now examine the properties that the vector field of system (4.1) must satisfy.

First of all, (P1) implies that $G(u, v; 0) = 0$, $F(u, T(u); 0) = 0$ and $F(u, v; 0) < 0$ for $v > T(u)$. Thus, we can write $G(u, v; \varepsilon) = \varepsilon G_1(u, v; \varepsilon)$, $F(u, v; \varepsilon) = T(u) - v + \varepsilon F_1(u, v; \varepsilon)$. The property (P2) implies that $G_1(u, T(u); \varepsilon) = T_u F_1(u, T(u); \varepsilon)$.

System (4.1) can be rewritten as

$$
\begin{aligned}
u' &= T(u) - v + \varepsilon F_1(u,v;\varepsilon),\\
v' &= \varepsilon T_u(u) F_1(u, T(u);\varepsilon) + \varepsilon (v - T(u)) G_2(u,v;\varepsilon).
\end{aligned}
\tag{4.2}
$$

System (4.2) is a singularly perturbed problem with ε as the singular parameter. As usual, the time t is called the *fast time*, which is the physical time of our problem. In terms of the *slow time* $\tau = \varepsilon t$, system (4.2) becomes

$$
\begin{aligned}
\varepsilon \dot{u} &= T(u) - v + \varepsilon F_1(u,v;\varepsilon),\\
\dot{v} &= T_u(u) F_1(u, T(u);\varepsilon) + (v - T(u)) G_2(u,v;\varepsilon).
\end{aligned}
\tag{4.3}
$$

where the overdot symbol indicates the derivative with respect to τ.

The slow manifold is

$$\mathcal{Z} = \{v = T(u)\}.$$

On the slow manifold $\mathcal{Z}$, the flow is given by

$$u' = \varepsilon F_1(u, T(u);\varepsilon).$$

It has a global sink at $u = u^0$.

We recall that the set $\mathcal{Z}$ is invariant under system (4.2) (or equivalently under system (4.3)) for all $\varepsilon \geq 0$. This property is so crucial in creating oscillations in the system. In fact, one will see late that there is a turning point on $\mathcal{Z}$ and, due to the invariance of $\mathcal{Z}$ for all ε, the turning point causes the delay of stability loss ([7, 17, 18, 23, 24, 25]). We believe that the delay of stability loss is one of the most important mechanism for the oscillation structure in biological population systems.

To describe the delay of stability loss, we define a map $P : (0, u_0) \to (u_0, u^0)$ via

$$\int_u^{P(u)} \frac{T_u(\xi)}{F_1(\xi, T(\xi))}\, d\xi = 0. \tag{4.4}$$

Also, for any $v > \min\{T(u) : u \in (0,\infty)\}$, let $l(v)$ and $r(v)$ be the two solutions of $v = T(u)$ for u with $l(v) < r(v)$ and set $v^0 = T(u^0)$.

Proposition 4.1 (Delay of stability loss). *Fix $\delta > 0$ small. For $\varepsilon > 0$ small, let $(u(\tau;\varepsilon), v(\tau;\varepsilon))$ be the solution of system (4.3) with the initial condition $(u(0), v(0))$ where $u(0) < u_0$ and $v(0) = T(u(0)) + \delta$. Let $\tau(\varepsilon) > 0$ be the time such that $v(\tau(\varepsilon);\varepsilon) = T(u(\tau(\varepsilon))) + \delta$. Then, as $\varepsilon \to 0$, $r(v(\tau(\varepsilon))) \to P(l(v(0))$.*

Note that $P(l(v^0)) < u^0$, and hence, $T(l(v^0)) = v^0 > T(P(l(v^0)))$.

Theorem 4.2. *For $\varepsilon > 0$ small, either the equilibrium $(u_\varepsilon, T(u_\varepsilon))$ is a global attractor of $\mathcal{R}$ for system (4.2) or there is a stable periodic relaxation oscillation. Furthermore,*

(i) *if there exists $u_1 \in (l(v^0), u_0)$ such that $T(u_1) < T(P(u_1))$ then, for $\varepsilon > 0$ small, system (4.2) has a stable periodic relaxation oscillation whose limiting orbit, as $\varepsilon \to 0$, is the union of the heteroclinic orbit from $(P(u^c), T(P(u^c)))$ to $(u^c, T(u^c))$ and the curve on $\{v = T(u)\}$ from $(u^c, T(u^c))$ to $(P(u^c), T(P(u^c)))$ for some $u^c \in (l(v^0), u_1)$ satisfying $T(u^c) = T(P(u^c))$;*

(ii) *if, for every $u \in (l(v^0), u_0)$, $T(u) > T(P(u))$, then, for $\varepsilon > 0$ small, the equilibrium $(u_\varepsilon, T(u_\varepsilon))$ is a global attractor of $\mathcal{R}$ for system (4.2).*

Proof. To prove the statement (i), note that the unstable manifold $W^u(u^0, v^0)$ will approach the left branch of the slow manifold $\{v = T(u)\}$ almost horizontally near the set $\{v = v^0\}$ toward the point $(l(v^0), v^0)$ and then follow the slow orbit through $(l(v^0), v^0)$ up to near the point $(P(l(v^0)), T(P(l(v^0)))$ and leave the slow manifold almost horizontally near the set $\{v = T(P(l(v^0)))\}$. Due to the fact that $T(l(v^0)) = v^0 > T(P(l(v^0)))$, upon leaving the slow manifold at near the point $(P(l(v^0)), T(P(l(v^0))))$, the unstable manifold $W^u(u^0, v^0)$ stays below its initial portion. Therefore, the unstable manifold spirals inward. By the same argument, the existence of u_1 with the property $T(u_1) < T(P(u_1))$ implies that the forward orbit starting from $(u_1 + \delta, T(u_1))$ for some $\delta > 0$ small spirals outward. This orbit together with the unstable manifold $W^u(u^0, v^0)$ encloses a positively invariant region. By the Poincaré-Bendixon Theorem, there is a stable periodic orbit. The above argument also shows that between any numbers $\hat{u}_1, \hat{u}_2 \in (l(v^0), u_0)$ with $\hat{u}_2 < \hat{u}_1$, $T(\hat{u}_1) < T(P(\hat{u}_1))$ and $T(\hat{u}_2) > T(P(\hat{u}_2))$, there is a periodic orbit strictly enclosed by the two orbits through respectively the points $(\hat{u}_1 + \delta, T(\hat{u}_1))$ and $(\hat{u}_2 + \delta, T(\hat{u}_2))$ for some small δ. Therefore, the limiting position of a periodic orbit is exactly as described in the statement.

The proof for the statement (ii) follows from the above argument and we will omit the details here. □

4.2. Statement of the main results for system (2.3). To translate Theorem 4.2 in terms of the original system (2.3), we recall that $H : (N_0, \infty) \to (0, N_0)$ is the function defined as: for $\varepsilon = 0$ and for $(dN/(d + p), 0, N) \in \mathcal{Z}_0$ with $N > N_0$, $(dH(N)/(d + p), 0, H(N)) \in \mathcal{Z}_0$ is the unique equilibrium so that there is a heteroclinic orbit from $(dN/(d+p), 0, N)$ to $(dH(N)/(d+p), 0, H(N))$. The map P defined in (4.4) is given by $P : (0, N_0) \to (N_0, \infty)$ by

$$\int_N^{P(N)} \frac{h(d\xi/(d + p), \xi) - a}{g(\xi)} \, d\xi = 0. \tag{4.5}$$

THEOREM 4.3. *Let $H(N)$ and $P(N)$ be defined as above. For $\varepsilon > 0$ small, either the endemic equilibrium $(S_\varepsilon, I_\varepsilon, N_\varepsilon)$ is a global attractor or there is a stable periodic relaxation oscillation. More precisely,*

(i) *if there exists $N_1 \in (H(N^*), N_0)$ such that $N_1 > H(P(N_1))$ then, for $\varepsilon > 0$ small, system* (2.3) *has a stable periodic relaxation oscillation whose limiting orbit, as $\varepsilon \to 0$, is the union of the heteroclinic orbit from the point $(dP(N^c)/(d + p), 0, P(N^c))$ to the point $(dN^c/(d + p), 0, N^c)$ and the segment on $\mathcal{Z}_0$ from the point $(dN^c/(d + p), 0, N^c)$ to the point $(dP(N^c)/(d + p), 0, P(N^c))$ for some $N^c \in (H(N^*), N_1)$ satisfying $N^c = H(P(N^c))$;*

(ii) *if, for every $N \in (H(N^*), N_0)$, $N < H(P(N))$, then, for $\varepsilon > 0$ small, the endemic equilibrium $(S_\varepsilon, I_\varepsilon, N_\varepsilon)$ is a global attractor for system* (2.3).

Proof. It suffices to show that, for $\varepsilon > 0$, M_ε attracts all solutions except the equilibria $(0, 0, 0)$ and $(S^*, 0, N^*)$. Since M_ε has a region attracting orbits on M_ε and M_ε is normally stable, there is neighbourhood $\mathcal{U}$ of M_ε independent of ε such that, for $\varepsilon > 0$ small enough, any solution entering $\mathcal{U}$ is attracted by the attracting region on M_ε. Therefore, we only need to show that any solution will enter $\mathcal{U}$.

First of all, we see that $N'(t) < 0$ if $N(t) > N^*$. Thus, all solutions are attracted by the domain $\mathcal{D}$ and the domain $\mathcal{D}$ is positively invariant. It can be verified that M_ε attracts all solution on $\{I = 0\}$ except $(0, 0, 0)$ and $(S^*, 0, N^*)$. Now, for a solution $(S(t), I(t), N(t))$ with the initial condition $(S(0), I(0), N(0)) \in \mathcal{D}$ and $I(0) > 0$, by continuity, for $\varepsilon > 0$ small independent of the solution starting in $\mathcal{D}$, the solution will approach a point $(\bar{S}, 0, \bar{N}) \in \mathcal{Z}_0$ with $\bar{N} \le N_0$ and then follow the slow orbit through $(\bar{S}, 0, \bar{N}) \in \mathcal{Z}_0$. Therefore, it enters a neighbourhood of $(S_0, 0, N_0)$ and hence into $\mathcal{U}$. ∎

4.3. Concrete conditions for the existence of relaxation oscillations of system (2.3).

PROPOSITION 4.4. *The map P satisfies $P(N_0) = N_0$, $P'(N_0) = -1$, and*

$$P''(N_0) = \frac{4g'(N_0)L - 2g(N_0)Q}{3g(N_0)L},$$

where L and Q are defined in (3.5)*.*

Proof. It follows from the definition of P that $P(N_0) = N_0$. Differentiating with respect to N on (4.5) we get

$$\frac{h(dP(N)/(d+p), P(N)) - a}{g(P(N))} P'(N) = \frac{h(dN/(d+p), N) - a}{g(N)}. \tag{4.6}$$

Note that

$$\begin{aligned}
P(N) =& P(N_0) + P'(N_0)(N - N_0) + \frac{1}{2}P''(N_0)(N - N_0)^2 + O(N - N_0)^3 \\
=& N_0 + P'(N_0)(N - N_0) + \frac{1}{2}P''(N_0)(N - N_0)^2 + O(N - N_0)^3, \\
P'(N) =& P'(N_0) + P''(N_0)(N - N_0) + O(N - N_0)^2, \\
g(N) =& g(N_0) + g'(N_0)(N - N_0) + \frac{1}{2}g''(N_0)(N - N_0)^2 + O(N - N_0)^3, \\
g(P(N)) =& g(N_0) + g'(N_0)(P(N) - N_0) + \frac{1}{2}g''(N_0)(P(N) - N_0)^2, \\
=& g(N_0) + g'(N_0)\left(P'(N_0)(N - N_0) + \frac{1}{2}P''(N_0)(N - N_0)^2\right) \\
&+ \frac{1}{2}g''(N_0)(P'(N_0))^2(N - N_0)^2 + O(N - N_0)^3,
\end{aligned}$$

and

$$\begin{aligned}
h(dN/(d+p), N) - a =& L(N - N_0) + \frac{1}{2}Q(N - N_0)^2 + O(N - N_0)^3, \\
h(dP(N)/(d+p), P(N)) - a =& L(P(N) - N_0) + \frac{1}{2}Q(P(N) - N_0)^2 + O(N - N_0)^3 \\
=& L\left(P'(N_0)(N - N_0) + \frac{1}{2}P''(N_0)(N - N_0)^2\right) \\
&+ \frac{1}{2}Q(P'(N_0))^2(N - N_0)^2 + O(N - N_0)^3.
\end{aligned}$$

Substituting these expansions into (4.6) and comparing the terms of like-powers in $(N - N_0)$ we get

$$\begin{aligned}
\text{for } N - N_0:& \quad gL(P')^2 = gL \Longrightarrow P' = -1, \\
\text{for } (N - N_0)^2:& \quad -\frac{1}{2}g(LP'' + Q) + g'L - gLP'' = -g'L + \frac{1}{2}gQ \\
&\Longrightarrow P'' = \frac{4g'L - 2gQ}{3gL}.
\end{aligned}$$

This completes the proof. □

Combining Propositions 3.2 and 4.4 we obtain the following result.

PROPOSITION 4.5. *The function $\bar{F} = H \circ P$ satisfies $\bar{F}(N_0) = N_0$, $\bar{F}'(N_0) = 1$, and*

$$\bar{F}''(N_0) = H''(N_0) - P''(N_0) = \frac{2\Delta_0}{(d+p)g(N_0)} + \frac{2}{3}\frac{g'(N_0)L + g(N_0)Q}{g(N_0)L},$$

where Δ_0 is defined in (2.4), and L and Q are defined in (3.5).

As a direct consequence of Theorem 4.3 and Proposition 4.5, we have

COROLLARY 4.6. *If $\bar{F}''(N_0) < 0$, then, for $\varepsilon > 0$ small, there is at least one stable relaxation oscillation.*

Example. We establish the existence of a stable relaxation oscillation in the case that E_ε is stable. More precisely, we take a special case of h that is biologically plausible and show that, for any g satisfying **(A1)**, there are parameter ranges for β and K, dependent on all other fixed parameters so that $\Delta_0 > 0$ in Theorem 2.4, for which the equilibrium E_ε is stable and $\bar{F}''(N_0) < 0$ holds. This guarantees the existence of a stable relaxation oscillation. Thus, a stable relaxation oscillation may exist even the equilibrium E_ε is stable. In this case, there exists at least an unstable periodic orbit between the stable relaxation and the equilibrium. In general, the unstable periodic orbit is not necessarily a relaxation oscillation but a small periodic orbit through a sub-critical Hopf bifurcation.

Consider $h(S,N) = \frac{\beta S}{K+S}$ with $\beta > a$, it can be verified that

$$\begin{aligned}
S_0 =& \frac{aK}{\beta - a} > 0, \quad N_0 = \frac{d+p}{d}\frac{aK}{\beta - a}, \\
L =& \frac{d}{d+p}\frac{\beta K}{(K+S_0)^2}, \quad Q = -\frac{d^2}{(d+p)^2}\frac{2\beta K}{(K+S_0)^3}, \\
\Delta_0 =& \Big(\frac{a}{\alpha} - \frac{d}{d+p}\Big)\frac{\beta K}{(K+S_0)^2} g(N_0) - (d+p)g_N(N_0), \\
\bar{F}''(N_0) =& \frac{2\Delta_0}{(d+p)g(N_0)} + \frac{2}{3}\frac{g'(N_0)L + g(N_0)Q}{g(N_0)L} \\
=& \frac{2}{(d+p)g(N_0)}\Bigg(\Big(\frac{a}{\alpha} - \frac{d}{d+p}\Big)\frac{\beta K}{(K+S_0)^2} g(N_0) \\
& -\frac{2}{3}(d+p)g_N(N_0) - \frac{2d}{3}\frac{K+S_0}{(K+S_0)^2} g(N_0)\Bigg).
\end{aligned}$$

We also note that,

$$\Big(\frac{a}{\alpha} - \frac{d}{d+p}\Big)\frac{\beta K}{(K+S_0)^2} - \frac{2d}{3}\frac{K+S_0}{(K+S_0)^2} = \Big(\frac{a}{\alpha} - \frac{d}{d+p} - \frac{2d}{3(\beta - a)}\Big)\frac{\beta K}{(K+S_0)^2}.$$

Choose $\beta^* > a$ such that

$$\frac{a}{\alpha} - \frac{d}{d+p} - \frac{2d}{3(\beta^* - a)} < 0,$$

and choose K^* such that, for $N_0 = N_0^* = \frac{d+p}{d}\frac{aK^*}{\beta^* - a}$, $g_N(N_0^*) = 0$ holds. Then

$$\begin{aligned}
\Delta_0 =& \Big(\frac{a}{\alpha} - \frac{d}{d+p}\Big)\frac{\beta^* K^*}{(K^* + S_0)^2} g(N_0) > 0, \\
\bar{F}''(N_0^*) =& \frac{2}{d+p}\Big(\frac{a}{\alpha} - \frac{d}{d+p} - \frac{2d}{3(\beta - a)}\Big)\frac{\beta K^*}{(K^* + S_0)^2} < 0.
\end{aligned}$$

This accomplishes the goal of this example.

We note that the construction of the above example strongly indicates that it may not be rare to have stable relaxation oscillations when the endemic equilibrium E_ε is stable. It is also possible to give a more detailed analysis, for fixed forms of h and g, on the parameter ranges for such co-existence of stable structures. It may reveal a more comprehensive understanding of the global dynamics of this model.

5. Numerical simulations and biological interpretations. In this section, we provide results from numerical simulations of model (2.3) that demonstrate and support our theoretical results on the existence of stable periodic solutions of relaxation oscillation type. Unless otherwise stated, we choose

$$g(N) = N\Big(1 - \frac{N}{N^*}\Big) \quad \text{and} \quad h(S, N) = \frac{\beta S}{K + S}.$$

It can be verified that $g(N)$ and $h(S, N)$ satisfy assumptions **(A1)**, **(A2)** and **(A3)**.

5.1. Existence of relaxation oscillations.

Case 1. Existence of relaxation oscillation when E_ε is unstable.

Choose $d = 0.2$, $p = 0.01$, $\alpha = 0.048$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$, $N^* = 400$ and $\varepsilon = 10^{-4}$. The endemic equilibrium $E_\varepsilon = (49.9, 0.09555, 52.84889)$ is unstable. In Figure 3, we show that a trajectory starting from $(35, 0.09555, 67)$ approaches a stable relaxation oscillation cycle with inter-epidemic period 5.6×10^4.

Case 2. Existence of relaxation oscillation when E_ε is stable.

Choose $d = 0.2$, $p = 0.01$, $\alpha = 0.049$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$, $\varepsilon = 10^{-4}$, and $N^* = 380$. In Figure 4, we show that a trajectory starting from $(197, 1.47, 204.4)$ approaches a stable relaxation oscillation cycle. We modified the function $h(S, N)$ in a small neighbourhood of E_ε so that becomes locally asymptotically stable. Such modification does not change the relaxation oscillation cycle since it is far away from E_ε. A trajectory starting from $(150, 1, 160)$ is shown in Figure 4 to approach the stable equilibrium E_ε. We note that there should be a second periodic orbit that is unstable (not shown in Figure 4).

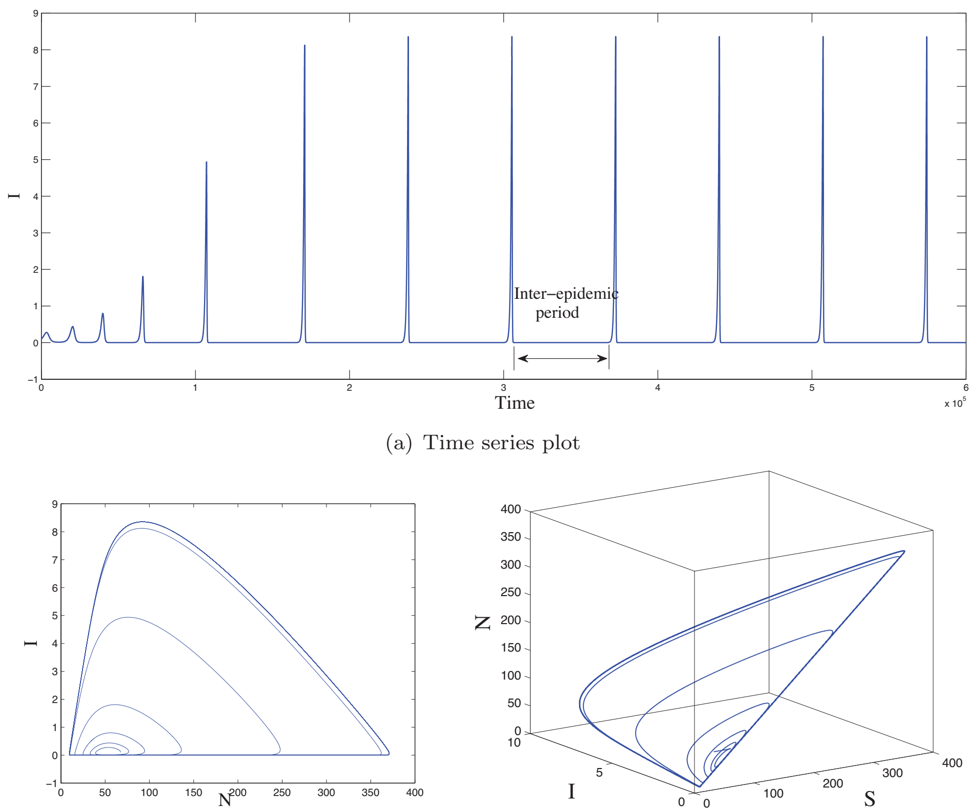
5.2. Dependence of inter-epidemic period (IEP) on physical parameters.

1. Dependence of IEP on the intrinsic growth rate ε.

We demonstrate using numerical evidence that the inter-epidemic period is of order $1/\varepsilon$. We choose $d = 0.2$, $p = 0.01$, $\alpha = 0.048$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$ and $N^* = 400$, and we vary the values of ε in the interval $[10^{-5}, 10^{-4}]$. For the simulations, we assume that the disease is in the inter-epidemic period if the number of the infected individuals is less than 10^{-7}. The plot of IEP against the values of ε and $1/\varepsilon$ are shown in Figure 5.

2. Dependence of IEP on parameters α and β.

In Figure 6, we show that the IEP decreases as the transmission coefficient β in creases, and the IEP increases as the rate α of disease-caused death increases. For the simulations, we choose $d = 0.2$, $p = 0.01$, $\gamma = 0.75$, $K = 0.1$, $N^* = 400$ and $\varepsilon = 10^{-4}$, and vary values of β when $\alpha = 0.048$ in Figure 6 (a), or vary the values of α when $\beta = 1$ in Figure 6 (b).



(a) Time series plot

(b) Projection in the (N, I) plane

(c) Plot in the 3D phase space

Fig. 3. *An orbit converging to a stable relaxation oscillation cycle when E_ε is unstable.*

Appendix I: Technical proofs.

Proof of Theorem 2.4. To show (i), note that the linearization of system (2.3) at $(0, 0, 0)$ is

$$J(0,0,0) = \begin{pmatrix} -(d+p) & 0 & d + \varepsilon g_N(0) \\ 0 & -a & 0 \\ 0 & -\alpha & \varepsilon g_N(0) \end{pmatrix},$$

whose eigenvalues are $-(d + p) < 0,\ -a < 0,\ \varepsilon g_N(0) > 0$, where $\varepsilon g_N(0) > 0$ follows from **(A1).** Similarly, the linearization at $(S^*, 0, N^*)$ is

$$J(S^*,0,N^*) = \begin{pmatrix} -(d+p) & -h & d + \varepsilon g_N(N^*) \\ 0 & h-a & 0 \\ 0 & -\alpha & \varepsilon g_N(N^*) \end{pmatrix},$$

with eigenvalues $-(d+p) < 0,\ h(S^*, N^*)-a > 0$, and $\varepsilon g_N(N^*) < 0$, where $h(S^*, N^*)-a > 0$ follows from **(A3)** and $\varepsilon g_N(N^*) < 0$ follows from **(A1)**.

The linearization at E_ε is

$$J = J(S_\varepsilon, I_\varepsilon, N_\varepsilon) = \begin{pmatrix} -(d+p+h_S I_\varepsilon) & -a & d - h_N I_\varepsilon + \varepsilon g_N \\ h_S I_\varepsilon & 0 & h_N I_\varepsilon \\ 0 & -\alpha & \varepsilon g_N \end{pmatrix},$$

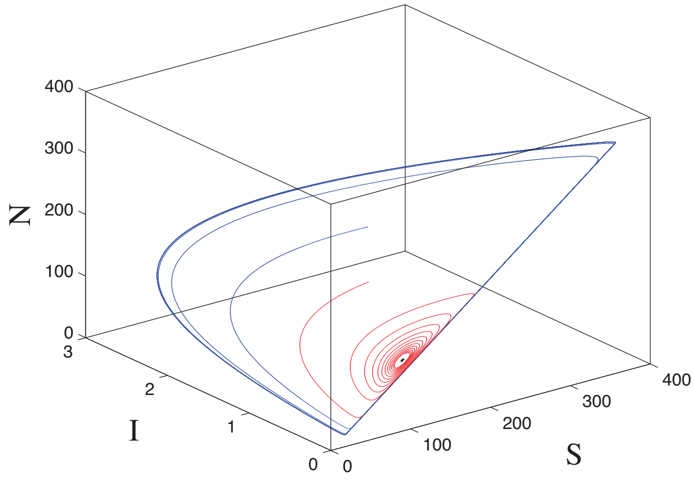
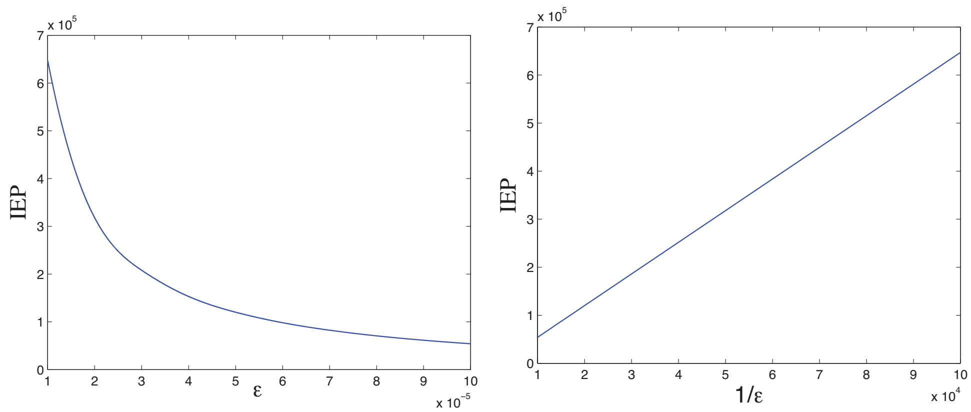


Fig. 4. *Numerical simulations show the existence of a stable periodic solution when the endemic equilibrium is stable. An oscillatory orbit with a large amplitude is shown to converge to a stable relaxation oscillation cycle, and an orbit with a smaller amplitude converges to the stable endemic equilibrium* E_ε.



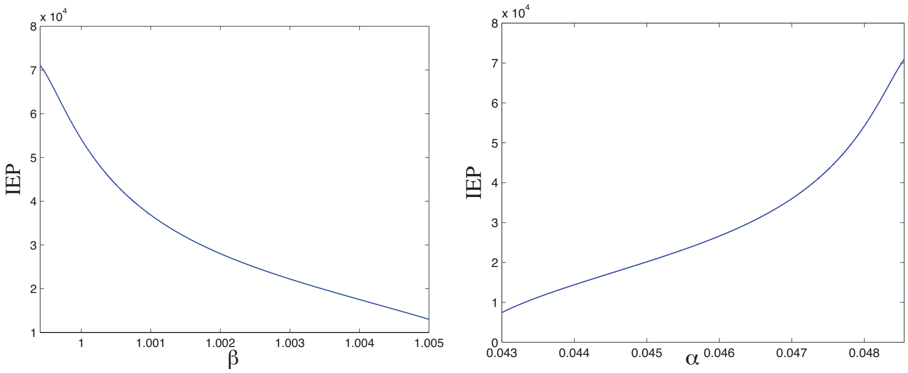
(a) Dependence of IEP on ε

(b) Dependence of IEP on $1/\varepsilon$

Fig. 5. *The inter-epidemic period (IEP) increases as the intrinsic growth rate* ε *decreases in (a), and the IEP is in proportion to* $1/\varepsilon$ *in (b).*

whose characteristic polynomial is given by

$$\begin{aligned} P_\varepsilon(\lambda) = {} & \lambda^3 + (d + p + \frac{\varepsilon h_S g}{\alpha} - \varepsilon g_N)\lambda^2 - \varepsilon((d+p)g_N - \frac{a h_S g}{\alpha} - h_N g - h_S g_N I_\varepsilon)\lambda \\ & + \alpha d h_S I_\varepsilon + \alpha(d+p)h_N I_\varepsilon - \varepsilon(a+\alpha)h_S g_N I_\varepsilon. \end{aligned}$$



(a) Dependence of IEP on β (b) Dependence of IEP on α

Fig. 6. *Dependence of inter-epidemic period (IEP) on the transmission coefficient β and on the rate of disease-caused death α.*

Hence,

(5.1)

$$
\begin{aligned}
\operatorname{tr}(J) =& -(d+p) - \frac{\varepsilon h_S g}{\alpha} + \varepsilon g_N < 0, \\
\det(J) =& -\alpha d h_S I_\varepsilon - \alpha(d+p) h_N I_\varepsilon + \varepsilon(a+\alpha) h_S g_N I_\varepsilon \\
=& -\varepsilon(d+p)\Big[\frac{d}{d+p} h_S(S_0,N_0) + h_N(S_0,N_0)\Big] g(N_0) + O(\varepsilon^2) < 0, \\
\operatorname{tr}(J) a_2 - \det(A) =& -\varepsilon\Big((d+p) g_N - \frac{a h_S g}{\alpha} - h_N g - h_S g_N I_\varepsilon\Big) \\
=& -\varepsilon(d+p)\Delta_0 + O(\varepsilon^2),
\end{aligned}
$$

where a_2 is the the coefficient of λ in $P_\varepsilon(\lambda)$, namely, the sum of all 2×2 principal minors of J.

When $\varepsilon = 0$, $P_0(\lambda) = \lambda^3 + (d+p)\lambda^2$. It has a negative root, $-(d+p)$. Therefore, when $\varepsilon > 0$ small, $P_\varepsilon(\lambda)$ has a negative root. We show that the remaining roots of $P_\varepsilon(\lambda)$ are always complex conjugates. To see this, write $P_\varepsilon(\lambda)$ as

$$P_\varepsilon(\lambda) = \lambda^3 - a_1\lambda^2 + a_2\lambda - a_3,$$

where $a_1 = \operatorname{tr}(A) < 0$, $a_3 = \det(A) < 0$, and a_2 is as above. The larger of the two critical points of $P_\varepsilon(\lambda)$ is

$$\lambda_1 = \frac{1}{3}\Big(a_1 + \sqrt{a_1^2 - 3a_2}\Big).$$

Straightforward calculation leads to

$$P_\varepsilon(\lambda_1) = \frac{1}{27}\Big[-2a_1^3 + 9a_1a_2 - 27a_3 - 2(a_1^2 - 3a_2)\sqrt{a_1^2 - 3a_2}\Big].$$

It can be verified that $P_\varepsilon(\lambda_1) > 0$ if and only if

$$27a_3^2 + 4a_1^3a_3 + 4a_2^3 - 18a_1a_2a_3 - a_1^2a_2^2 > 0. \tag{5.2}$$

When $\varepsilon = 0$, we have $\lambda_1 = 0$, which is a double root of $P_0(\lambda)$. Therefore, $P_\varepsilon(\lambda_1) = 0$ when $\varepsilon = 0$, and thus the sign of the expression in (5.2) is determined by the ε order terms, which is given by

$$4(d+p)^4 g(N_0)[\frac{d}{d+p} h_S(S_0, N_0) + h_N(S_0, N_0)]\varepsilon > 0,$$

by assumption **(A3)** and continuity. Hence $P_\varepsilon(\lambda_1) > 0$. This implies that $P_\varepsilon(\lambda)$ has only one real root. The signs of the real parts of the complex roots can be determined by the Routh-Hurwitz conditions, which state that all roots of $P_\varepsilon(\lambda)$ have negative real parts if and only if the following three conditions hold: $a_1 = \mathrm{tr}(A) < 0$, $a_3 = \det(A) < 0$ and $a_1 a_2 - a_3 < 0$. From relations in (5.1), we see that, for $\varepsilon > 0$ small, if $\Delta_0 > 0$, then all three eigenvalues have negative real parts, and if $\Delta_0 < 0$, then at least one eigenvalue has positive real parts. This establishes (ii).

Proof of Theorem 3.1. To show (i), we note that, for solution $(S(t), I(t), N(t))$ with initial condition $(S, 0, N) \in \mathcal{D}$, $I(t) \equiv 0$, $N(t) \equiv N$ and $S(t) \to dN/(d+p)$ as $t \to \infty$. Thus, $(S(t), I(t), N(t)) \to (dN/(d+p), 0, N)$ as $t \to \infty$. Now let $(S(t), I(t), N(t))$ be the solution with the initial condition $(S(0), I(0), N(0)) \in \mathcal{D}$ with $I(0) > 0$. From system (3.1), we have $I(t) > 0$ for all $t \geq 0$ and hence $N(t)$ monotonically decreases. Therefore $N(t) \to \bar{N}$ as $t \to \infty$ for some $\bar{N}$ dependent on the initial condition. We claim that $\bar{N} \leq N_0$.

First of all, note that the equilibrium $(S_0, 0, N_0)$ has two zero eigenvalues and one negative eigenvalue $-(d+p)$. Locally, there is a two dimensional center manifold $W^c(S_0, 0, N_0)$, and $W^c(S_0, 0, N_0)$ can be taken to consist of heteroclinic orbits from $(S, 0, N) \in \mathcal{Z}_0 \cap W^c(S_0, 0, N_0)$ with $S_0 < S < S_0 + \delta_0$, for some $\delta_0 > 0$ small, to a point $(\bar{S}, 0, \bar{N}) \in \mathcal{Z}_0 \cap W^c(S_0, 0, N_0)$ and there is a neighbourhood $\mathcal{N}(\delta_0)$ of $(S_0, 0, N_0)$ in $\mathcal{D}$ such that any solution entering in $\mathcal{N}(\delta_0)$ approaches a point $(S, 0, N) \in \mathcal{Z}_0 \cap W^c(S_0, 0, N_0)$ with $N < N_0$. Note that $\{I = 0\}$ is invariant and, on $\{I = 0\}$, any solution $(S(t), 0, N(t))$ is given by $N(t) = N(0)$ and $S(t) \to dN(0)/(d+p)$ with the rate $\exp\{-(d+p)t\}$. By continuity, for $\delta_1 > 0$ smaller than δ_0, any solution $(S(t), I(t), N(t))$ with $0 < I(0) \leq \delta_1$ and $N_0 - \delta_1 \leq N(0) \leq N_0 + \delta_1$ will follow the solution with the initial condition $(S(0), 0, N(0))$ to the neighbourhood $N(\delta_0)$ and hence approach a point $(S, 0, N) \in \mathcal{Z}_0 \cap W^c(S_0, 0, N_0)$ with $N < N_0$.

To establish the claim, we suppose on the contrary that $\bar{N} > N_0$. Then, $\bar{N} > N_0 + \delta_1$ from the above argument. For $\delta > 0$ small there exists $t_0 > 0$ such that $\bar{N} \leq N(t) < \bar{N} + \delta$ for $t \geq t_0$. Since $h(d\bar{N}/(d+p), \bar{N}) - a > 0$, there exist $\rho > 0$ small and $T > 0$ such that any solution that crosses the square $\{(S, \rho, N) : |S - d\bar{N}/(d+p)| < \rho, |N - \bar{N}| < \rho\}$ from below will stay above $\{I = \rho\}$ for a length of time greater than T. Now choose $\delta > 0$ such that $\alpha\rho T > \delta$. It is clear that there is an infinite sequence $t_n \to \infty$ such that $I(t_n) \to 0$. Thus, for some $t_n > t_0$, the forward orbit will cross the above square at some time t^*. We have $I(t) \geq \rho$ for $t \in [t^*, t^* + T]$, and hence

$$N(t^* + T) = N(t^*) - \alpha \int_{t^*}^{t^*+T} I(s)\, ds \leq \bar{N} + \delta - \alpha\rho T < \bar{N}.$$

This contradicts to that $\bar{N} \leq N(t)$ for $t \geq 0$ which establishes the claim.

We now show that $(S(t), I(t), N(t)) \to (d\bar{N}/(d+p), 0, \bar{N})$ as $t \to \infty$. From the existence of the sequence $t_n \to \infty$ such that $I(t_n) \to 0$, we know $(I(t_n), N(t_n)) \to (0, \bar{N})$ as $n \to \infty$. By continuity, for n large, the solution will follow the solution through the point $(S(t_n), 0, N(t_n))$ to a neighbourhood of the point $(d\bar{N}/(d+p), 0, \bar{N})$.

Since the set $\mathcal{Z}_0$ is normally stable near this point, the solution will approach some point on $\mathcal{Z}_0$ and it must be $(d\bar{N}/(d+p), 0, \bar{N})$ because $N(t) \to \bar{N}$ as $t \to \infty$.

To establish the statement (ii), we note that the unstable manifold of a point $(S, 0, N) \in \mathcal{Z}_0$ with $N > N_0$ is 1-dimensional and an orbit representing the unstable manifold with positive I-component, and hence, it converges to a point $(\bar{S}, 0, \bar{N})$ with $\bar{N} < N_0$ from statement (i). We now justify the properties of the function H in the statement. For $N_1 > N_2 > N_0$ with N_1 and N_2 close to N_0, it is clear that $H(N_1) < H(N_2)$ since the corresponding heteroclinic orbits lie on the local center manifold $W^c(S_0, 0, N_0)$ which is a disk-like. It is also clear that H is a continuous and one-to-one function. Therefore, the monotone decreasing property of H holds globally and $H(N) \to N_\infty$ as $N \to \infty$ exists. It remains to show that $N_\infty > 0$. It can be verified directly that the eigenvectors associated to the stable eigenvalues $\lambda_2 = -(d+p)$ and $\lambda_3 = h(0,0) - a = -a$ of $(0,0,0)$ are, respectively,

$$v_2 = (1,0,0) \quad \text{and} \quad v_3 = \left(\frac{\alpha d}{a(d+p-a)}, 1, \frac{\alpha}{a} \right).$$

Since $\alpha < a$, the vectors v_2 and v_3 at $(0,0,0)$ are pointing towards the exterior of the feasible region $\mathcal{D}$. Therefore, the *local* two dimensional stable manifold $W^s_{loc}(0,0,0)$ except $(0,0,0)$ stays outside of $\mathcal{D}$. By continuity, for some $\delta > 0$ small and for any equilibrium $(dN/(d+p), 0, N)$ with $N < \delta$, an orbit starting on the local stable manifold $W^s_{loc}(dN/(d+p), 0, N)$ except the equilibrium $(dN/(d+p), 0, N)$ will exit the region $\mathcal{D}$ backward, and will stay outside $\mathcal{D}$ in backward time upon the exit due to the positive invariance of $\mathcal{D}$. Hence, $H(N) \geq \delta$ for any $N > N_0$, which implies that $N_\infty \geq \delta > 0$.

Appendix II: Persistence of $M(\mathcal{Z}_0)$ for $\varepsilon > 0$ small. To establish the persistence of $M(\mathcal{Z}_0)$ claimed in Case 2 of Section 3.2, we make a change of variables. This change of variables is continuous but not everywhere smooth. Indeed, it is smooth everywhere except on $\{I = 0\}$. Nevertheless, the property that $\{I = 0\}$ is invariant for all $\varepsilon \geq 0$ makes the change of variables work.

Let m be a positive integer so that $a < m(d+p)$. We may assume that $m \geq 2$. Make the change of state variables: $S = x$, $I = y^m$ and $N = N$, for $y > 0$. In terms of the new variables (x, y, N), the equation for I in (2.3) becomes

$$my^{m-1}y' = (h(x,N) - a)y^m, \quad \text{or equivalently,} \quad y' = \frac{1}{m}(h(x,N) - a)y.$$

The model (2.3) becomes

$$\begin{aligned} x' &= dN + \varepsilon g(N) - h(x,N)y^m - dx - px, \\ y' &= \frac{1}{m}(h(x,N) - a)y, \\ N' &= \varepsilon g(N) - \alpha y^m. \end{aligned} \tag{5.1}$$

We note that this change of state variables is smooth for $y > 0$ and can be continued to $y = 0$. The new system (5.1) has exactly the same reduced dynamics on $\{y = 0\}$ as that of (2.3) on $\{I = 0\}$. We emphasize that, the naturally given property that $\{I = 0\}$ is invariant under (2.3) for $\varepsilon \geq 0$ is crucial for such a change of variables. The biological implication are commented and illustrated by examples in Section 5.

Recall that $m \geq 2$. The set $\mathcal{Z}_0$ corresponds, for (5.1), to

$$\mathcal{S}_0 = \left\{ y = 0, \ x = \frac{d}{d+p} N \right\}.$$

Let $M(\mathcal{S}_0)$ denote the corresponding invariant manifold $M(\mathcal{Z}_0)$.

The linearization at each equilibrium on $\mathcal{S}_0$ is

$$\begin{pmatrix} -(d+p) & 0 & d \\ 0 & \frac{1}{m}(h-a) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = -(d+p)$, and $\lambda_3 = (h(dN/(d+p), N) - a)/m$. The eigenvector v_1 associated with λ_1 is tangent to $\mathcal{S}_0$ and that v_2 associated with λ_2 is $(1, 0, 0)$, and v_1 and v_2 span the plane $\{y = 0\}$. The eigenvector v_3 associated with λ_3 is transversal to the plane $\{y = 0\}$. While the eigenvalue λ_2 stays negative, the eigenvalue λ_3 changes sign across $(S_0, 0, N_0) \in \mathcal{S}_0$. Nevertheless, $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_2$. The center manifold theory in [5, 6] implies that $M(\mathcal{S}_0)$ persists under system (5.1) for $\varepsilon > 0$ small.

REFERENCES

[1] R. M. Anderson and R. M. May, Population biology of infectious diseases: Part I, *Nature* **280** (1979), 361-367.

[2] R. M. Anderson and R. M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford Science Publications, 1992.

[3] N. T. J. Bailey, The Mathematical Theory of Infectious Diseases and its Application. London: Griffin, 1975.

[4] M. S. Bartlett, Deterministic and stochastic models for recurrent epidemics, in Berkeley Symp. on Math. Statist. and Prob., Proc. Third Berkeley Symp. on Math. Statist. and Prob., Vol. 4, Univ. of Calif. Press, 1956, pp 81-109.

[5] S.-N. Chow, W. Liu, and Y. Yi, Center manifold theory for smooth invariant manifolds. *Trans. Amer. Math. Soc.* **352** (2000), 5179-5211.

[6] S.-N. Chow, W. Liu, and Y. Yi, Center manifold theory for smooth invariant manifolds. *J. Differential Equations* **168** (2000), 355-385.

[7] F. Diener and M. Diener, Maximal delay. *Dynamic bifurcations (Luminy, 1990)*, 71-86. Lect. Notes in Math., **1493**, Springer, Berlin, 1991.

[8] J. D. Enderle, A stochastic communicable disease model with age-specific states and applications to measles, Ph.D. thesis, Rensselaer Polytechnic Institute, 1980.

[9] N. Fenichel, Persistence and smoothness of invariant manifolds for flows. *Indiana Uni. Math. J.* **21** (1971), 193-226.

[10] N. Fenichel, Geometric singular perturbation theory for ordinary differential equations. *J. Differential Equations* **31** (1979), 53-98.

[11] J. Graef, M. Li, and L. Wang, A study on the effects of disease caused death in a simple epidemic model, in Dynamical Systems and Differential Equations, W. Chen and S Hu, (Eds.), Southwest Missouri State University Press, 288-300, 1998.

[12] H. W. Hethcote, Asymptotic behavior in a deterministic epidemic model, *Bull. Math. Biol.* **35** (1973), 607-614.

[13] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* **42** (2000) 599-653.

[14] H.W. Hethcote, H.W. Stech and P. van den Driessche, Nonlinear oscillations in epidemic models, *SIAM J. Appl. Math.* **40** (1981), 1-9.

[15] M. Hirsch, C. Pugh, and M. Shub, Invariant Manifolds. Lect. Notes in Math., **583**, Springer-Verlag, New York, 1976.

[16] W. M. Liu, S. A. Levin, and Y. Iwasa, Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models, *J, Math. Biol.* **23** (1986), 187-204.

[17] W. Liu, Exchange Lemmas for singularly perturbation problems with certain turning points. *J. Differential Equations* **167** (2000), 134-180.

[18] W. Liu, Geometric singular perturbations for multiple turning points: invariant manifolds and exchange lemmas. *J. Dynam. Differential Equations* **18** (2006), 667-691.

[19] W. Liu, D. Xiao, and Y. Yi, Relaxation oscillations in a class of predator-prey systems. *J. Differential Equations* **188** (2003), 306-331.

[20] W. P. London and J. A. Yorke, Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates, *Am. J. Epidemiol.* **98** (1973), 453-468.

[21] R. Mãné, Persistent manifolds are normally hyperbolic. *Trans. Amer. Math. Soc.* **246** (1977), 261-283.
[22] J. Mena-Lorca and H. W. Hetheote, Dynamic models of infectious diseases as regulators of population sizes, *J. Math. Biol.* **30** (1992), 693-716.
[23] E. F. Mishchenko, Yu. S. Koleso, A. Yu. Kolesov, and N. Kh. Rozo, Asymptotic Methods in Singularly Perturbed Systems, Monographs in Contemporary Mathematics. Consultants Bureau, New York, 1994.
[24] A. Neishtadt, On stability loss delay for dynamical bifurcations, I. *Differential Equations* **23** (1987), 1385-1390.
[25] A. Neishtadt, On stability loss delay for dynamical bifurcations, II. *Differential Equations* **24** (1988), 171-176.
[26] United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects: The 2012 Revision, New York, 2013.