Complexity

Dynamical Analysis of Biological Systems

Lead Guest Editor: George V. Popescu
Guest Editors: Eberhard O. Voit and Constantin Udriste

Dynamical Analysis of Biological Systems

Complexity

Dynamical Analysis of Biological Systems

Lead Guest Editor: George V. Popescu
Guest Editors: Eberhard O. Voit and Constantin Udriste

Contents

Hindawi
Complexity
Volume 2020, Article ID 6402703, 11 pages
https://doi.org/10.1155/2020/6402703

Research Article

Dynamical Regulation of mRNA Distribution by Cross-Talking Signaling Pathways

Chunjuan Zhu,[1,2] Guosheng Han,[3] and Feng Jiao [1]

[1]*Center for Applied Mathematics, Guangzhou University, Guangzhou 510006, China*
[2]*Modern Business and Management Department, Guangdong Construction Polytechnic, Guangzhou 510440, China*
[3]*School of Mathematics and Computational Science, Xiangtan University, Hunan, China*

Correspondence should be addressed to Feng Jiao; jf1704@163.com

Received 30 March 2020; Revised 9 May 2020; Accepted 15 May 2020; Published 17 June 2020

Guest Editor: George V. Popescu

Gene transcription is a random process in single cells manifested by the observed distribution of mRNA copy numbers in homogeneous cell populations. A central question is to understand how mRNA distribution is modulated under environmental changes. In this work, we initiate a theoretical study on mRNA distribution dynamics for the stochastic transcription model that involves cross-talking signaling pathways to direct gene activation in response to external signals. We first express the distribution in mathematical dynamical formulas under both moderate and high transcriptional upregulations. In each scenario, our further numerical examples display an observed dynamical transition type among three distribution modes for stress genes in yeast. In particular, the intermediate bimodal stage sustains within a certain length of early time and lasts much longer than that generated by the single pathway. This shows the general and robust bimodal transcription regulated by the cross-talk of signaling pathways.

1. Introduction

Recent single-cell measurements have generated massive data on the histogram of mRNA copy numbers for the target gene in homogeneous cell populations [1, 2]. This provides a good approximation for data fitting by the probability mass function $P_m(t)$, the probability that there are exactly m mRNA molecules of the gene of our concern at time t in one cell [3, 4]. The distribution profile of $P_m(t)$ contains a panoramic information for distinct cellular phenotypes [5, 6]. The commonly observed modes are the decaying distribution that $P_m(t)$ decreases in m, the unimodal distribution that $P_m(t)$ peaks uniquely at some $m > 0$, and the bimodal distribution that $P_m(t)$ takes exactly two peaks with the first one at $m = 0$ and the other one at some $m > 0$. A decaying or unimodal distribution suggests a phenotypic homogeneity, while a bimodal distribution supports a binary process that steers cells into subpopulations with distinct cell identities [1, 7, 8].

It has been a central question to understand how varying external signals influence the profile of $P_m(t)$ for inducible genes [1, 2]. As suggested by the prevailing two-state model [1, 9], it may involve intrinsic mechanisms that regulate a random switching between gene-off (inactive) and gene-on (active) states [6, 10, 11]. As shown in the following equation:

$$\text{gene off} \underset{\gamma}{\overset{\lambda}{\rightleftarrows}} \text{gene on} \xrightarrow{\nu} \text{mRNA} \xrightarrow{\delta} \varnothing, \tag{1}$$

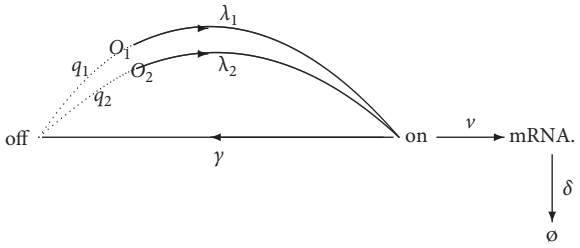
the dwell times in the off and on states are independent and exponentially distributed with the activation rate $\lambda > 0$ and inactivation rate $\gamma > 0$, respectively. In an active gene, RNA polymerase can bind to the promoter and traverses the template DNA strand to direct mRNA synthesis. The waiting time for the birth of a new mRNA molecule and its lifetime before death are independent and exponentially distributed with the synthesis rate $\nu > 0$ and degradation rate $\delta > 0$, respectively.

The two-state model (1) has emerged as a standard framework for quantifying the deviation of the mRNA level in individual cells from the mean, through calculating the noise (variance over mean squared), fano factor (variance over mean), and probability distribution $P_m(t)$ of random mRNA copy numbers [1, 2, 8]. These indexes have been expressed as analytical formulas for the two-state model or

its more elaborated extension regarding a larger degree of biological realism such as chromatin remodeling, mRNA maturation, and cell division [6, 12–14]. Together with experimental data, those formulas can be reversed to the variation of system parameters and, in turn, uncover a large spectrum of regulation modes that cells utilize under different cellular environments [2, 4, 14–16].

On the contrary, the two-state model (1) implicitly assumes that the gene activation from the off state to the on state is directed by a single signaling pathway. Such assumption may not be justified for a large class of inducible genes which maintain a low transcription level induced by the spontaneous basal pathway under normal cellular growth conditions [17, 18], while upregulating the transcription through specific signaling pathways in response to acute external changes [19, 20]. Also, the activation of other genes that are involved in stem cell renewal, development, and immunity is often mediated by two signal transduction pathways [21–23]. In all these cases, the downstream specific transcription factors (TFs) in one pathway compete with the other TFs in another pathway for binding at their shared promoter sites to direct the formation of the intermediate complex [19, 24]. Therefore, the two competitive cross-talking pathways could be generated to activate the target gene.

By integrating competitive cross-talking pathways into the gene activation process [24, 25], the two-state model (1) can be generalized as the following equation:



(2)

The transcription of the gene can be either induced by the first pathway, or alternatively, the second pathway. The sojourn times in the two pathways are independent and exponentially distributed, with the induction strength $\lambda_1 > 0$ for the first pathway and the strength $\lambda_2 > 0$ for the other, satisfying

$$0 < \lambda_1 \leq \lambda_2 < \infty. \tag{3}$$

We rename the gene-off state as O_1 if it is transformed to the gene-on state by the first pathway, and as O_2 otherwise. Then, the pathway selection probabilities, denoted by

$$\begin{aligned} q_1 &= \text{Prob}\,(O = O_1), \\ q_2 &= \text{Prob}\,(O = O_2), \end{aligned} \tag{4}$$

satisfy

$$\begin{aligned} &0 < q_1, q_2 < 1, \\ &q_1 + q_2 = 1. \end{aligned} \tag{5}$$

Also, the gene inactivation from the on state to the off state and the birth and the death of mRNA molecules are separately controlled by the first-order kinetic rates γ, ν, and δ, as modeled in the two-state model (1).

When $t \longrightarrow \infty$, the exact form for the stationary mass function $\lim_{t\longrightarrow\infty} P_m(t)$ was stated in [25]. The subsequent numerical examples showed that cross-talking pathways are more likely to generate bimodal distribution compared to the single pathway with $q_1 = 1$ or $\lambda_1 = \lambda_2$. This naturally gives rise to an interesting question of whether the cross-talking regulation still maintains the robust bimodal distribution as time develops. When time t is finite, however, the analytical formulas of $P_m(t)$ have remained elusive due to its intrinsic complexity. Only a few papers considered the case of a single pathway and expressed $P_m(t)$ or the corresponding generating function in the form of integrals [12], infinite series [26, 27], and hypergeometric functions [6, 28] under certain parameter regions. We shall derive $P_m(t)$ generated by cross-talking pathways in simple mathematical formulas in Section 2, and then discuss their dynamical profiles and implications in Section 3. The main conclusion and its discussion are given in the last section.

2. Analytical Formulas for Dynamical Distribution

In this section, we have endeavored to calculate $P_m(t)$ by solving the system of master equations. The standard approach is to transform it into a system of first-order partial differential equations through the method of generating functions [6, 12]. The analytical form of $P_m(t)$ has been found to be the solutions for several special types of third-order ordinary differential equations. Solving their corresponding initial value problems and then extracting $P_m(t)$ are in general formidable and pose a major obstacle in the study.

We have successfully derived many exact forms of $P_m(t)$ within a large range of system parameters. In particular, our first result assumes that the weaker signaling pathway is more frequently selected. In this case, the upregulation of the target gene may not be efficient due to the exposure of cells to mild external signals or the intrinsic mechanism that avoids overexuberant transcription of the target gene [17, 20, 29].

Theorem 1. *If $\gamma = 2\delta$, $q_1 > q_2$, and $\lambda_2 = \lambda_1 + \delta/(q_1 - q_2)$, then the probability mass function $P_m(t)$ in the cross-talking pathway model can be expressed as*

$$\begin{aligned} P_0(t) = {} & q_1 e^{-\lambda_1 t} + q_2 e^{-\lambda_2 t} + \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1} \int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right) \\ & \cdot e^{\nu\left(e^{-\delta t} - e^{-\delta s}\right)/\delta} \mathrm{d}s + \frac{q_1\lambda_1 + q_2\lambda_2}{\lambda_2 - \lambda_1} \\ & \cdot \int_0^t \left[(\delta - \lambda_1)e^{-\lambda_1 s} - (\delta - \lambda_2)e^{-\lambda_2 s}\right] e^{\delta(s-t)} e^{\nu\left(e^{-\delta t} - e^{-\delta s}\right)/\delta} \mathrm{d}s, \end{aligned} \tag{6}$$

and for $m = 1, 2, \ldots$,

$$P_m(t) = \frac{\nu^m(q_1\lambda_1 + q_2\lambda_2)}{\delta^m m!(\lambda_2 - \lambda_1)}\left[\frac{\lambda_1\lambda_2}{q_1\lambda_1 + q_2\lambda_2}\int_0^t\left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)\right.$$
$$\cdot\left(e^{-\delta s} - e^{-\delta t}\right)^m e^{\nu\left(e^{-\delta t} - e^{-\delta s}\right)/\delta}\mathrm{d}s$$
$$+\int_0^t\left[(\delta - \lambda_1)e^{-\lambda_1 s} - (\delta - \lambda_2)e^{-\lambda_2 s}\right]$$
$$\left.\cdot e^{\delta(s-t)}\left(e^{-\delta s} - e^{-\delta t}\right)^m e^{\nu\left(e^{-\delta t} - e^{-\delta s}\right)/\delta}\mathrm{d}s\right], \tag{7}$$

or alternatively, for $m = 0, 1, \ldots,$

$$P_m(t) = \frac{\nu^m}{\delta^m m!(\lambda_2 - \lambda_1)}\left[\int_0^t\left[(q_1\lambda_1 + q_2\lambda_2)\left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)\right.\right.$$
$$\cdot e^{-\delta t} - \lambda_2 e0^{-(\lambda_1+\delta)s} + \lambda_1 e^{-(\lambda_2+\delta)s}\Big]$$
$$\times\left[m\delta + \nu\left(e^{-\delta t} - e^{-\delta s}\right)\right]\left(e^{-\delta s} - e^{-\delta t}\right)^{m-1}$$
$$\left.\times e^{\nu\left(e^{-\delta t} - e^{-\delta s}\right)/\delta}\mathrm{d}s\right] + \frac{\nu^m\left(1 - e^{-\delta t}\right)^m}{\delta^m m! e^{\nu\left(1 - e^{-\delta t}\right)/\delta}}. \tag{8}$$

Proof. Let $P_{m,j}(t)$, $j = e, 1, 2$, be the respective probabilities of m mRNA molecules at time t that the gene is residing in the on state and the off state with the jth pathway being selected. Then, the total probability mass function is

$$P_m(t) = P_{m,1}(t) + P_{m,2}(t) + P_{m,e}(t). \tag{9}$$

Following the standard procedure in the theory of stochastic processes [13, 24], we find that the three partial mass functions in the cross-talking pathway model satisfy the following system of master equations:

$$P'_{m,1}(t) = q_1\gamma P_{m,e}(t) - (m\delta + \lambda_1)P_{m,1}(t) + (m+1)\delta P_{m+1,1}(t), \tag{10}$$

$$P'_{m,2}(t) = q_2\gamma P_{m,e}(t) - (m\delta + \lambda_2)P_{m,2}(t) + (m+1)\delta P_{m+1,2}(t), \tag{11}$$

$$P'_{m,e}(t) = \lambda_1 P_{m,1}(t) + \lambda_2 P_{m,2}(t) - (\nu + m\delta + \gamma)P_{m,e}(t) + (m+1)\delta P_{m+1,e}(t) + \nu P_{m-1,e}(t), \tag{12}$$

where, by convention, we set $P_{-1,e}(t) = 0$. Without loss of generality, we assume that the gene is inactive and the number of transcripts is zero at $t = 0$ with the initial conditions:

$$\begin{aligned} &P_{0,1}(0) = q_1, \\ &P_{0,2}(0) = q_2, \\ &P_{0,e}(0) = P_{m,i}(0) = 0, \quad \text{for } i = 1, 2, e,\ m \geq 1. \end{aligned} \tag{13}$$

Following the standard procedure [6, 12], we introduce the probability generating functions

$$V_i(z,t) = \sum_{m=0}^{\infty}(z+1)^m P_{m,i}(t), \quad i = 1, 2, e, \tag{14}$$

to transform the initial value problem for the master equations (10)–(12) into the problem for the system of first-order partial differential equations:

$$\frac{\partial V_1}{\partial t}(z,t) = -\lambda_1 V_1(z,t) + q_1\gamma V_e(z,t) - \delta z\frac{\partial V_1}{\partial z}(z,t), \tag{15}$$

$$\frac{\partial V_2}{\partial t}(z,t) = -\lambda_2 V_2(z,t) + q_2\gamma V_e(z,t) - \delta z\frac{\partial V_2}{\partial z}(z,t), \tag{16}$$

$$\frac{\partial V_e}{\partial t}(z,t) = \lambda_1 V_1(z,t) + \lambda_2 V_2(z,t) - \gamma V_e(z,t) + \nu z V_e(z,t) - \delta z\frac{\partial V_e}{\partial z}(z,t), \tag{17}$$

$$\begin{aligned} &V_1(z,0) = q_1, \\ &V_2(z,0) = q_2, \\ &V_e(z,0) = 0. \end{aligned} \tag{18}$$

If the initial value problem (15)–(18) can be solved analytically, then we may obtain $P_m(t)$ by adding the three solutions together and then applying the conversion formula:

$$P_m(t) = \frac{1}{m!}\left.\frac{\partial^z V(z,t)}{\partial z^m}\right|_{z=-1}, \quad \text{where } V(z,t) = V_1(z,t) + V_2(z,t) + V_e(z,t),\ m \geq 0. \tag{19}$$

We continue from there and transform (15)–(17) into ordinary differential equations through the method of characteristics [12, 28, 30]. Let z_0 be a parameter. Let

$$\begin{aligned} &u_i(t) = V_i\left(z_0 e^{\delta t}, t\right), \quad \text{for } i = 1, 2, e, \\ &u(t) = u_1(t) + u_2(t) + u_e(t), \end{aligned} \tag{20}$$

be the restriction of V_1, V_2, V_e, and V on the characteristic curve $z = z_0 e^{\delta t}$. Then,

$$\begin{aligned} u'_1(t) &= -\lambda_1 u_1(t) + q_1\gamma u_e(t), \\ u'_2(t) &= -\lambda_2 u_2(t) + q_2\gamma u_e(t), \end{aligned} \tag{21}$$

$$u'_e(t) = \lambda_1 u_1(t) + \lambda_2 u_2(t) - \gamma u_e(t) + \nu z_0 e^{\delta t} u_e(t), \tag{22}$$

$$\begin{aligned} u_1(0) &= q_1, \\ u_2(0) &= q_2, \\ u_e(0) &= 0. \end{aligned} \tag{23}$$

By introducing nondimensionalized system parameters,

$$\begin{aligned} \overline{\lambda}_1 &= \frac{\lambda_1}{\delta}, \\ \overline{\lambda}_2 &= \frac{\lambda_2}{\delta}, \\ \overline{\gamma} &= \frac{\gamma}{\delta}, \\ \overline{\nu} &= \frac{\nu}{\delta}, \end{aligned} \tag{24}$$

and from the further transformation,

$$\begin{aligned} w_i(x) &= u_i(t), \quad \text{for } i = 1, 2, e, \\ w(x) &= u(t), \quad \text{where } x = \overline{\nu} z_0 e^{\delta t}, \end{aligned} \tag{25}$$

we find $u_i'(t) = \delta x w_i'(x)$, and transform (21)–(23) into

$$\begin{aligned} x w_1'(x) &= -\overline{\lambda}_1 w_1(x) + q_1 \overline{\gamma} w_e(x), \\ x w_2'(x) &= -\overline{\lambda}_2 w_2(x) + q_2 \overline{\gamma} w_e(x), \end{aligned} \tag{26}$$

$$x w_e'(x) = \overline{\lambda}_1 w_1(x) + \overline{\lambda}_2 w_2(x) - \overline{\gamma} w_e(x) + x w_e(x), \tag{27}$$

$$\begin{aligned} w_1(\overline{\nu} z_0) &= q_1, \\ w_2(\overline{\nu} z_0) &= q_2, \\ w_e(\overline{\nu} z_0) &= 0. \end{aligned} \tag{28}$$

It is interesting to note that equations (26) and (27) are indeed identical with equations (17)–(19) in [25]. Therefore, by introducing two real numbers $\overline{\alpha} > \overline{\beta} > 0$ determined by algebra equations [24],

$$\begin{aligned} \overline{\alpha} + \overline{\beta} &= \overline{\lambda}_1 + \overline{\lambda}_2 + \overline{\gamma}, \\ \overline{\alpha}\overline{\beta} &= \overline{\lambda}_1 \overline{\lambda}_2 + q_1 \overline{\lambda}_2 \overline{\gamma} + q_2 \overline{\lambda}_1 \overline{\gamma}, \end{aligned} \tag{29}$$

the same calculation verifying (21) in [25] shows that $w(x)$ is the unique solution of the initial value problem:

$$\begin{aligned} \mathscr{L}_{\overline{\alpha},\overline{\beta}}^{\overline{\lambda}_1,\overline{\lambda}_2}(w) &= 0, \\ w(\overline{\nu} z_0) &= 1, \\ w'(\overline{\nu} z_0) &= 0, \\ w''(\overline{\nu} z_0) &= \frac{q_1 \overline{\lambda}_1 + q_2 \overline{\lambda}_2}{\overline{\nu} z_0}, \end{aligned} \tag{30}$$

for the linear operator $\mathscr{L}$ given by

$$\begin{aligned} \mathscr{L}_{c,d}^{a,b}(f) = x^2 f'''(x) + x(1 - x + c + d) f''(x) \\ + (c\,d - x(a + b + 1)) f\prime(x) - ab f(x), \end{aligned} \tag{31}$$

with real constants a, b, c, and d and a smooth function $f = f(x)$ of x.

To proceed, we define by

$$y(x) = x^2 w''(x) + x\left(\overline{\lambda}_1 + \overline{\lambda}_2 + 1\right) w'(x) + \overline{\lambda}_1 \overline{\lambda}_2 w(x), \tag{32}$$

and utilize notations (24) and (29) and the assumption of theorem to verify

$$\begin{aligned} \overline{\beta} &= \overline{\lambda}_1 + 1, \\ \overline{\alpha} &= \overline{\lambda}_2 + 1. \end{aligned} \tag{33}$$

We then reformulate $\mathscr{L}_{\overline{\alpha},\overline{\beta}}^{\overline{\lambda}_1,\overline{\lambda}_2}(w) = 0$ of (30) in the form

$$\begin{aligned} &\left[x^2 w'''(x) + x(1 + \overline{\alpha} + \overline{\beta}) w''(x) + \overline{\alpha}\overline{\beta} w'(x)\right] \\ &\quad = \left[x^2 w''(x) + x\left(\overline{\lambda}_1 + \overline{\lambda}_2 + 1\right) w'(x) + \overline{\lambda}_1 \overline{\lambda}_2 w(x)\right]. \end{aligned} \tag{34}$$

The substitution of (32) and (33) into the above equation gives $y'(x) = y(x)$, which readily converts (30) to the simple problem:

$$\begin{aligned} y'(x) &= y(x), \\ y(\overline{\nu} z_0) &= \overline{\lambda}_1 \overline{\lambda}_2 + \overline{\nu} z_0 \left(q_1 \overline{\lambda}_1 + q_2 \overline{\lambda}_2\right). \end{aligned} \tag{35}$$

The unique solution of (35) is $y = y(\overline{\nu} z_0)\exp(x - \overline{\nu} z_0)$. We replace its left side by (32) and the right side by the initial condition in (35). This gives an inhomogeneous equation:

$$\begin{aligned} &x^2 w''(x) + x\left(\overline{\lambda}_1 + \overline{\lambda}_2 + 1\right) w'(x) + \overline{\lambda}_1 \overline{\lambda}_2 w(x) \\ &\quad = \left[\overline{\lambda}_1 \overline{\lambda}_2 + \overline{\nu} z_0 \left(q_1 \overline{\lambda}_1 + q_2 \overline{\lambda}_2\right)\right] e^{x - \overline{\nu} z_0}, \end{aligned} \tag{36}$$

subject to the initial condition given in (30)

$$\begin{aligned} w(\overline{\nu} z_0) &= 1, \\ w'(\overline{\nu} z_0) &= 0. \end{aligned} \tag{37}$$

Taking into account that the homogeneous counterpart of (36) is Euler's equation which possesses two independent solutions $x^{-\overline{\lambda}_1}$ and $x^{-\overline{\lambda}_2}$, their linear combination, adding a particular solution of (36), constitutes the general solution of (36). Following the standard method of undetermined coefficients in the theory of ordinary differential equation [31], we find that one particular solution takes the following form:

$$\overline{w}(x) = \frac{\overline{\lambda}_1 \overline{\lambda}_2 + \overline{\nu} z_0 \left(q_1 \overline{\lambda}_1 + q_2 \overline{\lambda}_2\right)}{\overline{\lambda}_2 - \overline{\lambda}_1} \int_{\overline{\nu} z_0}^{x} \left(\frac{s^{\overline{\lambda}_1 - 1}}{x^{\overline{\lambda}_1}} - \frac{s^{\overline{\lambda}_2 - 1}}{x^{\overline{\lambda}_2}} \right) e^{s - \overline{\nu} z_0} \mathrm{d}s. \tag{38}$$

By fixing the coefficients in the linear combination through the initial condition (37), we obtain the unique solution of (36) and (37) in the following form:

$$w(x) = C_1 x^{-\overline{\lambda}_1} + C_2 x^{-\overline{\lambda}_2} + \overline{w}(x),$$
$$C_1 = \frac{\lambda_2 (\overline{\nu} z_0)^{\overline{\lambda}_1}}{\lambda_2 - \lambda_1}, \tag{39}$$
$$C_2 = -\frac{\lambda_1 \overline{\nu} z_0)^{\overline{\lambda}_2}}{\lambda_2 - \lambda_1}.$$

Now, the transformations $u(t) = V(z_0 e^{\delta t}, t)$ defined in (20) and $u(t) = w(\overline{\nu} z_0 e^{\delta t})$ defined in (25) convert (39) back to

$$V\left(z_0 e^{\delta t}, t\right) = \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} - \frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} + \frac{\overline{\lambda}_1 \overline{\lambda}_2}{\overline{\lambda}_2 - \overline{\lambda}_1} \int_{\overline{\nu} z_0}^{\overline{\nu} z_0 e^{\delta t}} \left(\frac{s^{\overline{\lambda}_1 - 1} e^{-\lambda_1 t}}{(\overline{\nu} z_0)^{\overline{\lambda}_1}} - \frac{s^{\overline{\lambda}_2 - 1} e^{-\lambda_2 t}}{(\overline{\nu} z_0)^{\overline{\lambda}_2}} \right) e^{s - \overline{\nu} z_0} \mathrm{d}s$$
$$+ \frac{\overline{\nu} z_0 \left(q_1 \overline{\lambda}_1 + q_2 \overline{\lambda}_2 \right)}{\overline{\lambda}_2 - \overline{\lambda}_1} \int_{\overline{\nu} z_0}^{\overline{\nu} z_0 e^{\delta t}} \left(\frac{s^{\overline{\lambda}_1 - 1} e^{-\lambda_1 t}}{(\overline{\nu} z_0)^{\overline{\lambda}_1}} - \frac{s^{\overline{\lambda}_2 - 1} e^{-\lambda_2 t}}{(\overline{\nu} z_0)^{\overline{\lambda}_2}} \right) e^{s - \overline{\nu} z_0} \mathrm{d}s. \tag{40}$$

By making the change of variable $s = \overline{\nu} z_0 \exp(\delta \tau)$ in the integral, we rewrite V as

$$V\left(z_0 e^{\delta t}, t\right) = \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} - \frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} + \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \int_0^t \left(e^{\lambda_1 (\tau - t)} - e^{\lambda_2 (\tau - t)} \right) e^{\overline{\nu} z_0 \left(e^{\delta \tau} - 1 \right)} \mathrm{d}\tau + \frac{\nu z_0 \left(q_1 \lambda_1 + q_2 \lambda_2 \right)}{\lambda_2 - \lambda_1} \int_0^t \left(e^{\lambda_1 (\tau - t)} - e^{\lambda_2 (\tau - t)} \right) e^{\overline{\nu} z_0 \left(e^{\delta \tau} - 1 \right)} \mathrm{d}\tau. \tag{41}$$

In turn, by using conversion $z_0 = z e^{-\delta t}$ and then letting $s = t - \tau$ in the integral, we obtain

$$V(z,t) = \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} - \frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} + \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s} \right) e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s + \frac{\nu z e^{-\delta t} \left(q_1 \lambda_1 + q_2 \lambda_2 \right)}{\lambda_2 - \lambda_1} \int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s} \right) e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s. \tag{42}$$

We divide the rest of the proof into two parts. We derive (6) and (7) in the first part and (8) in the second part.

Derivation of (6) and (7): we first note that the second integral in (42) can be expressed in an equivalent form through integration by parts:

$$\begin{aligned} & \nu z e^{-\delta t} \int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s} \right) e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s \\ &= e^{-\delta t} \int_0^t \left(e^{(\delta - \lambda_2) s} - e^{(\delta - \lambda_1) s} d e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \right) \\ &= e^{-\lambda_2 t} - e^{-\lambda_1 t} - \int_0^t \left[(\delta - \lambda_2) e^{-\lambda_2 s} - (\delta - \lambda_1) e^{-\lambda_1 s} \right] \\ &\quad \cdot e^{\delta (s - t)} e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s, \end{aligned} \tag{43}$$

which helps us rewrite the generating function $V(z,t)$ as

$$V(z,t) = q_1 e^{-\lambda_1 t} + q_2 e^{-\lambda_2 t} + \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s} \right) e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s + \frac{q_1 \lambda_1 + q_2 \lambda_2}{\lambda_2 - \lambda_1} \int_0^t \left[(\delta - \lambda_1) e^{-\lambda_1 s} - (\delta - \lambda_2) e^{-\lambda_2 s} \right] e^{\delta (s - t)} e^{\overline{\nu} z \left(e^{-\delta s} - e^{-\delta t} \right)} \mathrm{d}s. \tag{44}$$

To finally verify (6) and (7), we extract the probability mass function $P_m(t)$ from (44) through the conversion formula (19). Since the first two terms in (44) are eliminated by differentiation with respect to z, $P_0(t)$ in (6) needs to be derived from (44) directly as $P_0(t) = V(-1,t)$. For $m \geq 1$,

$$\frac{\partial^m V(z,t)}{\partial z^m} = \frac{\overline{\nu}^m(q_1\lambda_1 + q_2\lambda_2)}{\lambda_2 - \lambda_1}\left[\frac{\lambda_1\lambda_2}{q_1\lambda_1 + q_2\lambda_2}\int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)\left(e^{-\delta s} - e^{-\delta t}\right)^m e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s\right. \\ \left. + \int_0^t \left[(\delta - \lambda_1)e^{-\lambda_1 s} - (\delta - \lambda_2)e^{-\lambda_2 s}\right]e^{\delta(s-t)}\left(e^{-\delta s} - e^{-\delta t}\right)^m e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s.\right. \tag{45}$$

By substituting $\overline{\nu} = \nu/\delta$ and $z = -1$ into this expression and dividing it by $m!$, we obtain (7).

Derivation of (8): by changing the first integral in (42) in the form

$$\lambda_1\lambda_2\int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s = \int_0^t e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}d\left(\lambda_1 e^{-\lambda_2 s} - \lambda_2 e^{-\lambda_1 s}\right) \\ = \lambda_1 e^{-\lambda_2 t} - \lambda_2 e^{-\lambda_1 t} + (\lambda_2 - \lambda_1)e^{\overline{\nu}z\left(1 - e^{-\delta t}\right)} - \nu z\int_0^t \left(\lambda_2 e^{-(\lambda_1+\delta)s} - \lambda_1 e^{-(\lambda_2+\delta)s}\right)e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s, \tag{46}$$

we rewrite (42) as

$$V(z,t) = e^{\overline{\nu}z\left(1 - e^{-\delta t}\right)} - \frac{\nu z}{\lambda_2 - \lambda_1}\int_0^t \left(\lambda_2 e^{-(\lambda_1+\delta)s} - \lambda_1 e^{-(\lambda_2+\delta)s}\right)e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s \\ + \frac{\nu z e^{-\delta t}(q_1\lambda_1 + q_2\lambda_2)}{\lambda_2 - \lambda_1}\int_0^t \left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s. \tag{47}$$

With the help of the elementary identity

$$\frac{\partial^m (ze^{az})}{\partial z^m} = (m + az)a^{m-1}e^{az}, \tag{48}$$

we find

$$\frac{\partial^m V(z,t)}{\partial z^m} = \overline{\nu}^m\left(1 - e^{-\delta t}\right)^m e^{\overline{\nu}z\left(1 - e^{-\delta t}\right)} + \frac{\overline{\nu}^m}{\lambda_2 - \lambda_1}\left[\int_0^t \left[(q_1\lambda_1 + q_2\lambda_2)\left(e^{-\lambda_1 s} - e^{-\lambda_2 s}\right)e^{-\delta t} - \lambda_2 e^{-(\lambda_1+\delta)s} + \lambda_1 e^{-(\lambda_2+\delta)s}\right]\right. \\ \left. \times\left[m\delta + \nu z\left(e^{-\delta s} - e^{-\delta t}\right)\right]\left(e^{-\delta s} - e^{-\delta t}\right)^{m-1} e^{\overline{\nu}z\left(e^{-\delta s} - e^{-\delta t}\right)}\mathrm{d}s\right]. \tag{49}$$

Then, (8) follows from (19) immediately.

Next, we present a result for the case $\gamma = 0$, for which gene activation is irreversible [26]. It may give an approximation to the case when the gene-on state is regulated to be very stable ($\gamma \ll 1$) under the highest induction level [15, 16, 32]. □

Theorem 2. *If $\gamma = 0$, then the probability mass function $P_m(t)$ takes the form*

$$P_m(t) = \begin{cases} q_1 e^{-\lambda_1 t} + q_2 e^{-\lambda_2 t} + \int_0^t \left[q_1\lambda_1 e^{\lambda_1(s-t)} + q_2\lambda_2 e^{\lambda_2(s-t)}\right] e^{\overline{\nu}\left(e^{-\delta s}-1\right)} \mathrm{d}s, & \text{for } m = 0, \\ \dfrac{\overline{\nu}^m}{m!} \int_0^t \left[q_1\lambda_1 e^{\lambda_1(s-t)} + q_2\lambda_2 e^{\lambda_2(s-t)}\right] \left(1 - e^{-\delta s}\right)^m e^{\overline{\nu}\left(e^{-\delta s}-1\right)} \mathrm{d}s, & \text{for } m = 1, 2, \ldots, \end{cases} \tag{50}$$

or alternatively, for $m = 0, 1, \ldots,$

$$P_m(t) = \frac{\overline{\nu}^m}{m!}\left[\left(1 - e^{-\delta t}\right)^m e^{\overline{\nu}\left(e^{-\delta t}-1\right)} - e^{-\delta t}\int_0^t \left[q_1 e^{\left(\delta-\lambda_1\right)(t-s)} + q_2 e^{\left(\delta-\lambda_2\right)(t-s)}\right] \times \left[m\delta - \nu\left(1 - e^{-\delta s}\right)\right]\left(1 - e^{-\delta s}\right)^{m-1} e^{\overline{\nu}\left(e^{-\delta s}-1\right)} \mathrm{d}s\right]. \tag{51}$$

Proof. As in the proof of Theorem 1, our basic strategy is to solve the initial value problem (30). If $\gamma = 0$, then $\alpha = \lambda_2$ and $\beta = \lambda_1$ by definitions (24) and (29). Combining this with (31), we see that $\mathscr{L}_{\overline{\alpha},\overline{\beta}}^{\overline{\lambda}_1,\overline{\lambda}_2}(w) = 0$ reduces to

$$x^2\left[w'''(x) - w''(x)\right] + x\left(\overline{\lambda}_1 + \overline{\lambda}_2 + 1\right)\left[w''(x) - w'(x)\right] + \overline{\lambda}_1\overline{\lambda}_2\left[w'(x) - w(x)\right] = 0. \tag{52}$$

According to the initial conditions in (30), we denote by

$$\overline{y}(x) = w'(x) - w(x), \quad \text{with } \overline{y}'\left[\overline{\nu}z_0\right] = -1, \qquad \overline{y}''\left(\overline{\nu}z_0\right) = \frac{q_1\overline{\lambda}_1 + q_2\overline{\lambda}_2}{\overline{\nu}z_0}. \tag{53}$$

Then, (52) is Euler's equation of $\overline{y}(x)$ with two independent solutions $\overline{y}_1(x) = x^{-\overline{\lambda}_1}$ and $\overline{y}_2(x) = x^{-\overline{\lambda}_2}$. Hence, $\overline{y}(x)$ can be expressed as a linear combination of $\overline{y}_1(x)$ and $\overline{y}_2(x)$. By determining the coefficients through initial conditions of $\overline{y}(x)$, it readily follows that

$$\overline{y}(x) = -q_1\left(\overline{\nu}z_0\right)^{\overline{\lambda}_1} x^{-\overline{\lambda}_1} - q_2\left(\overline{\nu}z_0\right)^{\overline{\lambda}_2} x^{-\overline{\lambda}_2}. \tag{54}$$

Together with (53), we find that $w(x)$ can be obtained by solving the simple problem:

$$w'(x) - w(x) = -q_1\left(\overline{\nu}z_0\right)^{\overline{\lambda}_1} x^{-\overline{\lambda}_1} - q_2\left(\overline{\nu}z_0\right)^{\overline{\lambda}_2} x^{-\overline{\lambda}_2}, \qquad w\left(\overline{\nu}z_0\right) = 1. \tag{55}$$

By a routine calculation of the first-order linear ordinary differential equation [31], we obtain

$$w(x) = e^{x-\overline{\nu}z_0} - \int_{\overline{\nu}z_0}^{x}\left[q_1\left(\overline{\nu}z_0\right)^{\overline{\lambda}_1} s^{-\overline{\lambda}_1} + q_2\left(\overline{\nu}z_0\right)^{\overline{\lambda}_2} s^{-\overline{\lambda}_2}\right] e^{x-s}\mathrm{d}s. \tag{56}$$

By replacing x with $\overline{\nu}z$ and z_0 with $ze^{-\delta t}$, we convert $w(x)$ to the generating function $V(z,t)$:

$$V(z,t) = e^{\overline{\nu}z_0\left(e^{\delta t}-1\right)} - \int_{\overline{\nu}z_0}^{\overline{\nu}z_0 e^{\delta t}}\left[q_1\left(\overline{\nu}z_0\right)^{\overline{\lambda}_1} s^{-\overline{\lambda}_1} + q_2\left(\overline{\nu}z_0\right)^{\overline{\lambda}_2} s^{-\overline{\lambda}_2}\right] e^{\overline{\nu}z_0 e^{\delta t}-s}\mathrm{d}s. \tag{57}$$

Substituting $s = \overline{\nu}z_0\exp(\delta\tau)$ in the integral simplifies the expression to

$$V(z,t) = e^{\overline{\nu}z_0\left(e^{\delta t}-1\right)} - \nu z_0\int_0^t \left[q_1 e^{-\lambda_1\tau} + q_2 e^{-\lambda_2\tau}\right] e^{\delta\tau} e^{\overline{\nu}z_0\left(e^{\delta t}-e^{\delta\tau}\right)}\mathrm{d}\tau. \tag{58}$$

After substituting $z_0 = ze^{-\delta t}$ and then $s = t - \tau$, we finally obtain

$$V(z,t) = e^{\overline{\nu}z\left(1-e^{-\delta t}\right)} - \nu z e^{-\delta t}\int_0^t \left[q_1 e^{\left(\delta-\lambda_1\right)(t-s)} + q_2 e^{\left(\delta-\lambda_2\right)(t-s)}\right] \cdot e^{\overline{\nu}z\left(1-e^{-\delta s}\right)}\mathrm{d}s. \tag{59}$$

Derivation of (50): through integration by parts, we express the last integral of (59) in an equivalent form:

$$\begin{aligned} \nu z e^{-\delta t}\int_0^t \left[q_1 e^{\left(\delta-\lambda_1\right)(t-s)} + q_2 e^{\left(\delta-\lambda_2\right)(t-s)}\right] e^{\overline{\nu}z\left(1-e^{-\delta s}\right)}\mathrm{d}s &= e^{-\delta t}\int_0^t \left[q_1 e^{\left(\delta-\lambda_1\right)(t-s)} + q_2 e^{\left(\delta-\lambda_2\right)(t-s)}\right] e^{\delta s}\mathrm{d}e^{\overline{\nu}z\left(1-e^{-\delta s}\right)} \\ &= e^{\overline{\nu}z\left(1-e^{-\delta t}\right)} - q_1 e^{-\lambda_1 t} - q_2 e^{-\lambda_2 t} - \int_0^t \left[q_1\lambda_1 e^{\lambda_1(s-t)} + q_2\lambda_2 e^{\lambda_2(s-t)}\right] e^{\overline{\nu}z\left(1-e^{-\delta s}\right)}\mathrm{d}s. \end{aligned} \tag{60}$$

The substitution of this into (59) helps us rewrite the generating function $V(z,t)$ as

$$V(z,t) = q_1 e^{-\lambda_1 t} + q_2 e^{-\lambda_2 t} + \int_0^t \left[q_1\lambda_1 e^{\lambda_1(s-t)} + q_2\lambda_2 e^{\lambda_2(s-t)}\right] e^{\overline{\nu}z\left(1-e^{-\delta s}\right)} \mathrm{d}s. \tag{61}$$

By the conversion formula (19), $P_0(t)$ in (50) is derived from (61) directly as $P_0(t) = V(-1,t)$. For $m \geq 1$,

$$\frac{\partial^m V(z,t)}{\partial z^m} = \overline{\nu}^m \int_0^t \left[q_1\lambda_1 e^{\lambda_1(s-t)} + q_2\lambda_2 e^{\lambda_2(s-t)}\right]\left(1-e^{-\delta s}\right)^m e^{\overline{\nu}z\left(1-e^{-\delta s}\right)} \mathrm{d}s, \tag{62}$$

which leads to (50) by substituting $\overline{\nu} = \nu/\delta$ and $z = -1$ into this expression and dividing it by $m!$.

Derivation of (51): with the help of (59) and the elementary identity (48), we find

$$\frac{\partial^m V(z,t)}{\partial z^m} = \overline{\nu}^m\left(1-e^{-\delta t}\right)^m e^{\overline{\nu}z\left(1-e^{-\delta t}\right)} - \overline{\nu}^m e^{-\delta t} \int_0^t \left[q_1 e^{(\delta-\lambda_1)(t-s)} + q_2 e^{(\delta-\lambda_2)(t-s)}\right] \times \left[m\delta + \nu z\left(1-e^{-\delta s}\right)\right]\left(1-e^{-\delta s}\right)^{m-1} e^{\overline{\nu}z\left(1-e^{-\delta s}\right)} \mathrm{d}s. \tag{63}$$

Then, (51) follows from the conversion formula (19) immediately. The proof is completed. □

3. Dynamical Bimodal Distribution for Inducible Genes

We elucidate the regulation of cross-talking pathways on dynamical mRNA distribution for inducible genes under acute stresses, which have to fulfill two requirements. First, a default or spontaneous mechanism maintains transcription at the basal level when cells are under normal cellular environments to avoid the overexuberant expression [17, 29]. Second, the exposure of cells to the extracellular stresses results in the transient activation of dedicated intracellular signaling pathways, which enables cells to adapt immediately to the new environment for maximal cell survival [19, 20]. For instance, the osmotic shock induces the MAPK HOG pathway in *S. cerevisiae* through the phosphorylation of MAPK Hog1 that leads to the rapid translocation of Hog1 into the nucleus to launch a transcriptional program [19]. Therefore, the cross-talk between a weak basal pathway and an inducible signaling pathway could be generated to initiate the stress gene transcription.

We shall discuss the observed dynamical transition among three distribution modes for the *STL1* gene in *S. cerevisiae* in response to acute osmotic stress [19]. Under mild or high NaCl concentrations, no expression was detected in most cells at the initial time referred to as the decaying distribution and full expression was detected at a large time scale referred to as the unimodal distribution. The scenario becomes more fascinated within the intermediate time as both nonexpressing and fully expressing cell subpopulations were presented, characterized as the bimodal distribution. It is believed that such bimodal transcription output is controlled by both the retention time and concentration of Hog1 in the nucleus and results in a larger phenotypic variability for cell survival in harsh environments [19].

We demonstrate through numerical examples that cross-talking pathways can generate dynamical bimodality of osmoresponsive gene transcription under mild or high stress level. Most mRNA half-lives in *S. cerevisiae* range around a median of 11 min [33], so we fix $\delta = \ln(2)/11 \approx 0.063\,\mathrm{min}^{-1}$. We set the transcription rate as its upper bound $\nu/\delta_m = 10$ [34], and thus $\nu = 0.63\,\mathrm{min}^{-1}$. Also, we set a relative small strength rate $\lambda_1 = 0.1\,\mathrm{min}^{-1}$ for the basal pathway and a relative large strength rate $\lambda_2 = 3\,\mathrm{min}^{-1}$ for the signaling pathway. Such parameter set generates a prolonged region of the gene-off time (denoted by T_{off}):

$$T_{\mathrm{off}} = \frac{q_1}{\lambda_1} + \frac{q_2}{\lambda_2} \in \left(\frac{1}{\lambda_2}, \frac{1}{\lambda_1}\right), \tag{64}$$

that covers the lag time of 1 to 3 minutes for the activation of most typical osmoresponsive genes, such as STL1, GRE2, or GPD1 [20].

The mild stress level results in a moderate increasing nuclear accumulation of activated Hog1 [19]. This leads to the well-matched competition between two pathways with their selected probabilities $q_1 \approx q_2$. We set $q_1 = 0.51$ and $q_2 = 0.49$ that generates *transcriptional efficiency* $1/T_{\mathrm{off}} = 0.19\,\mathrm{min}^{-1}$. On the other hand, the high stress level induces saturated nuclear Hog1 concentration [19] and thus may direct gene activation more frequently through the signaling pathway with $q_2 > q_1$. We then set $q_1 = 0.3$ and $q_2 = 0.7$ that generates $1/T_{\mathrm{off}} = 0.3\,\mathrm{min}^{-1}$. Also, it was reported that the gene-on state could be significantly stabilized in response to increasing stress level, and thus $\gamma \ll 1$ [2, 15, 16, 32].

We next take advantage of exact forms (8) and (51) to quantify dynamical mRNA distribution under mild and high stress levels, respectively. In both scenarios, we observe obvious transitions from the decaying distribution to unimodal distribution as time develops, going through a bimodal stage within a 30 min length of early time (Figure 1). To test to what extent the cross-talking pathways modulate the bimodality, we reset $q_1 = 1$ so that the gene is activated by a single pathway with strength rate $\lambda = \lambda_1$. In turn, we let λ be equal to transcriptional efficiencies for both cases of mild and high stress levels. Interestingly, it shows that the intermediate bimodality becomes much more fragile that lasts less than 5 min, see Figure 1 (inset). This confirms that the cross-talking regulation indeed prolongs duration of mRNA bimodal distribution, reinforcing the general feature of bimodal transcription for stress genes [19, 35].

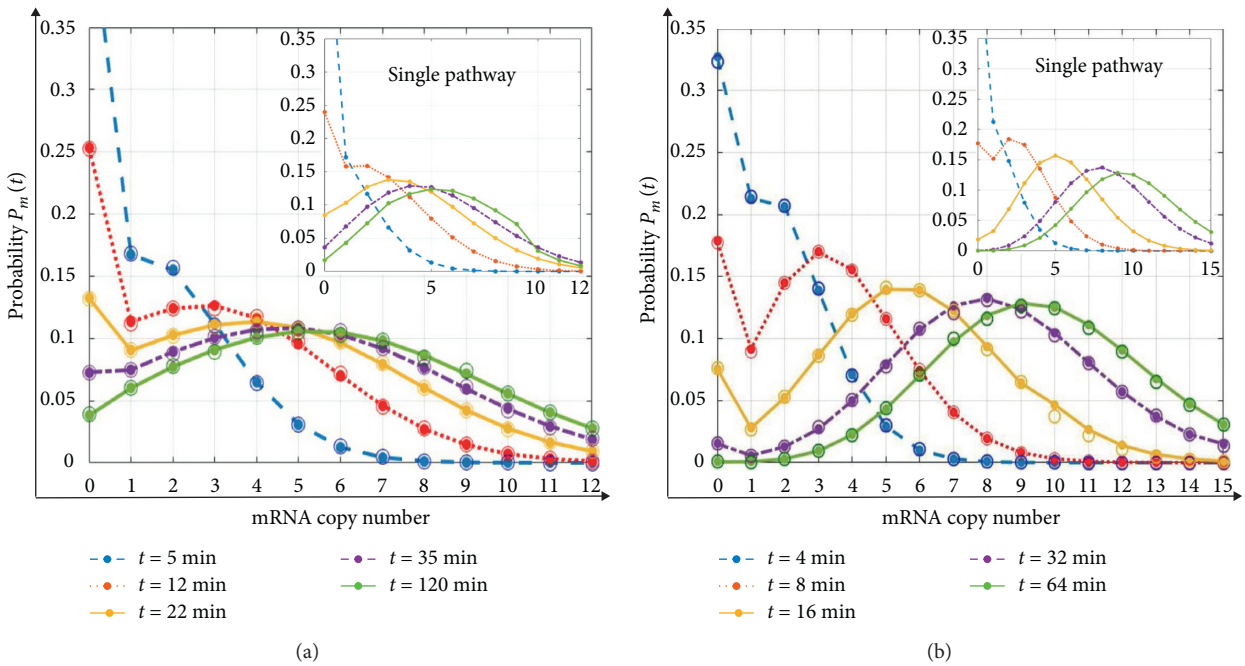


FIGURE 1: Cross-talking pathways generate typical dynamical transitions from mRNA decaying distribution to unimodal distribution undergoing the significant intermediate bimodal stage for stress genes in yeast with $\nu = 0.63, \delta = 0.063, \lambda_1 = 0.1$, and $\lambda_2 = 3$ (min^{-1}). (a) Dynamical distribution of the stress gene under mild stress level equipped with $q_1 = 0.51, q_2 = 0.49$, and $\gamma = 0.126\,\text{min}^{-1}$. (b) Distribution dynamics of the stress gene under high stress level with $q_1 = 0.3, q_2 = 0.7$, and $\gamma = 0\,\text{min}^{-1}$. In (a) and (b), the solid circles in lines represent analytical solution, whereas the hollow circles represent the numerical result obtained by Gillespie's algorithm. The embedded distribution dynamics are generated by the single pathway in which durations of bimodal distribution are much shorter than that generated by cross-talking pathways.

4. Conclusion and Discussion

Recent single-cell studies have generated massive data on the distribution of mRNA copy numbers for many genes under various experimental conditions [1–3]. The most commonly observed modes for the distribution are the decaying distribution, unimodal distribution, and bimodal distribution. It is reported that the bimodal distribution could help differentiate an isogenic cell population into two dynamically stable groups with distinct phenotypes. For instance, the bimodal *Nanog* expression plays a key role in mediating cell differentiation in embryonic stem cells [7]; the transcriptional bimodality of *Tat* generates one HIV highly expressed subpopulation and the other with HIV latency [36]; the bimodal transcription pattern of stress genes in yeast increases phenotypic variability in response to unpredictable environmental changes [19]. In the classical two-state model (1), the bimodal distribution has been suggested to originate from the random switch between gene-off and gene-on states as bimodality would disappear when the gene is always active [7, 28].

One of our major concerns is how the gene activation process modulates the bimodal distribution for inducible genes in face of external cues. We introduce a cross-talking pathway model (2) by integrating two competitive signaling pathways into the two-state model [24, 25]. Comparing to other statistical concepts such as the noise or fano factor that can be computed relatively easily, exploring transparent analytical formulas for mRNA distribution remains a challenging task. For instance, a traditional approach for calculating mRNA distribution is to express the corresponding generating function in the form of hypergeometric functions [6, 28]. However, this does not permit an easy tractable way to further transform the generating function to mRNA distribution when the time and system parameters are finite. We endeavor to express time-dependent mRNA distribution in simple mathematical formulas within a large range of finite parameters for the cross-talking pathway model. The formulas are obtained by solving the initial value problems of third-order ordinary differential equations, which are derived from the master equations of the cross-talking pathway model.

We consider stress-responsive genes under acute stresses as the implement of dynamical analytical formulas for mRNA distribution. The stress gene can be activated either by a weak spontaneous basal pathway with strength λ_1 and selection probability q_2, or alternately by an inducible signaling pathway with strength λ_2 and selection probability $q_2 = 1 - q_1$. The probability q_1 (or q_2) may be governed by the number and availability of activated transcription factors (TFs) near the binding sites [19, 20] and thus is related with the stress level; the inducible strength λ_2 can be governed by the binding strength between the TFs and the DNA sites [11, 20], and thus may be tightly related with the type of external signals. We suggest that the signaling pathway is efficiently induced with

$$\lambda_1 \ll \lambda_2. \quad (65)$$

On the assumption of Theorem 1, (65) leads to $q_1 \sim q_2$, which shows a mild stress level that may induce a similar concentration of activated TFs in each pathway. The assumption of Theorem 2 presents an active system in which the gene inactivation process is blocked. It suggests a very high stress level that may induce a significant decrease of the gene inactivation rate γ [2, 16, 32]. For instance, compared with normal cellular conditions, the inductions with the highest strength lead to the decrease of γ for over 10^6 folds for the *Lac* promoter in *E. coli* [32] and possible 10^5 fold change for the *POL1* promoter in yeast [34].

Under both mild and high stress levels, we utilize the parameter rates for stress genes in yeast to understand their dynamical distributions. We reveal a typical dynamical transition from the decaying distribution to unimodal distribution as time develops, going through an intermediate bimodal stage with over 30 minutes long. This result strongly suggests the general and robust bimodal transcription of stress genes regulated by the cross-talking pathways in response to acute stresses [19, 35]. First, noting that the system takes about 120 minutes to reach the steady state, bimodal distribution maybe clearly visible in the course of real-time imaging due to its relatively large time window of 30 minutes. Second, under the same durations for the gene-off and gene-on states, as well as the same mRNA synthesis and degradation rates, the two-state model (a single pathway) can only generate a much fragile intermediate bimodal stage that lasts less than 5 minutes.

The general intermediate bimodal stage is consistent with recent observations that the cross-talking pathway model (2) is more likely to generate high transcription noise and bimodal distribution at the steady state compared with the two-state model (1) [24, 25]. It suggests that natural selection may favor two or more parallel signaling pathways for gene activation to enhance expression variability which may provide a clear benefit in the face of an environmental stress [19, 37]. On the other hand, the cross-talking pathway model (2) assumes the constant system rates and cannot take into account some inducible genes and immune genes regulated by the oscillating signal [38], temporal feedback [39], or different cell cycle phases [14]. Interestingly, the dynamical bimodality can be easily generated through the other very different mechanisms such as the system involves positive or negative genetic feedback loops [39]. The extensions of our model that include those salient biological features may require assuming time-varying parameters with biological realism accordingly. Future work will focus on how those extended models modulate the dynamical bimodal or even multimodal transcription distributions.

Data Availability

The data used to support findings of this study are included within the article: (1) the degradation rate of the mRNA molecule in *S. cerevisiae* equals $\ln(2)/11 \approx 0.063$ (1/min), which is calculated through the 11 min average half-life obtained in reference [33]. (2) The mRNA synthesis rate in *S. cerevisiae* equals $0.063 * 10 = 0.63$ (1/min), which is calculated according to the mRNA degradation rate 0.063 (1/min) estimated in (1) and the upper bound of its transcription rate 10 (per mRNA lifetime) given in reference [34]. (3) The 1–3 min gene-off duration for most typical osmoresponsive genes in *S. cerevisiae* is reported in reference [20].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of China grants (nos. 11631005 and 11871174), Program for Changjiang Scholars and Innovative Research Team in University (no. IRT_16R16), Key Project of Hunan Educational Committee (no. 19A497), and China Postdoctoral Science Foundation (Grant no. 2019M660196).

References

[1] B. Munsky, G. Neuert, and A. van Oudenaarden, "Using gene expression noise to understand gene regulation," *Science*, vol. 336, no. 6078, pp. 183–187, 2012.

[2] A. Sanchez and I. Golding, "Genetic determinants and cellular constraints in noisy gene expression," *Science*, vol. 342, no. 6163, pp. 1188–1193, 2013.

[3] B. Munsky, Z. Fox, and G. Neuert, "Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics," *Methods*, vol. 85, pp. 12–21, 2015.

[4] G. Neuert, B. Munsky, L. Teytelman, M. Khammash, and A. van Oudenaarden, "Systematic identification of signal-activated stochastic gene regulation," *Science*, vol. 339, no. 6119, pp. 584–587, 2013.

[5] J. Zhang, L. Chen, and T. Zhou, "Analytical distribution and tunability of noise in a model of promoter progress," *Biophysical Journal*, vol. 102, no. 6, pp. 1247–1257, 2012.

[6] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, "Stochasticity of gene products from transcriptional pulsing," *Physical Review E*, vol. 79, no. 3, Article ID 031911, 2009.

[7] T. Kalmar, C. Lim, and P. Hayward, "Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells," *PLoS Biology*, vol. 7, no. 7, Article ID e1000149, 2009.

[8] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: from theories to phenotypes," *Nature Reviews. Genetics*, vol. 6, no. 6, pp. 451–464, 2005.

[9] A. J. M. Larsson, P. Johnsson, L. Hartmanis et al., "Genomic encoding of transcriptional burst kinetics," *Nature*, vol. 565, no. 7738, pp. 251–254, 2019.

[10] F. Jiao, Q. Sun, M. Tang, J. Yu, and B. Zheng, "Distribution modes and their corresponding parameter regions in stochastic gene transcription," *SIAM Journal on Applied Mathematics*, vol. 75, no. 6, pp. 2396–2420, 2015.

[11] D. R. Larson, "What do expression dynamics tell us about the mechanism of transcription," *Current Opinion in Genetics & Development*, vol. 21, no. 5, pp. 591–599, 2011.

[12] F. Jiao, M. Tang, and J. Yu, "Distribution profiles and their dynamic transition in stochastic gene transcription," *Journal of Differential Equations*, vol. 254, no. 8, pp. 3307–3328, 2013.

[13] T. Zhou and J. Zhang, "Analytical results for a multistate gene model," *SIAM Journal on Applied Mathematics*, vol. 72, no. 3, pp. 789–818, 2012.

[14] Z. Cao and R. Grima, "Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells," *Proceedings of the National Academy of Sciences*, vol. 117, no. 9, pp. 4682–4692, 2020.

[15] D. L. Jones, R. C. Brewster, and R. Phillips, "Promoter architecture dictates cell-to-cell variability in gene expression," *Science*, vol. 346, no. 6216, pp. 1533–1536, 2014.

[16] R. D. Dar, B. S. Razooky, A. Singh et al., "Transcriptional burst frequency and burst size are equally modulated across the human genome," *Proceedings of the National Academy of Sciences*, vol. 109, no. 43, pp. 17454–17459, 2012.

[17] J. Macia, S. Regot, T. Peeters, N. Conde, R. Sole, and F. Posas, "Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction," *Science Signaling*, vol. 2, no. 63, p. ra13, 2009.

[18] A. Jordan, P. Defechereux, and E. Verdin, "The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation," *The EMBO Journal*, vol. 20, no. 7, pp. 1726–1738, 2001.

[19] S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter, "Transient activation of the HOG MAPK pathway regulates bimodal Gene expression," *Science*, vol. 332, no. 6030, pp. 732–735, 2011.

[20] E. Nadal, G. Ammerer, and F. Posas, "Controlling gene expression in response to stress," *Nature Reviews Genetics*, vol. 12, no. 12, pp. 833–845, 2011.

[21] A. Aguirre, M. E. Rubio, and V. Gallo, "Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal," *Nature*, vol. 467, no. 7313, pp. 323–327, 2010.

[22] A. S. Yoo, C. Bais, and I. Greenwald, "Crosstalk between the EGFR and LIN-12/Notch pathways in C. elegans vulval development," *Science*, vol. 303, no. 5658, pp. 663–666, 2004.

[23] T. Tanji, X. Hu, A. N. R. Weber, and Y. T. Ip, "Toll and IMD pathways synergistically activate an innate immune response in Drosophila melanogaster," *Molecular and Cellular Biology*, vol. 27, no. 12, pp. 4578–4588, 2007.

[24] J. Yu, Q. Sun, and M. Tang, "The nonlinear dynamics and fluctuations of mRNA levels in cross-talking pathway activated transcription," *Journal of Theoretical Biology*, vol. 363, pp. 223–234, 2014.

[25] F. Jiao, Q. Sun, Q. Sun, and J. YuLin, "Distribution profiles in gene transcription activated by the cross-talking pathway," *Discrete & Continuous Dynamical Systems—B*, vol. 24, no. 6, pp. 2799–2810, 2019.

[26] M. Tabaka and R. Hołyst, "Binary and graded evolution in time in a simple model of gene induction," *Physical Review E*, vol. 82, no. 5, Article ID 052902, 2010.

[27] F. Jiao, J. Ren, J. Ren, and J. Yu, "Analytical formula and dynamic profile of mRNA distribution," *Discrete & Continuous Dynamical Systems—B*, vol. 25, no. 1, pp. 241–257, 2020.

[28] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression," *Proceedings of the National Academy of Sciences*, vol. 105, no. 45, pp. 17256–17261, 2008.

[29] D. S. Schneider, "How and why does a fly turn its immune system off?" *PLoS Biology*, vol. 5, no. 9, pp. 1847–1849, 2007.

[30] L. C. Evans, *Partial Differential Equations*, American Math. Society, Providence, USA, 2nd edition, 2010.

[31] M. W. Hirsch, S. Smale, and R. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, Academic Press, Cambridge, MA, USA, 2nd edition, 2003.

[32] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, "General properties of transcriptional time series in Escherichia coli," *Nature Genetics*, vol. 43, no. 6, pp. 554–560, 2011.

[33] C. Miller, B. Schwalb, K. Maier et al., "Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast," *Molecular Systems Biology*, vol. 7, no. 1, p. 458, 2011.

[34] L. B. Carey, D. V. Dijk, P. M. A. Sloot, J. A. Kaandorp, and E. Segal, "Promoter sequence determines the relationship between expression level and noise," *PLoS Biology*, vol. 11, no. 4, Article ID e1001528, 2013.

[35] S. Paliwal, P. A. Iglesias, K. Campbell, Z. Hilioti, A. Groisman, and A. Levchenko, "MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast," *Nature*, vol. 446, no. 7131, pp. 46–51, 2007.

[36] B. S. Razooky, A. Pai, K. Aull, I. M. Rouzine, and L. S. Weinberger, "A hardwired HIV latency program," *Cell*, vol. 160, no. 5, pp. 990–1001, 2015.

[37] J. Liu, H. Martin-Yken, F. Bigey, S. Dequin, J.-M. François, and J.-P. Capp, "Natural yeast promoter variants reveal epistasis in the generation of transcriptional-mediated noise and its potential benefit in stressful conditions," *Genome Biology and Evolution*, vol. 7, no. 4, pp. 969–984, 2015.

[38] S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert, "Single-cell NF-κB dynamics reveal digital activation and analogue information processing," *Nature*, vol. 466, no. 7303, pp. 267–271, 2010.

[39] J. Chen and R. Grima, "Dynamical phase diagram of an autoregulating gene in fast switching conditions," *The Journal of Chemical Physics*, vol. 152, no. 17, Article ID 174110, 2020.

Hindawi
Complexity
Volume 2020, Article ID 8536365, 15 pages
https://doi.org/10.1155/2020/8536365

Research Article

Optimal Design of Single-Cell Experiments within Temporally Fluctuating Environments

Zachary R. Fox,[1,2,3] Gregor Neuert,[4,5,6] and Brian Munsky [3,7]

[1]*Inria Saclay Ile-de-France, Palaiseau 91120, France*
[2]*Institut Pasteur, USR 3756 IP CNRS, Paris 75015, France*
[3]*School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, USA*
[4]*Department of Molecular Physiology and Biophysics, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA*
[5]*Department of Biomedical Engineering, School of Engineering, Vanderbilt University, Nashville, TN 37232, USA*
[6]*Department of Pharmacology, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA*
[7]*Department of Chemical and Biological Engineering, Colorado State University Fort Collins, CO 80523, USA*

Correspondence should be addressed to Brian Munsky; munsky@colostate.edu

Received 20 October 2019; Accepted 12 February 2020; Published 13 June 2020

Guest Editor: George V. Popescu

Modern biological experiments are becoming increasingly complex, and designing these experiments to yield the greatest possible quantitative insight is an open challenge. Increasingly, computational models of complex stochastic biological systems are being used to understand and predict biological behaviors or to infer biological parameters. Such quantitative analyses can also help to improve experiment designs for particular goals, such as to learn more about specific model mechanisms or to reduce prediction errors in certain situations. A classic approach to experiment design is to use the Fisher information matrix (FIM), which quantifies the expected information a particular experiment will reveal about model parameters. The finite state projection-based FIM (FSP-FIM) was recently developed to compute the FIM for discrete stochastic gene regulatory systems, whose complex response distributions do not satisfy standard assumptions of Gaussian variations. In this work, we develop the FSP-FIM analysis for a stochastic model of stress response genes in *S. cerevisiae* under time-varying MAPK induction. We verify this FSP-FIM analysis and use it to optimize the number of cells that should be quantified at particular times to learn as much as possible about the model parameters. We then extend the FSP-FIM approach to explore how different measurement times or genetic modifications help to minimize uncertainty in the sensing of extracellular environments, and we experimentally validate the FSP-FIM to rank single-cell experiments for their abilities to minimize estimation uncertainty of NaCl concentrations during yeast osmotic shock. This work demonstrates the potential of quantitative models to not only make sense of modern biological datasets but to close the loop between quantitative modeling and experimental data collection.

1. Introduction

The standard approach to design experiments has been to rely entirely on expert knowledge and intuition. However, as experimental investigations become more complex and seek to examine systems with more subtle nonlinear interactions, it becomes much harder to improve experimental designs using intuition alone. This issue has become especially relevant in modern single-cell-single-molecule investigations of gene regulatory processes. Performing such powerful, yet complicated, experiments involves the selection from among a large number of possible experimental designs, and it is often not clear which designs will provide the most relevant information. A systematic approach to solve this problem is model-driven experiment design, in which one combines existing knowledge or experience to form an assumed (and partially incorrect) mathematical model of the system to estimate and optimize the value of potential experimental settings. In practice, such preliminary models would be defined by existing data taken in simpler or more general settings such as inexpensive bulk experiments or would be estimated from literature values conducted on

similar genes, pathways, or organisms. When parameter or model structures are uncertain, these could be described according to a prior distribution, and experiments would need to be selected according to which performs best on average across the many possible model/parameter combinations.

In recent years, model-driven experiment design has gained traction for biological models of gene expression, whether in the Bayesian setting [1] or using Fisher information for deterministic models [2], and even in the stochastic, single-cell setting [3–7]. Despite the promise and active development of model-driven experiment design from the theoretical perspective, more general, yet biologically inspired, approaches are needed to make these methods suitable for the experimental community at large. In this work, we apply model-driven experiment design to an experimentally validated model of stochastic transcription that is activated by time-varying high osmolarity glycerol (HOG) mitogen-activated protein kinase (MAPK) induction in yeast [8–10]. To demonstrate a concrete and practical application of model-driven experiment design, we find the optimal *measurement schedule* (i.e., when measurements ought to be taken) and the appropriate *number of individual cells* to be measured at each time point.

In our computational analyses, we consider the experimental technique of single-molecule mRNA fluorescence *in situ* hybridization (smFISH), where specific fluorescent oligonucleotide probes are hybridized to mRNA of interest in fixed cells [11, 12]. Cells are then imaged, and the mRNA abundance in each cell is counted, either by hand or using automated software such as [13]. Such counting can be a cumbersome process, but little thought has been given typically to how many cells should be measured and analyzed at each time. Furthermore, when a dynamic response is under investigation, the specific times at which measurements should be taken (i.e., the times after induction at which cells should be fixed and analyzed) are also unclear. In this work, we use the newly developed finite state projection-based Fisher information matrix (FSP-FIM, [6]) to optimize these experimental quantities for osmotic stress response genes in yeast.

The first part of our current study introduces a discrete stochastic model to analyze time-varying MAPK-induced gene expression response in yeast and then demonstrates the use of FSP-based Fisher information to optimize experiments to minimize the uncertainty in model parameters. In the second part of this study, we expand upon this result to find and experimentally verify the optimal smFISH measurement times and cell numbers to minimize uncertainty about unknown environmental inputs (e.g., salt concentrations) to which the cells are subjected. In this way, we are presenting a new methodology by which one can optimally examine behaviors of natural cells to obtain accurate estimations of environmental changes.

2. Background

Gene regulation is the process by which small molecules, chromatin regulators, and general and gene-specific transcription factors interact to regulate the transcription of DNA into RNA and the translation of mRNA into proteins. Even within populations of genetically identical cells, these single-molecule processes are stochastic and give rise to cell-to-cell variability in gene expression levels. Adequate descriptions of such variable responses can only be achieved through the use of stochastic computational models [14–17]. In the following sections, we first introduce a nonequilibrium discrete stochastic model of HOG1-MAPK-induced gene expression, and we then discuss how this model can be analyzed and compared to data using finite state project analyses. All analysis codes are available at https://github.com/MunskyGroup/Fox_Complexity_2020.

2.1. Discrete Stochastic Model of HOG1-MAPK-Induced Gene Expression. To motivate and demonstrate our new approach, we focus our examination on the dynamics of the HOG1-MAPK pathway in yeast, which is a model system to study osmotic stress driven dynamics of signal transduction and gene regulation in single cells [18–23]. Discrete stochastic models of HOG1-MAPK-activated transcription have been used successfully to predict the variability in adaptive transcription responses across yeast cell populations [9, 10, 24]. In particular, the authors in [9] used smFISH data to fit and cross validate a number of different potential models with different numbers of gene states and time-varying parameters. They found that dynamics of two stress response genes, *STL*1 and *CTT*1, could each be described accurately by the model depicted in Figure 1(a).

In brief, the model [9] consists of transitions between four different gene states (S1, S2, S3, and S4). The probability of a transition from the i^{th} to the j^{th} gene state in the infinitesimal time dt is given by the propensity function, $k_{\text{ij}}\text{d}t$. Most of the rates $\{k_{\text{ij}}\}$ are constant in time, except for the transition from S2 to S1, which is controlled by the time-varying level of the HOG1-MAPK signal in the nucleus, $f(t)$. The resulting time-varying rate k_{21} is defined using a linear threshold function:

$$k_{21}(t) = \max[0, \alpha - \beta f(t)], \tag{1}$$

where α and β set the threshold for $k_{21}(t)$ activation/deactivation. The function $f(t)$ was calibrated at several NaCl concentrations by fitting the HOG1-MAPK nuclear localization signals as measured using a yellow fluorescence protein reporter [10]. Figure 1(b) shows $f(t)$ for osmotic stress responses to 0.2 M and 0.4 M NaCl, and Figure 1(c) shows the corresponding values of $k_{21}(t)$. In addition to the state transition rates, each i^{th} state also has a corresponding mRNA transcription rate, k_{ri}. All mRNA molecules degrade with rate γ, independent of gene state. Further descriptions and validations of this model are given in Supplementary Note 1 and in [9, 10, 24]. All experimentally determined parameters for the *STL*1 and *CTT*1 transcription regulation models are provided in Supplemental Table S1, and experimentally determined parameters for the HOG1-MAPK signal model are listed in Supplemental Table S2 [10].

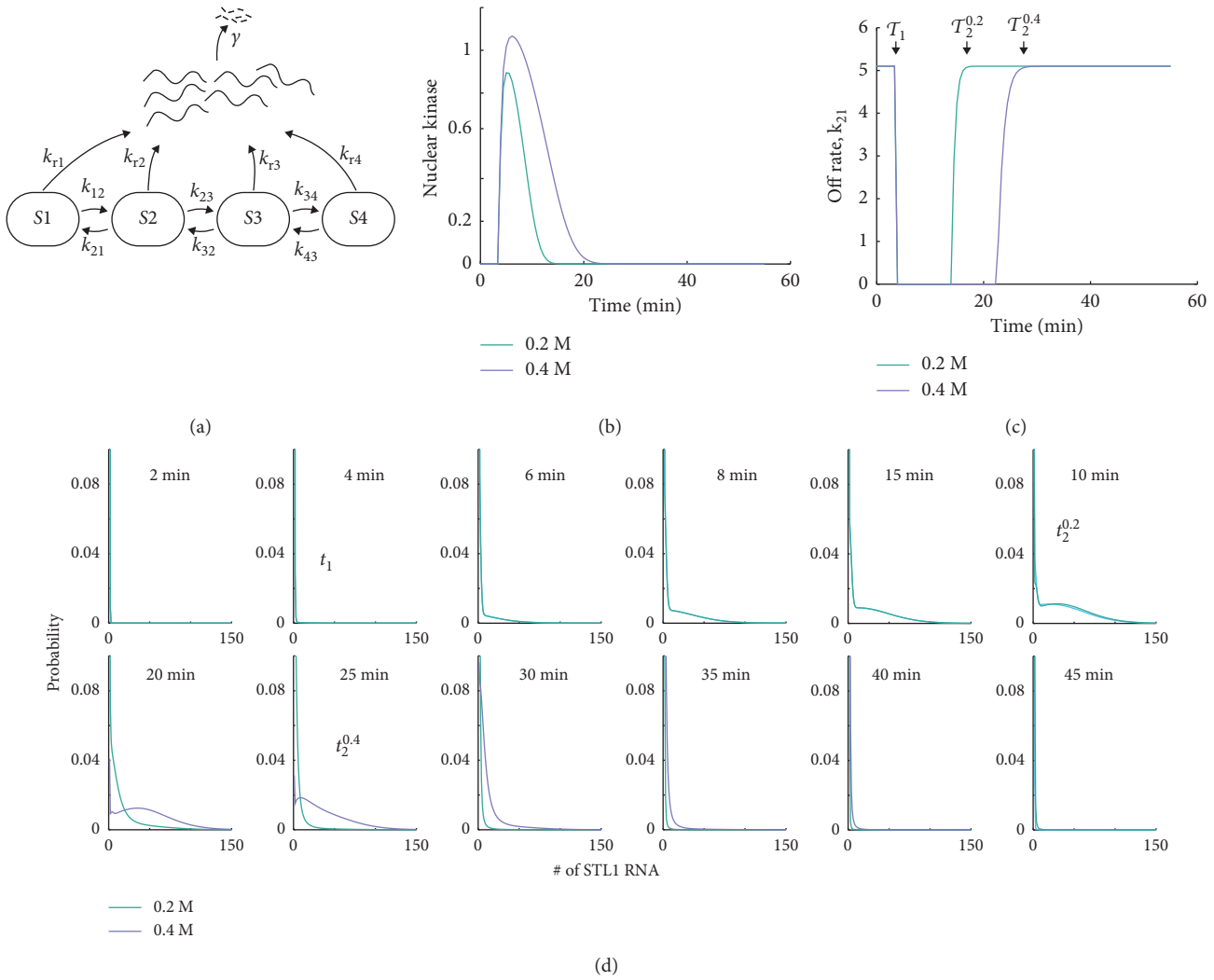


Figure 1: Stochastic modeling of osmotic stress response genes in yeast. (a) Four-state model of gene expression, where each state transcribes mRNA at a different transcription rate, but each mRNA degrades at a single rate γ. (b) Time-varying MAPK nuclear localization signal. (c) The rate of switching from gene activation state S2 to S1 (right) under 0.2 M or 0.4 M NaCl osmotic stress. The time at which k_{21} turns off is denoted with τ_1 and is independent of the NaCl level. The time at which k_{23} turns back on is given by τ_{NaCl} depending on the level of NaCl. (d) Time evolution of the STL1 mRNA in response to the 0.2 M and 0.4 M NaCl stress. Model and parameters from [10] are summarized in Supplementary Notes I and II and Supplementary Tables I and II.

2.2. The Finite State Projection Analysis of Stochastic Gene Expression. To analyze the model described above, we apply the chemical master equation (CME) framework of stochastic chemical kinetics [25]. Combining the time-varying and constant state transition rates $\{k_{ij}\}$, transcription rates $\{k_{ri}\}$, and degradation rate γ from above, the CME can be written in matrix form as a linear ordinary differential equation, $d\mathbf{p}/dt = \mathbf{A}(t)\mathbf{p}$, where the time-varying matrix $\mathbf{A}(t)$ is known as the infinitesimal generator (see Supplementary Note 1). The CME has been the workhorse of stochastic modeling of gene expression, and it is usually analyzed using simulated sample paths of its solution via the stochastic simulation algorithm [26] or with moment approximations [8, 27]. Alternatively, the CME can also be solved with guaranteed errors using the FSP approach [28, 29], which reduces the full CME only to describe the flow of probability among the most likely observable states of the system. Details of the FSP approach to solving chemical kinetic systems are provided in Supplementary Note 1. Application of the FSP analysis to the model (Figure 1(a)) with dynamic Hog1 (Figure 1(b)) modulates time-varying rates k_{21} (Figure 1(c)) and predicts time-evolving probability distributions at 0.2 M and 0.4 M NaCl, as shown in Figure 1(d) [10].

2.3. Likelihood of smFISH Data for FSP Models. Recently, it has come to light that for some systems, it is critical to consider the full distribution of biomolecules across cellular populations when fitting CME models [6, 10]. To match CME model solutions to single-cell smFISH data, one needs to compute and maximize the likelihood of the data given the CME model [9, 10, 24, 30]. Fortunately, the FSP approach allows for computation of the likelihood with

guaranteed accuracy bounds [28]. We assume that measurements at each time point $\mathbf{t} \equiv [t_1, t_2, \ldots, t_{N_t}]$ are independent, as justified by the fact that fixation of cells for measurement precludes temporal cell-to-cell correlations. Measurements of N_c cells can be concatenated into a matrix $\mathbf{D}_t \equiv [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{N_c}]_t$ of the observable mRNA species at each measurement time t.

The likelihood of making the independent observations for all N_c measured cells is the product of the probabilities of observing each cell's measured state. For most gene expression models, however, states are only partially observable, and we define the observed state $\mathbf{x}_i^L$ as the marginalization (or lumping) over all full states $\{\mathbf{x}_j\}_i$ that are indistinguishable from $\mathbf{x}_i$ based on the observation. For example, the model of *STL*1 transcription consists of four gene states (S1–S4, shown in Figure 1(a)), which are unobserved, and the measured number of mRNA, which is observed. If we let index i denote the number of mRNA, then the observed state $\mathbf{x}_i^L$ would lump together the full states (S1, i), (S2, i), (S3, i), and (S4, i). We next define y_i as the number of experimental cells that match $\mathbf{x}_i^L$ at time t. Under these definitions, the likelihood of the observed data (and its logarithm) given the model can be written as

$$\begin{aligned} \ell(\mathbf{D};\boldsymbol{\theta}) &= M \prod_{t=t_1}^{t_{N_t}} \prod_{i \in \mathcal{J}_D} p\left(\mathbf{x}_i^L; t, \boldsymbol{\theta}\right)^{y_i}, \\ \log \ell(\mathbf{D};\boldsymbol{\theta}) &= \sum_{t=t_1}^{t_{N_t}} \sum_{i \in \mathcal{J}_D} y_i \log\left(p\left(\mathbf{x}_i^L; t, \boldsymbol{\theta}\right)\right) + \log M, \end{aligned} \tag{2}$$

where $\mathcal{J}_D$ is the set of states observed in the data, M is a combinatorial prefactor (i.e., from a multinomial distribution) that comes from the arbitrary reordering of measured data, and $p(\mathbf{x}_i^L; t, \boldsymbol{\theta})$ is the marginalized probability mass of the observable species:

$$p\left(\mathbf{x}_i^L; t, \boldsymbol{\theta}\right) = \sum_{\mathbf{x}_j \in \mathbf{x}_i^L} p\left(\mathbf{x}_j; t, \boldsymbol{\theta}\right). \tag{3}$$

The vector of model parameters is denoted by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots]$. Neglecting the term $\log M$, which is independent of the model, the summation in equation (2) can be rewritten as a product $\mathbf{y} \log \mathbf{p}^L$, where $\mathbf{y} \equiv [y_0, y_1, \ldots]$ is the vector of the binned data and $\mathbf{p}^L = [p(\mathbf{x}_0^L), p(\mathbf{x}_1^L), \ldots]^{\mathrm{T}}$ is the corresponding marginalized probability mass vector. One may then maximize equation (2) with respect to $\boldsymbol{\theta}$ to find the *maximum likelihood estimate* (MLE) of the parameters, $\widehat{\boldsymbol{\theta}}$, which will vary depending on each new set of experimental data. We next demonstrate how this likelihood function and the FSP model of the HOG1-MAPK-induced gene expression system can be used to design optimal smFISH experiments using the FSP-based Fisher information matrix [6].

3. Results

3.1. The Finite State Projection-Based Fisher Information for Models of Signal-Activated Stochastic Gene Expression. The Fisher information matrix (FIM) is a common tool in engineering and statistics to estimate parameter uncertainties prior to collecting data, which allows one to find experimental settings that can make these uncertainties as small as possible [3, 4, 31–34]. Recently, it has been applied to biological systems to estimate kinetic rate parameters in stochastic gene expression systems [3–6, 35]. In general, the FIM for a single measurement is defined as

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}\left\{\left(\nabla_{\boldsymbol{\theta}} \log \mathbf{p}(\boldsymbol{\theta})\right)^T \left(\nabla_{\boldsymbol{\theta}} \log \mathbf{p}(\boldsymbol{\theta})\right)\right\}, \tag{4}$$

where the vector $\log \mathbf{p}(\boldsymbol{\theta})$ contains the log-probabilities of each potential observation and the expectation is taken over the probability distribution of states $\mathbf{p}(\boldsymbol{\theta})$ assuming the specific parameter set $\boldsymbol{\theta}$. As the number of measurements, N_c, is increased such that maximum likelihood estimates (MLE) of parameters are unbiased, the distribution of MLE estimates is known to approach a multivariate Gaussian distribution with a covariance given by the inverse of the FIM, i.e.,

$$\sqrt{N_c}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \xrightarrow{\text{dist}} \mathcal{N}\left(0, \mathcal{I}\left(\boldsymbol{\theta}^*\right)^{-1}\right). \tag{5}$$

In [6], we developed the FSP-based Fisher information matrix (FSP-FIM), which allows one to use the FSP solution $\mathbf{p}(t)$, and its sensitivity $\mathbf{s}_{\theta_j} \equiv \mathrm{d}\mathbf{p}/\mathrm{d}\theta_j$, to find the FIM for stochastic gene expression systems. For a general FSP model, the dynamics of the sensitivity to each j^{th} kinetic parameter $\mathrm{d}\mathbf{p}/\mathrm{d}\theta_j$ can be calculated according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbf{p} \\ \mathbf{s}_{\theta_j} \end{bmatrix} = \begin{bmatrix} \mathbf{A}(t) & 0 \\ \mathbf{A}_{\theta_j}(t) & \mathbf{A}(t) \end{bmatrix}\begin{bmatrix} \mathbf{p} \\ \mathbf{s}_{\theta_j} \end{bmatrix}, \tag{6}$$

where $\mathbf{A}_{\theta_j} = \partial \mathbf{A}/\partial \theta_j$. Solving equation (6) requires integrating a coupled set of ODEs that is twice as large as the original FSP system. The FSP-FIM at a single time t is then given by

$$\mathbf{F}(\boldsymbol{\theta}, t)_{j,k} = \sum_i \frac{1}{p\left(\mathbf{x}_i; t, \boldsymbol{\theta}\right)} \mathbf{s}_{\theta_j}^i(t) \mathbf{s}_{\theta_k}^i(t), \tag{7}$$

where the summation is taken over all states $\{\mathbf{x}_i\}$ included in the FSP analysis (or over all observed states $\{\mathbf{x}_i^L\}$ in the case of lumped observations). We note that the FSP computation of the FIM should be computationally tractable for problems for which the FSP solution itself is tractable. However, since the size of the FSP sensitivity matrix (equation (6)) scales exponentially with the number of species, practical applications of the presented formulation of the FSP-FIM are currently restricted to models that have, or can be reduced to have, three or fewer distinct chemical species.

The FIM for a sequence of measurements taken independently (e.g., for smFISH data) at times $\mathbf{t} = [t_1, t_2, \ldots, t_{N_t}]$ can be calculated as the sum across the measurement times:

$$\mathcal{I}(\boldsymbol{\theta}, \mathbf{t}, \mathbf{c}) = \sum_{l=1}^{N_t} c_l \mathbf{F}\left(\boldsymbol{\theta}, t = t_l\right), \tag{8}$$

where $\mathbf{c} = [c_1, c_2, \ldots, c_{N_t}]$ is the number of cells measured at each l^{th} measurement time. For smFISH experiments, the vector $\mathbf{c}$ plays an important role in the design of the study. By optimizing over all vectors $\mathbf{c}$ that sum to N_{total}, one can find how many cells should be measured at each time point and which time points should be skipped entirely (i.e., $c_l = 0$).

In the next section, we verify the FSP-FIM for this stochastic model with a time-varying parameter and later find the optimal **c** for *STL*1 mRNA in yeast cells.

3.2. The FSP-FIM Can Quantify Experimental Information for Stochastic Gene Expression under Time-Varying Inputs. Our work in [6] was limited to models of stochastic gene expression that had piecewise constant reaction rates. Here, we extend this to time-varying reaction rates that affect the promoter switching in the system and which lead to time-varying $\mathbf{A}(t)$ in equation (6). For example, in the model depicted in Figure 1(a), the temporal addition of osmotic shock causes nuclear translocation of HOG1-MAPK, according to the time-varying function in equation (1).

Model parameters simultaneously fit to experimentally measured 0.2 M and 0.4 M *STL*1 mRNA were adopted from [10] and used as a reference set of parameters (yellow dots in Figure 2(a) and S1), which we define as $\boldsymbol{\theta}^*$. These reference parameters were used to generate 50 unique and independent simulated datasets, and each n^{th} simulated dataset was fit to find the parameter set, $\widehat{\boldsymbol{\theta}}_n$, that maximizes the likelihood for that simulated dataset. This process was repeated for two different experiment designs, including the original intuitive design from [10] (results shown in Figure 2) and an optimized design discussed below (results shown in Figure S1). To ease the computational burden of this fitting, the four parameters with the smallest sensitivities and largest uncertainties (i.e., those parameters that had the least effect on the model predictions and which were most difficult to identify) were fixed at their baseline values. The resulting MLE estimates for the remaining five parameters were collected into a set of $\{\widehat{\boldsymbol{\theta}}_n\}$ and are shown as yellow dots in Figures 2 and S1. Using the asymptotic normality of the maximum likelihood estimator and its relationship to the FIM (equation (5)), we then compared the 95% confidence intervals (CIs) of the inverse of the Fisher information (i.e., the Cramér–Rao bound) to those of the MLE estimates (compare the purple and orange ellipses in Figures 2(a) and S1a). We also compared the eigenvalues of the inverse of the Fisher information, $\{v_i\}$, to the correspondingly ranked eigenvalues of the covariance matrix of MLE estimates, Σ_{MLE}, in Figures 2(b) and S1b. For further validation, we noted that the principle directions of the ellipses in Figures 2(a) and S1a also match for the FIM and MLE analyses, as quantified by the angle between the paired FIM and Σ_{MLE} eigenvectors (Figures 2(b) and S1b). For comparison, the angles between rank-matched eigenvectors of the FIM and Σ_{MLE} were all less than 12°, whereas non-rank-matched eigenvectors were all greater than 79.9°. With the FSP-FIM verified for the HOG1-MAPK-induced gene expression model, we next explore how the FSP-FIM can be used to optimally allocate the number of cells to measure at each time after osmotic shock.

3.3. Designing Optimal Measurements for the HOG1-MAPK Pathway in S. cerevisiae. To explore the use of the FSP-FIM for experiment design in a realistic context of MAPK-activated gene expression, we again utilize simulated time-course smFISH data for the osmotic stress response in yeast.

We start with a known set of underlying model parameters that were taken from simultaneous fits to 0.2 M and 0.4 M data in [10] (nonspatial model) to establish a baseline parameter set that is experimentally realistic. These parameters are then used to optimize the allocation of measurements at different time points $t = [1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55]$ minutes after NaCl induction. Specifically, we ask what fraction of the total number of cells should be measured at each time to maximize the information about a specific subset of important model parameters. We use a specific experiment design objective criteria referred to as D_s-optimality, which corresponds to minimizing the expected volume of the parameter space uncertainty for the specific parameters of interest [35] and which is found by maximizing the product of the eigenvalues of the FIM for those same parameters.

Mathematically, our goal is to find the optimal cell measurement allocation:

$$\mathbf{c}_{\text{opt}} = \arg\max_{c} |\mathcal{I}(\mathbf{c};\boldsymbol{\theta})|_{D_s} \text{such that} \sum_{l=1}^{N_t} c_l = 1, \tag{9}$$

where c_l is the fraction of total measurements to be allocated at $t = t_l$, and the metric $|\mathcal{I}(\mathbf{c};\boldsymbol{\theta})|_{D_s}$ refers to the product of the eigenvalues for the total FIM (equation (8)). The fraction of cells to be measured at each time point, **c**, was optimized using a greedy search, in which single-cell measurements were chosen one at a time according to which time point predicted the greatest improvement in the optimization criteria (see Supplementary Note 3 for more information).

To illustrate our approach, we first allocated cell measurements according to D_s-optimality as found through this greedy search. Figure 3 shows the optimal fraction of cells to be measured at each time following a 0.2 M NaCl input and compares these fractions to the experimentally measured number of cells from [10]. While each available time point was allocated a nonzero fraction of measurements, three time points at $t = [10, 15, 30]$ minutes were vastly more informative than the other potential time points. To verify this result, we simulated 50 datasets of 1,000 cells each and found the MLE estimates for each subsampled dataset. We compared the spread of these MLE estimates to the inverse of the optimized FIM, shown in Figure S1.

Comparing Figure S1 with Figure 2 illustrates the extent by which the design of optimal measurement times for a 0.2 M NaCl experiment can increase information collection and reduce parameter uncertainties compared to the intuitive measurment design from [10]. In addition to providing much higher Fisher information, the optimal experiment requires measurement of only three time points compared to the 16 time points that were measured in the original experiment. Furthermore, we note that the FIM prediction of the MLE uncertainty is more accurate for the simpler optimal design, which is likely related to our observation that MLE estimates converge more easily for the optimized experiment design than they do for the original intuitive design.

Figure 4 next compares the D_s-optimality criteria for the optimal (solid horizontal lines) and intuitive ([10], dashed horizontal lines) experiment designs to 1,000 randomly designed experiments for the 0.2 M (black) and 0.4 M (gray)

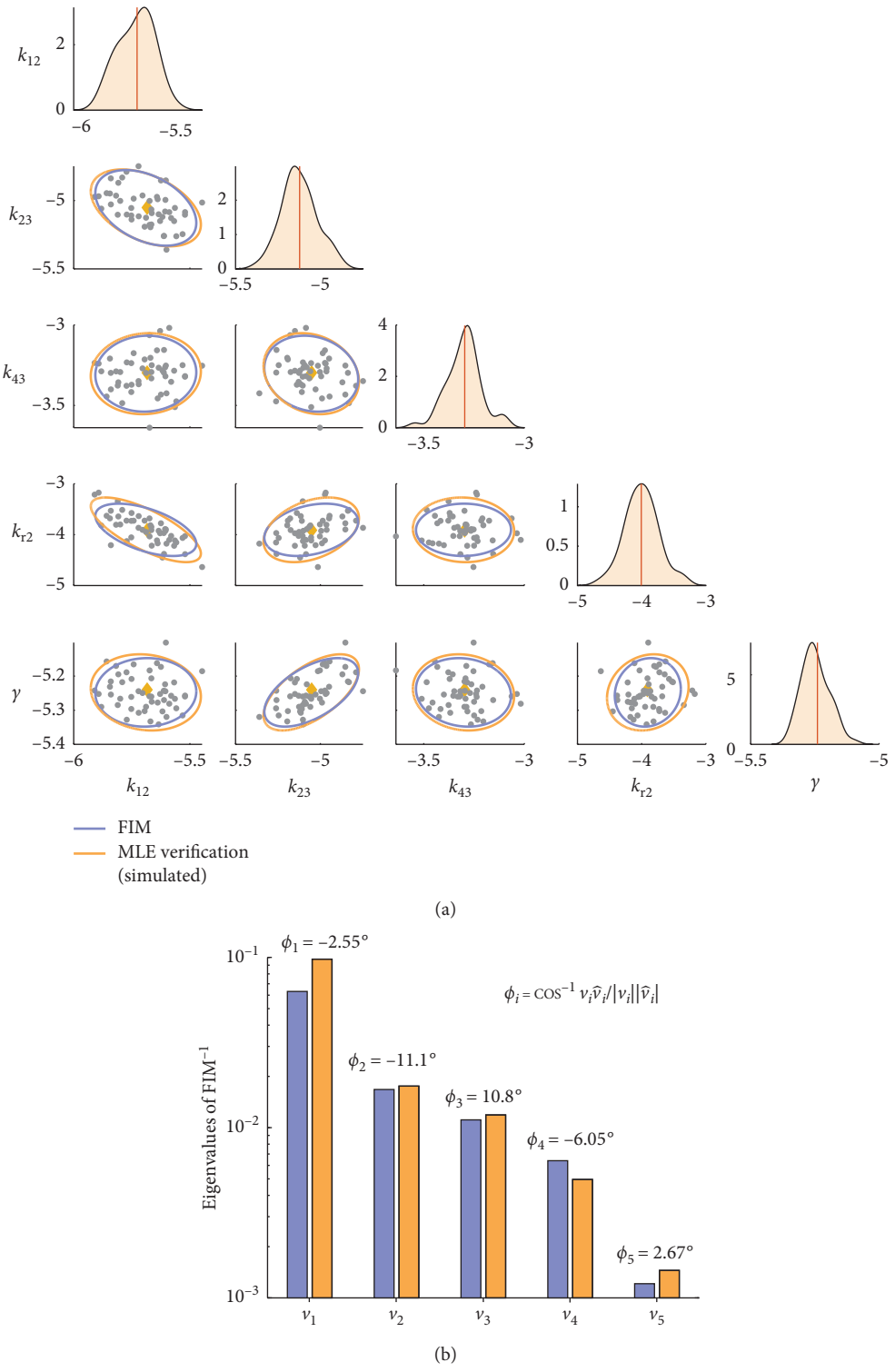


Figure 2: Verification of the FSP-FIM for the time-varying HOG1-MAPK model. (a) Marginal parameter histograms (top panels) and joint scatter plots (gray dots) for the MLE parameter estimates from 50 simulated datasets and for a subset of model parameters. All parameters are shown in logarithmic scale. The ellipses show the 95% CI for the inverse of the FIM (purple) and Gaussian approximation of MLE scatter plot (orange). The yellow dots indicate the "true" parameters at which the FIM and simulated datasets were generated. (b) Rank-paired eigenvalues (v_i) for the covariance of MLE estimates (orange) and inverse of the FIM (blue). The angles between corresponding rank-paired eigenvectors (ϕ_i) are shown in degrees.

conditions. To generate these random experiment designs, we selected a random subset of the measurement times and allocated the total 1,000 cells among chosen time points using a multinomial distribution with equal probability for each time point. Figure 4(a) shows that the intuitive experiment is more informative than most random

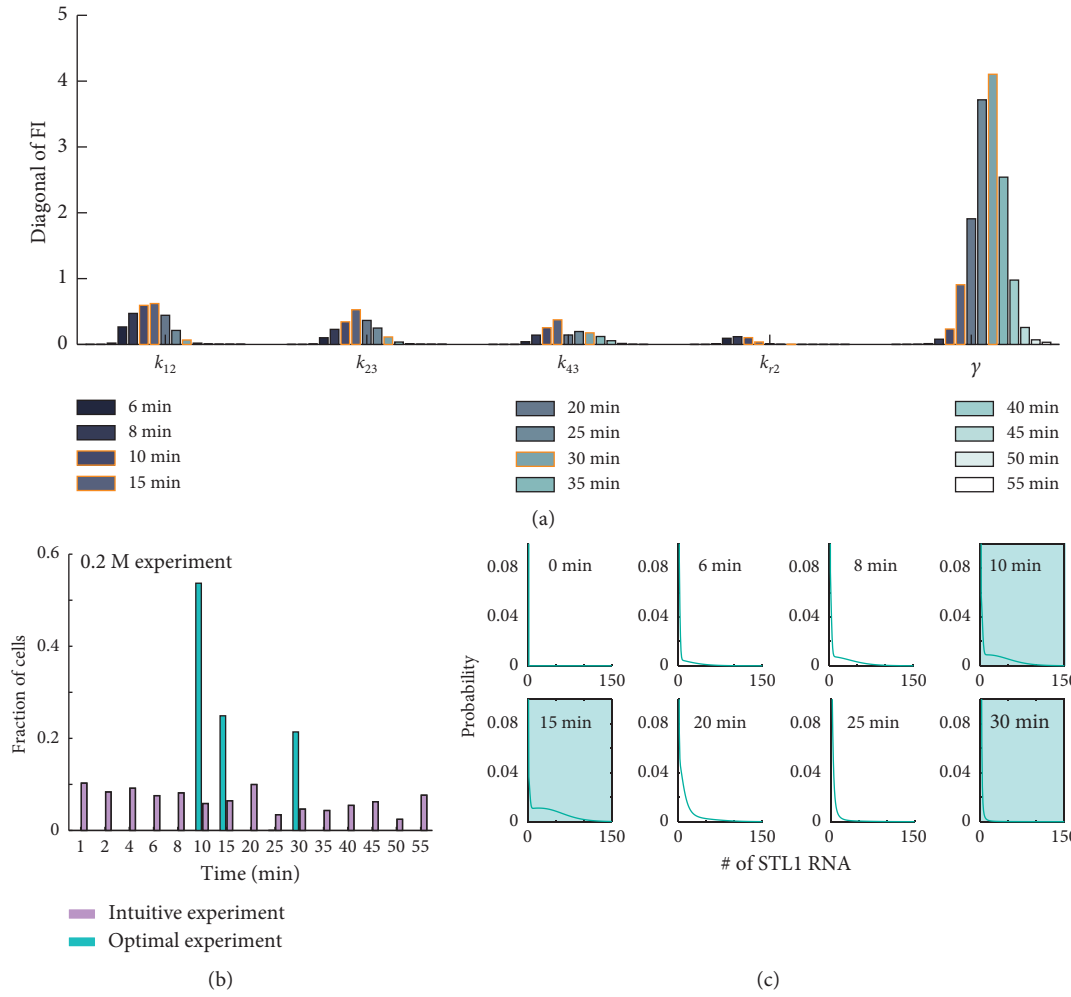


FIGURE 3: Optimizing the allocation of cell measurements at different time points. (a) Diagonal entries of the Fisher information at different measurement times. The optimal measurement times $t = [10, 15, 30]$ minutes are highlighted in orange. (b) Comparison of optimal fractions of cells to measure (blue) at different time points determined by the FSP-FIM compared to experimentally measured numbers of cells at 0.2 M NaCl (purple) from our work in [10]. (c) Probability distributions of *STL*1 mRNA at several of measurement times. The blue boxes denote the time points of optimal measurements.

experiments but is still substantially less informative than the optimal experiment.

In many practical applications, a scientist would be unlikely to have precise *a priori* knowledge of model parameters prior to conducting experiments. Rather, they would have some estimate of these parameters, such as rough knowledge of appropriate time scales or existing data from another type of experiment. Such estimates could come from previous analyses of the system response to simpler experimental conditions, for measurements taken on slightly different cell lines or organisms, or considering results from different genes in related regulatory pathways. To explore the importance of knowing the exact process parameters or input dynamics prior to designing the experiment, we asked how well an experiment design optimized using parameters from one gene at a given level osmotic shock (e.g., *STL*1 at 0.2 M NaCl) would do to estimate parameters for another gene in a different osmotic shock condition (e.g., *CTT*1 at 0.4 M NaCl). Figure 4(b) demonstrates the impact of such mismatched experiment designs, where each row corresponds to a different intuitive or optimized experiment design (i.e., a specific allocation of cells to be measured at each time), and each column corresponds to a specific gene and specific osmotic shock condition to which that design could be applied. In all cases, the much simpler FIM-based optimal experiment designs perform as well or better than the more difficult intuitive designs, even when these FIM designs were computed assuming different environmental conditions and assuming genes whose parameters differ considerably from one another (see Supplemental Tables 1 and 2 for parameter sets). In other words, these results suggest that if one can compute a simple yet optimal experiment design based on one well-analyzed gene in a previously studied environmental condition, then that design may be equally effective when applied to new investigations for related genes in similar biological contexts.

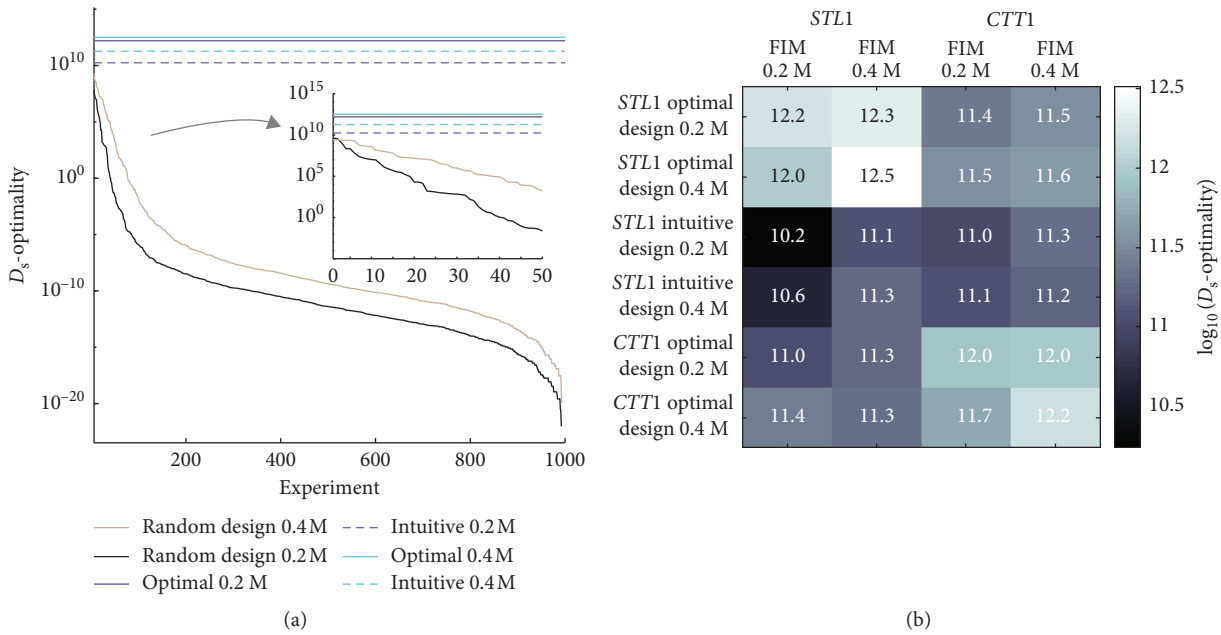


FIGURE 4: Information gained by performing optimal experiments compared to actual experiments. (a) D_s-optimality for optimal design using three time points compared to the intuitive experiment designs made using 16 time points is shown with horizontal lines (purple, 0.2 M, and blue, 0.4 M). Solid horizontal lines denote the optimal designs and dashed lines represent intuitive experiment designs. Randomly designed experiments with 0.2 M and 0.4 M NaCl are shown in black and orange. For the random experiments, the time points were selected by sampling them from the experimental measurement times, and then a random number of measurements were assigned to each selected time point. The inset shows the first 50 randomly designed experiments. (b) The D_s-metric for different experiment designs (different rows) when applied to different genes or different experimental levels of osmotic shock (different columns). Lighter shades (higher D_s-metrics) indicate experimental designs that are more suitable to identify parameters.

3.4. Using the FSP-FIM to Design Optimal Biosensor Measurements. Thus far, and throughout our previous work in [6], we have sought to find the optimal set of experiments to reduce uncertainty in the estimates of *model parameters*. In this section, we discuss how the FSP-FIM allows for the optimization of experiment designs to address a more general problem of inferring *environmental variables* from cellular responses. Toward this end, we assume a known and parametrized model (i.e., the model defined above, which was identified previously in [10]), but which is now subject to unknown environmental influences. We explore what would be the optimal experimental measurements to take to characterize these influences. Specifically, we ask how many cells should be measured using smFISH, and at what times, to determine the specific concentration of NaCl to which the cells have been subjected—or, equivalently, we ask what experiments would be best suited to measure the effective stress induction level caused by addition of an unknown solution to the cells.

Recall from above that in the HOG1-MAPK transcription model, extracellular osmolarity ultimately affects stress response gene transcription levels through the time-varying parameter $k_{21}(t)$ (equation (1)) as illustrated in Figure 1(c) for 0.2 M and 0.4 M salt concentrations. Higher salt concentrations delay the time at which $k_{21}(t)$ returns to its nonzero value. The function in equation (1) can be coarsely approximated by the sum of three Heaviside step functions, $u(t-\tau_i)$ as

$$k_{21}(t) = k_{21}^{0}\left(u(t) - u(t-\tau_1) + u(t-\tau_2)\right), \tag{10}$$

where τ_1 is the fixed delay of the time it takes for nuclear kinase levels to reach the k_{21} deactivation threshold (about 1 minute or less, [9, 10]) and τ_2 is the variable time it takes for the nuclear kinase to drop back below that threshold. In practice, the threshold-crossing time, τ_2, should be directly related to the salt concentration experienced by the cell under reasonable salinity levels. This relationship is shown in Figures 1(b), 1(c), and 5(b), where a 0.2 M NaCl input exhibits a shorter τ_2 than does a 0.4 M input. For our analyses, we assume a prior uncertainty such that time τ_2 can be any value uniformly distributed between $\tau_2^{\min} = 6$ and $\tau_2^{\max} = 31$ minutes, and our goal is to find the experiment that best reduces the posterior uncertainty in τ_2 (and therefore could provide an estimate for the concentration of NaCl).

To reformulate the FSP-FIM to estimate uncertainty in τ_2 given our model, the first step is to compute the sensitivity of the distribution of mRNA abundance to changes in the variable τ_2 using equation (5), in which $\mathbf{A}_{\theta_j}(t)$ is replaced with $\mathbf{A}_{\tau_2}(t) = \partial\mathbf{A}/\partial\tau_2$ as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{p}\\ \mathbf{s}_{\tau_2}\end{bmatrix} = \begin{bmatrix}\mathbf{A}(t) & 0\\ \mathbf{A}_{\tau_2}(t) & \mathbf{A}(t)\end{bmatrix}\begin{bmatrix}\mathbf{p}\\ \mathbf{s}_{\tau_2}\end{bmatrix}. \tag{11}$$

As $k_{21}(t)$ is the only parameter in $\mathbf{A}$ that depends explicitly on τ_2, all entries of $\partial\mathbf{A}/\partial\tau_2$ are zero except for those which depend on $k_{21}(t)$, and

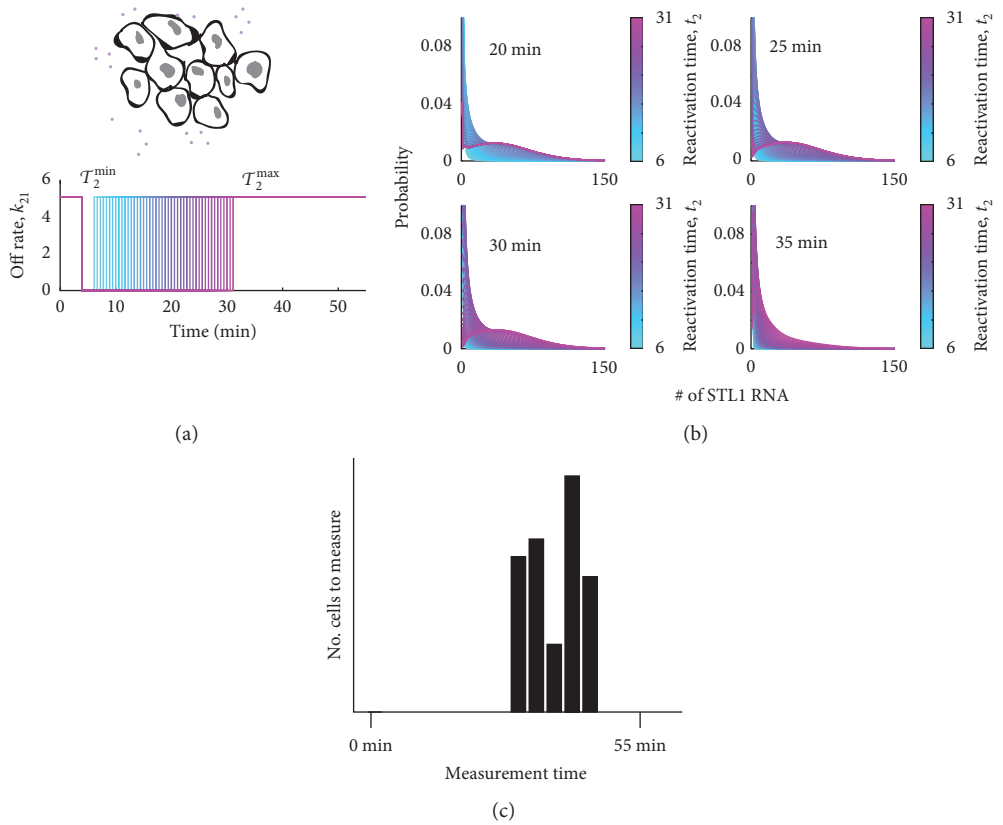


FIGURE 5: Overview of optimal design for biosensing experiments for the osmotic stress response in yeast. (a) Unknown salt concentrations (purple dots) in the environment give rise to different reactivation times, τ_2, which affect the gene expression in the model through the rate k_{21}. These different reactivation times cause downstream *STL*1 expression dynamics to behave differently as shown in (b). (c) Different responses can be used to resolve experiments that reduce the uncertainty in τ_2.

$$\mathbf{A}_{\tau_2}(t) = \frac{\partial \mathbf{A}}{\partial k_{21}} \frac{\partial k_{21}}{\partial \tau_2} = \mathbf{A}_{k_{21}} k_{21}^0 \delta(\tau_2), \tag{12}$$

and therefore $\mathbf{A}_{\tau_2} = \partial \mathbf{A}/\partial \tau_2$ is nonzero only at $t = \tau_2$. Using this fact, the equation for the sensitivity dynamics is uncoupled from the FSP dynamics for $t \neq \tau_2$ and can be written simply as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{s}_{\tau_2} = \begin{cases} 0 \text{ for } t < \tau_2 \text{ with } \mathbf{s}(0) = 0, \\ \mathbf{A}(t)\mathbf{s}_{\tau_2} \text{ for } t > \tau_2 \text{ with } \mathbf{s}_{\tau_2}(\tau_2) = k_{21}^0 \mathbf{A}_{k_{21}} \mathbf{p}(\tau_2). \end{cases} \tag{13}$$

If the Fisher information at each measurement time is written into a vector $\mathbf{f} = [f_1, f_2, \ldots, f_{N_t}]$ (noting that the Fisher information at any time t_l is the scalar quantity, f_l) and the number of measurements per time point is the vector, $\mathbf{c} = [c_1, c_2, \ldots, c_{N_t}]$, then the total information for a given value of τ_2 can be computed as the dot product of these two vectors:

$$\mathscr{I}(\tau_2) = \sum_{l=1}^{N_t} c_l f_l = \mathbf{c}^T \mathbf{f}. \tag{14}$$

Our goal is to find an experiment that is optimal to determine the value of τ_2, given an assumed prior that τ_2 is sampled from a uniform distribution between $\tau_2^{\min}$ and $\tau_2^{\max}$. To find the experiment $\mathbf{c}_{\text{opt}}$ that will reduce our posterior uncertainty in τ_2, we integrate the inverse of the FIM in equation (14) over the prior uncertainty in τ_2:

$$\begin{aligned} \mathbf{c}_{\text{opt}} &= \underset{\mathbf{c}, \sum c_l = 1}{\arg\min} \int_{\tau_2^{\min}}^{\tau_2^{\max}} \frac{1}{\tau_2^{\max} - \tau_2^{\min}} \mathscr{I}^{-1}(\mathbf{c}; \tau_2 = \tau, \boldsymbol{\theta}) \mathrm{d}\tau, \\ &= \underset{\mathbf{c}, \sum c_l = 1}{\arg\min} \int_{\tau_2^{\min}}^{\tau_2^{\max}} \mathscr{I}^{-1}(\mathbf{c}; \tau_2 = \tau, \boldsymbol{\theta}) \mathrm{d}\tau. \end{aligned} \tag{15}$$

For later convenience, we define the integral in equation (15) (i.e., the objective function of the minimization) by the symbol $\mathscr{J}$, which corresponds to the expected uncertainty about the value of τ_2 for a given $\mathbf{c}$.

Next, we apply the greedy search from above to solve the minimization problem in equation (15) to find the experiment design $\mathbf{c}_{\text{opt}}$ that minimizes the estimation error of τ_2. Figure 6 shows examples of seven different experiments to accomplish this task, ranked according to the FSP-FIM value $\mathscr{J}$ from most informative (top left) to least informative (bottom left), but all using the same number of measured cells. For each experiment, the FSP-FIM was used to estimate the posterior uncertainty (i.e., expected standard deviation) in the estimation of τ_2, which is shown by the orange bars in

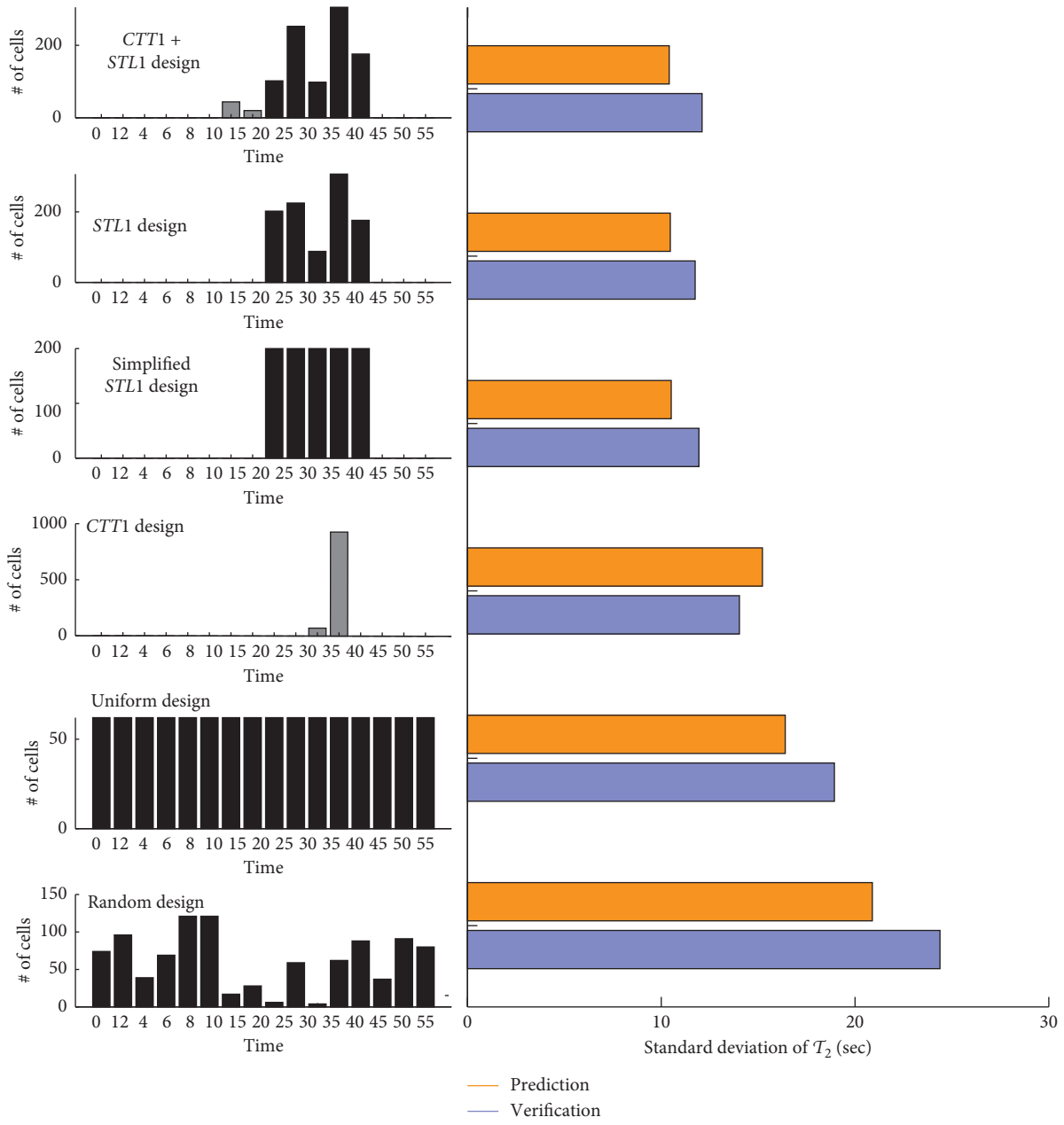


FIGURE 6: Verification of the uncertainty in τ_2 for different experiment designs. The left panel shows various experiment designs, where the sum of the bars (i.e., the total number of measurements) is 1,000. Gray bars represent the measurements of *CTT*1 and black bars *STL*1. The right panel shows the value of the objective function in equation (15) for each experiment design in orange, and the RMSE values for verification are shown in purple.

Figure 6. To verify these estimates, we then chose 64 uniformly spaced values of τ_2, which we denote as the set $\{\tau_2^{\text{true}}\}$, and for each τ_2^{true}, we simulated 50 random datasets of 1,000 cells distributed according to the specified experiment designs. For each of the 64×50 simulated datasets, we then determined the value τ_2^{MLE} between τ_2^{min} and τ_2^{max} that maximized the likelihood of the simulated data according to equation (2). The root mean squared estimate (RMSE) error over all random values of τ_2^{true} and estimates, $\sqrt{\langle(\tau_2^{\text{MLE}} - \tau_2^{\text{true}})^2\rangle}$, was then computed for each of the six different experiment designs. Figure 6 shows that the FIM-based estimation of uncertainty and the actual MLE-based uncertainty are in excellent agreement for all experiments (compare purple and orange bars). Moreover, it is clear that the optimal design selected by the FIM analysis performed much better to estimate τ_2 than did the uniform or random experimental designs. A slightly simplified design, which uses the same time points as the optimal, but with equal numbers of measurements at each time, performed nearly as well as the optimal design.

The set of experiment designs shown in Figure 6 includes the best design that only uses *STL*1 (second from top), the best design that uses only *CTT*1 (fourth from top), and the best design that uses some cells with *CTT*1 and some with

*STL*1 (top design). To find the best experiment design for measurement of two different genes, we assumed that at each time, either *STL*1 mRNA *or CTT*1 mRNA (but not both) could be measured, corresponding to using smFISH oligonucleotides for either *STL*1 or *CTT*1. To determine which gene should be measured at each time, we compute the Fisher information for *CTT*1 and *STL*1 for every measurement time and averaged this value over the range of τ_2. For each measurement time t_l, the gene is selected that has the higher average Fisher information for τ_2. The number of cells per measurement time was then optimized as before, except the choice to measure *CTT*1 or *STL*1 was based on which mRNA had the larger Fisher information (equation (14)) at that specific point in time. The best *STL*1-only experiment design was found to yield uncertainty of 10.5 seconds (standard deviation); the best *CTT*1-only experiment was found to yield an uncertainty of 15.2 seconds and the best mixed *STL*1/*CTT*1 experiment design was found to yield an uncertainty of 10.4 seconds. In other words, for this case, the *STL*1 gene was found to be much more informative of the environmental condition than was *CTT*1, and the use of both *STL*1 and *CTT*1 provides only minimal improvement beyond the use of *STL*1 alone. We note that although measurement times in the optimized experiment design were restricted to a resolution of five minutes or more, the value of τ_2 could be estimated with an error of only 10 seconds, corresponding to a roughly 30-fold improvement of temporal resolution beyond the allowable sampling rate.

3.5. Experimental Validation for FSP-FIM-Based Designs of Biosensor Measurements. To experimentally validate our FSP-FIM-based approach to design optimal measurement times, we next examined experimental smFISH data taken for the *STL*1 and *CTT*1 genes at different times following yeast osmotic shock [10]. These data include a total of 535–4808 cells measured at each of 16 time points following osmotic shocks of 0.2 M or 0.4 M NaCl. We asked how well could we identify the concentration of the osmotic shock from the experimental data using only 75 individual cells per experiment. We again proposed the six different potential experiments depicted in Figure 6, including the optimal *STL*1 and *CTT*1 design, the optimal *STL*1 design, the simplified *STL*1 design with 15 cells for each of the optimal five time points, the optimal *CTT*1 design, the uniform *STL*1 design, and the random STL1 design. For each design, we created 1,000 different experimental replica datasets, each consisting of 100 cells randomly chosen from the original data. For each replica dataset, we then used the CME model (Supplementary Note 1) with a parametrized form of the HOG1-MAPK nuclear localization signal (Supplementary Note 2) to find the NaCl concentration that maximizes the likelihood of the data given the model.

Figure 7 shows the resulting histograms for the estimated NaCl concentrations for each of the six experiment designs, when the cells were actually subjected to experimental osmotic shocks of 0.2 M NaCl (Figure 7(a)) or 0.4 M NaCl (Figure 7(c)). From Figures 7(a) and 7(c), it is clear that the FSP analysis provides an accurate estimate for the level of the osmotic shock input using a relatively small number of cells, despite the fact that producing such estimates was not an intended use of the model in its original formulation or parameter inference [9, 10]. Figures 7(b) and 7(d) show the uncertainty (standard deviation) in the experimental estimate of NaCl concentration (light bars), when cells are collected according to the six specific experiment designs, and compare these results to the FSP-FIM uncertainty estimates (dark bars) using the simplified step input function (equation (10)). With the exception of the suboptimal *CTT*1-only design, the close matches between the relative trends of the variance in experimental estimation of NaCl and the variance predicted by the FSP-FIM analysis with the approximated step-function input give further experimental validation that the FSP-FIM approach can be used to choose more informative experiment designs, even in cases where the FSP analyses use inexact assumptions for model kinetics. The single discrepancy in trends led us to more closely examine the model and experimental data for *CTT*1 expression at the 35-minute time point that dominates the *CTT*1-only design. By examining Supplemental Figure S7 from [10], we found that this specific combination of *CTT*1 at 35 minutes following 0.4 M NaCl osmotic shock showed a greater discrepancy between model and data than any of the other 63 combinations of 16 times, two genes, and two conditions, yet it is unclear if that difference was an artifact of the experiment or an actual transient effect that only affected that specific combination of gene, time, and environmental condition.

4. Discussion

The methods developed in this work present a principled, model-driven approach to allocate how many snapshot single-cell measurements should be taken at each time during analysis of a time-varying stochastic gene regulation system. We demonstrate and verify these theories on a well-established model of osmotic stress response in yeast cells, which is activated upon the nuclear localization of phosphorylated HOG1 [9, 10]. For this system, we showed how to optimally allocate the number of cells measured at each time so as to maximize the information about a subset of model parameters. We found that the optimal experiment design to estimate model parameters for the *STL*1 gene only required three time points. Moreover, these three time points ($t = [10, 15, 30]$ minutes, highlighted by blue in Figure 3(b)) are at biologically meaningful time points. At $t = 10$ and 15 minutes, the system is increasing to maximal expression, and the probability to measure a cell with elevated mRNA content is high, which helps reduce uncertainty about the parameters in the model that control maximal expression. Similarly, at the final experiment time of $t = 30$ minutes, the system is starting to shut down gene expression, and therefore this time is valuable to learn about the time scale of deactivation in the system as well as the mRNA degradation rate. These effects are clearly illustrated in Figure 3(a), which shows that times $t = 10$ and $t = 15$ minutes provide the most information about parameters k_{12}, k_{23}, and k_{43}, whereas measurements at $t = 30$ minutes provide the most

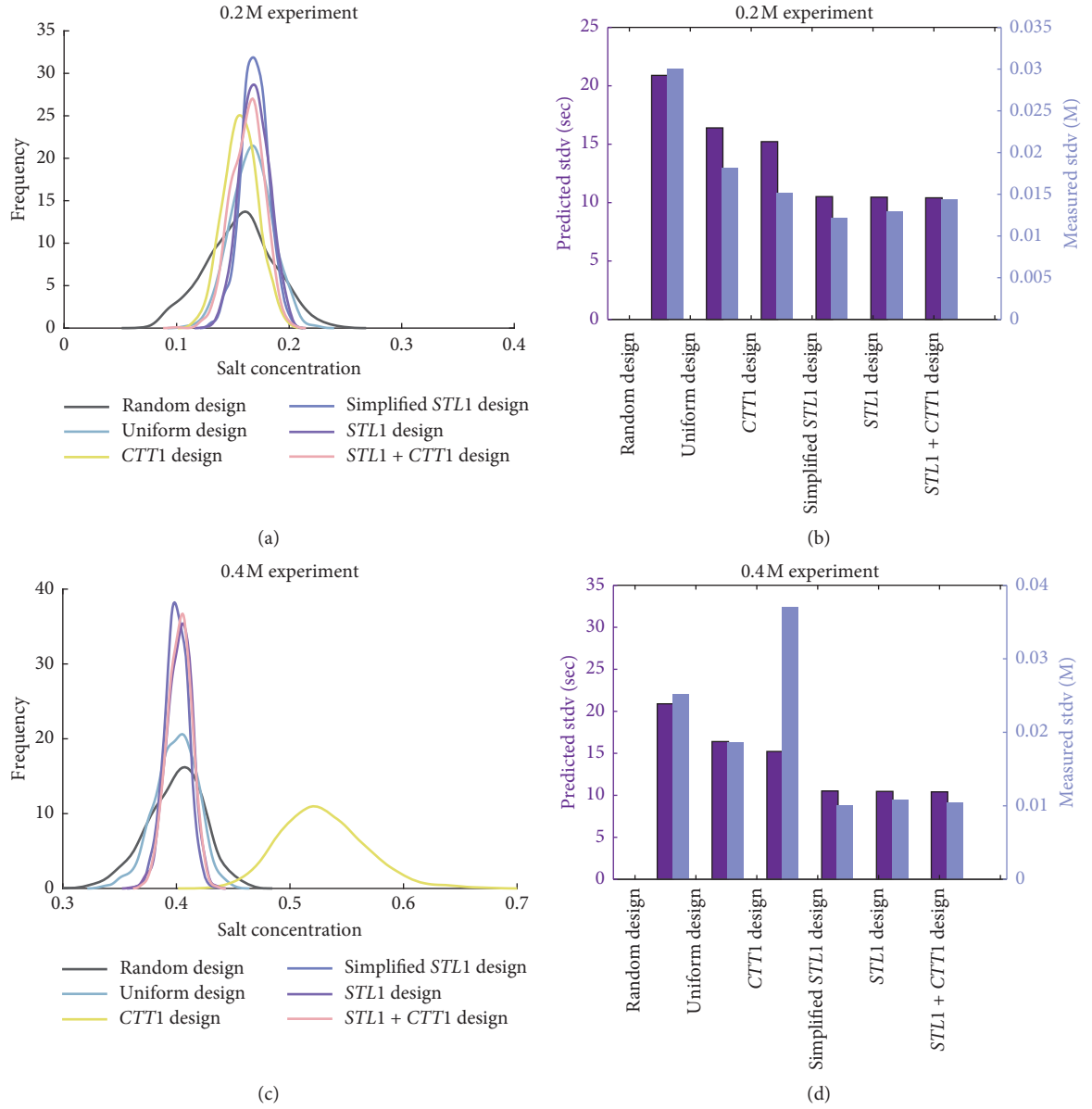


FIGURE 7: Experimental validation of FSP-FIM-based design for optimal biosensor measurements. (a) Distribution of FSP-based MLE estimates for NaCl concentration using the six experimental designs from Figure 6. Each distribution comes from 1,000 replicas of 75 cells per replica spread out over the possible 16 time points. Replica data were sampled randomly from published experimental data [10] that contain two or three biological replicas and 535–4808 cells per time point. The true experimentally applied level of osmotic shock was 0.2 M NaCl. (b) The MLE estimation standard deviation for each experiment design applied to a dataset taken at 0.2 M NaCl (blue). These deviations are compared to FSP-FIM deviation predictions using a piecewise constant model for HOG1 nuclear localization (purple). (c, d) Same as (a, b) but for a true NaCl concentration of 0.4 M.

information about γ. Because γ is the easiest parameter to estimate (e.g., its information is greater), not as many cells are needed at $t = 30$ minutes to constrain that parameter. Similarly, because k_{r2} is the most difficult parameter to estimate (e.g., it has the lowest information across all experiments) and because $t = 10$ minutes is one of the few time points to provide information about k_{r2}, the optimal experimental design selects a large number of cells at the time $t = 10$ minutes. This analysis demonstrates that the optimal experiment design can change depending upon which parameters are most important to determine (e.g., γ or k_{r2} in this case), a fact that we expect will be important to consider in future experiment designs.

Because we constrained all potential experiment designs to be within the subset of experiments performed in our previous work [10], we are able to compare the information of optimal experiment designs to intuitive designs that have actually been performed. We found that while the intuitive experiments were almost always better than could be expected by random chance, they still provided several orders of magnitude lower Fisher information than would be possible with optimal experiments (Figure 4(a)). Moreover,

in our analyses, we found that optimal designs could require far fewer time points than those designed by intuition (e.g., only three time points were needed in Figure 3), and therefore these designs can be much easier and less expensive to conduct. We also found that utility of optimal experiment designs could be relatively insensitive to variation in the experimental conditions or the specific model parameters used for the experiment design. For example, we found that experiments optimized for one gene at one level of osmotic shock were still at least as good—and in most cases better—than intuitive designs, even when conducted using different genes and at a different level of osmotic shock (Figure 4(b)). In practice, this fact would allow for effective experiment designs despite inaccurate prior assumptions.

In addition to suggesting optimal experiments to identify model parameters, we showed that the FSP approach could be used to infer parameters of fluctuating extracellular environments from single-cell data and that the FSP-FIM combined with an existing model could be used to design optimal experiments to improve this inference (Figures 5 and 6). We experimentally verified this potential by examining many small sets of single-cell smFISH measurements for different genes and different measurement times, and we showed that an FSP-FIM analysis could correctly rank which experiment designs would give the best estimates of osmotic shock environmental conditions. Along a very similar line of reasoning, one can also adapt the FSP-FIM analysis to learn what biological design parameters would be optimal to reduce uncertainty in the estimate of important environmental variables. For example, Figure 8 shows the expected uncertainty in τ_2 as a function of the degradation rate of the *STL*1 gene assuming that 50 cells could be measured at each experimental measurement time $t = [1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55]$ minutes using the smFISH approach. We found that the best choice for *STL*1 degradation rate to most accurately determine the extracellular fluctuations would be 2.4×10^{-3} mRNA/min, which is about half of the experimentally determined value of $5.3 \times 10^{-3} \pm 5.9 \times 10^{-5}$ from [10]. This result is consistent with our earlier finding that the faster degrading *STL*1 mRNA is a much better determinant of the HOG1 dynamics than the slower-degrading *CTT*1 mRNA and suggests that other less stable mRNA could be more effective still. We expect that similar, future applications of the FSP-based Fisher information will be valuable in other systems and synthetic biology contexts where scientists seek to explore how different cellular properties affect the transmission of information between cells or from cells to human observers. Indeed, similar ideas have been explored recently using classical information theory in [36–39], and recent work in [7, 40] has noted the close relationship between Fisher information and the channel capacity of biochemical signaling networks.

We expect that computing optimal experiment designs for time-varying stochastic gene expression will create opportunities that could extend well beyond the examples presented in this work. Modern experimental systems are making it much easier for scientists and engineers to precisely perturb cellular environments using chemical induction [41–43] or optogenetic control [44–46]. Many such experiments involve stochastic bursting behaviors at the mRNA or protein level [8–10, 45], and precise optimal experiment design will be crucial to understand the properties of stochastic variations in such systems. A related field that is also likely to benefit from such approaches is biomolecular image processing and feedback control, for which one may need to decide in real time which measurements to make and in what conditions.

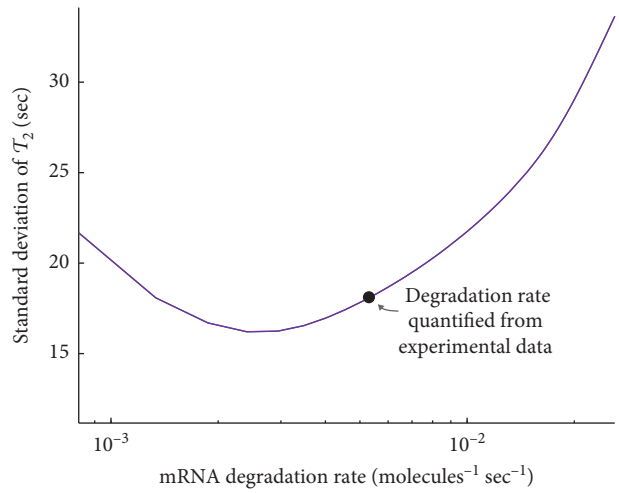


Figure 8: Optimal mRNA degradation rates to reduce uncertainty about the extracellular environment. Uncertainty in the time at which the *STL*1 gene turns off, τ_2, as a function of mRNA degradation rate (purple). The black dot corresponds to the degradation rate that was quantified from experimental data.

Data Availability

All data and codes associated with this article are available at https://github.com/MunskyGroup/Fox_Complexity_2020.

Disclosure

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

ZRF and BM were supported by National Institutes of Health (R35 GM124747). ZRF was also supported by the Agence Nationale de la Recherche (ANR-18-CE91-0002, CyberCircuits). GN was supported by the National Institutes of Health (DP2 GM11484901 and R01GM115892) and Vanderbilt Startup Funds. The presented analyses used the computational resources of the WM Keck High Performance Computing Cluster supported under a WM Keck Foundation Award.

Supplementary Materials

Supplementary note 1: stochastic model of yeast stress response. Supplementary note 2: nuclear localization of HOG-MAPK. Supplementary note 3: optimization of cell measurements. Table I: HOG-MAPK model parameters. Table II: HOG-signaling model parameters. Figure 1: verification of the FSP-FIM for the time-varying HOG-MAPK model. (*Supplementary Materials*)

References

[1] J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf, "Maximizing the Information Content of Experiments in Systems Biology," *PLoS Computational Biology*, vol. 9, no. 1, Article ID e1002888, 2013.

[2] J. F. Apgar, D. K. Witmer, F. M. White, and B. Tidor, "Sloppy models, parameter uncertainty, and the role of experimental design," *Molecular BioSystems*, vol. 6, no. 10, p. 1890, 2010.

[3] J. Ruess, A. Milias-Argeitis, and J. Lygeros, "Designing experiments to understand the variability in biochemical reaction networks," *Journal of The Royal Society Interface*, vol. 10, no. 88, Article ID 20130588, 2013.

[4] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models," *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. 8645–8650, 2011.

[5] C. Zimmer, "Experimental design for stochastic models of nonlinear signaling pathways using an interval-wise linear noise approximation and state estimation," *PLoS One*, vol. 11, no. 9, Article ID e0159902, 2016.

[6] Z. R. Fox and B. Munsky, "The finite state projection based Fisher information matrix approach to estimate information and optimize single-cell experiments," *PLoS Computational Biology*, vol. 15, no. 1, Article ID e1006365, 2019.

[7] V. Singh and I. Nemenman, "Universal properties of concentration sensing in large ligand-receptor networks," *Physical Review Letters*, vol. 124, no. 2, Article ID 028101, 2020.

[8] C. Zechner, J. Ruess, P. Krenn et al., "Moment-based inference predicts bimodality in transient gene expression," *Proceedings of the National Academy of Sciences*, vol. 109, no. 21, pp. 8340–8345, 2012.

[9] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, "Systematic identification of signal-activated stochastic gene regulation," *Science*, vol. 339, no. 6119, pp. 584–587, 2013.

[10] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, "Distribution shapes govern the discovery of predictive models for gene regulation," *Proceedings of the National Academy of Sciences*, vol. 115, no. 29, pp. 7533–7538, 2018.

[11] A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi, "Imaging individual mRNA molecules using multiple singly labeled probes," *Nature Methods*, vol. 5, no. 10, pp. 877–879, 2008.

[12] A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, "Visualization of single RNA transcripts in situ," *Science*, vol. 280, no. 5363, pp. 585–590, 1998.

[13] N. Tsanov, A. Samacoits, R. Chouaib et al., "smiFISH and FISH-quant—a flexible single RNA detection approach with super-resolution capability," *Nucleic Acids Research*, vol. 44, no. 22, p. e165, 2016.

[14] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, "Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings," *Nature Methods*, vol. 11, no. 2, pp. 197–202, 2014.

[15] R. M. Kumar, P. Cahan, A. K. Shalek et al., "Deconstructing transcriptional heterogeneity in pluripotent stem cells," *Nature*, vol. 516, no. 7529, pp. 56–61, 2014.

[16] L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer, "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity," *Cell*, vol. 122, no. 2, pp. 169–182, 2005.

[17] B. Munsky, G. Neuert, and A. van Oudenaarden, "Using gene expression noise to understand gene regulation," *Science*, vol. 336, no. 6078, pp. 183–187, 2012.

[18] H. Sharifian, F. Lampert, K. Stojanovski et al., "Parallel feedback loops control the basal activity of the HOG MAPK signaling cascade," *Integrative Biology*, vol. 7, no. 4, pp. 412–422, 2015.

[19] E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann, "Integrative model of the response of yeast to osmotic shock," *Nature Biotechnology*, vol. 23, no. 8, pp. 975–982, 2005.

[20] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," *Nature Biotechnology*, vol. 20, no. 4, pp. 370–375, 2002.

[21] D. Muzzey, C. A. Gómez-Uribe, J. T. Mettetal et al., "A systems-level analysis of perfect adaptation in yeast osmoregulation," *Journal of End-to-End-Testing*, vol. 138, no. 1, pp. 160–171, 2009.

[22] H. Saito and F. Posas, "Response to hyperosmotic stress," *Genetics*, vol. 192, no. 2, pp. 289–318, 2012.

[23] S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter, "Transient activation of the HOG MAPK pathway regulates bimodal gene expression," *Science*, vol. 332, no. 6030, pp. 732–735, 2011.

[24] B. Munsky, Z. Fox, and G. Neuert, "Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics," *Methods*, vol. 85, pp. 12–21, 2015.

[25] N. G. Van Kampen and N. Godfried, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, Netherlands, 1992.

[26] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[27] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, 2011.

[28] Z. Fox, G. Neuert, and B. Munsky, "Finite state projection based bounds to compare chemical master equation models using single-cell data," *The Journal of Chemical Physics*, vol. 145, no. 7, Article ID 074101, 2016.

[29] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *The Journal of Chemical Physics*, vol. 124, no. 4, Article ID 044104, 2006.

[30] M. Gomez-Schiavon, L.-F. Chen, A. E. West, and N. E. Buchler, "BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells," *Genome Biology*, vol. 18, no. 1, p. 164, 2017.

[31] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.
[32] G. Casella and R. L. Berger, *Statistical Inference*, Wadsworth and Brooks/Cole, Pacific Grove, CA, USA, 1990.
[33] C. Kreutz and J. Timmer, "Systems biology: experimental design," *FEBS Journal*, vol. 276, no. 4, pp. 923–942, 2009.
[34] B. Steiert, A. Raue, J. Timmer, and C. Kreutz, "Experimental design for parameter estimation of gene regulatory networks," *PLoS One*, vol. 7, no. 7, Article ID e40052, 2012.
[35] J. Ruess, F. Parise, A. Milias-Argeitis, M. Khammash, and J. Lygeros, "Iterative experiment design guides the characterization of a light-inducible gene expression circuit," *Proceedings of the National Academy of Sciences*, vol. 112, no. 26, pp. 8148–8153, 2015.
[36] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information transduction capacity of noisy biochemical signaling networks," *Science*, vol. 334, no. 6054, pp. 354–358, 2011.
[37] R. Suderman, J. A. Bachman, A. Smith, P. K. Sorger, and E. J. Deeds, "Fundamental trade-offs between information flow in single cells and cellular populations," *Proceedings of the National Academy of Sciences*, vol. 114, no. 22, pp. 5755–5760, 2017.
[38] J. Selimkhanov, B. Taylor, J. Yao et al., "Accurate information transmission through dynamic biochemical signaling networks," *Science*, vol. 346, no. 6215, pp. 1370–1373, 2014.
[39] G. Tkačik and A. M. Walczak, "Information transmission in genetic regulatory networks: a review," *Journal of Physics. Condensed Matter: An Institute of Physics Journal*, vol. 23, no. 15, Article ID 153102, 2011.
[40] T. Jetka, K. Nienałtowski, S. Filippi, M. P. H. Stumpf, and M. Komorowski, "An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling," *Nature Communications*, vol. 9, no. 1, p. 4591, 2018.
[41] A. H. Ng, T. H. Nguyen, M. Gómez-Schiavon et al., "Modular and tunable biological feedback control using a de novo protein switch," *Nature*, vol. 572, no. 7768, pp. 265–269, 2019.
[42] A. Thiemicke, H. Jashnsaz, G. Li, and G. Neuert, "Generating kinetic environments to study dynamic cellular processes in single cells," *Scientific Reports*, vol. 9, no. 1, Article ID 10129, 2019.
[43] J.-B. Lugagne, S. S. Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen, "Balancing a genetic toggle switch by real-time feedback control and periodic forcing," *Nature Communications*, vol. 8, no. 1, p. 1671, 2017.
[44] R. Chait, J. Ruess, T. Bergmiller, G. Tkačik, and C. C. Guet, "Shaping bacterial population behavior through computer-interfaced control of individual cells," *Nature Communications*, vol. 8, p. 2557, 2017.
[45] M. Rullan, D. Benzinger, G. W. Schmidt, A. Milias-Argeitis, and M. Khammash, "An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation," *Molecular Cell*, vol. 70, no. 4, pp. 745–756, 2018.
[46] S. M. Castillo-Hair, E. A. Baerman, M. Fujita, O. A. Igoshin, and J. J. Tabor, "Optogenetic control of Bacillus subtilis gene expression," *Nature Communications*, vol. 10, no. 1, p. 3099, 2019.

Hindawi
Complexity
Volume 2020, Article ID 5163704, 15 pages
https://doi.org/10.1155/2020/5163704

Research Article

A Nonoscillatory Second-Order Time-Stepping Procedure for Reaction-Diffusion Equations

Philku Lee,[1] **George V. Popescu**,[2,3] **and Seongjai Kim**[1]

[1]*Department of Mathematics and Statistics, Mississippi State University, Starkville, Mississippi State, MS 39762, USA*
[2]*Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, Mississippi State, MS 39762, USA*
[3]*The National Institute for Laser, Plasma & Radiation Physics, 077126 Măgurele, Ilfov, Romania*

Correspondence should be addressed to George V. Popescu; popescu@igbb.msstate.edu

Received 24 October 2019; Accepted 6 March 2020; Published 30 April 2020

Academic Editor: Lingzhong Guo

After a theory of morphogenesis in chemical cells was introduced in the 1950s, much attention had been devoted to the numerical solution of *reaction-diffusion* (RD) *partial differential equations* (PDEs). The *Crank–Nicolson* (CN) method has been a common second-order time-stepping procedure. However, the CN method may introduce spurious oscillations for nonsmooth data unless the time step size is sufficiently small. This article studies a nonoscillatory second-order time-stepping procedure for RD equations, called a *variable-θmethod*, as a perturbation of the CN method. In each time level, the new method detects points of potential oscillations to implicitly resolve the solution there. The proposed time-stepping procedure is nonoscillatory and of a second-order temporal accuracy. Various examples are given to show effectiveness of the method. The article also performs a sensitivity analysis for the numerical solution of biological pattern forming models to conclude that the numerical solution is much more sensitive to the spatial mesh resolution than the temporal one. As the spatial resolution becomes higher for an improved accuracy, the CN method may produce spurious oscillations, while the proposed method results in stable solutions.

1. Introduction

As molecular imaging and single cell analysis is advancing our understanding of spatial processes shaping the cellular dynamics, new models of nonlinear dynamics are necessary. Originating in study of organism development, spatial pattern formation has received a large amount of research over the past decade. Among the most studied, the *reaction-diffusion* (RD) systems are generating patterns that have been shown to represent well morphogenesis. A theory of morphogenesis based on a RD model was initially proposed by Turing [1]. Gierer and Meinhardt [2] were the first to explore pattern formation in biological systems using the RD equation. Subsequently, several equations of RD type have been studied to understand patterning in developmental biology. Some were derived from phenomenological models (Gierer–Meinhardt) while other modeled simple reaction schemes (Schnackenberg trimolecular autocatalytic reactions model [3], Gray–Scott model [4], Brusselator model [5], chlorite-iodide-malonic acid, CIMA model [6]). Recent work on RD systems demonstrates that it can be used to understand biological patterns formation [7], while [8] and [9] reviewed RD systems can be used to investigate spatial patterning in developmental systems.

The RD model for biological pattern formation is defined as follows [10]. Let Ω be a bounded domain in $\mathbb{R}^d$, $d = 1, 2, 3$, $\Gamma = \partial\Omega$ denote the boundary of Ω, and $J = (0, T]$ for some $T > 0$:

$$\begin{aligned} \frac{\partial u}{\partial t} - \mathscr{D}\Delta u &= f(u), \quad && \Omega \times J, \\ \frac{\partial u}{\partial \nu} &= 0, && \Gamma \times J, \\ u &= u^0, && \Omega \times \{t = 0\}, \end{aligned} \tag{1}$$

where $u = [u_1, u_2]^T$, $\mathscr{D} = \mathrm{diag}[D_1, D_2]^T$ is the diffusion tensor, Δ denotes the Laplace operator, $\partial/\partial\nu$ is the outward normal derivative on the boundary Γ, and $f(u)$ is the reaction kinetics of the system given as

$$f(u) = \begin{bmatrix} f_1(u_1, u_2) \\ f_2(u_1, u_2) \end{bmatrix}. \tag{2}$$

After Turing proposed a theory of morphogenesis in chemical cells in 1952 [1], much attention has been devoted to the numerical solution of RD problems of form (1); see [11–14] and references therein. Most of the numerical methods studied employed finite difference or finite element approximations for the spatial discretization, while some researchers use finite volume and collocation methods. Once the nonlinear reaction terms are treated (linearized or extrapolated), the *Crank–Nicolson* (CN) method can be applied as a second-order time-stepping procedure. Time-stepping procedures are required at each time step to solve a system of linear algebraic equations, which, although sparse, is compute intensive for multidimensional problems. In order to enhance efficiency of time-stepping procedures, one can adopt the *alternating direction implicit* (ADI) method as in [11–13, 15]. In particular, Fernandes et al. [12] introduce an ADI extrapolated CN orthogonal spline collocation method for RD problems.

ADI was invented as a perturbation of the CN method by Douglas, Peaceman, and Rachford in 1955 [16–18] and has been employed effectively for the calculation of numerical solution of various time-dependent multidimensional problems, either parabolic or hyperbolic [19, 20]. ADI reduces a multidimensional problem to multiple easy-to-solve one-dimensional problems, for an extra cost of a *splitting error* in $\mathrm{O}(\Delta t^2)$, where Δt is the time step size. However, the splitting error can be much larger than the sum of spatial and temporal discretization errors, unless the time step size is sufficiently small [21].

On the contrary, the CN method applied for nonsmooth data may introduce spurious oscillations to the numerical solution unless the time step size is sufficiently small to satisfy the maximum principle 12, which has been recognized in the original paper as well [22]. For this reason, whenever a larger time step or a higher spatial resolution is desirable/necessary, the (less accurate, first order) implicit method which is immune to oscillations has been used at least for several initial time steps with nonsmooth initial data [23]. The CN method and its perturbations (such as ADI) must be applied with care when the solution involves fast transitions or sharp edges; in particular, the time step size should be set very small, e.g., $\Delta t = \mathrm{O}(\Delta x^2)$, where Δx is the spatial grid size. In order to overcome the oscillation problem of the CN method applied for linear parabolic problems of nonsmooth data, the authors recently suggested a *variable-θmethod* in which the time-stepping parameter of the conventional θ-method, $\theta \in [0, 1]$, was determined based on local oscillatory characteristics of the solution and the data [24].

In this article, we apply the variable-θ method for the numerical solution of two-component nonlinear RD equations, as given in (1). The variable-θ method is a perturbation of the CN method which evolves the solution implicitly at points where the solution shows a certain portent of oscillations and maintains as a similar accuracy as the CN method with smooth data. The proposed method would be an adequate choice of time-stepping procedure for the numerical solution of RD *partial differential equations* (PDEs) when a larger time step or a higher spatial mesh resolution is desirable. We have performed a sensitivity analysis for the numerical solution of biological pattern forming models to the spatial and temporal grid sizes. It has been observed from various examples that accuracy of the numerical solution is much more sensitive to the spatial mesh resolution than the temporal one. When the spatial mesh resolution is set high for a higher accuracy, the method allows to keep the temporal resolution moderate or low. The suggested variable-θ method can result in a smooth/stable numerical solution by suppressing possible oscillations, unlike the CN method.

The article is organized as follows. Section 2 includes a brief review on the CN method and its spurious oscillatory behaviors, as preliminaries. In Section 3, a variable-θ method is presented for the numerical solution of two-component nonlinear RD equations. We adopt the *successive over-relaxation* (SOR) method to solve the resulting linear systems at each time level. Section 4 considers a heuristic technique for the choice of the optimal relaxation parameter for SOR. Section 5 gives numerical examples that show effectiveness of the variable-θ method applied to RD problems in 1D and 2D spaces. In Section 6, we summarize our experiments and present conclusions.

2. Preliminaries

In this section, we present a brief review of time-stepping procedures, for the numerical solution of *linear* parabolic equations of the form

$$\begin{aligned} \frac{\partial u}{\partial t} - \Delta u &= f, \quad \Omega \times J, \\ \frac{\partial u}{\partial \nu} &= 0, \quad \Gamma \times J, \\ u &= u^0, \quad \Omega \times \{t = 0\}, \end{aligned} \tag{3}$$

where $D > 0$ is a diffusion coefficient and f is a reaction/source term. We also consider difficulties arising when dealing with nonsmooth data (initial values, boundary conditions, and/or the source term).

2.1. The θ-Method: Difference Equation. Let Ω be a rectangular domain in $\mathbb{R}^2$: $\Omega = (a_x, b_x) \times (a_y, b_y)$. By partitioning $\Omega \times J$, we obtain the space-time grid points:

$$\left(x_{ij}, t^n\right) \coloneqq \left(x_i, y_j, t^n\right), \quad i = 0, 1, \ldots, n_x, j = 0, 1, \ldots, n_y, n = 0, 1, \ldots, n_t, \tag{4}$$

where n_x, n_y, and n_t are prescribed positive integers and

$$\begin{aligned} x_i &= a_x + i \cdot \Delta x, \\ y_j &= a_y + j \cdot \Delta y, \\ t^n &= n \cdot \Delta t, \\ \Delta x &= \frac{b_x - a_x}{n_x}, \\ \Delta y &= \frac{b_y - a_y}{n_y}, \\ t &= \frac{T}{n_t}. \end{aligned} \tag{5}$$

Define the discrete domain, the set of the spatial grid points, by

$$\Omega_d = \left\{\left(x_i, y_j\right) : 0 \le i \le n_x, 0 \le j \le n_y\right\}, \tag{6}$$

and denote the set of boundary grid points by $\Gamma_d = \Omega_d \cap \Gamma$ and the set of interior grid points by $\Omega_d^0 = \Omega_d \backslash \Gamma_d$.

Let $g_{ij}^n = g(x_{ij}, t^n)$ for all functions g defined in (x, t). Then, the second-order 5-point *finite difference* (FD) approximation $\mathscr{A}$ of $-\Delta$ at $t = t^n$ reads

$$\mathscr{A}u_{ij}^n = A_1 u_{ij}^n + A_2 u_{ij}^n, A_1 u_{ij}^n \coloneqq -\delta_{xx} u_{ij}^n, A_2 u_{ij}^n \coloneqq -\delta_{yy} u_{ij}^n, \tag{7}$$

where the FD operators are defined as

$$\begin{aligned} \delta_{xx} u_{ij}^n &= \frac{u_{i-1,j}^n - 2u_{ij}^n + u_{i+1,j}^n}{\Delta x^2}, \\ \delta_{yy} u_{ij}^n &= \frac{u_{i,j-1}^n - 2u_{ij}^n + u_{i,j+1}^n}{\Delta y^2}. \end{aligned} \tag{8}$$

For the temporal derivative $\partial u/\partial t$, a convenient FD approximation can give

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t}. \tag{9}$$

Expressing the spatial derivative by a weighted average $\theta \in [0, 1]$ of previous and current time values, we can formulate the θ-method for (3) as

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} + \mathscr{A}\left[\theta u_{ij}^{n+1} + (1 - \theta) u_{ij}^n\right] = f_{ij}^{n+\theta}, \tag{10}$$

where $f_{ij}^{n+\theta}$ is either $f(x_{ij}, t^{n+\theta})$ or $\theta f_{ij}^{n+1} + (1 - \theta) f_{ij}^n$. A simple algebraic rearrangement of (10) in a vector form becomes

$$(I + \theta \Delta t \mathscr{A}) u^{n+1} = [I - (1 - \theta)\Delta t \mathscr{A}] u^n + \Delta t f^{n+\theta}, \tag{11}$$

where $u^n = [u_{ij}^n]_{0 \le i \le n_x, 0 \le j \le n_y}$ and $f^{n+\theta} = [f_{ij}^{n+\theta}]_{0 \le i \le n_x, 0 \le j \le n_y}$, considered as column vectors. Popular choices of $\theta \in [0, 1]$ are 0, 1, and 1/2, which are, respectively, the explicit method (the forward Euler method), the implicit method (the backward Euler method), and the semi-implicit method (the Crank–Nicolson method).

(i) Forward Euler method: when $\theta = 0$, algorithm (11) is stable when it satisfies

$$\begin{aligned} \mu_x + \mu_y &\le \frac{1}{2}, \\ \mu_x &= \frac{\Delta t}{\Delta x^2}, \\ \mu_y &= \frac{\Delta t}{\Delta y^2}. \end{aligned} \tag{12}$$

Although the explicit method is efficient for each time step, its stability condition in (12) enforces, choosing a small time step size Δt; it may become less efficient compared with other implicit methods. It is elementary in numerical analysis that when $\theta \ge 1/2$, the θ-method (10) is unconditionally stable.

(ii) Crank–Nicolson method: when $\theta = 1/2$, (11) can be rewritten as

$$\left(I + \frac{\Delta t}{2}\mathscr{A}\right) u^{n+1} = \left(I - \frac{\Delta t}{2}\mathscr{A}\right) u^n + \Delta t f^{n+1/2}. \tag{13}$$

The CN method has been the most popular time-stepping procedure for the numerical solution of parabolic problems because it is stable and of second-order accuracy in both spatial and temporal directions. However, the CN method applied for nonsmooth data may introduce spurious oscillations to the numerical solution unless the algorithm parameters satisfy the maximum principle [22, 23]. As analyzed by the authors [24], the undesired oscillations are due to instability involved in the explicit half step of the CN method, the first term in the right side of (13). The *variable θ-method* proposed in [24] suppresses spurious oscillations, by evolving the solution implicitly ($\theta_{ij} = 1$) at points x_{ij} where the solution shows a certain portent of oscillations or reduced smoothness, and maintains as a similar accuracy as the CN method with smooth data.

(iii) Backward Euler method: when $\theta = 1$, algorithm (11) reads

$$(I + \Delta t \mathscr{A}) u^{n+1} = u^n + \Delta t f^{n+1}. \tag{14}$$

Although the implicit method shows a first-order accuracy in the temporal direction, it never introduces spurious oscillations to its numerical solutions.

2.2. Numerical Oscillations of the CN Method. Although the CN method is unconditionally stable and of second-order

accuracy in both spatial and temporal directions, it may introduce spurious oscillations into the numerical solution for nonsmooth data. For simplicity, consider a homogeneous diffusion equation with discontinuous initial values defined in the one-dimensional (1D) space:

$$\frac{\partial u}{\partial t} - u_{xx} = 0,$$

$$(x,t) \in (0,1) \times [0,T],$$

$$u(0,t) = u(1,t) = 0, \quad t \in [0,T],$$

$$u(x,0) = \begin{cases} 0 & \text{if } 0 < x < \dfrac{1}{4}, \\ 1 & \text{if } \dfrac{1}{4} \le x < \dfrac{3}{4}, \\ 0 & \text{if } \dfrac{3}{4} \le x < 1, \end{cases} \tag{15}$$

of which the exact solution is given by

$$u(x,t) = \sum_{k=1}^{\infty} \frac{4}{k\pi} \sin\left(\frac{k\pi}{2}\right) \sin\left(\frac{k\pi}{4}\right) \sin(k\pi x) e^{-k^2\pi^2 t}, \tag{16}$$

for $(x,t) \in (0,1) \times [0,T]$.

Figure 1 depicts the exact and numerical solutions evolved by the CN and the variable-θ method [24], while Figure 2 compares the numerical solutions at $T = 1.0$ obtained by the three numerical methods. The numerical solutions are obtained with $\Delta t = 0.01$ and $\Delta x = \Delta y = 0.025$. As one can see, spurious oscillations occur along with the step discontinuities in the numerical solution of the CN method, as shown in Figure 1(b). On the contrary, as in Figures 1(c) and 2, the variable-θ method results in an accurate numerical solution without any oscillations. One should notice from Figures 1(a) and 2 that the implicit method is immune to spurious oscillations, but its error is considerably large due to a first-order discretization error in the temporal direction. Although the CN method reveals spurious oscillations, it is quite accurate far from discontinuities. The variable-θ method results in numerical solutions which are *smooth* as for the implicit method and *accurate* as for the CN method associated with smooth data.

The variable-θ method is a hybrid time-stepping procedure that is based on the CN method ($\theta = 1/2$) and alternately using the implicit method ($\theta = 1$) at points, where the numerical solution shows a certain portent of oscillations or reduced smoothness (the wobble set).

3. A Variable-θ Method for Two-Component Nonlinear Problems

This section introduces an effective time-stepping procedure for the numerical solution of two-component RD problems (1).

3.1. Linearization through Extrapolation. Once the spatial derivatives are approximated by second-order finite difference schemes, as in Section 2.1, the semidiscrete problem for (1) is formulated as

$$\frac{\partial u}{\partial t} + \mathscr{D}\mathrm{A}u = f(u), \quad t \in (0,T], \qquad u = u^0, \quad t = 0. \tag{17}$$

Let numerical solutions be obtained up to the nth time level, $n > 0$. For the numerical solution in the $(n+1)$th level, we first extrapolate numerical solutions in the two previous levels to approximate the solution at the midpoint $t^{n+1/2}$:

$$\widetilde{u}^{n+1/2} := \frac{3}{2}u^n - \frac{1}{2}u^{n-1}. \tag{18}$$

See [12], for details of second-order extrapolation schemes for $n \ge 0$. Then, the θ-method for the two-component RD problem reads:

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} + \mathscr{D}\mathscr{A}\left[\theta u_{ij}^{n+1} + (I-\theta)u_{ij}^n\right] = f\left(\widetilde{u}^{n+1/2}\right)_{ij}, \tag{19}$$

where $u = [u_1, u_2]^T$, $\mathscr{D} = \mathrm{diag}[D_1, D_2]^T$, and $\theta = \mathrm{diag}[\theta_1, \theta_2]^T$.

The linearized problem (19) can be resolved by solving for two separate components $u^{n+1} = [u_1^{n+1}, u_2^{n+1}]^T$. Each component in (19) can be formulated as follows:

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} + D\mathscr{A}\left[\theta u_{ij}^{n+1} + (I-\theta)u_{ij}^n\right] = f_{ij}^{n+1/2}, \tag{20}$$

where u, D, and θ denote, respectively, u_k, D_k, and θ_k, for $k = 1$ or 2, and $f^{n+1/2}$ is a known source term. The θ-method (20) can be rewritten in a vector form as

$$(I + \theta\Delta t\, D\mathscr{A})u^{n+1} = [I - (1-\theta)\Delta t\, D\mathscr{A}]u^n + \Delta t\, f^{n+1/2}. \tag{21}$$

We present here the main steps of variable-θ method for a nonoscillatory solution of (20); a complete study of the method for diffusion equation was published in [24].

3.2. The Variable-θ Method. The method begins with defining the *wobble set*, the set of wobble points, as a collection of the grid points where the solution has high fluctuations so that the implicit method ($\theta = 1$) should be applied for the numerical solution not to develop oscillations.

One can easily verify that numerical oscillations of the CN method occur when its explicit half step produces spurious oscillations. Such nonphysical oscillations may happen particularly when the time step size Δt is larger than the stability limit of the explicit scheme. Thus, the wobble set may be formed to include points where the explicit half step of the CN method introduces undesired local extrema. It follows from (21) that the explicit half step of the CN method ($\theta = 1/2$) reads

$$u^{n,*} := \left(I - \frac{\Delta t}{2} D\mathscr{A}\right)u^n. \tag{22}$$

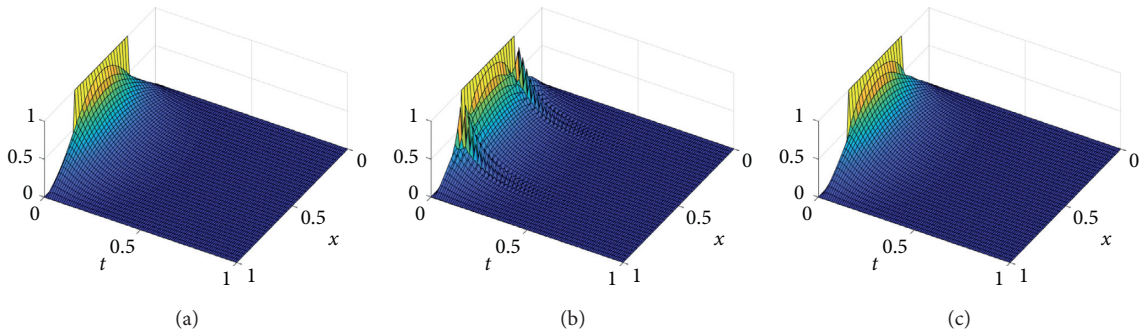


FIGURE 1: Propagation of the exact and numerical solutions for (15): (a) the exact solution and the numerical solution by (b) the CN method and (c) the variable-θ method [24], for $0 \le t \le T = 1.0$, when $\Delta t = 0.01$ and $\Delta x = 0.025$.

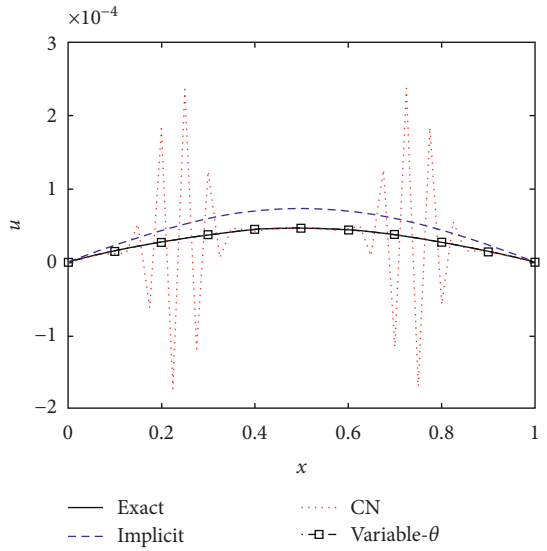


FIGURE 2: The numerical solutions of (15) at $T = 1.0$, compared with the exact solution.

Let x_{ij} be an interior grid point and consider the four partial directions (made with eight vicinal points of x_{ij}): four directions having 0°, 45°, 90°, and 135° from the positive x-direction. When spurious oscillations are observed in at least one direction, we select the point x_{ij} as a wobble point.

Define an index function for local extrema (idxt) as

$$\text{idxt}(a,b,c) = \begin{cases} 0, & \text{if } \min(a,c) < b < \max(a,c), \\ 1, & \text{if } b = \max(a,c), \\ -1, & \text{if } b = \min(a,c), \\ 2, & \text{if } \max(a,c) < b, \\ -2, & \text{if } b < \min(a,c). \end{cases} \tag{23}$$

Let P, Q, and R be point indices and define

$$\text{iswb}(P,Q,R,n) = \begin{cases} 1, & \text{if } \text{idxt}\left(u_P^{n,*}, u_Q^{n,*}, u_R^{n,*}\right) \neq 0 \quad \text{and} \\ & \left|\text{idxt}\left(u_P^{n,*}, u_Q^{n,*}, u_R^{n,*}\right) + \text{idxt}\left(u_P^{n}, u_Q^{n}, u_R^{n}\right)\right| < 4, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

Then, the wobble set (for the computation of $u^{n+1} = \{u_{ij}^{n+1}\}$) is defined as

$$\begin{aligned} W^n = \{x_{ij} \in \Omega_d^0 \mid & \text{iswb}[(i-1,j),(i,j),(i+1,j),n] = 1 \\ & \text{or iswb}[(i-1,j-1),(i,j),(i+1,j+1),n] = 1 \\ & \text{or iswb}[(i,j-1),(i,j),(i,j+1),n] = 1 \\ & \text{or iswb}[(i+1,j-1),(i,j),(i-1,j+1),n] = 1\}. \end{aligned} \tag{25}$$

Remark 1. The function `iswb` selects candidates for the wobble set from local extrema satisfying $\text{idxt}(u_P^{n,*}, u_Q^{n,*}, u_R^{n,*}) \neq 0$; however, the condition

$$\left|\text{idxt}\left(u_P^{n,*}, u_Q^{n,*}, u_R^{n,*}\right) + \text{idxt}\left(u_P^{n}, u_Q^{n}, u_R^{n}\right)\right| < 4, \tag{26}$$

excludes cases where a strict extremum in u^n becomes a strict extremum in the same sense in $u^{n,*}$. Thus, the wobble set (25) is the set of interior grid points x_{ij} where $u_{ij}^{n,*}$ becomes a local extremum while u_{ij}^n is either a nonextreme or an extreme in the opposite sense, for at least one of four partial directions.

Having the wobble set, the parameter θ for the computation of u^{n+1} can be assigned pointwise:

$$\theta_{ij}^{n+1} := \theta\left(x_{ij}, t^{n+1}\right) = \begin{cases} 1, & \text{if } x_{ij} \in W^n, \\ 1/2, & \text{otherwise.} \end{cases} \tag{27}$$

Thus, the *variable-θmethod* for (20) can be formulated as

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} + D\mathscr{A}\left[\theta_{ij}^{n+1} u_{ij}^{n+1} + \left(1 - \theta_{ij}^{n+1}\right) u_{ij}^n\right] = f_{ij}^{n+1/2}, \tag{28}$$

or, in a vector form after grouping variables:

$$\left(I + \theta^{n+1} \Delta t\, D\mathscr{A}\right) u^{n+1} = \left[I - \left(1 - \theta^{n+1}\right) \Delta t\, D\mathscr{A}\right] u^n + \Delta t\, f^{n+1/2}. \tag{29}$$

The variable-θ method is analyzed for its numerical stability and accuracy and verified for various examples [24]. It results in nonoscillatory numerical solutions of which the accuracy is *almost* second-order in time.

Remark 2. The ADI procedure was also applied to (19) for which the initial values show sharp transitions. It has been

observed that ADI may introduce undesirable/discontinuous features to its solution unless the time step size is sufficiently small, i.e., $\Delta t = \mathrm{O}(\Delta x^2, \Delta y^2)$. The main problem with ADI is that the diffusion becomes anisotropic, i.e., faster in the coordinate directions. When ADI is applied to the variable-θ formulation of (19), the anisotropic features are reduced significantly. However, it requires to set Δt small enough for a reliable numerical solution.

The algebraic system in (29) will be solved by applying the SOR method, with its initial value at the time level t^{n+1} being set as

$$u^{n+1,0} = 2u^n - u^{n-1}. \tag{30}$$

In particular, SOR converges quite fast for an appropriate choice of the relaxation parameter ω.

In the following, we will consider how to tune the optimal relaxation parameter $\widehat{\omega}$ for SOR.

4. The Optimal SOR Parameter $\widehat{\omega}$

In this section, we will try to find a relaxation parameter which is *heuristically optimal*. Let us begin with the 2D algebraic system of (21) with $\theta = 1/2$:

$$\mathscr{L}u = r \in \mathbb{R}^{m\times m}, \tag{31}$$

where $\mathscr{L} = I + (\Delta t/2)D\mathscr{A}$, $r = (I - (\Delta t/2)D\mathscr{A})u + \Delta t f^{n+1/2}$, and $m > 0$ is the dimension of the algebraic system. It is known that the optimal relaxation parameter for the SOR method can be determined as ([25], Section 4.3)

$$\widehat{\omega} = \frac{2}{1 + \sqrt{1 - \rho\left(\mathrm{T}_J\right)^2}}, \tag{32}$$

where $\rho(\mathrm{T}_J)$ is the spectral radius of the Jacobi iteration matrix T_J.

For simplicity, assume that the problem is defined on the unit square with a Dirichlet boundary condition. We further assume that the domain is partitioned into $N \times N$ grids so that $h = \Delta x = \Delta y = 1/N$. Then, the eigenvalues of the second-order 5-point FD coefficient matrix A read ([25], Section 6.5)

$$\lambda_{k,\ell}(\mathscr{A}) = \frac{1}{h^2}\left[4 - 2\cos\left(\frac{k\pi}{N}\right) - 2\cos\left(\frac{\ell\pi}{N}\right)\right], \quad 1 \le k, \ell \le N-1, \tag{33}$$

and therefore the eigenvalues of $\mathscr{L}$ can be formulated as

$$\lambda_{k,\ell}(\mathscr{L}) = 1 + \frac{\Delta t}{2}D\lambda_{k,\ell}(\mathscr{A}) = 1 + \frac{D\Delta t}{h^2}\left[2 - \cos\left(\frac{k\pi}{N}\right) - \cos\left(\frac{\ell\pi}{N}\right)\right], \tag{34}$$

for $1 \le k, \ell \le N-1$. Note that the diagonal element of $\mathscr{L}$ is

$$p \coloneqq 1 + \frac{\Delta t}{2}D\frac{4}{h^2} = 1 + 2\frac{D\Delta t}{h^2}. \tag{35}$$

So, the eigenvalues of the Jacobi iteration matrix T_J are given as

$$\lambda_{k,\ell}\left(\mathrm{T}_J\right) = \frac{p - \lambda_{k,\ell}(\mathscr{L})}{p} = \frac{D\Delta t}{h^2}\,\frac{[\cos(k\pi/N) + \cos(\ell\pi/N)]}{(1 + 2(D\Delta t/h^2))}. \tag{36}$$

In order to find the maximum of $|\lambda_{k,\ell}(\mathrm{T}_J)|$, we first obtain

$$\max_k\left|\cos\left(\frac{k\pi}{N}\right)\right| = \cos\left(\frac{\pi}{N}\right) = 1 - c_1h^2, \tag{37}$$

for some $c_1 > 0$. Here, we have used $h = 1/N$ and the approximation $\cos(x) \approx 1 - x^2/2$. Now the spectral radius of T_J reads

$$\rho\left(\mathrm{T}_J\right) = \max_{k,\ell}\left|\lambda_{k,\ell}\left(\mathrm{T}_J\right)\right| = \left(1 - c_1h^2\right)\left(1 + \frac{1}{2}\frac{h^2}{D\Delta t}\right)^{-1}. \tag{38}$$

Assuming that $c_1h^2 < h^2/(2D\Delta t) < 1$, we finally obtain

$$\rho\left(\mathrm{T}_J\right) = 1 - c_2\frac{h^2}{D\Delta t}, \tag{39}$$

for some $c_2 > 0$.

It follows from (32) and (39) that the optimal SOR parameter $\widehat{\omega}_{\Delta t,h}$, corresponding to the spatial grid size h and the time step size Δt, can be written as

$$\widehat{\omega}_{\Delta t,h} = \frac{2}{1 + c_0(h/\sqrt{D\Delta t})}, \tag{40}$$

for some $c_0 > 0$. The constant c_0 can be found *experimentally* from a selected set of $(\Delta t, h)$, as summarized in the following:

$$\begin{cases} \text{(a)Determine } \widehat{\omega}_{\Delta t_0,h_0} \text{ for prescribed grid sizes } (\Delta t_0, h_0), \text{ heuristically.} \\ \text{(b)Solve (40) for } c_0\text{:} \\ c_0 = \dfrac{\sqrt{D\Delta t_0}}{h_0}\left(\dfrac{2}{\widehat{\omega}_{\Delta t_0,h_0}} - 1\right). \end{cases} \tag{41}$$

Once c_0 is estimated as in (41), the parameter $\widehat{\omega}_{\Delta t,h}$ in (40) is *near-optimal* for various choices of $(\Delta t, h)$.

5. Numerical Experiments

In this section, we present numerical experiments which show effectiveness of the variable-θ method. The algorithms are implemented in MATLAB and carried out on a desktop computer of Intel Xeon CPU E5-1620 3.60 GHz processor.

To solve the algebraic system at each time level, the SOR method is employed with the near-optimal parameter $\widehat{\omega}$ calculated as in (40), with c_0 being estimated with a small grid problem. The SOR iteration is stopped when the maximum difference of consecutive iterates becomes smaller than a tolerance $\varepsilon = 10^{-6}$:

$$\left\|u^{n,k} - u^{n,k-1}\right\|_\infty < \varepsilon. \tag{42}$$

The L^∞-error $E_\infty[t^n]$, measured at $t = t^n$, is defined as follows:

$$E_\infty[t^n] := \left\| u^n - \widehat{u}(t^n) \right\|_\infty, \tag{43}$$

where $\widehat{u}$ is the exact solution.

5.1. One-Component RD System. To investigate accuracy of the variable-θ method, we consider the diffusion problem (15) studied earlier in Section 2.2. For a comparison purpose, we have implemented not only the θ-methods and the variable-θ method but also the *implicit predictor-corrector* $(0, 2)$-*Padé* (IPC-[0,2]) method [26] and *Lawson and Morris* (LM) local extrapolation method [23].

Table 1 presents the L^∞-error $E_\infty[T]$ at $T = 1.0$ when the five methods are applied for the numerical solution of (15), with various Δt and Δx. As one can see from the table, the CN method resolves its numerical solution poorly (due to oscillations), except for the case the method satisfies the maximum principle. On the contrary, the variable-θ method results in a second-order accuracy, with its errors being smallest among all the methods for most cases. The new method is a hybrid time-stepping procedure which assembles merits from the CN method (high-accuracy) and the implicit method (smoothness).

Now, consider a nonlinear RD problem of the form

$$\begin{aligned} &\frac{\partial u}{\partial t} - u_{xx} = u(1-u), \\ &(x,t) \in (0,1) \times [0,T], \end{aligned} \tag{44}$$

with the boundary and initial values, as given in (15).

Figure 3 presents the numerical solutions evolved by the implicit method, the CN method, and the variable-θ method, when $\Delta t = 0.01$ and $\Delta x = 0.025$ (the mesh is the same as the one selected in Figure 1). Similar to the linear problem in Figure 1, spurious oscillations are introduced into the numerical solution of the nonlinear problem by the CN method only. It should be noticed that spurious oscillations of the CN method are damped out much faster for the nonlinear problem than the linear problem, which is due to the reaction kinetic term $f(u) = u(1-u)$. For the nonlinear problem, It seems that the oscillations at early time steps do not affect the solution at later steps much. This observation explains a partial reason that the second-order CN method has been popular for the numerical solution of PDEs in mathematical biology. However, for other applications, such spurious oscillations at early moments may alter the numerical solution significantly so as for the CN method to be unstable; see Figure 4 below. It is important to develop an effective algorithm which can suppress spurious oscillations for convenient choices of algorithm parameters; the variable-θ method is effective and stable.

5.2. Two-Component Nonlinear RD Systems. Two-component RD systems enable to explain a much wider range of phenomena than their one-component counterparts. Many two-component models have been developed and numerically verified for dynamical patterning behaviors in biology and chemistry. In this section, we consider two-component models interested in the literature of biology and chemistry, to verify effectiveness of the variable-θ method.

5.2.1. Gray–Scott Model in 1D. We apply the numerical methods for the numerical solution of the Gray–Scott model [4, 27] defined as (1) associated with the following reaction kinetics:

$$f(u) = \left[F(1-u_1) - u_1 u_2^2, u_1 u_2^2 - (F+k) u_2 \right]^T, \tag{45}$$

for any constants F and k. Let $\Omega = (0, 1)$. We assign two sets of model constants and initial and boundary conditions as follows [28]:

$$\mathscr{D} = \left[10^{-4}, 10^{-6} \right]^T, F = 0.035, k = 0.049, \tag{46a}$$

$$u_1(x,0) = 1 - \frac{1}{2} \sin^{100}(\pi x), u_2(x,0) = \frac{1}{4} \sin^{100}(\pi x), x \in (0,1), \tag{46b}$$

$$u_1(0,t) = u_1(1,t) = 1, u_2(0,t) = u_2(1,t) = 0, \quad t \in [0,T], \tag{46c}$$

where (46b) is a *mid-pulse* initial condition, and

$$\mathscr{D} = \left[10^{-4}, 5 \cdot 10^{-5} \right]^T, F = 0.025, k = 0.0544, \tag{47a}$$

$$u_1(x,0) = 1 - \frac{1}{2} \cos^{100}\left(\frac{\pi x}{2}\right), u_2(x,0) = \frac{1}{4} \cos^{100}\left(\frac{\pi x}{2}\right), \quad x \in (0,1), \tag{47b}$$

$$\frac{\partial u_1}{\partial x}(0,t) = \frac{\partial u_1}{\partial x}(1,t) = 0, \frac{\partial u_2}{\partial x}(0,t) = \frac{\partial u_2}{\partial x}(1,t) = 0, \quad t \in [0,T], \tag{47c}$$

where (47b) is a *left-pulse* initial condition.

In Figure 5, we present the propagation of the numerical solution of u_2 by the variable-θ method associated with (46a)–(46c) at the grid sizes $\Delta t = 0.01$ and $\Delta x = 0.004$ over $0 \le t \le T = 2000$. The initial midpulse splits in early moments to travel in both directions (self-replication of the pulse). As each of the pulses travels, it becomes thicker (bigger) up to a certain width and begins to replicate itself recursively.

Table 1: L^∞-error $E_\infty[T]$ for (15) at $T = 1.0$.

$(\Delta t, \Delta x)$	Implicit	CN	IPC-[0, 2]	LM	Variable-θ
(0.1, 0.05)	$9.0 \cdot 10^{-4}$	$2.4 \cdot 10^{-1}$	$6.0 \cdot 10^{-5}$	$8.4 \cdot 10^{-5}$	$8.4 \cdot 10^{-5}$
(0.01, 0.05)	$2.8 \cdot 10^{-5}$	$4.8 \cdot 10^{-7}$	$1.6 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$8.5 \cdot 10^{-7}$
(0.1, 0.025)	$8.9 \cdot 10^{-4}$	$3.7 \cdot 10^{-1}$	$5.9 \cdot 10^{-5}$	$8.3 \cdot 10^{-5}$	$5.1 \cdot 10^{-5}$
(0.01, 0.025)	$2.7 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$	$9.1 \cdot 10^{-7}$	$2.5 \cdot 10^{-6}$	$1.5 \cdot 10^{-7}$
(0.1, 0.0125)	$8.9 \cdot 10^{-4}$	$4.3 \cdot 10^{-1}$	$5.8 \cdot 10^{-5}$	$8.3 \cdot 10^{-5}$	$2.7 \cdot 10^{-4}$
(0.01, 0.0125)	$2.7 \cdot 10^{-5}$	$3.9 \cdot 10^{-2}$	$7.5 \cdot 10^{-7}$	$2.4 \cdot 10^{-6}$	$1.2 \cdot 10^{-7}$

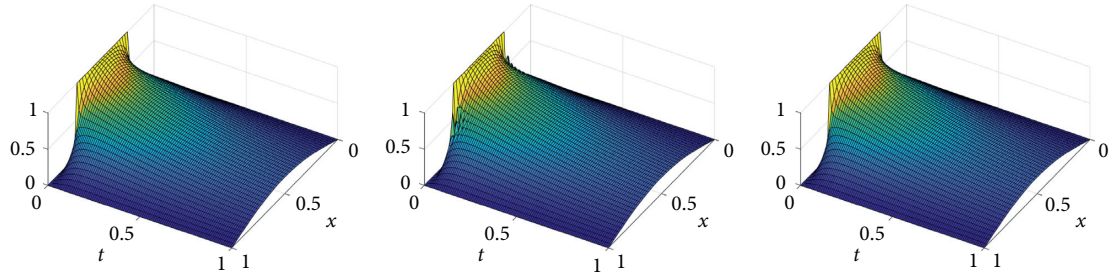


Figure 3: Propagation of the numerical solutions for (44): (a) the implicit method, (b) the CN method, and (c) the variable-θ method, for $0 \le t \le T = 1.0$, when $\Delta t = 0.01$ and $\Delta x = 0.025$.

Figure 6 depicts the propagation of the numerical solution of u_2 by the variable-θ method associated with (47a)–(47c) over $0 \le t \le T = 5000$, at the same resolution, as in Figure 5. One can clearly observe a traveling pulse which begins from the left edge point and reflects whenever it hits the boundary, due to the no-flux boundary condition (47c).

To investigate bifurcation in the RD system and our method numerical accuracy, we present numerical solutions of the wave-splitting problem (46a)–(46c) obtained with various spatial and temporal grid sizes, as shown in Figure 7. The image $I_{k\ell}$ represents the numerical solution obtained with the mesh resolution $(\Delta t, \Delta x) = (10^{-k}, 0.01/2^{\ell-1})$. For example, the image I_{23} is associated with the mesh resolution $(\Delta t, \Delta x) = (1/100, 1/400)$. One can easily point out from the images that the spatial resolution alters the numerical solution *dramatically* even with halved spatial grid sizes (compare the images horizontally), while the temporal resolution affects *little* the numerical solution even with one-order smaller temporal step sizes (compare them vertically).

The main reason for such a sensitivity to the spatial resolution is that the RD does not have as much time before growing to reach the margins of the mesh in low spatial resolutions (of large Δx's). When this happens, the RD pattern typically deteriorates and it does not travel in an appropriate speed nor reaches a condition to replicate itself on time, see ([29], Section 4.2) for similar observations. We summarize the experiments with the Gray–Scott model in 1D as follows:

(i) Accuracy of the numerical solution is much more sensitive to the spatial mesh resolution than the temporal one.

(ii) Thus, it is crucial to set a high spatial resolution (small Δx's) for a desirable accuracy.

(iii) As the spatial resolution becomes higher, the CN method may more likely produce spurious oscillations, while the variable-θ method results in stable solutions.

5.2.2. Gray–Scott Model in 2D. Note that the two-component Gray–Scott model is formulated as in (1) with the reaction kinetics $f(u)$ given in (45). We choose problem coefficients as follows [15]:

$$\begin{aligned} \Omega &= (-1, 1) \times (-1, 1), \\ \mathscr{D} &= [0.001, 0.001]^T, \\ F &= 1, \\ k &= 0. \end{aligned} \tag{48}$$

For the purpose of error analysis, we select a smooth solution $\widehat{u} = [\widehat{u}_1, \widehat{u}_2]$ defined as

$$\begin{aligned} \widehat{u}_1(x, y, t) &= \cos(2t)\cos(2\pi x)\cos(\pi y), \\ \widehat{u}_2(x, y, t) &= \cos(2t)\cos(\pi x)\cos(2\pi y), \end{aligned} \tag{49}$$

and replace the reaction kinetics $f(u)$ with $f_{\widehat{u}}(u)$:

$$f_{\widehat{u}}(u) \coloneqq f(u) + \frac{\partial \widehat{u}}{\partial t} - \mathscr{D}\Delta\widehat{u} - f(\widehat{u}). \tag{50}$$

Then, $\widehat{u} = [\widehat{u}_1, \widehat{u}_2]$ in (49) would be the exact solution of $\partial u/\partial t - \mathscr{D}\Delta u - f(u) = f_{\widehat{u}}(u)$ with the initial condition $u^0 = \widehat{u}(x, y, 0)$.

Table 2 summarizes the L^∞-error $E_\infty[T]$ with $T = 1.0$ and the elapsed time (CPU) for the implicit, CN, variable-θ methods for three different meshes refined by a factor of 2 in both spatial and temporal directions with $\Delta x = \Delta y$. Since the solution including the initial condition is smooth over the

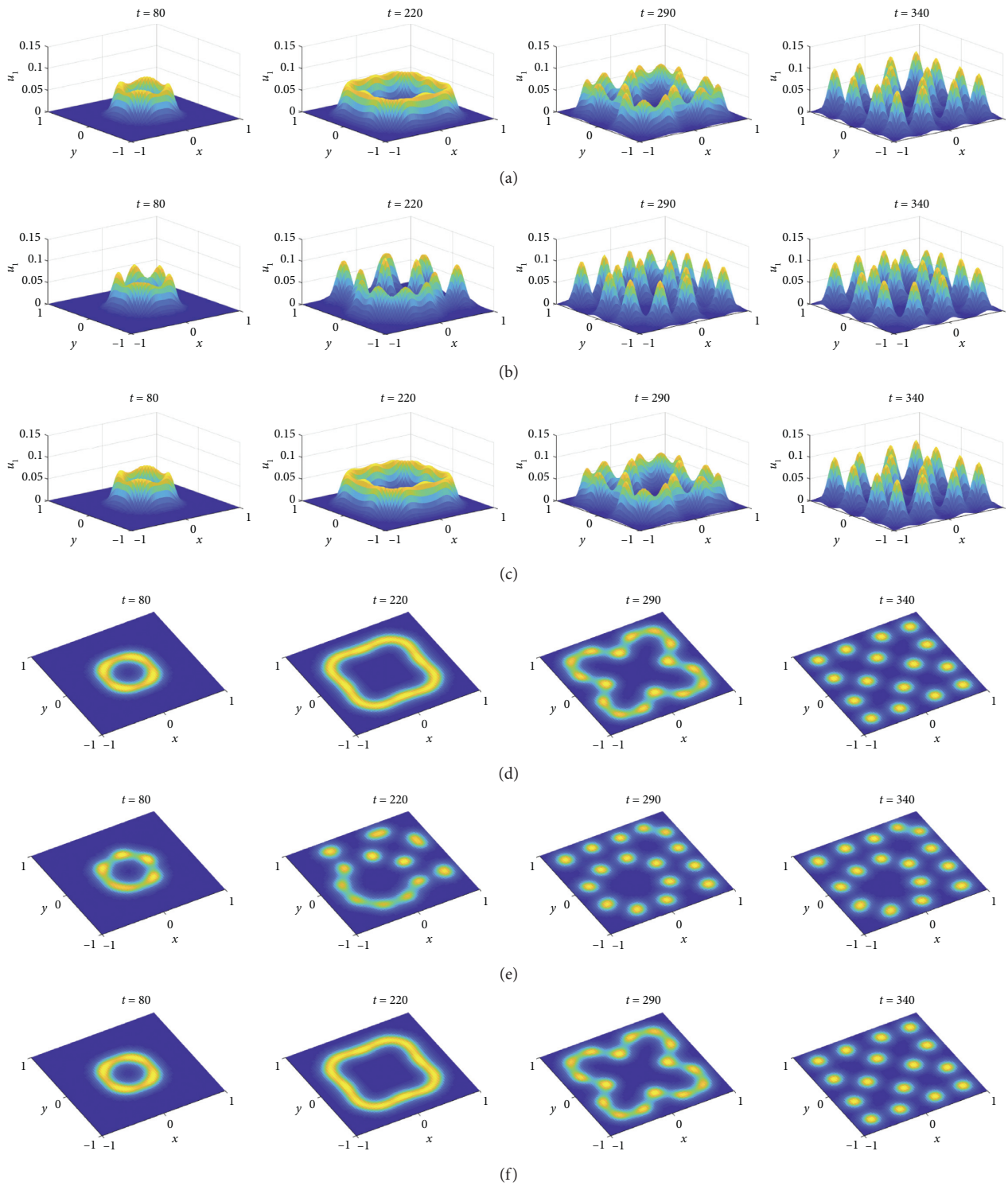
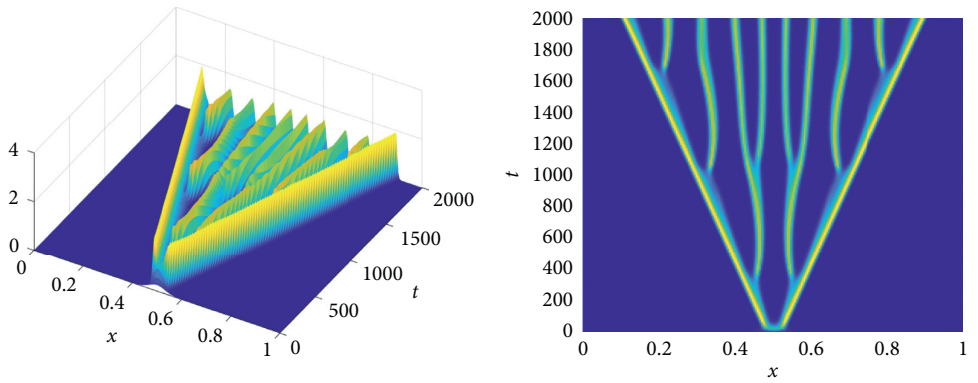


FIGURE 4: Numerical solutions for u_1 of the Gierer–Meinhardt model at the spatial resolution $\Delta x = \Delta y = 1/32$: (a) the variable-$\theta$ method with $\Delta t = 0.05$, (b) the CN method with $\Delta t = 0.05$, and (c) the CN method with $\Delta t = 0.005$; (d–f) are the aerial views of (a–c), respectively.

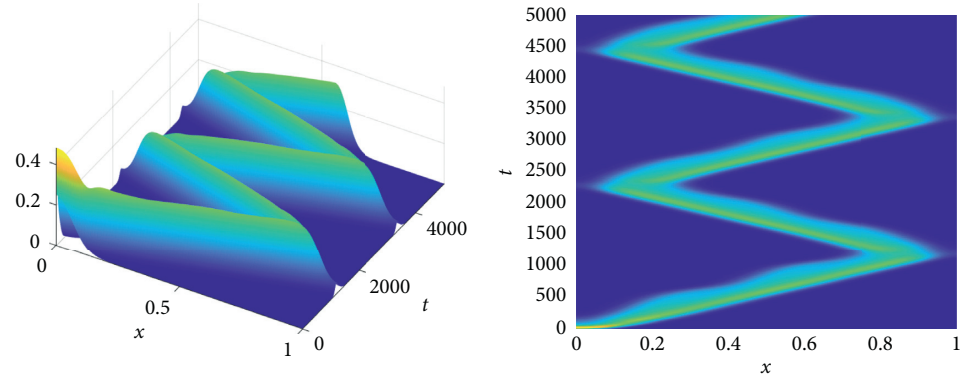
entire time interval $[0, T = 1]$, the CN method introduces no spurious oscillations into its numerical solution and proves a second-order accuracy for the two-component Gray–Scott model in the 2D space. One should notice that the variable-θ has also proved its second-order accuracy, the same as the CN method. On the contrary, the implicit method involves considerable errors due to its first-order convergence in temporal direction. Figure 8 shows the numerical solution by the variable-θ method, and its error at $T = 1.0$ for the Gray–Scott model, when $\Delta t = 0.025$ and $\Delta x = \Delta y = 0.05$.

For all the three methods, the algebraic system is solved by the SOR method with its optimal relaxation parameter



(a) (b)

Figure 5: Propagation of the numerical solution of u_2 for the 1D Gray–Scott model by the variable-θ method: (a) the wave-splitting (self-replication of the pulse) and (b) its aerial view over $0 \le t \le T = 2000$, when $\Delta t = 0.01$ and $\Delta x = 0.004$.



(a) (b)

Figure 6: Propagation of the numerical solution of u_2 for the 1D Gray–Scott model by the variable-θ method: (a) the pulse traveling and reflecting and (b) its aerial view over $0 \le t \le T = 5000$, when $\Delta t = 0.01$ and $\Delta x = 0.004$.

being calibrated from the lowest resolution, $(\Delta t, \Delta x = \Delta y) = (0.05, 0.1)$. That is, the constant c_0 in (41) is evaluated using the experimentally optimal $\widehat{\omega}_{\Delta t_0, h_0}$ with $(\Delta t_0, h_0) = (0.05, 0.1)$ and then (40) is utilized to estimate $\widehat{\omega}_{\Delta t, h}$ for other grid sizes $(\Delta t, h = \Delta x = \Delta y)$. With the *near-optimal* parameter and an effective initialization scheme in (30), for both the CN and variable-θ methods, the SOR method has converged in (59) iterations in average for solving the two algebraic systems (for u_1 and u_2) in a time level. SOR is comparable with ADI in efficiency when the parameter is set optimal and the initialization is carried out accurately; SOR has proven its efficiency for the numerical solution of elliptic obstacle problems [30]. For the Gray–Scott model in 2D, the variable-θ method becomes about a third most expensive computationally than the CN method, due to the wobble set processing.

5.2.3. Gierer–Meinhardt Model. The Gierer–Meinhardt model [2] is (1) defined in $\Omega = (-1, 1) \times (-1, 1) \subset \mathbb{R}^2$ with the following reaction kinetics and parameters:

$$\mathscr{D} = \left[\varepsilon^2, \frac{\kappa}{\mu}\right]^T,$$
$$f(u) = \left[\frac{u_1^2}{u_2} - u_1, \frac{1}{\mu}\left(\frac{u_1^2}{\varepsilon} - u_2\right)\right]^T, \tag{51}$$

for which various numerical methods have been developed [12, 31, 32]. We cast the experiment employing coefficients and the initial condition used in [32]:

The initial values are depicted in Figure 9. In this section, we restrict our attention to the dynamics of u_1 of the model.

$$\varepsilon = 0.04, \mu = 0.1, \kappa = 0.0128,$$
$$u_1(x, y, 0) = \frac{1}{2}\left[1 + 0.001 \sum_{k=1}^{20} \cos\left(\frac{k\pi y}{2}\right)\right] \operatorname{sech}^2\left(\frac{\sqrt{x^2 + y^2}}{2\varepsilon}\right),$$
$$u_2(x, y, 0) = \frac{\cosh\left(1 - \sqrt{x^2 + y^2}\right)}{3 \cos h(1)}. \tag{52}$$

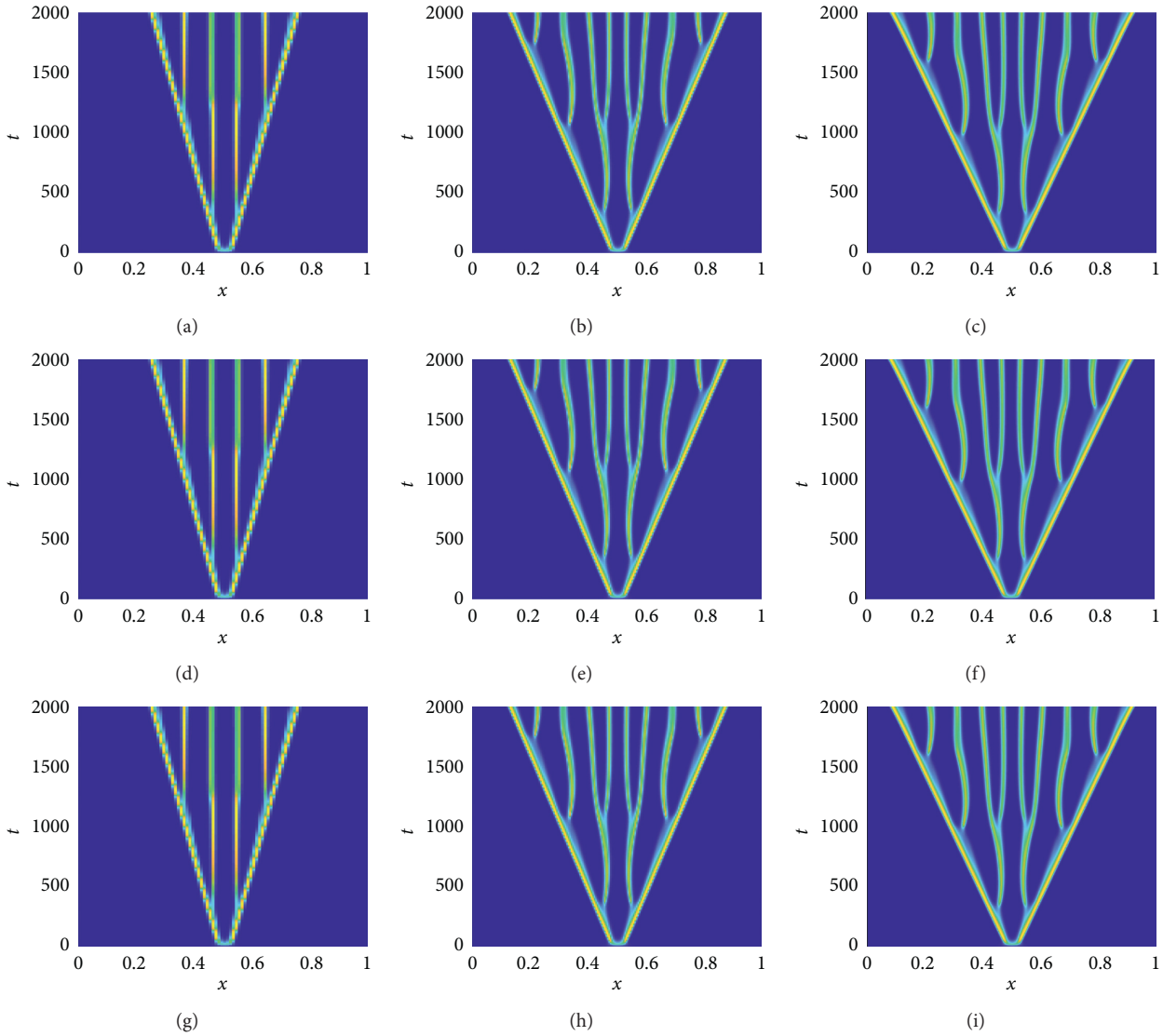


FIGURE 7: The wave-splitting (self-replication of the pulse) by the variable-θ method over $0 \le t \le 2000$ with various $(\Delta t, \Delta x)$. The image $I_{k\ell}$ represents the numerical solution obtained with the mesh resolution $(\Delta t, \Delta x) = (10^{-k}, 0.01/2^{\ell-1})$. (a) (I_{11}). (b) (I_{12}). (c) (I_{13}). (d) (I_{21}). (e) (I_{22}). (f) (I_{23}). (g) (I_{31}). (h) (I_{32}). (i) (I_{33}).

In order to investigate effectiveness of the variable-θ method and oscillatory behaviors of the CN method as well, we have carried out numerical experiments for the Gierer–Meinhardt model with a relatively low spatial resolution. Figure 4 presents numerical solutions at two different times for u_1 of the Gierer–Meinhardt model with the spatial resolution $\Delta x = \Delta y = 1/32$. When the time step size is set $\Delta t = 0.05$, the variable-θ method evolves the numerical solution as shown in Figures 4(a) and 4(d), for which the final steady-state pattern is the same as that in [32]. On the contrary, with the same $\Delta t = 0.05$, the CN method has produced a quite different pattern as in Figures 4(b) and 4(e), due to the nonsmooth initial values (Figure 9). However, the CN method can recover the correct steady-state pattern when it runs with $\Delta t = 0.005$, as depicted in Figures 4(c) and 4(f). For a similar accuracy, the variable-θ method (taking 170 s) is about 7 times more efficient than the CN method (taking 1242 s).

We summarize our experiments with the Gray–Scott and Gierer–Meinhardt models in 2D as follows:

(i) The variable-θ method shows the same accuracy as the CN method for problems of smooth data

(ii) For nonsmooth data, the variable-θ method evolves a smooth solution for all choices of Δt, while the CN method introduces spurious oscillations to alter the solution unless the time step size is sufficiently small

(iii) When a large time step size is desirable, the suggested method is a few times more efficient than the CN method for a similar accuracy

Table 2: L^∞-error $E_\infty[T]$ and the elapsed time (CPU) for the numerical solution of the Gray–Scott model in the 2D space: $T = 1.0$.

$\epsilon = 10^{-6}$	Implicit			CN			Variable-θ		
$(\Delta t, \Delta x = \Delta y)$	$E_\infty[T]$		CPU	$E_\infty[T]$		CPU	$E_\infty[T]$		CPU
	u_1	u_2		u_1	u_2		u_1	u_2	
$(0.05, 0.1)$	$4.3 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	0.036 s	$9.6 \cdot 10^{-4}$	$9.4 \cdot 10^{-4}$	0.047 s	$9.6 \cdot 10^{-4}$	$9.4 \cdot 10^{-4}$	0.062 s
$(0.025, 0.05)$	$2.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	0.225 s	$2.5 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	0.280 s	$2.5 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	0.394 s
$(0.0125, 0.025)$	$8.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-3}$	1.629 s	$6.4 \cdot 10^{-5}$	$5.5 \cdot 10^{-5}$	2.102 s	$6.4 \cdot 10^{-5}$	$5.5 \cdot 10^{-5}$	2.798 s
Conv. order	1.2	1.2		1.9	2.0		1.9	2.0	

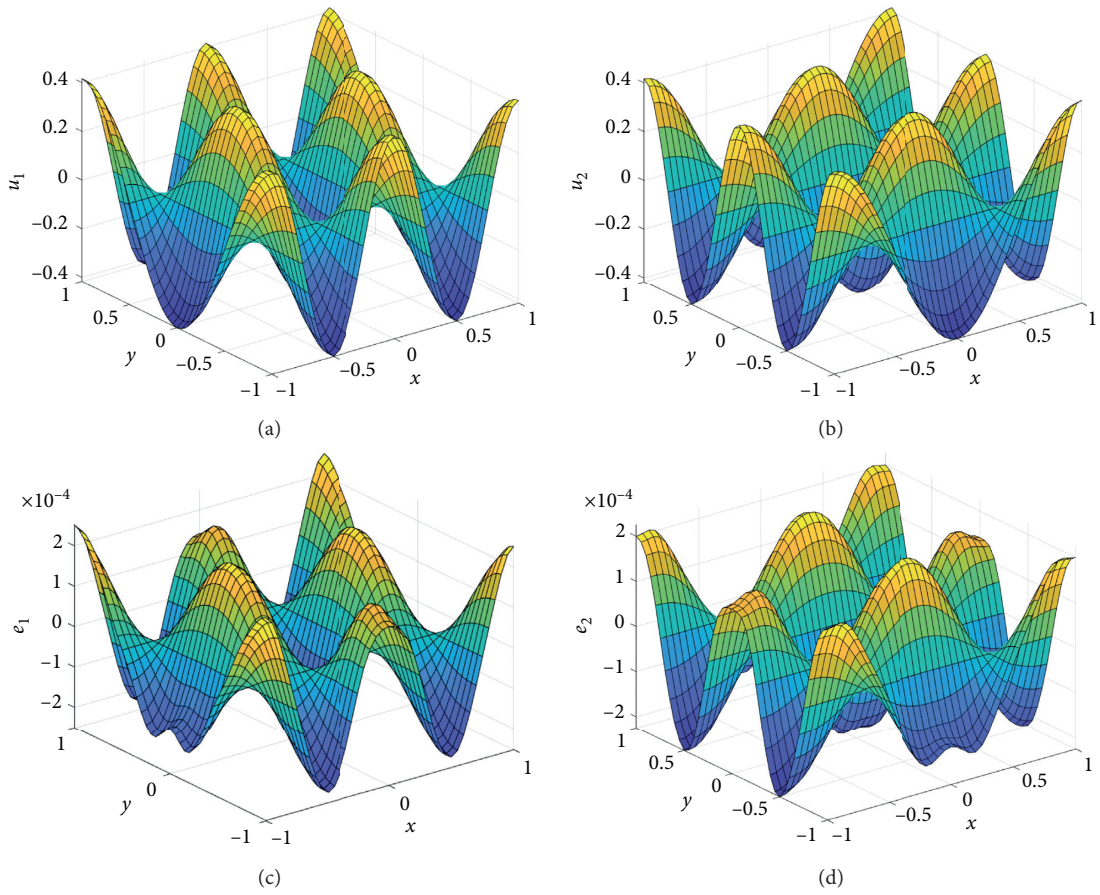


Figure 8: The numerical solution of the variable-θ method and its error at $T = 1.0$ for the Gray–Scott model, when $\Delta t = 0.025$ and $\Delta x = \Delta y = 0.05$: (a) u_1, (b) u_2, (c) $e_1 = u_1 - \widehat{u_1}$, and (d) $e_2 = u_2 - \widehat{u_2}$.

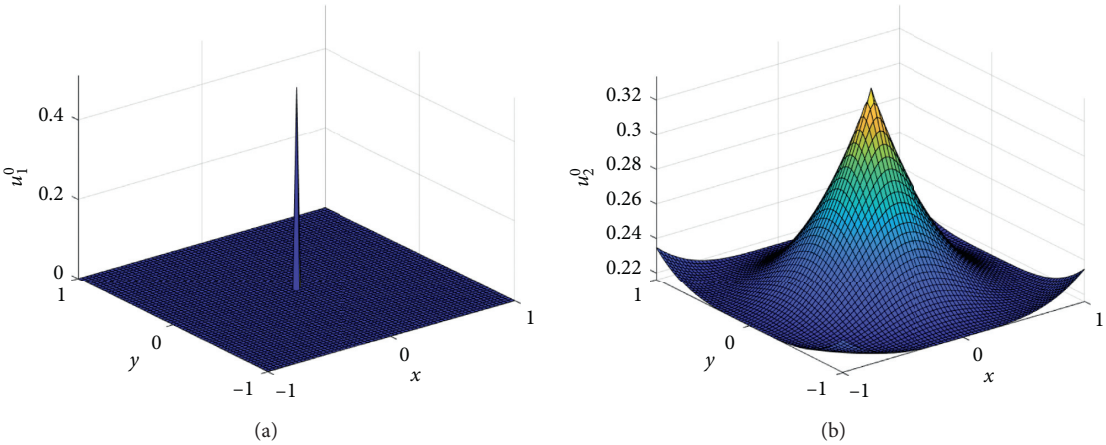


Figure 9: The initial values in (52) for the Gierer–Meinhardt model with $\epsilon = 0.04$ at the mesh resolution $\Delta x = \Delta y = 1/32$: (a) $u_1(x, y, 0)$ and (b) $u_2(x, y, 0)$.

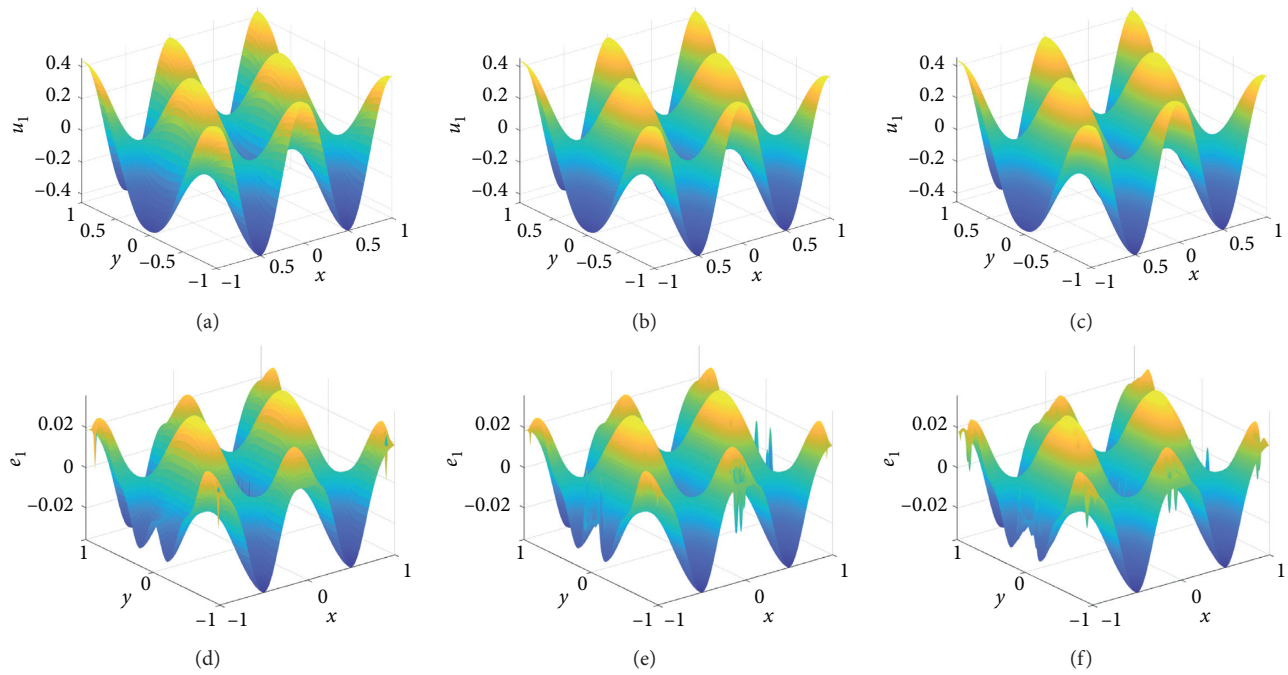


FIGURE 10: Numerical solutions for u_1 of the Gray–Scott model at $T = 1.0$ by the variable-θ method varying $\Delta x = \Delta y$ with fixed $\Delta t = 0.25$. The image I_ℓ represents the numerical solution obtained with the mesh resolution $(\Delta t, \Delta x = \Delta y) = (0.25, 0.02/2^{\ell-1})$ and I'_ℓ represents the error of I_ℓ. (a) (I_1). (b) (I_2). (c) (I_3). (d) (I'_1). (e) (I'_2). (f) (I'_3).

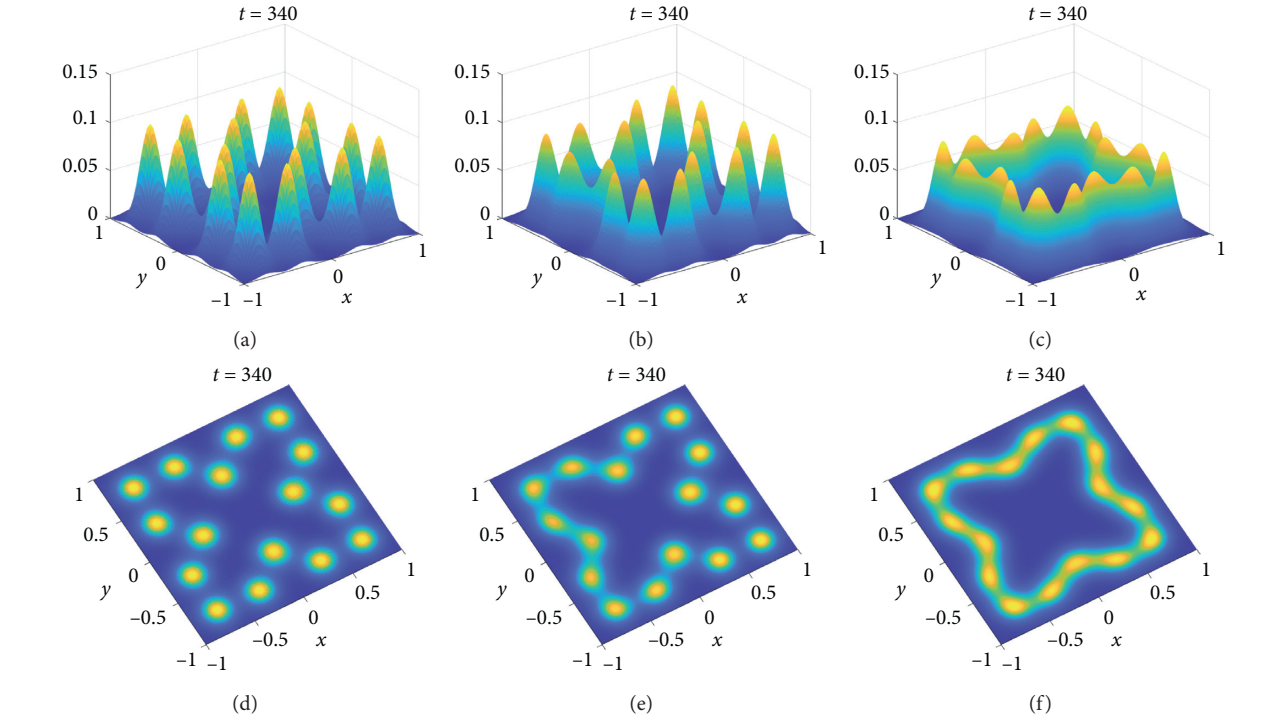


FIGURE 11: Numerical solutions for u_1 of the Gierer–Meinhardt model at $T = 340$ by the variable-θ method, varying $\Delta x = \Delta y$ with fixed $\Delta t = 0.1$. The image I_ℓ represents the numerical solution obtained with the mesh resolution $(\Delta t, \Delta x = \Delta y) = (0.1, 0.02/2^{\ell-1})$ and I'_ℓ represents the aerial view of I_ℓ. (a) (I_1). (b) (I_2). (c) (I_3). (d) (I'_1). (e) (I'_2). (f) (I'_3).

Remark 3. Although the variable-θ method can employ larger time step sizes than the CN method to get stable numerical solutions for problems of nonsmooth data, one may not set the time step size too large, due to an accuracy issue rather than the stability issue. Furthermore, for nonlinear problems, the overall stability of the numerical algorithm can be determined by not

only grid sizes but also numerical schemes including methods of dealing with the nonlinear terms.

Figure 10 presents numerical solutions and their errors of u_1 for the Gray–Scott model (48)–(50) at $T = 1.0$ by the variable-θ method, varying $\Delta x = \Delta y$ with fixed $\Delta t = 0.25$. Compared with Figure 8, the solutions show stability and a good accuracy, although the time step size is as large as $\Delta t = 0.25$. As shown in the bottom line in Figure 10, all three spatially different cases show the same level of errors since the entire errors are dominated by temporal direction errors. We conclude from this example that grid sizes in both temporal and spatial directions would not significantly affect the stability of the proposed method when the initial condition is smooth and the nonlinearity is not severe.

As an example of nonsmooth data and severe nonlinearity, we select the Gierer–Meinhardt model (51) and (52) to simulate with large temporal step sizes. When $\Delta t \geq 0.2$, the proposed algorithm introduced a rapid decay of solution values independently of the spatial grid size, so that the pattern is not formed appropriately. We believe that it is due to the error incorporated with the reaction term (19) when $u^{n+1/2}$ is approximated by the extrapolation scheme (18). However, when $\Delta t \leq 0.1$, our method produces stable solutions for all choices of spatial grid sizes. Figure 11 presents numerical solutions of u_1 for the Gierer–Meinhardt model at $T = 340$ by the variable-θ method with fixed $\Delta t = 0.1$ and various $\Delta x = \Delta y$. Note that for Gierer–Meinhardt model, the pattern forming is slow down as the spatial grid size becomes smaller, as shown in (I_2) and (I_3) of Figure 11; this tendency has been observed for all other choices of $\Delta t \leq 0.1$. This is another example that accuracy of the numerical solution is much more sensitive to the spatial mesh resolution than the temporal one.

6. Conclusions

The *Crank–Nicolson* (CN) method has been a popular second-order time-stepping procedure for the numerical solution of systems of nonlinear RD equations. However, the CN method may introduce spurious oscillations for nonsmooth data unless the time step size is sufficiently small. We have studied a nonoscillatory time-stepping procedure for RD equations, called a *variable-θ method*, as a perturbation of the CN method. In each time level, the new method detects points of potential oscillations and resolves the solution applying the implicit method locally at those points. The proposed time-stepping procedure has proven nonoscillatory and having a second-order temporal accuracy, although the initial conditions are nonsmooth. Various examples have been considered to show effectiveness of the method. We also have performed a sensitivity analysis for the numerical solution of biological pattern forming models to conclude that the numerical solution is much more sensitive to the spatial mesh resolution than the temporal one.

Data Availability

The experiment data and figures used to support the findings of this study are included within the article, and the MATLAB codes for the experiments are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by NSF-MCB (1714157) awarded to George V. Popescu.

References

[1] A. M. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 237, no. 641, pp. 37–72, 1952.

[2] A. Gierer and H. Meinhardt, "A theory of biological pattern formation," *Kybernetik*, vol. 12, no. 1, pp. 30–39, 1972.

[3] J. Schnakenberg, "Simple chemical reaction systems with limit cycle behaviour," *Journal of Theoretical Biology*, vol. 81, no. 3, pp. 389–400, 1979.

[4] P. Gray and S. K. Scott, "Autocatalytic reactions in the isothermal, continuous stirred tank reactor: isolas and other forms of multistability," *Chemical Engineering Science*, vol. 38, no. 1, pp. 29–43, 1983.

[5] I. Prigogine and G. Nicolis, "Self-organisation in nonequilibrium systems: towards a dynamics of complexity," in *Bifurcation Analysis*, pp. 3–12, Springer, Berlin, Germany, 1985.

[6] I. Lengyel and I. R. Epstein, "Modeling of turing structures in the chlorite–iodide–malonic acid–starch reaction system," *Science*, vol. 251, no. 4994, pp. 650–652, 1991.

[7] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation," *Science*, vol. 329, no. 5999, pp. 1616–1620, 2010.

[8] P. K. Maini and R. E. Baker, *Developmental Biology: Mathematical Modelling of Development*, Wiley, Hoboken, NJ, USA, 2001.

[9] K. U. Torii, "Two-dimensional spatial patterning in developmental systems," *Trends in Cell Biology*, vol. 22, no. 8, pp. 438–446, 2012.

[10] E. Crampin and P. Maini, "Reaction-diffusion models for biological pattern formation," *Methods and Applications of Analysis*, vol. 8, no. 3, pp. 415–428, 2001.

[11] C. Chiu and N. Walkington, "An ADI method for hysteretic reaction-diffusion systems," *SIAM Journal on Numerical Analysis*, vol. 34, no. 3, pp. 1185–1206, 1997.

[12] R. I. Fernandes, B. Bialecki, and G. Fairweather, "An ADI extrapolated Crank-Nicolson orthogonal spline collocation method for nonlinear reaction-diffusion systems on evolving domains," *Journal of Computational Physics*, vol. 299, pp. 561–580, 2015.

[13] I. Sgura, B. Bozzini, and D. Lacitignola, "Numerical approximation of Turing patterns in electrodeposition by ADI methods," *Journal of Computational and Applied Mathematics*, vol. 236, no. 16, pp. 4132–4147, 2012.

[14] F. Shakeri and M. Dehghan, "The finite volume spectral element method to solve Turing models in the biological pattern formation," *Computers & Mathematics with Applications*, vol. 62, no. 12, pp. 4322–4336, 2011.

[15] R. I. Fernandes and G. Fairweather, "An ADI extrapolated Crank-Nicolson orthogonal spline collocation method for nonlinear reaction-diffusion systems," *Journal of Computational Physics*, vol. 231, no. 19, pp. 6248–6267, 2012.

[16] J. Douglas Jr., "On the numerical integration of $^{\partial/2}u\partial x^2 + \partial^2 u/\partial y^2 = \partial u/\partial t$ by implicit methods," *Journal of the Society for Industrial and Applied Mathematics*, vol. 3, no. 1, pp. 42–65, 1955.

[17] J. Douglas Jr. and D. W. Peaceman, "Numerical solution of two-dimensional heat-flow problems," *AIChE Journal*, vol. 1, no. 4, pp. 505–512, 1955.

[18] D. W. Peaceman and H. H. Rachford Jr., "The numerical solution of parabolic and elliptic differential equations," *Journal of the Society for Industrial and Applied Mathematics*, vol. 3, no. 1, pp. 28–41, 1955.

[19] J. Douglas Jr., S. Kim, and H. Lim, "An improved alternating-direction method for a viscous wave equation," *Contemporary Mathematics*, vol. 329, pp. 99–104, 2003.

[20] H. Lim, S. Kim, and J. Douglas Jr., "Numerical methods for viscous and nonviscous wave equations," *Applied Numerical Mathematics*, vol. 57, no. 2, pp. 194–212, 2007.

[21] J. Douglas Jr. and S. Kim, "Improved accuracy for locally one-dimensional methods for parabolic equations," *Mathematical Models and Methods in Applied Sciences*, vol. 11, no. 9, pp. 1563–1579, 2001.

[22] J. Crank and P. Nicolson, "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 43, pp. 50–67, Cambridge University Press, Cambridge, UK, 1947.

[23] J. D. Lawson and J. L. Morris, "The extrapolation of first order methods for parabolic partial differential equations. I," *SIAM Journal on Numerical Analysis*, vol. 15, no. 6, pp. 1212–1224, 1978.

[24] P. Lee and S. Kim, "A variable-θ method for parabolic problems of nonsmooth data," *Computers & Mathematics with Applications*, vol. 79, no. 4, pp. 962–981, 2020.

[25] R. Varga, *Matrix Iterative Analysis*, Springer, Berlin, Germany, 2nd edition, 2000.

[26] A. Q. M. Khaliq, T. A. Biala, S. S. Alzahrani, and K. M. Furati, "Linearly implicit predictor-corrector methods for space-fractional reaction-diffusion equations with non-smooth initial data," *Computers & Mathematics with Applications*, vol. 75, no. 8, pp. 2629–2657, 2018.

[27] P. Gray, "Autocatalytic reactions in the isothermal, continuous stirred tank reactor: oscillations and instabilities in the system $a + 2b \longrightarrow 3b; b \longrightarrow c$," *Chemical Engineering Science*, vol. 39, no. 6, pp. 1087–1097, 1984.

[28] P. A. Zegeling and H. Kok, "Adaptive moving mesh computations for reaction–diffusion systems," *Journal of Computational and Applied Mathematics*, vol. 168, no. 1-2, pp. 519–528, 2004.

[29] C. Gingras and P. G. Kry, "Procedural modelling with reaction diffusion and growth of thin shells," in *Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019*, Canadian Human-Computer Communications Society, New York, NY, USA, pp. 1–7, 2019.

[30] P. Lee, T. W. Kim, and S. Kim, "Accurate and efficient numerical solutions for elliptic obstacle problems," *Journal of Inequalities and Applications*, vol. 34, no. 7, pp. 1–25, 2017.

[31] M. McCourt, N. Dovidio, and M. Gilbert, "Spectral methods for resolving spike dynamics in the gierer-meinhardt model," *Communications in Computational Physics*, vol. 3, pp. 659–678, 2008.

[32] Z. Qiao, "Numerical investigations of the dynamical behaviors and instabilities for the gierer-meinhardt system," *Communications in Computational Physics*, vol. 3, no. 2, pp. 406–426, 2008.

Hindawi
Complexity
Volume 2020, Article ID 5049825, 14 pages
https://doi.org/10.1155/2020/5049825

Research Article

Mathematical Model Analysis and Simulation of Visceral Leishmaniasis, Kashgar, Xinjiang, 2004–2016

Yateng Song,[1] Tailei Zhang,[2] Hui Li,[3] Kai Wang,[4] and Xiaobo Lu[5]

[1]*College of Public Health, Xinjiang Medical University, Urumqi 830011, China*
[2]*School of Science, Chang'an University, Xi'an 710064, China*
[3]*Central Laboratory, Xinjiang Medical University, Urumqi 830011, China*
[4]*College of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830011, China*
[5]*Department of Infectious Diseases, The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China*

Correspondence should be addressed to Kai Wang; wangkaimath@sina.com and Xiaobo Lu; xjykdluxiaobo@126.com

Received 12 September 2019; Revised 3 January 2020; Accepted 11 February 2020; Published 24 April 2020

Guest Editor: George V. Popescu

Visceral leishmaniasis (VL), known as kala-azar, is a serious parasitic disease. After malaria, VL is the second largest parasitic killer. This paper focuses on the VL transmission around sandflies, dogs, and people. Kashgar is located on the southwestern edge of Xinjiang, where kala-azar parasite infection occurs every year. According to the cases reported in the Kashgar Prefecture from 2004 to 2016, we proposed a dynamic model based on these three populations. The SEIR model was established for human population, the SI model was established for sandfly population, and the SI model was established for dog population. We fitted the model to cumulative cases from 2004 to 2016 for the epidemic in Kashgar and predicted that the cumulative incidence of kala-azar in Kashgar would continue to increase, but its growth rate would gradually slow down, which means that the number of cases would gradually decrease every year. We also estimated the basic reproduction number $R_0 = 1.76$ (95% CI: 1.49–1.93). The sensitivity analysis shows that the mutual infection between sandfly and dog contributes the most to the basic reproduction number, while the transmission proportion of sandfly to the susceptible person and the mutual infection between sandfly and dog contribute the most to the number of leishmaniasis human cases. Therefore, according to the sensitivity analysis results, reducing the contact between sandflies and dogs is an effective way to reduce kala-azar.

1. Introduction

Visceral leishmaniasis (VL) is also called kala-azar, which is a chronic infectious disease caused by *Leishmania infantum*. VL has ranked as the second largest parasitic killer after malaria and draws worldwide attention because of its severity; about 12 million people are affected by it around the world [1]. The pathogen causing visceral leishmaniasis are *L. donovani*, *L. infantum*, and *L. chagasi*. The incubation period for VL is generally 3 – 6 months, at least 10 days, while the longest is 9 years [1]. Humans and dogs are the main infection sources while sandflies are the main carriers.

The World Health Organization (WHO) lists leishmaniasis as the most easily neglected disease. It is reported that approximately 30% of new clinical cases and 58,000 deaths of leishmaniasis occur worldwide each year [2]. According to the characteristics of the source of infection, VL is mainly divided into three types: human-borne, canine-borne, and wildlife-borne in China [3]. In the early days of the People's Republic of China, human-borne kala-azar was the main disease in 16 provinces, such as Shandong and Henan. After prevention and treatment, it was basically eliminated. In recent years, VL has been prevalent in six provinces including Xinjiang, Inner Mongolia, Gansu, Sichuan, Shaanxi, and Shanxi [3]. Among six provinces, the typical VL in Xinjiang is wildlife-borne, and some areas have human-borne VL, while canine-borne is the main transmission way in Gansu.

Xinjiang is the largest provincial administrative region in China and is also a high-risk area for VL, especially in the Kashgar Prefecture in southern Xinjiang. The Kashgar Prefecture is located in the southwestern margin of the Xinjiang Uygur Autonomous Region. Due to its unique geographical

environment, there are two types of visceral leishmaniasis, human and desert [4]. Since the 1990s, the incidence of VL in Kashgar has been increasing year by year. From 2005 to 2015, Xinjiang's VL has experienced significant fluctuation. In 2007–2012, the incidence of VL in Kashgar accounted for more than 90% of the total population [5].

We utilized the data of human leishmaniasis cases from 2004 to 2016 reported by the Center for Disease Control and Prevention in Xinjiang, and then we plot bar diagrams about total, sex, and age which are presented in Figures 1(a)–1(c). From Figure 1(a), we could find that the reported cases of VL in Kashgar Prefecture of Xinjiang fluctuate considerably, with outbreaks occurring in October and November of 2008, 2009, and 2015, respectively. From Figure 1(b), it shows that there are more reported cases of VL in males than in females. Figure 1(c) shows that there are more reported cases of VL in the younger age group.

In recent years, a lot of mathematical models have been established around humans, dogs, and sandflies to understand the transmission dynamics of VL (see for instance [6–14]). For example, Muhammad and Ali [6] considered the uniform mixing of the population, established the SEIHR model for the population and the SEIR model for the dog, which used sandfly as the medium, modeled and backwarded the branching of zoonotic visceral leishmaniasis, and calculated the basic reproduction number R_0 for optimal control. Elmojtaba et al. [7] developed a mathematical model to study the dynamics of visceral leishmaniasis in Sudan, considering three different groups, analyzing the balance point and its stability and providing a basis for controlling and eliminating diseases. Zamir et al. [8] established SIR mathematical models for humans, hosts, and media, using the Routh–Hurwitz standard and next-generation methods to obtain threshold conditions and give numerical simulation results. There are three main control strategies in the controls of VL model: parameter control strategy, optimal control strategy, and control strategy selection using simulation. However, the most generalized control strategy is the parameter control strategy. When the parameters are adjusted, people can apply it to a real-world control strategy, and there is also a lot of research on optimal control strategy and control strategy selection using simulation. Simulation comparison is the most common method of VL mathematic control modeling in simulation. It can compare the human infected population with control and without control. Meanwhile, it also proves the effectiveness of the combined control strategies [9]. With reference to these established mathematical models, we have established a mathematical model based on logistic process.

The purpose of this paper is to know the transmission state about VL among humans, dogs, and sandflies in the Kashgar Prefecture of Xinjiang. Firstly, we proposed a model to simulate the cumulative data of the Kashgar Prefecture and estimate the basic reproduction number and the dynamic behavior of the model. Then, sensitivity analysis was performed on the number of leishmaniasis human/dog/sandfly cases and the basic reproduction number R_0 based on some key parameters. Finally, we explored some effective strategies for the prevention and control of VL in Kashgar.

The article is organized as follows. In Section 2, this paper introduces the model of VL and gives the expression of the basic reproduction number and the parameter value. The dynamical behaviors of the model are analyzed to better understand the transmission trends of the disease. Some mathematical analyses are given in Section 3. The numerical simulations, prediction of the epidemic trends for the next decades, estimation of the basic reproduction number, and sensitivity analysis of the basic reproduction number and the number of infected humans/sandflies/dogs are presented in Section 4. In Section 5, some brief summaries and discussion are given.

2. Mathematical Model

In order to study the transmission of VL in Kashgar, we developed a mathematical model based on humans, dogs, and sandflies, where the human population is divided into four groups: the susceptible, the exposed, the infected, and the recovered denoted by S_h, E_h, I_h, and R_h, respectively. The dog population is divided into two groups: the susceptible and the infected, denoted by S_r and I_r, respectively. We divide the sandfly population into two subclasses: the susceptible and the infected, denoted by S_v and I_v, respectively. And the total population for humans, dogs, and sandflies is $N_h = S_h + E_h + I_h + R_h$; $N_v = S_v + I_v$; $N_r = S_r + I_r$. The flowchart of VL transmission is illustrated in Figure 2.

The transmission process of VL is described by the following eight differential equations:

$$\begin{cases} \dfrac{dS_h}{dt} = r_1 S_h\left(1 - \dfrac{S_h + E_h + I_h + R_h}{K_1}\right) - \beta_1 S_h I_v, \\ \dfrac{dE_h}{dt} = \beta_1 S_h I_v - \mu_h E_h - \omega E_h, \\ \dfrac{dI_h}{dt} = \omega E_h - (\mu_h + \delta_1) I_h - \gamma I_h, \\ \dfrac{dR_h}{dt} = \gamma I_h - \mu_h R_h, \\ \dfrac{dS_v}{dt} = r_2 S_v\left(1 - \dfrac{S_v + I_v}{K_2}\right) - (\beta_2 I_h + \beta_3 I_r) S_v, \\ \dfrac{dI_v}{dt} = (\beta_2 I_h + \beta_3 I_r) S_v - \mu_v I_v, \\ \dfrac{dS_r}{dt} = r_3 S_r\left(1 - \dfrac{S_r + I_r}{K_3}\right) - \beta_4 I_v S_r, \\ \dfrac{dI_r}{dt} = \beta_4 I_v S_r - (\mu_r + \delta_2) I_r. \end{cases} \tag{1}$$

We assume that the susceptible individuals S_h and dogs S_r are bitten by infected sandflies with β_1 and β_4 infection

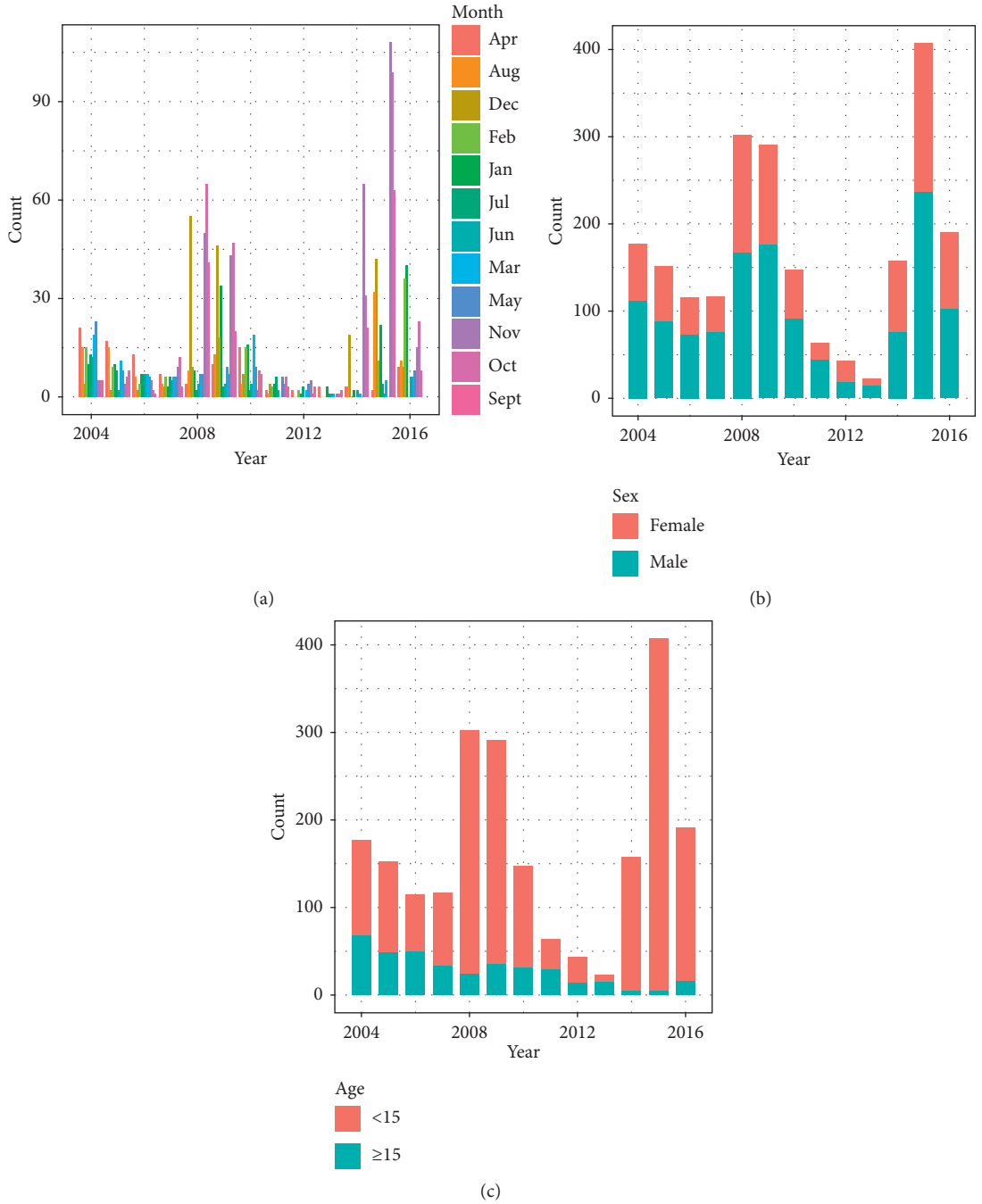


FIGURE 1: The data of human leishmaniasis cases from 2004 to 2016. (a) The reported data are described at different years. (b) The reported data are described at different genders and (c) different ages from 2004 to 2016.

rates, respectively. And the transmission probability from an infected dog or an infected individual to a susceptible sandfly is β_3 and β_2, respectively. $1/\omega$ and $1/\gamma$ indicate the latency and recovery rate of VL, respectively. The natural mortality rates of humans, vectors, and reserves are μ_h, μ_v, and μ_r, respectively. The death rate of humans and reserves caused by VL are δ_1 and δ_2, respectively. The susceptible human population is governed by the logistic growth with carrying capacity K_1 as well as intrinsic growth rate r_1. In the absence of disease, the vector population density grows according to a logistic curve carrying capacity K_2, with an intrinsic growth rate r_2. In the absence of disease, the dog population density grows according to a logistic curve carrying capacity K_3, with an intrinsic growth rate r_3.

3. Mathematical Analysis

3.1. Basic Reproduction Number. The basic reproduction number R_0 refers to the number of people infected by a patient during the average period of illness when all people are susceptible. R_0 was calculated using the next-generation

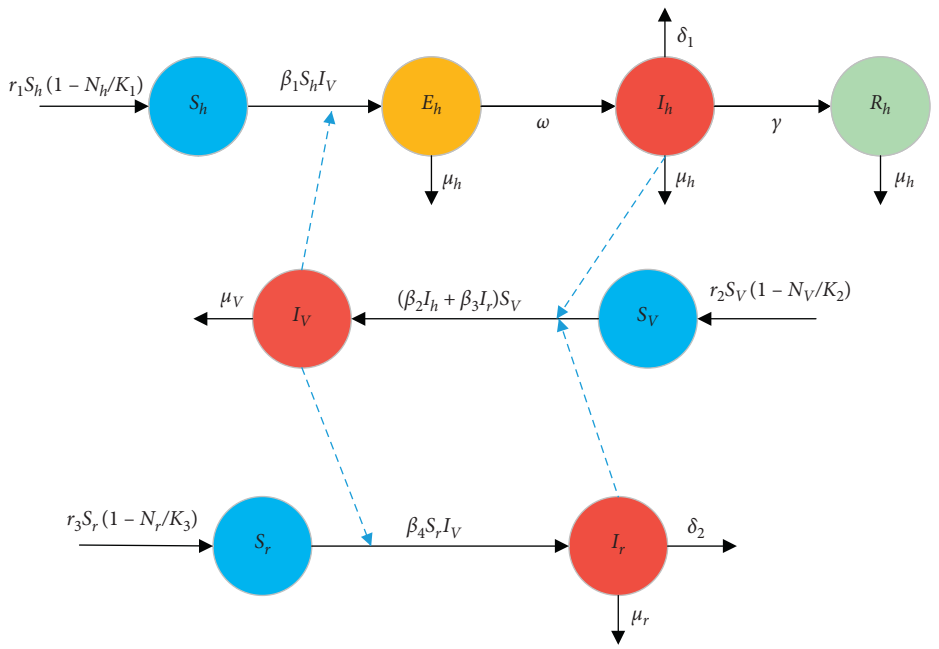


Figure 2: Flow diagram of the VL transmission model.

matrix [15, 16]. The infected compartments are E_h, I_h, I_v, and I_r, giving $m = 4$, so we have

$$\mathscr{F} = \begin{bmatrix} \beta_1 S_h I_v \\ 0 \\ (\beta_2 I_h + \beta_3 I_r) S_v \\ \beta_4 I_v S_r \end{bmatrix}, \quad \nu = \begin{bmatrix} (\mu_h + \omega) E_h \\ (\mu_h + \delta_1 + \gamma) I_h - \omega E_h \\ \mu_r I_r \\ (\mu_r + \delta_2) I_r \end{bmatrix}. \tag{2}$$

Then,

$$F = \begin{bmatrix} 0 & 0 & \beta_1 K_1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_2 K_2 & 0 & \beta_3 K_3 \\ 0 & 0 & \beta_4 K_3 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \mu_h + \omega & 0 & 0 & 0 \\ -\omega & \mu_h + \delta_1 + \gamma & 0 & 0 \\ 0 & 0 & \mu_v & 0 \\ 0 & 0 & 0 & \mu_r + \delta_2 \end{bmatrix}. \tag{3}$$

Thus,

$$\mathbf{FV}^{-1} = \begin{bmatrix} 0 & 0 & \dfrac{\beta_1 K_1}{\mu_v} & 0 \\ 0 & 0 & 0 & 0 \\ \dfrac{\omega\beta_2 K_2}{(\mu_h+\omega)(\mu_h+\delta_1+\gamma)} & \dfrac{\beta_2 K_2}{\mu_h+\delta_1+\gamma} & 0 & \dfrac{\beta_3 K_2}{\mu_r+\delta_2} \\ 0 & 0 & \dfrac{\beta_4 K_3}{\mu_v} & 0 \end{bmatrix}. \tag{4}$$

The reproduction number is given by $R_0 = \rho(\mathrm{FV}^{-1})$, and

$$R_0 = \sqrt{\frac{K_2\left(K_3\beta_3\beta_4\left(\mu_h^2 + \delta_1\mu_h + \gamma\mu_h + \delta_1\omega + \gamma\omega + \mu_h\omega\right) + \omega K_1\beta_1\beta_2\left(\delta_2 + \mu_r\right)\right)}{\mu_v\left(\delta_2 + \mu_r\right)\left(\mu_h + \omega\right)\left(\delta_1 + \gamma + \mu_h\right)}}. \tag{5}$$

R_0 can be used as a basic indicator in the study of VL propagation dynamics model. Usually, when $R_0 = 1$, it can be used as a threshold for the demise of the disease. When $R_0 > 1$, the disease will not die and eventually turn into endemic disease; when $R_0 < 1$, the disease will die out naturally.

3.2. Dynamic Behaviors for the Model. The dynamical behavior of model (1) is analyzed to better understand the transmission trends of the disease. We begin with elementary properties of solutions to the model (1). From biological considerations, we study model (1) in the set

$$\Gamma = \left\{ \left(S_h, E_h, I_h, R_h, S_v, I_v, S_r, I_r\right) \in \mathbb{R}_+^8 : 0 < S_h \le K_1, 0 < S_v \le K_2, 0 < S_r \le K_3, N_h \le \frac{\left(\mu_h + r_1\right)^2 K_1}{4\mu_h r_1}, N_v \le \frac{\left(\mu_v + r_2\right)^2 K_2}{4\mu_v r_2}, \right.$$
$$\left. N_r \le \frac{\left(\mu_r + r_3\right)^2 K_3}{4\mu_r r_3} \right\}. \tag{6}$$

In fact, by using

$$\begin{aligned} \frac{\mathrm{d}S_h}{\mathrm{d}t} &\le r_1 S_h \left(1 - \frac{S_h}{K_1}\right), \\ \frac{\mathrm{d}S_v}{\mathrm{d}t} &\le r_2 S_v \left(1 - \frac{S_v}{K_2}\right), \\ \frac{\mathrm{d}S_r}{\mathrm{d}t} &\le r_3 S_r \left(1 - \frac{S_r}{K_3}\right), \end{aligned} \tag{7}$$

we can deduce that $S_h(t) \le K_1$, $S_v(t) \le K_2$, and $S_r(t) \le K_3$ whenever $S_h(0) \le K_1$, $S_v(0) \le K_2$, and $S_r(0) \le K_3$. Adding the first four equations of (1) yields

$$\begin{aligned} \frac{\mathrm{d}N_h}{\mathrm{d}t} &= r_1 S_h \left(1 - \frac{N_h}{K_1}\right) - \mu_h N_h - \delta_1 I_h + \mu_h S_h \\ &\le \left(r_1 + \mu_h\right) S_h - \frac{r_1 S_h^2}{K_1} - \mu_h N_h \\ &\le \frac{\left(r_1 + \mu_h\right)^2 K_1}{4 r_1} - \mu_h N_h. \end{aligned} \tag{8}$$

In view of system (8), it is easy to check that $N_h'(t) \le 0$ wherever $N_h \ge \left(\left(r_1 + \mu_h\right)^2 K_1 / 4\mu_h r_1\right)$. The same conclusion can be drawn for N_v and N_r. According to the above analysis, Γ is a maximum positive invariant set of (1). There always exists a disease-free equilibrium P_0 (K_1, 0, 0, 0, K_2, 0, K_3, 0). In the following, we will show the stability for the disease-free equilibrium.

Theorem 1. *The disease-free equilibrium P_0 is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.*

Proof. The Jacobian matrix of model (1) at P_0 is given by

$$J\left(P_0\right) = \begin{bmatrix} -r_1 & -r_1 & -r_1 & -r_1 & 0 & -\beta_1 K_1 & 0 & 0 \\ 0 & -\left(\mu_h + \omega\right) & 0 & 0 & 0 & \beta_1 K_1 & 0 & 0 \\ 0 & \omega & -\left(\mu_h + \delta_1 + \gamma\right) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma & -\mu_h & 0 & 0 & 0 & 0 \\ 0 & 0 & -\beta_2 K_2 & 0 & -r_2 & -r_2 & 0 & -\beta_3 K_2 \\ 0 & 0 & \beta_2 K_2 & 0 & 0 & -\mu_v & 0 & \beta_3 K_2 \\ 0 & 0 & 0 & 0 & 0 & -\beta_4 K_3 & -r_3 & -r_3 \\ 0 & 0 & 0 & 0 & 0 & \beta_4 K_3 & 0 & -\left(\mu_r + \delta_2\right) \end{bmatrix}. \tag{9}$$

It is easy to see that there are four negative eigenvalues of $J(P_0)$: $-r_1, -r_2, -r_3, -\mu_h$. The other eigenvalues are determined by the following fourth-order equation:

$$\lambda^4 + a_1 \lambda^3 + a_2 \lambda^2 + a_3 \lambda + a_4 = 0, \tag{10}$$

where

$$
\begin{aligned}
a_1 &= m_1 + m_2 + n + \mu_v,\\
a_2 &= m_1 m_2 + \mu_v n - \beta_3\beta_4 K_2 K_3 + (m_1 + m_2)(\mu_v + n),\\
a_3 &= m_1 m_2 (\mu_v + n) + (m_1 + m_2)(\mu_v n - \beta_3\beta_4 K_2 K_3)\\
&\quad - \omega\beta_1\beta_2 K_1 K_2,\\
a_4 &= m_1 m_2 (\mu_v n - \beta_3\beta_4 K_2 K_3) - n\omega\beta_1\beta_2 K_1 K_2,
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
m_1 &= \mu_h + \omega,\\
m_2 &= \mu_h + \delta_1 + \gamma,\\
n &= \mu_r + \delta_2.
\end{aligned}
\tag{12}
$$

The inequality $R_0 < 1$ implies that

$$
\begin{aligned}
\beta_3\beta_4 K_2 K_3 &< \mu_v n,\\
\omega\beta_1\beta_2 K_1 K_2 &< m_1 m_2 \mu_v,\\
m_1 m_2 \beta_3\beta_4 K_2 K_3 + n\omega\beta_1\beta_2 K_1 K_2 &< m_1 m_2 \mu_v n.
\end{aligned}
\tag{13}
$$

From (13), we have $a_i > 0$ for $i = 1, 2, 3, 4$. On account of the above inequalities, we have

$$
\begin{aligned}
a_1 a_2 - a_3 &= (m_1 + m_2 + n + \mu_v)[m_1 m_2 + \mu_v n - \beta_3\beta_4 K_2 K_3] + (m_1 + m_2)(\mu_v + n)] - m_1 m_2 (\mu_v + n) - (m_1 + m_2)(\mu_v n - \beta_3\beta_4 K_2 K_3)\\
&\quad + \omega\beta_1\beta_2 K_1 K_2\\
&= m_1 m_2 (m_1 + m_2) + (\mu_v + n)(\mu_v n - \beta_3\beta_4 K_2 K_3) + (m_1 + m_2 + n + \mu_v)(m_1 + m_2)(\mu_v + n) + \omega\beta_1\beta_2 K_1 K_2 > 0.
\end{aligned}
\tag{14}
$$

Furthermore,

$$
\begin{aligned}
a_3(a_1 a_2 - a_3) - a_1^2 a_4 &= [m_1 m_2 (\mu_v + n) + (m_1 + m_2)(\mu_v n - \beta_3\beta_4 K_2 K_3) - \omega\beta_1\beta_2 K_1 K_2]\\
&\quad [m_1 m_2 (m_1 + m_2) + (\mu_v + n)(\mu_v n - \beta_3\beta_4 K_2 K_3) + (m_1 + m_2 + n + \mu_v)(m_1 + m_2)(\mu_v + n) + \omega\beta_1\beta_2 K_1 K_2]\\
&\quad - m_1 m_2 (\mu_v n - \beta_3\beta_4 K_2 K_3)(m_1 + m_2 + n + \mu_v)^2 + n\omega\beta_1\beta_2 K_1 K_2 (m_1 + m_2 + n + \mu_v)^2\\
&> (m_1 + m_2)(\mu_v n - \beta_3\beta_4 K_2 K_3)[m_1 m_2 (m_1 + m_2) + (m_1 + m_2 + n + \mu_v)(m_1 + m_2)(\mu_v + n) + \omega\beta_1\beta_2 K_1 K_2]\\
&\quad - m_1 m_2 (\mu_v n - \beta_3\beta_4 K_2 K_3)(m_1 + m_2 + n + \mu_v)^2\\
&= \left(m_1^2 + m_2^2\right)(\mu_v n - \beta_3\beta_4 K_2 K_3)(m_1 + m_2)(\mu_v + n) + \left(m_1^2 + m_2^2 + m_1 m_2\right)(\mu_v n - \beta_3\beta_4 K_2 K_3)(\mu_v + n)^2 > 0.
\end{aligned}
\tag{15}
$$

Therefore, each eigenvalue of equation (10) admits negative real part. When $R_0 < 1$, the disease-free equilibrium P_0 is locally asymptotically stable. When $R_0 > 1$, we conclude from $a_4 > 0$ that equation (10) has at least one positive real root; hence, the disease-free equilibrium P_0 is unstable.

Theorem 2. *As $R_0 < 1$, the disease-free equilibrium is globally asymptotically stable in Γ.*

Proof. Let us consider the following Lyapunov function:

$$
\begin{aligned}
V(t) &= \omega n\beta_2 K_2 E_h(t) + m_1 n\beta_2 K_2 I_h(t) + m_1 m_2 n I_v(t)\\
&\quad + m_1 m_2 \beta_3 K_2 I_r(t),
\end{aligned}
\tag{16}
$$

where

$$
\begin{aligned}
m_1 &= \mu_h + \omega,\\
m_2 &= \mu_h + \delta_1 + \gamma,\\
n &= \mu_r + \delta_2.
\end{aligned}
\tag{17}
$$

The derivative of $V(t)$ along solutions of (1) is

$$
\begin{aligned}
\dot{V}(t) &= \omega n\beta_1\beta_2 K_2 S_h I_v - m_1 m_2 n\beta_2 K_2 I_h + m_1 m_2 n S_v(\beta_2 I_h + \beta_3 I_r)\\
&\quad - m_1 m_2 \mu_v n I_v + m_1 m_2 \beta_3\beta_4 K_2 S_r I_v - m_1 m_2 n\beta_3 K_2 I_r\\
&\le [m_1 m_2 \beta_3\beta_4 K_2 K_3 + n\omega\beta_1\beta_2 K_1 K_2 - m_1 m_2 \mu_v n] I_v\\
&= m_1 m_2 \mu_v n\left(R_0^2 - 1\right) I_v\\
&\le 0.
\end{aligned}
\tag{18}
$$

By LaSalle's invariance principle, the omega limit set of each solution starting from Γ lies in an invariant set contained in

$$\Omega = \{(S_h, E_h, I_h, R_h, S_v, I_v, S_r, I_r) \in \Gamma: I_v = 0\}. \quad (19)$$

It can be verified that the only invariant set contained in Ω is the singleton $\{P_0\}$. Then, the disease-free equilibrium P_0 is globally asymptotically stable. This completes the proof. In the following, we will consider the dynamical behavior of model (1) when $R_0 > 1$.

Theorem 3. *For any solution $(S_h(t), E_h(t), I_h(t), R_h(t), S_v(t), I_v(t), S_r(t), I_r(t))$ of (1) with initial values $S_h(0) > 0$, $E_h(0) \geq 0$, $I_h(0) \geq 0$, $R_h(0) \geq 0$, $S_v(0) > 0$, $I_v(0) \geq 0$, $S_r(0) > 0$, $I_r(0) \geq 0$, and $I_h(0) + I_v(0) + I_r(0) > 0$, there exists $\nu > 0$ such that*

$$\begin{aligned} \liminf_{t\longrightarrow\infty} I_h(t) &> \nu, \\ \liminf_{t\longrightarrow\infty} I_v(t) &> \nu, \\ \liminf_{t\longrightarrow\infty} I_r(t) &> \nu, \end{aligned} \quad (20)$$

as $R_0 > 1$.

Proof. Let

$$\begin{aligned} X &= \{(S_h, E_h, I_h, R_h, S_v, I_v, S_r, I_r) \in \mathbb{R}_+^8: S_h > 0, S_v > 0, S_r > 0\}, \\ X_0 &= \{(S_h, E_h, I_h, R_h, S_v, I_v, S_r, I_r) \in X: I_h + I_v + I_r > 0\}, \\ \partial X_0 &= X \backslash X_0. \end{aligned} \quad (21)$$

Next, we will show that system (1) is uniformly persistent with respect to $(X_0, \partial X_0)$. Obviously, X is positively invariant with respect to (1). Set $a = \max\{\mu_h + \delta_1 + \gamma, \mu_v, \mu_r + \delta_2\}$. The inequality $(I_h(t) + I_v(t) + I_r(t))' \geq -a(I_h(t) + I_v(t) + I_r(t))$ and $I_h(0) + I_v(0) + I_r(0) > 0$ imply $I_h(t) + I_v(t) + I_r(t) > 0$ for all $t \geq 0$ as $I_h(0) + I_v(0) + I_r(0) > 0$. Therefore, X_0 is also a positive invariant set for system (1). Furthermore, by Theorem 3, there exists a compact set $\mathcal{C}$ in which all solutions of (1) initiated in X will enter and remain forever after. The compactness condition $(C_{4.2})$ in Thieme [17] is easily verified for this set $\mathcal{C}$. Denote

$$M_\partial = \{(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0)): (S_h(t), E_h(t), I_h(t), R_h(t), S_v(t), I_v(t), S_r(t), I_r(t)) \in \partial X_0, t \geq 0\}. \quad (22)$$

We first claim that

$$M_\partial = \{(S_h, 0, 0, R_h, S_v, 0, S_r, 0): S_h, R_h, S_v, S_r \geq 0\}. \quad (23)$$

Suppose that $(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0)) \in M_\partial$. It suffices to show $E_h(t) = 0$ for all $t \geq 0$. If it is not true, then there exists $t_0 > 0$ such that $E_h(t_0) > 0$. Then, $E_h'(t) \geq (\mu_h + \omega)E_h(g)$ implies $E_h(t) > 0$ for all $t \geq t_0$. From the third equation of (1), it may be concluded that

$$\begin{aligned} I_h(t) &= e^{-(\mu_h+\delta_1+\gamma)(t-t_0)} I_h(t_0) + \int_{t_0}^{t} e^{-(\mu_h+\delta_1+\gamma)(t-\theta)} \omega E_h(\theta) \mathrm{d}\theta \\ &\geq \int_{t_0}^{t} e^{-(\mu_h+\delta_1+\gamma)(t-\theta)} \omega E_h(\theta) \mathrm{d}\theta > 0, \end{aligned} \quad (24)$$

for all $t > t_0$. This is a contradiction with

$$(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0)) \in M_\partial. \quad (25)$$

Thus, (23) is valid.

Denote

$$\Omega = \bigcup \{\omega(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0)): (S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0)) \in M_\partial\}, \quad (26)$$

where $\omega(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0))$ is the omega limit set of the solution to (1) through $(S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), I_v(0), S_r(0), I_r(0))$. Restricting (1) on M_∂ yields

$$\begin{cases} \dfrac{\mathrm{d}S_h}{\mathrm{d}t} = r_1 S_h \left(1 - \dfrac{S_h + R_h}{K_1}\right), \\ \dfrac{\mathrm{d}R_h}{\mathrm{d}t} = -\mu_h R_h, \\ \dfrac{\mathrm{d}S_v}{\mathrm{d}t} = r_2 S_v \left(1 - \dfrac{S_v}{K_2}\right), \\ \dfrac{\mathrm{d}S_r}{\mathrm{d}t} = r_3 S_r \left(1 - \dfrac{S_r}{K_3}\right). \end{cases} \quad (27)$$

A trivial verification shows that system (27) has a unique equilibrium $(K_1, 0, K_2, K_3)$. Thus, $P_0(K_1, 0, 0, 0, K_2, 0, K_3, 0)$ is the unique equilibrium of (1) in M_∂. It is easily seen that $(K_1, 0, K_2, K_3)$ is globally asymptotically stable. Therefore, we have $\Omega = \{P_0\}$. And P_0 is a covering of Ω, which is isolated and is acyclic (since there exists no solution in M_∂ which links P_0 to itself). Finally, the proof will be done if P_0 is a weak repeller for X_0, i.e.,

$$\limsup_{t\longrightarrow\infty} \mathrm{dist}(\Psi(t), P_0) > 0, \quad (28)$$

where $\Psi(t) = (S_h(t), E_h(t), I_h(t), R_h(t), S_v(t), I_v(t), S_r(t), I_r(t))$ is an arbitrary solution with initial value in X_0. By Leenheer and Smith (Proof of Lemma 3.5, [18]), we only need to prove that $W^s(P_0) \cap X_0 = \phi$ where $W^s(P_0)$ is the stable manifold of E_0. Suppose it is not true, then there exists a solution $(S_h(t), E_h(t), I_h(t), R_h(t), S_v(t), I_v(t), S_r(t), I_r(t))$ in X_0, such that as $t \longrightarrow \infty$,

$$S_h(t) \longrightarrow K_1, E_h(t) \longrightarrow 0, I_h(t) \longrightarrow 0, R_h(t) \longrightarrow 0,$$
$$S_v(t) \longrightarrow K_2, I_v(t) \longrightarrow 0, S_r(t) \longrightarrow K_3, I_r(t) \longrightarrow 0. \tag{29}$$

When $R_0 > 1$,

$$\omega\beta_2 K_2(\mu_r + \delta_2)\beta_1 K_1 + \beta_3 K_2(\mu_h + \omega)(\mu_h + \delta_1 + \gamma)\beta_4 K_3 > (\mu_h + \omega)(\mu_h + \delta_1 + \gamma)(\mu_r + \delta_2)\mu_v. \tag{30}$$

On account of (30), we may choose $\rho > 0$, $\sigma > 0$, $\eta > 0$, and $\epsilon > 0$ such that

$$\begin{cases} 0, & < \omega\beta_2(K_2 - \epsilon)(\mu_r + \delta_2) - \rho \ll 1, \\ 0, & < \beta_3(K_2 - \epsilon)(\mu_r + \omega)(\mu_h + \delta_1 + \gamma) - \sigma \ll 1, \\ 0, & < \eta - (\mu_h + \omega)(\mu_h + \delta_1 + \gamma)(\mu_r + \delta_2) \ll 1. \end{cases} \tag{31}$$

$$\rho\beta_1(K_1 - \epsilon) + \sigma\beta_4(K_3 - \epsilon) - \eta\mu_v > 0. \tag{32}$$

From (31), we see that $((\mu_h + \omega)/\omega)\rho < (\mu_h + \omega)(\mu_r + \delta_2)$ $\beta_2(K_2 - \epsilon) < (\beta_2(K_2 - \epsilon))/(\mu_h + \delta_1 + \gamma)\eta$, $\eta\beta_3(K_2 - \epsilon) > \sigma(\mu_r + \delta_2)$. Therefore, there exists ξ such that

$$\frac{\mu_h + \omega}{\omega}\rho < \xi < \frac{\beta_2(K_2 - \epsilon)}{\mu_h + \delta_1 + \gamma}\eta. \tag{33}$$

For $\epsilon > 0$, by (29), there exists $T > 0$ such that

$$\begin{aligned} K_1 - \epsilon < S_h(t) < K_1 + \epsilon, \\ K_2 - \epsilon < S_v(t) < K_2 + \epsilon, \\ K_3 - \epsilon < S_r(t) < K_3 + \epsilon, \end{aligned} \tag{34}$$

for all $t \geq T$. Let

$$L(t) = \rho E_h(t) + \xi I_h(t) + \eta I_v(t) + \sigma I_r(t). \tag{35}$$

The derivative of L along the solution $(S_h(t), E_h(t), I_h(t), R_h(t), S_v(t), I_v(t), S_r(t), I_r(t))$ is given by

$$\begin{aligned} L'(t) &= [\xi\omega - \rho(\mu_h + \omega)]E_h + [\eta\beta_2 S_v - \xi(\mu_h + \delta_1 + \gamma)]I_h + [\rho\beta_1 S_h + \sigma\beta_4 S_r - \eta\mu_v]I_v + [\eta\beta_3 S_v - \sigma(\mu_r + \delta_2)]I_r \\ &\geq [\xi\omega - \rho(\mu_h + \omega)]E_h + [\eta\beta_2(K_2 - \epsilon) - \xi(\mu_h + \delta_1 + \gamma)]I_h + [\rho\beta_1(K_1 - \epsilon) + \sigma\beta_4(K_3 - \epsilon) - \eta\mu_v]I_v \\ &\quad + [\eta\beta_3(K_2 - \epsilon) - \sigma(\mu_r + \delta_2)]I_r \\ &\geq \varrho L(t), \end{aligned} \tag{36}$$

for all $t \geq T$, where

$$\varrho = \min\left\{\frac{\xi\omega - \rho(\mu_h + \omega)}{\rho}, \frac{\eta\beta_2(K_2 - \epsilon) - \xi(\mu_h + \delta_1 + \gamma)}{\xi}, \frac{\rho\beta_1(K_1 - \epsilon) + \sigma\beta_4(K_3 - \epsilon) - \eta\mu_v}{\eta}, \frac{\eta\beta_3(K_2 - \epsilon) - \sigma(\mu_r + \delta_2)}{\sigma}\right\} > 0. \tag{37}$$

Hence, $L(t) \longrightarrow \infty$ as $t \longrightarrow \infty$, which contradicts to the boundedness of $L(t)$. This completes the proof.

Remark 1. Theorems 1~3 state that the basic reproduction number R_0 is a sharp threshold value for model (1). As $R_0 < 1$, the disease-free equilibrium P_0 is globally stable, i.e., the disease will go to extinction. As $R_0 > 1$, the disease is uniformly persistent, i.e., the disease will become an endemic in the meaning of persistence.

4. Results

We applied model (1) to study the infection status of VL in Kashgar, Xinjiang. Most of the parameters were obtained from the literature, and some of them were assumed or simulated to have more realistic results. These parameter values are listed in Table 1. The interpretation of the parameters values are as follows:

(1) According to the population statistics in Kashgar Prefecture in Xinjiang Statistical Yearbook 2004–2016 [20], we have estimated the carrying capacity of population $K_1 = 5000000$.

(2) According to the terrain and climate factors of Kashgar and the survival habits of the sandflies [21–24], the carrying capacity of sandflies $K_2 = 7000000$ has been estimated in Kashgar. In 2015, the number of dogs in Kashgar Prefecture of Xinjiang was close to 80,000 [25], so we have assumed the carrying capacity of dogs in Xinjiang $K_3 = 100000$.

(3) The values of the parameter r_1 are obtained by numerical differentiation according to the population data in Kashgar, and the value of r_2 is calculated according to the life cycle of the sandflies (see [19]); the value of r_3 is assumed because there is no the specific data about dog.

Table 1: Parameter values (and their sources).

Parameters	Value	Interpretation*	Source
r_1	0.019	Intrinsic growth rate of human	Calculated
r_2	0.026	Intrinsic growth rate of vector	[19]
r_3	0.296	Intrinsic growth rate of reserve	Assumption
K_1	5000000	Human carrying capacity	Assumption
K_2	7000000	Vector carrying capacity	Assumption
K_3	100000	Reserve carrying capacity	Assumption
β_1	$1.35e-08$	Transmission probability from an infected sandfly to a susceptible human	Fitting
β_2	$3.565e-17$	Transmission probability from an infected human to a susceptible sand fly	Fitting
β_3	$3.515e-05$	Transmission probability from an infected reserve to a susceptible sandfly	Fitting
β_4	$1.23e-05$	Transmission probability from an infected sandfly to a susceptible reserve	Fitting
μ_h	0.0125	Natural mortality rate of human	[10]
μ_v	25.55	Natural mortality rate of vector	[11]
μ_r	0.1	Natural mortality rate of reserve	[12]
$1/\omega$	0.33	Human incubation period	[13]
δ_1	4.015	VL-induced death rate of humans	[8]
δ_2	3.65	VL-induced death rate of reserve	[13]
$1/\gamma$	2.5	Recovery period of VL	[14]

We model cumulative cases as a Poisson-distributed random variable because the Poisson distribution describes the number of observed events in an interval of time. We calibrate the model by sampling from the posterior distribution of parameter vector $\theta \mid \mathbf{y} = \{\beta_1, \beta_2, \beta_3, \beta_4\} \mid \mathbf{y}$, where vector $\mathbf{y}$ is derived from $(\mathrm{d}/\mathrm{d}t)Y(t) = \omega E_h$ and $Y(t)$ denotes the reported cumulative cases. We conduct sampling via MCMC (Markov Chain Monte Carlo) using the Metropolis–Hastings acceptance rule. The posterior density is

$$f_{\Theta|\mathbf{y}}(\theta \mid \mathbf{y}) = \prod_T \mathscr{L}(Y(t) \mid \theta) f_\Theta(\theta). \tag{38}$$

The prior density $f_\Theta(\theta)$ is the joint probability of four univariate priors. We consider that β_1, β_2, β_3, and β_4 are distributed according to $\mathcal{U}(0, 1)$. The program was implemented in R version 3.6.0. We sampled from 30,000 MCMC iterations and discarded the first 10,000 samples as a burn-in period. On the basis of these 20,000 samples, the point estimates and 95% confidence intervals for the transmission coefficients were calculated. The results are shown in Table 2.

The MCMC method was used to fit the cumulative incidence data of VL in Kashgar from 2004 to 2016, and the 95% confidence interval of the fitted curve was obtained (see Figure 3). It can be seen from the figure that the fitted values of the model matches are in accordance with the accumulated data values in the Kashgar Prefecture. Only some data fluctuate, but in any case, the cumulative data of VL in Kashgar are increasing year by year. This model is used to predict the prevalence of VL in the Kashgar Prefecture over the next decade (see Figure 4).

The basic reproduction number $R_0 = 1.76$(95% CI: 1.49–1.93). The result shows that $R_0 > 1$; according to the threshold theory, the disease will not disappear in the Kashgar Prefecture in a short period of time, and an endemic disease will be formed.

It is well known that the basic reproduction number R_0 is a very important parameter in the infectious disease model. In our model, R_0 is determined by the parameters of ω, δ_1, δ_2, γ, μ_h, μ_r, μ_v, β_1, β_2, β_3, and β_4. We use the Latin hypercube

Table 2: Parameter values for point estimation and 95% interval estimation.

Parameters	Point estimation	95% confidence interval
β_1	$1.35e-08$	$[1.06e-08,\ 1.94e-08]$
β_2	$3.565e-17$	$[2.87e-19,\ 2.44e-15]$
β_3	$3.515e-05$	$[2.067e-05,\ 1.002e-04]$
β_4	$1.23e-05$	$[3.18e-06,\ 2.36e-05]$

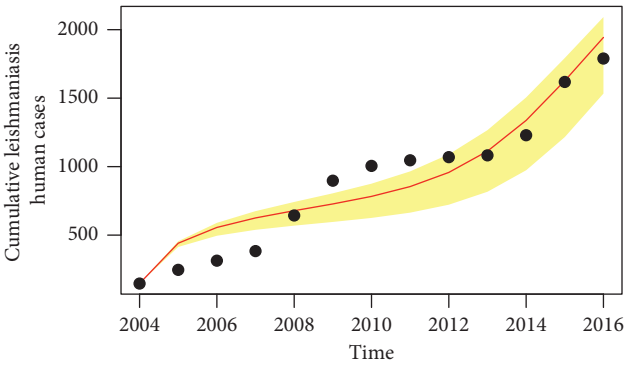


Figure 3: Cumulative incidence simulation of VL in Kashgar from 2004 to 2016.

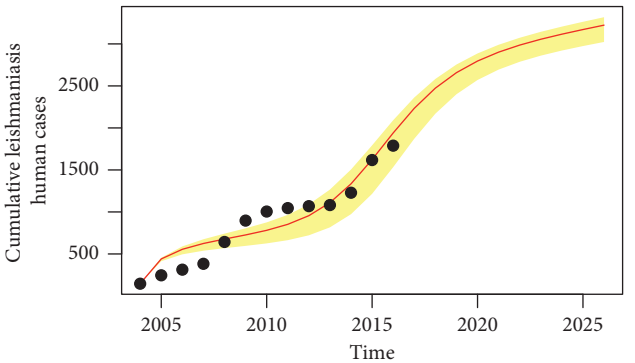


Figure 4: Prediction of cumulative incidence in 10 years from 2016 to 2026.

Table 3: Partial rank correlation coefficients (PRCCs) for R_0 on each input parameter variable.

Parameters	Distribution	2000 samples		3000 samples	
		PRCC	p value	PRCC	p value
ω	$N(3.03, 0.15^2)$	$9.295e-04$	0.504	$4.379e-04$	0.499
δ_1	$N(4.015, 0.2^2)$	$-1.995e-04$	0.493	$-1.619e-04$	0.496
δ_2	$N(3.65, 0.2^2)$	−0.247	<0.001	−0.2474	<0.001
γ	$N(0.4, 0.02^2)$	$-1.281e-03$	0.493	$3.011e-04$	0.502
μ_h	$N(0.0125, 0.0006^2)$	$8.376e-04$	0.511	$-1.041e-03$	0.505
μ_r	$N(0.1, 0.005^2)$	$-6.862e-03$	0.481	$-6.146e-03$	0.495
μ_v	$N(25.55, 3.28^2)$	−0.2327	<0.001	−0.2334	<0.001
β_1	$N(1.35e-08, 1e-08)$	$3.792e-04$	0.506	$-5.341e-05$	0.502
β_2	$N(3.565e-08, 2e-17^2)$	$-3.859e-04$	0.493	$-3.456e-04$	0.509
β_3	$N(3.515e-08, 1.5e-05^2)$	0.914	<0.001	0.914	<0.001
β_4	$N(1.23e-08, 5e-06^2)$	0.906	<0.001	0.906	<0.001

sampling (LHS) and partial rank correlation coefficients (PRCCs) to examine parameters which have a significant influence on the transmission of VL [26]. Using model (1), 2000 and 3000 samples are randomly generated by assuming a uniform distribution for each parameter based on values from Table 1.

We select eleven parameters as input variables and calculate the corresponding PRCC and p values for the eleven parameters. Table 3 and Figure 5 show the exact PRCC and p values of each input parameter and the effect on the basic reproduction number R_0, respectively. We assume the significance level $\alpha = 0.05$. The larger the absolute value of PRCC, the stronger the correlation between the input parameters and R_0. It can be seen from Table 3 that only δ_2, μ_v, β_3, and β_4 have significant impact on R_0. More concretely, parameters δ_2 and μ_v have negative impact on R_0 and parameters β_3 and β_4 have positive impact.

We perform sensitivity analysis of infected humans, infected sandflies, and infected dogs through evaluating the PRCCs with the parameters of interest of model (1) over time by choosing a normal distribution with mean value and standard deviation shown in Figure 6. In Figures 6(a)–6(c), we plot the PRCCs over time with respect to the infected humans, sandflies, and infected dogs, respectively. Figure 6(a) indicates that there are three PRCC values that are significantly different from zero. The first three parameters with most impact on the outcome (the number of infected humans) are the sandfly-to-human transmission (β_1), transmission from an infected dog to a susceptible sandfly (β_3), and transmission from an infected sandfly to a susceptible dog (β_4). Figure 6(c) indicates that there are two PRCC values that are significantly different from zero. The first two parameters with most impact on the outcome (the number of infected humans) are the transmission from an infected dog to a susceptible sandfly (β_3) and transmission from an infected sandfly to a susceptible dog (β_4). Figure 6(c) indicates that there are two PRCC values that are significantly different from zero. The first two parameters with most impact on the outcome (the number of infected humans) are the transmission from an infected sandfly to a susceptible dog (β_4) and transmission from an infected dog to a susceptible sandfly (β_3).

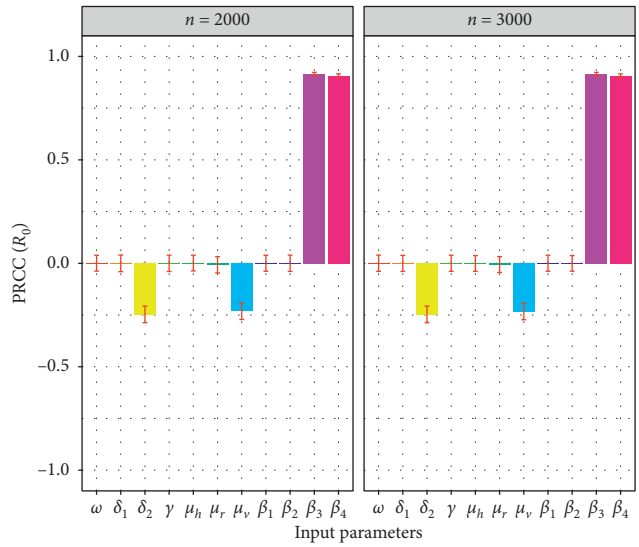


Figure 5: Partial rank correlation coefficient (PRCC) results for the dependence of R_0 on each parameter.

Through sensitivity analysis, we demonstrate that the infection rate from infected dogs to susceptible sandflies β_3 and infected sandflies to susceptible dogs β_4 is the most sensitive parameter of R_0. And β_1, β_3, and β_4 are the most sensitive parameters for the number of leishmaniasis human cases. Therefore, it is necessary to change the values of the parameters to observe their effects on the number of leishmaniasis human cases and R_0 (see Figure 7). From Figure 7(a), we observe that the number of leishmaniasis human cases decrease with the decrease of β_1. We change the values of β_3 and β_4 to 4, 2, 1/2, and 1/4 times the original values. We observe the effects of different values on the number of leishmaniasis human cases, and we find that the change of parameter β_3, β_4 can influence not only the number of leishmaniasis human cases, but also the peak time. As Figures 7(b) and 7(c) illustrate, when fixing other parameters at constant, the number of leishmaniasis human cases fall with the decrease of β_3 and β_4, respectively. And the peak of outbreak will be postponed.

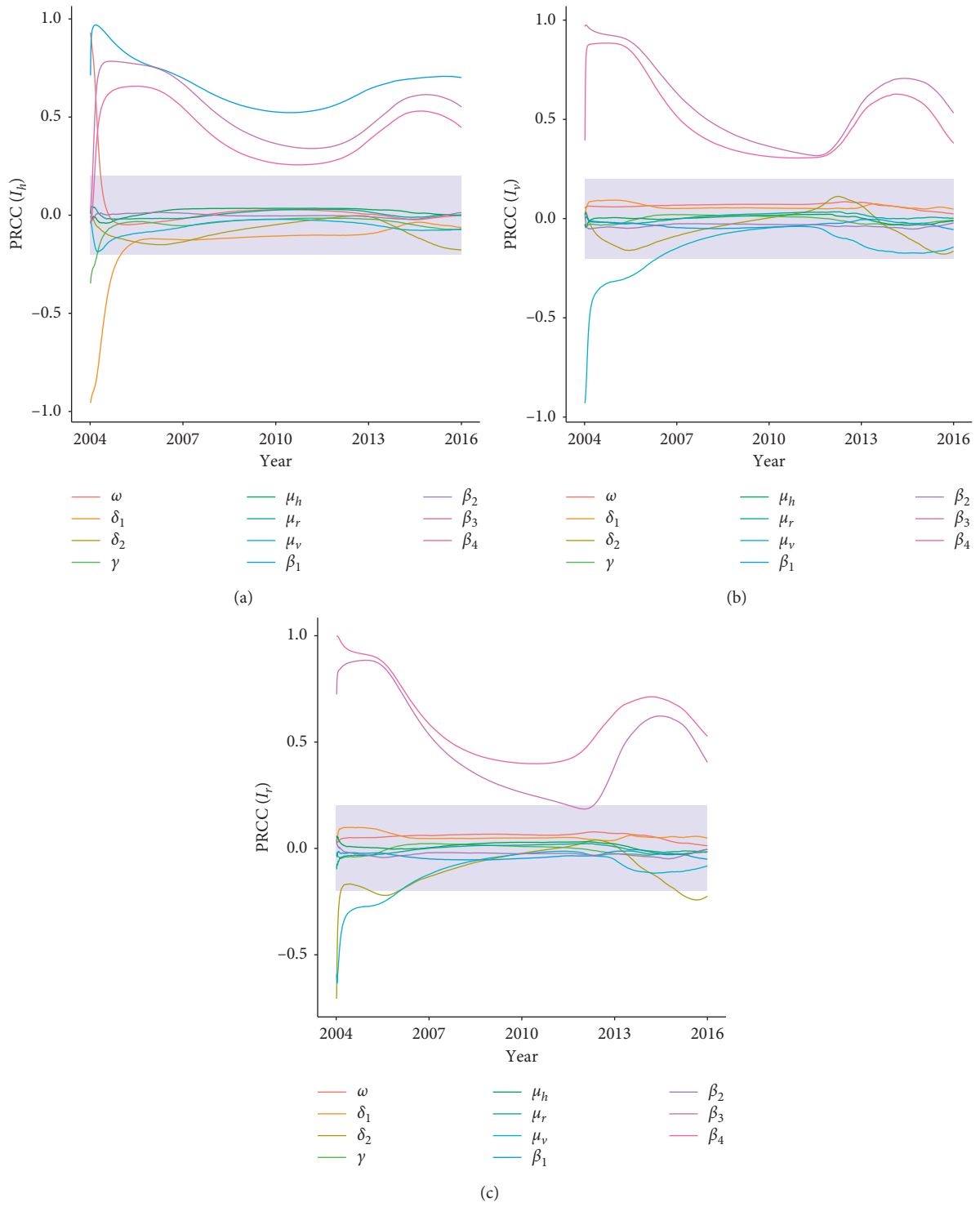


FIGURE 6: Sensitivity analysis: PRCC of the eleven parameters for (a) the number of infected humans, (b) the number of infected sandflies, and (c) the number of infected dogs.

Finally, in order to find better control strategies for VL transmission, we focus on changing the values of the parameters β_3 and β_4 to confirm the influence on the basic reproduction number R_0 (see Figure 8).

It can be seen from the figure that the parameters β_3 and β_4 have a strong influence on R_0, and the value of the basic reproduction number R_0 increased with the increases of β_3 and β_4; when β_3 is less than $6.854878e-06$ or β_4 is less than $2.39872e-06$, $R_0 < 1$, and the disease can be eliminated. This suggests that reducing the proportion rate between sandflies and dogs can effectively control the prevalence of VL.

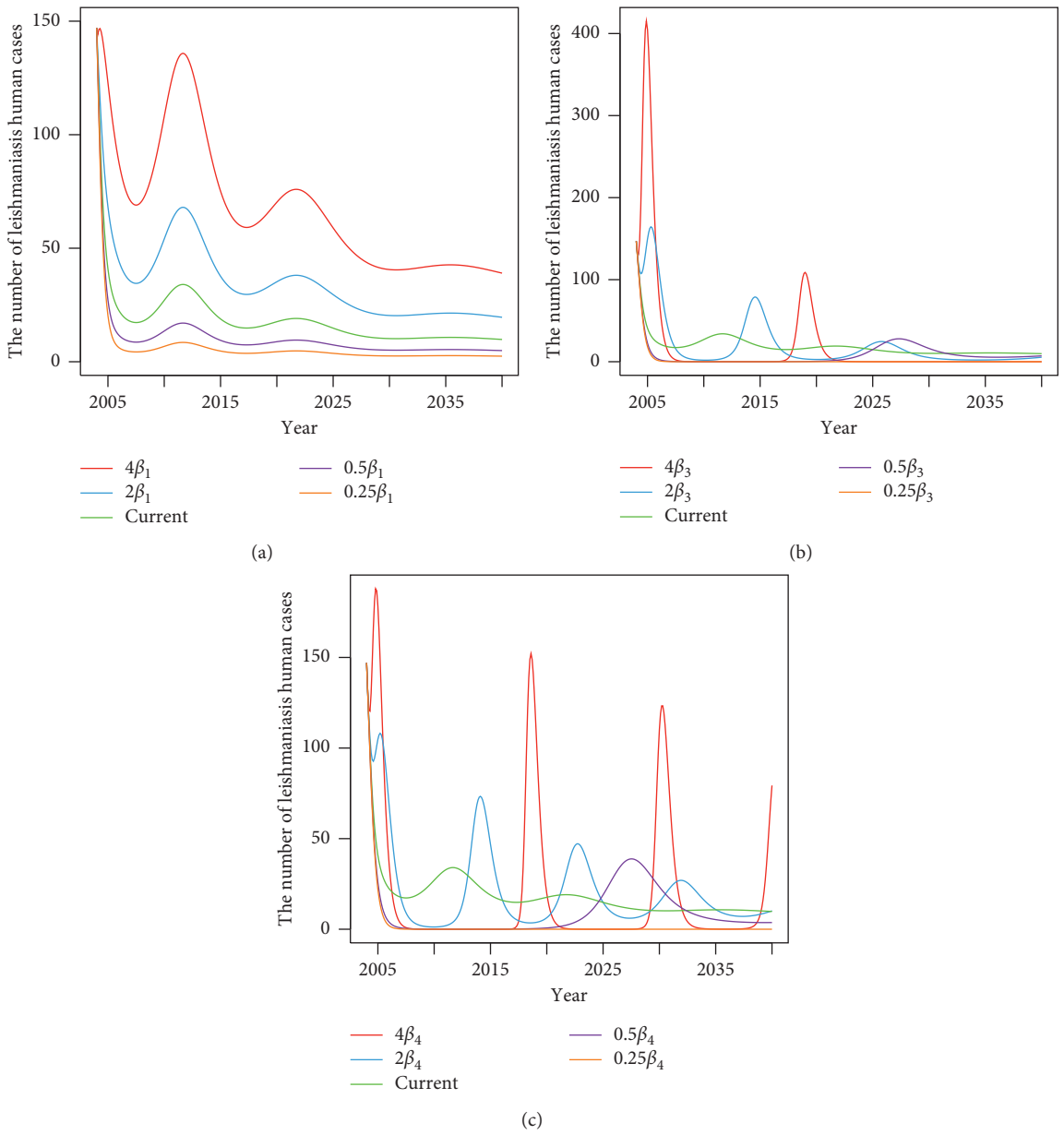


Figure 7: The influence of different values of parameters on the number of diseases of VL. (a) Versus β_1. (b) Versus β_3. (c) Versus β_4.

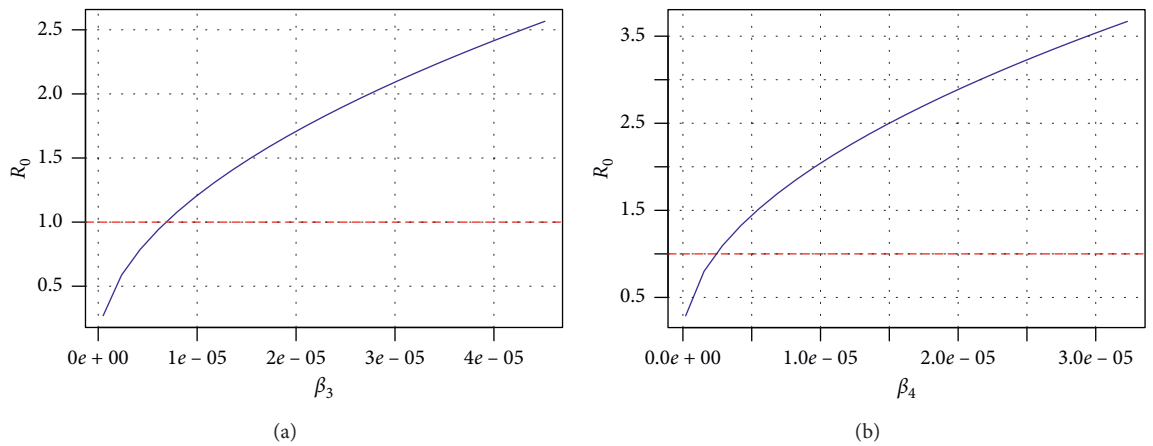


Figure 8: The influence of parameters on R_0 (red horizontal line is contour of $R_0 = 1$). (a) Versus β_3. (b) Versus β_4.

Through the above sensitivity analysis results, we find that the main cause of the outbreak of VL is the mutual contact infection between dogs and sandflies. Therefore, we can get some effective strategies to reduce the prevalence of VL: the effective way is to reduce the contact between dogs and sandflies, and thus we can give the dog a collar with impregnated insecticide, increase the control of dog and reduce its active area, spray insecticides vigorously, etc.

5. Conclusion and Discussion

VL is a serious parasitic disease. It has been endangered for several decades in Xinjiang and has become a major parasitic disease affecting the local social and economic development. Kashgar is a high-risk area for the occurrence of VL. In order to reveal the spread of VL in Kashgar and predict the prevalence of VL, this paper proposes a dynamic model of VL propagation with logistic growth. The model describes the transmission of VL among humans, dogs, and sandflies.

We use model (1) to fit the cumulative data of VL in Kashgar. As it can be seen from the simulation results in Figure 2, our model is consistent with the actual data of the cumulative cases of VL in Kashgar. The results show that there are certain reliability and rationality to study the prevalence of VL in Kashgar using logistic growth of the VL model. Using model (1), the basic reproduction number is estimated to be 1.76 (95% CI: 1.49–1.93) in the Kashgar Prefecture of Xinjiang. According to the threshold theory, it shows that VL will not disappear in the Kashgar area in a short time, and it may gradually become an endemic disease. According to the predictions for the next decade (see Figure 4), the cumulative incidence of VL in Kashgar is growing slowly, which means that the number of cases will gradually decrease over the next decade.

By selecting the sensitivity analysis of the parameters of interest for the basic reproduction number R_0 and the number of leishmaniasis human cases (see Figures 5 and 6), we can find that the most important factors affecting the basic reproduction number R_0 are β_3 and β_4, indicating that the infection between sandfly and dog has the greatest impact on R_0, and the most sensitive parameters affecting the number of leishmaniasis human cases are β_1, β_3, and β_4, which explains that the infection among the infected person, the dog, and the sandfly has the strongest impact on the number of diseases of VL. From Figure 7, we find that reducing the values of β_1, β_3, and β_4 can effectively reduce the value of the basic reproduction number R_0 and the number of leishmaniasis human cases. When the values of β_3 and β_4 drop to $6.854878e-06$ and $2.39872e-06$, respectively, the value of the basic reproduction number R_0 will decrease to 1, and VL will not be epidemic but gradually disappear in Kashgar. Therefore, in order to control the spread of the disease, effective strategies should be taken to prevent and control leishmaniasis in Kashgar; we can decrease the incidence of leishmaniasis in humans by reducing the contact between sandflies and dogs. Without considering costs, we can vector controls (e.g., environmental clean-up and insecticide sprayed around buildings) and dog controls (e.g., insecticide releasing dog collars and dog vaccinations). In order to reduce the bite of the sandflies, we can use insecticidal bed nets. Meanwhile, we should conduct public education on the dangers and prevention of leishmaniasis for people. In short, the model we have established now can reflect the dynamics of VL in Kashgar Prefecture.

Data Availability

The data will be available upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Yateng Song and Tailei Zhang contributed equally.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant no. 11961071).

References

[1] K. S. Rock, R. J. Quinnell, G. F. Medley, and O. Courtenay, "Chapter two-progress in the mathematical modelling of visceral leishmaniasis," *Advances in Parasitology*, vol. 94, pp. 49–131, 2016.

[2] A. Mubayi, C. Castillo-Chavez, G. Chowell et al., "Transmission dynamics and underreporting of Kala-azar in the Indian state of Bihar," *Journal of Theoretical Biology*, vol. 262, no. 1, pp. 177–185, 2010.

[3] C. J. Zheng, C. Z. Xue, W. P. Wu et al., "Analysis of epidemiological characteristics of black fever reported cases in China from 2005 to 2015," *Chinese Journal of Epidemiology*, vol. 38, no. 4, pp. 431–434, 2017.

[4] Kaisaier, C. Z. Zhu, W. P. Wu et al., "Reflections on prevention and control of VL in Kashgar, Xinjiang," *Chinese Journal of Disease Control and Prevention*, vol. 23, no. 5, pp. 23-24, 2008.

[5] L. Y. Wang, W. P. Wu, Y. Y. Guan et al., "Epidemiological characteristics of patients with high fever epidemic area in Kashgar, Xinjiang," *Chinese Journal of Pathogenic Biology*, vol. 8, no. 6, pp. 539-540, 2013.

[6] Z. Muhammad and R. Ali, "Zoonotic visceral Leishmania: modeling and control," *Journal of Applied & Computational Mathematics*, vol. 4, no. 4, 2015.

[7] I. M. Elmojtaba, J. Y. T. Mugisha, and M. H. A. Hashim, "Mathematical analysis of the dynamics of visceral leishmaniasis in the Sudan," *Applied Mathematics and Computation*, vol. 217, no. 6, pp. 2567–2578, 2010.

[8] M. Zamir, R. Sultana, R. Ali, W. Panhwar, and S. Kumar, "Study on the threshold conditions for infection of visceral leishmaniasis," *Sindh University Research Journal (Science Series)*, vol. 47, no. 3, pp. 619–622, 2015.

[9] K. Bi, Y. Chen, S. Zhaop, Y. Kuang, and C. H. John Wu, "Current visceral leishmaniasis research: a research review to inspire future study," *BioMed Research International*, vol. 2018, Article ID 9872095, 13 pages, 2018.

[10] Xinjiang Daily, http://www.xjdaily.com/.

[11] C. Dye, "The logic of visceral leishmaniasis control," *The American Journal of Tropical Medicine and Hygiene*, vol. 55, no. 2, pp. 125–130, 1996.
[12] M. I. ELmojtaba, "Mathematical model for the dynamics of visceral leishmaniasis-malaria co-infection," *Mathematical Methods in the Applied Sciences*, vol. 39, no. 16, pp. 4334–4353, 2016.
[13] S. Zhao, Y. Kuang, C. H. Wu et al., "Zoonotic visceral leishmaniasis transmission: modeling, backward bifurcation, and optimal control," *Journal of Mathematical Biology*, vol. 73, no. 6-7, pp. 1525–1560, 2016.
[14] A. D. P. Sevá, M. Martcheva, T. Necibe et al., "Efficacies of prevention and control measures applied during an outbreak in southwest Madrid, Spain," *PLoS One*, vol. 12, no. 10, 2017.
[15] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, "On the definition and the computation of the basic reproduction ratio R_0, in models for infectious diseases in heterogeneous populations," *Journal of Mathematical Biology*, vol. 28, no. 4, pp. 365–382, 1990.
[16] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1-2, pp. 29–48, 2002.
[17] H. R. Thieme, "Persistence under relaxed point-dissipativity (with application to an endemic model)," *SIAM Journal on Mathematical Analysis*, vol. 24, no. 2, pp. 407–435, 1993.
[18] P. D. Leenheer and H. Smith, "Virus dynamics: a global analysis," *SIAM Journal on Applied Mathematics*, vol. 63, no. 4, pp. 1313–1327, 2003.
[19] H. Y. Hu, "A simple method for solving the accurate value of intra-phase growth rate in population," *Anhui Agricultural Science Bulletin*, vol. 16, no. 3, pp. 173-174, 2010.
[20] Statistics Bureau of Xinjiang Uygur Autonomous Region, http://www.kashi.gov.cn/Category_1008/Index.aspx.
[21] G. Wang, J. Wang, J. S. Li et al., "Further study on the ecology of sandfly in the Kashgar prefecture of Xinjiang," *Chinese Journal of Disease Control and Prevention*, vol. 7, no. 1, pp. 72–76, 1992.
[22] L. Guan, J. J. Chai, and X. P. Zuo, "Advances in biology of Xinjiang sandfly," *Chinese Journal of Disease Control and Prevention*, vol. 14, no. 4, pp. 87–91, 1999.
[23] Z. B. Zhou, D. G. Gu, L. R. Guan et al., "Changes in the composition ratio of main species of Kashgar prefecture in Xinjiang and its relationship with visceral leishmaniasis," *International Journal of Medical Parasitic Diseases*, vol. 42, no. 2, 2015.
[24] A. A. Pereira Filho, M. da Conceição Abreu Bandeira, R. S. Fonteles et al., "An ecological study of sand flies (diptera: psychodidae) in the vicinity of Lençóis Maranhenses National park, Maranhão, Brazil," *Parasites &Vectors*, vol. 8, no. 1, pp. 8–442, 2015.
[25] China News Network, http://www.sohu.com/a/122899918_123753.
[26] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *Journal of Theoretical Biology*, vol. 254, no. 1, pp. 178–196, 2008.

Hindawi
Complexity
Volume 2020, Article ID 8428269, 15 pages
https://doi.org/10.1155/2020/8428269

Research Article

Periodic Solution of a Neutral Delay Leslie Predator-Prey Model and the Effect of Random Perturbation on the Smith Growth Model

Tongtong Li and Wencai Zhao

College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Wencai Zhao; zhaowencai@sdust.edu.cn

Received 18 October 2019; Revised 20 February 2020; Accepted 19 March 2020; Published 14 April 2020

Guest Editor: George V. Popescu

This paper puts forward a class of ratio-dependent Leslie predator-prey models. Firstly, a neutral delay predator-prey model with ratio dependence and impulse control is established and the existence of positive periodic solutions is proved by the coincidence degree theory. Secondly, a stochastic disturbance Leslie model of Smith growth is obtained when the interference of white noise is taken into consideration and the impact of delay is ignored. Applying Itô's formula, we get the conditions of system persistence and extinction. Finally we verify the correctness of theoretical analysis with numerical simulations.

1. Introduction

In the population dynamic systems, the relationship between population growth rate and population density is complex and diverse. The Verhulst–Pearl logistic equation [1]

$$\dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{K}\right), \tag{1}$$

is widely used to describe population growth for its concise mathematical expression and clear biological meaning, where $x(t)$ denotes the population density at time t, r is the intrinsic growth rate, and K is the environmental capacity. However, in 1963, Smith [2] studied the population dynamics of Daphnia magna and found that the growth of Daphnia magna population in the laboratory did not satisfy the above growth model, and he proposed the following growth equation:

$$\dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{K} - \frac{\dot{x}(t)}{cK}\right), \tag{2}$$

which is usually called the Smith growth equation, where c is a positive constant. Then, in 1991, Kuang [3] considered a neutral predator-prey model with time delays and the Smith growth as follows:

$$\begin{cases} \dot{x}(t) = rx(t)\left(1 - \dfrac{x(t-\tau) + \rho\dot{x}(t-\tau)}{K}\right) - y(t)p(x(t)), \\ \dot{y}(t) = y(t)(-\alpha + \beta p(x(t-\sigma))), \end{cases} \tag{3}$$

where τ and σ are the recovery time delay of the prey population and the digestion time delay of the predator population, respectively, α is the death rate of the predator, and β is the conversion factor of prey into the predator. $(\rho\dot{x}(t-\tau)/K)$ is the neutral delay term. $p(x)$ indicates the functional response of predator y to prey x and depends only on the density of the prey population. However, when the predators have to search for food, there will be competition among individual predators for limited prey. The amount of food the predator obtains depends not only on the density of the prey population but also on the density of the predator itself, which is influenced by the relative ratio of the number of predators and prey populations, reflecting the ratio-dependent characteristics. Leslie proposed the

following Leslie model [4, 5] in studying the relationship between the predator and prey:

$$\begin{cases} \dot{x}(t) = x(t)\left(r_1 - a_1 x(t) - c_1 y(t)\right), \\ \dot{y}(t) = y(t)\left(r_2 - a_2 \dfrac{y(t)}{x(t)}\right), \end{cases} \tag{4}$$

where $x(t)$ and $y(t)$ stand for prey and predator densities at time t, respectively. r_1 and r_2 are the intrinsic growth rates of the prey and predator, a_1 denotes the density-dependent coefficient of the prey, the parameter c_1 is the capturing rate, and a_2 is the conversion factor of prey into predators. Motivated by the works [3–5], we propose a ratio-dependent neutral delay Leslie predator-prey model as follows:

$$\begin{cases} \dot{x}(t) = rx(t)\left(1 - \dfrac{x(t-\tau_1)}{K} - \varepsilon \dot{x}(t-\tau_1)\right) - \dfrac{ex(t)y(t)}{ny(t) + x(t)}, \\ \dot{y}(t) = ay(t)\left(1 - b\dfrac{y(t-\tau_2)}{x(t-\tau_2)}\right), \end{cases} \tag{5}$$

where all of the parameters are positive and have the biological meanings listed in Table 1.

The first equation of the system states that the growth rate of prey depends not only on its own density but also on the neutral term. If there is no predator, the prey population will grow in Smith mode. The Smith model mainly describes the growth rates of the population under the condition of limited food. The relative growth rate of population size $(\dot{x}(t)/x(t))$ at time t is proportional to the amount of food left before the time τ_1. The food consumed by the population is mainly used for two purposes: the food needed to maintain the organism's own survival $(x(t-\tau_1)/K)$ and the food needed for population reproduction $\varepsilon\dot{x}(t-\tau_1)$. The denominator of functional response function $p = (ex/(ny + x))$ contains y, which reflects the interaction among individuals within the predator population. The second equation implies that the predator population grow in Leslie mode and the environmental capacity is proportional to prey population density.

Table 1: Biological meanings of parameters for model (5).

Parameters	Description
τ_1, τ_2	The discrete delays
r, a	The intrinsic growth rates of prey and predator
K	The environmental capacity of the prey
ε	The neutral coefficient of the prey
e	The conversion rate
n	The capture saturation rate
b	The conversion factor of prey into predator

Developing and utilizing biological populations in a proper way is an important issue facing humanity, which can make biological resources work for human beings and ensure the sustainability of ecosystems [6–12]. People's harvesting behavior is often not continuous, which may be once or several times at a certain time. For example, in fishery production, fishermen harvest fishes in few months and in agricultural production farmers spray pesticides at regular intervals to harvest (kill) pests. This harvesting method makes the population quantity or density be changed drastically in a short period of time, so it is more suitable to describe with impulsive differential equations. In the past 20 years, impulsive differential equations have been widely used in various biological models, for example, Zhao and Tang [13], Liu et al. [14], Li et al. [15], and Zhang et al. [16] introduced impulsive effects into the ordinary differential equation model. Li and Meng [17] and Qi et al. [18] established a class of impulsive stochastic differential equation models based on ordinary differential equations. Du and Feng [19] and Chen and Du [20] cooperated the impulsive control to the neutral predator-prey model, by using the coincidence degree theory, and they obtained the conditions for the existence of periodic solution. Then, based on [19, 20] and assuming that some parameters in system (5) are changed periodically and the population is pulse controlled, we will obtain the following neutral delay system with pulse effects:

$$\begin{cases} \left.\begin{aligned} \dot{x}(t) &= x(t)\left(r(t) - \dfrac{r(t)x(t-\tau_1(t))}{K(t)} - \rho(t)\dot{x}(t-\tau_1(t)) - \dfrac{e(t)y(t)}{ny(t) + x(t)}\right), \\ \dot{y}(t) &= a(t)y(t)\left(1 - b(t)\dfrac{y(t-\tau_2(t))}{x(t-\tau_2(t))}\right), \end{aligned}\right\} \quad t \neq t_k,\ k = 1, 2, \ldots, \\ \left.\begin{aligned} \Delta x(t_k) &= -\theta_{1k}x(t_k), \\ \Delta y(t_k) &= -\theta_{2k}y(t_k), \end{aligned}\right\} \quad t = t_k,\ k = 1, 2, \ldots, \end{cases} \tag{6}$$

with initial conditions

$$\begin{aligned} x(t) &= p(t),\ \ p \in C^1([-\tau, 0], [0, +\infty)),\ p(0) > 0, \\ y(t) &= q(t),\ \ q \in C^1([-\tau, 0], [0, +\infty)),\ q(0) > 0, \end{aligned} \tag{7}$$

where $\Delta x(t_k) = x(t_k^+) - x(t_k)$, $\Delta y(t_k) = y(t_k^+) - y(t_k)$, t_k is the pulse harvesting time, and $\theta_{ik}(i = 1, 2)$ is the pulse harvesting ratio. This implies that the harvest yield is in proportion to the biomass at that time. $r(t), K(t)$, $\tau_1(t), \tau_2(t), e(t), a(t), b(t)$, and $\rho(t)$ are continuous nonnegative T-periodic functions. n is positive constant and $\tau := \max_{t\in[0,T]}\{\tau_1(t), \tau_2(t)\}$.

On the contrary, biological populations multiply and thrive in nature, which will inevitably be affected by various environmental noise, so it is important to consider the effects of random disturbances on population dynamics [21–29]. In [30], the asymptotic stability of a stochastic May mutualism system was studied. In [31], Markov semigroup was used to study a stochastic ecological model of plants with infectious diseases. Authors in [32–34] focused on the long-term dynamic behavior of ecosystems affected by environmental noise. In particular, conditions for the existence of stationary distribution were obtained. Zhao et al. [35] presented an algal population growth model, where the authors not only gave the threshold conditions of permanence and extinction but also discussed the influence of environmental noise on the periodic blooms. Ignoring the influence of time delays and considering the disturbance of white noise, we construct the following stochastic Leslie predator-prey model with Smith growth:

$$\begin{cases} \mathrm{d}x(t) = \left[\dfrac{rx(t)}{1+\rho x(t)}\left(1 - \dfrac{x(t)}{K}\right) - \dfrac{ex(t)y(t)}{(ny(t)+x(t))(1+\rho x(t))}\right]\mathrm{d}t + \sigma_1 x(t) dB_1(t), \\ \mathrm{d}y(t) = \left[ay(t)\left(1 - \dfrac{by(t)}{x(t)}\right)\right]\mathrm{d}t + \sigma_2 y(t) dB_2(t), \end{cases} \tag{8}$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motions. σ_1^2 and σ_2^2 denote the intensity of white noise. $r, K, \rho, n, a, b, e, \sigma_1$, and σ_2 are all positive and the biological significances are the same as that in model (5).

To study the existence of periodic solutions for ordinary differential equations and delay differential equations, Gaines and Mawhin's continuation theorem based on coincidence degree theory [36] is an important tool. However, it is seldom applied in the study of the existence of positive periodic solutions of neutral delay system with pulse effects. It is generally considered difficult to prove that the nonlinear operator H is compact on set $\overline{\Omega}$ and the solutions of the equation have prior bound. In this work, by using analytical methods such as the mean value theorem of integral and inequality techniques, we obtain the sufficient conditions for the existence of positive periodic solutions of system (6). In addition, considering the influence of environmental disturbance on the population, we include the white noise into the Leslie predator-prey system, establish a new stochastic Smith growth model (8), and discuss the conditions for the persistence and extinction of the system.

2. Preliminaries

First, we convert the impulse system (6) into a nonimpulsive form. In system (6), we assume that

$[I_1]\ 0 < t_1 < t_2 < \cdots < t_k < \cdots$ are fixed points and $\lim_{k\longrightarrow\infty} t_k = +\infty$

$[I_2]\ \{\theta_{ik}\}$ are real sequences such that $\theta_{ik} < 1$ and $\prod_{0<t_k<t}(1 - \theta_{ik})$ $(i = 1, 2)$ are T-periodic functions

Assuming $[I_1]$ and $[I_2]$, we convert system (6) to the following system:

$$\begin{cases} \dot{D}_1(t) = D_1(t)\left[r(t) - A(t)D_1(t - \tau_1(t)) - \eta(t)\dot{D}_1(t - \tau_1(t)) - \dfrac{E(t)D_2(t)}{N(t)D_2(t) + \theta_1(t)D_1(t)}\right], \\ \dot{D}_2(t) = D_2(t)\left[a(t) - \dfrac{B(t)D_2((t - \tau_2(t))}{\theta_2(t)D_1(t - \tau_2(t))}\right], \end{cases} \tag{9}$$

with initial conditions

$$\begin{aligned} D_1(t) &= p(t),\quad p \in C^1([-\tau, 0], [0, +\infty)),\ p(0) > 0, \\ D_2(t) &= q(t),\quad q \in C^1([-\tau, 0], [0, +\infty)),\ q(0) > 0, \end{aligned} \tag{10}$$

where

$$D_1(t) = \prod_{0<t_k<t} (1-\theta_{1k})^{-1} x(t),$$

$$D_2(t) = \prod_{0<t_k<t} (1-\theta_{2k})^{-1} y(t),$$

$$A(t) = \frac{r(t)}{K(t)} \prod_{0<t_k<t-\tau_1(t)} (1-\theta_{1k}),$$

$$\eta(t) = \rho(t) \prod_{0<t_k<t-\tau_1(t)} (1-\theta_{1k}),$$

$$E(t) = e(t) \prod_{0<t_k<t} (1-\theta_{2k}),$$

$$N(t) = n \prod_{0<t_k<t} (1-\theta_{2k}),$$

$$\theta_1(t) = \prod_{0<t_k<t} (1-\theta_{1k}),$$

$$\theta_2(t) = \prod_{0<t_k<t-\tau_2(t)} (1-\theta_{1k}),$$

$$B(t) = a(t)b(t) \prod_{0<t_k<t-\tau_2(t)} (1-\theta_{2k}). \tag{11}$$

Lemma 1. *Suppose that* $[I_1]$ *and* $[I_2]$ *hold, then*

(1) If $(D_1(t), D_2(t))^T$ *is a solution of (9), then* $(x(t), y(t))^T$ *is a solution of (6)*

(2) If $(x(t), y(t))^T$ *is a solution of (6), then* $(D_1(t), D_2(t))^T$ *is a solution of (9)*

Proof. (1) Suppose $(D_1(t), D_2(t))^T$ is a solution of system (9). Next, we prove that $x(t) = \prod_{0<t_k<t}(1-\theta_{1k}) D_1(t)$, $y(t) = \prod_{0<t_k<t}(1-\theta_{2k}) D_2(t)$ is the solution of system (6). When $t \neq t_k$, we obtain

$$\begin{aligned}
&x'(t) - x(t)\left(r(t) - \frac{r(t)x(t-\tau_1(t))}{K(t)} - \rho(t)x'(t-\tau_1(t)) - \frac{e(t)y(t)}{ny(t)+x(t)}\right) \\
&= \prod_{0<t_k<t} (1-\theta_{1k}) D_1'(t) - \prod_{0<t_k<t} (1-\theta_{1k}) D_1(t)\left(r(t) - \frac{r(t)\prod_{0<t_k<t-\tau_1(t)}(1-\theta_{1k}) D_1(t-\tau_1(t))}{K(t)}\right. \\
&\quad \left. - \rho(t) \prod_{0<t_k<t-\tau_1(t)} (1-\theta_{1k}) D_1'(t-\tau_1(t)) - \frac{e(t)\prod_{0<t_k<t}(1-\theta_{2k}) D_2(t)}{n\prod_{0<t_k<t}(1-\theta_{2k}) D_2(t) + \prod_{0<t_k<t}(1-\theta_{1k}) D_1(t)}\right) \\
&= \prod_{0<t_k<t} (1-\theta_{1k}) \left[D_1'(t) - D_1(t)\left(r(t) - A(t)D_1(t-\tau(t)) - \eta(t)D_1'(t-\tau_1(t)) - \frac{E(t)D_2(t)}{N(t)D_2(t) + \theta_1(t)D_1(t)}\right)\right] \\
&= 0,
\end{aligned} \tag{12}$$

$$y'(t) - y(t)\left(a(t) - \frac{a(t)b(t)y(t-\tau_2(t))}{x(t-\tau_2(t))}\right) = \prod_{0<t_k<t} (1-\theta_{2k}) \left[D_2'(t) - D_2(t)\left(a(t) - \frac{B(t)D_2(t-\tau_2(t))}{\theta_2(t)D_1(t-\tau_2(t))}\right)\right]$$

$$= 0.$$

When $t = t_k$, one has

$$\begin{aligned}
x(t_k^+) - x(t_k) &= \prod_{0<t_j\le t_k} (1-\theta_{1j}) D_1(t_k) - \prod_{0<t_j<t_k} (1-\theta_{1j}) D_1(t_k) \\
&= \prod_{0<t_j<t_k} (1-\theta_{1j}) D_1(t_k)(-\theta_{1k}) \\
&= -\theta_{1k} x(t_k),
\end{aligned}$$

$$\begin{aligned}
y(t_k^+) - y(t_k) &= \prod_{0<t_j\le t_k} (1-\theta_{2j}) D_2(t_k) \\
&\quad - \prod_{0<t_j<t_k} (1-\theta_{2j}) D_2(t_k) \\
&= \prod_{0<t_j<t_k} (1-\theta_{2j}) D_2(t_k)(-\theta_{2k}) \\
&= -\theta_{2k} y(t_k).
\end{aligned} \tag{13}$$

Therefore, $(x(t), y(t))^T$ is the solution of system (6).

(2) Similar to the proof of case (1), we could show that if $(x(t), y(t))^T$ is a solution of (6), then $(D_1(t), D_2(t))^T$ is the solution of system (9). The proof of Lemma 1 is completed.

For convenience, throughout this paper we will use the symbols:

$$\overline{v} = \frac{1}{T}\int_0^T v(t)\mathrm{d}t,$$

$$\widehat{v} = \min_{t\in[0,T]} v(t), \tag{14}$$

$$\widehat{v} = \max_{t\in[0,T]} v(t),$$

where $v = v(t)$ is a periodic continuous function with period T. □

3. Existence of Positive Periodic Solution of System (6)

Before discussing the existence of periodic solution of system (6), we introduce the coincidence degree theory firstly.

Let U and V be two Banach spaces, G: Dom$G \subset U \longrightarrow V$ is a linear mapping and H: $U \longrightarrow V$ is a continuous mapping. If dim Ker$G$ = codim Im$G < +\infty$ and Im$G \subset V$ is closed, then we call the operator G is a Fredholm operator with index zero. If G is a Fredholm operator with index zero and there exist continuous projections P: $U \longrightarrow U$ and Q: $V \longrightarrow V$ such that ImP = Ker$G$ and ImG = ker$Q$ = Im$(I - Q)$; then, $G|_{\mathrm{Dom}G\cap \mathrm{Ker}P}$: $(I - P)U \longrightarrow$ ImG has an inverse function, and we set it as K_p. Assume that $\Omega \subset U$ is any open set, if $QH(\overline{\Omega})$ is bounded and $K_p(I - Q)H$: $\overline{\Omega} \longrightarrow U$ is relative compact, and then we say H is G-compact on $\overline{\Omega}$.

Lemma 2 (see [36]). *Let G:Dom $G \subset U \longrightarrow V$ be a Fredholm mapping with index zero and H is a G-compact on $\overline{\Omega}$. Furthermore, assume that*

(i) $Gx \neq \lambda Hx$, where $\lambda \in (0,1)$, $x \in \partial\Omega \cap DomG$

(ii) $QHx \neq 0$, for each $x \in \partial\Omega \cap \ker G$

(iii) $deg\{JQH, \Omega \cap Ker\, G, 0\} \neq 0$, where J: Im$Q \longrightarrow Ker\, G$ is an isomorphism

Then, the operator equation $Gx = Hx$ has at least one solution in Dom$G \cap \overline{\Omega}$.

Lemma 3 (see [37]). *If $\tau \in C^1(R,R)$ with $\tau(t+T) \equiv \tau(t)$ and $\tau'(t) < 1$ for $t \in [0,T]$, then function $\beta(t) = t - \tau(t)$ has a unique inverse $\beta^{-1}(t)$ satisfying $\beta^{-1} \in C^1(R,R)$ with $\beta^{-1}(s+T) \equiv \beta^{-1}(s) + T$ for $s \in [0,T]$.*

We are now in a position to state and prove our main result.

Theorem 1. *Assume that the following conditions hold:*

$[H_1]$ $\tau_1'(t) < 1$, $\tau_2'(t) < 1$ for any $t \in R$

$[H_2]$ $\mu'(t) < A(t)$ for any $t \in [0,T]$, where $\mu(t) = (\eta(t)/(1 - \tau_1'(t)))$

$[H_3]$ $\breve{\eta}e^W < 1$, $\overline{r} > \overline{M}$, where W and M are defined in the proof

$[H_4]$ The algebraic equations

$$\overline{r} - \overline{A}x - \frac{1}{T}\int_0^T \frac{E(t)y}{N(t)y + \theta_1(t)x}\mathrm{d}t = 0,$$
$$\overline{a} - \overline{\left(\frac{B}{\theta_2}\right)}\frac{y}{x} = 0, \tag{15}$$

have finite solutions. Then, system (6) has at least one T-periodic solution.

In the following proof, we first transform (9) into equivalent system. Then, we construct the mappings and open set Ω on Banach space and prove that the mappings satisfy Lemma 2 on $\overline{\Omega}$. Therefore, we obtain Theorem 1.

Proof

(1) Transform (9) into an equivalent system.

Let $D_1(t) = e^{w(t)}$ and $D_2(t) = e^{v(t)}$. Then, we can translate system (9) into

$$\begin{cases} w'(t) = r(t) - A(t)e^{w(t-\tau_1(t))} - \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))} - \dfrac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}, \\ v'(t) = a(t) - \dfrac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}}, \end{cases} \tag{16}$$

where all coefficient functions are defined in system (9). It is easy to see that if system (16) has one T-periodic solution $(w^*(t), v^*(t))^T$, then $(D_1^*(t), D_2^*(t))^T = (\exp\{w^*(t)\}, \exp\{v^*(t)\})^T$ is a positive T-periodic solution of system (9). So, system (6) has a T-periodic solution $(x^*(t), y^*(t))^T = (\prod_{0<t_k<t}(1-\theta_{1k})D_1^*(t), \prod_{0<t_k<t}(1-\theta_{2k})D_2^*(t))^T$. Therefore, we just have to prove system (16)s has one T-periodic solution.

(2) Construct Banach space and define mappings.

Take $U = V = \{u(t) = (w(t), v(t))^T \in C^1(R, R^2)$: $u(t+T) = u(t)\}$ and denote $|u|_\infty = \max_{t\in[0,T]}(|w(t)| + |v(t)|)$, $\|u\| = |u|_\infty + |u'|_\infty$. Then, U and V are both Banach spaces with the norm $\|\cdot\|$ and $|\cdot|_\infty$, respectively. Define operators G, P, Q, and H in the following form, respectively, G: $U \longrightarrow V$, $Gu(t) = (w'(t), v'(t))^T$; $P(u) = \overline{u}$, for $u \in U$; $Q(u) = \overline{u}$, for $u \in V$; and

$$Hu(t) = \begin{pmatrix} r(t) - A(t)e^{w(t-\tau_1(t))} - \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))} - \dfrac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}} \\ a(t) - \dfrac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}} \end{pmatrix}. \tag{17}$$

Then, $\mathrm{Ker}G = R^2$, $\mathrm{Im}G = \left\{u \in V: \int_0^T u(t)\mathrm{d}t = 0\right\}$ is closed in V, $\dim\mathrm{Ker}G = \mathrm{codim}\mathrm{Im}G$. P and Q are continuous projectors such that $\mathrm{Im}P = \mathrm{Ker}G$, $\mathrm{Ker}\,Q = \mathrm{Im}G = \mathrm{Im}\,(I-Q)$. Therefore, the Fredholm mapping G has a unique inverse. The generalized inverse(to G)K_p:$\mathrm{Im}G \longrightarrow \mathrm{Ker}P \cap \mathrm{Dom}G$ is given by the following form:

$$K_p(u) = \int_0^t u(s)\mathrm{d}s - \frac{1}{T}\int_0^T \int_0^t u(s)\mathrm{d}s\,\mathrm{d}t. \tag{18}$$

Then, QH: $U \longrightarrow V$ and $K_p(I-Q)H$: $U \longrightarrow U$ read

$$QHu(t) = \begin{pmatrix} \dfrac{1}{T}\displaystyle\int_0^T \left(r(t) - A(t)e^{w(t-\tau_1(t))} - \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))} - \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}\right)\mathrm{d}t \\ \dfrac{1}{T}\displaystyle\int_0^T \left(a(t) - \frac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}}\right)\mathrm{d}t \end{pmatrix}, \tag{19}$$

$$K_p(I-Q)Hu(t) = A - B - C,$$

where

$$A = \begin{pmatrix} \displaystyle\int_0^t \left(r(s) - A(s)e^{w(s-\tau_1(s))} - \eta(s)w'(s-\tau_1(s))e^{(ws-\tau_1(s))} - \frac{E(s)e^{v(s)}}{N(s)e^{v(s)} + \theta_1(t)e^{w(s)}}\right)\mathrm{d}s \\ \displaystyle\int_0^t \left(a(s) - \frac{B(s)e^{v(s-\tau_2(s))}}{\theta_2(t)e^{w(s-\tau_2(s))}}\right)\mathrm{d}s \end{pmatrix},$$

$$B = \begin{pmatrix} \dfrac{1}{T}\displaystyle\int_0^T\int_0^t \left(r(s) - A(s)e^{w(s-\tau_1(s))} - \eta(s)w(s-\tau_1(s))e^{w(s-\tau_1(s))} - \frac{E(s)e^{v(s)}}{N(s)e^{v(s)} + \theta_1(t)e^{w(s)}}\right)\mathrm{d}s\,\mathrm{d}t \\ \dfrac{1}{T}\displaystyle\int_0^T\int_0^t \left(a(s) - \frac{B(s)e^{v(s-\tau_2(s))}}{\theta_2(t)e^{w(s-\tau_2(s))}}\right)\mathrm{d}s\,\mathrm{d}t \end{pmatrix}, \tag{20}$$

$$C = \begin{pmatrix} \left(\dfrac{t}{T} - \dfrac{1}{2}\right)\displaystyle\int_0^T \left(r(s) - A(s)e^{w(s-\tau_1(s))} - \eta(s)w'(s-\tau_1(s))e^{w(s-\tau_1(s))} - \frac{E(s)e^{v(s)}}{N(s)e^{v(s)} + \theta_1(t)e^{w(s)}}\right)\mathrm{d}s \\ \left(\dfrac{t}{T} - \dfrac{1}{2}\right)\displaystyle\int_0^T \left(a(s) - \frac{B(s)e^{v(s-\tau_2(s))}}{\theta_2(t)e^{w(s-\tau_2(s))}}\right)\mathrm{d}s \end{pmatrix}.$$

Distinctly, it is easy to know that QH and $K_p(I-Q)H$ are both continuous by the Lebesgue theorem. And by using Arzela–Ascoli theorem, we know that $K_p(I-Q)H(\overline{\Omega})$ is compact for any open bounded set $\Omega \subset U$. Hence, H is G-compact on $\overline{\Omega}$.

(3) Construct an open set Ω.

For the sake of using Lemma 2, we need to look for an appropriate open-bounded subset $\Omega \subset U$. First, we prove that the periodic solution of equation $Gu = \lambda Hu$ is bounded. Corresponding to the operator equation $Gu = \lambda Hu$ for $\lambda \in (0,1)$, we have

$$\begin{cases} w'(t) = \lambda\left(r(t) - A(t)e^{w(t-\tau_1(t))} - \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))} - \dfrac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}\right), \\ v'(t) = \lambda\left(a(t) - \dfrac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}}\right). \end{cases} \tag{21}$$

Suppose that $u = (w(t), v(t))^T \in U$ is a solution of (21). Now we prove that $\|u\| \le S^*$, S^* is a constant. □

Step 1. Prove that $|w(t)| \le S_1$ and $|w'(t)| \le S_2$.

Integrating (21) over the interval $[0,T]$, we obtain

$$\overline{r}T = \int_0^T \left(A(t)e^{w(t-\tau_1(t))} + \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))} + \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}\right)dt, \tag{22}$$

and

$$\overline{a}T = \int_0^T \left(\frac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}}\right)dt. \tag{23}$$

In consideration of $[H_1]$ and $[H_2]$, we obtain

$$\begin{aligned} \int_0^T \eta(t)w'(t-\tau_1(t))e^{w(t-\tau_1(t))}dt &= \int_0^T \frac{\eta(t)}{1-\tau_1'(t)}\left(e^{w(t-\tau_1(t))}\right)dt \\ &= \int_0^T \frac{\eta(t)}{1-\tau_1'(t)}de^{w(t-\tau_1(t))} \\ &= -\int_0^T \mu\prime(t)e^{w(t-\tau_1(t))}dt, \end{aligned} \tag{24}$$

which together with (22) gives

$$\overline{r}T = \int_0^T \left((A(t) - \mu\prime(t))e^{w(t-\tau_1(t))} + \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}\right)dt. \tag{25}$$

In view of Lemma 3 and $[H_1]$, we obtain

$$\begin{aligned} &\int_0^T A(t) - \mu\prime(t)e^{w(t-\tau_1(t))}dt \\ &= \int_0^T \frac{\left(A\left(\beta_1^{-1}(t)\right) - \mu\prime\left(\beta_1^{-1}(t)\right)\right)e^{w(t)}}{1-\tau_1'\left(\beta_1^{-1}(t)\right)}dt, \end{aligned} \tag{26}$$

where $\beta_1(t) = t - \tau_1(t)$. From (25), we obtain

$$\begin{aligned} \overline{r}T &= \int_0^T \frac{\left(A\left(\beta_1^{-1}(t)\right) - \mu\prime\left(\beta_1^{-1}(t)\right)\right)e^{w(t)}}{1-\tau_1'\left(\beta_1^{-1}(t)\right)}dt \\ &\quad + \int_0^T \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}dt, \end{aligned} \tag{27}$$

which implies

$$\overline{r}T \ge \int_0^T \frac{\left(A\left(\beta_1^{-1}(t)\right) - \mu\prime\left(\beta_1^{-1}(t)\right)\right)e^{w(t)}}{1-\tau_1'\left(\beta_1^{-1}(t)\right)}dt. \tag{28}$$

Therefore,

$$\frac{1}{T}\int_0^T e^{w(t)}dt \le \frac{\overline{r}}{\widehat{R}}, \tag{29}$$

where $R(t) = ((A(\beta_1^{-1}(t)) - \mu\prime(\beta_1^{-1}(t)))/(1-\tau_1'(\beta_1^{-1}(t))))$. According to (25), we obtain

$$\overline{r}T \le \int_0^T A(t) - \mu\prime(t)e^{w(t-\tau_1(t))}dt + \int_0^T \frac{E(t)}{N(t)}dt, \tag{30}$$

together with $[H_3]$, which implies

$$\frac{1}{T}\int_0^T e^{w(t)}dt \ge \frac{\overline{r} - \overline{M}}{\check{R}}, \tag{31}$$

where $M(t) = (E(t)/N(t))$.

According to (29) and (31), we can get that there exists $\varphi_1 \in [0,T]$ such that

$$|w(\varphi_1)| \le \max\left\{\left|\ln\frac{\overline{r}}{\widehat{R}}\right|, \left|\ln\frac{\overline{r} - \overline{M}}{\check{R}}\right|\right\} := H_1. \tag{32}$$

By means of (25) and (29), we have

$$\int_0^T \left[(A(t) - \mu\prime(t))e^{w(t-\tau_1(t))} + R(t)e^{w(t)}\right]dt \le 2\overline{r}T. \tag{33}$$

On the basis of the mean value theorem of integral calculus, we know that there exists $\xi \in [0,T]$, such that

$$(A(\xi) - \mu\prime(\xi))e^{w(\xi-\tau_1(\xi))} + R(\xi)e^{w(\xi)} \le 2\overline{r}. \tag{34}$$

Since $\overline{r} > 0$, we have

$$w(\xi) \le \ln \frac{2\overline{r}}{\widehat{R}}, \qquad e^{w(\xi-\tau_1(\xi))} \le \frac{2\overline{r}}{\widehat{U}}, \tag{35}$$

where $U = A(t) - \mu\prime(t)$.

Through the first equation of (21) and (25), we obtain

$$\begin{aligned}\int_0^T \left|\frac{\mathrm{d}}{\mathrm{d}t}\left(w(t) + \lambda\mu(t)e^{w(t-\tau_1(t))}\right)\right|\mathrm{d}t &= \lambda\int_0^T \left|r(t) - \left(A(t) - \mu'(t)\right)e^{w(t-\tau_1(t))} - \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1(t)e^{w(t)}}\right|\mathrm{d}t \\ &\le 2\overline{r}T.\end{aligned} \tag{36}$$

Therefore,

$$\begin{aligned}w(t) + \lambda\mu(t)e^{w(t-\tau_1(t))} &\le w(\xi) + \lambda\mu(\xi)e^{w(\xi-\tau_1(\xi))} + \int_0^T \left|\frac{\mathrm{d}}{\mathrm{d}t}\left(w(t) + \lambda\mu(t)e^{w(t-\tau_1(t))}\right)\right|\mathrm{d}t \\ &\le \ln\frac{2\overline{r}}{\widehat{R}} + \frac{2\overline{r}\check{\mu}}{\widehat{U}} + 2\overline{r}T \coloneqq W.\end{aligned} \tag{37}$$

So, we obtain

$$w(t) \le W. \tag{38}$$

From (36), we know

$$\int_0^T |w(t)|\mathrm{d}t - \int_0^T \left|\frac{\mathrm{d}}{\mathrm{d}t}\lambda\mu(t)e^{w(t-\tau_1(t))}\right|\mathrm{d}t \le 2\overline{r}T. \tag{39}$$

Consequently,

$$\begin{aligned}\int_0^T \left|w'(t)\right|\mathrm{d}t &\le 2\overline{r}T + \int_0^T \left|\left(\mu'(t) + \eta(t)w'(t-\tau_1(t))\right)e^{w(t-\tau_1((t))}\right|\mathrm{d}t \\ &\le 2\overline{r}T + \left|\check{w}'\right|e^W T + \check{\eta}e^W\int_0^T \left|\mu'(t)\right|\mathrm{d}t.\end{aligned} \tag{40}$$

Through the transpose, we can obtain

$$\int_0^T \left|w'(t)\right|\mathrm{d}t \le \frac{2\overline{r}T + |\check{\mu\prime}|e^W T}{1 - \check{\eta}e^W} \coloneqq K. \tag{41}$$

Together with (32), we obtain

$$|w(t)| \le \left|w(\varphi_1)\right| + \int_0^T \left|w'(t)\right|\mathrm{d}t \le H_1 + K \coloneqq S_1. \tag{42}$$

Through the first equation of (21) and (38), we obtain

$$\left|w'(t)\right| = \left|\lambda\left(r(t) - A(t)e^{w(t-\tau_1(t))} - \eta(t)w'(t)e^{w(t-\tau_1(t))} - \frac{E(t)e^{v(t)}}{N(t)e^{v(t)} + \theta_1 e^{w(t)}}\right)\right|, \tag{43}$$

so we obtain

$$\max\left|w'(t)\right| \le \check{r} + \check{A}e^W + \check{\eta}e^W \max\left|w'(t)\right| + \frac{\check{E}}{\widehat{N}}. \tag{44}$$

Utilizing $[H_3]$, we have

$$\left|w'(t)\right| \le \frac{1}{1 - \check{\eta}e^W}\left(\check{r} + \check{A}e^W + \frac{\check{E}}{\widehat{N}}\right) \coloneqq S_2. \tag{45}$$

Step 2. Prove that $|v(t)| \le S_3$ and $|v'(t)| \le S_4$.

By means of (23) and (38), we obtain

$$\begin{aligned}\overline{a}T &= \int_0^T \frac{B(t)e^{v(t-\tau_2(t))}}{\theta_2(t)e^{w(t-\tau_2(t))}}\mathrm{d}t \\ &\ge \frac{\widehat{B}}{\check{\theta}_2}\int_0^T \frac{e^{v(t)}}{\left(1 - \tau_2'\left(\beta_2^{-1}(t)\right)\right)e^{w(t)}}\mathrm{d}t \ge \frac{\widehat{B}}{\check{\theta}_2\check{\sigma}e^W}\int_0^T e^{v(t)}\mathrm{d}t,\end{aligned} \tag{46}$$

where $\beta_2(t) = t - \tau_2(t)$ and $\sigma(t) = 1 - \tau_2'(\beta_2^{-1}(t))$, which implies that

$$\frac{1}{T}\int_0^T e^{v(t)}\mathrm{d}t \le \frac{\overline{a}\check{\theta}_2\check{\sigma}e^W}{\widehat{B}}. \tag{47}$$

According to (23) and (42), we obtain

$$\begin{aligned}\overline{a}T &= \int_0^T \frac{B(t)e^{v(t-\tau_2(t))}}{\theta_2 e^{w(t-\tau_2(t))}}\mathrm{d}t \\ &\le \frac{\check{B}}{\widehat{\theta}_2}\int_0^T \frac{e^{v(t)}}{\left(1 - \tau_2'\left(\beta_2^{-1}(t)\right)\right)e^{w(t)}}\mathrm{d}t \\ &\le \frac{\check{B}}{\widehat{\theta}_2\widehat{\sigma}e^{-S_1}}\int_0^T e^{v(t)}\mathrm{d}t.\end{aligned} \tag{48}$$

Thus,

$$\frac{1}{T}\int_0^T e^{v(t)}\mathrm{d}t \ge \frac{\overline{a}\widehat{\theta}_2\widehat{\sigma}e^{-S_1}}{\check{B}}. \tag{49}$$

From (47) and (49), there exists $\varphi_2 \in [0, T]$ such that

$$\left|v(\varphi_2)\right| \le \max\left\{\left|\ln\frac{\overline{a}\check{\theta}_2\check{\sigma}e^W}{\widehat{B}}\right|, \left|\ln\frac{\overline{a}\widehat{\theta}_2\widehat{\sigma}e^{-S_1}}{\check{B}}\right|\right\} \coloneqq H_2. \tag{50}$$

From (21) and (23), it is easy to see that

$$\int_0^T \left|v'(t)\right|\mathrm{d}t \le 2\overline{a}T. \tag{51}$$

Then,

$$|v(t)| \le |v(\varphi_2)| + \int_0^T v'(t)\mathrm{d}t \le H_2 + 2\overline{a}T := S_3. \tag{52}$$

Through the second equation of (21), (42), and (52), we have

$$\begin{aligned} |v'(t)| &= \left|\lambda\left(a(t) - \frac{B(t)e^{v\left(t-\tau_2(t)\right)}}{\theta_2 e^{w\left(t-\tau_2(t)\right)}}\right)\right| \\ &\le \check{a} + \frac{\check{B}e^{S_3}}{\widehat{\theta}_2 e^{-S_1}} := S_4. \end{aligned} \tag{53}$$

From (42), (45), (52), and (53), we obtain

$$\|u\| = |u|_\infty + |u'|_\infty \le S_1 + S_2 + S_3 + S_4 := S^*. \tag{54}$$

At this point, we have proved that the periodic solution of equation $Gu = \lambda Hu$ for $\lambda \in (0,1)$ is bounded.

In the following part, we will prove that the constant solution of equation $QHu = 0$ is bounded. According to condition $[H_4]$, the algebraic equation set

$$\begin{cases} \overline{r} - \overline{A}x - \dfrac{1}{T}\displaystyle\int_0^T \frac{E(t)y}{N(t)y + \theta_1(t)x}\mathrm{d}t = 0, \\ \overline{a} - \overline{\left(\dfrac{B}{\theta_2}\right)}\dfrac{y}{x} = 0, \end{cases} \tag{55}$$

has finite positive solutions $(x_i^*, y_i^*)(i = 1, 2, \ldots k)$. Therefore, for $u \in \mathrm{Ker}G$, the operator equation $QHu = 0$ has finite positive solutions $u_i^* = (w_i^*, v_i^*)(i = 1, 2, \ldots k)$. Suppose that $S_0 = \max_{1\le i\le k}\{|w_i^*| + |v_i^*|\}$, then $\|u_i^*\| \le S_0 (i = 1, 2, \ldots k)$. Therefore, the solution of equation $QHu = 0$ for $u \in \mathrm{Ker}G$ is bounded.

Let $S = S^* + S_0$, and take $\Omega = \{u = (w, v)^T : \|u\| < S\}$. It is easy to see that Ω verifies condition (i) in Lemma 2. When $u \in \partial\Omega \cap \mathrm{Ker}G = \partial\Omega \cap R^2$, u is a constant vector in R^2 with $\|u\| = S$, we obtain

$$QHu(t) = \begin{pmatrix} \overline{r} - \overline{A}x - \dfrac{1}{T}\displaystyle\int_0^T \frac{E(t)y}{N(t)y + \theta_1(t)x}\mathrm{d}t \\ \overline{a} - \overline{\left(\dfrac{B}{\theta_2}\right)}\dfrac{y}{x} \end{pmatrix} \neq 0. \tag{56}$$

Thus, condition (ii) of Lemma 2 holds.

(4) Calculate the degree of coincidence.

For each $u = (w, v)^T \in \mathrm{Im}Q$, define isomorphism $J{:}\mathrm{Im}Q \longrightarrow \ker G$ so that $J(u) = u$. By simply computing, we get the following Jacobian matrix:

$$\overline{J}(w, v) = \begin{pmatrix} -\overline{A}e^w + \dfrac{1}{T}\displaystyle\int_0^T \frac{\theta_1(t)E(t)e^{v+w}}{\left(N(t)e^v + \theta_1(t)e^w\right)^2}\mathrm{d}t & \dfrac{1}{T}\displaystyle\int_0^T \frac{-E(t)\theta_1(t)e^{w+v}}{\left(N(t)e^v + \theta_1(t)e^w\right)^2}\mathrm{d}t \\ \overline{\left(\dfrac{B}{\theta_2}\right)}\dfrac{e^v}{e^w} & -\overline{\left(\dfrac{B}{\theta_2}\right)}\dfrac{e^v}{e^w} \end{pmatrix}. \tag{57}$$

Then, we have

$$\begin{aligned} \det\overline{J}(w, v) &= \left(\overline{A}e^w - \frac{1}{T}\int_0^T \frac{\theta_1(t)E(t)e^{v+w}}{\left(N(t)e^v + \theta_1(t)e^w\right)^2}\mathrm{d}t\right)\left(\overline{\frac{B}{\theta_2}}\right)\frac{e^v}{e^w} \\ &\quad + \left(\overline{\frac{B}{\theta_2}}\right)\frac{e^v}{e^w}\cdot\frac{1}{T}\int_0^T \frac{E(t)\theta_1(t)e^{w+v}}{\left(N(t)e^v + \theta_1(t)e^w\right)^2}\mathrm{d}t \\ &= \left(\overline{\frac{B}{\theta_2}}\right)\overline{A}e^v > 0, \quad \forall (w, v) \in R^2, \end{aligned} \tag{58}$$

which implies

$$\deg\{JQH, \Omega \cap \mathrm{Ker}G, 0\} = \sum_{(w_i^*, v_i^*)\in QH^{-1}(0)} sgn\ \det\overline{J}, \quad (w_i^*, v_i^*) > 0. \tag{59}$$

By now, we have verified that Ω satisfies all the requirements in Lemma 2. Thus, system (16) has at least one T-periodic solution, which implies that system (6) has at least one positive T-periodic solution. This completes the proof.

Remark 1. If the delay $\tau_1(t)$ and $\tau_2(t)$ both are positive constants, that is to say, condition $[H_1]$ holds. At this time, the delay has no effect on the existence of periodic solutions.

Remark 2. If the neutral coefficient $\rho(t) = 0$, that is, $\eta(t) = 0$, at this time, condition $[H_2]$ holds. When $[H_1]$, $[H_3]$, and $[H_4]$ hold, then the system has at least one positive periodic solution.

Next, we will study the predator-prey model with random disturbances.

4. Persistence in Mean and Extinction of System (8)

In this section, we discuss the properties of stochastic system (8) with Smith growth, the sufficient conditions for persistence and extinction are obtained.

Lemma 4 (see [38]). *Let $(x(t), y(t))$ be a solution of system (8) with initial value $(x(0), y(0))$, then $\lim_{t\longrightarrow\infty}(x(t)/t) = 0$, $\lim_{t\longrightarrow\infty}(y(t)/t) = 0$, $\lim_{t\longrightarrow\infty}(\int_0^t x(s)dB_1(s)/t) = 0$, $\lim_{t\longrightarrow\infty}(\int_0^t y(s)dB_2(s)/t) = 0$ a.s.*

Theorem 2. *Let $(x(t), y(t))$ be a solution of system (8) with initial value $(x(0), y(0))$. If*

(1) $r - (\sigma_1^2/2) - (e/n) > 0$, $a - (\sigma_2^2/2) > 0$, then system (8) is persistent in mean

(2) $r - (\sigma_1^2/2) < 0$, $a - (\sigma_2^2/2) < 0$, then system (8) goes to extinction almost surely

Proof

(1) From the first equation of system (8) and using Itô's formula, we obtain

$$\begin{aligned} d(\ln x + \rho x) &= \left(\left(\frac{1}{x} + \rho\right)\left(\frac{rx}{1+\rho x}\left(1 - \frac{x}{K}\right) - \frac{exy}{(ny+x)(1+\rho x)}\right) - \frac{\sigma_1^2}{2}\right)\mathrm{d}t + \sigma_1(1+\rho x)\mathrm{d}B_1(t) \\ &= \left(r\left(1 - \frac{x}{K}\right) - \frac{ey}{ny+x} - \frac{\sigma_1^2}{2}\right)\mathrm{d}t + \sigma_1(1+\rho x)\mathrm{d}B_1(t). \end{aligned} \tag{60}$$

Integrating from 0 to t on both sides, we can obtain

$$\begin{aligned} \ln x(t) + \rho x(t) - \ln x(0) - \rho x(0) = &\left(r - \frac{\sigma_1^2}{2}\right)t - \frac{r}{K}\int_0^t x(s)\mathrm{d}s \\ &- \int_0^t \frac{ey(s)}{ny(s)+x(s)}\mathrm{d}s \\ &+ \int_0^t \sigma_1(1+\rho x(s))\mathrm{d}B_1(s). \end{aligned} \tag{61}$$

Then,

$$\begin{aligned} \ln x(t) = &\left(r - \frac{\sigma_1^2}{2}\right)t - \int_0^t \frac{ey(s)}{ny(s)+x(s)}\mathrm{d}s - \frac{r}{K}\int_0^t x(s)\mathrm{d}s \\ &- \rho x(t) + \ln x(0) + \rho x(0) + M(t), \end{aligned} \tag{62}$$

where $M(t) = \int_0^t \sigma_1(1+\rho x(s))\mathrm{d}B_1(s)$. Dividing t on both sides, we have that

$$\begin{aligned} \frac{1}{t}\ln x(t) \geq &\left(r - \frac{\sigma_1^2}{2} - \frac{e}{n}\right) - \frac{r}{Kt}\int_0^t x(s)\mathrm{d}s \\ &+ \frac{M(t)}{t} - \frac{\rho x(t)}{t} + \frac{\ln x(0)}{t} + \frac{\rho x(0)}{t}. \end{aligned} \tag{63}$$

According to Lemma 4, we obtain

$$\liminf_{t\longrightarrow\infty} \frac{\int_0^t x(t)\mathrm{d}t}{t} \geq \frac{(r - (\sigma_1^2/2) - (e/n))K}{r} > 0. \tag{64}$$

From the second equation of system (8) and using Itô's formula, we can obtain

$$\mathrm{d}(\ln y) = \left(a - \frac{\sigma_2^2}{2} - \frac{aby}{x}\right)\mathrm{d}t + \sigma_2\mathrm{d}B_2(t). \tag{65}$$

Integrating from 0 to t on both sides, we obtain

$$\ln y(t) - \ln y(0) = \left(a - \frac{\sigma_2^2}{2}\right)t - ab\int_0^t \frac{y(s)}{x(s)}\mathrm{d}s + \int_0^t \sigma_2\mathrm{d}B_2(s). \tag{66}$$

Thus,

$$ab\int_0^t \frac{y(s)}{x(s)}\mathrm{d}s = -\ln y(t) + \ln y(0) + \left(a - \frac{\sigma_2^2}{2}\right)t + \int_0^t \sigma_2\mathrm{d}B_2(s). \tag{67}$$

Then, we can obtain

$$\liminf_{t\longrightarrow\infty} \frac{1}{t}\int_0^t \frac{y(s)}{x(s)}\mathrm{d}s \geq \frac{a - (\sigma_2^2/2)}{ab} > 0. \tag{68}$$

Because x is persistent, $\liminf_{t\longrightarrow\infty}(1/t)\int_0^t y(s)\mathrm{d}s > 0$. We complete the proof of (1).

(2) By (61) and (66), we can obtain

$$\begin{aligned} \ln x(t) + \rho x(t) - \ln x(0) - \rho x(0) &= \left(r - \frac{\sigma_1^2}{2}\right)t - \frac{r}{K}\int_0^t x(s)\mathrm{d}s - \int_0^t \frac{ey(s)}{ny(s)+x(s)}\mathrm{d}s + \int_0^t \sigma_1(1+\rho x(s))\mathrm{d}B_1(s) \\ &\leq \left(r - \frac{\sigma_1^2}{2}\right)t + M(t), \\ \ln y(t) - \ln y(0) &= \left(a - \frac{\sigma_2^2}{2}\right)t - ab\int_0^t \frac{y(s)}{x(s)}\mathrm{d}s + \int_0^t \sigma_2\mathrm{d}B_2(s) \\ &\leq \left(a - \frac{\sigma_2^2}{2}\right)t + \int_0^t \sigma_2\mathrm{d}B_2(s). \end{aligned} \tag{69}$$

Therefore,

$$\limsup_{t\longrightarrow\infty}\frac{\ln x(t)}{t}\leq r-\frac{\sigma_1^2}{2}<0,$$
$$\limsup_{t\longrightarrow\infty}\frac{\ln y(t)}{t}\leq a-\frac{\sigma_2^2}{2}<0. \tag{70}$$

So, system (8) goes to extinction almost surely.

Theorem 2 shows that if the intensity of environmental noise is small, the prey and predator will survive for a long time, and vice versa, and they will become extinct. Therefore, environmental noise is not conducive to the survival of the population. □

5. Conclusion and Numerical Simulation

This paper mainly solves the following problems:

(1) For system (6) with neutral time delay and impulse control, first we transform it into nonpulse equivalent system by transformation: $x(t)=\prod_{0<t_k<t}(1-\theta_{1k})D_1(t)$ and $y(t)=\prod_{0<t_k<t}(1-\theta_{2k})D_2(t)$. Then, by using the coincidence degree theory, the sufficient conditions $[H_1]-[H_4]$ for the existence of periodic solution of the system is obtained.

(2) For the stochastic perturbation system (8) with Smith growth, the persistence and extinction of the system are discussed.

In the following part, we introduce numerical simulations to illustrate our main theoretical results. Firstly, we verify the existence of positive periodic solution of the neutral delay Leslie predator-prey model (16).

Example 1

(1) In system (16), we choose

$$\begin{aligned}
r(t)&=4+\sin(10\pi t),\\
A(t)&=0.4-\cos(10\pi t),\\
E(t)&=0.2-0.5\cos(10\pi t),\\
a(t)&=3+\sin(10\pi t),\\
B(t)&=0.5+\cos(10\pi t),\\
\eta(t)&=0.1+0.01\sin(10\pi t),\\
N(t)&=4+\sin(10\pi t),\\
\theta_1(t)&=0.8+\sin(10\pi t),\\
\theta_2(t)&=0.5+\sin(10\pi t),\\
\tau_1(t)&=0.2+0.01\cos(10\pi t),\\
\tau_2(t)&=0.2+0.01\sin(10\pi t).
\end{aligned} \tag{71}$$

By simple calculation, we know that conditions $[H_1]$ and $[H_2]$ hold, and $\check{\eta}e^W\approx 0.0987<1$, $\overline{r}=4>\overline{M}\approx 0.175$. Obviously $[H_3]$ holds. Moreover, the algebraic equations in $[H_4]$ have only one positive solution, which together with Theorem 1 yields that Example 1 has at least one periodic solution (see Figure 1).

(2) We choose $\tau_1(t)=0.3$ and $\tau_2(t)=0.3$ in system (16), and other parameters are the same as those in Figure 1.

In this case, condition $[H_1]$ is true because $\tau_1(t)$ and $\tau_2(t)$ are constants. Thus, there is at least one periodic solution as long as condition $[H_2-H_4]$ are true (see Figure 2). This shows that the existence of periodic solutions is not affected by the constant delays.

(3) We choose $\eta(t)=0$ in system (16), and other parameters are the same as those in Figure 1.

Because the neutral coefficient $\rho(t)=0$, condition $[H_2]$ must be true. The system has at least one periodic solution as long as other conditions are true. In this case, the prey population grows in logistic mode, which indicates that there is at least one periodic solution in general impulsive system (see Figure 3).

Example 2. Next we conduct simulations about the persistence and extinction of the stochastic Smith growth model (8). Obviously, when $r>(e/n+b)$, $a+((a\rho K/r)+1)(r-(e/(n+b)))>(be/(n+b)^2)$, the deterministic model corresponding to model (8) has a unique locally asymptotically stable positive equilibrium $P^*=(x^*,y^*)=((K/r)(r-(e/(n+b))),(K/rb)(r-(e/(n+b))))$. In system (8), we choose

$$\begin{aligned}
r&=0.4,\\
a&=0.2,\\
e&=0.4,\\
n&=2,\\
\rho&=0.1,\\
b&=0.5,\\
K&=0.8.
\end{aligned} \tag{72}$$

(1) Let $\sigma_1=0.1$, $\sigma_2=0.1$, then we can know that $r>(e/(n+b))$, $a+((a\rho K/r)+1)(r-(e/(n+b)))>(be/(n+b)^2)$, $r-(\sigma_1^2/2)-(e/n)>0$, $a-(\sigma_2^2/2)>0$. Both stochastic system and deterministic system exist for a long time. Thus, conclusion (1) of Theorem 2 holds (see Figure 4).

(2) Let $\sigma_1=0.9$ and $\sigma_2=0.8$. By simple calculation, we can know that $r>(e/(n+b))$, $a+((a\rho K/r)+1)(r-(e/(n+b)))>(be/(n+b)^2)$, and then the deterministic system is persistent. However, $r-(\sigma_1^2/2)<0$, $a-(\sigma_2^2/2)<0$, and then the random system is extinct. Thus, conclusion (2) of Theorem 2 holds. This shows that environmental noise is harmful to the long-term survival of the population (see Figure 5).

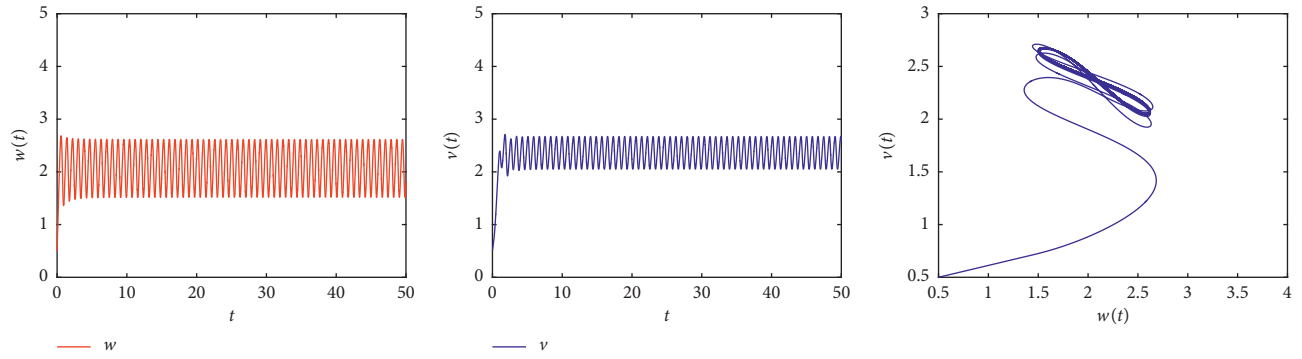


FIGURE 1: Computer simulation of the paths $w(t)$ and $v(t)$ for system (16) with the initial value $(w(0), v(0)) = (0.5, 0.5)$.

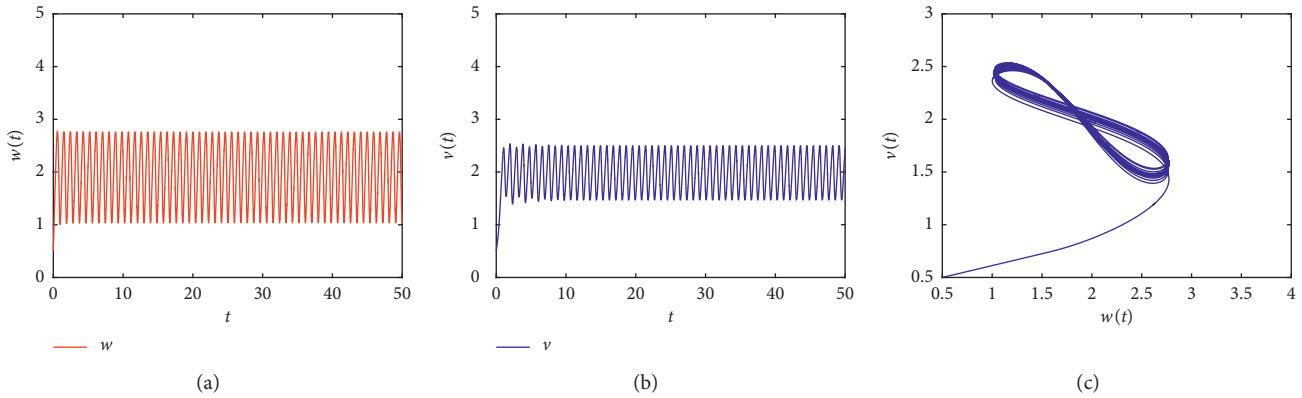


FIGURE 2: Computer simulation of the paths $w(t)$ and $v(t)$ for system (16) with the initial value $(w(0), v(0)) = (0.5, 0.5)$ and $\tau_1(t) = \tau_2(t) = 0.3$

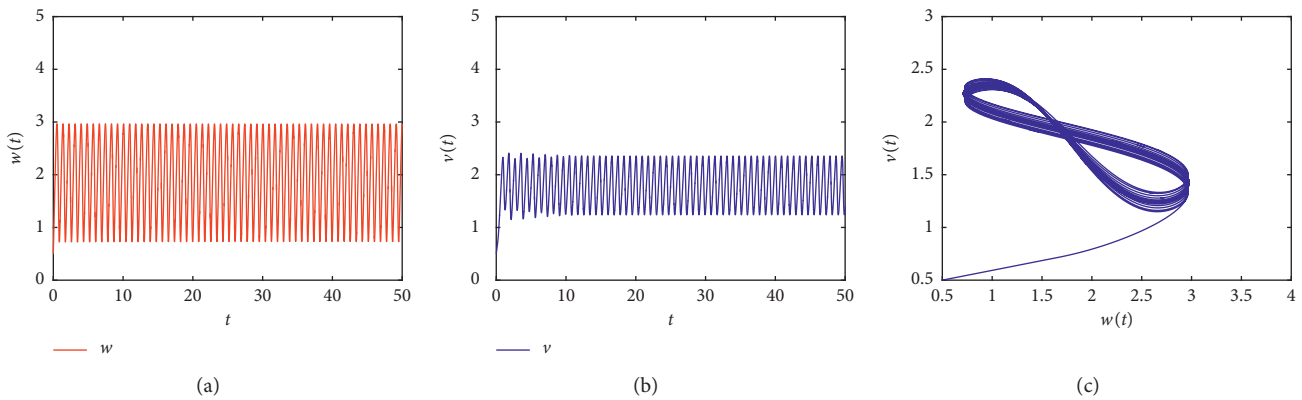


FIGURE 3: Computer simulation of the paths $w(t)$ and $v(t)$ for system (16) with the initial value $(w(0), v(0)) = (0.5, 0.5)$ and $\eta(t) = 0$.

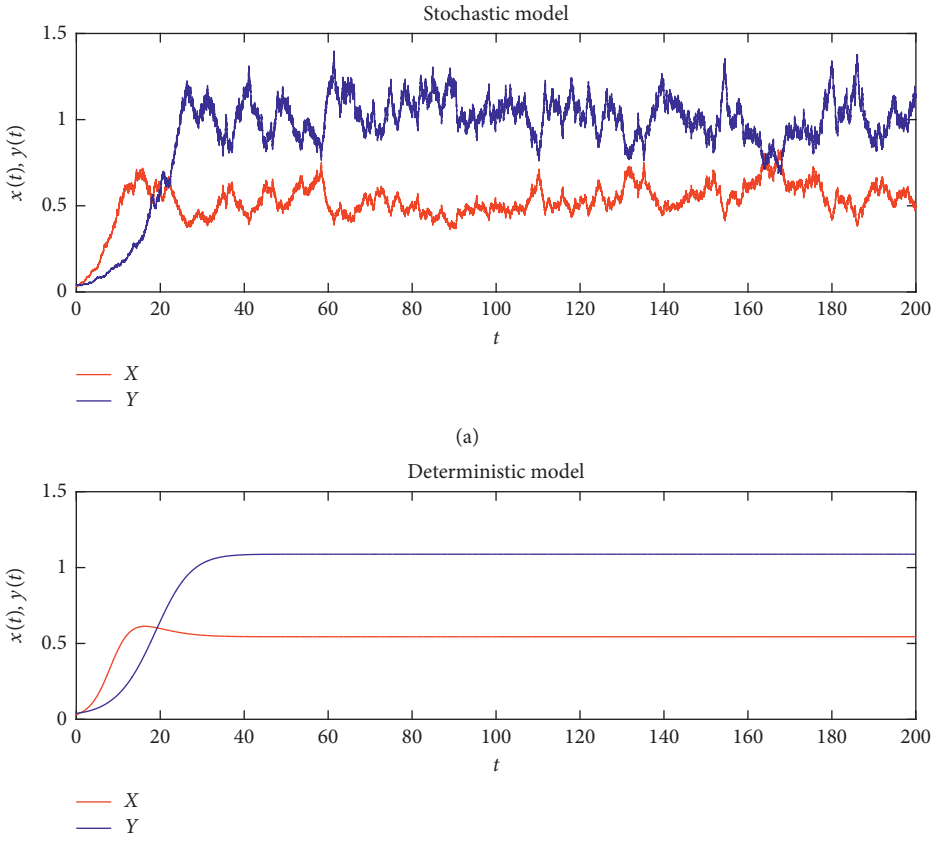


Figure 4: Computer simulations of the paths $x(t)$ and $y(t)$ for system (8) with the initial value $(x(0), y(0)) = (0.03, 0.04)$ and $\sigma_1 = 0.1$, $\sigma_2 = 0.1$, and its corresponding deterministic model with initial value $(x(0), y(0)) = (0.03, 0.04)$.

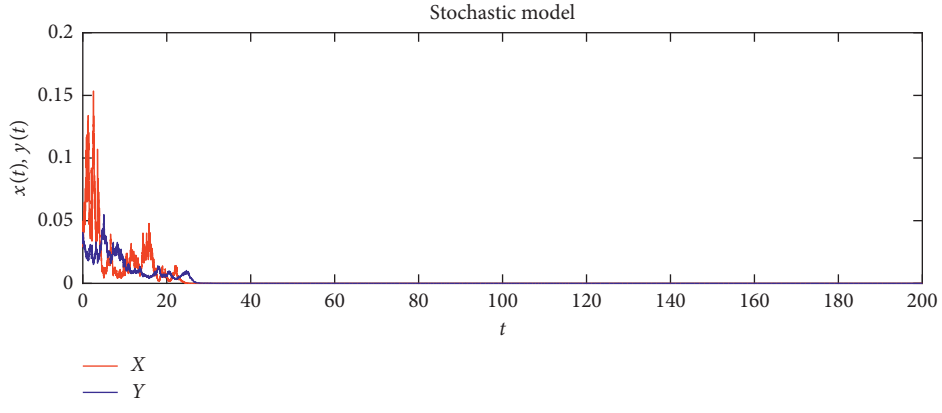
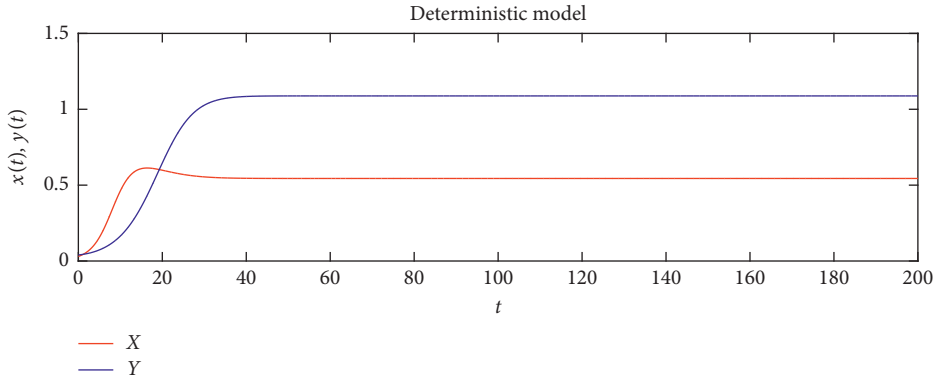


Figure 5: Continued.



(b)

Figure 5: Computer simulations of the paths $x(t)$ and $y(t)$ for system (8) with the initial value $(x(0), y(0)) = (0.03, 0.04)$ and $\sigma_1 = 0.9$, $\sigma_2 = 0.8$, and its corresponding deterministic model with initial value $(x(0), y(0)) = (0.03, 0.04)$.

Data Availability

Data sharing is not applicable to this article as all datasets are hypothetical during the current study.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

All authors contributed equally in writing this paper. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 11371230) and Shandong Provincial Natural Science Foundation of China (no. ZR2019MA003).

References

[1] R. Pearl and L. J. Reed, "On the rate of growth of the population of the united states since 1790 and its mathematical representation," *Proceedings of the National Academy of Sciences*, vol. 6, no. 6, pp. 275–288, 1920.

[2] F. E. Smith, "Population dynamics in Daphnia magna and a new model for population growth," *Ecology*, vol. 44, no. 4, pp. 651–663, 1963.

[3] Y. Kuang, "On neutral delay logistic Gause-type predator-prey systems," *Dynamics and Stability of Systems*, vol. 6, no. 2, pp. 173–189, 1991.

[4] P. H. Leslie, "Some further notes on the use of matrices in population mathematics," *Biometrika*, vol. 35, no. 3-4, pp. 213–245, 1948.

[5] P. H. Leslie and J. C. Gower, "The properties of a stochastic model for the predator-prey type of interaction between two species," *Biometrika*, vol. 47, no. 3/4, pp. 219–234, 1960.

[6] C.-q. Xu, S.-l. Yuan, and T.-h. Zhang, "Confidence domain in the stochastic competition chemostat model with feedback control," *Applied Mathematics—A Journal of Chinese Universities*, vol. 33, no. 4, pp. 379–389, 2018.

[7] T. Zhang, T. Zhang, and X. Meng, "Stability analysis of a chemostat model with maintenance energy," *Applied Mathematics Letters*, vol. 68, pp. 1–7, 2017.

[8] K. Liu, T. Zhang, and L. Chen, "State-dependent pulse vaccination and therapeutic strategy in an SI epidemic model with nonlinear incidence rate," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 3859815, 10 pages, 2019.

[9] T. Ma, X. Meng, and Z. Chang, "Dynamics and optimal harvesting control for a stochastic one-predator-two-prey time delay system with jumps," *Complexity*, vol. 2019, Article ID 5342031, 19 pages, 2019.

[10] M. Chi and W. Zhao, "Dynamical analysis of two-microorganism and single nutrient stochastic chemostat model with monod-haldane response function," *Complexity*, vol. 2019, Article ID 8719067, 13 pages, 2019.

[11] D. Jia, T. Zhang, and S. Yuan, "Pattern dynamics of a diffusive toxin producing phytoplankton-zooplankton model with three-dimensional patch," *International Journal of Bifurcation and Chaos*, vol. 29, no. 4, Article ID 1930011, 2019.

[12] Z. Jiang, X. Bi, T. Zhang et al., "Global hopf bifurcation of a delayed phytoplankton-zooplankton system considering toxin producing effect and delay dependent coefficient," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 3807–3829, 2019.

[13] T. Zhao and S. Tang, "Impulsive harvesting and by-catch mortality for the theta logistic model," *Applied Mathematics and Computation*, vol. 217, no. 22, pp. 9412–9423, 2011.

[14] B. Liu, Y. Zhang, and L. Chen, "The dynamical behaviors of a Lotka-Volterra predator-prey model concerning integrated pest management," *Nonlinear Analysis: Real World Applications*, vol. 6, no. 2, pp. 227–243, 2005.

[15] Y. Li, J. Cui, and X. Song, "Asymptotic behaviour of the non-autonomous competing two-species Lotka-Volterra models with impulsive effect," *Journal of Biological Dynamics*, vol. 3, no. 1, pp. 58–72, 2009.

[16] T. Zhang, N. Gao, T. Wang, and H. Liu, "Global dynamics of a model for treating microorganisms in sewage by periodically adding microbial flocculants," *Mathematical Biosciences and Engineering*, vol. 17, no. 1, pp. 179–201, 2020.

[17] Y. Li and X. Meng, "Dynamics of an impulsive stochastic nonautonomous chemostat model with two different growth rates in a polluted environment," *Discrete Dynamics in Nature and Society*, vol. 2019, Article ID 5498569, 15 pages, 2019.

[18] H. Qi, X. Meng, and F. Tao, "Dynamics analysis of a stochastic non-autonomous one-predator–two-prey system with beddington–deangelis functional response and impulsive perturbations," *Advances in Difference Equations*, vol. 2019, no. 1, p. 235, 2019.

[19] Z. Du and Z. Feng, "Periodic solutions of a neutral impulsive predator-prey model with Beddington-DeAngelis functional response with delays," *Journal of Computational and Applied Mathematics*, vol. 258, pp. 87–98, 2014.

[20] X. Chen and Z. Du, "Existence of positive periodic solutions for a neutral delay predator-prey model with hassell-varley type functional response and impulse," *Qualitative Theory of Dynamical Systems*, vol. 17, no. 1, pp. 67–80, 2018.

[21] X. Yu and S. Yuan, "Asymptotic properties of a stochastic chemostat model with two distributed delays and nonlinear perturbation," *Discrete Continuous Dynamical Systems-B*, vol. 22, no. 11, 2017.

[22] M. Liu and K. Wang, "Global stability of a nonlinear stochastic predator-prey system with Beddington-DeAngelis functional response," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 3, pp. 1114–1121, 2011.

[23] N. Gao, Y. Song, X. Wang, and J. Liu, "Dynamics of a stochastic SIS epidemic model with nonlinear incidence rates," *Advances in Difference Equations*, vol. 2019, no. 1, p. 41, 2019.

[24] Q. Liu, D. Jiang, T. Hayat, and B. Ahmad, "Stationary distribution and extinction of a stochastic predator-prey model with additional food and nonlinear perturbation," *Applied Mathematics and Computation*, vol. 320, pp. 226–239, 2018.

[25] T. Feng, Z. Qiu, X. Meng, and L. Rong, "Analysis of a stochastic HIV-1 infection model with degenerate diffusion," *Applied Mathematics and Computation*, vol. 348, pp. 437–455, 2019.

[26] W. Zhao, J. Liu, M. Chi, and F. Bian, "Dynamics analysis of stochastic epidemic models with standard incidence," *Advances in Difference Equations*, vol. 2019, no. 1, p. 22, 2019.

[27] F. Tao and Z. Qiu, "Global analysis of a stochastic TB model with vaccination and treatment," *Discrete Continuous Dynamical Systems-B*, vol. 24, no. 6, p. 2923, 2018.

[28] H. Ren and W. Zhao, "Dynamics analysis of a stochastic Leslie–Gower predator-prey model with feedback controls," *Mathematical Problems in Engineering*, vol. 2019, Article ID 8631272, 13 pages, 2019.

[29] T. Zhang, J. Wang, Y. Li, Z. Jiang, and X. Han, "Dynamics analysis of a delayed virus model with two different transmission methods and treatments," *Advances in Difference Equations*, vol. 2020, no. 1, pp. 1–17, 2020.

[30] G. Liu, H. Qi, Z. Chang, and X. Meng, "Asymptotic stability of a stochastic May mutualism system," *Computers & Mathematics with Applications*, vol. 79, no. 3, pp. 735–745, 2020.

[31] H. Qi, X. Meng, and Z. Chang, "Markov semigroup approach to the analysis of a nonlinear stochastic plant disease model," *Electronic Journal of Differential Equations*, vol. 2019, no. 116, pp. 1–19, 2019.

[32] H. Zhang and T. Zhang, "The stationary distribution of a microorganism flocculation model with stochastic perturbation," *Applied Mathematics Letters*, vol. 103, Article ID 106217, 2020.

[33] X. Yu, S. Yuan, and T. Zhang, "Survival and ergodicity of a stochastic phytoplankton-zooplankton model with toxin-producing phytoplankton in an impulsive polluted environment," *Applied Mathematics and Computation*, vol. 347, pp. 249–264, 2019.

[34] X. Yu, S. Yuan, and T. Zhang, "Asymptotic properties of stochastic nutrient-plankton food chain models with nutrient recycling," *Nonlinear Analysis: Hybrid Systems*, vol. 34, pp. 209–225, 2019.

[35] S. Zhao, S. Yuan, and H. Wang, "Threshold behavior in a stochastic algal growth model with stoichiometric constraints and seasonal variation," *Journal of Differential Equations*, vol. 268, no. 9, pp. 5113–5139, 2020.

[36] R. E. Gaines and J. L. Mawhin, "Coincidence degree and nonlinear differential equations," *Lecture Notes in Mathematics*, Vol. 568, Springer, Berlin, Germany, 1977.

[37] S. Lu, "On the existence of positive periodic solutions to a Lotka Volterra cooperative population model with multiple delays," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 68, no. 6, pp. 1746–1753, 2008.

[38] X. Meng, F. Li, and S. Gao, "Global analysis and numerical simulations of a novel stochastic eco-epidemiological model with time delay," *Applied Mathematics and Computation*, vol. 339, pp. 701–726, 2018.

Hindawi
Complexity
Volume 2020, Article ID 6403641, 11 pages
https://doi.org/10.1155/2020/6403641

Research Article

Separable Nonlinear Least-Squares Parameter Estimation for Complex Dynamic Systems

Itai Dattner,[1] **Harold Ship**,[1] **and Eberhard O. Voit**[2]

[1]*Department of Statistics, University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, Haifa 3498838, Israel*
[2]*The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 950 Atlantic Drive, Atslanta, GA 30332–2000, USA*

Correspondence should be addressed to Itai Dattner; idattner@stat.haifa.ac.il

Received 9 August 2019; Revised 10 November 2019; Accepted 5 December 2019; Published 2 April 2020

Academic Editor: Honglei Xu

Nonlinear dynamic models are widely used for characterizing processes that govern complex biological pathway systems. Over the past decade, validation and further development of these models became possible due to data collected via high-throughput experiments using methods from molecular biology. While these data are very beneficial, they are typically incomplete and noisy, which renders the inference of parameter values for complex dynamic models challenging. Fortunately, many biological systems have embedded linear mathematical features, which may be exploited, thereby improving fits and leading to better convergence of optimization algorithms. In this paper, we explore options of inference for dynamic models using a novel method of *separable nonlinear least-squares optimization* and compare its performance to the traditional nonlinear least-squares method. The numerical results from extensive simulations suggest that the proposed approach is at least as accurate as the traditional nonlinear least-squares, but usually superior, while also enjoying a substantial reduction in computational time.

1. Introduction

Nonlinear dynamic models are widely used for characterizing the processes that govern complex biological pathway systems. Of particular interest in this context are so-called canonical formats, which are very flexible in their possible responses, yet involve a very restricted domain of functional forms. Outside linear systems, the best-known canonical formats are Lotka–Volterra (LV) models [1–4], which use binomial terms, and power-law systems within the framework of Biochemical Systems Theory (BST), which exclusively use products of power functions. BST was originally devised for the analysis of biochemical and gene regulatory systems, but has subsequently found much wider application in various biomedical and other areas [5, 6]. Whereas it is easy to set up an LV or BST model for a complex biological system in a symbolic format, the identification of optimal parameter values continues to be a significant challenge. As a consequence, estimating parameters of systems of ordinary differential equations (ODEs) remains to be an active research area that attracts contributions from a variety of scientific fields (e.g., [7–12]. Indeed, numerous optimization methods for ODE models have been proposed in recent years, but none works exceptionally well throughout a wide range of applications, with reasons spanning the entire spectrum from intrinsic problems with biological data (sparseness, uncertainties, noise, ...) to technical and computational issues (numerous local minima, unidentifiability, sloppiness, ...). Methods like slope-based estimation (e.g., [13]) and dynamic flux estimation [14–16] alleviate these problems but are not panacea.

Here, we revisit, and bring to fruition, early ideas [17] of separating estimation tasks into linear and nonlinear aspects. However, our main focus is *not* really a new estimation method *per se*. Instead, we are interested in a more general and higher-level point of view regarding parameter estimation than that typically presented in technical articles. Specifically, this article addresses parameter estimation for dynamic models whose mathematical format contains linear features that allow a natural separation of parameters and

system states. A trivial example is a linear ODE where the vector field $x'(t) = \theta x(t)$ is linear in the parameter θ, with $x'(t)$ denoting the derivative of $x(t)$ with respect to t. As a more interesting example, the ODE vector field may be partially linear in the parameters, as it is the case for so-called S-system models in BST [5].

Example 1. An S-system [18] is defined as

$$x_j'(t) = \alpha_j \prod_{k=1}^{d} x_k^{g_{jk}}(t) - \beta_j \prod_{k=1}^{d} x_k^{h_{jk}}(t), \quad j = 1, \ldots, d. \tag{1}$$

Here α_j, β_j are non-negative rate constants, while g_{jk}, h_{jk} are real-valued kinetic orders that reflect the strength and directionality of the effect that a variable has on a given influx or efflux. Informally, one can view this system as a regression equation, where the "covariates" are the variables $x_j(t)$ on the right-hand side, whereas the "response" variables are the derivatives $x_j'(t)$ on the left-hand side. Note that the vector field is linear in the rate constants α_j, β_j, but nonlinear in the kinetic orders g_{jk}, h_{jk}.

Estimation methods that exploit separability of parameters and system states in dynamic models have a long history; see [19] for a special case. However, a rigorous statistical analysis of such a method has been achieved only recently [20]. In a classical paper on the inference for dynamic models, Varah [17] mentioned in passing that "one can use the idea of separability or variable projection (see [21] or [22]), in which the linear parameters are implicitly solved for, the resulting (fully) nonlinear least-squares problem is solved for the nonlinear parameters, and then the linear parameters are obtained using their representation in terms of the nonlinear parameters. Since this reduces the size of the nonlinear least-squares problem to be solved, it is worthwhile." Somewhat surprisingly, given that parameter estimation for ODEs is commonly thought as a computational bottleneck in modeling dynamic processes, Varah's suggestion has not been widely followed in practice. In fact, in the vast literature dedicated to parameter fitting techniques for dynamic models, we are aware of only two relevant references: using a direct integral approach, Dattner et al. [23] applied a separable nonlinear least-squares technique to the inference of parameters in a predator-prey system acting in a heterogeneous environment, while Wu et al. [24] used separability to estimate parameters of high-dimensional linear ODE systems. Moreover, Varah's idea of exploiting separability for estimating ODE parameters has been implemented only recently in a publicly available software package [25]. Pertinent details of this software will be discussed in a later section.

The analysis in this paper is hoped to convince the reader that Varah's idea is indeed worth pursuing. To support this claim, we explore and compare two general data fitting approaches for dynamic models: the traditional nonlinear least-squares method (NLS) and the proposed separable nonlinear least-squares method (SLS). Through extensive Monte-Carlo simulations of representative complex models, we identify and quantify significant statistical and computational gains obtained with this separation method. We will ultimately come to the conclusion that model separability can be very beneficial and that the SLS approach should be considered for any complex dynamic system that possess significant linear features.

The paper is organized as follows. In Section 2, we present details of the SLS methodology in the context of dynamic models. Section 3 describes the simulation setup, quantifies the statistical measures we use in order to compare the performance of SLS and NLS, and presents numerical results. In Section 4, we point out future research directions, while conclusions are provided in Section 5.

2. Separable Nonlinear Least-Squares (SLS) and Varah's Idea

2.1. Generalities. Following Varah's original idea within the context of inference in dynamic models, the main advantages of exploiting separability for parameter estimation are the following [26]:

(i) Fewer initial guesses are required for optimization

(ii) The optimization problem is better conditioned

(iii) Convergence is faster

These advantages have been convincingly demonstrated in several publications. For example, see Mullen [27] for an implementation and applications in physics and chemistry; Chung & Nagy [28] for a high-dimensional case, where the number of parameters is substantially larger than the number of observations; Gan et al. [29] who compared the performance of several algorithms for SLS problems; and Erichson et al. [30] who studied sparse principal component analysis via variable projection. Separable models are of broad practical applicability, and as such form a subject of active theoretical and applied research. For instance, when analyzing the "reduced" nonlinear optimization problem of a separable structure, simplified conditions are required for establishing a variety of theoretical results concerning numerical and statistical properties of the resulting estimators, compared to the original NLS problem (e.g., [20, 31]).

In the following, we focus on complex dynamic models that are formulated as systems of ordinary differential equations (e.g., [32]). Specifically, consider a system of equations given by

$$\begin{cases} x'(t) = F(x(t); \theta), & t \in [0, T], \\ x(0) = \xi, \end{cases} \tag{2}$$

where $x(t)$ takes values in $\mathbb{R}^d$, $\xi \in \Xi \subset \mathbb{R}^d$, and $\theta \in \Theta \subset \mathbb{R}^p$. For our purposes, we explicitly separate linear components from nonlinear ones in the function F by setting

$$F(x(t); \theta) = g(x(t); \theta_{\mathrm{NL}})\theta_{\mathrm{L}}, \tag{3}$$

where $\theta = (\theta_{\mathrm{NL}}^\top, \theta_{\mathrm{L}}^\top)^\top$, and the symbol $\top$ stands for the matrix transpose (cf., [20]). Here θ_{NL}, a vector of size p_{NL}, stands for the "nonlinear" parameters that are not separable from the state variables x, while θ_{L}, a vector of size p_{L}, contains the "linear" parameters; note that $p = p_{\mathrm{L}} + p_{\mathrm{NL}}$. As the vector field in (3) is separable in the linear parameter

vector θ_L, we refer to the corresponding ODE system as *linear in the parameter* θ_L (cf. the case of a linear regression model), although the *solution* to the system might be highly nonlinear in θ, or even implicit.

Example 2. Let

$$\theta_{NL} = (g_{11}, \ldots, g_{1d}, \ldots, g_{d1}, \ldots, g_{dd}, h_{11}, \ldots, h_{1d}, \ldots, h_{d1}, \ldots, h_{dd})^\top,$$

$$\theta_L = (\alpha_1, \beta_1, \ldots, \alpha_d, \beta_d)^\top.$$

Then, one sees that equation (1) is a special case of (2)–(3).

2.2. Solution Strategy. Let $x(t;\theta,\xi)$, $t \in [0,T]$, be the solution of the initial value problem (2). We assume that noisy measurements $Y_j(t_i)$ on the system are collected at time points $t_i \in [0,T]$. A common statistical formulation of this situation is

$$Y_j(t_i) = x_j(t_i;\theta,\xi) + \varepsilon_{ij}, \quad i = 1,\ldots,n,\ j = 1,\ldots,d. \tag{4}$$

Here the random variables ε_{ij} are unobservable, independent measurement errors (not necessarily Gaussian) with zero mean and finite variance.

Varah's approach to parameter estimation in ODE models works as follows. Let $\widehat{x}(t)$ stand for a smoother of the data, obtained, e.g., using splines or local polynomials (see e.g., [33, 34] and [35] for a treatment of various smoothing methods and an extensive bibliography). This smoother approximates the solution $x(t;\theta,\xi)$ to the ODE (2). Varah suggests to insert the smoother into equation (2), which will now be satisfied only approximately, and to minimize the resulting discrepancy over the parameters ξ and θ. A minimizer $(\widehat{\xi},\widehat{\theta})$ is then an estimator of (ξ,θ). This idea was put on a solid statistical foundation in Brunel [36] and Gugushvili and Klaassen [37]. Varah's original approach requires the use of the derivative $\widehat{x}'(t)$ as an estimator of $x'(t)$, which is a disadvantage, as it is well known that estimating derivatives from noisy and sparse data may be rather inaccurate; see e.g., Vilela et al. [38] and Chou and Voit [7] or more generally Fan and Gijbels [33]. Recent research [20, 23, 39–46] has shown that it is more fruitful to transplant Varah's idea to the solution level of equation (2). To accomplish this shift, we define an integral criterion function

$$\int_0^T \left\| \widehat{x}(t) - \xi - \int_0^t F(\widehat{x}(s);\theta)\mathrm{d}s \right\|^2 \mathrm{d}t, \tag{5}$$

as it is typical in estimation approaches based on integrals (see references above). Here, $||\cdot||$ is the Euclidean norm. A minimizer of (5) over (ξ,θ) yields a parameter estimator that typically has slightly different features than an estimator based on the differential equations themselves. In practice, the integral is discretized and replaced by a sum, so that minimization can be performed using a typical nonlinear least-squares method, such as *fminsearch* in Matlab. The discretized format is

$$\left(\widehat{\xi}_{NLS},\widehat{\theta}_{NLS}\right) = \arg\min_{\xi,\theta} \int_o^T \left\| \widehat{x}(t) - \xi - \int_0^t F(\widehat{x}(s);\theta)\mathrm{d}s \right\|^2 \mathrm{d}t. \tag{6}$$

The NLS solution does not take into account the specific linear form of the ODEs in (3), but uses the general form in (2).

It is at this stage that Varah suggested to utilize separability, without actually investigating such an approach. Here, we provide the necessary details (cf. [20]). Denote

$$\widehat{G}(t) \coloneqq \widehat{G}(t;\theta_{NL}) = \int_0^t g(\widehat{x}(s);\theta_{NL})\mathrm{d}s, \quad t \in [0,T],$$

$$\widehat{A} = \int_0^T \widehat{G}(t)\mathrm{d}t,$$

$$\widehat{B} = \int_0^T \widehat{G}^\top(t)\widehat{G}(t)\mathrm{d}t.$$

Then, with θ_{NL} kept fixed, a minimizer of (5) is given by

$$\widehat{\xi}(\theta_{NL}) = (TI_d - \widehat{A}\widehat{B}^{-1}\widehat{A}^\top)^{-1} \int_0^T (I_d - \widehat{A}\widehat{B}^{-1}\widehat{G}^\top(t))\widehat{x}(t)\mathrm{d}t,$$

$$\widehat{\theta}_L(\theta_{NL}) = \widehat{B}^{-1}\int_0^T \widehat{G}^\top(t)(\widehat{x}(t) - \widehat{\xi})\mathrm{d}t,$$

where I_d denotes the $d \times d$ identity matrix. The notation $\widehat{\xi}(\theta_{NL})$ and $\widehat{\theta}_L(\theta_{NL})$ emphasizes the dependence of the solution on the nonlinear parameters θ_{NL}. This solution $(\widehat{\xi}(\theta_{NL}),\widehat{\theta}_L(\theta_{NL}))$ is plugged back into (5), yielding the reduced integral criterion function (cf. [23]):

$$M(\theta_{NL}) \coloneqq \int_0^T \left\| \widehat{x}(t) - \widehat{\xi}(\theta_{NL}) - \widehat{G}(t;\theta_{NL})\widehat{\theta}_L(\theta_{NL}) \right\|^2 \mathrm{d}t. \tag{7}$$

Once $M(\theta_{NL})$ is minimized over θ_{NL} and a solution

$$\widehat{\theta}_{NL} = \arg\min_{\theta_{NL}} M(\theta_{NL}),$$

is obtained, estimators for ξ and θ follow immediately and are given (with mild abuse of the matrix transpose notation) by

$$\widehat{\xi}_{SLS} = \widehat{\xi}\left(\widehat{\theta}_{NL}\right),$$
$$\widehat{\theta}_{SLS} = \left(\widehat{\theta}_{NL},\widehat{\theta}_L\left(\widehat{\theta}_{NL}\right)\right), \tag{8}$$

respectively. Equations (7) and (8) are driven by Varah's [17] suggestion discussed above. Indeed, note that the nonlinear optimization is applied only for estimating the nonlinear parameters θ_{NL}, which, in comparison to the NLS approach, can substantially reduce the dimension of the nonlinear optimization problem.

From the above derivation, it is clear that SLS problems constitute a special class of NLS problems, with linear and nonlinear objective functions for different sets of variables. While the idea of using separability for improving parameter

estimation was presented already in Lawton and Sylvestre [47], it seems that much of the subsequent literature is based on the variable projection method proposed by Golub and Pereyra [21]. Golub and Pereyra [26] reviewed 30 years of research into this problem.

3. Simulation Framework and Results

In order to investigate the relative performance of SLS and NLS, we designed and performed a large Monte-Carlo simulation, whose results are presented in this section.

All computations were carried out in *R* on an Amazon EC2 m5a.4xlarge instance using the *simode* package of Yaari and Dattner [25] (Separable Integral Matching for Ordinary Differential Equations). The statistical methodologies applied in the package use smoothing and minimization of an integral-matching criterion function, taking advantage of the mathematical structure of the differential equations like separability of parameters from equations. Application of smoothing and integral-based methods to parameter estimation of ordinary differential equations was shown to yield more accurate and stable results comparing to derivative based ones [20]. Here, we used default smoothing and optimization settings in *simode*, and in that respect, both SLS and NLS received equal treatment. Specifically, *simode* uses cross validation (see, e.g., [35]) to determine the optimal amount of smoothing. A detailed guide for using the package can be found in Yaari and Dattner [25]. The code to reproduce our numerical results can be accessed on GitHub (see https://github.com/haroldship/complexity-2019-code/tree/master/Final Code First Submission). For plotting, we relied on the *ggplot2* package in *R*, see Wickham [48].

3.1. Monte-Carlo Study Design. We chose several representative and challenging ODE models arising in a variety of scientific disciplines. Those were

(i) An SIR model for simulating the spread of an infectious disease

(ii) A Lotka–Volterra population model with sinusoidal seasonal adjustment

(iii) A Generalised Mass Action (GMA) system within BST, e.g., for metabolic pathway systems

(iv) A FitzHugh–Nagumo system of action potentials along neuronal axons

Further mathematical details on these systems and the specific experimental setups we used are given below.

In each case, we generated observations by numerically integrating the system and then adding independent Gaussian noise to the time courses, as in (4). We considered various parameter setups, sample sizes, and noise levels, as specified below. The ODE parameters were estimated via both NLS and SLS, as defined in equations (6) and (8), respectively.

As performance criteria, the time required to perform optimization and the accuracy of the resulting parameter estimates were used. While comparing computation times is trivial, numerous options are available for comparing accuracy. We focused on the main difference between the two optimization schemes, namely the way they deal with the estimation of linear parameters. SLS does not require initial guesses for these parameters. By contrast, NLS does require a good initial guess for each linear parameter; otherwise, it might diverge or get stuck in a local minimum. Thus, finding "good" solutions to nonlinear optimization problems often requires "good" initial guesses in the parameter space. Clearly, some "prior information" regarding these parameters is of crucial importance for optimization purposes. The key insight is that this prior information is encapsulated in the mathematical form of the ODEs themselves, such as (3). Importantly, while NLS does not take into account the special mathematical features of the ODEs and treats all the parameters in a uniform manner, this is not the case for SLS. Thus, one might *a priori* expect SLS to be more efficient and possibly more accurate than NLS, when prior information regarding the linear parameters is of low quality. On the other hand, when one has high-quality prior information regarding the linear parameters, we expect that SLS and NLS will perform similarly. One might note that the nonlinear parameters in almost all GMA and S-systems are very tightly bounded, usually between −1 and +2, and that their sign is often known, whereas the linear parameters are unbounded in GMA systems and nonnegative in S-systems, and nothing is known about their magnitudes (see Chapter 5 of [18]). Thus, not needing prior information on the linear parameters in SLS can be a tremendous advantage.

For the Monte-Carlo study, we varied the prior information by using high-, medium-, and low-quality initial guesses for the parameter values. Here, higher quality means that the initial guesses were closer to the truth. To be more specific, the initial guesses for the linear parameters used by NLS were Gaussian random variables centered on the true parameter values and having standard deviations equal to the true parameter multiplied by a prior information value (in other words, the prior information value can also be understood as the coefficient of variation of the "prior distribution"). The specific quantification of "high," "medium," and "low" is admittedly somewhat subjective and varies across the different ODE models, as specified below. For the sake of better and faster convergence of the optimization algorithms (especially NLS), the nonlinear parameters were constrained to a given range, and this range was the same no matter how we varied the prior information on linear parameters. Further, in each Monte-Carlo iteration, we used exactly the same (pseudorandom) initial guess for nonlinear parameters for both NLS and SLS. Thus, as far as the information on nonlinear parameters is concerned, this was kept invariant for each benchmark model, irrespective of the prior on linear parameters. Consequently, both algorithms received the same prior information regarding nonlinear parameters, and neither one was treated preferentially.

The noise level (signal-to-noise ratio, SNR) we used is defined as follows. For a given solution $x(t)$ of an ODE equation, we calculate the standard deviation $\sigma_x = \mathrm{std}(x(t_1), \ldots, x(t_n))$. Then SNR of, say, 10% and 20% is given by $\sigma = \sigma_x/10$, and $\sigma = \sigma_x/5$, respectively, where σ is the standard deviation of a Gaussian measurement error ϵ as defined in equation (4). We will refer to these SNRs as "low noise" and "high noise," respectively (cf.

[49]; albeit in a different context). We then compared the mean square errors (MSE) of the resulting parameter estimates, which leads to a valid comparison in statistically identifiable ODE models (see e.g., [20] for relevant definitions and results). As another accuracy measure, we used the criteria (5) and (7) evaluated at optimal parameter values. The two criteria we propose, though reasonable, are different. Hence, they are not expected to be in agreement in every experimental setup. However, the global conclusions reached with them in Section 5 are coherent and favor SLS.

We now provide the mathematical details on the models and the experimental setups.

3.1.1. Age-Group SIR. The system of interest is an epidemiological model of SIR-type (Susceptible—Infected—Recovered) and includes age groups and seasonal components (e.g., [50]). The infectious process in each age group $1 \le a \le M$ and each season $1 \le y \le L$ is described using two equations for the proportion of susceptible (S) and infected (I) individuals within the population (the proportion of recovered individuals is given by $1 - S - I$):

$$\begin{aligned} S'_{a,y}(t) &= -S_{a,y}(t)\kappa_y \sum_{j=1}^{M} \beta_{a,j} I_{j,y}(t), \\ I'_{a,y}(t) &= S_{a,y}(t)\kappa_y \sum_{j=1}^{M} \left(\beta_{a,j} I_{j,y}(t)\right) - \gamma I_{a,y}(t). \end{aligned} \tag{9}$$

The parameters of the model are the $M \times M$ transmission matrix β, the recovery rate γ, and $\kappa_{2,\ldots,L}$, which signify the relative infectivity of, e.g., influenza virus strains circulating in seasons $2, \ldots, L$ compared to season 1 (κ_1 is used as a reference and fixed at 1). As shown in Yaari et al. [46], taking into account separability characteristics of this model is advantageous for data fitting purposes. Specifically, (9) is nonlinear in the initial value $S(0)$, which are typically unknown and have to be estimated. For our purposes, it suffices to consider a model with one age group and one season. The following parameter setup was used: $S(0) = 0.56, I(0) = 1e-04, \beta = 6, \gamma = 2.3$. We considered two sample sizes, 18 and 36, and two noise levels, 10% and 20%. The prior information used was $\{0.1, 0.2, 0.3\}$, corresponding to high, medium, and low quality, respectively. The size of the Monte-Carlo study was 500 simulations.

3.1.2. Lotka–Volterra with Seasonal Forcing. As another benchmark we considered an extension of a classical predator-prey model, namely, a Lotka–Volterra model including seasonal forcing of the predation rate, using two additional parameters that control the amplitude (ε) and phase (ω) of the forcing:

$$\begin{aligned} x'_1(t) &= \alpha x_1(t) - \beta(1 + \varepsilon \sin(2\pi(t/T + \omega)))x_1(t)x_2(t), \\ x'_2(t) &= \delta(1 + \varepsilon \sin(2\pi(t/T + \omega)))x_1(t)x_2(t) - \gamma x_2(t). \end{aligned}$$

The nonlinear parameters are ϵ and ω. We considered the dynamics within the time interval $t \in [0, 25]$. The parameter setup is given by

$$\begin{aligned} \theta &= \{\alpha, \beta, \gamma, \delta, \varepsilon, \omega\} \\ &= \{2/3, 4/3, 1.0, 1.0, 0.2, 0.5\}, \end{aligned}$$

and initial values are $\{x_1(0), x_2(0)\} = \{0.9, 0.9\}$. Four experimental scenarios were studied, corresponding to sample sizes of 100 and 200, and SNRs of 10% and 20%. The prior information values were $\{0.05, 0.1, 0.2\}$, corresponding to high, medium, and low quality, respectively. The size of the Monte-Carlo study was 500 simulations.

3.1.3. GMA System. The GMA system we analyzed consists of three differential equations in three variables ([18]; pp. 84–85). They are

$$\begin{aligned} x'_1(t) &= \gamma_{11} x_2^{f_{121}}(t) x_3^{f_{131}}(t) - \gamma_{12} \\ &\qquad x_1^{f_{112}} x_2^{f_{122}} - \gamma_{13} x_1^{f_{113}} x_3^{f_{133}}, \\ x'_2(t) &= \gamma_{12} x_1^{f_{112}} x_2^{f_{122}} - \gamma_{22} x_2^{f_{222}}, \\ x'_3(t) &= \gamma_{13} x_1^{f_{113}} x_3^{f_{133}} - \gamma_{32} x_3^{f_{332}}. \end{aligned}$$

Here the linear parameters are the rate constants γ, while the nonlinear ones are the indexed kinetic orders f. Note that the parameters f are allowed to become negative and their sign might or might not be known. We considered the dynamics of the system within the time interval $[0, 4]$. The parameter setup is the one presented in Voit [18]; namely,

$$\begin{aligned} \theta &= \{\gamma_{11}, f_{121}, f_{131}, \gamma_{12}, f_{112}, f_{122}, \gamma_{13}, f_{113}, \\ &\quad f_{133}, \gamma_{22}, f_{222}, \gamma_{32}, f_{332}\} \\ &= \{0.4, -1.0, -1.0, 3.0, 0.5, \\ &\quad -0.1, 2.0, 0.75, -0.2, 1.5, 0.5, 5.0, 0.5\}, \end{aligned}$$

and initial values are $\{x_1(0), x_2(0), x_3(0)\} = \{0.5, 0.5, 1.0\}$. Four experimental scenarios were studied: sample sizes of 100 and 200, with SNRs of 10% and 20%. The prior information values were $\{0.1, 0.3, 0.5\}$, corresponding to high, medium, and low quality, respectively. The size of the Monte-Carlo study was 500 simulations. Parameter estimation for GMA systems is considered to be a challenging numerical task [18].

3.1.4. FitzHugh–Nagumo System. The FitzHugh–Nagumo (FHN) system [51–53] models spiked action potentials in neuron transmission. It is given by

$$\begin{aligned} x'_1(t) &= c\left(x_1(t) - \frac{x_1^3(t)}{3} + x_2(t)\right), \\ x'_2(t) &= -\left(\frac{1}{c}\right)(x_1(t) - a + b x_2(t)). \end{aligned} \tag{10}$$

This system with two state variables was proposed as a simplification of a more complicated model presented in Hodgkin and Huxley's study [54] for studying and

simulating nerve function in giant squid axons. The FHN model is used in neurophysiology as an approximation of the observed action potential.

The system (10) is linear in parameters a, and b, but nonlinear in c. We considered two sample sizes, $n = 20$ and $n = 40$, and two SNRs of 10% and 20%. The parameters were set to $\{a, b, c\} = \{0.2, 0.2, 3\}$. The initial values were $\{x_1(0), x_2(0)\} = \{-1.0, 1.0\}$. The true solutions were obtained over the time interval $[0, 20]$. The prior information used here was $\{0.5, 1.0, 3.0\}$, corresponding to high, medium, and low quality, respectively (the initial guesses for parameters were assured to be positive). The size of the Monte-Carlo study was 500 simulations. Many researchers studied the problem of parameter estimation for the FHN model. In particular, Ramsay et al. [55], Campbell and Steele [56], and Ramsay and Hooker [11] pointed out several difficulties in estimating the parameters for this ODE system.

3.2. Results of the Monte-Carlo Analysis. Our findings are presented through charts and tables. The primary summaries are Tables 1 and 2, where we report the ratios of the mean square errors (square errors averaged over Monte-Carlo simulations) for estimates of linear parameters (for nonlinear parameters, see the discussion at the end of this section). Several conclusions can be gleaned from the tables.

(i) Given high-quality prior information, the accuracy of NLS and SLS is comparable, and neither is superior throughout the variety of experimental setups (at least some of the differences that one sees from the raw numbers in the tables are plausibly attributable to the Monte-Carlo simulation error and as such appear to be insignificant).

(ii) When the quality of prior information degrades to medium or low, the performance of SLS becomes in most cases significantly better than that of NLS (with an extent depending on the specific experimental setup).

(iii) For a fixed noise level, as the sample size increases, the advantage of SLS becomes more pronounced.

(iv) For a fixed sample size, as the noise level increases, the SLS is still better than NLS, but to a lesser extent.

These results can also be visualized through a combination of simple statistical charts. Thus, Figure 1 displays the line graphs that compare MSEs of the two methods under several experimental setups. Whereas the numbers in Tables 1 and 2 are ratios of MSEs, the figures here present absolute MSE values. From the graphs, an advantage of SLS over NLS is apparent for less than ideal prior information. Note that, in this specific setting, SLS performed worse than NLS for high-quality prior information. A plausible explanation may be the following: while, under our experimental setup, the amount of information used by SLS via (3) is fixed throughout simulations, NLS can in principle receive arbitrarily precise initial guesses on linear parameters. One may therefore envision a threshold, where using the latter kind of information outweighs the benefits of using the structural relationship (3). However, a precise quantification of the phenomenon is hardly possible beyond an observation that it appears to manifest itself in scenarios with excellent knowledge on likely parameter values. In reality, such ideal prior information is rare.

Table 1: MSE ratios for linear parameters (small samples).

Prior	Low noise				High noise			
	SIR	LV	GMA	FHN	SIR	LV	GMA	FHN
Low	8.3	5.8	4.0	2.3	2.8	2.2	2.2	1.3
Medium	3.9	1.8	3.1	1.7	1.4	1.2	1.8	1.2
High	0.9	1.3	0.9	2.0	0.4	1.0	0.6	1.1

The MSE ratios (computed by averaging square errors over Monte-Carlo simulation runs) of NLS and SLS for estimating the linear parameters in various benchmark models and under different experimental setups are displayed (see Section 3.1 for detailed specifications). To identify model names, self-explanatory abbreviations are used. The values in the table are rounded off to one significant digit. The sample size is $n = 100$ for the GMA and Lotka–Volterra models, $n = 20$ for the FitzHugh–Nagumo system, and $n = 18$ for SIR model. The noise levels are 10% and 20%. Values larger than 1 in the table correspond to the cases where SLS performs better than NLS. Note the decreasing pattern in the columns, reflecting the effect of the quality of prior information on the performance of NLS.

Table 2: MSE ratios for linear parameters (large samples).

Prior	Low noise				High noise			
	SIR	LV	GMA	FHN	SIR	LV	GMA	FHN
Low	12.0	9.6	3.9	3.9	3.8	3.2	2.2	2.2
Medium	4.1	4.3	2.8	1.9	1.6	1.6	1.6	1.3
High	1.0	2.2	0.7	2.3	0.7	1.2	0.5	1.5

The sample size is $n = 200$ for the GMA and Lotka–Volterra models; $n = 40$ for the FitzHugh–Nagumo system; and $n = 36$ for the SIR model. The noise levels are 10% and 20%. For an interpretation of the results, see Table 1. There is an increased advantage of SLS over NLS in comparison to Table 1.

Panel (a) of Figure 1 further suggests that in the specific scenarios we report, SLS improves when the noise level decreases, which is different from NLS in the same figure.

Figure 2 is a scatterplot of NLS and SLS losses (5) and (7) (on a log scale) evaluated at optimal parameter estimates. The figure highlights in yet another manner the importance of prior information for NLS: it is evident that the performance of the latter is strongly affected by the quality of initial parameter guesses. Again, NLS and SLS perform similarly when the prior information is of high quality. However, when the quality of prior information is less than ideal, as it is in most applications, NLS becomes substantially worse than SLS. The scatterplot also gives a quick impression of the variability of estimation results.

The conclusions that we drew from Figure 2 are confirmed by the panel (a) of Figure 3, which presents boxplots of NLS and SLS losses (on a log scale) measured according to criteria (5) and (7). The pattern is clear: SLS performs better than NLS, and the inferiority for NLS becomes more dramatic with degrading prior information.

Panel (b) of Figure 3 summarizes computation times. SLS is in general much faster. The execution time of NLS is affected by the quality of prior information and,

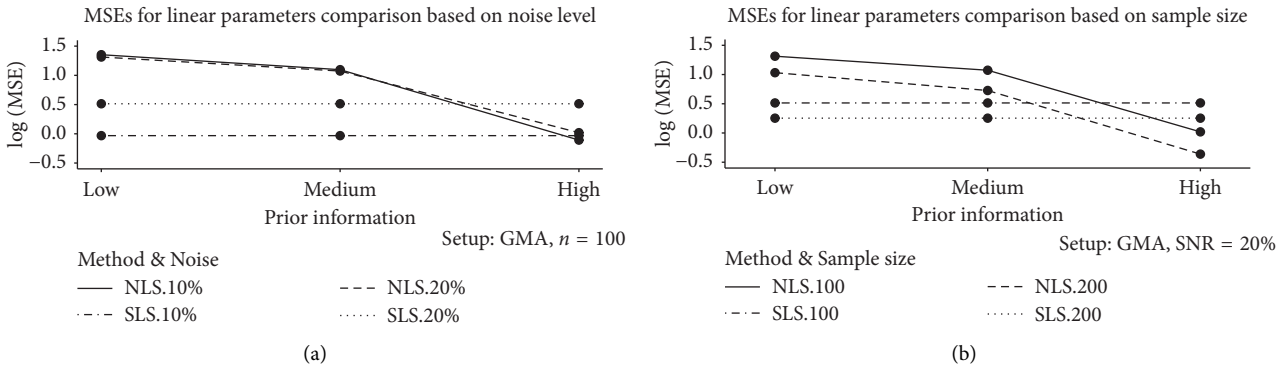


Figure 1: The plot gives MSEs on a log scale (computed as averages over Monte-Carlo simulation runs) for linear parameters plotted against the quality of prior information. In panel (a), the comparison is on the basis of the noise level. The graph indicates that the performance of NLS worsens with decreasing quality of prior information. On the other hand, the performance of SLS is not affected by the quality of prior information, in agreement with the experimental design. Except for the rare case of high-quality prior information, where NLS is better, SLS clearly outperforms NLS. In panel (b), the comparison is based on the sample size. The overall pattern is similar to that in panel (a).

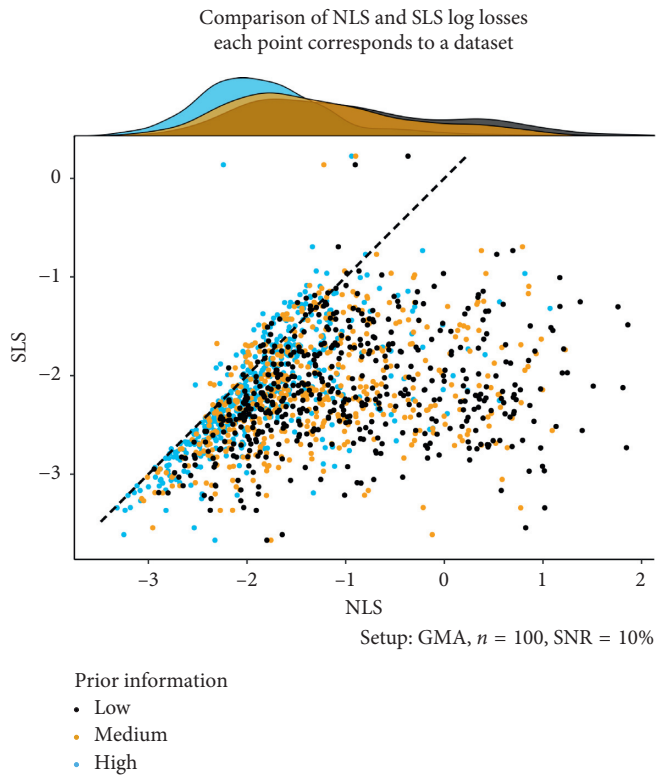


Figure 2: The plot visualizes the performance (on a log scale) of NLS and SLS according to criteria (5) and (7), which are evaluated at the optimal parameter estimates. Points in the scatterplot are colored according to the quality of prior information used to compute the NLS estimates. The 45° diagonal line passing through the origin has been added for reference and intuitive assessment. The scatterplot is supplemented with marginal density estimates using the same color coding. The density estimates indicate that, as the quality of prior information degrades, the quality of NLS results suffers, which manifests in longer right tails of the densities. By definition, performance of SLS is not affected by the quality of prior information on linear parameters. For high-quality prior information, clustering of losses in the scatterplot close to the reference line suggests that the overall performance of both NLS and SLS is comparable. As the quality of prior information decreases, the point clouds spread to the right, indicating that SLS starts to perform noticeably better than NLS. Furthermore, unlike Tables 1 and 2, the scatterplot and the range frame (see [57]; pp. 130–132) convey an impression of the variability in the estimation results over multiple datasets: NLS is visually more variable than SLS.

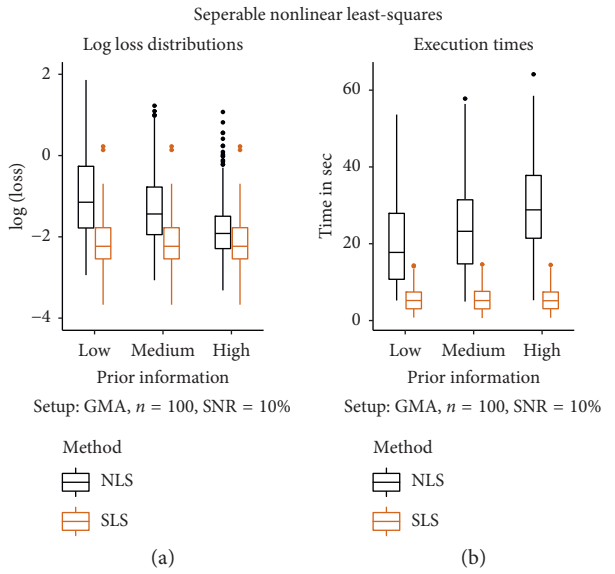


Figure 3: The plot presents a comparison of NLS and SLS. In panel (a) boxplots of the losses (5) and (7) (on a log scale) evaluated at the optimal parameter estimates are displayed. For high-quality prior information, the NLS and SLS loss distributions are close. As the quality of prior information degrades, NLS losses start to assume higher values compared to SLS, and their variability increases, as evidenced by the elongation of boxplots. In panel (b) the computation times are compared. The NLS computation times tend to be longer than those of SLS and increase as the quality of prior information increases. In both panels, the performance of SLS does not vary with the quality of prior information, in concordance with the experimental design.

interestingly, increases with this quality. The results for all other models (and noise levels) were similar and are therefore omitted.

Finally, Tables 3 and 4 provide information regarding the nonlinear parameters. In the case of NLS, one can observe how prior knowledge regarding linear parameters propagates into the estimation accuracy for nonlinear parameters. In particular, for less than ideal prior information on the linear parameters, SLS holds a pronounced edge over NLS, even in the case of nonlinear parameters.

4. Outlook

Data fitting in complex dynamical systems remains a challenging problem that cannot be treated in a cavalier fashion, even if one takes advantage of separability. For instance, in order to uncover the patterns in Section 3 of this work, we had to carefully design the experimental study, because otherwise simulations might not have converged or might have converged to poor solutions. This was true for both NLS and SLS, but whenever they were observed, convergence issues were much more severe for NLS; they were especially sensitive in the case of the Fitz-Hugh–Nagumo system. This result highlights the crucial role of *prior information* regarding the parameters or, expressed differently, the quality of the initial parameter guesses used for the optimization. We focused here primarily on the

Table 3: MSE ratios for nonlinear parameters (small samples).

Prior	Low noise				High noise			
	SIR	LV	GMA	FHN	SIR	LV	GMA	FHN
Low	23.0	1.7	1.1	1.1	6.6	1.1	1.0	1.0
Medium	8.7	1.2	1.0	1.0	2.6	0.9	1.0	1.0
High	1.0	1.0	0.9	1.0	0.4	0.9	0.9	1.0

The table displays the MSE ratios (computed through squared errors averaged over Monte-Carlo simulations) of NLS and SLS for estimating the nonlinear parameters. The experimental setup is as in Table 1. Values larger than 1 in the table correspond to the cases where SLS performs better than NLS. Since the prior information regarding nonlinear parameters stays invariant (see Section 3.1 for details), the table in particular shows the effects that the quality of initial guesses for linear parameters has on the estimation accuracy of NLS in the case of nonlinear ones. The results suggest that, in some settings, vague prior knowledge regarding linear parameters may have an adversary effect on the accuracy of NLS with respect to the nonlinear parameters.

Table 4: MSE ratios for nonlinear parameters (large samples).

Prior	Low noise				High noise			
	SIR	LV	GMA	FHN	SIR	LV	GMA	FHN
Low	29.0	4.1	1.1	1.6	6.5	1.3	1.0	1.3
Medium	8.2	2.1	0.9	1.0	2.2	1.1	0.9	1.0
High	0.9	1.2	0.8	1.0	0.7	0.9	0.8	1.0

The setup is as in Table 2. For an interpretation of the results, see Table 3.

effects of the prior information regarding the linear parameters. However, it also became clear that prior information on the nonlinear parameters has an equally crucial role for optimization purposes, and this was true for both NLS and SLS (data not shown).

As a result of our exploratory work, we envision the following promising research directions for the future.

4.1. Numerical Implementation of SLS for Dynamic Systems. All computations in our analysis were done in *R* using the *simode* package of Yaari and Dattner [25]. However, the idea of using separability properties of ODEs is independent of a particular programming language and can be implemented within other software packages quite as well. Indeed, much work has been done in the context of the variable projection method since it was first introduced by Golub and Pereyra [21]. In the context of nonlinear regression, the variable projection method of Golub and Pereyra [21] is implemented in *R* in the *nls* command; see Venables and Ripley [58] and pp. 218–220 for an example of its application. In addition, we are aware of the *TIMP* package of Mullen and van Stokkum [59]; which implements the variable projection method. Thus, a next step could be to combine the strengths of different packages, e.g., *simode* and *TIMP*, in order to develop advanced software for variable projection in the context of dynamic systems.

4.2. Customized Algorithms for Specific Classes of Complex Dynamical Systems. It is well known that the performance of an optimization scheme depends crucially on the underlying mathematical model used for description of the data. Thus, it

appears that different classes of dynamic models require specific algorithms tailored to their peculiarities. For instance, parameter estimation for GMA systems has different challenges than those encountered when working with SIR (see Section 3). We expect that there is much to gain from focusing future research on specific classes of models and developing stable algorithms for their parameter estimation.

4.3. Theoretical Properties of SLS in the Context of Dynamic Systems. Gugushvili and Klaassen [37] studied the statistical properties of NLS in the general context of smoothing, while Dattner and Klaassen [20] specifically addressed ODE systems that are linear in (functions of) the parameters. One might expect that some assumptions used in Gugushvili and Klaassen [37] can be relaxed when the problem is closer to the one considered in Dattner and Klaassen [20].

4.4. Extensions to Partially Observed, High-Dimensional, and Misspecified Dynamic Systems. Recent work dealing with inference in high-dimensional ODE models suggests that exploiting linearity in parameters is crucial for developing a successful estimation methodology (see e.g., [24, 39]). More generally, it would be interesting to use the variable projection method to study cases of partially observed, high-dimensional, and possibly misspecified dynamic systems. This work might additionally require the use of high-dimensional regularization techniques (e.g., [39]) for balancing data and model, and specifically take into account a potential model misspecification (see [55]).

5. Conclusions

In this work, we designed an extensive simulation study to explore the relative statistical and computational performance of two optimization schemes for inference in dynamic systems: the typical nonlinear least-squares (NLS) method and a novel, separable nonlinear least-squares (SLS) approach. As benchmarks, we considered several widely used ODE models arising in a variety of biological fields. We measured statistical performance of the two methods by the mean square error (MSE) of the estimates. As another performance criterion, we employed the loss function values at the optimal parameter estimates. Computational performance of the methods was also compared by the execution times required to complete each optimization.

Our overall recommendation is the following: whenever a complex dynamic system contains an appreciable number of linear parameters, estimation of its parameters should be addressed with the separable nonlinear least-squares method, rather than the more commonly used, generic nonlinear least-squares method. The general pattern emerging from our study is that SLS performs at least as well as, and frequently better than, NLS, especially if the prior information regarding the system is not ideal, which is typically the case in practice. This statement was found to be uniformly true over all models tested.

Data Availability

The code to reproduce our numerical results can be accessed at https://github.com/haroldship/complexity-2019-code/tree/master/Final_Code_First_Submission.

Disclosure

A preprint of the article is available on ArXiv at https://arxiv.org/abs/1908.03717.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Itai Dattner was supported by the Israeli Science Foundation grant number 387/15. Eberhard Voit was supported by grants NSF-MCB 1615373 (PI: Diana Downs) and NIH 2P30ES019776-05 (PI: Carmen Marsit). The funding agencies are not responsible for the content of this article.

References

[1] A. J. Lotka, *Elements of Mathematical Biology*, Dover Publications Inc., New York, NY, USA, 1956.

[2] R. M. May, *Stability and Complexity in Model Ecosystems. With a New Introduction by the Author*, Princeton University Press, Princeton, NJ, USA, 2nd edition, 2001.

[3] M. Peschel and W. Mende, *The Predator-Prey Model: Do We Live in a Volterra World?*, Springer, New York, NY, USA, 1986.

[4] V. Volterra, "Fluctuations in the abundance of a species considered mathematically," *Nature*, vol. 118, no. 2972, pp. 558–560, 1926.

[5] M. A. Savageau, "Biochemical systems analysis," in *A Study of Function and Design in Molecular Biology, Vol. 6739 of Advanced Book Program*, Addison-Wesley Publishing Company, Boston, MA, USA, 1976.

[6] E. O. Voit, "Biochemical systems theory: a review," *ISRN Biomathematics*, vol. 2013, Article ID 897658, 53 pages, 2013.

[7] I.-C. Chou and E. O. Voit, "Recent developments in parameter estimation and structure identification of biochemical and genomic systems," *Mathematical Biosciences*, vol. 219, no. 2, pp. 57–83, 2009.

[8] P. Gennemark and D. Wedelin, "Efficient algorithms for ordinary differential equation model identification of biological systems," *IET Systems Biology*, vol. 1, no. 2, pp. 120–129, 2007.

[9] M. Girolami and B. Calderhead, "Riemann manifold Langevin and Hamiltonian Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 2, pp. 123–214, 2011.

[10] K. McGoff, S. Mukherjee, and N. Pillai, "Statistical inference for dynamical systems: a review," *Statistics Surveys*, vol. 9, pp. 209–252, 2015.

[11] J. Ramsay and G. Hooker, *Dynamic Data Analysis. Modeling Data with Differential Equations*, Springer, New York, NY, USA, 2017.

[12] K. Schittkowski, *Numerical Data Fitting in Dynamical Systems. A Practical Introduction with Applications and Software*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2002.

[13] E. O. Voit and J. Almeida, "Decoupling dynamical systems for pathway identification from metabolic profiles," *Bioinformatics*, vol. 20, no. 11, pp. 1670–1681, 2004.

[14] S. Dolatshahi and E. O. Voit, "Identification of metabolic pathway systems," *Frontiers in Genetics*, vol. 7, p. 6, 2016.

[15] M. Faraji and E. O. Voit, "Stepwise inference of likely dynamic flux distributions from metabolic time series data," *Bioinformatics*, vol. 33, no. 14, pp. 2165–2172, 2017.

[16] G. Goel, I.-C. Chou, and E. O. Voit, "System estimation from metabolic time-series data," *Bioinformatics*, vol. 24, no. 21, pp. 2505–2511, 2008.

[17] J. M. Varah, "A spline least squares method for numerical parameter estimation in differential equations," *SIAM Journal on Scientific and Statistical Computing*, vol. 3, no. 1, pp. 28–46, 1982.

[18] E. O. Voit, *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge, UK, 2000.

[19] D. M. Himmelblau, C. R. Jones, and K. B. Bischoff, "Determination of rate constants for complex kinetics models," *Industrial & Engineering Chemistry Fundamentals*, vol. 6, no. 4, pp. 539–543, 1967.

[20] I. Dattner and C. A. J. Klaassen, "Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 1939–1973, 2015.

[21] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413–432, 1973.

[22] A. Ruhe and P. Å. Wedin, "Algorithms for separable nonlinear least squares problems," *SIAM Review*, vol. 22, no. 3, pp. 318–337, 1980.

[23] I. Dattner, E. Miller, M. Petrenko, D. E. Kadouri, E. Jurkevitch, and A. Huppert, "Modelling and parameter inference of predator-prey dynamics in heterogeneous environments using the direct integral approach," *Journal of The Royal Society Interface*, vol. 14, no. 126, Article ID 20160525, 2017.

[24] L. Wu, X. Qiu, Y.-X. Yuan, and H. Wu, "Parameter estimation and variable selection for big systems of linear ordinary differential equations: a matrix-based approach," *Journal of the American Statistical Association*, vol. 114, no. 526, pp. 657–667, 2019.

[25] R. Yaari and I. Dattner, "Simode: statistical inference for systems of ordinary differential equations using separable integral-matching, R package version 1.1.4," January 2019, https://CRAN.R-project.org/package=simode.

[26] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. 1–26, 2003.

[27] K. M. Mullen, *Separable nonlinear models: theory, implementation and applications in physics and chemistry*, Ph.D. thesis, Vrije Universiteit, Amsterdam, Netherlands, 2008.

[28] J. Chung and J. G. Nagy, "An efficient iterative approach for large-scale separable nonlinear inverse problems," *SIAM Journal on Scientific Computing*, vol. 31, no. 6, pp. 4654–4674, 2010.

[29] M. Gan, C. L. P. Chen, G.-Y. Chen, and L. Chen, "On some separated algorithms for separable nonlinear least squares problems," *IEEE Transactions on Cybernetics*, vol. 48, no. 10, pp. 2866–2874, 2018.

[30] N. B. Erichson, P. Zheng, K. Manohar, S. L. Brunton, J. N. Kutz, and A. Y. Aravkin, "Sparse principal component analysis via variable projection," 2018, https://arxiv.org/abs/1804.00341.

[31] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3426–3436, 2000.

[32] M. W. Hirsch, R. L. Devaney, and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Vol. 60, Academic Press, Cambridge, MA, USA, 1974.

[33] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman & Hall, London, UK, vol. 66 of Monographs on Statistics and Applied Probability, 1996.

[34] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London, UK, vol. 58 of Monographs on Statistics and Applied Probability, 1994.

[35] L. Wasserman, *All of Nonparametric Statistics*, Springer, New York, NY, USA, 2006.

[36] N. J.-B. Brunel, "Parameter estimation of ODEs via nonparametric estimators," *Electronic Journal of Statistics*, vol. 2, pp. 1242–1267, 2008.

[37] S. Gugushvili and C. A. J. Klaassen, "$\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing," *Bernoulli*, vol. 18, no. 3, pp. 1061–1098, 2012.

[38] M. Vilela, C. C. H. Borges, S. Vinga et al., "Automated smoother for the numerical decoupling of dynamics models," *BMC Bioinformatics*, vol. 8, no. 1, p. 305, 2007.

[39] S. Chen, A. Shojaie, and D. M. Witten, "Network reconstruction from high-dimensional ordinary differential equations," *Journal of the American Statistical Association*, vol. 112, no. 520, pp. 1697–1707, 2017.

[40] I. Dattner, "A model-based initial guess for estimating parameters in systems of ordinary differential equations," *Biometrics*, vol. 71, no. 4, pp. 1176–1184, 2015.

[41] I. Dattner and S. Gugushvili, "Application of one-step method to parameter estimation in ODE models," *Statistica Neerlandica*, vol. 72, no. 2, pp. 126–156, 2018.

[42] I. Dattner and A. Huppert, "Modern statistical tools for inference and prediction of infectious diseases using mathematical models," *Statistical Methods in Medical Research*, vol. 27, no. 7, pp. 1927–1929, 2018.

[43] F. Vissing Mikkelsen and N. R. Hansen, "Learning large scale ordinary differential equation systems," 2017, https://arxiv.org/abs/1710.09308.

[44] I. Vujačić, I. Dattner, J. González, and E. Wit, "Time-course window estimator for ordinary differential equations linear in the parameters," *Statistics and Computing*, vol. 25, no. 6, pp. 1057–1070, 2015.

[45] R. Yaari and I. Dattner, "Simode: R package for statistical inference of ordinary differential equations using separable integral-matching," 2018, https://joss.theoj.org/papers/10.21105/joss.01850.

[46] R. Yaari, I. Dattner, and A. Huppert, "A two-stage approach for estimating the parameters of an age-group epidemic model from incidence data," *Statistical Methods in Medical Research*, vol. 27, no. 7, pp. 1999–2014, 2018.

[47] W. H. Lawton and E. A. Sylvestre, "Elimination of linear parameters in nonlinear regression," *Technometrics*, vol. 13, no. 3, pp. 461–467, 1971.

[48] H. Wickham, *Ggplot2. Elegant Graphics for Data Analysis*, Springer, New York, NY, USA, 2009.

[49] I. M. Johnstone and B. W. Silverman, "Empirical Bayes selection of wavelet thresholds," *The Annals of Statistics*, vol. 33, no. 4, pp. 1700–1752, 2005.

[50] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.
[51] R. FitzHugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophysical Journal*, vol. 1, no. 6, pp. 445–466, 1961.
[52] R. FitzHugh, "Mathematical models of excitation and propagation in nerve," in *Biological Engineering*, H. P. Schwan, Ed., McGraw-Hill, 1969pp. 1–85, McGraw-Hill, New York, NY, USA, vol. 9 of Inter-university electronics series.
[53] J. Nagumo, S. Arimoto, and S. Yoshizawa, "An active pulse transmission line simulating nerve axon," *Proceedings of the IRE*, vol. 50, no. 10, pp. 2061–2070, 1962.
[54] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952.
[55] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao, "Parameter estimation for differential equations: a generalized smoothing approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 5, pp. 741–796, 2007.
[56] D. Campbell and R. J. Steele, "Smooth functional tempering for nonlinear differential equation models," *Statistics and Computing*, vol. 22, no. 2, pp. 429–443, 2012.
[57] E. R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, USA, 2nd edition, 2001.
[58] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S. Statistics and Computing*, Springer, New York, NY, USA, 4th edition, 2002.
[59] K. Mullen and I. van Stokkum, "TIMP: an R package for modeling multi-way spectroscopic measurements," *Journal of Statistical Software*, vol. 18, no. 3, pp. 1–46, 2007.

Hindawi
Complexity
Volume 2020, Article ID 3687961, 14 pages
https://doi.org/10.1155/2020/3687961

Research Article

The Dynamics of Canalizing Boolean Networks

Elijah Paul,[1] **Gleb Pogudin**,[2,3] **William Qin**,[4] **and Reinhard Laubenbacher**[5]

[1]*California Institute of Technology, Pasadena, CA, USA*
[2]*Courant Institute of Mathematical Sciences, New York University, New York, NY, USA*
[3]*Department of Computer Science, National Research University Higher School of Economics, Moscow, Russia*
[4]*Millburn High School, Millburn, NJ, USA*
[5]*Center for Quantitative Medicine, University of Connecticut Health Center and Jackson Laboratory for Genomic Medicine, Farmington, CT, USA*

Correspondence should be addressed to Reinhard Laubenbacher; laubenbacher@uchc.edu

Elijah Paul, Gleb Pogudin, and William Qin contributed equally to this work.

Received 13 September 2019; Accepted 6 November 2019; Published 20 January 2020

Guest Editor: George V. Popescu

Boolean networks are a popular modeling framework in computational biology to capture the dynamics of molecular networks, such as gene regulatory networks. It has been observed that many published models of such networks are defined by regulatory rules driving the dynamics that have certain so-called canalizing properties. In this paper, we investigate the dynamics of a random Boolean network with such properties using analytical methods and simulations. From our simulations, we observe that Boolean networks with higher canalizing depth have generally fewer attractors, the attractors are smaller, and the basins are larger, with implications for the stability and robustness of the models. These properties are relevant to many biological applications. Moreover, our results show that, from the standpoint of the attractor structure, high canalizing depth, compared to relatively small positive canalizing depth, has a very modest impact on dynamics. Motivated by these observations, we conduct mathematical study of the attractor structure of a random Boolean network of canalizing depth one (i.e., the smallest positive depth). For every positive integer ℓ, we give an explicit formula for the limit of the expected number of attractors of length ℓ in an n-state random Boolean network as n goes to infinity.

1. Introduction

Dynamic mathematical models are a key enabling technology in systems biology. Depending on the system to be modeled, the data and information available for their construction, the questions to be answered, and different modeling frameworks can be used. For kinetic models, systems of ordinary differential equations have a long tradition. Generally, they will have the very special structure of polynomial equations representing Michaelis–Menten kinetics, even in the case of systems, such as gene regulatory networks, that are not proper biochemical reaction networks. It is this special structure that gives models desirable properties and aids in model analysis. Besides continuous models, a range of discrete models are finding increasingly frequent use, in particular Boolean network models of a broad variety of biological systems, from intracellular molecular networks to population-level compartmental models (see e.g., [1–5]), going back to the work of Kauffman in the 1960s [6–8]. While Boolean network models, a collection of nodes, whose regulation by other nodes is described via a logical rule built from Boolean operators, are intuitive and mathematically simple to describe, their analysis is severely limited by the lack of mathematical tools. It generally consists of simulation results. Any set function on binary strings that takes on binary values can be represented as a Boolean function, so that the class of general Boolean networks is identical to the class of set functions on binary strings of a given length, making any general analysis impossible. The search for special classes of Boolean functions that are broad enough to cover all or most rules that occur in biology, but special enough to allow for mathematical approaches has a long history.

It was again Kauffman who proposed a class of functions [7] with properties inspired by the developmental biology

concept of *canalization*, going back to Waddington in the 1940s [9]. There is some evidence that canalizing Boolean functions do indeed appear disproportionately in published models and that the dynamics of Boolean network models consisting of canalizing functions has special properties, in particular a "small" number of attractors. This is important since, in the case of intracellular molecular network models, attractors correspond to the different phenotypes a cell is capable of. Here, again, the majority of available results are obtained by simulating large numbers of such networks. The main question of this paper is as follows: *What do the dynamics of a random canalizing Boolean network look like?* We approach this question using both computer simulations and analytical methods, with the main result of the paper being Theorem 2, which gives a provable formula for the number of expected attractors of a general Boolean network with a particular canalization property. In addition to providing important information about canalizing Boolean network models, this result can be viewed as a part of a growing body of mathematical results characterizing this class of networks that promises to be as rich as that for chemical reaction network models based on ordinary differential equations.

2. Background

The property of canalization for Boolean functions was introduced by Kauffman in [7], inspired by the concept of *canalization* from developmental biology [9]. A Boolean function is canalizing if there is a variable and a value of the variable such that if the variable takes the value, then the value of the function does not depend on other variables. It was shown that models defined by such functions often exhibit less chaotic and more stable behavior [10, 11]. Nested canalizing functions, obtained by applying the concept of canalization recursively, were introduced in [2]. They form a special subset of canalizing functions and have stable dynamics [11]. We note that there are other important properties shared by Boolean networks arising in modeling (for example, sparsity [7]). In this paper we focus only on canalization and its impact on the dynamics, and one of the natural future directions would be to consider several such properties simultaneously.

To cover more models arising in applications, the notion of nested canalizing function was relaxed by Layne et al. [12] by assigning to every Boolean function its *canalizing depth*. Noncanalizing functions have canalizing depth zero, and nested canalizing functions have the maximal possible canalizing depth equal to the number of variables. Canalizing depth of a Boolean network is defined as the minimum of the canalizing depths of the functions defining the network. In [12], activities and sensitivities of functions with different canalizing depths and stability and criticality of Boolean networks composed from such functions were investigated. It has been observed that Boolean networks of higher canalizing depth tend to be more stable and less sensitive. However, increasing the canalizing depth to the maximum does not improve the stability significantly compared to moderate positive canalizing depth. These observations give a strong indication of the biological utility of canalizing function, even with small canalizing depth.

Attractors in Boolean network models can be interpreted as distinct cell types [13, p. 202] and their lengths can be viewed as the variety of different gene expression patterns corresponding to the cell type. Thus, understanding the attractor structure of a random Boolean network defined by functions of a fixed canalizing depth is important for assessing biological relevance of such models. Analytic study of the attractor structure of nested canalizing Boolean networks has been carried out in [11]. For discussion about attractors of length one (i.e., steady state), we refer to [14].

3. Our Results

The main question of this paper is as follows: *What do the dynamics of a random canalizing Boolean network look like?* We approach this question using both computer simulations and analytical methods.

In our *computational experiments*, we generate approximately 30 million random Boolean networks of all possible canalizing depths with the number of variables ranging from 4 to 20. For each of these networks, we determine sizes of all the attractors and basins of attraction and analyze the obtained data. We discover the following:

(1) For a fixed number of variables, the sample mean of the number of attractors and average size of an attractor decrease when the canalizing depth increases

(2) The decrease of the average size of an attractor is much greater than the decrease of the number of attractors as the canalizing depth increases

(3) Both decreases from (8) are substantial when the canalizing depth changes from zero to small canalizing depths, but a further increase of the canalizing depth does not lead to a significant decrease for either the sample means or for the empirical distributions

(4) The relative decrease of the sample mean of the number of attractors and the average attractor size when the canalizing depth changes from zero to one becomes sharper when the number of variables increases

Observations (8) and (A.4) are consistent with the results obtained in [12] for sensitivity and stability. This provides new evidence that Boolean networks of small positive canalizing depth are almost as well-suited for modeling as those with nested canalizing functions, from the point of view of stability. Since there are many more canalizing functions of small positive canalizing depth than nested canalizing functions [15, Section 5], they provide a richer modeling toolbox.

Motivated by observation (A.4), we conduct a mathematical study of the attractor structure of a random Boolean network of canalizing depth one (that is, the minimal

positive depth). Our main theoretical result, Theorem 2, gives, for every positive integer ℓ, a formula for the limit of the expected number of attractors of length ℓ in a random Boolean network of depth one. The same formulas are valid for a random Boolean network defined by canalizing functions (see Remark 5). In particular, our formulas show that a large random network of depth one, on average, has more attractors of small sizes that an average Boolean network (Remark 6).

Formulas similar to the ones in our proofs (e.g., in Lemma A.4) have already appeared in the study of the average number of attractors of a given length in sparse Boolean networks, e.g., [16, equation (2)] and [17, equation (6)]. The results of [16, 17] are based on describing the asymptotic behavior of these formulas in terms of N, the number of nodes in the network, and the asymptotics is of the form $O(N^{\alpha})$. In our case, the average number of attractors of a given length simply approaches a constant as $N \longrightarrow \infty$ (that is, $O(1)$), but our methods allow us to find the exact value of this constant.

The source code we used for generating and analyzing data is available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics. The raw data are available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/tree/master/data.

Structure of the Paper. The rest of the paper is organized as follows. Section 4 contains necessary definitions about canalizing functions and Boolean networks. Outlines of the algorithms used in our computational experiments are in Section 5. The main observations are summarized in Section 6. Our main theoretical result about attractors in a random Boolean network of canalizing depth one (Theorem 2) is presented in Section 7. Section 8 contains conclusions. The proofs are located in the Appendix.

4. Preliminaries

Definition 1. A *Boolean network* is a tuple $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ of Boolean functions in n variables. For a state $\mathbf{a}_t = (a_{t,1}, a_{t,2}, \ldots, a_{t,n}) \in \{0,1\}^n$ at time t, we define the state $\mathbf{a}_{t+1} \coloneqq f(a_t) = (a_{t+1,1}, \ldots, a_{t+1,n}) \in \{0,1\}^n$ at time $t+1$ by

$$\begin{aligned} a_{t+1,1} &= f_1\left(a_{t,1}, \ldots, a_{t,n}\right), \\ &\vdots \\ a_{t+1,n} &= f_n\left(a_{t,1}, \ldots, a_{t,n}\right). \end{aligned} \tag{1}$$

Definition 2 (attractors and basins). Let $\mathbf{f} = (f_1, \ldots, f_n)$ be a Boolean network.

(i) A sequence $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \{0,1\}^n$ of distinct states is called an *attractor* of $\mathbf{f}$ if $\mathbf{f}(\mathbf{a}_i) = \mathbf{a}_{i+1}$ for every $1 \le i < \ell$ and $\mathbf{f}(\mathbf{a}_\ell) = \mathbf{a}_1$.

(ii) An attractor $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \{0,1\}^n$ is called a *steady state* if $\ell = 1$.

(iii) Let $A = (\mathbf{a}_1, \ldots, \mathbf{a}_\ell) \in (\{0,1\}^n)^\ell$ be an attractor of $\mathbf{f}$. The *basin* of A is the set

$$\{\mathbf{b} \in \{0,1\}^n \mid \exists N : \underbrace{\mathbf{f}(\mathbf{f}(\ldots(\mathbf{fb})\ldots))}_{N \text{ times}} \in A\}. \tag{2}$$

Definition 3. A nonconstant function $f(x_1, \ldots, x_n)$ is canalizing with respect to a variable x_i if there exists a canalizing value $a \in \{0,1\}$ such that

$$f\left(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n\right) \equiv \text{const}. \tag{3}$$

Example 1. Consider $f(x_1, x_2) = x_1 \cdot x_2$ (the product is understood modulo 2, that is, logical AND). It is canalizing with respect to x_1 with canalizing value 0 because $f(0, x_2) = 0$ regardless of the value of x_2. Analogously, it is canalizing with respect to x_2 with canalizing value 0.

Consider $g(x_1, x_2) = x_1 + x_2$ (summation is understood modulo 2, that is, logical XOR). It is not canalizing with respect to x_1 because

$$\begin{aligned} g(0, x_2) &= x_2 \neq \text{const}, \\ g(1, x_2) &= \overline{x}_2 \neq \text{const}. \end{aligned} \tag{4}$$

The same argument works for x_2 as well.

Definition 4. $f(x_1, \ldots, x_n)$ has canalizing depth [15, Definition 2.3] k if it can be expressed as

$$f = \begin{cases} b_1, & x_{i_1} = a_1, \\ b_2, & x_{i_1} \neq a_1, x_{i_2} = a_2, \\ \vdots & \\ b_k, & x_{i_1} \neq a_1, x_{i_2} \neq a_2, \ldots, x_{i_{k-1}} \neq a_{k-1}, x_{i_k} = a_k, \\ g \neq b_k, & x_{i_1} \neq a_1, \ldots, x_{i_k} \neq a_k, \end{cases} \tag{5}$$

where

(i) $i_1, \ldots, i_k$ are distinct integers from 1 to n

(ii) $a_1, \ldots, a_k, b_1, \ldots, b_k \in \{0,1\}$

(iii) g is a noncanalizing function in the variables $\{x_1, \ldots, x_n\} / \{x_{i_1}, \ldots, x_{i_k}\}$

Example 2. For example, if $f(x_1, x_2, x_3) = (x_1 + x_2)x_3$,

$$f\left(x_1, x_2, x_3\right) = \begin{cases} 0, & x_3 = 0, \\ x_1 + x_2, & x_3 \neq 0, \end{cases} \tag{6}$$

and $x_1 + x_2$ is noncanalizing. Therefore, f has canalizing depth 1.

Remark 1. Since g in Definition 4 is noncanalizing, every function has a single well-defined canalizing depth. In particular, a function of depth two is not considered to have depth one.

Definition 5. We say that a canalizing Boolean function $f(x_1, \ldots, x_n)$ is nested if f has canalizing depth n, that is, $g = 0$ or $g = 1$ (see Definition 4). For example, $f(x_1, x_2, x_3) = x_1 x_2 x_3$ is nested canalizing because

$$f = \begin{cases} 0, & x_3 = 0, \\ 0, & x_3 \neq 0, x_2 = 0, \\ 0, & x_2, x_3 \neq 0, x_1 = 0, \\ 1, & x_1, x_2, x_3 \neq 0, \end{cases} \tag{7}$$

so the canalizing depth of f is 3, which is equal to $n = 3$.

Definition 6. We say that a Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$ has canalizing depth k if $f_1, \ldots, f_n$ are Boolean functions of canalizing depth k.

5. Simulations: Outline of the Algorithms

In our computational experiment, we generated random Boolean networks of various canalizing depths. For each network, we store a list of pairs (a_i, b_i), where a_i is the size of the ith attractor of the network and b_i is the size of its basin. The generated data are available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/tree/master/data. To generate the data, we used two algorithms: one for generating a random Boolean network of a given canalizing depth and one for finding the sizes of attractors and their basins (Algorithm 1).

5.1. Generating Random Boolean Functions of a Given Canalizing Depth. [12, Section 5] contains a sketch of an algorithm for generating random Boolean functions that have canalizing depth at least k for a given k. Here, we generate functions of canalizing depth equal to k and take a different approach than [12]. In order to ensure that the probability distribution of possible outputs is uniform, we use the following structure theorem due to He and Macaulay [15].

Theorem 1 (see [15], Theorem 4.5). *Every Boolean function $f(x_1, \ldots, x_n) \not\equiv 0$ can be uniquely written as*

$$f(x_1, \ldots, x_n) = M_1(M_2(\cdots(M_{r-1}(M_r p_C + 1) + 1)\cdots) + 1) + b, \tag{8}$$

where $M_i = \prod_{j=1}^{k_i}(x_{i_j} + a_{i_j})$ for every $1 \leq i \leq r$, $p_C \not\equiv 0$ is a noncanalizing function, and $k = \sum_{i=1}^{r} k_i$ is the canalizing depth. Each x_i appears in exactly one of $\{M_1, \ldots, M_r, p_C\}$, and the only restrictions on equation (8) are the following "exceptional cases":

(E1) If $p_C \equiv 1$ and $r \neq 1$, then $k_r \geq 2$

(E2) If $p_C \equiv 1$ and $r = 1$ and $k_1 = 1$, then $b = 0$

Example 3. Consider $f(x_1, x_2, x_3, x_4) = x_1(x_2 + 1)(x_3 x_4 + x_3 + x_4)$ can be represented as

$$f = ((x_1 + 0)(x_2 + 1))(((x_3 + 1)(x_4 + 1))(1) + 1) + 0, \tag{9}$$

so $M_1 = (x_1 + 0)(x_2 + 1)$, $M_2 = (x_3 + 1)(x_4 + 1)$, $b = 0$, $k = 4$, and $p_C = 1$. This can be verified by expanding the brackets in the original and new representations of f.

Consider $g(x_1, x_2, x_3, x_4, x_5) = 1 + x_5(x_1 + x_2)(x_3 + 1)x_4$. It can be represented as

$$g = (x_5 + 0)(((x_3 + 1)(x_4 + 0))(x_1 + x_2) + 1) + 1, \tag{10}$$

so $M_1 = (x_5 + 0)$, $M_2 = (x_4 + 0)(x_3 + 1)$, $b = 1$, $k = 3$, and $p_C = x_1 + x_2$.

Our algorithm is summarized in Algorithms 2 and 3 below. Correctness of Algorithm 2 follows from Theorem 1, and correctness of Algorithm 3 can be proved directly by induction on k.

Remark 2. The complexity of Algorithm 2 is $O(n2^n)$ (see Proposition B.2). Given that the size of the output is $O(2^n)$, and this is nearly optimal.

We measured the runtimes of our implementation of Algorithm 2 on a laptop with a Core i5 processor (1.60 GHz) and 8 Gb RAM. Generating a single function with 20 variables (the largest number we used in our simulations) takes 4.9 − 5.5 seconds (faster for smaller canalizing depth). On a laptop, our implementation can go up to 24 variables (∼2 minutes to generate a function), and then hits memory limits. One can go further by using a lower level language and more careful packing. However, already a Boolean function in 40 variables would require at least 128 Gb of memory.

Remark 3. We generate a random noncanalizing function as follows. We generate a random Boolean function and test for canalization until we generate a noncanalizing one. Then, we return it. Since canalizing functions are rare [15, Section 5], this algorithm is fast enough for our purposes (see Lemma B.1).

6. Simulations: Results

Notation 1. For a Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$, let $N(\mathbf{f})$ and $S(\mathbf{f})$ denote the number of the attractors of $\mathbf{f}$ and the sum of the sizes of the attractors of $\mathbf{f}$, respectively. We define the average size of an attractor as $\mathrm{AS}(\mathbf{f}) := S(\mathbf{f})/N(\mathbf{f})$.

6.1. Sample Means of $N(\mathbf{f})$ and $AS(\mathbf{f})$. For every $n = 4, \ldots, 20$ and every $0 \leq k \leq n$, we generate random Boolean networks in n variables of canalizing depth k and compute the mean of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$. Figure 1 shows how these means depend on k for $n = 15$ (based on 50,000 samples for each k). The shape of the plots is similar for other values of n we did computation for (that is, $n = 4, \ldots, 20$). Note that although both means are decreasing, the decrease of the mean of $\mathrm{AS}(\mathbf{f})$ is more substantial.

In: A Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$ in n variables
Out: A list of pairs (a_i, b_i), where a_i is the size of the ith attractor of $\mathbf{f}$ and b_i is the size of its basin
(1) (Network $\longrightarrow$ Graph) Build a directed graph G with 2^n vertices corresponding to possible states and a directed edge from $\mathbf{a}$ to $\mathbf{f}(\mathbf{a})$ for every $\mathbf{a} \in \{0, 1\}^n$.
(2) (Attractors) Perform a depth-first search [18, § 22.3] traversal on G viewed as an undirected graph to detect the unique cycle in each connected component, these cycles are the attractors.
(3) (Basins) For each cycle from Step 2, perform a depth-first search traversal on G with all the edges reversed. The dfs trees will be the basins.
(4) Return the sizes of the attractors and basins found on Steps 2 and 3.

Algorithm 1: Finding the sizes of the attractors and their basins.

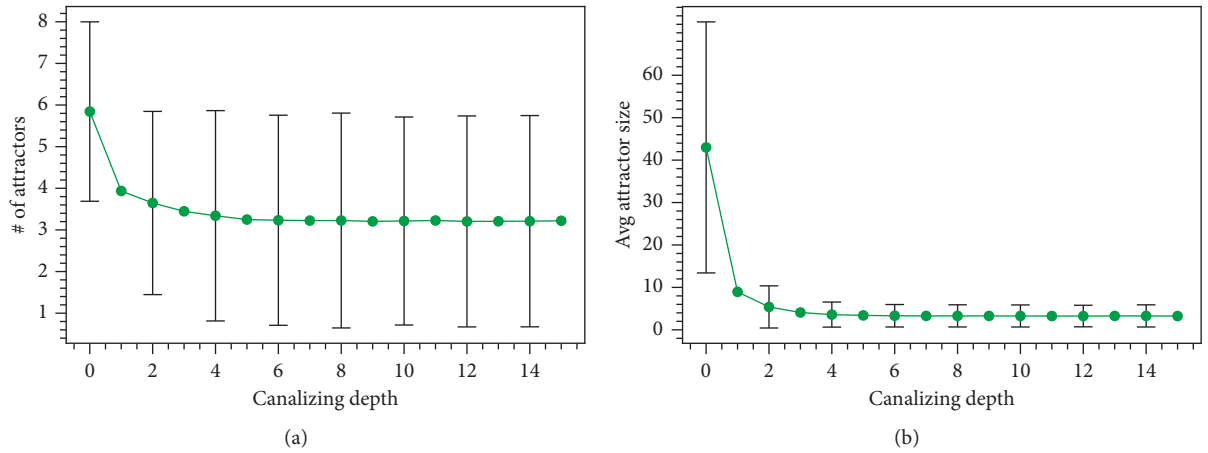


Figure 1: Dependencies of the sample means of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ on the canalizing depth. (a) The number of attractors $N(\mathbf{f})$. (b) Average size of an attractor $\mathrm{AS}(\mathbf{f})$.

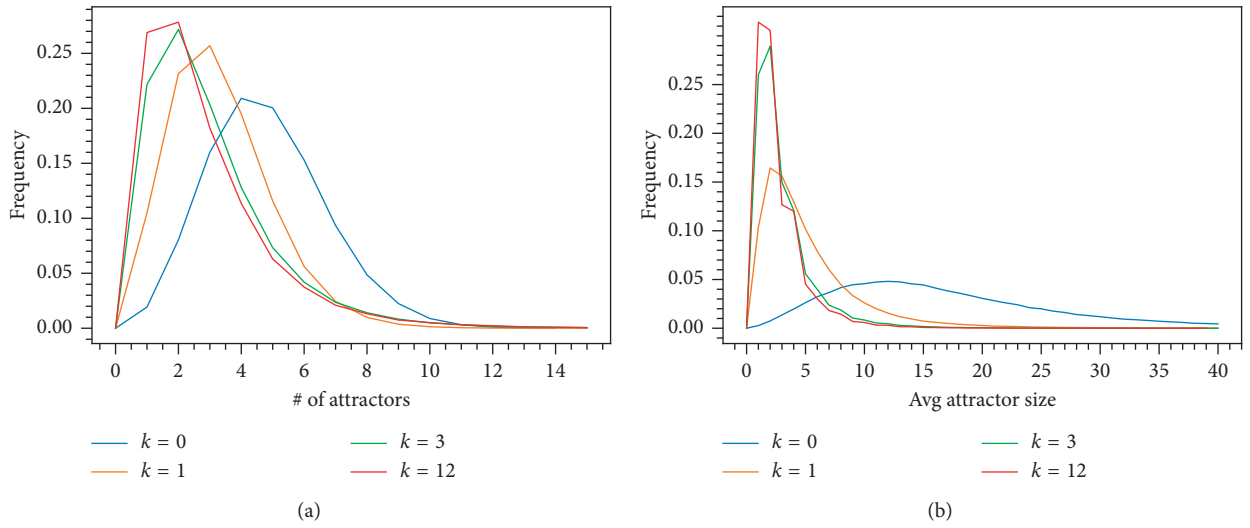


Figure 2: Empirical distributions of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ for $n = 12$ and $k = 0, 1, 3, 12$. (a) Distribution of the number of attractors $N(\mathbf{f})$. (b) Distribution of the average size of an attractor $\mathrm{AS}(\mathbf{f})$.

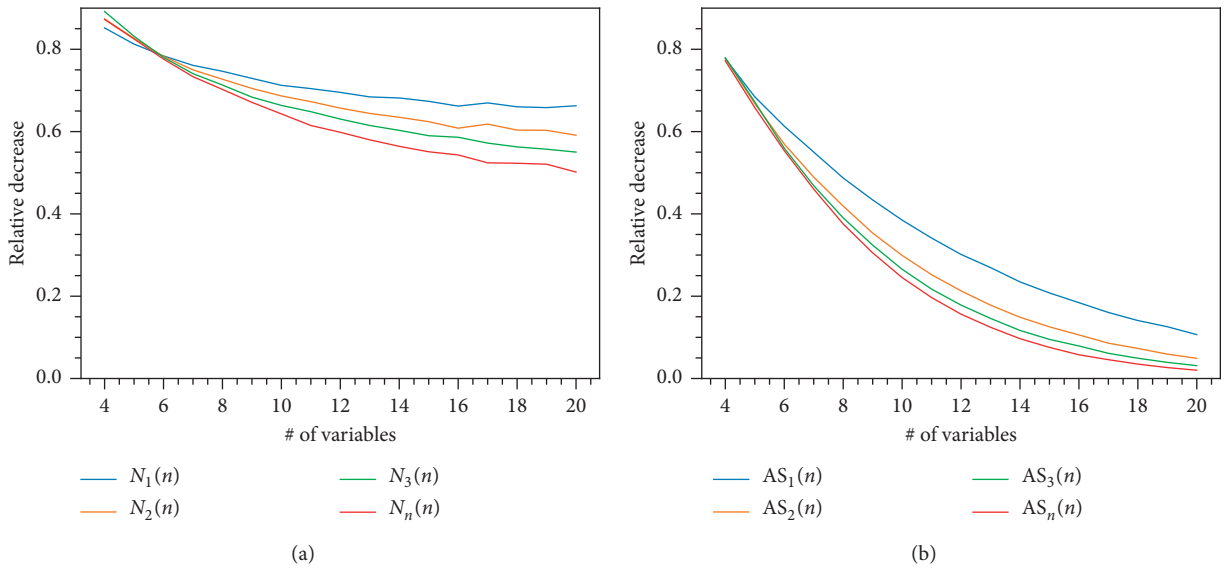


FIGURE 3: Dependence of the relative decreases of the sample means of $N(\mathbf{f})$ and $AS(\mathbf{f})$ on the number of variables n. (a) Relative decrease of the number of attractors. (b) Relative decrease of the average size of an attractor.

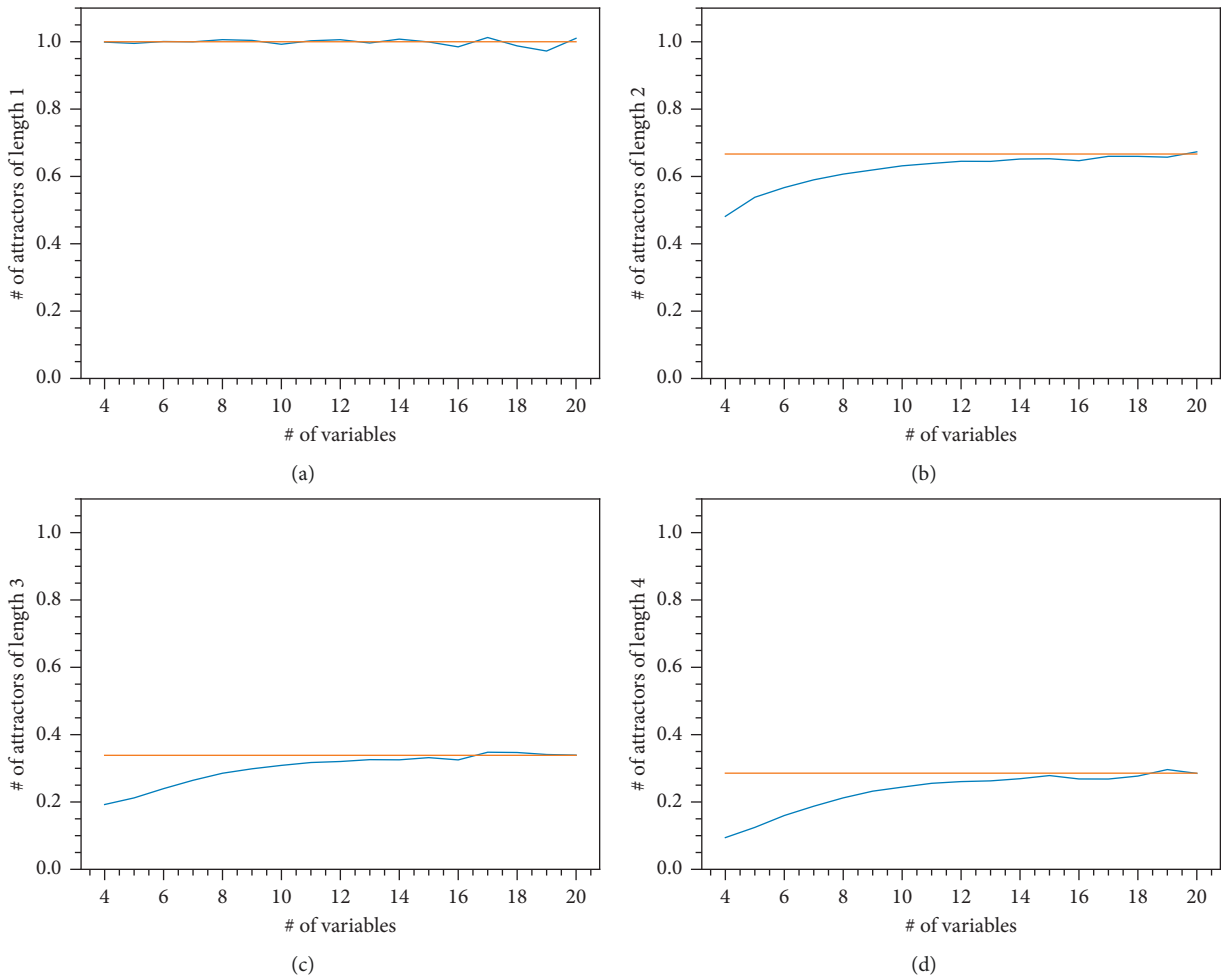


FIGURE 4: The average number of attractors of fixed length (blue plot) compared to the limiting value from Theorem 2 (orange plot). (a) Length 1. (b) Length 2. (c) Length 3. (d) Length 4.

In: Nonnegative integers k and n with $k \leq n$
Out: A Boolean function f in n variables of canalizing depth k such that, for fixed k and n, all possible outputs have the same probability
(1) In the notation of Theorem 1, generate the following:
(a) random bits $b, a_1, \ldots, a_n \in \{0, 1\}$;
(b) random subset $X \subset \{x_1, \ldots, x_n\}$ with $|X| = k$;
(c) random ordered partition $X = X_1 \sqcup \cdots \sqcup X_r$ of X (using Algorithm 2);
(d) random noncanalizing function $p_C \not\equiv 0$ in variables $\{x_1, \ldots, x_n\}/X$ (see Remark 3).
(2) Form a function $f(x_1, \ldots, x_n)$ using the data generated in Step 1 as in Theorem 1, where M_i involves exactly the variables from X_i for every $1 \leq i \leq r$.
(3) If f does not satisfy any of the conditions (E1) or (E2), discard it and run the algorithm again. Otherwise, return f.

Algorithm 2: Generating a random Boolean function of a given canalizing depth.

In: A finite set X with $|X| = k$
Out: An ordered partition $X = X_1 \sqcup \cdots \sqcup X_r$ into nonempty subsets $X_1, \ldots, X_r$ such that, for a fixed X, all possible outputs have the same probability
(1) Compute $p_0, \ldots, p_k$, where p_i is the number of ordered partitions of a set of size i, using the recurrence $p_j = \sum_{i=0}^{j-1} \binom{j}{i} p_{j-i}$, $p_0 = 1$ (see [19, equation (9)]).
(2) Generate an integer N uniformly at random from $[1, p_k]$.
(3) Find the minimum integer j between 1 and k such that $\sum_{i=0}^{j-1} \binom{k}{i} p_{k-i} \geq N$.
(4) Randomly select a subset $X_1 \subset X$ of size j.
(5) Generate an ordered partition $X_2 \sqcup \cdots \sqcup X_r$ of X/X_1 recursively.
(6) Return $X_1 \sqcup \cdots \sqcup X_r$.

Algorithm 3: Generating a random ordered partition of a given finite set.

6.2. Distributions of $N(\mathbf{f})$ and $AS(\mathbf{f})$. Figure 2 shows the empirical distributions of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ for $n = 12$ and $k = 0, 1, 3, 12$ based on 300,000 samples for each k. From the plot, we can make the following observations:

(i) The distributions become more concentrated and the peak shifts towards zero when k increases
(ii) The distributions for nonzero canalizing depths (especially for larger depths) are much closer to each other that to the distribution for zero canalizing depth. This agrees with the plots on Figure 1.

6.3. Relative Decreases. From Figure 1, we can observe that, for both $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$, the sample mean decreases rapidly for small canalizing depths. In order to understand how this decrease behaves for large n, we introduce

$$N_k(n) := \frac{\text{the sample mean of } N(f) \text{ for } n \text{ variables and canalizing depth } k}{\text{the sample mean of } N(f) \text{ for } n \text{ variables and canalizing depth } 0}. \tag{11}$$

$\mathrm{AS}_k(n)$ is defined analogously. Figure 3 plots $N_1(n)$, $N_2(n)$, $N_3(n)$, and $N_n(n)$ and $\mathrm{AS}_1(n)$, $\mathrm{AS}_2(n)$, $\mathrm{AS}_3(n)$, and $\mathrm{AS}_n(n)$ as functions of n. From the plots we see that

(i) The relative initial decrease from canalizing depth 0 to canalizing depth 1 becomes even more substantial when n increases
(ii) The relative decrease from canalizing depth 0 to canalizing depth 3 is already very close to the relative decrease from depth zero to the maximal depth (i.e., nested canalizing functions)

7. Theory: The Main Result

We will introduce notation needed to state the main theorem. Let us fix a positive integer ℓ. For a binary string $\alpha \in S := \{0, 1\}^\ell$, we define

(i) $|\alpha|$ denotes the number of ones
(ii) $\overline{\alpha}$ denotes component-wise negation
(iii) $s(\alpha)$ denotes a cyclic shift to the right

For binary strings $\alpha, \beta \in \{0, 1\}^\ell$, we define

$$f(\alpha,\beta) := \begin{cases} \frac{1}{2^{|\beta|}}, & \text{if } \alpha \vee \beta = \beta, \\ 0, & \text{otherwise,} \end{cases}$$

$$g(\alpha,\beta) := \frac{1}{4}\left(f(\alpha,\beta) + f(\overline{\alpha},\beta) + f(\alpha,\overline{\beta}) + f(\overline{\alpha},\overline{\beta})\right). \tag{12}$$

Then, we define a $2^\ell \times 2^\ell$ matrix G_ℓ by

$$(G_\ell)_{a,b} = g(a, s(b)), \tag{13}$$

where we interpret numbers $1 \le a$ and $b \le 2^\ell$ as binary sequences of length ℓ.

Theorem 2. *Let A_ℓ be the limit of the expected number of attractors of length ℓ in a random Boolean network of canalizing depth one (see Definition 6) when the number of variables n goes to infinity. Then,*

$$A_\ell = \frac{1}{\ell P'_{G_\ell}(1)}, \tag{14}$$

where P_{G_ℓ} is the characteristic polynomial of matrix G_ℓ introduced above. In particular, we have

$$\begin{aligned} A_1 &= 1, \\ A_2 &= \frac{2}{3} = 0.666\ldots, \\ A_3 &= \frac{64}{189} = 0.3386\ldots, \\ A_4 &= 0.2856\ldots, \\ A_5 &= 0.2004\ldots, \\ A_6 &= 0.1721\ldots. \end{aligned} \tag{15}$$

Remark 4. The plots below show that the result of Theorem 2 agrees with our simulations (Figure 4).

Remark 5. As explained in Remark A.1, Theorem 2 stills holds if we replace a random Boolean network of canalizing depth one with a random Boolean network defined by canalizing functions.

Example 4. Let $\ell = 2$. Then, for example, we have $f(0,2) = f(0,1) = 1/2$ and $g(0,1) = g(3,1) = 1/4$. In total, we have

$$G_2 = \begin{pmatrix} 3/8 & 1/4 & 1/4 & 3/8 \\ 1/8 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/8 \\ 3/8 & 1/4 & 1/4 & 3/8 \end{pmatrix}, \tag{16}$$

$$P_{G_2}(t) = t^4 - \frac{5}{4}t^3 + \frac{1}{2}t^2.$$

Remark 6. Theorem 2 and Corollary A.1 imply that $A_\ell > 1/\ell$ for every $\ell > 1$. On the other hand, a direct computation shows that the expected number of attractors of length ℓ in a random Boolean network (without any canalization requirements) is $1/\ell$. This is consistent with our observations from Section 6.1.

Remark 7. A sage script for computing numbers A_ℓ is available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/blob/master/core/theory.sage.

8. Conclusions

We conducted computational experiments to investigate the attractor structure of Boolean networks defined by functions of varying canalizing depth. We observed that networks with higher canalizing depth tend to have fewer attractors and the sizes of the attractors decrease dramatically when the canalizing depth increases moderately. As a consequence, the basins tend to grow when the canalizing depth increases. These properties are desirable in many biological applications of Boolean networks, so our results give new indications of the biological utility of Boolean networks defined by functions of positive canalizing depth.

We proved a theoretical result, Theorem 2, which complements the above observation as follows. The theorem implies that a large random Boolean network of canalizing depth one has on average more attractors of small size than a random Boolean network of the same size although it has less attractors in total. This also provides an explanation to the fact that the total size of attractors decreases faster than the number of attractors as the canalizing depth grows.

Furthermore, we observed that all the statistics we computed are almost the same in the case of the maximal possible canalizing depth (so-called nested canalizing Boolean networks) and in the case of moderate canalizing depth. This agrees with the results of Layne et al. [12]. This observation elucidates an interesting and powerful feature of canalization: even a very moderate canalizing influence in a Boolean network has a strong constraining influence on network dynamics. It would be of interest to explore the

prevalence of these features in published Boolean network models.

Finally, we provided evidence that the observed phenomena will occur for Boolean networks with larger numbers of state variables.

Appendix

A. Proofs

Notation A.1. We fix a positive integer ℓ.

(i) For every $1 \le i < j \le \ell$, we define a subset $S_{i,j} \subset S = \{0,1\}^{\ell}$ by

$$S_{i,j} := \left\{ (\alpha_1, \ldots, \alpha_\ell) \in S \,\middle|\, \alpha_i = \alpha_j \right\}. \tag{A.1}$$

(ii) For every $1 \le i < j \le \ell$, let $G_{\ell;i,j}$ be the submatrix of G_ℓ with rows and columns having indices from $S_{i,j}$.

Lemma A.1. *For every ℓ, we have*

(1) *G_ℓ^T is stochastic (see [20, § 8.5]), and G_ℓ has exactly one eigenvalue being equal to 1.*

(2) *For every $1 \le i < j \le \ell$, there exists a $2^{\ell-1} \times 2^{\ell-1}$ matrix $C_{\ell;i,j}$ with nonnegative entries such that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic and has exactly one of the eigenvalues being equal to 1.*

Proof. We will first show that G_ℓ^T is stochastic and irreducible (see [20, § 3.11]).

By definition, showing that G_ℓ^T is stochastic is equivalent to proving that, for every $\beta \in S := \{0,1\}^{\ell}$:

$$\sum_{\alpha \in S} g(\alpha, s(\beta)) = 1. \tag{A.2}$$

Since shift just permutes binary strings, this sum is equal to $\sum_{\beta \in S} g(\alpha, \beta)$. For a fixed β and $k \le |\beta|$, the number of $\alpha \in S$ such $\alpha \vee \beta = \beta$ and $|\alpha| = k$ is equal to $\binom{|\beta|}{k}$. Thus,

$$\sum_{\alpha \in S} h(\alpha, \beta) = \sum_{k=0}^{|\beta|} \binom{|\beta|}{k} \frac{1}{2^{|\beta|}} = 1 \Longrightarrow \sum_{\beta \in S} g(\alpha, \beta) = 1. \tag{A.3}$$

To prove irreducibility, we observe that, if $0 \in S$ denotes a zero binary string, then $g(\alpha, 0) \neq 0$ and $g(0, \alpha) \neq 0$ for every $\alpha \in S$. Then, [20, § 3.11, Exercise 12a] implies that G_ℓ^T is irreducible.

Since G_ℓ^T is stochastic, its largest eigenvalue is equal to 1 [20, § 8.5, p.156]. Since G_ℓ^T is irreducible, [20, Theorem 8.2] implies that 1 is a simple eigenvalue.

To prove the second part of the lemma, we fix $1 \le i < j \le \ell$. We will show that for every $\beta \in S_{i,j}$

$$\sum_{\alpha \in S_{i,j}} g(\alpha, s(\beta)) \le \frac{2^{\ell+2} - 1}{2^{\ell+2}}. \tag{A.4}$$

Indeed, let γ be a binary string with all zeroes and one at the ith position. Then, since $g(\gamma, s(\beta)) \ge (1/2^{|\beta|+2}) \ge (1/2^{\ell+2})$, we have

$$\sum_{\alpha \in S_{i,j}} g(\alpha, s(\beta)) \le \left(\sum_{\alpha \in S} g(\alpha, s(\beta)) \right) - g(\gamma, s(\beta)) \le 1 - \frac{1}{2^{\ell+2}}. \tag{A.5}$$

Inequality (A.4) implies that there exists a matrix $C_{\ell;i,j}$ with nonnegative entries such that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic.

Since $0 \in S_{i,j}$, the same argument as in the proof of the first part of the lemma shows that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic and has exactly one of the eigenvalues being equal to 1. □

Corollary A.1. *Let $P_\ell(t)$ be the charactersitic polynomial of G_ℓ. Then, for every $\ell > 1$, $|P'_\ell(1)| < 1$.*

Notation A.2. Fix a positive integer n. For vectors $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{Z}_{\ge 0}^n$ and $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{Z}_{\ge 0}^n$, we denote

$$\begin{aligned} \mathbf{a}! &:= a_1! \cdot \ldots \cdot a_n!, \\ \mathbf{a}^{\mathbf{b}} &:= a_1^{b_1} \cdot \ldots \cdot a_n^{b_n}, \\ |a| &:= a_1 + \cdots + a_n. \end{aligned} \tag{A.6}$$

Lemma A.2. *Let A be an $s \times s$ stochastic matrix with only one of the eigenvalues being one. We set*

$$C(A)_n := \sum_{\substack{\mathbf{m} \in \mathbb{Z}_{\ge 0}^s \\ |\mathbf{m}| = n}} \frac{(A\mathbf{m})^{\mathbf{m}}}{n^n}. \tag{A.7}$$

Let $P_A(t)$ be the characteristic polynomial of A. Then, $\lim_{n \longrightarrow \infty} C(A)_n = 1/P'_A(1)$.

Proof. We recall that the Lambert W function [21] is the principal branch of the inverse of xe^x. We will use the notation $y(z) = -W(-z)$ from [22] so that $y(z) = ze^{y(z)}$. Function $y(z)$ has a singularity of the square-root type at $z = 1/e$ and has the following expansion around this point (see [22, p. 107]):

$$y(z) = 1 - \varepsilon + \frac{1}{3}\varepsilon^2 - \cdots, \quad \text{where } \varepsilon = \sqrt{2 - 2ez}. \tag{A.8}$$

From this, we obtain

$$\frac{1}{y(z)} = 1 + \varepsilon + \frac{2}{3}\varepsilon^2 - \cdots, \quad \text{where } \varepsilon = \sqrt{2 - 2ez}. \tag{A.9}$$

The main result of [23] implies that, for every complex $s \times s$ matrix A, we have

$$\sum_{\mathbf{m} \in \mathbb{Z}_{\ge 0}^s} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} x^{|\mathbf{m}|} \exp\left(-x \sum_{i,j} m_j a_{i,j} \right) = \frac{1}{\det|E - xA|}. \tag{A.10}$$

Since A^T is stochastic, we have $\sum_{i=1}^{n} a_{i,j} = 1$, so

$$\sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} x^{|\mathbf{m}|} e^{-x|\mathbf{m}|} = \frac{1}{\det|E - xA|}. \tag{A.11}$$

If we perform a substitution $x = y(z)$ and use the definition of the Lambert W function, we obtain

$$\sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} z^{|\mathbf{m}|} = \frac{1}{\det|E - y(z)A|}. \tag{A.12}$$

From this, we obtain

$$\sum_{n=0}^{\infty} \frac{n^n C(A)_n}{n!} z^n = \sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} z^{|\mathbf{m}|} = \frac{1}{\det|E - y(z)A|} =: F(z). \tag{A.13}$$

$F(z)$ can be rewritten as

$$F(z) = \frac{1}{y(z)^s P_A(1/y(z))}. \tag{A.14}$$

Finding the asymptotic behavior of the Taylor coefficients of $F(z)$ would yield an asymptotic for $C(A)_n$. We will do this using singularity analysis [24, Chapter VI] (similarly to [22, Theorem 2]). Since $|y(z)| < 1$ for $|z| < 1/e$ (see [21, Figure 1]) and all roots of P_A lie in the unit circle due to the stochasticity of A, $1/e$ is the singularity of $F(z)$ with the smallest absolute value. Due to Lemma A.1, $P_A(t) = (1-t)Q_A(t)$, where $Q_A(1) \neq 0$. Using (A.8), we obtain the following expansion of $F(z)$ around $1/e$:

$$\begin{aligned} F(z) &= \frac{1}{(1-\varepsilon+\cdots)^s(-\varepsilon - 2/3\varepsilon^2 + \cdots)Q_A(1+\varepsilon+\cdots)} \\ &= \frac{-1}{Q_A(1)}\left(\frac{1}{\varepsilon} + \cdots\right), \text{ where } \varepsilon = \sqrt{2-2ez}. \end{aligned} \tag{A.15}$$

Singularity analysis [24, Corollary VI.1] implies that

$$\frac{n^n C(A)_n}{n!} \sim \frac{-e^n}{Q_A(1)\sqrt{2\pi n}}, \quad \text{as } n \longrightarrow \infty. \tag{A.16}$$

Using Stirling's formula, we get

$$C(A)_n \sim \frac{-n!e^n}{n^n Q_A(1)\sqrt{2\pi n}} \sim \frac{-1}{Q_A(1)}, \quad \text{as } n \longrightarrow \infty. \tag{A.17}$$

Using $P'_A = -Q'_A + (1-t)Q_\ell$, we deduce $P'_A(1) = -Q'_A(1)$, and this finishes the proof. □

Lemma A.3. *On the set of all Boolean networks with n states consider two probability distributions:*

(*A*) *All the networks with canalizing depth one have the same probability, and all others have probability zero*

(*B*) *ie probability assigned to each network is proportional to the product of the number of canalizing variables of the functions defining this network*

We fix a positive integer ℓ. By $A_{\ell,n}$ and $B_{\ell,n}$ we denote the average number of attractors of length ℓ in a random Boolean network with n states with respect to distributions (A) and (B), respectively. Then,

$$\lim_{n\longrightarrow\infty} A_{\ell,n} = \lim_{n\longrightarrow\infty} B_{\ell,n}. \tag{A.18}$$

Example A.1. We will illustrate the (*B*) distribution by an example. Consider the following three networks with two states:

$$\begin{aligned} \mathbf{f}_1 &= (x_1x_2 + 1, x_1 + x_2), \\ \mathbf{f}_2 &= (x_1x_2, x_1), \\ \mathbf{f}_3 &= (x_1x_2 + 1, x_1x_2). \end{aligned} \tag{A.19}$$

Since the canalizing depth of $x_1 + x_2$ is zero, $P_B(\mathbf{f}_1)$, the probability of $\mathbf{f}_1$ with respect to B, is zero. Since the canalizing depths of x_1x_2 and x_1 are 2 and 1, respectively, the ratio $P_B(\mathbf{f}_2)/P_B(\mathbf{f}_3)$ is equal to $(2 \cdot 1/2 \cdot 2) = (1/2)$.

Proof. Let F_n and F_n^* be the number of Boolean functions in n variables with canalizing depth exactly one and more than one, respectively. We will use the following bounds:

(1) $F_n^* \leq n^2 \cdot 4 \cdot 4 \cdot 2^{2^{n-2}}$: we look term-by-term. There are at most n^2 ways to choose first and second canalizing variables. There are at most 4 choices for the canalizing outputs and at most 4 choices for canalizing values for these two variables. There are at most $2^{2^{n-2}}$ core functions, since that is all possible functions, which may or may not be canalizing. Since redundant arrangements of canalizing variables are not accounted for, this must overcount.

(2) $F_n \geq 2^{2^{n-1}} - (n-1) \cdot 2 \cdot 2 \cdot 2^{2^{n-2}}$: this is a lower bound for the number of noncanalizing core function in $n-1$ variables because $(n-1) \cdot 2 \cdot 2 \cdot 2^{2^{n-2}}$ is an upper bound on the number of canalizing functions in $n-1$ variables (obtained in the same way as the bound above).

We also introduce

$$R_n := \frac{F_n^*}{F_n} \leq \frac{16n^2 2^{2^{n-2}}}{2^{2^{n-1}} - 4(n-1)2^{2^{n-2}}} = \frac{n^2}{2^{(2^{n-2})-4} - (1/4)(n-1)}. \tag{A.20}$$

For X being (*A*) or (*B*) and positive integer n, let $P_{X,n}$ denote the probability (it is always the same) of choosing a network from distribution X with all functions being of depth exactly one. Let P_n^* be the maximal probability of choosing a network from (*B*) with at least one function being of depth more than one, respectively. By S_n and S_n^* we denote the total number of attractors of length ℓ in networks with all functions being of depth exactly one and with at least one function being of depth more than one, respectively.

The statement of the lemma is equivalent to the statement that

$$\lim_{n\longrightarrow\infty} \left(A_{\ell,n} - B_{\ell,n}\right) = 0. \tag{A.21}$$

Using the notation introduced above, we can bound $A_{\ell,n} - B_{\ell,n}$ as

$$P_{n,A}S_n - P_{n,B}S_n - P_n^*S_n^* \le \left|A_{\ell,n} - B_{\ell,n}\right| \le P_{n,A}S_n + P_{n,B}S_n. \tag{A.22}$$

We set $U_n := S_n(P_{n,A} - P_{n,B})$ and $V_n := P_n^*S_n^*$. Then, (A.21) would follow from $\lim_{n\longrightarrow\infty}U_n = 0$ and $\lim_{n\longrightarrow\infty}V_n = 0$, so we will prove these two equalities.

Since any network has at most 2^n attractors of length ℓ, $S_n \le 2^nF_n^n$. Since the total sum of the products of canalizing depths over all Boolean networks does not exceed $(F_n + nF_n^*)^n$, we have $P_{n,B} \ge (1/(F_n + nF_n^*)^n)$. Since $P_{n,A} = 1/F_n^n$, we have

$$U_n \le 2^nF_n^n\left(\frac{1}{F_n^n} - \frac{1}{(F_n + nF_n^*)^n}\right) = 2^n\left(1 - \frac{1}{(1+nR_n)^n}\right)$$

$$= 2^n\frac{\binom{n}{1}nR_n + \binom{n}{2}(nR_n)^2 + \cdots + (nR_n)^n}{(1+nR_n)^n}. \tag{A.23}$$

(A.20) implies that $nR_n < 1$ for large enough n. Hence, for large enough n, we have

$$U_n \le 2^nnR_n\frac{2^n}{(1+nR_n)^n} \le 4^nnR_n \le \frac{4^nn^3}{2^{(2^{n-2})-4} - (1/4)(n-1)} \longrightarrow 0. \tag{A.24}$$

By similar arguments, $P_n^* \le n^n/F_n^n$ and $S_n^* \le 2^nn(F_n + F_n^*)^{n-1}F_n^*$, so

$$V_n \le 2^nn^{n+1}(F_n + F_n^*)^{n-1}F_n^*\frac{1}{F_n^n} \le 2^nn^{n+1}(1+R_n)^{n-1}R_n. \tag{A.25}$$

Since $R_n < 1$ for large enough n, using (A.20), we have

$$V_n \le 2^{2n-1}n^{n+1}R_n \le \frac{2^{2n-1}n^{n+3}}{2^{(2^{n-2})-4} - (1/4)(n-1)} \longrightarrow 0. \tag{A.26}$$

□

Remark 8. The proof of Lemma A.3 will be valid if we replace distribution (*B*) with any other distribution (*C*) such that, for every Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$

(i) If at least one of f_i's is noncanalizing, $P_C(\mathbf{f}) = 0$

(ii) There exists a constant $P_{n,C}$ such that, if the canalizing depth of every f_i is one, then $P_C(\mathbf{f}) = P_{n,C}$

(iii) We have $(P_C(\mathbf{f})/P_{n,C}) \le (P_B(\mathbf{f})/P_{n,B})$ (using notation from the proof of Lemma A.3)

The above properties hold, for example, for the following distribution.

(C) All the networks defined by canalizing functions have the same probability, and all others have probability zero.

Using this distribution instead of (*A*), we see that Theorem 2 holds also for a random Boolean network defined by canalizing functions.

Lemma A.4. *We will use Notation A.1 and notation from Lemma A.2. Then, for every positive integers ℓ and n, we have*

$$C(G_\ell)_n - \sum_{1\le i<j\le\ell} C\left(G_{\ell;i,j}\right)_n \le \ell B_{\ell,n} \le C(G_\ell)_n. \tag{A.27}$$

Proof. We fix n. Consider a tuple $\mathbf{X} = (X_1, \ldots, X_\ell)$ of ℓ distinct elements of $\{0,1\}^n$. For $1 \le i \le n$, we denote $\mathbf{X}_i := (X_{1,i}, \ldots, X_{n,i})$. For $\alpha \in S$, let

$$n_\alpha := \left|\left\{i \mid 1 \le i \le n, X_i = \alpha\right\}\right|. \tag{A.28}$$

Then, $\sum_{\alpha\in S}n_\alpha = n$. First, we will show that

$$P(X_1, \ldots, X_\ell \text{ form an attractor in this order})$$

$$= \prod_{\alpha\in S}\left(\sum_{\beta\in S}g(\alpha, s(\beta))\frac{n_\beta}{n}\right)^{n_\alpha} = \frac{(G_\ell\mathbf{n})^{\mathbf{n}}}{n^n}, \tag{A.29}$$

where $\mathbf{n} = (n_0, n_1, \ldots, n_{2^\ell-1})$.

To prove (A.29), we will use that the functions f_i ($i = 1, \ldots, n$) in the network are chosen independently to decompose the left-hand side as

$$P(X_1, \ldots, X_\ell \text{ form an attractor in this order})$$

$$= \prod_{i=1}^{n}P\left(f_i\left(X_j\right) = X_{j+1,i} \text{ for every } 1 \le j \le n\right), \tag{A.30}$$

where we use notation $X_{n+1} = X_1$ and the probability of each Boolean function to be chosen is assumed to be proportional to the number of its canalizing variables. We show that, for every $1 \le i \le n$,

$$P\left(f_i\left(X_j\right) = X_{j+1,i} \quad \text{for every } 1 \le j \le n\right) = \sum_{\beta\in S}g(\mathbf{X}_i, s(\beta))\frac{n_\beta}{n}. \tag{A.31}$$

Then, (A.29) would follow from multiplying (A.31) for all *i*. To prove (A.31), without loss of generality, we consider $i = 1$. Consider a set

$$\Omega = \left\{(f,k) \mid f : \{0,1\}^n \longrightarrow \{0,1\}, 1 \le k \le n, x_k \text{ is canalizing for } f\right\}, \tag{A.32}$$

with a uniform probability distribution P_Ω. Observe that for a function f with canalizing variables $x_{k_1}, \ldots, x_{k_s}$, we have

$$P(f) = P_\Omega((f,k_1)) + \cdots + P_\Omega((f,k_s)). \tag{A.33}$$

If we can show that, for every $1 \le k \le n$,

$$P_\Omega\left(f\left(X_j\right) = X_{j+1,1} \quad \text{for every } 1 \le j \le n \mid (f,k) \in \Omega\right) = g(\mathbf{X}_1, s(\mathbf{X}_k)), \tag{A.34}$$

then (A.31) would follow by summing up (A.34) over all k and using the law of total probability.

We consider one of the canalizing variables of f, say, x_k. Let c be the canalizing value of x_{k_1}, and let v be the value taken by f when $x_{k_1} = c$. Then, $(c, v) \in \{0,1\}^2$, and all these

four cases have the same probability due to the symmetry. As $g(\alpha, s(\beta)) = (1/4)(h(\alpha,\beta) + h(\overline{\alpha},\beta) + h(\alpha,\overline{\beta}) + h(\overline{\alpha},\overline{\beta}))$, it is sufficient to show that

$$P_\Omega\left(f\left(X_j\right) = X_{j+1,1} \quad \text{for every } 1 \le j \le n \mid (f,k) \in \Omega \text{ and } c = v = 0\right) = h\left(\mathbf{X}_1, s\left(\mathbf{X}_k\right)\right), \tag{A.35}$$

and then sum for all $(c, v) \in \{0,1\}^2$.

To prove (A.35), consider any j, say $j = 1$. There are then 4 cases for the values of $X_{1,k}$ and $X_{2,1}$:

(1) $X_{1,k} = 1$ and $X_{2,1}$ is 0 or 1. With probability 1/2, we have $f(X_1) = X_{2,1}$. This is true due to symmetry, as for any f_1 which takes on the value w at X_1, we can produce another function g that is equal to 0 if $X_{1,k} = 0$ and $\overline{f}_1$ if $X_{1,k} = 1$. Then, $g(X_1) = \overline{w}$.

(2) $X_{1,k} = 0$ and $X_{2,1} = 1$. Since $X_{1,k} = c$, the probability of $f(X_1) = X_{2,1} \neq v = 0$ is zero.

(3) $X_{1,k} = X_{2,1} = 0$. Since $X_{1,k} = c$ and $X_{2,1} = v$, the canalization property implies that $f(X_1) = X_{2,1}$ with probability one.

The only case in which $\mathbf{X}_1 \vee s(\mathbf{X}_k) \neq s(\mathbf{X}_k)$ is where there is at least one j such that case 2 is realized. In this case, the probability in the left-hand side of (A.35) will be zero. Otherwise, each occurrence of case 1 will multiply the total probability by 1/2 and each occurrence of case 3 will multiply the total probability by 1. Thus, we show that the left-hand side of (A.35) is indeed equal to $h(\mathbf{X}_1, s(\mathbf{X}_k))$. This finishes the proof of (A.29).

To finish the proof of the lemma, we set

$$U := \left\{ \mathbf{n} \in \mathbb{Z}_{\ge 0}^S \,\middle|\, \sum_{\alpha \in S} n_\alpha = n \;\&\; \text{the support of } \mathbf{n} \text{ does not belong to } \bigcup_{1 \le i < j \le \ell} S_{i,j} \right\}. \tag{A.36}$$

Summing (A.29) over all ℓ-tuples $(X_1, \ldots, X_\ell)$ of distinct elements of $\{0,1\}^n$, we obtain (see (A.7))

$$\ell B_{\ell,n} = \sum_{\mathbf{n} \in U} \frac{(G_\ell \mathbf{n})^{\mathbf{n}}}{n^n} \le C(G_\ell)_n. \tag{A.37}$$

On the other hand, if $\mathbf{n}$ is supported on one some $S_{i,j}$, then $G_\ell \mathbf{n} = G_{\ell;i,j}\mathbf{n}|_{S_{i,j}}$, where $\mathbf{n}|_{S_{i,j}}$ denotes the restriction of $\mathbf{n}$ on the coordinates from $S_{i,j}$. This implies that

$$C(G_\ell)_n - \ell B_{\ell,n} \le \sum_{1 \le i < j \le \ell} C\left(G_{\ell;i,j}\right)_n. \tag{A.38}$$

This finishes the proof of the lemma. □

Proof of Theorem 2. We fix positive integer ℓ. In the notation of Lemma A.3, we have $A_\ell = \lim_{n \longrightarrow \infty} A_{\ell,n}$. Lemma A.3 implies that $A_\ell = \lim_{n \longrightarrow \infty} B_{\ell,n}$. We fix any $1 \le i < j \le \ell$, and let $C_{\ell;i,j}$ be the matrix given by Lemma A.1. We set $M := (2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})$. Then,

$$0 \le C\left(G_{\ell;i,j}\right)_n \le C\left(G_{\ell;i,j} + C_{\ell;i,j}\right)_n = \left(\frac{2^{\ell+2} - 1}{2^{\ell+2}}\right)^n C(M)_n. \tag{A.39}$$

Lemma A.2 implies that $\lim_{n \longrightarrow \infty} C(M)_n$ is finite, thus we have that $\lim_{n \longrightarrow \infty} C(G_{\ell;i,j})_n = 0$. We finish the proof of the theorem by considering the limit of (A.27) and applying Lemma A.2 to G_ℓ. □

B. Complexity analysis

Proposition B.1. *Complexity of Algorithm 3 is $O(k^3)$.*

Proof. First, we show the complexity of a single run the algorithm, i.e., not taking into account the recursive call, is $O(k^2)$. Since the first k rows of the Pascal's triangle can be precomputed in $O(k^2)$, the complexity of step 1 is also $O(k^2)$. Similarly, the complexity of step 3 is $O(k^2)$. It remains to observe that step 2 takes $O(1)$ and step 4 takes $O(k^2)$ (indeed, selecting a subset of size j amounts to selecting and removing j indices). In total, we obtain $O(k^2)$.

The depth of the recursion calls is at most k. Since the complexity of each single call is $O(k^2)$, so the total complexity is $O(k^3)$. □

Lemma B.1. *The average complexity of the algorithm for generating a function in $n > 0$ variables which is either 1 or noncanalizing described in Remark 3 is $O(n2^n)$.*

Proof. [25, p. 116] implies that the proportion of functions which are canalizing in n variables is bounded from above by $4n/2^{2^{n-1}}$. Note that [25] considers constant functions canalizing which we do not. Thus, the probability P_n of choosing a function which is either 1 or noncanalizing is bounded from above by

$$\frac{4n}{2^{2^{n-1}}} - \frac{1}{2^{2^n}} = \frac{4n - \left(1/2^{2^{n-1}}\right)}{2^{2^{n-1}}}. \tag{B.1}$$

This bound is less than 3/4 for all values of n except 1 and 2, but we can compute directly that $P_1 = 3/4$ and $P_2 = 13/16$. Therefore, the number of times the generation of a function needs to be repeated averages to $1/1 - P_n$, which does not exceed 4, so the average complexity of the whole procedure is the same as of a single generation step.

The complexity of a single step consists of generating a random function (which is $O(2^n)$) and checking whether it is canalizing or not. We perform this check by running linearly through the table for each variable, so the complexity is $O(n2^n)$ time. Thus, the total complexity is indeed $O(n2^n)$. □

Lemma B.2. *There is a constant $c < 1$ such that the probability that a function generated in steps 1 and 2 of Algorithm 2 does not satisfy one of the conditions (E1) or (E2) is bounded by c for every n.*

Proof. Notice that

$$P((E1) \text{ or } (E2) \text{ is false}) = P(r \neq 1)P((E1) \text{ is false} \mid r \neq 1) + P(r = 1)P((E2) \text{ is false} \mid r = 1). \tag{B.2}$$

We will show that there is a constant $c < 1$ such that $P((E1) \text{ is false} \mid r \neq 1)$ and $P(5.1 \text{ is false} \mid r = 1)$ do not exceed c.

(i) $P((E1) \text{ is false} \mid r \neq 1)$: the probability of having $k_r = 1$ (the only possible $k_r < 2$) is just the proportion of ordered partitions with a single element at the end. We can construct all of these by picking an element and then picking a partition of the remaining elements, so this creates $k \cdot p_{k-1}$ possibilities. Thus, the probability of this occurring is kp_{k-1}/p_k. [19, equation (5)] implies that this approaches $\ln(2) < 1$ as n goes to infinity. Thus, there exists such c.

(ii) $P((E2) \text{ is false} \mid r = 1)$: the probability of ever picking $b = 1$ is just 1/2, so we can take $c = 1/2$. □

Proposition B.2. *Complexity of Algorithm 2 is $O(n2^n)$.*

Proof. Lemma B.2 implies that the average number of reruns in step 3 is constant. Thus, the complexity of the algorithm is the same as of a single run.

Proposition B.1 and Lemma B.1 imply that the complexity of step 1 is $O(k^3 + (n-k)2^{n-k})$. Step 2 generates a truth table for the function. There are 2^n input-output pairs, and computing the function takes at most k steps, so this is $O(k2^n)$. In step 3, the conditions (E1) or (E2) are verified in $O(2^n)$ time.

Summing everything, we obtain $O(k^3 + (n-k)2^{n-k} + k2^n) = O(n2^n)$ □

Data Availability

Python/sage code and the results of simulations used to support the findings of this study have been deposited at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful to Claus Kadelka, Christian Kratenthaler, and Doron Zeilberger for helpful discussions. GP was partially supported by NSF grants CCF-1564132, CCF-1563942, DMS-1760448, DMS-1853482, and DMS-1853650 by PSC-CUNY grants #69827-0047 and #60098-0048. RL was partially supported by Grants NIH 1U01EB024501-01 and NSF CBET-1750183. EP, GP, and WQ are grateful to the New York Math Circle, where their collaboration started.

References

[1] M. I. Davidich and S. Bornholdt, "Boolean network model predicts cell cycle sequence of fission yeast," *PLoS One*, vol. 3, no. 2, Article ID e1672, 2008.

[2] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14796–14799, 2003.

[3] Z. Mai and H. Liu, "Boolean network-based analysis of the apoptosis network: irreversible apoptosis and stable surviving," *Journal of Theoretical Biology*, vol. 259, no. 4, pp. 760–769, 2009.

[4] J. Saez-Rodriguez, L. Simeoni, J. A. Lindquist et al., "A logical model provides insights into T cell receptor signaling," *PLoS Computational Biology*, vol. 3, no. 8, pp. 1–11, 2007.

[5] A. Veliz-Cuba and B. Stigler, "Boolean models can explain bistability in the lac operon," *Journal of Computational Biology*, vol. 18, no. 6, pp. 783–794, 2011.

[6] S. A. Kauffman, "Sequential DNA replication and the control of differences in gene activity between sister chromatids—a possible factor in cell differentiation," *Journal of Theoretical Biology*, vol. 17, no. 3, pp. 483–497, 1967.

[7] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437–467, 1969.

[8] S. Kauffman, "Homeostasis and differentiation in random genetic control networks," *Nature*, vol. 224, no. 5215, pp. 177-178, 1969.

[9] C. H. Waddington, "Canalization of development and the inheritance of acquired characters," *Nature*, vol. 150, no. 3811, pp. 563–565, 1942.

[10] F. Karlsson and M. Hörnquist, "Order or chaos in Boolean gene networks depends on the mean fraction of canalizing functions," *Physica A: Statistical Mechanics and Its Applications*, vol. 384, no. 2, pp. 747–757, 2007.

[11] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Genetic networks with canalyzing Boolean rules are always stable," *Proceedings of the National Academy of Sciences*, vol. 101, no. 49, pp. 17102–17107, 2004.

[12] L. Layne, E. Dimitrova, and M. Macauley, "Nested canalyzing depth and network stability," *Bulletin of Mathematical Biology*, vol. 74, no. 2, pp. 422–433, 2012.

[13] S. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, UK, 1993.

[14] A. Veliz-Cuba, B. Aguilar, F. Hinkelmann, and R. Laubenbacher, "Steady state analysis of Boolean molecular network models via model reduction and computational algebra," *BMC Bioinformatics*, vol. 15, no. 221, 2014.

[15] Q. He and M. Macauley, "Stratification and enumeration of boolean functions by canalizing depth," *Physica D: Nonlinear Phenomena*, vol. 314, no. 1–8, 2016.

[16] B. Samuelsson and C. Troein, "Superpolynomial growth in the number of attractors in kauffman networks," *Physical Review Letters*, vol. 90, no. 9, Article ID 098701, 2003.

[17] B. Drossel, "Number of attractors in random boolean networks," *Physical Review E*, vol. 72, no. 1, Article ID 016110, 2005.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, 3 edition, 2009.
[19] O. A. Gross, "Preferential arrangements," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 4–8, 1962.
[20] D. Serre, *Matrices: Theory and Applications*, Springer, New York, NY, USA, 2 edition, 2010.
[21] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
[22] P. Flajolet, P. J. Grabner, P. Kirschenhofer, and H. Prodinger, "On Ramanujan's Q-function," *Journal of Computational and Applied Mathematics*, vol. 58, no. 1, pp. 103–116, 1995.
[23] L. Carlitz, "An application of MacMahon's Master theorem," *SIAM Journal on Applied Mathematics*, vol. 26, no. 2, pp. 431–436, 1974.
[24] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, New York, NY, USA, 1 edition, 2009.
[25] C. C. Walker and A. E. Gelfand, "A system theoretic approach to the management of complex organizations: management by exception, priority, and input span in a class of fixed-structure models," *Behavioral Science*, vol. 24, no. 2, pp. 112–120, 1979.

Hindawi
Complexity
Volume 2020, Article ID 1947809, 7 pages
https://doi.org/10.1155/2020/1947809

Research Article

Global Attractivity for Lasota–Wazewska-Type System with Patch Structure and Multiple Time-Varying Delays

Zhiwen Long [1,2] **and Yanxiang Tan** [3]

[1]*School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi, Hunan 417000, China*
[2]*School of Mathematics Science, Changsha Normal University, Changsha, Hunan 410100, China*
[3]*School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha, Hunan 410114, China*

Correspondence should be addressed to Zhiwen Long; longzw2005@126.com

Received 20 June 2019; Revised 12 August 2019; Accepted 26 August 2019; Published 17 January 2020

Academic Editor: Constantin Udriste

This paper aims to study the asymptotic behavior of Lasota–Wazewska-type system with patch structure and multiple time-varying delays. Based on the fluctuation lemma and some differential inequality techniques, we prove that the positive equilibrium is a global attractor of the addressed system with small time delay. Finally, we provide an example to illustrate the feasibility of the theoretical results.

1. Introduction

In 1988, in order to describe the survival of red blood cells in animals, Wazewska–Czyzewska and Lasota in [1] presented the following delayed differential equation model

$$x'(t) = -\delta x(t) + \sum_{j=1}^{m} p_j e^{-a_j x\left(t-\tau_j(t)\right)}, \tag{1}$$

where $x(t)$ represents the number of red blood cells at time t, δ denotes the death rate of red blood cells, p_j and a_j are related to the production of red blood cells per unit time, $\tau_j(t)$ represents the time required to produce a red blood cell. Since the model was proposed, there have been a large number of results about the dynamical behaviors for (1) and its modifications (see [2–7] and the references therein) due to their comprehensive practical application background.

As pointed out by Yao in [8], populations usually spread between different patches for survival and development. Recently, many scholars have paid attention to the population models with patch structure and time delays (see [9–16]). As far as we know, fewer works have been done concerning with the effect of time delay on dynamical behaviors of Lasota–Wazewska-type model with patch structure. The purpose of the present paper is to establish some sufficient conditions to guarantee the global attractivity of the following Lasota–Wazewska-type delay system with patch structure

$$\begin{aligned} x_i'(t) &= -\delta_i x_i(t) + \sum_{j=1, j\neq i}^{n} d_{ij} x_j(t) \\ &\quad + \sum_{j=1}^{m} p_{ij} e^{-a_{ij} x_i\left(t-\tau_{ij}(t)\right)}, \quad i \in J = \{1, 2, \ldots, n\}, \end{aligned} \tag{2}$$

where x_i denotes the number of species in the patch i, generation delay function $\tau_{ij} : \mathbb{R} \to (0, +\infty)$ is bounded and continuous, d_{ij} represents the dispersal coefficient of the species from patch j to patch i, δ_i, p_{ij}, and a_{ij} are all positive. In what follows, we always assume that

$$\begin{aligned} &a_i = \max_{1\le j\le m} a_{ij} > 0, \ r_i = \max_{1\le j\le m} \sup_{t\in\mathbb{R}} \tau_{ij}(t) > 0, \ \tau = \max_{1\le i\le n} r_i, \\ &0 < \delta_i - \sum_{j=1, j\neq i}^{n} d_{ij} < \sum_{j=1}^{m} p_{ij} \quad \text{for all } i \in J. \end{aligned} \tag{3}$$

For convenience, let $C = \prod_{i=1}^{n} C([-r_i, 0], \mathbb{R})$ and $C_+ = \prod_{i=1}^{n} C([-r_i, 0], [0, +\infty))$. If $x_i(t)$ is defined on $[t_0 - r_i, \sigma)$ with $t_0, \sigma \in \mathbb{R}$ and $i \in J$, then we write $x_t \in C$ as $x_t = \left(x_t^1, x_t^2, \ldots, x_t^n\right)$ where $x_t^i(\theta) = x_i(t+\theta)$ for all $\theta \in [-r_i, 0]$ and $t_0 \le t < \sigma$ and $i \in J$. Denote $x_t(t_0, \varphi)(x(t; t_0, \varphi))$ as an

admissible solution of (2) with the following admissible initial condition:

$$x_{t_0} = \varphi, \quad \varphi \in C_+, \quad \varphi_i(0) > 0, \quad i \in J. \tag{4}$$

Also, let $[t_0, \eta(\varphi))$ be the maximal right-interval of existence of $x_t(t_0, \varphi)$.

In the following, we further assume that there exists at least one positive constant $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ such that x^* is the positive equilibrium point of (2) satisfying

$$-\delta_i x_i^* + \sum_{j=1, j\neq i}^{n} d_{ij} x_j^* + \sum_{j=1}^{m} p_{ij} e^{-a_{ij} x_i^*} = 0, \quad \text{for all } i \in J. \tag{5}$$

2. Global Attractivity of the Positive Equilibrium Point $(x_1^*, x_2^*, \ldots, x_n^*)$

First, we will discuss the properties of the solution $x(t; t_0, \varphi)$ of the system (2) with (4).

Lemma 1. *$x(t; t_0, \varphi)$ is positive and bounded on $[t_0, \eta(\varphi))$, and $\eta(\varphi) = +\infty$. Moreover, $\liminf_{t\to+\infty} x_j(t; t_0, \varphi) > 0$, $j \in J$.*

Proof. For simplicity, we denote $x(t; t_0, \varphi)$ by $x(t)$. We first prove that

$$x_i(t) > 0 \quad \text{for all } t \in [t_0, \eta(\varphi)),\ i = 1, 2, \ldots, n. \tag{6}$$

Assume by contradiction that there exist $t^* \in [t_0, \eta(\varphi))$ and $k \in J$ such that

$$x_k(t^*) = 0, x_i(t) > 0 \quad \text{for all } t \in [t_0, t^*),\ i = 1, 2, \ldots, n. \tag{7}$$

It follows from (2) that

$$\begin{aligned} 0 &\geq D^- x_k(t^*) \\ &= -\delta_k x_k(t^*) + \sum_{j=1, j\neq k}^{n} d_{kj} x_j(t^*) + \sum_{j=1}^{m} p_{kj} e^{-a_{kj} x_k(t^* - \tau_{kj}(t^*))} \\ &> 0. \end{aligned} \tag{8}$$

This contradction implies that (6) holds. Furthermore, define

$$M(t) = \max\left\{\xi : \xi \leq t | \text{there exists } j_0 \in J \text{ such that } x_{j_0}(\xi) = \max_{t_0\leq s\leq t} x_{j_0}(s) = \max_{j\in J}\left\{\max_{t_0\leq s\leq t} x_j(s)\right\}\right\}. \tag{9}$$

We claim that $x_i(t)$ is bounded on $[t_0, \eta(\varphi)), i = 1, 2, \ldots, n$. Otherwise, we have $M(t) \to \eta(\varphi)$ as $t \to \eta(\varphi)$. Moreover, we can choose $k \in J$ and $\{l_p\}_{p\geq 1}$ with $\lim_{p\to+\infty} l_p = \eta(\varphi)$ such that

$$x_k\left(M\left(l_p\right)\right) = \max_{t_0\leq s\leq l_p} x_k(s) = \max_{j\in J}\left\{\max_{t_0\leq s\leq l_p} x_j(s)\right\} \tag{10}$$

and

$$\lim_{p\to+\infty} x_k\left(M\left(l_p\right)\right) = +\infty. \tag{11}$$

According to the definition of $M(l_p)$,

$$0 \leq D_- x_k\left(M\left(l_p\right)\right) = -\delta_k x_k\left(M\left(l_p\right)\right) + \sum_{j=1, j\neq k}^{n} d_{kj} x_j\left(M\left(l_p\right)\right) + \sum_{j=1}^{m} p_{kj} e^{-a_{kj} x_k\left(M\left(l_p\right) - \tau_{kj}\left(M\left(l_p\right)\right)\right)}. \tag{12}$$

Furthermore, we have

$$\delta_k \leq \sum_{j=1, j\neq k}^{n} d_{kj} \frac{x_j\left(M\left(l_p\right)\right)}{x_k\left(M\left(l_p\right)\right)} + \sum_{j=1}^{m} \frac{p_{kj}}{x_k\left(M\left(l_p\right)\right)} e^{-a_{kj} x_k\left(M\left(l_p\right) - \tau_{kj}\left(M\left(l_p\right)\right)\right)}. \tag{13}$$

Letting $p \to +\infty$ gives us that

$$\delta_k \leq \sum_{j=1, j\neq k}^{n} d_{kj}, \tag{14}$$

which contradicts with the inequality in (3). This shows that $x_i(t)$ is positive and bounded for all $t \in [t_0, \eta(\varphi)), i \in J$. From Theorem 2.3.1 in [17], we easily obtain $\eta(\varphi) = +\infty$.

Next we prove that any positive solution of (2) with (4) satisfies

$$\liminf_{t\to+\infty} x_j(t) > 0, \quad j \in J. \tag{15}$$

Denote $l = \min_{j\in J} \liminf_{t\to+\infty} x_j(t)$. We claim that $l > 0$. Suppose on the contrary that $l = 0$. Define

$$m(t) = \max\left\{\xi : \xi \leq t | \text{there exists } j_0 \in J \text{ such that } x_{j_o}(\xi) = \min_{t_0\leq s\leq t} x_{j_0}(s) = \min_{j\in J}\left\{\min_{t_0\leq s\leq t} x_j(s)\right\}\right\}. \tag{16}$$

Then $m(t) \to +\infty$ as $t \to +\infty$. Moreover, for a sequence $\{t_p\}_{p\geq 1}$ with $\lim_{p\to+\infty} t_p = +\infty$, we can choose $i_0 \in J$ and a subsequence $\{t_{p_k}\}_{k\geq 1} \subseteq \{t_p\}_{p\geq 1}$ such that

$$x_{i_0}\left(m\left(t_{p_k}\right)\right) = \min_{t_0\leq s\leq t_{p_k}} x_{i_0}(s) = \min_{j\in J}\left\{\min_{t_0\leq s\leq t_{p_k}} x_j(s)\right\} \tag{17}$$

and

$$\lim_{k\to+\infty} x_{i_0}\left(m\left(t_{p_k}\right)\right) = 0. \tag{18}$$

By virtue of the definition of $m(t_{p_k})$, we obtain that $D^- x_{i_0}(m(t_{p_k})) \leq 0$ or

$$\begin{aligned} \delta_{i_0} x_{i_0}\left(m\left(t_{p_k}\right)\right) &\geq \sum_{j=1, j\neq i_0}^{n} d_{i_0 j} x_j\left(m\left(t_{p_k}\right)\right) + \sum_{j=1}^{m} p_{i_0 j} \exp\left(-a_{i_0 j} x_{i_0}\left(m\left(t_{p_k}\right) - \tau_{i_0 j}\left(m\left(t_{p_k}\right)\right)\right)\right) \\ &\geq \sum_{j=1}^{m} p_{i_0 j} \exp\left(-a_{i_0 j} \sup_{t\geq t_0 - r_{i_0}} x_{i_0}(t)\right),\ m\left(t_{p_k}\right) > t_0 + \tau. \end{aligned} \tag{19}$$

Letting $k \to +\infty$ leads to

$$0 \geq \sum_{j=1}^{m} p_{i_0 j} \exp\left(-a_{i_0 j} \sup_{t\geq t_0 - r_{i_0}} x_{i_0}(t)\right) > 0. \tag{20}$$

This is a contradiction, and the claim holds. The proof of Lemma 1 is completed. □

Now, we show the global attractivity of $(x_1^*, x_2^*, \ldots, x_n^*)$ by the following two propositions:

Proposition 2. *If $y_i(t) = x_i(t) - x_i^*$ is eventually nonoscillating about zero, then $\lim_{t\to+\infty} x_i(t) = x_i^*$, for all $i \in J$.*

Proof. We only give the proofs for the case that $y_i(t) = x_i(t) - x_i^*$ is eventually nonnegative for all $i \in J$, since the eventually nonpositive case can be proved by a similar argument. In this case, we can choose $T > t_0$ such that

$$y_i(t) = x_i(t) - x_i^* \geq 0 \quad \text{for all } i \in J,\ t \geq T. \tag{21}$$

Let $i^* \in J$ such that $\limsup_{t\to+\infty} y_{i^*}(t) = \max_{j\in J} \limsup_{t\to+\infty} y_j(t)$. We claim that

$$\limsup_{t\to+\infty} y_{i^*}(t) = 0. \tag{22}$$

Assume the contrary that $\limsup_{t\to+\infty} y_{i^*}(t) > 0$. In view of the fluctuation lemma [18, Lemma A.1], there exists a sequence $\{t_k\}_{k\geq 1}$ such that

$$t_k \to +\infty,\ y_{i^*}(t_k) \to \limsup_{t\to+\infty} y_{i^*}(t),\ y'_{i^*}(t_k) \to 0 \text{ as } k \to +\infty. \tag{23}$$

It follows from (2) that

$$\begin{aligned} y'_{i^*}(t_k) &= -\delta_{i^*} x_{i^*}(t_k) + \sum_{j=1, j\neq i^*}^{n} d_{i^*j} x_j(t_k) \\ &\quad + \sum_{j=1}^{m} p_{i^*j} e^{-a_{i^*j} x_{i^*}\left(t_k - \tau_{i^*j}(t_k)\right)}, \quad t_k > T + \tau. \end{aligned} \tag{24}$$

Without loss of generality, we can pick a subsequence of $\{t_k\}$ (not relabelled) such that $\lim_{k\to+\infty} x_i(t_k)$, $\lim_{k\to+\infty} x_{i^*}(t_k)$ and $\lim_{k\to+\infty} x_{i^*}\left(t_k - \tau_{i^*j}(t_k)\right)$ exist for all $i \in J$, $j \in I = \{1, 2, \ldots, m\}$. Then,

$$\begin{aligned} x_{i^*}^* &\leq \lim_{k\to+\infty} x_{i^*}\left(t_k - \tau_{i^*j}(t_k)\right) \\ &\leq x_{i^*}^* + \limsup_{t\to+\infty} y_{i^*}(t) \\ &= \lim_{k\to+\infty} x_{i^*}(t_k), \quad j \in I. \end{aligned} \tag{25}$$

It follows from (24) that (taking limits)

$$\begin{aligned} 0 &\leq -\delta_{i^*}\left(\limsup_{t\to+\infty} y_{i^*}(t) + x_{i^*}^*\right) + \sum_{j=1, j\neq i^*}^{n} d_{i^*j}\left(\limsup_{t\to+\infty} y_{i^*}(t) + x_j^*\right) \\ &\quad + \sum_{j=1}^{m} p_{i^*j} e^{-a_{i^*j} x_{i^*}^*} < \left(-\delta_{i^*} + \sum_{j=1, j\neq i^*}^{m} d_{i^*j}\right) \limsup_{t\to+\infty} y_{i^*}(t) < 0, \end{aligned} \tag{26}$$

which leads to a contradiction. Hence, $\limsup_{t\to+\infty} y_{i^*}(t) = 0$. This completes the proof. □

Remark 3. It is worth noting that, from Proposition 2 the nonoscillating solutions of system (2) converge to the positive equilibrium point which does not depend on the delays.

Inspired by Theorem 4.1 in [19], we can obtain the following more general conclusion.

Lemma 4. *Let $v \leq 0 \leq u$ be such that $u \leq e^{-v} - 1$ and $kv \geq e^{-ku} - 1$ for some $k \geq 1$. Then $u = v = 0$.*

Proof. Since $u \leq e^{-v} - 1$ and $kv \geq e^{-ku} - 1$, we have

$$1 + u \leq e^{-v} \leq e^{(1/k) - (1/k)e^{-ku}}. \tag{27}$$

Let $f(x) = e^{(1/k) - (1/k)e^{-kx}} - x - 1$. Then,

$$f'(x) = e^{(1/k) - kx - (1/k)e^{-kx}} - 1 \tag{28}$$

and

$$f''(x) = e^{(1/k) - kx - (1/k)e^{-kx}}\left(-k + e^{-kx}\right) \leq 0, \quad \text{for all } x > 0. \tag{29}$$

Due to the fact that $f'(0) = 0$, it follows that

$$f'(x) < 0, \quad \text{for all } x > 0. \tag{30}$$

On the other hand, since $f(0) = 0$, we get $f(x) < 0$ for all $x > 0$. This implies that $u = 0$. According to $kv \geq e^{-ku} - 1$, we have $v \geq 0$. As $v \leq 0$, we must have $v = 0$. This finishes the proof. □

Next, we consider the attractivity of (2) on the premise that the conditions in Proposition 5 are not satisfied.

Let

$$y_i(t) = \frac{x_i(t)}{x_i^*} - 1, \quad i \in J. \tag{31}$$

Then, from (2), we get

$$\begin{aligned} y'_i(t) &= -\delta_i\left(y_i(t) + 1\right) + \sum_{j=1, j\neq i}^{n} d_{ij} \frac{x_j^*}{x_i^*}\left(y_j(t) + 1\right) \\ &\quad + \sum_{j=1}^{m} p_{ij} \frac{1}{x_i^*} e^{-a_{ij} x_i^*\left(y_i\left(t - \tau_{ij}(t)\right) + 1\right)} \\ &= -\delta_i\left(y_i(t) + 1\right) + \sum_{j=1, j\neq i}^{n} d_{ij}^*\left(y_j(t) + 1\right) \\ &\quad + \sum_{j=1}^{m} p_{ij}^* e^{-a_{ij} x_i^*\left(y_i\left(t - \tau_{ij}(t)\right) + 1\right)}, \end{aligned} \tag{32}$$

where $d_{ij}^* = d_{ij}\left(x_j^*/x_i^*\right)$, $p_{ij}^* = p_{ij}(1/x_i^*)$. It is easy to see that the global attractivity of the equilibrium $x^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$ for (2) is equivalent to the global attractivity of the trivial solution for (35).

Set

$$\lambda = \min_{i\in J} \liminf_{t\to+\infty} y_i(t), \quad \mu = \max_{i\in J} \limsup_{t\to+\infty} y_i(t). \tag{33}$$

By the fluctuation lemma [14, Lemma A.1], we can take $i_1^* \in J$ and a strictly monotonically increasing sequence $\left\{l_q^*\right\}_{q\geq 1}$ such that

$$l_q^* \to +\infty, \quad y_{i_1^*}\left(l_q^*\right) \to \mu, \quad \text{as } q \to +\infty. \tag{34}$$

Furthermore, choose $i_1 \in J$ and a subsequence $\left\{l_q^{**}\right\}_{q\geq 1} \subseteq \left\{l_q^*\right\}_{q\geq 1}$ satisfying

$$l_q^{**} > t_0,\ y_{i_1}(\bar{l}_q) = \max_{t\in[l_q^{**},l_{q+1}^{**}]} y_{i_1}(t) = \max_{i\in J}\left\{\max_{t\in[l_q^{**},l_{q+1}^{**}]} y_i(t)\right\} > 0, \quad \bar{l}_q \in [l_q^{**}, l_{q+1}^{**}]. \tag{35}$$

By the boundedness of $\{y'_{i_1}(\bar{l}_q)\}$, we pick a strictly monotonically increasing sequence $\{l_q\}_{q\geq 1} \subseteq \{\bar{l}_q\}_{q\geq 1}$ such that $\lim_{q\to+\infty} y'_{i_1}(l_q)$ exists. It follows from (35) that

$$y_{i_1}(l_q) > 0, \quad \lim_{q\to+\infty} l_q = +\infty, \quad \lim_{q\to+\infty} y_{i_1}(l_q) = \mu, \quad \lim_{q\to+\infty} y'_{i_1}(l_q) \geq 0. \tag{36}$$

Adopting the same procedure as in the proof of (12), there exist $i_2 \in J$ and a strictly monotonically increasing sequence $\{s_q\}_{q\geq 1}$ such that

$$y_{i_2}(s_q) < 0, \quad \lim_{q\to+\infty} s_q = +\infty, \quad \lim_{q\to+\infty} y_{i_2}(s_q) = \lambda, \quad \lim_{q\to+\infty} y'_{i_2}(s_q) \leq 0. \tag{37}$$

Subsequently, we prove that there is a positive integer Q_1 such that, for any $q \geq Q_1$, there exists $L_q \in [l_q - r_{i_1}, l_q)$ such that

$$y_{i_1}(L_q) = 0,\ y_{i_1}(t) > 0, \quad \text{for all } t \in (L_q, l_q). \tag{38}$$

In the contrary case, there exists a subsequence of $\{l_q\}$ (for convenience, we still denote by $\{l_q\}$) such that

$$y_{i_1}(t) > 0, \quad \text{for all } t \in [l_q - r_{i_1}, l_q),\ q = 1, 2, \ldots. \tag{39}$$

According to the definition of the l_q, we conclude from (32) that

$$\begin{aligned} y'_{i_1}(l_q) &= -\delta_{i_1}\left(y_{i_1}(l_q)+1\right) + \sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j}\left(y_j(l_q)+1\right) \\ &\quad + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{ij}x^*_{i_1}\left(y_{i_1}\left(l_q-\tau_{i_1 j}(l_q)\right)+1\right)} \\ &< -\delta_{i_1}\left(y_{i_1}(l_q)+1\right) + \sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j}\left(y_j(l_q)+1\right) \\ &\quad + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{ij}x^*_{i_1}}, \quad l_q > t_0^* + r_{i_1}. \end{aligned} \tag{40}$$

Assume that $\lim_{q\to+\infty} y_j(l_q)$ exists for all $j \in J\backslash\{i_1\}$, from the fact that $\limsup_{t\to+\infty} y_i(t) \leq \mu (i \in J)$, (40) implies that

$$\begin{aligned} 0 &\leq \lim_{q\to+\infty} y'_{i_1}(l_q) \\ &\leq -\delta_{i_1}\left(\lim_{q\to+\infty} y_{i_1}(l_q)+1\right) \\ &\quad + \sum_{j=1,j\neq i}^{m} d^*_{i_1 j}\left(\lim_{q\to+\infty} y_j(l_q)+1\right) + \sum_{j=1}^{m} p_{i_1 j} e^{-a_{i_1 j}x^*_{i_1}} \\ &\leq -\delta_{i_1}(\mu+1) + \sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j}(\mu+1) + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{ij}x^*_{i_1}} \\ &< (\mu+1)\left[-\delta_{i_1} + \sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j} + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{ij}x^*_{i_1}}\right] = 0, \end{aligned} \tag{41}$$

which is a contradiction and (38) is true.

Based on the above discussions, the following conclusion can be drawn.

Proposition 5. *Suppose that the assumption mentioned in Proposition 2 does not hold, and*

$$\frac{a_i\left(e^{\delta_i r_i}-1\right)\sum_{j=1}^{m} p_{ij} e^{-a_{ij}x_i^*}}{\delta_i + \left(\left(e^{\delta_i r_i}-1\right)/x_i^*\right)\sum_{j=1}^{m} p_{ij} e^{-a_{ij}x_i^*}} \leq 1, \quad \textit{for all } i \in J. \tag{42}$$

$$a_{i_1} x^*_{i_1} = \min_{1\leq i\leq n}\{a_i x_i^*\}. \tag{43}$$

Then $\lim_{t\to+\infty} x_i(t) = x_i^*, i \in J$.

Proof. Observe that, from Theorem 6, $-1 < \lambda \leq 0 \leq \mu$. Now, we show that $\lambda = \mu = 0$. Otherwise, either $\mu > 0$ or $\lambda < 0$ holds. We only consider the case that $\mu > 0$ holds (the situation is analogous for $\lambda < 0$).

For any given $\varepsilon > 0$ such that $\lambda - \varepsilon > -1$, by (33) there exists a positive integer $q^* > Q_2 + t_0^*$ such that

$$\lambda - \varepsilon < y_i(t) < \mu + \varepsilon \quad \text{for all } t > \min\left\{l_{q^*}, s_{q^*}\right\} - 2\tau,\ i \in J. \tag{44}$$

Furthermore, from the fact that

$$\begin{aligned} y'_i(t) &= -\delta_i(y_i(t)+1) + \sum_{j=1,j\neq i}^{n} d^*_{ij}\left(y_j(t)+1\right) \\ &\quad + \sum_{j=1}^{m} p^*_{ij} e^{-a_{ij}x_i^*\left(y_i\left(t-\tau_{ij}(t)\right)+1\right)}, \quad t \in [L_q, l_q], \end{aligned} \tag{45}$$

we have

$$\begin{aligned} &y_{i_1}(l_q)e^{\delta_{i_1} l_q} \\ &\leq -\left(e^{\delta_{i_1} l_q} - e^{\delta_{i_1} L_q}\right) + \sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j}\int_{L_q}^{l_q} (y_i(t)+1)e^{\delta_{i_1} t}dt \\ &\quad + \sum_{j=1}^{m} p^*_{i_1 j}\int_{L_q}^{l_q} e^{-a_{i_1 j}x_i^*\left(y_{i_1}\left(t_q-\tau_{i_1 j}(t_q)\right)+1\right)} e^{\delta_{i_1} t}dt \\ &\leq \left[-\delta_{i_1} + (\mu+\varepsilon+1)\left(\sum_{j=1,j\neq i_1}^{n} d^*_{i_1 j}\right) + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{i_1 j}x^*_{i_1}} e^{-a_{ij}x^*_{i_1 j}(\lambda-\varepsilon)}\right] \\ &\quad \times \frac{e^{\delta_{i_1} l_q} - e^{\delta_{i_1} L_q}}{\delta_{i_1}} \\ &= \left[-\delta_{i_1} + (\mu+\varepsilon+1)\left(\delta_{i_1} - \sum_{j=1}^{m} p^*_{i_1 j} e^{-x^*_{i_1}}\right) + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{i_1 j}x^*_{i_1}} e^{-a_{i_1 j}x^*_{i_1}(\lambda-\varepsilon)}\right] \\ &\quad \times \frac{e^{\delta_{i_1} l_q} - e^{\delta_{i_1} L_q}}{\delta_{i_1}} \\ &= \left[(\mu+\varepsilon)\delta_{i_1} + \sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{i_1 j}x^*_{i_1}}\left(e^{-a_{i_1 j}x^*_{i_1}(\lambda-\varepsilon)} - 1\right) - (\mu+\varepsilon)\sum_{j=1}^{m} p^*_{i_1 j} e^{-a_{i_1 j}x^*_{i_1}}\right] \\ &\quad \times \frac{e^{\delta_{i_1} l_q} - e^{\delta_{i_1} L_q}}{\delta_{i_1}}, \quad q > q^*, \end{aligned} \tag{46}$$

that is,

$$y_{i_1}(l_q) \le \left[(\mu+\varepsilon)\delta_{i_1} + \sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}\left(e^{-a_{i_1 j}x_{i_1}^{*}(\lambda-\varepsilon)}-1\right) - (\mu+\varepsilon)\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}\right]\frac{1-e^{-\delta_{i_1}r_{i_1}}}{\delta_{i_1}}, \quad q > q^{*}, \tag{47}$$

Letting $q \to +\infty$ and $\varepsilon \to 0^{+}$, (36) and (47) give us that

$$\mu \le \mu\left(1-e^{-\delta_{i_1}r_{i_1}}\right) + \left[\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}\left(e^{-a_{i_1 j}x_{i_1}^{*}\lambda}-1\right) - \mu\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}\right]\frac{1-e^{-\delta_{i_1}r_{i_1}}}{\delta_{i_1}}, \tag{48}$$

which implies

$$\mu \le \frac{\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}\left(1-e^{-\delta_{i_1}r_{i_1}}\right)/(\delta_{i_1})}}{e^{-\delta_{i_1}r_{i_1}} + \sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}\left(1-e^{-\delta_{i_1}r_{i_1}}\right)/(\delta_{i_1})}}\left(e^{-a_{i_1}x_{i_1}^{*}\lambda}-1\right) = \frac{\left(e^{\delta_{i_1}r_{i_1}}-1\right)\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}}{\delta_{i_1} + \left(e^{\delta_{i_1}r_{i_1}}-1\right)\sum_{j=1}^{m} p_{i_1 j}^{*} e^{-a_{i_1 j}x_{i_1}^{*}}}\left(e^{-a_{i_1}x_{i_1}^{*}\lambda}-1\right). \tag{49}$$

Let's assume that $\lambda < 0$ (if $\lambda = 0$, from (49) and $\mu \ge 0$ we have $\mu = 0$). Using the same arguments in the proof of (38), we can obtain that there is a positive integer $Q_2 > Q_1$ such that, for any $q \ge Q_2$, there exists $S_q \in \left[s_q - r_{i_2}, s_q\right)$ such that

$$y_{i_2}(S_q) = 0, \text{ and } y_{i_2}(t) < 0, \quad \text{for all } t \in (S_q, s_q). \tag{50}$$

Again from (32), we have

$$\begin{aligned}
&y_{i_2}(s_q)e^{\delta_{i_2}s_q}\\
&\ge -\left(e^{\delta_{i_2}s_q}-e^{\delta_{i_2}S_q}\right) + \sum_{j=1,j\ne i_2}^{m} d_{i_2 j}^{*}\int_{S_q}^{s_q}(y_i(t)+1)e^{\delta_{i_2}t}dt\\
&\quad + \sum_{j=1}^{m} p_{i_2 j}^{*}\int_{S_q}^{s_q} e^{-a_{i_2 j}x_{i_2}^{*}\left(y_i\left(s_q-\tau_{i_2 j}(s_q)\right)+1\right)}e^{\delta_{i_2}t}dt\\
&\ge \left[-\delta_{i_2} + (\lambda-\varepsilon+1)\left(\sum_{j=1,j\ne i_2}^{m} d_{i_2 j}^{*}\right) + \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}} e^{-a_{i_2 j}x_{i_2}^{*}(\mu+\varepsilon)}\right]\times\frac{e^{\delta_{i_2}s_q}-e^{\delta_{i_2}S_q}}{\delta_{i_2}}\\
&= \left[-\delta_{i_2} + (\lambda-\varepsilon+1)\left(\delta_{i_2} - \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\right) + \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}} e^{-a_{i_2 j}x_{i_2}^{*}(\mu+\varepsilon)}\right]\times\frac{e^{\delta_{i_2}s_q}-e^{\delta_{i_2}S_q}}{\delta_{i_2}}\\
&\ge \left[(\lambda-\varepsilon)\delta_{i_2} + \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\left(e^{-a_{i_2}x_{i_2}^{*}(\mu+\varepsilon)}-1\right) - (\lambda-\varepsilon)\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\right]\times\frac{e^{\delta_{i_2}s_q}-e^{\delta_{i_2}S_q}}{\delta_{i_2}}, \quad q > q^{*},
\end{aligned} \tag{51}$$

that is,

$$y_{i_2}(s_q) \ge \left[(\lambda-\varepsilon)\delta_{i_2} + \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\left(e^{-a_{i_2}x_{i_2}^{*}(\mu+\varepsilon)}-1\right) - (\lambda-\varepsilon)\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\right]\frac{1-e^{-\delta_{i_2}r_{i_2}}}{\delta_{i_2}}, \quad q > q^{*}. \tag{52}$$

Letting $q \to +\infty$ and $\varepsilon \to 0^{+}$, (37) and (52) give us that

$$\lambda \ge \lambda\left(1-e^{-\delta_{i_2}r_{i_2}}\right) + \left[\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\left(e^{-a_{i_2}x_{i_2}^{*}\mu}-1\right) - \lambda\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}\right]\times\frac{1-e^{-\delta_{i_2}r_{i_2}}}{\delta_{i_2}}, \tag{53}$$

which implies

$$\lambda \ge \frac{\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}\left(1-e^{-\delta_{i_2}r_{i_2}}\right)/(\delta_{i_2})}}{e^{-\delta_{i_2}r_{i_2}} + \sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}\left(1-e^{-\delta_{i_2}r_{i_2}}\right)/(\delta_{i_2})}}\left(e^{-a_{i_2}x_{i_2}^{*}\mu}-1\right) = \frac{\left(e^{\delta_{i_2}r_{i_2}}-1\right)\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}}{\delta_{i_2} + \left(e^{\delta_{i_2}r_{i_2}}-1\right)\sum_{j=1}^{m} p_{i_2 j}^{*} e^{-a_{i_2 j}x_{i_2}^{*}}}\left(e^{-a_{i_2}x_{i_2}^{*}\mu}-1\right). \tag{54}$$

This, combined with (42) and (49) imply that

$$a_{i_1}x_{i_1}^{*}\mu \le e^{-a_{i_1}x_{i_1}^{*}\lambda}-1, \; a_{i_2}x_{i_2}^{*}\lambda \ge e^{-a_{i_2}x_{i_2}^{*}\mu}-1. \tag{55}$$

Denote

$$\mu^{*} = a_{i_1}x_{i_1}^{*}\mu, \quad \lambda^{*} = a_{i_1}x_{i_1}^{*}\lambda, \quad k = \frac{a_{i_2}x_{i_2}^{*}}{a_{i_1}x_{i_1}^{*}}. \tag{56}$$

From (43), one can find that $k \ge 1$. By virtue of (55) and (56), we have

$$\mu^{*} \le e^{-\lambda^{*}}-1, \quad k\lambda^{*} \ge e^{-k\mu^{*}}-1. \tag{57}$$

According to Lemma 4, it is easy to see that $\lambda^{*} = \mu^{*} = 0$. Furthermore, we get $\lambda = \mu = 0$, which is a contradiction, and $\mu > 0$ is not true. This completes the proof. □

Combining Propositions 2 and 5, we have the following delay-dependent criterion of global attraction.

Theorem 6. *Assume that conditions (3), (42) and (43) are satisfied. Then the positive equilibrium $(x_1^{*}, x_2^{*}, \ldots, x_n^{*})$ is a global attractor of (2).*

Remark 7. Let us note that, from

$$\lim_{r_i \to 0^{+}} \frac{a_i\left(e^{\delta_i r_i}-1\right)\sum_{j=1}^{m} p_{ij}e^{-a_{ij}x_i^{*}}}{\delta_i + \left(\left(e^{\delta_i r_i}-1\right)/x_i^{*}\right)\sum_{j=1}^{m} p_{ij}e^{-a_{ij}x_i^{*}}} = 0 \quad \text{for all } i \in J, \tag{58}$$

one can find that condition (42) naturally holds under the sufficiently small delay, and the positive equilibrium point $(x_1^{*}, x_2^{*}, \ldots, x_n^{*})$ is a global attractor of (2) with small delays. Moreover,

$$\lim_{r_i \to +\infty} \frac{a_i\left(e^{\delta_i r_i}-1\right)\sum_{j=1}^{m} p_{ij}e^{-a_{ij}x_i^{*}}}{\delta_i + \left(\left(e^{\delta_i r_i}-1\right)/x_i^{*}\right)\sum_{j=1}^{m} p_{ij}e^{-a_{ij}x_i^{*}}} = a_i x_i^{*} \quad \text{for all } i \in J, \tag{59}$$

implies that condition (42) is not satisfied when the delays in (2) is sufficiently large and $a_i x_i^{*} > 1$ for $i \in J$.

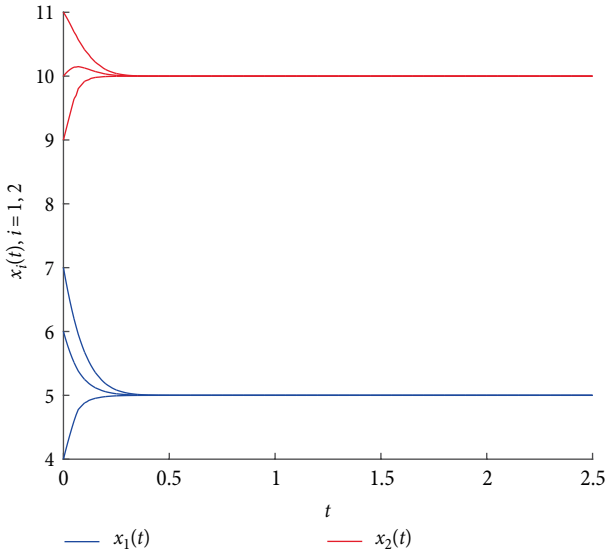


FIGURE 1: Numerical solution $(x_1(t), x_2(t))$ of Eq. (60) with different initial values $(x_1(s), x_2(s)) \equiv (4, 9), (6, 10), (7, 11), s \in [-0.05, 0]$.

3. A Numerical Example

Example 1. Consider the following Lasota–Wazewska-type delay system with two groups and patch structures:

$$\begin{cases} x_1'(t) = -8x_1(t) + 3x_2(t) + 4e^5 e^{-x_1\left(t-\tau_{11}(t)\right)} + 6ee^{-(1/5)-x_1\left(t-\tau_{12}(t)\right)}, \\ x_2'(t) = -5x_2(t) + 4x_1(t) + 10e^5 e^{-(1/2)x_2\left(t-\tau_{21}(t)\right)} + 20e^{(5/2)} e^{-(1/4)-x_2\left(t-\tau_{22}(t)\right)}. \end{cases} \tag{60}$$

It is easy to check that

$$a_1 = 1,\ a_2 = \frac{1}{2},\ 0 < 8 - 3 < 4e^5 + 6e,\ 0 < 5 - 4 < 10e^5 + 20e^{5/2}. \tag{61}$$

Therefore, (3) is true. Obviously, $(x_1^*, x_2^*) = (5, 10)$ is the positive equilibrium point of (60). In addition, $a_1 x_1^* = a_2 x_2^* = 5$ implies (43) holds. Now, we choose

$$\tau_{ij}(t) = 0.05\left|\sin(i + j)t\right|, \quad i, j = 1,\ 2, \tag{62}$$

such that (42) holds. By Theorem 6, we conclude that the positive equilibrium point (5, 10) is a global attractor of (60) with delays (62). This implies that small delays are harmless on the asymptotic behavior of system (60). Numerical runs with Matlab illustrate convergence of positive solutions to (x_1^*, x_2^*) (see Figure 1).

Remark 8. Observe that the methods used in [13, 14] are not suitable for (60) with (62) since the system (60) with time-varying delays (62) does not generate a semiflow. In addition, it is also worth pointing out that the components of the positive equilibrium point in this paper are not required to be equal, which is also different from the literature [12].

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was jointly supported by the Natural Science Foundation of Hunan Province (2018JJ2194) and Scientific Research Fund of Hunan Provincial Education Department of China (18B456, 15C0719, 16C0036).

References

[1] M. Wazewska-Czyzewska and A. Lasota, "Mathematical problems of the dynamics of red blood cells system," *Annals of the Polish Series I: Applied Mathematics*, vol. 17, pp. 23–40, 1988.

[2] K. Gopalsamy and S. Trofimchuk, "Almost periodic solutions of Lasota–Wazewska-type delay differential equation," *Journal of Mathematical Analysis and Applications*, vol. 237, no. 1, pp. 106–127, 1999.

[3] Z. Huang, S. Gong, and L. Wang, "Positive almost periodic solution for a class of Lasota–Wazewska model with multiple time-varying delays," *Computers & Mathematics with Applications*, vol. 61, no. 4, pp. 755–760, 2011.

[4] J. Shao, "Pseudo almost periodic solutions for a Lasota–Wazewska model with an oscillating death rate," *Applied Mathematics Letters*, vol. 43, pp. 90–95, 2015.

[5] L. Duan and C. Huang, "Existence and global attractivity of almost periodic solutions for a delayed differential neoclassical growth model," *Mathematical Methods in the Applied Sciences*, vol. 40, no. 3, pp. 814–822, 2017.

[6] X. Wang and Z. Li, "The existence, uniqueness and global attractively a positive almost periodic solution for a Lasota–Wazewska model," *Journal of Biomathematics*, vol. 23, pp. 449–456, 2008.

[7] L. Duan, L. Huang, and Y. Chen, "Global exponential stability of periodic solutions to a delay Lasota–Wazewska model with discontinuous harvesting," *Proceedings of the American Mathematical Society*, vol. 144, no. 2, pp. 561–573, 2016.

[8] L. Yao, "Global attractively of a delayed Nicholson-type system involving nonlinear density-dependent mortality terms," *Mathematical Methods in the Applied Sciences*, vol. 41, no. 6, pp. 2379–2391, 2018.

[9] D. Caetano and T. Faria, "Stability and attractivity for Nicholson systems with time-dependent delays," *Electronic Journal of Qualitative Theory of Differential Equations*, no. 63, pp. 1–19, 2017.

[10] Y. Takeuchi, W. Wang, and Y. Saito, "Global stability of population models with patch structure," *Nonlinear Analysis: Real World Applications*, vol. 7, no. 2, pp. 235–247, 2006.

[11] C. Huang, Z. Yang, T. Yi, and X. Zou, "On the basins of attraction for a class of delay differential equations with non-monotone bistable nonlinearities," *Journal of Differential Equations*, vol. 256, no. 7, pp. 2101–2114, 2014.

[12] R. Jia, Z. Long, and M. Yang, "Delay-dependent criteria on the global attractivity of Nicholson s blowflies model with patch structure," *Mathematical Methods in the Applied Sciences*, vol. 40, no. 11, pp. 4222–4232, 2017.

[13] L. Duan, H. Wei, and L. Huang, "Finite-time synchronization of delayed fuzzy cellular neural networks with discontinuous activations," *Fuzzy Sets and Systems*, vol. 361, pp. 56–70, 2019.

[14] T. Faria, "Asymptotic behaviour for a class of delayed cooperative models with patch structure," *Discrete and Continuous Dynamical Systems – Series B*, vol. 6, no. 6, pp. 1567–1579, 2013.

[15] C. Huang, H. Zhang, and L. Huang, "Almost periodicity analysis for a delayed Nicholson's blowflies model with nonlinear density-dependent mortality term," *Communications on Pure & Applied Analysis*, vol. 18, no. 6, pp. 3337–3349, 2019.

[16] L. Duan, X. Fang, and C. Huang, "Global exponential convergence in a delayed almost periodic Nicholson's blowflies model with discontinuous harvesting," *Mathematical Methods in the Applied Sciences*, vol. 41, no. 5, pp. 1954–1965, 2018.

[17] J. Hale and S. Verduyn Lunel, "Introduction to functional differential equations," *Applied Mathematical Sciences*, vol. 99, Springer-Verlag, New York, 1993.

[18] H. L. Smith, *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, Springer, New York, 2011.

[19] J. W.-H. So and J. S. Yu, "Global attractivity and uniform persistence in Nicholson blowflies," *Differential Equations Dynamics Systems*, vol. 2, pp. 11–18, 1994.

Hindawi
Complexity
Volume 2020, Article ID 8763126, 18 pages
https://doi.org/10.1155/2020/8763126

Research Article

Modeling Periodic HFMD with the Effect of Vaccination in Mainland China

Lei Shi,[1,2] **Hongyong Zhao**,[1] **and Daiyong Wu**[1,3]

[1]*Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*Department of Mathematics, Honghe University, Mengzi 661199, China*
[3]*Department of Mathematics, Anqing Normal University, Anqing 246133, China*

Correspondence should be addressed to Hongyong Zhao; hyzho1967@126.com

Received 30 August 2019; Revised 8 November 2019; Accepted 14 December 2019; Published 9 January 2020

Guest Editor: George V. Popescu

Hand, foot, and mouth disease (HFMD), associated with more than 20 disease-causing enteroviruses, is one of the major public health problems in mainland China, and the unique vaccine available is for enterovirus 71 (EV71). In this paper, we propose a new epidemic model to investigate the effect of EV71 vaccination on the spread of HFMD with multiple pathogenic viruses in mainland China. In addition, suitable periodic transmission functions are designed, with a two-year period and taking into consideration the effects of opening and closing of schools. After defining the basic reproduction number R_0, we prove that the disease-free equilibrium is globally asymptotically stable if $R_0 < 1$, and there exists at least one positive periodic solution and the disease is uniformly persistent if $R_0 > 1$. We use the model to simulate the HFMD reported data in mainland China from January 2008 to June 2019. The numerical experiments show that increasing the vaccinated rate can effectively control the spread of HFMD in mainland China, yet the disease does not become extinct. Moreover, if we can control the baseline contact rate of infectious individuals and the recovery rate of symptomatic infectious individuals under certain conditions, which can be achieved by improving protective measures and medical conditions, then the disease will be eliminated.

1. Introduction

Hand, foot, and mouth disease (HFMD) is a common infectious disease among infants and children caused by intestinal viruses of the Picornaviridae family. There exist many types of HFMD virus, such as coxsackievirus A5, A10, A16, A19 types and EV71 type, and so on. Among these viruses, the main viruses causing HFMD are coxsackievirus A16 (CVA16) and enterovirus 71 (EV71) [1]. HFMD mostly occurs in nursery schools or kindergartens, and its high incidence seasons are summer and autumn [1]. The usual incubation period is 2 to 7 days, and the recovery period is 7 to 10 days [2]. Since HFMD was first reported in New Zealand in 1957 [3], it has spread around the world, especially in Asia [4–7].

HFMD is a mild, self-limiting illness that primarily affects infants and young children. In recent years, HFMD is also a common illness in mainland China. Since 2008, the government has classified it as a class III infectious disease in the National Stationary Notifiable Communicable Diseases (NSNCD) [8], and the monthly cases have been archived by the Chinese Center for Disease Control and Prevention (CCDC) [9] (see Figure 1). The main viruses causing HFMD in mainland China are CVA16 and EV71, and the proportions of HFMD cases caused by CVA16 and EV71 are about 30% and 44%, respectively [10–12]. EV71 vaccine, which efficacy against EV71-associated HFMD was 97.4%, has been developed and put into use in mainland China beginning in 2016, so many young children would be protected from EV71 infections [13, 14]. As of May 2019, monovalent EV71 vaccines have not yet been included in the routine paediatric vaccination programme in China, which means that they are needed to be paid out-of-pocket by parents [15–17]. Meanwhile,

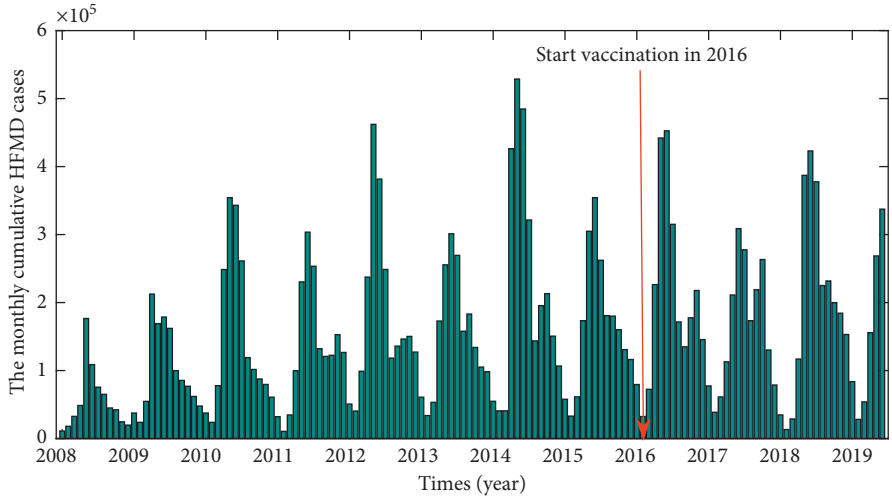


Figure 1: The confirmed monthly cumulative HFMD cases from the Chinese Center for Disease Control and Prevention (CCDC) from January 2008 to June 2019.

vaccine coverage of these monovalent EV71 vaccines among children aged 6 months to 5 years ranges from < 10% to 50% in different provinces in mainland China [15–17].

Compartmental differential equation modeling is an important tool for understanding infectious disease spread and control. There are two types of dynamic models for HFMD. One is an autonomous ordinary differential model where all the parameters such as transmission rate, birth population, recovery rate, and so on of whom are assumed as constants. For instance, Tiing and Labadin [2] used a simple SIR model to predict the number of the infected and the duration of an outbreak in Sarawak Malaysia; Roy and Halder [18, 19] proposed a deterministic susceptible-exposed-infectious-recovered (SEIR) model and a susceptible-exposed-infectious-quarantine-recovered (SEIQR) model of HFMD, and Viriyapong and Wichaino [20] further analyzed dynamic behaviors of SEIR; Yang et al. [21] and Li et al. [22] researched the optimal control of SEIQR and its dynamic behavior, respectively; Li et al. [23] further constructed a two-stage-structured SEIQR model to fit the HFMD data in mainland China from 2009 to 2014; Wang et al. [24] further considered indirect transmission coming from the contaminated environments to establish a $SEII_eRW$ model (I_e is the number of the asymptomatic infectious individuals), and Sharma and Samanta [25] considered the combined effect of asymptomatic infectious individuals and quarantine measures on the spread of HFMD. It should be noted that the reported monthly cases of HFMD from CCDC [9] exhibit the periodic pattern. From a practical point of view, clarifying the mechanisms that link seasonal environmental changes to disease dynamics may aid in forecasting the long-term health risks, crucial in developing an effective public health program and in setting objectives, and utilizing limited resources more effectively. For these reasons, the other approach is a nonautonomous differential equation model, which considers periodic transmission rates. Of these, Liu [26] constructed a periodic SEIQR model to simulate the dynamics of HFMD transmission and showed that quarantine has a positive impact on controlling the spread of HFMD; Zhu et al. [27] used the periodic SEIQR model to investigate the spread of seasonal HFMD in Wenzhou city, Zhejiang Province, China; Ma et al. [28], considering asymptomatic infectious individuals (I_e), proposed a periodic $SEII_eQR$ model to analyse HFMD in Shandong Province, China, which shows that asymptomatic infectious subpopulation plays an important role in the spread of HFMD. However, among the two types of dynamic models, fewer studies considered the influence of EV71 vaccination on the spread of HFMD.

Recently, Samanta [29] and Wang et al. [30] considered vaccination in modeling for HFMD. In these studies, the vaccinated individuals were assumed to be immunized against all HFMD-associated enteroviruses, in which a vaccinated individual would be transferred from the susceptible compartment to the recovered compartment in modeling (see system (1) in [29] or system (1) in [30]). However, in mainland China, only EV71 vaccine existed, and the vaccine only has well immunity for EV71, but not for others [13, 14, 31, 32]. Thus, while there are multiple HFMD-associated enteroviruses, only one vaccine for EV71 is available. A natural problem is how to build a suitable model to reflect the influence of EV71 vaccination on the spread of HFMD with multiple pathogenic viruses in mainland China.

From the practical point of view, periodic transmission functions are needed to consider in modeling. From Figure 1, it should be noted that the monthly reported cases of HFMD in mainland China from CCDC [9] exhibit periodic patterns. The annual cycle patterns or seasonal patterns may be attributed to these facts such as climatic, geographical, and demographic information but not

limited to them [28, 30, 33]. Although different seasonality across the whole country, Xing et al.'s epidemiological study of HFMD in China, 2008–2012 [33], implied that geographical differences in seasonal patterns were weakly associated with climate and demographic factors (variance explained 8–23% and 3–19%, respectively). Our analysis of previous studies for periodic HFMD models [4, 26, 28, 30] finds that all the chosen periodic transmission functions have a similar form, and its period is 1 year, which effectively can reflect the annual seasonal changes for HFMD. However, annual cycle patterns aside, the epidemics of HFMD in mainland China have displayed the following two phenomena: (1) the data from CCDC [8] (see Figure 1) show that HFMD has a two-year epidemic cycle because the epidemic period of partial HFMD-associated enteroviruses in some areas is two years [34, 35]; (2) Figure 1 shows that the number of cases suddenly rises in September and October each year because of the opening and closing of schools in mainland China [36]. The previous models based on the disease's transmission functions [4, 26, 28, 30] cannot simulate the above two phenomena. Hence, to better reveal the mechanism of the epidemic, it is very important to find a more suitable periodic transmission function for modeling.

Given the issues discussed above, we propose a new mathematical model to investigate the effect of EV71 vaccination on the spread of HFMD in mainland China. To analyze the effect of EV71 vaccination, we derive our modeling ideas from the previous modeling for influenza [37]. Then, in modeling, we divide the total individuals into two groups, the unvaccinated and the vaccinated. In addition, novel two-year periodic transmission functions, with considering the effects of school opening and closing, are also considered in our modeling. Then, we evaluate the basic reproduction number R_0 to analyze the dynamical behaviors of the model and use the proposed model to fit the monthly reported data of HFMD in mainland China from January 2008 to June 2019. Furthermore, through sensitivity analysis of the basic reproduction number R_0 in terms of key parameters, we explore some effective prevention and control measures for HFMD in mainland China.

The paper is organized as follows: in Section 2, we introduce a new periodic epidemiological model of HFMD. In Section 3, the dynamic behaviors of this model are analyzed theoretically. In Section 4, simulations of the model, sensitivity analysis of the basic reproduction number, and some prevention and control measures are performed. In Section 5, we discuss and summarize our conclusions.

2. Model Formulation

Recently, Samanta [29], Wang et al. [30] considered the effect of the vaccination in modeling for HFMD, in which it assumed that vaccine can resist all the HFMD-associated enteroviruses. However, there only exists EV71 vaccine to be used in mainland China, and the vaccine only has well immunity for EV71 but not for others [13, 14]. It notes that the modeling techniques in [29, 30] cannot be used to really reflect the spread of the disease with a single effective vaccine and multiple viruses. In order to model the influence of EV71 vaccination on the spread of HFMD with multiple pathogenic viruses in mainland China, we divide the total individuals into two different groups, the unvaccinated and the vaccinated, in modeling. Next, based on the above analyses, we denote the total number of unvaccinated and vaccinated individuals by $N_s(t)$ and $N_v(t)$, respectively, and classify each of them into four subclasses: susceptible, exposed, infectious with symptoms (symptomatic infectious), and infectious but not yet symptomatic (asymptomatic infectious), with the number of unvaccinated individuals denoted by $S(t), E_s(t), I_s(t), I_s^e(t)$ and $R_s(t)$ and the number of vaccinated individuals denoted by $V(t), E_v(t), I_v(t), I_v^e(t)$ and $R_v(t)$, respectively. Denote the total population size $N(t) = N_s(t) + N_v(t)$.

In addition, many epidemiological models for HFMD ([26–28, 30]), considering annual seasonal changes, were simulated by using the sinusoidal function $\beta_1(t) = a_0[1 + a_1 \sin((\pi/6)t)]$ with one year period (take a month as the unit time). As discussion in Section 1, seasonal changes and school holidays will change the contact rate of children, and the virus contagion changes periodically, some of which are one year for a cycle and some of which are two years for a cycle. Thus, taking a month as the unit time, we define periodic transmission functions with the period of 2 years as $\beta_1(t) = a_0[1 + a_1 \sin((\pi/2)t) + a_2 \sin((\pi/6)t) + a_3 \sin((\pi/12)t) + a_4 \sin(-1 + (\pi/3)t)]$ and $\beta_2(t) = b_0[1 + b_1 \sin(-1 + (\pi/2) + (\pi/4)t) + b_2 \sin((\pi/6)t) + b_3 \sin((\pi/12)t)]$ for symptomatic infectious individuals and asymptomatic infectious individuals, respectively. The studies in [31] showed that, in mainland China, the EV71 vaccine was targeted at infants or children aged 6 to 71 months. They were also a high-risk group for HFMD [31, 32]. Next, let Λ be the monthly number of children entering the age of 6–71 months, and d be progression rate leaving the children group aged 6–71 months. We also assume that p is the vaccinated rate for infants or children aged 6–71 months. Moreover, the results in [15, 17] showed that vaccine coverage of these monovalent EV71 vaccines among children aged 6 months to 5 years ranges from $<10\%$ to 50% in different provinces in mainland China in recent years. This implies that the current vaccination rate of the monovalent EV71 vaccine is very low in mainland China. In the case of the low vaccine coverage, we let k_s be the proportion of HFMD cases in mainland China, without vaccination, caused by HFMD-associated enteroviruses except EV71. Furthermore, we assume that a symptomatic or an asymptomatic infectious individual with vaccination has the same recovery rate as a symptomatic or asymptomatic infectious individual without vaccination. Then, the transmission dynamics associated with these subpopulations are shown in Figure 2.

The model is described by the following ordinary differential equations:

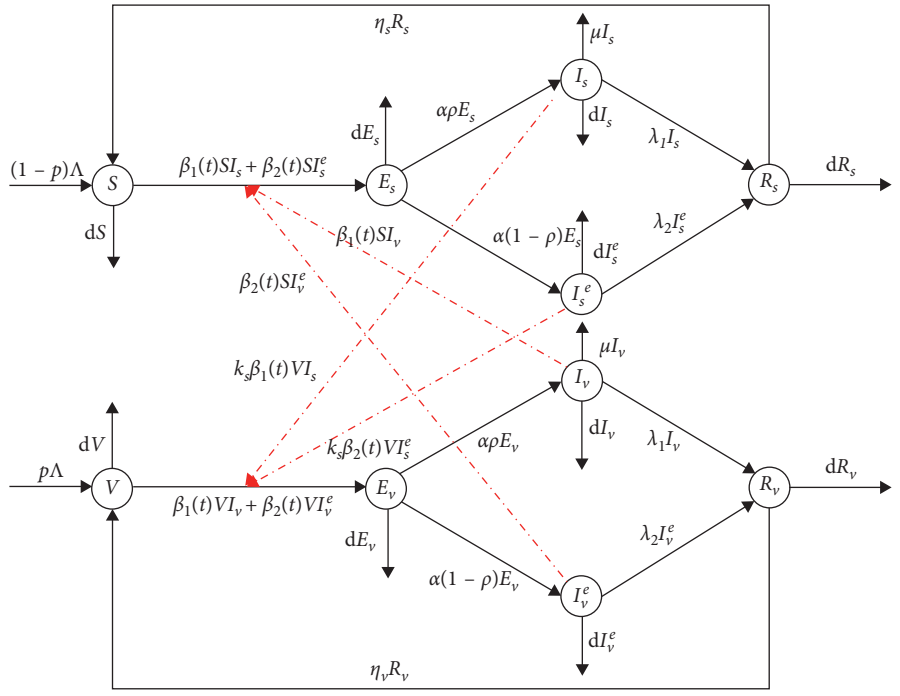


Figure 2: Flow diagram representing transmission routes and other processes modeled by system (1).

$$
\begin{cases}
\dfrac{dS}{dt} = (1-p)\Lambda - \beta_1(t)S(I_s + I_v) - \beta_2(t)S(I_s^e + I_v^e) \\
\quad -dS + \eta_s R_s, \\
\dfrac{dE_s}{dt} = \beta_1(t)S(I_s + I_v) + \beta_2(t)S(I_s^e + I_v^e) - (d+\alpha)E_s, \\
\dfrac{dI_s}{dt} = \alpha\rho E_s - (d+\mu+\lambda_1)I_s, \\
\dfrac{dI_s^e}{dt} = \alpha(1-\rho)E_s - (d+\lambda_2)I_s^e, \\
\dfrac{dR_s}{dt} = \lambda_1 I_s + \lambda_2 I_s^e - (d+\eta_s)R_s, \\
\dfrac{dV}{dt} = p\Lambda - \beta_1(t)V(k_s I_s + I_v) - \beta_2(t)V(k_s I_s^e + I_v^e) \\
\quad -dV + \eta_v R_v, \\
\dfrac{dE_v}{dt} = \beta_1(t)V(k_s I_s + I_v) + \beta_2(t)V(k_s I_s^e + I_v^e) - (d+\alpha)E_v, \\
\dfrac{dI_v}{dt} = \alpha\rho E_v - (d+\mu+\lambda_1)I_v, \\
\dfrac{dI_v^e}{dt} = \alpha(1-\rho)E_v - (d+\lambda_2)I_v^e, \\
\dfrac{dR_v}{dt} = \lambda_1 I_v + \lambda_2 I_v^e - (d+\eta_v)R_v,
\end{cases}
\tag{1}
$$

where p is nonnegative and other parameters are positive, and its biological meanings are listed in Table 1.

3. Stability Analysis and Persistence

In this section, we will investigate the dynamic behaviors of system (1). First, we introduce some notations which will be used throughout this paper. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, and $\mathbb{R}_+^n \triangleq \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0,\ i = 1, \ldots, n\}$. $A = (a_{ij})_{n\times n}$ denotes a $n \times n$ real matrix. The superscript T denotes matrix or vector transposition. Let

$$
\begin{aligned}
\varphi(0) = (&S(0), E_s(0), I_s(0), I_s^e(0), R_s(0), V(0), E_v(0), \\
&I_v(0), I_v^e(0), R_v(0))^T \in \mathbb{R}_+^{10},
\end{aligned}
\tag{2}
$$

be the initial condition of system (1), and

$$
\begin{aligned}
\varphi(t) = (&S(t), E_s(t), I_s(t), I_s^e(t), R_s(t), V(t), E_v(t), I_v(t), \\
&I_v^e(t), R_v(t))^T \in \mathbb{R}_+^{10},
\end{aligned}
\tag{3}
$$

be the solution of system (1) at time t for $t \geq 0$.

From the first equation of system (1), if $S(t) = 0$, then $(dS/dt) \geq 0$. According to function continuity and monotonicity, when $S(0) \geq 0$, we have $S(t) \geq 0$. Moreover, we have similar results for other state variables of system (1). Therefore, any solution of system (1) with nonnegative initial conditions is nonnegative. Next, the following lemma shows that the solutions are uniformly ultimately bounded.

Table 1: Descriptions and values of parameters in system (1).

Para.	Value	Interpretation (units)	Source
Λ	1.41×10^6	The monthly number of children entering the age of 6–71 months (/month)	[38]
D	1.30×10^{-2}	Progression rate leaving the children group aged 6–71 months (/month)	[38]
k_s	0.56	Proportion of infectious individuals, without vaccination, caused by HFMD-associated enteroviruses except EV71 (none)	[39]
μ	3.4×10^{-4}	Disease-related death rate of HFMD (/month)	[8, 9]
p	0.23	Vaccinated rate for children aged 6–71 months (/month)	Fitting
η_1	3.50×10^{-2}	Rate from $R_s(t)$ to $S(t)$ (/month)	[24]
η_2	3.50×10^{-2}	Rate from $R_v(t)$ to $V(t)$ (/month)	Assumption
$1/\alpha$	4/30	Average incubation period (/month)	[2]
ρ	0.23	Proportion of symptomatic infectious individuals (none)	Fitting
λ_1	3.53	Recovery rate of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (/month)	[2]
λ_2	3.46	Recovery rate of asymptomatic infectious individuals containing $I_s^e(t)$ and $I_v^e(t)$ (/month)	Fitting
a_0	1.09×10^{-7}	The baseline contact rate of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (none)	Fitting
a_1	0.12	The magnitude of forcing of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (none)	Fitting
a_2	0.22	The magnitude of forcing of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (none)	Fitting
a_3	0.03	The magnitude of forcing of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (none)	Fitting
a_4	0.15	The magnitude of forcing of symptomatic infectious individuals containing $I_s(t)$ and $I_v(t)$ (none)	Fitting
b_0	2.01×10^{-8}	The baseline contact rate of asymptomatic infectious individuals containing $I_s^e(t)$ and $I_v^e(t)$ (none)	Fitting
b_1	0.08	The magnitude of forcing of asymptomatic infectious individuals containing $I_s^e(t)$ and $I_v^e(t)$ (none)	Fitting
b_2	0.19	The magnitude of forcing of asymptomatic infectious individuals containing $I_s^e(t)$ and $I_v^e(t)$ (none)	Fitting
b_3	0.13	The magnitude of forcing of asymptomatic infectious individuals containing $I_s^e(t)$ and $I_v^e(t)$ (none)	Fitting

Lemma 1. *The solutions of system (1) eventually enter*

$$\Gamma = \left\{ \begin{array}{c} \left(S(t), E_s(t), I_s(t), I_s^e(t), R_s(t), V(t), E_v(t), I_v(t), I_v^e(t), R_v(t)\right)^T \in \mathbb{R}_+^{10}: \; 0 \le S(t) + E_s(t) \\ +I_s(t) + I_s^e(t) + R_s(t) \le \dfrac{(1-p)\Lambda}{d}, 0 \le V(t) + E_v(t) + I_v(t) + I_v^e(t) + R_v(t) \le \dfrac{p\Lambda}{d} \end{array} \right\}. \tag{4}$$

Moreover, Γ is a positively invariant set.

Proof. It is obvious that the unvaccinated population size $N_s(t) = S(t) + E_s(t) + I_s(t) + I_s^e(t) + R_s(t) \ge 0$ and the vaccinated population size $N_v(t) = V(t) + E_v(t) + I_v(t) + I_v^e(t) + R_v(t) \ge 0$. From system (1), we have

$$\begin{aligned} \frac{\mathrm{d}N_s}{\mathrm{d}t} &= (1-p)\Lambda - \mathrm{d}S - \mathrm{d}E_s - (d+\mu)I_s - \mathrm{d}I_s^e - \mathrm{d}R_s \\ &= (1-p)\Lambda - \mathrm{d}N_s - \mu I_s \\ &\le (1-p)\Lambda - \mathrm{d}N_s, \end{aligned} \tag{5}$$

$$\begin{aligned} \frac{\mathrm{d}N_v}{\mathrm{d}t} &= p\Lambda - \mathrm{d}V - \mathrm{d}E_v - (d+\mu)I_v - \mathrm{d}I_v^e - \mathrm{d}R_v \\ &= p\Lambda - \mathrm{d}N_v - \mu I_v \\ &\le p\Lambda - \mathrm{d}N_v. \end{aligned} \tag{6}$$

According to the comparison theorem, it follows from (5) and (6) that

$$N_s(t) \le \left(N_s(0) - \frac{(1-p)\Lambda}{d}\right)e^{-\mathrm{d}t} + \frac{\Lambda}{d}, \tag{7}$$

$$N_v(t) \le \left(N_v(0) - \frac{p\Lambda}{d}\right)e^{-\mathrm{d}t} + \frac{\Lambda}{d}. \tag{8}$$

Using (7) and (8), we obtain that $\limsup_{t\longrightarrow\infty} N_s(t) = ((1-p)\Lambda/d)$ and $\limsup_{t\longrightarrow\infty} N_v(t) = p\Lambda/d$. Hence, the solutions of system (1) are uniformly ultimately bounded. Moreover, if the initial condition $\varphi(0) \in \Gamma$, then the solution $\varphi(t) \in \Gamma$, i.e., Γ is a positively invariant set for system (1). This completes the proof.

It is obvious that system (1) always has a disease-free equilibrium:

$$\begin{aligned} &E_0\left(S^0, E_s^0, I_s^0, I_s^{e0}, R_s^0, V^0, E_v^0, I_v^0, I_v^{e0}, R_v^0\right) \\ &= \left(\frac{(1-p)\Lambda}{d}, 0, 0, 0, 0, \frac{p\Lambda}{d}, 0, 0, 0, 0\right)^T. \end{aligned} \tag{9}$$

Next, we first introduce a very important threshold basic reproduction number R_0. The basic reproduction number R_0 is the number of secondary cases in which one case would produce in a completely susceptible population [40]. We calculate the basic reproduction number R_0 for system (1) by using the general calculation procedure given by Wang and Zhao [41]. It is easy to verify that system (1) satisfies the conditions (A1)–(A7) given in [41], and we have

$$\mathscr{F} = \begin{pmatrix} \beta_1(t)S\left(I_s + I_v\right) + \beta_2(t)S\left(I_s^e + I_v^e\right) \\ 0 \\ 0 \\ \beta_1(t)V\left(k_s I_s + I_v\right) + \beta_2(t)V\left(k_s I_s^e + I_v^e\right) \\ 0 \\ 0 \end{pmatrix},$$

$$\mathscr{V} = \begin{pmatrix} (d+\alpha)E_s \\ -\alpha\rho E_s + (d+\mu+\lambda_1)I_s \\ -\alpha(1-\rho)E_s + (d+\lambda_2)I_s^e \\ (d+\alpha)E_v \\ -\alpha\rho E_v + (d+\mu+\lambda_1)I_v \\ -\alpha(1-\rho)E_v + (d+\lambda_2)I_v^e \end{pmatrix}. \tag{10}$$

Thus, we obtain that

$$\overline{F} = \begin{pmatrix} 0 & \beta_1(t)S^0 & \beta_2(t)S^0 & 0 & \beta_1(t)S^0 & \beta_2(t)S^0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_1(t)k_sV^0 & \beta_2(t)k_sV^0 & 0 & \beta_1(t)V^0 & \beta_2(t)V^0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\overline{V} = \begin{pmatrix} d+\alpha & 0 & 0 & 0 & 0 & 0 \\ -\alpha\rho & d+\mu+\lambda_1 & 0 & 0 & 0 & 0 \\ -\alpha(1-\rho) & 0 & d+\lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & d+\alpha & 0 & 0 \\ 0 & 0 & 0 & -\alpha\rho & d+\mu+\lambda_1 & 0 \\ 0 & 0 & 0 & -\alpha(1-\rho) & 0 & d+\lambda_2 \end{pmatrix}. \tag{11}$$

Let $\Phi_{\overline{V}}(t)$ be the monodromy matrix of the linear ω-periodic system $\mathrm{d}z/\mathrm{d}t = -\overline{V}(t)z$, and $\rho(\Phi_{\overline{V}}(\omega))$ be the spectral radius of $\Phi_{\overline{V}}(\omega)$, where ω is the period. Let $Y(t,s)$ be the 6×6 matrix solution of the following initial value problem:

$$\begin{cases} \dfrac{\mathrm{d}Y(t,s)}{\mathrm{d}t} = -\overline{V}Y(t,s), \quad \forall t \geq s, \\ Y(s,s) = I, \end{cases} \tag{12}$$

where I is the 6×6 identity matrix.

Suppose that $\phi(s)$ is the initial distribution of infectious individuals, then $F(s)\phi(s)$ is the rate of new infections produced by the infectious individuals introduced at time s, and $Y(t,s)F(s)\phi(s)$ represents the distribution of those infectious individuals who are newly infected at time s and still remain in the infected compartment at time t for $t \geq s$. Then, one has that

$$\begin{aligned} \psi(t) &:= \int_{-\infty}^{t} Y(t,s)F(s)\phi(s)\mathrm{d}s \\ &= \int_0^{\infty} Y(t,t-a)F(t-a)\phi(t-a)\mathrm{d}a, \end{aligned} \tag{13}$$

is the distribution of accumulative new infections at time t produced by all those infected individuals $\phi(s)$ introduced at time previous to t. Let C_ω be the ordered Banach space of all ω periodic functions from $\mathbb{R}^6$ to $\mathbb{R}$ with maximum norm $\|\cdot\|$ and positive cone $C_\omega^+ := \{\phi(t) \in C_\omega : \phi(t) \geq 0, \forall t \in \mathbb{R}\}$. Then, we can introduce a linear operator $L: C_\omega \longrightarrow C_\omega$ by

$$\begin{aligned} (L\psi)(t) &:= \int_0^{\infty} Y(t,t-a)F(t-a)\phi(t-a)\mathrm{d}a, \\ &\qquad \forall t \in \mathbb{R}, \phi \in C_\omega. \end{aligned} \tag{14}$$

Following the results obtained in [41], the basic reproduction number R_0 for model (1) is defined as the spectral radius of operator L, i.e., $R_0 := \rho(L)$.

In order to analyze the dynamic behaviors of system (1), we have the following result.

Lemma 2 (Theorem 2.2 in [41]). *The following statements are valid:*

(i) $R_0 = 1$ *if and only if* $\rho(\Phi_{\overline{F}-\overline{V}}(\omega)) = 1$

(ii) $R_0 > 1$ *if and only if* $\rho(\Phi_{\overline{F}-\overline{V}}(\omega)) > 1$

(iii) $R_0 < 1$ *if and only if* $\rho(\Phi_{\overline{F}-\overline{V}}(\omega)) < 1$

Therefore, the disease-free equilibrium E_0 of system (1) is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$. Next, we will analyze the global dynamics of the disease-free equilibrium E_0.

Theorem 1. *If $R_0 < 1$, then the disease-free equilibrium E_0 of system (1) is globally asymptotically stable.*

Proof. According to Lemma 2, one has that $R_0 < 1$ is equivalent to $\rho(\Phi_{\overline{F}-\overline{V}}(\omega)) < 1$. Hence, we only need to prove that when $\rho(\Phi_{\overline{F}-\overline{V}}(\omega)) < 1$, E_0 is globally asymptotically stable.

Choose $\varepsilon > 0$ small enough such that $\rho(\Phi_{\overline{F}-\overline{V}+\varepsilon M(\omega)}(\omega)) < 1$, where

$$M(\omega) = \begin{pmatrix} 0 & \beta_1(t) & \beta_2(t) & 0 & \beta_1(t) & \beta_2(t) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_1(t)k_s & \beta_2(t)k_s & 0 & \beta_1(t) & \beta_2(t) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{15}$$

From Lemma 1, for $\forall \varepsilon > 0$, there exists $t_1 > 0$ such that for $t > t_1$, $S(t) \leq ((1-p)\Lambda/d) + \varepsilon$ and $V(t) \leq (p\Lambda/d) + \varepsilon$. Thus, for $t > t_1$, it follows from system (1) that

$$\begin{cases} \dfrac{dE_s}{dt} \le \beta_1(t)\left(\dfrac{(1-p)\Lambda}{d}+\varepsilon\right)(I_s+I_v) \\ \quad +\beta_2(t)\left(\dfrac{(1-p)\Lambda}{d}+\varepsilon\right)(I_s^e+I_v^e)-(d+\alpha)E_s, \\ \dfrac{dI_s}{dt} = \alpha\rho E_s-(d+\mu+\lambda_1)I_s, \\ \dfrac{dI_s^e}{dt} = \alpha(1-\rho)E_s-(d+\lambda_2)I_s^e, \\ \dfrac{dE_v}{dt} \le \beta_1(t)\left(\dfrac{p\Lambda}{d}+\varepsilon\right)(k_sI_s+I_v)+\beta_2(t)\left(\dfrac{p\Lambda}{d}+\varepsilon\right) \\ \quad \cdot(k_sI_s^e+I_v^e)-(d+\alpha)E_v, \\ \dfrac{dI_v}{dt} = \alpha\rho E_v-(d+\mu+\lambda_1)I_v, \\ \dfrac{dI_v^e}{dt} = \alpha(1-\rho)E_v-(d+\lambda_2)I_v^e. \end{cases} \tag{16}$$

Consider the following auxiliary system:

$$\frac{du}{dt} = A(\varepsilon)h, \tag{17}$$

where $h=(h_1,\ldots,h_6)^T$, $A(\varepsilon)=\overline{F}-\overline{V}+\varepsilon M(\omega)$. Following Lemma 2.1 in [42], there exists a positive ω-periodic function $\widehat{h}(t)=(\widehat{h}_1,\ldots,\widehat{h}_6)^T$ such that $h(t)=e^{\mu t}\widehat{h}(t)$ is a solution of system (17), where $\mu=(1/\omega)\ln\rho(\Phi_{\overline{F}-\overline{V}+\varepsilon M(\omega)}(\omega))$. Due to the continuity of functions and following Lemma 2, when $\varepsilon\longrightarrow 0$ and $R_0<1$, then $\rho(\Phi_{\overline{F}-\overline{V}+\varepsilon M(\omega)}(\omega))<1$. Hence, we have $u(t)\longrightarrow 0$ as $t\longrightarrow\infty$, which implies that the zero solution of system (17) is globally asymptotically stable. Based on the comparison principle, it follows from system (17) that

$$\lim_{t\longrightarrow\infty}\left(E_s(t),I_s(t),I_s^e(t),E_v(t),I_v(t),I_v^e(t)\right)^T=0. \tag{18}$$

Using (18), it follows from the fifth and last equations in system (1) that

$$\begin{aligned} \lim_{t\longrightarrow\infty}R_s(t)&=0,\\ \lim_{t\longrightarrow\infty}R_v(t)&=0. \end{aligned} \tag{19}$$

Using (18) and (19), it follows from the first and sixth equations in system (1) that

$$\begin{aligned} \lim_{t\longrightarrow\infty}S(t)&=\frac{(1-p)\Lambda}{d},\\ \lim_{t\longrightarrow\infty}V(t)&=\frac{p\Lambda}{d}. \end{aligned} \tag{20}$$

Thus, E_0 is globally asymptotically stable when $R_0<1$. This completes the proof.

Next, we will analyze the persistence of system (1). Define

$$\begin{aligned} X_0=\big\{&(S,E_s,I_s,I_s^e,R_s,V,E_v,I_v,I_v^e,R_v)^T\in\Gamma:\\ &E_s>0,I_s>0,I_s^e>0,R_s>0,E_v>0,I_v>0,I_v^e>0,R_v>0\big\},\\ \partial X_0=&\Gamma\backslash X_0. \end{aligned} \tag{21}$$

Assume that $u(t,\varphi(0))$ is the solution of system (1) with the initial value $\varphi(0)$. According to the fundamental existence-uniqueness theorem in [43], $u(t,\varphi(0))$ is unique. Denote that P: $\Gamma\longrightarrow\Gamma$ is the Poincaré map with respect to system (1), i.e.,

$$P(\varphi(0))=u(\omega,\varphi(0)),\quad \forall\varphi(0)\in\Gamma, \tag{22}$$

where ω is the period. It is obvious that

$$P^m(E_0)=u(m\omega,E_0),\quad \forall m\ge 0. \tag{23}$$

It follows from Lemma 1 that Γ is positively invariant and P is point dissipative for system (1). In order to analyze the persistence for system (1), we give the following lemma.

Lemma 3. *If $R_0>1$, then there exists $\delta>0$ such that for any $\varphi(0)\in X_0$ with $\|\varphi(0)-E_0\|\le\delta$, we have*

$$\limsup_{m\longrightarrow\infty}D(P^m(\varphi(0)),E_0)\ge\delta, \tag{24}$$

where D is a distance function in X_0.

Proof. According to Lemma 1, if $R_0>1$, then $\rho(\Phi_{\overline{F}-\overline{V}}(\omega))>1$. Choose $\epsilon>0$ small enough such that $\rho(\Phi_{\overline{F}-\overline{V}-\varepsilon M(\omega)}(\omega))>1$. Now, we prove that

$$\limsup_{m\longrightarrow\infty}D(P^m(\varphi(0)),E_0)\ge\delta. \tag{25}$$

If it is false, then

$$\limsup_{m\longrightarrow\infty}D(P^m(\varphi(0)),E_0)<\delta, \tag{26}$$

for some $\varphi(0)\in X_0$. Assume that there exists $m>0$ such that $D(P^m(\varphi(0)),E_0)<\delta$. According to the continuity of the solutions associated with the initial conditions, when $\|\varphi(0)-E_0\|\le\delta$, it follows that

$$\left\|u(t_1,P^m(\varphi(0)))-u(t_1,E_0)\right\|<\varepsilon,\quad m\ge 0,t_1\in[0,\omega]. \tag{27}$$

For $\forall t\ge 0$, let $t=m\omega+t_1$, where $t_1\in[0,\omega]$ and $m=[t/\omega]$, which is the greatest integer less than or equal to t/ω. Then, for $\forall t\ge 0$, we have

$$\left\|u(t,\varphi(0))-u(t,E_0)\right\|=\left\|u(t_1,P^m(\varphi(0)))-u(t_1,E_0)\right\|<\epsilon,\quad t_1\in[0,\omega], \tag{28}$$

which implies that $((1-p)\Lambda/d)-\epsilon\le S(t)\le((1-p)\Lambda/d)+\epsilon$, $(p\Lambda/d)-\epsilon\le V(t)\le(p\Lambda/d)+\epsilon, t\ge 0$. Then, for $\|\varphi(0)-E_0\|\le\delta$, one has

$$\begin{cases} \dfrac{dE_s}{dt} \geq \beta_1(t)\left(\dfrac{(1-p)\Lambda}{d} - \varepsilon\right)(I_s + I_v) \\ \quad + \beta_2(t)\left(\dfrac{(1-p)\Lambda}{d} - \varepsilon\right)(I_s^e + I_v^e) - (d+\alpha)E_s, \\ \dfrac{dI_s}{dt} = \alpha\rho E_s - (d+\mu+\lambda_1)I_s, \\ \dfrac{dI_s^e}{dt} = \alpha(1-\rho)E_s - (d+\lambda_2)I_s^e, \\ \dfrac{dE_v}{dt} \geq \beta_1(t)\left(\dfrac{p\Lambda}{d} - \varepsilon\right)(k_s I_s + I_v) + \beta_2(t)\left(\dfrac{p\Lambda}{d} - \varepsilon\right) \\ \quad \cdot (k_s I_s^e + I_v^e) - (d+\alpha)E_v, \\ \dfrac{dI_v}{dt} = \alpha\rho E_v - (d+\mu+\lambda_1)I_v, \\ \dfrac{dI_v^e}{dt} = \alpha(1-\rho)E_v - (d+\lambda_2)I_v^e. \end{cases} \tag{29}$$

Consider the following auxiliary linear system:

$$\frac{dg}{dt} = \overline{A}(\epsilon)g. \tag{30}$$

As the similar proof process in Theorem 1, there exists a positive ω-periodic function $\widehat{g}(t) = (\widehat{g}_1, \ldots, \widehat{g}_6)^T$ such that $g(t) = e^{\mu t}\widehat{g}(t)$ is a solution of system (30), where $\mu = (1/\omega)\ln\rho(\Phi_{\overline{F}-\overline{V}-\varepsilon M(\omega)}(\omega))$. Since $\mu = (1/\omega)\ln\rho(\Phi_{\overline{F}-\overline{V}-\varepsilon M(\omega)}(\omega)) > 1$, it follows that if $g(0) > 0$, then $g(t) \longrightarrow \infty$ as $t \longrightarrow \infty$. From the comparison principle, we have

$$\begin{aligned} &E_s(t) \longrightarrow \infty, \\ &I_s(t) \longrightarrow \infty, \\ &I_s^e(t) \longrightarrow \infty, \\ &E_v(t) \longrightarrow \infty, \\ &I_v(t) \longrightarrow \infty, \\ &I_v^e(t) \longrightarrow \infty, \end{aligned} \tag{31}$$

as

$$\begin{aligned} &E_s(0) > 0, \\ &I_s(0) > 0, \\ &I_s^e(0) > 0, \\ &E_v(0) > 0, \\ &I_v(0) > 0, \\ &I_v^e(0) > 0, \end{aligned} \tag{32}$$

for $t \longrightarrow \infty$, which is a contradiction with Lemma 1. Therefore, $\limsup_{m\longrightarrow\infty} D(P^m(\varphi(0)), E_0) \geq \delta$. This completes the proof.

Theorem 2. *If $R_0 > 1$, then system (1) is uniformly persistent and has at least one positive periodic solution.*

Proof. First, we prove that system (1) is uniformly persistent, i.e., there exists $\delta > 0$ such that any solution $\varphi(t)$ of system (1) with the initial value $\varphi(0) \in X_0$ satisfies

$$\begin{aligned} &\liminf_{t\longrightarrow\infty} \left(E_s(t), I_s(t), I_s^e(t), R_s(t), E_v(t), I_v(t), I_v^e(t), R_v(t)\right)^T \\ &\geq (\delta, \delta, \delta, \delta, \delta, \delta, \delta, \delta)^T. \end{aligned} \tag{33}$$

For $\forall\varphi(0) \in X_0$, solving system (1), we obtain that

$$\begin{cases} S(t) = e^{-\int_0^t \left[d+\beta_1(t)\left(I_s(u)+I_v(u)\right)+\beta_2(t)\left(I_s^e(u)+I_v^e(u)\right)\right]du}\left[S(0) + \int_0^t ((1-p)\Lambda + \eta_s R_s)e^{\int_0^t \left[d+\beta_1(t)\left(I_s(u)+I_v(u)\right)+\beta_2(t)\left(I_s^e(u)+I_v^e(u)\right)\right]du}du\right], \\ E_s(t) = e^{-(\alpha+d)t}\left[E_s(0) + \int_0^t [\beta_1(t)S(u)\left(I_s(u)+I_v(u)\right) + \beta_2(t)S(u)\left(I_s^e(u)+I_v^e(u)\right)]du\right], \\ I_s(t) = e^{-(d+\mu+\lambda_1)t}\left[I_s(0) + \int_0^t \alpha\rho E_s(u)du\right], \\ I_s^e(t) = e^{-(d+\lambda_2)t}\left[I_s^e(0) + \int_0^t \alpha(1-\rho)E_s(u)du\right], \\ R_s(t) = e^{-(d+\eta_s)t}\left[R_s(0) + \int_0^t (\lambda_1 I_s(u) + \lambda_2 I_s^e(u))du\right], \\ V(t) = e^{-\int_0^t \left[d+\beta_1(t)\left(k_s I_s(u)+I_v(u)\right)+\beta_2(t)\left(k_s I_s^e(u)+I_v^e(u)\right)\right]du}\left[V(0) + \int_0^t (p\Lambda + \eta_s R_s)e^{\int_0^t \left[d+\beta_1(t)\left(k_s I_s(u)+I_v(u)\right)+\beta_2(t)\left(k_s I_s^e(u)+I_v^e(u)\right)\right]du}du\right], \\ E_v(t) = e^{-(\alpha+d)t}\left[E_v(0) + \int_0^t [\beta_1(t)V(u)\left(k_s I_s(u)+I_v(u)\right) + \beta_2(t)V(u)\left(k_s I_s^e(u)+I_v^e(u)\right)]du\right], \\ I_v(t) = e^{-(d+\mu+\lambda_1)t}\left[I_v(0) + \int_0^t \alpha\rho E_v(u)du\right], \\ I_v^e(t) = e^{-(d+\lambda_2)t}\left[I_v^e(0) + \int_0^t \alpha(1-\rho)E_v(u)du\right], \\ R_v(t) = e^{-(d+\eta_s)t}\left[R_v(0) + \int_0^t (\lambda_1 I_v(u) + \lambda_2 I_v^e(u))du\right]. \end{cases} \tag{34}$$

From (34), it is obvious that for $t \geq 0$, the solution $\varphi(t) \in X_0$ for system (1) as the initial value $\varphi(0) \in X_0$. This means that X_0 is positively invariant. In addition, ∂X_0 is relatively closed in Γ, and following Lemma 1, the discrete-time system $\{P^m\}_{m\geq 0}$ admits a global attractor in Γ. Then, we set

$$M_\partial := \{\varphi(0) \in \partial X_0 : P^m(\varphi(0)) \in \partial X_0, \forall m \geq 0\}. \tag{35}$$

Now, we prove that

$$M_\partial = \left\{(S, 0, 0, 0, 0, V, 0, 0, 0, 0)^T \in \partial X : S \geq 0, V \geq 0\right\} \triangleq M'_\partial. \tag{36}$$

It is easy to know that $M'_\partial \subseteq M_\partial$, then we only need to prove $M_\partial \subseteq M'_\partial$. That is, for any

$$\begin{aligned}\varphi(0) = (&S(0), E_s(0), I_s(0), I_s^e(0), R_s(0), V(0), E_v(0),\\ &I_v(0), I_v^e(0), R_v(0))^T \in \partial X_0,\end{aligned} \tag{37}$$

we have

$$\begin{aligned}I_s(m\omega) = I_s^e(m\omega) = R_s(m\omega) = E_v(m\omega) = I_v(m\omega)\\ = I_v^e(m\omega) = R_v(m\omega) = 0.\end{aligned} \tag{38}$$

If not, then there exists $m_1 \geq 0$ such that

$$\begin{aligned}&(I_s(m_1\omega), I_s^e(m_1\omega), R_s(m_1\omega), E_v(m_1\omega), I_v(m_1\omega),\\ &I_v^e(m_1\omega), R_v(m_1\omega))^T \geq 0.\end{aligned} \tag{39}$$

If we replace the initial time $t = 0$ with $t = m_1\omega$, then it follows from (34) that

$$\begin{aligned}E_s(t) > 0,\\ I_s(t) > 0,\\ I_s^e(t) > 0,\\ R_s(t) > 0,\\ E_v(t) > 0,\\ I_v(t) > 0,\\ I_v^e(t) > 0,\\ R_v(t) > 0,\end{aligned} \tag{40}$$

for $\forall t > m_1\omega$. This means that $P^m(\varphi(0)) \notin \partial X_0$, for $\forall m > m_1$, which contradicts the definition of M_∂. Therefore, it follows that $M_\partial \subseteq M'_\partial$. Moreover, there only exists one fixed point E_0 of P in M_∂.

According to Lemma 3, E_0 is an isolated invariant set in Γ. From the acyclicity theorem on uniform persistence for maps (Theorem 1.3.1 in [44]), it follows that P is uniformly persistent with respect to $(X_0, \partial X_0)$. Moreover, according to Theorem 1.3.1 in [44], the solutions of system (1) are uniformly persistent with respect to $(X_0, \partial X_0)$, i.e., there exists $\delta > 0$ such that any solution $\varphi(t)$ of system (1) with the initial value $\varphi(0) \in X_0$ satisfies

$$\begin{aligned}\liminf_{t\longrightarrow\infty} &(E_s(t), I_s(t), I_s^e(t), R_s(t), E_v(t), I_v(t), I_v^e(t), R_v(t))^T\\ &\geq (\delta, \delta, \delta, \delta, \delta, \delta, \delta, \delta)^T.\end{aligned} \tag{41}$$

Next, we prove that system (1) has a positive periodic solution. Theorem 1.3.6 in [44] implies that P has a fixed point

$$(S^*(0), E_s^*(0), I_s^*(0), I_s^{e*}(0), R_s^*(0), V^*(0), E_v^*(0), I_v^*(0), I_v^{e*}(0), R_v^*(0))^T \in X_0. \tag{42}$$

Moreover, we can also obtain that $S^*(0) > 0$ for system (1). Suppose not, assume $S^*(0) = 0$. By the periodicity of P, we have $S^*(0) = S^*(n\omega), n = 1, 2, 3, \ldots$. From the first equation of (34), one has

$$\begin{aligned}S^*(n\omega) = &e^{-\int_0^{n\omega}\left[d+\beta_1(t)\left(I_s(u)+I_v(u)\right)+\beta_2(t)\left(I_s^e(u)+I_v^e(u)\right)\right]\mathrm{d}u}\\ &\cdot\left[S^*(0) + \int_0^t ((1-p)\Lambda + \eta_s R_s)e^{\int_0^{n\omega}\left[d+\beta_1(t)\left(I_s(u)+I_v(u)\right)+\beta_2(t)\left(I_s^e(u)+I_v^e(u)\right)\right]\mathrm{d}u}\mathrm{d}u\right] > 0,\end{aligned} \tag{43}$$

which contradicts with $S^*(0) = S^*(n\omega) = 0$. Thus, it follows that $S^*(0) > 0$. Then, from (34), it is easy to obtain that $S^*(t) > 0, E_s^*(t) > 0, I_s^*(t) > 0, I_s^{e*}(t) > 0, R_s^*(t) > 0, V^*(t) > 0, E_v^*(t) > 0, I_v^*(t) > 0, I_v^{e*}(t) > 0, R_v^*(t) > 0$, for $t \geq 0$. Due to the definition of semiflow P, we obtain that

$$\begin{aligned}&(S^*(t), E_s^*(t), I_s^*(t), I_s^{e*}(t), R_s^*(t), V^*(t), E_v^*(t), I_v^*(t),\\ &I_v^{e*}(t), R_v^*(t))^T,\end{aligned} \tag{44}$$

is a positive ω-periodic solution of system (1). This completes the proof.

4. Simulations and Sensitivity Analysis

In this section, by using system (1), we simulate the reported data of HFMD in mainland China from January 2008 to June 2019, predict the trend of the disease, and seek for some prevention and control strategies. We obtain the monthly number of newly reported HFMD cases in mainland China

from the website of the Chinese Center for Disease Control and Prevention (CCDC) [9] and the National Health and Family Planning Commission of the People's Republic of China (NHFPC) [8]. We only consider the population of young children aged 6–71 months, who are a high-risk group for HFMD [31, 32] and get demographic information from the National Bureau of Statistics of China (NBSC) [38]. From NBSC [38], we obtain the monthly number of children entering the age of 6–71 months $\Lambda = 1.41 \times 10^6$ and the average monthly rate of leaving the children group aged 6–71 months $d = 1.30 \times 10^{-2}$. From CCDC [9] and NHFPC [8], we get the disease-related death rate $\mu = 3.40 \times 10^{-4}$. The proportion of infectious individuals without vaccination caused by HFMD viruses except EV71 is $k_s = 0.56$ [39]. Other parameters can be obtained from the literature or assumed on the basis of common sense (see Table 1). Unfortunately, there is no public data on the vaccinated rate p and the proportion of symptomatic infectious individuals ρ. Therefore, we estimate the values of unknown parameters containing p, ρ and β_1, β_2. By using the BP neural network algorithm [45] (see Appendix), we estimate parameters of system (1) by calculating the following error function:

$$\mathrm{ER} = \frac{1}{n}\sum_{i=1}^{n}\left(I_i - \widehat{I}_i\right)^2, \tag{45}$$

where $\widehat{I}_i$ is the reported data of HFMD at time t. I_i is the numerically computed solution of $I(t)$ of system (1) at time t. n is the total number of sample points collected at all times.

Next, we set the initial conditions for system (1). We take January 2008 as the initial time. It notes that the EV71 vaccine has been vaccinated in mainland China starting in 2016, then vaccinated rate p of system (1) is zero before 2016 [13, 14]. Hence, since the total number of vaccinated individuals $N_v(t) = 0$ before January 2016, then it follows that $V(0) = 0, E_v(0) = 0, I_v(0) = 0, I_v^e(0) = 0, R_v(0) = 0$. The initial number of the susceptible human population aged 6–71 months of 2008 is calculated as $S(0) = 9.23 \times 10^7$ from NBSC [8], and the number of the initial unvaccinated symptomatic infectious humans $I_s(0) = 2.10 \times 10^4$ is from CCDC [9]. However, the initial number of the unvaccinated exposed humans $E_v(0)$, the unvaccinated asymptomatic infectious humans $I_s^e(0)$, and the unvaccinated recovered humans $R_s(0)$ cannot be obtained. We derive $E_v(0)$ reversely by the parameter $\alpha\rho$; get $I_s^e(0)$ by using the proportion parameter ρ to compare with $I_s(0)$, and $R_s(0)$ is estimated roughly. The numerical simulation of the model on the number of HFMD cases in mainland China from January 2008 to June 2019 is shown in Figure 3, and its fitting error ER (45) is calculated as 6.71×10^{-4}. In addition, in order to test how well system (1) actually reflects the data, by using the method in [22, 46], we consider the following hypotheses.

Null hypothesis, H_0: the estimated parameters of system (1) are equal to actual values.

Alternative hypothesis, H_1: the estimated parameters of system (1) are not equal to actual values.

Since the EV71 vaccine was only started in 2016, we consider hypothesis test for the two time periods from January 2008 to December 2015 and January 2016 to June 2019, and the chi-square values and degrees of freedom are shown in Table 2. Then, from Table 2, we obtain that it cannot reject the null hypothesis at the 5% significant level by Pearson's criterion of chi-square test [47]. Therefore, it indicates that, with these parameter values, there is a good fit between the simulation of model (1) and the HFMD cases in mainland China. Furthermore, based on our parameters listed in Table 1, we calculate the basic reproduction number as $R_0 = \Phi_{\overline{F}-\overline{V}}(24) = 1.09$ after 2015, for this case where the vaccination is considered, which means that system (1) admits at least a positive periodic solution (see Theorem 2).

It notes the initial conditions except $I_s(0)$ adopted in model fitting are mostly assumed and back-extrapolated by parameters. Thus, it is necessary to analyze the influence of the initial values $S(0), E_s(0), I_s(0), I_s^e(0), R_s(0)$ on the number of infected cases $I_s(t) + I_v(t)$, which is illustrated in Figure 4. Figure 4 shows that the initial value $S(0)$ has a greater impact on $I_s(t) + I_v(t)$ while other initial values have little or no impact on $I_s(t) + I_v(t)$. This implies that the selection of the initial values $S(0)$ is very important for our simulation. According to [6, 7, 10, 11, 32], due to most of the infected individuals of HFMD are 6–71 months old, then we carefully select that the susceptible individuals are 6–71 months old.

In order to investigate the effect of the EV71 vaccination strategy on HFMD in mainland China, we compare the trends of HFMD between with vaccination and without vaccination by simulating model (1), which is presented in Figure 5. In Figure 5, the solution curve without considering vaccination is on top of the one with considering vaccination after 2015, which means that the EV71 vaccination strategy is effective in reducing the spread of HFMD in mainland China in the last two years. In addition, we calculate the basic reproduction number R_0 of system (1) without considering vaccination ($p = 0$) as 1.24 and with considering vaccination ($p = 0.23$) as 1.09, so system (1) in each case admits a positive periodic solution according to Theorem 2 (see the solution curves from 2030 to 2035 in Figure 5). In addition, Figure 6 shows that under the current vaccination measures, the total number of patients in 2016, 2017, and 2018 has reduced about 4.62×10^5, 5.16×10^5, and 7.39×10^5, respectively.

To compare the performance of the periodic transmission functions in this paper with some previous research studies, by using system (1), we consider the following two cases:

Periodic functions (I): choose the transmission functions in this paper such as $\beta_1(t) = a_0[1 + a_1 \sin((\pi/2)t) + a_2 \sin((\pi/6)t) + a_3 \sin((\pi/12)t) + a_4 \sin(-1 + (\pi/3)t)]$ and $\beta_2(t) = b_0[1 + b_1 \sin(-1 + (\pi/2) + (\pi/4)t) + b_2 \sin((\pi/6)t) + b_3 \sin((\pi/12)t)]$, where the values of $a_i, b_i, i = 1, \ldots, 3$ are given in Table 1.

Periodic functions (II): choose the transmission functions in the previous research studies [26–28, 30] such as $\overline{\beta}_1(t) = \overline{a}_0[1 + \overline{a}_1 \sin((\pi/6)t)]$ and $\overline{\beta}_2(t) = \overline{b}_0[1 + \overline{b}_1 \sin((\pi/6)t)]$. Based on our parameter values in Table 1, we estimate that the optimal values of $\overline{a}_0, \overline{a}_1, \overline{b}_0, \overline{b}_1$ are $1.19 \times 10^{-7}, 0.23, 0.21, 2.03 \times 10^{-8}$, respectively.

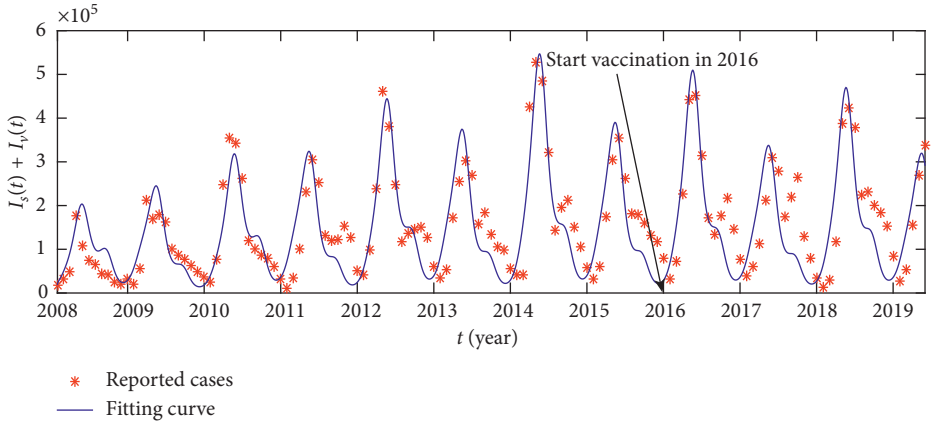


FIGURE 3: The comparison between the reported HFMD cases in mainland China and the simulation of $I_s(t) + I_v(t)$ of system (1) from January 2008 to June 2019. The values of parameters are given in Table 1. The initial conditions are $S(0) = 9.23 \times 10^7, E_s(0) = 8 \times 10^4, I_s(0) = 2 \times 10^4, I_s^e(0) = 6 \times 10^4, R_s = 1 \times 10^5, V(0) = 0, E_v(0) = 0, I_v(0) = 0, I_v^e(0) = 0, R_v(0) = 0$. It should be noted that we take $k = 0$ before 2016.

TABLE 2: Chi-square values and degrees of freedom from January 2008 to December 2015 and January 2016 to June 2019.

Year	Jan. 2008–Dec. 2015	Jan. 2016–June 2019
Chi-square value	68.07	33.19
DF	82	28
AV	104.14	41.34

Note: AV denotes the accepting value is at 5% significant level with degrees of freedom DF.

The studies in [48] showed that the trend of monthly new cases of infectious diseases is usually closely related to the transmission functions. It notes that the periods of $\overline{\beta}_1(t)$ and $\overline{\beta}_2(t)$ are 12 months, and these functions have only one peak in each year. Meanwhile, the periods of $\beta_1(t)$ and $\beta_2(t)$ are 24 months, and these functions have two peaks in each year. As explained in Introduction, the transmission functions $\beta_1(t), \beta_2(t)$ are closer to the trend of actual data change in mainland China than $\overline{\beta}_1(t), \overline{\beta}_2(t)$. Moreover, in Figure 7, the simulation of $I_s(t) + I_v(t)$ of system (1) is based on the periodic functions (I) (i.e., $\beta_1(t), \beta_2(t)$) but the periodic functions (II) (i.e., $\overline{\beta}_1(t), \overline{\beta}_2(t)$) can well fit the two phenomena, where HFMD has a two-year epidemic cycle and the number of the cases per year has a slight upward trend in September and October in mainland China (see the reasons in Introduction).

Moreover, the period of the solution for system (1) with periodic functions (II) is 12 months, which fails to reflect the 24 months periodic variation of the disease, while these biological phenomena are well presented by the solution for system (1) with periodic functions (I). Furthermore, we calculate the fitting error ER for system (1) with periodic functions (I) and with periodic functions (II) which is 4×10^{-3} and 1.20×10^{-2}, respectively. On the whole, our model, especially based on our selection of the transmission rate functions, can well reveal the objective laws of the spread of HFMD in mainland China.

According to Theorems 1 and 2, we have known that the basic reproduction number R_0, which determines the outbreak of the disease or extinction, is an important quantity in characterizing the spread of disease. Therefore, to find some effective prevention and control measures for HFMD in mainland China, we perform sensitivity analysis to determine the influence of some parameters on R_0. Due to some parameters of system (1) for HFMD are difficult or not to control such as Λ, d, α, ρ and so on, biologically, then we only consider the influence of the parameters $p, \beta_1, \beta_1, \lambda_1$ on R_0, which are depicted in Figure 8, and its related prevention and control measures are shown in Figure 9. Under the assumption of low vaccine coverage, choosing $p \le 0.3$, Figure 8(a) shows that R_0 decreases with the increase of p. This implies that the larger vaccinated rate p ($p \le 0.3$) is, the less number of symptomatic infectious individuals $I_s + I_v$ is (see Figure 9(a)). Moreover, Figure 8(a) also shows that R_0 is always larger than 1 as $p \le 0.3$, which means that without other prevention and control, the disease is not extinct. Figures 8(b) and 8(c) show that the less the baseline contact rate of β_1 or β_2 is, the less R_0 is. Moreover, when the baseline contact rate $a_0 < 9 \times 10^{-8}$ or $b_0 < 1.45 \times 10^{-8}$, R_0 can be less than 1, which means that the disease is extinct. However, based on our estimating parameter values, a_0 or b_0 is far greater than 9×10^{-8} or 1.45×10^{-8}, which means that HFMD becomes an endemic disease in mainland China. If the relevant departments can effectively control the contact rate between the healthy children and the infectious children and improve hygienic precautions and environmental cleaning, such as washing hands before meals and after using the toilet, and making air fresh indoors and so on, the disease can be effectively control or extinct (see Figures 9(b) and 9(c)). Figure 8(d) shows that R_0 decreases with the increasing of the recovery rate of symptomatic infectious individuals λ_1, i.e., if average recovery period $1/\lambda_1$ of symptomatic infectious individuals is less, then R_0 is less, and the disease is extinct if $\lambda_1 > 4.28$. Figure 9(d) shows that the number of symptomatic infectious cases decreases as the recovery rate λ_1 increases, especially the disease is extinct when $\lambda_1 = 4.73$. Therefore, we suggest that the patients should seek medical advice in time to achieve diagnosis and

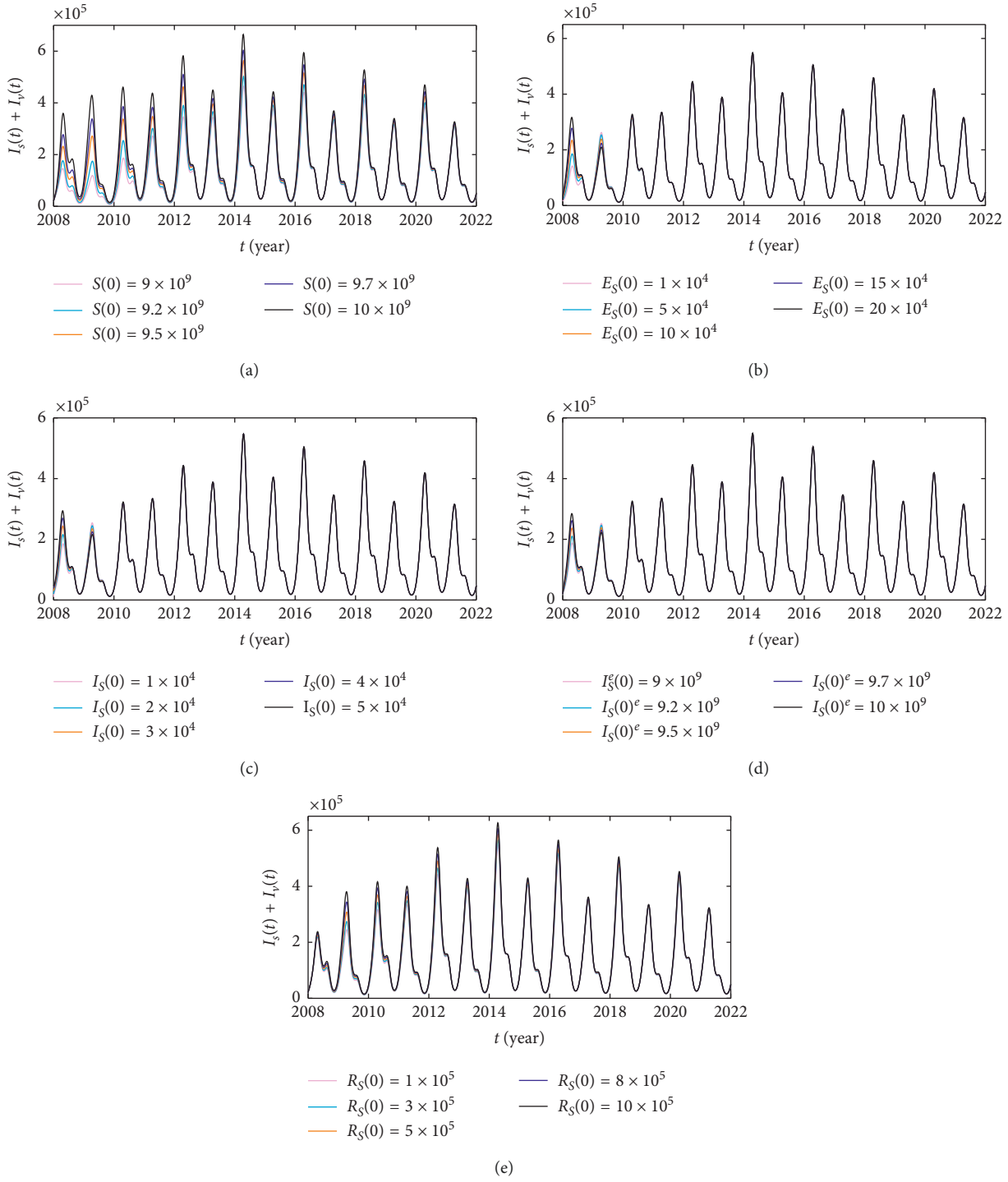


Figure 4: The influence of initial conditions of humans on the number of HFMD cases $I_s(t) + I_v(t)$: (a) different values of $S(0)$; (b) different values of $E_s(0)$; (c) different values of $I_s(0)$; (d) different values of $I_s^e(0)$; (e) different values of $R_s(0)$.

treatment early. At the same time, because the disease does not have targeted drugs, some targeted drugs need to be developed actively. Finally, in order to further discuss the credibility of our results, we will evaluate and discuss the consistency of the scope of R_0 with other studies in China. Although the ideas of dynamic modeling for different studies based on the HFMD cases in China are not the same, the range of R_0 of models may have similar characteristics. For instance, the study in [30], which assumed that the vaccine was against all HFMD-associated enteroviruses and analyzed the HFMD cases in mainland China from 2010 to 2014, showed that as the vaccination rate, the baseline contact rate of symptomatic infectious individuals and the baseline contact rate of asymptomatic infectious individuals change, the ranges of R_0 are $(1.15, 1.58)$, $(1.37, 1.58)$, and $(0.48, 1.58)$, respectively. The study in [28], based on the

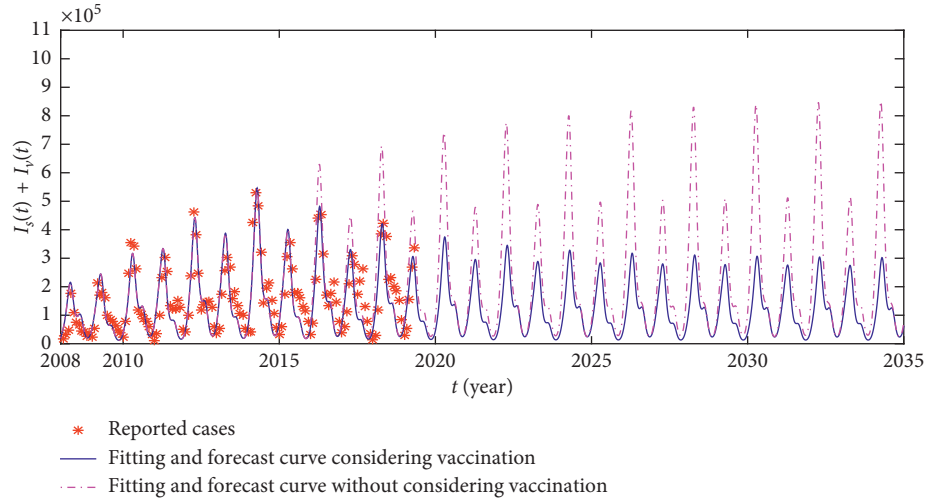


FIGURE 5: The tendency of HFMD from system (1) with vaccination ($p = 0$ before 2016 and $p = 0.23$ in others) and without vaccination in a long time ($p = 0$ in all times). The values of other parameters and the initial conditions are the same as in Figure 3.

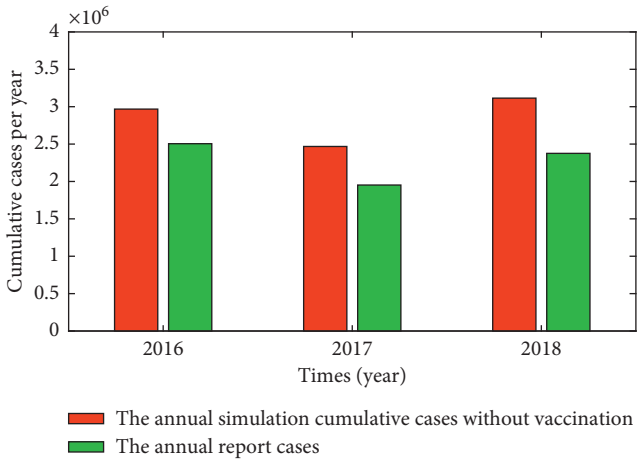


FIGURE 6: The comparison between the annual simulation cumulative cases by system (1) without vaccination ($p = 0$) and the annual reported cumulative cases for HFMD in mainland China from 2016 to 2018. The values of other parameters and the initial conditions are the same as in Figure 3.

HFMD cases in Shandong Province, China, which mainly discussed the impacts of quarantine measures and the fraction of asymptomatic infectious individuals on the spread of HFMD, indicated that, as the quarantine rate and the fraction of asymptomatic infectious individuals change, the ranges of R_0 are $(1.04, 1.14)$ and $(1.02, 1.05)$. In a word, most studies for HFMD in China (see paper [27, 28, 30, 49]) showed that the range of R_0 is around 1 and not far away from 1, which is similar to the change of R_0 shown in Figure 8. Therefore, these discussions can also indirectly improve the credibility of our results.

5. Discussion

In this paper, we have proposed and analyzed a HFMD model with periodic transmission rates to take into account seasonal outbreak of HFMD infection and to consider EV71 vaccination. The studies in [29, 30] have assumed that the vaccine is effective against all enteroviruses, which may not be in line with the current situation of HFMD transmission in mainland China. In fact, one vaccine is available in mainland China and it works only for EV71 but not for others. To investigate the effects of EV71 vaccination on HFMD in the presence of multiple pathogenic viruses, our proposed model divided the total individuals into the unvaccinated and the vaccinated, which were described in different compartments (see Section 2). Moreover, we have chosen novel periodic transmission functions, unlike the previous studies in [4, 26, 28, 30], to reflect the two-year periodic nature and the infections' change with the opening and closing of schools in mainland China (see Sections 1 and 2).

We have analyzed the global dynamics of system (1) by evaluating the basic reproduction number $R_0 := \rho(L)$: if $R_0 < 1$, then the disease-free equilibrium of system (1) is globally asymptotically stable (see Theorem 1), which means that the disease is extinct; if $R_0 > 1$, then system (1) uniformly persists and has at least one positive periodic solution (see Theorem 2), which means that the disease persists or has not been eliminated.

By fitting our model to the reported data on symptomatic cases of HFMD in mainland China, we estimated the basic reproduction number $R_0 = 1.24$ from 2008 to 2015 and $R_0 = 1.09$ from 2016 to June 2019. The major factors for this difference are that EV71 vaccination has been used since 2016, the reason of which can also be seen in the analyses for the influence of the vaccinated rate p on R_0 in Figure 8(a). Moreover, the periodic transmission functions chosen in our paper perform efficiently for system (1) to fit the reported data in mainland China, given the analyzes of Figure 7. Furthermore, to find effective prevention and control measures for HFMD in mainland China, we did sensitivity analysis to determine the influence of some artificial control parameters on R_0. It showed that under the assumption of low vaccine coverage, increasing the vaccinated rate with respect to EV71 vaccination can reduce the number of

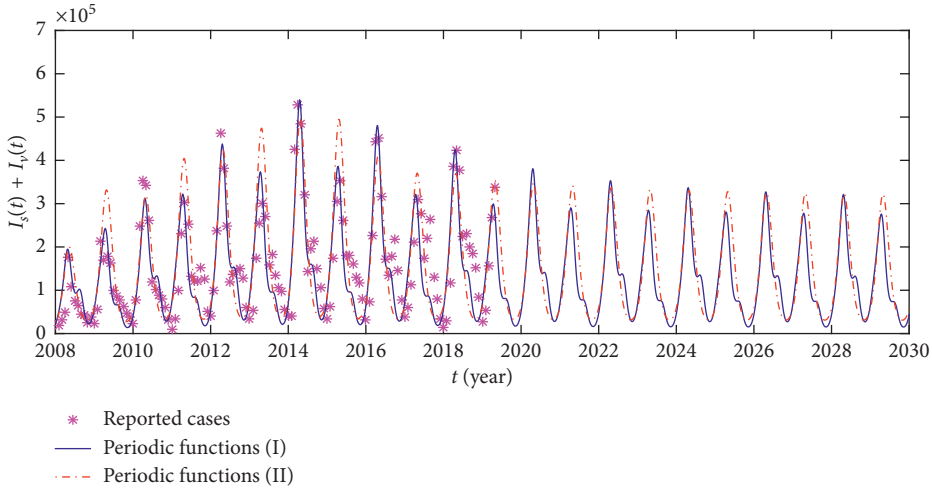


Figure 7: The tendency of HFMD from system (1)with periodic functions (I) and with periodic functions (II). The values of other parameters and the initial conditions are the same as in Figure 3.

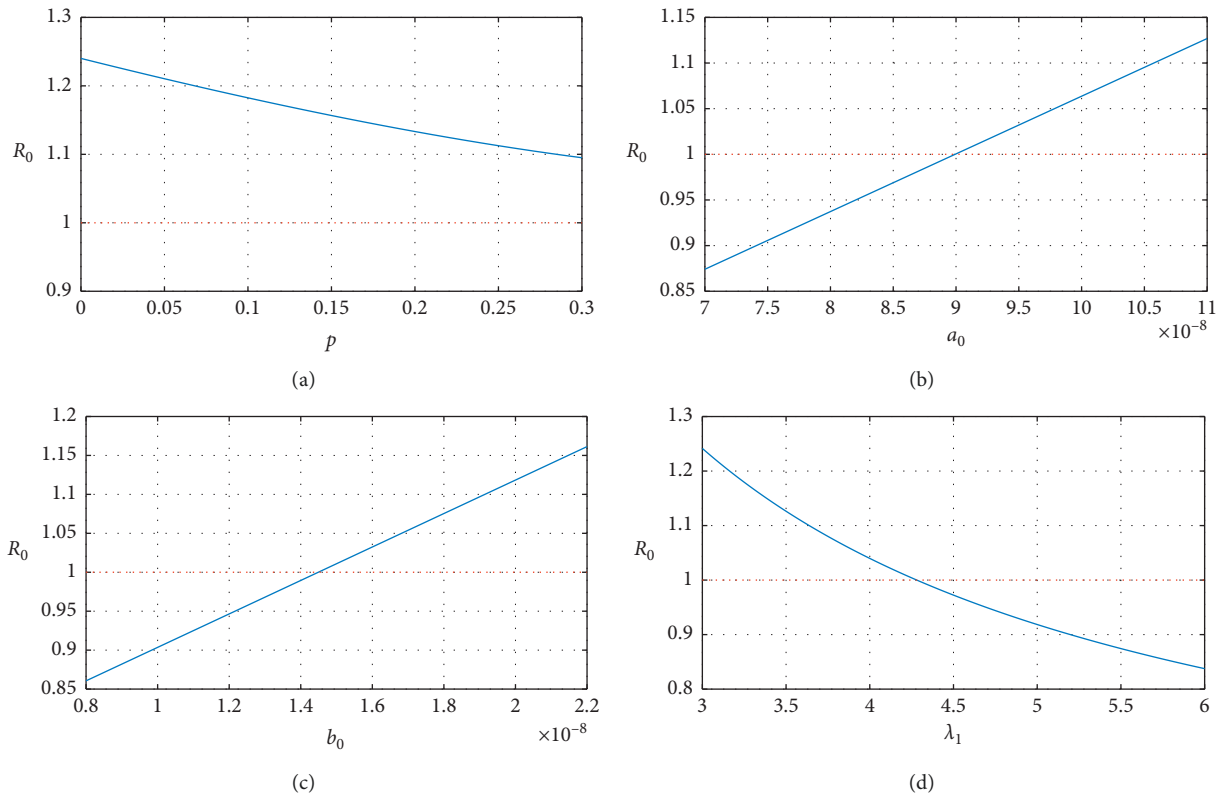


Figure 8: The influence of parameters on R_0 (a) versus p; (b) versus a_0, being the baseline contact rate of β_1; (c) versus b_0, being the baseline contact rate of β_2; (d) versus λ_1. Other parameter values are given in Table 1.

infections, but the disease does not go extinct. If we can effectively control the contact rate between the healthy children and the infectious children (see the analyses of Figures 8(b) and 8(c) and Figures 9(b) and 9(c)) and improve the medical conditions (see the analyses of Figures 8(d) and 9(d)), the disease may be eliminated. In a word, the proposed prevention and control measures may be useful for public health governance.

It notes that the observed aggregated monthly HFMD cases in Figure 1 are a mixture of different seasonality in mainland China. Also, Xing et al.'s study of HFMD in mainland China from 2008 to 2012 [33] showed that seasonal diversities (although this was not the only factor) made the spread of HFMD present different characteristics in different regions, such as the annual amplitude of HFMD epidemics increased with increasing latitude and semiannual

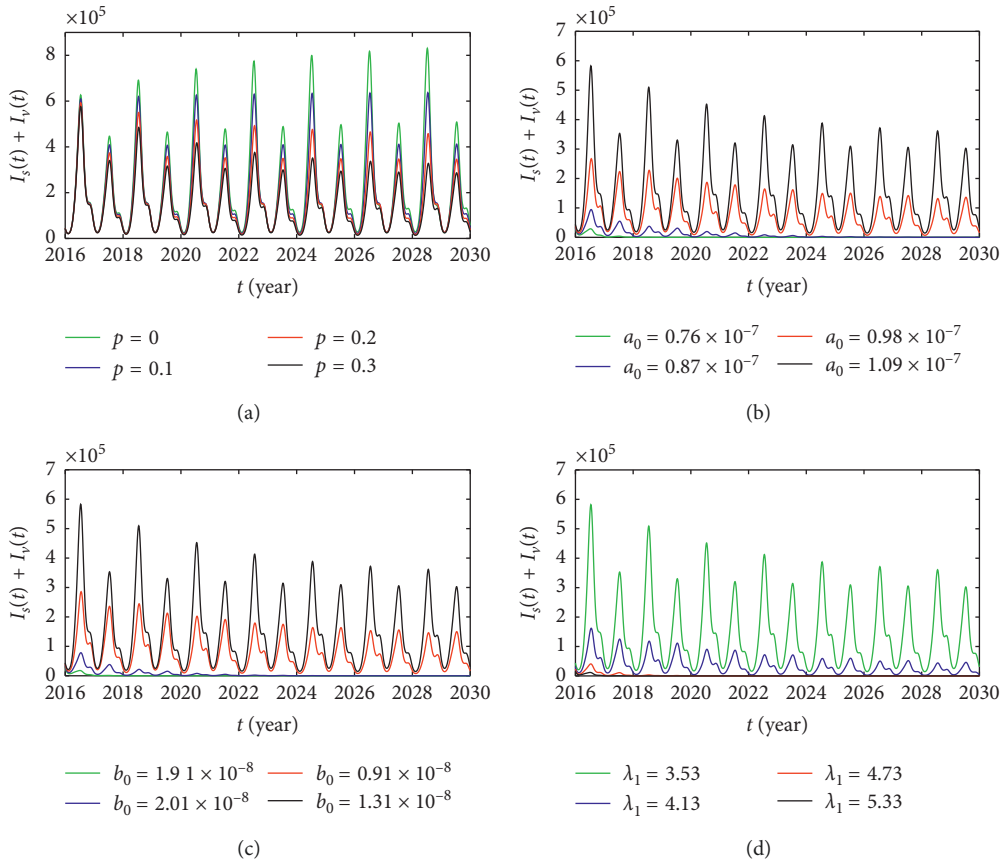


FIGURE 9: The influence of parameters on symptomatic infectious individuals $I_s + I_v$ (a) versus p; (b) versus a_0, being the baseline contact rate of β_1; (c) versus b_0, being the baseline contact rate of β_2; (d) versus λ_1. Other parameter values are given in Table 1, and the initial conditions are the same as in Figure 3.

periodicity was the strongest in the south, and so on. It is difficult for us to use the data in Figure 1 to discuss the differences in the spread of HFMD in different regions of the country. Thus, we used the national average transmission rate functions $\beta_1(t), \beta_2(t)$ in our modeling, where $\beta_1(t), \beta_2(t)$ are periodic functions. In future studies, in order to avoid the influence of seasonal differences in different regions on the research results, we will consider a specific region such as a province in China.

Appendix

The BP Neural Network Algorithm for Estimating Parameters

Based on the BP neural network, we design an algorithm to estimate unknown parameters of system (1). Denote the unknown parameter vector $\theta = (p, \rho, \lambda_2, a_0, a_1, a_2, a_3, a_4, b_0, b_1, b_2, b_3)^T$ for system (1) (see Table 1). Let $I = (I_1, I_2, \ldots, I_{138})^T$ and $\widehat{I} = (\widehat{I}_1, \widehat{I}_2, \ldots, \widehat{I}_{138})^T$ (the number 138 is the total months from January 2008 to June 2019). Then, the estimation algorithm is described as follows (the detailed process of the BP neural network can be seen in [45], and the topology of the BP neural network is shown in Figure 10):

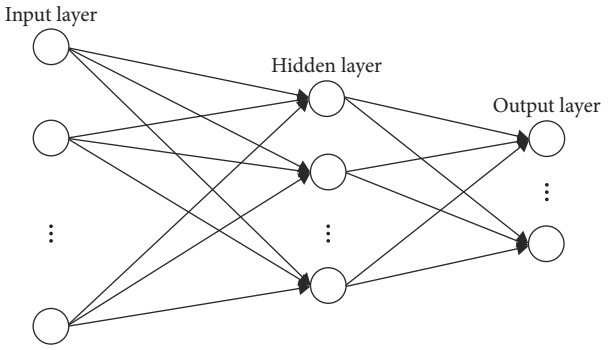


FIGURE 10: Structure chart of the BP neural network. The BP neural network consists of three layers: input layer, hidden layer, and output layer. The number of nodes in the input layer, the hidden layer, and the output layer is 120, 100, and 12, respectively.

Step 1: generate sample data: generate n group sample data $\theta^i = (p^i, \rho^i, \lambda_2^i, a_0^i, a_1^i, a_2^i, a_3^i, a_4^i, b_0^i, b_1^i, b_2^i, b_3^i)^T$, $i = 1, 2, \ldots, n$ by Latin hypercube sampling. Then, taking θ^i into system (1), we obtain n group numerically computed data $I^i = (I_1^i, I_2^i, \ldots, I_{138}^i)^T$, $i = 1, 2, \ldots, n$ (see Figure 11).

Step 2: use the BP neural network to train data: set $\overline{P} = (I^1, I^2, \ldots, I^n)$ and $\overline{T} = (\theta^1, \theta^2, \ldots, \theta^n)$, and each

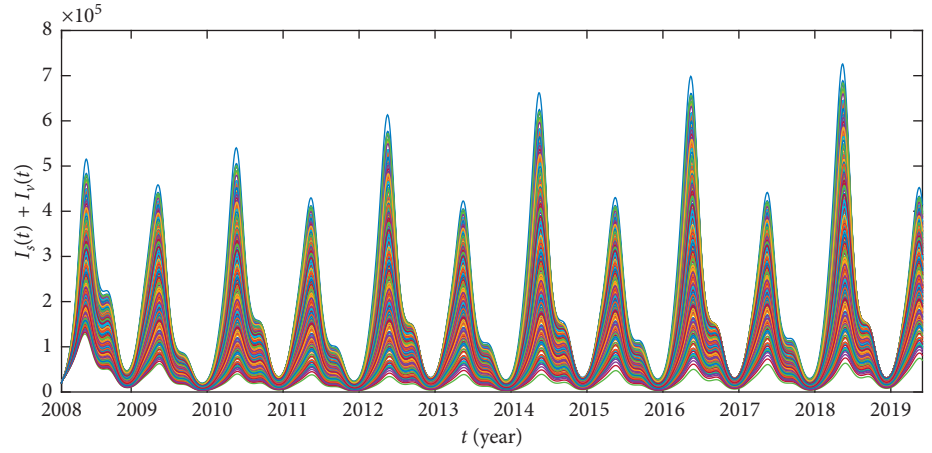


Figure 11: Multigroup numerically symptomatic infectious sample data I^i, $i = 1, 2, \ldots, n$ ($n = 400$) from system (1). θ^i, $i = 1, 2, \ldots, n$ ($n = 400$) is generated by Latin hypercube sampling. The values of other parameters and the initial conditions are the same as in Figure 3.

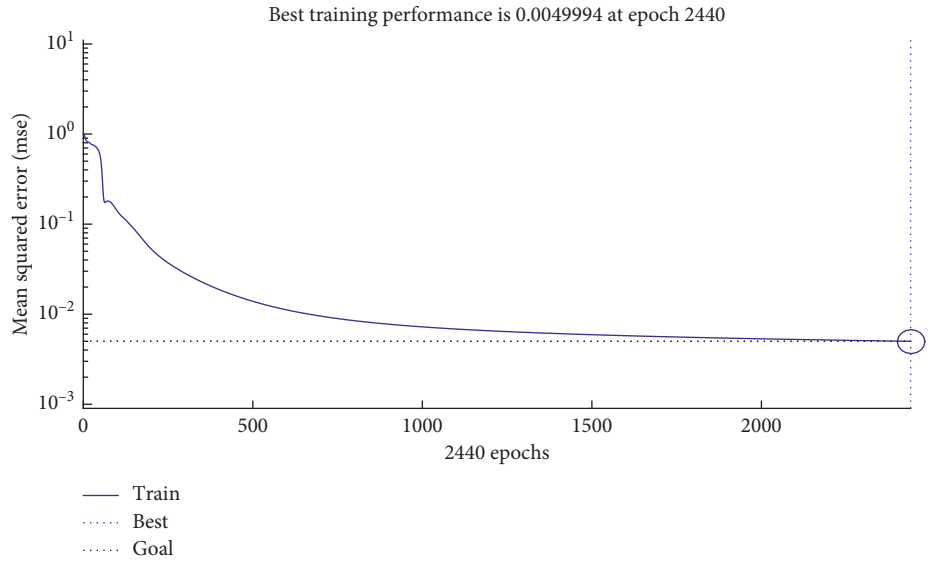


Figure 12: The neural training regression. The training goal mse (training error) is set as 0.005, and the maximum goal epoch is set as 5000. We obtain that the best training performance (training error) is 0.0049994 at epoch 2440.

element of $\overline{P}$ and $\overline{T}$ is normalized into interval $[-1, 1]$. Then, taking $\overline{P}$ as the input vector and $\overline{T}$ as the output vector, the BP neural network is used to train dataset, and let N_{et} be the training completed network, which the training error is less than the setting value (see Figure 12).

Step 3: estimate parameters: first, taking reported data $\widehat{I} = (\widehat{I}_1, \widehat{I}_2, \ldots, \widehat{I}_{138})^T$ of HFMD as the input vector into the training completed network N_{et}, we gain an output forecast vector $\theta^* = (p^*, \rho^*, \lambda_2^*, a_0^*, a_1^*, a_2^*, a_3^*, a_4^*, b_0^*, b_1^*, b_2^*, b_3^*)^T$. Then, taking θ^* into system (1), it gets the simulation of I_i. Finally, we set a small positive content $\varepsilon = 0.01$ to estimate the error function ER (45). If $\text{ER} < \varepsilon$, the estimated parameter vector θ^* is reasonable; otherwise, it goes to step 1 and reset the settings of the BP neural network.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (nos 11571170 and 11971013) and the Joint Special Funds for Basic Research in Local Undergraduate Universities (Part) of Yunnan Province of China (nos 2018FH001-113).

References

[1] Q. Cheng, L. Bai, Y. Zhang et al., "Ambient temperature, humidity and hand, foot, and mouth disease: a systematic review and meta-analysis," *Science of the Total Environment*, vol. 625, pp. 828–836, 2018.

[2] F. C. S. Tiing and J. Labadin, "A simple deterministic model for the spread of hand, foot and mouth disease (HFMD) in Sarawak," in *Proceedings of the 2008 Second Asia International Conference on Modelling & Simulation*, pp. 947–952, Washington, DC, USA, 2008.

[3] N. J. Schmidt, E. H. Lennette, and H. H. Ho, "An apparently new enterovirus isolated from patients with disease of the central nervous system," *Journal of Infectious Diseases*, vol. 129, no. 3, pp. 304–309, 1974.

[4] Y. C. Wang and F. C. Sung, *Modeling the Infections for Enteroviruses in Taiwan*, Institute of Environmental Health, Taipei, Taiwan, 2006, https://www.researchgate.net/publication/228559790.

[5] C. T. K. Lim, L. Jiang, S. Ma, L. James, and L. W. Ang, "Basic reproduction number of coxsackievirus type A6 and A16 and enterovirus 71: estimates from outbreaks of hand, foot and mouth disease in Singapore, a tropical city-state," *Epidemiology and Infection*, vol. 144, no. 5, pp. 1028–1034, 2016.

[6] G. Yu, Y. Li, J. Cai et al., "Short-term effects of meteorological factors and air pollution on childhood hand-foot-mouth disease in Guilin, China," *Science of the Total Environment*, vol. 646, no. 1, pp. 460–470, 2019.

[7] S. J. Chan, J. Labadin, and Y. Podin, "A dynamic SEIPR model for the spread of hand, foot and mouth disease in Sarawak," *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 9, no. 3, pp. 125–129, 2017.

[8] National Health and Family Planning Commission of the People's Republic of China (NHFPC), China Health Statistical Yearbook, August 2019, http://www.nhfpc.gov.cn/zwgk/rdts/ejlist.shtml.

[9] Chinese Center for Disease Control and Prevention (CCDC), *National Public Health Statistical Data*, Chinese Center for Disease Control and Prevention (CCDC), Beijing, China, 2019, http://www.nhfpc.gov.cn/jkj/s3578/new_list.shtml.

[10] Z. Chang, F. Liu, L. Bin et al., "Analysis on surveillance data of hand, foot and mouth disease in China, January–May 2017," *Disease Surveillance*, vol. 6, no. 32, pp. 447–452, 2017.

[11] Y. Cao, Z. Hong, L. Jin, M. Jian, and R. Hong, "Surveillance of hand, foot and mouth disease in China, 2011-2012," *Disease Surveillance*, vol. 28, no. 12, pp. 975–980, 2013.

[12] J. Sun, Z. Chang, L. Wang, J. Hong, L. Hui, and J. Zhang, "Analysis on the epidemic situation of hand, foot and mouth disease in China in January–March, 2013," *Practical Preventive Medicine*, vol. 21, no. 2, pp. 183–186, 2014.

[13] J. T. Wu, M. Jit, Y. Zheng et al., "Routine pediatric enterovirus 71 vaccination in China: a cost-effectiveness analysis," *PLoS Medicine*, vol. 13, no. 4, Article ID e1001975, 2016.

[14] S. Takahashi, Q. Liao, T. P. V. Boeckel et al., "Hand, foot, and mouth disease in China: modeling epidemic dynamics of enterovirus serotypes and implications for vaccination," *PLoS Medicine*, vol. 13, no. 2, Article ID e1001958, 2016.

[15] D. Liu, K. Leung, M. Jit et al., "Cost-effectiveness of Bivalent versus Monovalent vaccines against hand, foot and mouth disease," *Clinical Microbiology and Infection*, 2019, In press.

[16] Zhiyan Consulting, *China's New Vaccine Market In-Depth Survey and Market Demand Forecast Report for 2018–2024*, Zhiyan Consulting, Beijing, China, 2018, in Chinese.

[17] National Health Commission of China, *Text Record of Press Conference by National Health Commission of China on April 9, 2018*, National Health Commission of China, Beijing, China, 2018, in Chinese.

[18] N. Roy and N. Halder, "Compartmental modeling of hand, foot and mouth infectious disease (HFMD)," *Research Journal of Applied Sciences*, vol. 5, no. 3, pp. 177–182, 2010.

[19] N. Roy, "Mathematical modeling of hand-foot-mouth disease: quarantine as a control measure," *International Journal of Advanced Scientific Engineering and Technological Research*, vol. 1, no. 2, pp. 1–11, 2012.

[20] R. Viriyapong and S. Wichaino, "Mathematical modeling of hand, foot and mouth disease in the Northern Thailand," *Far East Journal of Mathematical Sciences (FJMS)*, vol. 100, no. 5, pp. 805–820, 2016.

[21] J. Yang, Y. Chen, and F. Zhang, "Stability analysis and optimal control of a hand-foot-mouth disease (HFMD) model," *Journal of Applied Mathematics & Computing*, vol. 41, no. 1-2, pp. 99–117, 2013.

[22] Y. Li, J. Zhang, and X. Zhang, "Modeling and preventive measures of hand, foot and mouth disease (HFMD) in China," *International Journal of Environmental Research and Public Health*, vol. 11, no. 3, pp. 3108–3117, 2014.

[23] Y. Li, L. Wang, L. Pang, and S. Liu, "The data fitting and optimal control of a hand, foot and mouth disease (HFMD) model with stage structure," *Applied Mathematics and Computation*, vol. 276, pp. 61–74, 2016.

[24] J. Wang, Y. Xiao, and R. A. Cheke, "Modelling the effects of contaminated environments on HFMD infections in mainland China," *BioSystems*, vol. 140, no. 1-2, pp. 1–7, 2016.

[25] S. Sharma and G. P. Samanta, "Analysis of a hand-foot-mouth disease model," *International Journal of Biomathematics*, vol. 10, no. 2, Article ID 1750016, 2017.

[26] J. Liu, "Threshold dynamics for a HFMD epidemic model with periodic transmission rate," *Nonlinear Dynamics*, vol. 64, no. 1-2, pp. 89–95, 2011.

[27] Y. Zhu, B. Xu, X. Lian, W. Lin, Z. Zhou, and W. Wang, "A hand-foot-and-mouth disease model with periodic transmission rate in Wenzhou, China," *Abstract and Applied Analysis*, vol. 2014, no. 3, pp. 1–11, 2014.

[28] Y. Ma, M. Liu, Q. Hou, and J. Zhao, "Modelling seasonal HFMD with the recessive infection in Shandong, China," *Mathematical Biosciences and Engineering*, vol. 10, no. 4, pp. 1159–1171, 2013.

[29] G. P. Samanta, "Analysis of a delayed hand-foot-mouth disease epidemic model with pulse vaccination," *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 61–73, 2014.

[30] J. Wang, Y. Xiao, and Z. Peng, "Modelling seasonal HFMD infections with the effects of contaminated environments in mainland China," *Applied Mathematics and Computation*, vol. 274, pp. 615–627, 2016.

[31] R. Li, L. Liu, Z. Mo et al., "An inactivated enterovirus 71 vaccine in healthy children," *New England Journal of Medicine*, vol. 370, no. 9, pp. 829–837, 2014.

[32] S. Esposito and N. Principi, "Hand, foot and mouth disease: current knowledge on clinical manifestations, epidemiology, aetiology and prevention," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 37, no. 3, pp. 1–8, 2018.

[33] W. Xing, Q. Liao, C. Viboud et al., "Hand, foot, and mouth disease in China, 2008–12: an epidemiological study," *The Lancet Infectious Diseases*, vol. 14, no. 4, pp. 308–318, 2014.

[34] M. Chen, Y. Ju, M. Chen et al., "Epidemiological and genetic characteristics of EV71 in hand, foot, and mouth disease in

Guangxi, southern China, from 2010 to 2015," *PLoS One*, vol. 12, no. 12, Article ID e0188640, 2017.

[35] Y. Chen and W. Chen, "Epidemic features and etiology of hand, foot and mouth disease in Huangshi from 2010 to 2013," *Chinese Journal of Experimental & Clinical Infectious Diseases*, vol. 10, no. 1, pp. 73–177, 2016.

[36] Z. Chang, J. Zhang, J. Sun, W. Zhang, and Z. Wang, "Epidemiological features of hand, foot and mouth disease in China, 2008-2009," *Chinese Journal of Epidemiology*, vol. 32, no. 7, pp. 676–680, 2011.

[37] S. M. Ashrafur Rahman and X. Zou, "Flu epidemics: a two-strain flu model with a single vaccination," *Journal of Biological Dynamics*, vol. 5, no. 5, pp. 376–390, 2011.

[38] National Bureau of Statistics of China (NBSC), *China Demographic Yearbook*, National Bureau of Statistics of China (NBSC), Beijing, China, 2019, http://www.stats.gov.cn/tjsj/ndsj/.

[39] Chinese Center for Disease Control and Prevention (CCDC), *Hand, Foot and Mouse Disease Description*, Chinese Center for Disease Control and Prevention (CCDC), Beijing, China, 2019, http://www.chinacdc.cn/zxdt/201606/t20160608_131032.html.

[40] K. Dietz, "The estimation of the basic reproduction number for infectious diseases," *Statistical Methods in Medical Research*, vol. 2, no. 1, pp. 23–41, 1993.

[41] W. Wang and X.-Q. Zhao, "Threshold dynamics for compartmental epidemic models in periodic environments," *Journal of Dynamics and Differential Equations*, vol. 20, no. 3, pp. 699–717, 2008.

[42] F. Zhang and X.-Q. Zhao, "A periodic epidemic model in a patchy environment," *Journal of Mathematical Analysis and Applications*, vol. 325, no. 1, pp. 496–516, 2007.

[43] L. Perko, *Differential Equations and Dynamical Systems*, Springer, New York, USA, 2001.

[44] X. Zhao, *Dynamical Systems in Population Biology*, Springer, New York, USA, 2003.

[45] S. S. Haykin, *Neural Networks and Learning Machines*, Pearson USA: Pearson Schweiz Ag, Zug, Switzerland, 2009.

[46] X. Zhang, Y. Zhao, and A. U. Neumann, "Partial immunity and vaccination for influenza," *Journal of Computational Biology*, vol. 17, no. 12, pp. 1689–1696, 2010.

[47] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*, Chapman and Hall/CRC, London, UK, 5th edition, 1992.

[48] L. Liu, X.-Q. Zhao, and Y. Zhou, "A tuberculosis model with seasonality," *Bulletin of Mathematical Biology*, vol. 72, no. 4, pp. 931–952, 2010.

[49] L. Shi, H. Zhao, and D. Wu, "Modelling and analysis of HFMD with the effects of vaccination, contaminated environments and quarantine in mainland China," *Mathematical Biosciences and Engineering*, vol. 16, no. 1, pp. 474–500, 2019.

Hindawi
Complexity
Volume 2020, Article ID 4760297, 14 pages
https://doi.org/10.1155/2020/4760297

Research Article

Walking Gait Phase Detection Based on Acceleration Signals Using Voting-Weighted Integrated Neural Network

Lei Yan,[1] **Tao Zhen**,[1] **Jian-Lei Kong**,[2] **Lian-Ming Wang**,[1] **and Xiao-Lei Zhou**[1]

[1]*Beijing Forestry University, Beijing 100083, China*
[2]*Beijing Technology and Business University, Beijing 100048, China*

Correspondence should be addressed to Lei Yan; mark_yanlei@bjfu.edu.cn and Jian-Lei Kong; kongjianlei@btbu.edu.cn

Received 12 September 2019; Revised 20 November 2019; Accepted 12 December 2019; Published 8 January 2020

Guest Editor: Eberhard O. Voit

Human gait phase recognition is a significant technology for rehabilitation training robot, human disease diagnosis, artificial prosthesis, and so on. The efficient design of the recognition method for gait information is the key issue in the current gait phase division and eigenvalues extraction research. In this paper, a novel voting-weighted integrated neural network (VWI-DNN) is proposed to detect different gait phases from multidimensional acceleration signals. More specifically, it first employs a gait information acquisition system to collect different IMU sensors data fixed on the human lower limb. Then, with dimensionality reduction and four-phase division preprocessing, key features are selected and merged as unified vectors to learn common and domain knowledge in time domain. Next, multiple refined DNNs are transferred to design a multistream integrated neural network, which utilizes the mixture-granularity information to exploit high-dimensional feature representative. Finally, a voting-weighted function is developed to fuse different submodels as a unified representation for distinguishing small discrepancy among different gait phases. The end-to-end implementation of the VWI-DNN model is fine-tuned by the loss optimization of gradient back-propagation. Experimental results demonstrate the outperforming performance of the proposed method with higher classification accuracy compared with the other methods, of which classification accuracy and macro-F1 is up to 99.5%. More discussions are provided to indicate the potential applications in combination with other works.

1. Introduction

As the most common form of human behavior, walking style is related to health status and individual differences, which can be shown by the differences of the gait phase [1]. Detecting results of the gait phase can provide references for disease diagnosis and rehabilitation [2, 3], which is of great significance to the patients' clinical rehabilitation. For example, an estimated gait disorder of 1.1 million children may have originated from different somatosensory disease in the United States [3]. In addition, researchers have managed to program humanoid robots to use human-based gait trajectories generated via gait classification [4], as well as consistently control wearable assistive devices such as robotic prostheses [5] and orthoses [6] for the recovery of lower-limb mobility. For instance, Yan et al. [4] proposed that gait phase detection can also be used to facilitate the development human auxiliary equipment, such as the medical ankle joint (AF), hip joint (HK), and knee ankle joint (KAF) orthopedic devices, as well as exoskeletons and other equipment. Similarly, gait phase detection plays an important role in sports medicine [7] and rehabilitation medicine [8].

Computational methods for gait phase recognition fall into two main categories. The first category is comprised of algorithms, which divide the gait phases based on the threshold selection of either raw or processed data [9]. Secondly, some deep-learning approaches have emerged in recent years to substitute the aforementioned techniques that rely on traditional classification algorithms. Some have applied deep-learning algorithms to different types of sensors to detect gait phases. For instance, Mukherjee et al. [10] present a fully automated frontal (i.e., employing front and back views only) gait phase recognition approach using the depth information captured by multiple Kinect RGB-D cameras. However, the captured image information is easily

disturbed by the external environment. Rosati et al. [11] proposed a method of hierarchical clustering to achieve recognition of the human gait phase by processing electromyography (EMG) data collected during gait, which has improved the abovementioned problems that are susceptible to environmental interference. However, muscle electrical signals are susceptible to factors such as sweat when collecting EMG data. Ding et al. [12] further improved the problems of the above EMG method and proposed a proportional fuzzy algorithm to achieve smooth recognition of the gait phase for foot pressure information processing, but the foot pressure will be affected by the wearer's weight, load, and other factors [13], and the pressure sensor also has a high failure rate. In recent years, researchers have started to study gait phase recognition methods based on inertial sensors (IMU). This is mainly due to the fact that more information can be obtained by adopting a small number of inertial sensor modules and most of the inertial sensor modules are installed on the legs and feet, so as to avoid damage or discomfort to the wearer [14]. At the same time, the information of IMU is basically unaffected by human body weight, belongings, clothing, sweat, and other factors, which is a prominent advantage compared to the method of plantar pressure or muscle electrical signal detection. In addition, inertial sensors are extremely cost effective [15] and acceleration signals acquired by inertial sensors exhibit typical waveform characteristics during the gait cycle. Previous studies have positioned inertial sensors on the instep, thigh, and calf [16–18]. This paper considers the position of the instep, lower leg, and thigh because the classifier has better classification performance at the lower extremity position [19].

For the recognition system, this paper designs an effective and adaptable gait detection method. Some research studies [13] indicate that a large amount of information can be obtained by using a small number of acceleration sensors that are located on the legs and feet to minimize sensor damage and discomfort to the person wearing the sensor. In this paper, we describe a system that uses three inertial sensor modules to obtain the acceleration information of the lower limbs of the human body. The collected acceleration data was reduced by the Principal Component Analysis (PCA) algorithm, which focuses on extracting the feature information of the original data and searches for a set of orthogonal low-wiki functions to represent a set of high-dimensional data, improving the recognition rate and recognition speed [20, 21]. Then, the paper divides the human gait into three phases and proposes a method of dividing the three gait phases. Finally, this paper proposes a VWF-DNN algorithm for detecting the gait phase, which is inspired by integrated learning. The core idea of the VWF-DNN algorithm is to use the three subneural networks with distinct differences to output the final classification result through the voting algorithm designed in this paper. The designed VWF-DNN with higher accuracy will be further evaluated with learned and unlearned data to test its suitability with acceleration classification.

This paper proposes an algorithmic model for detecting the gait phase, which uses the acceleration data from the instep, calf, and thigh to accurately detect two gait phase events. Finally, the effectiveness of the proposed VWI-DNN algorithm in gait phase detection is verified by the final recognition results.

2. Materials and Methods

2.1. Data Collection. Twenty volunteers with weight range in 46 kg to 88 kg and height range in 155 cm to 190 cm were recruited for the experimental data collection. The details of personal information are shown in Figure 1. The subjects have no physical or nerve injury to their legs or feet, which may affect walking gait phase detection.

With the improvement of the sensor manufacturing process, this study selected three IMU modules as portable devices for obtaining acceleration information. The inertial sensor modules were placed on the foot dorsum, the outer side of the lower leg, and the outer side of the thigh. The arrangement of the acceleration sensor on the instep, calf, and thigh monitoring the lower limb movement is shown in Figure 2. The acceleration resolution of the three-axis inertial sensor module used in the experiment is $6.1e-5$ g, the stability of attitude measurement is 0.01°, and the transmission baud rate in the experiment was set to 115200 bps.

In this experiment, all participants were asked to walk for at least 120 s on the configured treadmill with speeds at 0.78 m/s, 1.0 m/s, and 1.25 m/s, respectively. Participants walked normally three times on a treadmill at each speed, with all settings being the same in each state. In order to prevent the participants from affecting the gait due to fatigue, the experiment requires the participants to rest for 2 minutes for each walking test. In addition, data is only saved until the treadmill's running speed reaching the set speed. When the experiment was stopped and the treadmill began to slow down, we stopped collecting data. Moreover, each participant was asked to perform the same experiment under the same conditions to ensure the reliability and validity of the collecting process.

2.2. Data Preprocessing. Since each data sample contains multiple features from different sensors and each data in the same IMU module includes three acceleration data in X, Y, and Z directions, abundant data with different dimensions will lead to excessive complexity and easy overfitting of the detection model. In order to reduce the dimension of data set, the PCA method was adopted to synthesize the three-directional acceleration information a_x, a_y, and a_z of every IMU sensors into a new dimension variable Comp. PCA [22] is a general tool for dimensionality reduction and data analysis, and its essence is to project the data samples in the high-dimensional space into the low-dimensional space through linear transformation, while preserving the original data features as much as possible [23]. With the dimension reduction process, the compressed Comp can avoid the excessive information lose and adjust input dimension before passing acceleration data into the subsequent classifier. The Comp is calculated as follows:

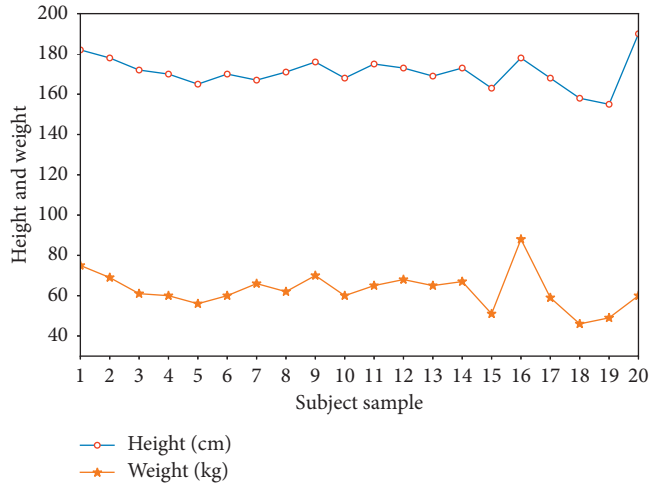


Figure 1: Information about volunteers participating in this experiment.

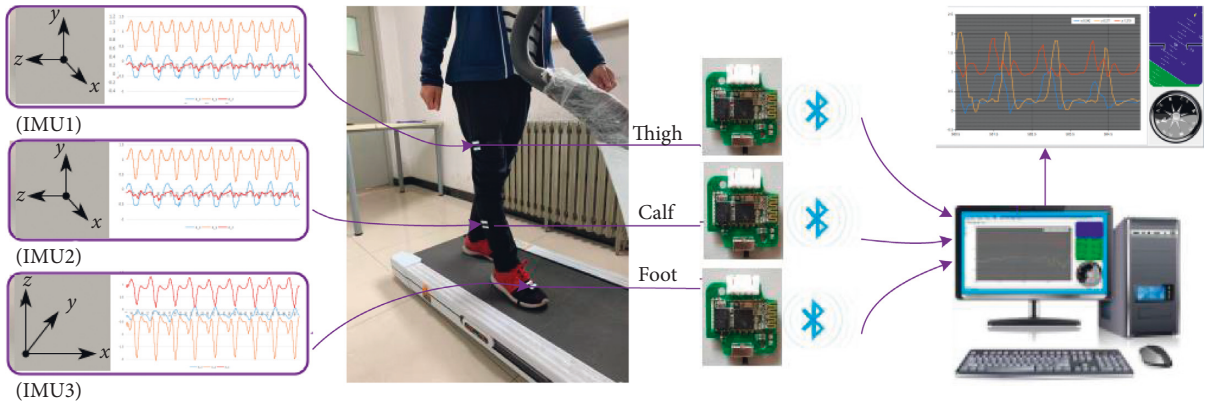


Figure 2: Human gait information acquisition system.

$$\text{Comp} = z\_1 * a_x + z\_2 * a_y + z\_3 * a_z, \quad (1)$$

where a_x, a_y, and a_z represent the acceleration in the X-, Y-, and Z-directions, respectively. Comp is a one-dimensional data by combining three-directions movement, which can improve the response performance of the algorithm and avoid overfitting during late training. The combined acceleration "Comp" of the instep, calf, and thigh together constitutes the input vector of the model, where z_1, z_2, and z_3 represent the coefficients of the acceleration in three directions. The distribution of z_1, z_2, and z_3 corresponding to different body parts at asynchronous speed is shown in Table 1.

Based on the above experiment, we can get the curve of the acceleration in the X, Y, and Z directions and the combined acceleration as shown in Figure 3.

The human walking process is a rhythmic movement, and a complete gait cycle definition is from the unilateral heel to the ipsilateral heel again [21]. A two-phase model has proven to be sufficient to control the knee module of an active orthosis [9]. Nonetheless, the most widespread approach relies on a four-phase model [24], which are independently written as (1) Heel Strike (HS), (2) the loading response phase or Flat Foot (FF), (3) the heel lifting or Heel-Off (HO) and (4) the initial Swing Phase (SW). This four-phase model of gait granularity has been used for the actuation of multiple robotic ankle-foot orthoses [25, 26].

Table 1: Acceleration data for different parts at different speeds using PCA synthesized parameter table.

Pace	Collection location	z_1	z_2	z_3
0.78 m/s	Calf	0.632	0.671	0.421
	Thigh	−0.652	0.569	0.355
	Foot	0.636	0.524	0.582
1.0 m/s	Calf	0.613	0.637	0.423
	Thigh	−0.479	0.601	0.673
	Foot	0.667	0.625	0.533
1.25 m/s	Calf	0.629	0.638	0.427
	Thigh	−0.565	0.623	0.641
	Foot	0.753	0.568	0.531

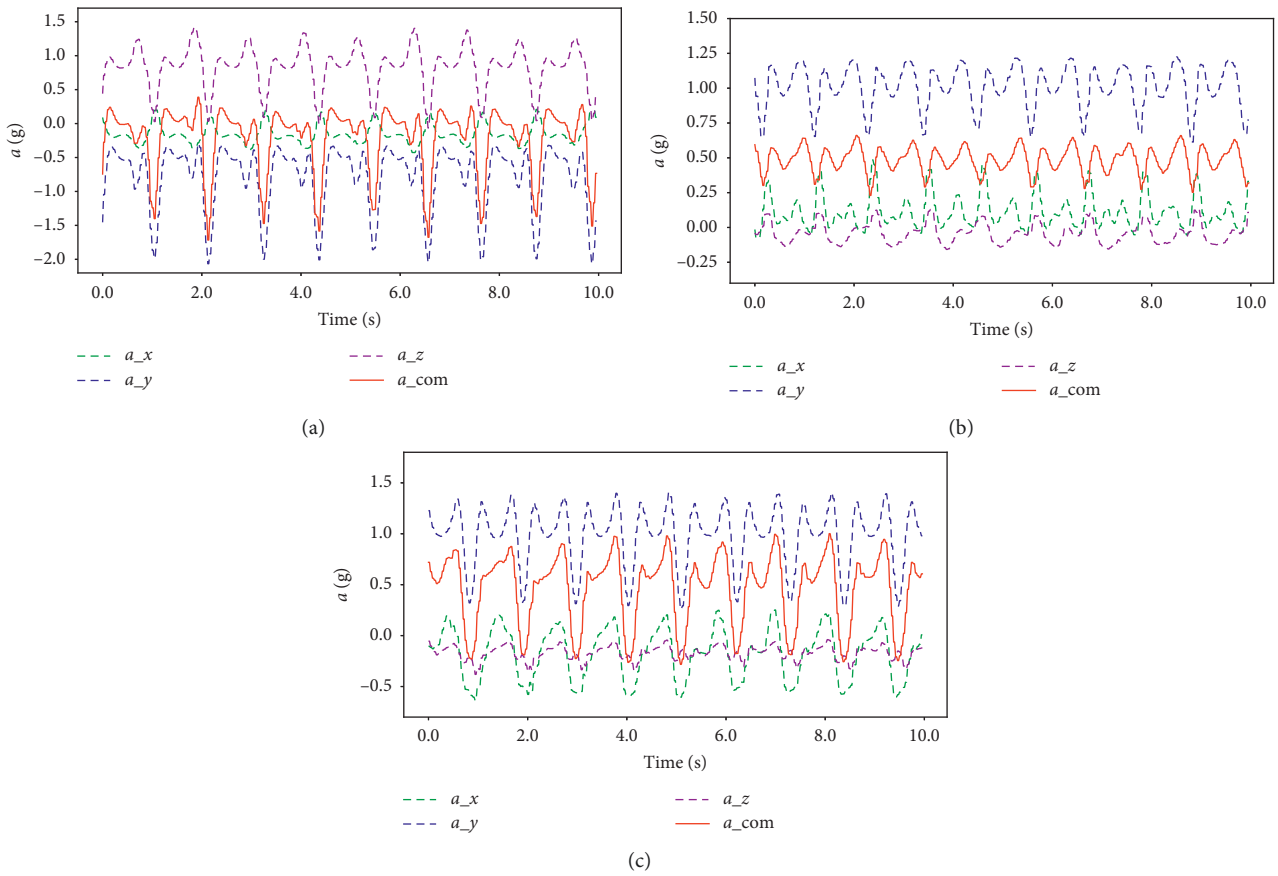


Figure 3: Acceleration data collected under the three body parts: foot (a), thigh (b), and calf (c). *a_x* represents the acceleration in the *x*-axis direction, *a_y* represents the acceleration in the *y*-axis direction, *a_z* represents the acceleration in the *z*-axis direction, and *a*_com represents the combined acceleration.

$$\text{soft max}(q)_i = \frac{e^{q_i'}}{\sum_{i=1}^{n} e^{q_i'}}. \tag{2}$$

To ensuring the scientificity of gait classification, the walking cycle in this paper was also divided into HS, FF, HO, and SW. During normal walking, the acceleration signals in the three directions of the foot, thigh, and calf exhibit periodicity. The sway phase accounts for approximately 40% of the total gait phase, and the stance phase accounts for approximately 60% of the total gait phase [27]. We can approximate that the stance phase is the biggest phase in the walking cycle. According to the division of gait phase, the phase division in this paper is shown in Figure 4.

Except for the gait phase division, feature selection is also used to extract meaningful information or noise from acceleration signals. After this processing, the key features effectively representing different gait phases are obtained from time domain for the subsequent recognition model. In this paper, the standard deviation (SD), mean absolute value (Mav), maximum value (Max), minimum value (Min), and median (Med) are selected to handle with acceleration signals as feature vectors. Since the vectors composed of single and multiple feature sets will produce different accuracy rates, we merged SD, Max, Min, Med, and Mav feature vectors to form the input feature vector in order to improve the recognition accuracy.

2.3. Voting-Weighted Integrated Neural Network. The next step is to design an algorithm to recognize the timings vectors related to the acceleration signal. As the commonly classifier with excellent performance [28], the DNN is a feedforward artificial neural network consisting of an input layer, an output layer, and at least two hidden layers [29]. Although the DNN is a strong classifier, sometimes the subneural network still misjudges certain situations, leading to the misclassification of results [30]. Thus, the voting fusion of the neural network is proposed to solve the instability of the subneural network, which may be insensitive to the data of some input layers due to a single network structure [31], and results in errors in the output layer. The output of the integrated neural network is determined by the output of each integrated neural network under the sample [32], which can improve the classification performance and generalization performance of the classifier to some extent [33, 34]. Therefore, inspired by the AdaBoosting algorithm [35] and Bagging [36] algorithm in current integrated learning, this paper further proposes a novel VWI-DNN algorithm (the entire structure is shown in Figure 5) by

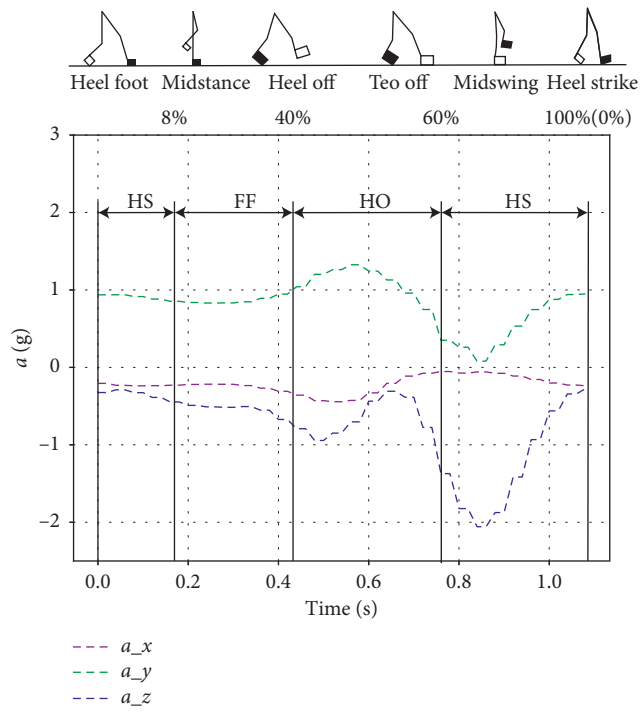


Figure 4: Phase division diagram of gait. Green curve represents the acquired foot acceleration data, brown curve represents the acquired calf acceleration data, and blue curve represents the collected thigh acceleration data.

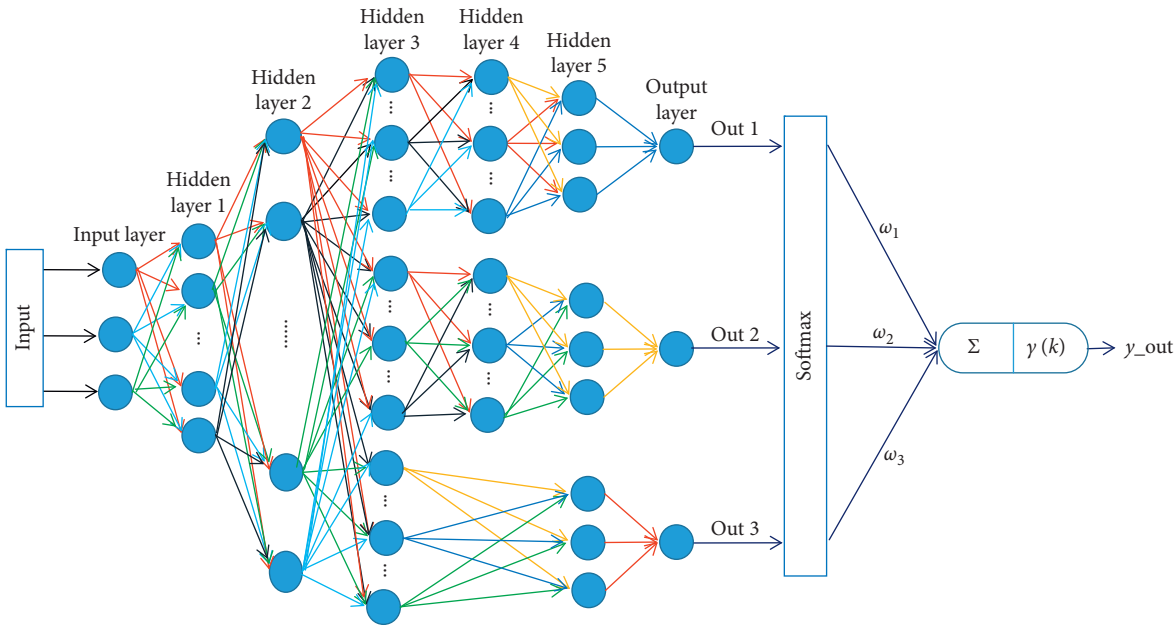


Figure 5: Structure diagram of integrated neural network algorithm based on vote weighting.

modifying the traditional DNN network. Our work aims to construct a general neural network structure allowing for different voting decisions at each submodel and demonstrate its use as a practical way to massively increase model capacity.

Firstly, we chose three advanced deep neural networks instead of the perceptual layer network as the classifiers and named them SNN_1, SNN_2, SNN_3, so as to improve the classification performance of the network. The research mainly focuses on the design of three subneural networks and the optimization of network parameters, as well as the fusion method of the output results of these three subneural networks. SNN_1, SNN_2, and SNN_3 have one common input layer and two hidden layers, so that they can have a

common network structure. In addition, three independent and different networks are set up in the third hidden layer. This design can simplify the network and share the weight information. The latter network structure maintains their independence, the number of neurons, and activation functions, and the number of hidden layers exist certain differences. The neural network has associated nonlinearities and are trained using greedy hierarchical supervision, and the final learning rate was manually specified and set to 0.05. Finally, the three subneural networks output their respective classification results through the Softmax regression layer. Some parameters of the structure of the entire network are shown in Table 2, where HL_x represents the xth hidden layer and the "Dropout column" represents the sparse rate that needs to be set.

The goal of the WWI-DNN algorithm is to identify the three gait phases of the human body, which solves the multiclassification task. However, the output of the neural network does not necessarily represent a probability distribution, so the output of the neural network must be transformed into a probability distribution through the Softmax regression layer, whose expression is shown in the following equation:

Then, the focus of the VWI-DNN algorithm is to establish a voting fusion mechanism. In the process of output fusion, the establishment of fusion algorithm is the core content of information fusion [37, 38]. The concept of voting fusion was therefore proposed to obtain more accurate results from multiple unreliable data. After years of development, there have been numerous voting algorithms, such as majority voting fusion, logical voting fusion, median voting fusion, and weighted average voting fusion. The voting weight of an excellent voting fusion algorithm should not be artificially set. It is more reasonable that its voting weight should be larger when a subneural network performs well and vice versa. In the AdaBoosting algorithm theory, the weight information is updated with the classification error rate, but the weight update is problematic when the error rate is higher than 50%. In order to avoid this problem, considering that good performance classifiers should be rewarded and poor performance classifiers should be punished, this paper proposes a weighted sum voting algorithm. The algorithm of weighted summation is to add the weights of the three networks to output the same result, respectively, and regard the sum of the weights as the fusion result. This is also a simple and effective information fusion algorithm.

In order to solve the multiple classifier weight problem, this paper introduces the weighting function. This paper used the classification accuracy to determine the weight coefficient of each classifier. Equation (3) shows how to solve the accuracy of each classifier, and we need to set the corresponding weighting function to determine the weight coefficient of the classification. The image of the weighting function is given in Figure 6, and its expression is as shown in equation (5), wherein the expression of the reward function is indicated in equation (4). It can be seen that the derivative value of the reward function (Deriv_reward) has a negative correlation with the value of the abscissa, which is why this paper chooses it as a reward function. Such a function can make the model with good performance get larger reward. When the accuracy ε_t is less than 40%, the weight should be reduced, so the weight coefficient is taken as $\omega_i \cdot \exp(\varepsilon_i - 1)$; similarly, when the accuracy ε_t is more than 40%, the classification performance of the weak classifier can be considered great, and its weight should be increased, so the weight should be taken reward(ε_t). Meanwhile, in order to make the new weight available in [0,1], this paper normalizes c by using equation (6). In addition, given that one classifier cannot be made large on its own, this paper sets a minimum threshold of 0.26 for each classifier's weight. The classification weight of each classifier is obtained according to equation (7). The given ω initial value ω_0 is 33.33% and the initial value c_0 of c is 1.0.

$$\varepsilon_i = \frac{n_{\text{correct}}}{N_{\text{total}}}, \quad i = 1, 2, 3, \tag{3}$$

$$\text{reward}(\varepsilon_i) = 0.5 + \frac{1}{1 + \exp(-5 * \varepsilon_i)}, \tag{4}$$

$$\begin{aligned} c_i = {} & \text{reward}(\varepsilon_i) \cdot I(\varepsilon_i > 0.4) \\ & + \exp(\varepsilon_i - 1) \cdot I(\varepsilon_i < 0.4), \end{aligned} \tag{5}$$

$$c_i = \frac{c_i}{\sum_{i=1}^{3}(c_i)}, \tag{6}$$

$$\omega_i = \begin{cases} \dfrac{\omega_{i-1} \cdot c_i}{\sum_{i=1}^{3}(\omega_{i-1} \cdot c_i)}, & \omega_i > 0.26, \\ 0.26, & \text{otherwise.} \end{cases} \tag{7}$$

The next step is to confirm the weight of the classifier. The classification result of each classifier is calculated by using equation (8). When the classification results are the same, the weights of these classifiers should be added and obtained the weight corresponding to each gait phase. This process can be expressed by equation (9), where γ_k indicates the probability output corresponding to the human gait phase k. The final classification result should be determined by the maximum weight value corresponding to each gait phase, thereby obtaining the final integrated output q, whose expression is as shown in equation (10), where n_{correct} denotes the number of samples correctly classified, N_{total} denotes the total number of samples, q_i' denotes the value output by the ith subneural network through the output layer, and ω_i denotes the weight of the ith neural network corresponding to the gait phase k, and when $k = 1$, it represents the starting phase; when $k = 2$, it indicates the swing phase; when $k = 3$, it indicates the foot phase.

Finally, the task of the VWI-DNN algorithm is to solve the problem of internal parameter update and optimization of each subneural network. The neural network generally updates the internal parameters of the network architecture by optimizing the loss function value. When using neural networks for classification, the usual function is the cross-entropy loss function [39] which characterizes the distance

Table 2: Parameter setting of VWI-DNN structure.

Layer	Number of neurons			Activation function			Dropout		
	SNN_1	SNN_2	SNN_3	SNN_1	SNN_2	SNN_3	SNN_1	SNN_2	SNN_3
HL_1	540	540	540	leaky_relu	leaky_relu	leaky_relu	1.0	1.0	1.0
HL_2	200	200	200	leaky_relu	leaky_relu	leaky_relu	1.0	1.0	1.0
HL_3	260	200	140	leaky_relu	relu	tanh	0.6	0.8	1.0
HL_4	120	90	60	leaky_relu	relu	tanh	1.0	1.0	1.0
HL_5	60	90	—	leaky_relu	relu	—	1.0	1.0	—

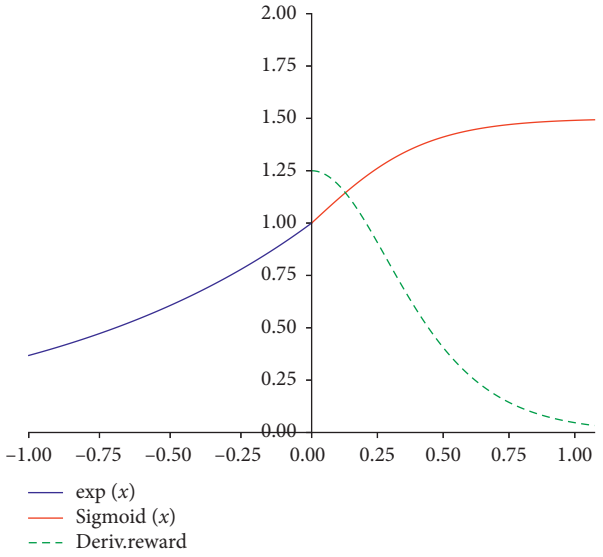


Figure 6: Weighting function diagram.

between two probability parts so that the cross-entropy loss function can be used to calculate the distance between the predicted probability distribution and the probability distribution of the real answer. According to the cross-entropy loss function equation (11), we can obtain the loss function equations (12)–(14) of the three subneural networks. When we train this network, we hope that the parameters of the first subneural network would not be changed when we train the other two neural networks. But the reality is that when you train any of the three subneural networks, the parameters of the shared layer will change. Therefore, if the optimizer optimizes the three loss functions separately, the shared layer parameters will be changed reciprocally and the ideal result will not be achieved. In order to improve this situation, this paper proposes a method to optimize the global loss by using a classifier. The optimizer no longer optimizes the three cross-entropy loss functions separately but only optimizes the sum of the three loss functions (loss_sum), which can ensure that the shared layer parameters can achieve a better result and the training speed can be improved. The loss_sum expression is shown in equation (15).

$$\mathrm{cf}_i = \arg\max(q_i), \quad i = 1, 2, 3, \tag{8}$$

$$\gamma_k = \sum_{i=1}^{N} \omega_i \cdot I(\mathrm{cf}_i = k), \quad k = 1, 2, 3, \tag{9}$$

$$q(\gamma_k) = \arg\max(\gamma_k), \quad k = 1, 2, 3, \tag{10}$$

$$\mathrm{loss} = -\sum p(x)\log q(x), \tag{11}$$

$$\mathrm{loss\_1} = -\sum p_1(x)\log q_1(x), \tag{12}$$

$$\mathrm{loss\_2} = -\sum p_2(x)\log q_2(x), \tag{13}$$

$$\mathrm{loss\_3} = -\sum p_3(x)\log q_3(x), \tag{14}$$

$$\mathrm{loss\_sum} = \mathrm{loss\_1} + \mathrm{loss\_2} + \mathrm{loss\_3}, \tag{15}$$

where $q_i(x)$ denotes the probability distribution of the three subneural networks predicting the phase of the three types of gaits after passing through the Softmax layer and $p_i(x)$ denotes the distribution of the real samples.

To avoid overfitting, 70% of the sample set was selected for training and 30% for testing. After training the three learning models 10,000 times with the same training set, the same test set was used to test the trained models, and the classification accuracy, macro-F value, and area ender curve (AUC) after the test were recorded. The entire process of this study is shown in Figure 7.

3. Results and Discussion

3.1. Evaluation Methods. Comparing the classification performance of different classifiers cannot determine the effectiveness of the algorithm by a single metric and there are many other methods commonly used to obtain a full census. The evaluation metrics including Precision, Recall, Accuracy, and F1-score are used to compare different methods. Precision and Recall are widely used in the fields of information retrieval and statistical classification to evaluate the quality of results, where the higher the Precision and Recall value, the better the method performs. F1 combines the results of P and R, and when F1 is high, it indicates that Precision and Recall are both high, and this evaluation index is relatively effective. However, the classifier of this paper performs multiclassification task. We hope to comprehensively investigate the Precision and Recall on several binary confusion matrices, and the most direct way is to calculate macro-F1 [40]. While, Accuracy reflects the ratio of the samples correctly classified by the classifier to the total samples for a given test data set. According to equations (16)–(21), we can calculate these evaluation factors, where TP, TN, FP, and FN, respectively,

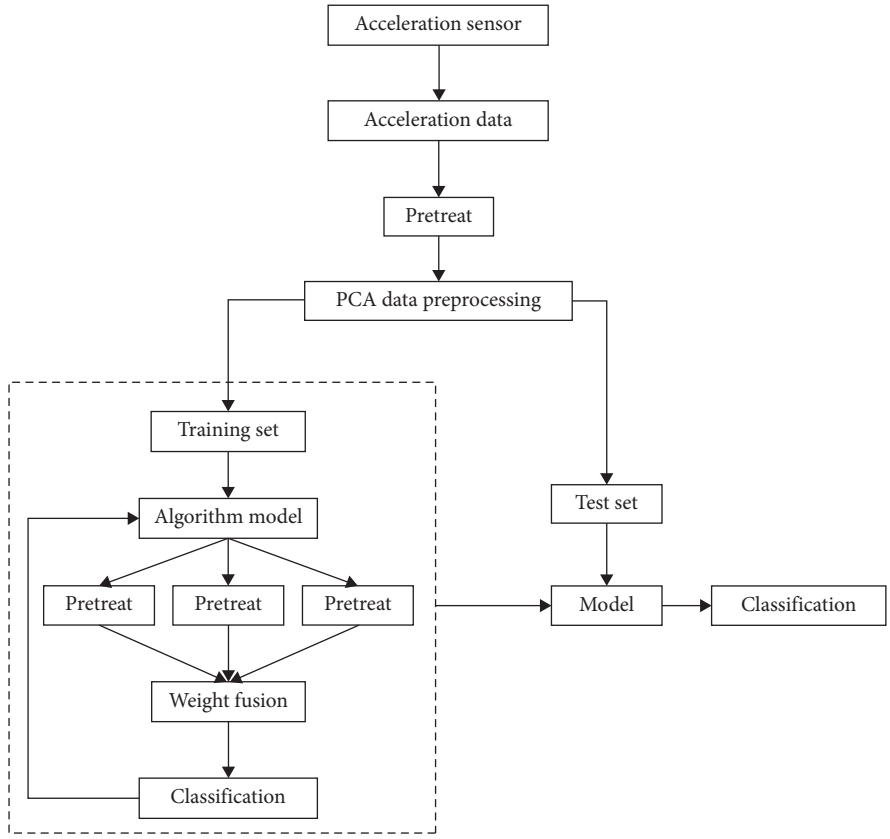


Figure 7: System block diagram.

represent true positive, true negative, false positive, and false negative.

In order to better analyze the performance of the classifier, this paper introduces the AUC under receiver operating characteristic (ROC) as the evaluation index of the algorithm. Spackman was the first to adopt ROC diagram for machine learning, and he proved the value of the ROC curve in evaluation [41]. In recent years, it has been applied more and more in machine learning and data mining research, partly because people realized that simple classification accuracy is usually not a good indicator to measure performance [42]. The AUC of each algorithm can be calculated to compare, and the algorithm that has the largest AUC will have the best diagnostic value:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{16}$$

$$P_i = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{17}$$

$$R_i = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{18}$$

$$\text{macro} - P = \frac{1}{n}\sum_{i=1}^{n} P_i, \tag{19}$$

$$\text{macro} - R = \frac{1}{n}\sum_{i=1}^{n} R_i, \tag{20}$$

$$\text{macro} - \text{F1} = \frac{2 \times \text{macro} - P \times \text{macro} - R}{\text{macro} - P + \text{macro} - R}. \tag{21}$$

3.2. Results. The joint confusion matrix of three gait phase recognition results at different synchronization speeds are, respectively, shown in Figures 8–10. According to Figures 8–10, we can easily get Tables 3–5 which, respectively, classifies the performance for each training function in terms of HS, FF, HO, and SW phase under three kinds of sync speed. According to Tables 3–5, it can be observed that all of Bagging, Boosting, and VWI-DNN have macro-F1 of HS and SW phase recognition nearly up to 100%. While from the results obtained by F1, Bagging, and AdaBoosting have poor recognition effects on FF and HO phases. In particular, Bagging has a case where macro-F1 of the FF and HO phases is 0. It can also be clearly seen from the observation of Figure 8 that the Bagging algorithm can easily recognize the FF phase as the HO phase and the HO phase as the FF phase. The Bagging algorithm is extremely poor in HO phase and FF phase recognition. It can be seen from Table 5 that although Bagging and AdaBoosting have poor recognition of FF phase and HO phase, VWI-DNN algorithm proposed in this paper still has more than 98% macro-F1 value for the FF and HO phase. The three training functions have higher accuracy for SW and HS phases recognition at three paces, generally reaching more than 98%. As aforementioned, macro-F1 can comprehensively

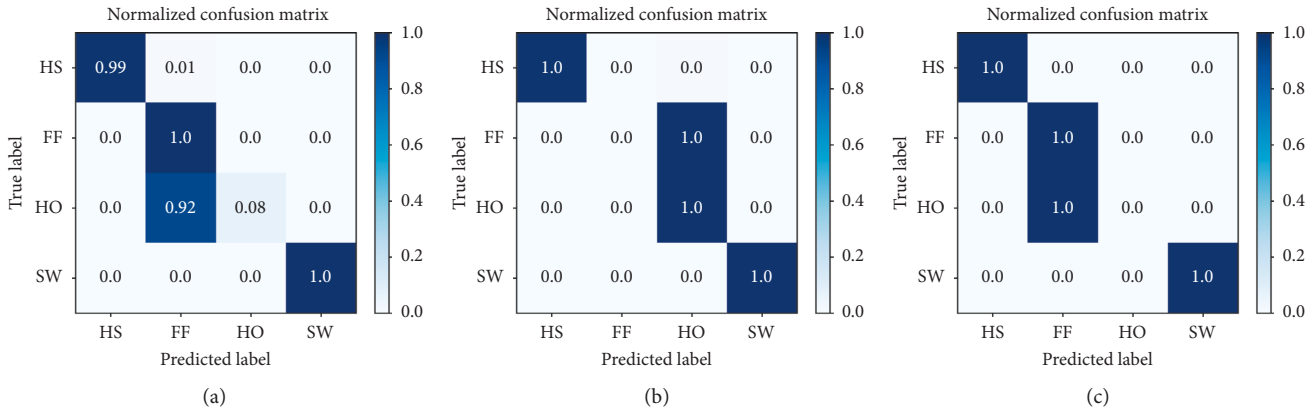


Figure 8: Confusion matrix of three gait patterns derived from bagging classification under three pace settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c) classes.

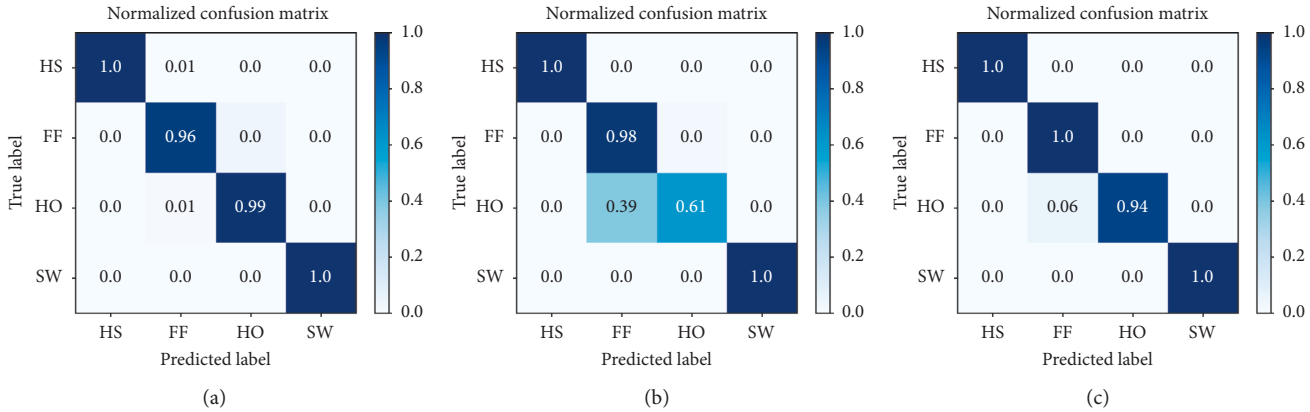


Figure 9: Confusion matrix of three gait patterns derived from AdaBoosting classification under three pace settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c) classes.

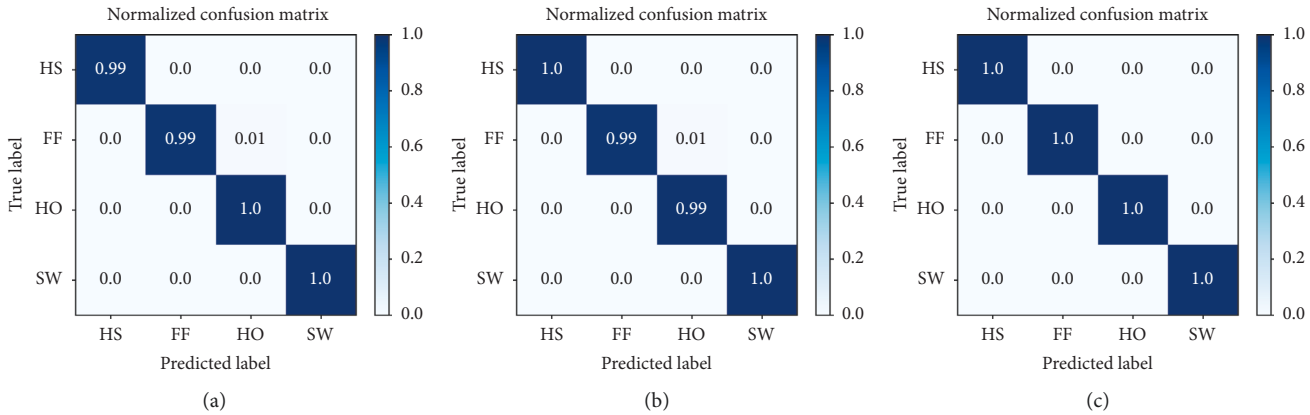


Figure 10: Confusion matrix of three gait patterns derived from VWI-DNN Classification under three pace settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c) classes.

measure the two indicators of Precision and Recall. It was observed that macro-F1 value of VWI-DNN is more than 98% for any phase at any pace.

In order to comprehensively measure macro-F1 in four gait phases, macro-F1 was introduced in this paper. According to equations (16)–(20), the corresponding macro-F1 can be calculated, and the corresponding Table 6 was obtained. According to Table 5, the recognition accuracy of the VWI-DNN algorithm is as high as 98% or more, while the other two training functions have lower accuracy. Especially, the Bagging algorithm has a phase recognition accuracy of less than 77% at three paces. It can be clearly seen

Table 3: Summary of classification performance of bagging at unsynchronized speed.

Pace	0.78 m/s				1.0 m/s				1.25 m/s			
Phase	HS	FF	HO	SW	HS	FF	HO	SW	HS	FF	HO	SW
Precision (%)	100.0	51.6	100.0	100.0	100.0	0	47.4	100.0	100.0	48.2	0	100.0
Recall (%)	98.9	100.0	8.2	100.0	100.0	0	100.0	100.0	100.0	100.0	0	100.0
F1 (%)	99.4	68.1	15.1	100.0	100.0	0	64.3	100.0	100.0	65.1	0	100.0

Table 4: Summary of classification performance of AdaBoosting at unsynchronized speed.

Pace	0.78 m/s				1.0 m/s				1.25 m/s			
Phase	HS	FF	HO	SW	HS	FF	HO	SW	HS	FF	HO	SW
Precision (%)	99.7	99.4	99.20	100.0	100.0	78.2	94.8	100.0	100.0	78.3	99.7	100.0
Recall (%)	99.5	99.4	99.40	100.0	99.7	96.9	71.4	99.5	88.0	99.7	86.9	100.0
F1 (%)	99.6	99.4	99.30	100.0	99.9	86.5	81.5	99.7	93.6	87.7	92.9	100.0

Table 5: Summary of classification performance of VWI-DNN at unsynchronized speed.

Pace	0.78 m/s				1.0 m/s				1.25 m/s			
Phase	HS	FF	HO	SW	HS	FF	HO	SW	HS	FF	HO	SW
Precision (%)	100.0	99.4	98.8	100.0	100.0	99.4	99.5	100.0	99.7	99.7	100.0	100.0
Recall (%)	99.5	99.1	99.7	100.0	100.0	99.4	99.5	100.0	100.0	99.7	99.7	100.0
F1 (%)	99.7	99.3	99.3	100.0	100.0	99.4	99.5	100.0	99.9	99.7	99.8	100.0

Table 6: Summary of classification performance for different training functions.

Pace	Training function	Classification rate		
		Accuracy (%)	Macro-F1 (%)	AUC
0.78 m/s	Bagging	76.9	70.7	0.98
	AdaBoosting	95.7	95.7	1.0
	VWI-DNN	99.6	99.5	1.0
1.0 m/s	Bagging	74.4	66.1	0.95
	AdaBoosting	92.4	91.9	1.0
	VWI-DNN	98.9	99.7	1.0
1.25 m/s	Bagging	74.4	66.3	0.98
	AdaBoosting	93.5	93.6	1.0
	VWI-DNN	99.1	99.7	1.0

from Figures 11 and 12 that the VWI-DNN algorithm is higher in accuracy and macro-F1 than the other two algorithms at any of the paces. By observing the AUC, we can also see that the VWI-DNN algorithm can reach 1.0, which is high. Figure 13 also illustrates the relationship between the three algorithms corresponding to the AUC at the unsynchronized paces. From Figure 14, it can be found that the Accuracy, macro-F1, and AUC of the VWI-DNN algorithm do not change much with the increase of the pace, which is relatively stable.

3.3. Discussion. This study demonstrated the capability of the proposed system to detect gait phases based on acceleration signals. To support this hypothesis, this paper proposes to use the voting-weighted integrated neural network to identify the gait phase and compares it with other integrated learning algorithms to verify the effectiveness of the algorithm.

3.3.1. Acceleration Signals Analysis. Walking activity emanated from the human body is important, and this information can be extracted through the use of acceleration signals. Although the VWF-DNN algorithm has shown certain validity in the classification of acceleration signals for gait event detection, it still needs to be further optimized in the future. In this study, the inertial sensor module needed to be placed at a designated location on the instep, lower leg, and thigh of each subject. However, due to the height, weight, gender, etc. of each subject, the sensors cannot be accurately placed in the specified position and can only be installed at an approximate designated position, which needs further investigation [43].

To characterize acceleration signals, there are three main cascaded modules which are data processing, feature extraction, and classification methods. It should be noted that the classification accuracy depends greatly on the features extracted. Moreover, four combined features is better than a single or two features for upper limb movement. This shows

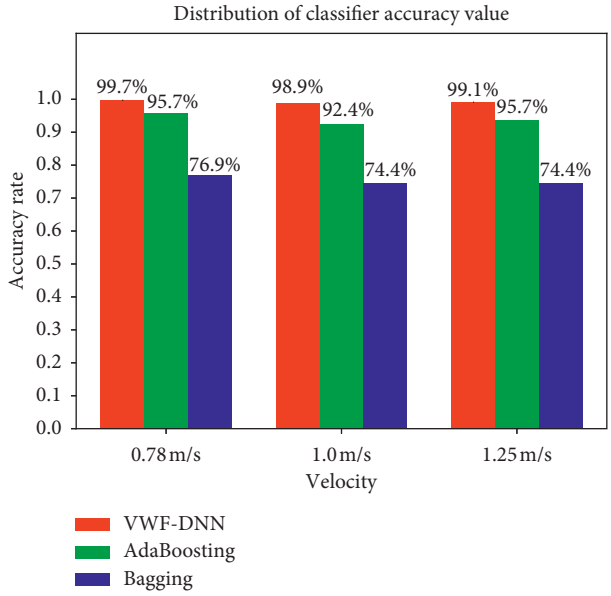


Figure 11: More tag set classifier accuracy distribution at three paces.

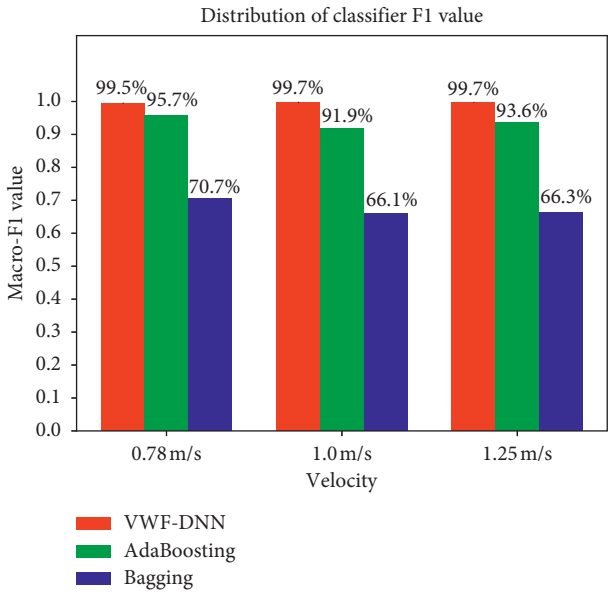


Figure 12: More tag data set classifier macro-F1 distribution under sync speed.

the network responses of five TD features. By observing Figures 8–10, it can be seen that all the training functions have clearly classified the HS and SW phases and there is very low classification error between the two. Interestingly, we can see through the confusion matrix that the Bagging and AdaBoosting misinterpreted the FF phase and HO phase, which results in a low-recognition precision for the FF and HO phases. Also, by analyzing the results of Tables 3–5, it can be shown that the model has a poor recognition effect on the FF and HO phases. This may be due to the fact that the two adjacent phase features are too similar. In addition, Figures 8–10 also illustrate that the phase difference between the FF and HO phase characteristics are not obvious, but the proposed VWI-DNN algorithm in this paper can significantly improve this situation.

Regarding the effective classification of acceleration signals, the five TD features provide better classification of gait events and the average accuracy of the VWF-DNN algorithm for HS, FF, HO, and SW phases is 99.2%. This study was then compared to some of the previous research studies. With regard to the gait event, different walking conditions by IMU located on foot show 82.2% accuracy using the ANN [39]. Similar result is achieved when five gait phases were classified using IMUs 82% accuracy [40]. In addition, this is important for the development of assistive devices for the lower leg, as they have strong relationships with the gait event [41]. To propose a system that may be applied to any individual, the generalization of trained VWF-DNN algorithm was tested on unlearned data of acceleration signals. This study found that the proposed system could predict the gait event successfully for the unlearned data. In general, the detection of HS, FF, HO, and SW phases based on acceleration signals seemed reliable.

In reviewing the literature, the reliability of IMUs was questioned. It is interesting to note that the percentage of stance and swing phases of footswitch data was in line with the theory of the gait cycle as 60% of the complete cycle was the stance phase while the remaining percent was swing phase 40% [42]. According to Observation Tables 3 and 4, it can be found that the recognition of FF and HO phases by Bagging and AdaBoosting is generally low, while other two phases is very high. In addition, according to observations of Figures 8–10, we can also see that Bagging and AdaBoosting divided FF phase into HO phase and also divide HO phases into FF phases. These results indicate that the features we extracted may require further discussion.

3.3.2. Gait Phase Detection. The purpose of this study was to apply a machine learning to predict HS, FF, HO, and SW from acceleration signals. In support of the hypothesis aforementioned, this study proposes the VWI-DNN algorithm and uses it to successfully predict HS, FF, HO, and SW phases. The results of learned data indicated that the acceleration signals had low variability and stability during walking on the ground. Compared with the research carried out by Nazmi et al. [1], this study can obtain higher recognition accuracy, but the neural network model used in this study was too complicated, which may lead to longer training time. Some portable gait event detection applications require functional electronic simulators, dynamic gait monitoring, and gait biofeedback [44], but currently no wearable sensors meet these requirements.

The VWI-DNN algorithm used a voting mechanism to fuse results and used three subneural networks to vote. Whether the three subneural networks are set up properly needs further exploration. The selection of the bonus penalty function takes the derivative changes and actual effects into account, but it still avoids the occasional occurrence of improper rewards and punishments, resulting in the final

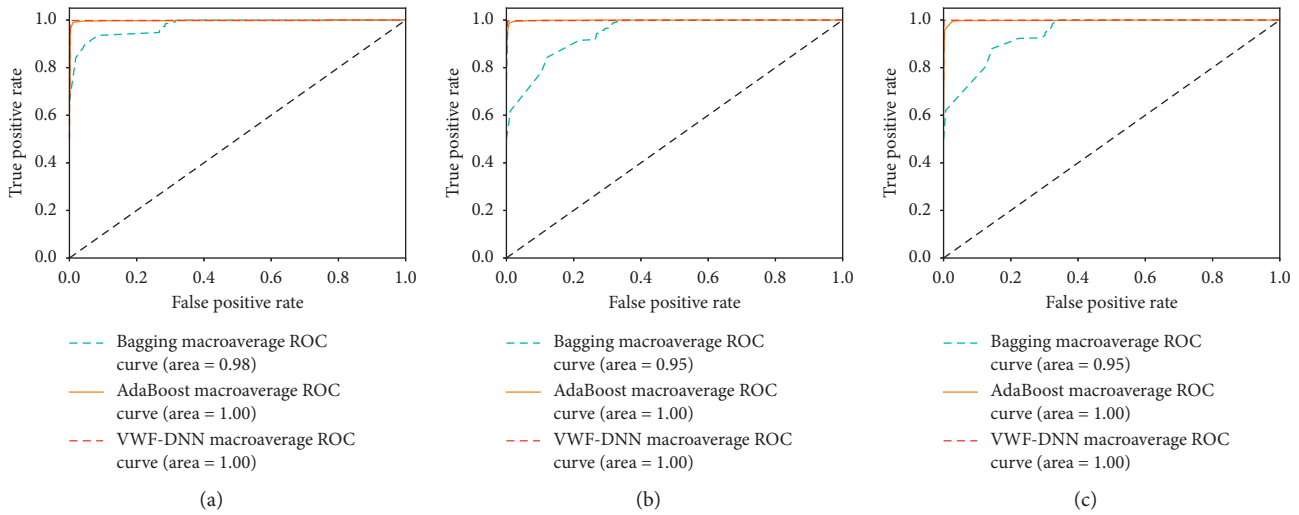


Figure 13: ROC curves of three classifiers at three paces: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c).

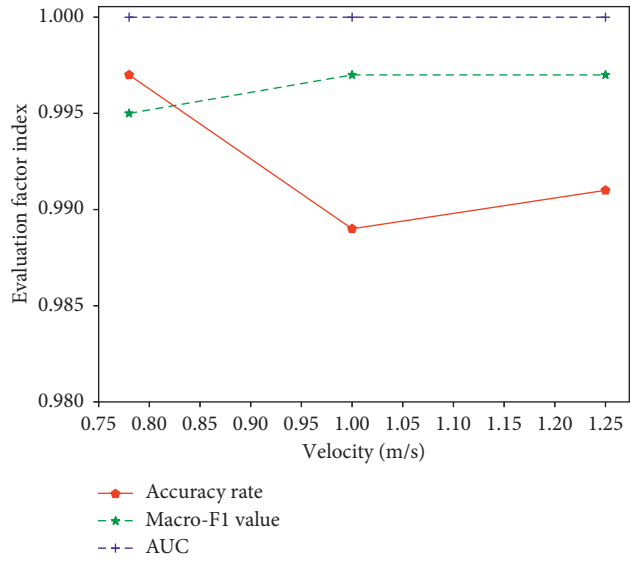


Figure 14: VWI-DNN algorithm evaluation factor changes with pace.

classification results are not ideal. Nevertheless, the model used in this paper can still achieve the recognition accuracy and the macro-F1 value, which were higher than 98%. And as can be seen from Figure 13, the AUC value of VWI-DNN algorithm reaches the maximum value of 1. It can be seen from the performance of each classifier in the AUC that the classification result obtained by the VWI-DNN algorithm is more reliable.

In terms of the generalization of the proposed system, this study revealed the VWI-DNN algorithm achieved better performance in gait phase recognition. The VWI-DNN algorithm based on the voting weighting mechanism detects HS, FF, HO, and SW phases with higher recognition accuracy, macro-F1, and AUC than existing Bagging and AdaBoosting. However, the macro-F1 of the FF and HO phases obtained by Bagging and AdaBoosting are very low. According to Figures 8–10, these results indicate that the HS and SW phase was reasonably accurate and FF and HO phases should yield more warnings. Even so, using the VWI-DNN algorithm proposed in this paper still obtains a good recognition effect on the FF and HO phases, and in some sense compensates for the defects of dividing the two phases. The study also shows that the proposed use of vector difference to distinguish between ST and IL phases does have a certain effect. In addition, it is possible to explain that the VWI-DNN algorithm proposed in this paper has strong performance in the field of gait phase division.

3.3.3. Limitations. There are some limitations to this study. Even though VWI-DNN has shown its usefulness in classifying acceleration signals for gait event detection, a further evaluation is needed using other machine learning

approaches. Future works should improve the classification accuracy by improving the method of extracting features, gait phase recognition algorithm, etc. In this study, the wearing three inertial sensor modules were assumed to be an acceptable wearable sensor compared to other wearable sensors. However, the wearing of the sensor in practice may have a potential impact on the gait of the person who has not yet been investigated.

The detection of gait phase in this study relied on data collected by IMUs. Although acceleration sensors have low cost and fast dynamic response, the accuracy might be affected by the circuit design and placement of sensors. Different walking styles could lead to erroneous placement of IMUs. However, as the sensor processing advances and algorithm innovations, these errors are further reduced. Therefore, IMU can possibly be applied in large scale in the direction of gait detection in the near future.

4. Conclusion

In order to recognize the walking gait phase accurately, this work proposed a VWI-DNN model to analyze multidimensional acceleration signals and detect different gait events including HS, FF, HO, and SW. It consists of three main parts, data preprocessing, multistream integrated neural network, and voting-weighted function, where data preprocessing employs PCA dimensionality reduction, four-phase division, and key feature selection in time domain. In addition, multiple refined DNNs are applied to design a multistream integrated neural network, which utilizes the mixture-granularity information to form a high-dimensional feature. Finally, a voting-weighted function is developed to fuse dissimilar submodels as a unified representation for distinguishing small discrepancy among different gait phases. It constructs a general neural network allowing for fusing different voting decisions at each submodel and demonstrates its use as a practical way to massively increase model capacity. Experiments and discussion demonstrate the effectiveness of the VWI-DNN with higher accuracy and macro-F1 up to 99.5%, which outperform other voting methods.

The network generates large number of parameters, which increases the time for model classification. Therefore, the model can only be trained on the GPU and the IMU data cannot be classified online on a convenient mobile device. Our future work is to try a lightweight network to compress the model parameters and speed. And we will attempt to combine our method with the newly updated work for other gait phase detection applications, such as rehabilitation training robot and medical Internet of Things.

Data Availability

The acceleration data used to support the findings of this study have not been made available since these data are related to the personal privacy of each volunteer involved in the experiment.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was financially supported by the Fundamental Research Funds for the Central Universities (No. 2015ZCQ-GX-03), National Key Research and Development Program of China (No. 2017YFC1600605), National Natural Science Foundation of China (No. 61673002), and Beijing Municipal Education Commission (No. KM201910011010).

References

[1] N. Nazmi, M. A. Abdul Rahman, S.-I. Yamamoto, and S. A. Ahmad, "Walking gait event detection based on electromyography signals using artificial neural network," *Biomedical Signal Processing and Control*, vol. 47, pp. 334–343, 2019.

[2] Y. Okubo, D. Schoene, and S. R. Lord, "Step training improves reaction time, gait and balance and reduces falls in older people: a systematic review and meta-analysis," *British Journal of Sports Medicine*, vol. 51, no. 102, pp. 586–593, 2017.

[3] S. Tian, M. Li, Y. Wang, and X. Chen, "Application of an improved correlation method in electrostatic gait recognition of hemiparetic patients," *Sensors*, vol. 19, no. 11, p. 2529, 2019.

[4] T. Yan, M. Cempini, C. M. Oddo, and N. Vitiello, "Review of assistive strategies in powered lower-limb orthoses and exoskeletons," *Robotics and Autonomous Systems*, vol. 64, pp. 120–136, 2015.

[5] H. Vu, F. Gomez, P. Cherelle, D. Lefeber, A. Nowé, and B. Vanderborght, "ED-FNN: A new deep learning algorithm to detect percentage of the gait cycle for powered prostheses," *Sensors*, vol. 18, no. 7, p. 2389, 2018.

[6] S. Murray and M. Goldfarb, "Towards the use of a lower limb exoskeleton for locomotion assistance in individuals with neuromuscular locomotor deficits," in *Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1912–1915, San Diego, USA, August 2012.

[7] S. Ounpuu, "The biomechanics of walking and running," *Clinics in Sports Medicine*, vol. 13, pp. 4843–4863, 1994.

[8] A. Ferrari, R. Brunner, S. Faccioli et al., "Gait analysis contribution to problems identification and surgical planning in CP patients: an agreement study," *European Journal of Physical & Rehabilitation Medicine*, vol. 51, no. 1, pp. 39–48, 2015.

[9] J. Taborri, E. Scalona, S. Rossi et al., "Real-time gait detection based on hidden Markov model: is it possible to avoid training procedure?," in *Proceedings of the 2015 IEEE International Symposium on Medical Measurements and Applications*, pp. 141–145, Torino, Italy, May 2015.

[10] J. Mukherjee, P. Chattopadhyay, and S. Sural, "Information fusion from multiple cameras for gait-based re-identification and recognition," *IET Image Processing*, vol. 9, no. 11, pp. 969–976, 2015.

[11] S. Rosati, V. Agostini, M. Knaflitz, and G. Balestra, "Muscle activation patterns during gait: a hierarchical clustering analysis," *Biomedical Signal Processing and Control*, vol. 31, pp. 1746–8094, 2017.

[12] S. Ding, X. Ouyang, Z. Li, and H. Yang, "Proportion-based fuzzy gait phase detection using the smart insole," *Sensors and Actuators A: Physical*, vol. 284, 2018.

[13] Y. Shimada, S. Ando, T. Matsunaga et al., "Clinical application of acceleration sensor to detect the swing phase of stroke gait in functional electrical stimulation," *The Tohoku Journal of Experimental Medicine*, vol. 207, no. 3, pp. 197–202, 2005.

[14] N. Chia Bejarano, E. Ambrosini, A. Pedrocchi, G. Ferrigno, M. Monticone, and S. Ferrante, "A novel adaptive, real-time algorithm to detect gait events from wearable sensors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 413–422, 2015.

[15] R. Caldas, M. Mundt, W. Potthast, F. Buarque de Lima Neto, and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait & Posture*, vol. 57, pp. 204–210, 2017.

[16] M. D. Sánchez Manchola, M. J. Pinto Bernal, M. Munera, and C. A. Cifuentes, "Gait phase detection for lower-limb exoskeletons using foot motion data from a single inertial measurement unit in hemiparetic individuals," *Sensors*, vol. 19, no. 13, 2019.

[17] M. Yuwono, S. W. Su, Y. Guo, B. D. Moulton, and H. T. Nguyen, "Unsupervised nonparametric method for gait analysis using a waist-worn inertial sensor," *Applied Soft Computing*, vol. 14, pp. 72–80, 2014.

[18] Y. Bai, X. Jin, X. Wang, T. Su, J. Kong, and Y. Lu, "Compound autoregressive network for prediction of multivariate time series," *Complexity*, vol. 2019, Article ID 9107167, 11 pages, 2019.

[19] J. Taborri, S. Rossi, E. Palermo, F. Patanè, and P. Cappa, "A novel HMM distributed classifier for the detection of gait phases by means of a wearable inertial sensor network," *Sensors*, vol. 14, no. 9, pp. 16212–16234, 2014.

[20] R. Liu, J. Zhang, L. Wang, and M. Zhang, "Research on face recognition method based on combination of SVM and LDA-PCA," in *The Proceedings of the Second International Conference on Communications, Signal Processing, and Systems*, pp. 62–66, Springer, Berlin, Germany, 2014.

[21] X. Lu and X. Xu, "Human behavior recognition based on acceleration and hga-bp neural network," *Computer Engineering*, vol. 41, no. 9, pp. 220–232, 2015.

[22] J. Ma and Y. Yuan, "Dimension reduction of image deep feature using PCA," *Journal of Visual Communication and Image Representation*, vol. 63, 2019.

[23] T. Juri, P. Eduardo, R. Stefano et al., "Gait partitioning methods: a systematic review," *Sensors*, vol. 16, no. 1, pp. 66–91, 2016.

[24] J. Kim, S. Hwang, R. Sohn, Y. Lee, and Y. Kim, "Development of an active ankle foot orthosis to prevent foot drop and toe drag in hemiplegic patients: a preliminary study," *Applied Bionics and Biomechanics*, vol. 8, no. 3-4, pp. 377–384, 2011.

[25] N. B. Bolus, C. N. Teague, O. T. Inan, and G. F. Kogler, "Instrumented ankle-foot orthosis: towards a clinical assessment tool for patient-specific optimization of orthotic ankle stiffness," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 6, pp. 2492–2501, 2017.

[26] C. Mummolo, L. Mangialardi, and J. H. Kim, "Quantifying dynamic characteristics of human walking for comprehensive gait cycle," *Journal of Biomechanical Engineering*, vol. 135, p. 91006, 2013.

[27] X. Jin, N. Yang, X. Wang, Y. Bai, T. Su, and J. Kong, "Integrated predictor based on decomposition mechanism for PM2.5 long-term prediction," *Applied Sciences*, vol. 9, no. 21, p. 4533, 2019.

[28] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, T.-L. Su, and J.-L. Wang, "Probability fusion decision framework of multiple deep neural networks for fine-grained visual classification," *IEEE Access*, vol. 7, pp. 122740–122757, 2019.

[29] Y. Y. Zheng, J. L. Kong, X. B. Jin, X.-Y. Wang, and M. Zuo, "Cropdeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture," *Sensors*, vol. 19, no. 5, p. 1058, 2019.

[30] A. N. M. Jaya Lakshmi and K. V. Krishna Kishore, "Performance evaluation of DNN with other machine learning techniques in a cluster using Apache Spark and MLlib," *Journal of King Saud University Computer and Information Sciences*, 2018.

[31] J. Liu, S. Xia, and W. Yu, "A neural network integrated incremental learning method for voting right value adjustment," *Signal Processing*, vol. 26, no. 1, pp. 46–50, 2010.

[32] N. Tommaso, C. Renzo, and A. Claudio, "An integrated artificial neural network–unscented Kalman filter vehicle sideslip angle estimation based on inertial measurement unit measurements," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 233, no. 7, pp. 1864–1878, 2019.

[33] D. Povey, X. Zhang, and S. Khudanpur, "Parallel training of deep neural networks with natural gradient and parameter averaging," 2014, https://arxiv.org/abs/1410.7455.

[34] G. Zhou, S. Ge, and L. Yang, "Fault diagnosis method of nuclear power plant based on neural network and vote fusion," *Journal of Atomic Energy Science and Technology*, vol. 44, no. B09, pp. 367–372, 2010.

[35] J. Zhu, A. Arbor, and T. Hastie, "Multi-class AdaBoost," *Statistics and Its Interface*, vol. 2, no. 3, pp. 349–360, 2006.

[36] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[37] W. W. Rodrigo and H. S. Lopes, "Neural networks for protein classification," *Applied Bioinformatics*, vol. 3, no. 1, pp. 41–48, 2004.

[38] R. H. W. Pinheiro, G. D. C. Cavalcanti, R. F. Correa, and T. I. Ren, "A global-ranking local feature selection method for text categorization," *Expert Systems with Applications*, vol. 39, no. 17, pp. 12851–12857, 2012.

[39] H.-Y. Lau, K.-Y. Tong, and H. Zhu, "Support vector machine for classification of walking conditions of persons after stroke with dropped foot," *Human Movement Science*, vol. 28, no. 4, pp. 504–514, 2009.

[40] D. Sanchez-Valdes, A. Alvarez-Alvarez, and G. Trivino, "Walking pattern classification using a granular linguistic analysis," *Applied Soft Computing*, vol. 33, pp. 100–113, 2015.

[41] J. L. Bartlett and R. Kram, "Changing the demand on specific muscle groups affects the walk-run transition speed," *Journal of Experimental Biology*, vol. 211, no. 8, pp. 1281–1288, 2008.

[42] J. Taborri, E. Palermo, S. Rossi, and P. Cappa, "Gait partitioning methods: a systematic review," *Sensors*, vol. 66, pp. 1–20, 2016.

[43] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *Proceedings of the Fifteenth International Conference (ICML 98)*, J. Shavlik, Ed., pp. 445–453, San Francisco, CA, USA, 1998.

[44] M. Hanlon and R. Anderson, "Real-time gait event detection using wearable sensors," *Gait & Posture*, vol. 30, no. 4, pp. 523–527, 2009.

Hindawi
Complexity
Volume 2019, Article ID 4876165, 19 pages
https://doi.org/10.1155/2019/4876165

Research Article

Dynamics of a Stochastic Three-Species Food Web Model with Omnivory and Ratio-Dependent Functional Response

Guirong Liu[1] **and Rong Liu**[2]

[1]*School of Mathematical Sciences, Shanxi University, Taiyuan, Shanxi 030006, China*
[2]*School of Applied Mathematics, Shanxi University of Finance and Economics, Taiyuan, Shanxi 030006, China*

Correspondence should be addressed to Guirong Liu; lgr5791@sxu.edu.cn

Received 25 February 2019; Revised 9 October 2019; Accepted 23 October 2019; Published 13 November 2019

Guest Editor: George V. Popescu

This paper is concerned with a stochastic three-species food web model with omnivory which is defined as feeding on more than one trophic level. The model involves a prey, an intermediate predator, and an omnivorous top predator. First, by the stochastic comparison theorem, we show that there is a unique global positive solution to the model. Next, we investigate the asymptotic pathwise behavior of the model. Then, we conclude that the model is persistent in mean and extinct and discuss the stochastic persistence of the model. Further, by constructing a suitable Lyapunov function, we establish sufficient conditions for the existence of an ergodic stationary distribution to the model. Then, we present the application of the main results in some special models. Finally, we introduce some numerical simulations to support the main results obtained. The results in this paper generalize and improve the previous related results.

1. Introduction

The dynamic relationship between predators and their preys has long been and will continue to be one of the dominant themes in both ecology and mathematical ecology due to its universal existence and importance [1]. During the past one hundred years, there have been many investigations on predator-prey models. To the best of our knowledge, in the predator-prey interaction, the functional response plays an important role in the population dynamics, and most of the predator-prey models with the functional responses only depend on the prey. However, laboratory experiments show that the ratio-dependent response function is more reasonable in characterizing the relationship between predators and their preys [2]. Arditi and Ginzburg [3] first proposed a ratio-dependent functional response of form $(\alpha x)/(x+\beta y)$. Kuang and Beretta [4] investigated the following ratio-dependent type predator-prey model:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1 x_1(t) - \dfrac{\alpha_{12}x_2(t)}{x_1(t)+\beta_{12}x_2(t)}\right]\mathrm{d}t, \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t)+\beta_{12}x_2(t)}\right]\mathrm{d}t, \end{cases} \tag{1}$$

where $x_1(t)$ and $x_2(t)$ represent population sizes of prey and predator at time t, respectively. r, a_1, and d_2 stand for the prey intrinsic growth rate, the intraspecific competition rate of the prey, and the predator death rate, respectively; α_{12}, β_{12}, and e_{12} represent the encounter rate, half capturing saturation constant, and conversion rate, respectively, that predator x_2 preys on prey x_1.

Long-term ecological research studies show that three-species predator-prey models are fundamental building blocks of large scale ecosystems. However, it was only in the 1970s that some scholars began to study the dynamics of

three-species predator-prey systems [5]. In particular, Hsu et al. [6] have classified all three-species predator-prey models into five types: two predators competing for one prey, one predator acting on two preys, food chain, food chain with omnivory, and food chain with cycle. Food chain architecture and strengths of species interactions are important determinants of trophic dynamics (see [7]). It is well known that tritrophic food chain model consists of one prey, one intermediate predator, and one top predator. Note that omnivory is a widespread mechanism in interacting populations. In [6], the authors investigated the following three-species predator-prey food chain model with an omnivory top predator:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_{11}x_1(t) - a_{12}x_2(t) - a_{13}x_3(t)\right]\mathrm{d}t, \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + a_{21}x_1(t) - a_{23}x_3(t)\right]\mathrm{d}t, \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 + a_{31}x_1(t) + a_{32}x_2(t)\right]\mathrm{d}t, \end{cases} \tag{2}$$

where x_1, x_2, and x_3 denote the number of prey, intermediate predator, and omnivorous top predator, respectively, r_1 is the growth rate of prey, r_i is the death rate of species x_i $(i = 2, 3)$, a_{11} is the intraspecific competition rate of prey, a_{12}, a_{13}, and a_{23} are the capture rates, and a_{21}, a_{31}, and a_{32} denote the efficiency of food conversion. Model (2) describes that the intermediate predator preys on only the prey and the omnivorous top predator preys on both the prey and the intermediate predator. This is a general part of marine or terrestrial food web ecological systems. Based on model (2), Namba et al. [8] considered the intraspecific competition of the intermediate predator and the intraspecific competition of the top predator. Moreover, the authors demonstrated the stabilizing role of intraspecific competition among intermediate and top predators when the growth rate of prey species is adequate to support both the predator species. Furthermore, Sen et al. [9] investigated the following three-species Lotka–Volterra model with intraguild predation and mixed functional responses:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_{11}x_1(t) - a_{12}x_2(t) - a_{13}x_3(t)\right]\mathrm{d}t, \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + a_{21}x_1(t) - a_{22}x_2(t) - \dfrac{a_{23}x_3(t)}{1 + \beta x_2(t)}\right]\mathrm{d}t, \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 + a_{31}x_1(t) + \dfrac{a_{32}x_2(t)}{1 + \beta x_2(t)} - a_{33}x_3(t)\right]\mathrm{d}t, \end{cases} \tag{3}$$

where x_1, x_2, and x_3 denote the number of prey, intermediate predator, and omnivorous top predator, respectively. Obviously, in [9], the authors considered Holling type-II functional response between the intermediate predator and top predator and other functional responses were assumed to be linear. All meanings of the parameters are exact to or similar as those for (2) except the following. Here, a_{ii} is the intraspecific competition rate of species x_i $(i = 2, 3)$ and β is the reciprocal of the half-saturation constant.

Note that the three-species food web models (2) and (3) with the functional responses only depend on prey density. However, in fact, the predator has to search and compete for food and the ratio-dependent function of the prey and the predator is more suitable to substitute for the model with complicated interaction between the prey and predator. Then, the ratio-dependent type three-species food web model with omnivory is expressed in the form:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1x_1(t) - \dfrac{\alpha_{12}x_2(t)}{x_1(t) + \beta_{12}x_2(t)} - \dfrac{\alpha_{13}x_3(t)}{x_1(t) + \beta_{13}x_3(t)}\right]\mathrm{d}t, \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 - a_2x_2(t) + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t) + \beta_{12}x_2(t)} - \dfrac{\alpha_{23}x_3(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t, \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 - a_3x_3(t) + \dfrac{e_{13}\alpha_{13}x_1(t)}{x_1(t) + \beta_{13}x_3(t)} + \dfrac{e_{23}\alpha_{23}x_2(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t, \end{cases} \tag{4}$$

where $x_1(t)$ stands for the total number of prey at time t, while $x_2(t)$ and $x_3(t)$ represent the total number of intermediate predators and omnivorous top predators at time t, respectively. Here, r is the intrinsic growth rate of prey; d_i represents the mortality rate of predator x_i $(i = 2, 3)$; a_i stands for the intraspecific competition rate of species x_i $(i = 1, 2, 3)$; α_{12}, β_{12}, and e_{12} are the encounter rate, half-saturation constant, and conversion rate, respectively, that x_2 preys on x_1; α_{13}, β_{13} and e_{13} stand for the same corresponding denotations that x_3 preys on x_1; and α_{23}, β_{23}, and e_{23} represent the same corresponding denotations that x_3 preys on x_2.

As mentioned above, we notice that population models (1)–(4) are described by the deterministic model. This is valid only at the macroscopic scale, that is, the stochastic effects can be neglected or averaged out, in view of the law of large numbers. However, in the real world, populations are actually subject to the environmental fluctuations. Generally speaking, such fluctuations could be modeled by a colored noise. It has been noted that if the colored noise is not strongly correlated,

then we can approximate the colored noise by a white noise $\dot{w}(t)$, and the approximation works quite well (see [10]). It turns out that the white noise $\dot{w}(t)$ is formally regarded as the derivative of a Brownian motion $w(t)$, i.e., $\dot{w}(t) = \mathrm{d}w(t)/\mathrm{d}t$ (see [11]). As a result, the study of stochastic ecological dynamics model has already become one of the important subjects in biological mathematics.

After taking the effect of randomly fluctuating environment into account, many researchers introduced stochastic environmental variation described by the Brownian motion into parameters in the deterministic model to establish the stochastic population model (see [12–15]). Liu and Bai [12] considered the optimal harvesting problem of a stochastic logistic model with time delay. In [13–15], the authors investigated the dynamics of stochastic predator-prey models. Ji et al. [13] discussed a stochastic predator-prey model with modified Leslie–Gowerand Holling-type II schemes. Jovanović and Krstić [14] investigated the extinction of a stochastic predator-prey model with the Allee effect on the prey. Liu and Jiang [15] considered the periodic solution and stationary distribution of stochastic predator-prey models with higher-order perturbation. In [16], considering that fluctuations in the environment would manifest themselves mainly as fluctuations in the intrinsic growth rate of the prey population and in the death rate of the predator population (see [17]), Ji et al. supposed parameters r and d_2 in model (1) were perturbed with

$$\begin{aligned} r &\longrightarrow r + \sigma_1 \dot{w}_1(t), \\ -d_2 &\longrightarrow -d_2 + \sigma_2 \dot{w}_2(t), \end{aligned} \tag{5}$$

where $w_1(t)$ and $w_2(t)$ are mutually independent Brownian motions and σ_i^2 represents the intensity of white noise $\dot{w}_i(t)$ $(i = 1, 2)$. Moreover, they investigated the long time behavior of the following stochastic ratio-dependent prey-predator model:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1 x_1(t) - \dfrac{\alpha_{12} x_2(t)}{x_1(t) + \beta_{12} x_2(t)}\right]\mathrm{d}t + \sigma_1 x_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + \dfrac{e_{12}\alpha_{12} x_1(t)}{x_1(t) + \beta_{12} x_2(t)}\right]\mathrm{d}t + \sigma_2 x_2(t)\mathrm{d}w_2(t). \end{cases} \tag{6}$$

Based on (6), Wu et al. [18] considered the corresponding nonautonomous stochastic ratio-dependent model. Lv et al. [19] introduced the intraspecific competition of the predator population, denoted by a_2, into model (6). Nguyen and Ta [20] considered a corresponding nonautonomous stochastic ratio-dependent prey-predator model, in which the white noise makes the effect on both the growth rates of species and the intraspecific competition coefficient of the species.

For the study of stochastic three-species models, consult [21–26] and the references therein. Geng et al. [21] investigated the stability of a stochastic one-predator-two-prey population model with time delay, while Liu et al. [22] studied the stability of a stochastic two-predator one-prey population model with time delay. In [23, 24], the authors discussed the dynamical behaviors of stochastic tri-trophic food-chain models. Li et al. [23] investigated the persistence and nonpersistence of a stochastic food-chain model, while Liu and Bai [24] considered the optimal harvesting problem of a stochastic three species food-chain model. Furthermore, in [25, 26], the stochastic three-species food-chain models with omnivory are discussed. Qiu and Deng [25] investigated the stationary distribution and global asymptotic stability of a stochastic food-web model with omnivory and linear functional response, while R. Liu and G. Liu [26] discussed the persistence in mean and extinction of a stochastic food-web model with intraguild predation and mixed functional responses. In [26], the authors considered Holling type-II functional response between the intermediate predator and the top predator and other functional responses were assumed to be linear.

To the best of our knowledge, so far there is no investigation on the dynamics of the stochastic three-species food web model with omnivory and ratio-dependent functional response. The purpose of this paper is to make some contribution in this direction. Recall that parameters r, d_2, and d_3 in model (4) represent the intrinsic growth rate of the prey population, the death rate of the intermediate predator, and the death rate of the omnivorous top predator, respectively. As done in [16], in this paper, we may replace r, d_2, and d_3 in model (4), respectively, by

$$\begin{aligned} r &\longrightarrow r + \sigma_1 \dot{w}_1(t), \\ -d_2 &\longrightarrow -d_2 + \sigma_2 \dot{w}_2(t), \\ -d_3 &\longrightarrow -d_3 + \sigma_3 \dot{w}_3(t), \end{aligned} \tag{7}$$

where $\dot{w}_i(t)$ is the white noise and σ_i^2 is the intensity of white noise $\dot{w}_i(t)$ $(i = 1, 2, 3)$. Then, the stochastic three-species food web model with omnivory and ratio-dependent functional response took the following form:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1 x_1(t) + \dfrac{\alpha_{12} x_2(t)}{x_1(t) + \beta_{12} x_2(t)} - \dfrac{\alpha_{13} x_3(t)}{x_1(t) + \beta_{13} x_3(t)}\right]\mathrm{d}t + \sigma_1 x_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 - a_2 x_2(t) + \dfrac{e_{12}\alpha_{12} x_1(t)}{x_1(t) + \beta_{12} x_2(t)} - \dfrac{\alpha_{23} x_3(t)}{x_2(t) + \beta_{23} x_3(t)}\right]\mathrm{d}t + \sigma_2 x_2(t)\mathrm{d}w_2(t), \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 - a_3 x_3(t) + \dfrac{e_{13}\alpha_{13} x_1(t)}{x_1(t) + \beta_{13} x_3(t)} + \dfrac{e_{23}\alpha_{23} x_2(t)}{x_2(t) + \beta_{23} x_3(t)}\right]\mathrm{d}t + \sigma_3 x_3(t)\mathrm{d}w_3(t), \end{cases} \tag{8}$$

with $(x_1(0), x_2(0), x_3(0)) = (x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+ = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_i > 0\, i = 1, 2, 3\}$. All meanings of the parameters are exact to or similar as those for (4) except the following. Here, $w = \{w_1(t), w_2(t), w_3(t) : t \geq 0\}$ represents the three-dimensional standard Brownian motion defined on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions. σ_i^2 represents the intensity of noise $w_i(t)$ $(i = 1, 2, 3)$. Throughout this paper, unless otherwise specified, we would rather assume that $a_1 > 0$, $a_2 \geq 0$, $a_3 \geq 0$, $\alpha_{13} \geq 0$, $\alpha_{23} \geq 0$, $\beta_{13} > 0$, $\beta_{23} > 0$, $e_{13} > 0$, and $e_{23} > 0$.

2. Existence and Uniqueness of Positive Solution

In this section, we consider the existence of the positive solution for all times. Typically, conditions assuring the nonexplosion of the solution involve local Lipschitz continuity and a linear growth condition. In our case, we miss this last condition, so it is necessary to prove that the solution does not explode at a finite time. To prove the solution is positive and does not explode at a finite time, we use the stochastic comparison theorem. For simplicity, we introduce the following notations:

$$\begin{aligned}
\kappa_1 &= r - \frac{\alpha_{12}}{\beta_{12}} - \frac{\alpha_{13}}{\beta_{13}};\\
\kappa_2 &= e_{12}\alpha_{12} - d_2 - \frac{\alpha_{23}}{\beta_{23}};\\
\kappa_3 &= e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3.
\end{aligned} \tag{9}$$

Theorem 1. *For any given initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$, model (8) has unique global positive solution $(x_1(t), x_2(t), x_3(t))$ for $t \geq 0$, that is, $(x_1(t), x_2(t), x_3(t)) \in \mathbb{R}^3_+$ with probability one for $t \in [0, \infty)$.*

Proof. Consider the following system:

$$\begin{cases}
\mathrm{d}X_1(t) = \left[r - a_1 e^{X_1(t)} + \dfrac{\alpha_{12}e^{X_2(t)}}{e^{X_1(t)} + \beta_{12}e^{X_2(t)}} + \dfrac{\alpha_{13}e^{X_3(t)}}{e^{X_1(t)} + \beta_{13}e^{X_3(t)}} - \dfrac{\sigma_1^2}{2}\right]\mathrm{d}t + \sigma_1 \mathrm{d}w_1(t),\\
\mathrm{d}X_2(t) = \left[-d_2 - a_2 e^{X_2(t)} + \dfrac{e_{12}\alpha_{12}e^{X_1(t)}}{e^{X_1(t)} + \beta_{12}e^{X_2(t)}} - \dfrac{\alpha_{23}e^{X_3(t)}}{e^{X_2(t)} + \beta_{23}e^{X_3(t)}} - \dfrac{\sigma_2^2}{2}\right]\mathrm{d}t + \sigma_2 \mathrm{d}w_2(t),\\
\mathrm{d}X_3(t) = \left[-d_3 - a_3 e^{X_3(t)} + \dfrac{e_{13}\alpha_{13}e^{X_1(t)}}{e^{X_1(t)} + \beta_{13}e^{X_3(t)}} + \dfrac{e_{23}\alpha_{23}e^{X_2(t)}}{e^{X_2(t)} + \beta_{23}e^{X_3(t)}} - \dfrac{\sigma_3^2}{2}\right]\mathrm{d}t + \sigma_3 \mathrm{d}w_3(t),
\end{cases} \tag{10}$$

with initial value $(X_1(0), X_2(0), X_3(0)) = (\ln x_{10}, \ln x_{20}, \ln x_{30})$. Obviously, the coefficients of (10) are locally Lipschitz continuous. Thus, there is a unique maximal local solution $(X_1(t), X_2(t), X_3(t))$ of (10) for $t \in [0, \tau_e)$, where τ_e denotes the explosion time. Let $x_i(t) = e^{X_i(t)}$ $(i = 1, 2, 3)$. Using Itô formula, it follows that $(x_1(t), x_2(t), x_3(t)) = (e^{X_1(t)}, e^{X_2(t)}, e^{X_3(t)})$ is the unique positive local solution of (8) with initial value (x_{10}, x_{20}, x_{30}) for $t \in [0, \tau_e)$.

Next, we show that $(X_1(t), X_2(t), X_3(t))$ is a global solution of (10), that is, $\tau_e = \infty$. Consider the following two stochastic differential systems:

$$\begin{cases}
\mathrm{d}\Phi_1(t) = \Phi_1(t)[r - a_1\Phi_1(t)]\mathrm{d}t + \sigma_1\Phi_1(t)\mathrm{d}w_1(t),\\
\mathrm{d}\Phi_2(t) = \Phi_2(t)[e_{12}\alpha_{12} - a_2\Phi_2(t)]\mathrm{d}t + \sigma_2\Phi_2(t)\mathrm{d}w_2(t),\\
\mathrm{d}\Phi_3(t) = \Phi_3(t)[e_{13}\alpha_{13} + e_{23}\alpha_{23} - a_3\Phi_3(t)]\mathrm{d}t + \sigma_3\Phi_3(t)\mathrm{d}w_3(t),
\end{cases} \tag{11}$$

with initial value $(\Phi_1(0), \Phi_2(0), \Phi_3(0)) = (x_{10}, x_{20}, x_{30})$ and

$$\begin{cases}
\mathrm{d}\phi_1(t) = \phi_1(t)[\kappa_1 - a_1\phi_1(t)]\mathrm{d}t + \sigma_1\phi_1(t)\mathrm{d}w_1(t),\\
\mathrm{d}\phi_2(t) = \phi_2(t)\left[\kappa_2 - \left(a_2 + \dfrac{e_{12}\alpha_{12}\beta_{12}}{\phi_1(t)}\right)\phi_2(t)\right]\mathrm{d}t + \sigma_2\phi_{23}(t)\mathrm{d}w_2(t),\\
\mathrm{d}\phi_3(t) = \phi_3(t)\left[\kappa_3 - \left(a_3 + \dfrac{e_{13}\alpha_{13}\beta_{13}}{\phi_1(t)} + \dfrac{e_{23}\alpha_{23}\beta_{23}}{\phi_2(t)}\right)\phi_3(t)\right]\mathrm{d}t + \sigma_3\phi_3(t)\mathrm{d}w_3(t),
\end{cases} \tag{12}$$

with initial value $(\phi_1(0), \phi_2(0), \phi_3(0)) = (x_{10}, x_{20}, x_{30})$.

Thanks to Lemma 4.2 in [27], systems (11) and (12) can be explicitly solved as follows:

$$\begin{cases} \Phi_1(t) = \dfrac{\exp\{(r-(\sigma_1^2/2))t+\sigma_1 w_1(t)\}}{(1/x_{10})+a_1\int_0^t \exp\{(r-(\sigma_1^2/2))s+\sigma_1 w_1(s)\}\mathrm{d}s}, \\ \Phi_2(t) = \dfrac{\exp\{(e_{12}\alpha_{12}-(\sigma_2^2/2))t+\sigma_2 w_2(t)\}}{(1/x_{20})+a_2\int_0^t \exp\{(e_{12}\alpha_{12}-(\sigma_2^2/2))s+\sigma_2 w_2(s)\}\mathrm{d}s}, \\ \Phi_3(t) = \dfrac{\exp\{(e_{13}\alpha_{13}+e_{23}\alpha_{23}-(\sigma_3^2/2))t+\sigma_3 w_3(t)\}}{(1/x_{30})+a_3\int_0^t \exp\{(e_{13}\alpha_{13}+e_{23}\alpha_{23}-(\sigma_3^2/2))s+\sigma_3 w_3(s)\}\mathrm{d}s}, \end{cases} \tag{13}$$

$$\begin{cases} \phi_1(t) = \dfrac{\exp\{(\kappa_1-(\sigma_1^2/2))t+\sigma_1 w_1(t)\}}{(1/x_{10})+a_1\int_0^t \exp\{(\kappa_1-(\sigma_1^2/2))s+\sigma_1 w_1(s)\}\mathrm{d}s}, \\ \phi_2(t) = \dfrac{\exp\{(\kappa_2-(\sigma_2^2/2))t+\sigma_2 w_2(t)\}}{(1/x_{20})+\int_0^t (a_2+((e_{12}\alpha_{12}\beta_{12})/(\phi_1(s))))\exp\{(\kappa_2-(\sigma_2^2/2))s+\sigma_2 w_2(s)\}\mathrm{d}s}, \\ \phi_3(t) = \dfrac{\exp\{(\kappa_3-(\sigma_3^2/2))t+\sigma_3 w_3(t)\}}{(1/x_{30})+\int_0^t (a_3+((e_{13}\alpha_{13}\beta_{13})/(\phi_1(s)))+((e_{23}\alpha_{23}\beta_{23})/(\phi_2(s))))\exp\{(\kappa_3-(\sigma_3^2/2))s+\sigma_3 w_3(s)\}\mathrm{d}s}. \end{cases}$$

Note that the local solution $(x_1(t), x_2(t), x_3(t))$ is positive on $[0, \tau_e)$. Then, from the comparison theorem of stochastic differential equations (see Theorem 3.1 in [28]), it follows that for $t \in [0, \tau_e)$

$$0 < \phi_i(t) \le x_i(t) \le \Phi_i(t), \quad \text{a.s., } i = 1, 2, 3. \tag{14}$$

Thus, for $t \in [0, \tau_e)$

$$\ln \phi_i(t) \le \ln x_i(t) \le \ln \Phi_i(t), \quad \text{a.s., } i = 1, 2, 3. \tag{15}$$

Since $\ln \phi_i(t)$ and $\ln \Phi_i(t)$ $(i = 1, 2, 3)$ exist for every $t \ge 0$, it follows that $\tau_e = \infty$. Thus, for any initial value $(X_1(0), X_2(0), X_3(0)) = (\ln x_{10}, \ln x_{20}, \ln x_{30}) \in \mathbb{R}^3$, and (10) has a unique global solution $(X_1(t), X_2(t), X_3(t))$ on $[0, \infty)$ a.s. Note that the coefficients of (8) are local Lipschitz continuous. Therefore, for any initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$, model (8) has a unique global positive solution $(x_1(t), x_2(t), x_3(t)) = (e^{X_1(t)}, e^{X_2(t)}, e^{X_3(t)})$ on $[0, \infty)$ a.s. The proof is therefore complete. □

3. Asymptotic Behaviors

Lemma 1 (see [13]). *Consider one-dimensional stochastic differential equation:*

$$\mathrm{d}x(t) = x(t)[a - bx(t)]\mathrm{d}t + \sigma x(t)\mathrm{d}w(t), \tag{16}$$

where a, b, and σ are positive constants and $w(t)$ is standard Brownian motion. For any $x_0 > 0$, let $x(t)$ be the solution of equation (16) with initial value x_0. If $a > (\sigma^2/2)$, then

$$\begin{aligned} &\lim_{t\to\infty} \frac{\ln x(t)}{t} = 0, \\ &\lim_{t\to\infty} \frac{1}{t}\int_0^t x(s)\mathrm{d}s = \frac{a-(\sigma^2/2)}{b}, \quad \text{a.s.} \end{aligned} \tag{17}$$

Theorem 2. *For any $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$, let $(x_1(t), x_2(t), x_3(t))$ be the solution of model (8) with initial value (x_{10}, x_{20}, x_{30}). If $a_2 > 0$, $a_3 > 0$, and $\kappa_i - (\sigma_i^2/2) > 0$ $(i = 1, 2, 3)$, then*

$$\lim_{t\to\infty} \frac{\ln x_i(t)}{t} = 0, \ \text{a.s.}, \quad i = 1, 2, 3. \tag{18}$$

Proof. From Theorem 1, it follows that

$$\phi_i(t) \le x_i(t) \le \Phi_i(t), \ \text{a.s.}, \quad i = 1, 2, 3. \tag{19}$$

Note that $\phi_1(t)$ and $\Phi_1(t)$ are the solutions of the following stochastic equations, respectively,

$$\begin{aligned} \mathrm{d}\phi_1(t) &= \phi_1(t)[\kappa_1 - a_1\phi_1(t)]\mathrm{d}t + \sigma_1\phi_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}\Phi_1(t) &= \Phi_1(t)[r - a_1\Phi_1(t)]\mathrm{d}t + \sigma_1\Phi_1(t)\mathrm{d}w_1(t). \end{aligned} \tag{20}$$

with initial value $x_{10} > 0$. Obviously, from Lemma 1, it follows that if $\kappa_1 - (\sigma_1^2/2) > 0$:

$$\begin{aligned} &\lim_{t\to\infty} \frac{\ln \phi_1(t)}{t} = 0, \\ &\lim_{t\to\infty} \frac{\ln \Phi_1(t)}{t} = 0 \ \text{a.s.} \end{aligned} \tag{21}$$

This, together with (19), yields

$$\lim_{t\longrightarrow\infty}\frac{\ln x_1(t)}{t}=0 \text{ a.s.} \tag{22}$$

Now, we show $\lim_{t\longrightarrow\infty}((\ln x_2(t))/t)=0$ a.s. Note that, $x_2(t)\le\Phi_2(t)$ a.s. and $\Phi_2(t)$ is the solution of equation

$$d\Phi_2(t)=\Phi_2(t)\left[e_{12}\alpha_{12}-a_2\Phi_2(t)\right]\mathrm{d}t+\sigma_2\Phi_2(t)\mathrm{d}w_2(t), \tag{23}$$

with initial $x_{20}>0$. Thus, from Lemma 1 and $\kappa_2-(\sigma_2^2/2)>0$, it follows that

$$\lim_{t\longrightarrow\infty}\frac{\ln\Phi_2(t)}{t}=0 \text{ a.s.} \tag{24}$$

Note that $\lim_{t\longrightarrow\infty}((\ln\phi_1(t))/t)=0$, a.s. Thus, for any $\varepsilon>0$, there exists $T_1>0$ such that

$$e^{-\varepsilon t}\le\phi_1(t)\le e^{\varepsilon t},\quad \text{for } t\ge T_1. \tag{25}$$

By the strong law of large numbers of local martingales (see Theorem 1.3.4 in [11]), it follows that $\lim_{t\longrightarrow\infty}((\sigma_2 w_2(t))/t)=0$ a.s. Thus, for any $\varepsilon>0$, there exists $T_2>0$ such that

$$-\varepsilon t\le\sigma_2 w_2(t)\le\varepsilon t,\quad \text{for } t\ge T_2. \tag{26}$$

From the expression of $\phi_2(t)$ that for any $t\ge T=T_1\vee T_2$, we have

$$\begin{aligned}\frac{1}{\phi_2(t)}&=\frac{1}{x_2(T)}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-T)-\sigma_2\left(w_2(t)-w_2(T)\right)\right]}+a_2\int_T^t e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)-\sigma_2\left(w_2(t)-w_2(s)\right)\right]}\mathrm{d}s\\&\quad+\int_T^t\frac{e_{12}\alpha_{12}\beta_{12}}{\phi_1(s)}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)-\sigma_2\left(w_2(t)-w_2(s)\right)\right]}\mathrm{d}s\\&\le\frac{1}{x_2(T)}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-T)+\varepsilon(t+T)\right]}+a_2\int_T^t e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)+\varepsilon(t+s)\right]}\mathrm{d}s\\&\quad+e_{12}\alpha_{12}\beta_{12}\int_T^t e^{\varepsilon s}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)+\varepsilon(t+s)\right]}\mathrm{d}s.\end{aligned} \tag{27}$$

Hence, from $\kappa_2-(\sigma_2^2/2)>0$, $\varepsilon>0$ and $t\ge T$, it follows that

$$\begin{aligned}\frac{e^{-3\varepsilon(t+T)}}{\phi_2(t)}&\le\frac{1}{x_2(T)}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-T)-2\varepsilon(t+T)\right]}+a_2\int_T^t e^{-\varepsilon(t-s)}e^{-\varepsilon t}e^{-3\varepsilon T}e^{-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)}\mathrm{d}s\\&\quad+e_{12}\alpha_{12}\beta_{12}e^{-3\varepsilon T}\int_T^t e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-s)-2\varepsilon(t-s)\right]}\mathrm{d}s\\&=\frac{1}{x_2(T)}e^{\left[-\left(\kappa_2-\left(\sigma_2^2/2\right)\right)(t-T)-2\varepsilon(t+T)\right]}+\frac{a_2}{\kappa_2+\varepsilon-\left(\sigma_2^2/2\right)}e^{-\varepsilon t}e^{-3\varepsilon T}e^{-\left(\kappa_2+\varepsilon-\left(\sigma_2^2/2\right)\right)T}\\&\quad+\frac{e_{12}\alpha_{12}\beta_{12}}{\kappa_2+2\varepsilon-\left(\sigma_2^2/2\right)}e^{-3\varepsilon T}e^{-\left(\kappa_2+2\varepsilon-\left(\sigma_2^2/2\right)\right)T}\le\frac{1}{x_2(T)}+\frac{a_2}{\kappa_2-\left(\sigma_2^2/2\right)}+\frac{e_{12}\alpha_{12}\beta_{12}}{\kappa_2-\left(\sigma_2^2/2\right)}\doteq K_1.\end{aligned} \tag{28}$$

That is $(1/(\phi_2(t)))\le K_1e^{3\varepsilon(t+T)}$ a.s., for $t\ge T$. Then, $-\ln\phi_2(t)\le\ln K_1+3\varepsilon(t+T)$. Thus, for any $\varepsilon>0$,

$$\liminf_{t\longrightarrow\infty}\frac{\ln\phi_2(t)}{t}\ge 0 \text{ a.s.} \tag{29}$$

Consequently,

$$\begin{aligned}0&\le\liminf_{t\longrightarrow\infty}\frac{\ln\phi_2(t)}{t}\le\liminf_{t\longrightarrow\infty}\frac{\ln x_2(t)}{t}\le\limsup_{t\longrightarrow\infty}\frac{\ln x_2(t)}{t}\\&\le\lim_{t\longrightarrow\infty}\frac{\ln\Phi_2(t)}{t}=0 \text{ a.s.}\end{aligned} \tag{30}$$

In addition,

$$\limsup_{t\longrightarrow\infty}\frac{\ln\phi_2(t)}{t}\le\limsup_{t\longrightarrow\infty}\frac{\ln x_2(t)}{t}\le 0\ \text{a.s.}\tag{31}$$

Thus,

$$\lim_{t\longrightarrow\infty}\frac{\ln\phi_2(t)}{t}=0,$$
$$\lim_{t\longrightarrow\infty}\frac{\ln x_2(t)}{t}=0,\ \text{a.s.}\tag{32}$$

Similarly, if $\kappa_1-(\sigma_1^2/2)>0$, $\kappa_2-(\sigma_2^2/2)>0$, and $\kappa_3-(\sigma_3^2/2)>0$, then

$$\lim_{t\longrightarrow\infty}\frac{\ln x_3(t)}{t}=0\ \text{a.s.}\tag{33}$$

The proof is therefore complete. □

4. Persistence in Mean and Extinction

In this section, we show that under some conditions, model (8) is persistent in mean and extinct.

Theorem 3. *Suppose that $a_2>0$, $a_3>0$, and $\kappa_i-(\sigma_i^2/2)>0$ $(i=1,2,3)$. Then, for any $(x_{10},x_{20},x_{30})\in\mathbb{R}_+^3$, the solution $(x_1(t),x_2(t),x_3(t))$ of model (8) with initial value (x_{10},x_{20},x_{30}) obeys*

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t x_1(s)\mathrm{d}s\ge\frac{\kappa_1-(\sigma_1^2/2)}{a_1}\ \text{a.s.},$$
$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_2x_2(s)+\frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)}\right]\mathrm{d}s\ge\kappa_2-\frac{\sigma_2^2}{2}\ \text{a.s.},$$
$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_3x_3(s)+\frac{e_{13}\alpha_{13}\beta_{13}x_3(s)}{x_1(s)}+\frac{e_{23}\alpha_{23}\beta_{23}x_3(s)}{x_2(s)}\right]\mathrm{d}s$$
$$\ge\kappa_3-\frac{\sigma_3^2}{2}\ \text{a.s.}\tag{34}$$

Proof. For prey x_1, from Theorem 1, it follows that

$$\phi_1(t)\le x_1(t)\ \text{a.s.},\tag{35}$$

and $\phi_1(t)$ is the solution of the following stochastic equation:

$$\mathrm{d}\phi_1(t)=\phi_1(t)[\kappa_1-a_1\phi_1(t)]\mathrm{d}t+\sigma_1\phi_1(t)\mathrm{d}w_1(t),\tag{36}$$

with initial value $x_{10}>0$. Obviously, from Lemma 1, it follows that if $\kappa_1-(\sigma_1^2/2)>0$,

$$\lim_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\phi_1(s)\mathrm{d}s=\frac{\kappa_1-(\sigma_1^2/2)}{a_1}\ \text{a.s.}\tag{37}$$

This, together with (35), yields

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t x_1(s)\mathrm{d}s\ge\lim_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\phi_1(s)\mathrm{d}s$$
$$=\frac{\kappa_1-(\sigma_1^2/2)}{a_1}>0\ \text{a.s.}\tag{38}$$

For intermediate predator x_2, using Itô formula, it follows that

$$\ln x_2(t)=\int_0^t\left[-d_2-a_2x_2(s)+\frac{e_{12}\alpha_{12}x_1(s)}{x_1(s)+\beta_{12}x_2(s)}-\frac{\alpha_{23}x_3(s)}{x_2(s)+\beta_{23}x_3(s)}-\frac{\sigma_2^2}{2}\right]\mathrm{d}s+\sigma_2w_2(t)+\ln x_{20}$$
$$\ge\left[\kappa_2-\frac{\sigma_2^2}{2}\right]t-a_2\int_0^t x_2(s)\mathrm{d}s-\int_0^t\frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)+\beta_{12}x_2(s)}\mathrm{d}s+\sigma_2w_2(t)+\ln x_{20}.\tag{39}$$

Hence,

$$\frac{1}{t}\int_0^t\left[a_2x_2(s)+\frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)}\right]\mathrm{d}s\ge\left[\kappa_2-\frac{\sigma_2^2}{2}\right]+\frac{\sigma_2w_2(t)}{t}+\frac{\ln x_{20}}{t}-\frac{\ln x_2(t)}{t}.\tag{40}$$

By the strong law of numbers of local martingales and Theorem 2, we get

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_2x_2(s)+\frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)}\right]\mathrm{d}s\ge\kappa_2-\frac{\sigma_2^2}{2}\ \text{a.s.}\tag{41}$$

For the omnivorous top predator x_3, it follows from Itô formula that

$$\ln x_3(t)=\int_0^t\left[-d_3-a_3x_3(s)+\frac{e_{13}\alpha_{13}x_1(s)}{x_1(s)+\beta_{13}x_3(s)}+\frac{e_{23}\alpha_{23}x_2(s)}{x_2(s)+\beta_{23}x_3(s)}-\frac{\sigma_3^2}{2}\right]\mathrm{d}s+\sigma_3w_3(t)+\ln x_{30}$$
$$\ge\left[\kappa_3-\frac{\sigma_3^2}{2}\right]t-\int_0^t\left[a_3x_3(s)+\frac{e_{13}\alpha_{13}\beta_{13}x_3(s)}{x_1(s)+\beta_{13}x_3(s)}+\frac{e_{23}\alpha_{23}\beta_{23}x_3(s)}{x_2(s)+\beta_{23}x_3(s)}\right]\mathrm{d}s+\sigma_3w_3(t)+\ln x_{30}.\tag{42}$$

Hence,

$$\frac{1}{t}\int_0^t\left[a_3x_3(s)+\frac{e_{13}\alpha_{13}\beta_{13}x_3(s)}{x_1(s)}+\frac{e_{23}\alpha_{23}\beta_{23}x_3(s)}{x_2(s)}\right]\mathrm{d}s\ge\left[\kappa_3-\frac{\sigma_3^2}{2}\right]+\frac{\sigma_3w_3(t)}{t}+\frac{\ln x_{30}}{t}-\frac{\ln x_3(t)}{t}.\tag{43}$$

By the strong law of numbers of local martingales and Theorem 2, we get

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_3x_3(s)+\frac{e_{13}\alpha_{13}\beta_{13}x_3(s)}{x_1(s)}+\frac{e_{23}\alpha_{23}\beta_{23}x_3(s)}{x_2(s)}\right]\mathrm{d}s \geq \kappa_3-\frac{\sigma_3^2}{2}\ \text{a.s.} \tag{44}$$

The proof is therefore complete. □

Theorem 4. *Suppose that* $r-(\sigma_1^2/2)<0$, $e_{12}\alpha_{12}-d_2-(\sigma_2^2/2)<0$ *and* $e_{13}\alpha_{13}+e_{23}\alpha_{23}-d_3-(\sigma_3^2/2)<0$. *Then, for any initial value* $(x_{10},x_{20},x_{30})\in\mathbb{R}_+^3$, *model (8) is extinct exponentially with probability one.*

Proof. From Itô formula, it follows that

$$\begin{aligned}
\ln x_1(t)&\leq\left[r-\frac{\sigma_1^2}{2}\right]t+\sigma_1w_1(t)+\ln x_{10},\\
\ln x_2(t)&\leq\left[e_{12}\alpha_{12}-d_2-\frac{\sigma_2^2}{2}\right]t+\sigma_2w_2(t)+\ln x_{20},\\
\ln x_3(t)&\leq\left[e_{13}\alpha_{13}+e_{23}\alpha_{23}-d_3-\frac{\sigma_3^2}{2}\right]t+\sigma_3w_3(t)+\ln x_{30}.
\end{aligned} \tag{45}$$

Note that $\lim_{t\longrightarrow\infty}[((\sigma_iw_i(t))/t)+((\ln x_{i0})/t)]=0$ $(i=1,2,3)$ and $r-(\sigma_1^2/2)<0$, $e_{12}\alpha_{12}-d_2-(\sigma_2^2/2)<0$, and $e_{13}\alpha_{13}+e_{23}\alpha_{23}-d_3-(\sigma_3^2/2)<0$. Then,

$$\begin{aligned}
\limsup_{t\longrightarrow\infty}\frac{\ln x_1(t)}{t}&\leq r-\frac{\sigma_1^2}{2}<0\ \text{a.s.},\\
\limsup_{t\longrightarrow\infty}\frac{\ln x_2(t)}{t}&\leq e_{12}\alpha_{12}-d_2-\frac{\sigma_2^2}{2}<0\ \text{a.s.},\\
\limsup_{t\longrightarrow\infty}\frac{\ln x_3(t)}{t}&\leq e_{13}\alpha_{13}+e_{23}\alpha_{23}-d_3-\frac{\sigma_3^2}{2}<0\ \text{a.s.}
\end{aligned} \tag{46}$$

Therefore, model (8) is extinct exponentially. The proof is complete. □

5. Stochastic Permanence

In this section, we discuss the stochastic permanence of model (8). The definition of stochastic permanence and stochastically ultimately boundness of model (8) were introduced in the literature [29, 30] as follows.

Definition 1 (see [29, 30]). Model (8) is called stochastically ultimate bounded, if for any $\varepsilon\in(0,1)$, there exist three positive constants $H_1=H_1(\varepsilon)$, $H_2=H_2(\varepsilon)$, and $H_3=H_3(\varepsilon)$ such that the solution $(x_1(t),x_2(t),x_3(t))$ of model (8) with any initial value $(x_{10},x_{20},x_{30})\in\mathbb{R}_+^3$ has the property that

$$\limsup_{t\longrightarrow\infty}\mathbb{P}\{x_i(t)>H_i\}<\varepsilon,\quad i=1,2,3. \tag{47}$$

Definition 2 (see [29, 30]). Model (8) is said to be stochastically permanent, if for any $\varepsilon\in(0,1)$, there exist positive constants $\delta_i=\delta_i(\varepsilon)$, $H_i=H_i(\varepsilon)$, and $\delta_i<H_i$ $(i=1,2,3)$, such that the solution $(x_1(t),x_2(t),x_3(t))$ of model (8) with any initial value $(x_{10},x_{20},x_{30})\in\mathbb{R}_+^3$ has the property that

$$\begin{aligned}
&\liminf_{t\longrightarrow\infty}\mathbb{P}\{x_i(t)\leq H_i\}\geq 1-\varepsilon,\\
&\liminf_{t\longrightarrow\infty}\mathbb{P}\{x_i(t)\geq\delta_i\}\\
&\quad\geq 1-\varepsilon,\quad i=1,2,3.
\end{aligned} \tag{48}$$

It is obvious that if stochastic model (8) is stochastically permanent, its solutions must be stochastically ultimately bounded.

5.1. Boundness. In this subsection, we investigate the stochastically ultimate boundness of model (8) in two different ways.

Lemma 2 (see [31]). *For any positive constants p, m, and n, the Bernoulli equation*

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=pmx(t)-pnx^{1+(1/p)}(t), \tag{49}$$

with the initial value $x(0)=x_0>0$, *has the solution*

$$x(t)=\left[\frac{m}{n\left(1-e^{-mt}+(m/n)x_0^{-(1/p)}e^{-mt}\right)}\right]^p. \tag{50}$$

Theorem 5. *For any* $(x_{10},x_{20},x_{30})\in\mathbb{R}_+^3$, *let* $(x_1(t),x_2(t),x_3(t))$ *be the solution of model (8) with initial value* (x_{10},x_{20},x_{30}). *If* $a_2>0$ *and* $a_3>0$, *then for any* $p\geq 0$,

$$\begin{aligned}
\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_1^p(t)\right]&\leq\left[\frac{r+(p/2)\sigma_1^2}{a_1}\right]^p,\\
\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_2^p(t)\right]&\leq\left[\frac{e_{12}\alpha_{12}+(p/2)\sigma_2^2}{a_2}\right]^p,\\
\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_3^p(t)\right]&\leq\left[\frac{e_{13}\alpha_{13}+e_{23}\alpha_{23}+(p/2)\sigma_3^2}{a_3}\right]^p.
\end{aligned} \tag{51}$$

That is, the solution of model (8) is uniformly bounded in the pth moment.

Proof. For Φ_1 in system (11), applying Itô formula to Φ_1^p leads to

$$\begin{aligned}
\Phi_1^p(t)=x_{10}^p&+\int_0^t p\Phi_1^p(s)\left[r+\frac{p-1}{2}\sigma_1^2-a_1\Phi_1(s)\right]\mathrm{d}s\\
&+\int_0^t p\sigma_1\Phi_1^p(s)\mathrm{d}w_1(s).
\end{aligned} \tag{52}$$

Taking the expectation on both sides of the above equation, we have

$$\mathbb{E}\left[\Phi_1^p(t)\right] = x_{10}^p + \mathbb{E}\int_0^t p\Phi_1^p(s)\left[r + \frac{p-1}{2}\sigma_1^2 - a_1\Phi_1(s)\right]\mathrm{d}s. \tag{53}$$

Then, using the Höder inequality, it follows that

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}\left[\Phi_1^p(t)\right]}{\mathrm{d}t} &= p\left(r + \frac{p-1}{2}\sigma_1^2\right)\mathbb{E}\left[\Phi_1^p(t)\right] - pa_1\mathbb{E}\left[\Phi_1^{p+1}(t)\right] \\ &\le p\left(r + \frac{p}{2}\sigma_1^2\right)\mathbb{E}\left[\Phi_1^p(t)\right] - pa_1\left(\mathbb{E}\left[\Phi_1^p(t)\right]\right)^{1+(1/p)} \\ &\doteq pb_1\mathbb{E}\left[\Phi_1^p(t)\right] - pa_1\left(\mathbb{E}\left[\Phi_1^p(t)\right]\right)^{1+(1/p)}.\end{aligned} \tag{54}$$

From Lemma 2 and the comparison theorem, it follows that

$$\mathbb{E}\left[\Phi_1^p(t)\right] \le \left[\frac{b_1}{a_1\left(1 - e^{-b_1 t} + (b_1/a_1)x_{10}^{-1}e^{-b_1 t}\right)}\right]^p. \tag{55}$$

Note that $b_1 = r + (p/2)\sigma_1^2 > 0$. Thus,

$$\limsup_{t\longrightarrow\infty}\mathbb{E}\left[\Phi_1^p(t)\right] \le \left[\frac{r + (p/2)\sigma_1^2}{a_1}\right]^p. \tag{56}$$

By a similar the discussion as in $\Phi_1(t)$, we also know that

$$\begin{aligned}&\limsup_{t\longrightarrow\infty}\mathbb{E}\left[\Phi_2^p(t)\right] \le \left[\frac{e_{12}\alpha_{12} + ((p-1)/2)\sigma_2^2}{a_2}\right]^p, \\ &\limsup_{t\longrightarrow\infty}\mathbb{E}\left[\Phi_3^p(t)\right] \le \left[\frac{e_{13}\alpha_{13} + e_{23}\alpha_{23} + ((p-1)/2)\sigma_3^2}{a_3}\right]^p.\end{aligned} \tag{57}$$

From Theorem 1, it follows that $0 < x_i(t) \le \Phi_i(t)$ a.s. $i = 1, 2, 3$. Then, for any $p \ge 0$, we have

$$0 < \mathbb{E}\left[x_i^p(t)\right] \le \mathbb{E}\left[\Phi_i^p(t)\right], \quad i = 1, 2, 3. \tag{58}$$

Now Theorem 5 follows immediately from the above analysis. The proof is complete. □

Theorem 6. *For any $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$, let $(x_1(t), x_2(t), x_3(t))$ be the solution of model (8) with initial value (x_{10}, x_{20}, x_{30}). Then,*

$$\begin{aligned}&\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_1(t)\right] \le \frac{K_2}{d^L\left(e_{12}e_{23} + e_{13}\right)}, \\ &\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_2(t)\right] \le \frac{K_2}{d^L e_{23}}, \\ &\limsup_{t\longrightarrow\infty}\mathbb{E}\left[x_3(t)\right] \le \frac{K_2}{d^L},\end{aligned} \tag{59}$$

where $d^L = d_2 \wedge d_3$ and $K_2 = ((e_{12}e_{23} + e_{13})(r + d^L)^2)/(4a_1)$.

Proof. Define $H(t) = (e_{12}e_{23} + e_{13})x_1(t) + e_{23}x_2(t) + x_3(t)$. By Itô formula, we have

$$\begin{aligned}\mathrm{d}H(t) &= \left(e_{12}e_{23} + e_{13}\right)\left[\left(rx_1 - a_1x_1^2 - \frac{\alpha_{12}x_1x_2}{x_1 + \beta_{12}x_2} - \frac{\alpha_{13}x_1x_3}{x_1 + \beta_{13}x_3}\right)\mathrm{d}t + \sigma_1x_1\mathrm{d}w_1(t)\right] + e_{23}\left[\left(-d_2x_2 - a_2x_2^2 + \frac{e_{12}\alpha_{12}x_1x_2}{x_1 + \beta_{12}x_2} - \frac{\alpha_{23}x_2x_3}{x_2 + \beta_{23}x_3}\right)\mathrm{d}t + \sigma_2x_2\mathrm{d}w_2(t)\right] + \left[\left(-d_3x_3 - a_3x_3^2 + \frac{e_{13}\alpha_{13}x_1x_3}{x_1 + \beta_{13}x_3} + \frac{e_{23}\alpha_{23}x_2x_3}{x_2 + \beta_{23}x_3}\right)\mathrm{d}t + \sigma_3x_3\mathrm{d}w_3(t)\right] \\ &= \Bigg[-e_{23}d_2x_2 - d_3x_3 + \left(e_{12}e_{23} + e_{13}\right)rx_1 - \left(e_{12}e_{23} + e_{13}\right)a_1x_1^2 - e_{23}a_2x_2^2 - a_3x_3^2\Big) - \frac{e_{12}e_{23}\alpha_{13}x_1x_3}{x_1 + \beta_{13}x_3} - \frac{e_{13}\alpha_{12}x_1x_2}{x_1 + \beta_{12}x_2}\Bigg]\mathrm{d}t \\ &\quad + \left(e_{12}e_{23} + e_{13}\right)\sigma_1x_1\mathrm{d}w_1(t) + e_{23}\sigma_2x_2\mathrm{d}w_2(t) + \sigma_3x_3\mathrm{d}w_3(t).\end{aligned} \tag{60}$$

Integrating it from 0 to t and taking expectation yields

$$\begin{aligned}\mathbb{E}[H(t)] = H(0) + \mathbb{E}\int_0^t\Bigg[&-e_{23}d_2x_2 - d_3x_3 + \left(e_{12}e_{23} + e_{13}\right)rx_1 - \left(e_{12}e_{23} + e_{13}\right)a_1x_1^2 - e_{23}a_2x_2^2 - a_3x_3^2 \\ &- \frac{e_{12}e_{23}\alpha_{13}x_1x_3}{x_1 + \beta_{13}x_3} - \frac{e_{13}\alpha_{12}x_1x_2}{x_1 + \beta_{12}x_2}\Bigg]\mathrm{d}s.\end{aligned} \tag{61}$$

Thus, using the Hölder inequality yields

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}[H(t)]}{\mathrm{d}t} &= -e_{23}d_2\mathbb{E}\left[x_2(t)\right] - d_3\mathbb{E}\left[x_3(t)\right] + \left(e_{12}e_{23} + e_{13}\right)r\mathbb{E}\left[x_1(t)\right] - \left(e_{12}e_{23} + e_{13}\right)a_1\mathbb{E}\left[x_1^2(t)\right] - e_{23}a_2\mathbb{E}\left[x_2^2(t)\right] - a_3\mathbb{E}\left[x_3^2(t)\right] \\ &\quad - e_{12}e_{23}\alpha_{13}\mathbb{E}\left[\frac{x_1(t)x_3(t)}{x_1(t) + \beta_{13}x_3(t)}\right] - e_{13}\alpha_{12}\mathbb{E}\left[\frac{x_1(t)x_2(t)}{x_1(t) + \beta_{12}x_2(t)}\right] \\ &\le -d^L\left(e_{12}e_{23} + e_{13}\right)\mathbb{E}\left[x_1(t)\right] - d^Le_{23}\mathbb{E}\left[x_2(t)\right] - d^L\mathbb{E}\left[x_3(t)\right] + \left(e_{12}e_{23} + e_{13}\right)\left(r + d^L\right)\mathbb{E}\left[x_1(t)\right] - \left(e_{12}e_{23} + e_{13}\right)a_1\left(\mathbb{E}\left[x_1(t)\right]\right)^2 \\ &= \left(e_{12}e_{23} + e_{13}\right)\left[\left(r + d^L\right)\mathbb{E}\left[x_1(t)\right] - a_1\left(\mathbb{E}\left[x_1(t)\right]\right)^2\right] - d^L\mathbb{E}[H(t)].\end{aligned} \tag{62}$$

It is clear that quadratic function $g(x) = (e_{12}e_{23} + e_{13})[(r + d^L)x - a_1x^2]$ reaches its maximum value at $x = ((r + d^L)/(2a_1)) > 0$. Thus, $g_{\max} = (((e_{12}e_{23} + e_{13})(r + d^L)^2)/(4a_1)) \doteq K_2$. Therefore, we have

$$\frac{\mathrm{d}\mathbb{E}[H(t)]}{\mathrm{d}t} \le K_2 - d^L\mathbb{E}[H(t)]. \tag{63}$$

Then, by the comparison theorem, we have

$$0 \le \limsup_{t\longrightarrow\infty} \mathbb{E}[H(t)] \le \frac{K_2}{d^L}. \tag{64}$$

Note that the solution of model (8) is positive. Thus,

$$\begin{aligned}
&\limsup_{t\longrightarrow\infty} \mathbb{E}[x_1(t)] \le \frac{K_2}{d^L(e_{12}e_{23} + e_{13})},\\
&\limsup_{t\longrightarrow\infty} \mathbb{E}[x_2(t)] \le \frac{K_2}{d^L e_{23}},\\
&\limsup_{t\longrightarrow\infty} \mathbb{E}[x_3(t)] \le \frac{K_2}{d^L}.
\end{aligned} \tag{65}$$

The proof is therefore complete. □

Theorem 7. *Model (8) is stochastically ultimate bounded.*

Proof. Let $(x_1(t), x_2(t), x_3(t))$ be solution of (8) with any initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$. For any $\varepsilon \in (0,1)$, let $H_1 = (K_2/(d^L(e_{12}e_{23} + e_{13})\varepsilon)) + 1$, $H_2 = (K_2/d^L e_{23}\varepsilon) + 1$, and $H_3 = (K_2/d^L\varepsilon) + 1$. Then, by Chebyshev's inequality

$$\mathbb{P}\{x_i(t) > H_i\} \le \frac{\mathbb{E}[x_i(t)]}{H_i}, \quad i = 1,2,3. \tag{66}$$

Hence, from Theorem 6

$$\limsup_{t\longrightarrow\infty} \mathbb{P}\{x_i(t) > H_i\} \le \limsup_{t\longrightarrow\infty} \frac{\mathbb{E}[x_i(t)]}{H_i} < \varepsilon, \quad i = 1,2,3. \tag{67}$$

The proof is therefore complete. □

5.2. Stochastic Permanence. In this section, we give some sufficient conditions to guarantee that model (8) is stochastically permanent. Denote $\gamma_i \doteq \kappa_i - \sigma_i^2$ $(i = 1,2,3)$. Define

$$\begin{aligned}
u_1(t) &= \frac{1}{\phi_1(t)},\\
u_2(t) &= \frac{1}{\phi_2(t)},\\
u_3(t) &= \frac{1}{\phi_3(t)}.
\end{aligned} \tag{68}$$

By the Itô formula, we have

$$\begin{cases}
\mathrm{d}u_1(t) = [a_1 - \gamma_1 u_1(t)]\mathrm{d}t - \sigma_1 u_1(t)\mathrm{d}w_1(t),\\
\mathrm{d}u_2(t) = [a_2 - \gamma_2 u_2(t) + e_{12}\alpha_{12}\beta_{12}u_1(t)]\mathrm{d}t\\
\quad - \sigma_2 u_2(t)\mathrm{d}w_2(t),\\
\mathrm{d}u_3(t) = [a_3 - \gamma_3 u_3(t) + e_{13}\alpha_{13}\beta_{13}u_1(t)\\
\quad + e_{23}\alpha_{23}\beta_{23}u_2(t)]\mathrm{d}t - \sigma_3 u_3(t)\mathrm{d}w_3(t),
\end{cases} \tag{69}$$

with initial value $(u_1(0), u_2(0), u_3(0)) = (1/x_{10}, 1/x_{20}, 1/x_{30}) \in \mathbb{R}^3_+$.

Lemma 3. *Let $(x_1(t), x_2(t), x_3(t))$ be the solution of model (8) with any initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$. If $\gamma_i > 0$ $(i = 1,2,3)$, then*

$$\limsup_{t\longrightarrow\infty} \mathbb{E}\left[\frac{1}{x_i(t)}\right] \le M_i, \quad i = 1,2,3, \tag{70}$$

where $M_1 = (a_1/\gamma_1)$, $M_2 = (a_2/\gamma_2) + ((a_1e_{12}\alpha_{12}\beta_{12})/(\gamma_1\gamma_2))$, and $M_3 = (a_3/\gamma_3) + ((a_1e_{13}\alpha_{13}\beta_{13})/\gamma_1\gamma_3) + +((e_{23}\alpha_{23}\beta_{23})/\gamma_3)[(a_2/\gamma_2) + ((a_1e_{12}\alpha_{12}\beta_{12})/\gamma_1\gamma_2)]$.

Proof. First, integrating both sides of the first equation of (69) from 0 to t yields

$$u_1(t) = \frac{1}{x_{10}} + \int_0^t [a_1 - \gamma_1 u_1(s)]\mathrm{d}s - \int_0^t \sigma_1 u_1(s)\mathrm{d}w_1(s). \tag{71}$$

Taking the expectation on both sides of the above equation, we have

$$\mathbb{E}[u_1(t)] = \frac{1}{x_{10}} + \mathbb{E}\int_0^t [a_1 - \gamma_1 u_1(s)]\mathrm{d}s. \tag{72}$$

Thus,

$$\frac{\mathrm{d}\mathbb{E}[u_1(t)]}{\mathrm{d}t} = a_1 - \gamma_1\mathbb{E}[u_1(t)], \tag{73}$$

with initial value $\mathbb{E}[u_1(0)] = 1/x_{10}$. By a simple computation, we can get

$$\mathbb{E}[u_1(t)] = \frac{1}{x_{10}}e^{-\gamma_1 t} + \frac{a_1}{\gamma_1}\left[1 - e^{-\kappa_1 t}\right]. \tag{74}$$

This, together with $\gamma_1 > 0$, yields

$$\lim_{t\longrightarrow\infty} \mathbb{E}[u_1(t)] = \frac{a_1}{\gamma_1} = M_1. \tag{75}$$

Next, integrating both sides of the second equation of system (69) from 0 to t yields

$$\begin{aligned}
u_2(t) = \frac{1}{x_{20}} &+ \int_0^t [a_2 - \gamma_2 u_2(s) + e_{12}\alpha_{12}\beta_{12}u_1(s)]\mathrm{d}s\\
&- \int_0^t \sigma_2 u_2(s)\mathrm{d}w_2(s).
\end{aligned} \tag{76}$$

Taking the expectation on both sides of the above equation, we have

$$\mathbb{E}[u_2(t)] = \frac{1}{x_{20}} + \mathbb{E}\int_0^t [a_2 - \gamma_2 u_2(s) + e_{12}\alpha_{12}\beta_{12}u_1(s)]\mathrm{d}s. \tag{77}$$

Thus,

$$\frac{\mathrm{d}\mathbb{E}[u_2(t)]}{\mathrm{d}t} = a_2 - \gamma_2\mathbb{E}[u_2(t)] + e_{12}\alpha_{12}\beta_{12}\mathbb{E}[u_1(t)], \tag{78}$$

with initial value $\mathbb{E}[u_2(0)] = (1/x_{20})$. By a simple computation, we can get

$$\begin{aligned}\mathbb{E}[u_2(t)] &= \frac{1}{x_{20}}e^{-\gamma_2 t} + \frac{a_2}{\gamma_2}\left[1 - e^{-\gamma_2 t}\right] \\ &\quad + e_{12}\alpha_{12}\beta_{12}\int_0^t e^{-\gamma_2(t-s)}\mathbb{E}[u_1(s)]\mathrm{d}s.\end{aligned} \tag{79}$$

From (75), it follows that

$$\begin{aligned}\lim_{t\longrightarrow\infty}\int_0^t e^{-\gamma_2(t-s)}\mathbb{E}[u_1(s)]\mathrm{d}s &= \lim_{t\longrightarrow\infty}\frac{\int_0^t e^{\gamma_2 s}\mathbb{E}[u_1(s)]\mathrm{d}s}{e^{\gamma_2 t}} \\ &= \lim_{t\longrightarrow\infty}\frac{\mathbb{E}[u_1(t)]}{\gamma_2} = \frac{a_1}{\gamma_1\gamma_2}.\end{aligned} \tag{80}$$

This, together with (79) yields

$$\lim_{t\longrightarrow\infty}\mathbb{E}[u_2(t)] \le \frac{a_2}{\gamma_2} + \frac{a_1 e_{12}\alpha_{12}\beta_{12}}{\gamma_1\gamma_2}. \tag{81}$$

At last, integrating both sides of the third equation of system (69) from 0 to t and taking the expectation, we have

$$\begin{aligned}\mathbb{E}[u_3(t)] &= \frac{1}{x_{30}} + \mathbb{E}\int_0^t [a_3 - \gamma_3 u_3(s) + e_{13}\alpha_{13}\beta_{13}u_1(s) \\ &\quad + e_{23}\alpha_{23}\beta_{23}u_2(s)]\mathrm{d}s.\end{aligned} \tag{82}$$

Thus,

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}[u_3(t)]}{\mathrm{d}t} &= a_3 - \gamma_3\mathbb{E}[u_3(t)] + e_{13}\alpha_{13}\beta_{13}\mathbb{E}[u_1(t)] \\ &\quad + e_{23}\alpha_{23}\beta_{23}\mathbb{E}[u_2(t)],\end{aligned} \tag{83}$$

with initial value $\mathbb{E}[u_3(0)] = (1/x_{30})$. By a simple computation, we can get

$$\begin{aligned}\mathbb{E}[u_3(t)] &= \frac{1}{x_{30}}e^{-\gamma_3 t} + \frac{a_3}{\gamma_3}\left[1 - e^{-\gamma_3 t}\right] \\ &\quad + e_{13}\alpha_{13}\beta_{13}\int_0^t e^{-\gamma_3(t-s)}\mathbb{E}[u_1(s)]\mathrm{d}s + e_{23}\alpha_{23}\beta_{23} \\ &\quad \cdot \int_0^t e^{-\gamma_3(t-s)}\mathbb{E}[u_2(t)]\mathrm{d}s.\end{aligned} \tag{84}$$

It follows from (75), (81), and (84) that

$$\begin{aligned}\lim_{t\longrightarrow\infty}\mathbb{E}[u_3(t)] &= \frac{a_3}{\gamma_3} + e_{13}\alpha_{13}\beta_{13}\lim_{t\longrightarrow\infty}\int_0^t e^{-\gamma_3(t-s)}\mathbb{E}[u_1(s)]\mathrm{d}s \\ &\quad + e_{23}\alpha_{23}\beta_{23}\lim_{t\longrightarrow\infty}\int_0^t e^{-\gamma_3(t-s)}\mathbb{E}[u_2(t)]\mathrm{d}s \\ &= \frac{a_3}{\gamma_3} + \frac{a_1 e_{13}\alpha_{13}\beta_{13}}{\gamma_1\gamma_3} + \frac{e_{23}\alpha_{23}\beta_{23}}{\gamma_3} \\ &\quad \cdot \left[\frac{a_2}{\gamma_2} + \frac{a_1 e_{12}\alpha_{12}\beta_{12}}{\gamma_1\gamma_2}\right].\end{aligned} \tag{85}$$

From the comparison theorem of stochastic differential equations, it follows that

$$\frac{1}{x_i(t)} \le \frac{1}{\phi_i(t)} = u_i(t), \quad i = 1, 2, 3. \tag{86}$$

Now, Lemma 3 follows immediately from the above analysis. The proof is complete. □

Theorem 8. *If $\gamma_i > 0$ $(i = 1, 2, 3)$, then model (8) is stochastically permanent.*

Proof. Let $(x_1(t), x_2(t), x_3(t))$ be solution of (8) with initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$. For any $\varepsilon \in (0, 1)$, let $\delta_i = (\varepsilon/M_i)$ $(i = 1, 2, 3)$, then

$$\begin{aligned}\mathbb{P}\{x_i(t) < \delta_i\} &= \mathbb{P}\left\{\frac{1}{x_i(t)} > \frac{1}{\delta_i}\right\} \le \frac{\mathbb{E}[1/x_i(t)]}{1/\delta_i} \\ &= \delta_i\mathbb{E}\left[\frac{1}{x_i(t)}\right], \quad i = 1, 2, 3.\end{aligned} \tag{87}$$

Thus, from Lemma 3, it follows that

$$\limsup_{t\longrightarrow\infty}\mathbb{P}\{x_i(t) < \delta_i\} \le \limsup_{t\longrightarrow\infty}\delta_i\mathbb{E}\left[\frac{1}{x_i(t)}\right] \le \varepsilon, \quad i = 1, 2, 3. \tag{88}$$

This implies

$$\liminf_{t\longrightarrow\infty}\mathbb{P}\{x_i(t) \ge \delta_i\} \ge 1 - \varepsilon, \quad i = 1, 2, 3. \tag{89}$$

Let $\varepsilon \in (0, 1)$ be sufficiently small such that $\delta_i < H_i$. From (67) and Definition 2, model (8) is stochastically permanent. The proof is therefore complete. □

6. Stationary Distribution and Ergodicity

In this section, we will show that there is an ergodic stationary distribution for the solution of (8). For the completeness of the paper, in this section, we list some theories about stationary distribution (see [32]). Let $X(t)$ be a homogeneous Markov process in E_d (denotes d-dimensional Euclidean space), described by the following stochastic differential equation:

$$\mathrm{d}X(t) = b(X(t))\mathrm{d}t + g(X(t))\mathrm{d}W(t), X(0) = X_0. \tag{90}$$

The diffusion matrix of the process $X(t)$ is defined as $J(X) = g(X)g^{\mathbb{T}}(X) = (a_{ij}(X))$.

Definition 3 (see [32]). Let $\mathbb{P}(t, X, \cdot)$ be the probability measure induced by $X(t)$ with initial value $X(0) = X_0$. That is, $\mathbb{P}(t, X_0, A) = \mathbb{P}(X(t) \in A \mid X(0) = X_0)$, for any Borel set $A \in \mathscr{B}(\mathbb{R}_+^d)$. If there exists a probability measure $\mu(\cdot)$ such that $\lim_{t\longrightarrow\infty}\mathbb{P}(t, X_0, A) = \mu(A)$ for all $X_0 \in \mathbb{R}_+^d$ and $A \in \mathscr{B}(\mathbb{R}_+^d)$, then we say that stochastic differential equation (90) has a stationary distribution $\mu(\cdot)$.

Lemma 4 (see [33, 34]). *Assume that there exists a bounded domain $D \subset E_d$ with regular boundary Γ and*

(i) (A1) There is a positive number M such that $\sum_{i,j=1}^d a_{ij}(X)\xi_i\xi_j \geq M|\xi|^2$, $X \in D$, and $\xi \in \mathbb{R}^d$;

(ii) (A2) There exists a nonnegative C^2-function V such that there exists a positive constant C, such that

$$LV \leq -C \text{ for any } X \in \frac{E_d}{D}. \tag{91}$$

Then, the Markov process $X(t)$ has a unique ergodic stationary distribution $\mu(\cdot)$. Moreover, if $f(\cdot)$ is a function integrable with respect to the measure μ, then

$$\mathbb{P}\left\{\lim_{T\longrightarrow\infty}\frac{1}{T}\int_0^T f(X(t))\mathrm{d}t = \int_{E_d} f(x)\mu(\mathrm{d}x)\right\} = 1. \tag{92}$$

Let $X(t) = (x_1(t), x_2(t), x_3(t))^{\mathbb{T}}$, $g(X) = \mathrm{diag}(\sigma_1 x_1, \sigma_2 x_2, \sigma_3 x_3)$, $W(t) = (w_1(t), w_2(t), w_3(t))^{\mathbb{T}}$, and

$$b(X) = \begin{pmatrix} x_1\left(r - a_1x_1 - \dfrac{\alpha_{12}x_2}{x_1 + \beta_{12}x_2} - \dfrac{\alpha_{13}x_3}{x_1 + \beta_{13}x_3}\right) \\ x_2\left(-d_2 - a_2x_2 + \dfrac{e_{12}\alpha_{12}x_1}{x_1 + \beta_{12}x_2} - \dfrac{\alpha_{23}x_3}{x_2 + \beta_{23}x_3}\right) \\ x_3\left(-d_3 - a_3x_3 + \dfrac{e_{13}\alpha_{13}x_1}{x_1 + \beta_{13}x_3} + \dfrac{e_{23}\alpha_{23}x_2}{x_2 + \beta_{23}x_3}\right) \end{pmatrix}. \tag{93}$$

Then, system (90) reduces to model (8) with diffusion matrix $J(X) = \mathrm{diag}(\sigma_1^2x_1^2, \sigma_2^2x_2^2, \sigma_3^2x_3^2)$.

Theorem 9. *If $a_2 > 0$, $a_3 > 0$, $\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2 > 0$, $\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2 > 0$, and $\kappa_3 - \sigma_3^2 > 0$, then for any $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$, model (8) has a stationary distribution and the solutions have ergodic property.*

Proof. Define C^2-function V_1: $\mathbb{R}_+^3 \longrightarrow \mathbb{R}_+$ by

$$V_1(X) = x_1 + x_2 + x_3, \tag{94}$$

for $X = (x_1, x_2, x_3) \in \mathbb{R}_+^3$. By Itô formula, we have

$$\begin{aligned} LV_1(X) &= x_1\left(r - a_1x_1 - \frac{\alpha_{12}x_2}{x_1 + \beta_{12}x_2} - \frac{\alpha_{13}x_3}{x_1 + \beta_{13}x_3}\right) \\ &\quad + x_2\left(-d_2 - a_2x_2 + \frac{e_{12}\alpha_{12}x_1}{x_1 + \beta_{12}x_2} - \frac{\alpha_{23}x_3}{x_2 + \beta_{23}x_3}\right) \\ &\quad + x_3\left(-d_3 - a_3x_3 + \frac{e_{13}\alpha_{13}x_1}{x_1 + \beta_{13}x_3} + \frac{e_{23}\alpha_{23}x_2}{x_2 + \beta_{23}x_3}\right) \\ &\leq -a_1x_1^2 + rx_1 - a_2x_2^2 + (e_{12}\alpha_{12} - d_2)x_2 - a_3x_3^2 \\ &\quad + (e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3)x_3. \end{aligned} \tag{95}$$

Define C^2-function V_2: $\mathbb{R}_+^3 \longrightarrow \mathbb{R}_+$ by

$$V_2(X) = x_1^{-1} + x_2^{-1} + x_3^{-1}, \tag{96}$$

for $X = (x_1, x_2, x_3) \in \mathbb{R}_+^3$. By Itô formula, we have

$$\begin{aligned} LV_2(X) &= -x_1^{-1}\left(r - a_1x_1 - \frac{\alpha_{12}x_2}{x_1 + \beta_{12}x_2} - \frac{\alpha_{13}x_3}{x_1 + \beta_{13}x_3}\right) + \sigma_1^2x_1^{-1} \\ &\quad - x_2^{-1}\left(-d_2 - a_2x_2 + \frac{e_{12}\alpha_{12}x_1}{x_1 + \beta_{12}x_2} - \frac{\alpha_{23}x_3}{x_2 + \beta_{23}x_3}\right) \\ &\quad + \sigma_2^2x_2^{-1} \\ &\quad - x_3^{-1}\left(-d_3 - a_3x_3 + \frac{e_{13}\alpha_{13}x_1}{x_1 + \beta_{13}x_3} + \frac{e_{23}\alpha_{23}x_2}{x_2 + \beta_{23}x_3}\right) \\ &\quad + \sigma_3^2x_3^{-1} \\ &\leq -x_1^{-1}(\kappa_1 - a_1x_1) + \sigma_1^2x_1^{-1} \\ &\quad - x_2^{-1}\left(\kappa_2 - a_2x_2 - \frac{e_{12}\alpha_{12}\beta_{12}x_2}{x_1 + \beta_{12}x_2}\right) + \sigma_2^2x_2^{-1} \\ &\quad - x_3^{-1}\left(\kappa_3 - a_3x_3 - \frac{e_{13}\alpha_{13}\beta_{13}x_3}{x_1 + \beta_{13}x_3} - \frac{e_{23}\alpha_{23}\beta_{23}x_3}{x_2 + \beta_{23}x_3}\right) \\ &\quad + \sigma_3^2x_3^{-1} \\ &\leq -x_1^{-1}(\kappa_1 - a_1x_1) + \sigma_1^2x_1^{-1} \\ &\quad - x_2^{-1}\left(\kappa_2 - a_2x_2 - \frac{e_{12}\alpha_{12}\beta_{12}x_2}{x_1}\right) + \sigma_2^2x_2^{-1} \\ &\quad - x_3^{-1}\left(\kappa_3 - a_3x_3 - \frac{e_{13}\alpha_{13}\beta_{13}x_3}{x_1} - \frac{e_{23}\alpha_{23}\beta_{23}x_3}{x_2}\right) + \sigma_3^2x_3^{-1} \\ &= -\left(\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2\right)x_1^{-1} + a_1 \\ &\quad - \left(\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2\right)x_2^{-1} + a_2 - \left(\kappa_3 - \sigma_3^2\right)x_3^{-1} + a_3. \end{aligned} \tag{97}$$

Let $V(X) = V_1(X) + V_2(X)$. Then,

$$\begin{aligned} LV(X) \leq & -a_1x_1^2 + rx_1 - \left(\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2\right)x_1^{-1} \\ & + a_1 - a_2x_2^2 + \left(e_{12}\alpha_{12} - d_2\right)x_2 \\ & - \left(\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2\right)x_2^{-1} + a_2 - a_3x_3^2 \\ & + \left(e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3\right)x_3 - \left(\kappa_3 - \sigma_3^2\right)x_3^{-1} + a_3 \\ = & f(x_1) + g(x_2) + h(x_3), \end{aligned} \tag{98}$$

where

$$\begin{aligned} f(x_1) = & -a_1x_1^2 + rx_1 - \left(\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2\right)x_1^{-1} \\ & + a_1, \\ g(x_2) = & -a_2x_2^2 + \left(e_{12}\alpha_{12} - d_2\right)x_2 - \left(\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2\right)x_2^{-1} \\ & + a_2, \\ h(x_3) = & -a_3x_3^2 + \left(e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3\right)x_3 - \left(\kappa_3 - \sigma_3^2\right)x_3^{-1} \\ & + a_3. \end{aligned} \tag{99}$$

Clearly, $f(x_1)$, $g(x_2)$, and $h(x_3)$ have upper bound on $\mathbb{R}_+$. Denote

$$\begin{aligned} f^u &= \sup_{x_1 \in \mathbb{R}_+} \{f(x_1)\}, \\ g^u &= \sup_{x_2 \in \mathbb{R}_+} \{g(x_2)\}, \\ h^u &= \sup_{x_3 \in \mathbb{R}_+} \{h(x_3)\}. \end{aligned} \tag{100}$$

From $\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2 > 0$, it follows that

$$\begin{aligned} LV(X) \leq & f(x_1) + g(x_2) + h(x_3) \leq f(x_1) + g^u \\ & + h^u \longrightarrow -\infty, \quad \text{a.s. } x_1 \longrightarrow 0^+ \text{ or } x_1 \longrightarrow +\infty. \end{aligned} \tag{101}$$

Similarly, from $\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2 > 0$ and $\kappa_3 - \sigma_3^2 > 0$, we have

$$\begin{aligned} LV(X) \leq & f(x_1) + g(x_2) + h(x_3) \leq f^u + g(x_2) \\ & + h^u \longrightarrow -\infty, \quad \text{a.s. } x_2 \longrightarrow 0^+ \text{ or } x_2 \longrightarrow +\infty, \\ LV(X) \leq & f(x_1) + g(x_2) + h(x_3) \leq f^u + g^u \\ & + h(x_3) \longrightarrow -\infty, \quad \text{a.s. } x_3 \longrightarrow 0^+ \text{ or } x_3 \longrightarrow +\infty. \end{aligned} \tag{102}$$

Consequently, there exists $\rho > 0$ (sufficiently small) such that

$$LV(X) \leq -1, \quad \text{for all } (x_1, x_2, x_3) \in \frac{\mathbb{R}_+^3}{D}, \tag{103}$$

where

$$D = \left\{(x_1, x_2, x_3) \in \mathbb{R}_+^3 \mid \rho < x_1 < \frac{1}{\rho}, \rho < x_2 < \frac{1}{\rho}, \rho < x_3 < \frac{1}{\rho}\right\} \subset \mathbb{R}_+^3. \tag{104}$$

Hence, (A2) in Lemma 4 is satisfied.

Denote $\sigma^2 = \sigma_1^2 \wedge \sigma_2^2 \wedge \sigma_3^2$. Then, for any $X = (x_1, x_2, x_3) \in D$ and $\xi = (\xi_1, \xi_2, \xi_2) \in \mathbb{R}^3$, we have

$$\sum_{i,j=1}^{3} a_{ij}(X)\xi_i\xi_j = \sigma_1^2x_1^2\xi_1^2 + \sigma_2^2x_2^2\xi_2^2 + \sigma_3^2x_3^2\xi_3^2 \geq M|\xi|^2, \tag{105}$$

where $M = \rho^2\sigma^2$. Thus, condition (A1) of Lemma 4 holds. According to Lemma 4, model (8) is ergodic and admits a unique stationary distribution. The proof is therefore complete. □

7. Application of Main Results

In this section, we present the application of the main results in some special models.

7.1. Two Species Predator-Prey Model. Let $\alpha_{13} = \alpha_{23} = 0$. Then, the first two equations of (8) form the following closed two-population system:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1x_1(t) + \dfrac{\alpha_{12}x_2(t)}{x_1(t) + \beta_{12}x_1(t)}\right]\mathrm{d}t \\ \quad + \sigma_1x_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 - a_2x_2(t) + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t) + \beta_{12}x_2(t)}\right]\mathrm{d}t \\ \quad + \sigma_2x_2(t)\mathrm{d}w_2(t), \end{cases} \tag{106}$$

with initial value $(x_{10}, x_{20}) \in \mathbb{R}_+^2$. This is also the stochastic predator-prey model discussed in [19]. From Theorems 3 and 4, we have the following result.

Corollary 1. *Let $(x_1(t), x_2(t))$ be solution of model (106) with initial value $(x_{10}, x_{20}) \in \mathbb{R}_+^2$.*

(i) If $a_2 > 0$, $r - (\alpha_{12}/\beta_{12}) - (\sigma_1^2/2) > 0$, and $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) > 0$, then

$$\begin{aligned} & \liminf_{t\longrightarrow\infty} \frac{1}{t}\int_0^t x_1(s)\mathrm{d}s \geq \frac{r - (\alpha_{12}/\beta_{12}) - (\sigma_1^2/2)}{a_1} > 0 \quad \text{a.s.}, \\ & \liminf_{t\longrightarrow\infty} \frac{1}{t}\int_0^t \left[a_2x_2(s) + \frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)}\right]\mathrm{d}s \\ & \quad \geq e_{12}\alpha_{12} - d_2 - \frac{\sigma_2^2}{2} > 0 \quad \text{a.s.} \end{aligned} \tag{107}$$

(ii) If $r - (\sigma_1^2/2) < 0$ and $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) < 0$, then (106) is extinct exponentially with probability one.

Remark 1. It is clear that Corollary 1 is consistent with Theorems 7 and 8 in [19]. Moreover, from Theorems 3 and 4, the persistence in mean and extinction conditions of the three-species model (8) are more complicated. Thus, our work can be seen as the extension of [19].

For model (106), similar to the proof of Theorem 6 (denote $H = e_{12}x_1 + x_2$), we have

$$\begin{aligned} \limsup_{t\longrightarrow\infty} \mathbb{E}\left[x_1(t)\right] &\leq \frac{K_2'}{e_{12}d_2}, \\ \limsup_{t\longrightarrow\infty} \mathbb{E}\left[x_2(t)\right] &\leq \frac{K_2'}{d_2}, \end{aligned} \tag{108}$$

where $K_2' = e_{12}(r + d_2)^2/4a_1$. Furthermore, from Theorems 7–9, for model (106), we have the following result.

Corollary 2

(i) Model (106) is stochastically ultimate bounded

(ii) If $r - (\alpha_{12}/\beta_{12}) - \sigma_1^2 > 0$ *and* $e_{12}\alpha_{12} - d_2 - \sigma_2^2 > 0$, *then model (106) is stochastically permanent*

(iii) If $a_2 > 0$, $r - (\alpha_{12}/\beta_{12}) - e_{12}\alpha_{12}\beta_{12} - \sigma_1^2 > 0$, *and* $e_{12}\alpha_{12} - d_2 - \sigma_2^2 > 0$, *then for any* $(x_{10}, x_{20}) \in \mathbb{R}_+^2$, *model (106) has a stationary distribution and the solutions have ergodic property*

If we do not consider the intraspecific competition of the predator, i.e., $a_2 = 0$ in model (106), then model (6) is available. From Theorems 4, 7, and 8, for model (6), we have the following result.

Corollary 3

(i) If $r - (\sigma_1^2/2) < 0$ *and* $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) < 0$, *then model* (6) *is extinct exponentially with probability one*

(ii) Model (6) *is stochastically ultimate bounded*

(iii) If $r - (\alpha_{12}/\beta_{12}) - \sigma_1^2 > 0$ *and* $e_{12}\alpha_{12} - d_2 - \sigma_2^2 > 0$, *then model* (6) *is stochastically permanent*

Remark 2. From Theorem 4.11 in [18], it follows that if $r - (\alpha_{12}/\beta_{12}) - (3/2)\sigma_1^2 > 0$ and $e_{12}\alpha_{12} - d_2 - (3/2)\sigma_3^2 > 0$, then model (6) is stochastically permanent. Obviously, if conditions of Theorem 4.11 in [18] hold, then conditions in Corollary 3 hold. On the contrary, it is not set up. Therefore, Corollary 3 generalizes and improves Theorem 4.11 in [18].

Remark 3. If $r - (\alpha_{12}/\beta_{12}) - \sigma_1^2 > 0$ and $e_{12}\alpha_{12} - d_2 - \sigma_3^2 > 0$, then by Theorem 3.3 in [16], model (6) is persistent in mean, but by Corollary 3, model (6) is stochastically permanent.

7.2. Three-Species Food-Chain Model. Let $\alpha_{13} = 0$. Then, we can get the following stochastic three-species food chain model:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1x_1(t) + \dfrac{\alpha_{12}x_2(t)}{x_1(t) + \beta_{12}x_2(t)}\right]\mathrm{d}t \\ \quad + \sigma_1x_1(t)dw_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 - a_2x_2(t) + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t) + \beta_{12}x_2(t)}\right. \\ \quad \left. - \dfrac{\alpha_{23}x_3(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t + \sigma_2x_2(t)\mathrm{d}w_2(t), \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 - a_3x_3(t) + \dfrac{e_{23}\alpha_{23}x_2(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t \\ \quad + \sigma_3x_3(t)\mathrm{d}w_3(t), \end{cases} \tag{109}$$

with initial value $(x_1(0), x_2(0), x_3(0)) = (x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$. Denote

$$\begin{aligned} \kappa_1' &= r - \frac{\alpha_{12}}{\beta_{12}}; \\ \kappa_2' &= e_{12}\alpha_{12} - d_2 - \frac{\alpha_{23}}{\beta_{23}}; \\ \kappa_3' &= e_{23}\alpha_{23} - d_3; \\ \gamma_i' &= \kappa_i' - \sigma_i^2, \\ & \qquad i = 1, 2, 3. \end{aligned} \tag{110}$$

For model (109), we have the following results.

Corollary 4. *For any* $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$, *let* $(x_1(t), x_2(t), x_3(t))$ *be the solution of model* (109) *with initial value* (x_{10}, x_{20}, x_{30}).

(i) If $a_2 > 0$, $a_3 > 0$, *and* $\kappa_I' - (\sigma_i^2/2) > 0$ $(i = 1, 2, 3)$, *then*

$$\begin{aligned} \lim_{t\longrightarrow\infty} \frac{\ln x_1(t)}{t} &= 0, \\ \lim_{t\longrightarrow\infty} \frac{\ln x_2(t)}{t} &= 0, \\ \lim_{t\longrightarrow\infty} \frac{\ln x_3(t)}{t} &= 0 \ \text{a.s.} \end{aligned} \tag{111}$$

(ii) If $a_2 > 0$, $a_3 > 0$, *and* $\kappa_i' - (\sigma_i^2/2) > 0$ $(i = 1, 2, 3)$, *then*

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t x_1(s)\mathrm{d}s \geq \frac{\kappa_1' - (\sigma_1^2/2)}{a_1} \text{ a.s.},$$

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_2x_2(s) + \frac{e_{12}\alpha_{12}\beta_{12}x_2(s)}{x_1(s)}\right]\mathrm{d}s \geq \kappa_2' - \frac{\sigma_2^2}{2} \text{ a.s.},$$

$$\liminf_{t\longrightarrow\infty}\frac{1}{t}\int_0^t\left[a_3x_3(s) + \frac{e_{23}\alpha_{23}\beta_{23}x_3(s)}{x_2(s)}\right]\mathrm{d}s \geq \kappa_3' - \frac{\sigma_3^2}{2} \text{ a.s.} \tag{112}$$

(iii) If $r - (\sigma_1^2/2) < 0$, $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) < 0$, *and* $e_{23}\alpha_{23} - d_3 - (\sigma_3^2/2) < 0$, *then model (109) is extinct exponentially with probability one.*

Corollary 5. *Model (109) is stochastically ultimate bounded. Furthermore, if* $\gamma_i' > 0$ $(i = 1, 2, 3)$, *then model (109) is stochastically permanent.*

Corollary 6. *If* $a_2 > 0$, $a_3 > 0$, $\kappa_1' - e_{12}\alpha_{12}\beta_{12} - \sigma_1^2 > 0$, $\kappa_2' - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2 > 0$, *and* $\kappa_3' - \sigma_3^2 > 0$, *then for any* $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$, *model (109) has a stationary distribution and the solutions have ergodic property.*

If we do not consider the intraspecific competition of the predator, i.e., $a_2 = a_3 = 0$ in model (109), then we obtain the following stochastic three-species food chain model:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_2x_1(t) - \dfrac{\alpha_{12}x_2(t)}{x_1(t) + \beta_{12}x_2(t)}\right]\mathrm{d}t + \sigma_1x_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t) + \beta_{12}x_2(t)} - \dfrac{\alpha_{23}x_3(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t + \sigma_2x_2(t)\mathrm{d}w_2(t), \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 + \dfrac{e_{23}\alpha_{23}x_2(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t + \sigma_3x_3(t)\mathrm{d}w_3(t), \end{cases} \tag{113}$$

with initial value $(x_1(0), x_2(0), x_3(0)) = (x_{10}, x_{20}, x_{30}) \in \mathbb{R}_+^3$. From Theorems 4, 7, and 8, for model (113), we have the following result.

Corollary 7

(i) If $r - (\sigma_1^2/2) < 0$, $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) < 0$, *and* $e_{23}\alpha_{23} - d_3 - (\sigma_3^2/2) < 0$, *then model (113) is extinct exponentially with probability one*

(ii) Model (113) *is stochastically ultimate bounded*

(iii) If $\gamma_i' > 0$ $(i = 1, 2, 3)$, *then model* (113) *is stochastically permanent*

7.3. Food-Web Model without Intraspecific Competition of Predators. If we do not consider the intraspecific competition of the predator, i.e., $a_2 = a_3 = 0$ in model (8), then we obtain the following stochastic three-species model:

$$\begin{cases} \mathrm{d}x_1(t) = x_1(t)\left[r - a_1x_1(t) - \dfrac{\alpha_{12}x_2(t)}{x_1(t) + \beta_{12}x_2(t)} - \dfrac{\alpha_{13}x_3(t)}{x_1(t) + \beta_{13}x_3(t)}\right]\mathrm{d}t + \sigma_1x_1(t)\mathrm{d}w_1(t), \\ \mathrm{d}x_2(t) = x_2(t)\left[-d_2 + \dfrac{e_{12}\alpha_{12}x_1(t)}{x_1(t) + \beta_{12}x_2(t)} - \dfrac{\alpha_{23}x_3(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t + \sigma_2x_2(t)\mathrm{d}w_2(t), \\ \mathrm{d}x_3(t) = x_3(t)\left[-d_3 + \dfrac{e_{13}\alpha_{13}x_1(t)}{x_1(t) + \beta_{13}x_3(t)} - \dfrac{e_{23}\alpha_{23}x_2(t)}{x_2(t) + \beta_{23}x_3(t)}\right]\mathrm{d}t + \sigma_3x_3(t)\mathrm{d}w_3(t), \end{cases} \tag{114}$$

with initial value $(x_{10}, x_{20}, x_{30}) \in \mathbb{R}^3_+$. For model (114), we have the following result.

Corollary 8

(i) If $r - (\sigma_1^2/2) < 0$, $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) < 0$, *and* $e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3 - (\sigma_3^2/2) < 0$, *then model (114) is extinct exponentially with probability one*

(ii) Model (114) is stochastically ultimate bounded

(iii) If $\gamma_i > 0$ $(i = 1, 2, 3)$, *then model (114) is stochastically permanent*

8. Numerical Simulations

In this section, we use the Milstein method (see [35]) to substantiate our main results. The numerical simulations of population dynamics are carried out for the academic tests with the arbitrary values of the vital rates and other parameters which do not correspond to some specific biological populations and exhibit only the theoretical properties of numerical solutions of the considered model. To illustrate the theoretical results, we take the parameter values as following with different noise intensities:

$$\begin{cases} r = 0.62, \\ d_2 = 0.0005, \\ d_3 = 0.0006, \\ a_1 = 0.0005, \\ a_2 = 0.0002, \\ a_3 = 0.0003, \\ \alpha_{12} = 0.2, \\ \alpha_{13} = 0.15, \\ \alpha_{23} = 0.08, \\ e_{12} = 0.8, \\ e_{13} = 0.5, \\ e_{23} = 0.6, \\ \beta_{12} = \beta_{13} = \beta_{23} = 1, \\ x_{10} = 800, \\ x_{20} = 300, \\ x_{30} = 200. \end{cases} \tag{115}$$

In Figure 1, we choose $\sigma_i = 0$ $(i = 1, 2, 3)$ and get the solutions of the corresponding deterministic model.

(i) Assume that $\sigma_1^2 = 1.4$, $\sigma_2^2 = 0.4$, and $\sigma_3^2 = 0.3$. By a simple computation, $r - (\sigma_1^2/2) = -0.08 < 0$, $e_{12}\alpha_{12} - d_2 - (\sigma_2^2/2) = -0.0405 < 0$, and $e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3 - (\sigma_3^2/2) = -0.0276 < 0$. Thus, the condition of Theorem 4 holds. From Theorem 4, model (8) will become extinct with probability one. As can be seen from Figure 2, all the population becomes extinct.

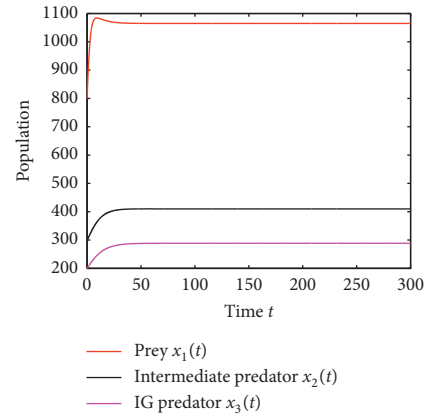


FIGURE 1: The trajectories of model (8) with $\sigma_1 = \sigma_2 = \sigma_3 = 0$.

(ii) Assume that $\sigma_1^2 = 0.44$, $\sigma_2^2 = 0.08$, and $\sigma_3^2 = 0.08$. Then, $\kappa_1 = r - (\alpha_{12}/\beta_{12}) - (\alpha_{13}/\beta_{13}) = 0.27$, $\kappa_2 = e_{12}\alpha_{12} - d_2 - (\alpha_{23}/\beta_{23}) = 0.0795$, and $\kappa_3 = e_{13}\alpha_{13} + e_{23}\alpha_{23} - d_3 = 0.1224$. Thus, $\kappa_1 - (\sigma_1^2/2) = 0.05 > 0$, $\kappa_2 - (\sigma_2^2/2) = 0.0395 > 0$, and $\kappa_3 - (\sigma_3^2/2) = 0.0824 > 0$. That is, the conditions of Theorem 3 hold. In view of Theorem 3, model (8) is persistent in mean. As can be seen from Figure 3, all the populations are permanent in mean. This is consistent to Theorem 3.

(iii) Assume that $\sigma_1^2 = 0.02$, $\sigma_2^2 = 0.01$, $\sigma_3^2 = 0.01$. From (ii), it follows that $\kappa_1 = 0.27$, $\kappa_2 = 0.0795$ and $\kappa_3 = 0.1224$. Thus, $\kappa_1 - \sigma_1^2 = 0.25 > 0$, $\kappa_2 - \sigma_2^2 = 0.0695 > 0$ and $\kappa_3 - \sigma_3^2 = 0.1124 > 0$. Hence, the conditions of Theorem 8 hold. In view of Theorem 8, model (8) is stochastically permanent. From Figure 4 that all the populations are stochastically permanent. This is consistent to Theorem 8.

(iv) Assume that $\sigma_1^2 = 0.02$, $\sigma_2^2 = 0.01$, and $\sigma_3^2 = 0.01$. From (ii), it follows that $\kappa_1 = 0.27$, $\kappa_2 = 0.0795$, and $\kappa_3 = 0.1224$. Furthermore, $\kappa_1 - e_{12}\alpha_{12}\beta_{12} - e_{13}\alpha_{13}\beta_{13} - \sigma_1^2 = 0.015 > 0$, $\kappa_2 - e_{23}\alpha_{23}\beta_{23} - \sigma_2^2 = 0.0215 > 0$, and $\kappa_3 - \sigma_3^2 = 0.1244 > 0$. Thus, the conditions of Theorem 9 hold. Therefore, model (8) has a stationary distribution according to Theorem 9 (see Figures 5 and 6).

9. Conclusions and Discussions

This paper is concerned with a stochastic three-species predator-prey food web model with omnivory and ratio-dependent functional response. First, by the comparison theorem of stochastic differential equations, we prove the existence and uniqueness of global positive solution of the model. Next, we investigate an important asymptotic property of the solution, which is crucial to the study of the dynamic behavior of the model. Then, under some conditions, we conclude that the model is persistent in mean and extinct. Moreover, we discuss the stochastic persistence of the model. Furthermore, by constructing a suitable Lyapunov function, we establish sufficient conditions for the

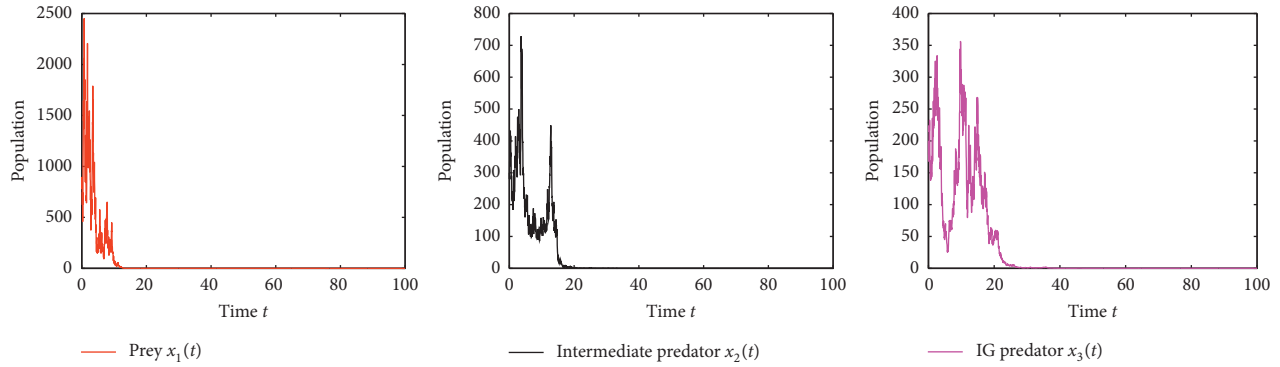


Figure 2: The trajectories of stochastic model (8) with $\sigma_1^2 = 1.4$, $\sigma_2^2 = 0.4$, and $\sigma_3^2 = 0.3$.

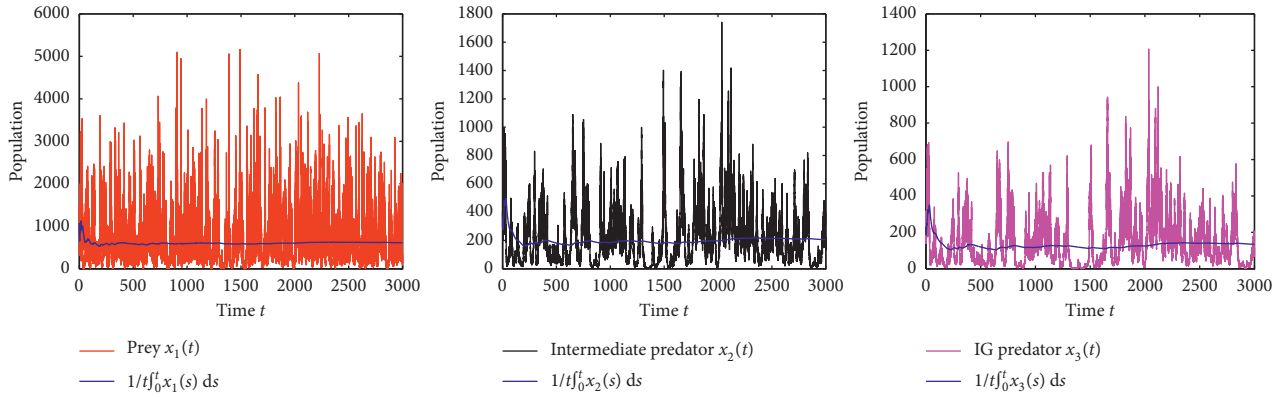


Figure 3: The trajectories of stochastic model (8) with $\sigma_1 = 0.44$, $\sigma_2 = 0.08$, and $\sigma_3 = 0.08$.

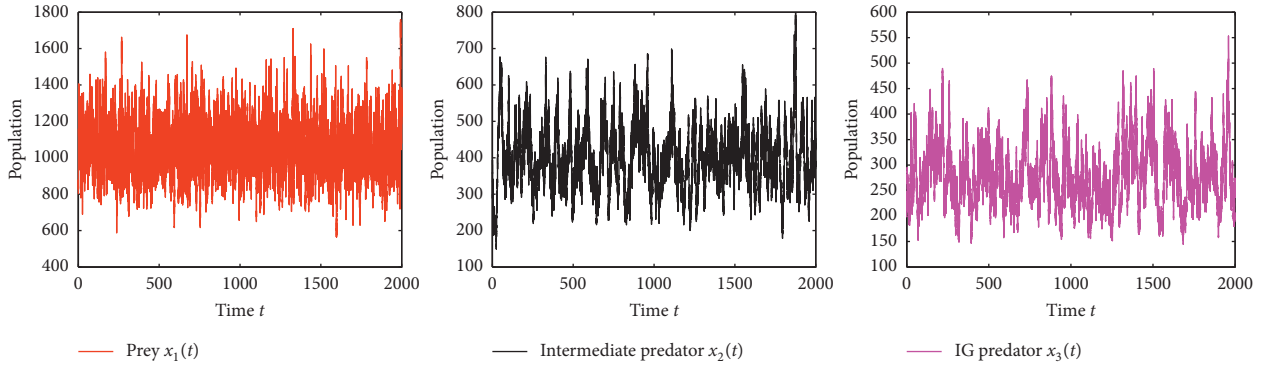


Figure 4: The trajectories of stochastic model (8) with $\sigma_1 = 0.02$, $\sigma_2 = 0.01$, and $\sigma_3 = 0.01$.

existence of an ergodic stationary distribution to the model. Then, we present the application of the main results in some special models. Finally, some numerical simulations are introduced to support the main results.

In Section 4, we prove that there are two typical phenomena arising in accordance with the relative values of the parameters of the model. In Theorem 3, we give the conditions on the parameters that informally can be stated by saying that the noise intensities σ_i^2 $(i = 1, 2, 3)$ are small compared to the other parameters, such that the species in model (8) are persistent in mean. From Theorem 4, it follows that in the case that the noise intensities σ_i^2 $(i = 1, 2, 3)$ are large with respect to the other parameters, then the solution of model (8) tends to extinction almost surely.

Later, in Section 5, we discuss on the stochastic permanence of the solution. This concept, which can be paraphrased by saying that the species in model (8) will survive forever, is one of the most important and interesting topics in the analysis of the model. From Theorem 8, if the

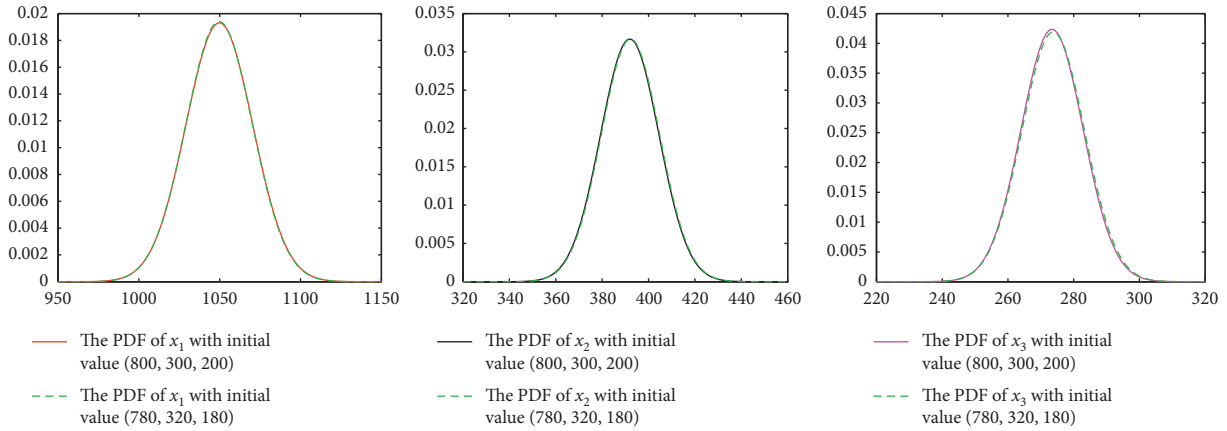


Figure 5: The density for each population at time $t = 10000$ with $\sigma_1 = 0.02$, $\sigma_2 = 0.01$, and $\sigma_3 = 0.01$.

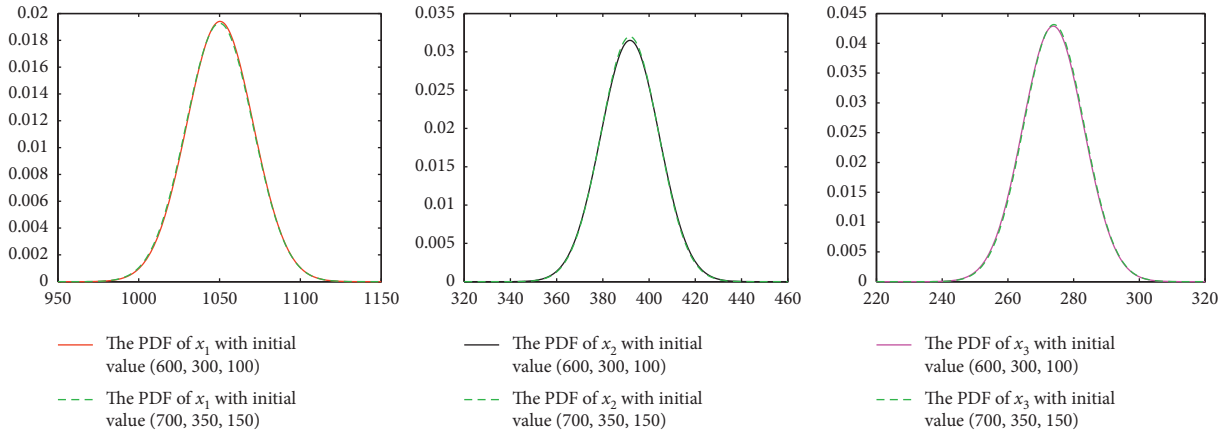


Figure 6: The density for each population at time $t = 50000$ with $\sigma_1 = 0.02$, $\sigma_2 = 0.01$, and $\sigma_3 = 0.01$.

noise intensities σ_i^2 $(i = 1, 2, 3)$ are small compared to the other parameters, such that $\kappa_i - \sigma_i^2 > 0$ $(i = 1, 2, 3)$, then model (8) is stochastically permanent.

Moreover, in Section 6, by constructing a suitable Lyapunov function, we show that there is an ergodic stationary distribution for the solution of model (8). In Theorem 9, we give the conditions on the parameters that can be stated by saying that the intensity σ_i^2 of white noise $\dot{w}_i(t)$ is sufficiently small, such that the solution model (8) has an ergodic stationary distribution.

The results in this paper generalize and improve the previous related results. From Remark 1, we know that our work can be seen as the extension of [19]. From Remark 2, we know that Theorem 8 generalizes and improves Theorem 4.11 in [18].

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All the authors contributed equally and significantly in writing this paper. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 11971279 and 11471197).

References

[1] H. I. Freedman, *Deterministic Mathematical Models in Population Ecology*, Marcel Dekker, New York, NY, USA, 1980.

[2] M. Ouyang and X. Li, "Permanence and asymptotical behavior of stochastic prey-predator system with Markovian switching," *Applied Mathematics and Computation*, vol. 266, pp. 539–559, 2015.

[3] R. Arditi and L. R. Ginzburg, "Coupling in predator-prey dynamics: ratio-Dependence," *Journal of Theoretical Biology*, vol. 139, no. 3, pp. 311–326, 1989.

[4] Y. Kuang and E. Beretta, "Global qualitative analysis of a ratio-dependent predator-prey system," *Journal of Mathematical Biology*, vol. 36, no. 4, pp. 389–406, 1998.

[5] H. I. Freedman and J. W.-H. So, "Global stability and persistence of simple food chains," *Mathematical Biosciences*, vol. 76, no. 1, pp. 69–86, 1985.

[6] S.-B. Hsu, S. Ruan, and T.-H. Yang, "Analysis of three species Lotka-Volterra food web models with omnivory," *Journal of Mathematical Analysis and Applications*, vol. 426, no. 2, pp. 659–687, 2015.

[7] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180–183, 2000.

[8] T. Namba, Y. Takeuchi, and M. Banerjee, "Stabilizing effect of intra-specific competition on prey-predator dynamics with intraguild predation," *Mathematical Modelling of Natural Phenomena*, vol. 13, no. 3, Aritcle Number 29, 41 pages, 2018.

[9] D. Sen, S. Ghorai, and M. Banerjee, "Complex dynamics of a three species prey-predator model with intraguild predation," *Ecological Complexity*, vol. 34, pp. 9–22, 2018.

[10] M. Liu, K. Wang, and Q. Wu, "Survival analysis of stochastic competitive models in a polluted environment and stochastic competitive exclusion principle," *Bulletin of Mathematical Biology*, vol. 73, no. 9, pp. 1969–2012, 2011.

[11] X. Mao, *Stochsatic Differential Equations and Applications*, Horwood, Chichester, UK, 2007.

[12] M. Liu and C. Bai, "Optimal harvesting of a stochastic Logistic model with time delay," *Journal of Nonlinear Science*, vol. 25, no. 2, pp. 277–289, 2015.

[13] C. Ji, D. Jiang, and N. Shi, "Analysis of a predator-prey model with modified Leslie-Gower and Holling-type II schemes with stochastic perturbation," *Journal of Mathematical Analysis and Applications*, vol. 359, no. 2, pp. 482–498, 2009.

[14] M. Jovanović and M. Krstić, "Extinction in stochastic predator-prey population model with Allee effect on prey," *Discrete and Continuous Dynamical Systems Series B*, vol. 22, no. 7, pp. 2651–2667, 2017.

[15] Q. Liu and D. Jiang, "Periodic solution and stationary distribution of stochastic predator-prey models with higher-order perturbation," *Journal of Nonlinear Science*, vol. 28, no. 2, pp. 423–442, 2018.

[16] C. Ji, D. Jiang, and X. Li, "Qualitative analysis of a stochastic ratio-dependent predator-prey system," *Journal of Computational and Applied Mathematics*, vol. 235, no. 5, pp. 1326–1341, 2011.

[17] P. K. Tapaswi and A. Mukhopadhyay, "Effects of environmental fluctuation on plankton allelopathy," *Journal of Mathematical Biology*, vol. 39, no. 1, pp. 39–58, 1999.

[18] Z. Wu, H. Huang, and L. L. Wang, "Dynamical behavior of a stochastic ratio-dependent predator-prey system," *Journal of Applied Mathematics*, vol. 2012, Article ID 857134, 17 pages, 2012.

[19] J. Lv, K. Wang, and D. Chen, "Analysis on a stochastic two-species ratio-dependent predator-prey model," *Methodology and Computing in Applied Probability*, vol. 17, no. 2, pp. 403–418, 2015.

[20] T. H. L. Nguyen and V. T. Ta, "Dynamics of a stochastic ratio-dependent predator-prey model," *Analysis & Applications*, vol. 9, no. 3, pp. 329–344, 2011.

[21] J. Geng, M. Liu, and Y. Zhang, "Stability of a stochastic one-predator-two-prey population model with time delays," *Communications in Nonlinear Science and Numerical Simulation*, vol. 53, pp. 65–82, 2017.

[22] M. Liu, C. Bai, and Y. Jin, "Population dynamical behavior of a two-predator one-prey stochastic model with time delay," *Discrete & Continuous Dynamical Systems—A*, vol. 37, no. 5, pp. 2513–2538, 2017.

[23] H. Li, F. Cong, D. Jiang, and H. Hua, "Persistence and non-persistence of a food chain model with stochastic perturbation," *Abstract and Applied Analysis*, vol. 2013, Article ID 125089, 9 pages, 2013.

[24] M. Liu and C. Bai, "Analysis of a stochastic tri-trophic food-chain model with harvesting," *Journal of Mathematical Biology*, vol. 73, no. 3, pp. 597–625, 2016.

[25] H. Qiu and W. Deng, "Stationary distribution and global asymptotic stability of a three-species stochastic food-chain system," *Turkish Journal of Mathematics*, vol. 41, pp. 1292–1307, 2017.

[26] R. Liu and G. Liu, "Analysis on stochastic food-web model with intraguild predation and mixed functional responses," *Physica A: Statistical Mechanics and Its Applications*, vol. 531, Article ID 121782, 2019.

[27] J. Bao, X. Mao, G. Yin, and C. Yuan, "Competitive Lotka-Volterra population dynamics with jumps," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 74, no. 17, pp. 6601–6616, 2011.

[28] S. Peng and X. Zhu, "Necessary and sufficient condition for comparison theorem of 1-dimensional stochastic differential equations," *Stochastic Processes and their Applications*, vol. 116, no. 3, pp. 370–380, 2006.

[29] X. Li, A. Gray, D. Jiang, and X. Mao, "Sufficient and necessary conditions of stochastic permanence and extinction for stochastic logistic populations under regime switching," *Journal of Mathematical Analysis and Applications*, vol. 376, no. 1, pp. 11–28, 2011.

[30] M. Liu and M. Fan, "Permanence of stochastic Lotka-Volterra systems," *Journal of Nonlinear Science*, vol. 27, no. 2, pp. 425–452, 2017.

[31] R. Liu and G. Liu, "Asymptotic behavior of a stochastict two-species competition model under the effect of disease," *Complexity*, vol. 2018, no. 15, Article ID 3127404, 2018.

[32] T. Caraballo, M. El Fatini, R. Pettersson, and R. Taki, "A stochastic SIRI epidemic model with relapse and media coverage," *Discrete & Continuous Dynamical Systems—B*, vol. 23, no. 8, pp. 3483–3501, 2018.

[33] Q. Liu, D. Jiang, T. Hayat, and B. Ahmad, "Stationary distribution and extinction of a stochastic predator-prey model with additional food and nonlinear perturbation," *Applied Mathematics and Computation*, vol. 320, pp. 226–239, 2018.

[34] Q. Liu, D. Jiang, T. Hayat, and B. Ahmad, "Stationary distribution and extinction of a stochastic SIRI epidemic model with relapse," *Stochastic Analysis and Applications*, vol. 36, no. 1, pp. 138–151, 2018.

[35] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Review*, vol. 43, no. 3, pp. 525–546, 2001.

Hindawi
Complexity
Volume 2019, Article ID 4134105, 8 pages
https://doi.org/10.1155/2019/4134105

Research Article

Bifurcation Analysis of Three-Strategy Imitative Dynamics with Mutations

Wenjun Hu,[1,2] Haiyan Tian,[1] and Gang Zhang[1]

[1]*College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China*
[2]*Department of Mathematics, Luliang University, Lishi 033000, China*

Correspondence should be addressed to Gang Zhang; gangzhang@hebtu.edu.cn

Received 11 April 2019; Accepted 13 September 2019; Published 30 October 2019

Academic Editor: Constantin Udriste

Evolutionary game dynamics is an important research, which is widely used in many fields such as social networks, biological systems, and cooperative behaviors. This paper focuses on the Hopf bifurcation in imitative dynamics of three strategies (Rock-Paper-Scissors) with mutations. First, we verify that there is a Hopf bifurcation in the imitative dynamics with no mutation. Then, we find that there is a critical value of mutation such that the system tends to an unstable limit cycle created in a subcritical Hopf bifurcation. Moreover, the Hopf bifurcation exists for other kinds of the considered mutation patterns. Finally, the theoretical results are verified by numerical simulations through Rock-Paper-Scissors game.

1. Introduction

Evolutionary game dynamics combines game theory and nonlinear dynamics to describe the evolution of the frequencies of strategies in one or more large population [1, 2]. It has edged into many fields such as networks population [3–6], economics [7, 8], biology [9, 10], management [11, 12], and cooperative behaviors [13–16]. There are many important evolutionary game dynamics such as replicator dynamics, imitative dynamics [17], best-response dynamics [18], and so on [19]. The Rock-Paper-Scissors (RPS) [20, 21] is a famous three-strategy game, which describes interactions among three competing species in ecology, sociological systems [22], and theoretical biology [23, 24]. Replicator dynamics is the best-known evolutionary dynamics, which was firstly defined by Taylor and Jonker [25], and has been researched in various fields [26, 27]. In practice, imitative dynamics is a generalized replicator dynamics, which investigates the spreading of strategies in the context of imitation instead of inheritance.

There are some research studies about the imitative dynamics [28–30]. Cheung [28] studied the imitative dynamics for games with continuous strategy space and obtained global convergence and local stability results for imitative dynamics. Wang et al. [29] investigated the imitation dynamics with delay, and they discussed the two-phenotype and three-phenotype model and obtained some relevant results for stability. Hu et al. [30] researched the imitative dynamics with discrete delay, and they discovered that the stability would be changed in the discrete delay dynamics and obtained some sufficient conditions. The emphasis of the literatures is the effect of delay in the imitative dynamics. However, the mutation is also a noticeable factor on the study of the stability of the evolutionary dynamics in reality.

Until now, many researchers have studied the effect of mutations in the replicator dynamics [31–34]. Mobilia [31] investigated the oscillatory dynamics in generic RPS games with mutations and found out the existence of the heteroclinic cycles in the RPS model. Nagatani et al. [32] studied a metapopulation model for RPS game with mutation, and they found that the mutation would lead to the phase transitions among three strategies. Toupo et al. [33, 34] researched the effect of mutations in the repeated prisoner's dilemma game and RPS game, and they found that the mutations would result in the Hopf bifurcation in the replicator dynamics. Their research studies illustrate that

mutation could change the stability of the dynamics, especially lead to bifurcation.

The bifurcation is an important behavior in dynamical systems [35], which has been researched by many scholars [36–39]. Wesson et al. [36, 37] investigated the Hopf bifurcation in two-strategy and three-strategy delayed replicator dynamics, and they demonstrated the existence of Hopf bifurcation and presented an analysis of the limit cycles through Lindstedt's method. Nesrine et al. [38] researched the Hopf bifurcation in RPS game with distributed delays. Umezuki [39] studied the bifurcation of RPS game with discrete-time logit dynamics and showed that some bifurcations would destroy the coexistence of the attractors in the RPS game.

According to the previous literatures, there are few research studies about the bifurcation in imitative dynamics. In this paper, we aim to discuss the Hopf bifurcation in imitative dynamics with mutation. Our research will illustrate that (i) the imitative dynamics appears to be a Hopf bifurcation at the parameter γ in the RPS game; (ii) the stability would be changed in the mutative imitation dynamics; and (iii) a subcritical Hopf bifurcation would be exhibited in this dynamics.

The rest of this paper is organized as follows. Section 2 sets the imitation dynamics model without mutation and analyses the stability and bifurcation. Section 3 researches the Hopf bifurcations with mutations in the imitative dynamics. Section 4 gives numerical simulations of the equilibrium and an unstable periodic solution. Section 5 offers concluding remarks.

2. RPS Model without Mutation

2.1. Derivation. We consider a symmetric three-phenotype model with pure strategies Rock (R), Scissors (S), and Paper (P) and with payoff matrix:

$$\begin{array}{c} \\ R \\ S \\ P \end{array} \begin{array}{c} \begin{array}{ccc} R & S & P \end{array} \\ \begin{pmatrix} 1 & 1+\gamma & 0 \\ 0 & 1 & 1+\gamma \\ 1+\gamma & 0 & 1 \end{pmatrix} \end{array}, \quad \gamma > 0. \tag{1}$$

The payoff matrix means that each strategy gets a payoff 1 when playing against itself, and the loser gets a payoff 0 while the winner gets $1+\gamma$. Let (x_1, x_2, x_3) denote the frequency of (R, S, P) and (f_1, f_2, f_3) the expected payoff of (R, S, P) with $f_i(x) = \sum_{j=1}^{3} x_j a_{ij}$, where $x = (x_1, x_2, x_3)$ with $\sum_{i=1}^{3} x_i = 1$ and a_{ij} denotes the payoff of S_i-individual plays against a S_j individual in which $i, j = 1, 2, 3$.

The classic imitation dynamics tacitly supposes that an individual is randomly selected from the population and awarded the same opportunity to change the strategy. That is, when an individual using S_i plays against an individual using S_j, the imitation rate that the S_j strategist switches to S_i is denoted by F_{ij} for $i, j = 1, 2, 3$. In the previous literature, it is assumed that the imitation rate F_{ij} depends on the expected payoffs $f_i(x)$ and $f_j(x)$:

$$F_{ij}(x) = F\big(f_i(x), f_j(x)\big), \quad (i, j = 1, 2, 3), \tag{2}$$

where the function $F(u, v)$ defines the imitation rule. Here, we take $F(u, v) = u/(u+v)$, i.e.,

$$F_{ij}(x(t)) = \frac{f_i(x(t))}{f_i(x(t)) + f_j(x(t))}, \quad (i, j = 1, 2, 3). \tag{3}$$

For ease of notation, write $(x_1, x_2, x_3) = (x, y, z)$. Under this condition, the imitation dynamics equations can be written as follows:

$$\begin{cases} \dot{x} = x\left(\dfrac{f_1 - f_2}{f_1 + f_2} y + \dfrac{f_1 - f_3}{f_1 + f_3} z\right), \\ \dot{y} = y\left(\dfrac{f_2 - f_1}{f_2 + f_1} x + \dfrac{f_2 - f_3}{f_2 + f_3} z\right), \\ \dot{z} = z\left(\dfrac{f_3 - f_1}{f_3 + f_1} x + \dfrac{f_3 - f_2}{f_3 + f_2} y\right). \end{cases} \tag{4}$$

Since x, y, and z are the frequencies of the three strategies, the region of interest is the three-dimensional simplex in R^3:

$$\sum \equiv \left\{(x, y, z) \in R^3 : x + y + z = 1, \ (x, y, z \geq 0)\right\}. \tag{5}$$

So, we can eliminate z using $z = 1 - x - y$ and the projection of $\sum$ into the $x - y$ plane: $S \equiv \{(x, y) \in R^2 : (x, y, 1 - x - y) \in \sum\}$. In this case, equation (4) can be written as

$$\begin{aligned} \dot{x} &= x\left[y \frac{(2+\gamma)x + (1+2\gamma)y - (1+\gamma)}{-\gamma x + (1+\gamma)y + (1+\gamma)} \right. \\ &\quad \left. + (1 - x - y) \frac{(1-\gamma)x + (2+\gamma)y - 1}{(1+\gamma)x + \gamma y + 1} \right], \\ \dot{y} &= y\left[-x \frac{(2+\gamma)x + (1+2\gamma)y - (1+\gamma)}{-\gamma x + (1+\gamma)y + (1+\gamma)} \right. \\ &\quad \left. + (1 - x - y) \frac{-x - (1+\gamma)y + 2 + \gamma}{-(1+\gamma)x - \gamma y + \gamma} \right]. \end{aligned} \tag{6}$$

2.2. Stability of Equilibria. System (6) has four equilibria:

$$\begin{aligned} e_1 &= (0, 0), \\ e_2 &= (0, 1), \\ e_3 &= (1, 0), \\ x^* &= \left(\frac{1}{3}, \frac{1}{3}\right). \end{aligned} \tag{7}$$

In order to discuss the stability of these equilibria, we linearize equation (6). As a result, we can analyze the stability of each point through the eigenvalues of the Jacobian. The eigenvalues of the three corner equilibria can be calculated as shown in Table 1.

From above analysis, in the nonmutation RPS equation, each corner of S is a saddle point.

Table 1: The eigenvalues of equilibria.

Points	Eigenvalues
$(1,0)$	$\lambda_1 = -1, \lambda_2 = \gamma/\gamma + 2$
$(0,1)$	$\lambda_2 = \gamma/\gamma + 2, \lambda_1 = -1$
$(0,0)$	$\lambda_1 = -1, \lambda_2 = \gamma/\gamma + 2$
$(1/3, 1/3)$	$\lambda_{1,2} = 1 - \gamma/4(\gamma+2) \pm (\sqrt{3}(1+\gamma)/4(\gamma+2))i$

Next, we consider an important equilibrium x^*; first, we discuss the stability in the nonmutation system. Since there are two imaginary eigenvalues at this equilibrium point, we think there might be a Hopf bifurcation at x^*.

2.3. Hopf Bifurcation. First, we introduce a lemma about Hopf bifurcation in the vector field.

Lemma 1 (see [40]). *Suppose that system*

$$\dot{x} = f_\mu(x), \quad x \in \mathbb{R}^n, \mu \in \mathbb{R}, \tag{8}$$

has an equilibrium (x_0, μ_0) at which the following properties are satisfied:

(H1) $D_x f_{\mu_0}(x_0)$ has a simple pair of pure imaginary eigenvalues and no other eigenvalues with zero real parts

(H2) $d = d/d\mu\,(\mathrm{Re}\lambda(\mu))|_{\mu=\mu_0} \neq 0$

(H3) $a = a(\mu_0) \neq 0$, where $a(\mu_0)$ is the first Lyapunov coefficient

Then, the system undergoes a Hopf bifurcation at $\mu = \mu_0$.

The coefficient $a(\mu)$ can be calculated as follows. On the center manifold, $f_\mu(x)$ has the following form near the origin:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \mathrm{Re}\lambda(\mu) & -\mathrm{Im}\lambda(\mu) \\ \mathrm{Im}\lambda(\mu) & \mathrm{Re}\lambda(\mu) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} f^1(x,y,\mu) \\ f^2(x,y,\mu) \end{pmatrix}, \tag{9}$$

where f^1 and f^2 are nonlinear functions in x and y and $\lambda(\mu)$ and $\overline{\lambda}(\mu)$ are the eigenvalues of the linearized system around the equilibrium at the origin. Especially, at the bifurcation point (i.e., $\mu = 0, \lambda_{1,2} = \pm i\omega$), the coefficient is given by

$$\begin{aligned} a(\mu) = {} & \frac{f^1_{xxx} + f^1_{xyy} + f^2_{xxy} + f^2_{yyy}}{16} \\ & + \frac{f^1_{xy}\left(f^1_{xx} + f^1_{yy}\right) - f^2_{xy}\left(f^2_{xx} + f^2_{yy}\right)}{16\omega} \\ & + \frac{f^1_{yy} f^2_{yy} - f^1_{xx} f^2_{xx}}{16\omega}. \end{aligned} \tag{10}$$

Lemma 2 (see [41]). *Consider the system form (8); for sufficiently small μ, the following four cases hold:*

(i) $d > 0, a > 0$: unstable equilibrium for $\mu > 0$ and asymptotically stable equilibrium for $\mu < 0$, with unstable periodic orbit (i.e., subcritical) for $\mu < 0$

(ii) $d > 0, a < 0$: unstable equilibrium for $\mu > 0$ and asymptotically stable equilibrium for $\mu < 0$, with asymptotically stable periodic orbit (i.e., supercritical) for $\mu > 0$

(iii) $d < 0, a > 0$: unstable equilibrium for $\mu < 0$ and asymptotically stable equilibrium for $\mu > 0$, with unstable periodic orbit (i.e., subcritical) for $\mu > 0$

(iv) $d < 0, a < 0$: unstable equilibrium for $\mu > 0$ and asymptotically stable equilibrium for $\mu < 0$, with asymptotically stable periodic orbit (i.e., supercritical) for $\mu < 0$

Next, we give a theorem to illustrate the bifurcation in dynamics (6).

Theorem 1. *The imitation dynamics (6) exhibits a subcritical Hopf bifurcation at $\gamma = 1$. Moreover, when $\gamma > 1$, the interior equilibrium is locally stable, and it is unstable when $\gamma < 1$.*

Proof. (i) When $\gamma \neq 1$, the sign of the real part of eigenvalues can be determined through Table 1. That is,

$$\begin{aligned} & \gamma > 1, \ \mathrm{Re}(\lambda) = \frac{1-\gamma}{4(\gamma+2)} < 0, \ \text{(stability)}, \\ & \gamma < 1, \ \mathrm{Re}(\lambda) = \frac{1-\gamma}{4(\gamma+2)} > 0, \ \text{(instability)}. \end{aligned} \tag{11}$$

(ii) When $\gamma = 1$, as the formula in Lemma 1, we obtain the nonlinear function form,

$$\begin{aligned} f^1(x,y) = {} & x\left[(1-x-y)\frac{3y-1}{2x+y+1} + y\frac{3x+3y-2}{-x+2y+2}\right] \\ & + \frac{\sqrt{3}}{6}y, \\ f^2(x,y) = {} & y\left[(1-x-y)\frac{-x-2y+3}{-2x-y+1} - x\frac{3x+3y-2}{-x+2y+2}\right] \\ & - \frac{\sqrt{3}}{6}x. \end{aligned} \tag{12}$$

We can obtain the Lyapunov coefficient through Matlab as follows:

$$\begin{aligned} a(\gamma) = {} & \frac{27(\gamma^3 + 4\gamma^2 + 2\gamma - 3)}{16(\gamma+2)^3}\Big|_{\gamma=1} = \frac{1}{4} > 0. \\ d(\gamma) = {} & \frac{d}{d(\gamma)}(\mathrm{Re}\lambda(\gamma)) = -\frac{3}{4(\gamma+2)^2}\Big|_{\gamma=1} < 0. \end{aligned} \tag{13}$$

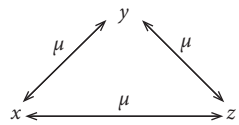


Figure 1: The global mutation in RPS.

According to Lemma 2, the Hopf bifurcation is subcritical at $\gamma = 1$. □

3. RPS Model with Mutations

In this section, we examine the imitative dynamics with all kinds of mutations, including global mutations, single mutation, double mutations, and so on.

3.1. Global Mutations in RPS Model. First, we discuss the global mutations in imitative dynamics. The relationship of mutations is shown in Figure 1.

In this case, the dynamics becomes the following form with mutant coefficient $\mu\,(\mu \geq 0)$:

$$\begin{cases} \dot{x} = x\left[(1-x-y)\dfrac{(1-\gamma)x+(2+\gamma)y-1}{(1+\gamma)x+\gamma y+1}\right. \\ \qquad \left. + y\dfrac{(2+\gamma)x+(1+2\gamma)y-(1+\gamma)}{-\gamma x+(1+\gamma)y+(1+\gamma)}\right] + \mu(1-3x), \\ \dot{y} = y\left[(1-x-y)\dfrac{-x-(1+\gamma)y+2+\gamma}{-(1+\gamma)x-\gamma y+\gamma}\right. \\ \qquad \left. - x\dfrac{(2+\gamma)x+(1+2\gamma)y-(1+\gamma)}{-\gamma x+(1+\gamma)y+(1+\gamma)}\right] + \mu(1-3y). \end{cases} \tag{14}$$

We give a theorem to illustrate the bifurcation in dynamics (14) as follows.

Theorem 2. *The following two conclusions are established for imitative dynamics (14):*

(i) There exists a subcritical Hopf bifurcation at $\mu = 1 - \gamma/12(\gamma+2)$ when $\gamma < 1$. Moreover, for $\mu > 1 - \gamma/12(\gamma+2)$, the interior equilibrium is locally stable, and for $\mu < 1 - \gamma/12(\gamma+2)$, it is unstable

(ii) The interior equilibrium is locally stable when $\gamma \geq 1$

Proof. The Jacobian matrix of dynamics (14) at $(1/3, 1/3)$ is

$$J\left(\frac{1}{3},\frac{1}{3}\right) = \begin{bmatrix} \dfrac{1-\gamma}{6(\gamma+2)} - 3\mu & \dfrac{2\gamma+1}{6(\gamma+2)} \\ -\dfrac{1}{6} & \dfrac{1-\gamma}{6(\gamma+2)} - 3\mu \end{bmatrix}, \tag{15}$$

and the conjugate complex eigenvalues are

$$\lambda_{1,2}(\mu) = \frac{1-\gamma}{4(\gamma+2)} - 3\mu \pm \frac{\sqrt{3}(\gamma+1)}{4(\gamma+2)}i. \tag{16}$$

Let $\mathrm{Re}(\lambda_{1,2}) = 0$, then $\mu = 1 - \gamma/12(\gamma+2)$. One can obtain the following:

(i) $\gamma < 1$: if $\mu = 1 - \gamma/12(\gamma+2)$, then the Jacobian matrix has a pair of pure complex eigenvalues

(ii) $\gamma \geq 1$: $\mathrm{Re}(\lambda) < 0$ is always correct, i.e., the interior equilibrium is locally stable

Similar to the proof in Theorem 1, the nonlinear functions can be obtained as follows:

$$\begin{cases} f^1(x,y) = x\left[y\dfrac{(2+\gamma)x+(1+2\gamma)y-(1+\gamma)}{-\gamma x+(1+\gamma)y+(1+\gamma)}\right. \\ \qquad \left. + (1-x-y)\dfrac{(1-\gamma)x+(2+\gamma)y-1}{(1+\gamma)x+\gamma y+1}\right] \\ \qquad + \dfrac{1-\gamma}{12(\gamma+2)}(1-3x) + \dfrac{\sqrt{3}(\gamma+1)}{4(\gamma+2)}y, \\ f^2(x,y) = y\left[x\dfrac{(1+\gamma)-(2+\gamma)x+(1+2\gamma)y}{-\gamma x+(1+\gamma)y+(1+\gamma)}\right. \\ \qquad \left. + (1-x-y)\dfrac{-x-(1+\gamma)y+2+\gamma}{-(1+\gamma)x-\gamma y+\gamma}\right] \\ \qquad + \dfrac{1-\gamma}{12(\gamma+2)}(1-3y) - \dfrac{\sqrt{3}(\gamma+1)}{4(\gamma+2)}x. \end{cases} \tag{17}$$

The Lyapunov coefficient can be calculated as follows:

$$a(\gamma) = \frac{27\left(\gamma^3+4\gamma^2+2\gamma-3\right)}{16(\gamma+2)^3} > 0, \quad (\text{for } 0 < \gamma < 1),$$

$$d = \frac{d}{d(\mu)}\left(\mathrm{Re}\lambda(\mu)\right) = \frac{d}{d(\mu)}\left(\frac{1-\gamma}{4(\gamma+2)} - 3\mu\right) = -3 < 0. \tag{18}$$

According to Lemma 2, the Hopf bifurcation is subcritical at $\mu = 1 - \gamma/12(\gamma+2)$.

The results in Theorem 1 and Theorem 2 show that the situation in dynamics (8) is different from dynamics (14). While the interior equilibrium is always unstable when $\gamma < 1$ in the former, the interior equilibrium is locally stable for $\mu > 1 - \gamma/12(\gamma+2)$ when $\gamma < 1$ in the latter. □

3.2. Other Mutations in RPS Model. In this section, we discuss the other mutations in imitative dynamics; the situation becomes complex as one adds more mutant pathways. For the ease of research, let us restrict attention to mutant forms that ensure $(x, y) = (1/3, 1/3)$ as the inner equilibrium for all values of γ and μ.

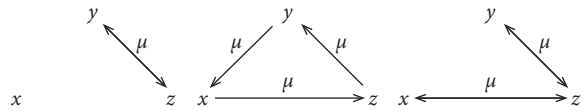


Figure 2: Three representative mutations in RPS.

Table 2: Hopf curve of different mutant forms.

Forms	Numbers	Hopf curve
$y \rightleftarrows z$	2	$\mu_c = 1 - \gamma/4(\gamma + 2)$
$x \longrightarrow y \longrightarrow z \longrightarrow x$	3	$\mu_c = 1 - \gamma/6(\gamma + 2)$
$x \leftrightarrows z, z \rightleftarrows y$	4	$\mu_c = 1 - \gamma/8(\gamma + 2)$

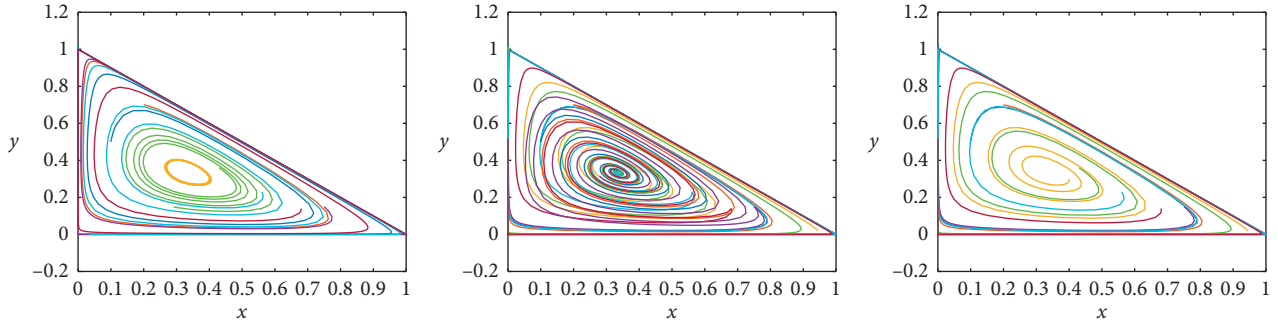


Figure 3: The interior equilibrium's changing situation with different γ. (a) When $\gamma = 1$, the limit cycle occurs around the interior equilibrium. (b) When $\gamma = 1.5$, the interior equilibrium is stable. (c) When $\gamma = 0.8$, the interior equilibrium is unstable.

Here, we discuss the following three kinds of mutations in RPS imitative dynamics (see Figure 2): (i) single mutation between two strategies; (ii) single-cycle mutations among three strategies; and (iii) double mutations between two strategies. These mutant forms are shown in Table 2. As the cycle symmetry of the RPS game, it suffices to consider one of the three possible single mutation and double mutations. In this case, we just consider the following representative mutations.

For the imitative dynamics with these three kinds of mutations, similar subcritical Hopf bifurcation at $\mu = \mu_c$ would be present. This is different from the stability in nonmutation imitative dynamics, which is always unstable when $\gamma < 1$.

4. Numerical Simulations

In this section, we propose to compare the properties of the bifurcating periodic solution. Here, we report two simulation results for imitative dynamics with nonmutation and mutation, respectively.

Example 1. In the imitative dynamics (6), we take

$$\begin{aligned} &\gamma = 1.5, \\ &\gamma = 1, \\ &\gamma = 0.8, \end{aligned} \tag{19}$$

into the equation. Through the Matlab software, one can obtain the following results (see Figure 3).

In Figure 3, the numerical simulation shows that the interior equilibrium x^* is asymptotically stable when $\gamma > 1$ (i.e., $\gamma = 1.5$). However, when $\gamma = 1$, the system state tends to a unstable periodic solution, and when $\gamma < 1$ (i.e., $\gamma = 0.8$), the interior equilibrium x^* is unstable.

Example 2. In the imitative dynamics (14), let

$$\begin{aligned} &\gamma = 1.5, \\ &\mu = 0.02, \\ &\mu = 0.002, \\ &\gamma = 0.8, \\ &\mu = 0.02, \\ &\mu = 0.002, \\ &\mu = 0.00595. \end{aligned} \tag{20}$$

Through the Matlab software, one can obtain the following results (see Figures 4 and 5).

In Figure 4, the numerical simulation shows that the interior equilibrium x^* is asymptotically stable when $\gamma = 1.5$ (i.e., $\gamma > 1$) for any value of μ, such as $\mu = 0.02$ and $\mu = 0.002$.

In Figure 5, the numerical simulation shows the following: (i) the Hopf curve in dynamics (14), i.e., the

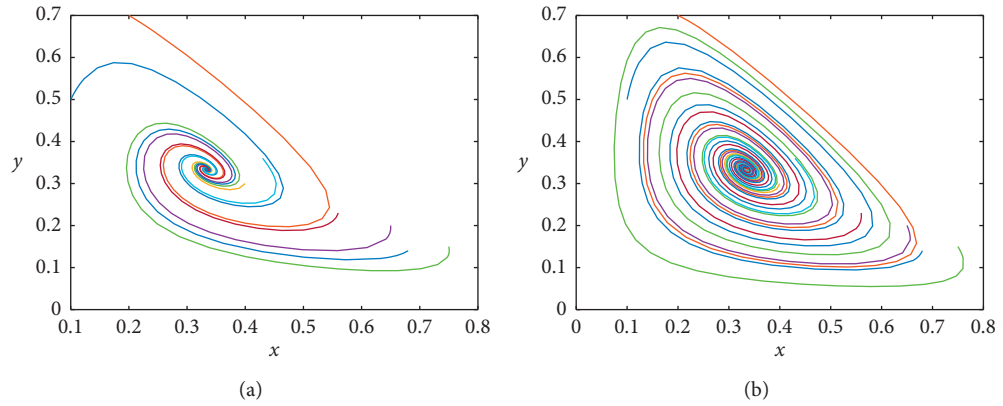


Figure 4: The interior equilibrium's changing situation with the same γ and different μ. (a) $\gamma = 1.5$ and $\mu = 0.02$. (b) $\gamma = 1.5$ and $\mu = 0.002$.

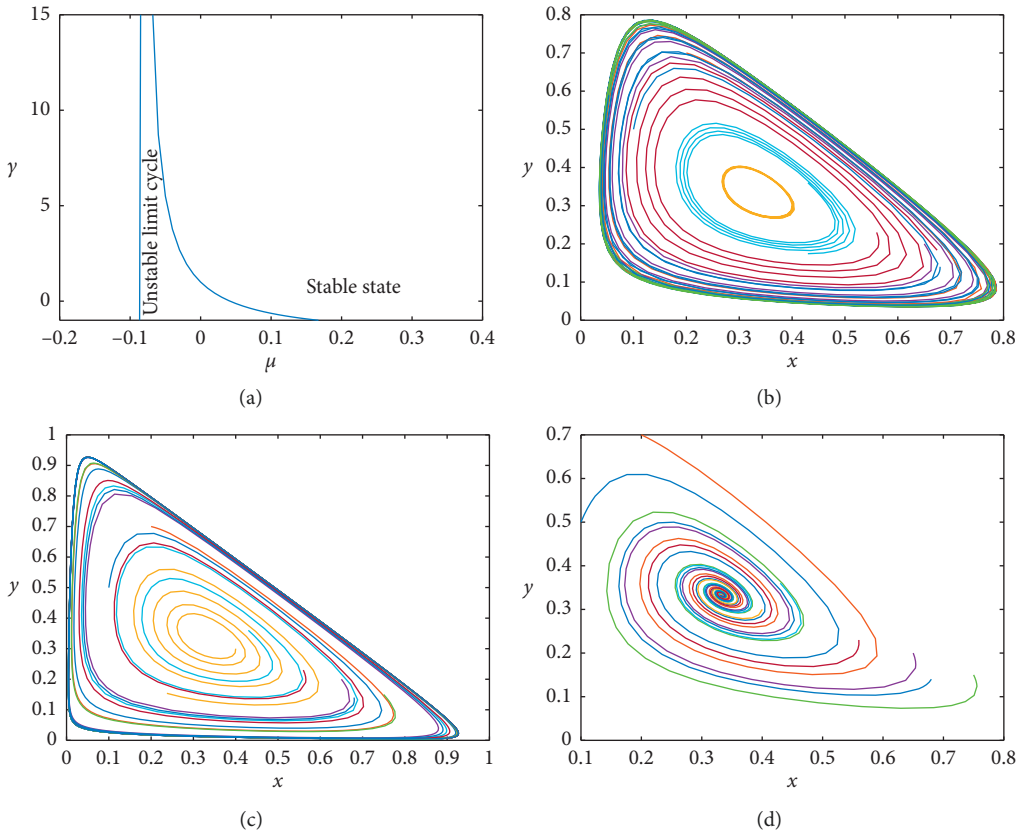


Figure 5: (a) The stability and limit cycle with different γ and μ (Hopf curve). (b) When $\gamma = 0.8$ and $\mu = 0.00595$, the limit cycle occurs around x^*. (c) When $\gamma = 0.8$ and $\mu = 0.002$, the interior equilibrium is unstable. (d) When $\gamma = 0.8$ and $\mu = 0.02$, the interior equilibrium is stable.

criticality of μ, changes with γ; (ii) the interior equilibrium x^* would be stable when $\gamma = 0.8$ (i.e., $\gamma < 1$); and (iii) when $\mu = 0.00595$, the system state tends to a unstable periodic solution, and when $\mu > 0.00595$, the interior equilibrium x^* is stable and is unstable when $\mu < 0.00595$. The marks in Figures 4 and 5 are same as the description in Figure 3.

5. Conclusion

In this paper, the stability of the interior equilibrium has been mainly investigated for imitative dynamics with mutations. Different from the result in replicator dynamics [31, 34], the stability of the interior equilibrium has been

changed with the mutations, and a subcritical Hopf bifurcation appears.

For the imitative dynamics in the RPS game, the stability is changed at the parameter γ in the payoff matrix with no mutation, and the mutation μ in global mutation. In the dynamics with no mutation, the interior equilibrium is locally stable when $\gamma > 1$ and is unstable when $\gamma < 1$, and a subcritical Hopf bifurcation appears at $\gamma = 1$ in given payoff matrix. In the imitative dynamics with global mutation, the interior equilibrium is stable when $\gamma \geq 1$, and it is different from the case when $\gamma < 1$. There is a subcritical Hopf bifurcation at $\mu = 1 - \gamma/12(\gamma + 2)$, and the interior equilibrium is locally stable when $\mu > 1 - \gamma/12(\gamma + 2)$ and is unstable when $\mu < 1 - \gamma/12(\gamma + 2)$.

If we change the number of parameters (i.e., from one to two), the stability and bifurcation would become much more complicated. Furthermore, some numerical examples have been given to illustrate the effectiveness of our results. As an extension to this work, we plan to discuss the imitation dynamics with delays and mutations.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed equally and significantly in writing this paper and read and approved the final manuscript.

Acknowledgments

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (NNSF) through grant no. 11562006, the Innovation Foundation for Graduate in Hebei Normal University (no. YJBS2019001), the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi Province (no. 2019L0940), the Science Foundation of Hebei Normal University (no. L2018B01), and the Natural Science Foundation of Shanxi Province (no. 201801D121009).

References

[1] J. M. Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, UK, 1982.

[2] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, UK, 1998.

[3] M. R. D'Orsogna and M. Perc, "Statistical physics of crime: a review," *Physics of Life Reviews*, vol. 12, pp. 1–21, 2015.

[4] D. Madeo and C. Mocenni, "Game interactions and dynamics on networked populations," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1801–1810, 2015.

[5] Z. Wang, M. A. Andrews, Z.-X. Wu, L. Wang, and C. T. Bauch, "Coupled disease–behavior dynamics on complex networks: a review," *Physics of Life Reviews*, vol. 15, pp. 1–29, 2015.

[6] J. M. Pacheco, A. Traulsen, and M. A. Nowak, "Coevolution of strategy and structure in complex networks with dynamical linking," *Physical Review Letters*, vol. 97, no. 25, Article ID 258103, 2006.

[7] H. Xu, C. Tian, X. Xiao, and S. Fan, "Evolutionary investors' power-based game on networks," *Applied Mathematics and Computation*, vol. 330, pp. 125–133, 2018.

[8] H. Xu, S. Fan, C. Tian, and X. Xiao, "Evolutionary investor sharing game on networks," *Applied Mathematics and Computation*, vol. 340, pp. 138–145, 2019.

[9] R. Liu and G. Liu, "Asymptotic behavior of a stochastic Two-Species competition model under the effect of disease," vol. 2018, Article ID 3127404, 15 pages, 2018.

[10] J. Yang and X. Wang, "Threshold dynamics of an SIR model with nonlinear incidence rate and age-dependent susceptibility," vol. 2018, Article ID 9613807, 15 pages, 2018.

[11] R. Lu, X. Wang, H. Yu, and D. Li, "Multiparty evolutionary game model in coal mine safety management and its application," *Complexity*, vol. 2018, Article ID 9620142, 10 pages, 2018.

[12] X. Wang, R. Lu, H. Yu, and D. Li, "Stability of the evolutionary game system and control strategies of behavior instability in coal mine safety management," *Complexity*, vol. 2019, Article ID 6987427, 14 pages, 2019.

[13] H.-X. Yang and X. Chen, "Promoting cooperation by punishing minority," *Applied Mathematics and Computation*, vol. 316, pp. 460–466, 2018.

[14] J. Gómez-Gardeñes, M. Campillo, L. M. Floría, and Y. Moreno, "Dynamical organization of cooperation in complex topologies," *Physical Review Letters*, vol. 98, no. 10, Article ID 108103, 2007.

[15] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, "Emergence of cooperation and evolutionary stability in finite populations," *Nature*, vol. 428, no. 6983, pp. 646–650, 2004.

[16] M. A. Nowak, "Five rules for the evolution of cooperation," *Science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[17] D. Vilone, J. J. Ramasco, Angel Sánchez, and M. S. Miguel, "Social and strategic imitation: the way to consensus," *Science Reports*, vol. 2, Article ID 686, 2012.

[18] C. H. Hommes and M. I. Ochea, "Multiple equilibria and limit cycles in evolutionary games with Logit Dynamics," *Games and Economic Behavior*, vol. 74, no. 1, pp. 434–441, 2012.

[19] R. Cressman and Y. Tao, "The replicator equation and other game dynamics," *Proceedings of the National Academy of Sciences*, vol. 111, no. S3, pp. 10810–10817, 2014.

[20] B. Sinervo and C. M. Lively, "The rock-paper-scissors game and the evolution of alternative male strategies," *Nature*, vol. 380, no. 6571, pp. 240–243, 1996.

[21] B. Xu, H. Zhou, and Z. Wang, "Cycle frequency in standard Rock-Paper-Scissors games: evidence from experimental economics," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 20, pp. 4997–5005, 2013.

[22] T. Platkowski and J. Zakrzewski, "Asymptotically stable equilibrium and limit cycles in the RPS game in a population of players with complex personalities," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 23-24, pp. 4219–4226, 2011.

[23] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," *Science*, vol. 303, no. 5659, pp. 793–799, 2004.

[24] E. Kukla and T. Platkowski, "Onset of limit cycles in population games with attractiveness driven strategy choice," *Chaos, Soliton and Fractals*, vol. 56, pp. 77–82, 2013.

[25] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, vol. 40, no. 1-2, pp. 145–156, 1978.
[26] Y. Tao and Z. Wang, "Effect of time delay and evolutionarily stable strategy," *Journal of Theoretical Biology*, vol. 187, no. 1, pp. 111–116, 1997.
[27] J. S. Weitz, C. Eksin, K. Paarporn, S. P. Brown, and W. C. Ratcliff, "An oscillating tragedy of the commons in replicator dynamics with game environment feedback," *Proceedings of the National Academy of Sciences of the United Atates of America*, vol. 113, no. 47, pp. E7518–E7525, 2016.
[28] M.-W. Cheung, "Imitative dynamics for games with continuous strategy space," *Games and Economic Behavior*, vol. 99, pp. 206–223, 2016.
[29] S.-C. Wang, J.-R. Yu, S. Kurokawa, and Y. Tao, "Imitation dynamics with time delay," *Journal of Theoretical Biology*, vol. 420, pp. 8–11, 2017.
[30] W. Hu, G. Zhang, and H. Tian, "The stability of imitation dynamics with discrete distributed delays," *Physica A: Statistical Mechanics and Its Applications*, vol. 521, pp. 218–224, 2019.
[31] M. Mobilia, "Oscillatory dynamics in rock-paper-scissors games with mutations," *Journal of Theoretical Biology*, vol. 264, no. 1, pp. 1–10, 2010.
[32] T. Nagatani, G. Ichinose, and K.-I. Tainaka, "Metapopulation model for rock-paper-scissors game: mutation affects paradoxical impacts," *Journal of Theoretical Biology*, vol. 450, pp. 22–29, 2018.
[33] D. F. P. Toupo, D. G. Rand, and S. H. Strogatz, "Limit cycles sparked by mutation in the repeated prisoner's Dilemma," *International Jounal of Bifurcation and Chaos*, vol. 24, no. 12, Article ID 1430035, 2014.
[34] D. F. P. Toupo and S. H. Strogatz, "Nonlinear dynamics of the rock-paper-scissors game with mutations," *Physical Review E*, vol. 91, no. 5, Article ID 052907, 2015.
[35] J. N. Webb, *Game Theory Decisions, Interaction and Evolution*, Springer Undergraduate Mathematics Series, New York, NY, USA, 2006.
[36] E. Wesson and R. Rand, "Hopf bifurcations in delayed rock-paper-scissors replicator dynamics," *Dynamic Games and Applications*, vol. 6, no. 1, pp. 139–156, 2016.
[37] E. Wessonn, R. Rand, and David Rand, "Hopf bifurcations in two-strategy delayed replicator dynamics," *International Jounal of Bifurcation and Chaos*, vol. 26, no. 1, Article ID 1650006, 2016.
[38] B. K. Nesrine, E. A. Rachid, and H. Yezekael, "Hopf bifurcations in replicator dynamics with distributed delays," 2017, https://arxiv.org/abs/1703.06721.
[39] Y. Umezuki, "Bifurcation analysis of the rock-paper-scissors game with discrete-time logit dynamics," *Mathematical Social Sciences*, vol. 95, pp. 54–65, 2018.
[40] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer, New York, NY, USA, 2002.
[41] S. Wiggins, *Intoduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer, New York, NY, USA, 2003.

Hindawi
Complexity
Volume 2019, Article ID 4590915, 17 pages
https://doi.org/10.1155/2019/4590915

Research Article

Stability and Hopf Bifurcation in a Three-Component Planktonic Model with Spatial Diffusion and Time Delay

Kejun Zhuang,[1] **Gao Jia**,[2] **and Dezhi Liu**[1]

[1]*School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China*
[2]*College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China*

Correspondence should be addressed to Kejun Zhuang; zhkj123@163.com

Received 21 March 2019; Accepted 28 May 2019; Published 11 July 2019

Academic Editor: Constantin Udriste

Due to the different roles that nontoxic phytoplankton and toxin-producing phytoplankton play in the whole aquatic system, a delayed reaction-diffusion planktonic model under homogeneous Neumann boundary condition is investigated theoretically and numerically. This model describes the interactions between the zooplankton and two kinds of phytoplanktons. The long-time behavior of the model and existence of positive constant equilibrium solution are first discussed. Then, the stability of constant equilibrium solution and occurrence of Hopf bifurcation are detailed and analyzed by using the bifurcation theory. Moreover, the formulas for determining the bifurcation direction and stability of spatially bifurcating solutions are derived. Finally, some numerical simulations are performed to verify the appearance of the spatially homogeneous and nonhomogeneous periodic solutions.

1. Introduction

Oceans have a major role in the global carbon cycling and so directly impact the pace and extent of climate change [1]. Marine organism can bring great economic and social values. It has the irreplaceable function in the global food processing, tourism, nutrient cycling, gas regulation, and so on. As the basis of the marine food chain system, plankton can supply food and oxygen to a myriad of marine life and can also absorb about half of the climate-warming carbon dioxide [2]. Besides, unlike fish or some intertidal creatures, plankton has rarely been commercially exploited. Moreover, this free-floating plankton can respond quickly to the temperature and change of oceanic system. As a consequence, mathematical modeling is a valuable tool for research fields of the marine ecology.

In general, plankton can be broadly divided into autotrophic phytoplankton and herbivorous zooplankton. The interactive process between phytoplankton and zooplankton is very complicated. It depends not only on the foraging style and feeding behavior, but also on other processes in the system. To describe the interactive relationship between phytoplankton and zooplankton, the following general model can be utilized:

$$\begin{aligned} \frac{dP}{dt} &= G(P) - F(P,Z) - M(P), \\ \frac{dZ}{dt} &= F(P,Z) - M(Z). \end{aligned} \tag{1}$$

Here, $P(t)$ and $Z(t)$ denote the concentration of phytoplankton and zooplankton respectively, $G(P)$ and $M(P)$ represent the growth rate and death rate of phytoplankton, respectively, $M(Z)$ is the death rate of zooplankton, and $F(P,Z)$ is the predation rate of zooplankton on phytoplankton. Recently, various particular cases of model (1) have been well studied [3–5]. Some interesting results about the stability and Hopf bifurcation type periodic oscillations have been obtained.

It should be noted that some phytoplankton can release toxic substances which will result in poisoning in both fish and shellfish. By accumulating in marine food webs, the toxins may have hazards on animal and human health. On the other hand, due to the rapid growth of plankton, harmful algal bloom may cause massive death of marine animals.

However, most previous studies did not directly consider the toxin-producing phytoplankton population and the role of toxic phytoplankton can not be ignored. While in the establishment of planktonic mathematical models, it is necessary to include the nontoxic phytoplankton, toxin-producing phytoplankton, and zooplankton. In 2004, Chattopadhyay et al. [6] proposed the following kinetic model which has three interacting components and is with an additional factor that the release of toxic substance reduces the growth of zooplankton:

$$\begin{aligned}\frac{dP_1}{dt} &= rP_1\left(1-\frac{P_1}{K}\right)-\alpha P_1 Z,\\ \frac{dP_2}{dt} &= sP_2\left(1-\frac{P_2}{K}\right)-\frac{\theta P_2 Z}{\gamma+P_2},\\ \frac{dZ}{dt} &= \beta P_1 Z-\mu Z-\frac{\theta_1 P_1 Z}{\gamma+P_2},\end{aligned} \tag{2}$$

where $P_1(t)$ denotes the concentration of the nontoxic phytoplankton at time t and $P_2(t)$ and $Z(t)$ denote the concentration of toxic phytoplankton population and zooplankton at time t, respectively. It is assumed that the two phytoplankton populations share the same resource. In model (2), all the coefficients are positive constants, r and s are the growth rates of two phytoplankton populations, respectively, K is the carrying capacity, α and β are the maximum zooplankton ingestion rate and maximum zooplankton conversion rate, respectively, μ is the death rate of zooplankton, θ is the rate of toxin liberation by toxic phytoplankton, and θ_1 is the specific predation rate of zooplankton population on toxic phytoplankton. This model shows that toxic substances released by phytoplankton have negative effects on the grazing pressure of zooplankton. It is finally concluded that toxin-producing phytoplankton may be used as a biocontrol agent for the harmful algal bloom problems. It is also mentioned that the role of time delay and environmental fluctuation in the planktonic dynamics may arouse some interesting results and needs further investigations.

Motivated by (2), some modified models have been proposed recently. For instance, Sarkar et al. [7] established a new model made of two harmful phytoplankton populations and one zooplankton population. Roy et al. [8] investigated the model where the two phytoplankton populations compete with each other. Pal et al. [9] further considered the three-component model with both nonlinear predation functions by zooplankton. Further, the study was also extended from the perspectives of stochastic dynamics and plankton-nutrient interactions, respectively [10, 11].

In view of the ocean current and monsoon, the plankton can freely drift and this spatial dispersal is subject to Fickian diffusion. So the effect of spatial diffusion has been investigated by many authors [12–17]. The results indicate that spatial diffusion has a vital role in the spatiotemporal dynamics of the planktonic model and spatial pattern may occur. Besides, the impact of time delay can not be ignored because it usually causes periodic oscillations, even chaotic behaviors, and time delay is ubiquitous in the real ecosystem [18–20].

According to the above factors, we consider the following three-component planktonic model with spatial diffusion and time delay:

$$\begin{aligned}\frac{\partial P}{\partial t} &= d_1\Delta P+r_1P\left(1-\frac{P}{K}\right)-\alpha PZ,\\ &\qquad (x,t)\in(0,L)\times(0,+\infty),\\ \frac{\partial T}{\partial t} &= d_2\Delta T+r_2T\left(1-\frac{T}{K}\right)-\frac{aTZ}{\gamma+T},\\ &\qquad (x,t)\in(0,L)\times(0,+\infty),\\ \frac{\partial Z}{\partial t} &= d_3\Delta Z+\beta PZ-dZ^2-\frac{bT(x,t-\tau)Z}{\gamma+T(x,t-\tau)},\\ &\qquad (x,t)\in(0,L)\times(0,+\infty),\\ \frac{\partial P}{\partial x} &= \frac{\partial T}{\partial x}=\frac{\partial Z}{\partial x}=0,\quad t\ge 0,\ x=0,L,\\ P(x,t) &= P_0(x,t)\ge 0,\\ T(x,t) &= T_0(x,t)\ge 0,\\ Z(x,t) &= Z_0(x,t)\ge 0,\\ &\qquad (x,t)\in(0,L)\times[-\tau,0],\end{aligned} \tag{3}$$

where $P(x,t)$, $T(x,t)$, and $Z(x,t)$ denote the densities of nontoxic phytoplankton, toxin-producing phytoplankton, and zooplankton at location x and time t, respectively, Δ is the usual Laplace operator, L denotes the depth of the water column, and the homogeneous Neumann boundary condition means that no plankton species is entering or leaving the column at the top or the bottom.

All the parameters are positive constants, $d_i, i=1,2,3$, are the three species' diffusion rates, respectively, d is the higher mortality of zooplankton, and τ is the time needed for zooplankton from ingesting toxic phytoplankton to dying. The other coefficients have the same meanings as in model (2). Note that the zooplankton may get eaten by higher predators, whose population is not being explicitly modelled [21–23]. So, we adopt the quadratic closure term dZ^2 to describe the higher mortality of zooplankton.

In this paper, we mainly investigate the spatiotemporal dynamics of delayed and diffusive system (3). The rest of the paper is organized as follows. In Section 2, the permanence and nonpersistence of system (3) are derived. In Section 3, the sufficient conditions for existence of positive constant equilibrium solution are obtained. In Section 4, the stability of equilibrium solution and delay-induced Hopf bifurcation are explored. In Section 5, the detailed formulae for determining the bifurcation properties are given by calculating the normal form on the center manifold. In Section 6, some numerical simulations are conducted to illustrate the theoretical results. Finally, some conclusions are given in Section 7.

2. Long-Time Behavior

In this section, we shall show that any nonnegative solution $(P(x,t), T(x,t), Z(x,t))$ of system (3) lies in a bounded region as $t\longrightarrow\infty$, for all $x\in\Omega$.

Theorem 1. *If $dr_1 > \alpha\beta K$ and $dr_1\beta\gamma - \alpha\beta^2\gamma K - bdr_1 > 0$ hold, then system (3) is permanent; that is, there exist positive constants m and M independent of solution such that*

$$\begin{aligned} m \le \liminf_{t\longrightarrow+\infty} P(x,t) \le \limsup_{t\longrightarrow+\infty} P(x,t) \le M,\\ m \le \liminf_{t\longrightarrow+\infty} T(x,t) \le \limsup_{t\longrightarrow+\infty} T(x,t) \le M,\\ m \le \liminf_{t\longrightarrow+\infty} Z(x,t) \le \limsup_{t\longrightarrow+\infty} Z(x,t) \le M \end{aligned} \tag{4}$$

for any nonnegative solution.

Proof. From the first equation of system (3), we have

$$\frac{\partial P}{\partial t} - d_1\Delta P \le r_1 P\left(1 - \frac{P}{K}\right). \tag{5}$$

The standard comparison principle implies

$$\limsup_{t\longrightarrow+\infty} P(x,t) \le K, \tag{6}$$

and thus for every real number $\varepsilon_1 > 0$, there exists a $T_1 > 0$ such that $P(x,t) \le K + \varepsilon_1$, for all $t > T_1$.

Similarly, from the second equation of (3), we have

$$\limsup_{t\longrightarrow+\infty} T(x,t) \le K, \tag{7}$$

And thus for every real number $\varepsilon_2 > 0$, there exists a $T_2 > 0$ such that $T(x,t) \le K + \varepsilon_2$, for all $t > T_2$.

The third equation of (3) can be reduced to

$$\begin{aligned} \frac{\partial Z}{\partial t} - d_3\Delta Z &\le \beta\left(K + \varepsilon_1\right)Z - dZ^2\\ &= Z\left[\beta\left(K + \varepsilon_1\right) - dZ\right], \end{aligned} \tag{8}$$

for all $t > T_1$. Then, we have $Z(x,t) \le \beta(K + \varepsilon_1)/d$, which indicates

$$\limsup_{t\longrightarrow+\infty} Z(x,t) \le \frac{\beta K}{d} \tag{9}$$

as $\varepsilon_1 \longrightarrow 0$. Therefore, for every real number $\varepsilon_3 > 0$, there exists a $T_3 > 0$ such that $Z(x,t) \le \beta K/d + \varepsilon_3$, for all $t > T_3$.

Again from the first equation of (3), we have

$$\frac{\partial P}{\partial t} - d_1\Delta P \ge r_1 P\left[1 - \frac{\alpha}{r_1}\left(\frac{\beta K}{d} + \varepsilon_3\right) - \frac{P}{K}\right], \tag{10}$$

for all $t > T_3$. This implies

$$P(x,t) \ge \frac{1 - (\alpha/r_1)\left(\beta K/d + \varepsilon_3\right)}{1/K}; \tag{11}$$

thus,

$$\liminf_{t\longrightarrow+\infty} P(x,t) \ge \frac{K\left(dr_1 - \alpha\beta K\right)}{dr_1} > 0. \tag{12}$$

Then, for every real number $\varepsilon_4 > 0$, there exists a $T_4 > 0$ such that $P(x,t) \ge K(dr_1 - \alpha\beta K)/dr_1 - \varepsilon_4$, for all $t > T_4$.

From the second equation of (3), we have

$$\begin{aligned} P(x,t) &\ge \frac{K\left(dr_1 - \alpha\beta K\right)}{dr_1} - \varepsilon_4\\ &= r_2 T\left[1 - \left(\frac{1}{K} + \frac{a}{r_2\gamma}\left(\frac{\beta K}{d} + \varepsilon_3\right)\right)T\right], \end{aligned} \tag{13}$$

which implies

$$\liminf_{t\longrightarrow+\infty} T(x,t) \ge \frac{dr_2\gamma K}{dr_2\gamma + a\beta K^2} > 0. \tag{14}$$

From the third equation of (3), we also have

$$\begin{aligned} \frac{\partial Z}{\partial t} - d_3\Delta Z \ge Z\left[\beta\left(\frac{K\left(dr_1 - \alpha\beta K\right)}{dr_1} - \varepsilon_4\right) - dZ\right.\\ \left. - \frac{b\left(K + \varepsilon_2\right)}{\gamma}\right], \end{aligned} \tag{15}$$

for all $t > \max\{T_2 + \tau, T_4\}$. Thus,

$$\liminf_{t\longrightarrow+\infty} Z(x,t) \ge \frac{dr_1\beta\gamma K - \alpha\beta^2\gamma K^2 - bdr_1 K}{d^2 r_1\gamma} > 0. \tag{16}$$

Finally, if we set

$$\begin{aligned} m = \min\left\{\frac{K\left(dr_1 - \alpha\beta K\right)}{dr_1}, \frac{dr_2\gamma K}{dr_2\gamma + a\beta K^2},\right.\\ \left.\frac{dr_1\beta\gamma K - \alpha\beta^2\gamma K^2 - bdr_1 K}{d^2 r_1\gamma}\right\}, \end{aligned} \tag{17}$$

and

$$M = \max\left\{K, \frac{\beta K}{d}\right\}, \tag{18}$$

then the proof is complete. □

Definition 2. System (3) is said to be not persistent if

$$\begin{aligned} &\min\left\{\liminf_{t\longrightarrow+\infty} P(x,t), \liminf_{t\longrightarrow+\infty} T(x,t), \liminf_{t\longrightarrow+\infty} Z(x,t)\right\}\\ &= 0 \end{aligned} \tag{19}$$

for some of its nonnegative solutions.

Next, we discuss the nonpersistence of system (3).

Theorem 3. *If $bdr_2\gamma \ge \beta(\gamma + K)(dr_2\gamma + a\beta K^2)$ holds, then system (3) is not persistent.*

Proof. According to the process of Theorem 1, for an arbitrary positive constant ε_5, there exists a $T_5 > 0$ such that

$$T(x,t) \ge \frac{dr_2\gamma K}{dr_2\gamma + a\beta K^2} - \varepsilon_5, \tag{20}$$

for all $t > T_5$. Further, we have

$$\frac{\partial Z}{\partial t} - d_3 \Delta Z \le Z\left[\beta\left(K+\varepsilon_1\right) - \frac{bdr_2\gamma K - b\varepsilon_5\left(dr_2\gamma + a\beta K^2\right)}{\left(\gamma+K+\varepsilon_2\right)\left(dr_2\gamma+a\beta K^2\right)} - dZ\right], \tag{21}$$

for all $t > \max\{T_1, T_2+\tau, T_5+\tau\}$. Then, by the arbitrariness of ε_1, ε_2, and ε_5, we have

$$\liminf_{t\longrightarrow+\infty} Z(x,t) \le \limsup_{t\longrightarrow+\infty} Z(x,t) \le \frac{\beta K}{d} - \frac{br_2\gamma K}{(\gamma+K)\left(dr_2\gamma+a\beta K^2\right)} \le 0 \tag{22}$$

as $t \longrightarrow +\infty$. The proof is complete. □

3. Steady State

In consideration of the biological significance of (3), we focus on the positive constant equilibrium solution. To determine the equilibrium solution of (3), we only need to solve the following algebraic equations:

$$\begin{aligned} r_1 P\left(1-\frac{P}{K}\right) - \alpha PZ &= 0,\\ r_2 T\left(1-\frac{T}{K}\right) - \frac{aTZ}{\gamma+T} &= 0,\\ \beta PZ - dZ^2 - \frac{bTZ}{\gamma+T} &= 0. \end{aligned} \tag{23}$$

Simplifying the first equation of (23) and substituting it into the third equation, we have

$$Z = \frac{r_1(K-P)}{\alpha K}, \tag{24}$$

and

$$\left(\beta + \frac{dr_1}{\alpha K}\right)P - \frac{dr_1}{\alpha} - \frac{bT}{\gamma+T} = 0. \tag{25}$$

From (24) and the second equation of (23), we have

$$\frac{aZ}{\gamma+T} = \frac{ar_1}{\alpha(\gamma+T)} - \frac{ar_1 P}{\alpha K(\gamma+T)} = r_2 - \frac{r_2}{K}T, \tag{26}$$

and

$$\frac{ar_1 P}{\alpha K(\gamma+T)} = \frac{ar_1}{\alpha(\gamma+T)} - r_2 + \frac{r_2}{K}T; \tag{27}$$

thus,

$$P = K - \frac{\alpha r_2 K(T+\gamma)}{ar_1} + \frac{\alpha r_2 T(T+\gamma)}{ar_1}, \tag{28}$$

which is equivalent to

$$P = \frac{\alpha r_2}{ar_1}\left[\left(T - \frac{K-\gamma}{2}\right)^2 + \frac{ar_1 K}{\alpha r_2} - \gamma K - \frac{(K-\gamma)^2}{4}\right]. \tag{29}$$

From (25) and (28), we have

$$\left(\beta + \frac{dr_1}{\alpha K}\right)\left[K - \frac{\alpha r_2 K(T+\gamma)}{ar_1} + \frac{\alpha r_2 T(T+\gamma)}{ar_1}\right] - \frac{dr_1}{\alpha} - \frac{bT}{\gamma+T} = 0, \tag{30}$$

and

$$a_3 T^3 + a_2 T^2 + a_1 T + a_0 = 0, \tag{31}$$

where

$$\begin{aligned} a_3 &= \frac{\alpha r_2}{ar_1}\left(\beta + \frac{dr_1}{\alpha K}\right) > 0,\\ a_2 &= \frac{\alpha r_2}{ar_1}(2\gamma - K)\left(\beta + \frac{dr_1}{\alpha K}\right),\\ a_1 &= \left(\beta + \frac{dr_1}{\alpha K}\right)\left[K + \frac{\alpha\gamma r_2}{ar_1}(\gamma - 2K)\right] - b - \frac{dr_1}{\alpha},\\ a_0 &= \gamma K\left(\beta + \frac{dr_1}{\alpha K}\right)\left(1 - \frac{\alpha\gamma r_2}{ar_1}\right) - \frac{dr_1\gamma}{\alpha}. \end{aligned} \tag{32}$$

According to Descartes' rule of signs, cubic equation (31) has at least one positive real root when $a_0 < 0$, that is,

$$K\left(\beta + \frac{dr_1}{\alpha K}\right)\left(1 - \frac{\alpha\gamma r_2}{ar_1}\right) < \frac{dr_1}{\alpha}. \tag{33}$$

For convenience, we denote any positive root of (31) by T^*. Combining (24) and (29), we can obtain the positive solution of (23) under the conditions $4ar_1K/\alpha r_2 > (K+\gamma)^2$ and $K > P^*$.

From above analyses, we can establish the existence of positive constant equilibrium solution of (3).

Theorem 4. *If the following assumption*

(*H1*) $K(\beta + dr_1/\alpha K)(1 - \alpha\gamma r_2/ar_1) < dr_1/\alpha$, $4ar_1K/\alpha r_2 > (K+\gamma)^2$ *and* $K > P^*$

holds, then system (3) has positive constant equilibrium solution $E^* = (P^*, T^*, Z^*)$.

Based on the aim of this study, we always assume that condition (H1) is satisfied in the following sections.

4. Hopf Bifurcation Induced by Time Delay

Here, we will regard time delay τ as the bifurcation parameter to investigate its effect on the stability of coexistence equilibrium solution $E^* = (P^*, T^*, Z^*)$.

Linearizing system (3) at E^* leads to the corresponding characteristic equation:

$$\begin{vmatrix} \lambda + d_1\left(\frac{k\pi}{L}\right) + \frac{r_1}{K}P^* & 0 & \alpha P^* \\ 0 & \lambda + d_2\left(\frac{k\pi}{L}\right) + \frac{r_2 T^*}{K} - \frac{aT^*Z^*}{(\gamma + T^*)^2} & \frac{aT^*}{\gamma + T^*} \\ -\beta Z^* & \frac{b\gamma Z^*}{(\gamma + T^*)^2} e^{-\lambda\tau} & \lambda + d_3\left(\frac{k\pi}{L}\right) + dZ^* \end{vmatrix} = 0, \tag{34}$$

which can be simplified as the following transcendental equation:

$$\lambda^3 + A_k\lambda^2 + B_k\lambda + C_k + (D_k\lambda + F_k)e^{-\lambda\tau} = 0, \tag{35}$$

where

$$\begin{aligned} A_k &= (d_1 + d_2 + d_3)\left(\frac{k\pi}{L}\right)^2 + \frac{r_1}{K}P^* + \frac{r_2}{K}T^* + dZ^* \\ &\quad - \frac{aT^*Z^*}{(\gamma + T^*)^2}, \\ B_k &= \alpha\beta P^* Z^* + \left[d_1\left(\frac{k\pi}{L}\right)^2 + \frac{r_1}{K}P^*\right]\left[d_2\left(\frac{k\pi}{L}\right)^2 \right. \\ &\quad \left. + \frac{r_2}{K}T^* - \frac{aT^*Z^*}{(\gamma + T^*)^2}\right] + \left[d_1\left(\frac{k\pi}{L}\right)^2 + \frac{r_1}{K}P^*\right] \\ &\quad \cdot \left[d_3\left(\frac{k\pi}{L}\right)^2 + dZ^*\right] + \left[d_2\left(\frac{k\pi}{L}\right)^2 + \frac{r_2}{K}T^* \right. \\ &\quad \left. - \frac{aT^*Z^*}{(\gamma + T^*)^2}\right]\left[d_3\left(\frac{k\pi}{L}\right)^2 + dZ^*\right], \\ C_k &= \left[d_2\left(\frac{k\pi}{L}\right)^2 + d_3\left(\frac{k\pi}{L}\right)^2 + \frac{r_2}{K}T^* - \frac{aT^*Z^*}{(\gamma + T^*)^2} \right. \\ &\quad \left. + dZ^*\right] \cdot \left[d_1\left(\frac{k\pi}{L}\right)^2 + \frac{r_1}{K}P^*\right] \\ &\quad + \alpha\beta P^* Z^*\left[d_2\left(\frac{k\pi}{L}\right)^2 + \frac{r_2}{K}T^* - \frac{aT^*Z^*}{(\gamma + T^*)^2}\right], \\ D_k &= -\frac{ab\gamma T^* Z^*}{(\gamma + T^*)^3} < 0, \\ F_k &= -\frac{ab\gamma T^* Z^*}{(\gamma + T^*)^3}\left[d_1\left(\frac{k\pi}{L}\right)^2 + \frac{r_1}{K}P^*\right] < 0. \end{aligned} \tag{36}$$

The special case of (35) with $\tau = 0$ is

$$\lambda^3 + A_k\lambda^2 + (B_k + D_k)\lambda + (C_k + F_k) = 0. \tag{37}$$

If all the roots of (37) have negative real parts for every nonnegative integer k, then the positive equilibrium solution E^* without time delay is asymptotically stable. With the help of Routh-Hurwitz criterion, E^* is stable if and only if $A_k > 0$, $B_k + D_k > 0$, $C_k + F_k > 0$, and $A_k(B_k + D_k) - (C_k + F_k) > 0$.

Assume that

(H2) $r_2/K > aZ^*/(\gamma + P_2^*)^2$, and $ab\gamma T^*/(\gamma + T^*)^3 < \min\{d, \alpha\beta P^*\}$;

then, we have the following stability conclusion.

Theorem 5. *If assumption (H2) is satisfied, then the equilibrium solution E^* of (3) is asymptotically stable for $\tau = 0$.*

Next, we shall discuss the distribution of characteristic roots when $\tau > 0$. Suppose $i\omega$ ($\omega > 0$) is a root of (35). Then, for any nonnegative integer k, we have

$$\begin{aligned} &-i\omega^3 - A_k\omega^2 + iB_k\omega + C_k \\ &\quad + (iD_k\omega + F_k)(\cos\omega\tau - i\sin\omega\tau) = 0, \end{aligned} \tag{38}$$

and

$$\begin{aligned} &-i\omega^3 - A_k\omega^2 + iB_k\omega + C_k + iD_k\omega\cos\omega\tau \\ &\quad + D_k\omega\sin\omega\tau + F_k\cos\omega\tau - iF_k\sin\omega\tau = 0. \end{aligned} \tag{39}$$

Separating the real and imaginary parts results in

$$\begin{aligned} \omega^3 - B_k\omega &= D_k\omega\cos\omega\tau - F_k\sin\omega\tau, \\ A_k\omega^2 - C_k &= F_k\cos\omega\tau + D_k\omega\sin\omega\tau; \end{aligned} \tag{40}$$

thus,

$$\begin{aligned} &\omega^6 + (A_k^2 - 2B_k)\omega^4 + (B_k^2 - 2A_kC_k - D_k^2)\omega^2 + C_k^2 \\ &\quad - F_k^2 = 0. \end{aligned} \tag{41}$$

For simplicity, we set $\sigma = \omega^2$; then, (41) can be rewritten in the form of

$$R(\sigma) = \sigma^3 + M_k\sigma^2 + N_k\sigma + P_k = 0, \tag{42}$$

where

$$\begin{aligned} M_k &= A_k^2 - 2B_k, \\ N_k &= B_k^2 - 2A_kC_k - D_k^2, \\ P_k &= C_k^2 - F_k^2. \end{aligned} \tag{43}$$

If the assumption

(H3) $M_k > 0$ and $M_k > 0$ for any $k > 0$

holds, then we have $P_k > 0$ by combining with (H2). And the cubic equation (42) of σ has no positive root, so (35) has no purely imaginary root. It can be concluded that the equilibrium solution E^* is always asymptotically stable for any $\tau > 0$ and that system (3) has no spatially nonhomogeneous periodic solution.

On the other hand, when $k = 0$, based on Lemma 2.2 in [24], the following condition is needed to ensure the existence of positive root of (42):

(H4) $\sigma_1 > 0$ and $R(\sigma_1)|_{k=0} \leq 0$, where $\sigma_1 = (-M_0 + \sqrt{M_0^2 - 3N_0})/3$.

Denote any positive root of (42) by σ^*; then, $\pm i\omega^*$ is a pair of purely imaginary roots of (35), where $\omega^* = \sqrt{\sigma^*}$.

When $k = 0$, rewrite (35) as

$$\lambda^3 + A_0\lambda^2 + B_0\lambda + C_0 + (D_0\lambda + F_0)e^{-\lambda\tau} = 0, \tag{44}$$

and we have

$$\sin\omega^*\tau = \frac{(A_0D_0 - F_0)\omega^{*3} + (B_0F_0 - C_0D_0)\omega^*}{D_0^2\omega^{*2} + F_0^2} \triangleq F_s, \tag{45}$$

$$\cos\omega^*\tau = \frac{D_0\omega^{*4} + (A_0F_0 - B_0D_0)\omega^{*2} - C_0F_0}{D_0^2\omega^{*2} + F_0^2} \triangleq F_c. \tag{46}$$

Define

$$\tau_0^j = \begin{cases} \frac{1}{\omega^*}\{2\pi - \arccos F_c + 2j\pi\}, & F_s \geq 0, \\ \frac{1}{\omega^*}\{\arccos F_c + 2j\pi\}, & F_c < 0, \end{cases} \tag{47}$$

where $j = 0, 1, 2, \ldots$ and $\tau_0 = \tau_0^0 = \min_{j=0,1,2,\ldots}\{\tau_0^j\}$. Then, (44) has a pair of purely imaginary roots $\pm i\omega^*$ when $\tau = \tau_0^j$.

We claim that if

(H5) $R'(\sigma^*) \neq 0$,

then

$$\left.\frac{\mathrm{d}\,\mathrm{Re}\lambda(\tau)}{\mathrm{d}\tau}\right|_{\tau=\tau_0^j,\omega=\omega^*} \neq 0. \tag{48}$$

In fact, differentiating both sides of (35) with respect to τ, it follows that

$$\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right)^{-1} = \frac{(3\lambda^2 + 2A_0\lambda + B_0)e^{\lambda\tau} + D_0}{\lambda(D_0\lambda + F_0)} - \frac{\tau}{\lambda}. \tag{49}$$

Thus, we have

$$\mathrm{Re}\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right)^{-1}\Bigg|_{\tau=\tau_0^j,\omega=\omega^*} = \frac{\omega^{*2}R'(\sigma^*)}{(D_0\omega^{*2})^2 + (F_0\omega^*)^2} \neq 0. \tag{50}$$

From what has been discussed above and the Hopf bifurcation Theorem by Hassard et al. [25], we can draw the conclusion on the existence of spatially homogeneous Hopf bifurcation.

Theorem 6. *Suppose that conditions (H1)-(H5) are satisfied.*

(i) The equilibrium solution E^* is locally asymptotically stable for $\tau \in [0, \tau_0)$.

(ii) The equilibrium solution E^* is unstable for $\tau > \tau_0$ and τ_0 is the Hopf bifurcation value.

(ii) System (3) undergoes spatially homogeneous periodic solutions at E^* when $\tau = \tau_0$.

5. Stability and Direction of the Bifurcation

In this section, we investigate the properties of spatially homogeneous periodic solutions, including bifurcation direction, stability of periodic solutions, monotonicity of periodic solutions. Here, we mainly apply the normal form theory and center manifold theorem for partial functional differential equations [25, 26].

For fixed $j = 0, 1, 2, \ldots$, denote bifurcation value τ_0^j by τ^* and introduce the new parameter $\mu = \tau - \tau^*$; then, $\mu = 0$ is the new Hopf bifurcation value. Let $\vartheta = t/\tau$ and rewrite ϑ as t; system (3) can be transformed into

$$\frac{\mathrm{d}U(t)}{\mathrm{d}t} = \tau^*D\Delta U(t) + L(\tau^*)(U_t) + F(U_t, \mu), \tag{51}$$

where $L(\mu)(\varphi) : C \longrightarrow X$ and $F(\cdot, \mu) : C \longrightarrow X$ are given by

$$\begin{aligned} &\varphi = (\varphi_1, \varphi_2, \varphi_3)^T \in C, \\ &D = \mathrm{diag}\{d_1, d_2, d_3\}, \\ &\Delta = \mathrm{diag}\left\{\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x^2}\right\}, \\ &U_t(\theta) = U(t + \theta), \quad -\tau \leq \theta \leq 0, \\ &L(\mu)(\varphi) \\ &= \mu\begin{pmatrix} -\frac{r_1}{K}P^*\varphi_1(0) - \alpha P^*\varphi_3(0) \\ \left(\frac{aT^*Z^*}{(\gamma + T^*)^2} - \frac{r_2}{K}T^*\right)\varphi_2(0) - \frac{aT^*}{\gamma + T^*}\varphi_3(0) \\ \beta Z^*\varphi_1(0) - \frac{b\gamma Z^*}{(\gamma + T^*)^2}\varphi_2(-1) - dZ^*\varphi_3(0) \end{pmatrix}, \\ &F(\varphi, \mu) = \mu D\Delta\varphi(0) + L(\mu)(\varphi) + f(\varphi, \mu), \end{aligned} \tag{52}$$

and

$$f(\varphi,\mu)=(\tau^*+\mu)\begin{pmatrix}-\frac{2r_1}{K}\varphi_1^2(0)-\alpha\varphi_1(0)\varphi_3(0)\\ \frac{2a\gamma Z^*}{(\gamma+T^*)^3}\varphi_2^2(0)-\frac{2r_2}{K}\varphi_2^2(0)-\frac{a\gamma}{(\gamma+T^*)^2}\varphi_2(0)\varphi_3(0)\\ \beta\varphi_1(0)\varphi_3(0)-2d\varphi_3^2(0)-\frac{b\gamma\varphi_2(-1)\varphi_3(0)}{(\gamma+T^*)^2}+\frac{2b\gamma Z^*}{(\gamma+T^*)^3}\varphi_2^2(-1)\end{pmatrix}+\text{h.o.t.} \tag{53}$$

Then, the linearized system of (51) at origin is

$$\frac{\mathrm{d}U(t)}{\mathrm{d}t}=\tau^*D\Delta U(t)+L(\tau^*)(U_t). \tag{54}$$

From the discussion in Section 4, we can find that characteristic equation (35) has a pair of purely imaginary roots $\Lambda_0=\{i\omega^*\tau^*,-i\omega^*\tau^*\}$ when $\tau=\tau^*$.

Let $\mathscr{C}:=C([-1,0],\mathbb{R}^3)$. Then, we consider the functional differential equation on $\mathscr{C}$:

$$\dot{z}=L(\tau^*)(z_t). \tag{55}$$

It is obvious that $L(\tau^*)$ is a continuous linear function mapping $C([-1,0],\mathbb{R}^3)$ into $\mathbb{R}^3$. By the Riesz representation theorem, there exists a 3×3 matrix function $\eta(\theta,\tau)$ $(-1\le\theta\le 0)$ such that

$$L(\tau^*)(\varphi)=\int_{-1}^{0}[\mathrm{d}\eta(\theta,\tau^*)]\varphi(\theta),\quad \text{for } \varphi\in\mathscr{C}; \tag{56}$$

here we choose

$$\begin{aligned}\eta(\theta,\tau^*)&=\tau^*\begin{pmatrix}-\frac{r_1}{K}P^* & 0 & -\alpha P^*\\ 0 & \frac{aT^*Z^*}{(\gamma+T^*)^2}-\frac{r_2}{K}T^* & -\frac{aT^*}{\gamma+T^*}\\ \beta Z^* & 0 & -dZ^*\end{pmatrix}\delta(\theta)\\ &\quad-\tau^*\begin{pmatrix}0&0&0\\0&0&0\\0&-\frac{b\gamma Z^*}{(\gamma+T^*)^2}&0\end{pmatrix}\delta(\theta+1).\end{aligned} \tag{57}$$

Let $A(\tau^*)$ be the infinitesimal generator of the semigroup induced by the solutions of (55) and A^* be the adjoint matrix of $A(\tau^*)$ under the bilinear pairing:

$$\begin{aligned}(\psi,\phi)&=(\psi(0),\phi(0))\\ &\quad-\int_{-1}^{0}\int_{\xi=0}^{\theta}\psi(\xi-\theta)\,\mathrm{d}\eta(\theta)\phi(\xi)\,\mathrm{d}\xi=(\psi(0),\phi(0))\\ &\quad+\tau^*\int_{-1}^{0}\psi(\theta+1)\cdot\begin{pmatrix}0&0&0\\0&0&0\\0&-\frac{b\gamma Z^*}{(\gamma+P_2^*)^2}&0\end{pmatrix}\phi(\theta)\,\mathrm{d}\theta,\end{aligned} \tag{58}$$

where $\varphi\in C,\psi\in C^*=C([0,1],\mathbb{R}^3)$. Then, $A(\tau^*)$ and A^* are a pair of adjoint operators and they both have characteristic roots $\pm i\omega^*\tau^*$. Let P and P^* be the generalized eigenspaces of $A(\tau^*)$ and A^*, respectively; then, P^* is the adjoint space of P and $\dim P=\dim P^*=2$.

By direct calculation, we have the following lemma.

Lemma 7. *Let*

$$\begin{aligned}\xi&=\frac{aT^*\zeta(\gamma+T^*)}{aKT^*Z^*-r_2T^*(\gamma+T^*)^2-i\omega^*K(\gamma+T^*)^2},\\ \zeta&=-\frac{i\omega^*K+r_1P^*}{\alpha KP^*},\\ \xi^*&=-\frac{aKT^*Z^*+bK\gamma\zeta^*Z^*e^{-i\omega^*\tau^*}}{r_2T^*(\gamma+T^*)^2},\\ \zeta^*&=\frac{i\omega^*K+r_1P^*}{\beta KZ^*};\end{aligned} \tag{59}$$

then a basis of P with Λ_0 is

$$\begin{aligned}p_1(\theta)&=e^{i\omega^*\tau^*\theta}(1,\xi,\zeta)^T,\\ p_2(\theta)&=\overline{p_1(\theta)},\\ &\quad -1\le\theta\le 0,\end{aligned} \tag{60}$$

and a basis a of P^* with Λ_0 is

$$\begin{aligned}q_1(s)&=e^{-i\omega^*\tau^*s}(1,\xi^*,\zeta^*)^T,\\ q_2(s)&=\overline{q_1(s)},\\ &\quad 0\le s\le 1.\end{aligned} \tag{61}$$

Let $\Phi = (\Phi_1, \Phi_2)$, $\Psi^* = (\Psi_1^*, \Psi_2^*)^T$, where

$$\Phi_1(\theta) = \frac{p_1(\theta) + p_2(\theta)}{2},$$
$$\Phi_2(\theta) = \frac{p_1(\theta) - p_2(\theta)}{2i}, \quad \theta \in [-1, 0], \tag{62}$$

and

$$\Psi_1^*(s) = \frac{q_1(s) + q_2(s)}{2},$$
$$\Psi_2^*(s) = \frac{q_1(s) - q_2(s)}{2i}, \quad s \in [0, 1]. \tag{63}$$

From (58), we can obtain (Ψ_1^*, Φ_2), (Ψ_1^*, Φ_1), and further

$$(q_1, p_1) = (\Psi_1^*, \Phi_1) - (\Psi_2^*, \Phi_2) + i[(\Psi_1^*, \Phi_2) - (\Psi_2^*, \Phi_1)]; \tag{64}$$

that is,

$$(q_1, p_1) = 1 + \xi\xi^* + \zeta\zeta^* - \tau^* \frac{bZ^*}{\gamma + P_2^*} \xi^* e^{-i\omega^*\tau^*} \triangleq D^*. \tag{65}$$

Therefore, we have

$$(\Psi_1^*, \Phi_1) - (\Psi_2^*, \Phi_2) = \mathrm{Re}\{D^*\},$$
$$(\Psi_1^*, \Phi_2) - (\Psi_2^*, \Phi_1) = \mathrm{Im}\{D^*\}. \tag{66}$$

Now, we define $(\Psi^*, \Phi) = (\Psi_l^*, \Phi_m)$ $(l, m = 1, 2)$ and construct a new basis for P^* by $\Psi = (\Psi_1, \Psi_2)^T = (\Psi^*, \Phi)^{-1}\Psi^*$. Moreover, we define $f_0 = (\xi_0^1, \xi_0^2, \xi_0^3)$, where

$$\xi_0^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \xi_0^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \xi_0^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{67}$$

Let $c \cdot f_0$ be defined by

$$c \cdot f_0 = c_1\xi_0^1 + c_2\xi_0^2 + c_3\xi_0^3 \tag{68}$$

for $c = (c_1, c_2, c_3)^T$, $c_j \in \mathbb{R}(j = 1, 2, 3)$. Then, the center space of linear equation (54) is given by $P_{CN}\mathscr{C}$, where

$$P_{CN}\varphi = \Phi(\Psi, \langle \varphi, f_0 \rangle) \cdot f_0, \quad \varphi \in \mathscr{C}, \tag{69}$$

and $C = P_{CN}\mathscr{C} \oplus P_S\mathscr{C}$, $P_S\mathscr{C}$ denotes the complementary subspace of $P_{CN}\mathscr{C}$.

If

$$A_{\tau^*}\varphi(\theta) = \dot{\varphi}(\theta) + X_0(\theta)[\tau^* D\Delta\varphi(0) + L(\tau^*)(\varphi(\theta) - \dot{\varphi}(0))], \quad \varphi \in B\mathscr{C}, \tag{70}$$

where $X_0 : [-1, 0] \longrightarrow B(X, X)$ and $X_0(\theta) = \begin{cases} 0, & -1 \le \theta \le 0, \\ I, & \theta = 0. \end{cases}$ then A_{τ^*} is the infinitesimal generator induced by the solutions of (51) and (54), which can be written as the following operator differential equation:

$$\dot{U}_t = A_{\tau^*}U_t + X_0 F(U_t, \mu), \tag{71}$$

Therefore, the solution of (51) can be written in the form of

$$U_t = \Phi\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \cdot f_0 + h(x_1, x_2, \mu), \quad \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = (\Psi, \langle U_t, f_0 \rangle), \tag{72}$$

where $h(x_1, x_2, \mu) \in P_S\mathscr{C}$, $h(0, 0, 0) = Dh(0, 0, 0) = 0$. Specifically, the solution of (51) on the center manifold is

$$U_t^* = \Phi\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \cdot f_0 + h(x_1, x_2, 0). \tag{73}$$

Let $z = x_1 - ix_2$, and notice that $p_1 = \Phi_1 + i\Phi_2$; then, we can rewrite (73) as

$$U_t^* = \frac{1}{2}\Phi\begin{pmatrix} z + \overline{z} \\ i(z - \overline{z}) \end{pmatrix} \cdot f_0 + W(z, \overline{z}) = \frac{1}{2}(p_1 z + \overline{p}_1 \overline{z}) \cdot f_0 + W(z, \overline{z}), \tag{74}$$

where $W(z, \overline{z}) = h((z + \overline{z})/2, -(z - \overline{z})/2i, 0)$. Furthermore, from [26], z also satisfies

$$\dot{z} = i\omega^*\tau^* z + g(z, \overline{z}), \tag{75}$$

where

$$g(z, \overline{z}) = (\Psi_1(0) - i\Psi_2(0))\langle F(U_t^*, 0), f_0 \rangle. \tag{76}$$

Let

$$W(z, \overline{z}) = W_{20}\frac{z^2}{2} + W_{11}z\overline{z} + W_{02}\frac{\overline{z}^2}{2} + \cdots \tag{77}$$

and

$$g(z, \overline{z}) = g_{20}\frac{z^2}{2} + g_{11}z\overline{z} + g_{02}\frac{\overline{z}^2}{2} + \cdots. \tag{78}$$

By (74), it is not difficult to compute that

$$\begin{aligned}\left\langle F\left(U_t^*,0\right),f_0\right\rangle &= \frac{\tau^* z^2}{4}\begin{pmatrix} -\frac{r_1}{K}-\alpha\zeta \\ \frac{a\gamma Z^*}{\left(\gamma+T^*\right)^3}\xi^2-\frac{r_2}{K}\xi^2-\frac{a\gamma}{\left(\gamma+T^*\right)^2}\xi\zeta \\ \beta\zeta-2d\zeta^2-\frac{2b\gamma}{\left(\gamma+T^*\right)^2}\xi\zeta e^{-i\omega^*\tau^*}+\frac{2b\gamma Z^*}{\left(\gamma+T^*\right)^3}\zeta^2 e^{-2i\omega^*\tau^*}\end{pmatrix}\\ &+\frac{\tau^* z\overline{z}}{4}\begin{pmatrix} -\frac{2r_1}{K}-\alpha\left(\zeta+\overline{\zeta}\right) \\ \frac{2a\gamma Z^*}{\left(\gamma+T^*\right)^3}\xi\overline{\xi}-\frac{2r_2}{K}\xi\overline{\xi}-\frac{a\gamma}{\left(\gamma+T^*\right)^2}\left(\xi\overline{\zeta}+\overline{\xi}\zeta\right) \\ \beta\left(\zeta+\overline{\zeta}\right)-2d\zeta\overline{\zeta}-\frac{b\gamma}{\left(\gamma+T^*\right)^2}\left(\xi\overline{\zeta}e^{-i\omega^*\tau^*}+\overline{\xi}\zeta e^{i\omega^*\tau^*}\right)+\frac{4b\gamma Z^*}{\left(\gamma+T^*\right)^3}\zeta\overline{\zeta}\end{pmatrix}\\ &+\frac{\tau^*\overline{z}^2}{4}\begin{pmatrix} -\frac{r_1}{K}-\alpha\overline{\zeta} \\ \frac{a\gamma Z^*}{\left(\gamma+T^*\right)^3}\overline{\xi}^2-\frac{r_2}{K}\overline{\xi}^2-\frac{a\gamma}{\left(\gamma+T^*\right)^2}\overline{\xi\zeta} \\ \beta\overline{\zeta}-2d\overline{\zeta}^2-\frac{2b\gamma}{\left(\gamma+T^*\right)^2}\overline{\xi\zeta}e^{i\omega^*\tau^*}+\frac{2b\gamma Z^*}{\left(\gamma+T^*\right)^3}\overline{\zeta}^2\end{pmatrix}\\ &+\frac{\tau^* z^2\overline{z}}{4}\begin{pmatrix} \left\langle -\frac{r_1}{K}\left(4W_{11}^{(1)}(0)+2W_{20}^{(1)}(0)\right)\right. \\ \left.-\alpha\left(W_{20}^{(3)}(0)+2W_{11}^{(1)}(0)+2\zeta W_{11}^{(1)}(0)+\overline{\zeta}W_{20}^{(1)}(0)\right),1\right\rangle \\ \left\langle\left(\frac{a\gamma Z^*}{\left(\gamma+T^*\right)^3}-\frac{r_2}{K}\right)\left(2\overline{\xi}W_{20}^{(2)}(0)+4\xi W_{11}^{(2)}(0)\right)\right. \\ \left.-\frac{a\gamma}{\left(\gamma+T^*\right)^2}\left(2\xi W_{11}^{(3)}(0)+\overline{\xi}W_{20}^{(3)}(0)+2\zeta W_{11}^{(2)}(0)+\overline{\zeta}W_{20}^{(2)}(0)\right),1\right\rangle \\ \left\langle\frac{-b\gamma}{\left(\gamma+T^*\right)^2}\left(2\xi e^{-i\omega^*\tau^*}W_{11}^{(3)}(0)+\overline{\xi}e^{i\omega^*\tau^*}W_{20}^{(3)}(0)+2\zeta W_{11}^{(2)}(-1)+\overline{\zeta}W_{20}^{(2)}(-1)\right)\right. \\ -2d\left(2\overline{\zeta}W_{20}^{(3)}(0)+4\zeta W_{11}^{(3)}(0)\right)+\beta\left(W_{20}^{(3)}(0)+2W_{11}^{(1)}(0)+2\zeta W_{11}^{(1)}(0)\right. \\ \left.\left.+\overline{\zeta}W_{20}^{(1)}(0)\right)+\frac{2b\gamma Z^*}{\left(\gamma+T^*\right)^3}\left(2\xi e^{-i\omega^*\tau^*}W_{11}^{(2)}(-1)+\overline{\xi}e^{i\omega^*\tau^*}W_{20}^{(2)}(-1)\right),1\right\rangle\end{pmatrix}\\ &+\cdots,\end{aligned} \tag{79}$$

where $\langle W_{ij}^{(n)}(\theta),1\rangle = (1/\pi)\int_0^{\pi} W_{ij}^{(n)}(\theta)(x)dx$, $i+j=2$, $n=1,2,3$. Let $(\psi_1,\psi_2,\psi_3)=\Psi_1(0)-i\Psi_2(0)$. We can get the following expressions:

$$\begin{aligned}g_{20} &= \frac{\tau^*}{2}\left[-\left(\frac{r_1}{K}-\alpha\zeta\right)\psi_1+\left(\frac{a\gamma Z^*}{\left(\gamma+T^*\right)^3}\xi^2-\frac{r_2}{K}\xi^2\right.\right.\\ &-\left.\frac{a\gamma}{\left(\gamma+T^*\right)^2}\xi\zeta\right)\psi_2+\left(\beta\zeta-2d\zeta^2-\frac{2b\gamma}{\left(\gamma+T^*\right)^2}\right.\\ &\cdot\left.\left.\xi\zeta e^{-i\omega^*\tau^*}+\frac{2b\gamma Z^*}{\left(\gamma+T^*\right)^3}\zeta^2 e^{-2i\omega^*\tau^*}\right)\psi_3\right],\end{aligned}$$

$$\begin{aligned}g_{11} &= \frac{\tau^*}{4}\left[-\left(\frac{2r_1}{K}+\alpha\left(\zeta+\overline{\zeta}\right)\right)\psi_1+\left(\frac{2a\gamma Z^*}{\left(\gamma+T^*\right)^3}\xi\overline{\xi}\right.\right.\\ &-\frac{2r_2}{K}\xi\overline{\xi}-\left.\frac{a\gamma}{\left(\gamma+T^*\right)^2}\left(\xi\overline{\zeta}+\overline{\xi}\zeta\right)\right)\psi_2+\left(\beta\left(\zeta+\overline{\zeta}\right)\right.\\ &-2d\zeta\overline{\zeta}-\frac{b\gamma}{\left(\gamma+T^*\right)^2}\left(\xi\overline{\zeta}e^{-i\omega^*\tau^*}+\overline{\xi}\zeta e^{i\omega^*\tau^*}\right)\\ &+\left.\left.\frac{4b\gamma Z^*}{\left(\gamma+T^*\right)^3}\zeta\overline{\zeta}\right)\psi_3\right],\end{aligned}$$

$$
\begin{aligned}
g_{02} &= \overline{g_{20}},\\
g_{21} &= \frac{\tau^*}{2}\Bigg[\left\langle -\frac{r_1}{K}\left(4W_{11}^{(1)}(0)+2W_{20}^{(1)}(0)\right)\right.\\
&\quad\left.-\alpha\left(W_{20}^{(3)}(0)+2(1+\zeta)W_{11}^{(1)}(0)+\overline{\zeta}W_{20}^{(1)}(0)\right),1\right\rangle\\
&\quad\cdot\psi_1+\left\langle\left(\frac{a\gamma Z^*}{(\gamma+T^*)^3}-\frac{r_2}{K}\right)\left(2\overline{\xi}W_{20}^{(2)}(0)\right.\right.\\
&\quad\left.+4\xi W_{11}^{(2)}(0)\right)-\frac{a\gamma}{(\gamma+T^*)^2}\left(2\xi W_{11}^{(3)}(0)\right.\\
&\quad\left.\left.+\overline{\xi}W_{20}^{(3)}(0)+2\zeta W_{11}^{(2)}(0)+\overline{\zeta}W_{20}^{(2)}(0),1\right)\right\rangle\psi_2\\
&\quad+\left\langle\frac{-b\gamma}{(\gamma+T^*)^2}\left(2\xi e^{-i\omega^*\tau^*}W_{11}^{(3)}(0)\right.\right.\\
&\quad+\overline{\xi}e^{i\omega^*\tau^*}W_{20}^{(3)}(0)+2\zeta W_{11}^{(2)}(-1)+\overline{\zeta}W_{20}^{(2)}(-1)\Big)\\
&\quad-2d\left(2\overline{\zeta}W_{20}^{(3)}(0)+4\zeta W_{11}^{(3)}(0)\right)+\beta\left(W_{20}^{(3)}(0)\right.\\
&\quad+2W_{11}^{(1)}(0)+2\zeta W_{11}^{(1)}(0)+\overline{\zeta}W_{20}^{(1)}(0)\Big)\\
&\quad+\frac{2b\gamma Z^*}{(\gamma+T^*)^3}\left(2\xi e^{-i\omega^*\tau^*}W_{11}^{(2)}(-1)\right.\\
&\quad\left.\left.+\overline{\xi}e^{i\omega^*\tau^*}W_{20}^{(2)}(-1)\right),1\right\rangle\psi_3\Bigg].
\end{aligned}
\tag{80}
$$

Since $W_{20}(\theta)$ and $W_{11}(\theta)$ appear in g_{21}, we should also establish them. According to (77), we have

$$
\dot{W}(z,\overline{z})=W_{20}z\dot{z}+W_{11}\left(\dot{z}\overline{z}+z\dot{\overline{z}}\right)+W_{02}\overline{z}\dot{\overline{z}}+\cdots, \tag{81}
$$

and

$$
A_{\tau^*}W=A_{\tau^*}W_{20}\frac{z^2}{2}+A_{\tau^*}W_{11}z\overline{z}+A_{\tau^*}W_{02}\frac{\overline{z}^2}{2}+\cdots. \tag{82}
$$

In addition, by [26], $W(z(t),\overline{z}(t))$ also satisfies

$$
\dot{W}=A_{\tau^*}W+H(z,\overline{z}), \tag{83}
$$

where

$$
\begin{aligned}
H(z,\overline{z}) &= H_{20}\frac{z^2}{2}+H_{11}z\overline{z}+H_{02}\frac{\overline{z}^2}{2}+\cdots\\
&= X_0F\left(U_t^*,0\right)-\Phi\left(\Psi,\left\langle X_0F\left(U_t^*,0\right),f_0\right\rangle\right)\\
&\quad\cdot f_0
\end{aligned}
\tag{84}
$$

with $H_{ij}\in P_S\mathscr{C}$ and $i+j=2$.

Consequently, (74) and (81)-(83) lead to

$$
\begin{aligned}
\left(2i\omega^*\tau^*-A_{\tau^*}\right)W_{20}&=H_{20},\\
-A_{\tau^*}W_{11}&=H_{11}.
\end{aligned}
\tag{85}
$$

Since A_{τ^*} has a pair of purely imaginary characteristic roots, (83) has the unique solution such that

$$
\begin{aligned}
W_{20}&=\left(2i\omega^*\tau^*-A_{\tau^*}\right)^{-1}H_{20},\\
W_{11}&=-A_{\tau^*}^{-1}H_{11}.
\end{aligned}
\tag{86}
$$

For $\theta\in[-1,0]$, it follows from (84) that

$$
\begin{aligned}
H(z,\overline{z}) &= -\Phi(\theta)\Psi(\theta)\left\langle F\left(U_t^*,0\right),f_0\right\rangle\cdot f_0\\
&= -\left(\frac{p_1(\theta)+p_2(\theta)}{2},\frac{p_1(\theta)-p_2(\theta)}{2i}\right)\\
&\quad\cdot\left(\Psi_1(0),\Psi_2(0)\right)^T\times\left\langle F\left(U_t^*,0\right),f_0\right\rangle\cdot f_0\\
&= -\frac{1}{2}\left[p_1(\theta)\left(\Psi_1(0)-i\Psi_2(0)\right)\right.\\
&\quad\left.+p_2(\theta)\left(\Psi_1(0)+i\Psi_2(0)\right)\right]\times\left\langle F\left(U_t^*,0\right),f_0\right\rangle\cdot f_0\\
&= -\frac{1}{4}\left[g_{20}p_1(\theta)+\overline{g}_{02}p_2(\theta)\right]z^2\cdot f_0-\frac{1}{2}\left[g_{11}p_1(\theta)\right.\\
&\quad\left.+\overline{g}_{11}p_2(\theta)\right]z\overline{z}\cdot f_0+\cdots.
\end{aligned}
\tag{87}
$$

Thus, for $\theta\in[-1,0]$, we have

$$
\begin{aligned}
H_{20}(\theta) &= -\frac{1}{2}\left[g_{20}p_1(\theta)+\overline{g}_{02}p_2(\theta)\right]\cdot f_0,\\
H_{11}(\theta) &= -\frac{1}{2}\left[g_{11}p_1(\theta)+\overline{g}_{11}p_2(\theta)\right]\cdot f_0,\\
H(z,\overline{z})(0) &= F\left(U_t^*,0\right)-\Phi\left(\Psi,\left\langle F\left(U_t^*,0\right)\right\rangle,f_0\right)\cdot f_0,\\
H_{20}(0) &= \frac{\tau^*}{2}\\
&\quad\cdot\begin{pmatrix}-\frac{r_1}{K}-\alpha\zeta\\ \frac{a\gamma Z^*}{(\gamma+T^*)^3}\xi^2-\frac{r_2}{K}\xi^2-\frac{a\gamma}{(\gamma+T^*)^2}\xi\zeta\\ \beta\zeta-2d\zeta^2-\frac{2b\gamma\xi\zeta e^{-i\omega^*\tau^*}}{(\gamma+T^*)^2}+\frac{2b\gamma Z^*\zeta^2e^{-2i\omega^*\tau^*}}{(\gamma+T^*)^3}\end{pmatrix}\\
&\quad-\frac{g_{20}p_1(0)+\overline{g}_{02}p_2(0)}{2}\cdot f_0,
\end{aligned}
\tag{88}
$$

and

$$H_{11}(0) = \frac{\tau^*}{4}\begin{pmatrix} -\frac{2r_1}{K} - \alpha\left(\zeta + \overline{\zeta}\right) \\ \frac{2a\gamma Z^*}{\left(\gamma + T^*\right)^3}\xi\overline{\xi} - \frac{2r_2}{K}\xi\overline{\xi} - \frac{a\gamma}{\left(\gamma + T^*\right)^2}\left(\xi\overline{\zeta} + \overline{\xi}\zeta\right) \\ \beta\left(\zeta + \overline{\zeta}\right) - 2d\zeta\overline{\zeta} - \frac{b\gamma}{\left(\gamma + T^*\right)^2}\left(\xi\overline{\zeta}e^{-i\omega^*\tau^*} + \overline{\xi}\zeta e^{i\omega^*\tau^*}\right) + \frac{4b\gamma Z^*}{\left(\gamma + T^*\right)^3}\zeta\overline{\zeta} \end{pmatrix} - \frac{g_{11}p_1(0) + \overline{g}_{11}p_2(0)}{2}\cdot f_0. \quad (89)$$

Combining with the definition of A_{τ^*} and (86), we have

$$\dot{W}_{20}(\theta) = 2i\omega^*\tau^* W_{20}(\theta) + \frac{g_{20}p_1(0) + \overline{g}_{02}p_2(0)}{2}\cdot f_0, \quad -1 \le \theta \le 0. \quad (90)$$

Since $p_1(\theta) = p_1(0)e^{i\omega^*\tau^*}$, we have

$$W_{20}(\theta) = \frac{i}{2}\left[\frac{g_{20}}{\omega^*\tau^*}p_1(\theta) + \frac{\overline{g}_{02}}{3\omega^*\tau^*}p_2(\theta)\right]\cdot f_0 + Ee^{2i\omega^*\tau^*\theta}, \quad (91)$$

where

$$E = W_{20}(0) - \frac{i}{2}\left[\frac{g_{20}}{\omega^*\tau^*}p_1(0) + \frac{\overline{g}_{02}}{3\omega^*\tau^*}p_2(0)\right]\cdot f_0. \quad (92)$$

From the definition of A_{τ^*} and (86) and (92), we have

$$\begin{aligned} &2i\omega^*\tau^*\left[\frac{ig_{20}}{2\omega^*\tau^*}p_1(0)\cdot f_0 + \frac{i\overline{g}_{20}}{6\omega^*\tau^*}p_2(0)\cdot f_0 + E\right] \\ &- \tau^* D\Delta\left[\frac{ig_{20}}{2\omega^*\tau^*}p_1(0)\cdot f_0 + \frac{i\overline{g}_{20}}{6\omega^*\tau^*}p_2(0)\cdot f_0 + E\right] \\ &- L(\tau^*) \\ &\cdot\left[\frac{ig_{20}}{2\omega^*\tau^*}p_1(0)\cdot f_0 + \frac{i\overline{g}_{20}}{6\omega^*\tau^*}p_2(0)\cdot f_0 + Ee^{2i\omega^*\tau^*\theta}\right] \\ &= \frac{\tau^*}{2} \\ &\cdot\begin{pmatrix} -\frac{r_1}{K} - \alpha\zeta \\ \frac{a\gamma Z^*}{\left(\gamma + T^*\right)^3}\xi^2 - \frac{r_2}{K}\xi^2 - \frac{a\gamma}{\left(\gamma + T^*\right)^2}\xi\zeta \\ \beta\zeta - 2d\zeta^2 - \frac{2b\gamma\xi\zeta e^{-i\omega^*\tau^*}}{\left(\gamma + T^*\right)^2} + \frac{2b\gamma Z^*\zeta^2 e^{-2i\omega^*\tau^*}}{\left(\gamma + T^*\right)^3} \end{pmatrix} \\ &- \frac{g_{20}p_1(0) + \overline{g}_{02}p_2(0)}{2}\cdot f_0. \end{aligned} \quad (93)$$

Noticing that

$$\begin{aligned} &\tau^* D\Delta\left[p_1(0)\cdot f_0\right] + L(\tau^*)\left[p_1(\theta)\cdot f_0\right] \\ &\quad = i\omega^*\tau^* p_1(0)\cdot f_0, \\ &\tau^* D\Delta\left[p_2(0)\cdot f_0\right] + L(\tau^*)\left[p_2(\theta)\cdot f_0\right] \\ &\quad = -i\omega^*\tau^* p_2(0)\cdot f_0, \end{aligned} \quad (94)$$

we have

$$\begin{aligned} &2i\omega^*\tau^* E - \tau^* D\Delta E - L(\tau^*)\left(Ee^{2i\omega^*\tau^*\theta}\right) = \frac{\tau^*}{2} \\ &\cdot\begin{pmatrix} -\frac{r_1}{K} - \alpha\zeta \\ \frac{a\gamma Z^*}{\left(\gamma + T^*\right)^3}\xi^2 - \frac{r_2}{K}\xi^2 - \frac{a\gamma}{\left(\gamma + T^*\right)^2}\xi\zeta \\ \beta\zeta - 2d\zeta^2 - \frac{2b\gamma\xi\zeta e^{-i\omega^*\tau^*}}{\left(\gamma + T^*\right)^2} + \frac{2b\gamma Z^*\zeta^2 e^{-2i\omega^*\tau^*}}{\left(\gamma + T^*\right)^3} \end{pmatrix}. \end{aligned} \quad (95)$$

From previous formulas, we can obtain

$$E = \frac{1}{2}\begin{pmatrix} 2i\omega^*\tau^* + \frac{r_1}{K}P^* & 0 & \alpha P^* \\ 0 & 2i\omega^*\tau^* - \frac{aT^*Z^*}{\left(\gamma + T^*\right)^2} + \frac{r_2}{K}P_2^* & \frac{aT^*}{\gamma + T^*} \\ -\beta Z^* & \frac{bZ^*}{\gamma + T^*}e^{-2i\omega^*\tau^*} - \frac{bT^*Z^*}{\left(\gamma + T^*\right)^2} & 2i\omega^*\tau^* + dZ^* \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} -\frac{r_1}{K}-\alpha\zeta \\ \frac{a\gamma Z^*}{(\gamma+T^*)^3}\xi^2-\frac{r_2}{K}\xi^2-\frac{a\gamma}{(\gamma+T^*)^2}\xi\zeta \\ \beta\zeta-2d\zeta^2-\frac{2b\gamma\xi\zeta e^{-i\omega^*\tau^*}}{(\gamma+T^*)^2}+\frac{2b\gamma Z^*\zeta^2 e^{-2i\omega^*\tau^*}}{(\gamma+T^*)^3} \end{pmatrix}. \tag{96}$$

With the same method, we also have

$$\dot{W}_{11}(\theta)=\frac{g_{11}p_1(0)+\overline{g}_{11}p_2(0)}{2}\cdot f_0,\quad -1\le\theta\le 0, \tag{97}$$

and

$$W_{11}(\theta)=\frac{i}{2\omega^*\tau^*}\left[-g_{11}p_1(\theta)+\overline{g}_{11}p_2(\theta)\right]\cdot f_0+F, \tag{98}$$

where

$$F=\frac{1}{4}\begin{pmatrix} \frac{r_1}{K}P^* & 0 & \alpha P^* \\ 0 & \frac{r_2}{K}T^*-\frac{aT^*Z^*}{(\gamma+T^*)^2} & \frac{aT^*}{\gamma+T^*} \\ -\beta Z^* & \frac{b\gamma Z^*}{(\gamma+T^*)^2} & dZ^* \end{pmatrix}^{-1} \times \begin{pmatrix} -\frac{2r_1}{K}-\alpha\left(\zeta+\overline{\zeta}\right) \\ \frac{2a\gamma Z^*}{(\gamma+T^*)^3}\xi\overline{\xi}-\frac{2r_2}{K}\xi\overline{\xi}-\frac{a\gamma}{(\gamma+T^*)^2}\left(\xi\overline{\zeta}+\overline{\xi}\zeta\right) \\ \beta\left(\zeta+\overline{\zeta}\right)-2d\zeta\overline{\zeta}-\frac{b\gamma}{(\gamma+T^*)^2}\left(\xi\overline{\zeta}e^{-i\omega^*\tau^*}+\overline{\xi}\zeta e^{i\omega^*\tau^*}\right)+\frac{4b\gamma Z^*}{(\gamma+T^*)^3}\zeta\overline{\zeta} \end{pmatrix}. \tag{99}$$

At this point, we are able to completely establish the value of g_{21} and thereby can determine the properties of Hopf bifurcation.

System (3) has the following Poincaré normal form:

$$\dot{\varpi}=i\omega^*\tau^*\varpi+C_1(0)\varpi|\varpi|^2+o\left(|\varpi|^5\right), \tag{100}$$

where

$$C_1(0)=\frac{i}{2\omega^*\tau^*}\left(g_{20}g_{11}-2|g_{11}|^2-\frac{|g_{02}|^2}{3}\right)+\frac{g_{21}}{3}. \tag{101}$$

Hence,

$$\begin{aligned} \sigma_2&=-\frac{\mathrm{Re}\left(C_1(0)\right)}{\mathrm{Re}\left(\lambda'(\tau^*)\right)}, \\ \beta_2&=2\mathrm{Re}\left(C_1(0)\right), \\ T_2&=-\frac{\mathrm{Im}\left(C_1(0)\right)+\sigma_2\mathrm{Im}\left(\lambda'(\tau^*)\right)}{\omega^*\tau^*}. \end{aligned} \tag{102}$$

By the Hopf bifurcation theory [25], we know that σ_2 determines the bifurcation direction: if $\sigma_2 > 0$ ($\sigma_2 < 0$), then the Hopf bifurcation is supercritical (subcritical) and the bifurcating periodic solutions exist for $\tau > \tau_0$ ($\tau < \tau_0$); β_2 determines the stability of bifurcating periodic solutions: if $\beta_2 > 0$ ($\beta_2 < 0$), then the periodic solutions are stable (unstable); T_2 determines the monotonicity of the period of periodic solutions: if $T_2 > 0$ ($T_2 < 0$), then the period increases (decreases).

6. Numerical Simulation

In this section, we conduct the numerical simulations with the help of MATLAB. We first choose the following parameter value:

$$\begin{aligned} d_1&=d_2=0.5, \\ d_3&=2, \\ \Omega&=(0,\pi), \\ r_1&=5, \end{aligned}$$

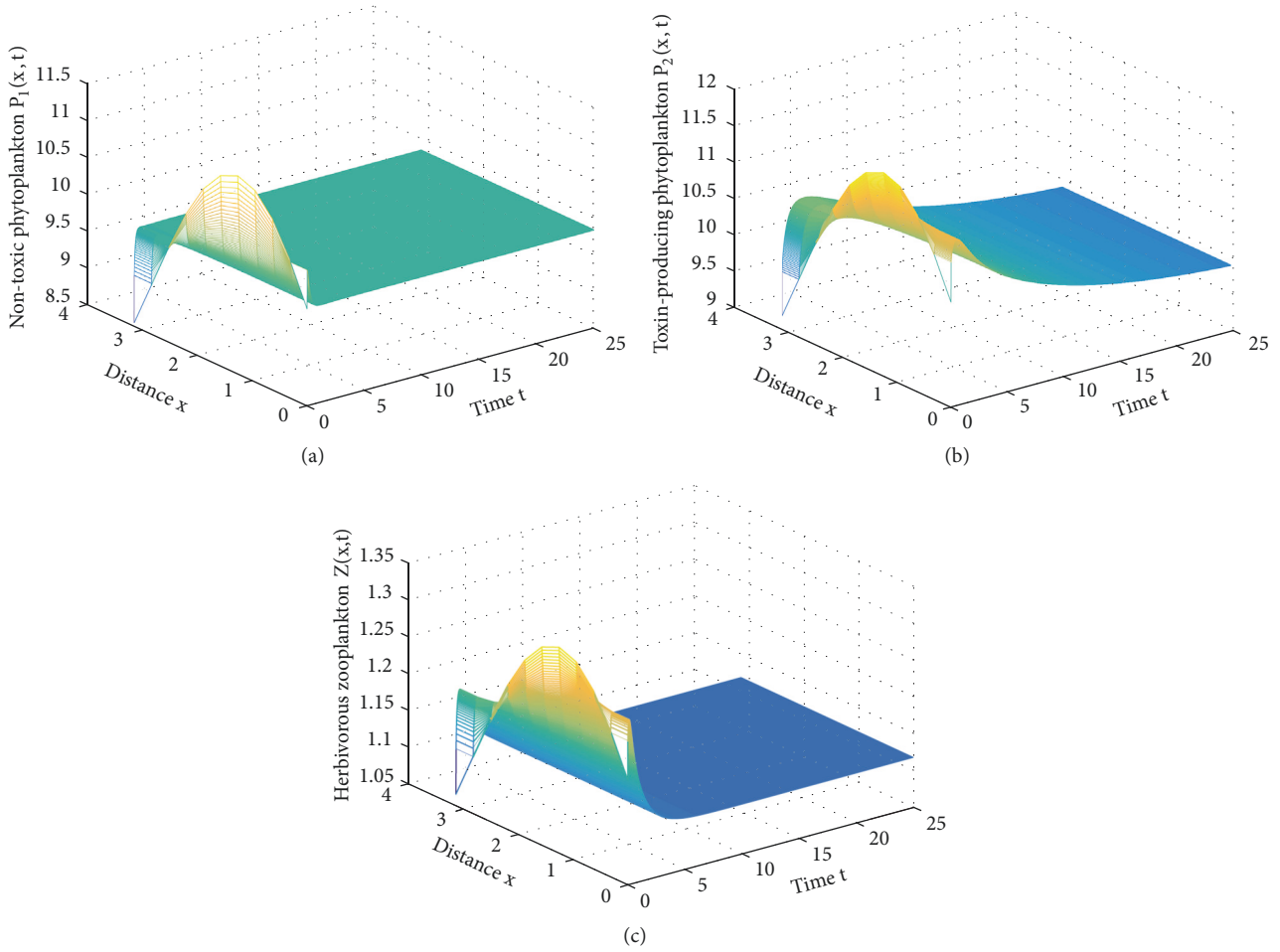


FIGURE 1: The equilibrium solution E^* of (3) is asymptotically stable when $\tau = 0$.

$$
\begin{aligned}
r_2 &= 0.2,\\
K &= 10,\\
\alpha &= 0.068,\\
\beta &= 0.08,\\
a &= 0.02,\\
b &= 0.005,\\
\gamma &= 0.5,\\
d &= 0.7.
\end{aligned}
\tag{103}
$$

By direct calculation, we have that system (3) has the unique positive equilibrium solution $E^* = (9.84786, 9.78241, 1.11867)$ and the corresponding Hopf bifurcation value is $\tau_0 \approx 22.3146$.

Figures 1–3 show that the constant equilibrium solution of (3) is asymptotically stable when the time delay is zero or appropriately small. On the other hand, once the time delay is larger than the critical value τ_0, the equilibrium solution would no longer be stable and spatially homogeneous periodic solution will bifurcate at the equilibrium solution (see Figure 4).

We reselect $d_1 = d_2 = 0.0005$, $d_3 = 0.005$, $d = 0.1$, and keep other coefficients the same. From Figures 5-6, it is shown that the equilibrium solution E^* is asymptotically stable when $\tau = 2.5$ and spatially inhomogeneous periodic solution exists when $\tau = 5.42$. It confirms that both spatial diffusion and toxin delay have significant effects on the spatiotemporal dynamics of system (3). Besides, if we set $\Omega = (0, 2\pi)$, then the spatially inhomogeneous periodic solution vanishes and the equilibrium solution becomes asymptotically stable (see Figure 7). It can be concluded that spatially inhomogeneous pattern is more prone to occurring in small space.

7. Conclusions

In this paper, we have proposed a delayed reaction-diffusion model incorporating three plankton populations. Mathematical analysis indicates that the three populations can coexist

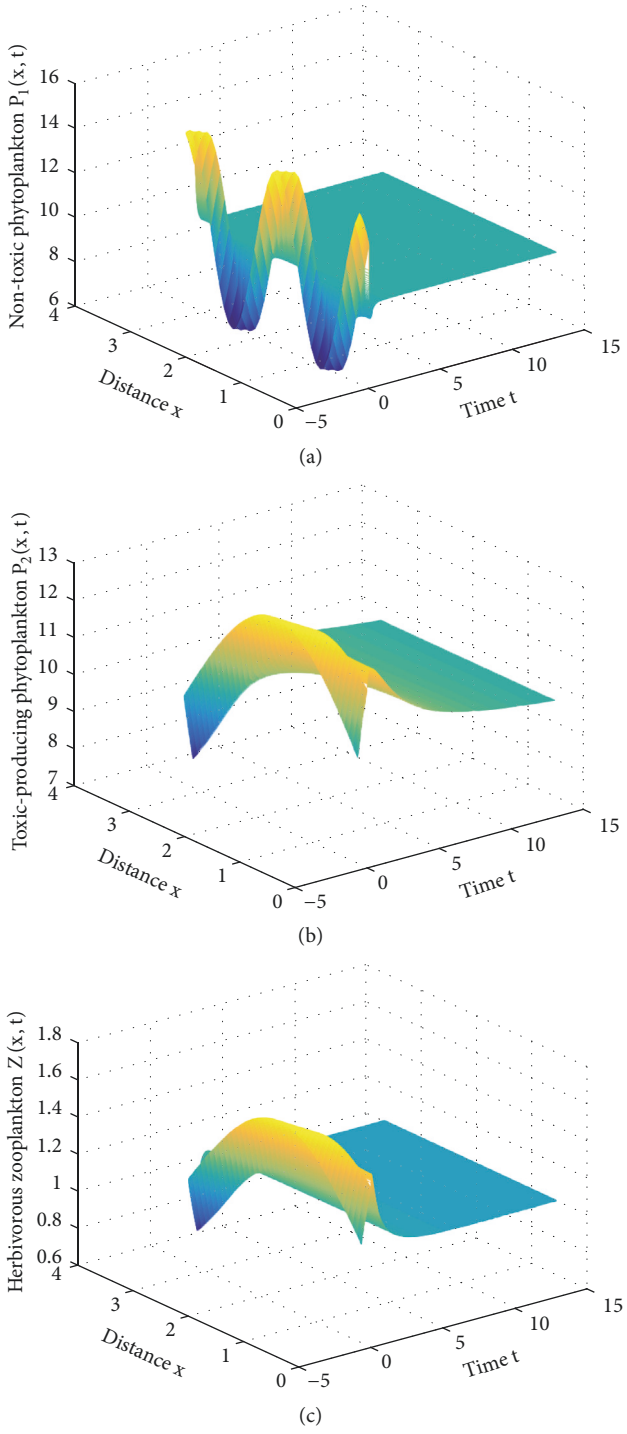


FIGURE 2: The equilibrium solution E^* of (3) is asymptotically stable when $\tau = 0.65 < \tau_0$.

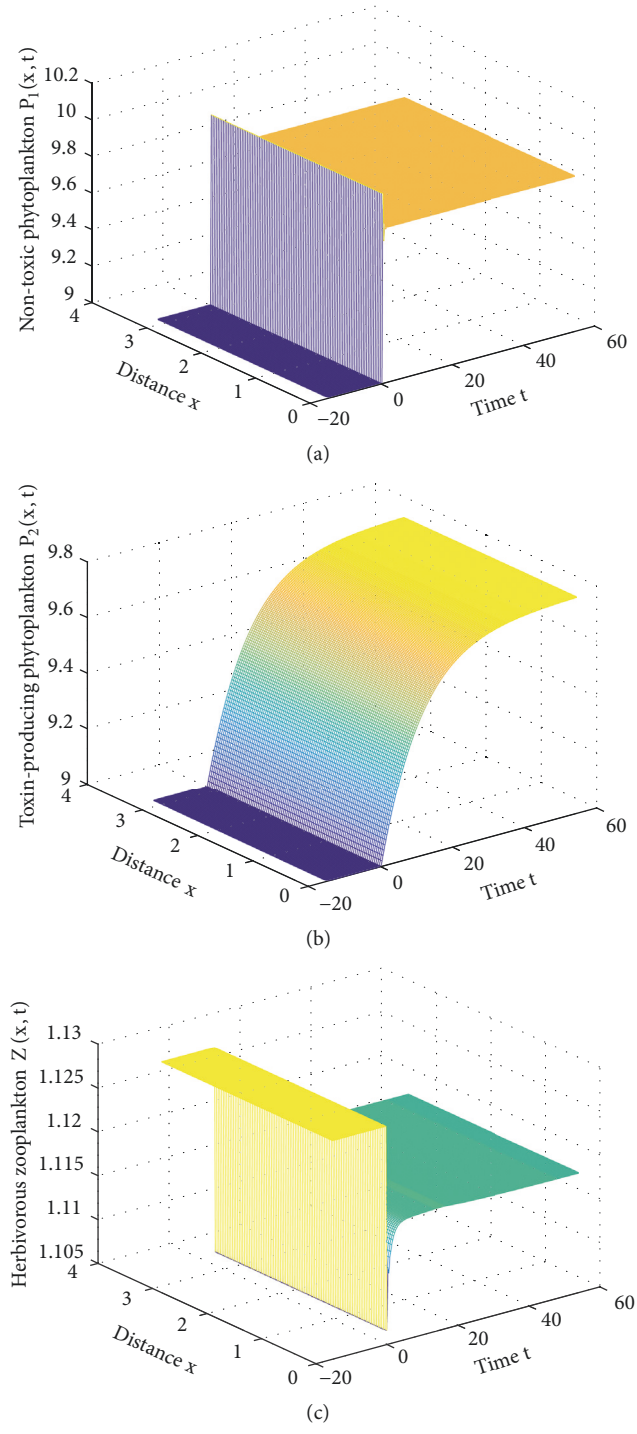


FIGURE 3: The equilibrium solution E^* of (3) is asymptotically stable when $\tau = 15 < \tau_0$.

when the natural growth rate of nontoxic phytoplankton is large and the death rate of zooplankton by toxin is small. In this case, the balance is finally achieved by interdependence and mutual restraint. Otherwise, the zooplankton will become extinct if the natural growth rate of nontoxic phytoplankton is small and the death rate of zooplankton by toxin is large. In the latter case, the biomasses of nontoxic phytoplankton and toxin-producing phytoplankton reach saturation and algae bloom occurs.

And when considering the toxin delay, the spatiotemporal dynamics of system is almost unaffected when the time delay is sufficiently small. However, when the time delay

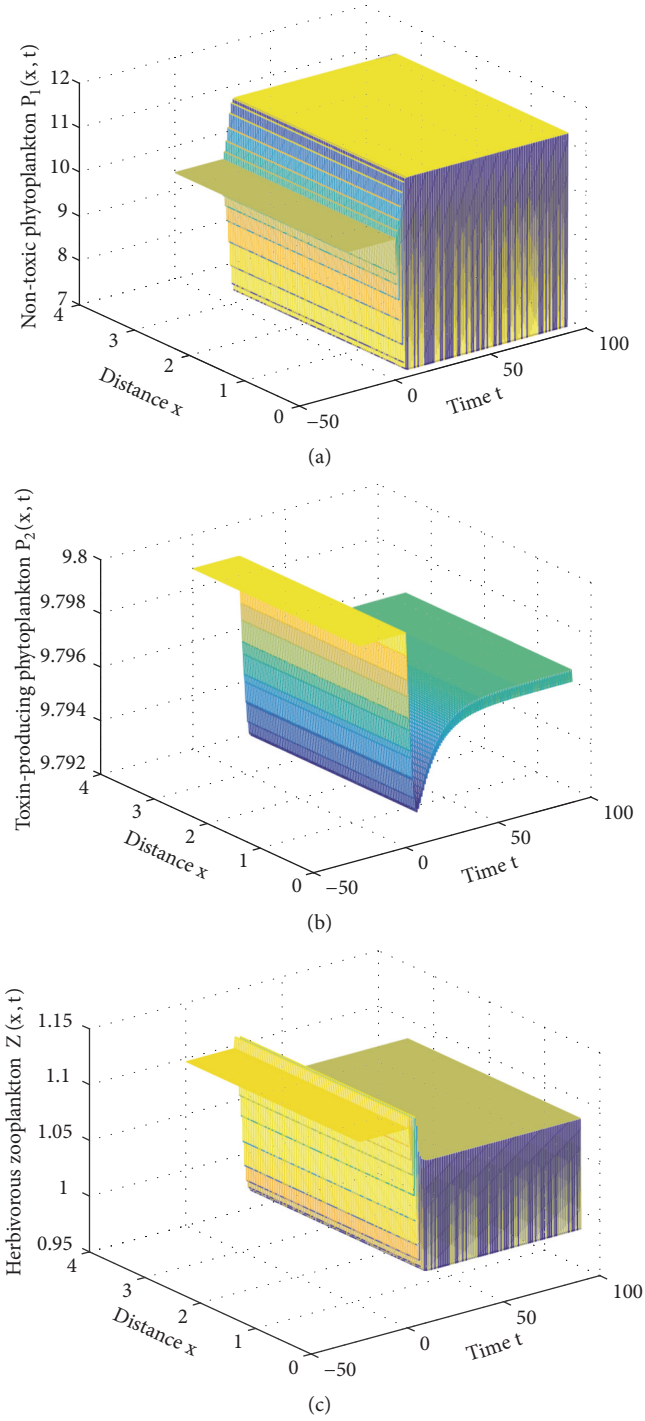


FIGURE 4: The spatially homogeneous periodic solution exists when $\tau = 25 > \tau_0$.

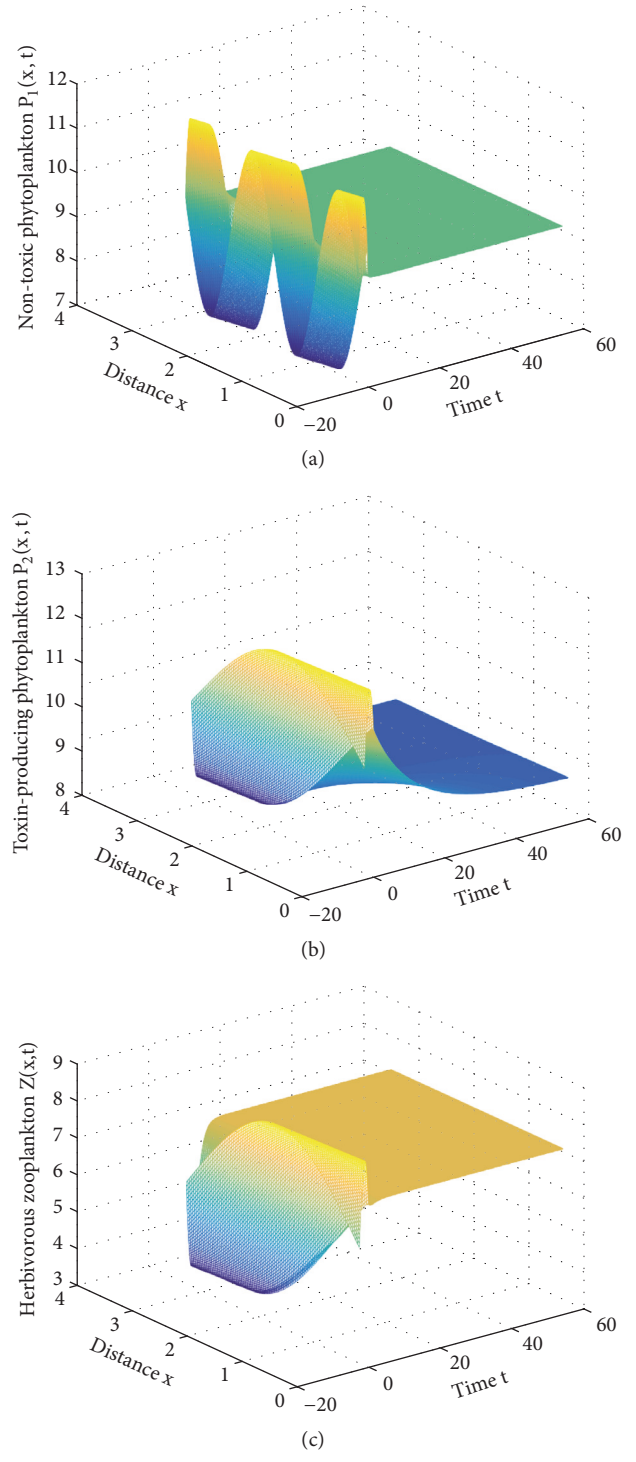


FIGURE 5: The equilibrium solution E^* of (3) is asymptotically stable when $\tau = 2.5$ with new coefficient values.

passes through some critical value, spatially homogeneous or inhomogeneous periodic solution may arise. This means that algal bloom erupts periodically under certain conditions. Therefore, some measures can be adopted to control or defer the occurrence of Hopf bifurcation, such as reducing the toxin delay or adding feedback control. Numerical simulations complementally illustrate that when there is no or only tiny time delay, the population distribution pattern is eventually spatially homogeneous even if the initial distribution is inhomogeneous. When the time delay is sufficiently large, the distribution pattern is time periodic. Moreover,

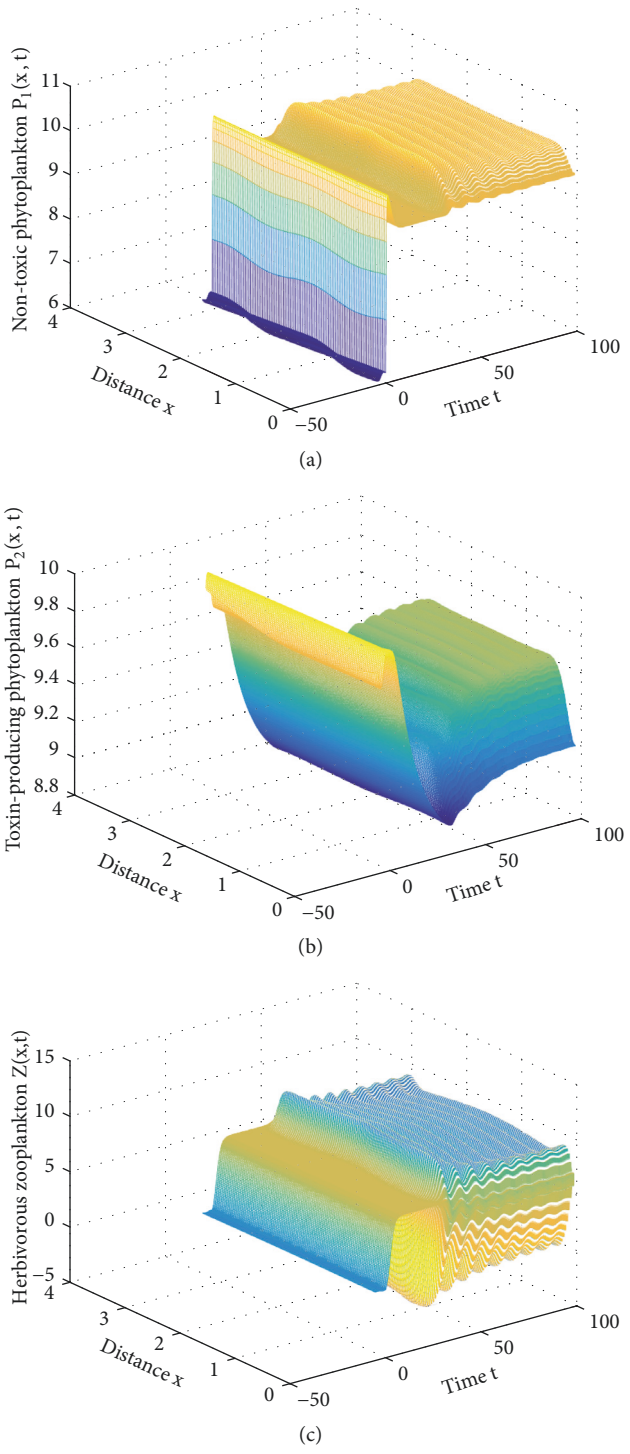


FIGURE 6: The spatially inhomogeneous periodic solution exists when $\tau = 5.42$ with new coefficient values.

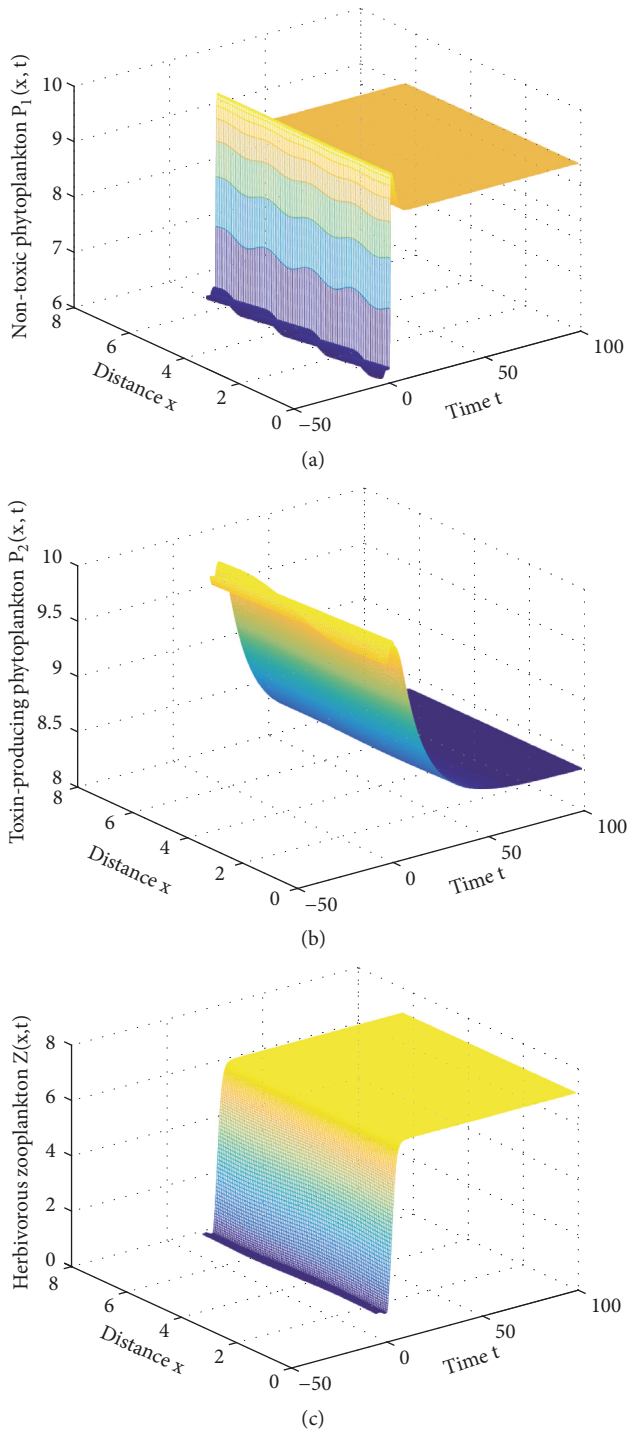


FIGURE 7: The equilibrium solution E^* of (3) is asymptotically stable when $\tau = 5.42$ and $\Omega = (0, 2\pi)$.

nontoxic phytoplankton has the largest oscillation amplitude, toxin-producing phytoplankton has the smallest oscillation amplitude, and the zooplankton has modest amplitude. These phenomena show that toxin delay has the greatest effects on nontoxic phytoplankton.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (61703001), the MOE Project of Humanities and Social Sciences of China (17YJC630175), the Natural Science Research Projects of Universities in Anhui Province (KJ2017ZD35, KJ2017A432), and Anhui Provincial Natural Science Foundation (1508085MA09).

References

[1] G. C. Hays, A. J. Richardson, and C. Robinson, "Climate change and marine plankton," *Trends in Ecology & Evolution*, vol. 20, no. 6, pp. 337–344, 2005.

[2] J. P. Connolly and R. B. Coffin, "Model of carbon cycling in planktonic food webs," *Journal of Environmental Engineering*, vol. 121, no. 10, pp. 682–690, 1995.

[3] A. M. Edwards and J. Brindley, "Zooplankton mortality and the dynamical behaviour of plankton population models," *Bulletin of Mathematical Biology*, vol. 61, no. 2, pp. 303–339, 1999.

[4] B. Mukhopadhyay and R. Bhattacharyya, "Role of gestation delay in a plankton-fish model under stochastic fluctuations," *Mathematical Biosciences*, vol. 215, no. 1, pp. 26–34, 2008.

[5] S. Rana, S. Samanta, S. Bhattacharya, K. Al-Khaled, A. Goswami, and J. Chattopadhyay, "The effect of nanoparticles on plankton dynamics: a mathematical model," *BioSystems*, vol. 127, pp. 28–41, 2015.

[6] J. Chattopadhyay, R. R. Sarkar, and S. Pal, "Mathematical modelling of harmful algal blooms supported by experimental findings," *Ecological Complexity*, vol. 1, no. 3, pp. 225–235, 2004.

[7] R. R. Sarkar, S. Pal, and J. Chattopadhyay, "Role of two toxin-producing plankton and their effect on phytoplankton-zooplankton system—a mathematical study supported by experimental findings," *BioSystems*, vol. 80, no. 1, pp. 11–23, 2005.

[8] S. Roy, S. Alam, and J. Chattopadhyay, "Competing effects of toxin-producing phytoplankton on overall plankton populations in the bay of bengal," *Bulletin of Mathematical Biology*, vol. 68, no. 8, pp. 2303–2320, 2006.

[9] S. Pal, S. Chatterjee, K. Das, and J. Chattopadhyay, "Role of competition in phytoplankton population for the occurrence and control of plankton bloom in the presence of environmental fluctuations," *Ecological Modelling*, vol. 220, no. 2, pp. 96–110, 2009.

[10] D. W. Huang, H. L. Wang, J. F. Feng, and Z. W. Zhu, "Modelling algal densities in harmful algal blooms (HAB) with stochastic dynamics," *Applied Mathematical Modelling*, vol. 32, no. 7, pp. 1318–1326, 2008.

[11] A. Chatterjee and S. Pal, "Plankton nutrient interaction model with effect of toxin in presence of modified traditional holling type II functional response," *Systems Science and Control Engineering: An Open Access Journal*, vol. 4, no. 1, pp. 20–30, 2016.

[12] H. Malchow, "Nonlinear plankton dynamics and pattern formation in an ecohydrodynamic model system," *Journal of Marine Systems*, vol. 7, no. 2-4, pp. 193–202, 1996.

[13] A. B. Medvinsky, S. V. Petrovskii, I. A. Tikhonova, H. Malchow, and B. L. Li, "Spatiotemporal complexity of plankton and fish dynamics," *SIAM Review*, vol. 44, no. 3, pp. 311–370, 2002.

[14] K. Chakraborty, K. Das, and T. K. Kar, "Modeling and analysis of a marine plankton system with nutrient recycling and diffusion," *Complexity*, vol. 21, no. 1, pp. 229–241, 2014.

[15] K. J. Richards, "Viral infections of oceanic plankton blooms," *Journal of Theoretical Biology*, vol. 412, pp. 27–35, 2014.

[16] R. Yang, M. Liu, and C. Zhang, "A diffusive toxin producing phytoplankton model with maturation delay and three-dimensional patch," *Computers & Mathematics with Applications*, vol. 73, no. 5, pp. 824–837, 2017.

[17] T. H. Zhang, X. Liu, X. Z. Meng, and T. Q. Zhang, "Spatio-temporal dynamics near the steady state of a planktonic system," *Computers & Mathematics with Applications*, vol. 75, no. 12, pp. 4490–4504, 2018.

[18] J. T. Zhao and J. J. Wei, "Dynamics in a diffusive plankton system with delay and toxic substances effect," *Nonlinear Analysis: Real World Applications*, vol. 22, pp. 66–83, 2015.

[19] Z. C. Jiang and T. Q. Zhang, "Dynamical analysis of a reaction-diffusion phytoplankton-zooplankton system with delay," *Chaos, Solitons & Fractals*, vol. 104, pp. 693–704, 2002.

[20] J. F. Zhang, S. L. Wang, and X. J. Kong, "Effects of toxin delay on the dynamics of a phytoplankton–zooplankton model," *Physica A*, vol. 505, pp. 1150–1162, 2018.

[21] A. M. Edwards and A. Yool, "The role of higher predation in plankton population models," *Journal of Plankton Research*, vol. 22, no. 6, pp. 1085–1112, 2000.

[22] A. M. Edwards and M. A. Bees, "Generic dynamics of a simple plankton population model with a non-integer exponent of closure," *Chaos, Solitons & Fractals*, vol. 12, no. 2, pp. 289–300, 2001.

[23] J. T. Zhao and J. J. Wei, "Persistence, turing instability and hopf bifurcation in a diffusive plankton system with delay and quadratic closure," *International Journal of Bifurcation and Chaos*, vol. 26, no. 3, Article ID 1650047, 2016.

[24] Y. L. Song, M. A. Han, and J. J. Wei, "Stability and Hopf bifurcation analysis on a simplified BAM neural network with delays," *Physica D: Nonlinear Phenomena*, vol. 200, no. 3-4, pp. 185–204, 2005.

[25] B. D. Hassard, N. Kazarinoff, and Y. Yan, *Theory and Applications of Hopf Bifurcation*, Cambridge University Press, England, UK, 1981.

[26] J. H. Wu, *Theory and Applications of Partial Functional-Differential Equations*, Springer, New York, NY, USA, 1996.