MODELING INFECTIOUS DISEASE SPREAD USING GLOBAL STOCHASTIC FIELD SIMULATION

Sangeeta Venkatachalam, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

August 2006

APPROVED:

Armin Mikler, Major Professor
Roy Jacob, Committee Member
Joseph Oppong, Committee Member
Krishna Kavi, Chair of the Department of Computer Science and Engineering
Oscar Garcia, Dean of the College of Engineering
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies

Venkatachalam, Sangeeta, Modeling Infectious Disease Spread Using Global Stochastic Field Simulation. Master of Science (Computer Science), August 2006, 65 pp., 9 tables, 31 illustrations, bibliography, 49 titles.

Susceptibles-infectives-removals (SIR) and its derivatives are the classic mathematical models for the study of infectious diseases in epidemiology. In order to model and simulate epidemics of an infectious disease, a global stochastic field simulation paradigm (GSFS) is proposed, which incorporates geographic and demographic based interactions. The interaction measure between regions is a function of population density and geographical distance, and has been extended to include demographic and migratory constraints. The progression of diseases using GSFS is analyzed, and similar behavior to the SIR model is exhibited by GSFS, using the geographic information systems (GIS) gravity model for interactions. The limitations of the SIR and similar models of homogeneous population with uniform mixing are addressed by the GSFS model. The GSFS model is oriented to heterogeneous population, and can incorporate interactions based on geography, demography, environment and migration patterns. The progression of diseases can be modeled at higher levels of fidelity using the GSFS model, and facilitates optimal deployment of public health resources for prevention, control and surveillance of infectious diseases.

ACKNOWLEDGEMENTS

This thesis work would not have been possible without the continuous support of my advisor, Dr. Armin Mikler. I take this opportunity to thank him for providing me with useful advice throughout my work on this thesis, and being always available for discussions. I am grateful to him and also to the members of the Computational Epidemiology Research Laboratory (CERL) for discussing the current research topics during weekly research meeting, which helped me in formalizing the thesis research idea. I would also like to thank Dr. Jacob for helping with reviewing the research and the thesis, and giving valuable input. I would also like to thank Dr. Joseph Oppong for providing me with all the help and data that I needed from the geography department. Finally, I would like to thank the GIS group at the Indian Institute of Technology, Mumbai for providing the help and software for GIS data conversions. Among the members of the lab I thank Kaja Abbas and Courtney Corley for being available to discuss any design or implementation issues with me, and Robert Gatti for keeping the machines up and running, and setting special resources that I needed for my experiments. My thanks also go to my husband, Sachin, who showed a tremendous amount of patience and support during the past two years. I sincerely thank him for being there whenever I needed him. Last but not least, I am grateful to my parents for supporting me in whatever I wished to do and for being a source of inspiration.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Globalization and the ever-increasing population diversity accelerates the spread of communicable diseases in the modern society [48, 49]. The World Health Organization (WHO) [46] and the Centers for Disease Control and Prevention (CDC) [18] involve in worldwide surveillance of infectious diseases, and prioritize prevention measures at the root cause of epidemics. As the significance of *public health* is being recognized, the role of epidemiologists has become more prominent. *Epidemiology* deals with the study of cause, spread, and control of diseases. The goal for epidemiologists is to implement mechanisms for surveillance, monitoring, prevention and control of diseases. Epidemiological studies may require large data sets of diseases to analyze and investigate in detail the cause and pattern of spread. Data collected from various sources aare usually are spatially and temporally distributed and contain details of different cases. It is in fact ironic that, for epidemiologists to study the dynamics of different diseases, it is imperative for an outbreak to occur. Epidemiologists have been studying and analyzing disease outbreak data by means of statistical tools. Epidemiology deals with data that are often sparse, widely distributed, incomplete (often due to confidentiality and other constraints), and frequently compromised by conflicting data that confound or disguise the evidence epidemiologists attempt to uncover. It is the ability to draw conclusions and make predictions from this type of information that identifies epidemiology.

In order for the epidemiologists to prepare for a sudden outbreak of an infectious disease or a bio-terror attack, they need tools to assist their planning. Hence, it is imperative to design new models that take advantage of today's cyber infrastructure and facilitate effective

solutions for disease surveillance, control and prevention. Our study focusses on the progression of infectious diseases in environments with diverse demographic and geographic settings. Infectious diseases are varied and specific expertise is needed for every disease.

Computational and mathematical models such as the SIR (susceptible infectives removed) model aid the epidemiologists in the analysis and understanding of the progression of an epidemic in a geographic region with specific demographic characteristics, thereby facilitating the optimal allocation of public health resources. Computational models also enhance the quality of information, accelerate the generation of answers to specific questions and facilitate prediction of disease outbreaks. The goal is to facilitate what-if-analyses that will allow the formulation of public health strategies, such as vaccine distribution in the event of infectious disease epidemics. The dynamics of an epidemic is tightly coupled with the geography and demographics of a region in which an outbreak has manifested itself.Also, this suggests that results that have been obtained by analyzing a disease outbreak in one geographic location may not be readily applicable to define control and prevention strategies in other regions. Although the framework for modeling and simulating infectious disease epidemics will be applicable to a variety of diseases, we will portray Influenza as the disease to justify and highlight the different methodologies. The choice of this diseases is deliberate as it exemplifies a dependence to specific demographic parameters.

The focus of this thesis is a new framework for modeling infectious disease epidemics in a given population. It focuses on the design, implementation and evaluation of global stochastic field simulation (GSFS) to simulate outbreaks of infectious diseases [43], thereby facilitating the optimal or at least adequate allocation of public health resources. We conjecture that the dynamics of an epidemic can be portrayed from simulation of interactions among local hosts. The results show that the model can replicate the results of existing epidemic models.

The remainder of this chapter introduces epidemics and concepts that are important while modeling a disease outbreak. Chapter 2 outlines the review of different mathematical and

computation models in epidemics also with related work and some of the existing approaches. Chapter 3 provides an overview of the Global Stochastic Field Simulation framework and the details of design and functioning of the model. Chapter 4 describes the experiments that were conducted to analyze the model and evaluate the performance of the model. Chapter 5 concludes this thesis with discussion and suggestions for future work.

1.1. Influenza

In an effort to prevent an influenza pandemic as the one witnessed in 1918, which killed an estimated 20 - 40 million people world wide, disease monitoring and syndromic surveillance methods have been deployed. These methods are designed to identify early cases of influenza and guide the allocation of public health resources to control and contain an outbreak. Nevertheless, the dynamics and progression of influenza in a given population remains elusive and cannot be easily derived. This is evident from the fact that estimates of expected influenza cases during a flu season in a particular region vary widely (5% - 40%) [18] . What is however known are medical facts that describe how influenza presents itself in individuals. It is further known that the communicability of influenza, i.e., the length of the infectious period is age dependent. The recent demand for influenza vaccination has resulted in ad-hoc mass-vaccination clinics held by public health departments throughout the nation. The lack of adequate supplies of vaccine has resulted in a selective prevention program specifically vaccinating individuals that were considered at higher risk. This included very young children, elderly, and individuals with compromised immune systems. The notion of risk, however, referred primarily to the medical consequences if the individual does indeed contract the disease. While this policy undoubtedly protects the individuals that have received the vaccine, it is by no means self-evident that this policy is optimal in general. An assessment of vaccination strategies (and other public health policies) necessitates their evaluation with respect to a specific geographic region with its corresponding demographics.

1.2. Epidemic Theory

1.2.1. *Definition of an Epidemic*

A disease outbreak usually occurs when the number of cases of infected individuals is higher than what is normally expected or endemic level of infection. Different disease outbreaks manifest themselves differently as a function of demographics and geographic dynamics of the region. An influenza epidemic, in the United States, is considered as an endemic in many parts of the world [19]. An epidemic when spreading through various continents is known as a *pandemic*. An example of Influenza pandemic is that of 1918 which roughly killed more than 30 million people around the world.

1.2.2. *Infection Life-Cycle*

1.2.3. *Transmission*

Transmission of the virus with respect to disease spread is from outside environment to a susceptible individual's body. Few important ways have been identified about transmission of the virus from outside environment into the human body. The influenza virus is transmitted primarily through direct-contact, which involves skin-to-skin contact and physical transfer of microorganisms to a susceptible host from an infected or colonized person. Direct-contact transmission occurs between two people through any form of physical contact (e.g., by hand contact). Transmission through indirect-contact involves contact of a susceptible host with a contaminated intermediate object, usually inanimate, known as fomites in the environment. Transmission of influenza may occur through either direct skin-to-skin contact or through indirect contact with virus in the environment. However, it is difficult to determine the proportion of influenza transmission that is attributable to direct or indirect contact. [26] Transmission can also occur through droplets, whereby contagious droplets produced by the infected host are propelled a short distance through coughing or sneezing and can be absorbed by a susceptible person. Influenza droplets, generally travel around 3 to 6 feet until they settle

to the ground. This determines which situations may by considered a contact, which is defined as an interaction that may facilitate the transmission of disease.

1.2.4. *Transmission Probability*

Transmission probability refers to the chance that there will be a successful transfer of the pathogen from one host to another during a contact. This probability helps in understanding the dynamics of an epidemic [37]. Transmission probability can be estimated by defining a secondary attack rate.

An attack rate is calculated as a measure of the virulence of an infectious disease, where virulence is defined as the capacity of the virus to cause an infection. Secondary Attack Rate is defined as the ratio of number of infectives to the total number of susceptible individuals in the population. The secondary attack rate is calculated in retrospect with the data collected as thus cannot be predicted. As a result, it cannot be used as a instantaneous measure of epidemic spread. It is averaged over the period of an epidemic. [23]

1.2.5. *Basic Reproductive Number R_0*

Basic reproductive number, R_0, is the average number of susceptibles that are infected by an infective host during its infectious period. This includes only the secondary infections and not tertiary ones. For instance, if R_0 for a disease is 4, then we would expect 4 individuals to get infected by each primary infected individual. If $R_0 = 1$, then the number of infectives remains constant and results in an endemic. For an epidemic to occur, $R_0 > 1$ is needed.

Public health policy and disaster preparedness has often relied on historical data of past epidemic. This is particularly true for the comparison of specific epidemics on the basis of the associated attack-rate (or reproduction number R_0).

Traditionally, R_0 was computed directly from the partial differential equation's that form the susceptible infectives removed (SIR-type) models 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.

However, one must recognize, that the comparison of attack rates of two or more epidemics for the purpose of deciding control measures must take demographic changes into consideration. That is, R_0 of a past outbreak must be adjusted to account for such changes. This however requires to predict how past disease models have manifested themselves in the correct demographics. Consequently, the need for a computational model arises, where in disease spread can be modeled given the demographics and can be used to predict how disease manifests itself in a population, given the demographics of the population.

CHAPTER 2

REVIEW OF COMPUTATIONAL AND MATHEMATICAL MODELS OF EPIDEMICS

Modeling infectious disease epidemics has been going on for several years. Models of epidemic spread exist in many forms, each with its own approach and set of assumptions. However, all these models use a similar parameter set for simulating disease outbreak in a reasonable amount of time.

Few parameters that are significant in determining epidemic spread as described by Mollison [33] include:

- size of susceptible population
- homogeneity or heterogeneity of population
- virulence or transmission probability
- immunity among individuals
- movement of individuals
- infection incubation period, latent period and recovery period

In this chapter computational and mathematical models are reviewed and discussed. Some of the existing epidemics models and reviewed by comparing their techniques and assumptions. Most of the past work uses ordinary and partial differential equations (ODE's and PDE's). Also mean field type approximation and cellular automata (CA) based models are analyzed.

2.1. Mathematical Epidemiology

The early 20^{th} century laid the foundations of the mathematical theory of epidemics [27],[35]. The initial work in epidemic modeling was mainly deterministic and probabilistic elements were not included. Around 1930, statistical distributions were introduced and stochastic modeling established its roots in the study of epidemics.

2.2. Deterministic and Stochastic Models

Deterministic models use partial and ordinary differential equations to model the spread of a disease epidemic, and describe the dynamic behavior without any stochastic terms. These models are likely to be very limited in terms of the parameters they consider, but can greatly overestimate likelihood of adverse effects. When considering a large population, the behavior of a stochastic model is similar to the deterministic model. Stochastic models provide a closer real life approach to modeling the spread of infectious diseases, by introducing probability metrics into the deterministic model. Stochastic models are better when there is inherent variation in the real system. Stochastic models and the analysis of the epidemic curves, have proved to be the primary mathematical framework of simulating infectious diseases outbreaks, even in the 21^{st} century [6].

2.3. Susceptible-Infectious-Removals Model

Mathematical models of infectious diseases are based on the principles of susceptibles, infectives, and removals, namely the SIR 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 by the disease and are infectious. Removals include all individuals that are incapable of transmitting the infection, and are either recovering, fully recovered, expired from the disease, or immune to the disease. Variations of this model have been developed, namely the SEIR, SIS and SEIS. In the susceptible exposed infectives removed (SEIR) model, a state of being exposed is added before becoming infectious. In the Susceptible Infectives Susceptible (SIS) model, the removals who recover may revert to susceptibles. The susceptible exposed infective susceptible (SEIS) model incorporates the exposed state and also the removals may revert to susceptibles. In case of influenza, a recovered individual can not be infected by the same influenza strain due to acquired immunity during the infection. Nevertheless, individuals may remain susceptible to other influenza strains. Figure 2.1 shows the transient curves for

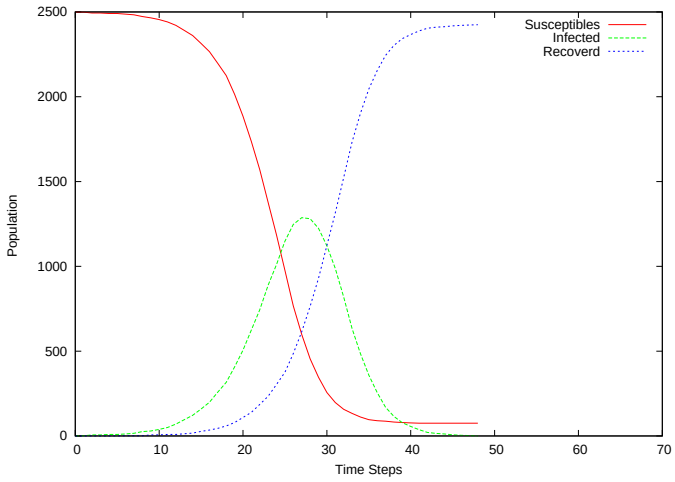


Figure 2.1. SIR Epidemic Curve for a Sample Population

the susceptibles, infectives and removals during the course of a disease epidemic in a given population.

The Kermack-McKendrick threshold theorem [8] is the basis for the SIR model. A continuous influx of susceptibles is a requisite for sustained infection in a population. This is the case of endemic diseases, such as tuberculosis, which prevail in a community at all times. The model is based on the presumption of a population equilibrium, assuming that the epidemic spreads rapidly enough that the changes brought in by births, deaths, migration and demographic changes are negligible [6]. During the start of a disease epidemic, the total population comprises of susceptibles, excluding those that have inherent immunity to the disease. The index case is the first infected individual and is the source of the infection. During the infectious period, the infection is transmitted to some susceptibles, who interact with the index case close enough to contract the infection. This triggers the cycle of infections spreading through the population. Once the infected individuals become non-infectious, they move over to the removals category. A point of interest is that the total number of susceptibles (S), infectives (I), and removals (R) is a constant (Eq. 1). The rising infection on reaching

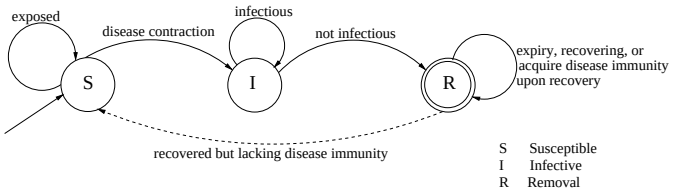


Figure 2.2. SIR/SIRS State Diagram

the peak starts to recede due to the decrease in the number of susceptibles, and diminishes eventually.

(1) $$S + I + R = constant$$

(2) $$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= +\beta SI - \gamma I \\ \frac{dR}{dt} &= +\gamma I \end{aligned}$$

The random mixing of susceptibles and infectives [6] is given by the multiplicative product, $S * I$. β defines the transmission coefficient [5] based on contact rate between susceptibles (S) and infectives (I), and infectivity of the disease. γ defines the rate of infectives (I) becoming non-infectious. Hence, the average duration of infectivity is given by $1/\gamma$ [6]. The set of differential equations used in classic SIR model for a closed population are shown in equation 2. The transfer rates of individuals from $S \rightarrow I$ and $I \rightarrow R$ are given by dS/dt and dR/dt respectively. The rate of change of infectives is given by dI/dt.

The SIR/SIRS state diagram (Fig. 2.2) illustrates the course of a disease in an individual. A susceptible individual may be exposed to a disease pathogen and continue to be in the susceptible state. A susceptible becomes an infective, once the susceptible is able to transmit the pathogen onto others. The recovery state begins once the ability to infect ceases. The

individual continues the state of recovery from the disease, or may expire. On full recovery, the individual may acquire full immunity from disease, and hence is no more susceptible to the disease (SIR model). The individual reverts to a susceptible on full recovery when lacking disease immunity (SIRS model).

The SIR model provides a simple framework for understanding the spread of a disease. However, it cannot be used to model a real epidemic for a specific population and region at sufficient fidelity. The SEIR model is an extension of the SIR model, in which the exposed/latent stage of a disease transmission is considered to account for the time period between the onset of the infection in the body and becoming infectious. The SIR and its related models do not take into consideration the geography or the spatial dimensions of a region. In general, interactions among individual is distance-dependent and it is often more likely to interact with individuals at closer proximity. Consequently, the probability of acquiring an infection from an infectious individual is inversely proportional to the interaction proximity. The spread of a disease is dependent on the levels of interaction in the given population of a specific region. The SIR model considers a uniform population with homogeneous mixing and null consideration of specific interaction measures. Also, it is assumed that the epidemic recedes to an end. The model cannot be used effectively for smaller population sizes. The SIR model can be extended to include geography and demographics, but makes it complicated and unwieldy.

2.4. Cellular Automata Models

Cellular automata modeling paradigm has been used for several decades [24] in the domain of computational and epidemiological models. Infectious disease modeling uses cellular automata to analyze the spatial progression and distribution of diseases [44]. The basic unit of a cellular automaton is a cell and may represent an individual or a sub-population. Each cell can be characterized with state and likelihood risks for exposure and contracting the

disease. The spatial disease progression is modeled via the cell neighborhoods, wherein each infected individual may diffuse and spread the disease to the adjacent neighbors. The naive cellular automata poses certain limitations due to the neighborhood restriction, which has been discussed in the next chapter.

2.5. Agent Based Models

Spatially delineated regions with a small ($\ll 10000$) population can be constructed, using an agent-based approach, in which each individual is represented by an autonomous agent [31]. Larger models with millions of agents necessitate the use of large computing clusters or grid computing that can provide the necessary computational power. The interaction parameters are pre-determined and population 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. The 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. Agent-based models have been used to analyze HIV/AIDS spread in the population and individual immune levels following the infection [17]. A survey of agent-based epidemic simulation models is available [7]. BioWar is an agent-based system that analyzes the disease spread, treatment, and recovery, by porting principles of interactions from social, knowledge and work networks [13]. Mikler et. al. [31] have applied agent-based models to analyze real world outbreaks of tuberculosis in factory and homeless shelter settings.

Stochastic agent based models have been widely used for modeling infectious disease spread such as influenza. Eubank et. al. [22] have developed a system for simulating spread of disease among individuals in large urban population over several weeks. The model relies of empirical data produced by TRANSIMS, a large scale simulation of transport system, to model contact patterns and mixing among individuals.

2.5.1. *Network Models*

Social networks or random networks have long been used in determining the rate and pattern of epidemic spread of infectious diseases. The focus has been particularly on the role of population heterogenities in the spread of sexually transmitted diseases, especially HIV/AIDS and HPV. However, little work has been done in the field of social networks and the spread of other infectious diseases.

2.6. Review of Literature

2.6.1. *Mathematical Models*

Most of the work in modeling infectious disease epidemics is mathematically inspired and based on differential equations and SIR/SEIR model [7, 6]. Differential equation SIR modeling rely on the assumption of closed population and neglect the spatial effects [14, 15]. The population sizes are constant throughout the simulation and do not account for births, deaths or immigration constraints. Population is uniformly distributed over the region. Such models, often fail to consider individual contact/interaction process and assume populations are homogeneously mixed and do not include variable susceptibility.Both partial and ordinary differential equation models are deterministic in nature and neglect stochastic or probabilistic behavior [40]. Nevertheless, according to Di Stefano [40] these approaches/models have been shown to be effective in regions of large populations.

Boccara and Cheong [14] study the SIS model for spread of infectious diseases in a population of mobile individuals, thereby introducing non-uniform population density. In the model susceptible individuals become infective with a probability of π if, and only if, it is the neighborhood of and infective. This hypothesis however neglects the incubation and latent period which have been identifies as factors influencing epidemic spread. After being infectious the infectives move from infected to susceptible with a certain probability at each time step. Length of time a individual stays infected is assumed to be 1 and is not variable.

Boccara and Cheong [15] concentrate on SIR epidemic models and take into consideration the fluctuation in the population by births and deaths, exhibiting a cyclic behavior with primary emphasis on moving individuals. This level of complexity has been incorporated by stochastic models, based on the same population categories as deterministic models, but relying upon probabilities variables and their relationship to relative uncertainties between the compartments. These models are better suited for larger communities because of use of probabilistic expressions.

2.6.2. *Cellular Automata Models*

The earliest example of use of cellular automata is Bailey's lattice model [9] for the spread of diseases from micro-level interactions. Schönfisch has analyzed varied cellular automata models to study the dynamics of epidemics [36].

Di Stefano *et al* [40] have developed a lattice gas cellular automata model to analyze the spread of epidemics of infectious diseases. The model is based on individuals who can change their state independent of others and can move from one cell to other. However, this approach does not consider the critical factor of the infection time-line.

Ahmed *et al* [4] study the susceptible-infected-susceptible (SIS) problem where susceptibility of the population is considered first as uniform then as random. Variations in population density is modeled by allowing cyclic host movement. The study claims that other models which include random motion, whereas realistic motion is usually cyclic e.g, home to work to home. In this model a susceptible site can be infected with the probability b and the number of occupied nearest neighbors, which may lead to saturation of neighbors. Incubation period of the infection is not considered as individuals become infected in just one time step.

Fu has used stochastic cellular automata to model epidemic outbreaks that take into account the heterogeneous spatiality [23]. The model takes into account some geographic and demographic factors such as neighborhood radius, natural birth, immigration and natural

death parameters. A morbidity parameter has been implemented that defines the probability that an infective will die from disease during a particular time step. However, the population distribution in the beginning of the simulation is uniformly distributed and there is no variance.

Situngkir has developed a dynamic model of spatial epidemiology to study avian influenza disease in Indonesia and uses cellular automata for computing analysis [38]. The CA model uses the von Neumann neighborhood system for terrestrial interaction and spread of infection, which is associated with a probability.

Bonabeau has studied the spatio-temporal characteristics of influenza outbreaks in France. The study infers that the global transportation systems of the modern world lend to propagation of influenza epidemics dominated by a global mixing process in comparison to local dynamic heterogenities [16].

Duryea has analyzed spatially detailed epidemic models using probabilistic cellular automata for heterogeneous population densities in a region [20]. The results suggest that increasing spatial heterogeneity in host density decreases the frequency of infection at endemic equilibrium. Spatial structure is given consideration assuming that individuals interact more frequently with nearby than with distant organisms. The iterations in the model are both local and global.

Benyoussef has used a one-dimensional lattice model and a two-dimensional automata network model to illustrate the spatial spread of rabies among foxes [12].

Fuks describes a SIR epidemic in the spatio-temporal domain via a lattice gas cellular automaton for both human and animal populations. Vaccination strategies are incorporated and dynamics of the disease spread are investigated in relation to the spatial distribution of the vaccinated individuals [25].

Mansilla and Gutierrez [30] have a developed a Deterministic site exchange cellular automata models for the spread of epidemics, rumors and news in a population of moving individuals. The model depends on parameters, which represent the mean length of motion

of individuals in population. They claim to have reproduced situations of perfect mixing and perfect diffusion which are often described by system of ordinary and partial differential equations. They assume that motion of individuals is periodic motion rather than random as assumed by most other similar models.

2.6.3. *MFT approximations*

Disease epidemics have been modeled using mean field type (MFT) approximations [29]. Most MFT models assume that susceptible population is uniform over the world, which makes it similar to differential equation models. Even though the MFT models are similar to the differential equations, they add a probabilistic nature by adding different probabilities for the mixing among individuals [14]. In the MFT models the decision to move around is independent among individuals, unlike the differential equation models, where all or none phenomena applies. According to Boccara [14], mean field approximations tend to neglect spatial dependencies and correlations and assume that the probability of the state of a cell being susceptible or infective is proportional to the density of the corresponding population.

MFT approximations sometimes use a lattice structure to simulate the spatial nature of the disease spread, where in an individual present at a site can exist in one of the states as specified by the model. Kleczkowski and Grenfell [29] describe that MFT and CA models converge when the MFT mixing parameters tend to infinity. This leads to say that the world contains more disorder than correlation.

Bayesian analysis of epidemiological data highlights the significance of analyzing demographics to uncover the higher risk spectrums of the population for infectious diseases [1]. A Monte Carlo simulation using a Markov model is implemented to study the infection models that occur naturally, such as influenza, whose viral pathogen spreads through a susceptible community, or induced deliberately, as in the case of bio-terror attacks [34].

Geographic-environmental re-infection modeling simulator (GERMS) [2] is a toolkit for modeling transmission of infectious diseases. The model takes into consideration heterogeneous population dynamics with varied socio-geographic characteristics, complex interactions among individuals, and infection specific features, such as transmission probabilities.

Viboud et.al [45] have analyzed spatial and temporal correlation of influenza epidemics in the United States, France, and Australia. The results indicate a high correlation between United States and France, but irregularity in the patterns between Australia and the other two countries. They highlight demography as one of the reasons for causing the discrepancies, and recommend further investigation using mathematical modeling.

CHAPTER 3

OVERVIEW OF GLOBAL STOCHASTIC FIELD SIMULATION

3.1. Cellular Automata

Cellular automata have been used for several decades [24] in the domain of computational models. Nevertheless, in modeling epidemics, this paradigm has rarely been utilized to its full potential [3, 39, 24, 40]. *Cellular automata*, as defined by Lyman Hurd, is a discrete dynamic system, where space, time, and the states of the system are distinct [47]. An automaton is best exemplified by representing a point in space as a cell C_i surrounded by other cells, thereby defining the neighborhood H_i of C_i. The cells are most often arranged to constitute a regular spatial lattice (see Fig. 3.1).

In general, we can define a cellular automaton of any dimension. One, two, and three dimensional automata are most often used in science. For a one dimensional automaton, $|H_i| = 2$, that is, cell C_i has a left and a right neighbor (ignoring edge conditions). A two dimensional automaton is best represented as a regular spatial lattice or grid. Here, cell $C_{i,j}$ is surrounded by cells that form its neighborhood $H_{i,j}$. Traditionally, there are two possible sizes of $C_{i,j}$'s neighborhood in a two dimensional automaton, namely, $|H_{i,j}| = 4$ in the *von Neumann neighborhood* and $|H_{i,j}| = 8$ in the *Moore neighborhood* [47] (see Fig. 3.1 Used with permission from World Scientific Publishing Company [32]). Table 3.1 (Used with permission from World Scientific Publishing Company [32]) specifies the neighboring cells for $C_{i,j}$ in both the neighborhoods.

At a particular time t, each cell C of the automaton is said to be in a specific state s(t), which depends on the specific application. $s(t) \in S$ where S is the state space of the cellular automaton. In a simple scenario, cells are assuming binary states $\langle 0,1 \rangle$. For more complex

Table 3.1. Neighborhood Specification

Neighborhood	Neighboring Cells for $C_{i,j}$
von Neumann	$C_{i+1,j}$, $C_{i-1,j}$, $C_{i,j+1}$, $C_{i,j-1}$
Moore	$C_{i+1,j}$, $C_{i-1,j}$, $C_{i,j+1}$, $C_{i,j-1}$, $C_{i+1,j+1}$, $C_{i-1,j-1}$, $C_{i-1,j+1}$, $C_{i+1,j-1}$

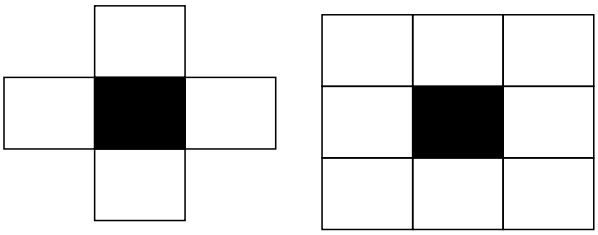


Figure 3.1. von Neumann and Moore Neighborhood

applications, any size of discrete (and even continuous) state space can be defined. The state of cell $C_{i,j}$ at time t is determined by the state of its neighborhood $H_{i,j}$ at time t-1 (see eq. 3). The function f can be considered as the rule that dictates how a particular state configuration of $H_{i,j}$ determines the next state of $C_{i,j}$. For a deterministic cellular automaton, the initial states of each cell and the update rule f completely describes the automaton. During a time step t, a new state s(t) is computed for every cell as described above. An initial state configuration will hence evolve, thus representing a dynamic system.

$$s_{i,j}(t) = f(H_{i,j}(t-1)) \tag{3}$$

An example of a cellular automata update rule is shown in Fig. 3.2. Here, the function f is defined by a majority rule. The state of the center cell transitions to a state, which is in majority among the cells in the neighborhood and itself. The update rule determines the deterministic or stochastic behavior of CA. Stochastic behavior is seen by probabilistic update rules in non-deterministic state transitions. For example, in stochastic CA, for every update,

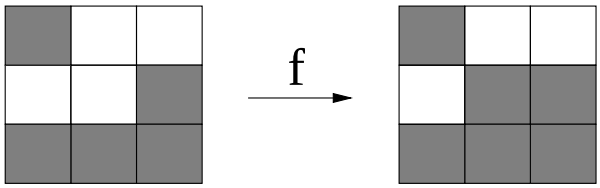


Figure 3.2. Cellular Automata Update from time step *t-1* to *t*

a cell can choose probabilistically from a set of update rules, or for a particular update rule, probabilistically choose from a set of states for the stochastic transition.

3.2. Disease Modeling with Cellular Automata

Modeling using traditional cellular automata paradigm incorporates the spatial distribution of the population using the Moore or Neumann neighborhood. The basic unit of cellular automata is a cell. In a disease model , a cell represents an individual or a sub-population. Each cell can be characterized with state and likelihood risks for exposure and contracting the disease. Unlike the SIR model, every cell comes in contact with the cells in its defined neighborhood. Similar to the SIR model, state *S* for susceptible is defined as the state in which the cell is capable of contracting a disease from its neighbors. In the infectious state *I*, the cell is capable of transmitting the infection to its neighbors. In the recovery state *R*, the cell is neither capable of passing on the infection nor capable of contracting the infection. On full recovery and acquisition of disease immunity, the cell shall continue in the removal state (R). The time-line for infection is illustrated in Fig. 3.4 Used with permission from World Scientific Publishing Company [32]).

3.3. Definitions

The following section describes the different states a cell can attain and parameters used in simulation. **States of an individual**

State *S* for susceptible is defined as the state in which, the individual is capable of contracting a disease from its neighbors.

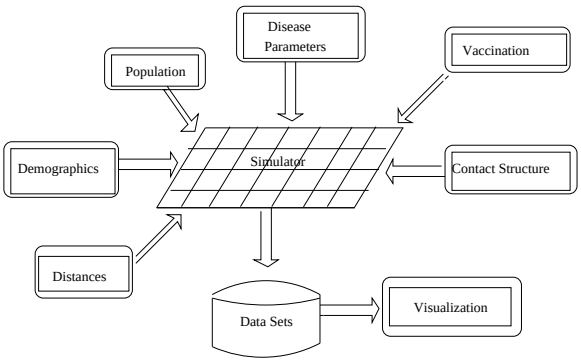


Figure 3.3. The GSFS Model

In the infectious state, I the individual is capable of passing on the infection to its neighbors.

In the recovery state, R the individual is neither capable of passing on the infection nor capable of contracting the infection.

Immunized individual is modeled with less susceptibility to acuire the disease.

Parameters for the simulator

Infectivity ψ of a disease is defined as the probability of a susceptible individual becoming infectious, when coming in contact with an neighboring infectious individual.

Latency λ is defined as the time period between exposure to an infectious organism and appearance of the disease.

Infectious period θ is the time period during which the infected individual is capable of transmitting the disease to other individuals.

Recovery period ρ is defined as the time period the individual takes to recover, wherein it is neither capable of transmission of the infection nor capable of contracting the infection.

The region over which the disease is to be simulated is given as a parameter together with population dynamics and age structure.

Individuals from the population may be immunized for a particular disease being simulated. The immunization is represented as a probability which as a results reduces the susceptibility

Table 3.2. Parameters

Parameter	Value
Incubation period	3 days
Latent period	3 days
Infectious period	4 days
Recovery period	6
Infectivity	0.035
Immunity	0.25
Population	Denton city
Contact Rate	8-14

of the individual. The immunization can be used to simulate various vaccination strategies or immunity from previous infections.

Contact rate CR is defined as the average number of contacts an individual is involved in during a 24 hour day. A contact is defined as a two way conversation without any physical barrier between the two parties, which may lead to a successful disease transmission. Contact rate is determined for individuals based on their demographics, which may result in different contact rates for individuals of different age groups.

The earliest documented case of a disease is known as the index case. Index cases are introduced in the beginning of the simulation at the initial state. These are the cases primarily responsible for starting the epidemic in the simulation.

Table 4.1 states the approximate parameter values used for a base experiment specific to influenza as the disease.

Rules for disease spread The rules described below determine the state transitions of individuals in the population for the SEIR and SEIRS models.

(i) An individual changes the state from *susceptible* to *latent* ($S \rightarrow L$) after coming in contact with an infectious individual in the defined neighborhood. The probability of acquiring the disease from an infectious neighbor is a function of infectivity ψ.

The individual remains in the latent state for the number of time units (days) as defined by the parameter latency λ.

(ii) The state of the individual changes from *latent* to *infectious* ($L \rightarrow I$) after being in state L for a given λ. In our model, we assume that every individual exposed to the pathogen will become infectious. In state I, the individuals are capable of passing on the infection to neighborhood individuals. For example for a disease D, with λ= 2 units the individual will enter the infectious state I after two time units of initial exposure.

(iii) After the infectious period θ, the individual changes the state from *infectious* to *recovered* or *removed* ($I \rightarrow R$). Once an individual enters the state R, they are no more capable of passing on the infection.

(iv) From the state R, the individual's state changes back to either susceptible S for the SEIRS model or it remains in state R, for the SEIR signifying complete immunity.

[1]

3.4. Restrictions of Classic Cellular Automata

3.4.1. *Limitations of Naive Cellular Automata:Neighborhood Saturation*

Figure 3.5 (Used with permission from World Scientific Publishing Company [32]) depicts the cell layers with respect to a central cell in $layer_1$. $Layer_1$ has 8 neighboring cells in its outer-line $layer_2$ in a Moore neighborhood model. The outer-line neighborhood of $layer_i$ is

[1]The next three sections are reproduced from: A. Mikler, S. Venkatachalam and K. Abbas, Modeling Infectious Diseases Using Global Stochastic Cellular Automata, Journal of Biological Systems, Vol. 13, No. 4, pp. 421-439, December 2005. with permission from World Scientific Publishing Co. Pte. Ltd, Singapore.

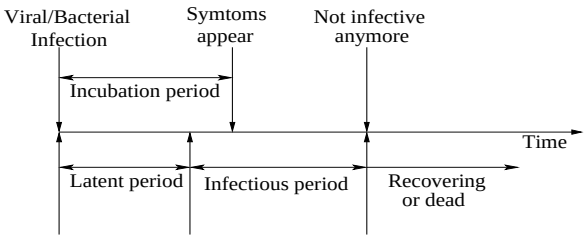


Figure 3.4. Infection Time-line

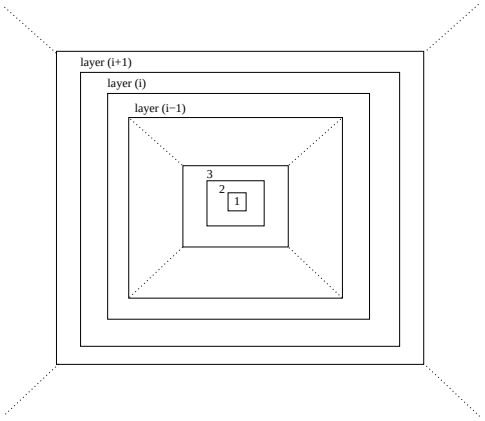


Figure 3.5. Cell Layers

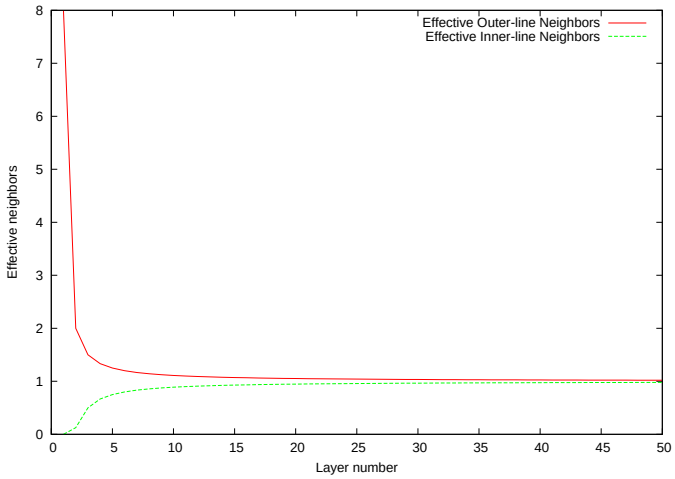


Figure 3.6. Effective Neighborhood

$layer_{i+1}$ and the inner-line neighborhood is $layer_{i-1}$. The total neighbors of a layer is defined by a summation of its outer-line and inner-line neighborhoods. The ratio of neighboring cells to the cells in the current layer defines the *effective neighbors* per cell of the current layer.

L_i is the number of cells in $layer_i$ and is defined in eq. 5. It can be visualized as the area enclosed by layer L_{i-1} subtracted from the area enclosed by layer L_i (see eq. 5). The effective outer-line neighbors of $layer_i$ are defined by L_{i+1}/L_i and the inner-line neighbors are L_{i-1}/L_i. Figure 3.6 illustrates the effective inner-line and outer-line neighbors from $layer_1$ up to $layer_{50}$. Even though the effective outer-line neighbors of $layer_1$ is 8, it converges to 1 for higher layers. The effective inner-line neighbors increase from 0 for $layer_1$ to 1 for higher layers.

$$
\begin{aligned}
L_i &= 1 && i = 1 \\
&= (2 * i - 1)^2 - (2 * i - 3)^2 && i > 1 \\
L_{i+1}/L_i &\rightarrow 1 && i \rightarrow \infty
\end{aligned}
\tag{4}
$$

In the context of epidemiology, we consider a disease progressing at 100% infectivity through neighboring layers. An index case at the central cell in $layer_1$ shall effectively infect 8 outer-line neighbors at $layer_2$. However, at higher layers, each cell at $layer_i$ is able to infect effectively only one outer-line cell at $layer_{i+1}$. This resulting *neighborhood saturation* is a primary limitation of naive cellular automata in depicting the spatial progression of a disease. The classic cellular automata methodology suffers from saturation of a limited neighborhood, as described above. This has been described in [32]. A neighborhood of 8 cells quickly saturates and thus reduces the number of susceptibles. In such a situation the increase of infectivity parameter plays no role and has the same effect on the spread of the disease. Neighborhood saturation dominates the effects of increasing infectivity and limits the spread of the disease. Further, the need to model a disease where an infective can spread the disease to an extended neighborhood in one time step can not be modeled. The movement of individuals, migration, or travel is not considered. Some of the models discussed

in the literature, deal with movement of individuals from one cell to another in the defined neighborhood. Clearly as discussed above they are deemed to be hampered by early saturation. In order to overcome the limitations posed by naive cellular automata, we introduce the global stochastic model for cellular automata, that shall incorporate the demographics of location and population density.

3.5. Towards a Global Model : Accounting for Spatial distributions

The traditional CA model described above may be used for simulating diseases over small regions with local interaction and global interaction respectively. As mentioned before, the model does not take into account the demographics of the region and may not be accurate for simulating disease spread over large geographic regions because of the neighborhood constriction posed by them. Hence, we can now look at a global stochastic cellular automata with demographics that will facilitate the understanding of effects of different demographics, the population density, socio-economics of a region and culture. It can also be used effectively for investigating different vaccination strategies and understanding the effects of travel.

3.5.1. *Interactions*

For simulating the spread of disease in such an environment, contacts need to established between cells. In this model, every cell has a chance of interacting with every other cell in the environment, but the probability of contact varies based on what is defined to be the interaction coefficient. The interaction coefficient reflects the factors which are important when considering contact between two cells. Such as distance, population and other demographics and socio-economic factors. The interaction coefficient is based on the distance between cells.

The **neighborhood** for a global SCA is defined using a fuzzy set:

Definition 1 : Fuzzy Neighborhood

The set $F \subset S$ where S is a set of all the cells

$F: \{\langle s, p\rangle | s \in S, 0 \leq p \leq 1\}$

$\langle$ s,1 $\rangle$: Total/Complete membership

$\langle$ s,0 $\rangle$: No membership

Interaction Coefficient i for a particular cell is defined as the strength or likelihood of interaction between two cells. As mentioned above the interaction coefficient depends on various factors which can affect contact between two subjects, such as individuals or cells. However, for this model distance is considered between cells as the factor influencing the interaction coefficient. It is calculated as the reciprocal of the Euclidean distance between the cells. Equation 5 shows the calculation for interaction coefficient based on distance. Experiments were conducted on calculating the coefficient based on distance and population. Equation 6 shows the calculation of interaction coefficient based on distance and population.

(5) $$i_{C_{i,j},C_{k,l}} = \frac{1}{\sqrt{\langle i-k\rangle^2 + \langle j-l\rangle^2}}$$

(6) $$i_{C_{i,j},C_{k,l}} = \frac{P_{C_{i,j}} \times P_{C_{k,l}}}{\sqrt{\langle i-k\rangle^2 + \langle j-l\rangle^2}}$$

The state of infection δ is defined for every cell as a number between 0 and 1, indicating the level of infection present in the cell. 0 indicates not infected, 1 indicates fully infected. This parameter is used in order to determine whether the subject or group is capable of transmitting the infection or not.

The global interaction coefficient Γ of cell $C_{i,j}$ is the sum of all the individual (n-1) interaction coefficients of the cell. This coefficient represents the overall interaction of the particular cell. It is different for every cell based on their position. Figure 3.7 shows the global interaction coefficient based on distance for every cell on a 50 $\times$ 50 grid. Naturally, the center cell has the maximum interaction coefficient. Figure 3.8 show the global interaction

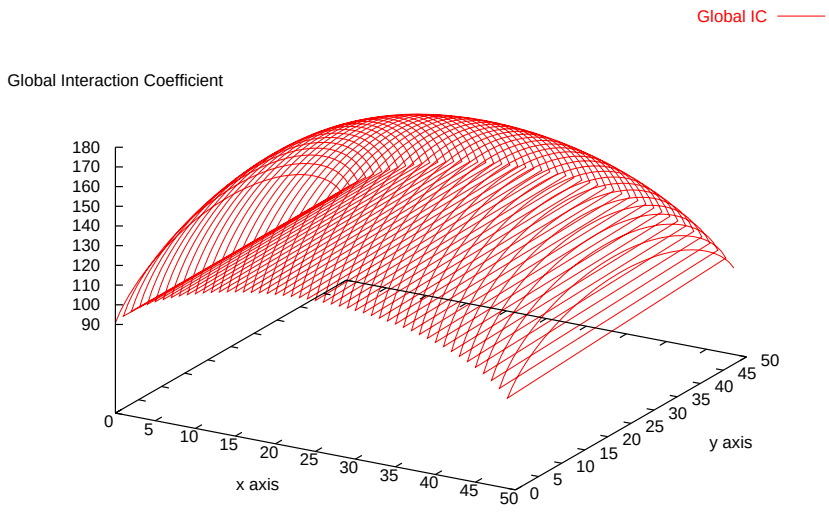


Figure 3.7. Global Interaction Coefficient Based on Distance

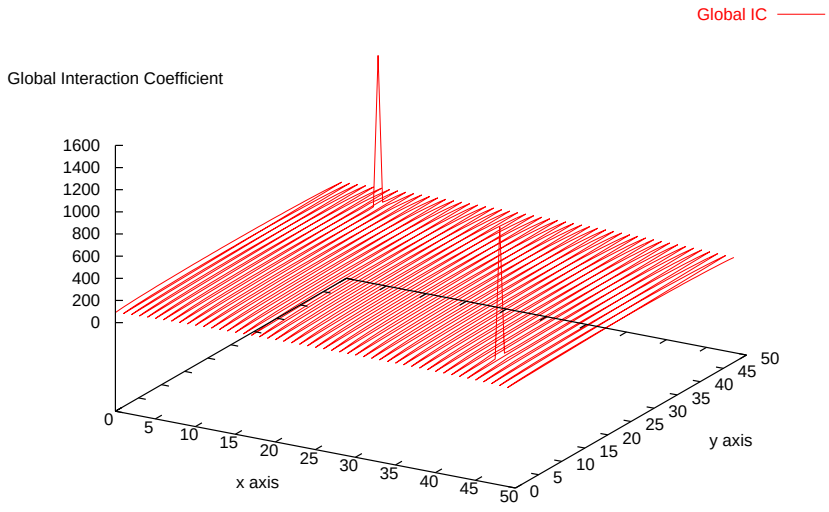


Figure 3.8. Global Interaction Coefficient Based on Distance and Population

coefficient based on distance and population for every cell on a 50 × 50 grid. Experiment was conducted with two cities of high population. In Figure 3.8, population size clearly dominates distance. This however may not hold true if the interaction coefficient incorporates measures of population and other demographic values.

The global interaction coefficient and the interaction coefficients are calculated based on the distance. As the distance between the cells reduces, the interaction coefficients increases, which indicates greater chances of interaction between them. Equations 7 and 8 show the calculation of the Γ based on distance and distance and population respectively.

$$\Gamma_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{1}{\sqrt{\langle i-k \rangle^2 + \langle j-l \rangle^2}} \tag{7}$$

$$\Gamma_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{1}{\sqrt{\langle i-k \rangle^2 + \langle j-l \rangle^2}} \times P_{C_{i,j}} \times P_{C_{i,j}} \tag{8}$$

The **infection factor I** is calculated as a fraction of the interaction coefficient to the global interaction coefficient Γ, for every cell to cell interaction. It is also based on the virulence of the disease and the state of infection of the infecting agent. Referring to definition 1 the parameter p is the ratio of interaction coefficient to the global interaction coefficient.

$$I_{C_{i,j}} = \sum_{\forall C_{k,l} \neq C_{i,j}} \frac{i_{C_{i,j},C_{k,l}}}{\Gamma_{C_{i,j}}} \times \delta_{C_{k,l}} \times \psi \tag{9}$$

3.6. Modeling with Global Stochastic Cellular Automata (GSCA)

In the GSCA model, the population is considered to be uniformly distributed over the gird. Each cell is considered as an individual, with certain structural properties. As defined earlier, contact rate for each cell is drawn from a Poisson distribution assuming that average contacts that individuals are involved in is Poisson distributed. To model the interactions between cells, we present two different algorithms for selection of contacts.

As mentioned earlier, the probability of interacting with cells closer is greater than the probability of interacting with cells farther. Considering a $n \times n$ grid with n^2 cells, each cell has $n-1$ possible interactions. The simulator selects k interactions, where k is the average

contacts per day, from the possible $n-1$ as controlled by the exponential decay function, which yields an inefficient computational solution. When n is considerably large as to 10^3 cells then the amount of data can approach terabyte range. Thus to solve the n^2 bottleneck problem we propose the following algorithms. We assume that the probability decays exponentially with distance. See Equation 10. The parameter α is a multiplicative factor used to scale the actual geographical distance parameter to match the distances on the grid.

$$P = \alpha \times e^{-\alpha d} \tag{10}$$

In order to calculate the exponential random variant, we draw a random number R $\in$ U[0,1] and transform equation 10 to obtain d. Once the distance is found two different ways are used to choose contact cells.

3.6.1. *Threshold Based Algorithm*

For the threshold method, cells are chosen at random and a rejection method is used. The cells with at a distance less than the threshold (calculated) distance are chosen, otherwise rejected. The drawback of the method is the number of rejections that might occur, until a cell at a distance less than the threshold is chosen. If a new distance is calculated after every rejection then the number of calls to the random number generator increases drastically, which may hamper the performance of the system. Figure 3.9 demonstrates that a large portion of cells chosen are at smaller distances. However, for this algorithm the number of calls to the random number generator is large, because of uniformly selecting from n^2-1 cells until a cell is found at a distance less than d.

The following pseudo code describes the threshold based algorithm for a cell x

```
1 choose distance d from the exponential decay function
```

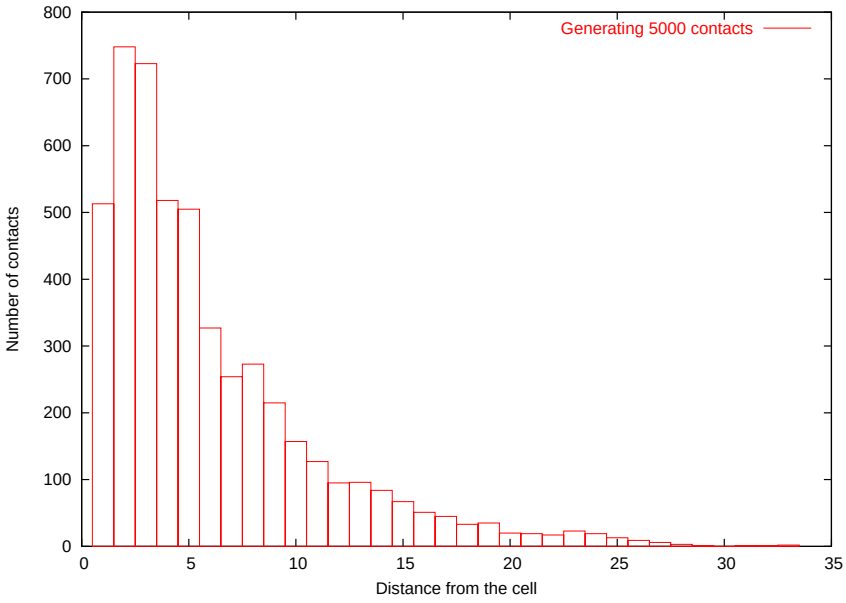


Figure 3.9. Frequency of Distances Based on Threshold

```
2 do
3   choose a cell y at random uniformly
4   calculate the distance of y from x
5 repeat until cell distance leq d
6 establish contact with y
7 repeat above steps until expected number of contacts are established.
```

Clearly, this method is based on trial and error. Consequently we propose the bounding box algorithm, where in instead of choosing from n^2-1 cells the choice is restricted by the bounding box, thus reducing the number of calls to random number generator. The bounding box based algorithm is described in the next section.

3.6.2. *The Bounding Box Algorithm*

The bounding box algorithm is proposed to reduce the overhead of calls to random number generator. The idea of this algorithm is reduce the sample space so that calls to the random number generator are limited and time complexity reduces. In this algorithm, with the cell as the center a virtual bounding box is drawn around the cell at a distance d, to signify the local

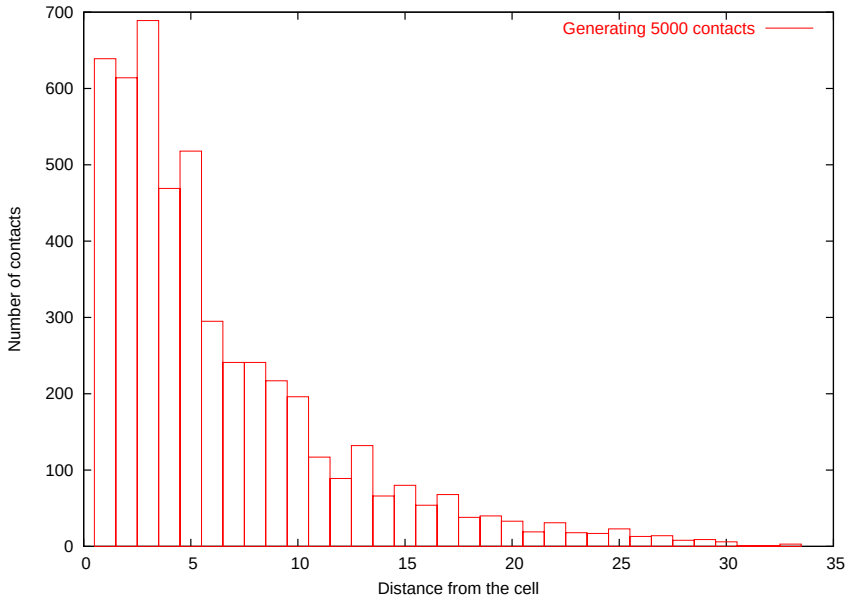


Figure 3.10. Frequency of Distances Based on Bounding Box

neighborhood of the cell. Cells within this neighborhood are chosen randomly for contact. For every new contact a new d is calculated.

The following pseudo code describes the algorithm for a cell x

```
1  choose distance d from the exponential decay function
2  draw a bounding box of distance d from x on all four sides
3  calculate the boundaries of the box
4  choose a cell y at random within the box boundaries
5  establish contact with y
6  repeat above steps until expected number of contacts are established.
```

Experiments were conducted by generating a relatively large set of contacts to obtain and test distribution of the contacts based on distance. Figure 3.10 shows the number of contacts generated for different intervals of distances, indicating that more contacts are generated with cells closer to the subject cell.

The two algorithms described above were used to implement contacts based on distance, assuming more contacts are local than global. To evaluate the effect of distance based interactions, experiments were conducted in the GSCA model. When distance based interactions are incorporated, the contacts between individuals are initiated assuming more contacts are made locally than globally. Figure 3.11 depicts the results that illustrate that the rate of disease progression is relatively slower in the global model, when the distance demographic parameter is incorporated. The disease progression slows down as a result of more local contacts being generated and interacting with same individuals repeatedly. Thus, it can be concluded that distance place an important role in modeling interactions.

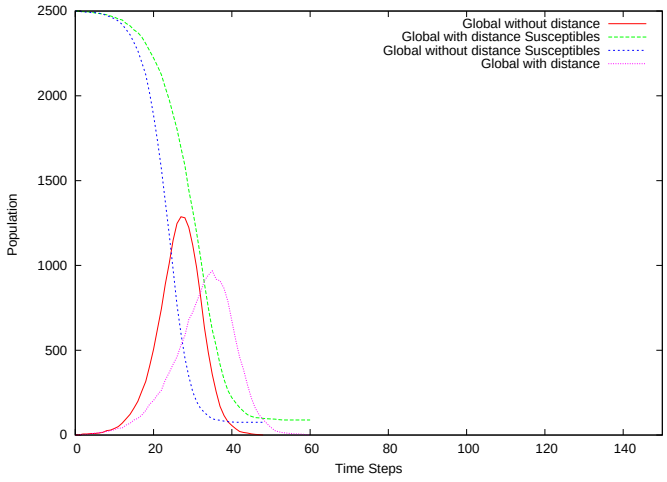


Figure 3.11. Disease Progression with and without Distance Demographic Parameter

3.7. Global Stochastic Field Simulation

Disease modeling over small regions with local interactions can be implemented using traditional cellular automata. However, its accuracy diminishes for simulating disease spread over large geographic regions because of neighborhood saturation. We propose global stochastic field simulation (GSFS) that includes demographic parameters of a given geographic

region [44]. This facilitates understanding of the effects of different demographics, the population density, socio-economics and culture of a region. It can also be used effectively for investigating different vaccination strategies and understanding the effects of travel.

3.7.1. *Approach*

Our approach to modeling spread of diseases is based on a geographic region represented as a grid. 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. Individuals belonging to specific locations can be characterized by a state and likelihood of risks for exposure and contracting the disease. A set of three possible states (S, I, R) has been defined to signify an individuals clinic disease stage. As opposed to a purely agent-based model, in a stochastic field simulator, each location maintains the statistics of the three states. These can be used to calculate the attack rate, relative risk of sub-populations. Disease spread is driven by contacts generated based on population statistics, unlike agent-based models where, individuals (agents) themselves indulge in contacts.

3.7.2. *Modeling Heterogeneous Populations*

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 means to derive the population dynamics that promotes the spread of diseases. This can be accomplished by exploiting 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,

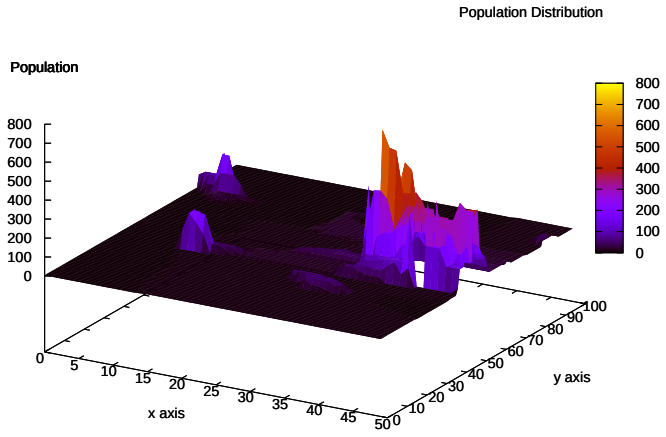


Figure 3.12. Heterogeneous Population Distribution

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. Age-structure of the population has been incorporated into the model as one of the demographic constraints. For simplicity in modeling behavior patterns among individuals the age-structure has been divided into four groups of under 9 years, 10 to 34 years, 35 to 59 years and 60 years and over.

The global stochastic field simulation (GSFS) model is implemented to incorporate heterogeneous populations. Rasterized GIS census block data of the area around city of Denton, Texas for the total population of 110000 is overlaid on a grid of size 50 * 98. The following section describes how census data is rasterized and ported into the simulator.

3.7.2.1. *Rasterized Input data.* Denton county regional census population data was imported in vector format to GRAM++ GIS package [42]. This vector file is then converted into a raster file with each block representing a unique location with specific demographic.

The following describes the process of converting the census information to simulator input format. This method distributes the population of census blocks into locations in the simulation field.

Let P_i be the population of ith census block.

Let N_i be the number of cells in that census block

Let C_{ij} be the population of the jth cell of the ith census block where

$C_{ij} = P_i/N_i$

where j= 1 to n and n is the number of cells in each census block

i = 1 to m and m is the number of census blocks in the county

Hence the population of each census block is assumed to be uniformly distributed among all the cells in that block Figure 3.12 shows the heterogeneous population distribution of area around Denton city.

3.7.3. *Demographics :Age-structure*

GSFS model has been extended to incorporate age-structure of the region as one of the demographic constraints. Together with the population age-structure data is obtained from census block data of area around city of Denton, Texas. To simplify in modeling behavior patterns among individuals, the age-structure has been divided into four groups of under 9years, 10 to 34 years, 35 to 59 years and 60 years and over. Each age group may be assigned different values for disease parameters such as infectious, latent and recovery periods based on how an individual of that age group might react to the disease in hand. They may also be assigned different contact rates to model different interaction patterns.

3.7.4. *Mixing Patterns*

As we have seen, essentially all mathematical models of spread of airborne infections were based on the assumption that the host population was homogeneously mixed. Therefore, each individual has an equal chance of contacting anyone in the population, and being

Table 3.3. Age Group Distribution

Age group	Range
1	0 - 9 years
2	10 - 34 years
3	35 - 59 years
4	60 + years

contacted by someone irrespective of whether or not that has person has been contacted before. Quantitative studies in social networks research has characterized relevant network structures, and has shown how specific aspects of contact patterns can alter the spread on infectious diseases and the evolution of infective agents [28]. For successful transmission of the virus, there is a need for a close contact between the susceptible host and the infected host. Consequently, mixing patterns and contact structure of the host population are of paramount importance, when dealing with spread of airborne infectious diseases. Studies [28] have shown that different structured networks lead to different types of epidemics. For instance, a densely connected host population is easier to invade than a sparsely connected host population.

The social contacts can vary in many different ways, such as the number of contacts per day, the context of a contact, the distance between the people in contact, duration of contact and age of contacts. Among these mixing patterns, the variation in mixing within and between age groups and between households and weekday to weekend variations is a point of interest. Edmunds et. al. [21] conducted a study to identify such mixing patterns which might lead to spread of infectious airborne diseases. A sample of 92 adults from an English university were asked to detail the individuals with whom they had a conversation over the period of on, randomly assigned, day. The data contained the age of the person, the age of the contact and the social context in which the contacts took place. Edmunds

et. al. [21] define a contact as a two way conversation in which at least wo words were spoken by each party and in which there was no physical barrier between the two parties. Statistical analysis was conducted on the collected data. The results showed that contacts were highly structured according to age. The average age of contacts increased with the age of the participant. Older individuals having more contacts with older adults, and considerably less contacts with younger individuals. Also, older individuals showed a tendency to mix with older children. Studies [21] have also shown that individuals are likely to interact more with individuals of the same age group as compared to individuals of other age groups. The statistical analysis of the data resulted in a mean number of contacts in a day of 16.8. There was no significant difference in the mean number of contacts between the different weekdays, irrespective of the age. Older adults had significantly lower contacts on weekends as compared to weekdays, whereas younger adults did not show any significant variation. Since we do not model weekdays and weekends, we averaged the contact rate for adults considering weekday and weekend contact rates. Regarding the context of the contacts, 12-13 contacts were work related, 2-4 social and 1-2 home contacts. Thus for individuals above the age of 80 who might have retired will not have the work related contacts. Similarly children under the age of 10 will also not have work/school related contacts. Thus contact structure is an essential aspect in explaining the level or intensity of invasion of a virus in a host population. In the GSFS model, a contact is modeled as any interaction which may lead to a successful disease transmission and is different from a regular social contact.

3.7.5. *Contact Structure*

GSFS models heterogeneous mixing in the population by establishing contacts in a random fashion. As defined before, a contact is any interaction between two individuals which can lead to a successful disease transmission. To generate contacts proportionally to the demographics we build probability distributions for the different age groups.

Each cell or a sub-region is involved in k contacts, where k is computed based on the population and contact rate of individuals per day. Assuming that contacts among individuals are Poisson distributed over time, and individuals make contacts at an average rate of λ, the effective contact rate for a sub-region is determined by a Poisson random variate. For a sub-region with population p, $k = p\lambda$. The probability of exposure along with infectivity determines the transmission of infection for a given contact. This leads to heterogeneous interactions in the population, thereby overcoming the presumption of homogeneous mixing in the SIR model. A contact is defined as an interaction between two individuals that may result in successful disease transmission.As derived by the GIS gravity model, the probability of contacts between individuals is inversely proportional to the domestic regional distance between them. The modeling of interactions among individuals in the GSFS model is based on this assumption.

To generate uniform contacts, while considering the age demographics and population distribution, probability distributions are generate. The following algorithms shows the steps involved in generating probability distribution for different age groups. The distribution represents the contacts of individuals in each age group, grouped by locations.

Let N_i be the number of individuals in age group i.

$N = \sum N_i$

Let C_i be the contacts established by the age group i and CR_i be the contact rate of age group i where

$C_i = N_i \times CR_i$

$C = \sum C_i$

Let P be the probability distribution.

$P_i := \{ P_{i-1} + C_i/C \ \ if \ i \geq 1 \}$

$P_i := \{ C_i/C \ \ if \ i = 1 \}$

$if\ i = 1\ to\ n\ P_n = 1$

To generate a contact between two individuals we choose two end points as representatives of these individuals. To identify their demographics and location, we randomly pick an age and location based on the demographics and population dynamics of the location. Subsequently, individual 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 infectious, susceptible, or recovered is decided using a random experiment. If either of individuals is infectious and the other susceptible, then the infection may be transmitted to the susceptible individual. The probability of transmission, referred in the model as infectivity, is the virulence of the virus strain being modeled along with considerations such as distance between individuals, duration of contact etc. Based on the infectivity, the infection may be transmitted to the susceptible individual. During each simulated time step 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. The steps involved in the generation of contacts are explained by the following steps.

3.7.6. *Generating Contacts*

- Randomly pick an age A from age probability distribution
- Pick a cell from the grid where population of A greater than 0
- Based on age A decide whether to make a local or a global contact
- Randomly pick another age B from age probability distribution
- Based on local or global contact pick cell from the respective neighborhoods with population of B greater than 0

- Contact is established between the two cells

3.7.7. *Spatial Interactions*

[2]

We have implemented a dichotomy of global and local interactions to model distance dependency and investigate the role of local and global contacts. This dichotomy helps analyze the transmission of infectious disease in a population where mixing takes place on the local level within small groups like households, and on the global level between random individuals in the population. Studies [10] have shown that transmission within local groups amplifies the total epidemic, which results in smaller periods of epidemics as compared to homogeneous mixing with global contacts only. For global interactions, contacts are initiated between any two cells in the grid, while for local interactions, the contacts are between neighboring cells. In general, locality can be defined as the set of cells (census blocks) within a specified distance range. The mixing patterns of the population are varied over different proportions of global and local interactions. Different age groups use different mixing patterns with different proportions of local and global contact, to model household, social, school or work related contacts.

The prevalence levels of influenza is witnessed to be the same, irrespective of the proportions of local and global mixing. This suggests that influenza prevalence is independent of the spatial domain, and correlates to the results of influenza prevalence in France [16]. The incidence of influenza is further analyzed for varied rates of local and global interactions to generate the corresponding epidemic curves, as shown in Fig 3.13. The incidence decreases with higher proportions of local interactions. The results indicate that although influenza

[2]This section is reproduced from: A. Mikler, S. Venkatachalam and K. Abbas, Modeling Infectious Diseases Using Global Stochastic Cellular Automata, Journal of Biological Systems, Vol. 13, No. 4, pp. 421-439, December 2005. with permission from World Scientific Publishing Co. Pte. Ltd, Singapore.

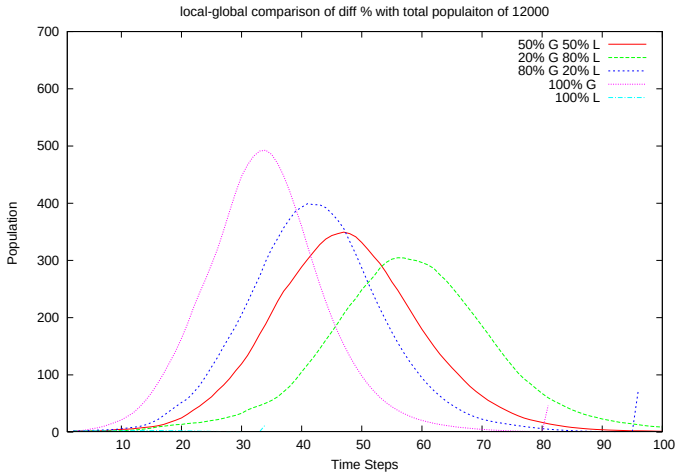


Figure 3.13. Epidemic Curves for Varied Rates of Global and Local Interactions in the Heterogeneous Population of Northern Denton County

prevalence is independent of the spatial domain, the incidence of the epidemic is lowered with higher proportions of local interactions.

CHAPTER 4

EXPERIMENTS

In the following section we discuss the experiments conducted with the model described in the previous chapter. While the main focus of the thesis was the design and implementation of the simulation framework, experiments were conducted to study how different parameters used in the simulation, effect the dynamics of disease spread in a population. Experiments were conducted to analyze the effects of change in demographic parameters and disease parameters. From the review of the literature and epidemiological studies few factors have been identified as being important in a disease manifestation, such as interactions, contact patterns, age-structure and the strain of virus. The model incorporates many of such factors with a simulation perspective and tries to model a real life scenario. The experiments were also used as a method of examining if this framework is a good starting point for a comprehensive model.

4.1. Experimental Setup

Experiments were conducted on a field of size 50 X 100, comprising a total of 5000 locations. Each location on the field has a population and corresponding age-structure mapped from a geographical region. The mapping of regional census data to grid data is explained in chapter 3. For experimental purposes, the population and age-structure data of Denton county, Texas was used. The total population size was around 100000. The results in this section represent the average over multiple similar experiments. The analysis of results in this section have been conducted with reference to to above definitions.

Table 4.1 states the approximate parameter values used for a base experiment. The values were changed for experiments on specific demographics or disease.

Table 4.1. Parameters

Parameter	Value
Incubation period	3 days
Latent period	3 days
Infectious period	4 days
Recovery period	6
Infectivity	0.035
Immunity	0.25
Population	Northern Denton county
Contact Rate	8-14

4.2. Disease Prevalence Distribution

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, on must rely of other means to derive the population dynamics that promote the spread of disease. This is accomplished by using publicly available datasets, that describe composition and behavior of the population of interest. We use data provided by US census. This section discusses the population distribution of the Denton county, Texas. The figures represent the data as obtained after rasterization of the US census data of denton county. For experimental purposes we used the data of northern part of the Denton county. The population of Denton county is around 431000 and population of northern Denton county is around 110000. The northern part encompasses the Denton city.

Denton County Figure 4.1 shows the heterogeneous population distribution of Denton county, with the population of 431,000 on a 200 *200 grid.

Northern Denton county

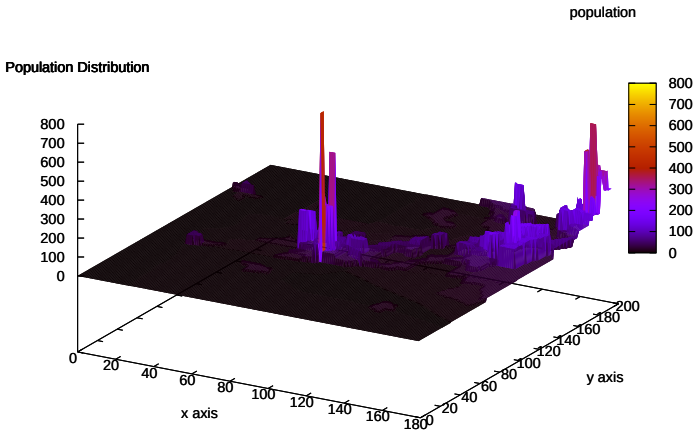


Figure 4.1. Heterogeneous Population Distribution

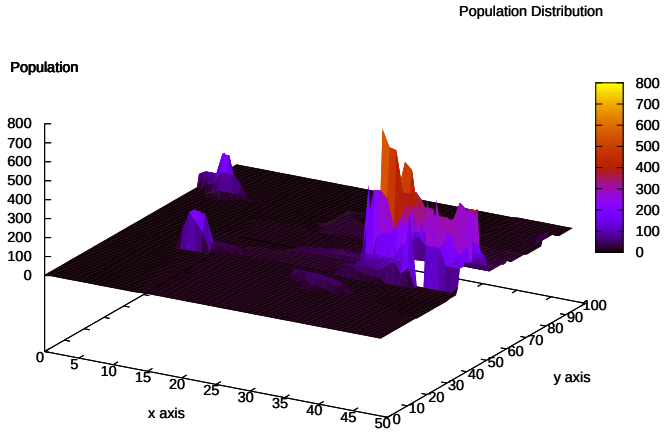


Figure 4.2. Heterogeneous Population Distribution

Figure 4.2 shows the heterogeneous population distribution of area around Denton city. Figure 4.3 illustrates the geographical map of the Northern Denton county area that we used for simulation and experiments. Northern denton county encompasses of Denton city. The total population of the region is 110000.

Figure 4.2 illustrates the heterogeneous population distribution for different age groups for the area around denton city. The different age groups being under 9 years, 10 to 34 years, 35 to 59 years and above 60 of age. It is clear that Denton city has a higher population of

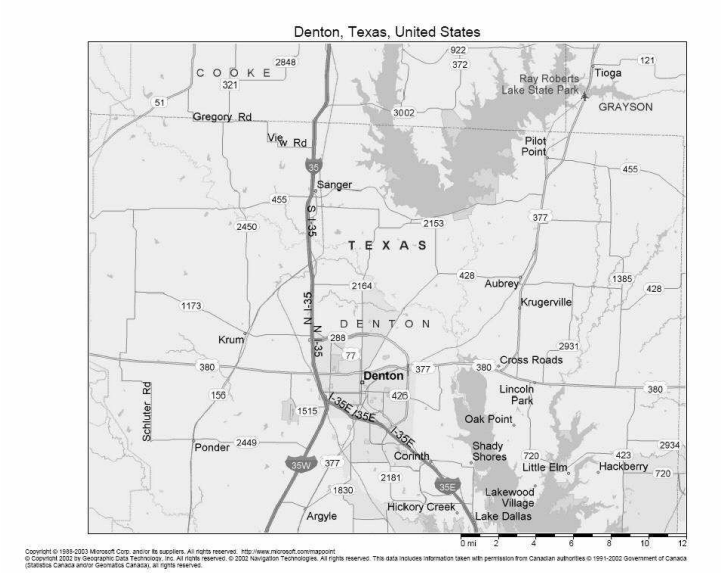


Figure 4.3. Northern Denton County Map

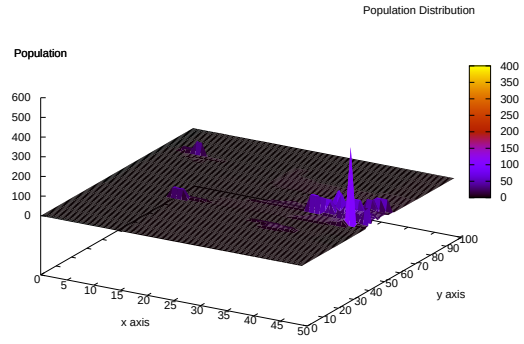
adults in the age group of 10 to 34 years and 35 to 59 years as compared to over 60 years. The child population is higher in the city as compared to areas farther than the city.

4.2.1. *Infected population distribution*

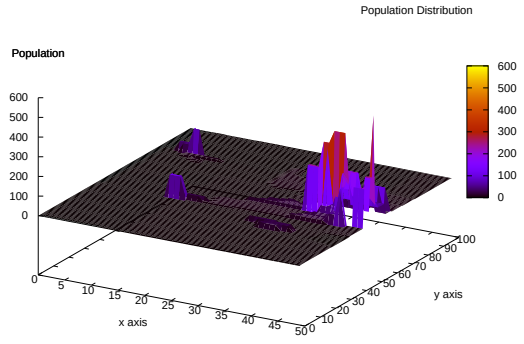
An experiment was conducted to analyze Influenza prevalence over Denton city. Similar parameters were used as mentioned for the base experiment. Disease prevalence data was analyzed for different age groups in the same population.

Fig. 4.5 illustrates the disease prevalence of influenza over that region. The total population of the region is 110000 and the total number of infected people is 48000.

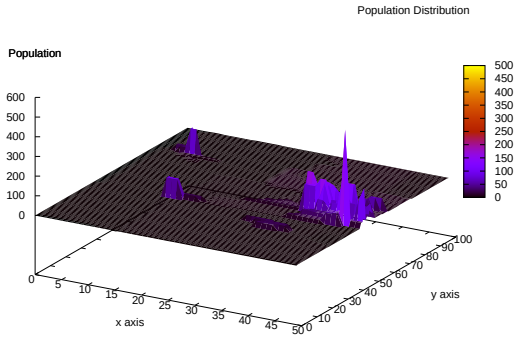
Figure 4.2.1 illustrates the disease prevalence of influenza over the region, for the different age groups being under 9 years, 10 to 34 years, 35 to 59 years and above 60 of age. Prevalence data on being compared with the population distribution of different age groups, we observe that the prevalence of influenza is higher in age groups of 10 to 34 and under 9 years of age. Such a prevalence may be thought of as a result of the population distribution and contact



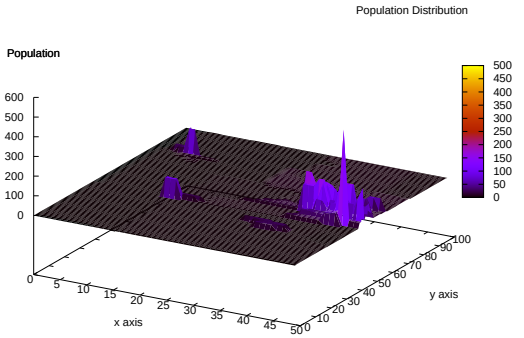
(a) Heterogeneous Population Distribution for age group 0-9



(b) Heterogeneous Population Distribution for age group 10-34



(c) Heterogeneous Population Distribution for age group 35-59



(d) Heterogeneous Population Distribution for age group 60 +

Figure 4.4. Heterogeneous Population for Different Age Groups

rate. Thus this experiment depicts the fact that a disease manifests itself highly in a densely populated region.

4.3. Age Structure Evaluation

To evaluate the age-structure impact on epidemic spread for influenza experiment was conducted with differing influenza disease parameters for each age group and also different contact rates for each age group. Table 4.2 shows the respective contact rates used for each age group. The results in figure 4.7 show normalized epi-curves for the four different age

Table 4.2. Contact Rates

Age group	Avg. Contact Rate
under 9	10
10 to 34	10 days
35 to 59	6
over 60	3

groups. Where the curves show the proportion of the population infected. The epi-curves show the number of infectives in the age group at any given time. The experiment was conducted with similar disease virulence parameter and the same strain of influenza virus. The difference in the curves for each age group is associated with their varying contact patterns and population distribution.

4.4. Contact Structure / Mixing Patterns

Networks of social contacts channel the transmission of airborne infections. For simulating an airborne infection like Influenza, it is important to simulate the contacts established in the population as most of the transmission takes place during contacts. Contact structure of a

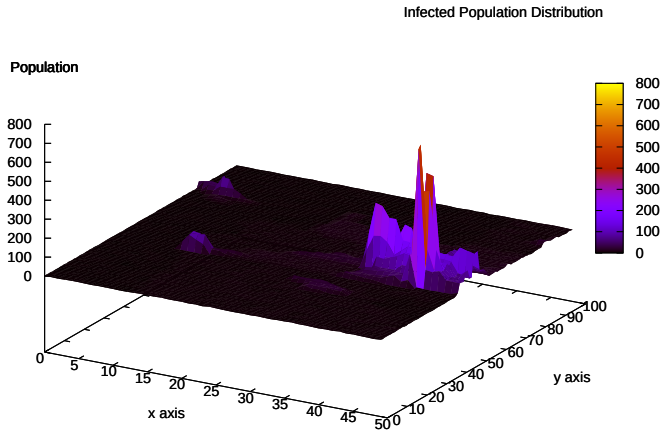
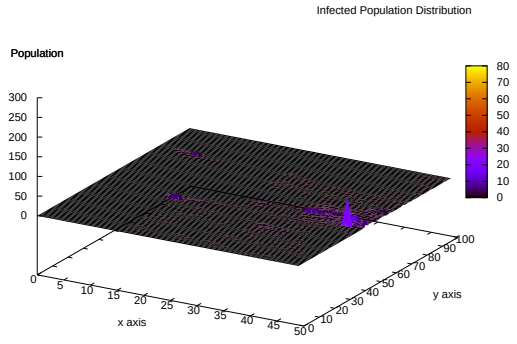
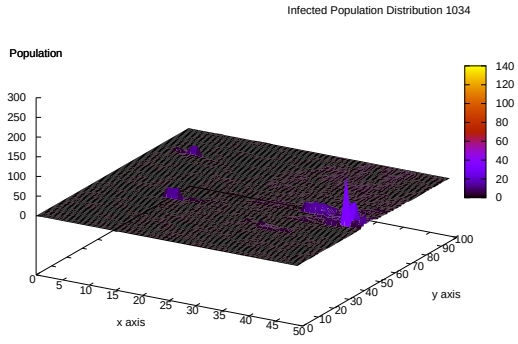


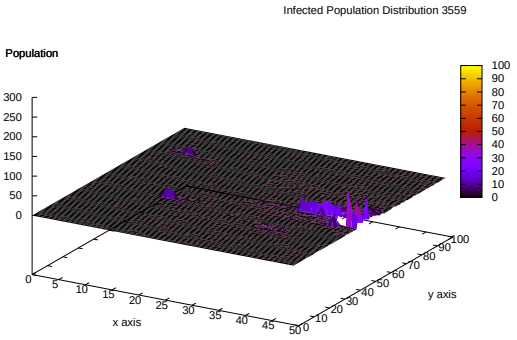
Figure 4.5. Disease Prevalence Distribution



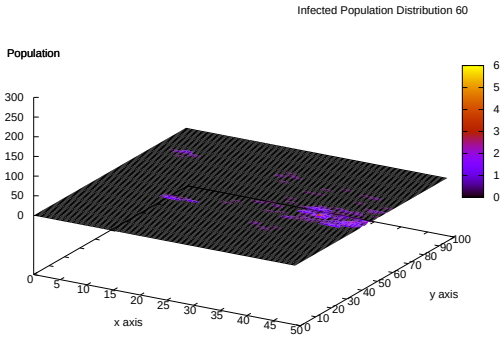
(a) Heterogeneous Population Distribution for age gropu 0-9



(b) Heterogeneous Population Distribution for age group 10-34



(c) Heterogeneous Population Distribution for age group 35-59



(d) Heterogeneous Population Distribution for age group 60+

Figure 4.6. Disease Prevalence Distribution for Different Age Groups

population varies based on the number of contacts made by individuals, the context of the contact, the age group of the contact and distance of the contact. To study the consequences of different contact patterns, different experiments were conducted. Each contact pattern had a specific contact rate distribution among the different age groups.

The GSFS models the behavior patterns based on different contact rates. Figure 4.8 depicts the fact that the incidence level of infection in the population varies depending on the contact rates of the age groups. Table 4.3 lists the contact rates for age groups 1, 2

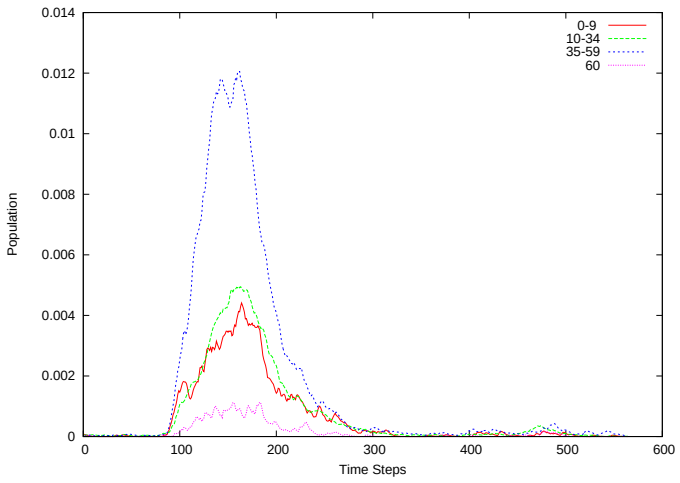


Figure 4.7. Epidemic Curves for Different Age Groups in the Heterogeneous Population of Northern Denton County: Represents Percentage of Population Infected

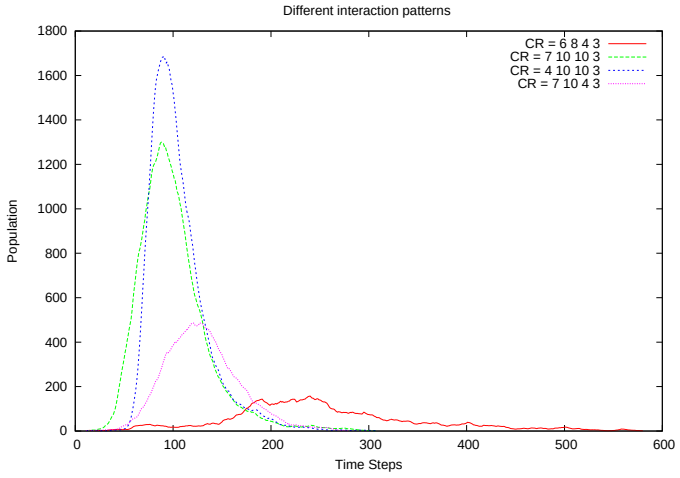


Figure 4.8. Incidence Level of Influenza for Different Contact/Interaction Patterns

,3 and 4 respectively. Higher contact rates for the second and the third age group result in higher incidence level of influenza. With moderate contact rates the infection seems to sustain for a longer time period infecting approximately the same number of people. This can be interpreted as a stretching effect where the duration of epidemic is longer than usual.

Table 4.3. Contact Rate

Age range	Avg. CR	Avg. CR	Avg. CR	Avg. CR
under 9	10	6	7	7
10 - 34	10	8	10	10
35 - 59	6	4	10	4
60 and over	3	3	3	4

4.5. Disease parameters

This section discusses the experiments conducted by changing various disease related parameters to understand the dynamics of disease spread. When studying disease spread in a population, along with the population and demographics dynamics, it is also important to understand the effect of the virus strain and its characteristics.

4.5.1. *Different Diseases*

Using the same metrics of population and grid size, experiments were conducted with the model for three different diseases, namely, common cold, conjunctivitis and influenza, under the assumption of similar virulence/Infectivity of disease.

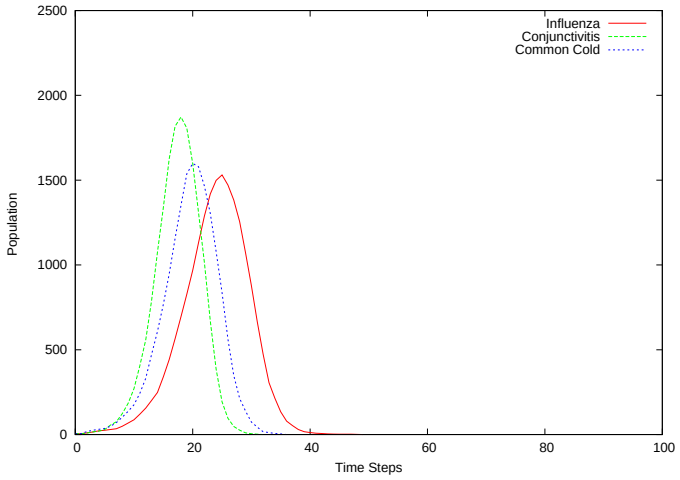


Figure 4.9. Comparison of Infection Spread for Different Diseases

Table 4.4. Infection Timelines for Common cold, Conjunctivitis and Influenza

Disease	Incubation Period	Latent Period	Infectious Period
Common cold	3 days	2 days	5 days
Conjunctivitis	3 days	1 day	6 days
Influenza	3 days	3 days	5 days

Table 4.5. Experiments with Different Index Cases

No. of Index Cases	Age Group	Total Infected
2	1,4	18601
1	2	12377
1	4	2
2	2,3	22384
2	1,3	22424

The infectious period, latency period and recovery period of the diseases, shown in Table 4.4 [11, 41] were used in the experiments. Due to the relatively smaller incubation period and higher infectious period of conjunctivitis, the rate of spread and the prevalence of conjunctivitis is relatively higher in comparison to common cold and influenza (see Fig. 4.9).

4.5.2. *Index Case*

The incidence level of influenza in each group differs based on their interaction patterns as shown by the experiments. To evaluate the effect of the age group of the index case experiment was conducted by introducing index cases in different age group. Table 4.5 shows the total number of infected individuals in the population for index cases in different age groups. Index case in the age group of above 60 years results in no infection spread as compared to two index cases in the age group of 10 to 34 years and 35 to 59 years, as they have a different behavior patterns.

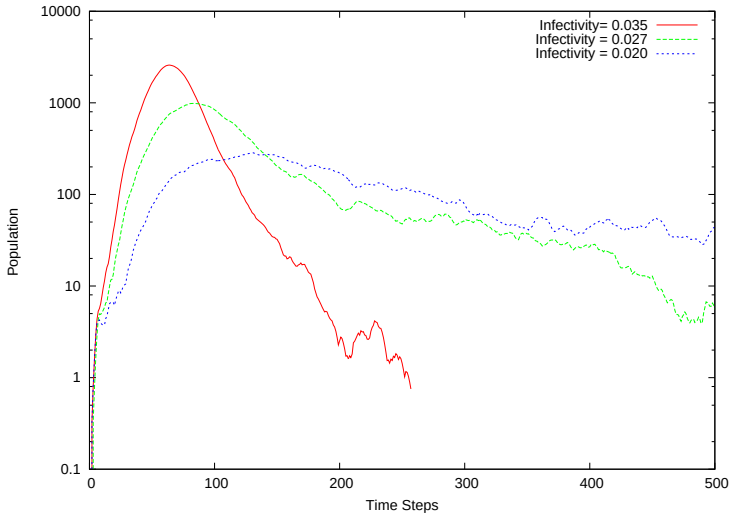


Figure 4.10. Epi-curves for Different Infectivity Levels shown in a Log Scale

4.6. Infectivity

The probability of a contact resulting in successful disease transmission depends on the disease infectivity which is the virulence of the disease and also the various considerations such as distance and duration of contact. When modeling virus strains the virulence can be thought of as the differentiating factor between the strains. The virulence of a strain identifies the degree of ability to cause the disease spread. The prevalence of influenza is analyzed for varied levels of infectivity. Figure 4.10 illustrates that incidence decreases for lower levels of infectivity. Rest of the parameters were similar to that of the base experiment. This experiment exemplifies the sensitivity of the infectivity parameter. As this parameter is more related to the disease dynamics, it is same for all the different age groups.

4.6.1. *Immunity*

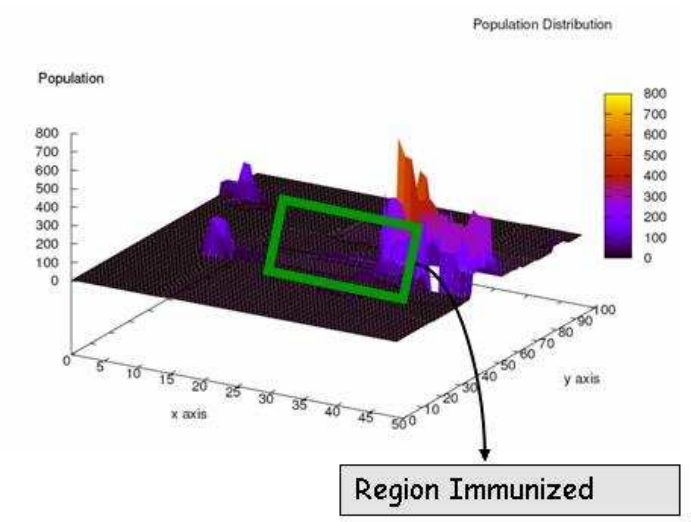
Influenza viruses go through change continually over time, due to mutation, resulting in different virus strains. This changing enables the virus strains to evade the immune system of the host, so that people are susceptible to influenza virus throughout life. Re-infection can occur in influenza when, an individual infected with influenza virus develops antibodies against that virus; and as the virus changes, the older antibodies fail to recognize the new

Table 4.6. Immunity Probabilities for Different Age Groups

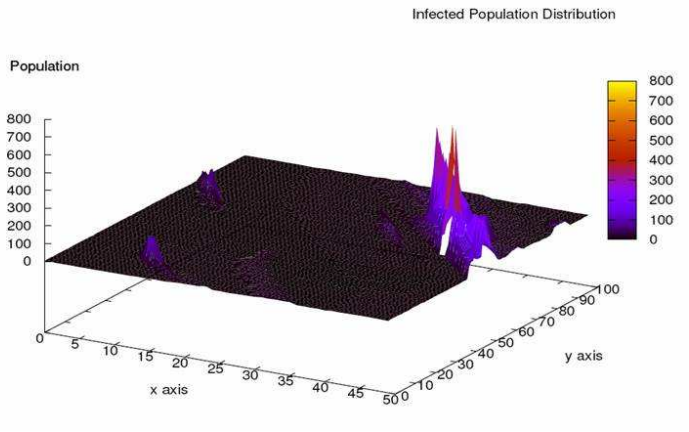
Age Group	Immunity
1	0.6
2	0.8
3	0.8
4	0.0

virus. There are possibilities where certain strains share common sub-sequences, because of that, individuals might develop cross immunity or heterosubtypic immunity. This immunity provides partial protection against re-infection. In order to model this an experiment was conducted where individuals in certain area were immunized. They were less susceptible than individuals in other areas. Fig. 11(a) illustrates the population distribution and region immunized. Result (Fig. 11(b)) indicates considerably lower prevalence as compared to (Fig. 12(a)) which illustrates prevalence for an experiment without immunity.

The constant change in the influenza virus, affects the pathogenictiy of the virus among different age groups. For instance a strain that comes early in the year, might affect children under the age of 15 more than young adults or older adults. This can be modeled by reduced susceptibility for certain age groups. GSFS models reduced susceptibility by introducing immunity probabilities for each age group. Based on the value of immunity probability, the chances of successful transmission of the virus to an individual, reduces or increases during each contact with a susceptible. Experiment was conducted with immunity values for age groups as listed in the table 4.6, where values represent the probability with which the susceptibility of an individual reduces. Figure 13(a) shows the disease prevalence distributions in the population with and without immunity being considered. It depicts that when population was immunized, it resulted in reduced susceptibility and a considerable lower rate of infection.

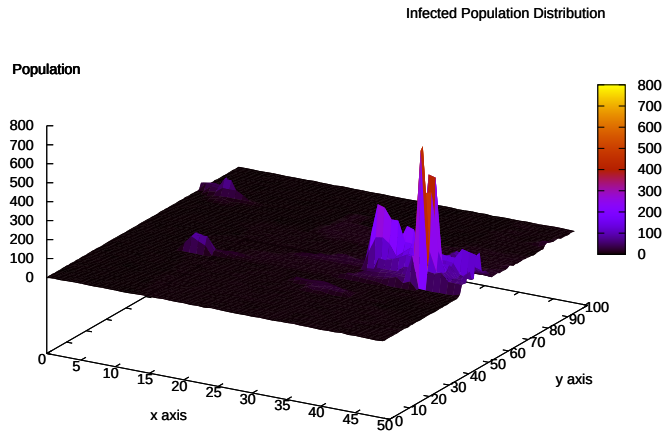


(a) Population Distribution with Immunization



(b) Disease Prevalence with Immunization

Figure 4.11. Population Distribution with immunization



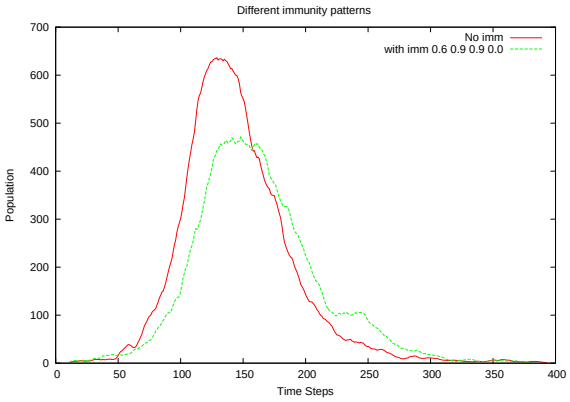
(a) Disease Prevalence without Immunization

Figure 4.12. Population Distribution without Immunization

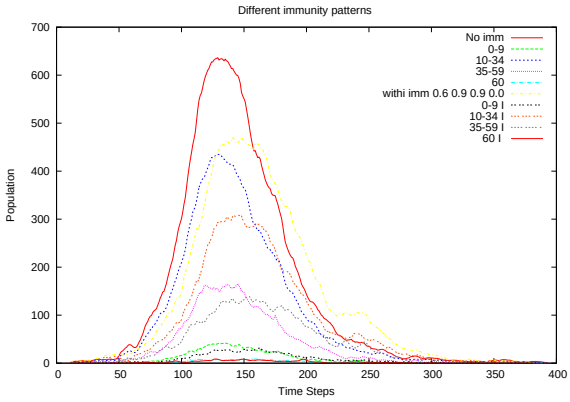
Figures 13(c), 13(d), 14(a), 14(b) show the disease prevalence comparison for total infection and infection for the particular age-groups respectively.

4.7. Composite Model

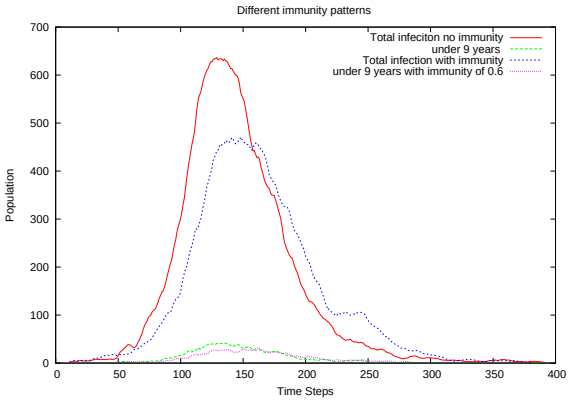
The GSFS model facilitates the modeling and simulation of a single disease outbreak in a large geographic region. The results obtained from this model represent the severity of an epidemic over time, allowing an epidemiologist to quantify the incidence and prevalence in response to employing different public health policies (e.g., vaccination strategies). A *composition model*, is developed which presumes 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 figure 4.15.In the context of influenza, 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 age breakdown are important demographic parameters that will determine how influenza will manifest itself in a particular sub-region. The influenza infectious period in young children is known to exceed that of



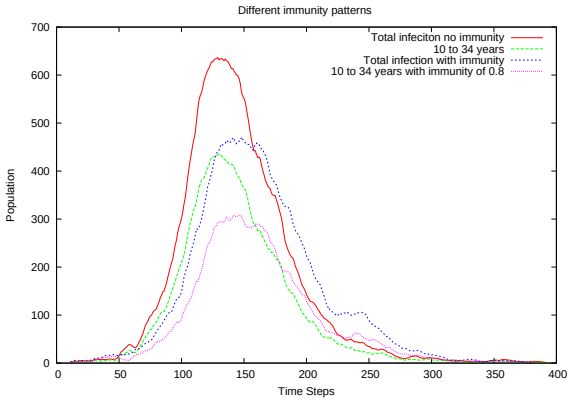
(a) Population Distribution with Immunization



(b) Disease Prevalence

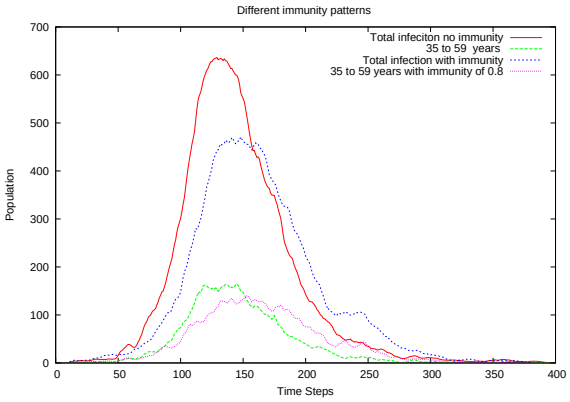


(c) Population Distribution with Immunization

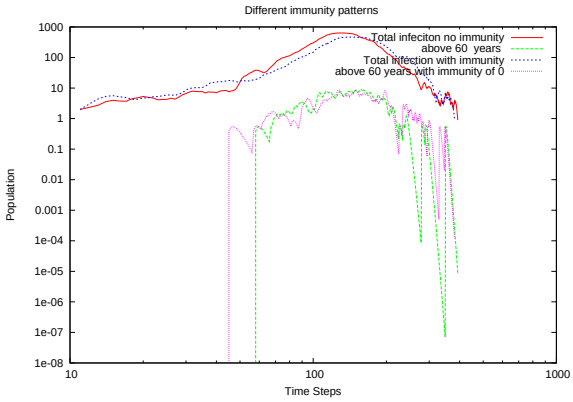


(d) Disease Prevalence

Figure 4.13. Population Distribution and Disease Prevalence with Immunization



(a) Population Distribution with Immunization



(b) Disease Prevalence

Figure 4.14. Population Distribution and Disease Prevalence with Immunization

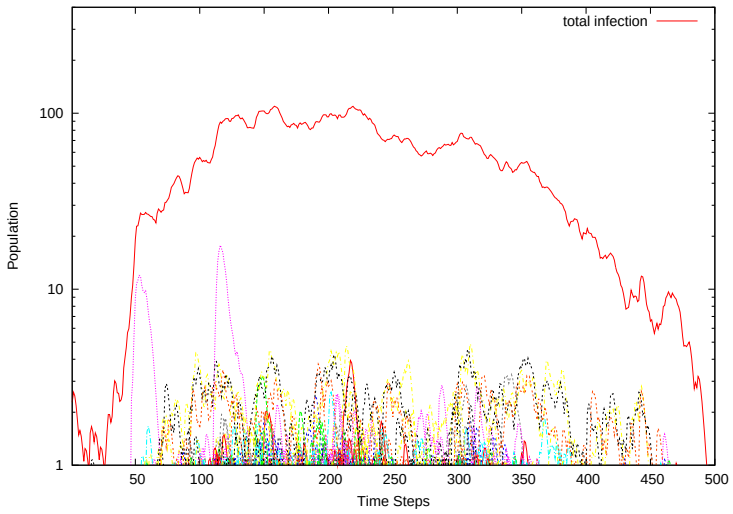


Figure 4.15. Observed Cumulative Epidemic caused by Temporally and Spatially Distributed Local Outbreaks

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 investigate the spread of disease in each location and spread of infection from one location to the other. A path of infection can be drawn by learning the data. The composition models helps in identifying high risk groups and risk rate among particular regions and can aid in applying public health policies.

CHAPTER 5

DISCUSSION

5.1. Conclusion

The modeling of disease progression through classic SIR and traditional CA are limited by the assumptions of homogeneous population and uniform mixing. These limitations are addressed by the GSFS model, which is oriented towards heterogeneous population. The cell interactions are currently based on population density, age and Euclidean distance, and can be extended to incorporate geography, demography, environment and migration patterns.

Modeling outbreaks of infectious diseases using the traditional cellular automata (CA) model is constrained by neighborhood saturation. The classic susceptibles-infectives-removals (SIR) model is oriented towards a homogeneous population with uniform mixing. The limitations of traditional CA and classic SIR models necessitates the need for new computational models to study the complexity of the spread of diseases in the real world. The global stochastic field simulation (GSFS) paradigm is used to model outbreaks of infectious diseases. The GSFS model supports modeling and 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. The GSFS model includes interactions based on population density, age as a demographic characteristic and Euclidean distance, and has been implemented to model the progression of three diseases, namely, common cold, conjunctivitis, and influenza. Rasterized GIS population data of Denton city is incorporated to model heterogeneous population through GSFS. The age structure is incorporated into GSFS by dividing the population into four different age groups. The different age groups being under 9 years, 10 to 34 years, 35 to 59 years and above 60

of age. Based on the experiments done on Denton city which has a higher population of adults in the age group of 10 to 34 years and 35 to 59 years as compared to over 60 years, we can notice that the prevalence is also different for these age groups. This shows that age is an important demographic when modeling disease spread. Along with the age group, other parameters that affect the spread of disease and are dependent on age of a person are infectivity and immunity. GSFS also facilitates modeling diseases with different values of infectivity and immunity. For simulating an airborne infection like Influenza, it is important to simulate the contacts established in the population as most of the transmission takes place during contacts. Contact structure of a population varies based on the number of contacts made by individuals, the context of the contact, the age group of the contact and distance of the contact. Current model of GSFS does simulate different contact patterns based on age groups. The spatial progression of influenza across the heterogeneous population reveals the independence of influenza prevalence for the spatial domain, while influenza incidence decreases with higher rates of local interactions. The GSFS models the behavior patterns based on different contact rates. Experiments have shown that the incidence level of infection in the population varies depending on the contact rates of the age groups. Higher contact rates for the 10 to 34 and 35 to 59 age group result in higher incidence level of influenza. With moderate contact rates the infection seems to sustain for a longer time period infecting approximately the same number of people. This can be interpreted as a stretching effect where the duration of epidemic is longer than usual. The stretching effect further also shows similar effect as seen by self quarantining. By self quarantine the contacts/interactions are reduced as a result of which the infection spread is reduced. GSFS can thus be used to model a situation of quarantine to study the effects on a disease outbreak. To facilitate surveillance, monitoring, prevention and control of different diseases, computational models must be developed. To this end, the GSFS model shall prove to be an valuable asset in the

analysis of progression of infectious diseases, thereby leading to optimal utilization of public health resources.

5.2. Future Work

Although the current GSFS model provides a framework for modeling infectious disease spread, there are lot of chances for future work and to enhance the model further. Based on the experiments shown in the previous chapter, it is obvious that interactions/contact structure is an important consideration while modeling infectious disease spread. The interactions among individuals currently is done based on population and distance and age proportions. However, instead of distance and population, the controlling factors for cell interaction are age proportions and population densities. One caveat, however, is that real-world interaction with individuals in the same neighborhood seldom causes the transmission of influenza. In fact, it is much more likely that influenza viruses are being transmitted at common mixing points, such as schools, work place, church, etc. Hence, we are proposing a modification of the composition model to supplement direct cell interaction with indirect cell interaction through common mixing points. These mixing points can have geographic/demographic associations to facilitate the interaction of specific sub-regions at specific. A model with few mixing points causes an accelerated epidemic over a shorter time as compared to a global interaction model. For example, children from a small sub-region,which is represented by a block of cells in the CA may interact with each other at the same elementary school, transmit influenza among themselves, and carry the disease to their corresponding locality. Teenagers from a larger, not necessarily connected geographic region may interact at the a high school. This may cause the spread of influenza across larger distance. Also as seen in the previous chapter the composition model can prove to be a very effective way to analyze and study the disease spread for all regions which are a part of the county or state. It can further by used to identify certain risk groups in the region for optimal utilization of public health resources.

BIBLIOGRAPHY

[1] K. Abbas, A. Mikler, A. Ramezani, and S. Menezes, *Computational Epidemiology: Bayesian Disease Surveillance*, Proc. of the International Conference on Bioinformatics and its Applications (FL, USA), December 2004.

[2] A. Adams, J. Koopman, S. Chick, and P. Yu, *GERMS: An epidemiologic simulation tool for studying geographic and social effects on infection transmission*, Proc. of the 31^{st} conference on Winter simulation: Simulation - a bridge to the future, vol. 2, 1999, pp. 1549–1556.

[3] E. Ahmed and H. Agiza, *On Modeling Epidemics, including Latency, Incubation and Variable Susceptibility*, Physica A 253 (1998), 347–352.

[4] E. Ahmed and A. Elgazzar, *On some Applications of Cellular Automata*, Physica A 296 (2002), 529–538.

[5] E. Allman and J. Rhodes, *Mathematical Models in Biology An Introduction*, Cambridge University Press, 2004.

[6] J. Aron, *Mathematical Modeling: The Dynamics of Infection*, ch. 6, Aspen Publishers, Gaithersburg, MD, 2000.

[7] R. Bagni, R. Berchi, and P. Cariello, *A Comparison of Simulation Models Applied to Epidemics*, Journal of Artificial Societies and Social Simulation 5 (2002), no. 3.

[8] N. Bailey, *The Mathematical Theory of Epidemics*, Hafner Publishing Company, NY, USA, 1957.

[9] ———, *The Simulation of Stochastic Epidemics in Two Dimensions*, Proc. of the 5^{th} Berkeley Symposium on Mathematics and Statistics (Berkeley and Los Angeles, CA), vol. 4, University of California, 1967.

[10] F. Ball, D. Mollison, and G. Scalia-Tomba, *Epidemics with mixing at two levels*, Ann. Appl. Prob. 7 (1997), 45–89.

[11] A. Benenson (ed.), *Control of Communicable Diseases Manual*, American Public Health Association, 1995.

[12] A. Benyoussef, N. Boccara, H. Chakib, and H. Ez-Zahraouy, *Lattice Three-species Models of the Spatial Spread of Rabies among Foxes*, International Journal of Modern Physics C 10 (1999), 1025–1038.

[13] *Biowar: Simulating disease outbreaks using social networks*, 2004.

[14] N. Boccara and K. Cheong, *Critical Behavior of a Probabilistic Automata Network SIS Model for the Spread of an Infectious Disease in a Population of Moving Individuals*, Journal of Physics A :Mathematical and General 26 (1993), no. 5, 3707–3717.

[15] N. Boccara, K. Cheong, and M. Oram, *A Probabilistic Automata Network Epidemic Model with Births and Deaths Exhibiting Cyclic Behavior*, Journal of Physics A: Mathematical and General 27 (1994), 1585–1597.

[16] E. Bonabeau, L. Toubiana, and A. Flahault, *Evidence for Global Mixing in Real Influenza Epidemics*, Journal of Physics A: Mathematical and General 31 (1998), L361–L365.

[17] A. Callaghan, *Agent-Based Modelling applied to HIV/AIDS*, ERCIM News (2005).

[18] *Centers for Disease Control and Prevention (CDC) website*, 2004.

[19] *Centers for Disease Control and Prevention (CDC). National center for health statistics.*, 2002.

[20] M. Duryea, T. Caraco, G. Gardner, W. Maniatty, and B. Szymanski, *Population Dispersion and Equilibrium Infection Frequency in a Spatial Epidemic*, Physica D 132 (1999), 511–519.

[21] W.J. Edmunds, C.J.Callaghan, and D.J. Nokes., *Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections.*, Proc. R. Soc. London B Biol. Sci. 264 (1997), 949–957.

[22] Stephen Eubank, *Scalable, Efficient Epidemiological Simulation*, Proc. of the 17^{th} Annual ACM Symposium on Applied Computing (SAC'02) (Madrid, Spain), 2002.

[23] S. Fu, *Modelling Epidemic Spread through Cellular Automata*, Master's thesis, The University of Western Australia, 2002.

[24] S. Fu and G. Milne, *Epidemic Modelling using Cellular Automata*, Proc. of the Australian Conference on Artificial Life, 2003.

[25] H. Fukś and A. Lawniczak, *Individual-based Lattice Model for Spatial Spread of Epidemics*, Discrete Dynamics in Nature and Society 6 (2001), 191–200.

[26] *Flu Transmission: Homeland Security website*, Tech. report, 2004.

[27] H. Hamer, *Epidemic disease in England*, Lancet 1 (1906), 733–739.

[28] Matt J. Keeling, *The effects of loacl spatial structure on epidemilogical invasions* , Proc. R. Soc. Lond. 266 (1999), 859–867.

[29] A. Kleczkowski and B. Grenfell, *Mean-field-type Equations for Spread of Epidemics: The 'Small World' Model*, Physica A 274 (1999), no. 355-360.

[30] R. Mansilla and J. Gutierrez, *Deterministic Site Exchange Cellular Automata Models for the Spread of Diseases in Human Settlements*, Complex Systems 13 (2001), no. 2.

[31] A. Mikler, R. Jacob, V. Gunupudi, and P. Patlolla, *Agent-based Simulation Tools in Computational Epidemiology*, Proc. of the International Conference on Innovative Internet Community Systems (Guadalajara, Mexico), June 2004.

[32] A. Mikler, S. Venkatachalam, and K. Abbas, *Modeling Infectious Diseases Using Global Stochastic Cellular Automata*, Journal of Biological Systems 13 (2005), no. 4, 421–439.

[33] D. Mollison, *Epidemic Models: Their Structure and Relation to Data*, Cambridge University Pres, 1995.

[34] D. O'Leary, *Models of Infection: Person to Person*, Computing in Science and Engineering 6 (2004), no. 1.

[35] R. Ross, *An application of the theory of probabilities to the study of a priori pathometry, I*, Proc. Roy. Soc. A (1916), no. 92, 204–230.

[36] B. Schönfisch, *Zelluäre Automaten und Modelle für Epidemien*, Ph.D. thesis, University of Tübingen, 1993.

[37] S. Scott and C. J. Duncan, *biology of Plagues: Evidence from Historical Populations.*, Cambridge University Press, 2001.

[38] H. Situngkir, *Epidemiology through Cellular Automata*, Tech. report, Bandung Fe Institute, March 2004.

[39] ———, *Epidemiology through Cellular Automata*, Tech. report, Bandung Fe Institute, 2004.

[40] D. Stefano, H. Fukś, and A. Lawniczak, *Object-oriented Implementation of CA/LGCA Modeling Applied to the Spread of Epidemics*, Canadian Conference on Electrical and Computer Engineering (Halifax), IEEE, 2000, pp. 26–31.

[41] T. Timmreck, *An Introduction to Epidemiology*, ch. 2, pp. 38–39, Jones and Bartlett, Boston, 2002.

[42] P. Venkatachalam and B. Krishna Mohan, *GRAM++ - A GIS Software For Resources Management*, Proc. of American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Conference (Washington DC, USA), May 2000.

[43] S. Venkatachalam and A. Mikler, *An Infectious Outbreak Simulator based on the Cellular Automata Paradigm*, Proc. of the International Conference on Innovative Internet Community Systems (Guadalajara, Mexico), vol. 3473, June 2004, pp. 198–211.

[44] ———, *Towards Computational Epidemiology: Using Stochastic Cellular Automata in Modeling Spread of Diseases*, Proc. of the 4^{th} Annual International Conference on Statistics, January 2005.

[45] C. Viboud, P. Bolle, K. Pakdaman, F. Carrat, A. Valleron, and A. Flahault, *Influenza epidemics in the United States, France, and Australia, 1972-1997*, Emerging Infectious Diseases 10 (2004).

[46] *World Health Organization (WHO) website*, 2004.

[47] S. Wolfram, *Statistical Mechanics of Cellular Automata*, Reviews of Modern Physics 55 (1983), 601–644.

[48] A. Yaganehdoost, E. Graviss, M. Ross, G. Adams, S. Ramaswamy, and G. Wanger et al, *Complex Transmission Dynamics of Clonally Related Virulent Mycobacterium Tuberculosis Associated with Barhopping by Predominantly Human Immunodeficiency Virus-positive Gay Men*, Journal of Infectious Diseases 180 (1999), no. 4, 1245–51.

[49] C. Youngblut, *Educational Uses of Virtual Reality Technology*, Tech. report, Institute for Defense Analyses, Alexandria, VA, 1998, Technical Report D-2128 IDA Document.