Nonlinear Analysis: Hybrid Systems 49 (2023) 101368

Contents lists available at ScienceDirect

Nonlinear Analysis: Hybrid Systems

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

Hybrid stochastic epidemic sir models with hidden states

Nguyen Du [a,1,2], Alexandru Hening [b,1,3], Nhu Nguyen [c,1,4], George Yin [d,1,5,*]

[a] Department of Mathematics-Mechanics-Informatics, Vietnam National University of Science, 34 Nguyen Trai, Thanh Xuan, Ha Noi, Viet Nam
[b] Department of Mathematics, Texas A&M University, Mailstop 3368, College Station, TX 77843-3368, United States
[c] Department of Mathematics and Applied Mathematical Sciences, University of Rhode Island, 5 Lippitt Road, Suite 200, Kingston, RI 02881-2018, United States
[d] Department of Mathematics, University of Connecticut, 341 Mansfield Road U1009, Storrs, CT 06269-1009, United States

ARTICLE INFO

Article history:
Received 8 August 2022
Received in revised form 16 February 2023
Accepted 11 April 2023
Available online xxxx

Keywords:
SIR model
Extinction
Permanence
Stationary distribution
Ergodicity

ABSTRACT

This paper focuses on realistic hybrid SIR models that take into account stochasticity. The proposed systems are applicable to most incidence rates that are used in the literature including the bilinear incidence rate, the Beddington–DeAngelis incidence rate, and a Holling type II functional response. Given that many diseases can lead to asymptomatic infections, this paper looks at a system of stochastic differential equations that also includes a class of hidden state individuals, for which the infection status is unknown. Assuming that the direct observation of the percentage of hidden state individuals being infected, $\alpha(t)$, is not given and only a noise-corrupted observation process is available. Using nonlinear filtering techniques in conjunction with an invasion type analysis, this paper shows that the long-term behavior of the disease is governed by a threshold $\lambda \in \mathbb{R}$ that depends on the model parameters. It turns out that if $\lambda < 0$ the number $I(t)$ of infected individuals converges to zero exponentially fast (extinction). However, if $\lambda > 0$, the infection is endemic and the system is persistent. We showcase our theorems by applying them in some illuminating examples.

1. Introduction

The SIR epidemic models introduced first by [1,2] look at the dynamics of susceptible, infected, and recovered individuals, whose densities at the time t are denoted by $S(t)$, $I(t)$, and $R(t)$, respectively. In the absence of random effects, the dynamics are described by the following system of differential equations

$$\begin{cases} dS(t) = \big[a_1 - \mu_S S(t) - F(S(t), I(t))\big]dt, \\ dI(t) = \big[-(\mu_I + r)I(t) + F(S(t), I(t))\big]dt, \\ dR(t) = \big[-\mu_R R(t) + rI(t)\big]dt. \end{cases} \tag{1.1}$$

* Corresponding author.
E-mail addresses: nhdu@viasm.edu.vn (N. Du), ahening@tamu.edu (A. Hening), nhu.nguyen@uri.edu (N. Nguyen), gyin@uconn.edu (G. Yin).

[1] The authors made equal contributions to the paper.

[2] The research of this author was supported in part by the Vietnam Foundation for Science and Technology Development of Vietnam's Ministry of Education and Training under grant B2023-TDV-02.
[3] The research of this author was supported in part by the National Science Foundation under grant DMS 2147903.
[4] The research of this author was supported in part by the AMS-Simons travel grant.
[5] The research of this author was supported in part by the National Science Foundation under grant DMS-2204240.

https://doi.org/10.1016/j.nahs.2023.101368
1751-570X/

Here $a_1 > 0$ is the recruitment rate of the population, $\mu_S, \mu_I, \mu_R > 0$ are the death rates of the susceptible, infected, and recovered individuals, $r > 0$ is the recovery rate of the infected individuals and $F(S(t), I(t))$ is the incidence rate. The dynamics of recovered individuals has no effect on that of the disease transmission. As a result, it is common not to consider the recovered individuals as part of the problem formulation. We adopt this practice throughout this paper. Various types of incidence rates $F(S, I)$ have been considered in the literature, for example, the Holling type II functional response $F(S, I) = \frac{\beta SI}{m_1+S}$ [3], the bilinear functional response $F(S, I) = \beta SI$ [4,5], the nonlinear functional response $F(S, I) = \frac{\beta SI^l}{1+m_2 I^h}$ [6–8], and the Beddington–DeAngelis functional response $F(S, I) = \frac{\beta SI}{1+m_1 S+m_2 I}$ [9–11].

It is by now widely known that in order to have a realistic model, one cannot ignore random environmental fluctuations (temperature, climate, water resources, etc.). This fact leads to extensive studies of stochastic epidemic models of the form

$$\begin{cases} dS(t) = \big[a_1 - b_1 S(t) - I(t)f(S(t), I(t))\big]dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \big[-b_2 I(t) + I(t)f(S(t), I(t))\big]dt + \sigma_2 I(t)dB_2(t), \end{cases} \tag{1.2}$$

where $B_1(t)$ and $B_2(t)$ are independent Brownian motions and $\sigma_1, \sigma_2 \neq 0$ are the noise intensities. Moreover, in the above we have rewritten the coefficients: $b_1 := \mu_S$, $b_2 := \mu_I + r$ and $F(S, I) = If(S, I)$ (compare this to (1.1)). This system has been analyzed in a general setting in [12].

It is well-known that there are diseases for which certain infected individuals are asymptomatic. Covid-19 is one such example — there have been many reports of infections where the infected exhibit no symptoms. We aim to capture this type of behavior in our model. In order to do this, we assume that the group of infected individuals that has incidence rate $I(t)f(S(t), I(t))$ (the rate that describes how the disease spreads from infected groups to susceptible groups) in the classical setting, now contains 2 sub-groups. The first group contains individuals who have been confirmed to be infected and with the incidence rate $I(t)f_1(S(t), I(t))$ (we will still denote this incidence rate by $I(t)f(S(t), I(t))$ for notational simplicity). The second group has incidence rate $I(t)h(S(t), I(t))$ and contains people whose infection status is unknown or hidden. Let $\alpha(t)$ be a Markov process taking values in $\mathcal{M} = [0, 1]$. We suppose that $\alpha(t)$ represents the percentage of individuals in the hidden-status class that are infected at time $t \geq 0$ and that only noise-corrupted observations of $\alpha(t)$ are available. More specifically, one can only observe $\alpha(t)$ with additive white noise.

It is natural to assume that the hidden status of potentially infected individuals affects the spread of the disease. As a result, we let the functions f and h depend on $\alpha(t)$. With the hidden state dynamics in (1.2), we obtain the following hybrid stochastic epidemic system

$$\begin{cases} dS(t) = \Big[a_1 - b_1 S(t) - I(t)f(\alpha(t), S(t), I(t)) - \alpha(t)I(t)h(\alpha(t), S(t), I(t))\Big]dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \Big[-b_2 I(t) + I(t)f(\alpha(t), S(t), I(t)) + \alpha(t)I(t)h(\alpha(t), S(t), I(t))\Big]dt + \sigma_2 I(t)dB_2(t). \end{cases} \tag{1.3}$$

Remark 1. One can understand the dynamics by looking at individuals from group S, in which susceptible individuals are infected at the rate $If(S, I)$. We assume that $\alpha(t)$ percent of the potentially infected individuals are actually infected. Then we can say that members of the hidden group infect susceptible individuals at the rate $\alpha Ih(S, I)$.

Remark 2. We could combine $I(t)f(\alpha(t), S(t), I(t)) + \alpha(t)I(t)h(\alpha(t), S(t), I(t))$ to produce a new function. However, we choose the current setup to make the formulation and motivation clear. Moreover, this will also be more convenient for later discussions.

Our results can be summarized as follows. Because the infection status of certain individuals is hidden, and $\alpha(t)$ is not directly available, the dynamics of (1.3) are difficult to study. To overcome the difficulty, we use nonlinear filtering techniques by considering the conditional distribution of the process $\alpha(t)$ given the observations. This enables us to replace the hidden Markov process $\alpha(t)$ in (2.3) with the corresponding conditional distribution. We start by studying the well-posedness of the equation under consideration together with the positivity of solutions, the Markov–Feller property, and some moment estimates. Next, we study the long-term behavior of the system. Under the assumption for ergodicity of nonlinear filtering [13,14] and using ideas from dynamical systems, by considering the boundary equation and growth rate (see e.g., [4,12,15]), we are able to prove that there is a threshold λ such that if $\lambda < 0$, the number of the infected individuals $I(t)$ tends to zero exponentially fast and if $\lambda > 0$, all invariant probability measures of the system concentrate on $\mathbb{R}^{2,\circ}_+ := (0, \infty)^2$, and then the system is permanent. We show that the threshold λ also characterizes the permanence and extinction of the original (hidden) system (1.3). We also study the case when the process $\alpha(t)$ is a hidden Markov chain taking values in a finite set. Next, we demonstrate our results using simple examples and numerical simulations.

The rest of the paper is organized as follows. We give the mathematical formulation of our problem in Section 2. Section 3 introduces the threshold λ. The sign of λ will be used to characterize the longtime behavior of the underlying system. Section 4 is devoted to the characterization of the longtime dynamics of our system. Section 5 offers some interpretations and implications of our results. Finally, Section 6 provides some simple examples and simulations to illustrate our theoretical results.

2. Problem formulation

Throughout this paper we use $\mathbb{R}_+ := [0, \infty)$, $\mathbb{R}_+^\circ := (0, \infty)$, $\mathbb{R}_+^2 := [0, \infty) \times [0, \infty)$, and $\mathbb{R}_+^{2,\circ} := (0, \infty) \times (0, \infty)$. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions, and $B_1(t)$, $B_2(t)$, and $W(t)$ be mutually independent standard Brownian motions. The process $\alpha(t)$ (termed a signal process) is assumed to be an adapted stochastic process taking values in [0, 1] that is independent of $B_1(t)$, $B_2(t)$, and $W(t)$. Moreover, $\mathcal{M}$ will denote the space of probability measures on ([0, 1], $\mathcal{B}$([0, 1])) endowed with the weak topology, and C[0, 1] the spaces of all real-valued continuous functions on [0, 1]. For any function $l \in C$[0, 1] and $\mu \in \mathcal{M}$, set

$$\mu(l) := \int_0^1 l(x)\mu(dx).$$

As discussed above, we consider the setting where the precise values $\alpha(t)$ are not available and only noisy observations are available. The observation process $y(t)$ of the signal process $\alpha(t)$ is given by

$$dy(t) = g(\alpha(t))dt + dW(t), \quad y(0) = 0, \tag{2.1}$$

where $g : [0, 1] \to \mathbb{R}$ is a continuous function. Let $\mathcal{F}_t^y := \sigma\{y(s) : 0 \leq s \leq t\} \bigvee \sigma(\alpha(0))$, where $\bigvee$ denotes the smallest σ-algebra generated by the union of some σ-algebras. Let $\Pi_t(\cdot) \in \mathcal{M}$ be the conditional distribution of the signal process $\alpha(t)$ given the observation $y(t)$ and the initial data, i.e.,

$$\Pi_t(A) = \mathbb{P}[\alpha(t) \in A|\mathcal{F}_t^y], A \in \mathcal{B}([0, 1]).$$

Such $\{\Pi_t(\cdot)\}$ is called nonlinear filtering.

The field of nonlinear filtering has a long history. The main idea stems from replacing the unknown state by its conditional distributions. The earliest result was the well-known Kushner's equation [16]. Subsequently, the Duncan–Mortensen–Zakai equation came into being [17–19]. In this paper, we make use of the version of filtering developed by Fujisaki–Kallianpur–Kunita [20]. We will also make use of the Wonham filter for hidden Markov chains, which is one of the handful finite-dimensional filters in existence [21].

To proceed, we next detail the results of Fujisaki–Kallianpur–Kunita [20] (see also [22]), which involve a differential equation for the nonlinear filtering $\Pi_t(\cdot)$. Define

$$\beta(t) = y(t) - \int_0^t \Pi_s(g)ds,$$

and note that the process $\beta(t)$ is a one-dimensional Wiener process, see e.g., [23, Theorem 7.2]. Moreover, $\sigma\{\beta(t_2)-\beta(t_2) : t_2 \geq t_1 \geq t\}$ and $\mathcal{F}_t^y$ are independent for all $t \geq 0$. If the signal process $\alpha(t)$ is a Markov process with infinitesimal generator A, then $\Pi_t(\cdot)$ is the solution of

$$\Pi_t(l) = \Pi_0(l) + \int_0^t \Pi_s(Al)ds + \int_0^t (\Pi_s(lg) - \Pi_s(l)\Pi_s(g))\, d\beta(s), \quad \forall l \in \mathcal{D}(A). \tag{2.2}$$

The interested reader is referred to the detailed analysis given in [20,22].

We will not make use of (2.2) often in our analysis, except for establishing some preliminary properties. The stochastic differential equation for $\Pi_t(\cdot)$ is rather complex and is not the main concern of the current paper. As will be seen in the next section, we only need to establish the related ergodicity. Thus, for us, it suffices to consider $\Pi_t(\cdot)$ as a stochastic process taking values in $\mathcal{M}$. In addition, we use the continuous measurable modification of $\Pi_t(\cdot)$; it is well known that such a modification always exists [14].

Under the premise that one only observes a noisy version of $\alpha(t)$, we proceed to study system (1.3) by using the nonlinear filter $\{\Pi_t(\cdot)\}$ with given the information of the observation process $y(t)$. More precisely, we consider the system

$$\begin{cases} dS(t) = \Big[a_1 - b_1 S(t) - \int_0^1 I(t) f(x, S(t), I(t))\Pi_t(dx) \\ \qquad\qquad - I(t)\int_0^1 x h(x, S(t), I(t))\Pi_t(dx) \Big]\, dt + \sigma_1 S(t) dB_1(t), \\ dI(t) = \Big[-b_2 I(t) + \int_0^1 I(t) f(x, S(t), I(t))\Pi_t(dx) \\ \qquad\qquad + I(t)\int_0^1 x h(x, S(t), I(t))\Pi_t(dx) \Big]\, dt + \sigma_2 I(t) dB_2(t), \end{cases} \tag{2.3}$$

where

$$\Pi_t(A) = \mathbb{P}[\alpha(t) \in A|\mathcal{F}_t^y], \forall A \in \mathcal{B}([0, 1]),$$
$$dy(t) = g(\alpha(t))dt + dW(t), \quad y(0) = 0.$$

Denote by $\mathbb{P}_{u,v,\pi}$ and $\mathbb{E}_{u,v,\pi}$ the probability and expectation corresponding to the initial values $S(0) = u$, $I(0) = v$, $\Pi_0 = \pi$, and the distribution of $\alpha(0)$, respectively. We next make some assumptions that will be used throughout this paper.

Assumption 2.1. *The following conditions hold:*

- *The function* $f : [0, 1]\times\mathbb{R}^2_+ \to \mathbb{R}_+$ *is nonnegative,* $f(x, 0, i) = 0$, $\forall x \in [0, 1]$, $i \geq 0$. *Furthermore,* f *is Lipschitz continuous, i.e., there exists a positive constant* L_1 *such that for all* $x_1, x_2 \in [0, 1]$, $s_1, s_2, i_1, i_2 \geq 0$

$$|f(x_1, s_1, i_1) - f(x_2, s_2, i_2)| \leq L_1(|x_1 - x_2| + |s_1 - s_2| + |i_1 - i_2|).$$

- *The function* $h : [0, 1] \times \mathbb{R}^2_+ \to \mathbb{R}_+$ *satisfies* $h(x, 0, i) = 0$, *and is Lipschitz continuous with Lipschitz constant* L_2, *i.e., for all* $s_1, s_2, i_1, i_2 \geq 0$, $x_1, x_2 \in [0, 1]$,

$$|h(x_1, s_1, i_1) - h(x_2, s_2, i_2)| \leq L_2(|x_1 - x_2| + |s_1 - s_2| + |i_1 - i_2|).$$

- *For each* $x \in [0, 1]$, *functions* $h(x, \cdot, 0)$ *and* $f(x, \cdot, 0)$ *are non-decreasing.*

Remark 3. We note that almost all of the incidence rate functions used in the literature (such as the bilinear incidence rate, the Beddington–DeAngelis incidence rate, the Holling type II functional response, etc.) satisfy these conditions. *Recall that the incidence rate in our setting is* $If(S, I)$ *rather than* $f(S, I)$.

The third condition is imposed because the incidence rate and the growth rate of the hidden class should increase when $S(t)$ and $I(t)$ increase. Since we have rewritten these rates as $If(S, I)$ and $Ih(\alpha, S, I)$, only an increasing condition on S is assumed.

Assumption 2.2. *The signal process* $\alpha(t)$ *is a Markov–Feller process that has a unique invariant measure* μ^* *and*

$$\lim_{t\to\infty} \|P(t, x, \cdot) - \mu^*(\cdot)\|_{TV} = 0,$$

where $P(t, x, \cdot)$ *is the transition probability and* $\|\cdot\|_{TV}$ *is the total variation norm.*

Remark 4. This assumption is needed to guarantee the ergodicity of the nonlinear filtering as discussed in Section 3.1 later. Using this, we can then define a threshold that fully characterizes the longtime behavior of the underlying system; see Section 3.2.

For $V(s, i) : \mathbb{R}^2 \to \mathbb{R}$, define the operator $\mathcal{L}V$ by

$$\begin{aligned}\mathcal{L}V[s, i, \pi] =& \frac{\partial V}{\partial s}\Big[a_1 - b_1 s - if(s, i) - i\int_0^1 xh(x, s, i)\pi(dx)\Big]\\ &+ \frac{\partial V}{\partial i}\Big[-b_2 i + if(s, i) + i\int_0^1 xh(x, s, i)\pi(dx)\Big]\\ &+ \frac{\sigma_1^2 s^2}{2}\frac{\partial^2 V}{\partial s^2} + \frac{\sigma_2^2 i^2}{2}\frac{\partial^2 V}{\partial i^2}.\end{aligned}$$

In the above, $(s, i, \pi) \in \mathbb{R}^2 \times \mathcal{M}$ represents the variable of $\mathcal{L}V$ rather than that of V.

Discrete state space and Wonham filter. If the Markov process $\alpha(t)$ takes values in a finite space $\{m_1, \ldots, m_{n^*}\} \subset [0, 1]$ and has generator $\{q_{ik}\}_{i,k\in\{1,\ldots,n^*\}}$, the formulation will be simpler and more explicit. We can formulate the problem as follows. Let

$$\begin{aligned}&e_k(t) := \mathbb{P}(\alpha(t) = m_k|\mathcal{F}_t^y) = \mathbb{E}[\mathbf{1}_{\{\alpha(t)=m_k\}}|\mathcal{F}_t^y], k = 1, \ldots, n^*,\\ &e(t) = (e_1(t), \ldots, e_{n^*}(t)),\\ &\mathcal{S}_{n^*} := \left\{e = (e_1, \ldots, e_{n^*}) \in \mathbb{R}^{n^*} : e_k \geq 0, \sum_{k=1}^{n^*} e_k = 1\right\},\\ &g_k := g(m_k), k = 1, \ldots, n^*, \quad \bar{g}(e) := \sum_{k=1}^{n^*} g_k e_k, e = (e_1, \ldots, e_{n^*}) \in \mathcal{S}_{n^*}.\end{aligned}$$

It was shown in [21] that the posterior probability $e(t)$ satisfies the following system of stochastic differential equations

$$\begin{cases} de_k(t) = \left[\sum_{i=1}^{n^*} q_{ik}e_i(t) - (g_k - \bar{g}(e(t)))\bar{g}(e(t))e_k(t)\right]dt + (g_k - \bar{g}(e(t)))e_k(t)dy(t), \ k = 1, \ldots, n^*,\\ e_k(0) = e_k^0, \quad k = 1, \ldots, n^*.\end{cases} \tag{2.4}$$

In this case, instead of considering system (2.3), one can study the following system of stochastic differential equation

$$\begin{cases} dS(t) = \Big[a_1 - b_1 S(t) - I(t)\sum_{k=1}^{n^*} f(m_k, S(t), I(t))e_k(t) \\ \qquad\qquad -I(t)\sum_{k=1}^{n^*} m_k h(m_k, S(t), I(t))e_k(t) \Big] dt + \sigma_1 S(t) dB_1(t), \\ dI(t) = \Big[-b_2 I(t) + I(t)\sum_{k=1}^{n^*} f(m_k, S(t), I(t))e_k(t) \\ \qquad\qquad +I(t)\sum_{k=1}^{n^*} m_k h(m_k, S(t), I(t))e_k(t) \Big] dt + \sigma_2 I(t) dB_2(t), \\ de_k(t) = \Big[\sum_{i=1}^{n^*} q_{ik} e_i(t) - (g_k - \bar{g}(e(t)))\bar{g}(e(t))e_k(t)\Big] dt + (g_k - \bar{g}(e(t)))e_k(t) dy(t), \quad k = 1, \dots, n^*. \end{cases} \tag{2.5}$$

Moreover, the process

$$\bar{W}(t) = y(t) - \int_0^t \bar{g}(e(s))ds$$

is a one-dimensional Brownian motion adapted to $\mathcal{F}_t^y$; see e.g., [23, Theorem 7.2]. Therefore, system (2.5) can be rewritten as

$$\begin{cases} dS(t) = \Big[a_1 - b_1 S(t) - I(t)\sum_{k=1}^{n^*} f(m_k, S(t), I(t))e_k(t) \\ \qquad\qquad -I(t)\sum_{k=1}^{n^*} m_k h(m_k, S(t), I(t))e_k(t) \Big] dt + \sigma_1 S(t) dB_1(t), \\ dI(t) = \Big[-b_2 I(t) + I(t)\sum_{k=1}^{n^*} f(m_k, S(t), I(t))e_k(t) \\ \qquad\qquad +I(t)\sum_{k=1}^{n^*} m_k h(m_k, S(t), I(t))e_k(t) \Big] dt + \sigma_2 I(t) dB_2(t), \\ de_k(t) = \sum_{i=1}^{n^*} q_{ik} e_i(t) + e_k(t)(g_k - \bar{g}(e(t)))d\bar{W}(t), \quad k = 1, \dots, n^*. \end{cases} \tag{2.6}$$

This system is easier to analyze than (2.3). However, in this case, we need to assume that the signal process $\alpha(t)$ representing the portion of the rate of the infection in the group of individuals with hidden infection status takes only finitely many values. This significantly limits the setting as well as the possible applications to real-world problems.

One may simplify the problem further by assuming that $\alpha(t)$ takes values in $\{0, 1\}$. This would mean that at any given time all individuals in the hidden status group are either susceptible or infected.

3. Ergodicity of nonlinear filter and threshold for permanence and extinction

3.1. Ergodicity of nonlinear filter

The study of the asymptotic properties of the nonlinear filter has a long history in the literature. We briefly summarize the developments. One of the first works is Kunita's paper [14]. We restate the main result (Theorem 3.3) of this reference as follows.

Proposition 3.1. *(Kunita 1971) Assume that the signal process $\alpha(t)$ taking values in a compact separable Hausdorff space is a Markov–Feller process with semigroup P_t that has a unique invariant measure μ^* and*

$$\limsup_{t\to\infty} \int |P_t l(x) - \mu^*(l)|\mu(dx) = 0, \forall l \in C([0, 1]).$$

Then the process $\Pi_t(\cdot)$ is an $\mathcal{M}$-valued Markov–Feller process that has unique invariant measure Φ^*. Moreover, μ^* is the barycenter of Φ^*, i.e.,

$$\mu^*(l) = \int_{\mathcal{M}} \nu(l)\Phi^*(d\nu), \forall l \in C[0, 1].$$

Unfortunately, it was pointed out in [24] that there was a serious gap in the proof of the main result in [14]. A key role in the verification of the uniqueness for the invariant measure of $\Pi_t(\cdot)$ is the following identity

$$\bigcap_{t\geq 0} \mathcal{F}_{[0,\infty)}^y \bigvee \sigma\{\alpha(s) : s \geq t\} = \mathcal{F}_{[0,\infty)}^y \bigvee \left(\cap_{t\geq 0}\sigma\{\alpha(s) : s \geq t\}\right), \tag{3.1}$$

where $\mathcal{F}_{[0,\infty)}^y := \bigvee_{t\geq 0} \mathcal{F}_t^y$. This identity is indispensable in the proof of the uniqueness of the invariant measure of nonlinear filtering; see the counterexample given in [24]. Moreover, the exchange of intersection and supremum is not always permitted in general.[6] However, in Kunita's proof, this identity was not proved. On the other hand, it is important to note that all the known counterexamples are based on the degeneracy of the observation, i.e., there is no added noise. Therefore, it was tempting to conjecture that the identity (3.1) still holds provided the nondegeneracy of the observation.

[6] According to Williams [25] this incorrect identity "...tripped up even Kolmogorov and Wiener"; see [26, p. 837], and [27, pp. 91–93]

In 2009, R. Handel [13] has partially solved this open problem in a general setting. In fact, [13, Theorem 4.2] proved that identity (3.1) does indeed hold under conditions of ergodicity of signal process [13, Assumption3.1] and nondegeneracy of the observation process [13, Assumption3.2], which are only mildly stronger than those in [14].[7] Finally, we state the following theorem on the ergodicity of the filter $\Pi_t(\cdot)$ under our setting and our assumption.

Proposition 3.2. *Under Assumption 2.2, the process $\Pi_t(\cdot)$ is an $\mathcal{M}$-valued Markov–Feller process and has a unique invariant measure Φ^*. Moreover, μ^* is the barycenter of Φ^*, i.e.,*

$$\mu^*(l) = \int_{\mathcal{M}} \nu(l)\Phi^*(d\nu), \forall l \in C[0,1].$$

Moreover, let $\mathcal{M}_{\Phi^*} \subset \mathcal{M}$ be the support of the invariant measure Φ^* of the nonlinear filter $\Pi_t(\cdot)$. In general, one should not expect that $\mathcal{M}_{\Phi^*} = \mathcal{M}$. In fact, this does not hold even in the simple setting of the Wonham filter when the state space has only 3 states [28, Section 4].

3.2. Threshold for permanence and extinction

We next use the ergodicity of the nonlinear filter developed in the previous section in conjunction with a Lyapunov exponent analysis (sometimes called invasion type analysis in population dynamics) coming from dynamical systems [4, 12,15]. This allows us to introduce a threshold λ, which characterizes the longtime behavior of system (2.3).

Consider the equation on boundary when the infected individuals are absent, i.e.,

$$d\varphi(t) = \big(a_1 - b_1\varphi(t)\big)dt + \sigma_1\varphi(t)dB_1(t), \quad \varphi(0) = u \geq 0. \tag{3.2}$$

By solving the Fokker–Planck equation, Eq. (3.2) has a unique stationary distribution $\hat{\mu}$ with density given by

$$\frac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\frac{b}{y}}, \ y > 0, \tag{3.3}$$

where $c_1 = b_1 + \frac{\sigma_1^2}{2}, a = \frac{2c_1}{\sigma_1^2}, b = \frac{2a_1}{\sigma_1^2}$ and $\Gamma(\cdot)$ is the Gamma function. The main idea is to determine whether $I(t)$ converges to 0 or not by looking at the Lyapunov exponent $\limsup_{t\to\infty} \frac{\ln I(t)}{t}$ when $I(t)$ is small. Using Itô's formula yields

$$\begin{aligned}\frac{\ln I(t)}{t} =& \frac{\ln v}{t} + \frac{\sigma_2 B_2(t)}{t} - c_2 + \frac{1}{t}\int_0^t \int_0^1 f(x, S(s), I(s))\Pi_s(dx)ds \\ &+ \frac{1}{t}\int_0^t \int_0^1 xh(x, S(s), I(s))\Pi_s(dx)ds,\end{aligned} \tag{3.4}$$

where $c_2 = b_2 + \frac{\sigma_2^2}{2}$. Intuitively, $\limsup_{t\to\infty} \frac{\ln I(t)}{t} < 0$ implies $\lim_{t\to\infty} I(t) = 0$. As a result, if $I(t)$ is small then $S(t)$ is close to $\varphi(t)$ provided $S(0) = \varphi(0)$. Therefore, when t is sufficiently large we have

$$\begin{aligned}&\frac{1}{t}\int_0^t \int_0^1 f(x, S(s), I(s))\Pi_s(dx)ds + \frac{1}{t}\int_0^t \int_0^1 xh(x, S(s), I(s))\Pi_s(dx)ds \\ &\quad\approx \frac{1}{t}\int_0^t \int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds + \frac{1}{t}\int_0^t \int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds.\end{aligned}$$

By the strong law of large numbers for $\varphi(t)$ and Π_t from (3.4), we obtain that the Lyapunov exponent of $I_{u,v}(t)$ can be approximated by

$$-c_2 + \int_0^\infty \int_{\mathcal{M}} (f(x, y, 0)\nu(dx))\, \Phi^*(d\nu)\hat{\mu}(dy) + \int_0^\infty \int_{\mathcal{M}} \left(\int_0^1 xh(x, y, 0)\nu(dx)\right) \Phi^*(d\nu)\hat{\mu}(dy). \tag{3.5}$$

Since μ^* is the barycenter of Φ^*, the Lyapunov exponent of $I(t)$ is approximated by

$$-c_2 + \int_0^\infty \int_0^1 f(x, y, 0)\mu^*(dx)\hat{\mu}(dy) + \int_0^\infty \int_0^1 xh(x, y, 0)\mu^*(dx)\hat{\mu}(y).$$

Therefore, we define the threshold λ by

$$\lambda := -c_2 + \int_0^\infty \int_0^1 f(x, y, 0)\mu^*(dx)\hat{\mu}(dy) + \int_0^\infty \int_0^1 xh(x, y, 0)\mu^*(dx)\hat{\mu}(dy). \tag{3.6}$$

[7] According to Handel [13] whether Kunita's condition is already sufficient to guarantee uniqueness of the invariant measure with barycenter μ^* remains an open problem.

In the next section, we prove that the sign of λ characterizes the longtime behavior of the system (2.3). It is also noted that when $\alpha(t)$ is available, so is (1.3), and the permanence or extinction of (1.3) is also determined by the sign of λ defined in (3.6) (see e.g., [29]).

4. Characterization of longtime properties: Permanence and extinction

4.1. The existence and uniqueness of the solution and preliminary results

We begin with the following theorem on the existence and uniqueness of the solution of (2.3) and then proceed with a complete characterization of its positivity and some other important properties.

Theorem 4.1. *For any $(u, v, \pi) \in \mathbb{R}^2_+ \times \mathcal{M}$, there exists a unique global solution to system (2.3) with initial value (u, v, π). The three-component process $\{(S(t), I(t), \Pi_t), t \geq 0\}$ is a Markov process.*

Proof. We prove the existence and uniqueness of the solution of (2.3) first. It is noted that although we have assumed $f(s, i)$ is Lipschitz continuous, the coefficient $if(s, i)$ in the system (2.3) is not globally Lipschitz in general. Since the coefficients of the equation are locally Lipschitz continuous, there is a unique solution $(S(t), I(t))$ with the initial value $(u, v, \pi) \in \mathbb{R}^2_+ \times \mathcal{M}$, defined on maximal interval $t \in [0, \tau_e)$, $\tau_e := \inf\{t \geq 0 : S(t) \vee I(t) = \infty\}$ with the convention $\inf \emptyset = \infty$; see e.g., [30, Theorem 3.8 and Remark3.10]. We need to show $\tau_e = \infty$ a.s. If we define

$$\tau_k = \inf\Big\{t \geq 0 : S(t) \vee I(t) > k\Big\},$$

then $\tau_e = \lim_{k\to\infty} \tau_k$. Consider $V_1(s, i) = s + i$, then we have from definition of $\mathcal{L}V_1$ that

$$\mathcal{L}V_1(s, i, \pi) = a_1 - b_1 s - b_2 i \leq a_1 \ \forall (s, i, \pi) \in \mathbb{R}^2_+ \times \mathcal{M}.$$

Hence, by applying Itô's formula and taking expectation, we obtain

$$\mathbb{E}_{u,v,\pi} V_1\big(S(\tau_k \wedge t), I(\tau_k \wedge t)\big) \leq V_1(u, v) + a_1 t,$$

which together with Markov's inequality implies that

$$\begin{aligned}\mathbb{P}_{u,v,\pi}\big\{\tau_k < t\big\} &\leq \mathbb{P}_{u,v,\pi}\Big\{V_1\big(S(\tau_k \wedge t), I(\tau_k \wedge t), \alpha(\tau_k \wedge t)\big) \geq k\Big\} \\ &\leq \frac{V_1(u, v) + a_1 t}{k} \to 0 \text{ as } k \to \infty.\end{aligned}$$

Therefore, we have $\mathbb{P}_{u,v,\pi}\{\tau_e \leq t\} = 0$ or $\mathbb{P}_{u,v,\pi}\{\tau_e > t\} = 1$ for all $t > 0$. As a consequence, $\mathbb{P}_{u,v,\pi}\{\tau_e = \infty\} = 1$. Hence, system (2.3) has a unique, global, and continuous solution.

We proceed to prove the Markov property. Since Π_t is a Markov process and is independent of $B_1(t)$ and $B_2(t)$, the Markov property of the joint process $(S(t), I(t), \Pi_t)$ follows by standard arguments; see for example, [30, Theorem 3.27 and Lemma 3.2] or [20, Lemma 6.1]. To see why the argument in [30] can be applied, note that Π_t satisfies the stochastic Eq. (2.2) driven by $\beta(t)$ and the σ-algebra generated by increments $\{\beta(t_2) - \beta(t_1) : t_2 \geq t_1 \geq t\}$ is independent of $\mathcal{F}^y_t$. □

Next, using Lyapunov functions, we estimate the moments of $S(t)$ and $I(t)$, and obtain some related results. Define $\sigma^2_* := \max\{\sigma^2_1, \sigma^2_2\}$.

Lemma 4.1. *The following assertions hold:*

(i) *For any $0 < p < \frac{2\kappa}{\sigma^2_*}$ there is a constant Q_1 such that*

$$\limsup_{t\to\infty} \mathbb{E}_{u,v,\pi}(S(t) + I(t))^{1+p} \leq Q_1 \ \ \forall (u, v, \pi) \in \mathbb{R}^2_+ \times \mathcal{M}.$$

(ii) *For any $\varepsilon > 0$, $H > 1$, $T > 0$, there is $\bar{H} = \bar{H}(\varepsilon, H, T)$ such that*

$$\mathbb{P}_{u,v,\pi}\left\{\frac{1}{\bar{H}} \leq S(t) \leq \bar{H}, \quad \forall t \in [0, T]\right\} \geq 1 - \varepsilon \ \ \text{if} \ \ (u, v, \pi) \in [H^{-1}, H] \times [0; H] \times \mathcal{M},$$

and

$$\mathbb{P}_{u,v,\pi}\{0 \leq S(t), I(t) \leq \bar{H}, \quad \forall t \in [0, T]\} \geq 1 - \varepsilon \ \ \text{if} \ \ (u, v, \pi) \in [0, H] \times [0; H] \times \mathcal{M}.$$

Proof. Consider the Lyapunov function $V_3(s, i) = (s + i)^{1+p}$. By directly calculation with the differential operator $\mathcal{L}V_3$ and using Assumption 2.1, we obtain

$$\begin{aligned}\mathcal{L}V_3(s, i, \pi) =& (1 + p)(s + i)^p(a_1 - b_1 s - b_2 i) + \frac{p(1 + p)}{2}(s + i)^{p-1}\Big(\sigma^2_1 s^2 + \sigma^2_2 i^2\Big) \\ \leq& -(1 + p)(s + i)^{p-1}\Big[-\frac{p}{2}\sigma^2_*(s + i)^2 - a_1(s + i)\Big], \ \forall (s, i, \pi) \in \mathbb{R}^2_+ \times \mathcal{M}.\end{aligned}$$

Let $0 < C_4 < \frac{p(1+p)}{2}\sigma_*^2$. By some standard calculations, we get

$$C_5 = \sup_{(s,i,\pi)\in\mathbb{R}^2_+\times\mathcal{M}} \{\mathcal{L}V_3(s,i,\pi) + C_4V_3(s,i)\} < \infty.$$

This implies

$$\mathcal{L}V_3 \le C_5 - C_4V_3. \tag{4.1}$$

Applying [31, Theorem 5.2, p.157] proves part (i) of the lemma. The proof of part (ii) follows from part (i) and standard arguments; see [12, Lemma 2.1]. □

Theorem 4.2. *The process $(S(t), I(t), \Pi_t)$ is a strong Markov and Feller process. Moreover, we have $\mathbb{P}_{u,v,\pi}\{S(t) > 0, t > 0\} = 1$ and $\mathbb{P}_{u,0,\pi}\{I(t) = 0, t > 0\} = 1$, $\mathbb{P}_{u,v,\pi}\{I(t) > 0, t > 0\} = 1$ provided $v > 0$.*

Proof. It is easily seen that the solution of (2.3) is a homogeneous strong Markov and Feller process provided that the coefficients are globally Lipschitz; see e.g., [31, Theorem 2.9.3] and [32, Section 2.5]. It is noted that the space $\mathcal{M}$ of probability measures in [0, 1] endowed with the weak topology can be metricized by the bounded Lipschitz metric defined by

$$\|\pi_1 - \pi_2\|_{\mathrm{BL}} := \sup\left\{|\pi_1(l) - \pi_2(l)| : \|l\| \le 1, \sup_{x\neq y\in[0,1]} \frac{|l(x) - l(y)|}{|x-y|} \le 1\right\}.$$

Therefore, by using the results in Lemma 4.1, we obtain from the local Lipschitz property of coefficients of (2.3) and a truncation argument that $(S(t), I(t), \Pi_t)$ is a homogeneous strong Markov and Feller process. The details of this truncated argument and this result can be found in [33, Theorem 5.1].

Next, we establish the positivity of solutions. First, suppose that $u, v > 0$. Let us consider the Lyapunov function $V_2 : \mathbb{R}^2_+ \to \mathbb{R}_+$

$$V_2(s,i) = (s - 1 - \ln s) + (i - 1 - \ln i).$$

By direct calculations, we have

$$\begin{aligned}\mathcal{L}V_2(s,i,\pi) = &\left(1 - \frac{1}{s}\right)\left(a_1 - b_1 s - if(x,s,i) - i\int_0^1 xh(x,s,i)\pi(dx)\right) + \frac{\sigma_1^2 s^2}{2s^2}\\ &+ \left(1 - \frac{1}{i}\right)\left(-b_2 i + if(x,s,i) + i\int_0^1 xh(x,s,i)\pi(dx)\right) + \frac{\sigma_2^2 i^2}{2i^2}.\end{aligned}$$

It follows from Assumption 2.1 that $f(x,s,i) = |f(x,s,i) - f(x,0,i)| \le L_1 s$ and $h(x,s,i) = |h(x,s,i) - h(x,0,i)| \le L_2 s$. Therefore, it is easily seen that

$$\begin{aligned}\mathcal{L}V_2(s,i) &\le C_1 + \frac{f(s,i)i}{s} + \frac{i\int_0^1 h(x,s,i)\pi(dx)}{s}\\ &\le C_1 + (L_1 + L_2)(s+i),\end{aligned}$$

where $C_1 = a_1 + b_1 + \frac{\sigma_1^2}{2} + b_2 + \frac{\sigma_2^2}{2}$. As a result, if we let $C_2 = L_1 + L_2 + 1$ and $C_3 = C_1 + 2C_2\ln C_2 + 2C_2$ then

$$\begin{aligned}\mathcal{L}V_2(s,i,\pi) - C_2V(s,i) &\le C_1 - s - i + C_2(\ln s + \ln i) + 2C_2\\ &\le C_1 + 2C_2\ln C_2 + 2C_2 = C_3.\end{aligned} \tag{4.2}$$

For $k > 1$, denote

$$\eta = \inf\left\{t \ge 0 : S(t) \wedge I(t) \le 0\right\},$$

$$\eta_k = \inf\left\{t \ge 0 : S(t) \wedge I(t) < \frac{1}{k}\right\}.$$

Then $\eta = \lim_{k\to\infty}\eta_k$. Therefore, by using the same argument as above, we obtain from (4.2) that

$$\mathbb{P}_{u,v,\pi}\{\eta_k < t\} \le \frac{V_2(u,v) + C_3 t}{e^{-C_2 t}(\ln k - 1)} \to 0 \text{ as } k \to \infty.$$

As a result, $\mathbb{P}_{u,v,\pi}\{\eta_\infty = \infty\} = 1$. This implies that

$$\mathbb{P}_{u,v,\pi}\left\{S(t) > 0 : t > 0\right\} = \mathbb{P}_{u,v,\pi}\left\{I(t) > 0 : t > 0\right\} = 1 \ \forall u, v > 0. \tag{4.3}$$

If $u > 0, v = 0$, the result $\mathbb{P}_{u,v,\pi}\left\{S(t) > 0 : t > 0\right\} = 1$ can be shown similarly. Moreover, it is obvious that $\mathbb{P}_{u,0,\pi}\left\{I(t) = 0 : t > 0\right\} = 1$.

We are in a position to consider the case $u = 0$ and $v \geq 0$ and prove the positivity of $S(t)$. Let $\varepsilon > 0$ be sufficiently small such that

$$a_1 - b_1\tilde{u} - \tilde{v} \sup_{x\in[0,1]} \big(f(x,\tilde{u},\tilde{v}) + h(x,\tilde{u},\tilde{v})\big) \geq \frac{a_1}{2}, \tag{4.4}$$

for any $(\tilde{u},\tilde{v}) \in \mathbb{R}^2$ satisfying $\tilde{u} + |\tilde{v} - v| < \varepsilon$. Such an ε exists due to Assumption 2.1. Set

$$\tilde{\tau}_1 = \inf\{t > 0 : S(t) + |I(t) - v| \geq \varepsilon\}.$$

By the continuity of $(S(t), I(t))$ it is clear that $\mathbb{P}_{0,v,\pi}\{\tilde{\tau}_1 > 0\} = 1$. It follows from (4.4) that

$$a_1 - b_1 S(t) - I(t)\int_0^1 f(x,S(t),I(t))\Pi_t(dx) + I(t)\int_0^1 xh(x,S(t),I(t))\Pi_t(dx) > 0 \text{ if } t \in (0,\tilde{\tau}_1].$$

This and the variation of constants formula (see [31, Chapter 3]) imply that

$$\mathbb{P}_{0,v,\pi}\{S(t) > 0, t \in (0,\tilde{\tau}_1]\} = 1,$$

which combined with (4.3) and the strong Markov property of $(S(t), X(t), \Pi(t))$ yields that

$$\mathbb{P}_{0,v,\pi}\{S(t) > 0, t \in (0,\infty)\} = 1.$$

The theorem is therefore proved. □

4.2. Extinction

Consider the case $\lambda < 0$. We shall show that the number of the infected individuals $I(t)$ tends to zero at an exponential rate while the number of the susceptible individuals $S(t)$ converges to $\varphi(t)$.

Theorem 4.3. *Assume that $\lambda < 0$. Then for any initial point $(u, v, \pi) \in \mathbb{R}^{2,\circ}_+ \times \mathcal{M}$, the number of the infected individuals $I(t)$ tends to zero at an exponential rate, i.e.,*

$$\mathbb{P}_{u,v,\pi}\left\{\limsup_{t\to\infty} \frac{\ln I(t)}{t} = \lambda\right\} = 1,$$

and the susceptible class $S(t)$ converges weakly to the solution $\varphi(t)$ on the boundary.

In order to prove Theorem 4.3, we need the following auxiliary results.

Lemma 4.2. *For any $T, H > 1$, $\varepsilon > 0, \theta > 0$, there is a $\delta = \delta(H, T, \varepsilon, \theta)$ such that*

$$\mathbb{P}_{u,v,\pi}\{\tau_\theta \geq T\} \geq 1 - \varepsilon, \quad \forall\, (u, v, \pi) \in [0, H] \times (0, \delta] \times \mathcal{M},$$

where $\tau_\theta = \inf\{t \geq 0 : I_{u,v}(t) > \theta\}$.

Proof. By the exponential martingale inequality [31, Theorem 7.4, p. 44], we have $\mathbb{P}_{u,v,\pi}(\Omega_1) \geq 1 - \frac{\varepsilon}{2}$, where

$$\Omega_1 = \left\{\sigma_2 B_2(t) \leq \frac{\sigma_2^2 t}{2} + \ln\frac{2}{\varepsilon}\ \forall t \geq 0\right\}.$$

In view of part (ii) Lemma 4.1, there exists a $\bar{H} = \bar{H}(T, H, \varepsilon)$ such that $\mathbb{P}_{u,v,\pi}(\Omega_2) \geq 1 - \frac{\varepsilon}{2}$, where

$$\Omega_2 = \{0 \leq S(t), I(t) \leq \bar{H}\ \ \forall t \in [0, T]\}.$$

Applying Itô's formula to Eq. (2.3) yields that

$$\begin{aligned}\ln I(t) = \ln i - c_2 t + \int_0^t\int_0^1 f(x, S_{u,v}(s), I_{u,v}(s))\Pi_s(dx)ds \\ + \int_0^t\int_0^1 xh(x,S(s),I(s))\Pi_s(dx)ds + \sigma_2 B_2(t).\end{aligned} \tag{4.5}$$

Therefore, for any $(s, i, \pi) \in [0, H] \times (0, H] \times \mathcal{M}$ and $\omega \in \Omega_1 \cap \Omega_2$ we have from (4.5) and the Lipschitz continuity of f and h that

$$\ln I(t) < \ln i - b_2 T + T\big(2L_1\bar{H} + 2L_2\bar{H}\big) + \ln\frac{2}{\varepsilon}, \quad \forall t \in [0, T].$$

Hence, we can choose a sufficiently small $\delta = \delta(H, T, \varepsilon, \theta) < H$ such that for all $(s, i, \pi) \in [0, H] \times (0, \delta] \times \mathcal{M}$ and $0 \leq t \leq T$, $\ln I(t) < \ln\theta, \forall\omega \in \Omega_1 \cap \Omega_2$. The proof is complete. □

Proposition 4.1. *Suppose that the assumptions from Theorem 4.3 hold. For any* $0 < \varepsilon < \min\{\frac{1}{5}, -\frac{\lambda}{5}\}$ *and* $H > 1$, *there exists* $\widehat{\delta} = \widehat{\delta}(\varepsilon, H) \in (0, H^{-1})$ *such that*

$$\mathbb{P}_{u,v,\pi}\left\{\limsup_{t\to\infty}\frac{\ln I(t)}{t} = \lambda\right\} \geq 1 - 4\varepsilon,\ \forall (u, v, \pi) \in [H^{-1}; H] \times (0; \widehat{\delta}] \times \mathcal{M}_{\Phi^*}.$$

Proof. Let $\theta_0 = \theta_0(\varepsilon) < \frac{\varepsilon}{L_1+L_2} \wedge \frac{b_1}{L_1+L_2}$ be such that

$$\frac{a_1 + \theta_0(L_1 + L_2 + L_1\theta_0 + L_2\theta_0)}{b_1 - (L_1 + L_2)\theta_0} - \frac{a_1}{b_1} < \frac{\varepsilon}{L_1 + L_2}. \tag{4.6}$$

Consider the following stochastic differential equation

$$d\bar{\varphi}(t) = \big(\bar{a}_1 - \bar{b}_1\bar{\varphi}(t)\big)dt + \sigma_1\bar{\varphi}(t)dB_1(t), \tag{4.7}$$

where $\bar{a}_1 = a_1 + L_1 + L_2 + L_1\theta_0 + L_2\theta_0$, $\bar{b}_1 = b_1 - (L_1 + L_2)\theta_0 > 0$. A comparison result shows that $\varphi(t) \leq \bar{\varphi}(t)$, $t \geq 0$ *a.s.* provided that $\varphi(0) = \bar{\varphi}(0)$. Moreover, the strong law of large numbers yields that

$$\begin{aligned}
&\lim_{t\to\infty}\frac{1}{t}\Big|\int_0^t\int_0^1 f(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds + \int_0^t\int_0^1 xh(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds \\
&\qquad - \int_0^t\int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds - \int_0^t\int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds\Big| \\
&\leq \lim_{t\to\infty}\frac{1}{t}\int_0^t (L_1 + L_2)(\bar{\varphi}(s) - \varphi(s))ds \\
&= (L_1 + L_2)\left(\frac{\bar{a}_1}{\bar{b}_1} - \frac{a_1}{b_1}\right)\ a.s.
\end{aligned} \tag{4.8}$$

Combining (4.6) and (4.8) implies that

$$\begin{aligned}
&\lim_{t\to\infty}\frac{1}{t}\Big|\int_0^t\int_0^1 f(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds + \int_0^t\int_0^1 xh(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds \\
&\qquad - \int_0^t\int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds - \int_0^t\int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds\Big| < \varepsilon\ a.s.
\end{aligned} \tag{4.9}$$

Therefore, there exists $T_1 = T_1(\varepsilon)$ such that $\mathbb{P}(\Omega_3) \geq 1 - \varepsilon$, where

$$\begin{aligned}
\Omega_3 := \Big\{ &\Big|\frac{1}{t}\int_0^t\int_0^1 f(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds + \frac{1}{t}\int_0^t\int_0^1 xh(x, \bar{\varphi}(s), 0)\Pi_s(dx)ds \\
&- \frac{1}{t}\int_0^t\int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds - \frac{1}{t}\int_0^t\int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds\Big| < 2\varepsilon, \forall t \geq T_1\Big\}
\end{aligned}$$

By definition of λ and ergodicity of $\varphi(t)$, Π_t, we obtain

$$\mathbb{P}_{H,\pi}\left\{-c_2 + \lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds + \lim_{t\to\infty}\frac{1}{t}\int_0^t\int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds = \lambda\right\} = 1,$$

where $\mathbb{P}_{H,\pi}$ indicates the initial values of $(\varphi(t), \Pi_t)$. As a result, there exists $T_2 = T_2(H, \varepsilon) > 1$ such that $\mathbb{P}_{H,\pi}(\Omega_4) \geq 1 - \varepsilon$, where

$$\Omega_4 = \left\{-c_2 + \frac{1}{t}\int_0^t\int_0^1 f(x, \varphi(s), 0)\Pi_s(dx)ds + \frac{1}{t}\int_0^t\int_0^1 xh(x, \varphi(s), 0)\Pi_s(dx)ds \leq \lambda + \varepsilon,\ \forall t \geq T_2\right\}.$$

In view of the uniqueness of solutions, we have for all $u \in [0, H]$ that $\varphi_u(s) \leq \varphi_H(s)$, $s \geq 0$ almost surely where the subscript of $\varphi(s)$ indicates the initial value $\varphi(0)$. This implies that $\mathbb{P}_{u,\pi}(\Omega_4) \geq 1 - \varepsilon$ for all $(u, \pi) \in [0, H] \times \mathcal{M}_{\Phi^*}$.

Since $\lim_{t\to\infty}\frac{B_2(t)}{t} = 0$ a.s., there is a $T_3 = T_3(\varepsilon) > 1$ such that $\mathbb{P}(\Omega_5) \geq 1 - \varepsilon$ where

$$\Omega_5 = \left\{\sigma_2\frac{B_2(t)}{t} < \varepsilon\ \forall t \geq T_3\right\}.$$

Let $T = \max\{T_1, T_2, T_3\}$. By Lemma 4.2, there is a $\hat{\delta} = \hat{\delta}(\varepsilon, H) < \theta_0$ such that for all $(u, v, \pi) \in [0, H] \times (0, \hat{\delta}] \times \mathcal{M}$, $\mathbb{P}_{u,v,\pi}(\Omega_6) \geq 1 - \varepsilon$, where

$$\Omega_6 = \{\tau_{\theta_0} \geq T\}.$$

Now, it follows from (4.5) that $\forall (u, v, \pi) \in [0, H] \times (0, \delta] \times \mathcal{M}_{\Phi^*}$ we have in $\bigcap_{j=3}^{6} \Omega_j$, for all $t \in [T, \tau_{\theta_0}]$ that

$$\begin{aligned}
\ln I(t) &= \ln v - c_2 t + \int_0^t \left(\int_0^1 f(x, S(s), I(s)) \Pi_s(dx) + \int_0^1 x h(x, S(s), I(s)) \Pi_s(dx) \right) ds + \sigma_2 B_2(t) \\
&= \ln v - c_2 t + \int_0^t \left(\int_0^1 f(x, S(s), 0) \Pi_s(dx) + \int_0^1 x h(x, S(s), 0) \Pi_s(dx) \right) ds + \sigma_2 B_2(t) \\
&\quad + \int_0^t \Bigg(\int_0^1 f(x, S(s), I(s)) \Pi_s(dx) + \int_0^1 x h(x, S(s), I(s)) \Pi_s(dx) \\
&\qquad - \int_0^1 f(x, S(s), 0) \Pi_s(dx) - \int_0^1 x h(x, S(s), 0) \Pi_s(dx) \Bigg) ds \\
&\leq \ln v - c_2 t + \int_0^t \left(\int_0^1 f(x, \bar{\varphi}(s), 0) \Pi_s(dx) + \int_0^1 x h(x, \bar{\varphi}(s), 0) \Pi_s(dx) \right) ds \\
&\qquad + \varepsilon t + (L_1 + L_2) \int_0^t I(s) ds \\
&\leq \ln v - c_2 t + \int_0^t \left(\int_0^1 f(x, \varphi(s), 0) \Pi_s(dx) + \int_0^1 x h(x, \varphi(s), 0) \Pi_s(dx) \right) ds + 3\varepsilon t \\
&\leq \ln v + (\lambda + 4\varepsilon) t < \ln \hat{\delta} < \ln \theta_0.
\end{aligned} \tag{4.10}$$

In the above, we have used the fact of that whenever $I(t) \leq \theta_0$, one has

$$\begin{aligned}
I(s) \int_0^1 x h(x, S(s), I(s)) \Pi_s(dx) &\leq \theta_0 (h(0, 0, 0) + L_2 + L_2 S(s) + L_2 I(s)) \\
&\leq \theta_0 (L_2 + L_2 \theta_0) + L_2 \theta_0 S(s),
\end{aligned}$$

so $S(s) \leq \bar{\varphi}(s)$ for all $t \in [0, \tau_{\theta_0}]$.

As a result of (4.10), we must have $\tau_{\theta_0} = \infty$ for $\omega \in \bigcap_{j=3}^{6} \Omega_j$. We obtain this claim by a contradiction argument as follows. If the claim is false then we have a set $\Omega_7 \subset \bigcap_{i=3}^{6} \Omega_i$ with $\mathbb{P}(\Omega_7) > 0$ and $\tau_{\theta_0} < \infty$ for any $\omega \in \Omega_7$. We already proved that $\tau_{\theta_0} > T$ for $\omega \in \bigcap_{i=3}^{6} \Omega_i$. Moreover, in view of (4.10), we have $I(t) \leq \hat{\delta} < \theta_0$ for any $t \in [T, \tau_{\theta_0}]$. Because $I(t)$ is continuous almost surely, for almost all $\omega \in \Omega_7$ we have that $\lim_{t \to \tau_{\theta_0}} I(t) = I(\tau_{\theta_0}) < \hat{\delta} < \theta_0$, which is a contradiction. So, $\tau_{\theta_0} = \infty$ for $\omega \in \bigcap_{j=1}^{4} \Omega_j$. We deduce from $\tau_{\theta_0} = \infty$ and (4.10) that $\lim_{t \to \infty} I(t) = 0$ for almost $\omega \in \bigcap_{j=3}^{6} \Omega_j$.

Next, because $I(t) \leq \theta_0$ for any $t \geq 0$ for almost all $\omega \in \bigcap_{j=3}^{6} \Omega_j$, we have shown that $S(t) \leq \bar{\varphi}(t)$, $\forall t \geq 0$ almost surely in $\bigcap_{j=3}^{6} \Omega_j$. Similar to (3.2), since $\bar{b}_1 > 0$, the solution to (4.7) has a unique invariant measure, say $\bar{\mu}$. Then we have from the ergodicity of $\bar{\varphi}(t)$ that for some small $\hat{p} > 0$,

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t S^{1+\hat{p}}(u) du \leq \lim_{t \to \infty} \frac{1}{t} \int_0^t \bar{\varphi}^{1+\hat{p}}(u) du = \int x^{1+\hat{p}} \bar{\mu}(dx) < \infty \text{ almost surely in } \bigcap_{j=3}^{6} \Omega_j. \tag{4.11}$$

Using (4.11), the fact that $\lim_{t \to \infty} I(t) = 0$, and the compactness of $\mathcal{M}$, the family of random occupation measures

$$\widetilde{U}^t_{u,v,\pi}(\cdot) := \frac{1}{t} \int_0^t \mathbf{1}_{\{(S(s), I(s), \Pi_s) \in \cdot\}} ds$$

is tight for almost all $\omega \in \bigcap_{j=1}^{3} \Omega_j$. From [15, Lemma 5.6], with probability 1, any weak limit of $\widetilde{U}^t_{u,v,\pi}(\cdot)$ as $t \to \infty$ (if it exists) is an invariant probability measure, which has support on $[0, \infty) \times \{0\} \times \mathcal{M}$. Because $\hat{\mu} \times \boldsymbol{\delta} \times \Phi^*$, where $\boldsymbol{\delta}$ is the Dirac measure concentrated at 0, is an invariant probability measure on $[0, \infty) \times \{0\} \times \mathcal{M}$, the family $\widetilde{U}^t_{u,v,\pi}(\cdot)$ converges weakly to $\boldsymbol{\mu}_0 \times \boldsymbol{\delta}$ almost surely in $\bigcap_{j=3}^{6} \Omega_j$ as t tends to ∞. One has from the weak convergence and the uniform integrability in

(4.11) that

$$\begin{aligned}\lim_{t\to\infty}\frac{\ln I(t)}{t} &= \lim_{t\to\infty}\frac{1}{t}\Big[-b_2 t-\frac{\sigma_2^2 t}{2}-\int_0^t\int_0^1 f(x,S(s),I(s))\Pi_s(dx)ds\\ &\qquad +\int_0^t\int_0^1 xh(x,S(s),I(s))\Pi_s(ds)ds\Big]+\lim_{t\to\infty}\frac{\sigma_2 B_2(t)}{t}\\ &=\lim_{t\to\infty}\int_{\mathbb{R}_+^2\times\mathcal{P}(\mathcal{M})}\Big[-c_2+\int_0^1 f(x,y,i)\nu(dx)+\int_0^1 xh(x,y,i)\nu(dx)\Big]\widetilde{U}^t_{u,v,\pi}(dy,di,d\nu)\\ &=-c_2+\int_0^\infty\int_{\mathcal{M}}\Big[\int_0^1 f(x,y,0)\nu(dx)\Big]\Phi^*(d\nu)\hat{\mu}(dy)\\ &\quad+\int_0^\infty\int_{\mathcal{M}}\Big[\int_0^1 xh(x,y,0)\nu(dx)\Big]\Phi^*(d\nu)\hat{\mu}(dy)\\ &=\lambda<0,\end{aligned}$$

for almost every $\omega\in\bigcap_{j=3}^6\Omega_j$, $(u,v,\pi)\in[0,H]\times(0,\hat{\delta}]\times\mathcal{M}_{\Phi^*}$. The proof is complete by noting that $\mathbb{P}(\bigcap_{j=3}^6\Omega_j)>1-4\varepsilon$. □

Proof of Theorem 4.3. Let $\varepsilon>0$ be arbitrary. In view of Proposition 4.1, the process $(S(t),I(t))$ is transient (see e.g., [34] for definition) in $\mathbb{R}_+^{2,\circ}$. Thus, the process has no invariant probability measure in $\mathbb{R}_+^{2,\circ}$. Thus $\hat{\mu}\times\boldsymbol{\delta}\times\Phi^*$ is the unique invariant probability measure of $(S(t),I(t),\Pi_t)$ in $\mathbb{R}_+^2\times\mathcal{M}_{\Phi^*}$.

Let H be sufficiently large that $\hat{\mu}((0,H))>1-\varepsilon$. Thanks to Lemma 4.1 part (i) and compactness of $\mathcal{M}$, the process $(S(t),I(t),\Pi_t)$ is tight. Consequently, the occupation measure

$$U^t_{u,v,\pi}(\cdot):=\frac{1}{t}\int_0^t\mathbb{P}_{u,v,\pi}\{(S(s),I(s),\Pi_s)\in\cdot\}\,ds$$

is tight in $\mathbb{R}_+^2\times\mathcal{M}$. Since any weak limit of $U^t_{u,v,\pi}$ as $t\to\infty$ must be an invariant probability measure of $(S(t),I(t),\Pi_t)$ (see [15]), we have that $U^t_{u,v,\pi}$ converges weakly to $\hat{\mu}\times\boldsymbol{\delta}\times\Phi^*$ as $t\to\infty$. As a result, for any $\delta>0$, there exists a $\hat{T}>0$ such that

$$U^{\hat{T}}_{u,v,\pi}((0,H)\times(0,\delta)\times\mathcal{M}_{\Phi^*})>1-\varepsilon,$$

or equivalently,

$$\frac{1}{\hat{T}}\int_0^{\hat{T}}\mathbb{P}_{u,v,\pi}\{(S(t),I(t),\Pi_t)\in(0,H)\times(0,\delta)\times\mathcal{M}_{\Phi^*}\}dt>1-\varepsilon.$$

As a result, we have

$$\mathbb{P}_{u,v,\pi}\{\hat{\tau}\le\hat{T}\}>1-\varepsilon,$$

where $\hat{\tau}=\inf\{t\ge0:(S(t),I(t),\Pi_t)\in(0,H)\times(0,\delta)\times\mathcal{M}_{\Phi^*}\}$. Using the strong Markov property and Proposition 4.1, we have that

$$\mathbb{P}_{u,v,\pi}\left\{\lim_{t\to\infty}\frac{\ln I(t)}{t}\le\lambda+4\varepsilon\right\}\ge1-\varepsilon,$$

for any $(u,v,\pi)\in\mathbb{R}_+^{2,*}\times\mathcal{M}$. Therefore, since $\varepsilon>0$ is arbitrary, the assertion in convergence of $I(t)$ follows. Once we have the exponentially fast convergence to 0 of $I(t)$, the convergence of $S(t)$ to $\varphi(t)$ follows from standard arguments; see, for example, [4,12]. □

4.3. Permanence

In this section, we deal with the case $\lambda>0$ and prove that the system is permanent in the following sense.

Definition 4.1. We say that system (2.3) is permanent (in mean) if for any initial value $(u,v,\pi)\in\mathbb{R}_+^{2,\circ}\times\mathcal{M}$

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t\mathbb{E}_{u,v,\pi}S(s)ds>0,\quad\liminf_{t\to\infty}\frac{1}{t}\int_0^t\mathbb{E}_{u,v,\pi}I(s)ds>0.$$

Theorem 4.4. *Assume that $\lambda>0$. Then for any initial data $(u,v,\pi)\in\mathbb{R}_+^{2,\circ}\times\mathcal{M}$, system (2.3) is permanent (in mean).*

Proof. We first prove that all invariant measures of $(S(t), I(t), \Pi(t))$ concentrate on $\mathbb{R}^{2,\circ}_+ \times \mathcal{M}$. We assume by contradiction that there is no invariant measure on $\mathbb{R}^{2,\circ}_+$ of $(S(t), I(t))$. Therefore, there is no invariant measure on $\mathbb{R}^{2,*}_+$ since the solutions starting in $\mathbb{R}^{2,*}_+$ will enter and remain in $\mathbb{R}^{2,\circ}_+$ due to Theorem 4.1. As a result, $\hat{\mu} \times \boldsymbol{\delta} \times \Phi^*$ (the unique invariant on the boundary $\mathbb{R}_+ \times \{0\} \times \mathcal{M}$) is the unique invariant probability measure of the process $\{S(t), I(t), \Pi_t\}$ on $\mathbb{R}^2_+ \times \mathcal{M}$. Therefore, by applying [15, Lemma 3.4], we have

$$\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t}\mathbb{E}_{u,v,\pi} \int_0^t \int_0^1 f(x, S(s), I(s))\Pi_s(dx)ds \\
&\quad = \int_{\mathbb{R}^2_+\times\mathcal{M}} \int_0^1 f(x, y, i)\hat{\mu}(dy)\boldsymbol{\delta}(di)\Phi^*(d\nu) \\
&\quad = \int_0^\infty \int_0^1 f(x, y, 0)\mu^*(dx)\hat{\mu}(dy).
\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t}\mathbb{E}_{u,v,\pi} \int_0^t \int_0^1 xh(x, S(s), I(s))\Pi_s(dx)ds \\
&\quad = \int_{\mathbb{R}^2_+\times\mathcal{M}} \int_0^1 xh(x, y, i)\nu(dx)\hat{\mu}(dy)\boldsymbol{\delta}(di)\Phi^*(d\nu) \\
&\quad = \int_{\mathbb{R}_+\times\mathcal{M}} \int_0^1 xh(x, y, 0)\nu(dx)\hat{\mu}(dy)\Phi^*(d\nu) \\
&\quad = \int_0^\infty \int_0^1 xh(x, y, 0)\mu^*(dx)\hat{\mu}(dy).
\end{aligned} \tag{4.13}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}_{u,v,\pi} \frac{\ln Y(t)}{t} =& \mathbb{E}_{x,y} \frac{\ln v}{t} - c_2 + \frac{1}{t}\mathbb{E}_{x,y} \left(\int_0^t \int_0^1 f(x, S(s), I(s))\Pi_s(dx)ds \right. \\
&\left. + \int_0^t \int_0^1 xh(x, S(s), I(s))\Pi_s(dx)ds \right) + \mathbb{E}_{x,y} \frac{\sigma_2 W_2(t)}{t}.
\end{aligned}$$

As a result, we have

$$\begin{aligned}
&\lim_{t\to\infty} \mathbb{E}_{u,v,\pi} \frac{\ln I(t)}{t} \\
&= -c_2 + \lim_{t\to\infty} \mathbb{E}_{x,y} \frac{1}{t} \left(\int_0^t \int_0^1 f(x, S(s), I(s))\Pi_s(dx)ds + \int_0^t \int_0^1 xh(x, S(s), I(s))\Pi_s(dx)ds \right) \\
&= -c_2 + \int_0^\infty \int_0^1 f(x, y, 0)\mu^*(dx)\hat{\mu}(dy) + \int_0^\infty \int_0^1 xh(x, y, 0)\mu^*(dx)\hat{\mu}(dy) \\
&= \lambda > 0.
\end{aligned}$$

This contradicts the fact that

$$\lim_{t\to\infty} \mathbb{E}_{u,v,\pi} \frac{\ln I(t)}{t} \le \lim_{t\to\infty} \mathbb{E}_{u,v,\pi} \frac{I(t)}{t} = 0$$

because $\ln y \le y$ while Lemma 4.1 implies $\lim_{t\to\infty} \mathbb{E}_{u,v,\pi} \frac{I(t)}{t} = 0$. As a result, all invariant measures of $(S(t), I(t), \Pi(t))$ concentrate on $\mathbb{R}^{2,\circ}_+ \times \mathcal{M}_{\Phi^*}$ (the existence of an invariant measure follows from Lemma 4.1).

By the moment boundedness in Lemma 4.1, we obtain that there exists a sequence $T_k \to \infty$ as $k \to \infty$ such that

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{E}_{u,v,\pi} S(s)ds = \lim_{k\to\infty} \frac{1}{T_k} \int_0^{T_k} \mathbb{E}_{u,v,\pi} S(s)ds,$$

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{E}_{u,v,\pi} I(s)ds = \lim_{k\to\infty} \frac{1}{T_k} \int_0^{T_k} \mathbb{E}_{u,v,\pi} I(s)ds.$$

Therefore, by applying [15, Lemma 3.4] and the fact that all invariant measures of $(S(t), I(t), \Pi(t))$ concentrate on $\mathbb{R}^{2,\circ}_+ \times \mathcal{M}_{\Phi^*}$, we obtain the desired result. □

4.4. Hidden Markov chain

This section is devoted to the case when the signal process $\alpha(t)$ takes values in a finite set $\{m_1, \dots, m_{n^*}\} \subset [0, 1]$ and admits a unique invariant measure $(\mu_1^*, \dots, \mu_n^*) \in \mathcal{S}_{n^*}$. In this case, system (2.3) is replaced by (2.6). We first have the following well-posedness and other preliminary results.

Theorem 4.5. *Consider system (2.6). For any $(u, v, e) \in \mathbb{R}_+^2 \times \mathcal{S}_{n^*}$, there exists a unique global solution to system (2.6) with initial data (u, v, e). The three-component process $\{(S(t), I(t), \Pi_t) : t \geq 0\}$ is a Markov–Feller process. Moreover, we have $\mathbb{P}_{u,v,e}\{S(t) > 0, t > 0\} = 1$ and $\mathbb{P}_{u,0,\pi}\{I(t) = 0, t > 0\} = 1$, $\mathbb{P}_{u,v,e}\{I(t) > 0, t > 0\} = 1$ provided $v > 0$.*

Proof. The proof of this Theorem is as same as that of Theorems 4.1 and 4.2, and is thus omitted. □

To proceed, we classify the persistence and extinction of system (2.6) by the following threshold λ:

$$\lambda := -c_2 + \sum_{k=1}^{n^*} \mu_k^* \int_0^\infty f(m_k, y, 0)\hat{\mu}(dy) + \sum_{k=1}^{n^*} m_k \mu_k^* \int_0^\infty h(m_k, y, 0)\hat{\mu}(dy). \tag{4.14}$$

Theorem 4.6. *The following results hold.*

(i) *If $\lambda < 0$ then for any initial point $(u, v, e) \in \mathbb{R}_+^{2,\circ} \times \mathcal{S}_{n^*}$, the number of the infected individuals $I(t)$ tends to zero at an exponential rate, that is,*

$$\mathbb{P}_{u,v,e}\left\{\limsup_{t\to\infty} \frac{\ln I(t)}{t} = \lambda\right\} = 1,$$

and the susceptible class $S(t)$ converges weakly to the solution $\varphi(t)$ on the boundary.

(ii) *If $\lambda > 0$, then for any initial point $(u, v, e) \in \mathbb{R}_+^{2,\circ} \times \mathcal{S}_{n^*}$, system (2.6) is permanent (in mean).*

(iii) *Moreover, the observable system (2.6) and original system (1.3) share the same threshold for the persistence and extinction.*

Proof. The proofs of part (i) and (ii) in this Theorem are the same as that of Theorems 4.3, and 4.3 and, therefore are omitted. Part (iii) follows immediately from the formulation of λ given in (4.14) and the threshold for system (1.3) given in [29]. □

Remark 5. Note that in Theorem 4.6, we have collected several results. These results may be presented in separate theorems. Given that they are all related to the same hidden process and that we have carried out an extensive analysis in the last section, it seemed reasonable to collect these results in one theorem.

In Section 6.2, we present numerical examples for Theorem 4.6. The reader can see that although the observable system (2.6) may not approximate well the original at every point in time, it preserves the longtime properties (permanence or extinction).

5. Discussion and interpretation

When a pandemic arises, there are usually multiple options one can take in order to control it. If we apply an extreme policy to control the disease transmission (i.e., we try to reduce the infection rate to be very small or almost 0), we can certainly control the pandemic. This can be seen by looking at λ defined in the previous section and noting that if $f(S, I) \approx 0$ then $\lambda \approx -c_2 < 0$. However, this type of *highly restrictive policy* may hurt the economy. It is important to balance public health and the economic issues. In our context, this is equivalent to ensuring that $\lambda < 0$, which ensures the pandemic is under control, but not making λ too negative, since in the process of doing this (lockdowns, bankruptcy of businesses, unemployment) the economy could suffer significantly.

Definition 5.1. We say a proposed threshold λ_1 is overcautious if it is greater than the exact threshold, that is $\lambda_1 > \lambda$. We say a proposed threshold λ_1 is incautious if it is less than the exact threshold, that is $\lambda_1 < \lambda$.

Remark 6. We say a threshold $\lambda_1 > \lambda$ (the exact one) is an overcautious proposed threshold because if this threshold is implemented, we tend to apply a policy to make $\lambda_1 < 0$. But this may not be necessary because the exact threshold is $\lambda < \lambda_1$. Conversely, a threshold $\lambda_1 < \lambda$ is an incautious one because reducing λ_1 to be less than 0 may not be enough and the pandemic would still not be controlled.

Let us consider the case when $\alpha(t)$ takes the finitely many values $0 = m_1 < m_2 < \cdots < m_{n^*} = 1$. We note that our analysis is still true for the general case when α takes values in $[0, 1]$. Recall that we assume $\alpha(t)$ has a unique invariant (discrete) measure $(\mu_1^*, \dots, \mu_{n^*}^*)$.

Since $\alpha(t)$ is not directly available, we have two options. One option is to use filtering to estimate $\alpha(t)$ and then consider the corresponding system with filtering. It was shown that this method preserves the longtime behavior of the original system. In particular, from Section 4.4 the exact threshold for this method is

$$\lambda = -c_2 + \sum_{k=1}^{n^*} \mu_k^* \int_0^\infty f(m_k, y, 0)\hat{\mu}(dy) + \sum_{k=1}^{n^*} m_k \mu_k^* \int_0^\infty h(m_k, y, 0)\hat{\mu}(dy).$$

Another possible option is to estimate some prediction for $\alpha(t)$ and give that value for $\alpha(t)$. Let $k_0 \in \{1, 2, \dots, n^*\}$. Assume that we estimate $\alpha(t) = m_{k_0}$ and consider the system

$$\begin{cases} dS(t) = \Big[a_1 - b_1 S(t) - I(t) f(m_{k_0}, S(t), I(t)) \\ \qquad\qquad -\alpha_0 I(t) h(m_{k_0}, S(t), I(t)) \Big] dt + \sigma_1 S(t) dB_1(t), \\ dI(t) = \Big[-b_2 I(t) + I(t) f(m_{k_0}, S(t), I(t)) \\ \qquad\qquad +m_{k_0} I(t) h(m_{k_0}, S(t), I(t)) \Big] dt + \sigma_2 I(t) dB_2(t). \end{cases} \tag{5.1}$$

Using the results from [29], the threshold for persistence and extinction of (5.1) is given by

$$\lambda_{pre} = -c_2 + \int_0^\infty f(m_{k_0}, y, 0)\hat{\mu}(dy) + m_{k_0} \int_0^\infty h(m_{k_0}, y, 0)\hat{\mu}(dy).$$

The following results tell us when λ_{pre} is an incautious or overcautious threshold.

Proposition 5.1.

(i) *If*

$$\mu_{k_0}^* \int_0^\infty (f(m_{k_0}, y, 0) + m_{k_0} h(m_{k_0}, y, 0))\hat{\mu}(dy) < \sum_{k=1}^{n^*} \mu_k^* \int_0^\infty f(m_k, y, 0)\hat{\mu}(dy) + \sum_{k=1}^{n^*} m_k \mu_k^* \int_0^\infty h(m_k, y, 0)\hat{\mu}(dy)$$

then $\lambda_{\text{pre}} < \lambda$. As a result, λ_{pre} is an incautious threshold.

(ii) *If*

$$\mu_{k_0}^* \int_0^\infty (f(m_{k_0}, y, 0) + m_{k_0} h(m_{k_0}, y, 0))\hat{\mu}(dy) > \frac{1}{1 - \mu_{n^*}^*} \Big(\sum_{k=1}^{n^*-1} \mu_k^* \int_0^\infty f(m_k, y, 0)\hat{\mu}(dy) + \sum_{k=1}^{n^*-1} m_k \mu_k^* \int_0^\infty h(m_k, y, 0)\hat{\mu}(dy) \Big)$$

then $\lambda_{\text{pre}} > \lambda$. As a result, λ_{pre} is an overcautious threshold.

It is natural to assume that $f(x, y, 0)$ is an increasing function w.r.t. x because the higher infection rate in the group of potentially infected individuals would make the disease spread faster. Note that this intuition is natural but not always true because we are examining functions at the boundary, i.e., there is no infected group.

Assumption 5.1. *For each fixed y, the functions $f(x, y, 0)$ and $h(x, y, 0)$ are increasing in x.*

Under this natural assumption, we will see that assuming that all potentially infected individuals are infected will lead to an overcautious policy. Conversely, it will be incautious if we assume that all individuals with hidden infection status are free of the disease. We will make this analysis clearer in the following two subsections.

5.1. The overcautious case: assuming that all individuals in the hidden group are infected

Suppose we do not use the filtering to consider the observable problem, make the assumption $\alpha(t) = 1, t \geq 0$ and consider the system (5.1) with $m_{k_0} = m_{n^*} = 1$. Under Assumption 5.1, this is an overcautious prediction. The following theorem is consistent with this intuition.

Proposition 5.2. *Under Assumption 5.1,*

$$\int_0^\infty (f(1,y,0)+h(1,y,0))\hat{\mu}(dy) > \frac{1}{1-\mu^*_{n^*}}\Big(\sum_{k=1}^{n^*-1}\mu^*_k\int_0^\infty f(m_k,y,0)\hat{\mu}(dy)+\sum_{k=1}^{n^*-1}m_k\mu^*_k\int_0^\infty h(m_k,y,0)\hat{\mu}(dy)\Big) \tag{5.2}$$

As a result, $\lambda_{\text{pre}} > \lambda$ and λ_{pre} is an overcautious threshold.

Proof. The estimate (5.2) follows immediately from Assumption 5.1. We omit the details here. □

5.2. The incautious case: assuming that all individuals in the hidden group are not infected

If we make the assumption that $\alpha(t) = 0$, $t \geq 0$ and consider the system (5.1) with $m_{k_0} = m_1 = 0$, this is an incautious prediction (under Assumption 5.1). The following result is consistent with this fact.

Proposition 5.3. *Under Assumption 5.1,*

$$\int_0^\infty (f(0,y,0)+h(0,y,0))\hat{\mu}(dy) < \frac{1}{1-\mu^*_1}\Big(\sum_{k=2}^{n^*}\mu^*_k\int_0^\infty f(m_k,y,0)\hat{\mu}(dy)+\sum_{k=2}^{n^*}m_k\mu^*_k\int_0^\infty h(m_k,y,0)\hat{\mu}(dy)\Big)$$

As a result, $\lambda_{\text{pre}} < \lambda$ and λ_{pre} is an incautious threshold.

6. Examples and simulations

6.1. A simple example

In this section, we consider a simple example. Assume that all the individuals in the hidden class have the same status (susceptible or infected). In other words, the signal Markov process $\alpha(t)$ has only two possible states, 0 or 1, (corresponding to the case that all individuals in the hidden class are disease-free or infected, respectively). Assume that $\alpha(t)$ has the generator

$$Q = \begin{bmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{bmatrix}.$$

We can only observe $\alpha(t)$ through the observation process $y(t)$ given by

$$dy(t) = g(\alpha(t))dt + dW(t), \quad y(0) = 0,$$

where $g : \{0, 1\} \to \mathbb{R}$ and set $g_1 = g(0)$, $g_2 = g(1)$. Let $e(t) = \mathbb{E}[\mathbf{1}_{\{\alpha(t)=0\}}|\mathcal{F}^y_t]$.

The dynamics of this epidemic system under the (Wonham) filter is described by

$$\begin{cases} dS(t) = \Big[\, a_1 - b_1S(t) - I(t)f(0,S(t),I(t))e(t) - I(t)f(1,S(t),I(t))(1-e(t)) \\ \qquad\qquad -I(t)h(1,S(t),I(t))(1-e(t)) \,\Big]\, dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \Big[\, -b_2I(t) + I(t)f(0,S(t),I(t))e(t) + I(t)f(1,S(t),I(t))(1-e(t)) \\ \qquad\qquad +I(t)h(1,S(t),I(t))(1-e(t)) \,\Big]\, dt + \sigma_2 I(t)dB_2(t), \\ de(t) = [q_2 - (q_1+q_2)e(t)]dt + (g_1-g_2)e(t)(1-e(t))d\bar{W}(t). \end{cases} \tag{6.1}$$

We will also assume that $q_1, q_2 > 0$ and $g := g_1 - g_2 \neq 0$ since the other cases are trivial. Consider the equation of the third component

$$de(t) = [q_2 - (q_1+q_2)e(t)]dt + ge(t)(1-e(t))d\bar{W}(t), \quad e(0) \in [0,1]. \tag{6.2}$$

By using the Lyapunov functional method as in Theorem 4.1, we obtain the following result.

Theorem 6.1. *For any initial value in* $[0, 1]$*, Eq. (6.2) has a unique solution and*

$$\mathbb{P}\{e(t) \in (0,1), \forall t > 0\} = 1.$$

Now, by solving the Fokker–Planck equation, Eq. (6.2) has a unique invariant measure supported in [0, 1] with density given by

$$\phi^*(x) = Ce^{-\frac{d_1}{1-x}}(1-x)^{2(d_1-d_2-1)}e^{-\frac{d_2}{x}}x^{2(d_2-d_1-1)}, \ x \in (0,1),$$

where $d_1 = \frac{q_1}{g^2}$, $d_2 = \frac{q_2}{g^2}$, and C is a normalizing constant.

As developed in the main results, the threshold λ is defined by

$$\begin{aligned}\lambda =& -c_2 + \frac{q_2}{q_1+q_2}\int_0^\infty f(0,y,0)\hat{\mu}(dy)\\ &+ \frac{q_1}{q_1+q_2}\int_0^\infty f(1,y,0)\hat{\mu}(dy) + \frac{q_1}{q_1+q_2}\int_0^\infty h(1,y,0)\hat{\mu}(dy),\end{aligned} \tag{6.3}$$

where $\hat{\mu}$ is the invariant measure with the density given by (3.3).

Theorem 6.2. *Consider system (6.1) and λ as in (6.3).*

- *If $\lambda < 0$ then for any initial point $(u, v, e_0) \in \mathbb{R}^{2,\circ}_+ \times \mathcal{S}_2$, the number of the infected individuals $I(t)$ tends to zero at an exponential rate, i.e.,*

$$\mathbb{P}_{u,v,e_0}\left\{\limsup_{t\to\infty}\frac{\ln I(t)}{t} \le \lambda\right\} = 1,$$

  and the susceptible class $S(t)$ converges weakly to the solution $\varphi(t)$ on the boundary.
- *If $\lambda > 0$, then for any initial point $(u, v, e_0) \in \mathbb{R}^{2,\circ}_+ \times \mathcal{S}_2$, all invariant probability measures of the solution $(S(t), I(t), e(t))$ concentrate on $\mathbb{R}^{2,\circ}_+ \times [0,1]$. Moreover, the system (6.1) is permanent.*

6.2. Numerical examples

In this section, we consider a simple example and provide some numerical simulations. Consider Eq. (6.1) with $f(x,s,i) = m_1(x)s$, $h(x,s,i) = \frac{m_2(x)s}{1+s+i}$, $g(x) = x$, $(x \in \{0,1\})$, i.e., consider

$$\begin{cases} dS(t) = \Big[\, a_1 - b_1S(t) - m_1(0)S(t)I(t)e(t) - m_1(1)S(t)I(t)(1-e(t)) \\ \qquad\qquad - \dfrac{m_2(1)(1-e(t))S(t)I(t)}{1+S(t)+I(t)} \,\Big]\, dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \Big[\, -b_2I(t) + m_1(0)S(t)I(t)e(t) + m_1(1)S(t)I(t)(1-e(t)) \\ \qquad\qquad + \dfrac{m_2(1)(1-e(t))S(t)I(t)}{1+S(t)+I(t)} \,\Big]\, dt + \sigma_2 I(t)dB_2(t), \\ de(t) = [q_2 - (q_1+q_2)e(t)]dt + e(t)(1-e(t))d\bar{W}(t). \end{cases} \tag{6.4}$$

The above is the corresponding system with a filtering of hidden Markov chain $\alpha(t)$, whereas the following system is one under hidden process $\alpha(t)$

$$\begin{cases} dS(t) = \Big[\, a_1 - b_1S(t) - m_1(\alpha(t))S(t)I(t) \\ \qquad\qquad - \dfrac{\alpha(t)m_2(\alpha(t))S(t)I(t)}{1+S(t)+I(t)} \,\Big]\, dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \Big[\, -b_2I(t) + m_1(\alpha(t))S(t)I(t) \\ \qquad\qquad + \dfrac{\alpha(t)m_2(\alpha(t))S(t)I(t)}{1+S(t)+I(t)} \,\Big]\, dt + \sigma_2 I(t)dB_2(t). \end{cases} \tag{6.5}$$

Here $\alpha(t)$ is the Markov chain taking values in $\{0,1\}$ with the generator Q.

Suppose one does not use the filtering to consider the corresponding observable system (6.4), and considers instead the system under some predictions for $\alpha(t)$. If one uses the incautious prediction and assumes that $\alpha(t) = 0$ (i.e., considers all individuals in the hidden group not to be infected), the corresponding system is

$$\begin{cases} dS(t) = \Big[a_1 - b_1S(t) - m_1(0)S(t)I(t)\Big]dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \Big[-b_2I(t) + m_1(0)S(t)I(t)\Big]dt + \sigma_2 I(t)dB_2(t). \end{cases} \tag{6.6}$$

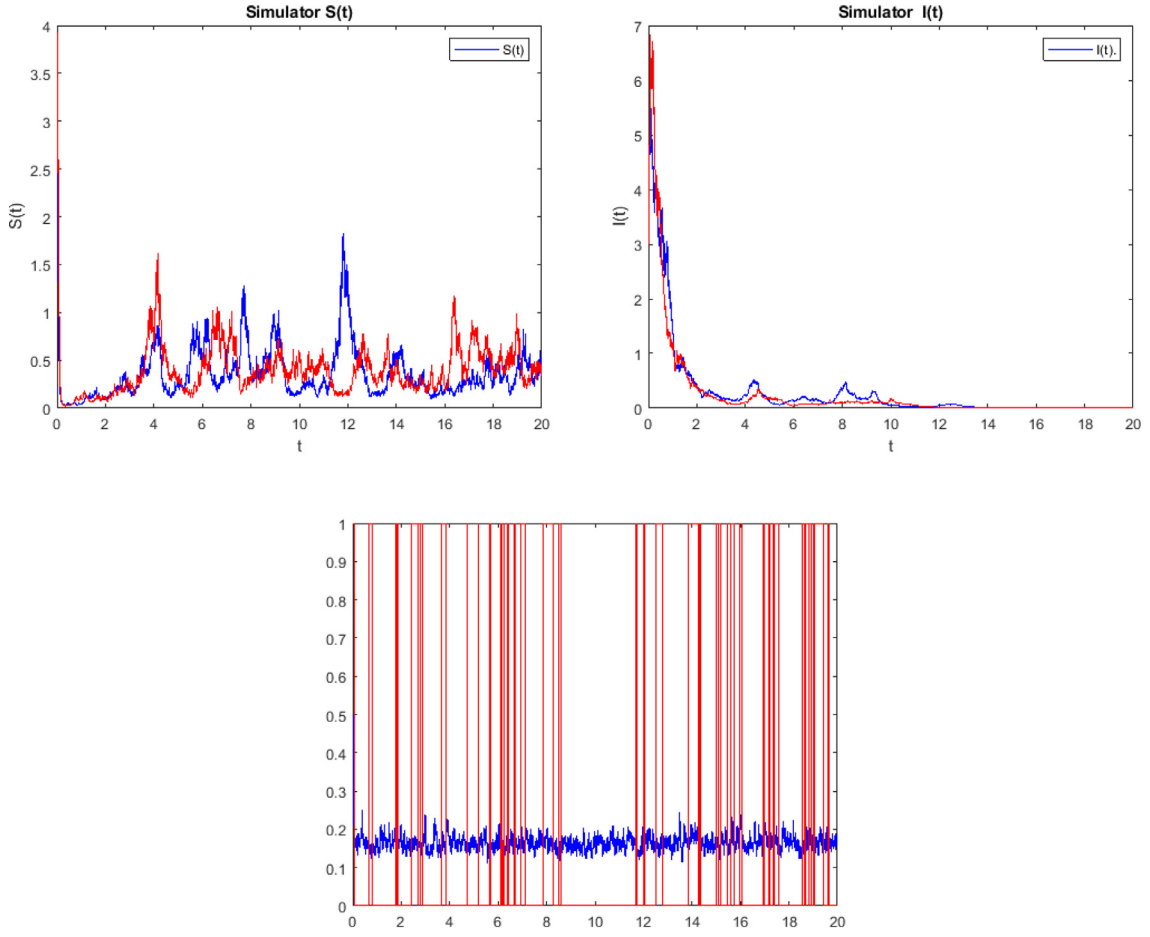

Fig. 1. Left top subfigure: Sample paths of $S(t)$, in which the red curve is for system under the hidden $\alpha(t)$ in (6.5), the blue curve corresponding to the known $\alpha(t)$ in (6.4). Right top subfigure: Sample paths of $I(t)$, the red curve is for system (6.5) with hidden state, the blue curve is that of (6.4) with known $\alpha(t)$. Bottom subfigure: the sample path of the Markov chains, the red is of the signal process $\alpha(t)$, the blue is its filter $e(t)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

If one uses the overcautious prediction and assumes that $\alpha(t) = 1$ (i.e., considers all individuals in the hidden group to be infected), the corresponding system is

$$\begin{cases} dS(t) = \left[a_1 - b_1 S(t) - m_1(1)S(t)I(t) - \dfrac{m_2(1)S(t)I(t)}{1+S(t)+I(t)}\right]dt + \sigma_1 S(t)dB_1(t), \\ dI(t) = \left[-b_2 I(t) + m_1(1)S(t)I(t) + \dfrac{m_2(1)S(t)I(t)}{1+S(t)+I(t)}\right]dt + \sigma_2 I(t)dB_2(t). \end{cases} \tag{6.7}$$

Example 6.1. Consider (6.4) with $a_1 = 0.5, b_1 = 1, \sigma_1 = 1, b_2 = 2, \sigma_2 = 0.5, m_1(0) = 0.1, m_1(1) = 4, m_2 = 0.1, q_1 = 5, q_2 = 25$. Our computation shows that $\lambda = -1.7252$, the (exact) threshold determining the longtime behavior (persistence and extinction) for both systems (6.4) and (6.5). Similarly, we can compute $\lambda_0 = -2.0750$, the threshold for system (6.6) and $\lambda_1 = 0.0241$, the threshold for system (6.7).

Note that $\lambda_1 > \lambda > \lambda_0$ and as a result λ_1 is an overcautious threshold. In this case, if we use system (6.7), we may conclude that the disease may not be controlled although it will indeed be controlled. Some unnecessarily restrictive policy might be chosen and this might lead to economic downturns. Conversely, λ_0 is an incautious threshold which would lead to overly optimistic expectations.

As our theoretical results show the number of infected $I(t)$ will tend to 0 as $t \to \infty$, i.e., the infected group goes extinct. We have also shown that systems (6.4) and (6.5) have the same longtime behavior. We provide the numerical simulations for this example in Fig. 1.

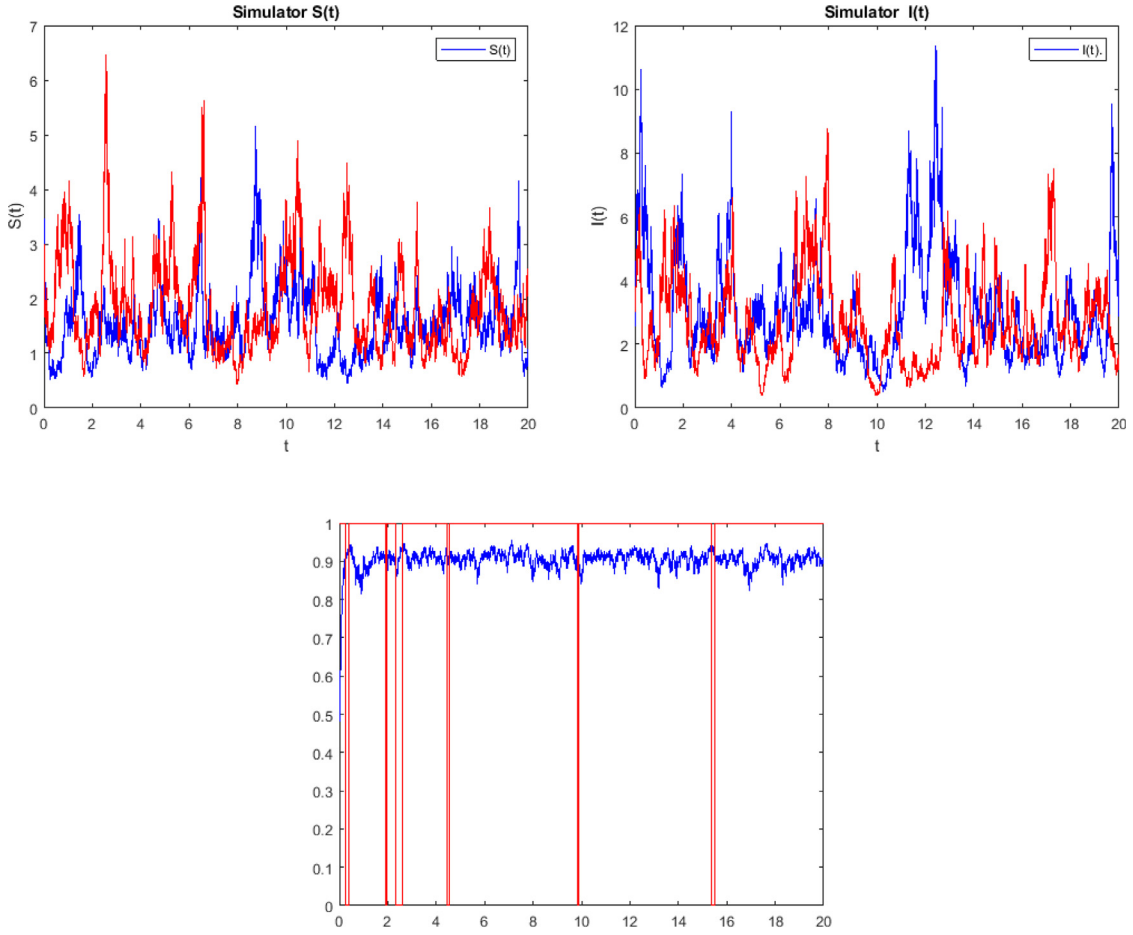

Fig. 2. Left top subfigure: Sample path of $S(t)$, the red curve is for the system under noisy observation of $\alpha(t)$, the blue is the system with precise known value of $\alpha(t)$. Right top subfigure: Sample path of $S(t)$, the red curve is for system under noisy observation, the blue curve is for system with precise known value of $\alpha(t)$. Bottom subfigure: the sample path of the Markov chain, the red curve is of the signal process, the blue curve is its filter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Example 6.2. Consider (6.4) with $a_1 = 10, b_1 = 1, \sigma_1 = 1, b_2 = 3, \sigma_2 = 1, m_1(0) = 0.1, m_1(1) = 2, m_2(1) = 0.1, q_1 = 10, q_2 = 1$. Direct computations yield $\lambda = 14.8522$, the (exact) threshold determining the longtime behavior (persistence and extinction) for both systems (6.4) and (6.5). Similarly, we can compute $\lambda_0 = -2.5000$, the threshold for system (6.6) and $\lambda_1 = 16.5874$, the threshold for system (6.7). Because $\lambda_1 > \lambda > \lambda_0$, we note that λ_1 is an overcautious threshold. Conversely, λ_0 is an incautious threshold. [It is readily seen that the system is actually permanent, i.e., the disease will not be controlled but λ_0 recommends the disease will be extinct. Conversely, λ_1 is an overcautious threshold which would lead to overly pessimistic expectations.]

Applying our theoretical results to this example, we get that $I(t)$ never converges to 0 and the system is permanent. As before, the systems (6.4) and (6.5) have the same longtime behavior. The numerical simulations of this example are provided in Figs. 2 and 3.

Declaration of competing interest

The authors declare no competing interests.

Data availability

Data will be made available upon request.

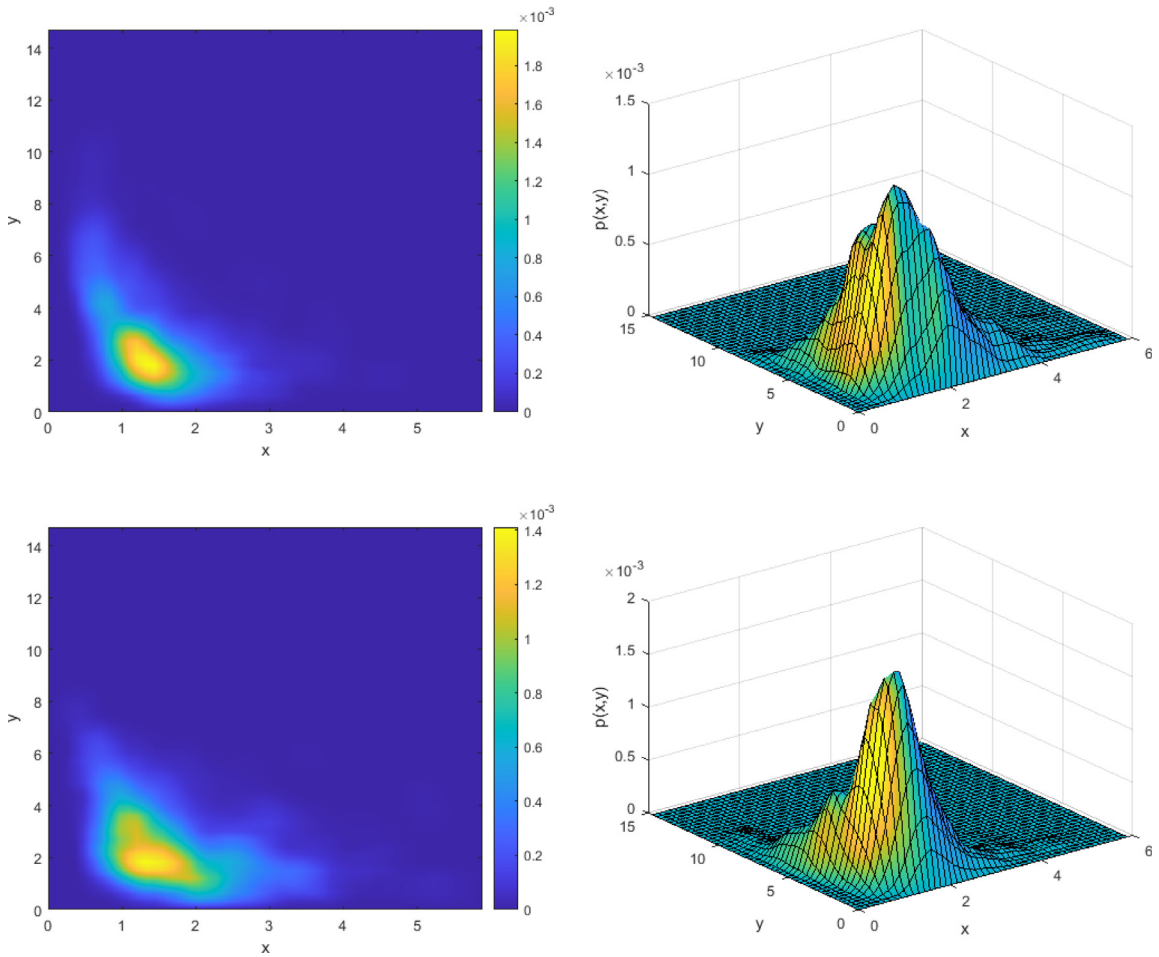

Fig. 3. Left top and right top subfigures: The density of the invariant probability measure (the marginal one in the space of $(S(t), I(t))$) of (6.5) in 2D and 3D settings, respectively. Bottom subfigure: The density of the invariant probability measure (the marginal one in the space of $(S(t), I(t))$) of (6.6) in 2D and 3D settings, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

[1] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics I, Proc. R. Soc. Lond. Ser. A 115 (1927) 700–721.
[2] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics II, Proc. Roy. Soc. Lond. Ser. A 138 (1932) 55–83.
[3] L. Huo, J. Jiang, S. Gong, B. He, Dynamical behavior of a rumor transmission model with Holling-type II functional response in emergency event, Physica A 450 (2016) 228–240.
[4] N.T. Dieu, D.H. Nguyen, N.H. Du, G. Yin, Classification of asymptotic behavior in a stochastic SIR model, SIAM J. Appl. Dyn. Syst. 15 (2016) 1062–1084.
[5] L. Zhang, Z.C. Wang, X.Q. Zhao, Threshold dynamics of a time periodic reaction–diffusion epidemic model with latent period, J. Differential Equations 258 (2015) 3011–3036.
[6] S. Ruan, W. Wang, Dynamical behavior of an epidemic model with a nonlinear incidence rate, J. Differential Equations 188 (2003) 135–163.
[7] Q. Yang, D. Jiang, N. Shi, C. Ji, The ergodicity and extinction of stochastically perturbed SIR and SEIR epidemic models with saturated incidence, J. Math. Anal. Appl. 388 (2012) 248–271.
[8] X.-B. Zhang, H.-F. Huo, H. Xiang, X.-Y. Meng, An SIRS epidemic model with pulse vaccination and non-monotonic incidence rate, Nonlinear Anal. Hybrid Syst. 8 (2013) 13–21.
[9] N.H. Du, N.N. Nhu, Permanence and extinction of certain stochastic SIR models perturbed by a complex type of noises, Appl. Math. Lett. 64 (2017) 223–230.
[10] N.H. Du, N.T. Dieu, N.N. Nhu, Conditions for permanence and ergodicity of certain SIR epidemic models, Acta Appl. Math. 160 (2019) 81–99.
[11] T.D. Tuong, D.H. Nguyen, N.T. Dieu, K. Tran, Extinction and permanence in a stochastic SIRS model in regime-switching with general incidence rate, Nonlinear Anal. Hybrid Syst. 34 (2019) 121–130.
[12] N.H. Du, N.N. Nguyen, Permanence and extinction for the stochastic SIR epidemic model, J. Differential Equations 269 (2020) 9619–9652.
[13] R. Handel, The stability of conditional Markov processes and Markov chains in random environments, Ann. Probab. 37 (2009) 1876–1925.
[14] H. Kunita, Asymptotic behavior of the nonlinear filtering errors of Markov processes, J. Multivariate Anal. 1 (1971) 365–393.
[15] A. Hening, D. Nguyen, Coexistence and extinction for stochastic Kolmogorov systems, Ann. Appl. Probab. 28 (2018) 1893–1942.

[16] H.J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes with applications, J. SIAM Control Ser. A 2 (1964) 106–119.
[17] T.E. Duncan, Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory (Ph.D. Diss), Stanford Univ., 1967.
[18] R.E. Mortensen, Maximum-likelihood recursive nonlinear filtering, J. Optim. Theory Appl. 2 (1968) 386–394.
[19] M. Zakai, On the optimal filtering of diffusion processes, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 11 (1969) 230–243.
[20] M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic differential equations for the non linear filtering problem, Osaka J. Math. 9 (1972) 19–40.
[21] W.M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, SIAM J. Control 2 (1965) 347–369.
[22] G. Kallianpur, C. Striebel, Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes, Theor. Probability Appl. 4 (1969) 597–622.
[23] R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes, I: Applications of Mathematics, (New York), vol. 5, Springer, Berlin, 2001, General theory, Translated from the 1974 Russian original by A.B. Aries, Stochastic Modelling and Applied Probability.
[24] P. Baxendale, P. Chigansky, R. Liptser, Asymptotic stability of the Wonham filter: Ergodic and nonergodic signals, SIAM J. Control Optim. 43 (2004) 643–669.
[25] D. Williams, Probability with Martingales, Cambridge University Press, Cambridge, UK, 1991.
[26] Y.G. Sinaı, Kolmogorov's work on ergodic theory, Ann. Probab. 17 (1989) 833–839.
[27] P. Masani, Wiener's contributions to generalized harmonic analysis prediction theory and filter theory, Bull. Amer. Math. Soc. 72 (1996) 73–125.
[28] T. Tamura, Y. Watanabe, Hypoellipticity and ergodicity of the Wonham filter as a diffusion process. (English summary), Appl. Math. Optim. 64 (2011) 13–36.
[29] D. Nguyen, N. Nguyen, G. Yin, General nonlinear stochastic systems motivated by chemostat models: complete characterization of long-time behavior, optimal controls, and applications to wastewater treatment, Stochastic Process. Appl. 130 (8) (2020) 4608–4642.
[30] X. Mao, C. Yuan, Stochastic Differential Equations with Markovian Switching, Imperial College Press, London, 2006.
[31] X. Mao, Stochastic differential equations and their applications, Horwood Publishing chichester, 1997.
[32] C. Zhu, G. Yin, On strong feller, recurrence, and weak stabilization of regime-switching diffusions, SIAM J. Control Optim. 48 (2009) 2003–2031.
[33] D. Nguyen, G. Yin, C. Zhu, Certain properties related to well posedness of switching diffusions, Stochastic Process. Appl. 127 (2017) 3135–3158.
[34] W. Kliemann, Recurrence and invariant measures for degenerate diffusions, Ann. Probab. 15 (1987) 690–707.