

Research Article

Epidemic Modelling by Ripple-Spreading Network and Genetic Algorithm

Jian-Qin Liao,^{1,2} Xiao-Bing Hu,^{1,3} Ming Wang,¹ and Mark S. Leeson³

¹ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

² Angel Women's & Children's Hospital, Chengdu 610041, China

³ School of Engineering, University of Warwick, Coventry CV4 7AL, UK

Correspondence should be addressed to Xiao-Bing Hu; dr_xiaobinghu@hotmail.co.uk

Received 20 July 2013; Revised 8 September 2013; Accepted 8 September 2013

Academic Editor: J. A. Tenreiro Machado

Copyright © 2013 Jian-Qin Liao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mathematical analysis and modelling is central to infectious disease epidemiology. This paper, inspired by the natural ripple-spreading phenomenon, proposes a novel ripple-spreading network model for the study of infectious disease transmission. The new epidemic model naturally has good potential for capturing many spatial and temporal features observed in the outbreak of plagues. In particular, using a stochastic ripple-spreading process simulates the effect of random contacts and movements of individuals on the probability of infection well, which is usually a challenging issue in epidemic modeling. Some ripple-spreading related parameters such as threshold and amplifying factor of nodes are ideal to describe the importance of individuals' physical fitness and immunity. The new model is rich in parameters to incorporate many real factors such as public health service and policies, and it is highly flexible to modifications. A genetic algorithm is used to tune the parameters of the model by referring to historic data of an epidemic. The well-tuned model can then be used for analyzing and forecasting purposes. The effectiveness of the proposed method is illustrated by simulation results.

1. Introduction

Mathematical representation and analysis of infectious diseases has been central to infectious disease epidemiology since its inception as a discipline more than a century ago [1–7]. In recent years, detailed electronic surveillance of infectious diseases has become widespread owing to the advent of improved computing, electronic data management, the ability to share and deposit data over the internet, and rapid diagnostic tests and genetic sequence analysis. These ongoing developments have increased the application of mathematical models to both the generation and testing of basic scientific hypotheses and to the design of practical strategies for disease control. Mathematical analyses and models have provided successful explanations of previously puzzling observations and played a central part in public health strategies in many countries [3, 4].

Fundamental to the growing importance of mathematical epidemiology has been the integration of mathematical models with rigorous statistical methods to estimate

key parameters of these models and test hypotheses using available data. In the absence of reliable data, mathematics can be used to help formulate hypotheses, inform data-collection strategies, and determine sample sizes, which can permit discrimination of competing hypotheses. In this way, mathematics is “no more, but no less, than a way of thinking clearly about the problem in hand” [5]. The extent and quality of available data can be variable. Ideally, data should be analysed using models that adequately describe the observed dynamics and patterns of interest and the mechanisms that generate these observations. Models should be as simple as possible, but not so simple that the conclusions drawn are altered by the consideration of additional realistic complexity. Unnecessary complexity can obscure fundamental results and is almost as undesirable as oversimplification. Indeed, model choice—the process of deciding which model complexities are necessary—is a central part of mathematical modelling of infectious diseases.

The most recent survey on epidemic modelling can be found in [7]. Basically, existing epidemic models can be

classified into two categories [8–20]: (i) top-down models which are deterministic and based on systems of differential equations [20], Markov Chain [8], mean field type equations [9], and (ii) bottom-up models which are stochastic and based on computer simulations [10], agent based methods [11], cellular automata [12], and network theory [13]. For top-down models, diffusive or perfect mixing and random motion are assumptions that are not always fulfilled, at least in the human population. These models also tend to incorporate many parameters to explain reality, which increase their complexity and make them computationally intensive and difficult to analyze. Contrary to what happens to top-down models, the complex systems approach is the foundation of bottom-up models, where it is considered that spatial extended systems are capable of nontrivial collective behaviour—unexpected behaviour which is observed in macroscopic quantities. It is assumed that there are several levels of reality: at a microscopic level, interactions may be described by complicated potentials, but, at a macroscopic level, the properties of the system are dominated by the aggregated effect of all microscopic interactions. Human epidemics are strongly related to the dynamics of populations and to the network of social contacts. In particular, network theory has proven a promising new method applicable to epidemiology. For instance, the influence of small-world and scale-free topologies on the breakout of plagues was investigated in [13, 15]; a random network was used to conduct multiscale analysis on epidemic dynamics [16]; a growing network model was reported to develop a population-level epidemic model [17]. As stated in [18], the combination of network theory and epidemic modelling can deliver an improved understanding of disease dynamics and better public health through effective disease control. However, as pointed out in a recent perspective paper [19], new theories and methods are still needed to study interacting dynamics, amplification, and cascading effects in complex network systems such as epidemic dynamics.

Inspired by the natural ripple-spreading phenomenon on liquid surfaces, this paper reports a novel complex system based bottom-up epidemic model: ripple-spreading epidemic model (RSEM), which is an application-focused extension of our recent work on general ripple-spreading network models [21, 22]. As widely acknowledged, random contacts and movements of individuals impose a big challenge to epidemic modelling. Defining neighbourhoods and/or introducing transport rules at a microscopic level are often measures adopted to simulate social contacts and physical movements of individuals. In contrast, the new model proposed here takes account of the effect of such interactions between individuals via a ripple-spreading process and the reaction of nodes to ripples, whilst all nodes can be fixed without the need for a predefined neighbourhood. Basically, the infection probability is reflected by the point energy of ripples, and the social activeness of individuals can be associated with the threshold and the amplifying factor of nodes. Actually, the proposed ripple-spreading model can intuitively capture many spatial and temporal factors which matter in the outbreak of plagues. Therefore, this new model, when in combination with an effective parameter tuning method such

as genetic algorithms (GAs), possesses excellent potential for studying epidemic dynamics.

The remainder of this paper is organized as follows. Firstly, a general ripple-spreading network model of epidemic is proposed in Section 2. Then a genetic algorithm based method is reported in Section 3 to tune the model, so that it can simulate a specific epidemic. Some simulation results are illustrated in Section 4, and the paper ends with its main conclusions in Section 5.

2. Ripple-Spreading Epidemic Model (RSEM)

2.1. The Basic Idea of Ripple-Spreading Network Model. The basic natural ripple-spreading phenomenon is as follows. Suppose a collection of stakes is randomly distributed in a quiet pond, and suddenly a stone is thrown into the pond generating an initial ripple from the point where the stone hits the quiet water surface. When the ripple reaches a near stake, a new ripple is generated around the stake due to the reflection effect. Hereafter, for the sake of consistency, we denote such a new ripple as a responding ripple or outgoing ripple and the ripple which triggers the responding ripple as a stimulating ripple or incoming ripple. As the initial stimulating ripple is spreading, more and more responding ripples are stimulated around stakes. However, since the point energy on the initial stimulating ripple decays as it spreads out, those responding ripples triggered at a late phase will hardly be noticed. Let a node in a network stand for a stake in the pond, and an edge will be established between two nodes if a stake's ripple triggers a new ripple around the other stake. Then, after all ripples decay, we will get a network according to which stake's ripple has caused which stake to generate a new ripple. This is the basic idea of the ripple-spreading network model. Figure 1 gives an illustration of the development of a ripple-spreading network. For more details, readers are referred to [21, 22].

We can liken the outbreak of plagues to the above natural ripple-spreading phenomenon. Replace the stakes in the pond with a population of susceptible individuals in a community where an epidemic may break out. An initial infective case in the community is likened to the stone which hits the water surface. The influence of this initial infective case on other individuals is analogous to the initial stimulating ripple. The probability of infection can be related to the point energy of the ripple (but as will be explained later, the point energy is not exactly the probability of infection). The point energy decays as the ripple spreads out, and this can effectively capture the fact that long distance and few contacts often mean a low probability of infection. The process that a stimulating ripple triggers a responding ripple is related to a susceptible individual being infected and then becoming infective. Stakes of different textures may have different reflection effects; for instance, a rigid texture will cause strong reflection whilst a soft texture absorbs the most energy of the incoming ripple and therefore has no reflection. This is likened to the difference in the physical fitness, immunity, and social activities of individuals. Therefore, we can set a threshold for each individual. If and only if the point energy of an incoming ripple is above the threshold,

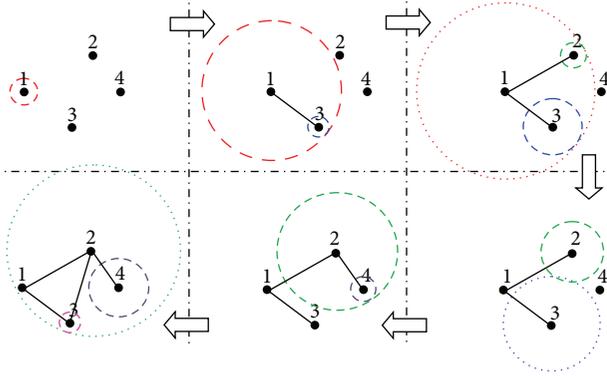


FIGURE 1: An illustration of the development of a ripple-spreading network.

it can be possible for an individual to be infected. Once an individual is infected, how infective s/he will become largely depends on his/her social activities. We can set an amplifying factor for each individual, which will determine the initial point energy of the responding ripple based on the point energy of the stimulating ripple. Therefore, the responding ripple of a socially active individual will have high point energy, which means a high probability of infecting other individuals. A stake can generate new reflection every time when it is reached by an incoming ripple that has enough point energy. This corresponds to the fact that an individual can be infected again when s/he becomes susceptible once again following a period of postrecovery immunity. Which stake's ripple causes which subsequent stake to generate a new ripple is analogous to who infects whom, indicated by an established edge between two nodes in a network. Who has infected whom during an epidemic outbreak can be illustrated by a growing network that is simulated by which stake's ripple has triggered which following stake to generate a new ripple. Now, one may get a feeling that the ripple-spreading network model invented in [21, 22] can be used to simulate the outbreak of plagues. To this end, we first need to, based on the work reported in [21, 22], develop a mathematical ripple-spreading network model of epidemic, which hereafter is called ripple-spreading epidemic model, denoted as RSEM.

2.2. Mathematical Formulation of RSEM. In the proposed RSEM, there are two groups of parameters. The first group is those of the general ripple-spreading network model as reported in [21, 22], but some modifications may be necessary in order to fit them in the scope of epidemiology. In this group, first we have parameters for N_{EISR} epicenters of initial stimulating ripples (EISRs), which are related to those initial cases in an epidemic outbreak. In this study, we only focus on the infectious disease transmission between human hosts, and other hosts such as rats and mosquitoes are not considered. Therefore, each EISR is actually a node in the network, that is, an individual in the community. The i th EISR, $i = 1, 2, \dots, N_{\text{EISR}}$, has an initial point energy of $E_{\text{EISR}}(i)$, and it is not active, that is, not infective, until time

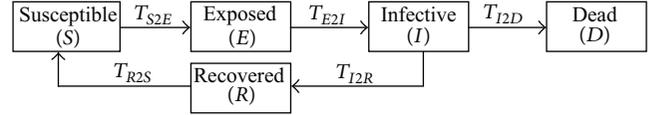


FIGURE 2: The cycle of epidemic states.

instant $T_{\text{EISR}}(i)$. Suppose there are N_N nodes in the network. Then, for node i , there is a threshold $\beta(i)$ to determine whether it is possible for the node to be infected by a certain infective node and an amplifying factor $\alpha(i)$ to calculate the initial energy when the node is infected. Basically, we can assume all ripples have the same spreading speed s and the same energy decaying coefficient vector η .

The second parameter group is epidemic specific. Because the nodes of network represent individuals, every node will have a certain epidemic state at each time point. Let $S_N(i, t)$ denote the epidemic state of node i at time t . Table 1 lists the epidemic states used in this study. Figure 2 shows how an individual will go through these epidemic states in an epidemic cycle. Simply speaking, in the epidemic cycle, an individual is initially susceptible. Once s/he is infected, s/he becomes exposed and then s/he develops to be infective; an infected individual may either die or recover from the infection. A recovered individual usually gets postrecovery immunity. If this is temporary, then after a period of time of immunity, s/he will become susceptible again. As one can see from Figure 2, there is a time period for an individual to transfer from one epidemic state to its following epidemic state. Except T_{S2E} , all other state transfer times, that is, T_{E2I} , T_{I2D} , T_{I2R} , and T_{R2S} , are independent of the ripple-spreading process, being instead mainly determined by the nature of the epidemic, the physical fitness of individuals, and/or relevant public health policies/measures. Basically, from statistical study of historical data, we may get an estimate of each state transfer time distribution (excluding T_{S2E}). In this study, we assume they all have Poisson distributions. For instance, T_{E2I} may have a Poisson distribution with a mean of \bar{T}_{E2I} . Therefore, with \bar{T}_{E2I} as a parameter, we can roughly know how long it will take for an individual to transfer from state “E” to state “I”. When running the RSEM, once a node becomes exposed, we then randomly assign a T_{E2I} to the node according to a Poisson distribution with \bar{T}_{E2I} as the mean. Similarly, we deal with other state transfer times except T_{S2E} . Therefore, in the second group, we have four means as parameters, namely, \bar{T}_{E2I} , \bar{T}_{I2D} , \bar{T}_{I2R} , and \bar{T}_{R2S} . For the sake of generality, T_{E2I} , T_{I2D} , T_{I2R} , and T_{R2S} are allowed to be zero, which means one or more epidemic states may not be experienced by an individual in an epidemic cycle. For example, if $T_{E2I} = 0$ and $T_{R2S} = 0$, then there is no latency period or postrecovery immunity.

Besides the two groups of parameters, there are three kinds of dynamics in the RSEM: the ripple-spreading process, the reaction of nodes to ripples, and the state transfer of nodes. The first two originate from the general work reported in [21, 22], but some minor epidemic-specific modifications

are introduced. The dynamics of state transfer is a brand-new concept for the general ripple-spreading network model.

The ripple-spreading process in the RSEM is mathematically described as follows. Suppose ripple i is associated with the i th node, $i = 1, 2, \dots, N_N$. Then, let $E_N(i)$ be the initial point energy of the ripple i , $e_N(i, t)$ the point energy of the ripple i at time t , and $r_N(i, t)$ the radius of ripple i . Initialize $E_N(i) = 0$, $e_N(i, 0) = 0$ and $r_N(i, 0) = 0$. Once ripple i is triggered at a certain time point, $E_N(i)$ will be given a value larger than zero. If $E_N(i) > 0$ and $S_N(i, t) = "I"$ (i.e., the state of node i is "Infective"), then ripple i will spread out as follows:

$$r_N(i, t) = r_N(i, t - 1) + s, \quad (1)$$

$$e_N(i, t) = f_{\text{Decay}}(E_N(i), r_N(i, t)), \quad (2)$$

where f_{Decay} is a function defining how the point energy decays as the ripple spreads out. A typical decaying function may be

$$f_{\text{Decay}}(E, r) = \frac{\eta(k)E}{2\pi r}, \quad (3)$$

and k is an integer index calculated as

$$k = \frac{r_N(i, t)}{s}. \quad (4)$$

Once the $S_N(i, t)$ changes from "I" to "D" or "R", reset $E_N(i) = 0$, $e_N(i, 0) = 0$, and $r_N(i, 0) = 0$.

The reaction of nodes to ripples defines how a node is infected. Suppose node i is reached by ripple j at time t ; that is,

$$r_N(j, t) \geq D_N(i, j) \geq r_N(j, t - 1), \quad (5)$$

where $D_N(i, j)$ is the distance between node i and node j . If $S_N(i, t) = "S"$ and $e_N(j, t) \geq \beta(i)$, that is, the point energy of ripple j is above the threshold of node i , then node i will be infected with a probability which may be defined based on an arctangent function of $e_N(j, t)$ as follows:

$$p_R(i) = \frac{\tan^{-1}(\kappa((e_N(j, t) - \beta(i))/\beta(i)) - \delta) + \tan^{-1}(\delta)}{\pi/2 + \tan^{-1}(\delta)}, \quad (6)$$

where $\kappa > 0$ and δ are coefficients which can adjust the shape and location of the arctangent function. Once node i is infected, we establish a directional link from node j and node i by modifying the adjacency matrix:

$$M_A(j, i) = M_A(j, i) + 1, \quad (7)$$

set $S_N(i, t + 1) = "E"$, set the initial point energy of ripple i as

$$E_N(i) = \alpha(i)e_N(j, t), \quad (8)$$

record the time of becoming exposed $t_E(i) = t$, and assign values to $T_{E2I}(i)$, $T_{I2D}(i)$, $T_{I2R}(i)$, and $T_{R2S}(i)$ according to relevant Poisson distributions defined by \bar{T}_{E2I} , \bar{T}_{I2D} , \bar{T}_{I2R} , and \bar{T}_{R2S} .

Please note that, although ripple i has $E_N(i) > 0$ at time $t_E(i)$, it will not start spreading until the state of node i has become "Infective", which is determined by the following state transfer dynamics: if $S_N(i, t) = "E"$ and $t - t_E(i) \geq T_{E2I}(i)$, then set $S_N(i, t + 1) = "I"$ and start the ripple-spreading process for ripple i ; if $S_N(i, t) = "I"$ and $t - t_E(i) \geq T_{E2I}(i) + T_{I2D}(i)$, randomly decide whether node i will die according to a preset death rate R_D ; if node i has died, set $S_N(i, t + 1) = "D"$, $r_N(i, t) = 0$, and $E_N(i) = 0$, and stop the ripple-spreading process of ripple i ; if $S_N(i, t) = "I"$ and $t - t_E(i) \geq T_{E2I}(i) + T_{I2R}(i)$, set $S_N(i, t + 1) = "R"$, $r_N(i, t) = 0$, and $E_N(i) = 0$, and stop the ripple-spreading process of ripple i ; if $S_N(i, t) = "R"$ and $t - t_E(i) \geq T_{E2I}(i) + T_{I2R}(i) + T_{R2S}(i)$, set $S_N(i, t + 1) = "S"$. So, node i has gone through an epidemic cycle.

By integrating the above parameters and dynamics, the proposed RSEM can be finally described as a whole by the following steps.

Step 1. Initialization, that is, set $S_N(i, 0) = "S"$, $E_N(i) = 0$, $r_N(i, 0) = 0$, $t = 0$ and $M_A(j, i) = 0$ for all $i = 1, 2, \dots, N_N$ and $j = 1, 2, \dots, N_N$, randomly choose N_{EISR} nodes as initial cases, set their state as "E", set their T_{E2I} as T_{EISR} , and randomly set their T_{I2D} , T_{I2R} , and T_{R2S} according to the relevant Poisson distributions defined by \bar{T}_{I2D} , \bar{T}_{I2R} , and \bar{T}_{R2S} .

Step 2. While the termination criteria are not satisfied, let $t = t + 1$, do for $i = 1, 2, \dots, N_N$.

Substep 2.1. Let $S_N(i, t) = S_N(i, t - 1)$.

Substep 2.2. If $E_N(i) > 0$ and $S_N(i, t) = "I"$, let ripple i spread for one time step by following the ripple-spreading process described by (1) to (4).

Substep 2.3. If node i has $S_N(i, t) = "S"$ and is reached ripple j according to (5), calculate the reaction of node i as described by (6) to (8).

Substep 2.4. If $S_N(i, t) \neq "S"$, following the state transfer dynamics to process node i for one time step.

When an RSEM run is terminated, a network will appear based on all established directional links which indicate how the infectious disease has transmitted between individuals. Figure 3 illustrates how the infectious disease transmission can be simulated by the proposed RSEM.

2.3. Further Analysis of RSEM. As discussed in the Introduction section, social contacts and physical movements of individuals impose a big challenge to epidemic modeling. Fortunately, thanks to ripple-spreading dynamics, the proposed RSEM can effectively simulate the effect of social contacts and physical movements without actually applying such activities to each individual. Basically, the strength of social contacts and physical movements of an individual is largely reflected by the amplifying factor $\alpha(i)$ and the threshold $\beta(i)$. For a socially active individual, s/he is more

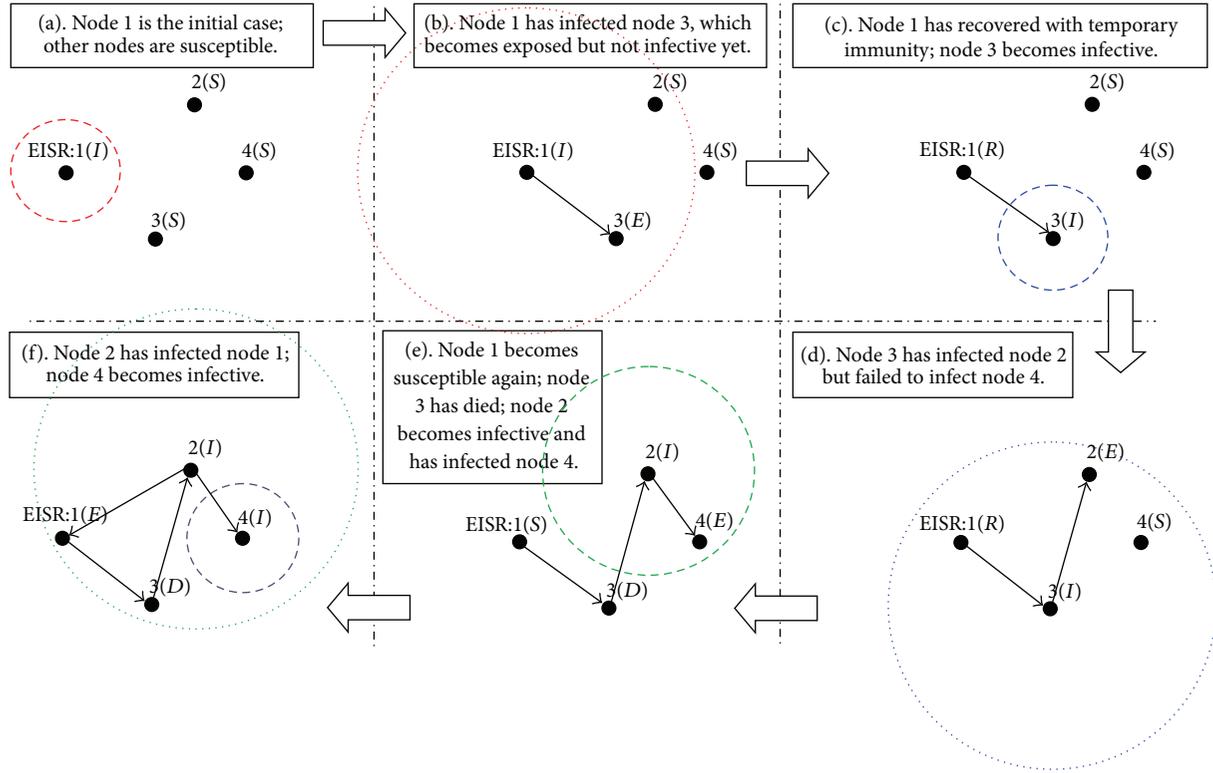


FIGURE 3: An illustration of RSEM.

likely to be infected, so s/he should have a smaller $\beta(i)$. According to (6), one can see that a smaller $\beta(i)$ means a higher probability of being infected. At the same time we set a larger $\alpha(i)$ for a socially active individual, and, from (8), one can see that his/her ripple will then have more point energy to infect others, which is in line with the reality. According to (2) and (3), as a ripple spreads out, its point energy decays, which in general means that the infection probability becomes lower as the distance between two individuals increases. This is also the case in reality: for an individual, usually most of his/her social contacts and physical movements happen within a small group of people (compared with the whole community) and a limited physical space (compared with the entire space). Therefore, most secondary infections caused by him/her will be spatially limited. Now it is clear that, by associating the point energy of ripples to the probability of infection and the amplifying factor and the threshold of nodes to the social contacts and physical movements of individuals, the proposed RSEM provides a new approach to simulate and study the epidemic dynamic.

One may argue that, according to Table 1 and Figure 2, the proposed RSEM misses the state of “Admitted”, which usually means that individuals who are infective are quarantined in hospital and therefore have no chance to transmit the infection to others. Obviously, this state is not an original state in the natural epidemic cycle, but a man-made state due to the public health service. Actually, the effect of the state of “Admitted” can be equivalent to shortening \bar{T}_{I2D} and \bar{T}_{I2R} , and increasing \bar{T}_{R2S} . Therefore, in this study, we do

not consider the state of “Admitted” explicitly. This will be further discussed later in the simulation section.

It should be noted that, in the ripple-spreading process, (3) uses an energy decaying coefficient vector rather than a single coefficient scalar. This is because the infectiousness over time after infection is not a constant, as shown by the examples in Figure 4. Therefore, we can use a piecewise function defined by the vector η to describe the time-varying infectiousness of a given epidemic.

In the reaction of node to ripples, according to (6), the curve of the probability $p_R(i)$ for node i to be infected has a shape of an arctangent function like the solid line in Figure 5. Basically, once $e_N(j, t) \geq \beta(i)$, the larger the point energy $e_N(j, t)$ is, the higher the probability for node i to be infected by node j . The probability $p_R(i)$ does not go up linearly as $e_N(j, t)$ increases. Actually, $p_R(i)$ increases very slowly before $e_N(j, t)$ reaches a certain critical value, around which $p_R(i)$ goes up sharply and then gets almost saturated no matter how large $e_N(j, t)$ is. This is reasonable, because, according to (3), the value of $e_N(j, t)$ largely reflects the distance (spatial, social, or both) between two nodes, and in reality infection occurs mainly within a certain range of spatial and/or social distance, whilst beyond that range the probability of infection is quite low. The coefficient κ determines how sharply the probability curve changes around the critical value of $e_N(j, t)$: the larger the value of κ , the sharper the probability curve. The coefficient δ determines how likely it is that an $e_N(j, t)$ will cause a large probability of infection: the smaller the value of δ , the more likely a large probability. To simplify

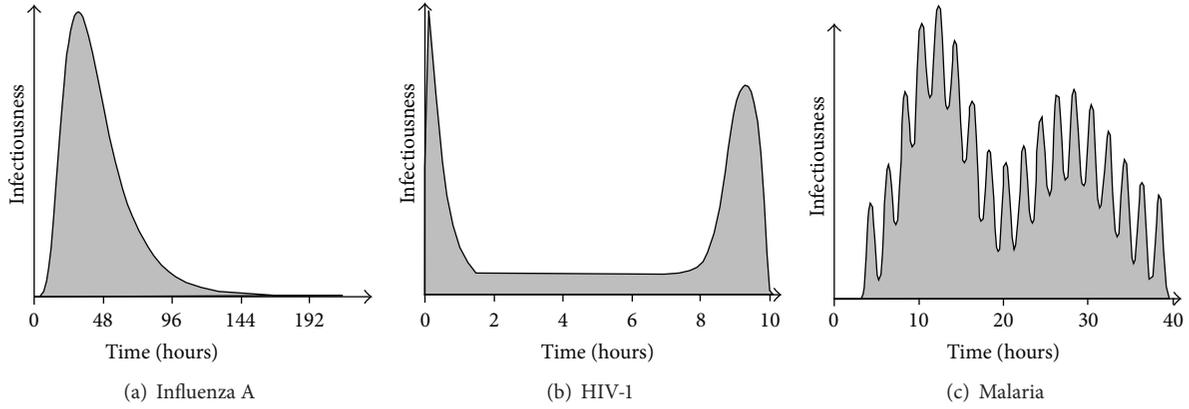


FIGURE 4: Biological infectiousness over time after infection for three different human pathogens [14].

TABLE 1: States used in epidemic model.

Symbol	Epidemic state	Description
S	Susceptible	Healthy Individuals who could potentially develop the disease.
E	Exposed	Individuals who have been infected with the disease, but who are still in the latent period (with or without symptoms of the disease) and who cannot transmit the disease to others.
I	Infective	Individuals who are infected with the disease (with or without symptoms of the disease) and who are capable of transmitting the infection to others.
R	Recovered	Individuals who have recovered from infection thereby acquiring immunity (temporary or permanent) from infection.
D	Dead	Individuals who have died from infection

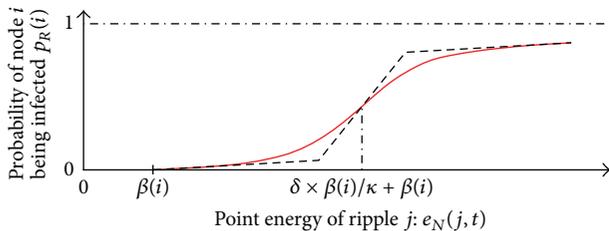


FIGURE 5: Probability of infection and point energy of ripple.

the RSEM, one may sometimes use a piecewise probability function, such as the dotted line in Figure 5, to approximate the arctangent function.

From the description above, it is apparent that the proposed RSEM is more complicated than many existing

epidemic models. This also means that it is more flexible to modifications. For instance, the space where nodes distribute and ripples spread out is not necessarily the real geographic space. Instead, it can be an artificial multidimensional space which includes social relationships, shopping patterns, and habits. The nodes may not always be fixed; in other words, by referring to relevant statistic studies of long-distance travelling patterns, some randomly chosen nodes may occasionally jump in the space. The proposed RSEM can easily incorporate such modifications, but this is beyond the scope of this paper.

3. Genetic Algorithm for Tuning Model Parameters

Like all other models, the proposed RSEM is supposed to be able to (i) simulate historic epidemic outbreaks, (ii) analyze epidemic dynamics/mechanism and health policies/measures, and (iii) make forecasts to some extent. Simulating historic epidemic outbreaks is mainly used for model verification building confidence so that the model will be employed for analysis and forecasting. In the simulation of historic epidemic outbreaks, the output of a model should match the relevant historic data as closely as possible. Basically, this can be assessed by comparing the number of daily infections $N_{DI}(t)$, the number of daily deaths $N_{DD}(t)$, the number of daily recoveries $N_{DR}(t)$, and the basic reproduction ratio R_0 , which is defined as the average number of secondary infections produced when one infected individual is introduced into a population where everyone else is susceptible [14]. In short, the quantity R_0 governs whether an infection can spread and be sustained within a population. If R_0 is greater than one, then the number of infections in a susceptible population will increase and the infection will be sustained, whereas if R_0 is less than one, the infection should fail to take hold and will die out very quickly.

For a simple deterministic model with a few parameters such as those based differential equations, it is possible to tune the parameters by hand based on experience, in order to match relevant historic data. However, for a complicated stochastic model with many parameters, which is the case for the proposed RSEM, hand tuning is very difficult, if

not impossible. Actually, the output of the RSEM does not simply rely on the values of parameters but is more determined by whether such values together satisfy certain conditions, known or unknown. For instance, [21] derived some conditions which involve all parameters to determine whether all nodes will be connected together in the general ripple-spreading network model. Such complicated relationships between parameters reflect the fact that in reality it is usual for many factors to work together in a complex, intermixed fashion. Tuning the parameters for a complicated model/system is often far beyond the capability of even human experts. Fortunately, we can turn to methods such as genetic algorithms (GAs) for help. A GA is a large-scale parallel stochastic searching and optimizing method inspired by the biological mechanisms of evolution and heredity. In recent years, GAs have been widely used for resolving various complex problems, particularly those including parameter optimization [23–26]. The basic idea of a GA is that, given a population of chromosomes, the environmental pressure causes natural selection (survival of the fittest), and hereby the fitness of the population grows. It is easy to see such a process as optimization. Given an objective function to be maximized, we can randomly create a set of candidate solutions (chromosomes) and use the objective function as an abstract fitness measure (the higher the better). Based on this fitness, some of the better chromosomes are chosen to seed the next generation by applying crossover and/or mutation. Crossover is applied to two selected chromosomes, the so-called parents, and results in one or two new chromosomes, the children. Mutation is applied to one chromosome and results in one new chromosome. Applying crossover and mutation leads to a set of new chromosomes, the offspring. Based on their fitness, these offsprings compete with the old chromosomes for a place in the next generation (in some GA implementations, the population is all replaced by the offspring). This process can be iterated until a solution is found or a previously set time limit is reached. Many components of such an evolutionary process are stochastic. According to Darwin, the emergence of new species, adapted to their environment, is a consequence of the interaction between the survival of the fittest mechanism and undirected variations. Variation operators must be stochastic, the choice of which pieces of information will be exchanged during crossover, as well as the changes in a chromosome during mutation, is random. On the other hand, selection operators can be either deterministic or stochastic. In the latter case fitter chromosomes have a higher chance of being selected than less fit ones, but typically even the weak chromosomes have a chance to become a parent or to survive. For more theoretical details of GAs, readers may refer to [23–26].

To apply a GA in this study, firstly, we need to decide, in the RSEM, which parameters can be set up by hand and which ones need to be tuned by the GA. Basically, parameters such as the number of initial cases N_{EISR} , the means of state transfer times \bar{T}_{E2I} , \bar{T}_{I2D} , \bar{T}_{I2R} , and \bar{T}_{R2S} , energy decaying coefficient vector η , and death rate R_D can be easily set up according to relevant historic data and statistic studies. Coefficients κ and δ can also be tuned by hand mainly to

change the overall shape of the infection probability curve as shown in Figure 5. Most ripple-spreading related parameters are purpose-designed and have no real-world meanings or references, and therefore they should be tuned by the GA. These parameters specifically include the initial point energy $E_{EISR}(i)$ of initial cases, the ripple spreading speed s , the threshold $\beta(i)$, and the amplifying factor $\alpha(i)$. It should be noted that we do not need to tune $\beta(i)$ and $\alpha(i)$ for every node, but we just need to tune their means, that is, $\bar{\beta}$ and $\bar{\alpha}$, and then generate $\beta(i)$ and $\alpha(i)$ according to a Poisson distribution with $\bar{\beta}$ and $\bar{\alpha}$ as the means, respectively.

Then, we construct a fitness function based on $N_{DI}(t)$, $N_{DD}(t)$, $N_{DR}(t)$, and R_0 as follows:

$$J = w_1 \left| \bar{R}_0 - R_0^* \right| + w_2 \sum \left| \bar{N}_{DI}(t) - N_{DI}^*(t) \right| + w_3 \sum \left| \bar{N}_{DD}(t) - N_{DD}^*(t) \right| + w_4 \sum \left| \bar{N}_{DR}(t) - N_{DR}^*(t) \right|, \quad (9)$$

where w_1 to w_4 are weights, the variables with “*” are the approximated values of relevant historical data, and the over barred variables are the relevant mean results of a number of random RSEM runs with a given set of parameter values. Usually, the curve of historical data is not smooth and nor is the curve of a single RSEM run. If we directly compare the historical curve with the curve of any single RSEM run, the summed error will usually be very large, even though the two curves are actually similar to each other. Therefore, we need to use approximated values and average values, both of which are much smoother for accurate comparison. Another reason for utilizing the average results of many RSEM runs is that the proposed RSEM is stochastic in nature, which means that the model will deliver different results in different runs even for the same set of parameter values. Therefore, to better assess the effect of a given set of parameter values, a sufficient number of RSEM runs need to be conducted for a parameter value set.

Thus, we can use GA to tune the parameters for the RSEM as follows.

Step 1. Initialize a population of sets of parameter values.

Step 2. For a set of parameter values, conduct a number of RSEM runs, and then calculate the fitness based on the average results. Repeat the same for every set of parameter values in the population.

Step 3. If the termination criterion is satisfied, go to Step 4. Otherwise, perform selection, crossover, and mutation to generate a new population of sets of parameter values. Go to Step 2.

Step 4. Output the best set of parameter values, and stop.

Once the RSEM is well tuned by the GA, we can then use it to analyze epidemic details and policy effects, as well as forecasting new trend or future development of an epidemic. Some relevant technical details concerning GA design and implementation may be found in [22]. It should be noted,

however, that [22] focuses on applying a GA to optimize the basic RSNM of [21] so as to generate small-world and scale-free network topologies. As discussed in Section 2, the basic RSNM is somewhat different from the RSEM proposed in this paper. Nevertheless, the use of a GA to optimize model parameters in [22] is helpful to tune the RSEM in this paper.

4. Illustrative Simulation Results

As explained in Section 3, the first step in applying the proposed RSEM is to use some historical data to tune the parameter values in order to simulate the associated real epidemic outbreak. Here we consider the March 2003 Severe Acute Respiratory Syndrome (SARS) outbreak in Hong Kong. From the empirical data for Hong Kong, we know that a severe localized epidemic occurred in Amoy Gardens. Through contact tracing, it has been estimated that 332 cases were involved in this outbreak, a large proportion of which could be traced to a single person. Outbreaks involving the direct spread of the disease from one primary case to a very large number of secondary cases are often referred to as “Super-Spreading Events”, which are often ideally used to test epidemic models. As one can see from Figures 6, 7, and 8, the real historical data form no smooth curves. Because the SARS case in Amoy Gardens is a unique single case, we cannot get any smooth curve of average historical data, which makes it difficult to calculate the fitness of chromosome as defined in (9). Thus, to be able to apply a GA to tune the reported RSEM, we adopt the differential equations based model in [20] to generate an approximated smooth curve of average historical data. With 3 smooth curves (of daily infections, death, and recovery, resp.) generated by the model in [20], we can calculate the chromosome fitness according to (9) and then apply GA to tune the RSEM parameters. It should be noted that there are actually no daily infection data, and we hence use the daily admissions data in [20] to approximate them. In this study, the weights w_1 to w_4 in (9) are set to 100, 1, 10, and 2, respectively. These values for w_1 to w_4 are chosen largely by referring to the data of basic reproduction ratio, daily admissions, daily death, and daily recovery in [20], so that the four items in (9) will have roughly equal contributions to the fitness function. In the test, all SARS-specific disease related parameters are set up according to [20], and the number of nodes is $N_N = 2396$ (i.e., the susceptible population in Amoy Gardens). For the GA, the crossover probability is set to 0.6, the mutation probability is set to 0.1, the population size is 200, and the number of evolving generations is 500. For each chromosome in a population, we use the parameter values represented by it to run the RSEM for 100 times, and then use the average results to calculate its fitness according to (9). Finally, we use the fittest chromosome in the last generation of GA to set the RSEM, in order to simulate the SARS case in Amoy Gardens.

Figures 6, 7, and 8 plot the final simulation results of applying RSEM to study the SARS case in Amoy Gardens; the standard deviations (SDs) of the results of the model in [20] and the RSEM from the historical data are given in Table 2. From Figures 6, 7, and 8 and Table 2, one can see that (i) the average curves of RSEM are similar to those

of [20]; (ii) the SDs of the RSEM are close to those of the model in [20]; (iii) the results of a single RSEM run have patterns (peaks and valleys) fairly similar to those of historical data; (iv) the ranges and tendencies of RSEM data all match those of historical data and [20]. Therefore, Figures 6, 7, and 8 and Table 2 illustrate that the proposed RSEM and GA can deliver a fairly satisfactory performance in the case study of SARS and therefore can be useful to model epidemic dynamics. However, due to lack of relevant data on the physical fitness, immunity, and social activities of the population in Amoy Gardens, at the moment it is still difficult to explain or make sense of the values of RSEM parameters in this SARS case. Therefore, more cases studies based on different historical data and social survey data are still needed, which is beyond the scope of this paper and demands effort in future research.

Due to the early stage of this work, we have not collected sufficient historical data for parameter tuning purposes. However, we can still investigate the model based on synthetic data sets, in order to understand the properties of the model itself. Here, we can give some simulation results to see how the parameters can affect the output of the model. Now we set $N_N = 250$; that is, the population of original susceptible individuals is 250. We assume there are 8 initial cases randomly distributed in the population. Firstly, we suppose that the population is uniformly randomly distributed in a square area defined by coordinates $(-1000, -1000)$ and $(1000, 1000)$, and all initial cases have the same initial point energy $E_{EISR} = 10000$, the Poisson means for generating $T_{E2I}(i)$, $T_{I2D}(i)$, $T_{I2R}(i)$ and $T_{R2S}(i)$, $\beta(i)$ and $\alpha(i)$ are $\bar{T}_{E2I} = 10$, $\bar{T}_{I2D} = 15$, $\bar{T}_{I2R} = 15$ and $\bar{T}_{R2S} = 20$, $\bar{\beta} = 5$, and $\bar{\alpha} = 1000$, and coefficients $\eta = 1$, $\kappa = 1$, and $\delta = 10$. Then, in the simulation, we want to study the influence of three sets of parameters: (i) the square area (which is related to the density of population), (ii) the pair of $\bar{\beta}$ and $\bar{\alpha}$ (which are related to the overall social activeness of the population), and (iii) the triple of \bar{T}_{I2D} , \bar{T}_{I2R} , and \bar{T}_{R2S} (which are related to the general public health service standard). Each time we only change the value(s) of one set of parameters by 20% of the associated default values, while all the other parameters have the default values as given above. Please note that $\bar{\beta}$ and $\bar{\alpha}$ change in opposite directions ($\bar{\beta}$ increases whilst $\bar{\alpha}$ decreases) and \bar{T}_{R2S} and the pair of $(\bar{T}_{I2D}, \bar{T}_{I2R})$ change in opposite directions (\bar{T}_{R2S} increases whilst the other two decrease). For each given whole set of parameter values, we conduct 10 RSEM runs. Then, we only check the number of total infections, and the average results are given in Table 3, from which we have the following observations.

- (i) The default values for parameters result in a network where all nodes are connected, which means the whole population is eventually infected. This default case is used as a benchmark to assess the influence of some model parameters.
- (ii) The density of population plays an important role in the transmission of infectious diseases. A high density will contribute to the outbreak of plagues.

TABLE 2: Comparative results between RSEM and the model in [20].

Standard deviations	Number of infections		Number of death		Number of recovery	
	Daily	Cumulative	Daily	Cumulative	Daily	Cumulative
Model in [20]	12.7012	26.2656	1.5375	4.2721	7.9375	37.0366
RSEM	12.7217	27.2463	1.5573	3.5244	8.0686	40.9718

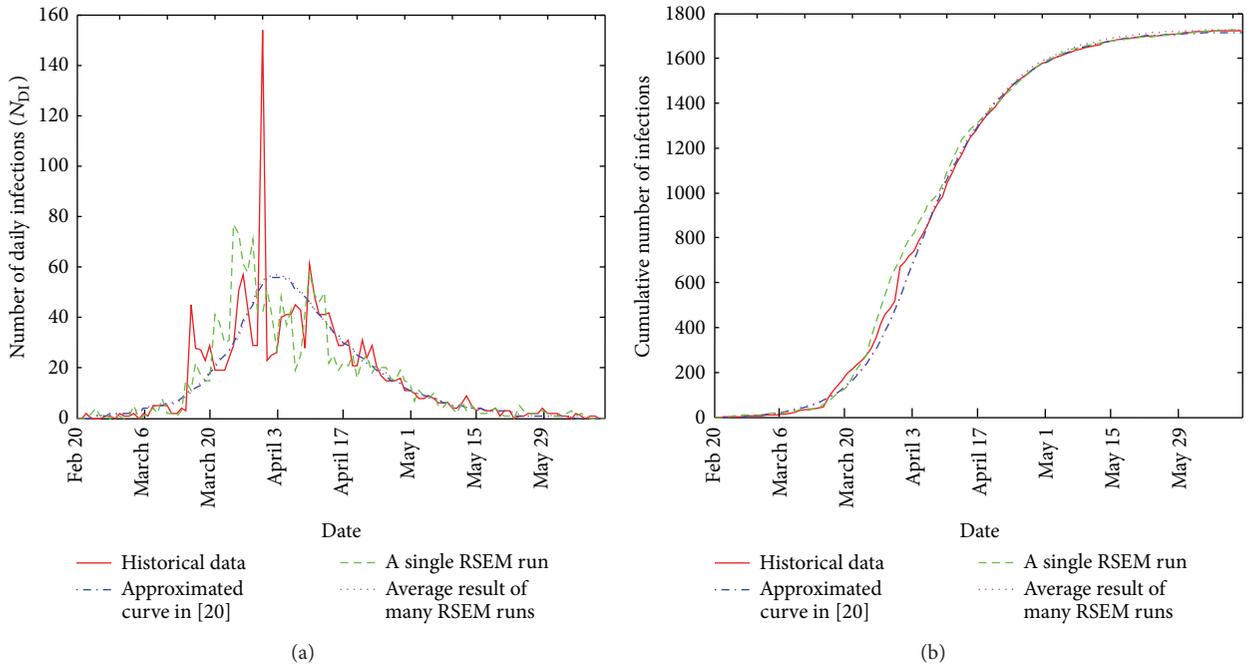


FIGURE 6: Simulation results for Amoy Gardens, Hong Kong: number of infections.

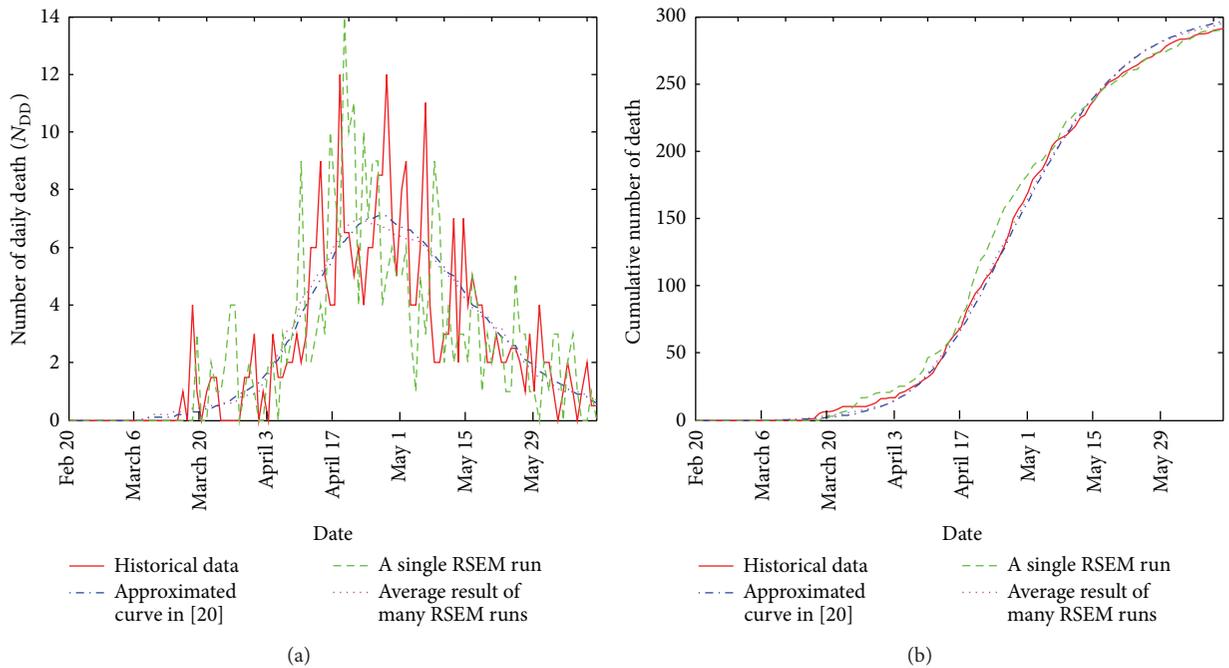


FIGURE 7: Simulation results for Amoy Gardens, Hong Kong: number of deaths.

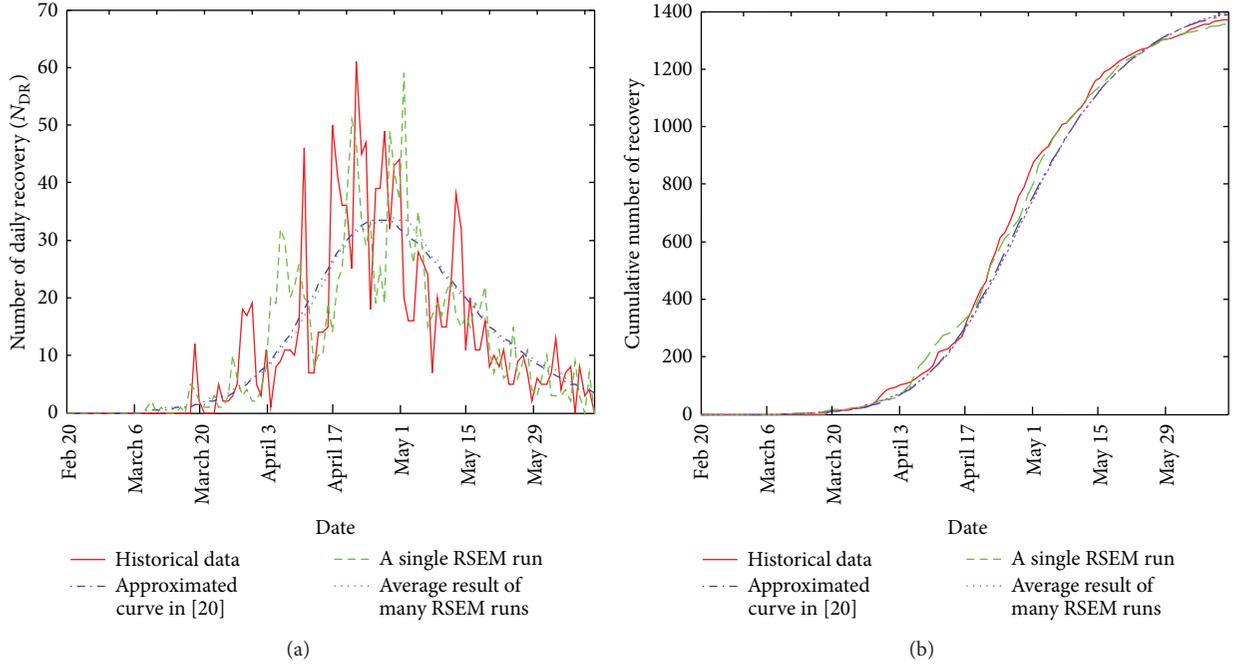


FIGURE 8: Simulation results for Amoy Gardens, Hong Kong: number of recovery.

- (iii) Large threshold and small amplifying factor have a negative influence on the transmission of infectious diseases. This means a community where the overall social activeness is low will have a better resistance to epidemic outbreaks. When an 80% change applies to the default values, usually only the initial cases will cause a few infections, and the epidemic will die out soon.
- (iv) Similarly, the general public health service standard plays a crucial role in the control of epidemic outbreaks. If infected individuals can be quarantined in a timely manner, say, if \bar{T}_{I2D} and \bar{T}_{I2R} are reduced by 80%, then infections are largely limited to the initial cases and the nearby.
- (v) The above simulation results clearly reveal the trends of the influence of different RSEM parameters. However, again, to fully understand the performance of RSEM, extensive efforts are still needed based on various real epidemics data in future research.

5. Conclusions and Future Work

This paper reports a novel ripple-spreading network model for the study of infectious disease transmission. We term the new model the ripple-spreading epidemic model (RSEM). By mimicking the ripple-spreading phenomenon on a calm liquid surface, the new model has many natural advantages to describe spatial and temporal factors in the outbreak of plagues. The effect of individuals' social activities, such as contacts and movements, which are often difficult for many existing models to simulate, can be effectively described by

TABLE 3: Influence of model parameters on total infections.

Change % default values	Square area	$(\bar{\beta}, \bar{\alpha})$	$(\bar{T}_{R2S}, \bar{T}_{I2D}, \bar{T}_{I2R})$
0%	250.0	250.0	250.0
20%	228.7	219.3	232.5
40%	205.7	185.8	191.6
60%	136.4	108.2	93.8
80%	51.9	16.5	21.9

the amplifying factor of nodes and the point energy of ripples, which largely determines the probability of infection. As the point energy of a ripple decays as the ripple spreads out, the probability of infection decreases. The threshold of nodes is another important parameter determining the dynamic process of infectious disease transmission, and basically this parameter is well associated with the physical fitness and immunity of individuals. The proposed model is highly flexible to modifications and is therefore capable of incorporating new real factors which matter in the study of epidemic. The general model can be tuned by genetic algorithm according to the historic data of a specific plague, and the well-tuned model can then be used to analyze and forecast the associated infectious disease. The effectiveness of the proposed model and method is illustrated by some simulation results.

Future work will include (i) carrying out more theoretical analysis of the proposed model, (ii) comparing with other models, and (iii) modifying and tuning the proposed model in order to study some real outbreaks of plagues. In particular,

as a new epidemic model, the RSEM needs relevant data on the physical fitness, immunity, and social activities of the population in a community/society, in order to better set up and understand the RSRPs (ripple-spreading related parameters). To get such data, a purpose-designed social survey needs to be conducted within a certain time window after a plague breaks out. Therefore, conducting relevant social survey for a recent plague event to collect sufficient data will be a crucial part of future work. Once such data are collected for a specific plague event, more comparisons with different relevant existing models will be made. Such application-oriented study needs to be conducted for different diseases, in order to explore the full potential of the proposed RSEM.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the China “973” Project under Grant 2012CB955404, the project Grant 2012-RC-02 from Beijing Normal University, China, and the Seventh Framework Programme (FP7) of the European Union under Grant PIOF-GA-2011-299725. Some early results of this paper were presented at the *2012 the 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012)*, October 16–18, 2012, Chongqing, China.

References

- [1] H. Heesterbeek, “The law of mass-action in epidemiology: a historical perspective,” in *Ecological Paradigms Lost: Routes of Theory Change*, K. Cuddington and B. Beisner, Eds., pp. 81–105, Elsevier, Burlington, Mass, USA, 2005.
- [2] K. Dietz and J. A. P. Heesterbeek, “Daniel Bernoulli’s epidemiological model revisited,” *Mathematical Biosciences*, vol. 180, pp. 1–21, 2002.
- [3] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.
- [4] J. Glasser, M. Meltzer, and B. Levin, “Mathematical modeling and public policy: responding to health crises,” *Emerging Infectious Diseases*, vol. 10, no. 11, pp. 2050–2051, 2004.
- [5] R. M. May, “Uses and Abuses of Mathematics in Biology,” *Science*, vol. 303, no. 5659, pp. 790–793, 2004.
- [6] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, Hafner Press, New York, NY, USA, 2nd edition, 1975.
- [7] C. I. Siettos and L. Russo, “Mathematical modeling of infectious disease dynamics,” *Virulence*, vol. 4, no. 4, pp. 295–306, 2013.
- [8] P. D. O’Neill, “A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods,” *Mathematical Biosciences*, vol. 180, pp. 103–114, 2002.
- [9] A. Kleczkowski and B. T. Grenfell, “Mean-field-type equations for spread of epidemics: the “small world” model,” *Physica A*, vol. 274, no. 1, pp. 355–360, 1999.
- [10] “Computer simulation model provides tactics to deal with flu epidemic,” <http://www.healthjockey.com/2009/05/04/computer-simulation-model-provides-tactics-to-deal-with-flu-epidemic/>.
- [11] R. Connell, P. Dawson, and A. Skvortsov, “Comparison of an agent-based model of disease propagation with the generalised SIR epidemic model,” Tech. Rep. DSTO-TR-2342, 2009.
- [12] M. Toffoli, *Cellular Automata Machines*, Scientific Computation, MIT Press, Cambridge, Mass, USA, 1991.
- [13] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: structure and dynamics,” *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [14] N. C. Grassly and C. Fraser, “Mathematical models of infectious disease transmission,” *Nature Reviews Microbiology*, vol. 6, no. 6, pp. 477–487, 2008.
- [15] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.
- [16] A. I. Reppas, K. G. Spiliotis, and C. I. Siettos, “Epidemiology: from the host-host interactions to the systematic analysis of the emergent macroscopic dynamics of epidemic networks,” *Virulence*, vol. 1, no. 4, pp. 338–349, 2010.
- [17] R. Breban, R. Vardavas, and S. Blower, “Linking population-level models with growing networks: a class of epidemic models,” *Physical Review E*, vol. 72, no. 4, part 2, Article ID 046110, p. 8, 2005.
- [18] M. J. Keeling and K. T. D. Eames, “Networks and epidemic models,” *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 295–307, 2005.
- [19] D. Helbing, “Globally networked risks and how to respond,” *Nature*, vol. 497, pp. 51–59, 2013.
- [20] N. Jia and L. Tsui, “Epidemic modelling using SARS as a case study,” *North American Actuarial Journal*, vol. 9, no. 4, pp. 28–42, 2005.
- [21] X.-B. Hu, M. Wang, M. S. Leeson, E. L. Hines, and E. Di Paolo, “Deterministic ripple-spreading model for complex networks,” *Physical Review E*, vol. 83, no. 4, Article ID 046123, 2011.
- [22] X. B. Hu, M. Wang, and M. S. Leeson, “Ripple-spreading network model optimization by genetic algorithm,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 176206, 15 pages, 2013.
- [23] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.
- [24] E. Goldberg, *Genetic Algorithms in Search Optimization & Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.
- [25] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, Germany, 2nd edition, 1994.
- [26] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Natural Computing Series, Springer, Berlin, Germany, 2003.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

