# Randomness accelerates the dynamic clearing process of the COVID-19 outbreaks in China

Sha He[a], Dingding Yan[a], Hongying Shu[a], Sanyi Tang[a], Xia Wang[a,*], Robert A. Cheke[b]

[a]School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, 710119, China
[b]Natural Resources Institute, University of Greenwich at Medway, Central Avenue, Chatham Maritime, Chatham, Kent, ME4 4TB, UK

**Abstract**

During the implementation of strong non-pharmaceutical interventions (NPIs), more than one hundred COVID-19 outbreaks induced by different strains in China were dynamically cleared in about 40 days, which presented the characteristics of small scale clustered outbreaks with low peak levels. To address how did randomness affect the dynamic clearing process, we derived an iterative stochastic difference equation for the number of newly reported cases based on the classical stochastic SIR model and calculate the stochastic control reproduction number (SCRN). Further, by employing the Bayesian technique, the change points of SCRNs have been estimated, which is an important prerequisite for determining the lengths of the exponential growth and decline phases. To reveal the influence of randomness on the dynamic zeroing process, we calculated the explicit expression of the mean first passage time (MFPT) during the decreasing phase using the relevant theory of first passage time (FPT), and the main results indicate that random noise can accelerate the dynamic zeroing process. This demonstrates that powerful NPI measures can rapidly reduce the number of infected people during the exponential decline phase, and enhanced randomness is conducive to dynamic zeroing, i.e. the greater the random noise, the shorter the average clearing time is. To confirm this, we chose 26 COVID-19 outbreaks in various provinces in China and fitted the data by estimating the parameters and change points. We then calculated the MFPTs, which were consistent with the actual duration of dynamic zeroing interventions.

*Keywords:* Stochastic difference model, Newly reported cases, Change point, Mean first passage time, Data fitting.

## 1. Introduction

Emerging and re-emerging infectious diseases pose threats to human life and are of major concern to public health authorities througout the world [1]. COVID-19, which has been prevalent for three years and is still spreading at a high level, has had a devastating impact on life and livelihood worldwide [2]. Epidemiological mathematical models play important roles in understanding the transmission dynamics of the disease. Through modelling research the spread of the virus can be evaluated, and the effects of prevention measures can be analyzed [3]. The classic SIR (Susceptibles-Infectious-Recovered) model for the study of infectious diseases was proposed by Kermack-MacKendrick nearly a century ago [4]. On the basis of this model, there are a series of studies that expand the model by focusing on random factors and prevention measures, and then address their impact on the process of disease transmission [5–8]. The main results of many stochastic differential equation (SDE) epidemic models reveal that the threshold of disease persistence or extinction is lower than the threshold of the corresponding deterministic model under the influence of random noise [9–13].

In order to estimate parameters from daily incidence and mortality series, Lekone et al. developed a stochastic discrete-time susceptible-exposed-infectious-recovered (SEIR) model for infectious diseases [14]. This stochastic model can more realistically reflect the transitions between different compartmental populations by introducing probability densities, and can better investigate the effect of control measures than previous methods. For different data sets, including incomplete data on infectious disease, Markov chain Monte Carlo methods are used for the analyses. The method is flexible and can applied on a wide range of models [15]. During an epidemic of infectious diseases, the number of daily reported cases is one of the most commonly available and useful types of surveillance data. The curve of daily reported case counts contains a lot of useful information. This trend not only reveals changes in transmissibility, but also predicts the follow-up

*Corresponding author
*Email address:* xiawang@snnu.edu.cn (Xia Wang)

of the epidemic through indicators such as inflection points and epidemic sizes [16]. Therefore, research on the changes in the number of newly reported cases is meaningful.

To contain the spread of emerging infectious diseases, various non-pharmaceutical interventions (NPIs) will be adopted such as case isolation, contact tracing and quarantine, school closures, travel restrictions, etc. [17, 18], such that the spread of the virus is no longer natural transmission [19]. The effective reproductive number $R_t$, a quantity varying with time, is the expected number of new infections caused by an infectious individual in a population, which can assess the effectiveness of interventions. In general, when $R_t > 1$ the intervention measurs are needed to cope with the growing incidence, whereas $R_t > 1$ is a sign of the epidemic has enter a controlled phase [20]. The disease can normally be eliminated after a single wave if strict and sustained control measures are taken. Multiple waves may occur as a result of behavioral changes, mutation of virus strains or medical practices. Before 1 September 2022, there were more than 100 outbreaks induced by different COVID-19 virus strains in China. Under the strong NPIs, the goal of zero clearing could usually be achieved in about 40 days. Therefore, each outbreak was characterized by being of small scale and short duration, which permitted multiple random factors to have huge impacts on the numbers of daily reported cases.

For many outbreaks, especially the COVID-19 epidemic, there have been multi-wave transmission pattern [21–24]. The changing intensity of epidemic prevention measures, vaccinations, information updates and behavioral changes drive waves of the COVID-19 pandemic or rebound of the disease [25]. Various studies have proposed mechanisms to shap multiple waves of epidemic [26–28]. Considering the incubation-period from infection to diagnosis for the fraction of population infected, Blonigan et al. present a Bayesian method to infer and forecast a multiwave outbreak of COVID-19 pandemic. They fitted one wave and multi-wave models to the data from New Mexico, California, and provide 7-10 days ahead short-term forecasts [29]. Xu et al. have presented a series of models to infer the mechanisms driving two-wave outbreaks. They analyzed factors involved including heterogeneities in host population, pathogens, space and time and combinations of them for the predicting the second wave of infection [30].

A key question arise when the transmission ability changes in one or more waves of an epidemic, but how can this time point be obtained? Inferring changes of the reproductive number given a reported daily cases incidence curve is crucial. Lim et al. have established a Bayesian regime switching (BRS) model to infer epidemic and endemic regimes of dengue transmissions. The growth and decline regimes of dengue cases are specific autoregressive process [31]. In our previous research on air pollution and respiratory diseases, we estimated the series of change points in the air quality index (AQI) data between 2010 and 2016 using a Bayesian method [32]. However, if the regime switch can be considered on the basis of stochastic compartmental epidemic model and the switch points are unknown, the changes of disease transmissibility can be reflected and understand more intuitively. In this study, we will consider that the reproductive number varies in stages during a wave in an epidemic, and one of the problems we want to study is how to derive the time change point based on the data on the numbers of daily reported cases. Another problem we are concerned about is that the length of time for disease extinction after the curve of daily reported cases entering a declining phase. The pathes for the stochastic model are random and the time when the disease extinction is obtained according to each path is different. Therefore, when we study the extinction time of the stochastic infectious disease model, we use the definition of mean first passage time (MFPT) to analyze it. First passage problems can be most simply described as finding the distribution of times according to which a random process first exceeds a prescribed threshold [33]. There are many applications of first passage time (FPT) in stochastic processes and such research has been conducted on solving to FPT problems in theoretical and computational studies [34–36]. The FPT of a stochastic nonlinear generalization of the logistic model was analysed by Floris et al [37]. Srivastava et al. studied a multistage stochastic process with two absorbing boundaries. By using martingale theory, they analysed the properties of the mean decision time, hitting probabilities, and FPT densities of the process [38].

In this paper, we consider a stochastic SIR epidemic model with NPIs implemented. Based on the dynamic properties of the SIR model, we will establish a stochastic model to characterize changes in the number of daily reported cases and investigate the dynamic of incidence curve. In order to for the model to better correspond with the actual data, the discrete state stochastic model structure is deduced and then the threshold for disease outbreak or extinction is obtained. The epidemic process will deviate from the original trajectory and change due to the implementation of control measures. It is therefore useful to find the time point when the curve of the number of daily reported cases enters the declining stage. To this aim, it is interesting to consider switching dynamical systems in which the switching time of the declining stage caused by the effectiveness of the control measures is unknown. Determining the point in time when switching occurs is one of the focuses of this study. Furthermore, we will calculate the average time from the change point to disease extinction in the stochastic model using the relevant theory of FPT. The spread of COVID-19 in various provinces in China has shown multiple waves. With the mutation of the virus, the epidemic has repeatedly broken out, but the epidemic waves experienced are different due to the different time and intensity of the prevention measures taken. As

a case study, we will illustrate the main factors affecting the process of the epidemic through data fitting and numerical analysis of multi-wave data in China, including the use of generalized additive models and sensitivity analysis methods.

## 2. Model formulation

### 2.1. Deterministic model

Take the prevention and control of COVID-19 in China as an example, most infected people will not have experienced a complete process from infection to incubation period as a result of the powerful high-frequency nucleic acid testing. Thus, considering the contact tracing and quarantine measures strictly taken to prevent infectious disease, we develop the model based on the simplest Susceptible-Infected-Recovered (SIR) epidemiological model. The susceptible are divided into two parts: quarantined and unquarantined. The infected population are also divided into two parts: isolated and not isolated. The total population is assumed to be a constant denoted as $N$, since we do not consider death here. We denote the quarantined proportion of individuals is $q$. The population are divided into the following groups: susceptible ($S$), infected ($I$), quarantined susceptible ($S_q$), quarantined infected ($I_q$) and recovered ($H$). The epidemic model can be constructed as follows,

$$\begin{cases} \frac{dS}{dt} = -\frac{(\beta c + cq(1-\beta))SI}{N}, \\ \frac{dI}{dt} = \frac{\beta c(1-q)SI}{N} - \gamma I, \\ \frac{dS_q}{dt} = \frac{(1-\beta)cqSI}{N}, \\ \frac{dI_q}{dt} = \frac{\beta cqSI}{N} - \delta_q I_q, \\ \frac{dH}{dt} = \gamma I + \delta_q I_q, \end{cases} \tag{1}$$

In the model, $\beta$ is the transmission probability per contact and $c$ is the contact rate. The confirmation rate of the quarantined infected people is assumed as $\delta_q$. $\gamma$ represents the actual transmission period of infected cases, which is closely related to the intensity of the control measures implemented. Due to the implementation of the contact tracing and quarantine measures, and the powerful high-frequency nucleic acid testing, infected individuals are usually diagnosed before experiencing a complete process from infection to incubation period. Therefore, new infections will be immediately detected and reported. We use the number of reported cases replace the number of new infections in our model. Further, we assume that the isolated population will no longer return to the susceptible population. When the proportion of infected individuals to the total population is small, the number of susceptible individuals is very close to the total population. It can be considered that the susceptible population is approximately equal to the total population $N$. By calculating the data infection cases in multiple waves in China, it can be found that the ratio $S/N$ is close to 1 with strictly NPIs. Therefore, we assume that $S/N \approx 1$ and the model can be simplified as

$$\begin{cases} \frac{dI}{dt} = \beta c(1-q)I - \gamma I, \\ \frac{dI_q}{dt} = \beta cqI - \delta_q I_q. \end{cases} \tag{2}$$

Our model is applied to characterize situations where the infection scale is small compared to the total population size. Each outbreak experienced in mainland China before the relaxation has this characteristic. We focus on the change of the number of infected cases, and the equations of $S$, $S_q$ and $H$ are independent of $I$ and $I_q$, so they can be omitted in the model. The reproductive number of model (1) denote as $R_d = \frac{\beta c(1-q)}{\gamma}$ can be calculated. Then the model (1) can also be rewritten as follows

$$\begin{cases} \frac{dI}{dt} = \gamma(R_d - 1)I, \\ \frac{dI_q}{dt} = \frac{\gamma q}{1-q}R_d I - \delta_q I_q. \end{cases} \tag{3}$$

Since the actual reported data are usually the number of daily confirmed cases, next we will derive the equation for newly reported cases. Solving the first equation we have

$$I(t) = I_0 \exp[\gamma(R_d - 1)t].$$

Substituting it into the second equation yields

$$\frac{dI_q}{dt} = \beta cqI_0 \exp[\gamma(R_d-1)t] - \delta_q I_q.$$

Solving it with initial value $I_q(0)=0$, gives

$$I_q(t) = \left[\frac{\beta cqI_0 \exp[(\gamma R_d - \gamma + \delta_q)t] - \beta cqI_0}{\gamma R_d - \gamma + \delta_q}\right] \exp(-\delta_q t).$$

There are two sources for obtaining the numbers of newly reported cases: people diagnosed from the isolated population and those screened from the unisolated population. Data from the isolated population can be fitted by $\frac{\gamma q}{1-q} R_d I$ and from the unisolated population by $\gamma R_d I$. Therefore, the total number of reported cases denoted as $M(t)$ can be calculated as follows

$$M(t) = \gamma R_d I(t) + \frac{\gamma q}{1-q} R_d I(t) = \frac{\gamma}{1-q} R_d I(t) = \frac{\gamma R_d I_0}{1-q} \exp[\gamma(R_d-1)t].$$

Then we have the form of the difference iteration as

$$\begin{aligned} M(t+1) &= \frac{\gamma R_d I_0}{1-q} \exp[\gamma(R_d-1)(t+1)] \\ &= M(t) \exp[\gamma(R_d-1)]. \end{aligned} \tag{4}$$

It can be clearly seen from the above equation that $M(t)$ is an increasing sequence when $R_d > 1$, and conversely $M(t)$ is a decreasing sequence when $R_d < 1$. However, the transmission process of infectious diseases is inherently random, being affected by random factors such as environmental noise. Next we will derive a stochastic model to further describe the changes in the number of newly reported cases.

*2.2. Stochastic model*

Suppose that the number of potentially infectious contacts between an infectious individual and another individual has a normal distribution whose mean is denote as $c$ and variance scale is $\sigma^2$. Using the properties of Brownian motion, we can express the change of contact rate in a small subsequent time interval $cdt + \sigma dB(t)$. Therefore we replace $cdt$ by $cdt + \sigma dB(t)$ in (1) to get the corresponding SDE model as follows

$$\begin{cases} dI = [\beta c(1-q) - \gamma] I dt + \sigma\beta(1-q) I dB(t), \\ dI_q = [\beta cqI - \delta_q I_q] dt + \sigma\beta qI dB(t). \end{cases} \tag{5}$$

We first discuss the threshold of the extinction and persistence of the SDE model (5). Because only those who are not isolated are contagious, we only need to study the behavior of $I(t)$. Define

$$R_s = \frac{\beta c(1-q)}{\gamma} - \frac{\sigma^2\beta^2(1-q)^2}{2\gamma}$$

as the stochastic control reproduction number (SCRN) for SDE model (5), then we have the following property:

**Theorem 2.1.** *If $R_s < 1$, then for any given initial value $I(0) = I_0 \in (0,N)$, $I(t)$ tends to zero exponentially almost surely. That is, the disease become extinct with probability one.*

*Proof.* Apply Itô formula, we have

$$\begin{aligned} \log(I(t)) &= \log I_0 + \int_0^t [\beta c(1-q) - \gamma - \frac{1}{2}\sigma^2\beta^2(1-q)^2] ds + \int_0^t \sigma\beta(1-q) dB(s), \\ &= \log I_0 + [\beta c(1-q) - \gamma - \frac{1}{2}\sigma^2\beta^2(1-q)^2] t + \int_0^t \sigma\beta(1-q) dB(s). \end{aligned}$$

Then we can get

$$\limsup_{t\to\infty} \frac{1}{t} \log I(t) \le \beta c(1-q) - \gamma - \frac{1}{2}\sigma^2\beta^2(1-q)^2 + \limsup_{t\to\infty} \frac{1}{t} \int_0^t \sigma\beta(1-q) dB(s). \tag{6}$$

Making use of the large number theorem for martingales, we have

$$\limsup_{t\to\infty}\frac{1}{t}\int_0^t \sigma\beta(1-q)dB(s)=0.$$

Therefore,

$$\limsup_{t\to\infty}\frac{1}{t}\log I(t)\le \beta c(1-q)-\gamma-\frac{1}{2}\sigma^2\beta^2(1-q)^2, \tag{7}$$

then it is shown that when $R_s<1$, $\limsup_{t\to\infty}\frac{1}{t}\log I(t)\le 0$. That means the solution of the fist SDE in model (5) tends to zero exponentially almost surely. This completes the proof. □

Now we derive the stochastic equation for the number of newly reported cases base on the equation in (5). Solving the first SDE of $I(t)$, we have

$$I(t)=I(0)\exp\{(\beta c(1-q)-\gamma-\frac{1}{2}\sigma^2\beta^2(1-q)^2)t+\sigma\beta(1-q)B(t)\}. \tag{8}$$

As previously explained, the number of newly reported cases has two sources, one from the isolated population and the other from the unisolated population. Newly reported cases diagnosed from the isolated population can be fitted to $\beta c(1-q)I(t)dt+\sigma\beta qI(t)dB(t)$ and newly reported cases from the screened unisolated population can be fitted to $\beta cqI(t)dt+\sigma\beta qI(t)dB(t)$. Thus, the total number of daily reported cases at time $t$ obtained according to the stochastic model (5) is denoted as $Ms(t)$ and can be calculated when the time difference $dt$ equals to 1. $Ms(t)$ is calculated as follows

$$\begin{aligned} M_S(t) &= \beta cI+\sigma\beta I(t)(B(t)-B(t-1)) \\ &= \beta cI+\sigma\beta I(t)\varepsilon_t \\ &= (\beta c+\sigma\beta\varepsilon_t)I(0)\exp\{(\beta c(1-q)-\gamma-\frac{1}{2}\sigma^2\beta^2(1-q)^2)t+\sigma\beta(1-q)B(t)\}. \end{aligned}$$

Similarly, we have the form of difference iteration as

$$\begin{aligned} M_S(t+1) &= (\beta c+\sigma\beta\varepsilon_{t+1})I(0)\exp\{(\beta c(1-q)-\gamma \\ &\quad -\frac{1}{2}\sigma^2\beta^2(1-q)^2)(t+1)+\sigma\beta(1-q)B(t+1)\} \\ &= M_S(t)\frac{c+\sigma\varepsilon_{t+1}}{c+\sigma\varepsilon_t}\exp\{(\beta c(1-q)-\gamma-\frac{1}{2}\sigma^2\beta^2(1-q)^2)+\sigma\beta(1-q)\varepsilon_t\}, \end{aligned}$$

where $\varepsilon_t=B(t)-B(t-1)$, so that $\varepsilon_t$ $(t=1,2\cdots)$ are independent and all obey a standard normal distribution with mean value of 0 and variance of 1, i.e. $\varepsilon_t\sim N(0,1)$. From the expression of $R_s$, we have

$$\beta(1-q)=\frac{c\pm\sqrt{c^2-2\sigma^2R_s\gamma}}{\sigma^2}.$$

From the limit state when $\sigma$ tends to 0, it can be judged that only the minus sign can be taken in the above formula. Then the above equation can be rewritten as

$$M_S(t+1)=M_S(t)\frac{c+\sigma\varepsilon_{t+1}}{c+\sigma\varepsilon_t}\exp\{\gamma(R_s-1)+\frac{c-\sqrt{c^2-2\sigma^2R_s\gamma}}{\sigma^2}\varepsilon_t\}. \tag{9}$$

Moving $c+\sigma\varepsilon_t$ to the left of the equation, we can get the iterative equation

$$\frac{M_S(t+1)}{c+\sigma\varepsilon_{t+1}}=\frac{M_S(t)}{c+\sigma\varepsilon_t}\exp\{\gamma(R_s-1)+\frac{c-\sqrt{c^2-2\sigma^2R_s\gamma}}{\sigma^2}\varepsilon_t\}. \tag{10}$$

Define $Z_t = \dfrac{M_S(t)}{c+\sigma\varepsilon_t}$, and the above equation can be written as

$$Z_{t+1} = A_t Z_t, \tag{11}$$

where $A_t = \exp\{\gamma(R_s-1) + \frac{c-\sqrt{c^2-2\sigma^2 R_s\gamma}}{\sigma^2}\varepsilon_t\}$. $A(t)$ is a sequence of identically distributed (i.i.d) $\mathbb{R}$-valued random variables. Starting from a real-valued random variable $Z_0$, which is always assumed to be independent of the i.i.d sequence $A_t$. We want to investigate in which cases $Z_t$ converges in distribution, study its limit distribution and investigate the limit property of $M_S(t)$.

**Lemma 2.1.** *If $Z_t \xrightarrow{d} Z$ for some $\mathbb{R}$-valued random variable $Z$, then $Z$ satisfies the stochastic equation*

$$Z \overset{d}{=} AZ, \;\; Z \text{ and } A \text{ independent}, \tag{12}$$

*where $A$ is a random variable with the same distribution as $A_n$. $\xrightarrow{d}$ is denoted as convergence in distribution, and $\overset{d}{=}$ is denoted as equality in distribution.*

*Proof.* For $\mathbb{R}^2$, we have $(A_t, Z_t) \xrightarrow{d} (A, Z)$. Since $Z$ and $A$ are independent, $A_t Z_t$ converges in distribution to $AZ$. It is easy to get the desired conclusion (12). □

In order to investigate the limit property of $Z(t)$, we iterate (12) and obtain

$$Z_n = A_n A_{n-1} \cdots A_1 Z_0. \tag{13}$$

This shows $Z_n$ depend on the initial value $Z_0$. We write $Z_n = Z_n(Z_0)$ to display the relevance on $Z_0$. With different starting $Z_0^{'}$ and $Z_0^{''}$, we obtain

$$Z_n(Z_0^{'}) - Z_n(Z_0^{''}) = A_t A_{t-1} \cdots A_1 (Z_0^{'} - Z_0^{''}). \tag{14}$$

The limiting behaviour of $Z_t$ can be decided by that of the random sequence $(\sum_{k=1}^{t} \log A_k)_{t=1}^{\infty}$. Assume that

$$\mu := E \log |A|,$$

in case the expectation exists, finite or infinite. By the strong law of large numbers

$$\frac{1}{n}\sum_{k=1}^{t} \log |A_k| \to \mu \quad w.p.1.$$

**Theorem 2.2.** *If $R_s < 1$, $M_S(t)$ converges in degenerate distribution 0 for all $M_S(0)$.*

*Proof.* By summing the logarithms of $A_t$, it is easy to get

$$\sum_{k=1}^{t} \log |A_k| = \gamma(R_s-1)t + \frac{c-\sqrt{c^2-2\sigma^2 R_s\gamma}}{\sigma^2}\sum_{k=1}^{t}\varepsilon_k.$$

Since $\varepsilon_t \sim N(0,1)$, $-\infty \le \mu < 0$. If $R_s < 1$, then $\sum_{k=1}^{t} \log |A_k| \xrightarrow{d} -\infty$. We have

$$A_t A_{t-1} \cdots A_1 Z_0 \xrightarrow{d} 0. \tag{15}$$

The solution of (12) is unique in distribution and $Z_t(Z_0)$ converges in distribution for all $Z_0$. The distribution is degenerate distribution 0. This completes the proof. □

The threshold $R_s$ we refer to here as the reproductive number of the stochastic model. In general, the SCRN will change due to the intervention of factors such as the implementation of control measures. The magnitude relationship of this threshold to 1 can account for disease outbreaks and extinctions. Therefore, we next explore the characteristics and switching between different stages of epidemic process.

## 3. Phase of epidemic and Change points

Typically, a wave of outbreaks has two main phases from start to end, an exponential rise and decline. Sometimes, such a wave will have a plateau, the number of cases fluctuates around a mean level. Therefore, when we use model (10) to describe the change in the number of new cases, a wave can be divided into three stages through changes in the parameter $R_s$. The three phases are identified as follows:

(1) Increasing (exponential) phase. The spread of the epidemic is an exponential process at the beginning, and the SCRN is greater than one, i.e. $R_s > 1$;

(2) Plateau phase. In this stage, the numbers of new cases remains at about the same level, and the SCRN at this time is approximately equal to 1, i.e. $R_s = 1$;

(3) Extinction phase. After the control measures are implemented, the epidemic is under control, and the number of new infections will decrease. At this time, the SCRN is less than 1, i.e. $R_s < 1$.

The actual data shows volatility, it is hard to determine the time point when a wave switches between phases based on the actual data. According to the above analysis, the SCRN changes with time, for convenience we denote it as $R(t)$. To investigate the change point of the reproductive number, we consider that the reproductive number is affected by environmental noise and suppose that the reproductive number $R(t)$ is perturbed over time. We establish the SDE model as follows

$$dI(t) = \gamma(R(t)-1)Idt + \sigma_1\gamma I dB(t). \tag{16}$$

The switch of reproductive number occurs when the implementation control intensity changes. We denote $T_1$ and $T_2$ as the change points, and $R(t)$ is a piecewise function which has the following form

$$R(t) = \begin{cases} R_1, & t \le T_1 \\ 1, & T_1 < t \le T_2 \\ R_2, & t > T_2. \end{cases} \tag{17}$$

If a wave does not have a plateau, there is only one change point, and $T_1$ is equal to $T_2$ in the piecewise function. When we use the equation (16) to estimate the change point in the data, it is difficult to obtain the data corresponding to the number of existing cases represented by $I(t)$ in the equation. In order to better use the reported data to estimate the change point of the SCRN, we do not consider the recovered cases in the model. Then the cumulative number of infected cases $I_c(t)$ satisfies

$$dI_c(t) = \gamma R(t) I_c dt + \sigma_1\gamma I_c dB(t) \tag{18}$$

The cumulative number of cases is easily obtained from reports. We use the method of Bayesian statistical inference to estimate the change point. To get the posterior distribution of the change point, we start with the difference form of the solution of (18). The approximate solution satisfies the following discrete equation:

$$I_{c,k+1} = I_{c,k} + \alpha(I_{c,k},\Theta)\Delta t + \hat{\sigma}(I_{c,k},\Theta)\omega_k, \quad k = 1,\cdots,N', \tag{19}$$

where $N'$ denotes the length of the data set and $\alpha = \gamma R(t) I_c$, $\hat{\sigma} = \sigma_1\gamma I_c$, $\omega_k \sim \mathrm{N}(0,\Delta t)$. $\omega_k$ represents the increment of the Wiener process and obeys the normal distribution with mean zero and variance $\Delta t$. The time step $\Delta t$ depends on the interval between data observations. The estimation of the unknown parameter vector is based on the time series of the observed cumulative number of infected cases, which is represented by $I_c = (I_{c,1}, I_{c,2}, \cdots, I_{c,N'})'$. The likelihood function of $I_c$ is:

$$p(I_c \mid \Theta) = \prod_{k=1}^{N'-1} \frac{1}{\sqrt{2\pi\Delta t}\sigma_1\gamma I_{c,k}} \exp\Big(-\frac{1}{2(\sigma_1\gamma I_{c,k})^2}\frac{(I_{c,k+1} - I_{c,k} - \gamma R(t) I_{c,k}\Delta t)^2}{I_{c,k}\Delta t}\Big), \tag{20}$$

In general, the cumulative number of infected cases is collected every day and so the time step $\Delta t = 1$ is small enough to guarantee that the approximation of the SDE equation is good. The likelihood function can be simplified to the same form as the Bayes' linear regression model. Then we use the probability density function of the prior distribution and the derived posterior distribution of the change point in reference [32].

## 4. Extinction time

In fact, after the development of the epidemic enters the declining phase, what we are concerned by is when the disease can be cleared, that is, how long it takes for the number of infected cases $I(t)$ to reach 0. This can be understood as the

time when a certain state or threshold is reached for the first time in the stochastic model. We define $T_{\xi,x_0}$ as the time when $I(t)$ first crosses a threshold $\xi$ or in other words a barrier. We denote this as

$$T_{\xi,x_0} = \inf\{t > t_0 : I(t) \leq \xi; I(t_0) = x_0\}, \tag{21}$$

the FPT of $I(t)$ through the constant boundary $\xi < x_0$. We focus on calculating the MFPT from a state $x_0$ to the final state $\xi$, which is denoted as $\tau_{\xi,x}$. Refer to [39], $\tau_{\xi,x}$ for the stochastic model (5) is the solution to the backward equation

$$[\beta c(1-q) - \gamma]x\frac{\partial \tau_\xi}{\partial x} + \frac{1}{2}\sigma\beta(1-q)x\frac{\partial^2 \tau_\xi}{\partial x^2} = -1. \tag{22}$$

The boundary conditions are $\tau_\xi(\xi) = 0$, $\lim_{x\to\infty}\frac{\partial \tau_\xi}{\partial x} = 0$. Since $\sigma\beta(1-q)x > 0$ for all $x > 0$. Then equation (22) can be rewritten:

$$\frac{\partial^2 \tau_\xi}{\partial x^2} + \frac{2[\beta c(1-q)-\gamma]}{\sigma\beta(1-q)}\frac{\partial \tau_\xi}{\partial x} = -\frac{2}{\sigma\beta(1-q)x}. \tag{23}$$

Denote $y_\xi(x) = \frac{\partial \tau_\xi}{\partial x}$ and $a = \frac{2|\beta c(1-q)-\gamma|}{\sigma\beta(1-q)}$, then we have

$$\frac{\partial y_\xi(x)}{\partial x} + a y_\xi(x) = -\frac{2}{\sigma\beta(1-q)x}. \tag{24}$$

The solution of the first-order equation is

$$y_\xi(x) = \exp(a(\xi - x))\left[c_1 - \int_\xi^x \frac{2}{\sigma\beta(1-q)s}\exp(a(x-\xi))ds\right]. \tag{25}$$

Setting $c_1 = \int_\xi^\infty \frac{2}{\sigma\beta(1-q)s}\exp(a(s-\xi))ds$, then

$$y_\xi(x) = \exp(a(\xi - x))\int_x^\infty \frac{2}{\sigma\beta(1-q)s}\exp(a(s-\xi))ds. \tag{26}$$

Now, to recover the FPT integrate $y_\xi(x)$:

$$\tau_\xi(x) = \int_\xi^x \int_\theta^\infty \frac{2}{\sigma\beta(1-q)s}\exp(a(s-\theta))dsd\theta. \tag{27}$$

So far, we have obtained the explicit expression MFPT. We will calculate its value in the later part of the data analysis based on the fitting of the parameters and compare it with the numerical results. Here, we further explore the difference in the time to reach the threshold between MFPT obtained from the stochastic model and obtained by the deterministic model. The following theorem gives a conclusion for the comparison of the stochastic model with the corresponding deterministic model.

**Theorem 4.1.** *Assume that the number of infective cases at the change point $T_c$ is $I_p$, and the time required to reduce the number of infective cases from $I_p$ to $I_{end}$ based on the deterministic model and stochastic model are denoted as $T_{o,end}$ and $T_{s,end}$, respectively, then we have the mean time from $I_p$ to $I_{end}$ taken from the stochastic model is shorter than in the deterministic model, i.e. $T_{o,end} > E[T_{s,end}]$.*

*Proof.* For the deterministic model, we have

$$\frac{1}{I}dI = [\beta c(1-q) - \gamma]dt.$$

By integrating both sides of the equation,

$$\int_{I_p}^{I_{end}} \frac{1}{I}dI = \int_{T_c}^{T_{o,end}} [\beta c(1-q) - \gamma]dt.$$

Then, we can get

$$T_{o,end} = T_c + \frac{\ln(I_{end}) - \ln(I_p)}{\beta c(1-q) - \gamma} = T_c + \alpha_1(\ln(I_{end}) - \ln(I_p)), \tag{28}$$

where $\alpha_1 = \dfrac{1}{\beta c(1-q)-\gamma}$.

For the stochastic model, we define a function $\Phi : (0,N) \to \mathbb{R}_+$ by $\Phi(I) = \ln([\beta c(1-q)-\gamma]I)$. By the Itô formula, we have

$$d\Phi = (1 - \frac{\sigma^2\beta^2(1-q)^2}{2[\beta c(1-q)-\gamma]})dt + \frac{\sigma\beta(1-q)}{\beta c(1-q)-\gamma}dB(t). \tag{29}$$

Integrating both sides of equation (29) from $T_c$ to $T_{s,end}$ shows that

$$\begin{aligned} &\frac{1}{[\beta c(1-q)-\gamma]}(\ln(I_{end}) - \ln(I_p)) = \\ &(1 - \tfrac{\sigma^2\beta^2(1-q)^2}{2[\beta c(1-q)-\gamma]})(T_{s,end} - T_c) + \frac{\sigma\beta(1-q)}{\beta c(1-q)-\gamma}[B(T_{s,end}) - B(T_c)]. \end{aligned} \tag{30}$$

Then we can obtain the explicit expression of the extinction time $T_{s,end}$ as

$$T_{s,end} = T_c + \frac{\ln(I_{end}) - \ln(I_p) - \sigma\beta(1-q)(B(T_{s,end}) - B(T_c))}{\beta c(1-q) - \gamma - \frac{1}{2}\sigma^2\beta^2(1-q)^2}. \tag{31}$$

By transformation, we have

$$T_{s,end} + \alpha_3 B(T_{s,end}) = \alpha_2(\ln(I_{end}) - \ln(I_p)) + T_c + \alpha_3 B(T_c), \tag{32}$$

where $\alpha_2 = \dfrac{1}{\beta c(1-q)-\gamma-\frac{1}{2}\sigma^2\beta^2(1-q)^2}$, $\alpha_3 = \dfrac{\sigma\beta(1-q)}{\alpha_2}$. According to the properties of Brownian motion, we can calculate the mean of $T_{s,end}$ as

$$E[T_{s,end}] = \alpha_2(\ln(I_{end}) - \ln(I_p)) + T_c. \tag{33}$$

It can be seen that $\alpha_2 < \alpha_1$ and $I_{end} < I_p$ in the decreasing phase. Then we can judge that $T_{o,end} > E[T_{s,end}]$. This completes the proof. □

**Remark 4.1.** *The extinction time of the number of infected cases $I(t)$ is consistent with that of the number of the newly confirmed cases $Ms(t)$.*

The theorem illustrates that the random noise in the model has an effect on disease extinction and this effect is positive. The random noise can speed up disease extinction. This will also be verified in the following numerical experiments.

## 5. Results from the data

So far, the COVID-19 pandemic has lasted for three years, and health authorities have taken many measures against the disease. In China, large or small outbreaks occurred in various provinces and cities at different times. The common feature of the many waves is that after the early spread, the government took strict intervention measures to suppress the spread of the virus. Our stochastic infectious disease model with reproductive number change has characterized such patterns well. So far, there have been more than 100 outbreaks in China. Since the initial epidemic of the original strain, the virus continues to mutate. The Delta strain first appeared and spread in Guangdong on 21 May 2021, and the Omicron strain first appeared and spread in Tianjin on 8 January 2022. We collected data on the numbers of daily confirmed cases in Guangdong, Shaanxi and other provinces from the National Health Commission to explore the characteristics of the transmission of different virus strains and the main factors affecting the transmission and control of the epidemic through the analysis of multi-wave epidemic data based on the stochastic model.

*5.1. Data fitting*

We sorted the data on multiple waves and counted the time when the number of daily confirmed cases in each wave dropped to zero. This gives the extinction time obtained from the data, as shown in Table 2. However, there may be a rebound after the number of daily confirmed cases become zero for the first time. We performed parameter estimation from the data on the number of cases from the start of reporting to the first zero. Based on the posterior distribution form of the change points, we inferred the time point $T_1$ at which the SCRN changes. When we do parameter estimation, the value of $\gamma$ in the model is fixed. Due to the implementation of nucleic acid screening, tracking and isolation and other control measures, there is a difference between the actual transmission period before and after the measures are implemented. Thus, we assume that $\gamma$ is $1/5$ during the increasing phase and $1/2$ during the extinction phase. The reproductive number $R_1$ for the increasing phase and $R_2$ for the extinction phase are obtained by fitting the data. Other parameters including noise intensity $\sigma_1$ and $\sigma_2$, contact rates, $c_1$ and $c_2$ in the increasing phase and the extinction phase respectively, were also estimated by fitting the data. The estimated values of the unknown parameters and change points for 26 waves selected from different provinces in China are listed in Table 1.

Fig.1 shows the fitting results for three waves in Guangdong, Shaanxi and Hebei. The first wave in Guangdong occurred from 22 January to 25 March, 2020, and the prevalent strain was the original strain. Our estimated change point is 9, which is marked with a vertical red line in the Fig.1 (a). The number of daily confirmed cases began to gradually decrease on the 9th day, and dropped to 0 for the first time on the 34th day. Our parameter estimation results show that the reproductive number in the increasing phase is 2.04, and in the extinction phase is 0.62. The Fig.1 (a) shows the results of 100 stochastic simulations and the corresponding average values. The fitting of the data of daily confirmed cases shows that $Ms(t)$ in the stochastic difference model (9) is in good agreement with the actual data. Our simple model with switching can well depict the development trend of the epidemic. The area behind the green line is the range of the small-scale rebound of the epidemic. A numerical result of the mean extinction time and its confidence interval can be obtained through multiple stochastic simulation paths. We take the starting point of the extinction phase, that is, the value of the change point, as the initial value parameter, and bring the estimated values of other parameter into the explicit expression of the MFPT, then we can obtain a set of theoretical results for MFPT. The mean value and confidence interval are listed in Table 2. By comparing the results shown in Table 2, we can see that the results from three different waves are consistent. Since the Delta virus entered Guangdong on 22 May, 2021, the second wave of the epidemic began in that region. The following third wave in Guangdong lasted from 12 February to 25 March, 2022, and its prevalent strain was the omicron variant with stronger transmission ability than the other two variants. This wave was the longest among the three waves, and its rebound lasted the longest. Fig.1 (b) and (c) show the fitting results of Delta and Omicron waves, respectively. In the same way, we fitted and analyzed the 26 waves of Inner Mongolia and other provinces and cities. The fitting results are shown in Appendix S1. The duration of the rebound can reflect the difficulty of controlling the outbreaks. By comparing the duration of the rebound in 26 waves of the epidemic, we found that five out of six provinces experienced a longer rebound time for outbreaks of Omicron strain than both the original and alpha strains. The difference in rebound time under different strains shows that as the virus mutates, it becomes more difficult to control the epidemic.

*5.2. Results of regression analysis*

Multiple linear regression was used to determine which factors affected the epidemic situation significantly. The four variables including initial daily infected number $Ms_0$, the reproductive number in the increasing phase $R_1$, the reproductive number in the decreasing phase $R_2$, the noise intensity in increasing phase $\sigma_1$, the noise intensity in the decreasing phase $\sigma_2$, and the change point $T$ we chosen for analysis. For convenience, we denote these six independent variables as $x_1, x_2, \cdots, x_6$ in turn. The dependent variable is the cumulative number of cases in a wave, which is denoted by $y$. In building the multiple linear regression model, the training data set consisted of the data on 26 waves shown in Table 1. The results of regression analyses are shown in Table 3. From the regression output, it can be seen that only the independent variable $x_3$ has a significant linear effect on the dependent variable, for which the regression coefficient is 0.75. The rest of the variables have a low degree of explanation for the cumulative number of cases. In order to determine the specific relationships between the other variables and the cumulative number of cases, we propose a generalized additive model for which the specific expressions are as follows,

$$\begin{aligned} y &= x_3 + ns(x_1, df = 2) + ns(x_2, df = 2) + ns(x_4, df = 1) \\ &\quad + ns(x_5, df = 3) + ns(x_6, df = 3). \end{aligned} \tag{34}$$

Table 1: Parameter estimates for various provinces and cities in China.

| Provinces | Virus strain | Contact rate | | Noise density | | Reproductive number | | Initial infected | Change points |
|---|---|---|---|---|---|---|---|---|---|
| | | $c_1$ | $c_2$ | $\sigma_1$ | $\sigma_2$ | $R_1$ | $R_2$ | $Ms_0$ | $T_1$ |
| Guangdong | Alpha | 6.92 | 1.07 | 1.03 | 0.30 | 2.04 | 0.62 | 20 | 9 |
| | Delta | 6.99 | 1.07 | 1.16 | 0.32 | 2.12 | 0.52 | 1 | 13 |
| | Omicron | 6.90 | 1.05 | 1.33 | 0.26 | 1.96 | 0.72 | 1 | 29 |
| Shaanxi | Alpha | 7.01 | 1.15 | 1.51 | 0.28 | 1.81 | 0.43 | 3 | 13 |
| | Delta | 6.94 | 1.09 | 1.25 | 0.31 | 2.42 | 0.58 | 1 | 19 |
| | Omicron | 7.04 | 1.58 | 1.38 | 0.10 | 2.89 | 0.66 | 1 | 13 |
| Hebei | Alpha | 7.00 | 1.08 | 1.49 | 0.35 | 1.97 | 0.32 | 1 | 18 |
| | Delta | 6.94 | 1.09 | 1.86 | 0.21 | 3.61 | 0.61 | 1 | 9 |
| | Omicron | 6.96 | 1.04 | 1.33 | 0.30 | 2.72 | 0.61 | 1 | 19 |
| Inner Mongolia | Alpha | 6.98 | 1.15 | 1.59 | 0.35 | 1.76 | 0.43 | 1 | 13 |
| | Delta | 6.97 | 1.08 | 1.25 | 0.33 | 2.33 | 0.36 | 1 | 13 |
| | Omicron | 6.67 | 0.79 | 0.69 | 0.20 | 3.38 | 0.32 | 7 | 7 |
| Tianjin | Alpha | 7.04 | 1.14 | 1.61 | 0.32 | 1.21 | 0.71 | 4 | 17 |
| | Delta | 6.99 | 1.14 | 1.06 | 0.21 | 4.31 | 0.72 | 3 | 5 |
| | Omicron | 6.97 | 1.01 | 1.05 | 0.36 | 2.02 | 0.63 | 1 | 22 |
| Jiangsu | Alpha | 7.03 | [1.13, 0.10] | 1.14 | [0.30, 0.33] | 2.94 | [1, 0.47] | 1 | [10, 17] |
| | Delta | 7.00 | [1.06, 1.07] | 1.01 | [0.20, 0.25] | 1.92 | [1, 0.30] | 13 | [7, 17] |
| | Omicron | 6.99 | 1.39 | 0.84 | 0.51 | 1.33 | 0.61 | 1 | 74 |
| Heilongjiang | Alpha | 6.99 | 1.13 | 0.72 | 0.43 | 2.09 | 0.52 | 2 | 16 |
| | Delta | 7.00 | 1.10 | 1.87 | 0.23 | 3.78 | 0.31 | 3 | 8 |
| | Omicron | 6.98 | 1.11 | 1.53 | 0.35 | 1.73 | 0.90 | 5 | 25 |
| Beijing | Alpha | 7.01 | 1.10 | 1.42 | 0.30 | 1.10 | 0.57 | 14 | 18 |
| | Delta | 6.98 | 1.19 | 1.08 | 0.27 | 7.01 | 0.69 | 1 | 4 |
| Henan | Alpha | 6.99 | 1.04 | 1.49 | 0.31 | 2.20 | 0.51 | 5 | 14 |
| Fujian | Alpha | 7.01 | 1.42 | 1.50 | 0.22 | 3.53 | 0.71 | 1 | 7 |
| | Delta | 7.05 | 1.16 | 1.56 | 0.21 | 3.54 | 0.34 | 3 | 7 |

The two waves in Jiangsu are have three phases, so the values in brackets are two change values.

Table 2: Confidence intervals of MFPTs and simulated extinction times.

| | True extinction time | Simulated extinction time | | Calculated values of MFPT | |
|---|---|---|---|---|---|
| | From data | Mean | Confidence Interval | Mean | Confidence Interval |
| Guangdong | 33 | 33.91 | [30.59 37.23] | 32.79 | [31.25 34.34] |
| | 26 | 24.65 | [22.42 26.88] | 23.27 | [21.78 24.76] |
| | 73 | 68.7 | [63.07 74.33] | 67.75 | [63.87 71.63] |
| Shaanxi | 27 | 26.10 | [24.46 27.74] | 24.75 | [23.35 26.15] |
| | 41 | 43.37 | [39.88 46.86] | 42.60 | [40.23 44.97] |
| | 41 | 41.34 | [38.17 44.51] | 39.38 | [36.29 42.47] |
| Hebei | 29 | 28.54 | [26.49 30.61] | 27.41 | [25.79 29.03] |
| | 31 | 31.85 | [27.57 36.13] | 30.11 | [25.90 34.32] |
| | 47 | 48.92 | [44.03 53.81] | 48.69 | [46.07 51.31] |
| Inner Mongolia | 19 | 20.34 | [18.50 22.18] | 19.12 | [17.57 20.68] |
| | 25 | 24.21 | [22.20 26.22] | 22.78 | [21.56 24.01] |
| | 22 | 22.45 | [20.74 24.16] | 21.06 | [20.20 21.91] |
| Tianjin | 32 | 31.17 | [27.35 34.99] | 30.02 | [27.34 32.70] |
| | 31 | 33.47 | [29.49 37.45] | 31.12 | [27.89 34.35] |
| | 43 | 45.02 | [39.73 50.31] | 44.49 | [42.12 46.85] |
| Jiangsu | 30 | 31.55 | [29.13 33.97] | 30.18 | [27.91 32.45] |
| | 34 | 33.91 | [32.74 35.08] | 32.39 | [31.41 33.38] |
| | 98 | 97.1 | [91.69 102.51] | 98.51 | [96.73 100.30] |
| Heilongjiang | 32 | 32.97 | [29.04 36.89] | 32.05 | [30.96 33.14] |
| | 20 | 20.69 | [18.50 22.88] | 19.12 | [16.84 21.40] |
| | 76 | 88.34 | [62.05 114.63] | 97.94 | [85.32 110.56] |
| Beijing | 33 | 33.18 | [30.52 35.84] | 31.35 | [29.59 33.11] |
| | 27 | 27.71 | [24.00 31.42] | 26.02 | [22.44 29.60] |
| Henan | 34 | 34.03 | [31.11 36.95] | 32.93 | [30.85 35.02] |
| Fujian | 29 | 28.25 | [24.12 32.38] | 26.56 | [22.63 30.49] |
| | 20 | 20.73 | [18.92 22.54] | 19.19 | [17.31 21.08] |

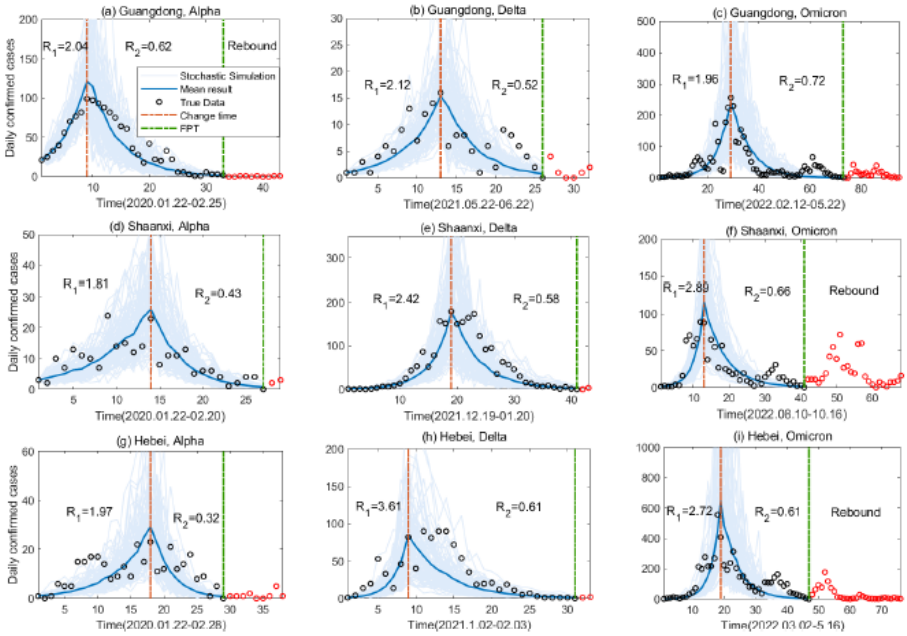


Fig. 1: The fitting results of the three waves in Guangdong, Shaanxi, and Hebei provinces. The Black circles are the actual data on the number of daily confirmed cases and the red circles are the data from the first extinction to the complete extinction. The pale blue background curves are the results of 100 stochastic simulations. The dark blue curves are the mean results of the stochastic simulations. The data fitting includes two phases: the increase and decrease of the number of daily confirmed cases, and the rebound period after the number of daily confirmed cases is less than 1. The vertical red line on the left is the estimated time change point at which reproductive number is switched, and the vertical green line on the right is the time obtained from the data at which daily confirmed cases were less than 1 for the first time.

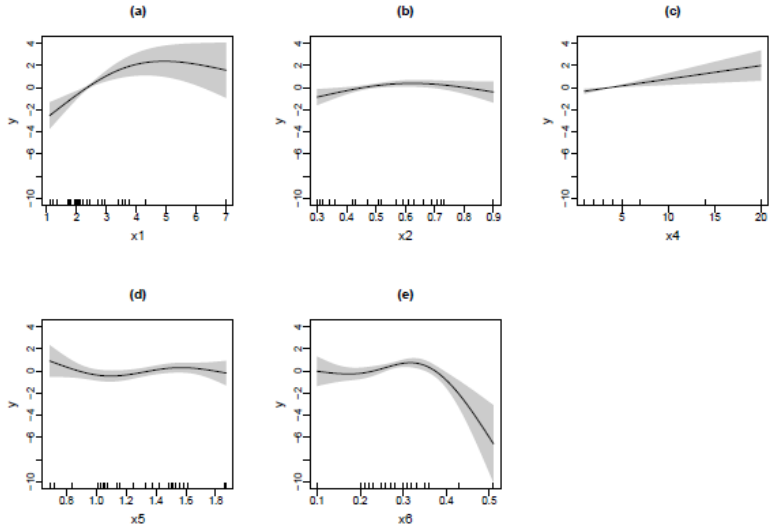


Fig. 2: Non-linear relationships between the cumulative number of cases with six variables. They are (a) the reproductive number in increasing phase, (b) the reproductive number in the decreasing phase, (c) the initial daily confirmed case number, (d) the noise intensity in the increasing phase, (e) the noise intensity in the decreasing phase, respectively. The black line represents the curve of the smoothed regression, the gray shaded area is the 95% confidence interval.

The regression results show that the change point of the reproductive number has a significant linear relationship with the cumulative number of cases. This suggests that taking control measures early to reduce the value of the reproductive number is a key way to control the scale of the epidemic. The results from the generalized additive model are shown in Table 4. The cumulative number of infected cases has a significant non-linear relationship with the reproductive number in the increasing phase. As shown in Fig.2 (a), the overall trend is that the reproductive number in the increasing phase has a positive impact on the cumulative number of infected cases. The stronger the virus transmission, the larger the scale of the epidemic. The reason for the flatness in the last part of the curve can be found from the data and fitting results of the second wave in Beijing. It can be seen that even if the reproductive number is large, the spread of the epidemic can be well suppressed if measures can be taken quickly. The non-linear relationship of the reproductive number in the decreasing phase with the cumulative number of infected cases shows the reproductive number in the decreasing phase after taking control measures has no obvious impact on the scale of the entire epidemic Fig.2 (b). This shows that if the number of cases at the change point is not large and the reproductive number is less than 1, the epidemic can be controlled, which can be seen from the data and fitting results of the third wave in Heilongjiang. The initial daily confirmed case number is significant in affecting the scale of the outbreak Fig.2 (c). The non-linear relationship of the noise intensity in the increasing phase with the scale of the outbreak shows it has little impact on epidemic scale Fig.2 (d). However, the relationship between the noise intensity in the decreasing phase and the cumulative number of infected cases has a negative impact on the scale of the outbreak when the noise is relatively large, that is, the greater the noise, the smaller the scale of the outbreak Fig.2 (e). This result illustrates that high noise intensity is conducive to controlling the spread of the disease.

### *5.3. Dynamic changing patterns*

The newly emerged SARS CoV-2 omicron variant causes milder symptoms, has a shorter incubation period, shorter recovery time and wider spread than the early strains, thus posing enhanced challenges to COVID-19 pandemic prevention.

Table 3: Results of linear regression.

| | Estimate | Std.Error | t-value | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| Intercept | $-2.44 * 10^{(-16)}$ | 0.17 | 0 | 1 |
| $x_1$ | 0.08 | 0.24 | 0.33 | 0.75 |
| $x_2$ | 0.16 | 0.18 | 0.90 | 0.38 |
| $x_3$ | 0.75 | 0.26 | 2.88 | 0.0096 |
| $x_4$ | 0.01 | 0.20 | 0.05 | 0.96 |
| $x_5$ | -0.04 | 0.20 | -0.19 | 0.85 |
| $x_6$ | -0.23 | 0.25 | -0.92 | 0.37 |
| Multiple R-squared: 0.4315, Adjusted R-squared: 0.252 | | | | |

Table 4: Results of generalized additive model.

| | edf | Ref.df | F | p-value |
|---|---|---|---|---|
| $s(x_1)$ | 1.959 | 1.998 | 9.968 | 0.003 |
| $s(x_2)$ | 1.935 | 1.993 | 3.688 | 0.062 |
| $s(x_4)$ | 1 | 1 | 8.487 | 0.012 |
| $s(x_5)$ | 2.776 | 2.966 | 1.44 | 0.234 |
| $s(x_6)$ | 2.957 | 2.997 | 6.154 | 0.007 |
| R-sq.(adj)=0.737, Deviance explained =85.9% | | | | |

In China, the quarantine measures are constantly adjusted according to the characteristics of the virus. For example, the isolation period is reduced due to a shortened incubation period, and the frequency of nucleic acid screening is also increased. Since the first outbreak of the Omicron strain in Tianjin in January 2022, the trend of the number of new cases in most places is complicated, and an epidemic wave will not end after going through a increasing and extinction phase. We collected the numbers of daily reported cases for Beijing city and five provinces including Henan, Zhejiang, Sichuan, Yunnan and Guangxi since February 2022. Using Bayesian parameter estimation, we estimated change points in the time series of daily reported cases numbers, including when the data begin to rise or fall sharply. As shown in Fig.3 (a), from March 2022 to July 2022 in Beijing, we have found that the change points for the daily reported cases data are the 47th, 78th, 95th and 100th days, respectively. These time nodes are usually accompanied by the adjustment of the local government's policy on epidemic prevention. For example, on 23 May, the 77th day in the Fig.3 (a), a press conference on the prevention of COVID-19 held in Beijing reported the requirements for working at home. Due to the strengthening of the policy, the number of daily reported cases began to decline on the 78th day. Starting from 6 June, corresponding to the 93rd day in the Fig.3 (a), the areas that previously worked from home could go to work normally, schools started face-to-face teaching, and restaurants will resume serving diners. However, the number of cases began to increase later, which showed that under the condition that the cases were not completely eliminated, the relaxation of the policy would lead to a rebound of the epidemic. Similarly, the change point of the daily reported cases in Henan, Zhejiang, Sichuan, Yunnan and Guangxi provinces are also shown in Fig.3. The mutation of the virus and the dynamic adjustment of policies make the trend of the number of people infected with the virus complicated and changeable over time. This makes it very difficult and challenging to predict when the epidemic will peak and become extinct.

## 6. Sensitivity analysis

The analysis in the previous section illustrates the significance of each parameter on an epidemic wave. To investigate how the parameters affect the time to extinction, we performed an uncertainty analysis of eight parameters in the increasing and decline phases that may affect the extinction time. We used the Partial rank correlation coefficient (PRCC) method to evaluate the impact of eight parameters on the MFPT. It was noted that the contact rates $c_1$ and $c_2$ had a high positive correlation with the reproductive number $R_1$ and $R_2$, respectively, so here we only analyze the effect of parameter $R_1$ and $R_2$. The distributions of each parameter are chosen from a uniform distribution and the range of parameter values is given according to the conditions in the model. The PRCC results which illustrate the dependence of MFPT on various input parameters are shown in Fig.4. The magnitude of the PRCCs implies the sensitivity of these parameters. The higher the bar in the Fig.4, the more sensitive the value of MFPT is to this parameter. The results also show a positive or

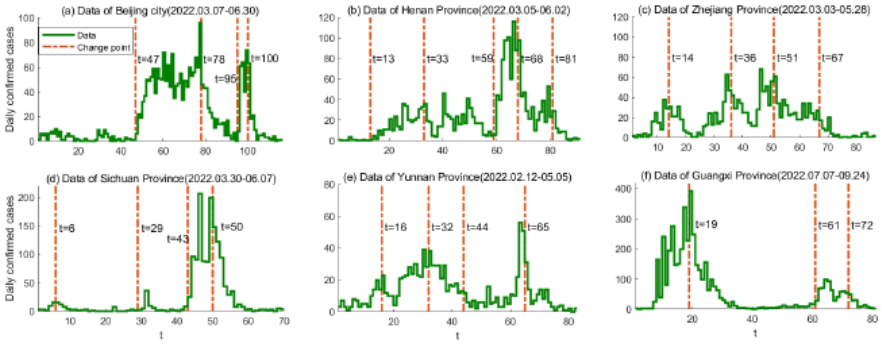


Fig. 3: The data of daily reported cases and change point estimation in Beijing city and five provinces including Henan, Zhejiang, Sichuan, Yunnan and Guangxi after February 2022.

negative correlation between the input parameters (i.e. $\gamma_1$, $\gamma_2$, $\sigma_1$, $\sigma_2$, $R_1$, $R_2$, $Ms0$, $T_1$) and MFPT. Among all these input parameters, the removal rate $\gamma_1$ in the increasing phase, the reproductive number $R_1$ and $R_2$, the initial number of infected cases and the change point have positive signs for their PRCCs, which indicates that larger values of these parameters will increase the transmission time of the disease. The removal rate $\gamma_2$, the noise parameters $\sigma_1$ in the increasing phase and $\sigma_2$ in the decline phase have negative signs for their PRCCs. These results illustrate that greater control efforts can speed up the elimination of the disease. An interesting conclusion is that both the noise in the increasing and decline phases favor earlier disease extinction.

To further explore the time required for the number of new cases to be cleared after entering the decline phase, we allowed one of the epidemiological parameters in the model to vary within a certain range, with the values of other parameters being fixed. Let $X_0$ indicate the number of daily confirmed cases at the change point, we selected the reproductive number set within the range of 0.2 to 0.8 and the set of $X_0$ within the range of 20 to 200. We plotted the contour map of $R_2$ and $X_0$ with respect to MFPT. From Fig.5 (a), we can see the comprehensive impact of the number of newly infected cases at the change point and the reproductive number in the extinction phase on the time of disease extinction. It is worth noting that the variations of MFPT with $R_2$ and $X_0$ are not linear. We found that large values of both $X_0$ and $R_2$ have a negative impact on disease extinction. We can also obtain from the explicit expression (27) of MFPT that MFPT decreases with the decrease of $x_0$ and $R_2$. Fig.5 (b) shows four different scenarios of the means of the numbers of daily confirmed cases when the reproductive number $R_2$ is 0.3, 0.5, 0.7 and 0.8, respectively. For smaller values of $R_2$, the faster the number of cases decreases, and the less is the time to reach a state of less than 1. Similarly, we present the mean curves of the stochastic solution under four different values $X_0$ shown in Fig.5 (c).

Fig.6 shows the effect of noise intensity on the extinction time in the declining phase, for which we conducted multiple random numerical experiments with different noise intensity. Fig.6 (a) shows some statistical characteristics of the stochastic simulations including the distribution of extinction times when the noise takes four different values from small to large. The white points in the figure are the average values of 5000 runs under four cases, and the mean values from left to right are 24.49, 24.3, 23.75 and 22.26. The average extinction time is shorter as the noise intensity becomes larger. In the Fig.6 (b-e), we present the curves of the stochastic solutions and the mean curves of 5000 simulations under four different noise intensities. It can be seen that the greater the randomness, the greater the possibility of shortening the time required for disease extinction.

## 7. Discussion and Conclusion

Most research on the dynamics of infectious diseases uses compartment models established by classifying the population. The established model can reflect different situations by focusing on various factors such as the different transmission characteristics of different diseases. The infected population in the compartment model represents the number of existing infected people, but these data are difficult to obtain in practice. In general, the reported information on the epidemic

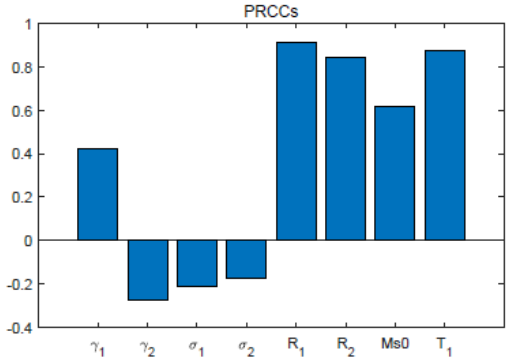


Fig. 4: Sensitivity analysis of the MFPT using PRCC. In the analysis of the eight parameters of the MFPT, we choose $c_1 = 5$, $c_2 = 2$, $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\sigma_1 = 1.5$, $\sigma 2 = 0.5$, $R_1 = 2$ $R2 = 0.2$, $Ms0 = 10$, $T_1 = 14$ as base values, and [0.1, 0.3], [0.45, 0.55], [0.5, 2], [0.1, 0.8], [1, 4], [0.1, 0.7], [1, 20], [8, 20] as their ranges. The obtained PRCCs of the eight parameters are 0.65, -0.23, -0.20, -0.15, 0.89, 0.81, 0.58, 0.86.

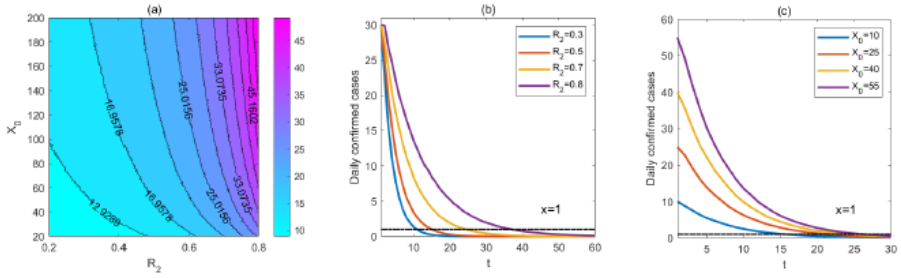


Fig. 5: Sensitivity analysis of the MFPT after the daily number of confirmed cases has entered a decreasing phase. $R_2$ is the basic reproduction number in the descending phase. $X_0$ is assumed as the number of daily confirmed cases at the change point for entering the descending phase. The point crossing the dotted line is the time when the solution of $X(t)$ first reaches 1. (a) Contour map of $R_2$ and $X_0$ with respect to MFPT; (b) Epidemic curves corresponding to different $R_2$, given $\gamma_2 = 0.5$, $c_2 = 1$, $X_0 = 30$, $\sigma_2 = 1$; (c) Epidemic curves corresponding to different $X_0$, given $R_2 = 0.6$.

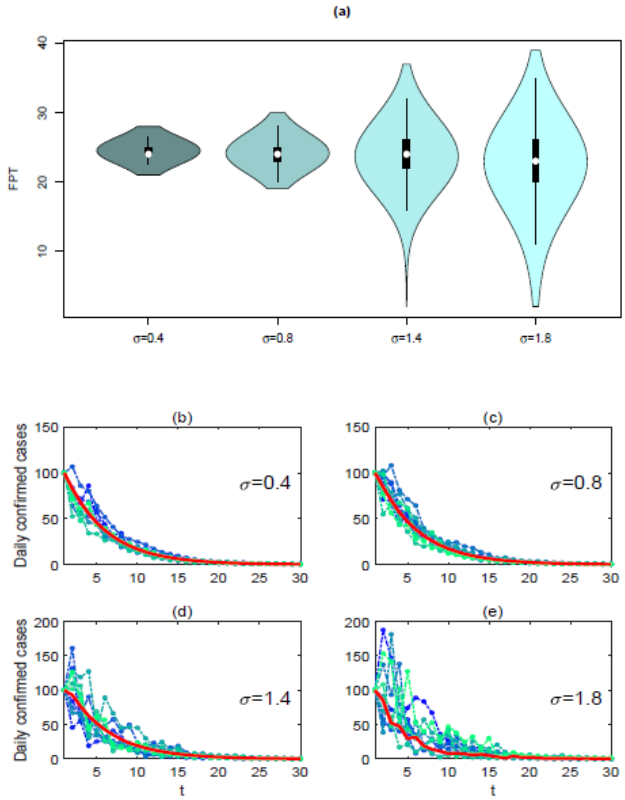


Fig. 6: Statistical analysis of the effect of noise intensity on MFPT. (a) Violin plots of 1000 samples of extinction times under four different levels of noise generated by our stochastic model; (b-e) Stochastic solution curves with four different noise intensities, and the red lines are the mean values of 5000 stochastic simulations.

mainly includes the number of newly confirmed cases per day, the number of new deaths, the cumulative number of infected cases, etc. Therefore, the purpose of our study was to build a model to explore the characteristics of the number of newly confirmed cases. Based on a SIR model, we derived the equation and the iterative form for the number of newly confirmed cases. We also obtained a one-dimensional model of the number of newly infected cases in the deterministic model. Further we established a corresponding stochastic model by considering the influence of random noise and derived the random process form of the number of newly infected cases. This model is based on an assumption that the infection scale of the epidemic is relatively small, and the number of susceptible people is assumed to be approximately equal to the total population. We calculated the reproductive number of the deterministic model denoted as $R_d$, and the threshold for disease outbreak and extinction from the corresponding stochastic model denoted as $R_s$. The ordinary and stochastic difference equation of numbers of newly confirmed cases can be expressed by $R_d$ and $R_s$. By analyzing the properties of the stochastic difference equation, we prove that $R_s < 1$ is the condition that the number of newly confirmed cases finally tends to zero.

The reproductive number can reflect the trend of disease transmission. The implementation of control measures will reduce the spread of the disease, and the number of infected cases will change from rising to falling. The process of disease transmission can generally be divided into two phases: free transmission and controlled transmission. In particular, there will be a plateau phase, that is, the number of cases fluctuates around a mean level. Therefore, we consider the three phases of the disease infection process and the critical time point for switching between phases based on a change in the reproductive number. During the decreasing phase of the number of newly confirmed cases, what we are concerned with is how long will the disease take to go extinction. We use the FPT to define the extinction time in the stochastic model. The explicit expression of MFPT is obtained by building and solving the backward equation of the stochastic model. Since the number of infections cannot reach 0 in a finite time in the differential model, we assume that the first time the number of cases reaches 1 indicates the extinction of the disease. By comparison of MFPT in the stochastic model with FPT in the deterministic model, one fascinating conclusion is that random noise can accelerate the extinction of diseases, which indicates that it has a good effect on disease control.

From the data on the numbers of daily confirmed cases, it can be seen that COVID-19 has been spreading in many provinces and cities in China, but the outbreak scale has mostly been small. After the occurrence of cases, contact tracing and quarantine and other control measures can effectively control the spread of the epidemic. Our stochastic model is suitable for studying data like the multiple waves of the COVID-19 epidemic in many places in China. To apply the stochastic difference model on the analysis of the data on COVID-19 disease, we collected the daily number of confirmed cases for each of 26 waves in Guangdong, Shaanxi, Jiangsu, Tianjin, Hebei and other provinces. Most of the places with multi-wave data have experienced the epidemic dominated by all three different strains (Alpha, Delta and Omicron). Using a Bayesian method, we estimated the change point when the reproductive number in each wave changed. Based on the stochastic difference model, we fitted data from 26 waves and obtained the values of the parameters including the contact rate, noise density and reproductive number in both the increasing and decreasing phases. Not only that, we also calculated the values and confidence intervals of the MFPT based on the actual data and a theoretical expression, respectively. The development of a wave process can be described comprehensively with these indicators. Therefore, we used regression analysis and a generalized additive model to analyze the linear or nonlinear impact of each indicator on the cumulative number of cases in detail. The results indicate that the time point of the reproductive number change is crucial for disease control.

Our sensitivity analysis explores the influence of reproductive number, initial value and noise intensity in the decreasing phase on the MFPT, and the results show that larger amounts of noise are conducive to the extinction of diseases, and the other two factors have a negative effect on disease extinction. One characteristic of each outbreak that we have studied is that they only went through a rising phase and a falling phase. However, with virus mutation and dynamic adjustment of policies, the data for many places in 2022 showed the number of newly confirmed cases continues to rebound, showing a pattern of multiple waves. We collected data from cities with these characteristics as shown in Fig.3. We only estimated change points from the data and showed that how they relate to policy implementation. We did not fit such multi-wave data, which requires future research to explore the pattern of the epidemic that has rebounded with multiple waves. The stochastic difference model that we derived can be used to model changes in the number of newly confirmed cases well. This paper focuses on the switching of the reproductive number and the mean extinction time. In future research, more specific factors can be considered to further promote the application of stochastic difference models.

**Acknowledgment**

This research was partially supported by the National Natural Science Foundation of China (12171295(XW) 12031010(ST), 12201377(SH)) and the Project of Science and Technology Young Star in Shaanxi Province of China (2022KJXX29(XW)).

**Declarations**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] D. M. Morens, G. K. Folkers, A. S. Fauci, The challenge of emerging and re-emerging infectious diseases, Nature, 430(6996) (2004) 242-249.

[2] R. E. Baker, A. S. Mahmud, I. F. Miller, et al. Infectious disease in an era of global change, Nat. Rev. Microbiol., 20(4) (2022) 193-205.

[3] S. Cobey. Modeling infectious disease dynamics, Science, 2020, 368(6492): 713-714.

[4] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Math. Phys. Eng. Sci., 115 (1927) 700-721.

[5] T. Britton, Stochastic epidemic models: a survey, Math. Biosci., 2010, 225(1): 24-35.

[6] H. C. Tuckwell, R. J. Williams, Some properties of a simple stochastic epidemic model of SIR type, Math. Biosci., 208(1) (2007) 76-97.

[7] J. R. Artalejo, M. J. Lopez-Herrero, On the exact measure of disease spread in stochastic epidemic models, Bull. Math. Biol., 75(7) (2013) 1031-1050.

[8] L. J. S. Allen, A. M. Burgin, Comparison of deterministic and stochastic SIS and SIR models in discrete time, Math. Biosci., 163(1) (2000) 1-33.

[9] A. Gray, D. Greenhalgh, L. Hu, et al. A stochastic differential equation SIS epidemic model, SIAM J. Appl. Math., 71(3) (2011) 876-902.

[10] Y. L. Cai, J. J. Jiao, Z. J. Gui, et al. Environmental variability in a stochastic epidemic model, Appl. Math. Comput., 329 (2018) 210-226.

[11] Y. N. Zhao, D. Q. Jiang, The threshold of a stochastic SIRS epidemic model with saturated incidence, Appl. Math. Lett., 34 (2014) 90-93.

[12] C. Y. Ji, D. Q. Jiang, Threshold behaviour of a stochastic SIR model, Appl. Math. Model., 38(21-22) (2014) 5067-5079.

[13] Q. Liu, D. Q. Jiang, T. Hayat, et al. Dynamics of a stochastic multigroup SIQR epidemic model with standard incidence rates, J. Franklin Inst., 356(5) (2019) 2960-2993.

[14] P. E. Lekone, B. F. Finkenstädt, Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study, Biometrics, 62(4) (2006) 1170-1177.

[15] P. D. O'Neill, A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods, Math. Biosci., 180(1-2) (2002) 103-114.

[16] K. V. Parag, Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves, PLoS Comput. Biol., 17(9) (2021) e1009347.

[17] J. Sun, Z. Shi, H. Xu, Non-pharmaceutical interventions used for COVID-19 had a major impact on reducing influenza in China in 2020, J Travel Med., 27(8) (2020) taaa064.

[18] N. Haug, L. Geyrhofer, A. Londei, et al. Ranking the effectiveness of worldwide COVID-19 government interventions, Nat. Hum. Behav., 4(12) (2020) 1303-1312.

[19] Y. Liu, C. Morgenstern, J. Kelly , et al. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories, BMC Med., 19(1) (2021) 1-12.

[20] B. Shayaka, M. M. Sharmab, M. Gaurc, A. K. Mishra, Impact of reproduction number on the multiwave spreading dynamics of COVID-19 with temporary immunity: A mathematical model, Int. J. Infect. Dis., 104 (2021) 649-654.

[21] D. Rios-Doria, G. Chowell, Qualitative analysis of the level of cross-protection between epidemic waves of the 1918-1919 influenza pandemic, J. Theor. Biol., 261 (2009) 584-592.

[22] A. Flahault, E. Vergu, P. Boëlle, Potential for a global dynamic of Influenza A (H1N1), BMC Infect. Dis., 9 (2009) 129.

[23] Y. H. Hsieh, C. W. S. Chen, Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks, Trop. Med. Int. Health, 14(6) (2009) 628-638.

[24] K. Ghosh, A. K. Ghosh, Study of COVID-19 epidemiological evolution in India with a multi-wave SIR model, Nonlinear Dynam., 109 (2022) 47-55.

[25] S. Y. Su, W. C. Lee, Monitoring the peaks of multiwave COVID-19 outbreaks, J. Microbiol. Immunol., 55(2) (2022) 350-352.

[26] G. Perakis, D. Singhvi, O. S. Lami, L. Thayaparan, COVID-19: a multiwave SIR-based model for learning waves, Prod. Oper. Manag., (2022) 1-19.

[27] M. J. Beira, P. J. Sebastião, A differential equations model-ftting analysis of COVID-19 epidemiological data to explain multi-wave dynamics, Sci. Rep., 11 (2021) 16312.

[28] E. V. M. Dos Reis, M. A. Savi, A dynamical map to describe COVID-19 epidemics, Eur. Phys. J. Spec. Top., 231(5) (2022) 893-904.

[29] P. Blonigan, J. Ray, C. Safta, Forecasting multi-wave epidemics through Bayesian inference, Arch. Comput. Method. E., 28 (2021) 4169-4183.

[30] B. Xu, J. Cai, D. H. He, G. Chowell, B. Xu, Mechanistic modelling of multiple waves in an influenza epidemic or pandemic, J. Theor. Biol., 486 (2020) 110070.

[31] J. T. Lim, B. S. Dickens, S. Haoyang, et al. Inference on dengue epidemics with Bayesian regime switching models, PLoS Comput. Biol., 16(5) (2020) e1007839.

[32] S. He, S. Y. Tang, Y. N. Xiao, R. A. Cheke, Stochastic modelling of air pollution impacts on respiratory infection risk, Bull. Math. Biol., 80 (2018) 3127-3153.

[33] T. Chou, M. R. D'Orsogna, First passage problems in biology, In First-passage phenomena and their applications. (2014) 306-345.

[34] L. M. Ricciardi, A. D. Crescenzo, V. Giorno, et al. An outline of theoretical and algorithmic approaches to first passage time problems with applications to biological modeling, Mathematica Japonica, 50 (1999) 247-322.

[35] H. C. Tuckwell, F. Y. M. Wan, First passage time to detection in stochastic population dynamical models for HIV-1, Appl. Math. Lett., 13(5) (2000) 79-83.

[36] C. Skiadas, C. H. Skiadas, Development, simulation, and application of first-exit-time densities to life table data, Commun. Stat. Theory Methods, 39(3) (2010) 444-451.

[37] C. Floris, First-passage time study of a stochastic growth model, Nonlinear Dynam., 98(2) (2019) 861-872.

[38] V. Srivastava, S. F. Feng, J. D. Cohen, et al. A martingale analysis of first passage times of time-dependent Wiener diffusion models, J. Math. Psychol., 77 (2017) 94-110.

[39] A. G. Strang, K. C. Abbott, P. J. Thomas, How to avoid an extinction time paradox, Theor. Ecol., 12(4) (2019) 467-487.