Commun Nonlinear Sci Numer Simulat 37 (2016) 163–192

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

Dynamical analysis and simulation of a 2-dimensional disease model with convex incidence

Pei Yu*, Wenjing Zhang, Lindi M. Wahl

Department of Applied Mathematics, Western University, London, Ontario, N6A 5B7, Canada

ARTICLE INFO

Article history:
Received 4 February 2015
Revised 20 November 2015
Accepted 31 December 2015
Available online 26 January 2016

Keywords:
Recurrent infection
HIV viral blips
Recurrent autoimmune disease
Dynamical system
Stability and bifurcation
Hopf and generalized Hopf bifurcations
Bogdanov–Takens (BT) bifurcation
Homoclinic orbit

ABSTRACT

In this paper, a previously developed 2-dimensional disease model is studied, which can be used for both epidemiologic modeling and in-host disease modeling. The main attention of this paper is focused on various dynamical behaviors of the system, including Hopf and generalized Hopf bifurcations which yield bistability and tristability, Bogdanov–Takens bifurcation, and homoclinic bifurcation. It is shown that the Bogdanov–Takens bifurcation and homoclinic bifurcation provide a new mechanism for generating disease recurrence, that is, cycles of remission and relapse such as the viral blips observed in HIV infection.

1. Introduction

Mathematical models in epidemiology and in-host disease share common features, dividing a population of individuals (epidemiology) or cells (in-host) into discrete classes relevant to the disease dynamics, and typically describing their dynamics with a system of ordinary differential equations (ODEs). A key feature of such systems is the incidence function, which defines the spread of the infection to susceptibles.

For example, in classical epidemiological models, the incidence rate is often assumed to take the form $\frac{\beta SI}{N}$, where $S(t)$ is the number of susceptible individuals, $I(t)$ is the number of infectives and β is a constant, the transmission rate [1]. When N, the population size, is constant, this incidence function is also simply written as βSI. Similarly, for in-host models, the rate at which uninfected cells become infected is often described as βxy, where $x(t)$ reflects the uninfected cell density and $y(t)$ denotes the density of infected cells [2].

Bilinear incidence functions of this form have been used extensively and are well-studied in the mathematical literature. As described in greater detail elsewhere [3], a number of possibilities for non-linear incidence functions have also been studied in some detail, including the general form $\beta I^p S^q$, where p and q are positive constants [4–9], and several more complex forms [4,10].

Because of physical limitations on the number of new infections possible as disease prevalence increases, a common feature of many incidence functions is their concavity with respect to the number of infectives. In particular, the incidence

* Corresponding author. Tel.: +1 519661 2111x88783; fax: +1 519 661 3523.
E-mail address: pyu@uwo.ca (P. Yu).

http://dx.doi.org/10.1016/j.cnsns.2015.12.022
1007-5704/

rate $f(S, I, N)$ typically satisfies the condition

$$\frac{\partial^2 f(S,I,N)}{\partial I^2} \leq 0.$$

Taking advantage of this common feature, Korobeinikov and Maini [3] derived elegant results for all concave incidence functions, showing the global asymptotic stability of the disease-free equilibrium when the basic reproduction number $R_0 \leq 1$, and global asymptotic stability of the endemic equilibrium when $R_0 > 1$, for the standard SIRS model [1] with a constant population size. In other words, the concavity of the incidence rate guarantees the uniqueness and stability of the endemic equilibrium in these models, and these powerful results apply to any concave incidence function.

In contrast, we have recently analyzed a number of ODE models with convex incidence functions. If incidence is convex, or "synergistic", the rate at which new infections occur can increase supralinearly with disease prevalence. This situation can arise in a number of realistic scenarios. For example, in in-host models of the human immunodeficiency virus (HIV), increasing the extent of the infection involves greater damage to the immune system, and can thus increase the incidence rate [11]. Similarly, in autoimmune disease, increases to the autoimmune response against self tissue can cause a positive feedback loop which will further increase the incidence rate [12]. While these two examples both arise in in-host disease modeling, catastrophic outbreak or pandemic conditions could also result in convex epidemiological incidence. In particular, an outbreak that is severe enough to compromise health care infrastructure (increasing hospital crowding and front-line worker exposure rates, for example) could involve a supralinear increase in incidence rates with disease prevalence.

In this contribution, we analyze in detail the possible dynamical behaviors of a simple 2-dimensional disease model with a convex, or synergistic, incidence function. The system we analyze is a standard non-dimensionalized SI model which arises in both epidemiology and in-host modeling: it assumes a birth rate into the susceptible population, death rates for both populations, and an incidence rate between the two. The incidence function we study has an analytical form which has arisen in a number of models previously analyzed [11–15]. Its behavior is such that when the infective population I is small, incidence increases linearly with I; when I is large, incidence also increases linearly, but with a steeper slope. A convex region of the function connects these limiting behaviors.

In marked contrast to the powerful general conclusions obtained for concave incidence functions [3], we find that a wide range of dynamical behaviors are possible when incidence is synergistic. In particular, as previously analyzed in related higher-dimensional models [13–15], we note the appearance of *recurrent infection*, that is, cycles consisting of long periods close to the disease free equilibrium, punctuated by brief bursts of disease. This pattern of recurrence occurs in many diseases, including the intriguing pattern of "viral blips" in HIV, as well as the recurrent episodes characteristic of autoimmune diseases, such as multiple sclerosis [16], multifocal osteomyelitis [17,18], lupus [19], eczema [20], and psoriasis [21]. In this contribution, we explore several mechanisms which can underly these physiologically relevant patterns of infection, finding that when the incidence function is convex, bistable equilibrium solutions, Hopf and generalized Hopf bifurcations and, in particular, homoclinic bifurcations may all contribute to disease recurrence.

In related work, Ruan and Wang [22] analyzed a reduced SI model, which has a zero disease-free equilibrium and a positive endemic equilibrium. In this model, $R_0 = 0$, although it can be shown that the disease can still persist. In [22], the authors also considered Hopf bifurcation, Bogdanov–Takens bifurcation and homoclinic orbits. The structure of the model in [22] is mathematically appealing, such that the authors could transform the model to a Liénard system and then prove the uniqueness of the limit cycle from Hopf bifurcation. Moreover, their analysis of the homoclinic orbit takes the standard form (e.g., see [23]). In contrast, the model we study in this contribution has been derived from physical considerations and has known realistic parameter ranges, however this model cannot be transformed to a Liénard system, and the analysis of homoclinic orbits does not follow the standard form.

The rest of the paper is organized as follows. In next section, we give a detailed dynamical analysis of the simple 2-dimensional disease model. In Section 3, Hopf and generalized Hopf bifurcations are studied in detail, which may be the main features underlying complex dynamical behaviors. Then, in Section 4, Bogdanov–Takens bifurcation and homclinic bifurcation are investigated, giving rise to another scenario/mechanism for generating blips. Finally, conclusions and discussion are given in Section 5.

2. Dynamics of the 2-D disease model

Consider the 2-dimensional system:

$$\begin{aligned}\frac{dX}{d\tau} &= 1 - DX - \left(B + \frac{AY}{Y+C}\right)XY,\\ \frac{dY}{d\tau} &= \left(B + \frac{AY}{Y+C}\right)XY - Y,\end{aligned} \tag{2.1}$$

where all parameters, A, B, C and D take positive real values. This system was originally derived as an in-host model of HIV dynamics [11], but has been reduced in dimension and non-dimensionalized using quasi-steady state assumptions as described in [13,14]. Although arising from in-host disease modeling, the reduced 2-D system is also equivalent to the SIRS model studied in [3], taking the recovery rate, α of [3], to be zero. At appropriate parameter values, system (1) thus represents either an in-host infection (susceptible and infected cells), or an SIR epidemiological model (susceptible and infected

individuals). The key difference between system (1) and the class of models studied in [3] is that the incidence function in system (1), $XY(B+AY/(Y+C))$, is convex. Our goal is to understand the dynamical behaviors made possible by this convexity.

In [13,14], this 2-dimensional model is not analyzed in detail. For example, well-posedness of solutions of this system and the global stability of the disease-free equilibrium were not considered; and a trapping region was proved only for fixed parameter values when $B > D$. In the following subsections, we will provide general proofs for the above mentioned problems with no additional restriction on the positive parameter values.

2.1. Well-posedness and trapping region of solutions

We first prove the positiveness of solutions of system (2.1), and further show that the solutions are bounded, attracted to a region. We have the following result.

Theorem 2.1. *Solutions of system (2.1) are non-negative provided the initial conditions are non-negative, and further these solutions are eventually attracted to a bounded region G, in the shape of a right triangle, bounded by the X-axis, the Y-axis and the line L:* $X+Y=\max\{1,\frac{1}{D}\}+\varepsilon$ $(0<\varepsilon\ll 1)$.

Proof. Using the first equation of system (2.1), with the formulae of variation of parameters, we obtain the solution for $X(\tau)$ as

$$X(\tau)=X(0)e^{-\int_0^\tau[D+(B+\frac{AY(s)}{Y(s)+C})Y(s)]ds}+\int_0^\tau e^{-\int_s^\tau[D+(B+\frac{AY(u)}{Y(u)+C})Y(u)]du}ds, \tag{2.2}$$

which clearly indicates that $X(\tau)>0$ for $\tau>0$ if $X(0)\geq 0$.

For the solution of $Y(\tau)$, we note that $Y=0$ (i.e., the X-axis) is an invariant solution trajectory of the system. Thus, any solution starting from an initial point with $X(0)>0, Y(0)>0$ must be kept in the first quadrant of the X–Y plane due to the uniqueness of solutions of the system.

Next, we show that all solutions are eventually attracted into the bounded region G defined in the theorem. To achieve this, first note that system (2.1) has two equilibrium solutions obtained by setting $\frac{dX}{d\tau}=\frac{dY}{d\tau}=0$: one is the disease-free equilibrium, $\mathrm{E}_0=(\frac{1}{D},0)$, which is a boundary equilibrium, and other is the endemic equilibrium, $\mathrm{E}_1=(X_1,Y_1)$, which is an interior equilibrium, where

$$Y_1=1-DX_1, \tag{2.3}$$

and X_1 is determined from the quadratic polynomial equation:

$$\begin{aligned}Q(X)&=D(A+B)X^2-(A+B+D+BC)X+C+1\\&=\frac{1}{D}[(A+B)(1-DX)^2-(A+B-D-BC)(1-DX)-C(B-D)]\\&=0.\end{aligned} \tag{2.4}$$

The existence of E_1 depends on the values of the parameters.

First, consider the X-axis. Note that $\mathrm{E}_0=(\frac{1}{D},0)$ is located on the X-axis, with two eigenvalues, $\xi_1=-D$ and $\xi_2=\frac{B}{D}-1$, and their corresponding eigenvectors are $v_1=(1,0)$ and $v_2=(1,\frac{D}{B}(1-D)-1)$, respectively. Moreover, v_1 is in the direction of the X-axis, which can be shown to be a solution trajectory of the system. With a negative eigenvalue, the trajectory along the X-axis converges to the point E_0. Thus, the X-axis is a separator (invariant manifold) of the dynamical system, and so no trajectory can cross it due to the uniqueness of solutions. Hence, every trajectory entering the region G cannot escape from this boundary – the X-axis.

On the Y-axis, it is easy to obtain $\frac{\mathrm{d}X}{\mathrm{d}\tau}=1$ and $\frac{\mathrm{d}Y}{\mathrm{d}\tau}=-Y$, showing that all trajectories cross the Y-axis from left to right.

Finally, we want to prove that all trajectories which cross the line L actually move into the region G. To show this, note that the direction of the line L is $(1,-1)$, and so the normal direction of the line in its gradient direction is $(1,1)$. Define

$$S(Y)=(1,1)\bullet\left(\frac{\mathrm{d}X}{\mathrm{d}\tau},\frac{\mathrm{d}Y}{\mathrm{d}\tau}\right)=\frac{dX}{d\tau}+\frac{dY}{d\tau},$$

where the dot denotes inner product (or dot product). We need to show $S(Y)<0$ for $0<Y<\max\{1,\frac{1}{D}\}+\varepsilon$. Simplifying $S(Y)$ yields

$$\begin{aligned}S(Y)&=\frac{\mathrm{d}X}{\mathrm{d}\tau}+\frac{\mathrm{d}Y}{\mathrm{d}\tau}=\left[1-DX-\left(B+\frac{AY}{Y+C}\right)XY\right]+\left(B+\frac{AY}{Y+C}\right)XY-Y\\&=1-DX-Y=1-DX-\left[\max\left\{1,\frac{1}{D}\right\}+\varepsilon-X\right]\\&=-\varepsilon+1-\max\left\{1,\frac{1}{D}\right\}+(1-D)X\end{aligned}$$

$$= \left.\begin{cases} -\varepsilon + \left(1 - \frac{1}{D}\right) + (1-D)X, & D < 1 \\ -\varepsilon, & D = 1 \\ -\varepsilon - (D-1)X, & D > 1 \end{cases}\right\} \text{ for } 0 < X < \max\left\{1, \frac{1}{D}\right\} + \varepsilon$$
$$\leq -\varepsilon \min\{1, D\} < 0.$$

Note that one may set $\varepsilon = 0$ for $D \neq 1$. Hence, for all positive parameter values, there always exists a trapping region G, bounded by the X-axis, the Y-axis, and the line L, and all trajectories move into G when crossing the Y-axis and the line L, and once they enter G, they cannot escape from the X-axis. □

In the following, we consider the dynamical behavior of system (2.1) according to the conditions: $B < D$, $B > D$ and $B = D$. Note that system (2.1) is actually equivalent to the model studied in [3] when Y is small so that $Y^2 \approx 0$. In this case, system (2.1) has bilinear incidence, which is concave, and the local $\mathrm{R}_0 = \frac{B}{D}$. Thus, we expect that the disease-free equilibrium, E_0, when Y is in fact small, is locally stable when $B < D$, and becomes a saddle point when $B > D$.

2.2. Dynamical behavior of (2.1) when B < D

First, we study the dynamical behavior of system (2.1) when $B < D$. In particular, we want to investigate the global stability of the disease-free equilibrium E_0. For convenience, define

$$\mathrm{H}_1 \triangleq A + B - D - BC - 2\sqrt{C(A+B)(D-B)}, \quad (B < D). \tag{2.5}$$

We have the following result.

Theorem 2.2. *When $B < D$, the disease-free equilibrium E_0 of system (2.1) is globally asymptotically stable if $\mathrm{H}_1 < 0$, under which the endemic equilibrium E_1 does not exist. Otherwise, there exist two disease equilibria – one of them is a saddle point while the other may be a stable (or an unstable) node or focus – and no definite conclusion can be made regarding the global stability of E_0.*

Proof. First, it is easy to see that when $B < D$, the disease-free equilibrium E_0 is a stable node since both eigenvalues are negative. In order to prove this theorem, we also need the information about the disease equilibrium E_1. Solving Eq. (2.4) yields two roots:

$$X_\pm = \frac{(A+B+D+BC) \pm \sqrt{\Delta}}{2D(A+B)}, \tag{2.6}$$

where

$$\begin{aligned} \Delta &= (A+B+D+BC)^2 - 4(C+1)D(A+B) \\ &= (A+B-D-BC)^2 - 4C(A+B)(D-B), \end{aligned} \tag{2.7}$$

which implies that the existence condition for $X_\pm$ when $B < D$ is given by

$$\begin{aligned} \Delta &= (A+B-D-BC)^2 - 4C(A+B)(D-B) \\ &= [A+B-D-BC+2\sqrt{C(A+B)(D-B)}]\,\mathrm{H}_1 > 0. \end{aligned} \tag{2.8}$$

Now, based on H_1, we discuss the existence condition of biologically meaningful solutions $X_\pm$.

(i) When $\mathrm{H}_1 \geq 0$, it yields $\Delta \geq 0$, for which $0 < X_- \leq X_+ < \frac{1}{D}$, implying that the disease equilibrium E_1 has two solutions E_{1+}: (X_+, Y_+) and E_{1-}: (X_-, Y_-). In particular, when $\mathrm{H}_1 = 0$, $0 < X_- = X_+ < \frac{1}{D}$, indicating a saddle-node bifurcation to occur from the equilibrium E_1.

(ii) When $\mathrm{H}_1 < 0$, there are two cases.

(iia) If $-2\sqrt{C(A+B)(D-B)} < A+B-D-BC < 2\sqrt{C(A+B)(D-B)}$, then $\Delta < 0$, and so there is no real solution for $X_\pm$. Thus, equilibrium E_1 does not exist.

(iib) If $A+B-D-BC \leq -2\sqrt{C(A+B)(D-B)}$ under which $\Delta \geq 0$, we then have $X_+ \geq X_- \geq \frac{1}{D}$, showing that there do not exist biologically meaningful equilibria E_1.

The above discussions show that a biologically meaningful equilibrium E_1 does not exist if $\mathrm{H}_1 < 0$ (with $B < D$), and in this case, there exists only one stable equilibrium E_0 on the boundary of the trapping region G. By index theory, this means that all trajectories of system (2.1) must converge to the stable node E_0, and so the disease-free equilibrium E_0 is globally asymptotically stable if $\mathrm{H}_1 < 0$ when $B < D$.

Remark 2.1. The condition $B \geq D$ guarantees the existence of unique disease equilibrium E_1, for which the disease-free equilibrium E_0 is a saddle point. (When $B = D$, E_0 is a degenerate saddle point, which will be proved later in Section 2.4.) When $B < D$, the disease equilibrium E_1 may or may not exist. The additional condition $H_1 \geq 0$ (with $B < D$) guarantees the existence of two disease equilibria $\mathrm{E}_{1\pm}$ ($\mathrm{E}_{1-} = \mathrm{E}_{1+}$ when $B = D$). It can be easily seen from (2.5) that when $B < D$, $H_1 \geq 0$ implies $A + B - D - BC > 0$, i.e., $A > (D-B) + BC$, indicating that A must pass through a threshold value to generate the

disease equilibrium solution E_1. This is clear from the second equation of (2.1), which can be rewritten as $\frac{dY}{d\tau} = [(BX-1) + \frac{AXY}{Y+C}]Y$, that the first term $BX-1<0$ for $X<\frac{1}{D}$ and $B<D$. Thus, $\frac{dY}{d\tau}<0$ with small values of A for all values of X, implying that Y will die out. When the value of A exceeds its threshold, $\frac{dY}{d\tau}$ becomes positive at least for some values of X, which makes Y gain a steady state and thus the disease equilibrium E_1 exists. Biologically, the threshold value of the contact rate, A, means that the interaction between X and Y produces sufficient infection such that Y persists.

In the remainder of the proof, we assume that $B<D$ and $H_1 \geq 0$. If $H_1>0$, then $0<X_-<X_+<\frac{1}{D}$, which implies that two biologically meaningful equilibrium solutions exist for E_1. When $H_1=0$, we have $0<X_-=X_+<\frac{1}{D}=X_0$, which means that there is only one solution for equilibrium E_1. To find the stability of the equilibrium E_1, evaluating the Jacobian matrix of system (2.1) at E_1 results in

$$\begin{aligned} J(E_1) &= \begin{bmatrix} -D-\left(B+\frac{AY}{Y+C}\right)Y & -\left(B+\frac{AY}{Y+C}\right)X-\frac{ACXY}{(Y+C)^2} \\ \left(B+\frac{AY}{Y+C}\right)Y & \left(B+\frac{AY}{Y+C}\right)X+\frac{ACXY}{(Y+C)^2}-1 \end{bmatrix}_{(X,Y)=(X_1,Y_1)} \\ &= \begin{bmatrix} -\frac{1}{X_1} & -1-\frac{ACX_1Y_1}{(Y_1+C)^2} \\ \frac{1}{X_1}-D & \frac{ACX_1Y_1}{(Y_1+C)^2} \end{bmatrix}. \end{aligned} \tag{2.9}$$

Then, the characteristic equation of E_1 is given by

$$\xi^2 - \mathrm{Tr}(J)\,\xi + \det(J) = 0, \tag{2.10}$$

where

$$\begin{aligned} \det(J) &= -\frac{ACY_1}{(Y_1+C)^2} + \left(\frac{1}{X_1}-D\right)\left(1+\frac{ACX_1Y_1}{(Y_1+C)^2}\right) \\ &= \frac{1}{X_1} - D - \frac{CD}{Y_1+C}\frac{AY_1}{Y_1+C}X_1 \\ &= \frac{1}{X_1} - D - \frac{CD}{Y_1+C}\left(\frac{1}{X_1}-B\right)X_1 \\ &= \frac{1}{(1-DX_1+C)X_1}\frac{1}{D}[(D+BC)(1-DX_1)^2 + 2C(D-B)(1-DX_1) + C(B-D)] \\ &= \frac{-(1-DX_1)}{(1-DX_1+C)X_1}[(A+B+D+BC)X_1 - 2(1+C)] \quad \text{(by using (2.4))} \\ &= \frac{-(1-DX_1)}{2D(A+B)(1-DX_1+C)X_1}[\sqrt{\Delta}(\sqrt{\Delta} \pm (A+B+D+BC)], \end{aligned} \tag{2.11}$$

in which $Y_1 = 1-DX_1$ and (2.6) have been used. Since it is assumed that $H_1>0$, i.e. $\Delta = (A+B+D+BC)^2 - 4(C+1)D(A+B) > 0$, we have

$$\det(J) < 0 \quad \text{for} \quad X_1 = X_+, \quad \text{and} \quad \det(J) > 0 \quad \text{for} \quad X_1 = X_-. \tag{2.12}$$

When $\det(J)<0$, the two eigenvalues of the characteristic polynomial (2.10) are real with opposite signs, and thus the equilibrium point $E_{1+} = (X_+, Y_+)$ is a saddle point.

To consider the property of another equilibrium point $E_{1-} = (X_-, Y_-)$, we need to calculate $\mathrm{Tr}(J)$ as follows:

$$\begin{aligned} \mathrm{Tr}(J) &= -\frac{1}{X_1} + \frac{ACX_1Y_1}{(Y_1+C)^2} \\ &= -\frac{1}{X_1} + \frac{CX_1}{Y_1+C}\left(\frac{1}{X_1}-B\right) \\ &= \frac{-(1-DX_1)-C+CX_1(1-BX_1)}{X_1(Y_1+C)} \\ &= -\frac{1}{X_1(Y_1+C)}[BCX_1^2 - (C+D)X_1 + C + 1] \\ &= -\frac{1}{(Y_-+C)}[(BC-DA-DB)X_- + (A+B+BC-C)] \\ &= -\frac{1}{2D(A+B)(Y_-+C)}[AD(A+B-C) + BC(DA+DB+BC) \\ &\quad -(D-B)(A+B)(C+D) + (DA+DB-BC)\sqrt{\Delta}], \end{aligned} \tag{2.13}$$

Table 1
Classification of E_{1-} ($H_1 \geq 0$).

	$H_2 < 0$	$H_2 > 0$	$H_2 = 0$
$H_3 < 0$	SF	UF	Elementary center
$H_3 > 0$	SN	UN	-—
$H_3 = 0$	DSN	DUN	Double-zero

which can be positive or negative, depending upon the values of parameters. Therefore, the equilibrium point E_{1-} may be a stable (or an unstable) node or focus.

Summarizing the above results, we have shown that when $B < D$, the boundary equilibrium E_0 is a stable node. Moreover, when $H_1 < 0$, a biologically meaningful disease equilibrium E_1 does not exist and E_0 is the unique equilibrium solution, so it is globally asymptotically stable by applying Theorem 2.1. When $H_1 \geq 0$, there exist two disease equilibria, E_{1+} and E_{1-} (E_{1-} coincides E_{1+} if $H_1 = 0$, giving rise to a saddle-node bifurcation), and E_{1+} is a saddle point, while E_{1-} may be a stable (or an unstable) node or focus. In this case, no conclusion can be made regarding for the global stability of the disease-free equilibrium E_0. □

When $\det(J) > 0$, we may use $\mathrm{Tr}(J)$ and $\det(J)$ to further classify the equilibrium point E_{1-}. For convenience, let

$$H_2 \triangleq (D-B)(A+B)(C+D) - AD(A+B-C) - BC(DA+DB+BC) - (DA+DB-BC)\sqrt{\Delta}, \quad (A+B+D+BC-\sqrt{\Delta} > 0), \tag{2.14}$$

and

$$H_3 \triangleq \mathrm{Tr}^2(J) - 4\det(J), \quad (A+B+D+BC-\sqrt{\Delta} > 0). \tag{2.15}$$

Thus, $\mathrm{Tr}(J)$ has the same sign of H_2, and $\det(J)$ has the same sign of $A+B+D+BC-\sqrt{\Delta}$, but H_2 and $A+B+D+BC-\sqrt{\Delta}$ only depends upon the parameters A, B, C and D.

Then, E_{1-} can be classified according to the signs of H_2 and H_3, as shown in Table 1, where SF, UF, SN, UN, DSN and SUN stand for Stable Focus, Unstable Focus, Stable Node, Unstable Node, Degenerate Stable Node and Degenerate Unstable Node, respectively.

So far, the above analysis and conclusions are for the 4-dimensional parameter space (A, B, C, D), where the parameters are assumed to take arbitrary positive real values. However, in real applications these parameters are not free to be chosen due to certain physical restrictions. The typical realistic parameter values used in [13,14] to demonstrate different dynamical behaviors of system (2.1) are taken as

$$A = 0.364, \quad C = 0.823, \quad D = 0.057, \quad B = 0.060, \tag{2.16}$$

for which viral blips. Although these parameter values are used to study the blips phenomenon, we should consider possible parameter values near these values, rather than considering the all values in the positive cone of the parameter space. Searching the whole 4-dimensional parameter space may show some interesting dynamical behavior, but may most likely not realistic.

Note that $B = D = 0.057$ is the transcritical point between the equilibrium solutions E_0 and E_1, and the oscillating behavior (blips) shown in [13,14] is for $B > D$. Here, we want to change the parameter values near the above set of values for $B < D$ to demonstrate more interesting dynamical behaviors, in particular, the bistable equilibrium solutions, Hopf and generalized Hopf bifurcations, and Bogdanov–Takens (BT) bifurcation. In order to give a more clear view of analysis, rather than giving a more generic picture for the 4-dimensional parameter space, we will choose two parameters and fix the other two parameters to explore all possible dynamical behaviors in the two parameter space. Since in [13,14] the parameter B ($B > D$) has been treated as a bifurcation parameter to explore the blips phenomenon, and D is related to B by defining the reproductive number as $R_0 = \frac{B}{D}$. Therefore, in this paper, we will take the parameters A and C as bifurcation parameters and investigate their effects on dynamical behavior, since these two parameters involved in the $\beta(X, Y)$ function play a very important role in the modeling. To achieve this, we fix $B = 0.054$, and choose two values for $D = 0.057, 0.087$, and then plot the three curves $H_1 = H_2 = H_3 = 0$ on the A–C plane, as shown in Figs. 1 and 2, where the red curve, blue curve and green curve correspond to $H_1 = 0$, $H_2 = 0$ and $H_3 = 0$, respectively. Figures 1 and 2 clearly indicate the regions corresponding to the classification shown in Table 1. If we vary the parameters B and D, we will obtain more such figures, showing rich patterns of dynamical behaviors. It should be noted from Fig. 1(a) that the very narrow region bounded by the red curve and green curve corresponds to $H_1 > 0, H_2 > 0$ and $H_3 > 0$, and thus taking parameter values from this region generate an unstable node E_{1-}. Each point on the curve $\mathrm{Tr}(J) = 0$ yields a Hopf critical point, leading to bifurcation of limit cycles. At the intersection point of the blue curve ($H_2 = 0$) and the green curve ($H_3 = 0$), as shown in Figs 1(b) and 2(b), $\mathrm{Tr}(J) = \det(J) = 0$, giving rise to a BT bifurcation, characterized by a double-zero eigenvalue. Thus, by using Figs. 1 and 2, we can easily find different values of A and C to get different types of the equilibrium E_{1-}. Also note from these two figures that the BT bifurcation point, marked by a circle, is actually the intersection point of all three curves $H_1 = H_2 = H_3 = 0$. A number of sets of these parameter values and their corresponding classification of E_{1-} are given in Table 2. In this section, we present the results for the non-degenerate cases ($H_2 H_3 \neq 0$), and leave the degenerate cases, leading to Hopf and generalized Hopf bifurcations, and BT bifurcation, to be considered later in Sections 3 and 4.

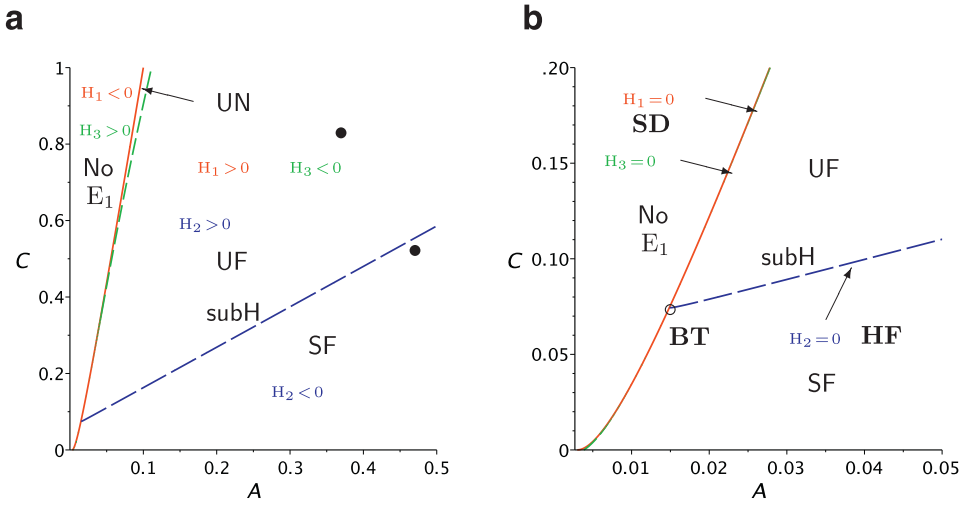

Fig. 1. (a) Plot of the three curves $H_1 = 0$ (in red), $H_2 = 0$ (in blue) and $H_3 = 0$ (in green), on the A-C plane for $B = 0.054$, $D = 0.057$, with signs of H_1, H_2 and H_3 indicated; and (b) a zoomed in region near the origin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

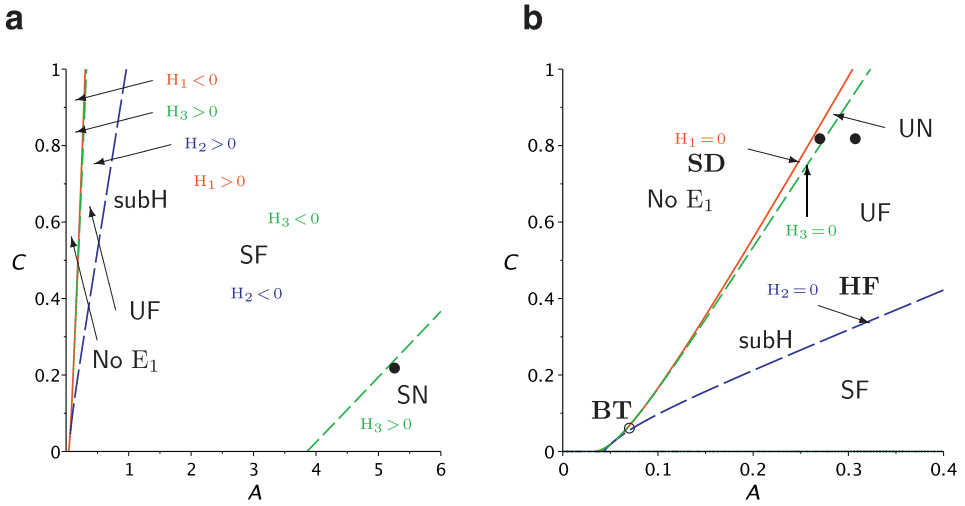

Fig. 2. (a) Plot of the three curves $H_1 = 0$ (in red), $H_2 = 0$ (in blue) and $H_3 = 0$ (in green), on the A-C plane for $B = 0.054$, $D = 0.087$, with signs of H_1, H_2 and H_3 indicated; and (b) a zoomed in region near the origin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Hence, when $B < D$ and $H_1 \geq 0$, for positive parameter values, there may exist bistable equilibrium solutions E_0 and E_1, and bifurcation of limit cycles or even homoclinic orbits from the BT bifurcation.

In the following, we will further investigate the bistable equilibrium solutions in more details using simulation, and then try to provide some biological explanation. For completeness, we also show the results for the cases $H_1 < 0$ and $H_1 = 0$, see Table 2, where $A^{(1)} = 0.09559649$, $A^{(2)} = 0.26302225$. Note that the results for the two sets of values in rows three and eight (see Table 2) are obtained by taking a point from the narrow region of Fig. 1(a) and a point from the narrow region of Fig. 1(b), respectively. We shall present the simulations for the sets of values in Table 2 in the rows 4, 5, 6, 7, 8 and 11, and the corresponding points in the (A, C) parameter space are marked by the black points in Figs. 1 and 2. Also, in Figs. 1(b) and 2(b), the saddle-node (SD) bifurcation (usually saddle-node bifurcation is denoted by SN, which has been used for Stable Node in this paper), determined by $H_1 = 0$, and the Hopf (HF) bifurcation, determined by $H_2 = 0$, are indicated, and the BT bifurcation is marked by a circle.

2.2.1. $A = 0.364$, $C = 0.823$, $D = 0.057$, $B = 0.054$

For this set of parameter values, system (2.1) has three equilibrium solutions: $E_0 = (X_0, Y_0) = (17.5439, 0)$, $E_{1+} = (X_{1+}, Y_{1+}) = (17.4056, 0.0079)$ and $E_{1-} = (X_{1-}, Y_{1-}) = (4.3959, 0.7494)$. It can be shown that E_0 is a stable node, E_{1+} is a saddle point, while E_{1-} is an unstable focus. The phase portrait is shown in Fig. 3, indicating that there do not exist limit cycles, and the disease-free equilibrium E_0 attracts all solution trajectories in the first quadrant of the X–Y plane, which do not start from E_{1-}, or E_{1+}, or the stable manifold of E_{1+},

Table 2
Classification of E_{1-} for given parameter values ($D > B = 0.054$).

A	C	D	E_{1-}	Eigenvalues	H_1	Class
0.100	1.050	0.057	No E_1, E_0 exist	$-0.0570, -0.0526$	< 0	SN
$A^{(1)}$	0.950	0.057	(15.122, 0.1380)	$0.0940, 0$	$= 0$	DUN
0.100	0.950	0.057	(13.901, 0.2076)	$0.0999, 0.0327$	> 0	UN
0.364	0.823	0.057	(4.3959, 0.7494)	$0.0858 \pm 0.3747i$	> 0	UF
0.464	0.523	0.057	(2.9509, 0.8318)	$-0.0072 \pm 0.5132i$	> 0	SF
0.260	0.823	0.087	No E_1, E_0 exist	$-0.0870, -0.3793$	< 0	SN
$A^{(2)}$	0.823	0.087	(8.1300, 0.2927)	$0.2908, 0$	$= 0$	DUN
0.264	0.823	0.087	(7.8326, 0.3186)	$0.2719, 0.0165$	> 0	UN
0.364	0.823	0.087	(4.9202, 0.5719)	$0.1150 \pm 0.2556i$	> 0	UF
0.364	0.250	0.087	(3.0732, 0.7326)	$-0.0566 \pm 0.4655i$	> 0	SF
5.200	0.223	0.087	(0.2331, 0.9797)	$-1.8817, -2.2251$	> 0	SN

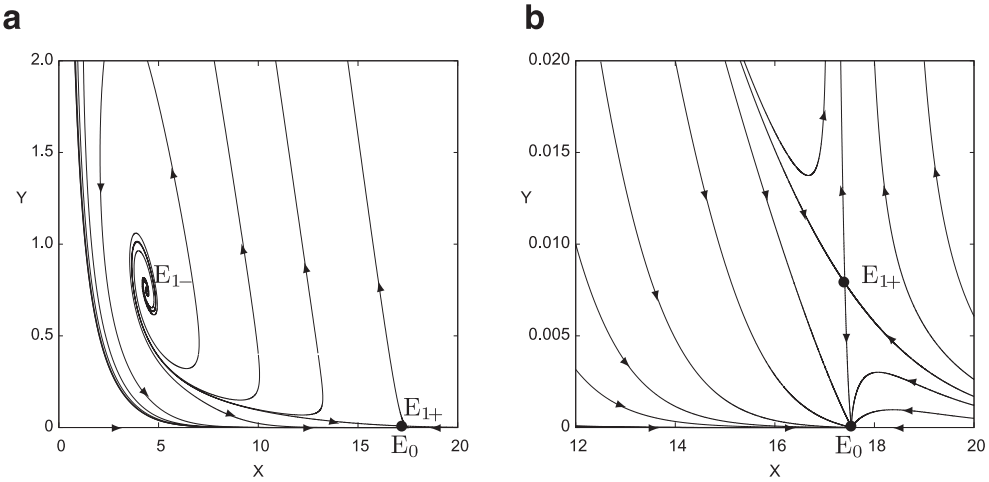

Fig. 3. Simulated phase portrait of system (2.1) for $A = 0.364$, $C = 0.823$, $D = 0.057$, $B = 0.054$, showing the global stability of E_0: (a) depicting three equilibrium points E_0, E_{1+}, E_{1-}; and (b) showing E_0, E_{1+} in a zoomed in region.

2.2.2. $A = 0.464$, $C = 0.523$, $D = 0.057$, $B = 0.054$

For this set of parameter values, system (2.1) still has three equilibrium solutions: $E_0 = (X_0, Y_0) = (17.5439, 0)$, remains unchanged from the previous case since D is not changed, and is a stable node; $E_{1+} = (X_{1+}, Y_{1+}) = (17.4800, 0.0036)$ is still a saddle point, but now $E_{1-} = (X_{1-}, Y_{1-}) = (2.9509, 0.8318)$ becomes a stable focus. The phase portrait for this case is depicted in Fig. 4, which shows an unstable limit cycle enclosing the stable focus E_{1-}. Thus, for this set of parameter values, there exist bistable equilibrium solutions E_0 and E_{1-}. The attracting region for E_{1-} is the region inside the limit cycle, while the area outside the limit cycle is the attracting region for E_0.

To view the bistable equilibrium solutions, we plot the bifurcation diagram in the A-X plane for fixed values: $C = 0.523$, $D = 0.057$, $B = 0.054$, as shown in Fig. 5, where the solid red line and blue curve denote the stable equilibria E_0 and E_{1-}, respectively, while the dashed blue line represents the unstable equilibrium E_{1+}. A saddle-node bifurcation point is seen between E_{1-} and E_{1+}, which is actually the underlying cause for the existence of bistable equilibrium solutions. In fact, the saddle-node bifurcation point is the turning point on the solution curve E_1.

2.2.3. $A = 0.264$, $C = 0.823$, $D = 0.087$, $B = 0.054$

For this set of parameter values, system (2.1) has three equilibrium solutions: $E_0 = (X_0, Y_0) = (11.4943, 0)$, a stable node; $E_{1+} = (X_{1+}, Y_{1+}) = (8.4127, 0.2681)$, a saddle point; and $E_{1-} = (X_{1-}, Y_{1-}) = (7.8326, 0.3186)$, an unstable node. The phase portrait for this case is given in Fig. 6, showing that there do not exist limit cycles, and the disease-free equilibrium E_0 attracts all solution trajectories in the first quadrant of the X–Y plane, which do not start from E_{1-}, or E_{1+}, or the stable manifold of E_{1+},

2.2.4. $A = 5.200$, $C = 0.223$, $D = 0.087$, $B = 0.054$

For this set of parameter values, system (2.1) still has three equilibrium solutions: $E_0 = (X_0, Y_0) = (11.4943, 0)$, a stable node; $E_{1+} = (X_{1+}, Y_{1+}) = (11.4778, 0.0014)$, a saddle point; and $E_{1-} = (X_{1-}, Y_{1-}) = (0.2331, 0.9797)$, a stable node. The phase portrait for this case is depicted in Fig. 7, which shows no limit cycles to exist, but there still exist bistable equilib-

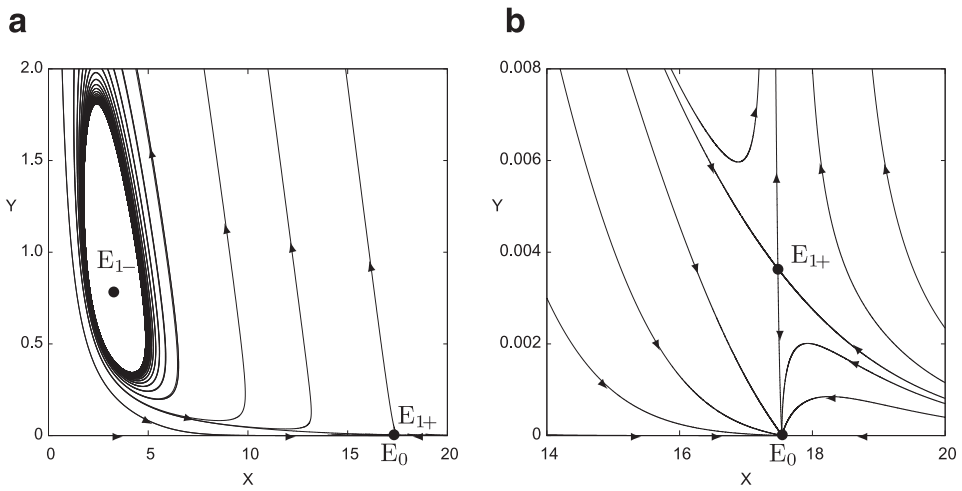

Fig. 4. Simulated phase portrait of system (2.1) for $A = 0.464$, $C = 0.523$, $D = 0.057$, $B = 0.054$, showing the bistable equilibria E_0 and E_{1-} and an unstable limit cycle: (a) depicting three equilibrium points E_0, E_{1+}, E_{1-}; and (b) showing E_0, E_{1+} in a zoomed in region.

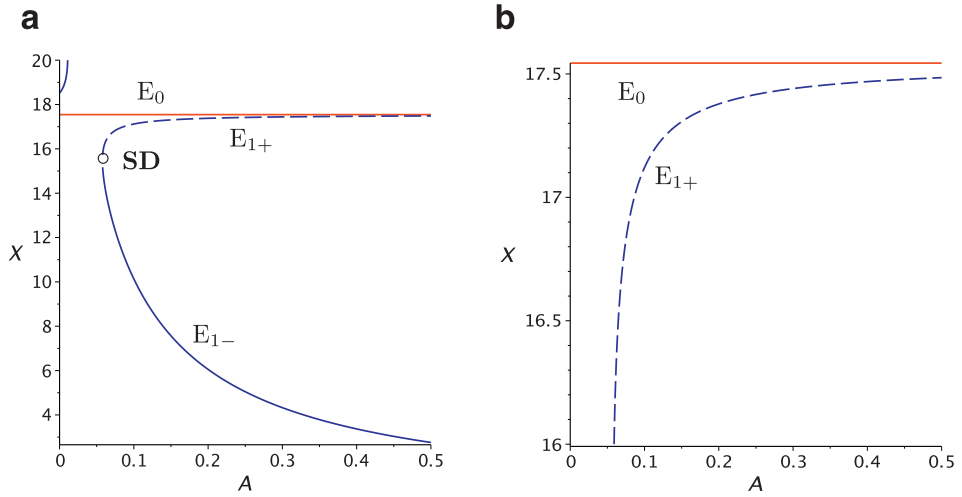

Fig. 5. (a) Bifurcation diagram for the bistable equilibrium solutions for $B = 0.054$, $C = 0.823$, $D = 0.057$; and (b) a zoomed in region near the equilibrium E_0.

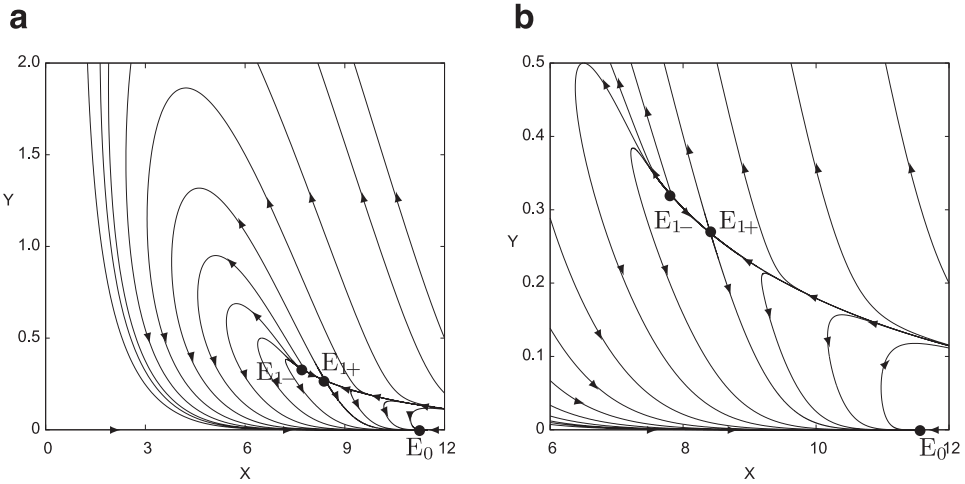

Fig. 6. Simulated phase portrait of system (2.1) for $A = 0.264$, $C = 0.823$, $D = 0.087$, $B = 0.054$, showing the global stability of E_0: (a) depicting three equilibrium points E_0, E_{1+}, E_{1-}; and (b) showing E_0, E_{1+} in a zoomed in region.

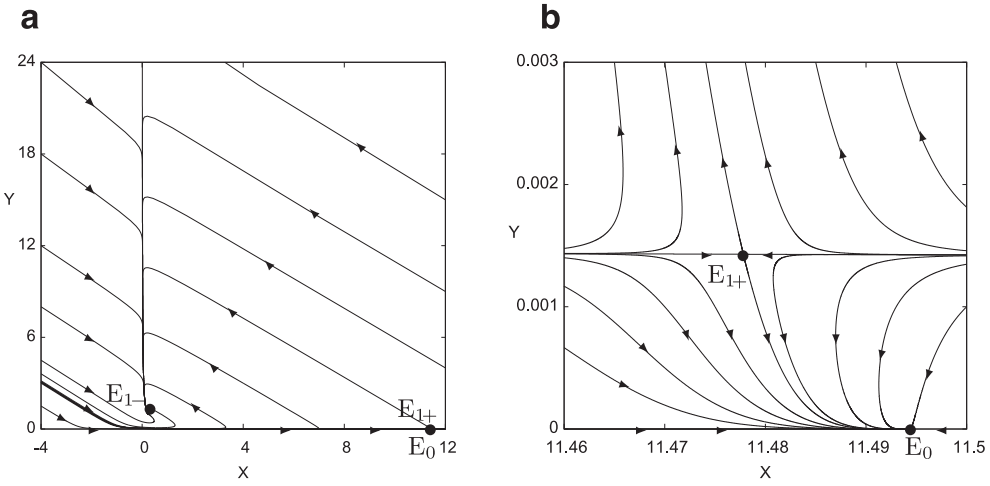

Fig. 7. Simulated phase portrait of system (2.1) for $A = 5.2$, $C = 0.223$, $D = 0.097$, $B = 0.064$, showing the bistable equilibria E_0 and E_{1-}: (a) depicting three equilibrium points E_0, E_{1+}, E_{1-}; and (b) showing E_0, E_{1+} in a zoomed in region.

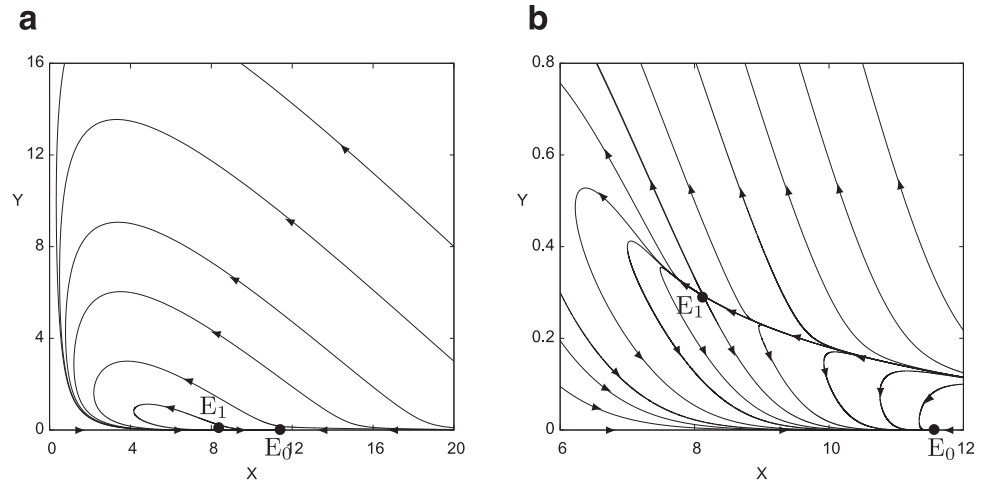

Fig. 8. Simulated phase portrait of system (2.1) for $A = 0.26302225$, $C = 0.823$, $D = 0.087$, $B = 0.054$, showing the global stability of E_0: (a) depicting two equilibrium points E_0, E_1; and (b) showing E_0, E_1 in a zoomed in region.

rium solutions E_0 and E_{1-}. The attracting regions for E_0 and E_{1-} are separated by the two trajectories passing through the saddle point E_{1+}. A similar bifurcation diagram like that given in Fig. 5 can be obtained.

2.2.5. $A = 0.26302225$, $C = 0.823$, $D = 0.087$, $B = 0.054$ $(H_1 = 0)$

For this set of parameter values, $H_1 = 0$ under which $E_{1+} = E_{1-} = E_1 = (8.1300, 0.2927)$, which is a degenerate node consisting of a stable manifold, as shown in Fig. 8, and thus the disease-free equilibrium $E_0 = (11.4943, 0)$ attracts all solution trajectories in the first quadrant of the X-Y plane, which do not start from E_1 and its stable manifold.

2.2.6. $A = 0.260$, $C = 0.823$, $D = 0.087$, $B = 0.054$ $(H_1 < 0)$

For this set of parameter values, $H_1 = -0.002136 < 0$ under which E_1 does not exist, and so the disease-free equilibrium $E_0 = (11.4943, 0)$ is globally asymptotically stable. The simulated phase portrait is similar to Fig. 8(a) but without the existence of E_1.

The most interesting phenomenon found in this section for $B < D$ is the bistable equilibrium solutions E_0 and E_1. E_0 is always a stable node, while E_{1-} may be a stable focus (see Fig. 4) or a stable node (see Fig. 7). The separator between the two attracting regions of the two stable equilibria is either an unstable limit cycle (see Fig. 4) or the saddle trajectories. Dynamically, this bistable phenomenon is due to the existence of a saddle-node bifurcation on the equilibrium solution E_1, which has two branches, one of them is stable and the other is unstable. Biologically, this phenomenon is not fully understood. System (2.1) was developed from an in-host model of HIV infection, and there has been evidence of possible bistability in this disease. In particular, the equilibrium viral load, or "viral set point" can differ by orders of magnitude among patients. Several authors have previously suggested bistable equilibrium solutions as an explanation for the phenomenon [24].

It is also noted from Figs. 3, 6 and 8 that when E_{1-} is unstable (either focus or node), the equilibrium E_0 seems globally asymptotically stable. This may be explained as follows: first, it can be seen from Fig. 1 that when the parameters A and C are varied to cross the blue curve, defined by $H_2 = 0$, from the bottom-right to the top-left (e.g., in the negative direction of the A-axis), the equilibrium E_{1-} changes from a stable focus (SF) to an unstable focus (UF). Hopf bifurcation occurs when the parameters are varied to cross the blue curve. The simulations shown in Figs. 4 and 3 correspond to the two points chosen from the SF region and UF region, respectively, implying that the Hopf bifurcation is subcritical. This is why an unstable limit cycle is shown in Fig. 4, while there is no limit cycle in Fig. 3 and so all trajectories converge to the stable node E_0. Similarly, the simulations shown in Figs. 6 and 8 imply that when the parameters A and C are varied to cross the blue curve ($H_2 = 0$ in Fig. 2) from the bottom-right to the top-left, a subcritical Hopf bifurcation occurs. The proof for the two subcritical Hopf bifurcations will be given in Section 3.

2.3. Dynamical behavior of (2.1) when B > D

Now, we discuss the dynamical behavior of system (2.1) for $B > D$. In this case, E_0 becomes a saddle point, while E_1 always exists, since Eq. (2.4) always has two roots for

$$\begin{aligned}\Delta &= (A+B+D+BC)^2 - 4(C+1)D(A+B)\\ &= (A+B-D-BC)^2 + 4C(A+B)(B-D)\\ &> (A+B-D-BC)^2 \quad (\text{due to } B>D)\\ &\geq 0,\end{aligned}$$

and thus $0 < X_- < X_+$. Further, noticing from (2.4) that $Q(0) = C+1 > 0$ and $Q(\frac{1}{D}) = -C(\frac{B}{D} - 1) < 0$ $(B > D)$, we have

$$0 < X_- < \frac{1}{D} < X_+.$$

Thus,

$$X_1 = X_- = \frac{(A+B+D+BC) - \sqrt{\Delta}}{2D(A+B)}, \quad \text{since} \quad X_1 \in \left[0, \frac{1}{D}\right],$$

which guarantees that $0 \leq Y_1 = 1 - DX_1 \leq 1$.

Since E_0 is a saddle point (unstable), and $X_{1+} > \frac{1}{D}$ (which yields $Y_{1+} < 0$) is not biologically meaningful, bistable equilibria cannot exist bistable equilibria for this case $B > D$. To find the stability of E_1 (i.e., E_{1-}) when $B > D$, we first show that $\det(J) > 0$. This can be obtained using (2.11) as follows:

$$\begin{aligned}\det(J) &= \frac{1}{X_1} - D + \frac{CD}{Y_1 + C}(BX_1 - 1)\\ &> \frac{1}{X_1} - D + \frac{CD}{Y_1 + C}(DX_1 - 1) \quad (B > D)\\ &= \left(\frac{1}{X_1} - D\right)\left(1 - \frac{CDX_1}{Y_1 + C}\right) \quad (0 < DX_1 < 1,\ 0 < Y_1 < 1)\\ &> \left(\frac{1}{X_1} - D\right)\left(1 - \frac{CDX_1}{C}\right)\\ &= \frac{1}{X_1}(1 - DX_1)^2 > 0.\end{aligned}$$

Therefore, all the formulae derived in the previous section for E_{1-} (when $B < D$) and the results shown in Table 1 can be applied here to classify the type of the equilibrium E_{1-} (when $B > D$). Similarly, we may fix B and D and then plot the two curves $H_2 = H_3 = 0$ on the A–C plane to identify the possible parameter values which yield different qualitative behavior of system (2.1). Note that now for $B > D$ we do not need the condition $H_1 > 0$ since $\Delta > 0$ is guaranteed when $B > D$. Two sets of values for $(B, D) = (0.057, 0.060)$, $(0.087, 0.090)$ are chosen to plot the figures. However, it is found that these two figures are quite similar, implying that, unlike the case $B < D$, here slightly varying B and D does not change the behavior of the system. Hence, we only present the result for $(B, D) = (0.057, 0.060)$, as shown in Fig. 9. It can be seen that for this case, there is no saddle-node bifurcation, nor BT bifurcation, since $H_1 > 0$ for all parameter values.

It is also seen from Fig. 9 that for most of the parameter values, $H_3 < 0$, in particular for not very large values of A. This means that for most of parameter values, E_1 is a focus. Further, it can be shown that for the points bounded by the blue curve ($H_2 = 0$, i.e. $\mathrm{Tr}(J) = 0$) the equilibrium E_1 is an unstable focus. Therefore, for these parameter values, by Theorem 2.1, we can conclude that there exists at least one stable limit cycle inside the trapping region G. When the parameter values are taken from the region outside the region bounded by the blue curve, the equilibrium E_1 is the unique equilibrium inside the trapping region G, and thus the equilibrium E_1 is globally asymptotically stable.

Now, we are ready to prove the following theorem.

Theorem 2.3. *When $B > D$ and $H_2 > 0$, system (2.1) has at least one stable limit cycle, and the limit cycle must not bifurcate from a homoclinic orbit.*

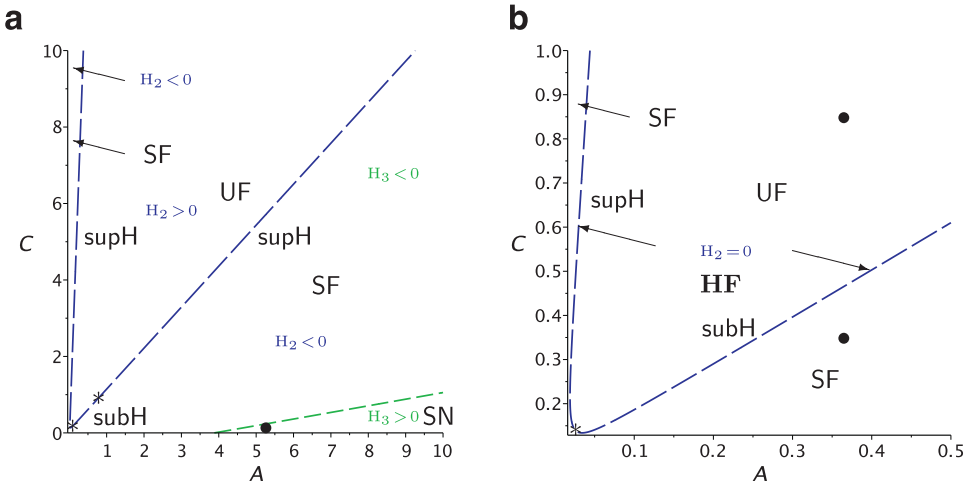

Fig. 9. (a) Plot of two curves $H_2 = 0$ (in blue) and $H_3 = 0$ (in green), on the A-C plane for $B = 0.060$, $D = 0.057$, with signs of H_2 and H_3 indicated; and (b) a zoomed in region near the origin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

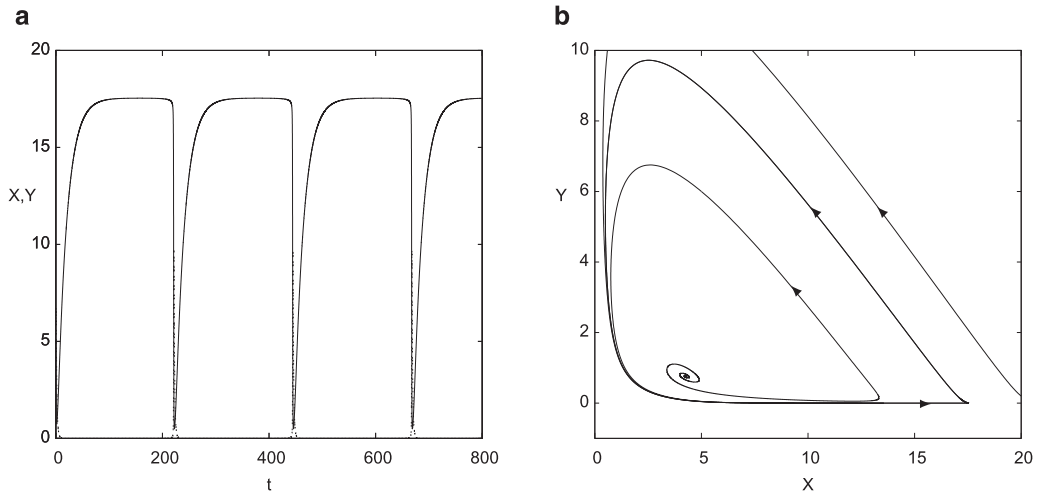

Fig. 10. Simulated blips of system (2.1) for $A = 0.364$, $C = 0.823$, $D = 0.057$, $B = 0.060$: (a) time history showing blips; and (b) phase portrait showing a limit cycle.

Proof. First, we show that the positive equilibrium $E_1 = E_{1-} = (X_1, Y_1)$ is inside the trapping region G, defined in Theorem 2.1. That is, the point E_1 should be below the line L: $X + Y = \max(1, \frac{1}{D}) + \varepsilon$. Note that $Y_1 = 1 - DX_1$ for $0 < X_1 < \frac{1}{D}$, implying that the point (X_1, Y_1) is on the line, defined by $DX + Y = 1$, which is obviously below the line L.

To prove that limit cycles do not bifurcate from a homoclinic orbit, first note that the only possible homoclinic orbit comes from the saddle point E_0 when $B > D$. Thus, it suffices to show that there do not exist homoclinic orbits passing through this singular point. Otherwise, suppose there exists a homoclinic orbit passing through this point, then the homoclinic orbit must leave this point along the direction of the eigenvector $v_2 = (1, \frac{D}{B}(1 - D) - 1)$ and return to this point along the direction of the eigenvector $v_1 = (1, 0)$, that is, the direction of the X-axis. In other words, the homoclinic orbit must return to the saddle point along the X-axis. But we have already shown that the X-axis itself is a solution trajectory, and thus other trajectories, in particular, the one leaving the saddle point along the v_2 direction, cannot connect to the X-axis due to the uniqueness of solutions.

The proof of Theorem 2.3 is complete. □

To end this section, we present three simulations for the common parameter values: $D = 0.057$, $B = 0.060$; but for $(A, C) = (0.364, 0.823)$, $(0.364, 0.350)$ and $(5.2, 0.2)$, respectively. The first simulation is shown in Fig. 10, which yields a blip-like oscillation, as has been discussed in [13,14]. The simulations for the second and third cases are depicted in Fig. 11(a) and (b), respectively. Fig. 11(a) shows that E_1 is asymptotically stable and all trajectories starting from the initial points inside an unstable limit cycle converge to this equilibrium E_1; while trajectories outside the unstable limit cycle converge to a separator of the saddle point E_0. Fig. 11(b) indicates that E_1 is globally asymptotically stable without the existence of limit cycles.

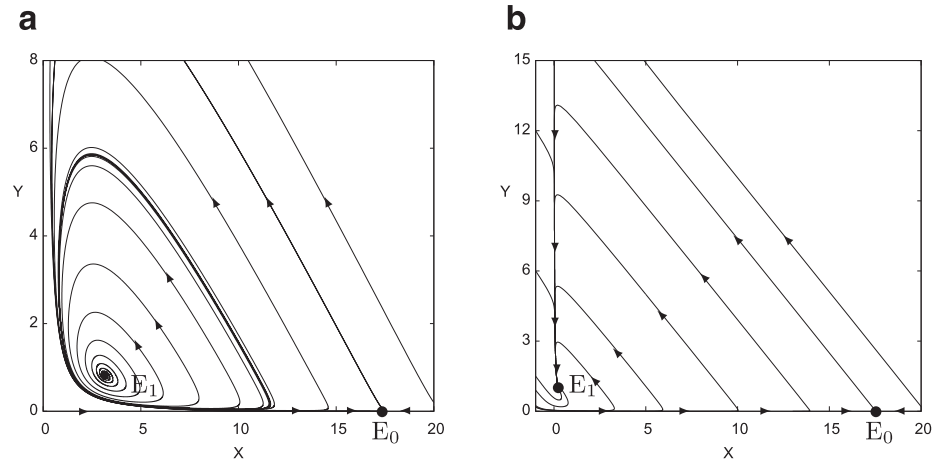

Fig. 11. Simulated phase portrait of system (2.1) when $(B, D) = (0.060, 0.057)$: (a) for $(A, C) = (0.364, 0.352)$ showing the trajectories inside an unstable limit cycle convergent to the stable equilibrium E_1; and (b) for $(A, C) = (5.2, 0.2)$ showing the global stability of E_1 without existence of limit cycles.

The results shown in Figs. 10 and 11 clearly indicate that the Hopf bifurcation which occurs on the left branch of the blue curve in Fig. 9 is supercritical (when, say, A is increasing to cross the blue curve), generating the stable limit cycle (blips) shown in Fig. 10, and the bifurcation which occurs on the right branch of the blue curve (see Fig. 9) is subcritical (when, say, A is decreasing to cross the blue curve), leading to the unstable limit cycle shown in fig. 11(a). The proof for the supercritical and subcritical Hopf bifurcations will be given in Section 3.

2.4. Dynamical behavior of (2.1) when $B = D$

We now turn to the case $B = D$. First note that when $B = D$, the equilibrium $E_{1+} = (X_+, Y_+)$ coincides with the disease-free equilibrium E_0, while the other equilibrium $E_{1-} = (X_-, Y_-) = (\frac{1+C}{A+D}, \frac{A-DC}{A+D})$. In order to have $X_- < \frac{1}{D}$, we require $A + D > D + DC$, or $A > DC$. Note that when $A < DC$, the equilibrium E_{1-} does not exist; and when $A = DC$, the equilibrium E_{1-} also coincides with E_0. So for the generic case, we assume $A > DC$ in this subsection.

To find the stability of E_0 for this case, we note that the two eigenvalues associated with this equilibrium now become $-D$ and 0, which is a critical case and the application of center manifold theory is required to determine its stability. To achieve this, we introduce an affine transformation, given by

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{1}{D} \\ 0 \end{pmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & -D \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{2.17}$$

into (2.1) to obtain a system, expanded around $(u_1, u_2) = (0, 0)$, as

$$\begin{aligned} \frac{du_1}{d\tau} &= -Du_1 + D(D-1)u_1u_2 + \frac{1}{C}(A - DC)(1 - D)u_2^2 + \frac{AD}{C}(1 - D)u_1u_2^2 + \cdots, \\ \frac{du_2}{d\tau} &= Du_1u_2 - \frac{1}{C}(A - DC)\,u_2^2 - \frac{AD}{C}\,u_1u_2^2 + \cdots, \end{aligned} \tag{2.18}$$

whose linear part is now in the Jordan canonical form with eigenvalues $-D$ and 0. To find the center manifold, let $u_1 = h(u_2) = a_2u_2^2 + O(u_2^3)$ and then use (2.18) to find $a_2 = \frac{(1-D)(A-DC)}{DC}$. Therefore, the center manifold up to second order is given by

$$W^C = \left\{ (u_1, u_2) \,|\, u_1 = \frac{(1 - D)(A - DC)}{DC}\, u_2^2 + O(u_2^3) \right\},$$

and the differential equation describing the dynamics on the center manifold is

$$\frac{du_2}{d\tau} = -\frac{1}{C}(A - DC)\,u_2^2 + \frac{(1 - D)(A - DC)}{C}\, u_2^3 + O(u_2^4). \tag{2.19}$$

Since $Y = -Du_2 > 0$, we only consider $u_2 < 0$. Note that the leading term in (2.19) is $-\frac{1}{C}(A - DC)\,u_2^2$ with a negative coefficient, implying that u_2 is decreasing from a negative initial value (and so Y is increasing from a positive initial value). Hence, the equilibrium E_0 is a degenerate saddle point, similar to the case when $B > D$.

Next, we consider the stability of E_{1-}. Evaluating the Jacobian (2.9) at this equilibrium yields

$$J(E_{1-}) = \begin{bmatrix} -\dfrac{A + D}{1 + C} & -\dfrac{A + 2AC - DC^2}{A(1 + C)} \\ \dfrac{A - DC}{1 + C} & \dfrac{C(A - DC)}{A(1 + C)} \end{bmatrix},$$

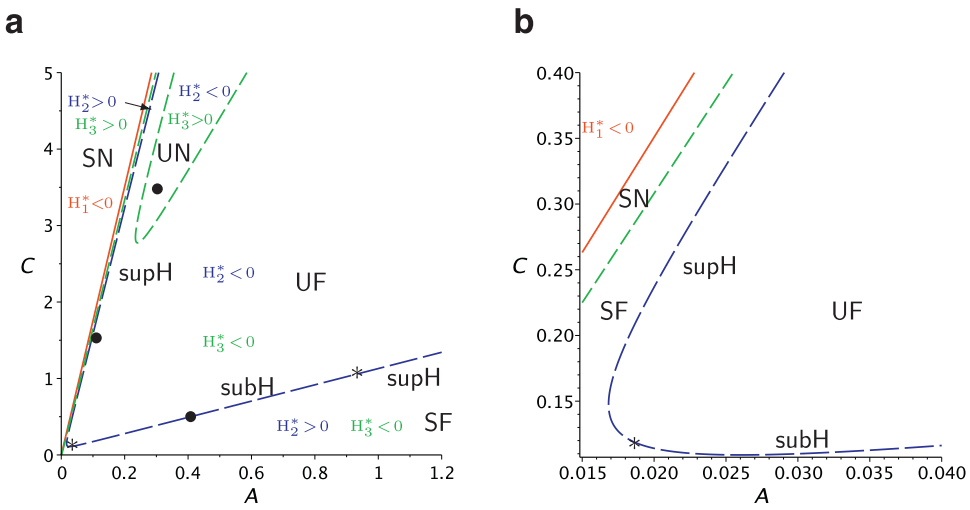

Fig. 12. (a) Plot of the three curves $H_1^* = 0$ (in red), $H_2^* = 0$ (in blue) and $H_3^* = 0$ (in green), on the A-C plane for $B = D = 0.057$, with the regions indicated for the classifications SF, UF, SN and UN; and (b) a zoomed in region near the turning point. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

which in turn results in two eigenvalues, given by

$$\xi_{\pm} = \frac{-[C(A-DC)-A(A+D)] \pm \sqrt{[C(A-DC)-A(A+D)]^2 - 4A(1+C)(A-DC)^2}}{2A(1+C)}. \tag{2.20}$$

Hence, under the condition $A > DC$, the equilibrium E_{1-} is asymptotically stable (unstable) if $C(A-DC)-A(A+D) < 0$ (>0), which is a node (focus) when $[C(A-DC)-A(A+D)]^2 - 4A(1+C)(A-DC)^2 > 0$ (<0). In order to find parameter values for these four categories, let

$$\begin{aligned} H_1^* &\triangleq H_1^{B=D} = A - DC, \\ H_2^* &\triangleq H_2^{B=D} = C(A-DC) - A(A+D), \\ H_3^* &\triangleq H_3^{B=D} = [C(A-DC)-A(A+D)]^2 - 4A(1+C)(A-DC)^2. \end{aligned}$$

Then choosing $B = 0.057$, we plot the three curves $H_1^* = H_2^* = H_3^* = 0$ on the A-C plane, as shown in Fig. 12, from which it is easy to find the parameter values which correspond to different classifications of the equilibrium E_{1-}. Since the equilibrium E_0 is a degenerate saddle node and only one solution exists for E_1, this case $B = D$ is similar to the case $B > D$. Thus, in general, if E_1 is unstable (either a focus or a node), then there must exist stable limit cycles; if E_1 is stable, then it is globally asymptotically stable. When the parameter values of A and C are chosen from the blue curve (see Fig. 12) defined by $H_2^* = 0$, Hopf bifurcation occurs, leading to limit cycles. This will be further discussed in the next section.

3. Hopf and generalized Hopf bifurcations

In this section, we consider bifurcation of limit cycles due to Hopf and generalized Hopf bifurcations. There are three types of Hopf bifurcations, which occur from the critical blue line $H_2 = 0$ for $B < D$ (see Figs. 1 and 2) and $B > D$ (see Fig. 9), and from the critical blue line $H_2^* = 0$ for $B = D$ (see Fig. 12). First we give a detailed analysis for the case $B = D$, and then summarize the results for other cases with representative simulations.

3.1. Hopf bifurcation

We first consider Hopf bifurcation, starting from the case: $B = D = 0.057$, for which the Hopf critical points are located on the blue curve defined by $H_2^* = 0$ (see Fig. 12) is determined from the equation $A(A+D) - C(A-DC) = 0$, from which we solve for C to obtain

$$C_{\pm} = \frac{500 \pm \sqrt{A(19300A - 3249)}}{57A}, \quad \left(A > \frac{3249}{193000}\right), \tag{3.1}$$

where we use $B = D = \frac{57}{1000}$ to facilitate symbolic computation. Note that the leftmost point on the blue curve is given by $(A, C) = (\frac{3249}{193000}, \frac{57}{386})$. The solutions C_- and C_+ correspond to the points (see Fig. 12(b)) on the upper and lower branches of the $H_2^* = 0$ curve, respectively. In order to apply normal form theory to calculate the first-order focus value (or the first

Lyapunov constant), we introduce an affine transformation, given by

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{1000(1+C)}{1000A+57} \\ \frac{1000A-57C}{1000A+57} \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ \frac{-A(1000A+57)}{1000+2000AC-57C^2} & \frac{-1000A(1+C)\omega_c}{1000+2000AC-57C^2} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

where $\omega_c = \frac{1000A-57C}{1000\sqrt{A(1+C)}} > 0$ (since $1000A - 57C > 0$ due to $Y > 0$), into (2.1) to yield a system to be expanded around $(u_1, u_2) = (0, 0)$ up to third-order terms, and then apply the Maple program for computing the normal forms associated with Hopf and generalized Hop bifurcations [25] to this system to obtain the normal form in polar coordinates up to third-order terms as follows:

$$\frac{dr}{d\tau} = r\,[v_0\,\mu + v_1\,r^2 + o(r^4)], \quad \frac{d\theta}{d\tau} = \omega_c + t_0\,\mu + t_1\,r^2 + o(r^4), \tag{3.2}$$

where μ is a perturbation parameter to measure the distance from a critical point on the blue curve $H_2^* = 0$ along the positive direction of the A-axis. v_0 and v_1 are the zero-order and the first-order focus values. The first equation of (3.2) can be used to perform bifurcation analysis and the sign of v_1 determines whether the Hopf bifurcation is supercritical or subcritical. The values v_0 and t_0 can be found from a linear analysis, while v_1 and t_1 are obtained by applying the Maple program. The calculation shows that

$$v_0 = \frac{57C^2 - 1000A^2}{2000A^2(1+C)}, \quad t_0 = \frac{1000A + 57C}{4000A\sqrt{A(1+C)}}, \tag{3.3}$$

and the output of the Maple program gives v_{1-} and v_{1+}, corresponding to C_- and C_+, respectively, as

$$\begin{aligned} v_{1\pm} = {} & \frac{-3249(1000A+57)^3}{8000000000A(500A+A_m)(500A+57-A_m)^3(557000A+60249-1000A_m)^3} \\ & \times [(3864999285035000000000000A^5 + 8614082577809850000000A^4 \\ & + 7051942944965614500000A^3 + 2233569477660976755 00A^2 \\ & - 3214238968494000000A + 38317671392498001) \\ & \pm (8796766369990000000000A^4 + 2033715969208290000000A^3 \\ & + 178485978671452530000A^2 + 759905488695261807A + 24859340130996000)\,A_m]. \end{aligned}$$

where $A_m = \sqrt{A(193000A - 3249)}$. It can be shown that $v_{1+} < 0$ for $A > \frac{3249}{193000} \approx 0.0168$. For v_{1-}, it has two real roots: $A = 0.0184$ and $A = 0.9210$ such that $v_{1-} > 0\ \forall A \in (0.0184, 0.9210)$ and $v_{1-} < 0\ \forall A \in (0.0168, 0.0184) cup (0.9210, \infty)$. Moreover, it can be shown that $v_0 > 0$ when $C = C_+$ for any values of $A > 0.0168$, and there is a critical point on C_-, defined by $A = 0.0260$, such that when $C = C_-$, $v_0 > 0$ for $A \in (0.0168, 0.0260)$ but $v_0 < 0$ for $A > 0.0260$. Therefore, we can combine the information on the signs of v_0 and v_1 to precisely determine whether a Hopf bifurcation is supercritical or subcritical. In fact, on the upper branch C_+ of the blue curve $H_2^* = 0$, all Hopf bifurcations are supercritical, while on the lower branch C_-, the Hopf bifurcation is supercritical for $A \in (0.0168, 0.0184) \cup (0.9210, \infty)$, and subcritical for $A \in (0.0184, 0.9210)$, as shown in Fig. 12, where the two points on the blue curves, at $A = 0.0184$ and $A = 0.9210$ are marked by $*$, where 'supH' and 'subH' represent supercritical and subcritical Hopf bifurcations, respectively.

It should be pointed out that since E_0 is a degenerate saddle point, for any point inside the region bounded by the blue curve, there must exist stable limit cycles due to Poincaré–Bendixson theory no matter whether E_1 is an unstable focus or node. This seems to imply a contradiction for the subcritical Hopf bifurcation from the lower branch of the blue curve for $A \in (0.0184, 0.9210)$, giving rise to unstable limit cycles below the curve. But on the other side, there exist stable limit cycles. This is because the unstable limit cycle is from a local (Hopf) bifurcation, while the stable limit cycle comes from a global bifurcation. Several representative parameter sets (A, C) are chosen for this case when $B = D = 0.057$ as follows:

$$(A, C) = (0.1, 1.55), \quad (0.3, 3.5), \quad (0.42, 0.50), \quad (0.39, 0.50),$$

which are marked on Fig. 12 by black points (the last two are at the same place), and the corresponding simulations are shown in Figs. 13 and 14. Note that all of them show the existence of limit cycles. The first two cases confirm that the Hopf bifurcations emerging from the upper branch of the blue curve are indeed supercritical (with the focus value $v_{1+} < 0$), and so the bifurcating limit cycles are stable (see Fig. 13). The last two points are very close, with one below the curve and one above the curve. The third one yields a typical subcritical Hopf bifurcation and the bifurcating limit cycle is unstable (see Fig. 14(a)). The last one is not generated by Hopf bifurcation though the critical point is near the blue curve. It is a big limit cycle, generated due to Poincaré–Bendixon theory, and it is stable since it encloses an unstable focus (see Fig. 14(b)).

Similarly, we can consider the cases $B < D$ and $B > D$ and determine whether the Hopf bifurcations are supercritical or subcritical. Without giving detailed calculations, we summarize the results as follows. For the case with $B = 0.054 < D = 0.057$, the blue curve actually has a turning point at $A = \frac{729}{49000} \approx 0.014878$ while the BT bifurcation point is above this point at $A = 0.014881$, as shown in Fig. 16(a) in the next section. On the lower branch of the blue curve, the focus value for the Hopf bifurcation (see the blue curve in Fig. 1) is shown to have the property that $v_1 > 0$ for $A \in (0.014981, 0.9455)$

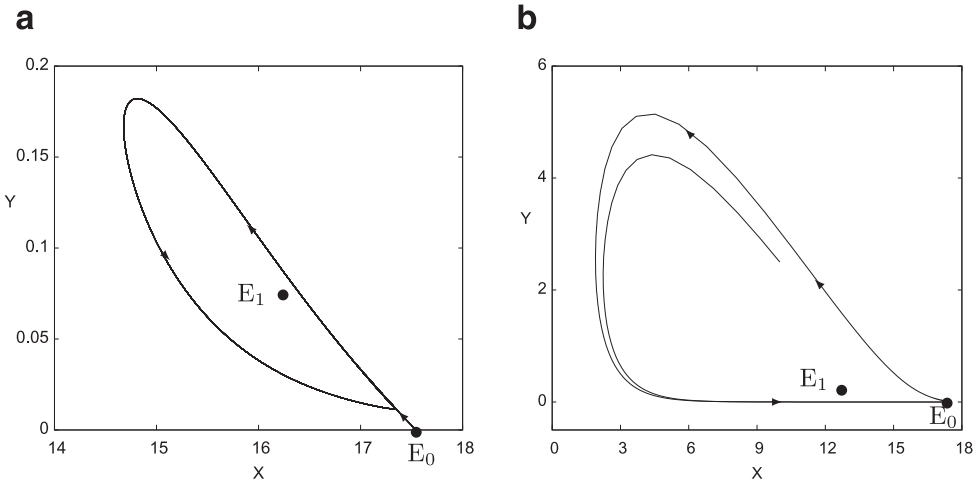

Fig. 13. Simulations of system (2.1) when $B = D = 0.057$, showing stable limit cycles: (a) $(A, C) = (0.1, 1.55)$ with E_1 being an unstable focus; and (b) $(A, C) = (0.3, 3.5)$ with E_1 being an unstable node.

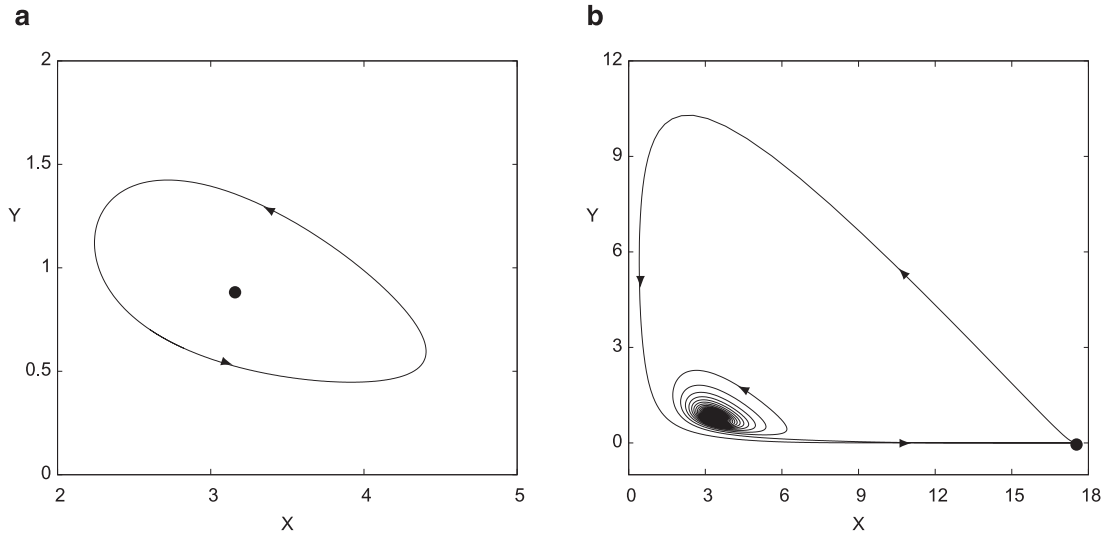

Fig. 14. Simulations of system (2.1) when $B = D = 0.057$, showing (a) an unstable limit cycle for $(A, C) = (0.42, 0.50)$ with E_1 being a stable focus; and (b) a stable limit cycle for $(A, C) = (0.39, 0.50)$ with E_1 being an unstable focus.

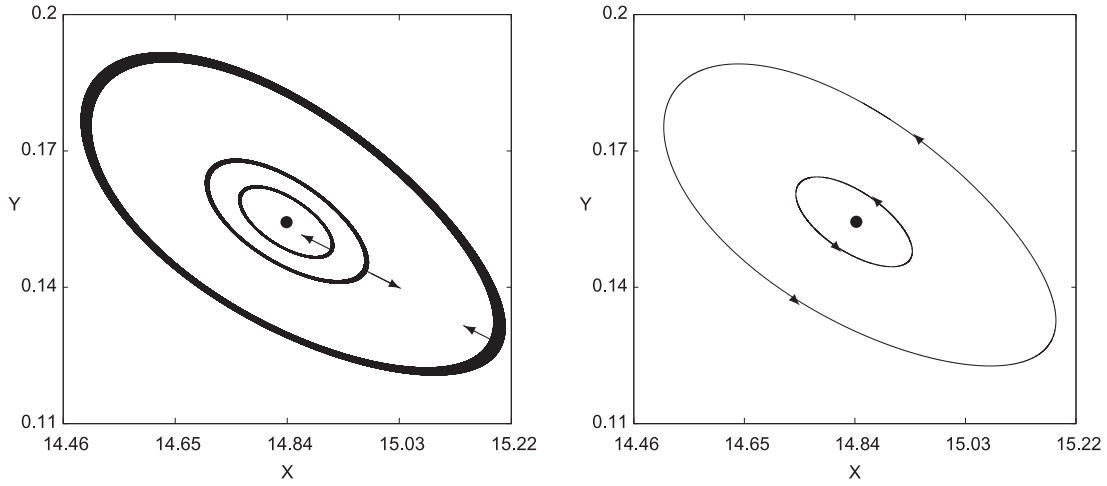

Fig. 15. Simulation of two limit cycles for system (2.1) when $B = D = 0.057$, $A = 0.01846287$, $C = 0.11969000$: (a) three trajectories with moving directions indicated; and (b) two limit cycles with the inner unstable and outer stable.

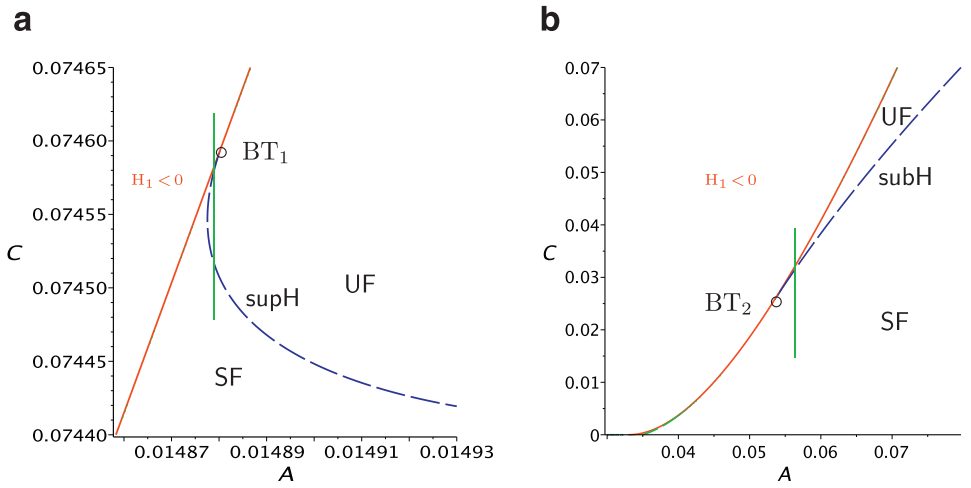

Fig. 16. The BT bifurcation diagram around the critical points: (a) $(B, D, A, C) = (\frac{27}{500}, \frac{57}{1000}, \frac{3078507}{206879500}, \frac{61731}{827518})$, and (b) $(B, D, A, C) = (\frac{27}{500}, \frac{87}{1000}, \frac{118428267}{2237439500}, \frac{219501}{8949758})$.

and $v_1 < 0$ for $A \in (0.014878, 0.014981) \cup (0.9455, \infty)$. On the upper branch of the blue curve, $v_1 < 0$ for $A \in (0.014878, 0.014881)$. Hence, when $(A, C) = (0.364, 0.823)$, the Hopf bifurcation is subcritical, and the bifurcating limit cycle is unstable, as the example shown in Fig. 4. We expect that a Hopf bifurcation is supercritical when choosing a point with $A > 0.9455$. For the case with $B = 0.054 < D = 0.087$ (see Fig. 2), only the upper branch of the blue curve is the solution, which does not contain the turning point, as shown in Fig. 16(b) (in the next section). It is found that the focus value $v_1 > 0$ for $A \in (0.0393, 1.1708)$ and $v_1 < 0$ for $A > 1.1708$. But for this case, the BT bifurcation point is at $A = 0.0529$, and the portion for $A < 0.0529$ yields $H_1 < 0$. Therefore, for this case, $v_1 > 0$ for $A \in (0.0529, 1.1708)$. Several typical simulations can be seen in Figs. 3, 4, 6, 7 and 8.

Finally, we consider the case $B = 0.060 > D = 0.057$ and confirm the conclusion that we made at the end of Section 2.3. Note that for this case $H_1 > 0$ for all positive parameter values. Compared to the case $B < D$, now there are two branches on the blue curve (see Fig. 9). For the upper branch, it can be shown that $v_{1+} < 0$ for $A > 0.0189$, and the Hopf bifurcation emerging from the upper branch of the blue curve is supercritical and so the bifurcating limit cycles are stable (see the blips example in Fig. 10). For the lower branch of the blue curve, it can be shown that the focus value $v_{1-} > 0$ for $A \in (0.0214, 0.8964)$ and $v_{1-} < 0$ for $A \in (0.0189, 0.0214) \cup (0.8964, \infty)$. Hence, the Hopf bifurcation from the lower branch of the blue curve is subcritical for $A \in (0.0214, 0.8964)$, giving rise to unstable limit cycles (an example is shown in Fig. 11). When $A \in (0.0189, 0.0214) \cup (0.8964, \infty)$, the Hopf bifurcation becomes supercritical and so the bifurcating limit cycles are stable. This is similar to the case $B = D$ (see Fig. 12 where supercritical and subcritical Hopf bifurcations are indicated), and thus we omit the details.

By comparing the Figs. 1, 2, 9 and 12, we have observed an important difference between the different cases: although all the blue curves are defined by a quadratic polynomial in A and C, the case $B < D$ shows no turning point on the blue curve, while the cases $B \geq D$ do have a turning point on the blue curve. As a matter of fact, if we zoomed in the area around the BT point in Figs. 1 and 2 (see Fig. 16 in the next section), we will see the turning point for the case $B = 0.054, D = 0.057$ since the blue curve contains the turning point, while the blue curve for the case $B = 0.054, D = 0.087$ does not include the turning point. Summarizing the above results, we have the following theorem.

Theorem 3.1. *For system (2.1), there always exists Hopf bifurcation which occurs from the disease equilibrium* E_1, *for suitable positive parameter values. The bifurcations may be supercritical or subcritical, and a limit cycle bifurcating from a supercritical (subcritical) Hopf critical point is stable (unstable), which encloses an unstable (a stable) focus point – the equilibrium* E_1.

3.2. Generalized Hopf bifurcation

Now we consider possible generalized Hopf bifurcations which may occur from system (2.1), leading to bifurcation of multiple (two) limit cycles from a Hopf bifurcation point. The condition for generalized Hopf bifurcation is that the first-order focus value vanishes, i.e., $v_1 = 0$. In other words, on the Hopf bifurcation curve (the blue curves in Figs. 1, 2, 9 and 12), such a critical point is identified when the Hopf bifurcation changes from supercritical to subcritical, or vice versa.

Again, we first consider the case $B = D = \frac{57}{1000}$, for which there are two generalized Hopf critical points located on the lower branch of the blue curve (see Fig. 12): $A_{gH}^{(1)} = 0.0184$ and $A_{gH}^{(2)} = 0.9210$, where the subscript 'gH' denotes 'generalized Hopf'. Note that in computation we take the accuracy up to 30 decimal points ($A_{gH}^{(1)} = 0.018412874626407510689934961140 4$, $A_{gH}^{(2)} = 0.921012043225272084984762668632$), but only present the results up to 4 decimal points for brevity. The corresponding critical values of C are given by $C_-(A)$ in Eq. (3.1) as $C_{gH}^{(1)} = C_-(A_{gH}^{(1)}) = 0.1199$

and $C_{gH}^{(2)} = C_-(A_{gH}^{(2)}) = 1.0456$. Then the first equation of the normal form (3.2) associated with the critical point $(A_{gH}^{(1)}, C_{gH}^{(1)})$ is given by

$$\frac{dr}{d\tau} = r\,[v_0\,\mu + v_1\,r^2 + v_2\,r^4 + o(r^6)], \tag{3.4}$$

where $v_1 = 0$ and $v_2 = -0.1076 \times 10^{-3}$, called the second-order focus value, is obtained by using the Maple program [25]. Note that we now take the unfolding term from perturbing the parameter C as $C = C_-(A) + \mu$. Thus, we can perturb A from $A_{gH}^{(1)}$ to get $v_1 > 0$ such that $v_1 \ll |v_2|$, and then find $v_0\mu < 0$ satisfying $|v_0\mu| \ll v_1$. This gives two limit cycles bifurcating from the critical point $(A_{gH}^{(1)}, C_{gH}^{(1)})$. For this case, by perturbing C we have

$$v_0 = \frac{A(1057 + 1000A) - 57C(2 + C)}{2000A(1 + C)^2}.$$

To obtain $v_1 > 0$, we perturb $A = A_{gH}^{(1)}$ to $A^* = A_{gH}^{(1)} + 0.00005 = 0.01846287$, for which $C^* = C_-(A^*) = 0.11969100$ and so $v_0 = 0.11653286$. Now for the Hopf bifurcation associated with the critical values (A^*, C^*), we obtain $v_1 \approx 0.13257095 \times 10^{-4}$ and $v_2 \approx -0.10838198 \times 10^{-3}$. Further, we choose $\mu = -10^{-6} < 0$, i.e. C is decreased to pass through the critical point (A^*, C^*), yielding $v_0\mu \approx -0.11653286 \times 10^{-6}$. Finally, we obtain the normal form for this generalized Hopf bifurcation, up to 5th-order terms, in the form of

$$\frac{dr}{d\tau} = r\,[0.11653286 \times (-10^{-6}) + 0.13257096 \times 10^{-4} r^2 - 0.10838198 \times 10^{-3} r^4],$$

giving two real positive roots, $r_1 \approx 0.09763824$ and $r_2 \approx 0.33583483$, which approximate the amplitudes of the two limit cycles. Since $v_2 < 0$, the larger limit cycle is stable while the smaller limit cycle is unstable, and the equilibrium solution at this critical point is a stable focus.

In order to show the existence of the two limit cycles predicted above, first note that at the parameter values $B = D = 0.057$, $A = A^*$, $C = C^* - 10^{-6}$, the Jacobin matrix evaluated at the fixed point $E_- = (14.8376281, 0.1542552)$ has eigenvalues $-0.11918442 \times 10^{-6} \pm 0.08096077\,i$, confirming that this fixed point is a stable focus. But the convergence speed of nearby trajectories to this stable focus is very very slow. Next, we only need to show that there exists a stable limit cycle around this point since $v_2 < 0$, and expect that the convergence speed is also very slow. Therefore, there exists one unstable limit cycle between the stable focus and the stable limit cycle, as shown in Fig. 15. It can be seen from this figure that the analytical predictions, $r_1 \approx 0.10$ and $r_2 \approx 0.34$, give very good approximations for the amplitudes of the two simulated limit cycles, see Fig. 15(b).

Following the above procedure, we can also obtained two limit cycles bifurcating from the other critical point $A_{gH}^{(2)}$. We give the normal form for this case below without giving details for brevity. Taking $A = A^* = A_{gH}^{(2)} - 10^{-9}$, $C = C^* = C_-(A^*)$ yields

$$\frac{dr}{d\tau} = r\,[0.21278281 \times (-10^{-9}) + 0.93716102 \times 10^{-7} r^2 - 0.87730535 \times 10^{-5} r^4],$$

which has two real positive roots, $r_1 \approx 0.05721775$ and $r_2 \approx 0.08607204$, approximating the amplitudes of the two limit cycles bifurcating from this critical point $(A = A^*, C = C^*)$. Again, since $v_2 < 0$, the larger limit cycle is stable and the smaller limit cycle is unstable, and the equilibrium point is a stable focus. For simulation, we should take $A = A^* = 0.9210120422$, $C = C^* - 10^{-9} = 1.0456736673$, which yields the eigenvalues at the equilibrium point $E_- = (2.09166511, 0.88077509)$ as $-0.22645814 \times 10^{-5} \pm 0.62756454i$, and a similar figure to Fig. 15.

Similarly, we can obtain the five normal forms corresponding to the two critical points for the case of $B = 0.054 < D = 0.057$, one critical point for the case of $B = 0.054 < D = 0.087$, and two critical points for the case of $B = 0.060 > B = 0.057$. We first define the five cases followed by the corresponding five normal forms.

(a) $B = 0.054 < D = 0.057$ with $A = A_{gH}^{(1)} = 0.0149805591$

(b) $B = 0.054 < D = 0.057$ with $A = A_{gH}^{(2)} = 0.9454739030$

(c) $B = 0.054 < D = 0.087$ with $A = A_{gH} = 1.1708464105$

(d) $B = 0.060 > D = 0.057$ with $A = A_{gH}^{(1)} = 0.0213860900$

(e) $B = 0.060 > D = 0.057$ with $A = A_{gH}^{(2)} = 0.8963921091$

$$\text{(a)}\quad \frac{dr}{d\tau} = r[0.20278804 \times (-10^{-6}) + 0.89169329 \times 10^{-4} r^2 - 0.14900851 \times 10^{-2} r^4]$$
$$= 0 \implies r_1 = 0.04866092\ (\text{US}),\ r_2 = 0.23973712\ (\text{S});$$

$$\text{(b)}\quad \frac{dr}{d\tau} = r[0.21515679 \times (-10^{-10}) + 0.94588780 \times 10^{-8} r^2 - 0.22142107 \times 10^{-6} r^4]$$
$$= 0 \implies r_1 = 0.04909881\ (\text{US}),\ r_2 = 0.20076919\ (\text{S});$$

$$\text{(c)}\quad \frac{dr}{d\tau} = r[0.21521113 \times 10^{-9} - 0.93765555 \times 10^{-6} r^2 + 0.12177368 \times 10^{-3} r^4]$$

$$= 0 \quad \Longrightarrow \quad r_1 = 0.01538841 \text{ (S)}, \ r_2 = 0.08638971 \text{ (US)};$$

$$\text{(d)} \quad \frac{dr}{d\tau} = r[-0.04825749 \times 10^{-9} + 0.15893286 \times 10^{-6} r^2 - 0.58166912 \times 10^{-4} r^4]$$

$$= 0 \quad \Longrightarrow \quad r_1 = 0.01865320 \text{ (US)}, \ r_2 = 0.04883049 \text{ (S)};$$

$$\text{(e)} \quad \frac{dr}{d\tau} = r[-0.00236277 \times 10^{-8} + 0.92766615 \times 10^{-7} r^2 - 0.16897622 \times 10^{-4} r^4]$$

$$= 0 \quad \Longrightarrow \quad r_1 = 0.01636337 \text{ (US)}, \ r_2 = 0.07226452 \text{ (S)},$$

where US and S denote unstable limit cycle and stable limit cycle, respectively.

Summarizing the above results we have the following result.

Theorem 3.2. *For system (2.1), there always exists generalized Hopf bifurcation leading to two limit cycles bifurcating from the disease equilibrium* E_1, *for suitable positive parameter values. One of the two limit cycles is stable while the other is unstable.*

This theorem indicates that regardless whether $B < D$ or $B = D$ or $B > D$, the system can always exhibit complex dynamics including different types of bistability or even tristablity. More precisely, for Cases (a) and (b) (for which $B < D$), the disease-free equilibrium E_0 is a stable node, the disease equilibrium E_{1-} is a stable focus (another disease equilibrium E_{1+} is a saddle point), and there exist a stable limit cycle, as well as an unstable limit cycle between the stable limit cycle and the stable focus. This indeed shows tristability involving two stable equilibrium solutions and one stable periodic solution. Therefore, the first quadrant of the X–Y plane can be divided into three trapping regions, each corresponding to one of the three stable solutions. Case (c) (again $B < D$) shows a bistable situation, since for this case the disease equilibrium E_{1-} is an unstable focus, and there exist two limit cycles enclosing this unstable focus, with the inner one stable. The disease-free equilibrium E_0 is still a stable node. For Cases (d) and (e) (for which $B > D$) and the two cases when $B = D$, we can see that the disease-free equilibrium E_0 now becomes a saddle point (a degenerate saddle point for $B = D$) and there is only one disease equilibrium E_1 which is a stable focus. There are two limit cycles enclosing the stable focus and the outer one is stable. So this again shows a bistability but it involves one stable equilibrium solution and one stable periodic solution, different from the Hopf bifurcation case.

The above discussion implies that the real situation could be very complex, showing the co-existence of a stable disease-free equilibrium, stable disease equilibria, and even stable oscillating motion, all of which are possible depending upon the initial conditions. Moreover, note that the above seven cases (five cases plus two cases for $B = D$) are obtained for fixed parameter values of B and D. Hence, such phenomena are not uncommon, but quite rich if the parameters B and D are also allowed to be varied.

4. Bogdanov–Takens bifurcation

Finally, we consider possible Bogdanov–Takens (BT) bifurcations in system (2.1), characterized by a critical point with a double-zero eigenvalue. First, we have noticed that it is not possible to have a double-zero singularity at $E_0 = (\frac{1}{D}, 0)$ since it has eigenvalues $\xi_1 = -D$ and $\xi_2 = \frac{B}{D} - 1$, implying that it can have at most one zero eigenvalue when $B = D$. Secondly, for the case $B > D$, on the equilibrium solution E_{1-}, $\det(J) > 0$ which cannot have a double-zero eigenvalue. Thirdly, for the case $B = D$, again the equilibrium solution E_{1-} cannot have a double-zero critical point since when $\mathrm{Tr}(J) = 0$, $\det(J) = \frac{A^3(1+C)}{(A+C^2)^2} > 0$. Thus, the only possibility comes from the case $B < D$ on the equilibrium E_{1-}, which is observed from Figs. 1 and 2. In fact, it can be seen from (2.11) that $\det(J) = 0$ requires $\Delta = 0$, together with (2.14) to solve A and C to obtain the solutions for $B = \frac{27}{500}$, $D = \frac{57}{1000}$ as

$$\mathrm{BT}_1 = (B_1, D_1, A_1, C_1) = \left(\frac{27}{500}, \frac{57}{1000}, \frac{3078507}{206879500}, \frac{61731}{827518}\right) \tag{4.1}$$

which is marked as a circle in Fig. 1, and for $B = \frac{27}{500}, D = \frac{87}{1000}$ as

$$\mathrm{BT}_2 = (B_2, D_2, A_2, C_2) = \left(\frac{27}{500}, \frac{87}{1000}, \frac{118428267}{2237439500}, \frac{219501}{8949758}\right). \tag{4.2}$$

which is marked by a circle in Fig. 2. For a clear view, the zoomed areas around the two BT bifurcation points in Figs. 1 and 2 are shown in Fig. 16(a) and (b), respectively. As has been discussed in Section 3 that near the BT bifurcation points, the Hopf bifurcation is supercritical when $B = 0.054$, $D = 0.057$; while it is subcritical when $B = 0.054$, $D = 0.087$, which result in stable and unstable limit cycles, respectively. Thus, we will present the results for both cases.

4.1. Case $B = 0.054$, $D = 0.057$

We first consider the case $B = 0.054$, $D = 0.057$. We will derive the normal form associated with the BT_1 bifurcation, and then use the normal form to carry out bifurcation analysis. To achieve this, we introduce the following transformations:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{943}{57} \\ \frac{57}{1000} \end{pmatrix} + \begin{bmatrix} -\frac{1000}{943} & 0 \\ \frac{57}{943} & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} \frac{3078507}{206879500} \\ \frac{61731}{827518} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \tag{4.3}$$

into (2.1) and expanding the resulting system around the point $(u_1, u_2, \mu_1, \mu_2) = (0, 0, 0, 0)$ up to second order terms yields the system:

$$\begin{aligned} \frac{du_1}{d\tau} &= u_2 + f(u_1, u_2, \mu_1, \mu_2), \\ \frac{du_2}{d\tau} &= f(u_1, u_2, \mu_1, \mu_2) \\ &\equiv \frac{390174737}{1013000000}\mu_1 - \frac{413759}{9500000}\mu_2 - \frac{171196510081}{58491633000}\mu_1\mu_2 + \frac{171196510081}{517272748500}\mu_2^2 \\ &\quad + \left(\frac{315698117}{513084500}\mu_1 - \frac{22326849399}{450404618500}\mu_2\right)u_1 + \left(\frac{619297507081}{58491633000}\mu_1 - \frac{827518}{955259}\mu_2\right)u_2 \\ &\quad - \frac{1624500}{900809237}u_1^2 + \frac{3990000}{900809237}u_1u_2 + \frac{541500}{955259}u_2^2. \end{aligned} \tag{4.4}$$

Next, we apply the near-identity nonlinear transformation (up to second order), given by

$$\begin{aligned} u_1 &= y_1 + \frac{2685896921}{19000000}\bar{\beta}_1 + \left(\frac{2774136507391115169073729}{22529079489117000000000}\bar{\beta}_1 - \frac{5624557927591883}{413759000000000}\bar{\beta}_2\right)y_1 \\ &\quad + \frac{257312250}{900809237}y_1^2 + \frac{19238822633}{2865777000}y_1y_2 + \frac{6868613670961379723}{72599684000000000}y_2^2 \\ u_2 &= y_2 - \bar{\beta}_1 - \frac{739146260609762232437}{2514827202000000000}\bar{\beta}_1^2 + \frac{6014732664591883}{413759000000000}\bar{\beta}_1\bar{\beta}_2 \\ &\quad + \left(\frac{17614322633}{2865777000}\bar{\beta}_1 - \frac{57}{1000}\bar{\beta}_2\right)y_1 + \left(\frac{739146260609762232437}{2514827202000000000}\bar{\beta}_1 - \frac{6014732664591883}{413759000000000}\bar{\beta}_2\right)y_2 \\ &\quad + \frac{1624500}{900809237}y_1^2 + \frac{541500}{955259}y_1y_2 + \frac{17614322633}{2865777000}y_2^2, \end{aligned} \tag{4.5}$$

and the parameterization,

$$\mu_1 = -\frac{55157609919}{171196510081}\bar{\beta}_1 + \frac{55157609919}{171196510081}\bar{\beta}_2 \quad \mu_2 = -\frac{8836997403671}{342393020162}\bar{\beta}_1 + \frac{975576403671}{342393020162}\bar{\beta}_2,$$

to (4.4) to obtain the normal form:

$$\begin{aligned} \frac{dy_1}{d\tau} &= y_2, \\ \frac{dy_2}{d\tau} &= \bar{\beta}_1 + \alpha\,\bar{\beta}_2\,y_1 + \bar{\beta}_2\,y_2 - a_1\,y_1^2 + a_2\,y_1y_2, \end{aligned} \tag{4.6}$$

where

$$\alpha = \frac{57}{1000}, \quad a_1 = \frac{1624500}{900809237}, \quad a_2 = \frac{741000}{900809237}.$$

In order to further simplify system (4.6), we introduce the following scalings with an additional shift:

$$y_1 = m_1\,x_1 + y_{10}, \quad y_2 = m_2\,x_2, \quad \tau_1 = m_3\,\tau,$$

into (4.6) to obtain

$$\begin{aligned}\frac{dx_1}{d\tau_1} &= x_2,\\ \frac{dx_2}{d\tau_1} &= \beta_1 + \beta_2 x_2 - x_1^2 + x_1 x_2.\end{aligned} \tag{4.7}$$

Here,

$$\begin{aligned}y_{10} &= \frac{\alpha\bar{\beta}_2}{2a_1} = \frac{900809237}{57000000}\bar{\beta}_2,\\ m_1 &= \frac{a_1}{a_2^2} = \frac{900809237}{338000}, \quad m_2 = \frac{a_1^2}{a_2^3} = \frac{51346126509}{8788000}, \quad m_3 = \frac{a_1}{a_2} = \frac{57}{26},\\ \beta_1 &= \frac{a_2^4(4a_1\bar{\beta}_1 + \alpha^2\bar{\beta}_2^2)}{4a_1^4} = \frac{228488000}{2926729211013}\bar{\beta}_1 + \frac{28561}{812250000}\bar{\beta}_2^2,\\ \beta_2 &= \frac{a_2(2a_1 + a_2\alpha)}{2a_1^2}\,\bar{\beta}_2 = \frac{13169}{28500}\bar{\beta}_2.\end{aligned} \tag{4.8}$$

Thus, the relation between the original perturbation parameters (μ_1, μ_2) and the new perturbation parameters (β_1, β_2) is given by

$$\begin{aligned}\beta_1 &= \frac{11817370799}{392997970672875}\,\mu_1 - \frac{23634741598}{6950981876155875}\,\mu_2\\ &\quad + \frac{4889543524423441\,(465105126509\,\mu_1 - 5806064202\,\mu_2)^2}{247115847924796669728417225000000000000000000},\\ \beta_2 &= \frac{2501738616510085303}{1551818245500000000}\,\mu_1 - \frac{5448792271}{270750000000}\,\mu_2,\end{aligned} \tag{4.9}$$

It should be noted that due to the large values of m_1 and m_2, very small values of (x_1, x_2) can result in very large values of (y_1, y_2) and so (u_1, u_2), which are perturbations from the BT critical point (A_T, C_T). Therefore, we should take small values of x_1 and x_2 when solving system (4.7). Also note in (4.9) that the coefficients of μ_1 and μ_2 are small, so we should choose very small values for the perturbation parameters β_1 and β_2. Moreover, since in general μ_2 should take negative values (see Fig. 16(a)), we will show in the following that β_1 must take positive values.

Now, we use the normal form (4.7) to analyze the BT bifurcation. First, we note that in almost all existing articles or books, the unfolding terms (i.e. the terms with the coefficient β_1 or β_2) are usually taken as in a generic form with no direct relation to the original physical system parameters, which may cause difficulty in bifurcation analysis when solving practical problems. Here, we involve perturbation parameters in the nonlinear transformation to obtain the explicit unfolding terms (in terms of β_1 and β_2), which have a direct relation to the original system parameters A and C, and thus facilitate a realistic dynamical study. It is also noted that the standard normal form for BT bifurcations, given in the existing literature, either in the form of [23],

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= \beta_1 + \beta_2 x_2 + x_1^2 + x_1 x_2,\end{aligned} \tag{4.10}$$

or in the form of [26]

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= \beta_1 + \beta_2 x_1 + x_1^2 - x_1 x_2.\end{aligned} \tag{4.11}$$

Our system (4.7) is in the line of (4.10) though there is sign difference for the term x_1^2. We use a negative sign in our system is to keep the sign of the original system, which does not affect the analysis.

The two equilibrium solutions of (4.7) are given by

$$\mathrm{E}_\pm = (x_{1\pm}, 0), \quad \text{where} \quad x_{1\pm} = \pm\sqrt{\beta_1}, \quad (\beta_1 \geq 0). \tag{4.12}$$

The condition $\beta_1 \geq 0$ can be actually determined from (2.4) and (4.9) as

$$\mathrm{H}_1 \approx \frac{3063807}{206879500}\mu_1 - \frac{173223}{103439750}\mu_2 \approx \frac{9632559468266793081}{19558174097693764}\,\beta_1,$$

indicating that $\mathrm{H}_1 \geq 0$ must yield $\beta_1 \geq 0$.

To find the stability of the two equilibrium solutions, we use the Jacobian of (4.7) to obtain the characteristic polynomial $\lambda^2 - \mathrm{Tr}\,\lambda + \det$, where

$$\mathrm{Tr} = \beta_2 + x_1 \quad \text{and} \quad \det = 2x_1.$$

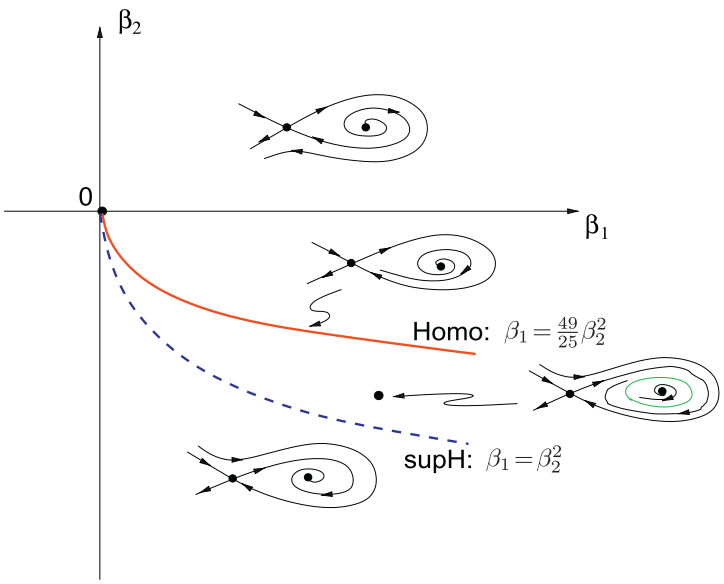

Fig. 17. Bifurcation set and phase portraits of system (4.7).

Defining $\Delta = \mathrm{Tr}^2 - 4\,\mathrm{det}$, we have

$$\mathrm{Tr}^+ = \beta_2 + \mathrm{x}_{1+} = \beta_2 + \sqrt{\beta_1}, \quad \mathrm{det}^+ = 2\sqrt{\beta_1} > 0, \tag{4.13}$$

implying that the equilibrium E_+: $(x_{1+}, 0)$ is either a focus or node, which is stable (unstable) when $\beta_2 + \sqrt{\beta_1} < 0$ (>0). Similarly, for the equilibrium E_-: $(x_{1-}, 0)$ we have

$$\mathrm{Tr}^- = \beta_2 + \mathrm{x}_{1-} = \beta_2 - \sqrt{\beta_1}, \quad \mathrm{det}^- = 2\,\mathrm{x}_{1-} = -2\sqrt{\beta_1} < 0, \tag{4.14}$$

indicating that E_- is always a saddle point. The bifurcation set (only for $\beta_1 \geq 0$) and corresponding phase portraits are shown in Fig. 17 (the phase portrait for $\beta_1 < 0$ is trivial). Note that the Hopf bifurcation near the critical point (denoted by the dashed blue curve in Fig. 17) is obtained from $\mathrm{Tr}^+ = 0$ as

$$\beta_1 = \beta_2^2 \quad (\beta_2 < 0). \tag{4.15}$$

There is another curve in Fig. 17, shown in red, which denotes the bifurcation of homoclinic loop (see, for example [23,26]).

Before we derive the equation for the bifurcation of the homoclinic loop, we consider the Hopf bifurcation which occurs from the dashed blue curve. The Hopf critical point on this curve can be defined as $\beta_{2\mathrm{H}} = -\sqrt{\beta_1}$, and then introducing the transformation: $x_1 = \sqrt{\beta_1} + \tilde{x}_1$, $x_2 = \omega_c \tilde{x}_2$ into (4.7) results in the system:

$$\begin{aligned}
\frac{d\tilde{x}_1}{d\tau_1} &= \omega_c \tilde{x}_2 \equiv \tilde{f}(\tilde{x}_1, \tilde{x}_2),\\
\frac{d\tilde{x}_1}{d\tau_1} &= -\omega_c \tilde{x}_1 - \frac{1}{\omega_c}\tilde{x}_1^2 + \tilde{x}_1\tilde{x}_2 \equiv \tilde{g}(\tilde{x}_1, \tilde{x}_2),
\end{aligned}$$

where $\omega_c = (2\sqrt{\beta_1})^{1/2}$. Thus, the first focus value v_1 is given by

$$v_1 = -\frac{1}{16\omega_c^2}(-\tilde{g}_{\tilde{x}_1\tilde{x}_1}\tilde{g}_{\tilde{x}_1\tilde{x}_2}) = -\frac{1}{16\omega_c} \times \frac{2}{\omega_c} = -\frac{1}{8\omega_c^2} < 0,$$

indicating that the Hopf bifurcation is supercritical, and bifurcating limit cycles are stable, as shown in Fig. 17 (see the ellipse in green). The Hopf bifurcation near the BT critical point is not surprising since the original system does have Hopf bifurcations which occur from the blue curve, as shown in Fig. 16. In fact, as discussed in Section 3, we can similarly use the original system to show that the Hopf bifurcations from the blue curve (see Fig. 16) are indeed supercritical, which agrees with the conclusion obtained above, and so the bifurcating limit cycles are stable.

Next, we consider homoclinic loops which may bifurcate near the BT critical point. Here, we apply the technique of rescaling, as used in [23] to find the approximation equation for the homoclinic curve. Set

$$x_1 = \varepsilon^2 w_1, \quad x_2 = \varepsilon^3 w_2, \quad \beta_1 = \varepsilon^4 \nu_1, \quad \beta_2 = \varepsilon^2 \nu_2, \quad (0 \leq \varepsilon \ll 1), \tag{4.16}$$

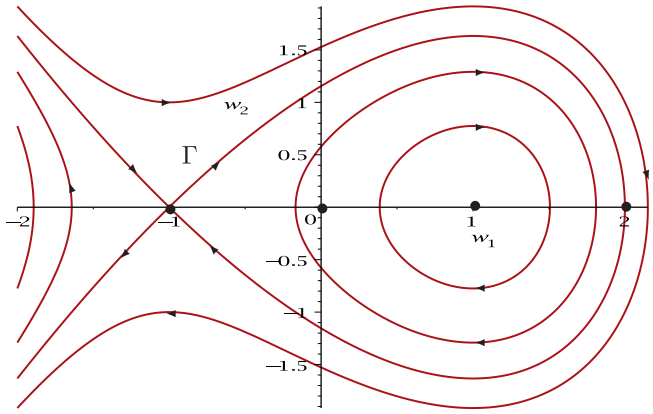

Fig. 18. The phase portrait of (4.18) with $\nu_1 = 1$, showing a homoclinic loop Γ.

and rescale time $t = \varepsilon\tau_1$, so that (4.7) can be rewritten (up to ε order) as

$$\begin{aligned} \frac{dw_1}{dt} &= w_2, \\ \frac{dw_2}{dt} &= \nu_1 + \varepsilon\nu_2 w_2 + \varepsilon w_1 w_2 - w_1^2. \end{aligned} \tag{4.17}$$

Now, letting $\varepsilon = 0$ in (4.17) yields a Hamiltonian system:

$$\begin{aligned} \frac{dw_1}{dt} &= w_2, \\ \frac{dw_2}{dt} &= \nu_1 - w_1^2. \end{aligned} \tag{4.18}$$

with Hamiltonian

$$H(w_1, w_2) = -\nu_1 w_1 + \frac{1}{3} w_1^3 + \frac{1}{2} w_2^2. \tag{4.19}$$

Taking $\nu_1 = 1$, which corresponds to $\beta_1 \geq 0$, we have two fixed points: $(w_1, w_2) = (\pm 1, 0)$, with (1, 0) being a center and $(-1, 0)$ a saddle point, as shown in Fig. 18.

The solution on the saddle loop Γ based at the point $(w_1, w_2) = (2, 0)$ is given by

$$(w_1(t), w_2(t)) = (3\,\mathrm{sech}^2 t - 1,\ 3\sqrt{2}\,\mathrm{sech}^2 t\ \tanh t). \tag{4.20}$$

Thus, the first-order Melnikov function $M(t_0)$ on the vector filed $\varepsilon\, q(w_1, w_2)\frac{\partial}{\partial w_2}$, where $q(w_1, w_2) = \nu_2 w_2 + w_1 w_2$, is independent of time, and can be calculated as [27]

$$\begin{aligned} M(\nu_2) &= \oint_\Gamma H_{w_2}\, q(w_1, w_2)\, dt = \int_{-\infty}^{\infty} w_2(t)\,[\nu_2 w_2(t) + w_1(t) w_2(t)]\, dt \\ &= 18\,\nu_2 \int_{-\infty}^{\infty} \mathrm{sech}^4 t \tanh^2 t\, dt + 18 \int_{-\infty}^{\infty} (3\,\mathrm{sech}^2 t - 1)\,\mathrm{sech}^4 t \tanh^2 t\, dt. \end{aligned}$$

Then, solving $M \equiv 0$ for the saddle connection yields

$$\begin{aligned} \nu_2 &\approx -\frac{\int_{-\infty}^{\infty} (3\,\mathrm{sech}^2 t - 1)\,\mathrm{sech}^4 t \tanh^2 t\, dt}{\int_{-\infty}^{\infty} \mathrm{sech}^4 t \tanh^2 t\, dt} \\ &= -\frac{\int_{-\infty}^{\infty} (2\tanh^2 t - 5\tanh^4 t + 3\tanh^6 t)\,\mathrm{sech}^2 t\, dt}{\int_{-\infty}^{\infty} (\tanh^2 t - \tanh^4 t)\,\mathrm{sech}^2 t\, dt} \\ &= = -\frac{5}{7}, \end{aligned}$$

which implies $\beta_2 \leq 0$ as expected. Here, the formula:

$$\int_{-\infty}^{\infty} \tanh^k t\,\mathrm{sech}^2 t\, dt = \frac{\tanh^{k+1} t}{k+1}\bigg|_{-\infty}^{\infty} = \begin{cases} \dfrac{2}{k+1}, & k = \text{even}, \\ 0, & k = \text{odd}, \end{cases}$$

has been used. Finally, noticing $\nu_1 = 1$, and $\beta_1 = \varepsilon^4$, $\beta_2 = \varepsilon^2\,\nu_2$, we obtain the approximate bifurcation curve for the homoclinic loop as

$$\text{Homo:}\quad \beta_1 = \frac{49}{25}\,\beta_2^2, \quad \beta_2 \leq 0. \tag{4.21}$$

The true bifurcation curve is tangent to the semi-parabola at $\beta_1 = \beta_2 = 0$. Combining this with Eq. (4.15) for Hopf bifurcation, we indeed see that a second bifurcation curve, denoted as 'Homo', is located above the Hopf bifurcation curve and tangent to it (and to $\beta_1 = 0$) at $(\beta_1, \beta_2) = (0, 0)$, and the phase portrait on this bifurcation set has a saddle loop, as shown in Fig. 17. The sign taken by the Melnikov function M for $\beta_1 < \frac{49}{25}\beta_2^2$ (or $> \frac{49}{25}\beta_2^2$, respectively) gives the relative position of the stable and unstable manifolds (separators of the saddle). More precisely, one can use (4.7) to show that the eigenvalues of the system evaluated on the saddle point E_- are

$$\lambda_{1,2} = \frac{1}{2}\left[\beta_2 - \sqrt{\beta_1} \pm \sqrt{\left(\beta_2 - \sqrt{\beta_1}\right)^2 + 8\sqrt{\beta_1}}\right],$$

which yields the "saddle quantity", given by

$$\lambda_1 + \lambda_2 = \beta_2 - \sqrt{\beta_1} < 0 \quad (\beta_2 < 0).$$

This implies that the homoclinic orbit is stable (an ω-limit set, attracting the nearby points). Alternatively, we may apply the approach given in [27] to obtain an expansion of the Melnikov function near the homoclinic orbit, which can be used to prove the stability of the homoclinic orbit. Further, it can be shown (see [23]) that in the region between the Hopf bifurcation curve 'supH' and the Homoclinic bifurcation curve 'Homo' (see Fig. 17) the system has a unique attracting limit cycle for each pair of parameter values (β_1, β_2).

To demonstrate the bifurcation phenomena discussed above, we show simulations using the original system (2.1), rather than the normal form Eq. (4.7), which gives a more realistic observation. We take seven sets of perturbations on the parameters A and C near the BT_1 critical point (see Fig. 16(a)) as $A = A_1 + \mu_1$, $C = C_1 + \mu_2$, where A_1 and C_1 are given in (4.1). These seven sets of perturbations denote seven points in the bifurcation diagram (see Fig. 16) on a same vertical line (see the green line in Fig. 16(a)) with the same coordinate $A = A_1 - 0.000001$, and different coordinates $C = C_1 + \mu_2$ with μ_2 given from top to the bottom as follows:

$$\mu_2 = -0.0000094,\quad -0.000098,\quad -0.0000106,\quad -0.00003, -0.0000414239,\quad -0.0000875,\quad -0.0001.$$

It is noted that the equilibrium E_{1-} is a stable focus at the top and the bottom points, but is an unstable focus at the other five points. Here, we have found an interesting phenomenon that since the Hopf bifurcation curve has a turning point and all nearby points can lead to stable limit cycles in the region where the equilibrium E_{1-} is an unstable focus, there exist two homoclinic loops when one goes through the five points along the vertical line starting with the second point from the top. However, the above normal form theory for the BT bifurcation and the result given in Fig. 17 only show one homoclinic loop. This is not surprising since the normal form for the BT_1 bifurcation is only applicable for the study of dynamics around the BT_1 point and thus it only predicts the top homoclinic loop. Due to the perturbations being very small, the convergence of the simulating trajectories is very slow. Moreover, the direction of the trajectories near the saddle point is hard to distinguish. Therefore, in order to give a clear view, we, based on the simulating phase portraits which have been rotated by a angle of $\frac{\pi}{55}$, present seven schematic diagrams with exaggerated convergence speed and the part near the saddle point. Since the simulations for the top and bottom points are similar, we will only present one figure for these two points (see Fig. 19(a)). Of course, they are different quantitatively and the simulation for the bottom point is much clearer than that of the top one. The Fig. 19(b)–(f) correspond to the other five points from top to the bottom. The relation between the original coordinates (X, Y) and the new coordinates $(\bar{X}, \bar{Y})$ shown in Fig. 19 is given by

$$\bar{X} = \cos\left(\frac{\pi}{55}\right)X - \sin\left(\frac{\pi}{55}\right)Y, \quad \bar{Y} = \sin\left(\frac{\pi}{55}\right)X + \cos\left(\frac{\pi}{55}\right)Y.$$

In the next example for $B = 0.054$, $D = 0.087$, we will see true simulating phase portraits, which clearly show the Hopf bifurcation and homoclinic bifurcation.

4.2. Case $B = 0.054$, $D = 0.087$

Now we turn to study the case $B = 0.054$, $D = 0.087$. As we have discussed, a particular difference between this case and previous case is that now the Hopf bifurcation near the BT_2 critical point is subcritical, and thus the bifurcating limit cycles are unstable. This difference can cause dramatically different meanings in the biological explanation of this phenomenon.

Since the solution procedure is similar to the previous case, we will skip some detailed steps and only present the main results in the following. Using a series of transformations, similar to (4.3), (4.5) and (4.8), we obtain the following normal form:

$$\begin{aligned}\frac{dx_1}{d\tau_1} &= x_2,\\ \frac{dx_2}{d\tau_1} &= \beta_1 + \beta_2 x_2 - x_1^2 - x_1 x_2.\end{aligned} \tag{4.22}$$

The solution formulae are the same as that given in (4.12). Again, we can similarly argue that $\beta_1 \geq 0$, and as a matter of fact, $H_1 \geq 0$ implies $\beta_1 \geq 0$. The linear stability of these two equilibrium solutions show that $(x_{1-}, 0)$ is a saddle point,

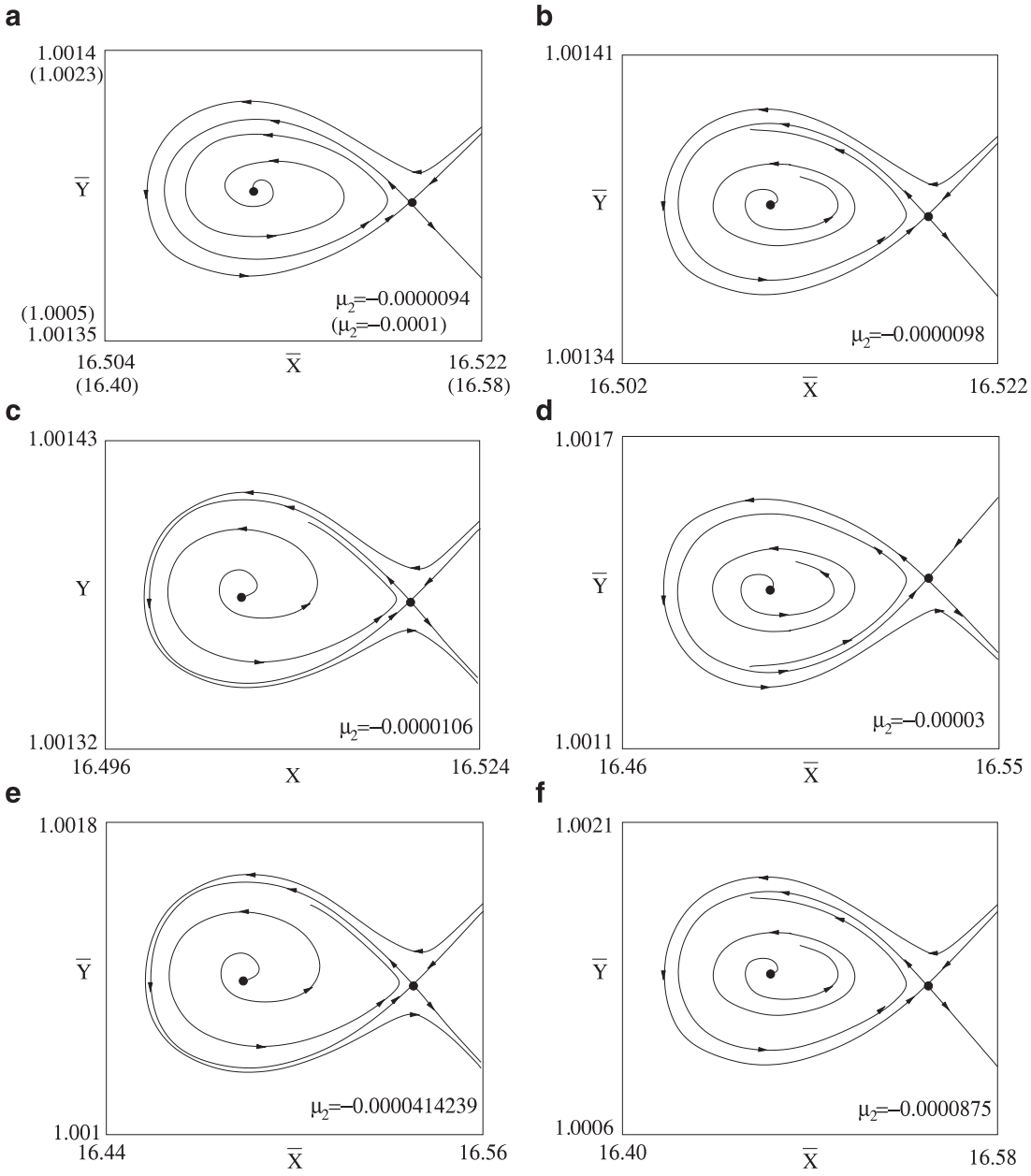

Fig. 19. Simulations of system (2.1) when $B = 0.054$, $D = 0.057$, $A = 0.01487968$ for (a) $C = 0.07458837$ or $C = 0.07449777$, showing stable focus E_{1-}, (b) $C = 0.07458797$, showing unstable focus E_{1-} and a stable limit cycle, (c) $C = 0.07458717$, showing unstable focus E_{1-} and a stable homoclinic loop, (d) $C = 0.07456777$, showing unstable focus E_{1-} and a stable limit cycle, (e) $C = 0.0745563461$, showing unstable focus E_{1-} and a stable homoclinic loop, (f) $C = 0.07449777$, showing unstable focus E_{1-} and a stable limit cycle.

while $(x_{1+}, 0)$ is stable (unstable) if $\beta_2 - \sqrt{\beta_1} < 0$ (>0). The Hopf bifurcation near the BT_2 critical point is determined as

$$\beta_1 = \beta_2^2 \quad (\beta_2 \geq 0), \tag{4.23}$$

and the bifurcation is subcritical, since the first focus value can be obtained as $v_1 = \frac{1}{8\omega_c^2} > 0$. Similarly we can obtain the homoclinic bifurcation which occurs from the curve:

$$\text{Homo:} \quad \beta_1 = \frac{49}{25}\beta_2^2, \quad \beta_2 \geq 0. \tag{4.24}$$

The bifurcation set and corresponding phase portraits are depicted in Fig. 20, which is quite different from the case $B = 0.054$, $D = 0.057$ (see Fig. 17). Simulations based on the original system (2.1) for this case are shown in Fig. 21, where the perturbation (μ_1, μ_2), on the parameters A and C, take the following values:

$$(0.004, 0.0085), \quad (0.004, 0.00807), \quad (0.004, 0.0073813), \quad (0.004, 0.007),$$

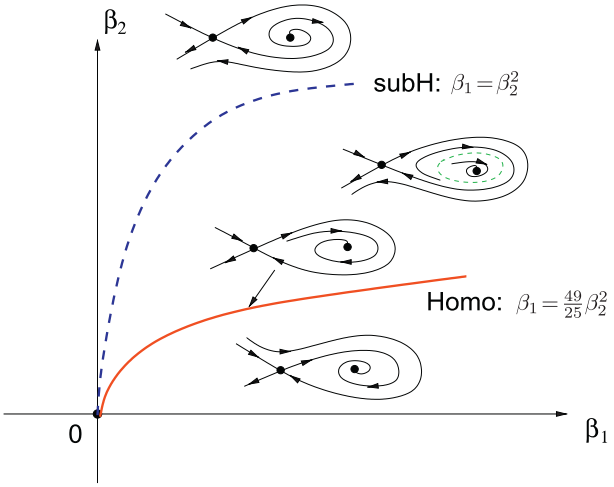

Fig. 20. Bifurcation set and phase portraits of system (4.22).

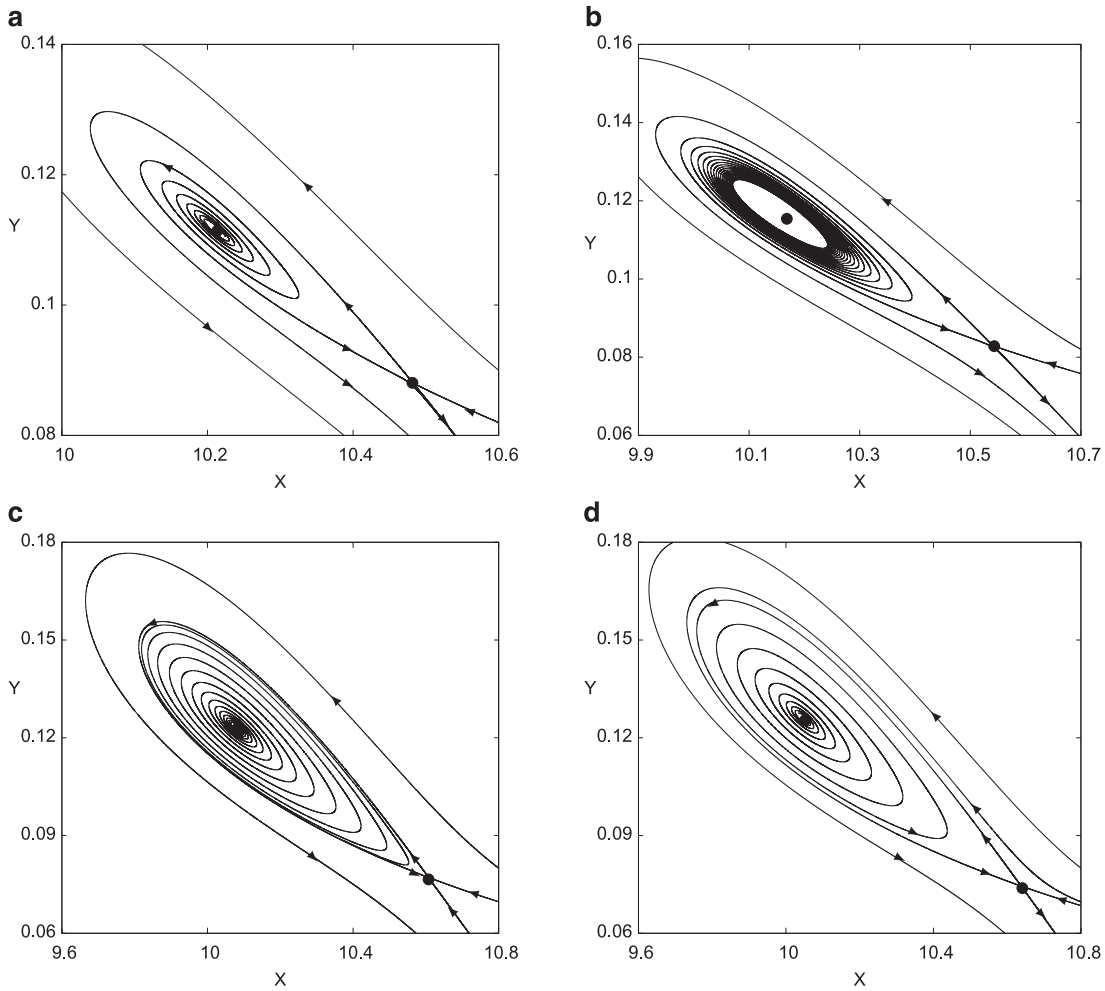

Fig. 21. Simulations of system (2.1) when $B = 0.054$, $D = 0.087$ for (a) $A = 0.0569302656$, $C = 0.0330259146$, showing an unstable focus E_{1-} with one trajectory divergent to the saddle point E_{1+}, (b) $A = 0.0569302656$, $C = 0.0325959146$, showing stable focus E_{1-} enclosed by an unstable limit cycle, (c) $A = 0.0569302656$, $C = 0.0319072146$, showing a homoclinic loop enclosing a stable focus, and (d) $A = 0.0569302656$, $C = 0.0315259146$, showing convergence of the trajectory starting from the saddle point E_{1+} to the stable focus E_{1-}.

which represent four points on the same vertical green line in the bifurcation diagram, shown in Fig. 16(b).

It is seen from Fig. 21(a) that the phase portrait for the first perturbation, corresponding to a point above the Hopf bifurcation curve (see the blue curve in Fig. 20), shows a unstable focus E_{1-} and there exists one trajectory starting from the saddle point E_{1+} and converging to this focus as $\tau \to -\infty$.

Figure 21(b) shows the phase portrait for the second perturbation, corresponding to a point between the Hopf bifurcation curve (the blue curve in Fig. 20) and the homoclinic bifurcation curve (the red curve in Fig. 20), shows an unstable limit cycle (see the green curve in Fig. 20) and trajectories starting near this limit cycle either converge to the stable focus E_{1-} or to the stable node E_0 (which is not shown in Figs. 20 and 21) as $\tau \to +\infty$.

Figure 21(c) shows a homoclinic loop under the third perturbation, corresponding to a point on the homoclinic bifurcation curve, which encloses the stable focus E_{1-}, and all trajectories inside this homoclinic loop coverage to the focus as $\tau \to +\infty$. We can use a very similar approach as that used for system (4.7) to consider the homoclinic orbit of system (4.22). In fact, it is easy to use (4.22) to obtain the eigenvalues of the system evaluated on the saddle point E_- as

$$\lambda_{1,2} = \frac{1}{2}\left[\beta_2 + \sqrt{\beta_1} \pm \sqrt{\left(\beta_2 + \sqrt{\beta_1}\right)^2 + 8\sqrt{\beta_1}}\right],$$

from which we obtain the saddle quantity,

$$\lambda_1 + \lambda_2 = \beta_2 + \sqrt{\beta_1} > 0 \quad (\beta_2 > 0),$$

implying that the homoclinic orbit is unstable (an α-limit set).

Finally, Fig. 21(d) shows a phase portrait for the fourth perturbation, corresponding to a point below the homoclinic bifurcation curve, which encloses the stable focus E_{1-}. It is seen from Fig. 21 that the saddle connection before and after the homoclinic loop (Fig. 21(c)) change the way to connect the focus or the limit cycle. Note that unlike the bifurcation shown in Fig. 16(a) where there are two homoclinic loops which occur from the green line, here there is only one homoclinic loop since no more Hopf bifurcation happens when the parameter C is decreased to cross the Hopf critical line along the green line (see Fig. 16(b)).

Summarizing the results obtained in this section we have the following theorem.

Theorem 4.1. *For system (2.1), when $B < D$ and $H_1 > 0$, there always exists Bogdanov–Takens bifurcation, which occurs from the precritical disease bifurcation solution, leading to homoclinic bifurcation near a Hopf bifurcation, with homoclinic loop being either stable or unstable.*

4.3. A new mechanism for generating blips

A detailed study for a 4-dimensional system has been given in [13,14], shows a mechanism for generating the blips phenomenon, and four conditions are proposed in a hypothesis, which guarantee the existence of blips. In [13,14], blips are also shown to exist in two 3-dimensional models as well as in the 2-dimensional model (2.1). An important condition for the existence of blips is Hopf bifurcation, which is the source of oscillation. Very recently, another mechanism has been identified in [15], which is also related to Hopf bifurcation. These two mechanisms have a common property that both of them generate oscillations with large changes in both amplitude and frequency, and they both appear on the post-critical disease bifurcation solution. It has also been noted that these two mechanisms have a fundamental difference: the former guarantees blips to occur near a transcritical bifurcation point; while the later yields blips far away from a transcritical bifurcation point, which are not guaranteed. The second mechanism needs further investigation.

In order to discuss a new mechanism of generating blips, in the following we list Hypothesis 1 from [13,14], and propose a second Hypothesis based on the results obtained in [15].

Hypothesis 1. [13,14] The following four conditions are needed for an in-host infection model to generate viral blips:

(i) there exist at least two equilibrium solutions;
(ii) there exists a transcritical bifurcation at an intersection of the two equilibrium solutions;
(iii) there is a Hopf bifurcation which occurs from one of the equilibrium solutions; and
(iv) large oscillations (or, more generally, global, persistent motions) can occur near the transcritical critical point.

Hypothesis 2. [15] The following four conditions are needed for an in-host infection model to generate viral blips: conditions (i)–(iii) are the same as that given in Hypothesis 1; and

(iv) large oscillations (or, more generally, global, persistent motions) can occur far away from the transcritical and Hopf critical points.

We use the bifurcation diagrams shown in Fig. 22(a) and (b) (which are Fig. 3.3(a) and (b) in [14]) to illustrate Hypothesis 1, and the bifurcation diagram in Fig. 22(c) (which is Fig. 3.1(a) in [15]) to explain Hypothesis 2, where R and A are state variables, B and α are parameters. E_0 and E_1 denote the disease-free and disease equilibrium solutions. The green lines indicate where the blip-like oscillations occur. It is clear from Fig. 22(a) and (b) that the blips appear near the transcritical point, and may or may not appear near the Hopf critical point, where both E_0 and E_1 are unstable, illustrating

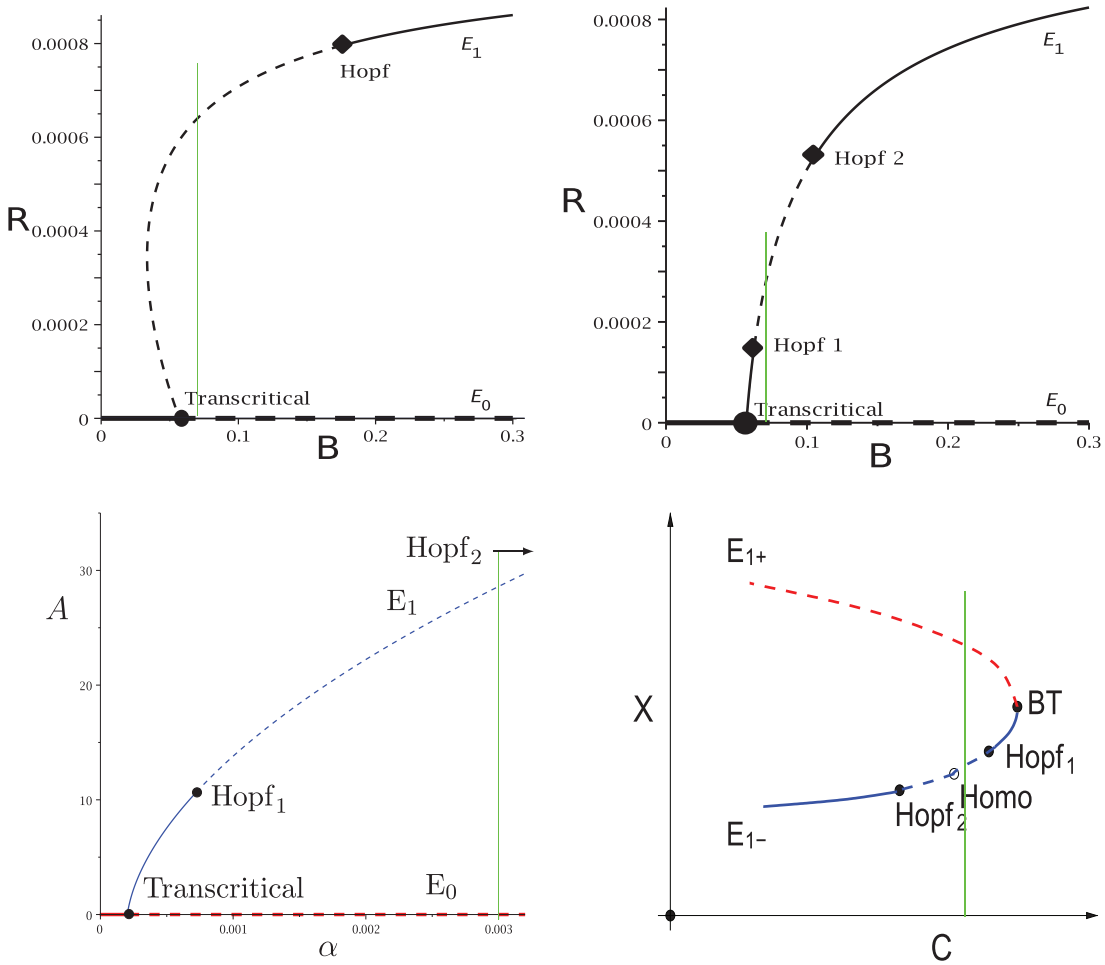

Fig. 22. Bifurcation diagrams illustrating Hypotheses: (a) and (b) for Hypothesis 1, (c) for Hypothesis 2 and (c) Hypothesis 3.

condition (iv) in Hypothesis 1. Fig. 22(c) (where the second Hopf critical point '$Hopf_2$' is outside the figure) shows that the blips occur far away from the transcritical and Hopf bifurcation points.

Through the study given in this section on the BT bifurcation, we have found a third mechanism for generating blips, due to the BT bifurcation, explained as follows. First of all, note that the trajectory starting from a point on the homoclinic loop will reach the saddle point either as $\tau \to +\infty$ or $\tau \to -\infty$. Therefore, it can be seen from Fig. 17 that near the homoclinic bifurcation curve, for certain parameter values, the bifurcating stable limit cycles can be large close to the saddle separators and thus such a stable limit cycle will move extremely slowly near the saddle point but will move fast when it is away from the saddle point – giving rise to the blips phenomenon. A schematic bifurcation diagram for the case, which is depicted in Fig. 19 when $B = 0.054, D = 0.057, A = 0.01487968,$ is shown in Fig. 22(d). Also note from Figs. 20 and 21 that when the limit cycle inside the saddle separators is unstable, the trajectories starting near the unstable limit cycle may converge to the stable focus E_{1-}, or to the stable node E_0 but will take very long time since it will go through a route close to the saddle point though not generating blips in this case.

The big difference between the first two mechanisms and the new mechanism is that the first two mechanisms result in very large oscillations in both amplitude and frequency, while the new mechanism only causes significant changes in frequency, but very little variation in the amplitude. The biological implication of the new mechanisms is interesting and may explain some real situations, namely, in some situations a patient may not feel obvious changes nor will measurable changes in disease progression be apparent, but nonetheless the patient may be experiencing recurrent disease without any significant observation. In other situations, neither the infected individual nor the clinician may be able to detect whether the infection has been cured, since complete recovery may take an extremely long time. In both cases, the patient is in an uncertain situation. To describe these scenarios, we have the following hypothesis.

Hypothesis 3. The following four conditions are needed for an in-host infection model to generate viral blips or to take an extremely long time to recover (converge to the disease-free equilibrium): conditions (i)–(iii) are the same as that given in Hypothesis 1; and

(iv) there exists Bogdanov–Takens bifurcation, leading to homoclinic loops near a Hopf bifurcation, which may yield blips with very small changes in amplitude, or extremely slow convergence to the disease-free equilibrium.

5. Conclusion and discussion

In this paper, we have given a detailed dynamical study of a 2-dimensional disease model, which can be used not only for in-host disease modeling, but also for epidemiologic modeling. We have shown that when the reproduction number, $R_0 = \frac{B}{D}$, is varied near $R_0 = 1$, the system exhibits rich dynamical behaviors, including equilibrium solutions which exchange their stability at the transcritical point $R_0 = 1$. Both Hopf and generalized Hopf bifurcations can occur regardless whether $R_0 < 1$ or $R_0 \geq 1$, which lead to bistability or even tristability. In particular, our study has indicated that when $R_0 < 1$, the system can have Bogdanov–Takens bifurcation leading to more complex dynamical behavior such as homoclinic orbit bifurcation. This special bifurcation may provide a new scenario/mechanism for generating recurrence or the viral blips phenomenon, summarized in Hypothesis 3.

Hypothesis 3 is completely different from Hypotheses 1 and 2, and may provide an explanation for interesting clinical phenomena. In many disease models, the concept of R_0 is straightforward, i.e. if $R_0 < 1$, the disease cannot invade or persist, and the disease only exists for $R_0 > 1$. In reality, disease dynamics are more complex, and our model indeed reflects this complexity. Hypothesis 3 allows for the possibility that even if control or therapy reduces R_0 below one, a disease may persist indefinitely with low level oscillations, or may die out, but with an extremely slow time course of decay. The possibility of disease persistence when $R_0 < 1$ is a feature of backward bifurcation [28–30], an issue which we are investigating for this model and related disease models as well [31].

Mathematically, the most interesting dynamical behavior of our model is the Bogdanov–Takens bifurcation leading to homoclinic loops, which in turn provides a new mechanism for explaining a very different blips phenomenon. In particular, this phenomenon does not have obvious changes in the amplitude of the oscillating motion. This can only happen when $B < D$ (i.e. $R_0 < 1$). However, this condition, $B < D$, is not enough, the additional condition $H_1 \geq 0$, which guarantees the existence of disease equilibrium, E_1, must also be satisfied. Intuitively, if $B < D$, then the epidemic cannot get started because near the disease-free equilibrium, E_0, the behavior of the model is similar to that studied in [3], and thus no oscillation can occur with $R_0 < 1$. However, $H_1 \geq 0$, as mentioned in Remark 2.1, implies that the contact rate A exceeds its threshold such that the infected cells, denoted by Y, are sufficiently infectious such that the epidemic can sustain itself once started even if $B < D$. Therefore, this leads, after getting over an initial threshold, to potential bistable equilibrium solutions and even more complex dynamical behavior.

One question which is not discussed in this paper but is closely related to our study is the so-called fast-slow motion arising from singular perturbation problems which contain "fast" and "slow" manifolds (e.g., see [32] and references therein). The oscillations appearing in a singular perturbed system are specific periodic solutions, which are similar to the blips motion studied in [13–15]. Then, a question arises: what result can we expect if we use the singular perturbation method to analyze our system, compared to that obtained using the simple method introduce in [13]? In fact, the 2-dimensional model (2.1) studied in this paper is reduced from a 3-dimensional model by applying the quasi-steady state assumption [13]. The 3-dimensional model indeed contains a 1-dimensional fast manifold and 2-dimensional slow manifold. In [13], we used our method to perform a detailed parametric study to show the blips phenomenon in a very large parameter region, confirmed by simulating the 3-dimensional dynamical system. Thus, we believe that applying the singular perturbation method to the 3-dimensional system will unlikely show more complex dynamical behavior than what we obtained using the simple approach. The 2-dimensional model (2.1), on the other hand, describes dynamical behavior of the system on the slow manifold. For some parameter values, we have shown [13] that the system exhibits blips oscillation. If we want to apply the singular perturbation technique to study the 2-dimensional model (2.1) to find the blips (fast-slow motion), we must carefully choose some parameter scaling to make an ε to appear in this equation so that one equation becomes a fast manifold and the other becomes slow manifold. But this process is quite technique and requires some particular condition to be satisfied. Certainly, what kind of problems which can be analyzed by the singular perturbation method but not our approach are remain open, and will be our future research project. Also, we would like to extend the study in this work for the 2-dimensional (A, C) parameter space to the 4-dimensional (A, B, C, D) parameter space in order to get a more generic picture, and find out what is the maximal number of limit cycles which can bifurcate from a Hopf critical point due to the increase of the number of parameters.

Finally, we would mention that the ideas and methodologies presented in this paper can be used to analyze other types of in-host disease models as well as epidemiologic models. We hope that they can also be generalized to study functional differential systems (e.g. with time delays), or even other physical or engineering systems which exhibit similar "blips-like" phenomenon.

Acknowledgment

Insightful comments from the anonymous referees are greatly appreciated. This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC no. R2686A02 and no. 238389-RGPIN).

References

[1] Anderson RM, May RM. Infectious diseases of humans: Dynamics and control. Oxford University Press; 1992.
[2] Nowak MA, May RM. Virus dynamics. New York: Oxford University Press; 2000.
[3] Korobeinikov A, Maini PK. Non-linear incidence and stability of infectious disease models. Mathematical Medicine and Biology 2005;22(2005):113–28.
[4] Liu W, Levin S, Iwasa Y. Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. Journal of Mathematical Biology 1986;23:187–204.
[5] Liu W, Hethcote HW, Levin SA. Dynamical behavior of epidemiological models with nonlinear incidence rates. Journal of Mathematical Biology 1987;25:359–80.
[6] Hethcote HW, Lewis MA, van den DP. An epidemiological model with a delay and a nonlinear incidence rate. Journal of Mathematical Biology 1989;27:49–64.
[7] Hethcote HW, van den DP. Some epidemiological models with nonlinear incidence. Journal of Mathematical Biology 1991;29:271–87.
[8] Derrick WR, Driessche P. A disease transmission model in a nonconstant population. Journal of Mathematical Biology 1993;31:495–512.
[9] Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. Mathematical Biosciences 1995;125:155–64.
[10] Briggs CJ, Godfray HCJ. The dynamics of insect-pathogen interactions in stage-structured populations. American Naturalist 1995;145:855–87.
[11] van Gaalen RD, Wahl LM. Reconciling conflicting clinical studies of antioxidant supplementation as HIV therapy: a mathematical approach. BMC Public Health 2009;9:1–18.
[12] Alexander HK, Wahl LM. Self-tolerance and autoimmunity in a regulatory T cell model. Bulletin of mathematical biology 2011;73:33–71.
[13] Zhang W, Wahl LM, Yu P. Conditions for transient viremia in deterministic in-host models: viral blips need no exogenous trigger. SIAM Journal on Applied Mathematics 2013;73:853–81.
[14] Zhang W, Wahl LM, Yu P. Viral blips may not need a trigger: how transient viremia can arise in deterministic in-host models. SIAM Review 2014a;56:127–55.
[15] Zhang W, Wahl LM, Yu P. Modelling and analysis of recurrent autoimmune disease. SIAM Journal on Applied Mathematics 2014b;74:1998–2025.
[16] Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. New England Journal of Medicine 2000;343:1430–8.
[17] Girschick HJ, Zimmer C, Klaus G, Darge K, Dick A, Morbach H. Chronic recurrent multifocal osteomyelitis: what is it and how should it be treated? Nature Clinical Practice Rheumatology 2007;3:733–8.
[18] Iyer RS, Thapa MM, Chew FS. Chronic recurrent multifocal osteomyelitis: review. American Journal of Roentgenology 2011;196:S87–91.
[19] Munro DD. Recurrent subacute discoid lupus erythematosus. Proceedings of the Royal Society of Medicine 1963;56:78–9.
[20] Fergusson DM, Horwood LJ, Shannon FT. Early solid feeding and recurrent childhood eczema: a 10-year longitudinal study. Pediatrics 1990;86:541–6.
[21] Farber EM, Mullen RH, Jacobs AH, Nall L. Infantile psoriasis: a follow-up study. Pediatric Dermatology 1986;3:237–43.
[22] Ruan S, Wang W. Dynamical behavior of an epidemic model with a nonlinear incidence rate. Journal of Differential Equations 2003;188:135–63.
[23] Guckenheimer J, Holmes P. Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. 4th ed. New York: Springer-Verlag; 1993.
[24] Krakauer DC, Komarova NL. Levels of selection in positive-strand virus dynamics. Journal of Evolutionary Biology 2003;16:64–73.
[25] Yu P. Computation of normal forms via a perturbation technique. Journal of Sound and Vibration 1998;211:19–38.
[26] Kuznetsov Yuri A. Elements of applied bifurcation theory. 2nd ed. New York: Springer-Verlag; 1998.
[27] Han M, Yu P. Normal forms, Melnikov functions and bifurcations of limit cycles. New York: Sringer-Verlag; 2012.
[28] Dushoff J, Huang W, Castillo-Chavez C. Backwards bifurcations and catastrophe in simple models of fatal diseases. Journal of Mathematical Biology 1998;36:227–48.
[29] Blayneh KW, Gumel AB, Lenhart S, Clayton T. Backward bifurcation and optimal control in transmission dynamics of west nile virus. Bulletin of Mathematical Biology 2010;72:1006–28.
[30] Arino J, McCluskey CC, van den DP. Global results for an epidemic model with vaccination that exhibits backward bifurcation. SIAM Journal on Applied Mathematics 2003;64:260–76.
[31] Zhang W, Wahl LM, Yu P. Backward bifurcations, turning points and rich dynamics in simple disease models. Journal of Mathematical Biology (revised version has been submitted).
[32] Krupa M, Szmolyan P. Relaxiation oscillation and canard explosion. Journal of Differentical Equations 2001;174:312–68.