

Decisions under uncertainty: a computational framework for quantification of policies addressing infectious disease epidemics

Armin R. Mikler · Sangeeta Venkatachalam ·
Suhasini Ramisetty-Mikler

Published online: 17 April 2007
© Springer-Verlag 2007

Abstract Emerging infectious diseases continue to place a strain on the welfare of the population by decreasing the population's general health and increasing the burden on public health infrastructure. This paper addresses these issues through the development of a computational framework for modeling and simulating infectious disease outbreaks in a specific geographic region facilitating the quantification of public health policy decisions. Effectively modeling and simulating past epidemics to project current or future disease outbreaks will lead to improved control and intervention policies and disaster preparedness. In this paper, we introduce a computational framework that brings together spatio-temporal geography and population demographics with specific disease pathology in a novel simulation paradigm termed, global stochastic field simulation (GSFS). The primary aim of this simulation paradigm is to facilitate intelligent what-if-analysis in the event of health crisis, such as an influenza pandemic. The dynamics of any epidemic are intrinsically related to a region's spatio-temporal characteristics and demographic composition and as such, must be considered when developing infectious disease control and intervention strategies. Similarly, comparison of past and current epidemics must include demographic changes into any effective public health policy for control and intervention

strategies. GSFS is a hybrid approach to modeling, implicitly combining agent-based modeling with the cellular automata paradigm. Specifically, GSFS is a computational framework that will facilitate the effective identification of risk groups in the population and determine adequate points of control, leading to more effective surveillance and control of infectious diseases epidemics. The analysis of past disease outbreaks in a given population and the projection of current or future epidemics constitutes a significant challenge to Public Health. The corresponding design of computational models and the simulation that facilitates epidemiologists' understanding of the manifestation of diseases represents a challenge to computer and mathematical sciences.

Keywords Global stochastic field simulation · Infectious diseases

1 Introduction

Epidemics of infectious diseases have plagued humankind since historical times. There are accounts of epidemics dating back to the times of Hippocrates (459–377 B.C.) and the ancient Greeks (Bailey 1957). The fourteenth century Europe lost a quarter of its 100 million people to Black Death. The fall of the Aztecs empire in 1521 was due to smallpox that eradicated half of its 3½ million population. The pandemic influenza of 1918 caused over 20 million excess deaths in 12 months. More recently, the severe acute respiratory syndrome (SARS) outbreak of 2003 highlighted the rapid spread of an epidemic at the global level. The outbreak, emanating from a small Guangzhou province in China, spread around the world requiring a concerted response from public health admin-

A. R. Mikler (✉) · S. Venkatachalam
Computational Epidemiology Research Laboratory,
Department of Computer Science and Engineering,
University of North Texas, Denton, TX 76203, USA
e-mail: mikler@cs.unt.edu

S. Ramisetty-Mikler
Dallas Regional Campus,
University of Texas School of Public Health,
Dallas, TX 75390, USA

istrations around the world and the World Health Organization (WHO) to curtail the epidemic (Heymann and Rodier 2004). The WHO (2004) and Centers for Disease Control and Prevention (CDC) (2004) actively engage in worldwide surveillance of infectious diseases, and prioritize prevention and control measures at the root cause of epidemics.

The lurking threat of emerging and re-emerging diseases, and the necessity to prepare for disaster in the wake of bioterrorism raise complex issues for Public Health researchers in general and Epidemiologists in particular. This research requires computational support to facilitate policy and decision-making under uncertainty to allocate limited public health resources. This paper addresses these requirements through the development of a computational framework for modeling and simulating infectious disease outbreaks in a geographic region that allows for the quantification of public health policies. The proposed framework is based on a novel concept of GSFS and utilizes information about regional demographics, geography, and disease parameters. GSFS is a hybrid approach to modeling, implicitly combining agent-based modeling with the cellular automata paradigm. Specifically, the GSFS will facilitate the effective identification of high risk groups in the population and adequate points of control, leading to more effective surveillance and control of infectious disease epidemics. Preliminary results for the incidence and prevalence of a simulated influenza-like infectious disease outbreak in parts of Denton County, Texas, elucidate the utility of the proposed modeling and simulation approach.

1.1 Modeling epidemics

In what follows, we are introducing a new framework for modeling infectious disease epidemics in a given population. The primary goal is to facilitate what-if-analysis that will allow the formulation of public health strategies in the event of infectious disease epidemics. It is essential to recognize that the dynamics of an epidemic are tightly coupled with the geography and demographics of a region in which an outbreak has manifested itself. This suggests that results that have been obtained by analyzing a disease outbreak in one particular geographic location may not be readily applicable when defining control and prevention strategies in other regions. Similarly, one must recognize that comparison of morbidity/mortality of two or more past epidemics for the purpose of deciding control measures must take demographic changes during these years into consideration. Further, this necessitates the availability of computational tools, which enable epidemiologists to model an outbreak by bringing together knowledge of disease, geography, and demographics from past and present.

1.2 Epidemics in time and space

The spatial and temporal dynamics of the distribution of the population in the United States is of great concern to planners, service providers and epidemiologists in both public and private sectors. For example, changes in socioeconomic status, lifestyle, and demand for appropriate living conditions and health-care services, as the geographical pattern of the elderly in many American cities has changed considerably in the past decade. Unfortunately the nature of this change is not clearly documented and still being studied. Some states are more attractive to elderly. However, limited research exists on association between an extensive set of location-specific factors and the migration of retirement-age individuals. For instance, Duncombe et al. (2003) estimated an individual-level location-choice model by using a combination of place-characteristics data and Census county-to-county migration data, identifying income taxes to have the largest relative effects. However, other factors, including climate, economic conditions, and population characteristics, appear to play much larger roles in migration and location decisions. Such findings may provide clues to understanding geographic distribution and the change in the level of health care demands, particularly in the context of aging America. Understanding of the temporal pattern of the geographic distribution of the population, including the urban elderly, will provide insight into the future trends and facilitate strategic planning of needed services. Disease surveillance systems based on antiquated demographic data are inadequate. Thus, examining and evaluating the appropriateness of current community services available to the elderly is critical for identifying and eliminating gaps in service.

In anticipation of the baby boomers beginning to turn into the aged cohort in 2010, a thorough study on the spatial and temporal changes of urban elderly seem imperative. Human migration, changes in economy, births, and deaths, cause the demographic characteristic of geographic region to change continually. Certain life cycle changes—marriage, birth of a child, age of current children—all impact a family's housing needs and causes people to migrate to different places. These *natural* changes in the demographic characteristics necessitate demographic data to be continually updated.

Rogerson and Han (2002) report that migration can have a serious effect on the detection of geographical differences in disease risk. In general, areas of high in-migration are also characterized by high out-migration, and there is substantial regional variation in mobility rates. Hence, disease data and disease prevalence projections based on human interaction patterns, are easily and quickly outdated if there are significant changes in demographics resulting from migration or urbanization. Population structure plays

an important role in determining the spreading patterns of infectious diseases among humans, forcing us to consider meta-population models making an explicit distinction between the intra- and inter-community interactions (Sattenspiel and Dietz 1995). Environmental factors associated with the location of contact can have important effects on transmission risk. For instance, a natural disaster such as Hurricane Katrina has led to drastic changes in the demography in the New Orleans as well as in surrounding states.

Public health policy and disaster preparedness has often relied on historical data of past epidemics. This is particularly true for the comparison of specific epidemics on the basis of the associated attack-rate or reproduction number R_0 , as used by Longini et al. (2005) to develop strategies for containing pandemic influenza. R_0 is the expected (average) number of new infectious individuals in a completely susceptible population produced by a single infectious individual during their entire period of infectiousness. Accordingly, when $R_0 > 1$, an epidemic occurs in a completely susceptible population; if $R_0 < 1$, the disease dies out and cannot establish itself in the population (Ferguson et al. 2005). Traditionally, R_0 was computed directly from the equations that form the SIR-type models used to describe disease dynamics in a homogeneous population. More recently, the attack rate R_0 of a specific infectious disease has been determined by analyzing data that has been collected during the epidemic (Glass et al. 2006). As discussed above, interaction patterns are a function of changing demographics and infrastructure in response to migration and urbanization. Thus, to be able to analyze and compare past epidemics to present ones, the R_0 of a past outbreak must be adjusted to account for such changes. This necessitates the analysis of how past disease outbreaks may have manifested themselves in the corresponding demographics. The displacement of individuals caused by Hurricane Katrina, for instance, is therefore an important factor that must be taken into consideration when predicting or planning for potential epidemics. GSFS facilitates the prediction of how infectious disease outbreaks will manifest in the current demographics through the simulation of contacts, which are deemed the primary determinant of an epidemic's dynamics.

2 Approaches to modeling epidemics

With the ever-present risk of infectious disease outbreaks, it has become imperative to develop new methodologies that facilitate the preparedness and training of public health professionals. Recent examples of epidemics possibly pandemics include SARS and Avian Influenza. Further, the threat of bio-terrorism forces epidemiologists to develop

disaster preparedness plans that outline explicit responses to possible disease outbreaks. Newly emerging or re-emerging infectious diseases continue to occur regularly (Heymann and Rodier 2004). Some diseases have changed their appearance, some have become resistant to drug treatment, while others are so new that no previous outbreaks have ever been studied.

Medical research has enhanced the understanding of disease characteristics in an individual. For example, the epidemiologic stages of influenza as described by *Latent Period*, *Infectious Period*, and *Recovery Period* (Benenson 1995) are well known (Benenson 1995). So are the symptomatic stages of influenza (i.e., *incubation period* until symptoms occur) as shown in Fig. 1. The *Susceptibles–Infectives–Removals* (SIR) state diagram Fig. 2. illustrates the course of a disease in an individual. The manifestation and spread of many infectious diseases in the population remain elusive and are dependent on socio-behavioral interaction patterns and population dynamics.

To gain insight into the intricacies of disease dynamics in a specific population, statistical and mathematical models of infectious disease epidemics have been developed. Recently, some computational disease models have emerged, which facilitate the simulation and investigation of different disease characteristics. These include models that exploit SIR paradigm, *Cellular Automata (CA)* methodology, *Agent-Based Modeling* and *Bayesian Reasoning*.

Most of the work in modeling infectious disease epidemics is mathematically inspired and based on differential equations and SIR/SEIR model (Aron 2000; Bagni et al. 2002). Differential equations and SIR modeling rely on the assumption of closed population and neglect the spatial effects (Boccara and Cheong 1993; Boccara et al. 1994).

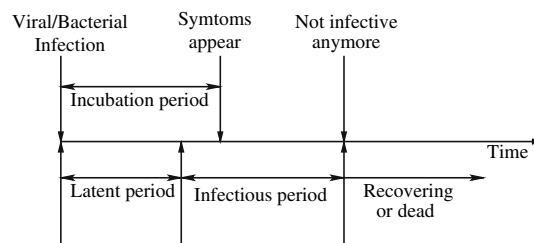


Fig. 1 Infection time-line

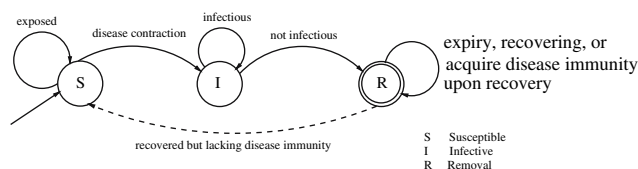


Fig. 2 SIR/SEIR state diagram

They often fail to consider individual contact/interaction processes, assume populations to mix homogeneously, and do not include variable susceptibility. Both partial and ordinary differential equation models are deterministic in nature and neglect stochastic or probabilistic behavior (Stefano et al. 2000). Ahmed and Agiza (1998) introduce incubation and latency time that lead to an accelerating impact on the spread of a disease epidemic.

Most mathematical models are based on the interaction principles between groups of *susceptible*(*S*), *exposed*(*E*)/*infective*(*I*), and *removed*(*R*) individuals, i.e., the *SIR/SEIR* model. *Susceptibles* are those individuals in a population who can be infected by the disease under study. *Infectives* are those individuals who have been infected and are infectious. *Removals* include all individuals that are incapable of transmitting the infection, and are either recovering, fully recovered or expired from the disease. In complex models, the removals who recover may revert to susceptibles, leading to a *SIS* model, if the exposure to the disease does not result in lifelong immunity. The Kermack–McKendrick Threshold Theorem (Bailey 1957) is the basis for the *SIR* model. A continuous influx of susceptibles is a requisite for sustained infection in a population. The model is based on the presumption of a closed homogeneous population, assuming that the epidemic spreads sufficiently fast such that the changes brought in by births, deaths, migration and demographic changes are negligible (Aron 2000).

The spatial and temporal correlation of influenza epidemics in the United States, France, and Australia from 1972 to 1997 has been analyzed using the *SIR* model (Viboud et al. 1972). The results indicate a high correlation between United States and France, but irregularity in the patterns between Australia and the other two countries. Geography, demography, cultural diversity and the resulting varied socio-behavioral interactions are highlighted as the reasons for the discrepancies, and call for computational modeling for further investigation.

The *SIR* model provides a simple framework to represent the spread of a disease. However, it does not provide sufficiently accurate insight into the composition of an epidemic to be used as a policy and planning tool for the allocation of public health resources. The *SIR* model does not take into account the geography or the spatial dimensions of a region, i.e., it does not model the fact that the probabilities of contacts may be distance dependent. Further, the spread of a disease may depend on the specifics of geography and demographics of a region. While the *SIR* model could potentially be extended to include geography and demographics this would drastically increase its intrinsic complexity, thus rendering the model computationally infeasible.

Cellular automata have been used for several decades for computational models (Fu and Milne 2003). A two

dimensional automaton is used in epidemic models utilizing cellular automata (Ahmed and Agiza 1998; Fu and Milne 2003; Situngkir 2004; Stefano et al. 2000). Each cell may represent an individual or a sub-population, and is characterized with state and likelihood risks for exposure and contraction of the disease. The disease progression is studied through its diffusion across the neighboring cells.

The earliest example of use of cellular automata is Bailey's lattice model (Bailey 1967) for the spread of diseases from micro-level interactions. Di Stefano et al. (2000) have developed a lattice gas cellular automata model to analyze the spread of epidemics of infectious diseases. This model, however, does not consider the critical factor of infection time-line. Fu (2002, 2003) has used stochastic cellular automata to model epidemic outbreaks that take into account the spatial heterogeneity. Situngkir (2004) has developed a dynamic model of spatial epidemiology to study avian influenza disease in Indonesia and uses cellular automata for computational analysis. Naive cellular automata are impeded by a limited neighborhood, and the social interactions based on demographics are not readily incorporated. The authors have introduced the global stochastic cellular automata paradigm, addressing the issue of limited neighborhood in a classical CA (Mikler et al. 2005). In order to overcome the limitations posed by naive cellular automata, we introduce GSFM, which incorporates the demographics of location and population density. The current models can be potentially extended to include geography, demographics, and social dynamics, nevertheless, the drastic increase in intrinsic complexity may render the model computationally infeasible. The restrictions and scalability limitations of the current models will be addressed by the proposed computational framework for modeling infectious diseases. This framework will complement the current existing methodologies with studies of heterogeneous population, including dynamic interactions based on geography, demography, environment and migration patterns.

Spatially delineated regions with a small ($< < 10,000$) population can be constructed using an **agent-based** approach, in which each individual is represented by an autonomous agent (Mikler et al. 2004). Larger models with millions of agents necessitate the use of large computing clusters or grid computing that can provide the necessary computational power. The parameters that control interactions among individuals are generally predetermined through social science research when a population's real-world mixing patterns are studied. The agent based model is then used to understand the progression of diseases in a simulated agent society by observing the emergent behavior of the epidemic. Real work mixing patterns and social interactions can be modeled by **social networks**, which have become increasingly important in

our understanding of complex networks and the epidemic spread of diseases in the real world. A social network is a social structure made of nodes each representing individuals or organizations. Links indicate the ways in which individuals or groups are connected through various social familiarities ranging from casual acquaintance to close familial bonds. The term social network was first coined in 1954 by Barnes (1954). Much research has been conducted in the past half century on social networks; however, only in the last decade have researchers in a variety of domains (i.e., computer scientists, physicists, mathematicians) become interested in this field. Complex networks are comprised of lattice-type, small-world and scale-free network structures. Social networks have many of the same properties as other real world networks such as degree distribution. However, one of the large differences between a social network and other complex networks (i.e. topology of the Internet) is network transitivity (Barabasi et al. 2002). The clustering in social networks occur with greater frequency than pure chance, or more casually “party people party together”. Examples of social networks include scientific collaboration networks, friendship networks in the blogosphere, and networks of human contacts (Barabasi et al. 2002; Flake et al. 2000; Liljeros et al. 2001]. Applying methods in social network analysis to public health and epidemiology has grown in the recent years. Some of these methods include agent-based simulation to model the spread of infection on a population (Eubank 2002) and targeted social distancing to mitigate influenza attack rates. An interesting extension to the work by Glass et al. (2006) would be to not only mitigate attack rate by targeted social distancing but to borrow the concept of min-cuts from graph theory to determine an epidemic distance in the social network and target responses in those areas with maximum flow of infection. Applying social network analysis to problems in population-health has many exciting open research opportunities in the future. However, a cumulative modeling error that may be introduced when the number of individuals increases may grow prohibitively and thus it is essential to represent members of society with high fidelity, as attainable by behavioral statistics. Agent-based models have been used to analyze HIV/AIDS spread in the population and individual immune levels following the infection (Callaghan 2005). A survey of agent-based epidemic simulation models is available (Bagni et al. 2002). BioWar (2004) is an agent-based system that analyzes the disease spread, treatment, and recovery, by porting principles of interactions from social, knowledge and work networks. The authors have applied agent-based models to analyze real world outbreaks of tuberculosis in factory and homeless shelter settings (Mikler et al. 2004; Oppong et al. 2004).

The **Bayesian paradigm** incorporates the capabilities of probabilistic reasoning and reasoning under uncertainty. Probabilistic and stochastic analysis are integral to Bayesian methodologies and give a closer view of the real world compared with rule-based systems. Bayesian learning has been successfully applied in the areas of medical diagnosis, weather forecasting, gaming, and fault diagnosis. Nevertheless, in the field of modeling epidemics and their analysis, the Bayesian paradigm has been rarely utilized to its full potential. In the Amazon region, where onchocerciasis (river blindness) is endemic, Bayesian reasoning has been used to identify communities that needed priority ivermectin treatment (Carabin et al. 2003). Spiegelhalter et al. (1999) investigated the utility of Bayesian analysis for health technology assessment and highlighted its practical advantages in handling complex inter-related problems. Bayesian monitoring of critical factors in cancer related clinical trials, such as toxicity and quality of life measures, led to higher accuracy (Fayers et al. 1997). An epidemiological model using Bayesian analysis has been developed for malaria in Ndiop, Senegal (Cancre et al. 2000). The authors have used Bayesian learning to infer the dependency of disease incidence on the demographics in different geographic regions (Abbas et al. 2004).

3 Global stochastic field simulation framework

The GSFS paradigm is a hybrid of agent-based simulation and cellular automata. Rather than restricting the interactions between geographic regions to a well defined neighborhood as in the CA paradigm, GSFS models the spread of diseases based on a field representation of a geographic region. A field is an overlay of the geographic region encompassing the spatial distribution of population and interaction distributions. Each location in the field is assumed to contain a population of n individuals with associated demographics as obtained from US Census data. Individuals belonging to specific locations can be characterized by a state and likelihood of risks for exposure and contracting the disease. A set of four possible states (S, L, I, R) has been defined to signify individuals’ clinical disease stage. As opposed to a purely agent-based model, a stochastic field simulator, maintains the statistics of the three states for each location. Disease spread is driven by contacts generated based on population statistics, unlike agent-based models, for which individuals (agents) themselves engage in contacts. Further, if a model of a spatially delineated region with a small population is to be constructed, an agent-based approach, by which each individual is represented by an autonomous agent is possible (Barfoot and D’Eleuterio 2001; Billari and Prskawetz 2003). However, the composite modeling error introduced

when the the number of individuals increases, is prohibitive. Further, it would require collecting information about individuals' behavior, which can only be coarsely described.

In order to model spatial spread of disease over a geographic region with a large population, it is important to understand the underlying population and demographic dynamics of the region. Consequently, one must rely on other means to derive the population dynamics that promote the spread of diseases. This can be accomplished by utilizing publicly available datasets, that describe composition and behavior of the population of interest. For example, US census information provides necessary data that describes the population in terms of socio-demographic, race/ethnicity, age, gender, etc., at different levels of geographic aggregation. Geographic information systems (GIS) facilitates the integration of information from different sources for a specific geographic region or location. Any larger geographic region, such as a city, county, or state can be decomposed into individual census blocks. We are proposing to use this structure as an overlay to a global stochastic field, which will use the associated census information to define its corresponding interactions among individuals and places. For demonstration purposes, an age-structure of the population has been incorporated into the model as one of the demographic constraints. In order to model different behavior patterns among individuals, the age-structure has been categorized into four groups, which have been chosen to model the following interaction characteristics:

- **Children** are more likely to interact locally, making contact with other children in daycare settings or pre-K and elementary school environments.
- **Youths and Young Adults** represent the sub-population that interacts across small distances in the context of schooling or employment.
- **Adults** represent individuals with well established contact patterns. Contacts for this group may be across different distances for everyday activities such as shopping and employment.
- **Elderly** form a mostly isolated group whose members may have fewer contacts and limited interaction distances.

Global stochastic field simulation model is implemented to incorporate heterogeneous populations. Regional census block data is imported in vector format to a GIS package. This vector file is then converted into a raster file with each block representing a unique location with specific demographic. Let P_i and N_i be the population and number of cells in the i th census block respectively. Let C_{ij} be the population of the j th cell of the i th census block. The population of each census block is assumed to be uniformly

distributed among all the cells in that block i.e., $C_{ij} = P_i/N_i$; where $j = 1$ to N_i and $i = 1$ to m and m is the number of census blocks in the county.

Upon introducing the infection to the population at a specific location, interactions drive the spread of the disease. Rather than simulating all possible interactions, GSFS models contacts, which are a subset of all possible interactions that facilitate the spread of disease. In general, a contact is any interaction between two individuals, which has the potential for successful disease transmission. While the concept of contact modeling seems intuitive, one must carefully consider what constitutes a contact when modeling epidemics. Specifically, we must acknowledge that contacts necessary to transmit different types of diseases differ greatly. For instance, pathogens of Influenza, Syphilis, and Athlete's Foot utilize completely different modes of transmission, thereby defining the type of contact necessary to spread the disease. The proportion of contacts emitted by a location is a function of the population and demographic characteristics of that location. Two individuals that are participating in a simulated contact are representatives of their corresponding location and demographic characteristics. Consequently, GSFS models heterogeneous mixing in the population by establishing contacts non-deterministically. To generate a contact, age and location are randomly chosen based on the demographics and population proportions of the location. Subsequently, individuals belonging to locations with larger populations and age group with higher contact rates have a higher probability of being chosen for a contact as compared to the ones belonging to locations with smaller populations and lower contact rates. Once the two end points are identified, based on the proportions of infectious and susceptible populations present in the locations they belong to, their clinical stage of being susceptible, latent, infectious, or recovered is decided using a random experiment. The probability of transmission referred to in the model as infectivity is the virulence of the virus strain being modeled. Based on the infectivity, the infection may be transmitted to the susceptible individual. During each simulation day individuals are moved from infectious to recovered state and from latent to infectious state based on the latent period and infectious period of the disease.

3.1 Simulating contact structure

Global stochastic field simulation models heterogeneous mixing in the population by establishing contacts randomly but not arbitrary. A contact, as defined above is any interaction between two individuals which can lead to a successful disease transmission. To generate contacts proportional to the demographics, GSFS utilizes probability distributions for contact frequencies for different sub

groups. Assuming that contacts among individuals are Poisson distributed over time, the effective contact rate for a location is determined by a Poisson random variate. Thus, the number of contacts to be simulated is determined independently for each time frame of the simulation. For example, if contacts are modeled as a function of age, the contact rates associated with different age groups will determine how many contacts are to be simulated in each time step.

3.1.1 Generating contacts proportional to demographics

Let N_i be the number of individuals in age group i and $N = \sum N_i$. Let C_i be the contacts established by the age group i where $C_i = N_i \times CR_i$. Here, CR_i is the contact rate specific to each demographic subgroup; C is the total number of contacts established in the region $C = \sum_i C_i$. Let P be the probability distribution and p_i be the contact probability associated with each age group, given by: $p_i = \{p_{i-1} + C_i/C \text{ if } i > 1\}$; $p_i = \{C_i/C \text{ if } i = 1\}$; where $\sum_{i=1}^n p_i = 1$. This approach generalizes for any categorical parameter and thus provides a means for modeling of contacts as a function of specific demographic characteristics.

Modeling outbreaks of infectious diseases using the traditional cellular automata (CA) model is constrained by neighborhood saturation. The classic SIR model is oriented towards a homogeneous population with uniform mixing. GSFM can be used to model outbreaks of infectious diseases. It facilitates the analysis of disease progression in heterogeneous environments and can incorporate geography, demography, environment, and migration patterns into the interaction measure between cells on a global neighborhood level. To facilitate surveillance, monitoring, prevention and control of different diseases, computational models must be developed. As compared to agent-based models, contacts generated in GSFS models are not driven by individual agents but rather as a function of a Monte Carlo process that distributes contacts across the entire

population according to the underlying demographic and geographic characteristics.

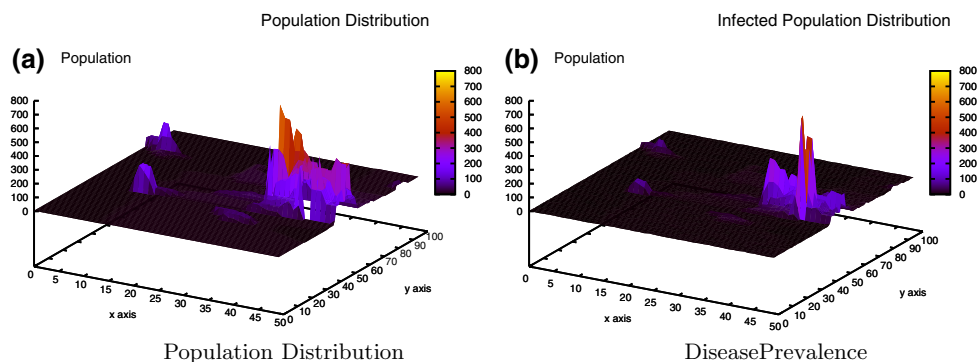
4 Experimental results

The features of GSFS have been evaluated through the simulation of outbreak models based on population data for the northern part of Denton County, Texas. The disease-specific parameters have been selected to approximate an influenza-like disease. Figure 3a depicts the population distribution of the areas in Denton County that have been modeled, whereas Fig. 3b illustrates the disease prevalence over that region as determined by GSFS. The total population of the region is 110,000 and the disease prevalence is 48,000. The contact patterns between individuals have been based on several parameters. These include population density, Euclidean distances between regions, and age-structure. GSFS models are easily extended to incorporate other geographic, demographic, environmental, and migration patterns.

Networks of social contacts channel the transmission of airborne infections. The dynamics of an airborne infectious disease like influenza hinges on the simulation of contacts that are established in the population, as most of the transmissions take place during these contacts. The contact structure of a population generally varies as a function of the number of contacts made by individuals, the context of the contact, the age group of the contact, and distance of the contact. An experiment was conducted during which different contact patterns were generated by associating specific contact rates among the four different age groups.

Global stochastic Field Simulation models different interaction patterns based on different contact rates. From Fig. 4 it is evident that the incidence level of infection in the population varies as a function of the contact rates of age groups. Table 1 summarizes the contact rates used in the experiments for each of the age groups. Higher contact rates for the second and the third age group resulted in higher incidence level of the simulated influenza-like

Fig. 3 Disease prevalence in a heterogeneous population



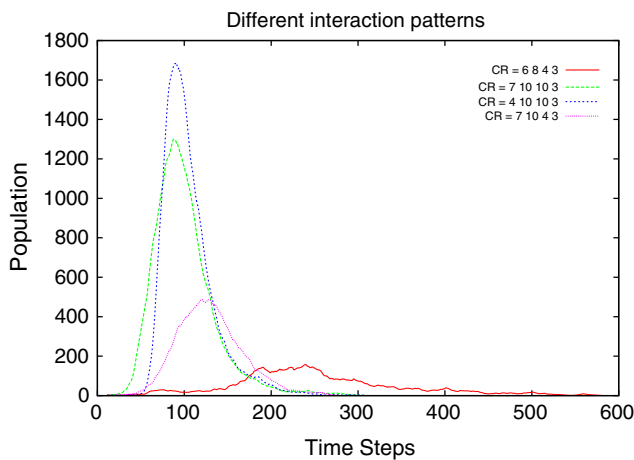


Fig. 4 Incidence of an influenza-like epidemic for different interaction patterns

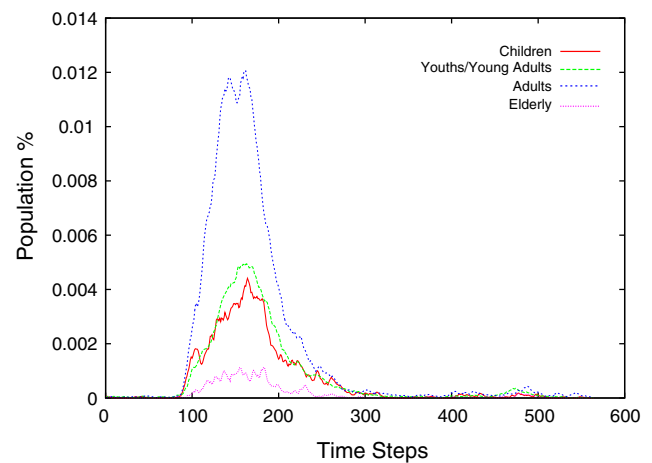


Fig. 5 Epidemic curves for different age groups in the population

Table 1 Contact rate

Age range	Average CR	Average CR	Average CR	Average CR
Children	6	7	4	7
Youths and young adults	6	10	10	10
Adults	4	10	10	4
Elderly	3	3	3	3

illness. It is noteworthy, however, that the reduction in contact rate *CR* for *children* from seven to four results in an epidemic curve that peaks at a lower level. However, the number of infections increases earlier in the epidemic time line. This experiment highlights the importance of modeling disease dynamics for different contact rates as it depicts the effects of health policies that result in reduced contact rates, thereby delaying or even eliminating the onset of an epidemic.

The incidence level of the disease in each group is a function of group-specific interaction modeled by different contact rates. Table 2 shows the respective contact rates used for each age group. The results in Fig. 5 show normalized outbreak-graphs of the resulting epidemic for the four different age groups. The graphs show the dynamics of the epidemic for each age group as opposed to the cumu-

Table 2 Contact rates

Age group	Average contact rate
Children	10
Youths and young adults	10
Adults	6
Elderly	3

lative incidence as observed in Fig. 4. GSFS allows for the analysis of the difference in incidence across demographic groups, which is an important feature that facilitates the formulation of affective public health policy.

To evaluate the effect of the age as a demographic characteristic experiments were conducted by introducing index cases in different age groups. Table 3 shows the total number of infected individuals in the population for index cases in different age groups. An index case in the *elderly* age group does not result in a spread of the disease, whereas one index case in the *Youths/Young Adult* age group triggers an epidemic. We postulate that the different behavior as represented by different contact rate in these groups facilitate the spread of the disease and thus the manifestation of an epidemic. Nevertheless, it is important to note that GSFS is based on a Monte Carlo methodology, thus any results should be interpreted as trends, rather than expected outcome as result of specific parameter choices.

The probability of a contact resulting in successful disease transmission depends on the disease infectivity. *Infectivity* is defined as a pathogen’s ability to manifest an infection. When modeling virus strains the infectivity can be thought of as the differentiating factor between the strains. The infectivity of a strain is an important factor that determines the dynamics of disease spread. Experiments have been conducted to demonstrate the prevalence of

Table 3 Experiments with different index cases

No. of index cases	Age group	Total infected
2	Children and elderly	18601
1	Youths/young adults	12377
1	Elderly	2
2	Youths/young adults and adults	22384
2	Children and adults	22424

influenza-like illness for varied levels of infectivity. Figure 6 illustrates that incidence decreases for lower levels of infectivity. The remaining parameters were similar to that of the basic experiments described above. This experiment exemplifies the sensitivity of the infectivity parameter. As this parameter is primarily related to the disease dynamics, it has been maintained uniform across different age groups.

The GSFS models facilitate the modeling and simulation of a single disease outbreak in a large geographic region. The results obtained from this model represent the severity of an epidemic over time, allowing an epidemiologist to quantify the incidence and prevalence, as computed by GSFS, in response to employing different public health policies (e.g., vaccination strategies). In addition to comparing and quantifying different outbreak scenarios, the experimental analysis of GSFS revealed that an epidemic as observed by health care providers and public health officials, is the cumulative effect of multiple spatially and temporally distributed small outbreaks as shown in Fig. 7. In the context of influenza-like illnesses, the temporal-spatial progression of the disease account for cases that are observed by health care providers during a flu season. Clearly, population density and specific age strata are important demographic parameters that will determine how the disease will manifest itself in a particular sub-region. For instance, the infectious period for influenza in young children is known to exceed that of adults. Hence, one could expect cells (or sub-regions) with a larger proportion of children to display an increased prevalence of influenza as compared to regions with a larger proportion of adult population. Further, it is known that children are the primary transmitters of influenza. Consequently, one might hypothesize that the composition model will yield results that reflect an accelerated spread among regions with larger proportion of children. The model can be used to investi-

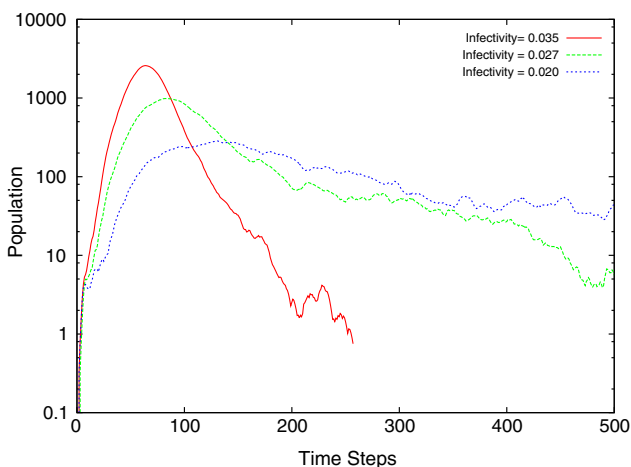


Fig. 6 Differences in disease dynamics as a function of infectivity

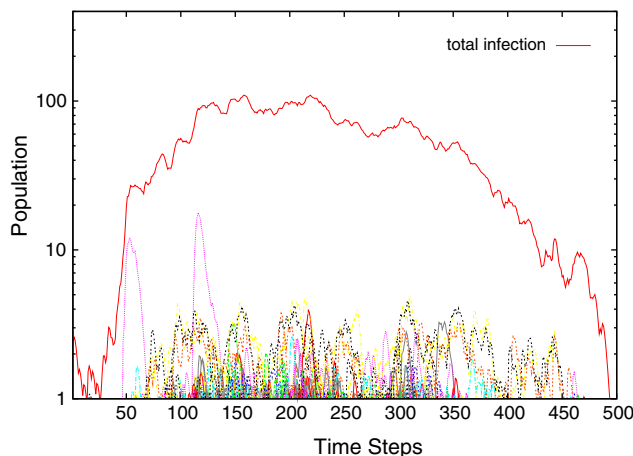


Fig. 7 Observed cumulative epidemic caused by temporally and spatially distributed local outbreaks

gate the spread of disease in each location and the spread of infection from one location to the other. Analyzing the order of local outbreaks may help identify a set of highly probable paths the epidemic may take in the population at large. This will facilitate the identification of high risk groups and prevalence among particular regions, which will aid in the formulation of surveillance and control measures.

5 Summary

The prediction of how an infectious disease outbreak may manifest itself in a given population and whether or not prevalence will rise to epidemic or pandemic proportions has eluded public health experts. Most predictions are based on past records of epidemics that have occurred in different regions with different demographics. Modeling such outbreaks either mathematically or by means of simulation paradigms such as cellular automata and other methodologies have primarily focused on disease spread with respect to the infection or virulence of the disease in homogeneous populations. The GSFS paradigm is the basis for modeling outbreaks of infectious diseases in populations with varied demographic characteristics. That is, GSFS facilitates the analysis of disease progression in heterogeneous environments, and can incorporate geographic and demographic parameters into the model. In order to quantify public health policies in preparation for potential epidemics, predictive tools are required that can facilitate the what-if-analysis of different surveillance and control methodologies. This requires the analysis of how past disease epidemics have manifested themselves in their region with corresponding demographics. These results must then be mapped to predict the disease dynamics in current and future demo-

graphics. This paper demonstrates some of the features of GSFS through an experimental analysis of the incidence and prevalence of an influenza-like illness for parts of Denton County, Texas. From the analysis it is evident that regional demographics are an important aspect when modeling disease outbreak and making decisions about policies to address infectious diseases. GSFS facilitates the identification of high risk groups in the population and adequate points of control, leading to more effective surveillance and control of infectious disease epidemics. Analysis of age as one of the population characteristics has revealed that disease spread patterns changes with different demographics. To this end, GSFS shall prove to be a valuable asset in the analysis of progression of infectious diseases, thereby leading to optimal utilization of public health resources. As the computational demand of simulating epidemics across multiple regions grows significantly, the use of high-performance computing infrastructure must be considered and public health professionals must be prepared to incorporate computational science methodologies into their repertoire of tools.

References

- Abbas K, Mikler A, Ramezani A, Menezes S (2004) Computational epidemiology: Bayesian disease surveillance. In: Proceedings of the international conference on bioinformatics and its applications, Florida
- Ahmed E, Agiza H (1998) On modeling epidemics, including latency, incubation and variable susceptibility. *Physica A* 253:347–352
- Aron J (2000) Mathematical modeling: the dynamics of infection. Aspen Publishers, Gaithersburg, chap 6
- Bagni R, Berchi R, Cariello P (2002) A comparison of simulation models applied to epidemics. *J Artif Soc Social Simul* 5(3)
- Bailey N (1957) The mathematical theory of epidemics. Hafner Publishing Company, New York
- Bailey N (1967) The simulation of stochastic epidemics in two dimensions. In: Proceeding of the 5th Berkeley symposium on mathematics and statistics, vol 4. University of California, Berkeley and Los Angeles
- Barabasi A, Jeong H, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. *Physica A Stat Mech Appl* 311(3–4):590–614
- Barfoot T, D’Eleuterio G (2001) Multiagent coordination by stochastic cellular automata. In: Proceedings of the international joint conference on artificial intelligence
- Barnes J (1954) Class and communities in a Norwegian island parish. *Hum Relat* 7:39–58
- Benenson A (Ed) (1995) Control of communicable diseases manual. American Public Health Association
- Billari F, Prskawetz A (2003) Agent-based computational demography. *Physica-Verlag, Heidelberg*
- BIOWAR (2004) Simulating disease outbreaks using social networks. [Online]. Available: <http://www.casos.ece.cmu.edu/projects/BioWar/biowar.doc>
- Boccaro N, Cheong K (1993) Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. *J Phys A Math Gen* 26(5):3707–3717
- Boccaro N, Cheong K, Oram M (1994) A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behavior. *J Phys A Math Gen* 27:1585–1597
- Callaghan A (2005) Agent-Based Modelling applied to HIV/AIDS. ERCIM News
- Cancre N, Tall A, Rogier C, Faye J, Sarr O, Trape J, Spiegel A, Bois F (2000) Bayesian analysis of an epidemiologic model of *Plasmodium falciparum* Malaria Infection in Ndiop, Senegal. *Am J Epidemiol* 152(8):760–770
- Carabin H, Escalona M, Marshall C, Vivas-Martinez S, Botto C, Joseph L, Basanez M (2003) Prediction of community prevalence of human onchocerciasis in the amazonian onchocerciasis focus: Bayesian approach. *Bull World Health Organ* 81(7):473–550
- The CDC website (2004) (Online). Available: <http://www.cdc.gov/>
- Duncombe W, Robbins M, Wolf DA (2003) Place characteristics and residential location choice among the retirement-age population. *J Gerontol Social Sci* 58(4):S244–52
- Eubank S (2002) Scalable, efficient epidemiological simulation. In: Proceedings of the 17th annual ACM symposium on applied computing (SAC’02), Madrid
- Fayers P, Ashby D, Parmar M (1997) Tutorial in biostatistics: Bayesian data monitoring in clinical trials. *Stat Med* 16:1413–1430
- Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke D (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437(7056):209–214
- Flake G, Lawrence S, Giles C (2000) Efficient identification of web communities. In: Proceedings of the 6th ACM SIGKDD international conference on knowledge discovery and data mining, Boston, pp 150–160
- Fu S (2002) Modelling epidemic spread through cellular automata. Master’s thesis, The University of Western Australia, Perth
- Fu S, Milne G (2003) Epidemic modelling using cellular automata. In: Proceedings of the Australian conference on artificial life
- Glass R, Glass L, Beyeler W, Min H (2006) Targeted social distancing design for pandemic influenza. *Emerg Infect Dis* 12(11)
- Heymann D, Rodier G (2004) Global surveillance, national surveillance, and sars. *Emerg Infect Dis* 10(2)
- Longini IM Jr, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings CA, Halloran ME (2005) Containing pandemic influenza at the source. *Science* 309(5737):1083–1087
- Liljeros F, Edling C, Amaral L, Stanley H, Aberg Y (2001) Social networks: the web of human sexual contacts. *Nature* 6840:907
- Mikler A, Jacob R, Gunupudi V, Patlolla P (2004) Agent-based Simulation Tools in Computational Epidemiology. In: Proceedings of the International Conference on Innovative Internet Community Systems, Guadalajara, Mexico
- Mikler A, Venkatachalam S, Abbas K (2005) Modeling infectious diseases using global stochastic cellular automata. *J Biol Syst* 13(4):421–439
- Oppong J, Mikler A, Moonan P, Weis S (2004) From medical geography to computational epidemiology—dynamics of tuberculosis transmission in enclosed spaces. In: Proceedings of the international conference on innovative internet community systems, Guadalajara
- Rogerson PA, Han D (2002) The effects of migration on the detection of geographic differences in disease risk. *Soc Sci Med* 55(10):1817–1828
- Sattenspiel L, Dietz K (1995) A structured epidemic model incorporating geographic mobility among regions. *Math Biosci* 128(43):1–2
- Situngkir H (2004) Epidemiology through cellular automata, Bandung Fe Institute, Technical Report

- Spiegelhalter D, Myles J, Jones D, Abrams K (1999) An introduction to Bayesian methods in health technology assessment. *BMJ* 319:508–512
- Stefano D, Fuks H, Lawniczak A (2000) Object-oriented implementation of CA/LGCA modeling applied to the spread of epidemics. In: Canadian Conference on electrical and computer engineering. IEEE, Halifax, pp 26–31
- Viboud C, Boëlle P, Pakdaman K, Carrat F, Valleron A, Flahault A (2004) Influenza epidemics in the United States, France, and Australia, 1972–1997. *Emerg Infect Dis* 10
- (2004) The WHO website. [Online]. Available: <http://www.who.org/>