

UNIVERSITY OF SOUTHAMPTON

**Simulation the Transmission of Airborne
Infectious Disease by Individual
Space-Time Activity-based Model**

by

Yong Yang

Thesis for the degree of Doctor of Philosophy

in the

School of Geography

August 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF GEOGRAPHY

Doctor of Philosophy

**SIMULATION THE TRANSMISSION OF AIRBORNE INFECTIOUS DISEASE BY
INDIVIDUAL SPACE-TIME ACTIVITY-BASED MODEL**

By Yong Yang

The 21st century has seen a boom of the application of individual-based models for the simulation of infectious disease transmission. How to consider the contact patterns and spatial effects at the individual level are two of the key questions. Considering the contact patterns and spatial effects, ISTAM (Individual space-time activity-based model) was designed and developed. The model is based on the integration of an individual-based model and GIS under the framework of time geography. ISTAM is a bottom-up model in which the transmission network is built on the physical contacts between individuals at a fine space-time scale. At this scale, human social behaviors, the environment's physical conditions and specific infectious disease's transmission modes are considered. The two-level structure of ISTAM (between-AB and within-AB levels) makes the model flexible to be applied to different circumstances. At the within-AB level, raster space AB simulation, vector space AB simulation and rôle-based AB simulation were designed and applied to different application cases. Activity bundle (AB) simulation is a method to obtain a specific contact network (specific to target infectious disease) from the space-time dynamics of individuals at fine scales constrained by both the individuals' social activity and the space's physical condition. Parameters of raster space AB were explored. The human contact network, a by-product during the simulation, was used for model calibration and validation. ISTAM was applied to simulate hypothetical influenza outbreaks in the campus of the University of Southampton and the city of Eemnes (the Netherlands), respectively. Different control measures were tested. The results show that the model behavior is approximately consistent with expectations. By the simulation of ISTAM, simulated data can be accumulated at micro-level and aggregated to several different higher levels. This can be used for the model validation and calibration purposes. Also further analyses can be based on these simulated data.

Contents

Table of Contents	ii
List of Figures	vi
List of Tables	x
Declaration of Authorship	xii
Acknowledgements	xiv
1 Introduction	1
1.1 Background of this research	1
1.2 Infectious disease transmission	4
1.2.1 Parameters for infectious disease transmission	6
1.2.1.1 Time lines	6
1.2.1.2 Transmission probability	7
1.2.1.3 Basic reproductive number	7
1.2.1.4 Proportion of infections before the onset of symptoms	9
1.2.1.5 Virulence	9
1.2.2 Control measures	9
1.3 Thesis outline	10
2 Modelling Infectious Disease Transmission	11
2.1 Mathematical models	11
2.1.1 SIR models	12
2.1.2 SEIR models	13
2.1.3 SIS models	13
2.1.4 Reed-Frost epidemic models	14
2.1.5 Problems of mathematical models	15
2.2 Individual-based models	16
2.2.1 Definition of individual-based models	16
2.2.2 General benefits and limitations of individual-based models	17
2.2.3 Individual-based models for the simulation of infectious disease transmission	19
2.3 Contact patterns	21
2.4 Spatial effects	23
2.5 Summary	25

3	Modelling Humans' Space-Time Dynamics	26
3.1	Time geography	26
3.2	Concept of activity bundle and its relation to infectious disease transmission	29
3.3	Between-AB simulation	30
3.4	Within-AB simulation and properties of humans' space-time dynamics . .	32
3.4.1	Air within an AB	33
3.4.2	Space within an AB	34
3.4.3	Distance between individuals within an AB	36
3.4.4	Two kinds of contacts within an AB	36
3.4.5	Individual's rôle in their activities	37
3.4.6	Properties of humans space-time dynamics within an AB	37
3.5	Geographical information systems	38
3.6	Network theory	39
3.7	Summary	42
4	Conceptual Model of ISTAM	44
4.1	Objects in ISTAM	45
4.1.1	Infectious diseases	45
4.1.2	Activity bundles	48
4.1.3	Individuals	49
4.2	Two levels of simulation	50
4.3	Within-AB simulation	52
4.3.1	Raster space AB simulation	53
4.3.2	Vector space AB simulation	53
4.3.3	Rôle-based AB simulation	57
4.4	Input and output	61
4.5	Summary	61
5	Implementation of ISTAM	62
5.1	Tools for implementation of ISTAM	62
5.1.1	IBM library of RePast	62
5.1.2	Java Topology Suite	63
5.1.3	Object-oriented programming and Java	64
5.2	Main packages and classes of ISTAM	65
5.2.1	Class of Student and Citizen	67
5.2.2	Class of DiseaseEvolution	67
5.2.3	Class of ActivityPattern	68
5.2.4	Classes for spatial and social environment	68
5.2.5	Class of Contact	69
5.2.6	Class of DiseaseControlStrategy	69
5.3	Simulation process of ISTAM	69
5.3.1	Preparation stage	69
5.3.2	Simulation stage	70
5.4	Summary	71
6	Application to the Campus of the University of Southampton	73

6.1	Background and survey	73
6.2	Students and ABs	74
6.3	Influenza	75
6.4	Simulation results	78
6.4.1	Calibration by the basic reproductive number	78
6.4.2	Base line simulation	80
6.4.3	Number of index cases	81
6.4.4	Distribution type of the population and index cases	83
6.4.5	Control measure tests	87
6.4.6	Contact network analysis	89
6.4.7	AB simulation analysis	93
6.5	Summary	98
7	Parameter Exploration of Raster space AB simulation	99
7.1	Introduction	99
7.2	Proportion occupied	101
7.2.1	Type 1	101
7.2.2	Type 2	103
7.2.3	Type 3	104
7.2.4	Type 4	105
7.3	Number of index cases	107
7.4	Size of AB	110
7.5	Ratio of AB	112
7.6	Size of cell and ratio of cell	115
7.7	Movement proportion and movement frequency	117
7.8	Rank importance of parameters	120
7.9	Risk assessment for ABs within the application to the campus of the University of Southampton	121
7.10	Summary	127
8	Application to Eemnes	129
8.1	Overview of Eemnes	130
8.2	Background and data sources	131
8.3	Preparation process of building the population and city	132
8.3.1	Process 1: selection of activity types	133
8.3.2	Process 2: selection of activity bundles	134
8.3.3	Process 3: building of city	134
8.3.4	Process 4: population synthesis	135
8.3.5	Process 5: assign activity patterns to individuals	136
8.4	Simulation process	137
8.4.1	Simulation of individuals' movements between ABs	137
8.4.2	Simulation of individuals' interactions within ABs	139
8.5	Simulation results	141
8.5.1	Calibration	141
8.5.1.1	Calibration of the contact frequency index	141
8.5.1.2	Calibration of the contact intimacy index	142

8.5.2	Dynamics of population during the whole day	142
8.5.3	Infection PC6 distribution	143
8.5.4	Infection distribution at different types of AB	145
8.5.5	Network analysis	147
8.6	Summary	148
9	Discussion	149
9.1	Reflection on two application cases of ISTAM	149
9.2	Calibration and validation of ISTAM	151
9.3	Selection of AB	153
9.4	Within-AB simulation	155
9.5	Possibility of further analysis based on the simulated data of ISTAM . . .	158
9.5.1	Introduction to geostatistics	159
9.5.2	Pilot model description	160
9.5.3	Initial analysis results	161
9.6	Scaling up and scaling down at different scales	161
9.7	Changing of humans' daily activity patterns	163
9.8	Problems of current ISTAM and future work	164
10	Conclusion	166
A	Interfaces of the applications of ISTAM	169
B	Interfaces of the pilot project	173
C	Questionnaire of student activity survey	175
	Bibliography	177

List of Figures

1.1	Epidemiological triad	5
1.2	Time lines of infectious diseases	6
1.3	R_0 estimated from various studies for SARS and other diseases.	8
3.1	Elements of time geography	27
3.2	Simplified activity space.	31
3.3	Raster space of a homeless shelter (The shelter contains mats and beds, with the mats shown in light-grey and the beds shown in dark-grey. People are shown by dots with different colours for different states).	35
3.4	Vector space of a hospital ward (comprised of nine rooms visited by three nurses three times per day. Patients are shown as squares and nurses as triangles. Patients and nurses have different states shown by different colours).	35
3.5	Random rewiring procedure for interpolating between a regular ring lattice and a random network	41
3.6	The percentage of people infected as a function of time by a disease which starts with a single person and spreads through a community with the topology of a small world graph	41
4.1	Structure of ISTAM	46
4.2	Simulation at two levels: (a) an example of student A and B's activities between 11:30 and 13:30, (b) an example of the spatial distribution of students inside a lecture room between 12:00 and 13:00, (c) an example of the spatial distribution of students inside a refectory between 13:00 and 13:30	51
4.3	Proportion infected plotted against the distance between individuals for different values of the mobility index. In (a) the total number of individuals is 10 and in (b) it is 5.	55
4.4	Spatial display of a shopping mall by vector space AB simulation.	56
4.5	The spatial distribution pattern within a group: (a) static random, (b) static even, (c) queue and (d) loop (The points represent individuals. The solid lines between two points represent the possible contacts existing between these two individuals. The dotted circle around individual a shows all the individuals within this area who could come into contact with individual a).	58
4.6	Classification by rôles of individuals inside a restaurant.	60
5.1	Packages and classes within an example application of ISTAM.	66

6.1	The relation between infectivity index and the basic reproductive number from simulation results.	79
6.2	The relation between attack distance and the basic reproductive number from simulation results.	79
6.3	Example plot of base line simulation.	80
6.4	Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different number of index cases.	82
6.5	Distribution of (a) duration, (b) peak day, (c) peak number and (d) remaining number for different allocations of the population with random allocated index cases.	85
6.6	Distribution of (a) start day, (b) duration, (c) peak number and (d) remaining number for different distribution of population with same school index cases.	86
6.7	Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different control measures.	88
6.8	Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different alert values.	90
6.9	Contact network analysis: degree distribution of all students for one day's contact.	92
6.10	Simulation plots of within-AB for six different ABs: (a) gym, (b) computer room, (c) lecture room, (d) bar, (e) student union and (f) library room. Black dots correspond to simulated points, straight black lines correspond to trend lines, and straight grey lines are computed on the basis of the 'density' assumption (see text). Simulation starts with one index case and an attack distance 2 m.	95
6.11	Percentage of infections from different types of ABs with an attack distance of 2 m.	96
6.12	Percentage of infections from different types of ABs with different attack distances: (a) 1 m, (b) 1.5 m, (c) 2.5 m, (d) 3 m, (e) 3.5 m and (f) 4 m.	97
7.1	Different sizes of infection area based on the relative location of infectious individuals within an AB.	100
7.2	Proportion infected plotted against proportion occupied with default values for other variables.	102
7.3	The relation of number of newly added effective contacts with the location of the n^{th} individual (individual N is the n^{th} individual and individual A, B, C and D are individuals whose effective contact numbers increase by 1 due to the newly added individual).	103
7.4	The shape of clusters with different g values (2-9) and the sequence of adding individuals around the key member.	104
7.5	Proportion infected plotted against proportion occupied for a Type 3 AB with values of (a) $g = 2$, (b) $g = 3$, (c) $g = 4$, (d) $g = 5$, (e) $g = 6$, (f) $g = 7$, (g) $g = 8$ and (h) $g = 9$ (solid dots are the simulated results and the solid lines are based on Equation 7.3).	106
7.6	Proportion infected plotted against proportion occupied with different numbers of index cases for Type 1, 2 and 3 ABs.	108

7.7	Proportion infected plotted against proportion occupied with different numbers of index cases for Type 4 AB.	109
7.8	Proportion infected plotted against number of index cases for Type 4 AB (solid dots are the simulated results while the solid line is based on Equation 7.5).	109
7.9	Proportion infected plotted against size of AB with different numbers of index cases and proportion occupied for Type 1, 2, 3 and 4 ABs.	110
7.10	Proportion infected plotted against size of AB for Type 4 AB (solid dots are the simulated results while the solid line is based on Equation 7.6).	111
7.11	Proportion infected plotted against size of AB for (a) Type 1, (b) Type 2 and (c) Type 3 ABs (solid dots are the simulated results while the solid lines are based on regression formulae).	113
7.12	Proportion infected plotted against ratio of AB for Type 1, 3 and 4 ABs with different numbers of index cases and proportion occupied.	114
7.13	Proportion infected plotted against ratio of AB for Type 2 ABs with different numbers of index cases and proportion occupied.	115
7.14	Proportion infected plotted against size of cell for Type 1, 2, 3 and 4 ABs with different numbers of index cases and proportion occupied.	116
7.15	Proportion infected plotted against movement proportion with different numbers of index cases and movement frequency.	118
7.16	Proportion infected plotted against movement percentage for dynamic ABs (solid dots are the simulated results while the solid line is based on Equations 7.8 and 7.9).	119
7.17	Proportion infected plotted against movement frequency for dynamic ABs (solid dots are the simulated results while the solid line is based on Equations 7.8 and 7.9).	119
7.18	Simulated proportion infected plotted against the expected proportion infected based on Equations 7.8 and 7.9.	120
7.19	Simulated changing number of individuals within an AB during a whole day: (a) bar, (b) off campus, (c) refectory, (d) library room, (e) student union, (f) gym, (g) in campus, (h) kitchen and (i) computer room.	124
7.20	Simulated changing probability of infection for individuals within an AB during a whole day (assume one infectious individual present): (a) bar, (b) refectory, (c) library room, (d) student union, (e) gym, (f) kitchen and (g) computer room.	127
8.1	Preparation process of building the population and city of Eemnes.	133
8.2	Spatial distribution of households and schools at PC6 levels (solid black circles indicate PC6 with school).	138
8.3	Spatial distribution of AB for work, relaxation and maintenance at PC6 levels.	139
8.4	Distribution of number of contacts for all individuals per day.	141
8.5	The changing number of susceptible, infected and immune individuals over time for one simulation of 100 days.	143
8.6	Individuals' spatial distribution at PC6 levels, at (a) 4 am, (b) 10 am, (c) 4 pm and (d) 10 pm.	144

8.7	Spatial distributions of infections at PC6 level: (a) based on the AB where infections occurred; (b) based on the households of infected persons	145
8.8	The changing of number of infected but not infectious, infectious without symptoms, infectious with symptoms and the total infected individuals over time.	146
8.9	The contact network between individuals for one day.	147
9.1	Process of calibration and validation of ISTAM.	152
9.2	Relation between different types of static AB and dynamic AB.	155
9.3	Variograms for days (a) 65, (b) 149 and (c) 308 showing some of the changes in spatial structure which occurs through simulation sequences.	162
A.1	Initial interface of ISTAM.	169
A.2	Main interface of ISTAM.	170
A.3	Parameters interface of ISTAM.	170
A.4	Interface of initial infectious individuals setting.	171
A.5	Spatial display of ISTAM application to the campus of University of Southampton.	171
A.6	Spatial display within flat of ISTAM application to the campus of University of Southampton.	172
A.7	Interface for displaying of individuals' records within ISTAM.	172
A.8	Interface of infection tree within ISTAM.	172
B.1	Main interface of the variogram program.	173
B.2	Spatial display of the variogram program (for a cell, the darker the color, the more infected cases).	174
B.3	Variograms for many days.	174
B.4	Number of infected individuals by day.	174
C.1	Questionnaire of survey on students' activity pattern.	176

List of Tables

1.1	Means of transmission.	5
3.1	Distance rules.	36
4.1	Model parameters in ISTAM.	47
4.2	Four types of spatial distribution of individuals within an AB	54
4.3	Parameters for vector space AB simulation.	55
4.4	Spatial patterns within group.	59
4.5	Interaction types.	59
4.6	Possible contacts between individuals inside a restaurant.	60
5.1	Main tables within database used by ISTAM.	70
5.2	The schedule for each day within ISTAM.	72
6.1	List of ABs in an example application of ISTAM.	75
6.2	Parameters for influenza in an example application of ISTAM.	77
6.3	Value and description of five parameters for base line simulation.	81
6.4	ANOVA analysis of the five characteristic for the combination of population allocations and index cases allocations.	84
6.5	Percentage of infections from different types of ABs with different attack distances.	96
7.1	Input variables, their default values and value ranges.	101
7.2	Average number of effective contacts for individuals within clusters with different g values (2-9).	105
7.3	Regression analysis results between proportion infected and ratio of AB for Type 1, 2 and 3 ABs with P_o equal to 0.5 and number of index cases equal to 1.	112
7.4	Different sum of covered areas for different shapes of cell.	117
7.5	Four variables and ranges of values for dynamic AB simulation.	117
8.1	Comparison between the application of ISTAM to University of Southampton campus and the city of Eemnes.	130
8.2	Data sources for the application of ISTAM to Eemnes.	132
8.3	All activity types which take more than 1% in the average time distribution.	133

8.4	Activity bundle types within ISTAM for the application to Eemnes (The number in brackets is the expected number of staff working in this type of work place while * means one and only one AB of this type within the current PC6.	134
8.5	Age structure for the population of Eemnes.	135
8.6	Working index according to the average number of working (study) hours per week.	136
8.7	Rôle types within household.	136
8.8	Contacts within AB.	140
8.9	Percentage of infections at different types of AB.	147
9.1	Subgrouping mixing.	154
9.2	Daily smallpox transmission probabilities in a simulated community. . . .	154
9.3	Transmission probabilities for different types of AB	158

Acknowledgements

I am indebted to my supervisor, Prof. Peter Atkinson, for his support from almost all aspects of my life in Southampton, endless encouragement and guidance for this research. I gained much from his intelligence, experience, knowledge and motivation.

I give a big thank you to Dr. Dick Ettema, who supported part of this research during my three months' visit to the University of Utrecht and provided related data sources. I am grateful for the scholarship from the global exchange programme within Worldwide Universities Network which made this visit possible.

I gratefully acknowledge the funding support for this research through the graduate research scholarship from School of Geography and the overseas research students award from the Universities UK.

I thank Prof. David Martin, Ms. Sam Cockings, Dr. Peter Gething and all other related staff and students within School of Geography for their kindly help and support.

Special thanks goes to my wife, Xin. She came to the UK to support me in 2004. We shared so many things in the last six years. I am looking forward to share more with you in the rest of our life.

Finally, thanks to my parents, the rest of my family and all my friends for their love and support.

Chapter 1

Introduction

1.1 Background of this research

To predict the future is very hard, even for Lord Kelvin (1824-1907). In 1900, some of his words in his lecture titled “Nineteenth Century Clouds over the Dynamical Theory of Heat and Light” to the Royal Institution of Great Britain may seem funny now: “There is nothing new to be discovered in physics today. All that remains is more and more precise measurement”. However, this is probably unfair, since the two dark clouds (the problem of the ether and the problem of specific heats) predicted by him proved to be so dark that they were beyond his imagination (Kleppner 1998). In the middle of the last century, similar and improper optimism about infectious diseases happened again in the field of medicine and public health. Unfortunately, even today we humans still do not know how many dark clouds are on the way.

Infectious diseases afflict humans at all times. The impacts of infectious diseases on humans are enormous, both in terms of suffering and in terms of social and economic consequences. From the 1950s, a period that saw unprecedented development of new

vaccines and antimicrobial agents, the optimistic view that infectious diseases were no longer threats (at least to developed countries) kept growing and was boosted by the eradication of smallpox in the late 1970s (Fenner et al. 1988). The Surgeon General of the United States (1965-1969), William H. Stewart, declared that it was time to close the book on infectious diseases (Patlak 1996) and chronic ailments such as cancers and heart diseases should be paid more attention to. This encouraged a transfer of resources and public health specialists away from infectious disease control.

The above optimistic view proved to be wrong. Even in the same decade, Acquired Immune Deficiency Syndrome (AIDS) emerged as a pandemic disease. West Nile virus, Ebola and Severe Acute Respiratory Syndrome (SARS) kept emerging. Old diseases like measles, malaria and influenza still threaten large numbers of human beings. Besides humans themselves, livestock are affected severely by foot-and-mouth disease (FMD), Bovine Spongiform Encephalopathy (BSE, also known as 'mad cow disease') and avian flu, and crops and plants are infested by pests. Prevention of bio-terror (such as anthrax or smallpox) is also a concern of governments and the public. Population growth, increasing mobility, changes of geographic distribution of the human population, and other global changes such as land conversion, agricultural intensification and climate change, all have an impact upon the epidemiological environment (Mayer 2000, Daily & Ehrlich 1996). At the same time, immune suppression, the loss of biodiversity and indigenous knowledge, and the evolution of antibiotic resistance, all decrease human ability to defend against diseases (Daily & Ehrlich 1996).

Simulating the processes of infectious disease transmission amongst the human population, both in time and in space, can not only lead to better understanding of the transmission mechanisms, but also provide practical utilities such as building and testing of theories, design and analysis of epidemiological surveys, forecasting trends and testing of control measures (Hethcote 2000). Faced with diseases such as AIDS, SARS, FMD and avian flu, powerful simulation models for infectious disease transmission are becoming increasingly important tools for public health researchers and practitioners.

The transmission of an infectious disease amongst the human population is a complex process. The time lines of infectious diseases, the infection probability, the contacts between individuals, the demographic dynamics that determine the contact pattern and the occasional imported infection can all be considered as stochastic. It is hard to build a satisfactory model. The reasons rest with two aspects: theory and data (Koopman 2005). On one hand, human knowledge at both the micro and macro levels of the infectious disease transmission process is insufficient. At the micro level, the exact process of a single infection of most types of infectious diseases is not clear, even today (Koopman 2004). At the macro level, disease diffusion amongst the human population is a complex process, as many factors from both social and physical aspects can affect this process with different magnitudes. On the other hand, it is hard to get all the data that the model requires in terms of model building, validation and application.

Considering the above difficulties, it is crucial to choose appropriate tools for modelling the diffusion of a certain infectious disease amongst the human population within a target study area. Mathematical models and computer simulations are two general categories of tools for simulating the processes of infectious disease transmission (Koopman 2004, Hethcote 2000). Mathematical models have been used from the earlier part of the 20th century, and both their merits and weaknesses have been identified (see Section 2.1). Individual-based models (IBM), as one sub category of computer simulations, have been paid more attention recently (see Section 2.2.3).

The most important advantages of IBMs are that an IBM can consider the heterogeneity of both individuals and their environment and incorporate the stochastic nature of infectious disease transmission. However, this advantage also means that the applications of IBM for transmission of infectious disease are strongly data-dependent. And quite a number of assumptions have to be made to make the simulation plausible. How to consider the contact patterns and spatial effects at the individual level are two of the key questions.

During this PhD research, ISTAM (Individual space-time activity-based model) was designed and developed to simulate the transmission of airborne infectious disease (influenza) amongst the human population based on individuals' space-time activities. The model is based on the integration of an individual-based model and GIS under the framework of time geography. ISTAM is a bottom-up model in which the transmission network is built on the physical contacts between individuals at a fine space-time scale. At this scale, human social behaviors, the environment's physical conditions and specific infectious disease's transmission modes are considered. The two-level structure of ISTAM (between-AB and within-AB levels) makes the model flexible to be applied to different circumstances. At the within-AB level, raster space AB simulation, vector space AB simulation and rôle-based AB simulation were designed and applied to different application cases. Activity bundle (AB) (Hägerstrand 1975) simulation is a method to obtain a specific contact network (specific to target infectious disease) from the space-time dynamics of individuals at fine scales constrained by both the individuals' social activity and the space's physical condition. Parameters of raster space AB were explored. The human contact network, a by-product during the simulation, was used for model calibration and validation. ISTAM was applied to simulate hypothetical influenza outbreaks in the campus of the University of Southampton and the city of Eemnes (the Netherlands), respectively. Different control measures were tested. The results show that the model behavior is approximately consistent with expectations. By the simulation of ISTAM, simulated data can be accumulated at micro-level and aggregated to several different higher levels. This can be used for the model validation and calibration purposes. Also further analyses can be based on these simulated data.

1.2 Infectious disease transmission

The traditional model of infectious disease causation can be expressed by the epidemiological triad (Last 1988). It includes three components: an external agent (infectious disease), a susceptible host (human), and an environment that brings the host and agent

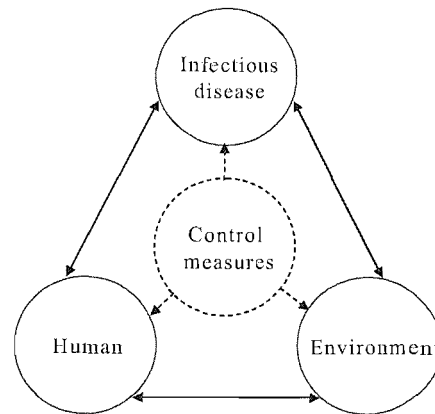


FIGURE 1.1: Epidemiological triad

together to make the infection possible (see Figure 1.1). Possible control measures are applied to target populations or the environment. The classification of infectious diseases can be by causative organism, means of transmission or disease reservoir. The causative organism is the main way to classify infectious diseases: pathogenic organism, bacteria, viruses, fungi, and parasites have properties to cause disease and infection. Using the means of transmission, infectious diseases are classified as contact, food or water, airborne, vector-borne or perinatal (see Table 1.1).

TABLE 1.1: Means of transmission.

(see http://bioterrorism.slu.edu/bt/products/bio_epi/scripts/mod3.pdf)

Means	Descriptions
Contact	Requires direct or indirect contact (fomite, blood, or body fluid)
Food or Water	Ingestion of contaminated food or water
Airborne	Inhalation of contaminated air
Vector-borne	Dependent on biology of vector as well as infectivity of organism
Perinatal	Similar to contact infection, however, the contact may occur inutero or during delivery

A set of parameters to express disease transmission and disease control measures are presented in this Section.

1.2.1 Parameters for infectious disease transmission

1.2.1.1 Time lines

The time lines of infection within the host can be described in terms of the dynamics of infectiousness and disease (see Figure 1.2).

- Latent period is the time interval from infection to development of infectiousness.
- Incubation period is the time interval between infection to the development of clinical disease (i.e. the appearance of signs and symptoms of the diseases).
- Infectious period is the time during which the host can infect another susceptible host or vector they come in contact with. The infectious period may not necessarily be associated with the presence of clinical signs or symptoms (e.g. AIDS).

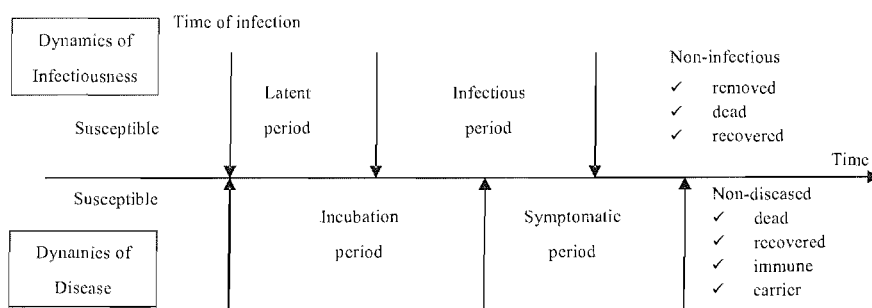


FIGURE 1.2: Time lines of infectious diseases
(Halloran 1998)

The latent period reflects the pathogen's perspective and the incubation period reflects the host's perspective. These two time periods are seldom the same (i.e. humans may become infectious to other humans before they show symptoms themselves) and the latent period can be longer or shorter than the incubation period.

1.2.1.2 Transmission probability

The transmission probability of infection refers to the chance that there is a successful infection from one host to another. The secondary attack rate (SAR) is a measure of transmission probability and is defined as the ratio of individuals who develop an infection to the total number of susceptible individuals. It is calculated by identifying the infective hosts, tracking which healthy hosts come in contact with them, and then noting which become infective as well. SAR is a static figure that is averaged over the entire epidemic (Halloran 1998).

1.2.1.3 Basic reproductive number

The basic reproductive number (R_0) is a measure of infectivity. It is defined as “the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible” (Anderson & May 1992, pp 17). In person-to-person transmission, R_0 is a measure of the ability of a disease agent to cause infection. The meaning of R_0 is the expected number of new infectious hosts that one infectious host will produce during his or her period of infectiousness in a large population that is completely susceptible. R_0 does not include new cases produced by the secondary cases, or further down the chain. R_0 equals the product of the number of contacts per unit time c , the transmission probability per contact p and the duration of infectiousness d (Halloran 1998).

$$R_0 = cpd \tag{1.1}$$

R_0 is taken as the threshold quantity that determines whether an epidemic will occur or not. In a population, if $R_0 = 1$, each person who becomes infected with the disease will pass it on to exactly one other person (on average). Thus, the infection is maintained in the population without the need for external inputs. This is said to be endemic; if

$R_0 > 1$, the infection will grow in a population, and there is some possibility of a major epidemic; if $R_0 < 1$, disease transmission will terminate.

Generally, the larger the value of R_0 , the harder it is to control the epidemic. R_0 may vary considerably not only for different infectious diseases but also for the same disease in different populations or during different time periods. For example, R_0 of measles was between 13.7 and 18.0 in the infant population in England and Wales between 1944 and 1979, and the value was about 12.5 in the infant population in parts of North America between 1919 and 1928 (Anderson & May 1982). Another example is that the R_0 for mumps varied between 4.3 and 7.1 (Anderson & May 1982). Especially for some ‘new’ infectious diseases, the value of R_0 varies significantly, such as SARS, Figure 1.3 is from Bauch et al. (2005) which shows R_0 estimated from various studies for SARS and other diseases.

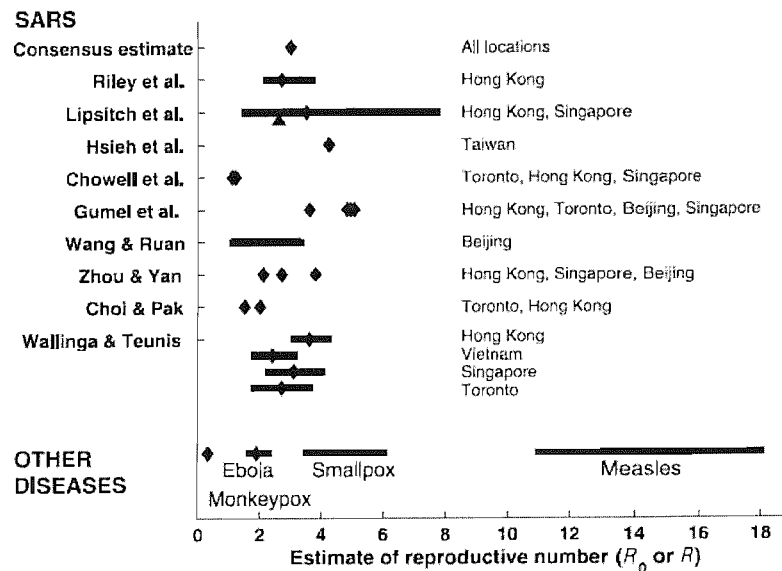


FIGURE 1.3: R_0 estimated from various studies for SARS and other diseases. (Bauch et al. 2005)

1.2.1.4 Proportion of infections before the onset of symptoms

The proportion of infections that occur before the onset of symptoms is defined to be θ (Riley & Ferguson 2006). For interventions that rely on the recognition of symptoms to reduce transmission, such as case isolation and contact tracing, the smaller θ , the more likely the intervention to be successful.

1.2.1.5 Virulence

Virulence refers to the severity of the disease after infection occurs. It is measured by the case fatality rate or the proportion of clinical cases that develop severe disease.

1.2.2 Control measures

All control measures should try to control the infectious individual, control the susceptible individual or prevent the contact between them. The basic control measures are vaccination, isolation, quarantine and so on. The target population for control are symptomatic individuals and the individuals who have had contact with the symptomatic cases. Most of the control measures may be undertaken voluntarily or compelled by public health authorities.

It is generally assumed that an individual is protected from an infectious disease when he or she has the requisite disease immunity. Immunity can be present due to either genetic makeup, vaccination or past exposure. Vaccination is the primary policy of public health departments in combating the spread of infectious diseases. Since it is hard (if not impossible) to control all susceptible individuals, tracing the individuals who have contacts with infectious individuals can be an alternative. Contact tracing is one form of target control focusing on the potential cases of the next generation. Contact tracing is

a successful strategy when the number of infectious cases is low (Eames & Keeling 2003). Ring vaccination, one form of spatially explicit contact tracing, controls an outbreak by vaccinating and monitoring a ring of people around each infected individual. In practice, this means the vaccination of all susceptible individuals in a prescribed area around an outbreak of an infectious disease (Kretzschmar et al. 2004).

Both the isolation and quarantine aim to control exposure to infected or potentially infected persons. The two measures differ in that isolation applies to persons who are known to have an illness, and quarantine applies to those who have been exposed to an illness but who may or may not become ill (see <http://www.cdc.gov/ncidod/dq/facts.htm>). The movement of people who under controls are restricted to stop the spread of the disease. People in isolation or quarantine may be cared for in their homes, in hospitals, or in designated health care facilities.

1.3 Thesis outline

Chapters 2 and 3 provide a literature review on the modelling of infectious disease transmission and humans' dynamics in space and time, respectively. The concept of an individual space-time activity-based model (ISTAM) is presented in Chapter 4, and its implementation in Chapter 5. Activity bundle (AB), as a key concept in ISTAM, is introduced. Within-AB and between-AB simulations are also discussed. Chapter 6 provides one example case study of ISTAM: application to a hypothetical influenza outbreak in the campus of the University of Southampton. Raster-based AB simulation and its parameters are explored in Chapter 7. Another example case study of ISTAM, a hypothetical influenza outbreak in the city of Eemnes, is presented in Chapter 8. Discussions and conclusions are made in Chapter 9 and 10, respectively.

Chapter 2

Modelling Infectious Disease

Transmission

In this chapter, mathematical models for infectious disease transmission are introduced. Then, the definition, application and implementation of individual-based models are elaborated. Sections 3 and 4 provide a discussion of simulation methods at the individual level in terms of contact patterns and spatial effects, respectively.

2.1 Mathematical models

Traditional mathematical models of disease transmission are population-based and expressed by deterministic mathematical equations (Anderson & May 1992). From the most classic mathematical model, susceptible-infectious-recovered (SIR) (Kermack & McKendrick 1927), many mathematical models such as susceptible-exposed-infectious-recovered (SEIR) and susceptible-infectious-susceptible (SIS) have been presented. They

all share the same basic idea as the SIR model (separate the population into several compartments and use differential equations to express the flow patterns between the compartments) (Anderson & May 1992).

2.1.1 SIR models

Based on the following basic assumptions: (1) population size is large and constant (no birth, death, immigration or emigration); (2) no latency (the latency period is omitted); (3) homogeneous mixing (i.e., each pair of individuals has equal probability of coming into contact with one another), the SIR (susceptible-infected-removed) models divide the population into three fundamental compartments as: (1) susceptible (S), that is, subjects susceptible to contracting the infection or those who may catch the disease but currently are not infected; (2) infected (I), that is, subjects that have already contracted the infection, to the stage of full development or incubation, so they are vectors of contagion or those who are infected with the disease and currently contagious; (3) removed or recovered (R), that is, subjects that have died or that have developed immunity to the contagion and, therefore, are no longer able to infect or to contract infection.

Then the expressions of the SIR model are:

$$\partial S(t)/\partial t = -\beta S(t)I(t), S(0) = S_0 \geq 0 \quad (2.1)$$

$$\partial I(t)/\partial t = \beta S(t)I(t) - \alpha I(t), I(0) = I_0 \geq 0 \quad (2.2)$$

$$\partial R(t)/\partial t = \alpha I(t), R(0) = R_0 \geq 0 \quad (2.3)$$

(where β is the rate of new infections in the unit time and α is the rate of mortality or immunity in the unit time. $S(t)$, $I(t)$, and $R(t)$ are the numbers in these classes, so that $S(t) + I(t) + R(t) = N$, N is the total number of individuals in the population).

2.1.2 SEIR models

The SEIR (susceptible-exposed-infectious-recovered) models include four compartments represented by the susceptibles (S), exposeds (E) (or infecteds), infectious (I), and recovered or immunized (R) (Anderson & May 1992). The SEIR model accounts for the infecteds (E). This means that the model accounts for the latent period of the disease. This model is needed when infected individuals (exposed) go through a latent period before becoming infectious.

$$\partial S(t)/\partial t = -\beta S(t)I(t) \quad (2.4)$$

$$\partial E(t)/\partial t = \beta S(t)I(t) - \alpha E(t) \quad (2.5)$$

$$\partial I(t)/\partial t = \alpha E(t) - \gamma I(t) \quad (2.6)$$

$$\partial R(t)/\partial t = \gamma I(t) \quad (2.7)$$

(where β is the rate of infection per unit time, α is the rate at which an infected individual becomes infectious per unit time, γ is the rate at which an infectious individual recovers per unit time. $S(t)$, $I(t)$, $E(t)$ and $R(t)$ are the numbers in these classes, so that $S(t) + E(t) + I(t) + R(t) = N$, N is the total number of individuals in the population).

2.1.3 SIS models

The SIS (susceptible-infected-susceptible) models describe diseases for which the infections do not confer any long lasting immunity. Thus, infected individuals do not have a recovered state and become susceptible again after infection. Two compartments are represented by the susceptibles (S) and infected (I).

$$\partial S(t)/\partial t = -\partial I(t)/\partial t \quad (2.8)$$

(where $S(t) + I(t) = N$, N is the total number of individuals in the population).

2.1.4 Reed-Frost epidemic models

Lowell Reed and Wade H. Frost, developed a mathematical model to describe how diseases spread through populations in the 1920s. The model known as the ‘Reed-Frost Epidemic Model’ is described by Abbey (1952) as follows:

Model assumptions:

- The infection is spread directly from infected individuals to susceptible individuals by a certain kind of contact (‘adequate’ or ‘effective’ contact) and in no other way.
- Any susceptible individuals in the group, after such contact with an infectious person in a given period, will develop the infection and will be infectious to others only within the following time period, after which he or she is wholly immune.
- Each individual has a fixed probability of coming into adequate contact with any other specified individual in the group within one time interval, and this probability is the same for every member of the group.
- The individuals are wholly segregated from others outside the group.
- These conditions remain constant during the epidemic.

In other words, each individual in the study population is in one of three possible states during any time period. These are: susceptible state, infectious state and immune state. Susceptible individuals change to infectious if and only if they come into ‘effective contact’ with infectious individuals. Infectious state always changes to the immune state in subsequent time periods. Immune individuals remain immune and are sometimes omitted from consideration. The notation for the model is as follows:

$$I_{t+1} = S_t(1 - q^{I_t}) \quad (2.9)$$

(where p is the probability of an individual making effective contact, $q = 1 - p$, t is the time period, I_{t+1} is the number of infectious cases in time period $t + 1$, S_t is the number of susceptible individuals in time period t).

The Reed-Frost model is population-based. In fact, the SIR model can be simplified to the Reed-Frost model if the infectious period is assumed to be constant.

2.1.5 Problems of mathematical models

Recently, mathematical models have been devised that divide the heterogeneous population into several sub populations or groups. This division can be based on the purpose of the model, the specific disease being modelled, the process of transmission, the age structure of the population or on the environmental heterogeneity and spatial structure (Hethcote 2000). One example is the meta-population models (Hanski 1998) which increase heterogeneity in the population by dividing the population amongst a number of distinct patches, as well as a coupling term allowing influence proportional to the magnitude of infection at the other sites. However, the conventional ‘mixing homogeneously’ assumption is still retained within patches and there is no *a priori* way of assigning the coupling magnitudes between patches (Rhodes & Anderson 1997). This makes the stochastic property of the transmission process difficult to model well.

Another problem of mathematical models are the rapid growth of the mathematical complexity of the models used to describe in sufficient detail various phenomena and the difficulty in solving them in an analytical form (Bagni et al. 2002). “The complexity of differential equations increases exponentially as the complexity of behavior increases” (Bonabeau 2002, pp 7281).

2.2 Individual-based models

Computer simulation models emerged later than mathematical models because the former rely on evolving computer technology. With sufficient knowledge and data and by faithful modelling and model space exploration, computer simulation can provide detailed insights on complex and realistic social systems (Prietula et al. 1998). The common steps for computer simulation are: (1) set up a computer model to mimic part of the real world; (2) use random number generators to simulate the random components of the model; (3) simulate repeatedly to observe outcomes on average, and the variability in outcomes. Individual-based models (IBM), one sub category of computer simulation, have been applied increasingly due to the advance of computer technology.

2.2.1 Definition of individual-based models

IBMs are simulations based on the global consequences of local interactions of members of a population (see <http://www.red3d.com/cwr/ibm.html>). IBMs originate in mathematics and computer science, especially in artificial intelligence, and are also affiliated with complex adaptive systems. IBMs provide a means to connect interactions between individuals and environmental and other influences, taking account of differences between individuals (Gimblett 2002). IBMs can be used to estimate phenomena at different organizational levels from the actions of individuals.

An agent-based model (ABM) is defined as “a system is modeled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions on the basis of a set of rules” (Bonabeau 2002, pp 7280). IBMs and ABMs are similar concepts used in different fields. IBMs are applied widely in ecology and biology. Important related literatures are DeAngelis and Gross (1992) and Grimm (1999, 2005). ABMs are used commonly in social science and computer science (Epstein & Axtell 1996). The ABMs are distinguished from IBMs by the fact that

'agents' in ABMs are autonomous and have flexible behavior (e.g., reactive, proactive and social) (Wooldridge 2002) and 'individuals' in IBMs do not necessarily do so. In this specific sense, IBMs are used in this thesis.

An individual person is not necessarily an 'individual' within an IBM. It may be convenient in concept or computation to "break up an individual into the different stages of its life cycle" or to take a group of individuals to be an 'individual' (Ginot et al. 2002, pp 24).

Overlap exists between IBMs and cellular automata (CA). CAs are similar to spatially explicit, grid-based, immobile individual-based models. But CAs are homogeneous and dense, and a grid-based IBM might occupy only a few grid cells, and more than one distinct type of individual may live on the same grid (see <http://www.red3d.com/cwr/ibm.html>).

IBMs have been applied to many fields, such as to simulate vehicles and pedestrians in traffic, people in crowds, artificial characters in computer games, agents in financial markets, and humans and machines in the battlefields. In the context of spatial or geographic models, IBMs are popular for urban study, transportation, tourism and so on.

2.2.2 General benefits and limitations of individual-based models

Major advantages of IBMs are the possibility to incorporate individual behaviors and micro processes in the models. The heterogeneity of information can be represented fully in the model and maintained during simulation. The output will consequently contain a great variety of information about general and specific conditions at the micro level; information that can be aggregated easily to the level suitable for answering research and applied questions. This facilitates a detailed analysis of micro process or sequences of individual's actions and provides opportunities for a more thorough understanding

of the mechanisms behind the macro processes and of the consequences at aggregate or disaggregate levels (Fotheringham & Wegener 2000).

Generally, the benefits of application of IBMs (or ABMs) (Bonabeau 2002) are as follows:

1. IBMs can capture emergent phenomena. Emergent phenomena result from the interactions of individual entities. For a complex system, the whole is more than the sum of its parts because of the interactions between the parts.
2. IBMs provide a natural description of a system.
3. IBMs are flexible. For example, the category of agents, the complexity of the agents and the levels of description and aggregation can be changed.
4. IBMs can contain both deterministic and stochastic relations (Holm et al. 2000). Deterministic relations are the unavoidable rules or strong logical or structural constraints. Deterministic relations produce the same result in each simulation if the initial conditions are same. Stochastic relations are the unknown part of the study, so the outcome of each simulation is different even with the same initial conditions. The simulation is executed many times with different random seeds to obtain the distribution of the results.

IBMs are especially useful within the following context (Bonabeau 2002):

1. When the interactions between the agents are complex, nonlinear, discontinuous, or discrete.
2. When space is crucial and the agents's positions are not fixed.
3. When the population is heterogeneous, that is, when each individual is (potentially) different.
4. When the topology of the interactions is heterogeneous and complex.

5. When the agents exhibit complex behavior, including learning and adaptation.
6. When the model includes soft factors that are difficult to quantify, calibrate, and sometimes justify (examples are human intelligence).

Obvious limitations of IBMs are as follows (Nagel & Marchal 2003):

1. IBMs require many simulations to evaluate any particular situation as it is based upon an underlying stochastic model. Simulating the behavior of all of the units can be extremely computationally intensive and, therefore, time consuming (Bonabeau 2002).
2. Knowledge at the micro level, for example, about human behavior, may not be sufficient. IBMs typically require assumptions about what aspects of behavior are important and what can be ignored.
3. The necessary input data for all the details, in many circumstances, are not available.

2.2.3 Individual-based models for the simulation of infectious disease transmission

For infectious disease transmission modelling, the individual-based model is not a new concept. Elveback et al. (1976) presented a highly agent-specific stochastic simulation epidemic model. The population in this model was highly structured, allowing for five age groups and for sub-group mixing in families, neighbourhoods, schools, and preschool playgroups as well as total community mixing. Partly due to the lack of both data and advanced computation ability, the potential of IBMs was not applied widely (Koopman 2002).

The 21st century has seen a boom of the application of IBMs for the simulation of infectious disease transmission from different disciplines, such as computer science (Aschwanden 2004, Carley et al. 2004, Brouwers 2005), geography (Bian 2004, Bian & Liebner 2005, Dibble & Feldman 2004), epidemiology (Eichner 2003, Ferguson et al. 2003, Ferguson et al. 2005, Ferguson et al. 2006, Longini et al. 2005, Riley & Ferguson 2006) and interdisciplinary fields (Eubank 2002, Eubank et al. 2004, Halloran et al. 2002, Huang et al. 2004). Many advantages can be achieved by IBMs simulation, but the most important advantage is that it can consider the heterogeneity of both individuals and environment, and also the stochastic essence of infectious disease transmission. IBMs can express explicitly the differences between individuals in terms of the attributes that influence the process of disease transmission such as physical, social, economic and environmental characteristics. For example, age, gender, occupation and lifestyle variables all contribute to the subsequent disease experience of an individual (Elliott et al. 2000). The interaction between individuals, which is one of the key components determining infectious disease transmission, can be expressed explicitly in the model. By IBMs, the time lines of infectious diseases, the infection probability, the demographic dynamics that determine the contact pattern, and the occasional imported infection, can all be considered as stochastic.

For infectious disease transmission, individual-based models have the following advantages over the traditional population-based models (Eubank 2002): (1) subpopulation is based on a few demographics while individuals can carry many demographics; (2) subpopulation mixing rate is unknown while individual contact rates can be estimated independently; (3) reproductive number is not directly observable for traditional models while it can emerge from the simulation of individual-based transmission; (4) the complex interaction between individuals (or ‘agents’) is out of the reach of pure mathematical methods (Bonabeau 2002, Epstein & Axtell 1996).

2.3 Contact patterns

Transmission is the most fundamental characteristic of infectious diseases distinguishing them from non-infectious diseases. The occurrence of infection in individuals depends on the occurrence of that disease in other members of the population (Halloran 1998). This dependence of disease events in infectious diseases was called ‘dependent happening’ by Sir Ronald Ross (1916), that is, implying the need for contact between susceptible and infectious individuals, even if indirect. Infectious diseases with different transmission modes require different kinds of contact between individuals. For example, some diseases, such as AIDS, require blood-to-blood or sexual contacts. In the context of this thesis, interest is confined to those airborne diseases transmitted by droplet, such as influenza and measles.

Airborne diseases (like many other infectious diseases) transmitted between humans are spread through populations via the networks formed by physical contacts amongst individuals (Meyers et al. 2005). Such networks channel the transmission of infections through the host population (Wallinga et al. 1999). Although the network assumption is radically simplifying and unrealistic (Koopman 2004), it is a great advance on mathematical models which assume individuals mix homogeneously and every individual has the same probability of coming into contact with another. A model of the contact network between individuals can be an important component in a model of an infection transmission system. Significant benefits of modelling the contact pattern are:

- The ability to assess the effects of the network on infectious disease propagation at the whole population level (Yahja 2002).
- The ability to assess how the location of individuals with different risks within the contact structure alters their importance on the population level of infection (Koopman 2004).
- The ability to evaluate specific disease control measures such as contact tracing (Eames & Keeling 2003, Read & Keeling 2003, Eichner 2003).

For infectious diseases that are transmitted between individuals, changing the pattern of connections between exposed and unexposed individuals can often affect infection levels more than changing the exposure status of individuals (Koopman & Lynch 1999).

IBMs are bottom-up models that require a profound understanding of the low-level processes and elements to generate aggregate system behavior by simulating the individual entities in the system (Epstein & Axtell 1996). Understanding of the transmission process at individual level is crucial for building an IBM. Spatial adjacency and social relations are the two most prevalent bases for building a contact network at the individual level. Spatial adjacency assumes that individuals come into contact with other individuals who are physically adjacent (e.g., based on their residential addresses). Examples of models that utilize spatial adjacency are the lattice model and CA model (Rhodes & Anderson 1997, Sirakoulis et al. 2000, Fuks & Lawniczak 2001). Individuals perhaps have greater probability to come into contact with the people who are spatially adjacent but individuals can also come into contact with other individuals over a wide range of distances. Firstly, there is no accepted positive relation between the probability of contact and the distance between residential addresses. Secondly, residential addresses themselves reflect only partially the spatial location of the activities of most of the population. In reality, humans are dynamic not static, as reflected in their social activities. Therefore, geographical distance (e.g., spatial adjacency or distance between residences) alone is inadequate to measure infection risk amongst humans (Eames & Keeling 2003, Bian 2004). The latter basis for building a contact network assumes that individuals come into contact with their social relations (e.g., family, colleagues or friends). Social relations alone are also inadequate to build the contact network. Firstly, individuals do not make physical contact with all their social relations (people can retain their social relations by email or telephone). Secondly, individuals also come into contact with other individuals (not social relations) by chance (e.g., strangers on the public bus or tube; individuals sit in proximity in the theatre). In conclusion, the contacts between individuals are functions of social behaviors while infection *per se* is a physical process. The consideration of both social and physical dynamics is, thus, a prerequisite for modelling the contact network.

Age is an important variable that indicates the risk of acquiring and transmitting a disease (Anderson & May 1992). Besides age-specific infectiousness and susceptibility, age-specific contact rate is another important concept. The study by Pool and Kochen (1978) showed that contact patterns tend towards being assortative with respect to age. It is obvious that people have more contacts with people similar to their own ages.

The population consists of groups that have frequent contacts within groups (local contacts) and relatively rare contacts between groups (global contacts) (Wallinga et al. 1999). Separating the population into smaller sub populations is a way of investigating the rôle of global and local contacts. Meta-population models are based on this idea. The point is: individuals do not belong to only one group; they can belong to different groups at the same time. For example, one person belongs to his or her family *and* his or her company or school. And sometimes the separation between local and global contacts is not obvious.

2.4 Spatial effects

There are two types of diffusion: expansion diffusion and relocation diffusion (Cliff et al. 1981, Kuby et al. 2004). The difference is that during the former process, the item being diffused remains, and often intensified, in the originating region, but new areas are also occupied by the item in subsequent time periods; while during the latter process, the items being diffused leave the areas where they originated as they move to new areas. Cliff et al. (1981) took the contagious spread (that places near the origin are usually affected first) and hierarchical spread (the spread that occurs through an ordered sequence of classes or places) as two ways in which expansion diffusion occurs. Kuby et al. (2004) summarized that the above two types of spread have two spatial regularities which can be exhibited by all diffusion processes, and named them as contagious effect and hierarchical effect.

“The spread of disease is unavoidably spatial” (Holmes 1997, pp 111), and yet most traditional transmission models are non-spatial. Traditional geographical models treat the dispersion of disease like wildfire or are wavelike (Haggett 2000). In reality, the spread of infectious disease amongst humans is different from the spread of wildfire or disease amongst plants or vegetables because humans are not static in terms of their spatial locations. The movement patterns of individuals, together with infection processes and the disease evolution process within an individual make the spread of infectious disease amongst humans exhibit complex spatial effects at different regions, different spatial levels and different infectious diseases. In fact, the spatial spread of infectious diseases amongst humans can not be classified to be expansion or relocation diffusion and it can exhibit both contagious effect and hierarchical effect as discussed above. As reviewed by Bian (2004), the disease spreading process can be separated into both local transmission and long-distance dispersion (Holmes 1997). Local transmission exhibits the dynamics of expansion diffusion and long-distance dispersion exhibits the dynamics of relocation diffusion. So the whole dynamics may involve the integration of both of them. Contagious diffusion is valid at the micro-scale. At the macro-scale, such as the regional level, the hierarchical diffusion pattern may be exhibited. In conclusion, infectious disease transmission is a dynamic process involving interactions between people both spatially and temporally, and different space-time dynamics are exhibited at different space-time levels such as the global level, regional level, city level, community level and individual level.

The geographical space represented in an individual level simulation needs to be simplified to express humans’ daily activities well. The humans’ activities are restricted by both the physical landscape and human-made structures so it is not appropriate to consider the space where the activities of humans occur as a Euclidean continuous space without constraints. Rather, the simulation should be confined to the space where transmission of a specific disease is most likely to occur. Contacts between individuals, along with other conditions, can be used to delimit such a space. It needs to be pointed out that a likely infection space is specific to one infectious disease. For example, a high mobility pavement can be part of the likely infection space if the targeted infectious

disease has a shorter attack duration but can not be part of the likely infection space if the targeted infectious disease has a longer attack duration.

The probability of contact varies throughout the most likely infection space. For example, the probability of effective contact amongst elementary school children in a classroom can differ from amongst the same children on a soccer field or from the same number of adults in an office setting. As Koopman pointed out one needs to consider “how often and when individuals go to specific places where they might become infected from other individuals” (Koopman et al. 2000, pp 317) and “geographic or social locations where contacts are made could be used to reflect many different types of contact” (Koopman & Lynch 1999, pp 1171). To express the variation in the probability of contact, both the dynamics of individuals and the space’s physical condition need to be considered.

2.5 Summary

This chapter provides a review about the mathematical models and individual-based models in terms of their application to the transmission of infectious disease. At the individual level, the contact patterns and spatial effects are discussed. This review can be taken as the traditional theory background and research methods which is part of the foundation of this PhD research. Besides, theories and research methods from other fields were imported as another part of the foundation which are presented in the next chapter.

Chapter 3

Modelling Humans' Space-Time

Dynamics

In this chapter, Section 3.1 introduces the classic framework for humans' space-time dynamics: time geography. Then a concept from time geography, the activity bundle, is elaborated in the following section. Section 3.3 describes between-AB simulation, while Section 3.4 describes within-AB simulation and the properties of humans' space-time dynamics at fine scales. Sections 3.5 and 3.6 give a brief introduction of two related techniques: geographical information systems (GIS) and network theory. In this research, the former is used for spatial display and spatial analysis and the latter is used to analyse and display the physical contacts between individuals.

3.1 Time geography

Time geography describes and defines how individuals perform activities under the constraints of time and space, and these activities are conditioned not only by physical but

also by social constraints that prevent individuals from doing certain things but enable them to do others (Hägerstrand 1975). It introduces a conceptual framework for analyzing social micro-level interactions in time-space. Time geography offers a set of visual tools. The environment in which activities are performed is taken as a cube, with the two horizontal axes representing the two spatial dimensions of x and y , and the vertical axis representing the time dimension t . Hägerstrand has termed this an 'aquarium', within which are placed time geography elements such as lifelines, stations, bundles, domains and prisms (see Figure 3.1). A lifeline represents the path taken by an indi-

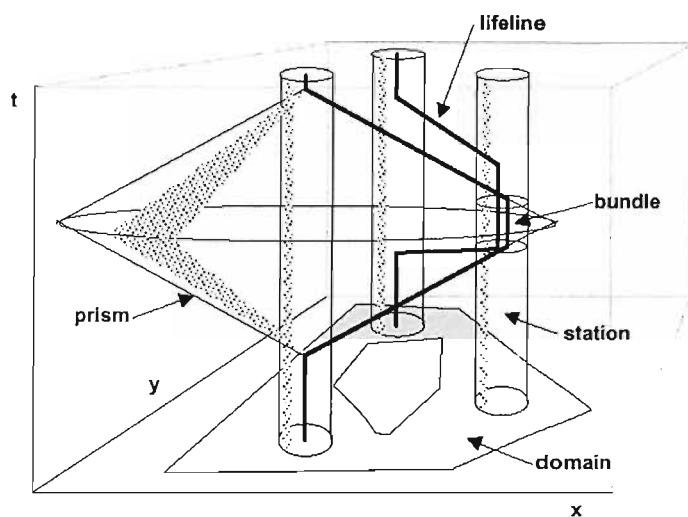


FIGURE 3.1: Elements of time geography
(Moore et al. 2003)

vidual in time-space. Stations are spaces when individuals stay some time in the path. If individuals congregate with other individuals, bundles will be created (Miller 2005). Domain is a concept to describe the physical boundaries within which an individual's movements are constrained. Another concept of prism represents the total area of space reachable by an individual during the available time.

Individuals are not uniformly distributed in space and time. They move through space as part of their daily lives or because of migration. During these movements, indexed by time, individuals will travel through numerous exposure surfaces. The characteristics of individuals, such as age, gender, occupation and generic factors, lifestyle variables such

as smoking and diet, and the lifelong dose due to an exposure of interest, all contribute to the subsequent disease experience of an individual (Elliott et al. 2000). The activity pattern can be obtained from tracking data which normally consist of a sequence of tuples (x, y, t) ordered by t , indicating the location (x, y) of a moving individual at intervals of time denoted by t . Time geography can provide the necessary intellectual framework for the time-space lag effect because of human mobility: many diseases are believed to show symptoms after periods of latency, ranging from seconds to decades (Schaerstrom 1996).

Individual space-time activity data can be collected by questionnaire. One data collection example is to ask respondents to recall their mobility over a 24-hour period, and this day should not be a holiday or a special day of the week (Meade et al. 1988). Since 1978 in Finland the national health institute has been sending out annually a postal questionnaire to a random sample (5000) of the Finnish adult population between the ages of 15 and 64. This survey is designed to elicit information on health behavior, such as dietary habits, smoking, alcohol consumption, physical exercise and health status, together with a set of individual background factors describing the respondents (Antikainen 1999), and importantly locational coordinates at the spatial resolution of one metre.

The advance of information technology makes it possible to obtain data on individual's daily movements. A location based service (LBS), is an information service that exploits real-time positioning in individual level decision-making. LBS can provide abundant tracking data on the daily movements of people. This means that it is possible to analyze the space-time characteristics of an individual's recent movements to build epidemiological models at the individual level.

3.2 Concept of activity bundle and its relation to infectious disease transmission

In the movie *Turn Left Turn Right* (see http://en.wikipedia.org/wiki/Turn_Left,_Turn_Right, 2003), two lovers live in adjoining apartment buildings separated by one wall. They never meet because every day one turns right and the other turns left. For infectious disease transmission, it means these two lovers have no contact even though they live right next door because their daily activities do not overlay in space and time. From the view of time geography, airborne disease infection can occur only inside and during the bundle. If lifelines of every individual in the population are available, their potential interaction within a bundle can be aggregated. Thus, it is possible to simulate the transmission of specific infectious diseases (Forer 2002).

Although from the time of Hippocrates, it has been known that location can influence disease and health, the importance of 'place', 'location' or 'venue' in the study of transmission has not been recognized widely. Some studies have been undertaken on the probabilities of infection at particular 'places' such as within an airplane (Mangili & Gendreau 2005), but generally, few studies have considered the rôles that 'place' plays in the disease spreading process (i.e., transmission between 'places' by the movements of humans). Klovdahl et al. (2001) discussed the importance of 'place' in a study of a community-level tuberculosis outbreak. They suggested that 'person-oriented' methods should be supplemented by 'place-oriented' methods. Koopman stated that "data on how often and when individuals go to specific places where they might get infected from other individuals can be used to describe contact patterns by making assumptions or by gathering data regarding the nature of contact at these different places" (Koopman 2004, pp 317). It is well known that commonly people repeat similar daily activities and certain activities occur at certain 'places'. Thus, the patterns of people present over time also exist for 'place'. Based on the simulation model TRANSIMS (see <http://transims.tsasa.lanl.gov/>), Eubank et al. (2004) displayed the changing number

of people present within locations such as the home, workplace, school and so on during a whole day.

To model humans' contacts which could cause infections at fine spatial scales, the concept of activity bundle (AB) is defined as a semantic space where contact probability varies as a function of the dynamics of humans inside. To avoid confusion, the following thesis uses 'activity bundle' to replace 'place', 'location' or 'venue'.

The process of infectious disease transmission and its corresponding study can be divided into two stages: the occurrence of an AB in space-time and the occurrence of infections inside and during the AB. The former involves studies related with human activity patterns (between-AB) while the latter involves studies of dynamics of humans at the fine-scale (within-AB).

3.3 Between-AB simulation

Individuals' movements between ABs are restricted by both locational supply and individuals' choices. The subset of all locations within which an individual has direct contact as a result of his or her daily activities is defined as an activity space (Golledge & Stimson 1997). As stated by Golledge and Stimson, the activity space of a typical individual is composed of three parts:

- Movements within and near the home.
- Movements to and from regular activity locations, such as journeys to work, to shop, to socialize and so on.
- Movements in and around the locations where those activities occur.

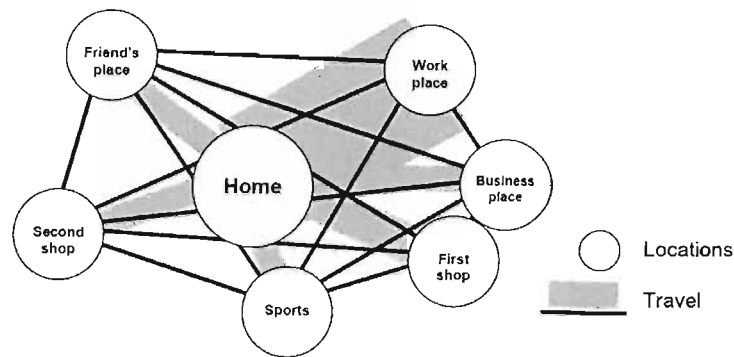


FIGURE 3.2: Simplified activity space.
(adopted from Schönfelder and Axhausen (2002) and Maier et al. (1977))

Figure 3.2 represents one individual's simplified activity space. Schönfelder and Axhausen (2002; 2003) described and discussed three methods for describing individual activity spaces: (1) confidence ellipses, (2) kernel densities and (3) minimum spanning trees (networks). A similar concept of 'personal network for usual places' was presented by Flamm and Kaufmann (2006) to identify a certain number of daily life centers. A large amount of research has been undertaken on individuals' daily activities at between-AB level, especially in the fields of transportation and urban planning. Both utility maximisation (Bowman & Ben-Akiva 2001) and rule-based approaches (Pendyla & Kitamura 1998, Arentze & Timmermans 2007) have been used to describe how households and individuals decide about their daily 'lifeline' including decisions about which activities to undertake, at what locations, at what times, with which persons and how to travel to those locations. These decisions are made subject to space-time constraints as defined by time-geography (Hägerstrand 1975). Activity-based models at this level have been applied widely in transportation (Miller et al. 2004, Batty 2005a, Batty 2005b, Ettema 2006). These models provide a starting point for modelling individuals' presence in ABs, in which diseases are transmitted.

3.4 Within-AB simulation and properties of humans' space-time dynamics

Although it is well known that the transmission probabilities in different types of places may be different, factors such as environmental conditions and the spatial locations of individuals that affect transmission probabilities at particular places are poorly understood (Koopman 2004). It is natural to begin from the infection process at fine level: within-AB level.

In approximate terms, the infection process of airborne infectious disease is as follows: infectious individuals shed infectious substance to the air, the substance stays and/or diffuses up to a certain distance and until a certain duration according to the AB's physical condition (such as ventilation and sanitation conditions), and then if susceptible individuals happen to be within the 'contaminated' air and absorb the infectious substance, then infection is possible.

In the above process, the space-time dynamics of the infectious substance depends on airflow rates, heating and cooling, and the architectural properties of the AB. The more challenging question here is, at the within-AB level, how best to model humans' space-time interactions (or to be more precise, the space-time relations between susceptible individuals and infectious individuals or the infectious substance) which are crucial for infection.

Research on the dynamics of humans inside a small space can be traced to 'room geography' (Jackle et al. 1976) which studied how individuals distribute across a specific small area. But research at fine-scales has been slow due to the lack of systematic data and corresponding tools and methods (Batty et al. 2003). Recently, IBMs and object-oriented technology have facilitated further study. One example is the pedestrian model proposed by Batty and his colleagues (Batty 2003, Batty et al. 2003, Batty 2005a) to

study how behavior emerges from the aggregated interactions between fine-scale objects. Applications include traffic modelling, disaster evacuation and the spreading of infectious diseases.

Sommer (1983) claimed that: “when people interact, their spacing is regulated by many factors, including their relationship, amount of previous contact, their backgrounds, the activities in which they are engaged, and environmental factors including the size and layout of the room, noise, lighting levels, and other sources of background stimulation”. A similar statement from Batty (2003) is “...behavior in human systems is not simply determined by preferences, intentions, desires but by the environment which reflects the spatial or geometric structure in which the agents function as well as variability between agents, in terms of their intrinsic differences and the uncertainty that they have to deal with in making any response”. In this section, reviews are focused on factors such as the air and space in an AB, the distances between individuals and individuals' rôles in their activities.

3.4.1 Air within an AB

The work of Wells (1955), which assumes complete air mixing, is the basis for models such as the Mass Action model (Riley 1974), Riley, Murphy and Riley's model (Riley et al. 1978) and Gammaitoni and Nucci's model (Gammaitoni & Nucci 1997, Beggs et al. 2003). In the above models, the infectious substances are evenly distributed throughout the place so that the proximity of susceptible individuals to infectious individuals, and the duration of exposure are neglected. But in reality, the air in confined spaces is seldom mixed completely.

One example is provided by the EpiSims model (Eubank 2002) which maintains a list of individuals who are present and a disease load for every AB. Load here means the viral concentration in an individual or AB. The load grows or decays with time depending on

the specific infectious disease. Contamination by shedding from infectious individuals may be restricted to a small region near the infected person, or may spread to the entire location. Due to the lack of data for the proximity of people, Eubank et al. (2004) split a large location into sub-locations which are small enough to assume that all individuals within the same sub-location are within the distance for infection. One problem is that this split assumes that no infection can occur between individuals in adjacent sub-locations even when they are close enough for infection. Another problem is that the allocation of individuals to these sub-locations cannot be avoided.

3.4.2 Space within an AB

If internal physical information of an AB such as its geometry and spatial layout are available, the geographical space inside an AB can be represented using either (1) the raster data model (e.g., a rectangular or hexagonal grid) or (2) the vector data model (i.e., a continuous space involving the movement of individuals). Figure 3.3 is an example of the former and Figure 3.4 is an example of the latter. Further, the raster or vector space can be two or three dimensional.

One example of within-AB simulation was provided by Epstein et al. (2002). In this model, spaces of all ABs are structured as raster spaces and the whole day is discretized into a specific number of rounds. In each round and each social unit, one infectious agent interacts with one of their Moore neighbour agents. The locations of individuals within an AB change every day. Fewer contacts are assumed to occur at the workplace or school than at the home or hospital. The likelihood of an interaction resulting in a contact at home is 1.0 and at work is 0.3. These values were based on intuition instead of estimated from data. In fact, the cell occupied by an individual does not reflect the individuals' spatial location within an AB since there is no 'spatial distance' meaning between cells. It can be taken as a graph to describe possible contacts between individuals in different types of AB (with different weights such as the above values of 1.0 and 0.3).

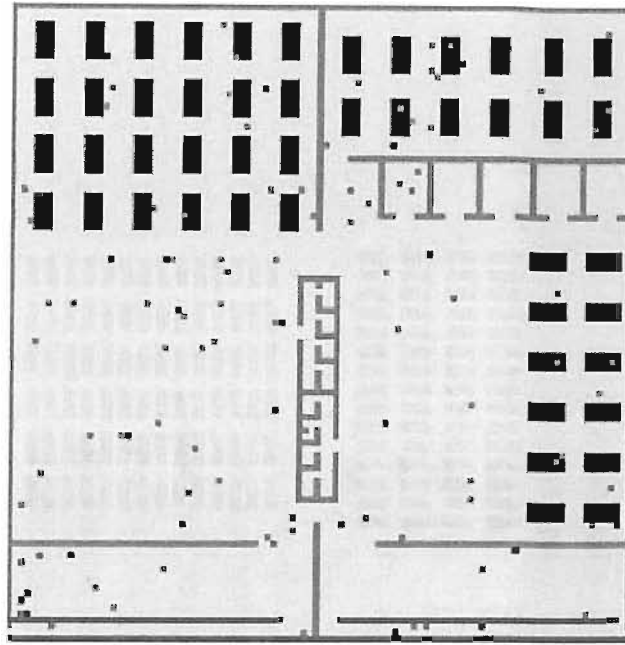


FIGURE 3.3: Raster space of a homeless shelter (The shelter contains mats and beds, with the mats shown in light-grey and the beds shown in dark-grey. People are shown by dots with different colours for different states).

(Patlolla et al. 2006)

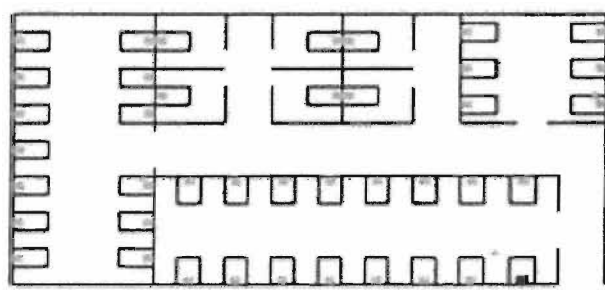


FIGURE 3.4: Vector space of a hospital ward (comprised of nine rooms visited by three nurses three times per day. Patients are shown as squares and nurses as triangles. Patients and nurses have different states shown by different colours).

(Jacquez et al. 2005)

Beltran, Quera and their colleagues (Quera et al. 2000, Beltran et al. 2006) presented the minimum-dissatisfaction (MD) model to study the spatial behavior of a small group of people interacting in a closed space. In the MD model, individuals move on a lattice in order to minimize their dissatisfaction which was defined as a function of the discrepancy between the real distance and the ideal or desired distance between agents. A computer program P-Space was implemented.

3.4.3 Distance between individuals within an AB

It is accepted that the possibility of infection for susceptible individuals increases with proximity to infectious individuals (Hutton et al. 1990, Noakes et al. 2006, Oppong et al. 2006). There are social rules about how close humans can approach each other. Hall (1966) identified four distances: intimate distance, personal distance, social distance and public distance (see Table 3.1). It needs to be pointed out that the values of the above distances vary between populations from different cultures, ages, genders and etc. Although the above reviews are general they provide a basis for within-AB simulation: focusing on the changing distances between individuals, specifically the circumstance when the distances are less than the distance required for infection.

TABLE 3.1: Distance rules.
(Hall 1966)

<i>Distance type</i>	<i>Value</i>	<i>Meaning</i>
Public	>3 m	The range of non-involvement
Social	3-1.5 m	The range in which most public interactions are observed. This is a comfortable distance for people who are standing in a group, but maybe not talking directly with one another. People sitting in chairs or gathered in a room will tend to prefer this distance
Personal	1.5-0.6 m	This is a protected area, where strangers would not be welcome. At its inner limit it holds other people "at arms length"
Intimate	<0.6 m	This range is reserved for lovers, family, small children and very close friends

3.4.4 Two kinds of contacts within an AB

The contacts between individuals are driven by social purposes or constrained by physical conditions or both while infection *per se* is a physical process. Some contacts are indispensable for undertaking some activities. For example, an individual who goes shopping will generally make contact with the salesperson when checking out. Some contacts, although not purposeful, occur because of the restriction of the environment. For example, in a crowded bus, passengers have to be close to each other when there is no spare space. The first type of contact is voluntary and determined wholly by the

individuals. Such contacts are explicit to the individuals, that is, individuals are aware of these contacts. The second type of contact is not voluntary and is determined only partially by the individuals. Individuals may be or may be not aware of such contacts.

As reviewed by Ellegård, concepts of 'series' and 'group' were pointed out by Jensen and Vestergaard (1979). A number of individuals "who occasionally meet at the same place where they intend to do the same thing constitute a series" and individuals in a 'group' "form a social entity which persists over a period of time and in which the individuals share a long-term purpose" (Ellegård 1999, pp 167). Most of the contacts between individuals within a series are the second type contacts while most of the contacts occurring within a group are the first type contacts.

3.4.5 Individual's rôle in their activities

It is believed that rôle performance and social involvement are influenced by the spatial environment (Baldassare 1978). Baldassare claimed that "individuals must obtain varying degrees of physical separation or closeness to others in order to engage in their activities" (Baldassare 1978, pp 46). Similarly, Gatrell claimed that "the spacing of individuals depends not only on how the environment has been designed but also on the task or nature of interaction that the individuals must perform" (Gatrell 1983, pp 75).

3.4.6 Properties of humans space-time dynamics within an AB

Three properties of humans' space-time dynamics within ABs which have been observed in reality are suggested to be considered when modelling the space-time dynamics of individuals within ABs: (1) individuals' static spatial distribution pattern, (2) individuals' movement pattern and (3) minimum distances between individuals. The spatial distribution pattern of individuals can be observed in most types of AB. One example

is provided by a restaurant: people sit in clusters which reflect the existence of different groups. Another example is provided by individuals visiting a library: they try to find an empty table and sit as far as possible from each other (Given & Leckie 2003). In some ABs, individuals are assumed to remain static during the simulation time unit, such as in a lecture room. In other ABs, movements must be considered. At fine spatial scales, individuals' movement patterns are strongly confined by the physical condition of the current AB and the status of other individuals.

3.5 Geographical information systems

A geographical information system (GIS) is a tool that allows the processing of spatial data into information, generally information tied explicitly to, and used to make decisions about, some portion of the Earth (Demers 2000). GIS is useful in all areas of health research, from the description and explanation of spatial variation of disease and illness to the planning and use of health services (Gatrell & Senior 1999).

GIS provides an excellent means for visualizing and analyzing epidemiological data, revealing trends, dependencies and inter-relationships. The power of GIS to combine data from many sources, using many different scales, projections and data models is one of its major strengths. The second advantage of GIS over traditional cartographic and statistical techniques, besides the obvious speed advantages for handling large volumes of data, is in the improvement of hypothesis generating capabilities (Peuquet 1999). This power makes it possible to assess and account for even some rather complicated causal relationships between disease and different kinds of environmental and societal factors (Antikainen 1999).

Both spatial and temporal information is crucial to get meaningful inference about the causes of disease (Jacquez 2000), but GIS is not good at providing a dynamic representation of spatial phenomena at a micro level (Holm et al. 2000). One of the major

challenges of GIS is to formulate a conceptual framework and to integrate theories of individual's behavior, micro-level interaction and space-time constraints in order to model spatial micro-level dynamics. Another difficulty is to generate and analyze empirical data of micro processes in order to specify and calibrate dynamic spatial micro models (Fotheringham & Wegener 2000).

Integration of IBM and GIS is an issue (Brown et al. 2005, Gimblett 2002). The object-oriented nature of IBMs and GIS means they can complement each other, considering IBMs' temporal representations and GIS' spatial data representations (Brown et al. 2005). Brown (2000) pointed out four relationships between IBMs and GIS: (1) identity relationships, (2) causal relationships, (3) temporal relationships and (4) topological relationships.

3.6 Network theory

Complex networks describe a wide range of systems in nature and society. Social networks are complex networks. Technically, most individual-based models are based on some pre-defined (always computer-generated) contact network, and others build this network from real data. In the latter case, a sample is used because a real human transmission network would require a vast amount of detailed sociological data, and would be difficult to collect, analyze and understand. However, even a sample can require a huge effort. EpiSims (see <http://www.ccs.lanl.gov/ccs5/projects/episims.shtml>) is a good example of using real contact data. EpiSims takes advantage of human mobility information derived from TRANSIMS (see <http://transims.tsasa.lanl.gov/>). TRANSIMS estimates the movement of people as constrained by transportation infrastructure based on census data and activity surveys taken from a small sample (2000 households) of the population. Computer-generated simulation can, in certain circumstances, present a viable alternative.

Random graphs, small world models (Moore & Newman 2000, Small & Tse 2005) and scale free models (May & Lloyd 2001, Pastor-Satorras & Vespignani 2001) are the most widely used network models. Random graphs can be built as follows: start with N nodes and connect each pair of nodes with probability p , just a graph with approximately $pN(N - 1)/2$ edges distributed randomly. In scale free networks, some nodes act as "highly connected hubs" (high degree), although most nodes are of low degree. Scale free networks' structure and dynamics are independent of the system's size, the number of nodes the system has. In other words, a network that is scale free will have the same properties no matter what the number of its nodes is. Their defining characteristic is that their degree distribution follows a power law relationship. Watts and Strogatz (1998) defined the small world network: "As highly clustered, like regular lattices, yet having small characteristic path lengths, like random graphs". The most important properties of the small world network are as follows:

- Average vertex-to-vertex distances increase only logarithmically with the total number N of vertices.
- Two neighbours of a vertex will often also be neighbours of one another.

Watts and Strogatz (1998) designed a rewiring procedure to construct a small world network between regular and random networks: starting from a ring lattice with n vertices and k edges per vertex, rewire each edge at random with probability p ($p = 0$, regularity; $p = 1$, disorder; $0 < p < 1$, small world network), see Figure 3.5. As Newman (2000) stated: "Disease spreading on a small world graph reaches a number of people which increases initially as a power of time, then changes to an exponential increase, and then flattens off as the graph becomes saturated" (Newman 2000, pp 839), see Figure 3.6.

Infectious diseases spread rapidly on both the small world and scale free networks because of the short average vertex-to-vertex distances along the links of the networks. Further,

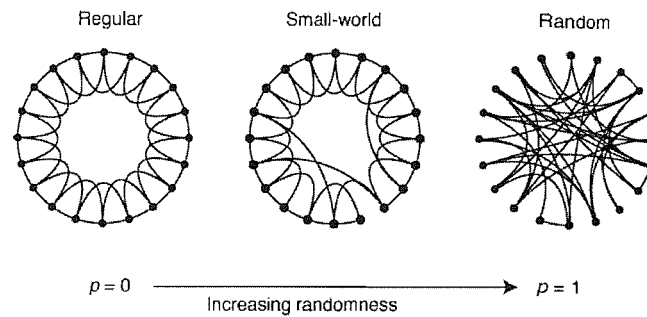


FIGURE 3.5: Random rewiring procedure for interpolating between a regular ring lattice and a random network (Watts & Strogatz 1998)

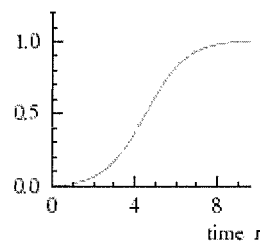


FIGURE 3.6: The percentage of people infected as a function of time by a disease which starts with a single person and spreads through a community with the topology of a small world graph (Newman 2000)

spreading is accelerated in scale free networks because of the power-law distribution of connections per individual.

Three network characteristics have an important bearing on the spread of infectious diseases (Wallinga et al. 1999):

- The average number of people contacted per person indicates how many secondary cases might potentially acquire the infection from one index case (Wallinga et al. 1999). Edmunds et al. (1997) made a questionnaire survey on a sample of staff and students at two British universities and their family and friends to get their daily number of contacts. It was observed that the average number of daily contacts was 17. In the study by Pool and Kochen (1978), the participants had an average of 23 daily contacts (including telephone conversations and letters) but the total new contacts over 100 days was only 360.

- The network's transitivity. Transitivity is the average proportion of secondary contacts that also contact each other (i.e. how many of your friends are also friends of each other) and is a measure of how clustered a network is. Pool and Kochen (1978) estimated that any person contacted by one of their study participants knew between 8% and 36% of the other people contacted by that participant. Of course, the exact value will depend on the period over which contacts are recorded and the degree of intimacy of the contact.
- The characteristic path length. The characteristic path length of a network is defined as the average number of contacts in the shortest route between two individuals in the network, with the average taken over all possible pairs of individuals. The characteristic path length describes the global network structure.

Read and Keeling (2003) pointed out the four properties of human disease transmission networks:

- The finite number and variability of potential contacts.
- Small world property (on average, any two individuals are connected by a small number of social or transmission steps) (Watts & Strogatz 1998).
- The clustering of social contacts such that adjacent individuals in contact space are likely to have many shared social contacts.
- The existence of local networks (highly clustered networks where connected individuals are likely to share common contacts) and global networks (unclustered networks with a high proportion of long-range connections).

3.7 Summary

This chapter introduced the modelling of humans' space-time dynamics, especially at fine space-time scales. Time geography provided a framework, and theories and concepts

were borrowed from social science. Geographical information systems and network theory were introduced as related tools. The basic idea behind this review was that the physical contact between individuals is a prerequisite condition for the transmission of airborne infectious diseases. The concept of the activity bundle, key to linking individuals' activity and the environment, is used throughout this thesis.

Chapter 4

Conceptual Model of ISTAM

An individual space-time activity-based model (ISTAM) was developed, which integrates the contact patterns of individuals, an infectious disease process model and a stochastic infection model together by simulating ABs. The model is stochastic and time-discrete.

In this chapter, firstly, key objects within ISTAM, such as infectious disease, individual and activity bundle are introduced. ISTAM's two-level (between-AB and within-AB) structure is presented later. At within-AB levels, three different simulation approaches, that is, raster space AB simulation, vector space AB simulation and rôle-based AB simulation, were adopted during the application of ISTAM to different cases. Vector space AB simulation is introduced in terms of its concept, implementation, running and initial analysis results and the other two approaches are introduced only at the conceptual level (because the applications of the other two approaches are presented fully in other related Chapters). At last, the input and output of ISTAM are presented.

4.1 Objects in ISTAM

As reviewed by Bian (2003), four components should be represented explicitly in an IBM: (1) individuals, (2) environment, (3) interactions between individuals in the environment, and (4) interactions between individuals and the environment. In ISTAM, it is assumed that interactions between individuals and the environment have no direct effects on the transmission of infectious diseases. The other three components are represented in ISTAM.

The structure of ISTAM is shown in Figure 4.1. Original data sources about infectious diseases, the population and the environment are used to 'feed' the model for different application cases. Parameters can be changed to suit different simulation scenarios to answer 'What-If' questions. Statistics can be generated by aggregation at different levels.

4.1.1 Infectious diseases

In ISTAM, three parameter groups are defined to describe a specific infectious disease: (1) evolution of the infection and disease within the host, (2) disease severity and (3) effective contact.

Different infectious diseases need different groups of time lines to describe the evolution of the infection and disease within the host. Generally, time lines include the latent period, infectious period, incubation period and symptomatic period. There are no widely accepted distributions for the time lines for one specific infectious disease because these parameters are dependent on age, race and many other differences within the human population and on variation in the infectious disease itself. In ISTAM, the time lines are simplified to be uniformly distributed between the minimum and maximum value (see Table 4.1).

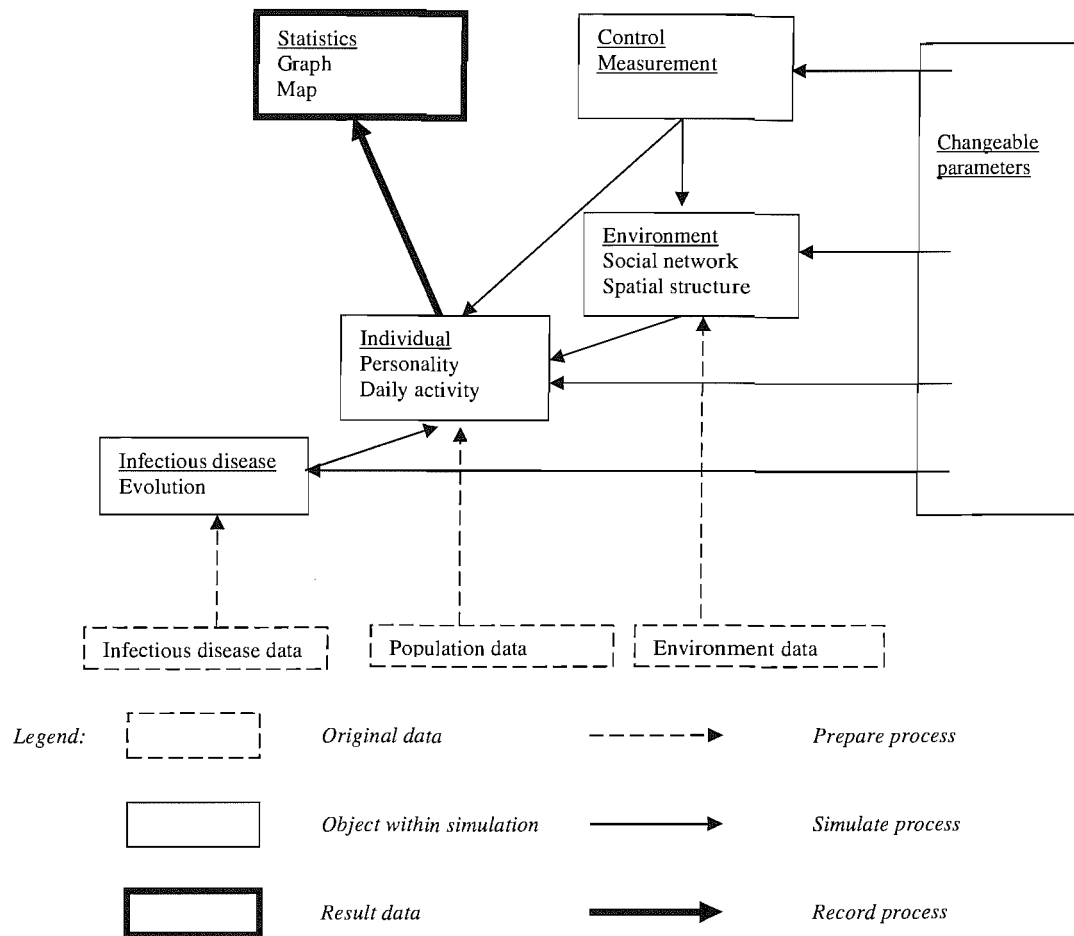


FIGURE 4.1: Structure of ISTAM

In ISTAM, infectivity and virulence are two properties used to describe the severity of a specific disease. Infectivity means the ability of a disease agent to cause infection. An infectivity index is used in ISTAM to measure infectivity. For one contact, the larger the infectivity index, the greater the likelihood of transmission between a pair of infectious and susceptible individuals. Virulence refers to the severity of the disease after infection occurs. It is measured by the case fatality rate or the proportion of clinical cases that develop severe disease. In ISTAM, mortality is used to measure virulence.

There is no clear and widely accepted definition of effective contact, since infection is a complex process and different infectious diseases have different transmission modes. The definition of an effective contact depends on the current understanding of the infection process for the specific disease. For airborne diseases, Edmunds et al. (1997) defines

TABLE 4.1: Model parameters in ISTAM.

	Sub-group	Parameter name
Infectious disease	Time lines	Latent min period
		Latent max period
		Incubation min period
		Incubation max period
		Infectious min period
		Infectious max period
	Severity	Infectivity index
		Mortality
	Effective contact	Attack duration (minutes)
		Attack distance (metres)
Activity bundle	Geometry	Width (metres)
		Length (metres)
	Dynamics	Spatial distribution type
		Minimum X (metres)
		Minimum Y (metres)
		Movement frequency
		Movement proportion
		Movement pattern
Individual	Disease dynamics	Latent period
		Incubation period
		Infectious period
	Activity pattern	
	Health level	
	Personal information	Accommodation
		School

the at risk contact as a two-way conversation. Essentially, if two individuals are close enough to have a conversation then they are probably close enough to transmit disease. This is easy to understand but hard to quantify. For some diseases such as measles, the virus in droplets sprayed by infectious people can remain active and contagious for up to two hours (see <http://www.cdc.gov/nip/diseases/measles/faqs.htm>). In ISTAM, it is assumed that transmission probability is inversely related to the distance between the infectious and susceptible individuals if there are no obstacles between them and also is directly related to the duration of the contact. Thus, if an infectious individual stays within a specific distance (attack distance R) of another susceptible individual longer than a specific period (attack duration T), the two individuals come into effective contact (i.e., infection is possible). In ISTAM, the contact index $C_{A,B}$ is used to express

the degree of effective contact between individual A and individual B , as follows:

$$C_{A,B} = I(t) \text{Min}(1, e^{1-\frac{r}{R}}) \quad (4.1)$$

Where: $I(t) = 1$, if $t \geq T$; $I(t) = 0$, if $t < T$ (r =distance between two individuals, t =duration). The concepts of effective contact and contact index are applicable directly to airborne infectious diseases.

If individual A is infectious and individual B is susceptible, then the probability of infection $P_{A,B}$ is as estimated as:

$$P_{A,B} = \frac{I_A C_{A,B}}{H_B} \quad (4.2)$$

Where I_A is the infectivity index of individual A , H_B is the health level of individual B , and $C_{A,B}$ is the contact index between them. Thus, the process of disease transmission is separated into the occurrence of contact and occurrence of infection: the two key factors that need to be captured in a disease transmission model (Zheng et al. 2005).

4.1.2 Activity bundles

Space in ISTAM is divided into a number of ABs where most human interactions occur. An AB can be one room, a whole building or a building complex. According to the type of human activities and the function of the space, ABs can be categorized into different groups described by a set of parameters.

The basic parameters required to describe an AB are the geometry of the space (such as size and spatial layout) and space-time dynamics of individuals within it (see Table 4.1). The space can be a two or three-dimensional grid (raster) or a continuous space (vector) depending on need and the availability of data.

In ISTAM, the parameters required to describe individuals' space-time dynamics are: (1) the minimum possible distance between individuals in the x and y directions (i.e., MinX and MinY); (2) spatial distribution types: four types of individuals' spatial distribution within an AB were implemented in ISTAM (see Table 4.2); (3) movement. The minimum distance can vary with direction to reflect such differences in reality (e.g., in a lecture room or computer room, the distance between individuals side by side may be less than that between adjacent rows). In some ABs, individuals are assumed to remain static during the simulation time unit, such as in a lecture room. In some other ABs, movements must be considered. Three parameters are used to describe individuals' movements inside an AB during one time unit: movement frequency, movement proportion and movement pattern. Movement frequency (M_f) refers to how often individuals move while movement proportion (M_p) refers to the proportion of the individuals who choose to move instead of staying where they are. Movement pattern (M_a) is used to describe how individuals choose their next position inside the AB (currently, only one option is implemented, i.e., randomly).

4.1.3 Individuals

In ISTAM, one individual has three groups of properties: static, dynamic and intelligent properties. Static properties include health level, disease dynamics, activity pattern or other related personal information which does not change during the simulation and should be input at the initial stage of one simulation (see Table 4.1). Dynamic properties can be changed and the change can affect the simulation. Examples are current AB and current disease state. Intelligent properties include the individual's reaction to the current situation. Currently, intelligent properties include changing their daily activities after control measures (such as isolation) have been triggered, and it is assumed that one individual will stay at home during his or her symptomatic period then come back to their normal daily activity after this period.

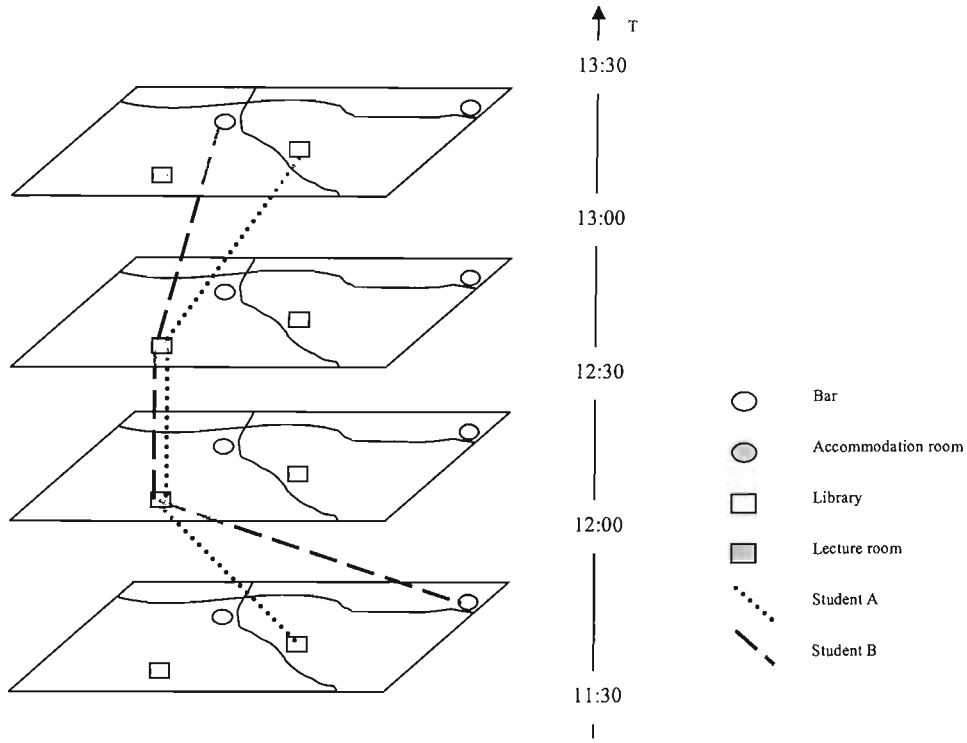
In ISTAM, an individual's space-time model is expressed at two levels. At the between-AB level, an individual's space-time model consists of a sequence of 2-tuples (AB, t) , where t can be a discrete time variable (currently, one time unit is 15 or 30 minutes). At the within-AB level (i.e., during one time unit), one individual's position within an AB is described by one tuple (x, y, t) (if static) or a sequence of tuples (x, y, t_i) (if dynamic, where t_i is a sub-period within the time unit).

The specific time lines of disease and infection within a host (infected individual) within ISTAM are obtained by drawing a realization from the distributions for the time lines. It needs to be pointed out that dependency exists between the time lines. For example, the incubation period for influenza within a host is modelled as one day longer than the latent period (Longini et al. 2005).

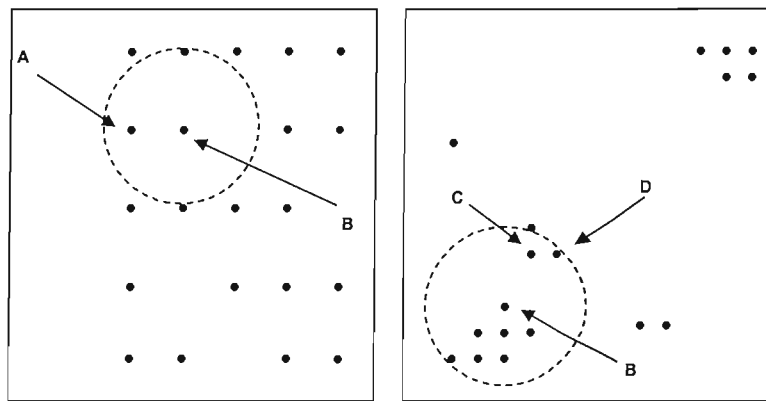
4.2 Two levels of simulation

AB simulation is a process by which the contact network (specific to a target infectious disease) is generated from the space-time dynamics of individuals constrained by both the individuals' activities and the space's physical condition. Through AB simulation, the activity pattern, contact network and spatial effect are considered at two levels: (1) between-AB and (2) within-AB. At the between-AB level, the expectation is that individuals are more likely to visit locations nearby their current location. This can be reflected by their activity patterns: for example, individuals prefer to visit buildings with particular functions (e.g., shops) near to their household. At the within-AB level, the concept of effective contact is applied. In short, space-time coidentity at the between-AB level is a prerequisite for effective contact at the within-AB level, which is a prerequisite for infection occurring.

At the between-AB level, individuals are placed into their residential accommodation at the beginning of the simulation. Each individual has an activity pattern to determine



(a)



(b)

(c)

FIGURE 4.2: Simulation at two levels: (a) an example of student A and B's activities between 11:30 and 13:30, (b) an example of the spatial distribution of students inside a lecture room between 12:00 and 13:00, (c) an example of the spatial distribution of students inside a refectory between 13:00 and 13:30

his or her whole day activity. At each time step, the individual moves to the next AB or stays in the current AB. As an example, Figure 4.2a shows the activities of student A and student B between 11:30 until 13:30 during a particular weekday. Student A was at the library while student B was in an accommodation room before they both went to the same lecture room. After a one-hour lecture (between 12:00 and 13:00), Student A returned to the library and student B went to a bar. If student B is infectious, then infection is possible depending on the AB simulation inside the lecture room and bar.

At the within-AB level, such as during the lecture, student A and student B can come into contact or not depending on the simulation inside the lecture room (see Figure 4.2b). Figure 4.2c shows an example of the spatial position of student B amongst other students inside the bar between 13:00 and 13:30. It is clear that 17 students sit as five clusters. Effective contact is still possible from student B to student C or student D, even though they are not in the same cluster, because the minimum distance between students in a bar is modeled as only 0.5 m with the attack distance modeled as 2 m. It is clear that in the bar student B came into contact with more students than in the lecture room at the same attack distance (2 m) because of the clustered distribution and the smaller distances between students inside the bar than in the lecture room.

4.3 Within-AB simulation

It is obvious that for infection to be possible, the physical activities of the infectious and susceptible individuals must overlay in both space and time. If all the physical activities (at the space-time scale required by the infection model) of all individuals belonging to the population are traced in the model, all possible infections are captured. While straightforward in theory, such physical activities, which are driven by social purpose and constrained by the physical world, are complex. Thus, such knowledge is rarely possible in practice because it is difficult to trace all such physical activities at a meaningful

space-time scale. It is, therefore, important to identify the factors that play primary rôles and choose appropriate simulation tools.

Three types of within-AB simulation approaches were designed and applied as raster space AB simulation, vector space AB simulation and rôle-based AB simulation.

By raster space or vector space AB simulation, individuals' spatial locations within an AB are explicit and both the first and the second types of contacts can be simulated directly and no need to consider them separately. Under most circumstances, internal physical information of ABs is not available, rôle-based AB simulation is the choice. The spatial pattern of individuals is explicit but individuals' spatial locations are not. And the first and second types of contacts need to be simulated separately.

4.3.1 Raster space AB simulation

For raster space AB simulation in ISTAM, the space inside an AB can be represented using a rectangular or hexagonal grid. Although research has shown that the shape of the grid can affect disease transmission (Morris 1994), to simplify, a rectangular grid was used to represent the space inside each AB, and consequently, the position of one individual is represented as an explicit discrete grid coordinate (integer values for x and y) at one snapshot in time.

Four types of space-time dynamic inside an AB were implemented in ISTAM (see Table 4.2). An exploration of the effect of the parameters required for raster space AB simulation is provided in Chapter 7.

4.3.2 Vector space AB simulation

The space within an AB can be represented as continuous and the simulation can be based on the vector data model. Continuous vector space allows representation of indi-

TABLE 4.2: Four types of spatial distribution of individuals within an AB

<i>Distribution type</i>	<i>Distribution description</i>	<i>Example</i>	<i>Mathematical description</i>	<i>Parameter value</i>
1	Try to place each individual outside the neighbourhood of all others.	Computer room	$L \sim Rl ((L_m - L) > r_{\min}) \quad \forall m$	L is the spatial position of a new individual; L_m is the location of each <i>already</i> allocated individual m ; r_{\min} is the neighbourhood distance (here, $r_{\min} = 2$ m).
2	Individuals are located sequentially, by row and column.	Lecture room	$L_n = (\text{mod}(n, W), n/W)$	W is Width; n is an individual index (in this model, $n=1,2,\dots,N$); N is the total number of current individuals inside an AB.
3	Clustered	Bar	$L_n \sim R,$ $L_m \sim Rl (L_m - L_0) = \min,$ where $L_0 \sim R(L_1, L_2, \dots, L_{N/g})$	g is the expected cluster size (here, $g=2$); n and m represent individuals (in this model, $n=1,2,\dots,N/g$ and $m=((N/g)+1),\dots,N$).
4	Random	Gym	$L \sim R$	

R means random spatial distribution ($x \in R[0, \dots, \text{Width}-1], y \in R[0, \dots, \text{Length}-1]$); $R(L_1, L_2, \dots, L_i)$ means that the spatial position is selected from L_1, L_2, \dots, L_i randomly

vidual movements at a finer spatial scale than raster space. In vector space, the distance between any two individuals within an AB can be represented as a real value of a continuous variable (rather than in terms of pixels). Importantly, the distance rules can be applied directly.

A continuous-space model was built to represent a more general AB. The target area is a $10 \times 10m^2$ continuous space, where a certain number of persons (one of them is an infectious individual, all others are susceptible individuals) stay or move around inside the AB for 10 minutes. The time unit of the simulation is 5 seconds and the total simulation lasts 120 time steps. For influenza, this duration is far less than the duration for one individual to change from infected to being infectious and from being infectious to being immune, so during the simulation the only possible state change is from susceptible to infected. Three parameters were used to describe the interaction between individuals (see Table 4.3). The result of the simulation is the final number of infected individuals or the rate of infected individuals at the end (i.e., number of infected individuals divided by the total number of susceptible individuals at the beginning).

The vector space AB simulation model was used to determine the effect of total number of individuals and mobility index on the relation between the rate of infection and

TABLE 4.3: Parameters for vector space AB simulation.

Parameters	Description
Number of individuals (N)	The total number of individuals
Distance type (D)	Public, social, personal, intimate
Mobility index (M)	The probability for individuals to move (it is assumed that the next spatial position is assigned randomly if a move is made). The higher the value, the higher the average percentage of individuals that move during the next step.

distance between individuals (Figure 4.3). The results show that the mobility index is the primary factor affecting the rate of infection. If the total number of individuals and mobility index are not too high such that people inside the AB can maintain larger distances between each other, the effect of the total number of individuals is small. It is clear that the infection rate falls abruptly when the distance type changes from personal to social distance. The reason is that the diffuse distance parameter is assumed to be 2 m. The practical meaning is that if people can maintain separations of more than this distance, the infection rate decreases remarkably.

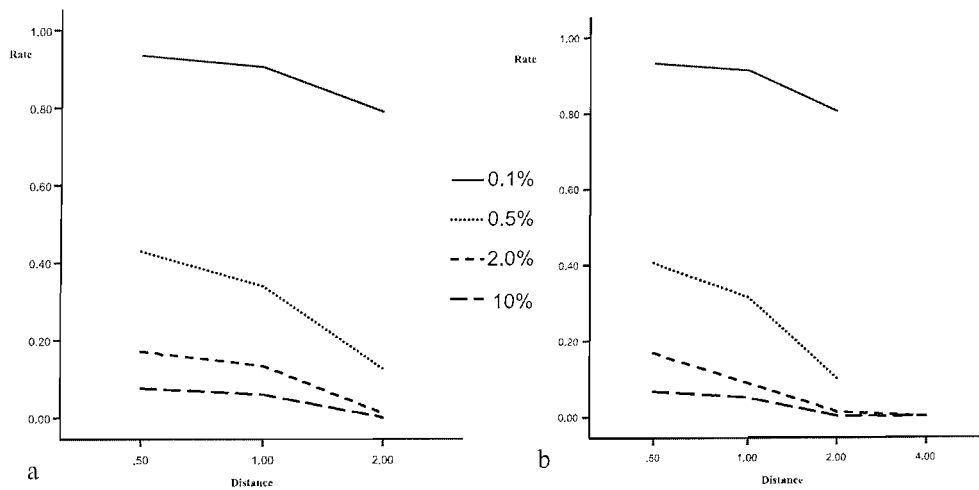


FIGURE 4.3: Proportion infected plotted against the distance between individuals for different values of the mobility index. In (a) the total number of individuals is 10 and in (b) it is 5.

Vector space simulation highlights the effect of the mobility index. The model can be used to adjust the parameters of the mobility index, the total number of individuals and the distance type. However, the size and spatial layout of each AB can also be

expressed and individuals inside can have different movement patterns to express the activities that they are involved in or the different roles that they play.

An example is provided here which is used to model influenza transmission amongst customers and salespeople inside one shopping mall (Figure 4.4). The lines show the outer and inner wall of this shopping mall. 13 salespeople are placed (ordered in an anticlockwise direction) and move only inside their sales zone of responsibility. Customers enter into the shopping mall from the entrance at the left-top corner and cross all salespeople's sales zones clockwise or anticlockwise. During each simulation step, customers can stay or move within the current sales zone or move to the next sales zone. The exact spatial location of each individual within each sales zone is assigned randomly. One person is assigned randomly to be infectious from the beginning. The AB-based infection simulation model for influenza is then applied. The smaller circular attack zone which is built by the release of the infectious virus from the infectious individual will diffuse to be a larger circular diffuse zone which can persist for a short duration.

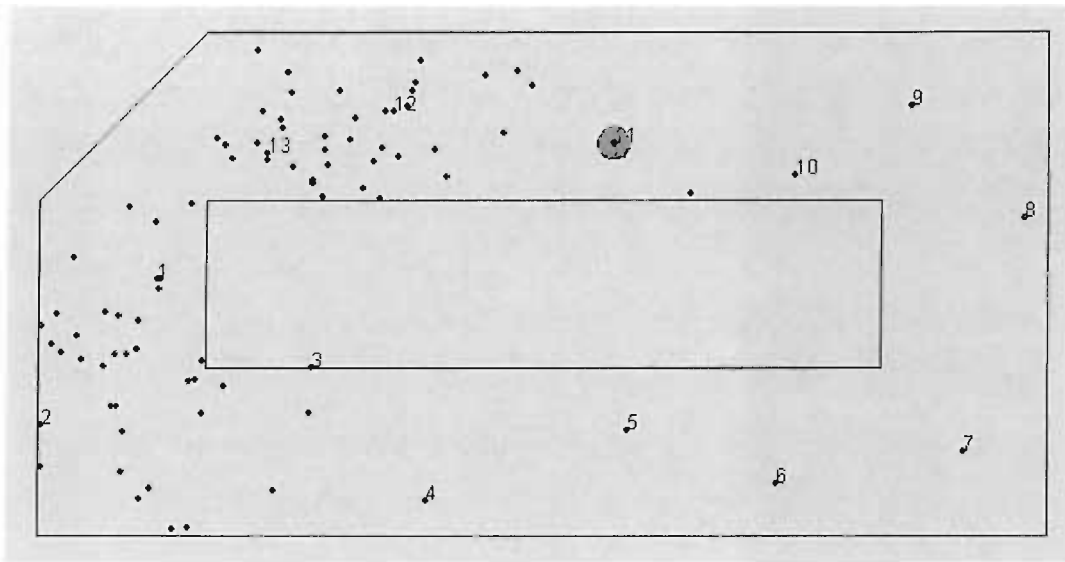


FIGURE 4.4: Spatial display of a shopping mall by vector space AB simulation.

The example above, although simple, shows two merits of vector space simulation: (1) the spatial layout can be expressed explicitly by the use of vector data (e.g., within a geographical information system, GIS); (2) the individuals' movement patterns inside the

AB can be as detailed as required such as to reflect their activities and rôles. However, a fundamental question is: how much detail is needed for this simulation such as to generate a valid conclusion which can be applied to more general situations? Both the raster space and vector space simulations are used to generate individuals' contacts based on their space-time relations at a fine spatial scale which requires an individual space-time model to be defined at fine space-time resolutions. Unfortunately, such models are seldom available.

4.3.3 Rôle-based AB simulation

Individuals' rôles within their joint activity are the basis for rôle-based AB simulation. The assumption of rôle-based simulation is that when people come together for a certain joint activity, their interactions will depend strongly on the different rôles which individuals play. For example, it is obvious that in a shop, the contacts between salespeople, between salespeople to customers and between customers are different not only in frequency but also in intensity.

The infection process between two individuals, as discussed before, can be separated into two stages: coming into contact and becoming infected. The probability of coming into contact (P_c) and probability of infection during one contact (P_i) were used to express the random properties of these two processes. As discussed before, different distance rules can be applied to determine the distances between individuals who are involved in a variety of activities and play different rôles. It is assumed that P_c can be determined by individuals' rôles in their joint activities.

The procedure for simulating interactions within an AB can be summarized as follows: (1) all individuals presented within the same AB are divided into a certain number of groups based on each individual's rôle in the joint activity; (2) a certain spatial pattern (without explicit spatial location at fine spatial scales) is assumed to exist within groups.

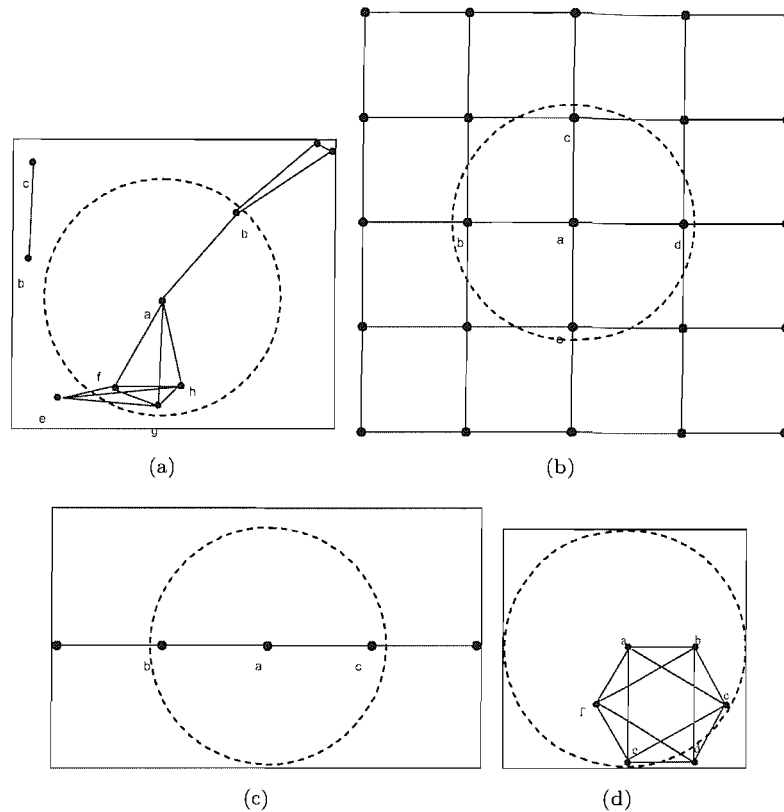


FIGURE 4.5: The spatial distribution pattern within a group: (a) static random, (b) static even, (c) queue and (d) loop (The points represent individuals. The solid lines between two points represent the possible contacts existing between these two individuals. The dotted circle around individual a shows all the individuals within this area who could come into contact with individual a).

Interactions between individuals within a group are simulated based on these spatial patterns. Table 4.4 lists a few general spatial patterns within a group and Figure 4.5 shows four of these spatial patterns; (3) interactions between individuals from different groups are simulated based on the activity properties. Table 4.5 lists a few general interaction types between two individuals belonging to different groups within the same AB.

Five levels of P_c were used to represent different probabilities of coming into contact. The lowest level was assumed to be a constant value F , and the ratio between two adjacent levels were fixed to be 2. That is: $P_c^1 = 2P_c^2 = 4P_c^3 = 8P_c^4 = 16P_c^5$. For each specific application, the value of F will be calibrated by simulation according to related research. An example can be found in Section 8.5.1.

TABLE 4.4: Spatial patterns within group.

<i>Pattern name</i>	<i>Value of P_i</i>	<i>Description</i>
Full	P_i^2	Every person is in contact with all other persons
Loop	P_i^1 with first above and below, and P_i^2 with second above and below.	Static. Fixed before the simulation, and does not change. Individuals distribute like a loop, so one person contact with persons first and second above or below with himself / herself in the loop
Queue	P_i^2	Static. Fixed linear distribution. One person makes contact with another person just before or after him / her in the queue
Static random	P_i^2	Persons are distributed randomly. Each individual comes into contact with her / his four Von Neumann neighbours. One additional parameter: Density
Static even	P_i^1	Persons are distributed evenly in a grid space. Each individual comes into contact with her / his four Von Neumann neighbours
Dynamic random	P_i^3	Every person comes into contact with a number of randomly selected other persons. This process repeats a number of times. Two additional parameters: Density and Mobility

TABLE 4.5: Interaction types.

<i>Relation type</i>	<i>Default distance</i>	<i>Contact distance</i>	<i>Description or example</i>
Cooperation	Social distance	Personal distance	Teamwork or teachers and students, managers and workers
Service	Public distance	Personal distance	Service environment. One individual serves another
Care	Social distance, Personal distance	Intimate distance	Health care facilities. Patients and nurse
Random contact	Public distance	Social distance, Personal distance	Restriction by physical environment

Three levels of P_i were used to represent different probability types of getting infection with the contact distance set to be intimate, personal and social distance. The lowest level was assumed to a constant value I , and the ratio between two adjacent levels were fixed to be 2. That is: $P_i^1 = 2P_i^2 = 4P_i^3$. For each specific application, the value of I will be calibrated by simulation according to related research. An example can be found in Section 8.5.1.

As an example, let us consider the possible contacts within a restaurant. According to their rôles, the individuals inside the restaurant can be classified to groups such as the manager, kitchen staff, waiters and customers. The customers can be divided further

into several customer groups (Figure 4.6). Alternatively, to simplify the computation, the above individuals could be divided into only two groups (i.e., workers and clients). Different types of contacts (with different values of P_c) could occur between and within different groups (see Table 4.6). For example, it can be assumed that all customers belonging to the same customer group prefer to maintain personal distances or intimate distances between each other while the waiters prefers to maintain a social distance with customers when they are not providing service and personal distance when they are providing service. As the customers belong to different customer groups, they prefer to maintain public distances to retain privacy, but in many circumstances this cannot be satisfied due to the restriction of the physical environment of the restaurant.

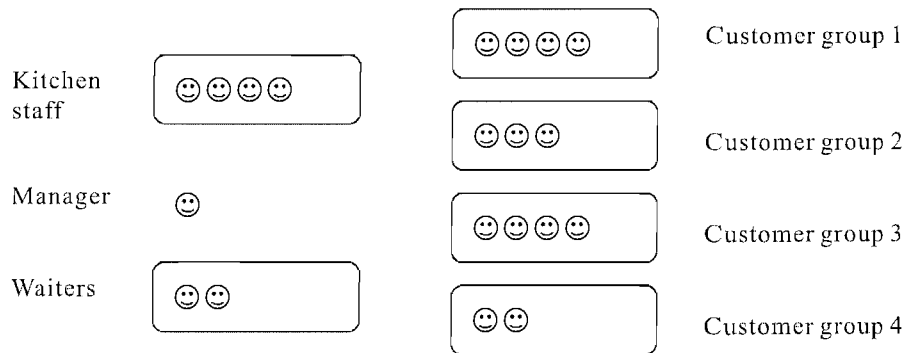


FIGURE 4.6: Classification by rôles of individuals inside a restaurant.

TABLE 4.6: Possible contacts between individuals inside a restaurant.

	<i>Contacts</i>	<i>Spatial pattern / interaction type</i>
Within groups	Between kitchen staff	Full
	Between waiters	Static random
	Customers belong to the same customer group	Full
Between different groups	Manager with kitchen staff	Cooperation
	Manager with waiters	Cooperation
	Waiter with customers	Service
	Manager with customers	Service
	Customers belong to different customer groups	Static random

4.4 Input and output

To simulate the transmission of a specific disease, information is needed on (1) the infectious disease, (2) the environment and (3) the population. In ISTAM, the corresponding parameters needed represent the infectious disease, activity bundles and individuals (Table 4.1). Parameters representing the infectious disease can be obtained from the relevant literature, while parameters representing the activity bundles and individuals have to be obtained from a survey targeted to the corresponding population and environment. After defining the model parameters, several running parameters need to be set for one simulation including the number of initial infected individuals, number of initial immune individuals, simulation duration, disease control measures and so on.

Information on changes in the activity and disease state at both the individual level and AB-level are recorded into a database during each ISTAM simulation. Therefore, subsequent analysis can be done after each simulation or based on a considerable number of repeated simulations.

4.5 Summary

In this chapter, the model of ISTAM and three within-AB simulation approaches were introduced at the conceptual level. The implementation is presented in next chapter.

Chapter 5

Implementation of ISTAM

5.1 Tools for implementation of ISTAM

The implementation of ISTAM is based partly on RePast (an agent-based simulation toolkit, see <http://repast.sourceforge.net>) and Java Topology Suite (JTS, an API which provides an implementation of the spatial data model, see <http://www.vividsolutions.com/jts/>). Java (an object-oriented language, see <http://www.java.sun.com>) was used for development.

5.1.1 IBM library of RePast

The University of Chicago's Social Science Research Computing's Recursive Porous Agent Simulation Toolkit (RePast) is a software framework for creating agent-based simulations using the Java language (requires version Java 1.4 or greater). It provides a library of classes for creating, running, displaying and collecting data from an agent-based simulation. In addition, RePast can take snapshots of running simulations, and

create quick time movies of simulations. RePast borrows much from the Swarm simulation toolkit and can properly be termed ‘Swarm-like’.

Swarm (see <http://www.swarm.org>) is the earliest IBM software library and is a software package for multi-agent simulation of complex systems. The basic architecture of Swarm is the simulation of collections of concurrently interacting agents. Models are developed as object-oriented programs which give flexible ways of specifying model behaviors. Swarm is taken as the most exhaustive object-oriented modelling environment developed and one that is most commonly applied to social and urban simulations (Agarwal & Abraham 2003). Swarm is provided in both C and Java.

An evaluation of free Java-libraries for agent-based simulation by Tobias and Hofmann (2004) compared RePast, Swarm (Java version) and two other relatively less-used software platforms (Quicksilver and VSEit). The evaluation was based on official program documentation, statements by developers and users, and the experiences and impressions of the evaluators. The conclusion of the evaluation was that RePast was the most suitable simulation framework for the applied modelling of social interventions based on theories and data.

In another evaluation, Railsback et al. (2006) compared software platforms for IBM including Netlogo, MASON, RePast, Swarm (Java version) and Swarm (Object C version). RePast was taken as “the most complete Java platform” and “its execution speed to be good compared to the other platforms” (Railsback et al. 2006, pp 624). Also, RePast includes classes for geographical information systems and network functions.

5.1.2 Java Topology Suite

The JTS is a Java API that implements a core set of spatial data operations using an explicit precision model and robust geometric algorithms. It provides a complete model

for specifying 2-D linear geometry. Many common operations in computational geometry and spatial data processing such as Boolean functions (equals, disjoint, intersects, touches, crosses, within, contains, overlaps), overlay, buffer, polygonization and merging a set of line strings functions are exposed in a clear, consistent and integrated API. JTS is intended to be used in the development of applications that support the validation, cleaning, integration and querying of spatial datasets.

As stated in the previous chapter, space in ISTAM is considered as two levels. At between-AB level, JTS is used to compute the distances between ABs mainly. At within-AB level, JTS is used to compute and select a location for individuals according to the current AB's space-time dynamics.

5.1.3 Object-oriented programming and Java

The popularity of IBMs relies on the increases in computer power and the use of object-oriented languages (Ginot et al. 2002). IBM programming can be done in any language, but Object-Oriented Programming (OOP) languages are the most appropriate languages since the concept of an object is similar to the concept of an individual (or agent).

OOP languages provide data structures which naturally allow for efficient individual-based modelling. An object is a software entity containing attributes plus methods that act on these attributes. An object controls access to its attributes and methods by declaring them as public, private or protected (which indicates different levels of accessibility by other objects).

Fundamental concepts within OOP are inheritance, encapsulation and polymorphism. A class **B** can inherit the attributes and methods of another class **A**. The class **B** is then called the subclass of class **A**, and class **A** is called the superclass of class **B**. A subclass can include specialized attributes and methods that are not present in the superclass. Encapsulation is the process of determining which aspects of a class are not needed by

other classes, and hiding these aspects from other classes. Polymorphism is the ability for two separate yet related classes to receive the same message but to act on it in their own way. In other words, two different (but related) classes can have the same method name, but they implement the method in different ways. The most popular OOP languages are C++, C# and Java.

Java is an object-oriented programming language developed by Sun Microsystems in the early 1990s. One of the most important characteristics of Java is its platform independence, that is, programs written in the Java language can run similarly on any supported hardware/operating-system platform. In Java, the ‘class’ is the key concept which defines the object’s characteristics (named its attributes or properties) and its behaviors (named methods or features). Objects are created as instances of a class. In this chapter, all names of class are highlighted by bold font.

5.2 Main packages and classes of ISTAM

To facilitate code reuse, Java allows users to group several class definitions together in a logical grouping called a package. Figure 5.1 shows the main packages and classes within an example application of ISTAM (see Chapter 6). There are mainly three types of packages being applied within this application.

- Packages from the RePast library for which names begin with ‘uchicago.src.sim’. These packages are used mainly for schedule control (such as `uchicago.src.sim.engine`), within-AB space management (such as `uchicago.src.sim.space`) or the generation of random numbers (such as `uchicago.src.sim.util` and `cern.jet.random`).
- Packages from the Java language. The names of these packages begin with ‘java’, these general packages are used mainly for data input and output from file or database and some common utility functions.

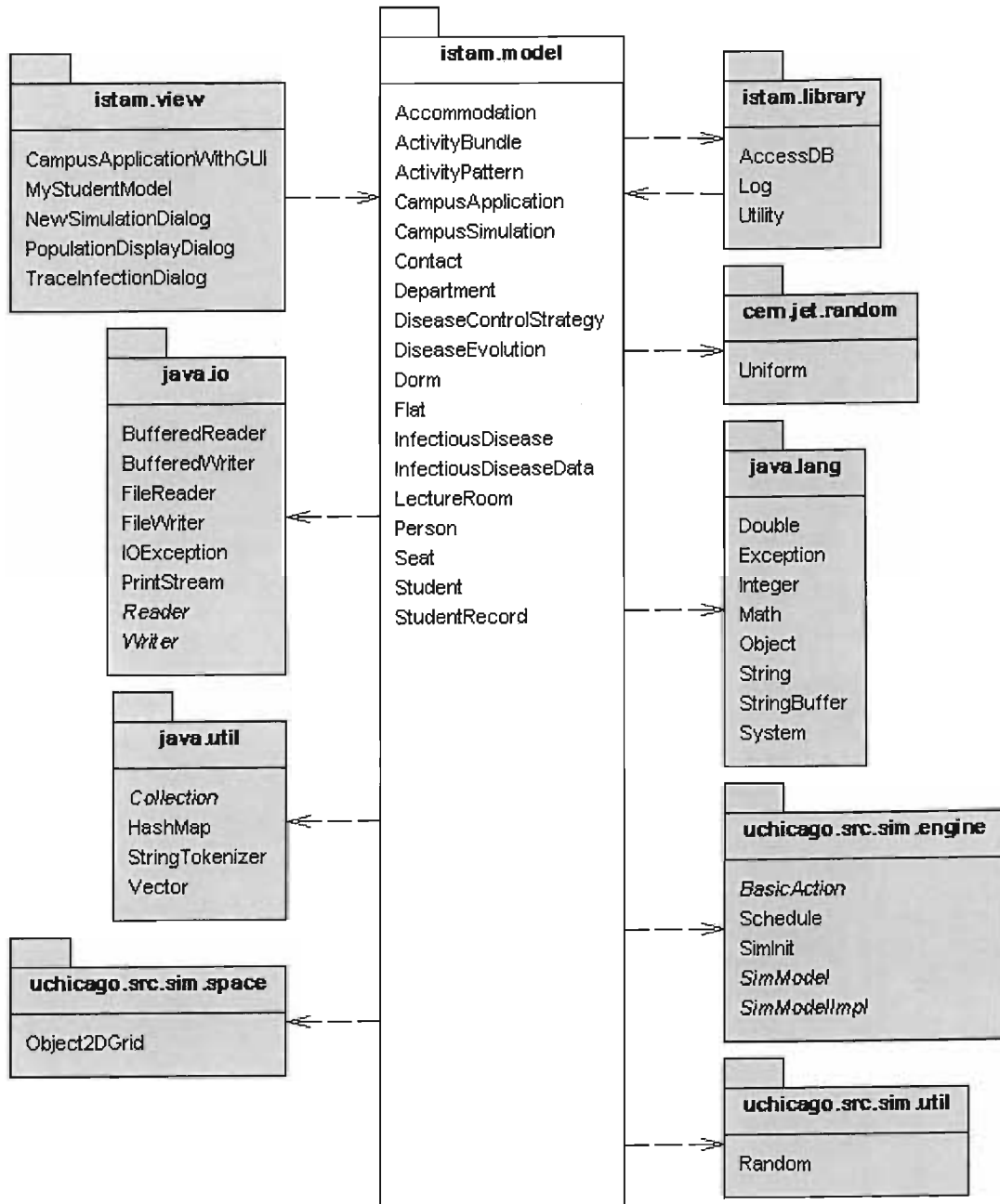


FIGURE 5.1: Packages and classes within an example application of ISTAM.

- Packages for which names begin with ‘istam’ were developed by the author of this thesis. The ‘istam.model’ package contains classes corresponding to the objects within the conceptual model (see Figure 4.1), the ‘istam.view’ package contains classes which are used for interface display, the ‘istam.library’ package contains classes for database input and output function (**AccessDB**), utility functions (**Utility**) and log information (**Log**).

Three main classes **InfectiousDisease**, **ActivityBundle** and **Person** were implemented according to the description in Section 4.1. Here, other classes are described.

5.2.1 Class of Student and Citizen

Class of **Person** is the basic class for an individual, and Class of **Student** and **Citizen** extend the class of **Person** for different applications (for application of the former see Chapter 6 and for the latter see Chapter 8). Beyond properties of **Person**, **Student** has properties such as **Department**, **Flat** and **Dorm**, **Citizen** has properties such as **Household**, **WorkPlace** or **StudyPlace**. The classes of **Student** and **Citizen** are subclasses of the class of **Person**, and **Person** is the superclass of **Student** and **Citizen**. This is an example of the application of inheritance.

5.2.2 Class of DiseaseEvolution

The class of **DiseaseEvolution** was used to implement the process of infectious disease evolution with one specific individual. Three properties of **DiseaseEvolution** are latent, incubation and infectious periods. Each instance of **Person** (**Student** or **Citizen**) has an instance of **DiseaseEvolution** which has fixed values for the above three time lines (generated by **InfectiousDisease**). Along the simulation, if one individual gets infected, the **DiseaseEvolution** will begin to record her/his disease state and the time

of staying in the current disease state. If staying in the current disease state is equal to the threshold (by the value of the time lines), her/his state will be changed. Thus, the disease will evolve according to the time lines within an individual.

5.2.3 Class of ActivityPattern

As ISTAM is a time discrete model, the time unit was set to be a duration (15 or 30 minutes), such that one day is divided into a number of units. **ActivityPattern** is used to assign each unit with certain activity types. The properties of **ActivityPattern** consist of two parts: some fixed activities and some non-fixed activities. Fixed activities are obligatory activities for individuals during certain times such as attending lectures for students, going to work for adult citizens and sleeping for all individuals. For example, **Department** has a property of lecture timetable (implemented by an array of integers. Non-fixed activities are optional activities and the probability of the occurrence of these activities is based on survey data. A multi-dimensional double array was used to express this part; it can be relatively simple (see Section 6.2) or relatively complex (see Section 8.3.5).

5.2.4 Classes for spatial and social environment

ActivityBundle is the basic unit used to represent the whole spatial and social environment within which the population lives. According to the different within-AB simulation methods, the subclasses of the class of **ActivityBundle** such as **LectureRoom** and **StudentUnion** were implemented in different ways. This is an example of the application of polymorphism.

Other classes are used to represent the environment, such as **Accommodation**, **Flat**, **Department** for the application in Chapter 6, **PC6Zone** for the application in Chapter 8. All the instances of these spatial or social units have a collection of individuals. This is implemented by the class of **Vector** or **HashMap**.

5.2.5 Class of Contact

Contact is a class to record all the contacts between individuals at fine levels. Each instance consists of six properties: two persons who came into contact, contact time, contact index, contact result and the activity bundle where this contact occurred. By this class, all information related to contacts can be recorded, and saved into a database for future analysis.

5.2.6 Class of DiseaseControlStrategy

The class of **DiseaseControlStrategy** allows the implementation of disease control measures. During the simulation, the summary disease information can be accumulated by class of **Utility**. If the number of individuals who are infectious with symptoms is larger than a certain threshold (Alert value), the disease control will be applied upon related individuals within the model depending on different control strategies. If one individual is set to be quarantined, her/his activity pattern will be changed.

5.3 Simulation process of ISTAM

Currently, all related data sources (original data sources, simulation information and simulated results) are saved in a Microsoft Access database. The class of **AccessDB** allows input and output data functions which are based on the package of 'java.sql'. The Table 5.1 shows the main database tables used by applications of ISTAM.

5.3.1 Preparation stage

As shown in Figure 4.1, before the main simulation, a series of tables (such as PRE_RAN_population, CON_activityBundles and CON_activityPatterns in Table 5.1) recording information

TABLE 5.1: Main tables within database used by ISTAM.

Name of table	Description
SIMULATION_INDEX	An index table. Each simulation is saved as a record. Parameters of the simulation are saved as columns.
SIM_compartments	Simulation results table. For each simulation, the numbers of individuals belonging to different disease states (compartments, e.g., susceptible, latent) at certain time units is saved as a record.
SIM_contacts	Simulation results table. Each contact is saved as a record. All contacts or part of contacts according to certain criteria (e.g., according to contact results or contact index)
PRE_***_population	Constant table. Each individual is presented in the table as a record. There could be a few different strategies for synthesizing the population (indicated by ***).
CON_activityBundles	Constant table. Each activity bundle is presented in the table as a record.
CON_activityPatterns	Constant table. Each type of activity pattern is presented in the table as a record.

about individuals and ABs need to be built based on population data and activity survey. Parameters about the target disease are obtained from statistical data about the previous epidemic outbreaks.

5.3.2 Simulation stage

In this Section, the process of the application of ISTAM within the first year undergraduate student body in the University of Southampton is given as an example of the simulation process.

After the initial interface, the main interface will be shown. To begin a simulation, a set of parameters need to be set or chosen in the parameters interface. The population can be chosen from different synthesizing methods. Initial index cases can be set randomly or by users. By the latter, the different kinds of distribution of index cases can be tested. Related analysis can be seen in Section 6.4.4.

Simulation duration can be set as: (1) a fixed value. After this duration, the simulation will be terminated; (2) limitless. The simulation will only be terminated by the

interruption of the user; (3) auto stop. The program can judge if there are no potential infections within the model (if all individuals are susceptible or immune), the simulation can be terminated automatically. This option makes the simulation flexible to the need of users.

If the user chooses the display option then during the main simulation a spatial display for the whole environment or a certain AB can be shown dynamically (i.e., updated every unit). The example of dynamic spatial display at both between-AB and within-AB levels for the application of ISTAM to the campus of University of Southampton can be seen in Appendix A (see Figure A.5 and Figure A.6). Spatial display at both between-AB and within-AB level during the simulation is optional.

After the setting of parameters, all information about this simulation will be saved into the table 'SIMULATION_INDEX' before the main simulation.

ISTAM benefits most from RePast in terms of its scheduling mechanism. RePast behaves as a discrete event simulator whose quantum unit of time is known as a tick. The tick exists only as a hook on which the execution of events can be hung, ordering the execution of the events relative to each other (Collier 2001). The schedule within ISTAM is run as tick by tick. A certain number of ticks makes up a day. The schedule for each day can be simplified as in Table 5.2. After the simulation, simulated results at individual level and AB-level are recorded into a database for future analysis. An example to display a certain student's activity during a certain day (see Figure A.7) and another example to display an infection tree (for one individual, the individuals who had been infected by him/her and the time and the AB of this infection are displayed, see Figure A.8) can be found in Appendix A.

5.4 Summary

The implementation of ISTAM was presented in this chapter. After the programming work, the entity of the model existed as a number of Java classes being organized within

TABLE 5.2: The schedule for each day within ISTAM.

1. IF a day begins	FOR each Person	Make a day plan according to activity pattern
2. IF meet the condition for control measures	Control related Persons	
3. FOR each Person	Move between ABs according to his/her day plan Update disease state according to disease evolution time lines	
4. FOR each AB	Interaction between all individuals within AB	
5. Update summary information related to current tick and save to database		
6. Update graph or map according to updated summary information		
7. IF a day ends		Update summary information related to current day and save to database

a group of packages. These packages were used as the basic library. When ISTAM is applied to a specific application, additional application classes will be implemented. Two application cases of ISTAM, are presented in Chapters 6 and 8.

Chapter 6

Application to the Campus of the University of Southampton

6.1 Background and survey

ISTAM was applied to simulate a hypothetical influenza epidemic in the University of Southampton. The University of Southampton is a research-led University with 13032 undergraduate students during 2004-2005 (see <http://www.soton.ac.uk>). In ISTAM, the first year undergraduate students from the University of Southampton were taken as a hypothetical population. Why choose this population? Firstly, most of the first-year undergraduate students live in the accommodation provided by the University. Thus, the University, as a semi-closed system covering a small geographical space, provides a useful basis for model development and evaluation. Secondly, students live within a limited space with a semi-predictable daily activity pattern (especially for first year undergraduate students, most of their study activities involve taking lectures which are known) (Tomlinson et al. 1973, Huisman & Forer 1998) and relatively less interaction

with the outer world (weekends are neglected). Thirdly, age structure need not be considered and there are no births or deaths and marginal immigration and emigration.

To acquire data on students' space-time activity patterns, a questionnaire survey was taken by the author in March 2005 at the University of Southampton. The questionnaire included three parts (see Figure C.1): (1) personal information, (2) time points and (3) activity preferences and durations. 315 valid questionnaires were collected from first year undergraduate students from ten University schools. These data were input into databases to build activity patterns per individual. According to the survey, most of the students prefer to visit the library, computer room or Student's Union between two discontinuous lecture periods. At night (no lectures), they prefer to go to the student's bar, gym or stay at accommodation more than go to the library or out of campus. During lunch (mid-day) and dinner (evening), some students cook in the accommodation kitchen, while others go to a refectory or buy food.

6.2 Students and ABs

4000 first year undergraduate students belonging to six accommodation houses and 26 schools were modelled in ISTAM, and it was assumed that no other students or staff existed. The weekends and holidays were also ignored. For each student, one activity pattern was selected randomly from the surveyed 315 activity patterns. However, since the survey data represent only a small sample of the first year student population, some assumptions have to be made to build a complete student space-time activity pattern. The activity pattern includes two parts: (1) time points and (2) activity preferences and durations. Some time points are fixed such as the lecture times (it is assumed that every student takes part in every lecture on which he or she is enrolled) while other time points have a fluctuating range (such as the time for a student to leave or return to his or her accommodation). Activity preferences are weighted (for example, one student may have 75% probability to go to the library and 25% probability to go to the gym during

one hour's vacant time between two lectures). Again, some activities have a relatively fixed duration (such as lunch at refectory) while others do not (such as studying in the library or a computer room).

Twelve types of AB were designed to describe the spatial environment (see Table 6.1). The spatial structure of the accommodation (building, flat, room, kitchen) and the

TABLE 6.1: List of ABs in an example application of ISTAM.

Name	Number	Distribution type	Geometry ($L \times W, x_{min} \times y_{min}$) ^a	Movement ^b
Bar	15	3	$5 \times 5, 0.5 \times 0.5$	
Computer room	10	1	$20 \times 20, 1 \times 2$	
Gym	1	4	$45 \times 45, 1.5 \times 1.5$	
In campus	1	4	$1000 \times 1000, 2 \times 2$	$M_f=3, M_p=100\%$
Kitchen	576	1	$3 \times 3, 1 \times 1$	
Lecture room (a)	4	2	$8 \times 15, 1 \times 1.5$	
Lecture room (b)	9	2	$15 \times 18, 1 \times 1.5$	
Lecture room (c)	12	2	$20 \times 20, 1 \times 1.5$	
Off campus	1	4	$10000 \times 10000, 2 \times 2$	$M_f=3, M_p=100\%$
Refectory	2	3	$20 \times 20, 1 \times 1$	
Library room	10	4	$20 \times 20, 1 \times 1$	$M_f=3, M_p=20\%$
Student union	1	4	$40 \times 40, 1 \times 1$	$M_f=3, M_p=50\%$

^a All measures given in meters.
^b M_f : movement frequency; M_p : movement percentage.

lecture timetables of schools were obtained from real data. Here, library room and student union are the two ABs that consider the movements of students inside. It was assumed that movement pattern inside the above two ABs can be modelled as random (i.e., the next position of one individual has no relation with the current position). 'Off campus' and 'in campus' are two special ABs here to include all other spaces where the interaction between students are possible but have not been covered by other ABs.

6.3 Influenza

This research focuses on the infectious diseases that can be transmitted by contact or airborne transmission, such as influenza, measles and mumps. ISTAM was applied to

simulate hypothetical influenza outbreaks. Influenza (commonly called ‘the flu’) was used to test the model and to compare different disease control measures.

Influenza is a contagious respiratory illness caused by influenza viruses. There are three types of influenza virus: A, B and C. Influenza B almost exclusively infects humans (Hay et al. 2001) and influenza A is the most virulent human pathogen amongst the three types and causes the most severe disease. Infection with influenza viruses can result in illness ranging from mild to severe and life-threatening complications (see <http://www.cdc.gov/flu/pdf/keyfacts.pdf>). Symptoms include fever, respiratory symptoms, nasal discharges, cough, headache and sore throat.

It is believed that time lines for influenza have been relatively invariant over the past half a century (Longini et al. 2005). The latent period of influenza varies from 1 to 3 days, and the incubation period is one day longer than the latent period (Steinhoff 2000). The infectious period varies from 3 to 6 days (Longini et al. 2004). Elveback et al. (1975) estimated the latent and infectious period of influenza as: latent period of 30% 1 day, 50% 2 days, and 20% 3 days; infectious period of 30% 3 days, 40% 4 days, 20% 5 days, and 10% 6 days. It is during the intersection of the incubation period and the infectious period where viruses spread the most (Scott & Duncan 2001). One reason may be because without symptoms the infectious individuals are unaware of their infectious status and continue normal contact with other susceptible individuals.

For influenza, airborne spread predominates amongst crowded populations in enclosed spaces. It is generally accepted that influenza viruses are spread primarily by small-particle aerosols of virus-laden respiratory secretions that are expelled into the air (generally up to 1 m) and deposited on the mouth or nose of people nearby by infected persons during coughing, sneezing, or talking (Cox & Subbarao 2000). It was assumed that the attack distance is 2 m the attack duration is 10 minutes (see Table 6.2). The infection may also be spread from person to person by direct contact with infected secretions.

TABLE 6.2: Parameters for influenza in an example application of ISTAM.

Parameters	Value
Latent period	48-144 units ^a
Incubation period	96-192 units ^a
Infectious period	144-288 units ^a
Infectivity Index ^b	0.04
Mortality ^c	0.0008
Attack distance	2.0 m
Attack duration	10 minutes

^a 1 unit=0.5 hours
^b the infectivity index was calibrated from R_0 [for influenza, $R_0 = 1.68$ (Longini *et al.*, 2004)].
^c Mortality was estimated from CDC (2004).

Related studies demonstrated high rates of infection in school-age children and the importance of schoolchildren as vehicles of infection within families (Cox & Subbarao 2000). Detailed statistics of an influenza epidemic at a boys' boarding school were reported in an article from the British Medical Journal (4th March, 1978). One infected boy initiated the epidemic on January 22 and it ended on February 4 with 512 of the 763 boys in the school contracting the disease. The SIR model was applied by Murray (1989) to study the spread of the flu epidemic in this school and he found that $R_0 = 3.78$. While within the whole population, the R_0 was estimated to be less than 2 (Longini et al. 2004).

Three pandemics have occurred in the twentieth century: the 1918 H1N1 pandemic, the 1957 H2N2 pandemic, and the 1968 H3N2 pandemic. A characteristic pattern for localized epidemics within a community is: begins abruptly, peaks within 2 to 3 weeks, ends within 5 to 10 weeks, and the overall attack rates are estimated to be 10-20%, or even 40-50% in populations such as schoolchildren or nursing home residents (Cox & Subbarao 2000).

It is possible to be vaccinated against influenza. However, due to the high mutability of the virus a particular flu vaccine formulation usually only works for about a year. The World Health Organization co-ordinates the contents of the vaccine each year to contain the most likely strains of the virus to attack the next year. The flu vaccine is usually recommended for anyone in a high-risk group who would be likely to suffer complications from influenza. There are no effective measures to control the spread of influenza in schools except to keep students who are symptomatic at home.

As reviewed by Cox (2000), quantification of deaths caused by influenza is complicated by the fact that death can not only arise from pneumonia and influenza but also from cardiopulmonary and other chronic diseases that can be exacerbated by influenza. The mortality occurs primarily in the elderly, but it can occur in all age groups. As reviewed by Cox (2000), at the global level, influenza activity peaks during the winter months in temperate regions, and in tropical regions influenza can occur throughout the year.

6.4 Simulation results

6.4.1 Calibration by the basic reproductive number

For diseases that have been studied extensively such as influenza and measles, R_0 can be found from public sources. As R_0 cannot be applied directly in an individual-based model, the infectivity index is used here to describe the infectivity of the disease agent in ISTAM instead.

The number of contacts per unit time can be obtained by micro-simulation inside the AB, the duration of infectiousness as a parameter can be obtained from the literature, and the transmission probability per contact is in direct proportion to the infectivity index (see Equation 1.1 and Equation 4.2). This means that the infectivity index can be calibrated by R_0 in ISTAM. For influenza, R_0 is estimated to be 1.68 (Longini et al. 2004). Thus, by fixing the attack distance to be 2.0 m and simulating with different infectivity indices, then the corresponding R_0 can be computed (see Figure 6.1). In reality, individuals who are infectious with symptoms will change their daily activities, such as staying home or visiting health care facilities. So in this application, the infectivity index was estimated to be 0.04 with a corresponding R_0 of 1.84.

When the infectivity index is fixed to be 0.04, simulated R_0 from different attack distances are shown in Figure 6.2. It is clear the larger the attack distance, the larger of the basic reproductive number.

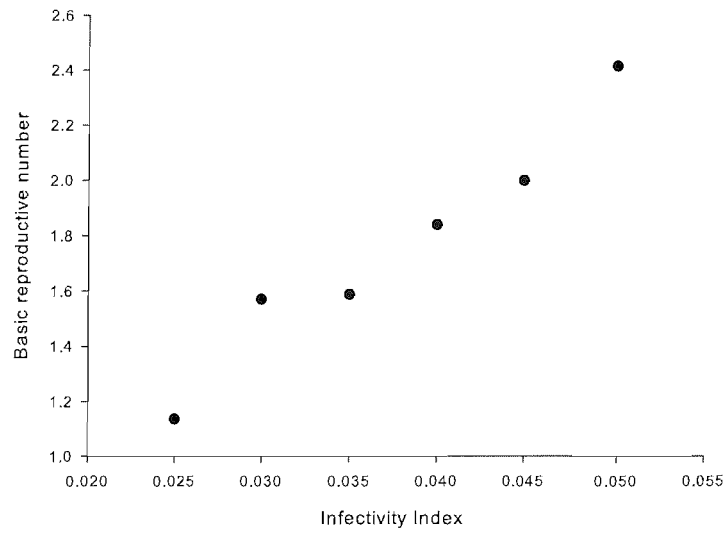


FIGURE 6.1: The relation between infectivity index and the basic reproductive number from simulation results.

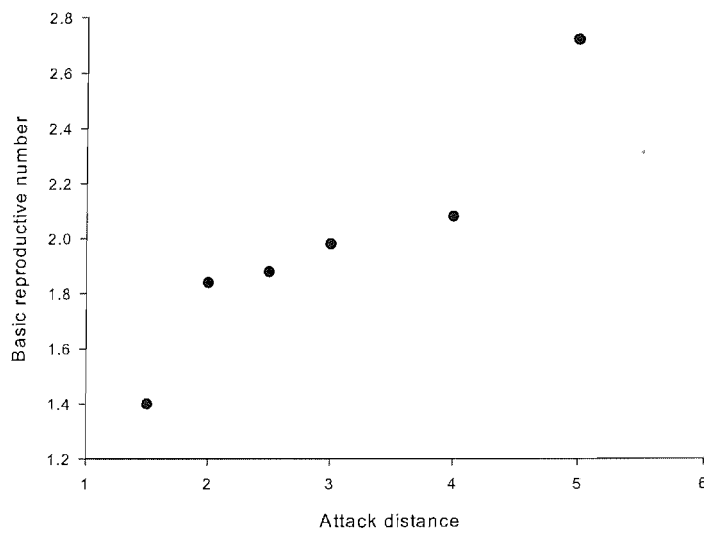


FIGURE 6.2: The relation between attack distance and the basic reproductive number from simulation results.

6.4.2 Base line simulation

The conditions for base line simulation are with 5 randomly distributed index cases, 3995 randomly distributed susceptible individuals, and no control measures, 0.04 infectivity index and 2 m attack distance. Figure 6.3 is an example plot obtained by base line simulation

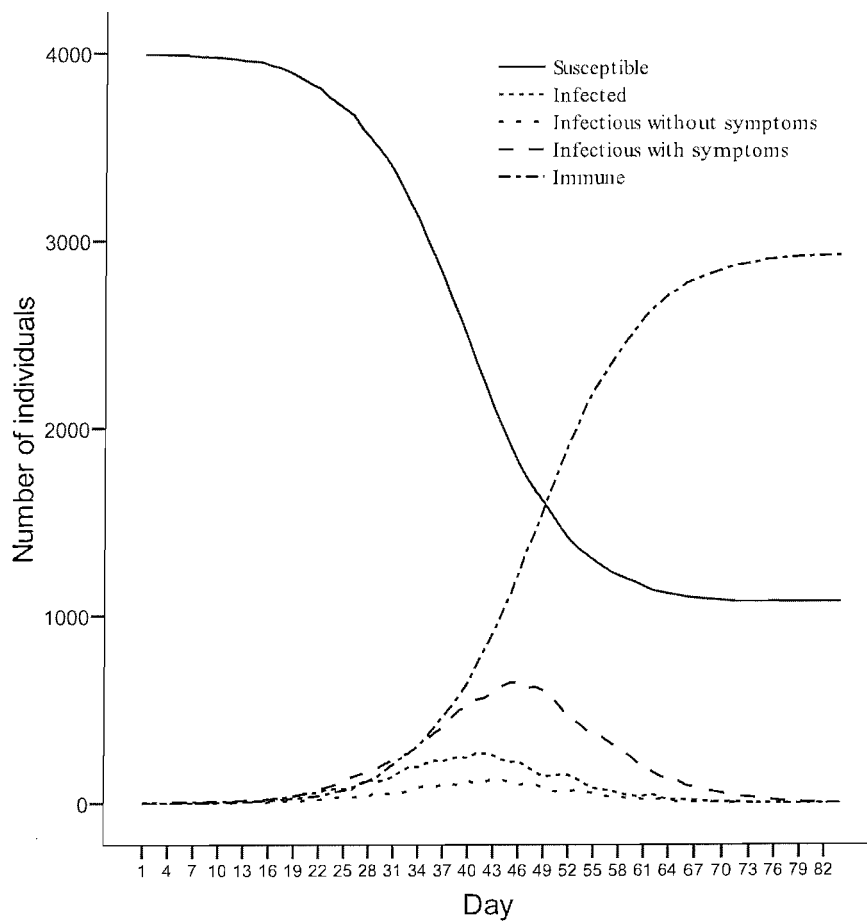


FIGURE 6.3: Example plot of base line simulation.

Due to the complexity of real world, it is almost impossible to predict precisely an epidemic outbreak, both the size and duration of the outbreak are highly variable (Kretzschmar et al. 2004). For one epidemic outbreak, five parameters are selected to describe its characteristics: (1) start day, the day when the number of infectious individuals with symptoms begins to exceed 10; (2) peak day, the day when the number of infectious individuals with symptoms reaches a maximum value during the whole

outbreak; (3) peak number, the number of infectious individuals with symptoms at the peak day; (4) duration, the duration between the start day and the last day when the number of infectious individuals with symptoms is larger than 10; (5) remaining number, the number of susceptibles after the outbreak (see Table 6.3).

TABLE 6.3: Value and description of five parameters for base line simulation.

Parameters	Description	Min	Max	Mean	Std. Deviation
Start day	The day when the number of infectious individuals with symptoms begins to exceed 10	6	59	11.51	5.4
Duration	The duration between start day and the last day when the number of infectious individuals with symptoms is larger than 10	62	84	70.14	3.8
Peak day	The day when the number of infectious individuals with symptoms reaches a maximum value during the whole outbreak	29	46	35.62	2.7
Peak number	The number of infectious individuals with symptoms at the peak day	545	722	645.95	31.75
Remaining number	The number of susceptibles after the outbreak	919	1169	1052.9	41.2

From Table 6.3, it is clear that the peak time is between 29-46 days, the whole period is around 70 days, and the overall attack rate is more than 70%. All these results are approximately higher (or longer) compared to the characteristic pattern above. Considering the high density of the University student population and without any control measures, this difference is deemed acceptable.

6.4.3 Number of index cases

It is interesting to explore the effect of the number of index cases on the characteristics of an outbreak. Here, the five parameters for one outbreak are compared with the number of index cases from 1 to 5 (with 100 simulations). When the number of index cases was set to 1, 36% of the simulations were terminated without any infections, that is, no outbreak. When the number of index cases was set to 2, 12% of the simulations had no infections. When the number of index cases was set to be 3, 3% of the simulations had no outbreak. When the number of index cases was more than 3, an outbreak occurs within most of the simulations.

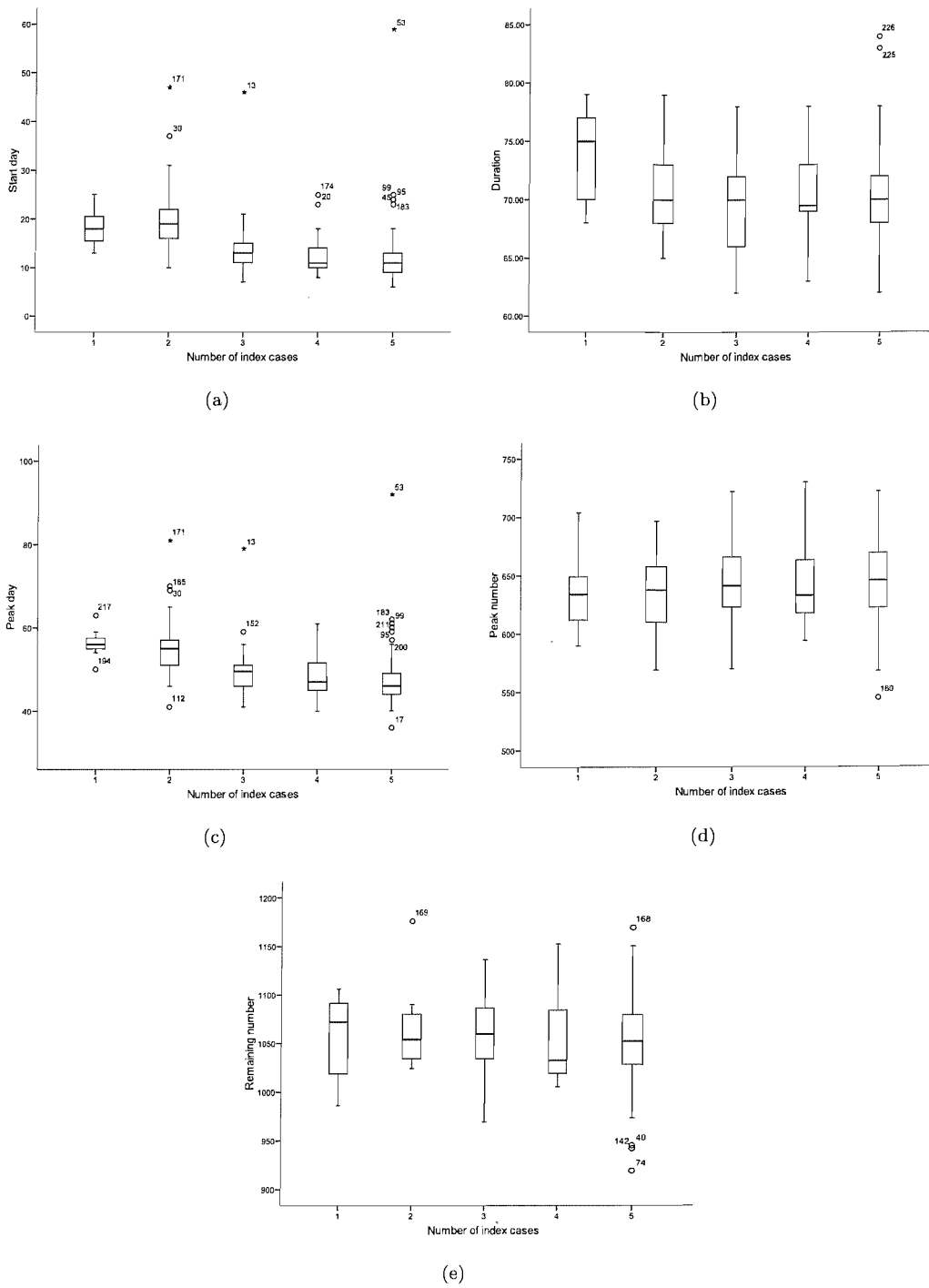


FIGURE 6.4: Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different number of index cases.

The distribution of each parameter for outbreaks are shown in Figure 6.4 based on the simulation results when the outbreak occurred. The most obvious difference are the parameters of start day and peak day: with the increase in the number of index cases, the start day and peak day come earlier. There are no significant differences in terms of the other three parameters: duration, peak number and remaining number.

The conclusion is that when no control measures are applied, the smaller the number of index cases, the larger the probability of no outbreak. But if the outbreak occurred, the larger the number of index cases, the earlier the start and peak of the outbreak. The number of index cases has no significant effect on the duration, peak number and remaining number.

6.4.4 Distribution type of the population and index cases

The base line simulation assumes students' allocation within accommodation has no correlation with their allocation within schools, that is, the probability of the students belonging to the same school to be within the same accommodation is the same as that for students belonging to different schools. In this research, this student population is named the 'normal population'. In reality, some Universities may try to allocate the students from the same school to be within the same flat (the basic unit of accommodation) or at least same accommodation if possible for the convenience of management (such as most of the Universities in China). In this research, this student population is named the 'clustered population'. On the contrary, some Universities may encourage students to live with students from different schools for the purpose of widening their horizons (such as the University of Oxford and the University of Cambridge). In this research, this student population is named the 'dispersed population'. It is interesting to explore the effects of the above three types of population allocation on the infectious disease transmission.

The base line simulation assumes the five index cases are randomly distributed, that is, they are randomly selected from all students without considering their accommodations, schools and relations between each other. It is assumed that all index cases become infected from outside at the same time. In reality, it is also possible for these index cases to have ‘linkages’ between each other. In this research, another two distribution types were assumed: the index cases are from (i) the same flat and (ii) the same school. It is interesting to explore the effects of the above three types of index cases allocation on infectious disease transmission.

The combination of the allocations of population and index cases produced nine sets of simulations (3×3). Each set was repeat 100 times.

For each of the five characteristic parameters, one way analysis of variance (ANOVA) (with 95% confidence) was used to test the differences in the means between different sets of simulation results (see Table 6.4). Interest is focused on the difference between different allocations of the population when the allocation of index cases is fixed (left part of Table 6.4) and the difference between different allocations of index cases when the allocation of the population is fixed (right part of Table 6.4).

TABLE 6.4: ANOVA analysis of the five characteristic for the combination of population allocations and index cases allocations.

Allocation of population	All types						Normal	Clustered	Dispersed			
	Random		Same school		Same flat		All types					
Allocation of index cases	F	p	F	p	F	p	F	p	F	p	F	p
Start day	1.99	0.14	<i>4.62</i>	<i>0.01</i>	1.93	0.16	1.45	0.24	0.74	0.48	<i>4.02</i>	<i>0.02</i>
Duration	<i>10.91</i>	<i>4E-05</i>	3.27	0.046	0.73	0.49	1.23	0.30	0.09	0.92	0.79	0.46
Peak day	<i>5.13</i>	<i>0.007</i>	0.99	0.38	1.40	0.26	1.79	0.17	0.05	0.96	1.51	0.23
Peak number	<i>21.69</i>	<i>4E-09</i>	<i>13.34</i>	<i>2E-05</i>	1.93	0.16	0.31	0.73	1.18	0.32	0.27	0.76
Remaining number	<i>28.25</i>	<i>3E-11</i>	<i>8.09</i>	<i>0.0008</i>	2.15	0.14	0.28	0.75	0.67	0.52	0.89	0.42

When the p value is < 0.05 , the larger the value of the test statistic (F), the larger the significant difference in means for the parameter within related sets (highlighted by italic font for F and p). It is clear that there are significant differences between different allocations of population when the allocation of index cases is fixed to be random or within the same school.

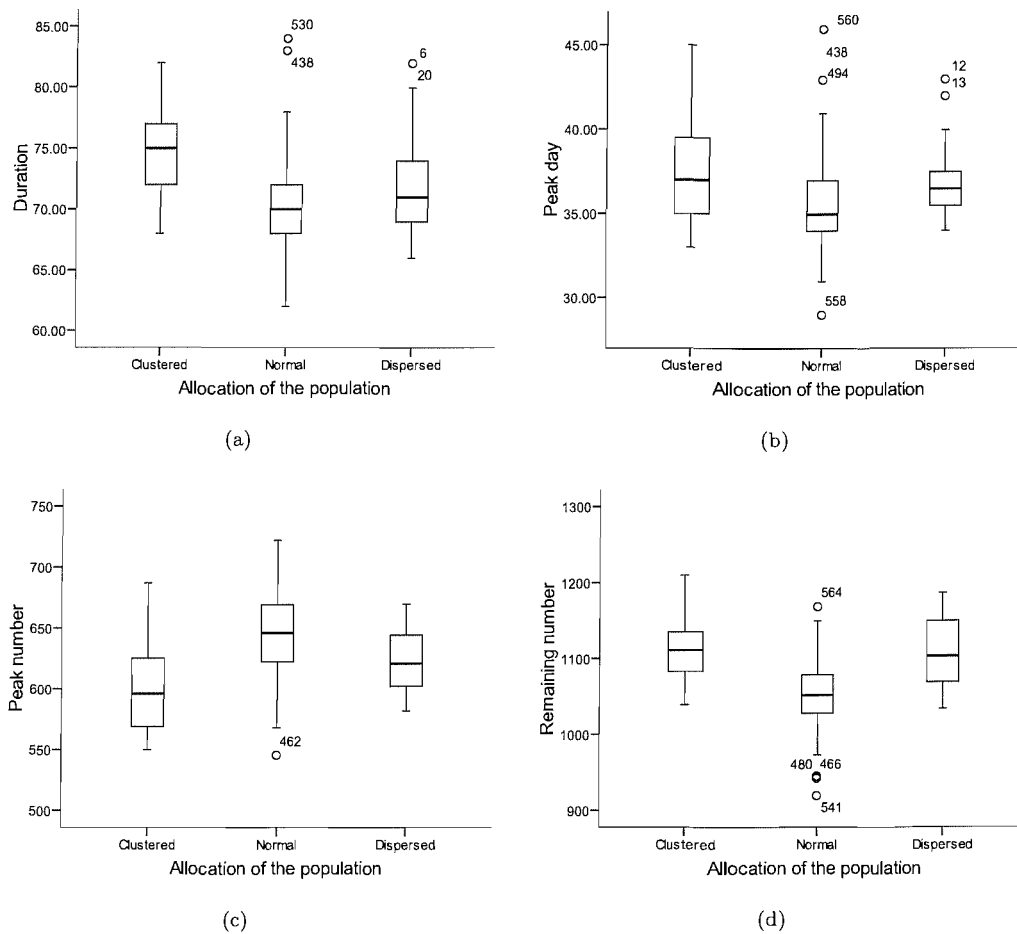


FIGURE 6.5: Distribution of (a) duration, (b) peak day, (c) peak number and (d) remaining number for different allocations of the population with random allocated index cases.

Figure 6.5 shows the boxplot of the four parameters for different allocations of population when the allocation of index cases is random.

It seems when the index cases are randomly allocated, clustered population may be a little ‘stronger’ than normal population in terms of the resistance to disease outbreak as the former has a larger remaining number and smaller peak number. A likely explanation is that within a clustered population, students have less other students to contact during a day (their flatmates are also classmates) compared with a normal population.

Figure 6.6 shows the boxplot of the four parameters for different allocations of population when the allocation of index cases is within the same school. In terms of the

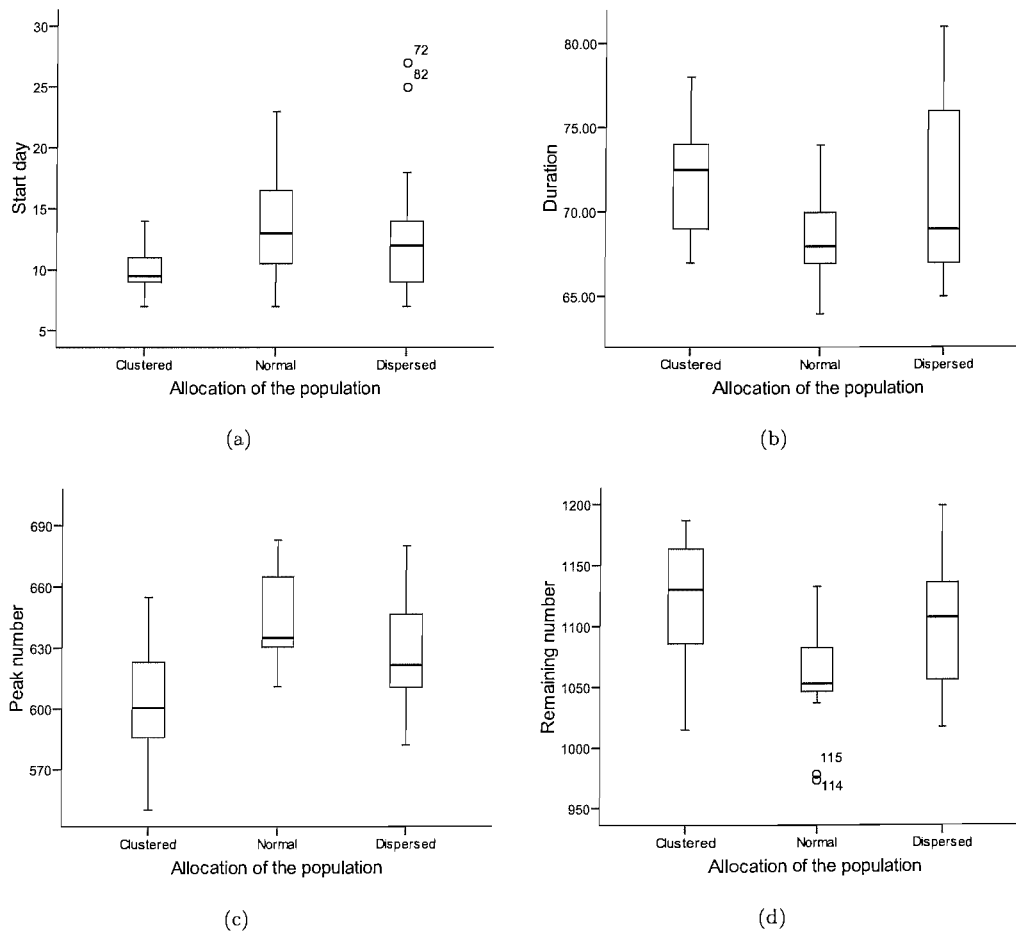


FIGURE 6.6: Distribution of (a) start day, (b) duration, (c) peak number and (d) remaining number for different distribution of population with same school index cases.

parameters of duration, peak number and remaining number, the difference patterns between different allocations of population when the index cases are within the same school is the same as when the index cases are randomly allocated, and the difference becomes a little more significant. A likely explanation is that when index cases are from the same school, this makes some effective contacts ‘overlay’, that is, there is more chance that one infectious student comes into contact with students who have just been infected by another infectious student (they all belong to the same school).

6.4.5 Control measure tests

Contact tracing was tested by ISTAM. As Eames (2003) claimed, ‘Contact tracing has proved to be a highly successful strategy when the number of infectious cases is low’.. Simple control measures have been implemented in ISTAM. ISTAM provides the option to quarantine the classmates or flatmates of symptomatic individuals. The implementation of the control measures is triggered by alert value (a pre-established value, when the number of symptomatic students is more than this value, this control measure will be applied). From the AB analysis in Section 6.4.7, it is clear that most of the infections occur in lecture rooms, bars, the student union, kitchens and refectories (see Table 6.5). A period of infectiousness without symptoms exists for influenza. This makes isolating symptomatic students and tracing and quarantining their contacts (flatmates and classmates here) a possible control measure. A 3-digit number is used to label different control measures, each digit indicating one target control population (from high to low sequence, the target control population is all individuals who are infectious with symptoms, all flatmates of individuals who are infectious with symptoms and all classmates of individuals who are infectious with symptoms). A value of 1 means quarantine control measure and value of 0 means no control measures. If one individual is quarantined, her/his activity pattern will be changed, she/he will stay within their accommodation for the whole day, and the only chance to come into contact with other students is to cook in the kitchen.

Figure 6.7 shows the difference in five parameters between four types of control measures (i.e., 100, 110, 101 and 111). It is clear that the effect of control measures 100 and 110 are similar in terms of almost all five parameters. The effect of control measures 101 and 111 are also similar in terms of four parameters (the only exception is the start day). Generally, 111 and 101 are more effective than 100 and 110. By control measures 111 and 101, the outbreak terminates earlier (between 10 to 15 days) than by control measures 110 and 100 (between 20-100 days). At the end of the outbreak, by control measures 111 and 101, the remaining number is higher (about 99% students were not

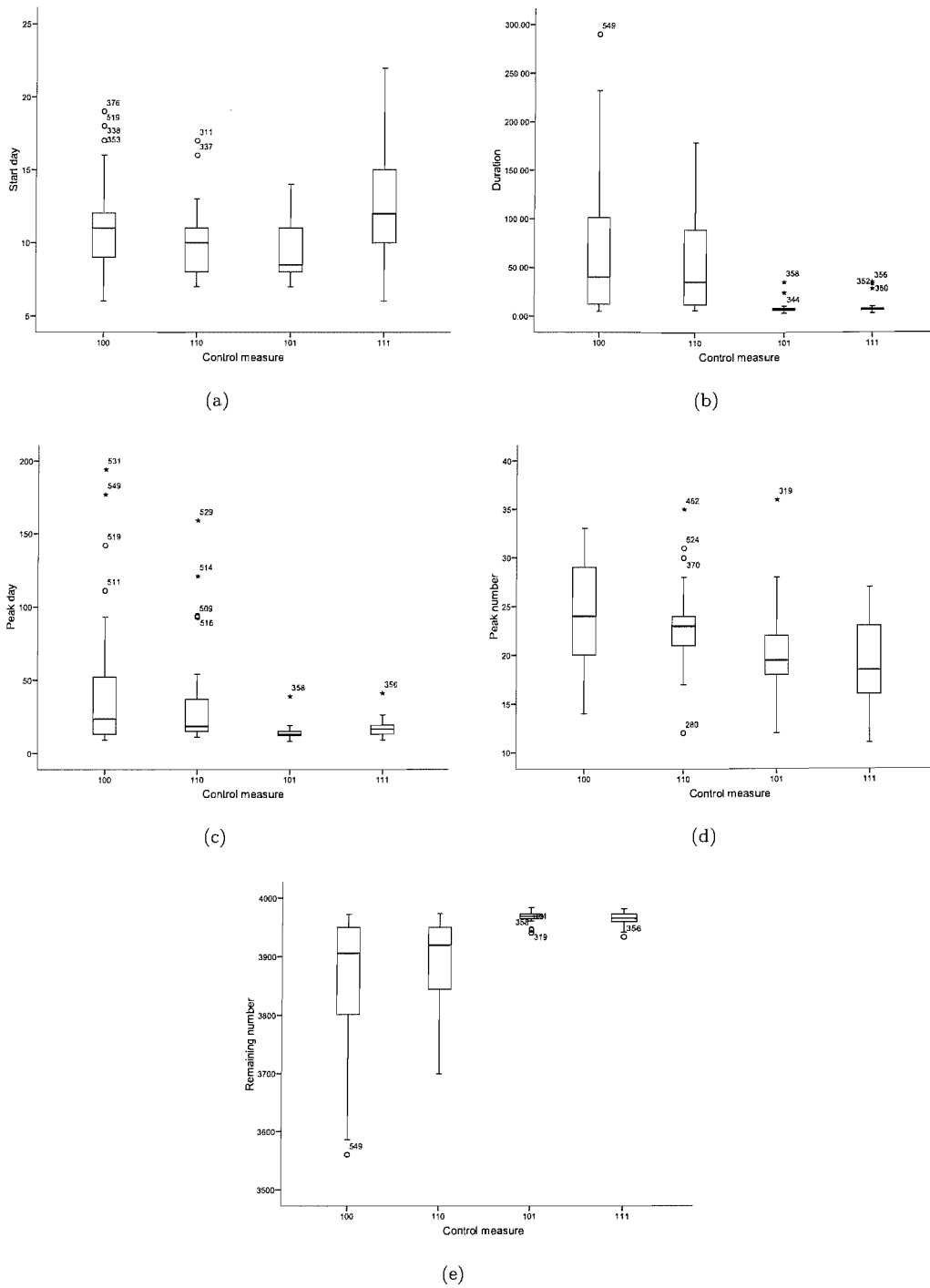


FIGURE 6.7: Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different control measures.

infected on average) than than by control measures 110 and 100 (about 97.5% students were not infected on average). These analyses also show that to quarantine all flatmates of individuals who are infectious with symptoms is not as effective as to quarantine all classmates of individuals who are infectious with symptoms. A possible explanation is that students do not spend much time within their accommodation and the relatively smaller number of flatmates (between 5 and 9) compared with the number of classmates (between 50 and 300).

In reality, delays will occur between a patient becoming ill and being isolated. This delay obviously will reduce the efficacy of the control measures. To control an infectious disease outbreak, the earlier the control measure can be applied, the better. Different values of alert values (values as: 1, 5, 10, 15, 20 and 25) were tested (see Figure 6.8). The results are well matched with expectation. Generally, the larger of the alert value, the longer the duration, the later the start day and peak day, the larger the peak number and the smaller the remaining number. Also, the value of R_0 when the alert value is set to be 1 was computed to be 0.41, which is consistent with an R_0 value of 1.84 when no control measures were applied. The reason is from the time lines: average infectious period is assumed to be 4.5 days, average incubation period 3 days and average latent period 2 days. This means that the average infectious without symptoms period is 1 day and average infectious with symptoms period is 3.5 days. So if any individuals who are infectious with symptoms will be quarantined at once, the decrease of R_0 is in proportion to the decrease of the infectious period during which she/he can make contacts with other susceptible individuals.

6.4.6 Contact network analysis

The contact network can be obtained from ISTAM simulation. Taking individuals as nodes and the contacts between them as edges, a contact network can be built from ISTAM. There can be multiple-contacts with different times and contact indices between

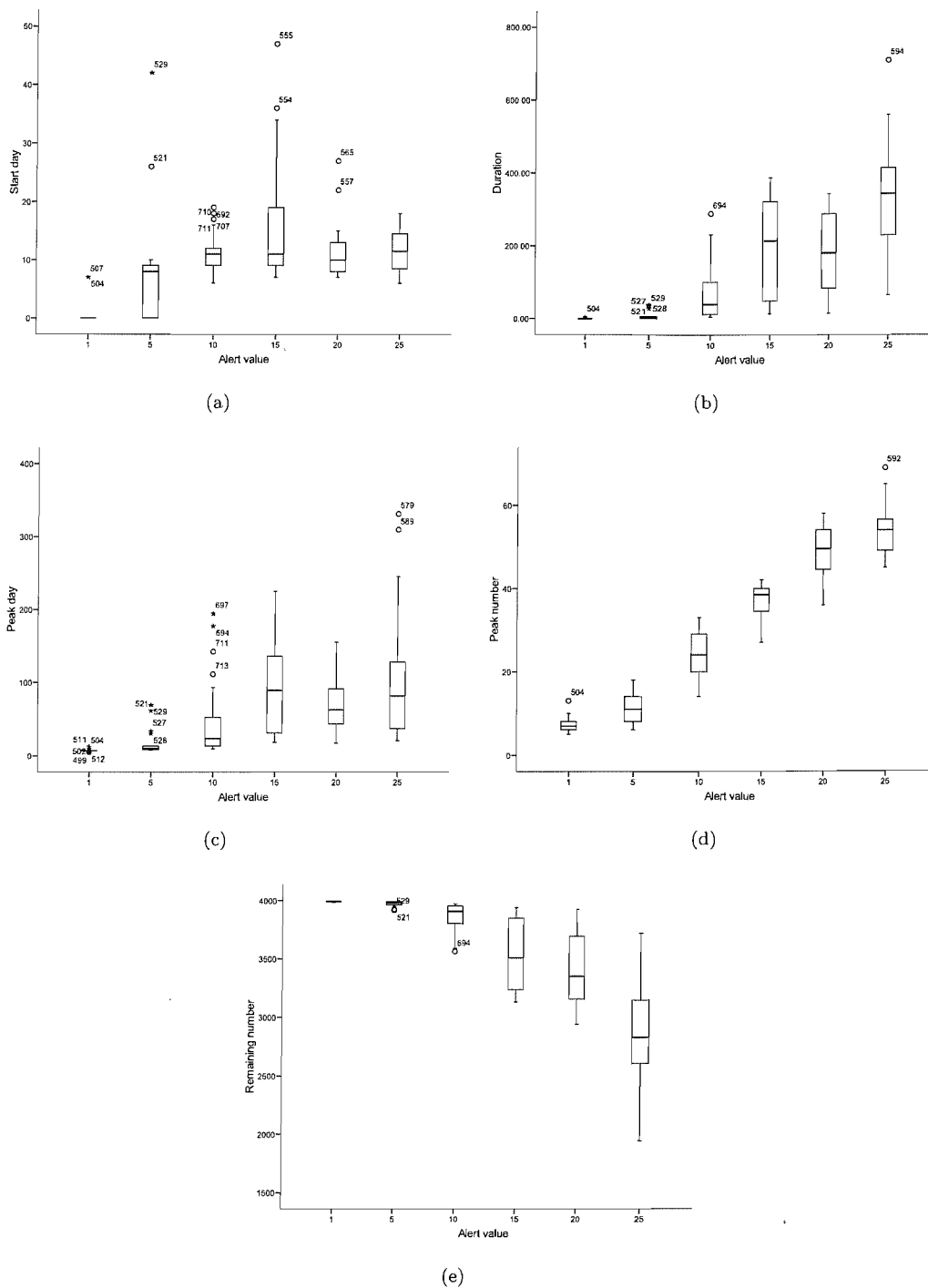


FIGURE 6.8: Distribution of (a) start day, (b) duration, (c) peak day, (d) peak number and (e) remaining number for different alert values.

two students. Therefore, the contact network is a dynamic network. Here, the sum of all contacts between two individuals is taken as the value for the edge and the network is simplified to a static network corresponding to one specific period. A key advantage is that the individual's activity can be simulated during any period to provide the contact network without the influence of infectious diseases (since a person's activities are affected by the disease situation, actively or passively).

As Read and Keeling (2003) pointed out, the three most important properties of human disease-transmission are (1) finite number and variability of contacts, (2) small path lengths and (3) highly clustered contacts. Here an analysis (focusing on the three properties above) of the contact network obtained from ISTAM simulation is provided as a method to validate the ISTAM simulation model (Huang et al. 2004).

To describe the number of contacts (or degree) adequately, both the average and distribution (characterized by a distribution function $P(k)$, which gives the probability that a randomly selected node has exactly k edges) of number of contacts are required. One of the most accepted characteristic distribution patterns is provided by the scale free models with a power-law distribution (Albert & Barabasi 2002, Wong et al. 2006). Take a 50-day simulation result from ISTAM as an example: there are 2287189 contacts in total, so the average number of contacts for every individual every day will be $2287189/(50 \times 4000 \times 0.5) = 22.87$ (including multi-contact between the same two students), which is close to the average contact number per day of 16.8 estimated by Edmunds (1997) from a survey of 92 adults by questionnaire, and average of 23 daily contacts (including telephone conversations and letters) by the study of Pool and Kochen (1978). The degree distribution for one day is given in Figure 6.9. The distribution has a right-sided power-law tail.

Path length is the average number of contacts in the shortest route between two individuals in the network, with the average taken over all possible pairs of individuals. Small path length is accepted as one property of human disease-transmission (Read & Keeling 2003, Watts & Strogatz 1998). From the simulation results of ISTAM, for a one

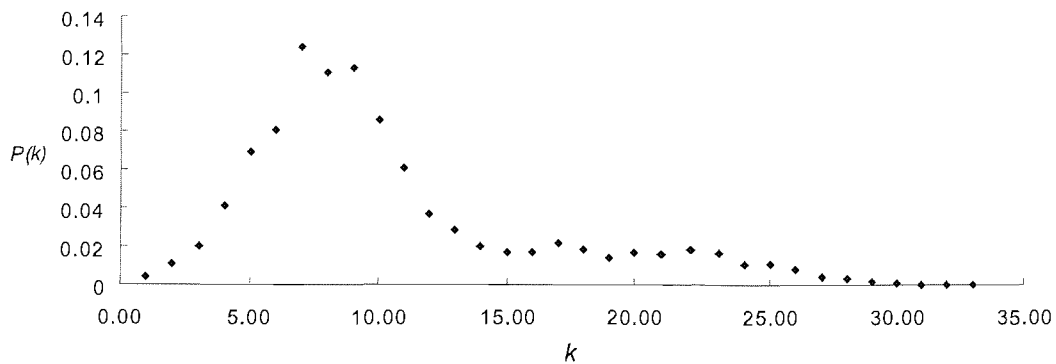


FIGURE 6.9: Contact network analysis: degree distribution of all students for one day's contact.

day period, characteristic path length was computed as 4.07 (averaged from 50 single days) amongst 4000 students. The value is relatively small, and therefore, consistent with the expectations of Read and Keeling (2003).

The clustering of a network can be measured by the clustering coefficient (defined by Watts and Strogatz, 1998). For one single day, the average clustering coefficient was 0.35 (range from 0 to 1, and standard deviation of 0.20); for a 50-day period, the average clustering coefficient was 0.35 (range from 0.10 to 0.96, and standard deviation of 0.13). Pool and Kochen (1978) estimated that any person contacted by one of their study participants knew between 8% and 36% of the other people contacted by that participant. This is consistent with the present study.

The conclusion can be drawn that the contact network from ISTAM exhibits the characteristic properties of a small world network of being highly clustered and having small characteristic path lengths (Watts & Strogatz 1998) and the characteristic properties of a scale free network as having degree distribution with a power-law tail (Albert & Barabasi 2002). This is consistent with most related research (Newman 2002, Newman et al. 2002, Keeling & Eames 2005).

6.4.7 AB simulation analysis

Few disease transmission models consider the effects of individual contact within a small-scale space. The simplest approach is to assume individuals are distributed uniformly in small-scale space. Thus, the probability of contact is determined by the human density. For a continuous space with width of W and length of L , if individuals are distributed randomly (assume that one individual's location does not affect others and the duration is longer than the attack duration and movements are not considered), the number of effective contacts is directly proportional to the number of susceptible and infectious individuals inside. To simplify the problem, if the number of infectious individuals is fixed to be 1 and the number of susceptible individuals is n , then the number of effective contacts $P(n)$ is:

$$P(n) = kn \quad (6.1)$$

Where: $k = \frac{\pi R^2}{WL}$ (R is the attack distance of a specific infectious disease).

The above 'density' assumption is easy to understand and compute but overly simple compared with the real situation. Representing reality more closely, the AB simulation presented here considers the dynamics of individuals, physical structures and stochastic factors. Figure 6.10 shows the number of effective contacts plotted against the total number of individuals (with one index case, attack distance R is 2 m) within six types of AB obtained by within-AB simulation. Due to different width, length, and minimum x and y values, the gym, computer room, lecture room, bar, student union and library room have different capacities in ISTAM. To compare with the 'density' assumption made at the beginning of this section, in Figure 6.10, grey straight lines show the number of effective contacts against the number of susceptible individuals inside a continuous space of the same size as the corresponding AB under the same conditions. It is clear from Figure 6.10 that the simulation for a gym (Figure 6.10a) results in a plot that is closest to the grey line produced by the 'density' assumption (and for which the fitted line crosses the ordinate close to the origin) because it is assumed that individuals are distributed randomly. The simulation for a computer room (Figure 6.10b) results in

a plot that is lower than the grey line. Further, a threshold of susceptible individuals is needed for the occurrence of contacts. This can be accounted for by the decision to model the distribution of individuals based on individuals' trying to be as far as possible from each other. The simulation for a lecture room (Figure 6.10c) results in a different plot to that for the density assumption because a special spatial distribution is applied for a lecture room. The simulation for a bar (Figure 6.10d) is also similar with the grey line. However, the slope of the plotted points is less than that of the grey line. The reason rests with the modelling of individuals as clustered and the small minimum distance in both x and y directions. For the simulation for student's union and the library room (Figure 6.10e and 6.10f), the slope of the plotted points is greater than that for the density assumption line. This is because the value of M_f is 3. The movement of individuals increases the probability of effective contacts.

The contact probability should be sensitive to both the attack distance and attack duration of the target disease in any disease transmission model. AB simulation can achieve more than this: changing the attack distance or attack duration has effects of different magnitude for different types of ABs. Taking influenza as an example, if the attack distance is set to be 1 m, 1.5 m, 2 m, 2.5 m, 3 m, 3.5m or 4 m, different infected numbers (including the latent and infectious without or with symptoms) over time are gained. The larger the attack distance, the more (and more quickly) that individuals are infected. It is interesting to see how the number of infections which result from different locations (ABs) vary as a function of attack distance (see Table 6.5, Figure 6.11 and Figure 6.12), the most prominent characteristic is that when the attack distance increases from 1 m to 2 m, the percentage of infections occurring at a bar decreases while the percentage of infections occurring at a lecture room increases. When the attack distance is set to 1 m, no infections occurring in a gym. With any value for attack distance, infection seldom occurs in campus.

The population consists of groups that have frequent contacts within groups (local contacts) and relatively rare contacts between groups (global contacts) (Wallinga et al. 1999). In ISTAM, two kinds of social network exist between students: classmates and

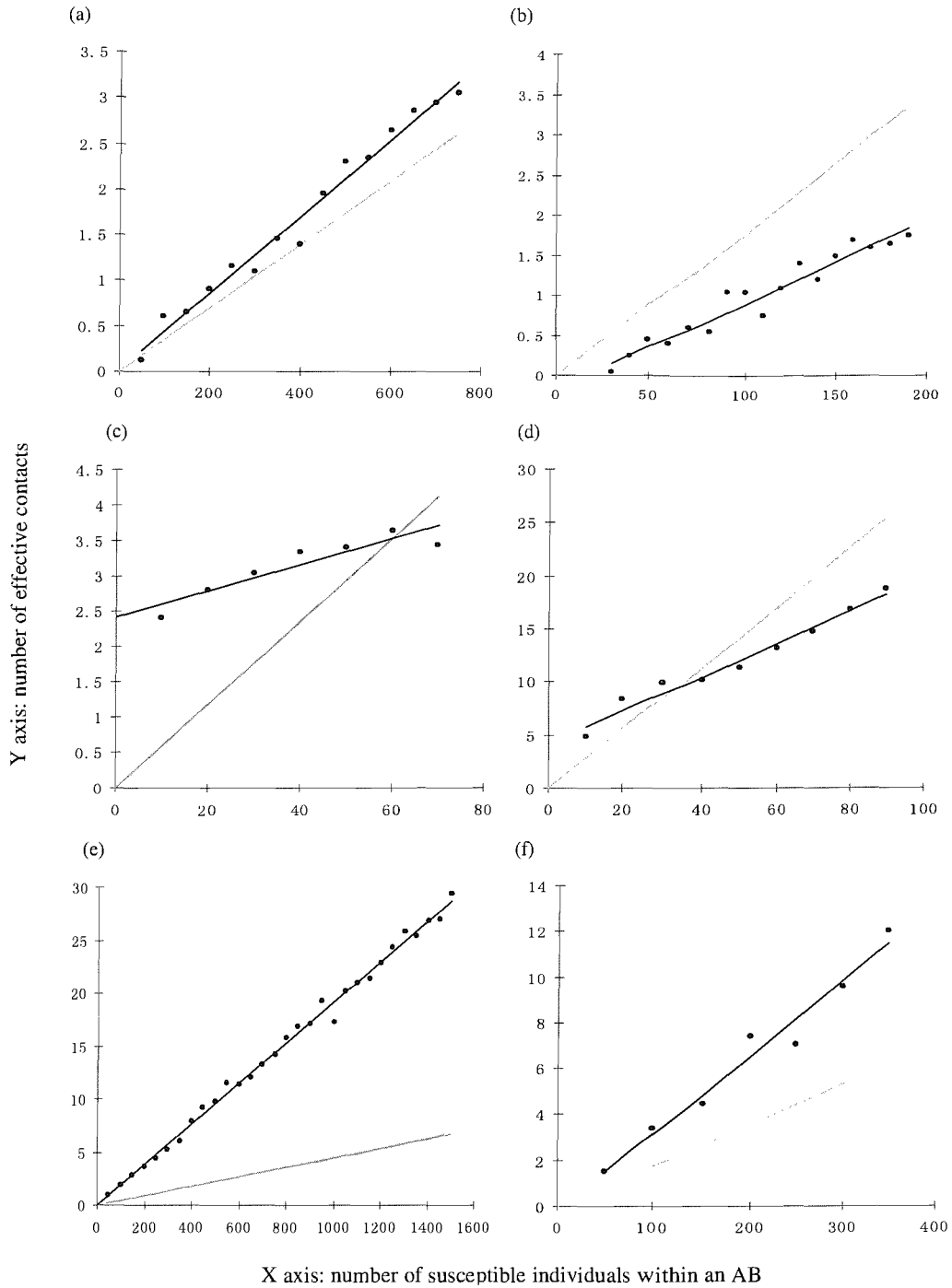


FIGURE 6.10: Simulation plots of within-AB for six different ABs: (a) gym, (b) computer room, (c) lecture room, (d) bar, (e) student union and (f) library room. Black dots correspond to simulated points, straight black lines correspond to trend lines, and straight grey lines are computed on the basis of the 'density' assumption (see text).

Simulation starts with one index case and an attack distance 2 m.

TABLE 6.5: Percentage of infections from different types of ABs with different attack distances.

Attack distance \ Activity bundle	1 m	1.5 m	2 m	2.5 m	3 m	3.5 m	4 m
Bar	0.547466	0.430082	0.372247	0.373264	0.375009	0.383031	0.380802
Computer room	0.003319	0.002876	0.006875	0.00825	0.007865	0.008226	0.007714
Gym	0	0.001607	0.000924	0.00161	0.001163	0.000952	0.001342
In campus	0	0	0.000205	6.71E-05	0	6.8E-05	0.000268
Kitchen	0.043815	0.031639	0.039746	0.039708	0.03878	0.038684	0.039576
Lecture room	0.248064	0.429659	0.461759	0.455363	0.458929	0.4536	0.449021
Refectory	0.071255	0.044328	0.048331	0.049031	0.050544	0.048814	0.047894
Library room	0.018809	0.014973	0.018436	0.018512	0.016483	0.016113	0.019453
Student union	0.067272	0.044835	0.051478	0.054195	0.051228	0.050513	0.053931

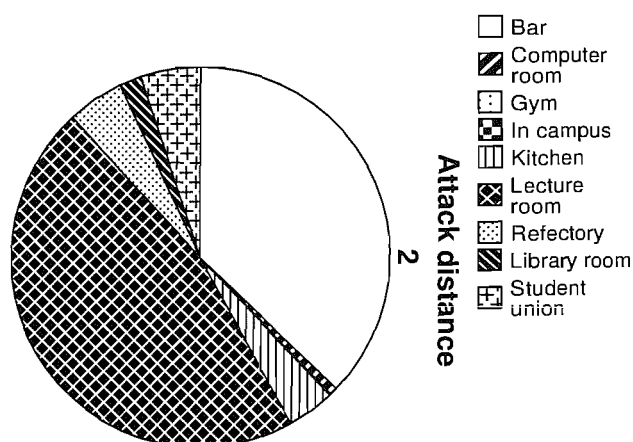


FIGURE 6.11: Percentage of infections from different types of ABs with an attack distance of 2 m.

flatmates. Infection occurring within the accommodation (kitchen, shared by flatmates) and the lecture room (shared by classmates) can be taken as infection from local contacts and other sources of infection can be taken as from global contacts. However, these local contacts account for only about 50% (Table 6.5, attack distance of 2 m) of all infections. This conflicts with common sense that most human contacts tend to remain in a limited social group (Wallinga et al. 1999). The first possible reason for the discrepancy is in fact, that more social networks exist, such as friends, partners, or society generally. The second reason is that when students go to the bar or other places, they still prefer to go with the people within their network, but this has not been expressed in ISTAM.

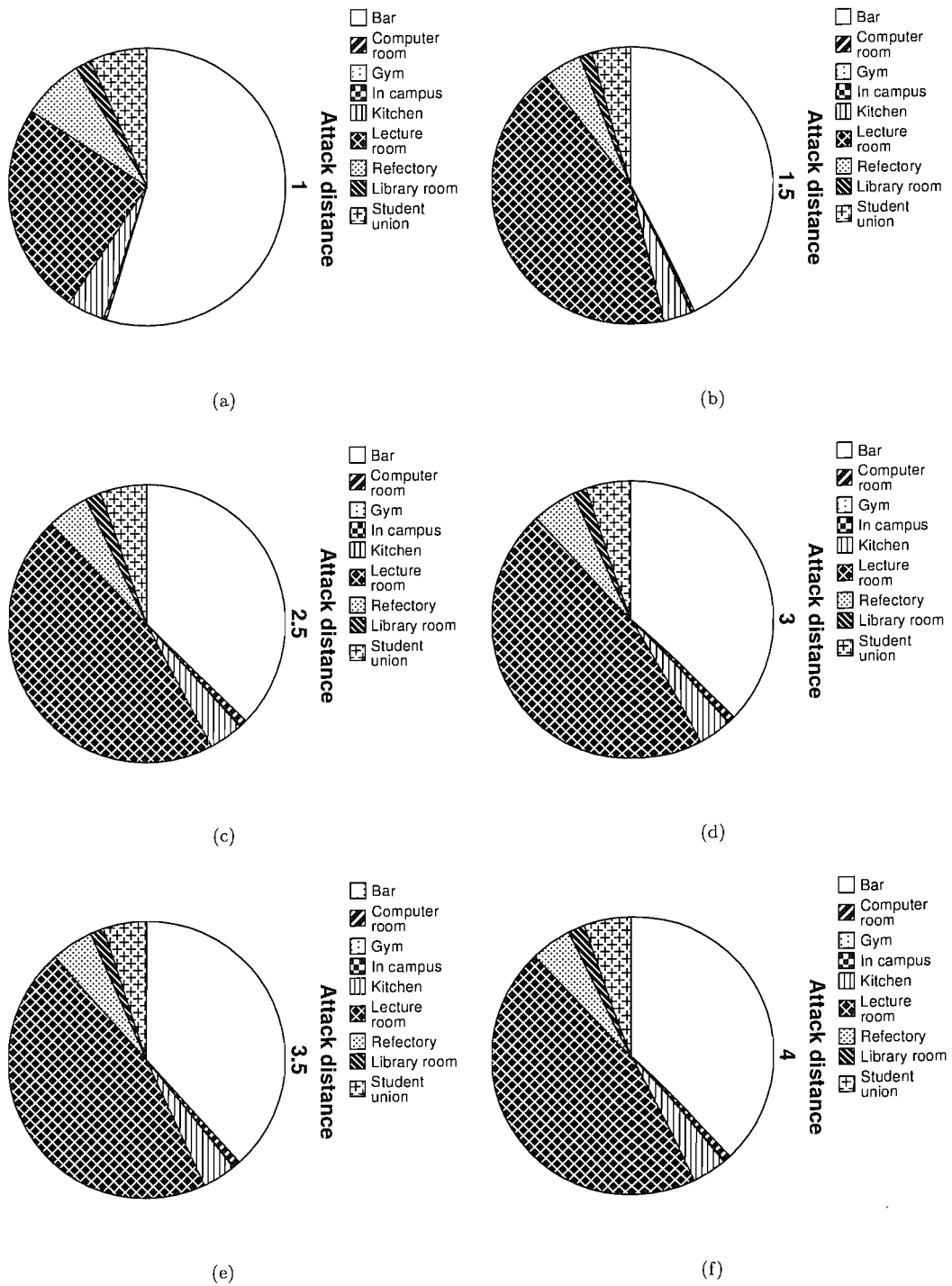


FIGURE 6.12: Percentage of infections from different types of ABs with different attack distances: (a) 1 m, (b) 1.5 m, (c) 2.5 m, (d) 3 m, (e) 3.5 m and (f) 4 m.

Understanding the main source of infection is a prerequisite to the design of effective control measures. From the analysis above, if not considering expense, control tracing should be extended to both the flatmates (with 4.0 % infections) and classmates (with 46.2 % infections) of the symptomatic students and the bar (with 37.2 % infections) could be closed.

6.5 Summary

The first-year undergraduate student population body was chosen to be the first test bed of ISTAM. The speciality of the targeted population and environment make some of the assumptions plausible and 'simplifies' the complexity of the real world. On the other side, this speciality means the analysis results from the above application are hard to apply to the more general cases. In Chapter 8, another application case of ISTAM (to a city) is presented.

Chapter 7

Parameter Exploration of Raster space AB simulation

7.1 Introduction

The aim of this chapter is to explore the relation between transmission of airborne infectious diseases and humans' space-time dynamics at fine space-time scales by parameter exploration of raster space AB simulation which was implemented during the application of ISTAM to the campus of University of Southampton (see Chapter 6). To quantify and compare the significance of the properties of humans' space-time dynamics to infectious disease infection can further understanding of the transmission process at within-AB level and effective quantitative comparison of the infection probabilities within different types of AB can assist in the design and application of control measures before and during epidemics. Influenza, one common airborne infectious disease was used for this exploration. To simplify the question, some assumptions were made. It was assumed that if the contact distance between infectious and susceptible individuals was less than

2 m, then infection will occur. For example, if the MinX and MinY are set to be 1 m, one infectious individual can have effective contacts with his or her eight neighbours within the Moore neighbourhood (see Figure 7.1, location A). Here, the contact duration was assumed to be less than the duration of one step in the simulation, and, therefore, did not need to be considered further. To simplify, if two individuals come into contact more than once during a time unit (such as within a dynamic AB), it is taken as one effective contact. An AB is simplified to be a rectangular space and the effect of the physical conditions within an AB on the occurrence of infection was ignored, such as the air circulation, temperature and humidness.

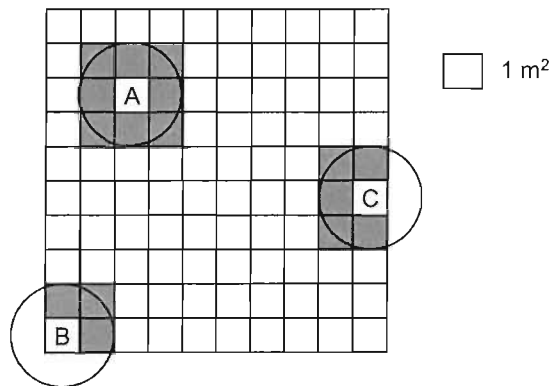


FIGURE 7.1: Different sizes of infection area based on the relative location of infectious individuals within an AB.

Here, the output variable was set to be the proportion infected (P_i) which is the ratio of (1) the number of susceptible individuals who make effective contacts with infectious individuals (infection may occur) during the simulation duration to (2) the total number of susceptible individuals before the simulation. The input variables (see Table 7.1) included the parameters required to describe individuals' space-time dynamics as mentioned in Section 4.3.1 and two additional variables; the number of index cases (N) and proportion occupied (P_o). Proportion occupied is the ratio of (1) the number of current individuals to (2) the maximum number of individuals that can be accommodated (which is determined by the size of AB and size of cell). Proportion occupied is used to describe the saturation of the AB. To simplify, it was assumed that no immune individuals existed within the target AB. To make the result of the analysis more generally applicable,

the parameters of Width (W), Length (L), MinX and MinY were transformed to be functions of the size of AB (WL), ratio of AB ($\frac{W}{L}$), size of cell ($MinX \times MinY$) and ratio of cell ($\frac{MinX}{MinY}$). Every simulation was repeated 1000 times to obtain the average value.

TABLE 7.1: Input variables, their default values and value ranges.

<i>Variables (Symbol)</i>	<i>Default value</i>	<i>Range of value</i>
Spatial distribution type (T)		1, 2, 3, 4
Proportion occupied (P_o)	0.5	0.05, 0.1, 0.15, ..., 0.95
Number of index cases (N)	1	1, 2, ..., 10
Ratio of AB (R_{AB})	Width or Length = 10 m	Width or Length: 1, 2, ..., 10, 11, 13, 14, 17, 20, 25.
Size of AB (S_{AB})		33, 50, 100 m
Ratio of Cell (R_c)	MinX or MinY = 1 m	MinX or MinY: 0.5, 1, 1.5 m
Size of Cell (S_c)		
Movement frequency (M_f)		1, 2, ..., 5
Movement proportion (M_p)		0.1, 0.2, ..., 0.9

7.2 Proportion occupied

This section explores the relation between proportion occupied and proportion infected for all types of spatial distribution. The value of proportion occupied was varied from 0.05 to 0.95 with a step value of 0.05. All other variables were fixed to the default values in Table 7.1. Figure 7.2 shows the relation of proportion infected with the proportion occupied with default values for all other parameters.

7.2.1 Type 1

It is clear that Type 1 AB (e.g., computer room) displays a positive asymmetric and non-linear relation which is quite straightforward to interpret: individuals in Type 1 try to be as far as possible from each other and when the proportion occupied is large

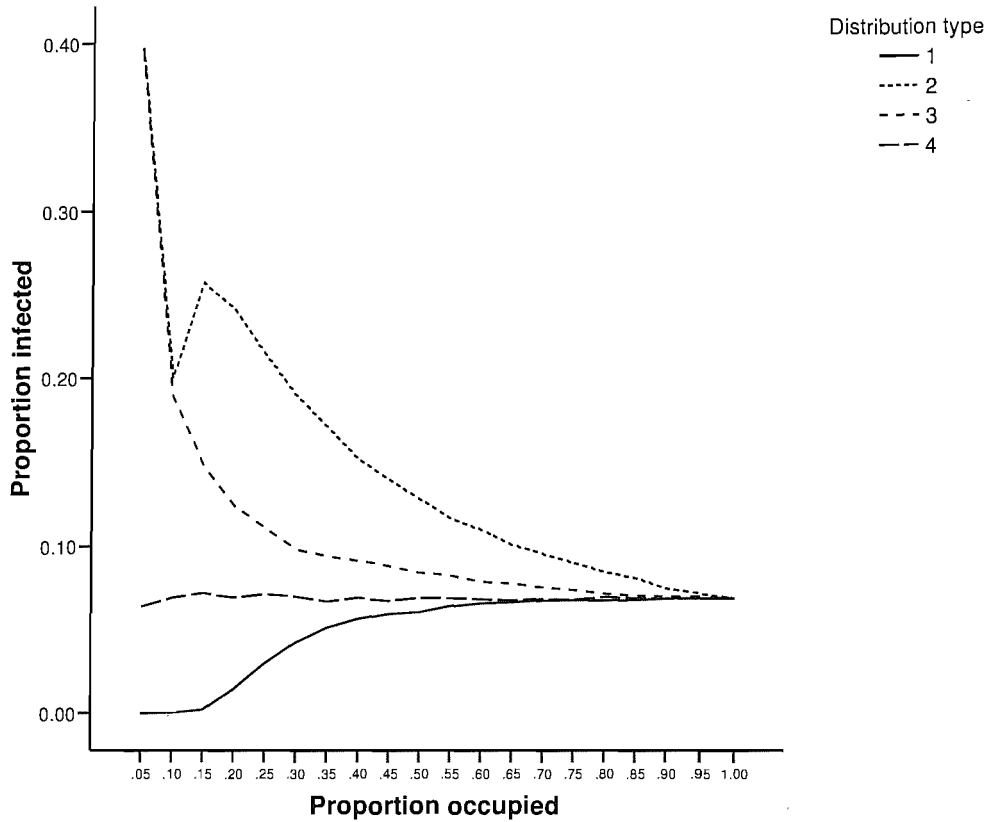


FIGURE 7.2: Proportion infected plotted against proportion occupied with default values for other variables.

enough there is no space for individuals to avoid each other so the proportion infected approaches a constant asymmetrically. When $P_o < 0.15$, individuals within the AB can always find a location where no effective contacts can occur, so P_i remains as 0. When $P_o > 0.6$, most individuals locate themselves randomly. When $P_o \in (0.15, 0.6)$, the percentage of individuals who can keep the neighbourhood distance decreases with an increase in P_o . Regression analysis confirmed an inverse relationship between P_i and P_o (with $R^2 = 0.992$):

$$P_i = 0.094 - \frac{0.016}{P_o} \quad (7.1)$$

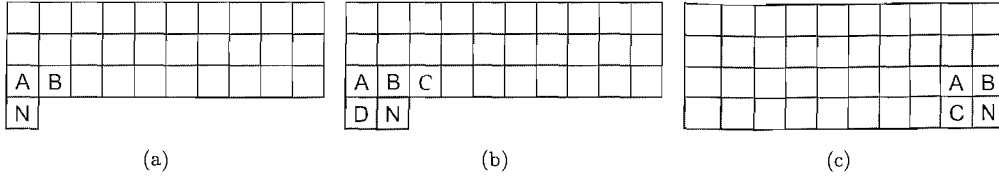


FIGURE 7.3: The relation of number of newly added effective contacts with the location of the n^{th} individual (individual N is the n^{th} individual and individual A, B, C and D are individuals whose effective contact numbers increase by 1 due to the newly added individual).

7.2.2 Type 2

Type 2 AB (e.g., lecture room) displays a periodicity which ranges between 0.10 and 0.20 of the proportion occupied. The reason is that the width of this AB is 10 m and MinX is 1 m such that 10% proportion occupied implies ten individuals which take up just one row in a Type 2 AB. In fact, according to the mathematical definition of Type 2, the plot can be computed without real simulation as following: If the total size of individuals within the AB is n , C_n is the total number of effective contacts.

1. When $n \leq \text{Width}$, this means all individuals are in a line one by one. Individuals who are at the front and the end of this line have one effective contact and all others have two effective contacts. So $C_n = 2 + 2(n - 2) = 2n - 2$.
2. When $n > \text{Width}$, this means individuals are in at least two lines. For every time the total size of individuals increases from $(n - 1)$ to n , the number of newly added effective contacts (i.e., $C_n - C_{n-1}$) can be considered according to the location of the n^{th} individual with the line (Figure 7.3):

(a) At the front, $C_n = C_{n-1} + 4$;

(b) In the middle, $C_n = C_{n-1} + 8$;

(c) At the end, $C_n = C_{n-1} + 6$.

3. For any values of n , the $P_i(n) = \frac{C_n}{n}$.

As this can be certified from theory, there is no need to calculate R^2 for the simulated result.

7.2.3 Type 3

Type 3 AB (e.g., a bar) displays a negative relation. This can be explained by the fact that it is easy for infectious individuals to infect susceptible individuals within his or her own cluster, but hard to infect individuals from other clusters. Regression analysis confirmed an inverse relationship between P_i and P_o (with $R_2 = 0.992$):

$$P_i = 0.049 + \frac{0.016}{P_o} \quad (7.2)$$

For Type 3 AB, firstly, the cluster number was computed as the total individual number divided by g . Each cluster was allocated one individual (selected randomly) as a key member. The location of the key member was selected randomly from the whole AB. Secondly, all other individuals were allocated randomly to a cluster, with a location within the Moore neighbourhood of the key member as shown in the sequence of Figure 7.4. These two steps ensure that the average size of clusters equals g , and every cluster has at least one individual.

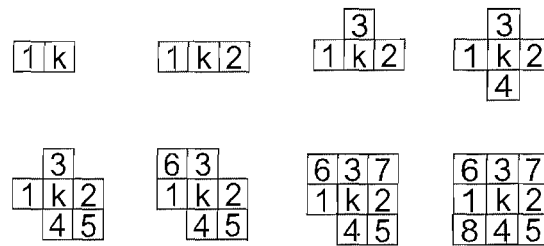


FIGURE 7.4: The shape of clusters with different g values (2-9) and the sequence of adding individuals around the key member.

If P_o is small enough that there are no effective contacts between different clusters, then all effective contacts occur between individuals within the same cluster. Based on the definition above, the shape of the clusters was determined by g (Figure 7.4). The average number of effective contacts ($\overline{C_n}$) can be computed based on the value of g . Table 7.2 lists the average number of effective contacts for individuals within clusters with different g values from 2 to 9. P_i can be computed as:

$$P_i = \frac{0.01\overline{C_n}}{P_o} \quad (7.3)$$

TABLE 7.2: Average number of effective contacts for individuals within clusters with different g values (2-9).

g	Number of effective contacts (according the sequence of $k, 1, 2, \dots$)	Average number of effective contacts
2	1, 1	1
3	2, 1, 1	1.33
4	3, 2, 2, 3	2.5
5	4, 3, 3, 3, 3	3.2
6	5, 3, 4, 3, 4, 3	3.67
7	6, 4, 4, 4, 4, 3, 3	4
8	7, 4, 5, 5, 4, 3, 3, 3	4.25
9	8, 5, 5, 5, 5, 3, 3, 3, 3	4.44

It is clear from Figure 7.5 that with an increase in g , the simulated data were closer to the theoretical curve provided by Equation 7.3. The reason is that with an increase in g , the rate of effective contacts between individuals from the same cluster to effective contacts between individuals from different clusters increases.

7.2.4 Type 4

For Type 4 AB (e.g., a gym), the proportion infected is approximately independent of proportion occupied. In this situation, the likelihood of one susceptible individual being within the infection distance of an infectious individual is determined only by the ratio of the infection area to the whole area of the AB (in this case the area of the AB is $100 m^2$). Figure 7.1 shows the difference in the size of infection area (the area where if individuals are present, effective contacts will occur with infectious individuals) according to the relative location of infectious individuals. Infection area is $8 m^2$ when an infectious individual is not at the border or corner of the AB (e.g., the location A) with a probability of $64/100$; Infection area is $5 m^2$ when an infectious individual is at the border of the AB (e.g., location B) with a probability of $32/100$; Infection area is $3 m^2$ when an infectious individual is at the corner of the AB (e.g., the location C) with a probability of $4/100$.

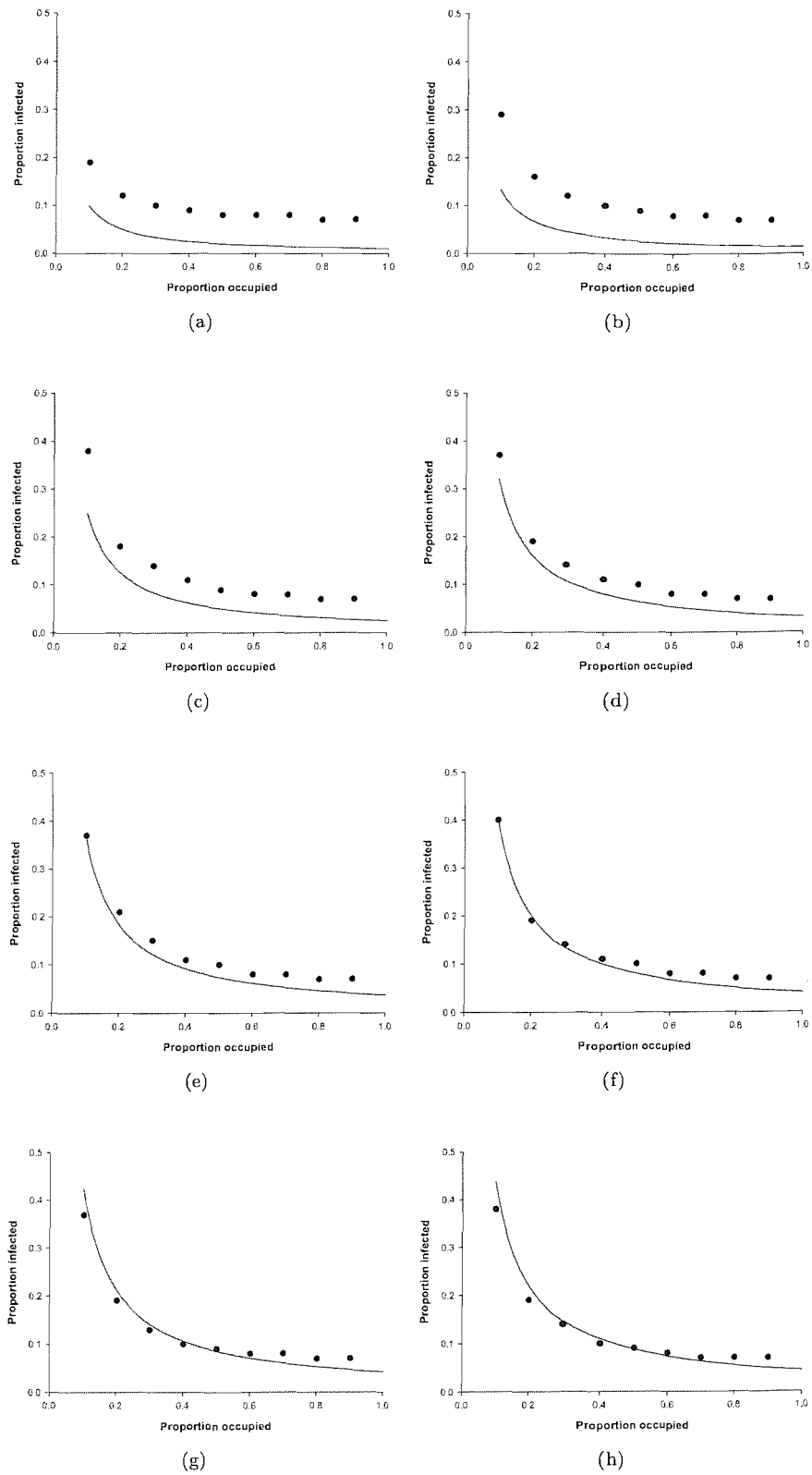


FIGURE 7.5: Proportion infected plotted against proportion occupied for a Type 3 AB with values of (a) $g = 2$, (b) $g = 3$, (c) $g = 4$, (d) $g = 5$, (e) $g = 6$, (f) $g = 7$, (g) $g = 8$ and (h) $g = 9$ (solid dots are the simulated results and the solid lines are based on Equation 7.3).

If the number of index cases is fixed to be 1, the probability of one susceptible individual having an effective contact will be $\frac{8}{100}0.64 + \frac{5}{100}0.32 + \frac{3}{100}0.04 = 0.0684$. According to the definition of the proportion infected and proportion occupied, the proportion infected is also 0.0684 which is very close to the average value of Figure 7.2. This value was defined to be a constant k as it was used frequently in the following sections. More generally, for a Type 4 rectangular AB with 1 m for MinX and MinY, the proportion infected will be computed by: $(\frac{4}{WL}3 + \frac{2((W-2)+(L-2))}{WL}5 + \frac{(W-2)(L-2)}{WL}8)(\frac{8}{WL})$. The equation is as follows:

$$P_i = \frac{8WL - 6W - 6L + 4}{W^2L^2} \quad (7.4)$$

7.3 Number of index cases

This section explores the relation between the number of index cases and proportion infected for all types of spatial distribution. The value of number of index cases was varied from 1 to 10 with a step value of 1 for the Type 4 AB (Figure 7.6) and varied from 1 to 5 with a step value of 1 for all other types (Figure 7.7). All other variables were fixed to the default values in Table 7.1.

It is not surprising that for every type of AB, the proportion infected increases with an increase in the number of index cases. Also, Figure 7.7 shows the independent relation between proportion infected and proportion occupied for Type 4 AB. The interesting point here is the increase in magnitude of the proportion infected over the increase in the number of index cases. Omitting the irregularity of proportion infected when the proportion occupied is close to 0 or 1, it is clear that in both Figure 7.6 and 7.7, the magnitude of increase in the proportion infected decreases for each increase in the number of index cases.

The Type 4 AB was selected to be an example for the quantification of the relation between the proportion infected and the number of index cases. The value of proportion

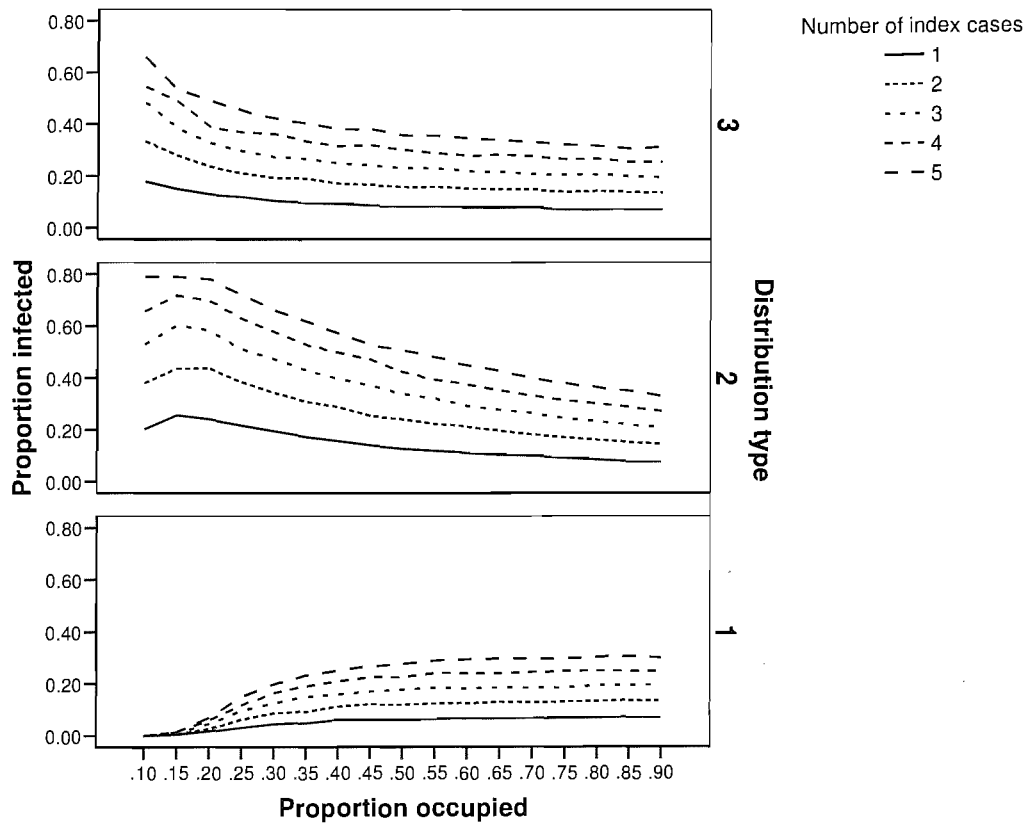


FIGURE 7.6: Proportion infected plotted against proportion occupied with different numbers of index cases for Type 1, 2 and 3 ABs.

infected was averaged over the proportion occupied (from 0.25 until 0.75) for each of the number of index cases from 1 to 10 (Figure 7.7). These values are displayed (solid dots) in Figure 7.8. Then, a mathematical relation was obtained for Type 4 ABs. The effect of an additional index case will make the $P_i(N + 1) = P_i(N) + (1 - P_i(N))P_i(N = 1)$, and $P_i(N = 1) = k$ (see Section (3.1.4)). So the formula is deduced from theory as following:

$$P_i = 1 - (1 - k)^N \quad (7.5)$$

Where N is the number of index cases, k is a constant with the condition that all other variables take the default values. The mathematical curve plotted based on Equation 7.5 (Figure 7.8) fits the values obtained by simulation very well (with R^2 equals to 0.9958).

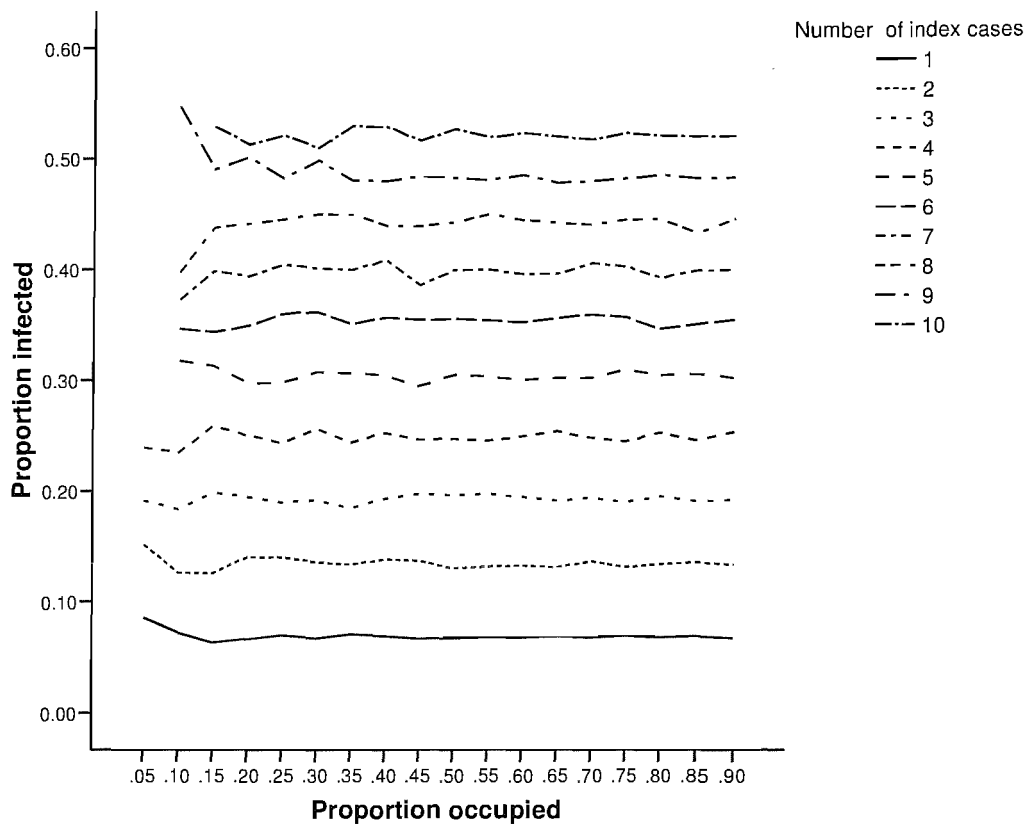


FIGURE 7.7: Proportion infected plotted against proportion occupied with different numbers of index cases for Type 4 AB.

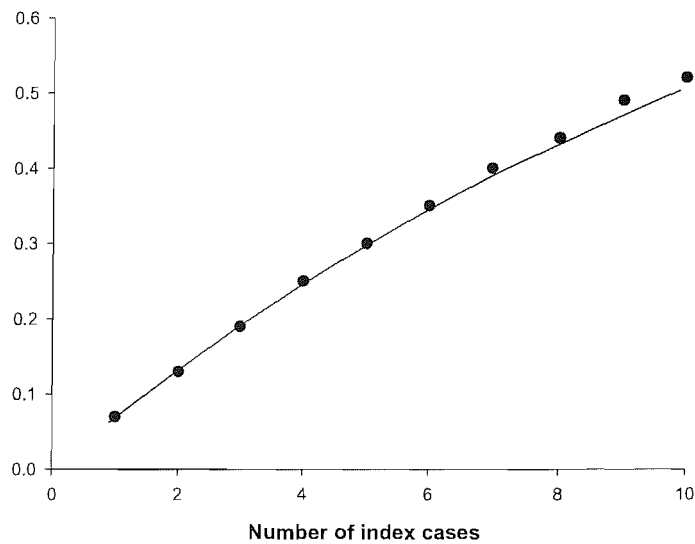


FIGURE 7.8: Proportion infected plotted against number of index cases for Type 4 AB (solid dots are the simulated results while the solid line is based on Equation 7.5).

7.4 Size of AB

This section explores the relation between the size of AB and proportion infected for all types of spatial distribution. The R_{AB} was set to be 1, and value of Width (W) and Length (L) varied from 3 to 15 m with a step value of 1 m (Figure 7.9). All other variables were fixed to the default values in Table 7.1.

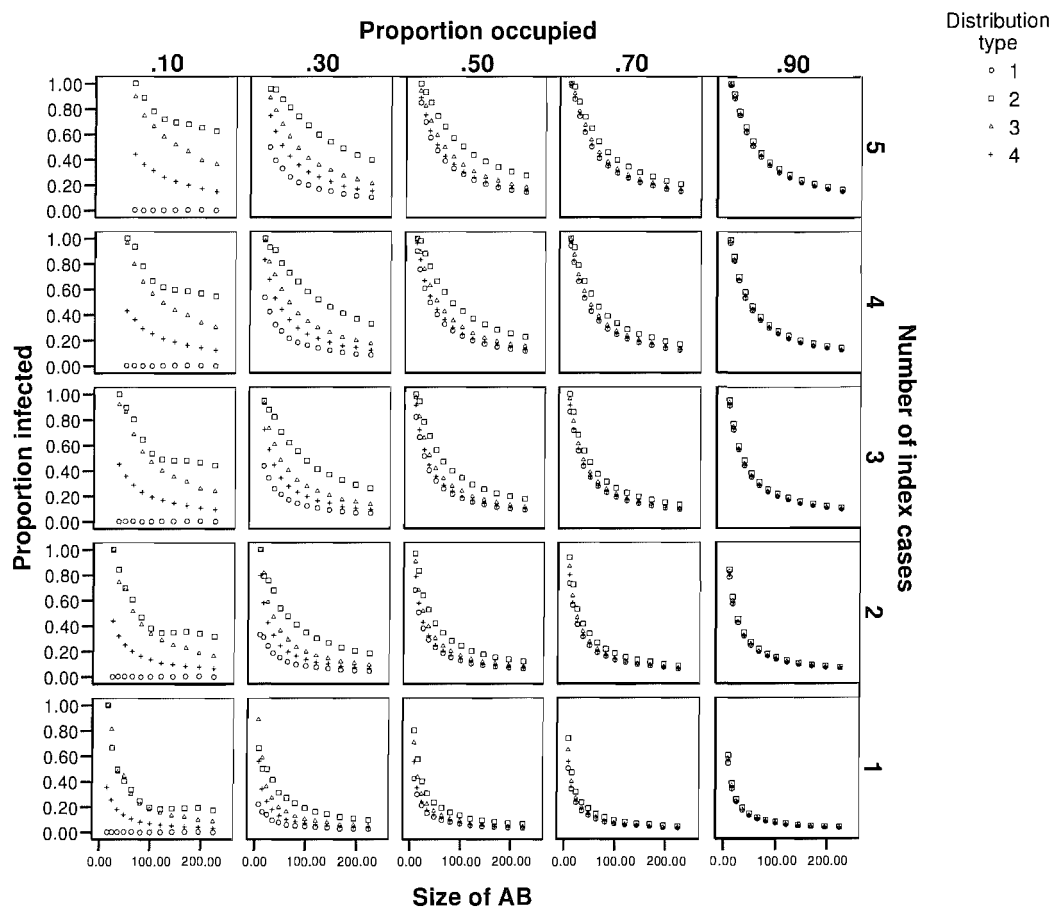


FIGURE 7.9: Proportion infected plotted against size of AB with different numbers of index cases and proportion occupied for Type 1, 2, 3 and 4 ABs.

Three properties are apparent from Figure 7.9: (1) for almost all types of AB, P_i decreased with an increase in the size of AB with 25 combinations of number of index cases and P_o . One exception is when $P_o = 0.1$ for the Type 1 ABs: individuals have enough space to keep away from each other, and $P_i = 0$; (2) in almost all circumstances,

$P_i(\text{Type2}) \geq P_i(\text{Type3}) \geq P_i(\text{Type4}) \geq P_i(\text{Type1})$; (3) with an increase in P_o , the differences between the different types decreases. Especially when $P_o = 0.9$, all types have a similar P_i . It is clear that for the Type 4 AB, the plots of P_i against the size of AB are almost exactly the same for different P_o with the same number of index cases. This demonstrates the independence of P_i on P_o for Type 4 AB again.

Equation 7.4 from the above section can be applied here for analysis of the Type 4 AB.

When $W = L$, since $S_{AB} = WL$, Equation 7.4 is transformed to be:

$$P_i = \frac{4(2S_{AB} - 3\sqrt{S_{AB}} + 1)}{(S_{AB})^2} \quad (7.6)$$

A theoretical curve was computed and plotted in Figure 7.10 to compare with simulated data. It is clear that it fits very well, demonstrating a strong inverse relationship between P_i and the size of AB for Type 4 ABs.

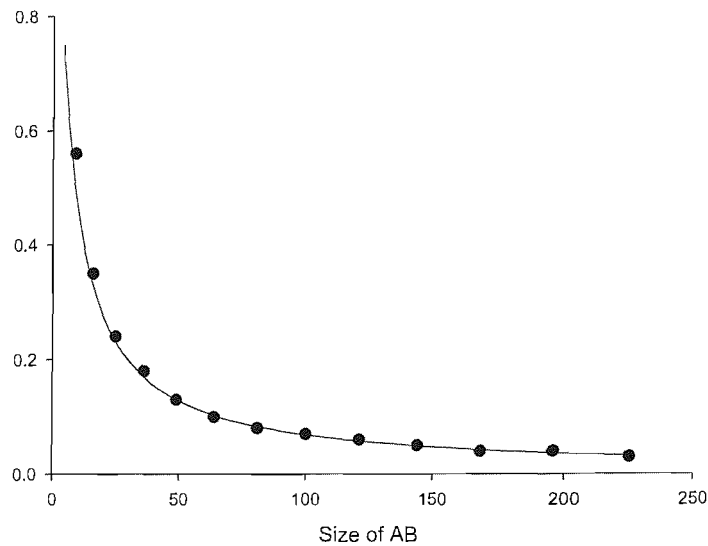


FIGURE 7.10: Proportion infected plotted against size of AB for Type 4 AB (solid dots are the simulated results while the solid line is based on Equation 7.6).

For $P_o = 0.5$ and number of index cases = 1, the regression analysis for Type 1, 2 and 3 are as follows (Figure 7.11 and Table 7.3). The R^2 value for the power curve model

is larger than for the inverse model. However, the power values for Type 1, 2 and 3 AB are -0.846, -0.817 and -0.905, which are close to -1 indicating that the inverse model fits the relation between P_i and size of AB for all types of AB well.

TABLE 7.3: Regression analysis results between proportion infected and ratio of AB for Type 1, 2 and 3 ABs with P_o equal to 0.5 and number of index cases equal to 1.

Type	First estimation model	Parameters	R^2	Second estimation model	Parameters	R^2
1	Power	$P_i = 3.033S_{AB}^{-0.846}$	0.997	Inverse	$P_i = 0.028 + \frac{3.831}{S_{AB}}$	0.977
2	Power	$P_i = 5.398S_{AB}^{-0.817}$	0.997	Inverse	$P_i = 0.063 + \frac{7.197}{S_{AB}}$	0.974
3	Power	$P_i = 5.425S_{AB}^{-0.905}$	0.999	Inverse	$P_i = 0.025 + \frac{6.311}{S_{AB}}$	0.995

7.5 Ratio of AB

This section explores the relation between the R_{AB} and proportion infected for all types of spatial distribution. The size of AB was set to be 100. Since for Type 1, 3 and 4 (see Figure 7.12), each AB's two dimensions are interchangeable, the value of W was varied from 1 to 10 m with a step value of 1 m. Thus, Widths of AB are no larger than Lengths of AB, and the ratio of AB was constrained within the range 0 and 1. For Type 2 ABs, since the AB's dimensions cannot be interchanged, the value of W was varied from 1 to 10 m with a step value of 1, 11, 13, 14, 17, 20, 25, 33, 50 and 100 m. Thus, when $W < L$, R_{AB} is between 0 and 1; when $W > L$, R_{AB} is between 1 and 100. To display the relation, a transformation was made: when $W > L$, R_{AB} was replaced by $R_{AB}^* = (2 - \frac{1}{R_{AB}})$. The range of R_{AB}^* is from 1 to 2 and also R_{AB} and $\frac{1}{R_{AB}}$ are symmetrical to the point of $R_{AB} = 1$. All other variables were fixed to the default values in Table 7.1.

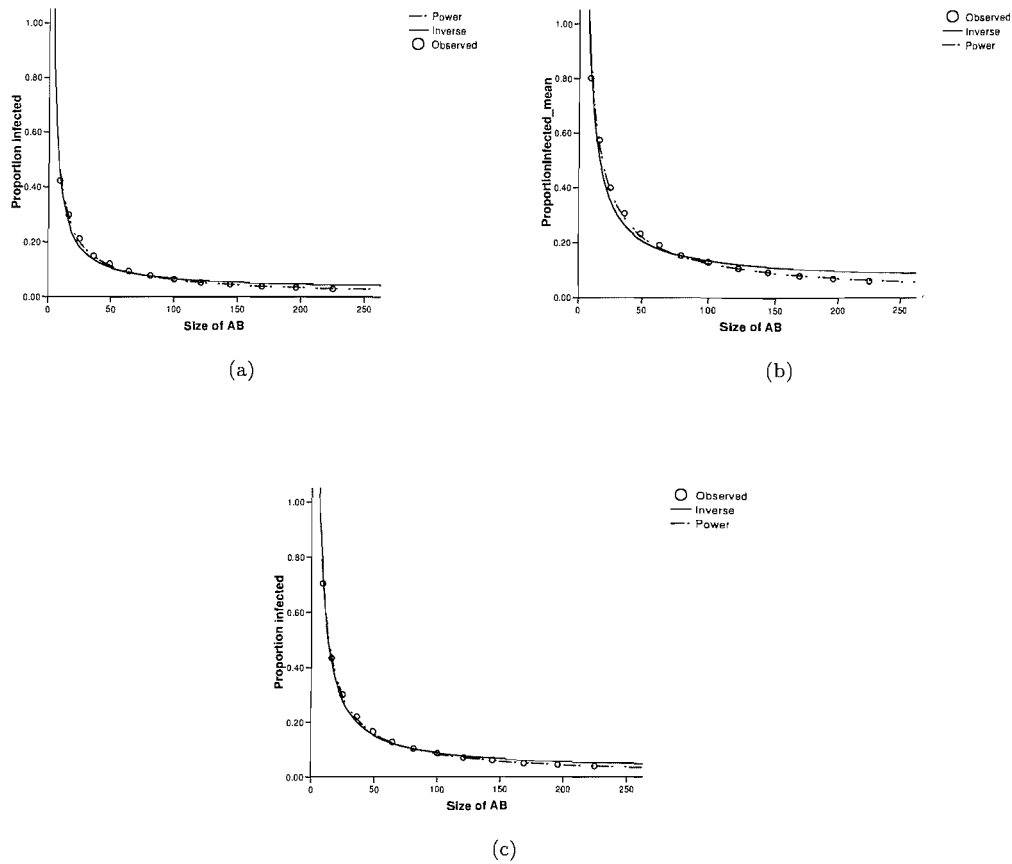


FIGURE 7.11: Proportion infected plotted against size of AB for (a) Type 1, (b) Type 2 and (c) Type 3 ABs (solid dots are the simulated results while the solid lines are based on regression formulae).

From Figure 7.12, the first property of interest is that the relation between the proportion infected and ratio of AB is positive generally. The reason is that for rectangular ABs with a fixed size, the closer the ratio of AB to 1, the smaller the proportion of the border and corner area in the whole AB. As discussed before, index cases at the border or corner positions of an AB can infect fewer individuals than at other positions. Take the Type 4 AB as an example. Equation 7.4 can be transformed to:

$$P_i = \frac{8S_{AB} - 6W - 6L + 4}{(S_{AB})^2} \quad (7.7)$$

It is clear that when S_{AB} is fixed, the value of $W + L$ reaches the minimum value

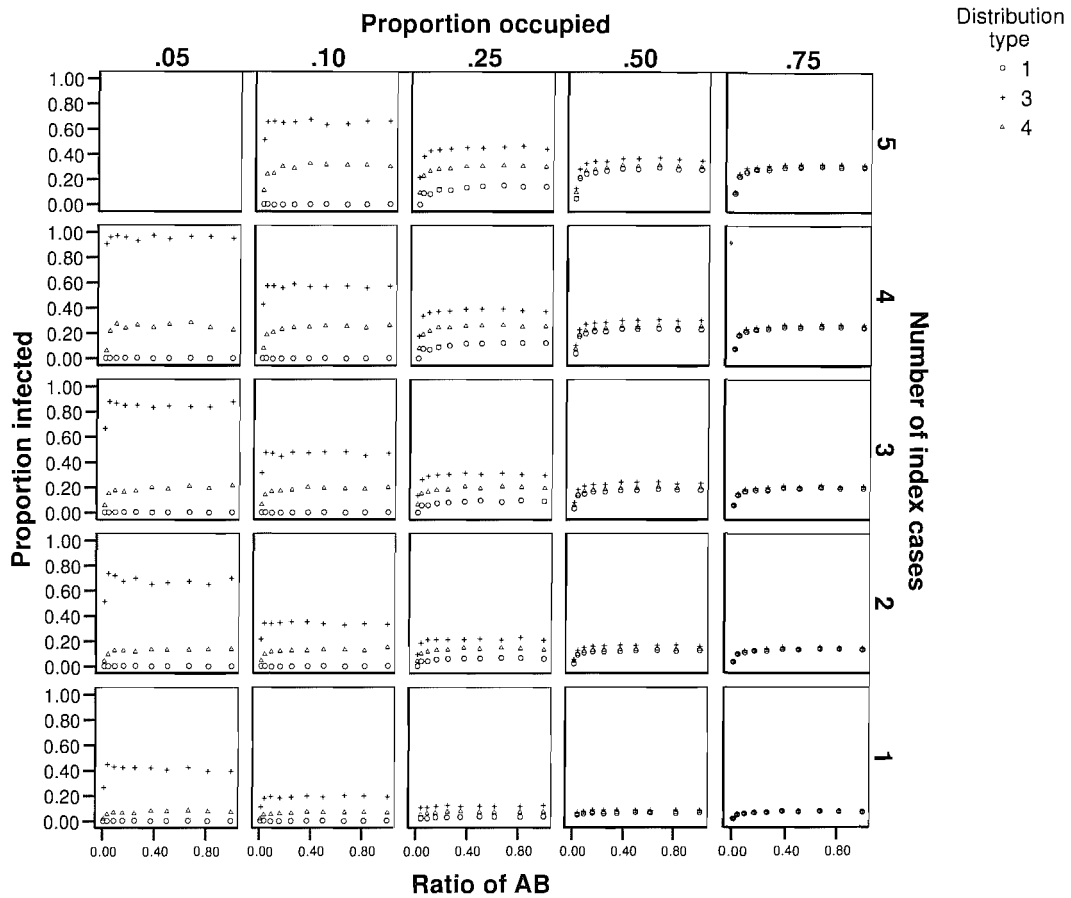


FIGURE 7.12: Proportion infected plotted against ratio of AB for Type 1, 3 and 4 ABs with different numbers of index cases and proportion occupied.

when $W = L$. Secondly, almost in all circumstances, $P_i(\text{Type3}) \geq P_i(\text{Type4}) \geq P_i(\text{Type1})$. Thirdly, with an increase in P_o , the differences between the different types of AB decrease. Especially when $P_o = 0.75$, all AB types take a similar P_i value.

For Type 2 ABs, with an increase in P_o , the curve becomes more symmetrical to the point of $S_{AB} = 1$ (Figure 7.13). The reason is that when the AB is full of individuals, the two dimensions can be interchanged. Secondly, when P_o is small (5% or 10% which means 5 or 10 individuals in total in this simulation) and the width is large enough, all individuals arrange in a line and there is no difference, even when W is larger (the ratio of AB is larger). This accounts for the flat tail which curves display when P_o is smaller than 0.5. The size of the tail decreases within an increase in P_o .

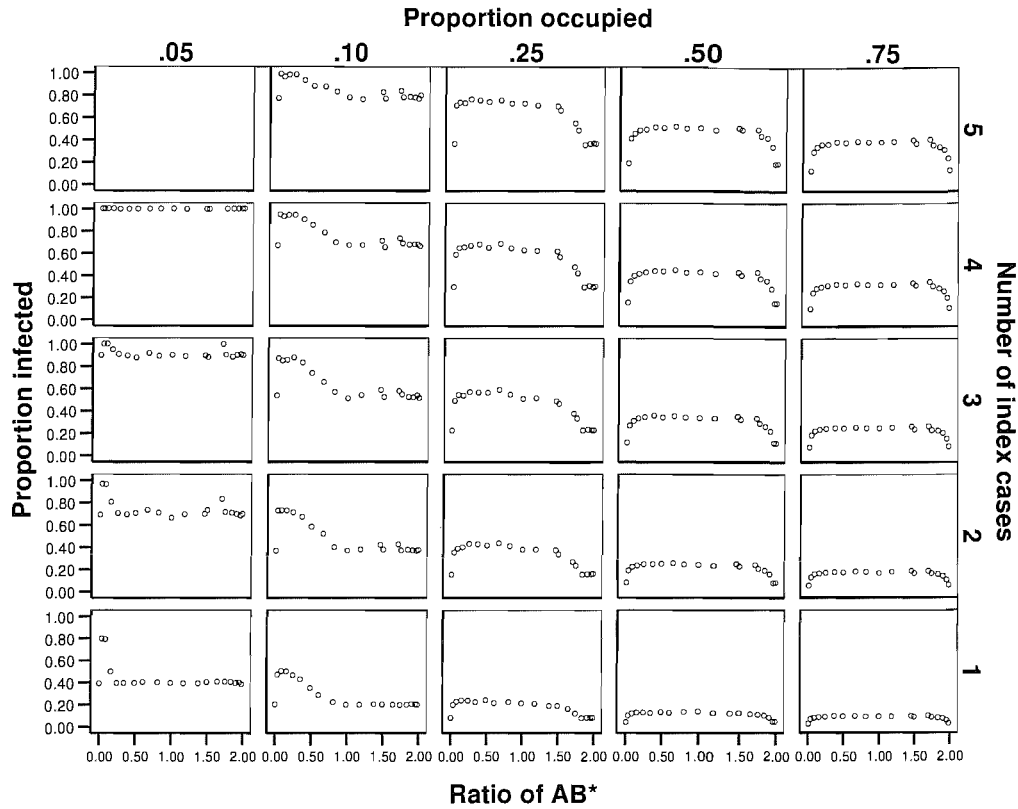


FIGURE 7.13: Proportion infected plotted against ratio of AB for Type 2 ABs with different numbers of index cases and proportion occupied.

7.6 Size of cell and ratio of cell

This section explores the relation between P_i and the size and ratio of cell for all types of spatial distribution. In the current raster-based AB simulation, possible values for minX or minY are 0.5, 1.0 and 1.5 m. Further, minX and minY are interchangeable. Thus, all possible shapes of cell (considering both the ratio and size of the cell) are $0.25 m^2$ (0.5×0.5), $0.5 m^2$ (0.5×1), $0.75 m^2$ (0.5×1.5), $1 m^2$ (1×1), $1.5 m^2$ (1×1.5) and $2.25 m^2$ (1.5×1.5).

From Figure 7.14, the most obvious characteristic to note is that all AB types for different proportions occupied show clearly two ‘peaks’ (i.e., smaller values when the size of cell is 0.25, 0.75 or $1.5 m^2$ and larger values when the size of cell is 0.5, 1 or $2.25 m^2$). This can be accounted for by the relation between the circular shape of the infection area with

the rectangular shape of the cells. The value of the circular infection area is determined only by the infection distance which is assumed to be 2 m for influenza. When applying this circular area to ABs with different shapes of cells, the sum area of all cells which overlays with the circular area is different since it is assumed that if only the center point of the cell is within the infection area then the whole cell is included (Table 7.4). Also almost in all circumstances, $P_i(\text{Type2}) \geq P_i(\text{Type3}) \geq P_i(\text{Type4}) \geq P_i(\text{Type1})$. With an increase in P_0 , the differences between the different types decreases.

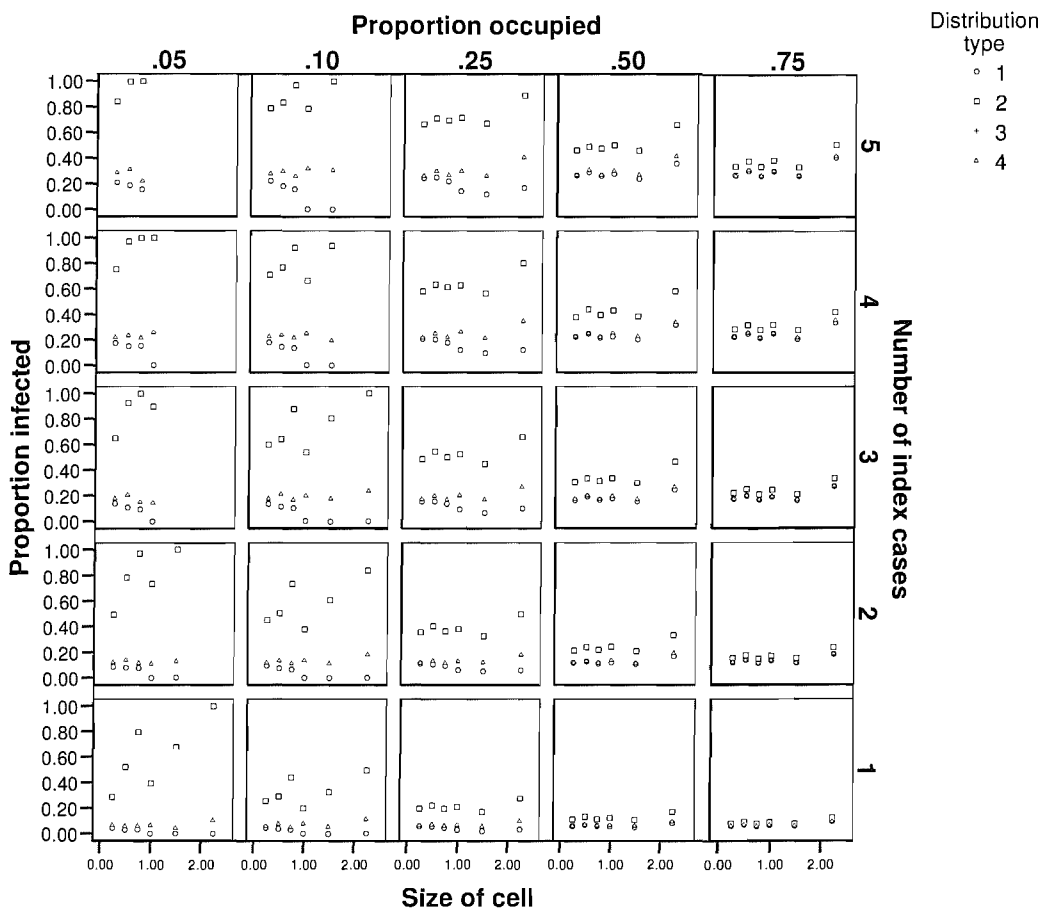


FIGURE 7.14: Proportion infected plotted against size of cell for Type 1, 2, 3 and 4 ABs with different numbers of index cases and proportion occupied.

TABLE 7.4: Different sum of covered areas for different shapes of cell.

<i>Size of cell</i>	0.25	0.5	0.75	1	1.5	2.25
<i>Number of cells covered by the circular infection area (not including the cell occupied by the infectious individual).</i>	28	17	9	9	4	4
<i>Sum covered area</i>	7	8.5	6.75	9	6	9

7.7 Movement proportion and movement frequency

This section analyses the parameters of dynamic ABs. The spatial distribution type is assumed to be Type 4, that is, randomly distributed for a dynamic AB both for the first allocation and subsequent movements. W , L , MinX and MinY take the default values from Table 7.1. Simulations were repeated based on the combination of the different values of four variables within Table 7.5. The simulation results showed again that P_i has no relation with P_o . Figure 7.15 shows the relations of P_i with the other three variables.

TABLE 7.5: Four variables and ranges of values for dynamic AB simulation.

<i>Variables</i>	<i>Range of value</i>
P_o	0.1, 0.2, ..., 0.9
N	1, 2, ..., 10
M_f	1, 2, ..., 9
M_p	0.1, 0.2, ..., 0.9

Initial regression analysis between P_i and these three variables showed similar exponential relations. It is straightforward to suspect that for dynamic ABs, the number of index cases affects P_i in the same way as it does for a static AB, as summarised in Equation 7.5. M_f and M_p are suspected to affect P_i jointly by their product form, that is, $H(M_f)F(M_p)$. Here, the whole relation was suspected to be as following:

$$P_i = 1 - (1 - G(n))^{1+H(M_f)F(M_p)}, \quad \text{where } G(n) = 1 - (1 - k)^n \quad (7.8)$$

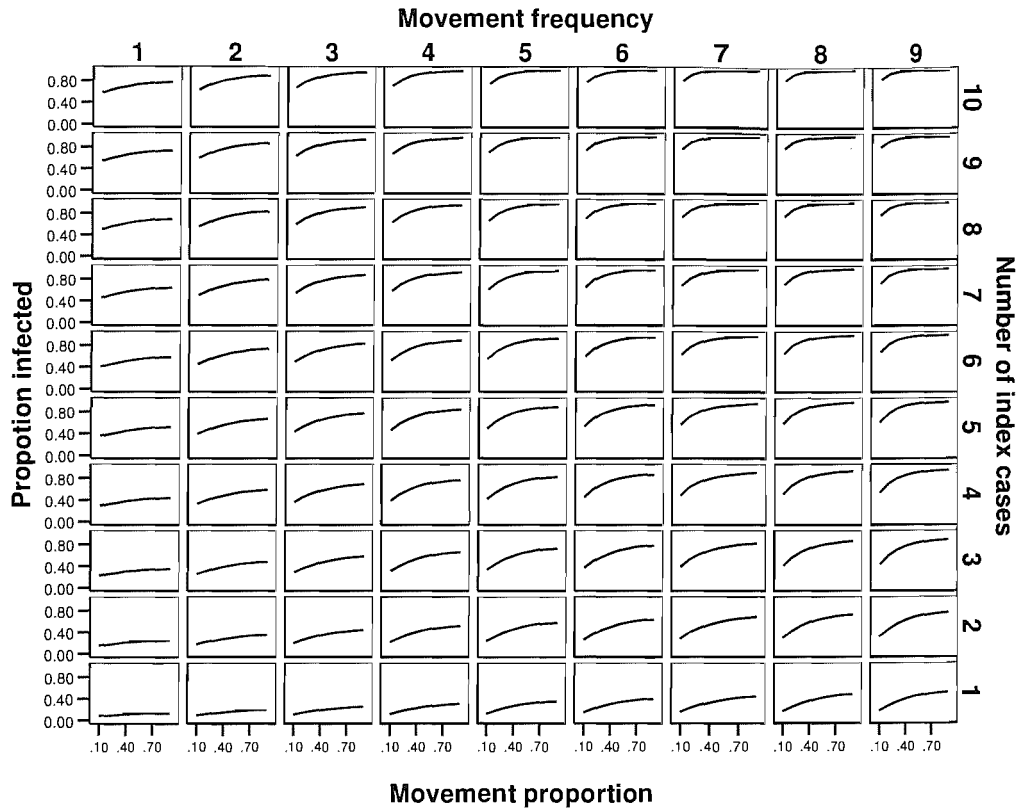


FIGURE 7.15: Proportion infected plotted against movement proportion with different numbers of index cases and movement frequency.

Firstly, it was assumed that $H(M_f) = M_f$, then $F(M_p)$ was generated from regression analysis. Simulated data with $M_f = 1$ and $N = 1$ were used for regression. The analysis generated Equation 9 with $R^2 = 0.996$:

$$F(M_p) = 0.007 + 2.075M_p - 1.061(M_p)^2 \quad (7.9)$$

Figure 7.16 shows the regression curve and simulated data.

Secondly, simulated data with $N = 1$ and $M_p = 0.5$ were used to compare with the mathematical curves based on Equations 7.8 and 7.9 (Figure 7.17). The value of R^2 was computed to be 0.998.

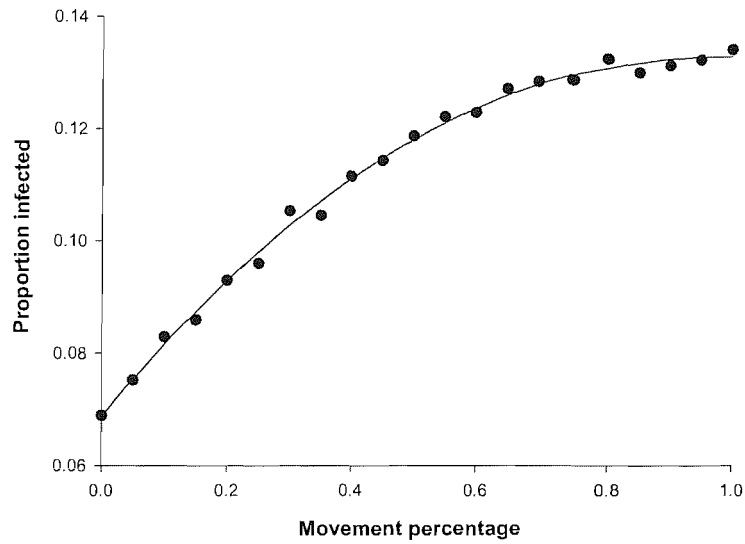


FIGURE 7.16: Proportion infected plotted against movement percentage for dynamic ABs (solid dots are the simulated results while the solid line is based on Equations 7.8 and 7.9).

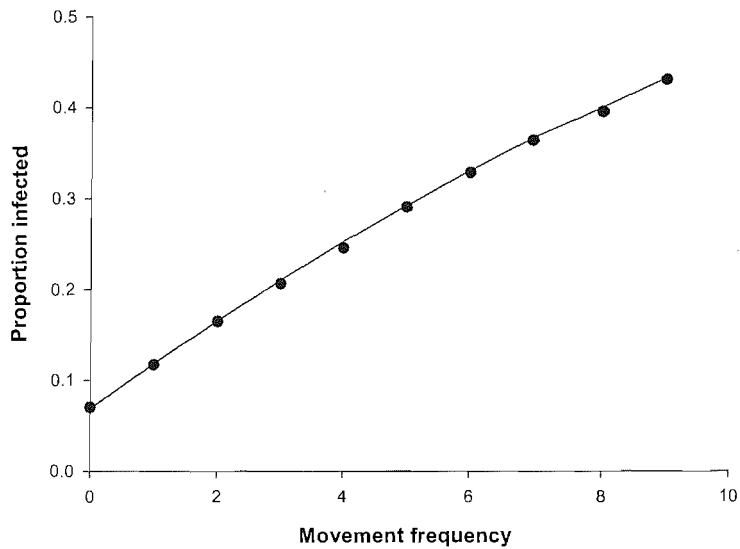


FIGURE 7.17: Proportion infected plotted against movement frequency for dynamic ABs (solid dots are the simulated results while the solid line is based on Equations 7.8 and 7.9).

Finally, for every case of the combination of different values of these four variables, the computed P_i was plotted against the simulated P_i (Figure 7.18). The result showed that Equation 7.8 and 7.9 fitted with the simulated result very well.

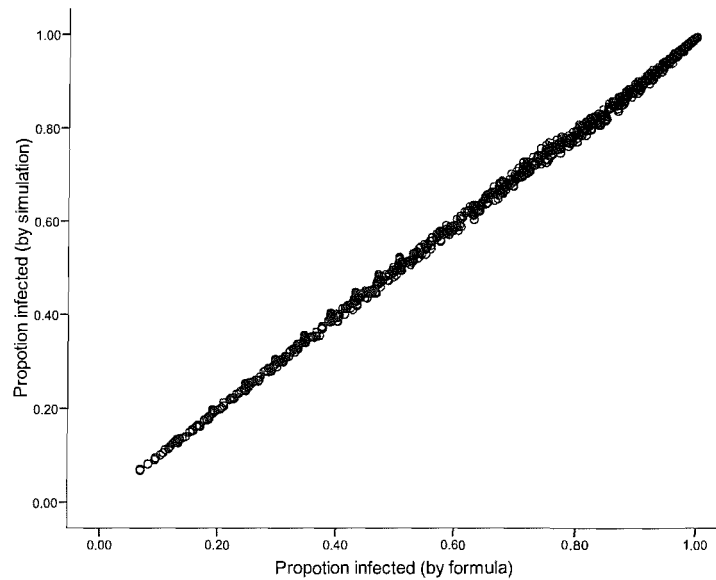


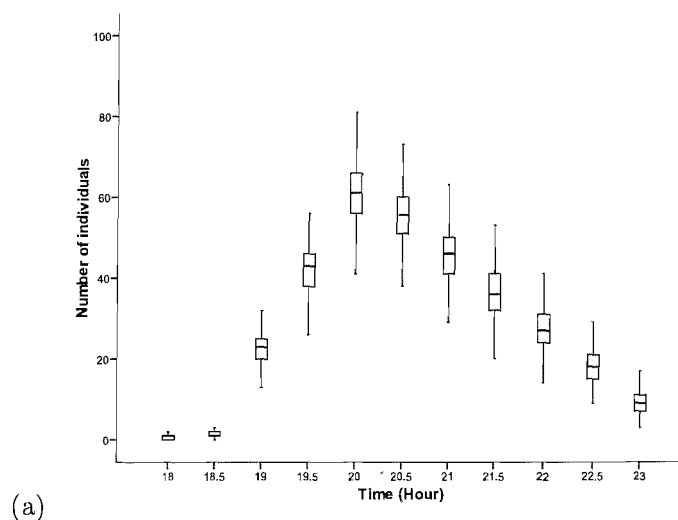
FIGURE 7.18: Simulated proportion infected plotted against the expected proportion infected based on Equations 7.8 and 7.9.

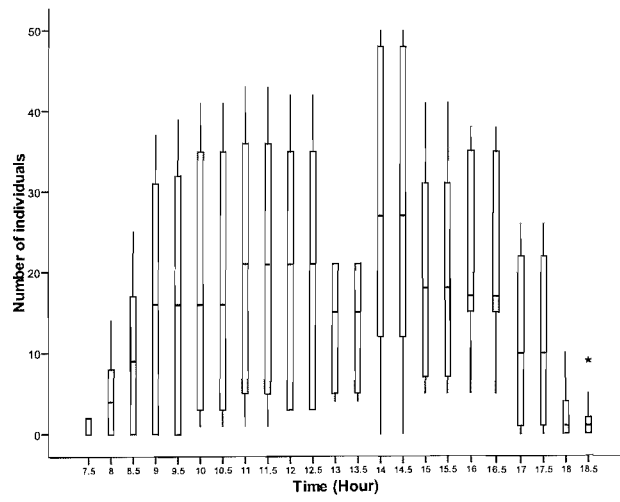
7.8 Rank importance of parameters

For static ABs, the correlation coefficients between the above six parameters and P_i were computed. The results show that the important sequences for any type of static ABs are very similar. That is: number of index cases is the most important one and the size of AB and the proportion occupied are the second and third (one exception is Type 1 AB, for which the size of AB is the third and the proportion occupied is the second). Size of cell, shape of cell and ratio of AB are less important. Compared with static AB, movement within dynamic AB can increase the proportion infected significantly. For dynamic ABs, the correlation coefficient between movement frequency and P_i is larger than between movement proportion and P_i .

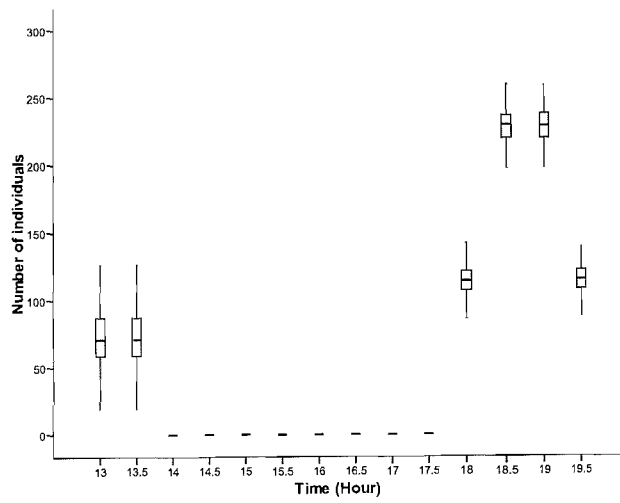
7.9 Risk assessment for ABs within the application to the campus of the University of Southampton

Raster space AB simulation was applied within ISTAM to simulate a hypothetical influenza outbreak amongst the first year undergraduate student body in the University of Southampton. Twelve types of ABs (see Table 6.1) were designed to describe the spatial environment of the part of the campus relevant to these students. Students' daily activities were simulated at between-AB and within-AB levels. At between-AB level, during the simulation, the change in the numbers of individuals within a certain AB during the whole day can be recorded. Figure 7.19 shows how the number of individuals changes within nine types of AB during a whole day (based on 1000 days simulation). The computer room and library room have very similar patterns because in ISTAM, students were assumed to have the same probability to go to the computer room or library room. It needs to be pointed out that some of simulated results are not perfectly realistic. Examples are that University library may not open at 6 am and also the change in the numbers of individuals at the gym seems not consistent with reality. This need more detailed survey on the daily activities of students.

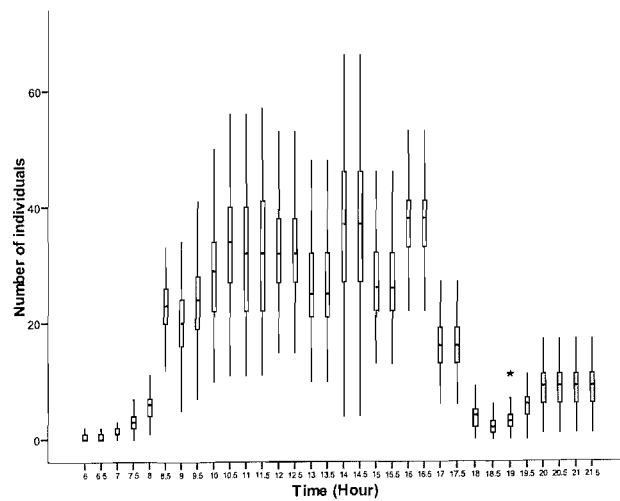




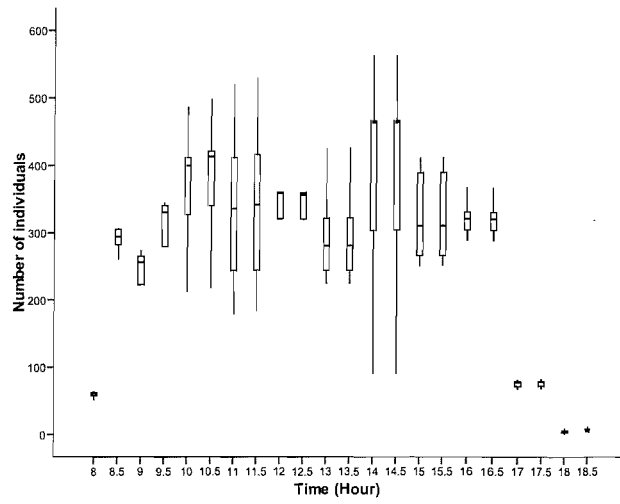
(b)



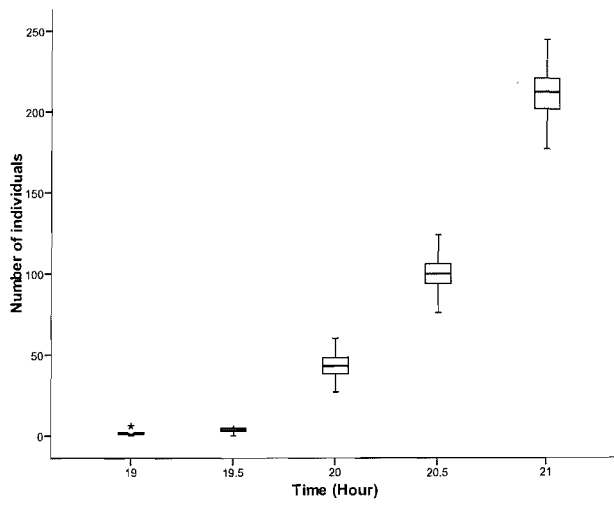
(c)



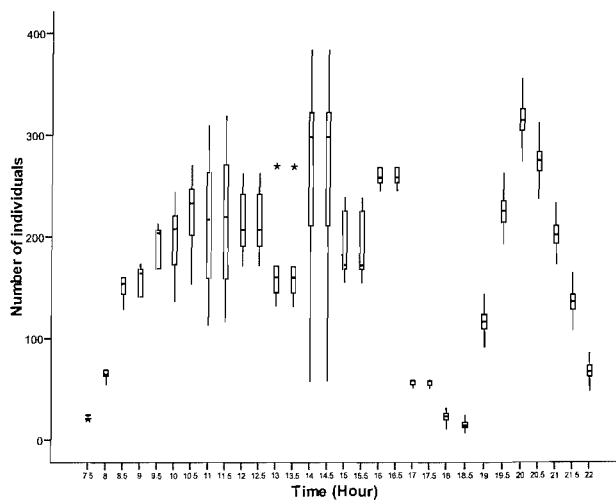
(d)



(e)



(f)



(g)

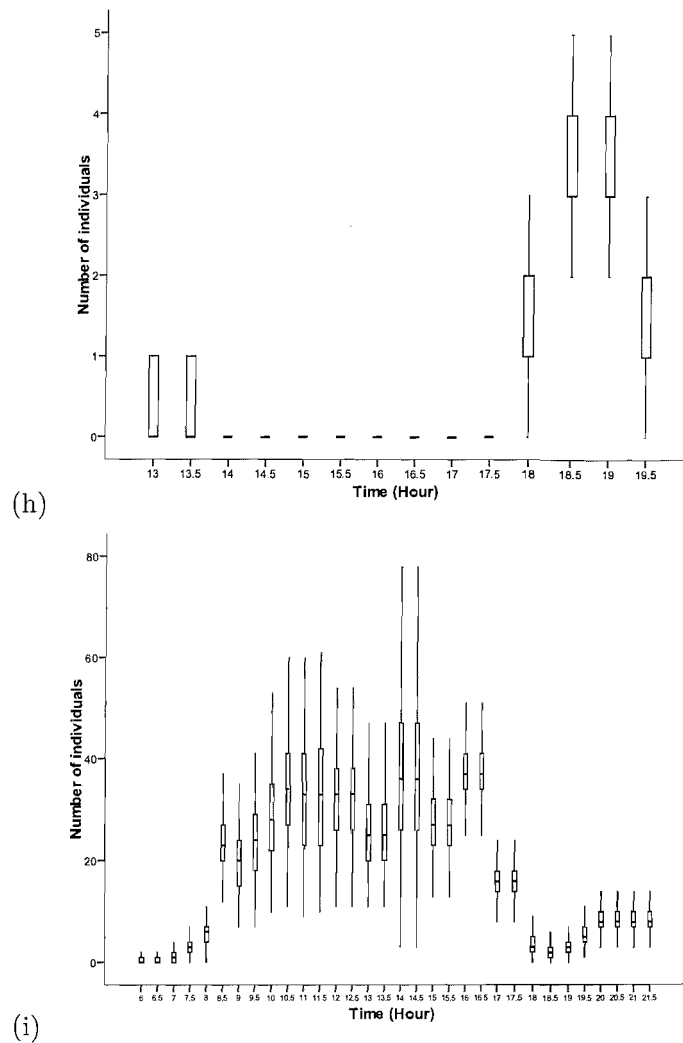
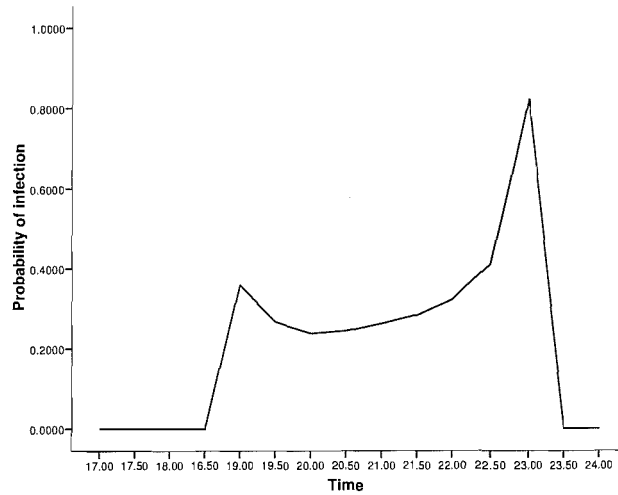


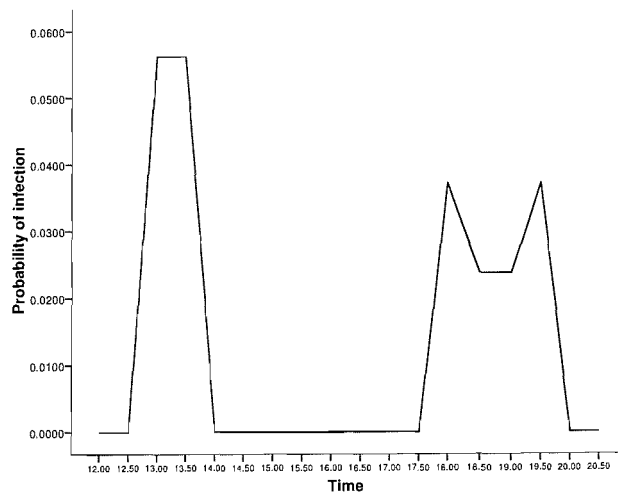
FIGURE 7.19: Simulated changing number of individuals within an AB during a whole day: (a) bar, (b) off campus, (c) refectory, (d) library room, (e) student union, (f) gym, (g) in campus, (h) kitchen and (i) computer room.

The change in the number of individuals within a certain AB can be converted to be proportion occupied since the size of AB and cell of AB are known. Figure 7.20 shows the averaged proportion infected (based on 1000 simulations) for seven ABs during the whole day (the proportion occupied is based on the average value from Figure 7.19, and it is assumed that one individual is infectious and all others are susceptible within the AB). In-campus and off-campus were omitted as the chance of infections at these two ABs is small. It is interesting to see that the computer room and library room have

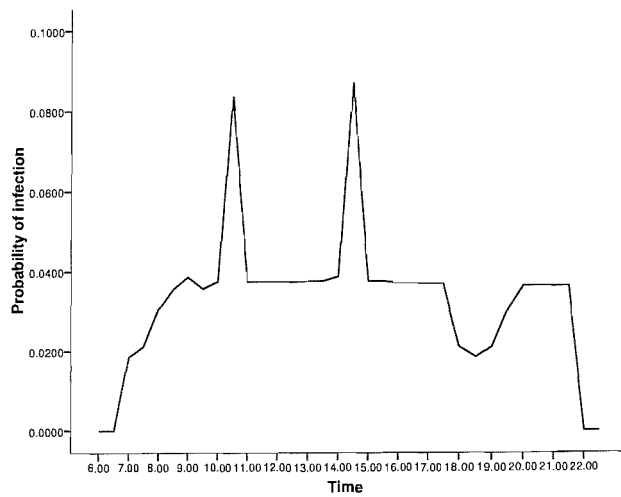
the same patterns of number of individuals and same physical size, but the proportions infected during the whole day are quite different.



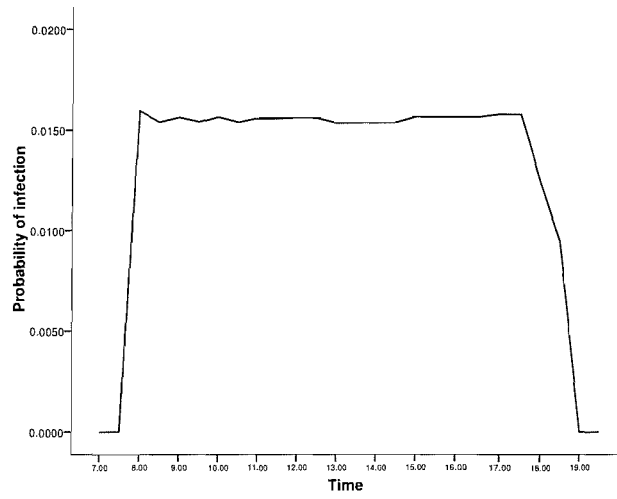
(a)



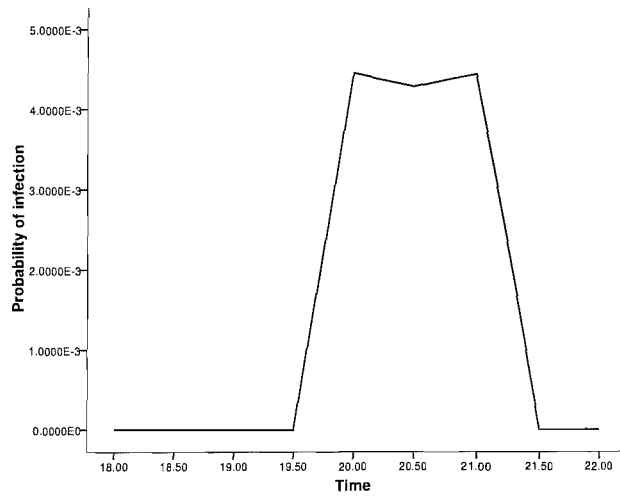
(b)



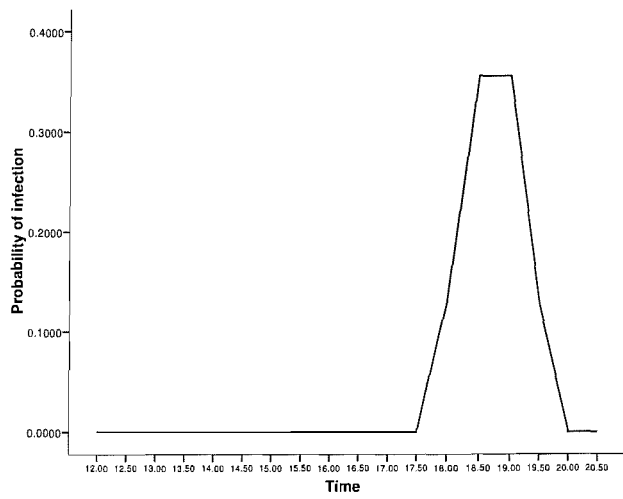
(c)



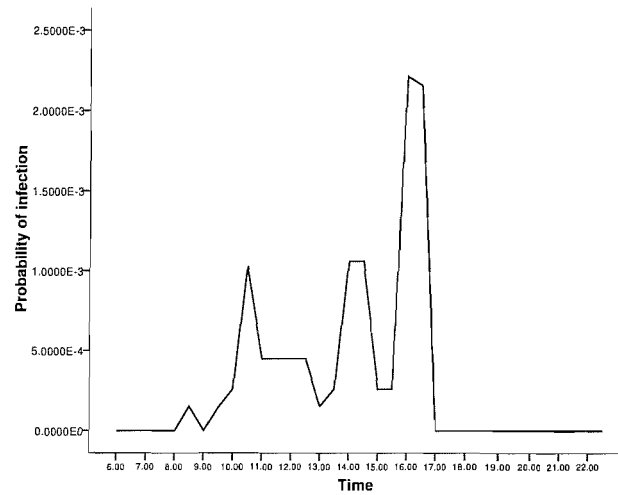
(d)



(e)



(f)



(g)

FIGURE 7.20: Simulated changing probability of infection for individuals within an AB during a whole day (assume one infectious individual present): (a) bar, (b) refectory, (c) library room, (d) student union, (e) gym, (f) kitchen and (g) computer room.

7.10 Summary

The exploration of the raster space AB simulation's parameters and related analysis were presented. For static ABs, four types of spatial distribution were assumed to exist amongst individuals. The analyses were focused on the relations between proportion occupied, number of index cases, size of AB, ratio of AB, size of cell, ratio of cell and proportion infected respectively. For dynamic ABs, the analyses were focused on the relations between proportion occupied, number of index cases, movement proportion, movement frequency and proportion infected respectively.

This chapter explored and discussed the parameters of raster space AB simulation within ISTAM. The parameter exploration showed that:

1. For different types of AB, P_o has a different effect on P_i . A negative inverse relation exists for Type 1 while a positive inverse relation exists for Type 3. A negative

relation exists for Type 2 while P_o has no relation with P_o for Type 4. With an increase in P_o , the difference between different AB types decreases as the number of index cases increases.

2. For all types of AB, P_i increases with an increase in the number of index cases while the magnitude of the increase decreases.
3. For all types of AB, inverse relations exist between P_i and size of AB.
4. The relation between the ratio of AB and P_i is positive generally for all types.
5. For static ABs, number of index cases is most important and size of AB and proportion occupied are the second and the third parameters in terms of the effect on the proportion infected.
6. For dynamic ABs, the relations between movement frequency, movement proportion and number of index cases with P_i are exponential and a regression model fitted the simulated data very well.

Although the above analysis results were based on some assumptions which are not always consistent with reality, this provides a starting point for future analysis. Some of the relations revealed by the above analysis do exist, while some others maybe depend on the simulation approaches (for example, simulation by raster space or vector space). Further understanding of the infection process at within-AB level and effective quantitative comparison of the infection probabilities for different types of AB may assist in the design and application of control measures before and during epidemics.

Chapter 8

Application to Eemnes

By 2020, 55% of the world's population is projected to live in urban areas (Leautier 2006). Fortunately, the epidemiological environment can be improved by urbanization. For example, urbanization processes can create improved nutrition, housing less vulnerable to vermin and cleaner drinking water (Daily & Ehrlich 1996). However, the high density of population living in urban areas increases the probability of intimate contact between people. As reviewed by Daily and Ehrlich (1996), the urban area's 'amplification' effect in terms of infectious disease transmission was noted by DeCock and McCormick (1988) and urban centers may be considered as 'graveyards of mankind' (Garrett 1994). The need to investigate the relationship between urbanisation patterns, individuals' space-time dynamics and infectious disease transmission has led to an interest in the development of simulation models of both disease transmission and space-time behaviour at the individual-level.

8.1 Overview of Eemnes

ISTAM has been applied to simulate a hypothetical influenza outbreak amongst a simulated first year undergraduate students within the campus of the University of Southampton. This case provides an example which was relative simple in terms of daily activity patterns, population structure and AB structure. It is a very special case comparing with a general human community (e.g., the whole population within a city). Table 8.1 summarizes the main difference between the above example and a general city. Considering

TABLE 8.1: Comparison between the application of ISTAM to University of Southampton campus and the city of Eemnes.

	<i>Campus of the University of Southampton</i>	<i>City of Eemnes</i>
<i>Study case</i>	First year undergraduate students in the University of Southampton between 2004 and 2005	The whole population within a city
<i>Individuals</i>	Assume only first year undergraduate students exist, only consider their genders	Need to consider gender, age, family rôle, income level, employment and other properties
<i>Activity patterns</i>	Relatively simple: assume all students take lectures so their activities are semi-fixed	Complex: different groups of individuals have different types of pattern
<i>Spatial area</i>	Assume all students live in the accommodation provided by the University; the campus is semi-closed; assume no interaction with the outside; relatively small area; the travel time can be neglected easily	Within the whole city; assume no interaction with the outside; the travel time spent within vehicles could be considerable
<i>ABs</i>	Relatively small number of ABs. The spatial distribution of all ABs and their physical condition is known or can be estimated	A large number of ABs. the composition and spatial distribution of all ABs are hard to obtain or may be available only at a higher level. Physical conditions of all ABs are especially hard to obtain
<i>Within-AB Simulation methods</i>	Assume all students play the same role; raster space AB simulation is suitable	Rôle-based AB simulation

these differences, to apply ISTAM at city level, four additional questions need to be answered:

1. How to build the whole population in terms of their demographic, social and economic properties?
2. How to build all ABs within a city in terms of their category, percentage and spatial distribution?
3. How to assign daily activity pattern to each individual?

4. How to simulate individuals' activities at with-AB levels without detailed information about the these ABs' physical condition?

This chapter explores the above four questions and attempts to provide answers through application of the ISTAM model to the city of Eemnes (see <http://www.eemnes.nl>), the Netherlands. The research aim was to apply ISTAM at the city level to model a hypothetical influenza epidemic outbreak amongst the population.

8.2 Background and data sources

The target city was Eemnes in the Netherlands. The municipality of Eemnes is located in the north of the province of Utrecht, it has less than 9000 inhabitants and covers about 3362 hectares. The Netherlands is one of the most densely populated countries in the world. A characteristic of ISTAM (and any epidemiologic model) is that it assumes a closed area, without individuals leaving or entering the area. Of course, this is not realistic, and future work will be devoted to solving this problem (e.g., by allowing visitors to enter and leave the area). In order to demonstrate the usefulness of ISTAM, Eemnes was selected since it is not located in the direct vicinity of larger cities, limiting the exchange with 'outside' areas.

The main data sources used in this application are listed in Table 8.2. The activity survey data were assumed to be representative for people residing within Eemnes although in fact, it is not surprising that the people living in large cities such as Amsterdam or Utrecht have different activity patterns compared to people living in a rural area. The land use data was PC6-based (in the Netherlands, a complete postcode contains 6 positions and relates to an average of 17 addresses or delivery points, hence 'PC6'). The Netherlands has about 420,000 postcodes), while synthesized household data were grid-based. Preparation was needed to aggregate the household data to the PC6 level.

TABLE 8.2: Data sources for the application of ISTAM to Eemnes.

<i>Data source (with short name)</i>	<i>Description</i>
Activity survey data (ACT)	One activity survey named AMADEUS was taken in 2000 (Ettema 2005). In this data collection, 1997 households (with 3499 people) (in the Amsterdam-Utrecht region) filled out a two-day travel/activity diary.
Synthesized household data (HH)	Synthesized household data (at grid-space level), above 1,500,000 cases, cover the area of Amsterdam-Utrecht corridor. Every household within this area is recorded with properties such as number of persons and workers, age of the head household member, household income and car availability (Ettema 2006).
Land use data (Land)	PC6 based, 433689 cases, the whole Netherlands. For every PC, the numbers of people engaged with each vocation (such as Offices, Education, Healthcare, Industry, Transport and etc) are recorded.
PC6 statistical data (PC6)	Statistical data about the age and household structure for every PC6 of the whole of the Netherlands.

The general procedures were as follows: (1) preparation, that is, to build the population and the city of Eemnes. For the population, the key point was to assign properties such as family structure, number of cars and income level to every household, and assign properties such as age and gender and activity patterns to every individual. For the city, the social and spatial structures needed to be built. These simulated data were saved into a database for the next step; (2) simulation, that is, to generate individuals' daily activities from their activity patterns, then to model individuals' movements between ABs and the interaction of individuals within ABs. During the interaction, possible infection occurred from contacts between individuals.

8.3 Preparation process of building the population and city

The preparation processes is shown in Figure 8.1. Five sub processes are represented below.

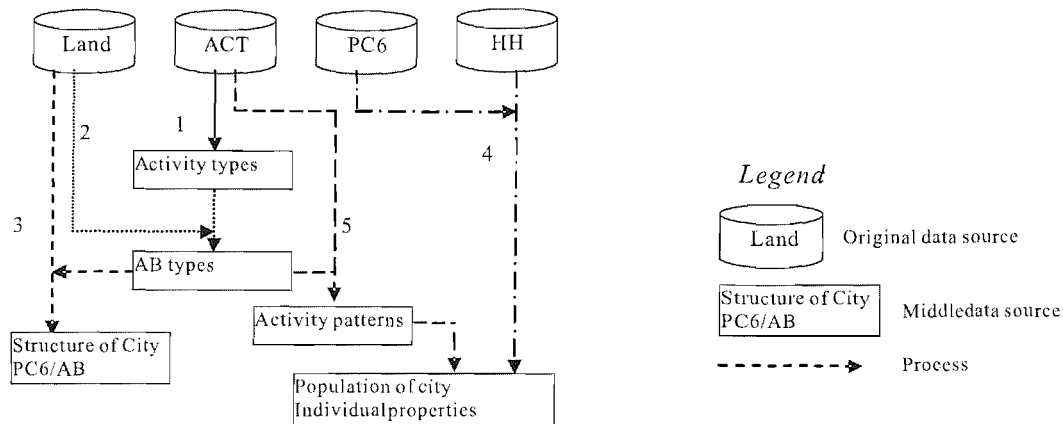


FIGURE 8.1: Preparation process of building the population and city of Eemnes.

8.3.1 Process 1: selection of activity types

The selection of the activity types depends strongly on the activity data (i.e., the ACT data source). Here, it was assumed that the survey activity data were sampled from the whole population (in fact, the sample population did not include children less than 10 years old). Some types of activity are neglected if they are not explicitly related to contact between individuals (and, thus, are not important for infectious disease transmission). Some other activity types need to be included even if their percentage contribution is small because these activities bring individuals into contact intimately (such as health care). It needs to be mentioned that trip activity was neglected in this study to simplify the research and also due to the limited availability of related data sources. All activity types which take more than 1% in the average time distribution are listed in Table 8.3.

TABLE 8.3: All activity types which take more than 1% in the average time distribution.

<i>Activity type</i>	<i>Percentage (%)</i>
Sleep	32.20
Work	14.40
Others at home	13.60
Trip	7.50
Housekeeping	5.40
Eating	4.40
Waiting	2.90
Personal care	2.70
Social activity/family, friends	2.20
Education (school, study...)	1.80
Child care	1.30
Tele-community	1.30
Receiving visitors	1.00

8.3.2 Process 2: selection of activity bundles

Specific objects in this application (e.g., different land use units inside one PC6) were classified to reflect the relations between the types of individuals' activities and the types of land use. It was assumed that within one PC6, no more than one school or industry could exist. Individuals visit certain ABs such as office, industry and farm for work. Other ABs such as post office and bank, some individual work there and others visit for service. All ABs types and corresponding activity types within ISTAM for this application are listed in Table 8.4.

TABLE 8.4: Activity bundle types within ISTAM for the application to Eemnes (The number in brackets is the expected number of staff working in this type of work place while * means one and only one AB of this type within the current PC6.

<i>AB type</i>	<i>Activity type</i>
Office (5)	Work
Industry (*)	Work
Farm (5)	Work
Shop (3)	Shopping/ work
Post office / Bank (3)	Go to post office / go to bank/ work
Healthcare place (3)	Health care/ work
Sport place (3)	Sport / work
Household	Sleep / activities within home / visit friend's home
Social place (3)	Dinner out or other social activities in caf�, bar, club <i>etc.</i>
Relax place	Cultural or recreational activities
Service place	All other personal service
School (*)	Study / work

8.3.3 Process 3: building of city

Land use data record the numbers of individuals engaged with every vocation. It was assumed that the expected size of a work place of a given vocation type within all PC6 of Eemnes is constant. Then from the number of individuals engaged with a given vocation within each PC6, the number of corresponding work places can be computed. For example, if the number of individuals engaged within the health care vocation within a given PC6 is ten, and the expected size of the health care work place is set to be five, then it is computed that there are two health care places within this PC6.

8.3.4 Process 4: population synthesis

As the synthesized household data are grid-based, which are not compatible with PC6 statistical data, all households within all grid cells of Eemnes were pooled together and then each household with its all family members were allocated randomly to all PC6 zones of Eemnes as follows.

Firstly, all individuals were classified according to age structure (Table 8.5). Different sub-classifications, and subsequently simulation of activities, were applied to the above four classifications. It is believed that daily patterns are related to individuals' socio-economic characteristics such as household rôle, lifestyle, and life cycle (Vaughn et al. 1997). In Janelle et al. (1998), the whole population was divided into 14 rôle groups based on dimensions of gender, marital status, job, child care, residence tenure and mobility. In

TABLE 8.5: Age structure for the population of Eemnes.

<i>Classification</i>	<i>Description</i>
Children under 5	For children under 5, it is assumed that they do not have independent activities and always stay inside households
Children between 5-10	Since the activity survey data do not cover children in this age range, their activity patterns are assumed to be simple: go to school at school hours and stay inside households at all other times
Children between 11-18	Their activity patterns are based on survey data, but delete all work parts. They will be further sub-categorized by four dimensions
Adult	Same with above, based on survey, but delete all study parts (their studies are taken as work), they will be further sub-categorized by four dimensions

this research, average working (study) hours per day (Table 8.6), rôles in household (Table 8.7), car availability index (1: no car; 2: yes and always; 3: yes and sometimes) and everyday in a week (Sunday to Saturday) are selected as four dimensions. Based on the above four dimensions, both children between 11-18 and adults in the population were further sub-classified, and their corresponding activity patterns were generated from the activity survey data. If the survey data cannot generate a corresponding activity pattern from a specific value of these four dimensions, a replacement activity pattern was assigned at a more aggregated level, that is, the number of dimensions was decreased. The sequence of the dimensions to be excluded was car availability index, rôle in household and working (study) hours. For one individual, her/his daily activity pattern was further classified according to the seven weekdays.

TABLE 8.6: Working index according to the average number of working (study) hours per week.

<i>Index</i>	<i>Number of working (study) hours per week</i>
0	0
1	1-15
2	16-30
3	31-45
4	>45

TABLE 8.7: Rôle types within household.

<i>Index</i>	<i>Description</i>
1	no family
2	single parent with children
3	child with single head
4	parent in couple with no children
5	parent in couple with children
6	child in couple with children

8.3.5 Process 5: assign activity patterns to individuals

Within this application, the time unit was set to be 15 minutes, such that one day is divided into 96 units. Thus, the objective of simulating a person's daily activity is actually to assign the 96 units to certain activity types.

For each value from the combination of the four dimensions mentioned in the above Section, aggregated data about the daily distribution of time spent on the main types of activity were computed from the activity survey data (if corresponding data exist), then these aggregated data were saved into the database for further use. This activity pattern included not only the average duration for every possible activity, but also the probability of the time of commencement for certain activities. Three types of activities were dealt with as exceptions: study time (applicable to children, assuming all children go to school during school hours), working time (applicable to all adult workers, assuming all workers go to work during working hours) and sleeping time. These three types of time were assigned to individuals as personal properties during the generation of the whole population. For the daily activity of one person of one week day, firstly, the above three special types of activities were fixed. Then from the beginning to the end of the time sequence, the vacant time units were assigned to activity types by the probability weights from the assigned activity pattern in relation to the socio-demographic segment.

For spatial location, for every person, her/his household and workplace (or school) were fixed from the beginning, and other locations (ABs) were selected randomly (or the spatially closest one was selected). At every time step, individuals can move between different locations.

For both children between 11 and 18 and adults, their activity patterns are assigned according to their properties at the four dimensions. After these processes, the whole population of this city was built and all types of AB are distributed to all PC6 zone. Figure 8.2 shows the spatial distribution of households and schools at PC6 levels for the whole Eemnes city. It is clear that most of households are within the boundary of the city center, so in the chapter, the spatial displays are zoomed within the boundary of city center to highlight the area where most of contacts between individuals could occur. Figure 8.3 shows the spatial distribution of ABs for work (includes office, industry, farm and school), relaxation (includes sports, social and relaxation places) and maintenance (includes shop, post office, bank and healthcare places) at PC6 levels.

8.4 Simulation process

8.4.1 Simulation of individuals' movements between ABs

As individuals' movement between ABs is not the focus of this work, a simple method was implemented. For every working individual, the locations (PC6) of her/his household and work place were fixed before the simulation and distributed randomly across the whole city. Due to the relatively small size of Eemnes, the locations of the only four schools with services for different stages of pupils were fixed before the simulation. Then, for others activities, two methods for selecting ABs were implemented. By the first method, during the simulation, an individual selects randomly from all ABs of specific types from all PC6s (including the current PC6) within a certain distance from the

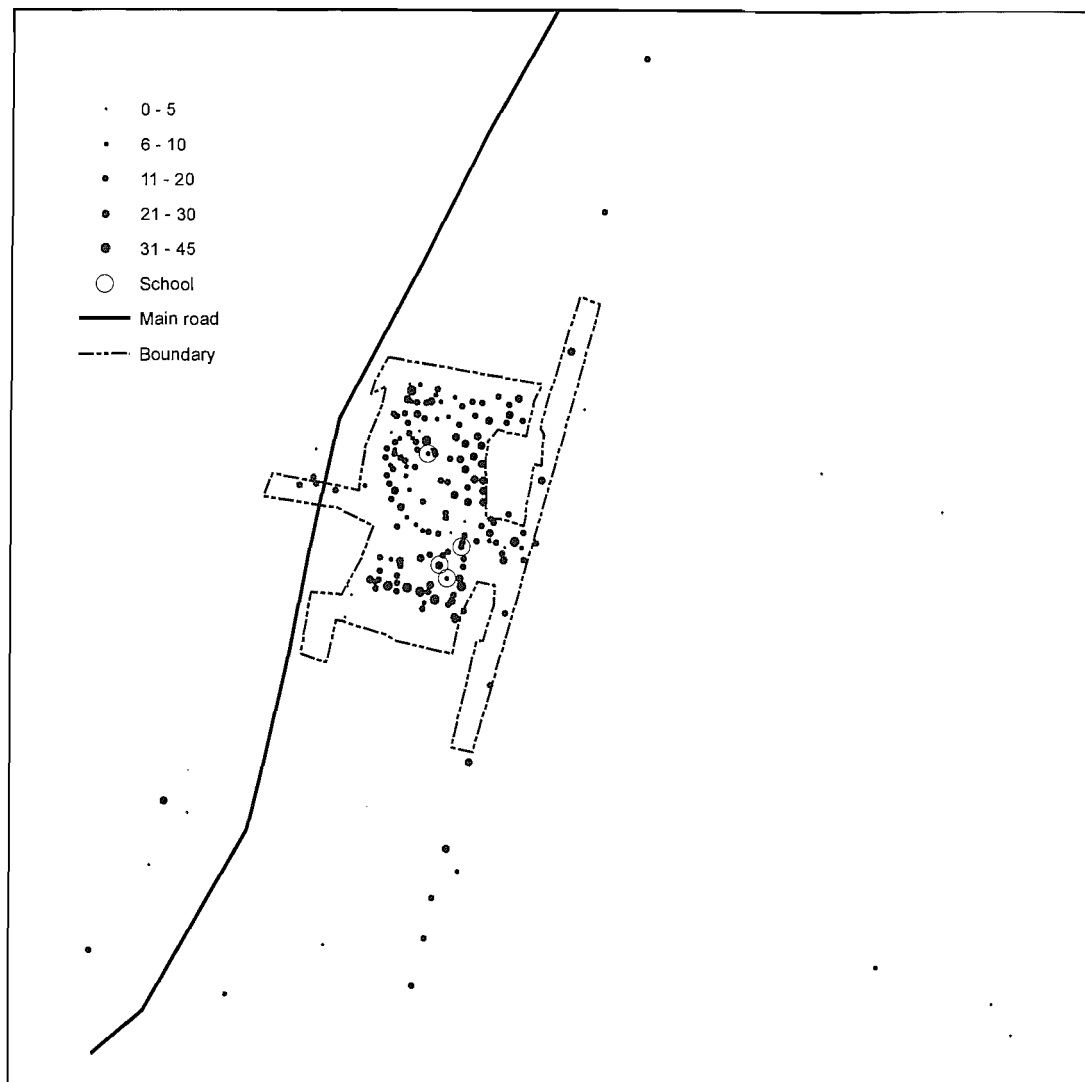


FIGURE 8.2: Spatial distribution of households and schools at PC6 levels (solid black circles indicate PC6 with school).

current PC6 if she/he needs to take another activity. By the second method, before the simulation, for every type of activity, every individual is assigned to a fixed AB which has the minimum distance from her/his household. As mentioned before, trip activities are neglected in this research which means that there are no infections between individuals during the travel time.

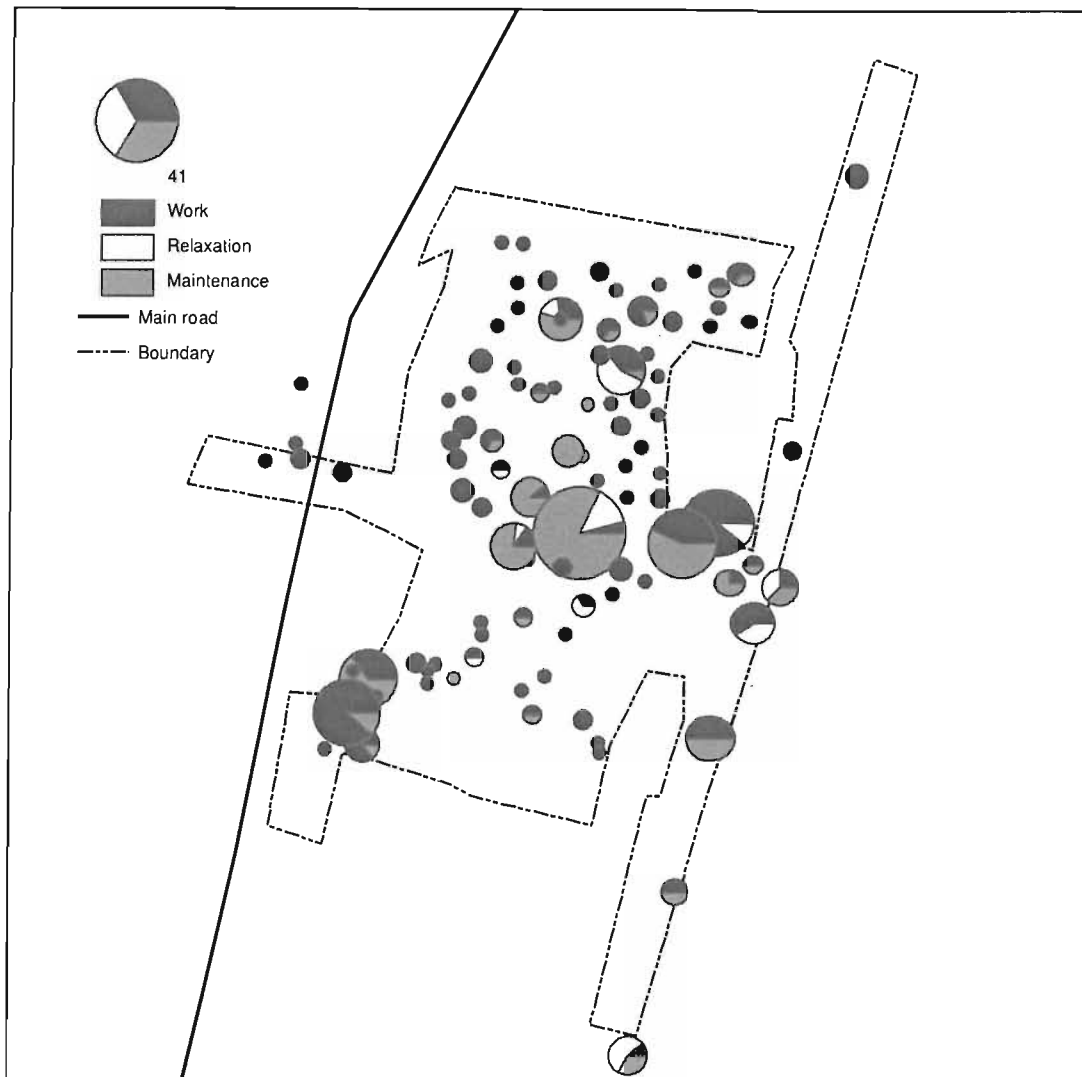


FIGURE 8.3: Spatial distribution of AB for work, relaxation and maintenance at PC6 levels.

8.4.2 Simulation of individuals' interactions within ABs

Eleven types of AB were defined for rôles-based within-AB simulation. For each type of AB, its description and values of parameters are shown in Table 8.8. The simulation results can be recorded into a database during the simulation or after the whole simulation. Currently, simulation results can be recorded as individual-based, AB-based, contact-based and PC6-based. Further analysis can be based on these simulation data.

TABLE 8.8: Contacts within AB.

<i>AB type</i>	<i>Description</i>	<i>Values of parameters (default: P_c)</i>
Office	Loop between all staff	P_c^2
Industry / Farm	The whole staff are divided into a few subgroups (with one as key subgroup, the size of the key subgroup is the same as the total number of other subgroups). Individuals within the same subgroup come into contact with each other as a Loop . Each individual from the key subgroup comes into contact with one key number of her / his corresponding subgroup	Loop (P_c^4) Between key subgroup and other subgroup (P_i^2, P_c^5)
Shop	Loop between workers Dynamic Random between all customers 20% of customers come into contact with one staff member randomly	Loop (P_c^3) Dynamic random (Density=0.5, Mobility=1, P_c^4) Randomly selected 20% of customers come into contact with one staff (P_c^3)
Post office / Bank	Loop between workers Queue between clients	Loop (P_c^4) Queue (P_c^1)
Health care place	One worker come into contacts with one client If clients > workers, then Static Random between the other clients; if clients < workers, then the other workers stay alone.	Worker with client (P_i^1, P_c^1) Static random (Density=1.0, P_c^2)
Sport place	Loop between workers Dynamic Random between customers	Loop (P_c^4) Dynamic random (Density=0.2, Mobility=1.0, P_c^2)
Household	Full between family members	P_c^2
Social place	Loop (1) between workers Clients are subgrouped, Loop (2) contact within each subgroup One client randomly selected from each subgroup to come into contact with one worker randomly selected.	Loop 1 (P_c^4) Loop 2 (P_c^2) Between clients and workers (P_i^3, P_c^4)
Relaxation place	Static Random between both workers and clients	Static Random (Density=1.0, P_c^5)
Service place	Every client come into contact with one worker randomly	Clients with workers (P_i^3, P_c^3)
School	Loop between supervisors (all teachers) Students are subgrouped. Static Even within every subgroup One teacher has a fixed subgroup and comes into contact with 5 students randomly selected from his / her subgroup	Loop (P_c^4) Static Even (Density=1.0, P_c^1); Teacher with five randomly selected students (P_i^3, P_c^1)

8.5 Simulation results

8.5.1 Calibration

8.5.1.1 Calibration of the contact frequency index

Since all contacts (whether infection is possible or not) can be recorded, the average contact number per person one day can be computed for comparison with related research. According to the related researches by Pool and Kochen (1978) and Edmunds et al. (1997), F was calibrated to be 0.03, and at this value, the simulated average contact number for one person per day is 14.7. Figure 8.4 shows the distribution of the number of contacts of all individuals for one day.

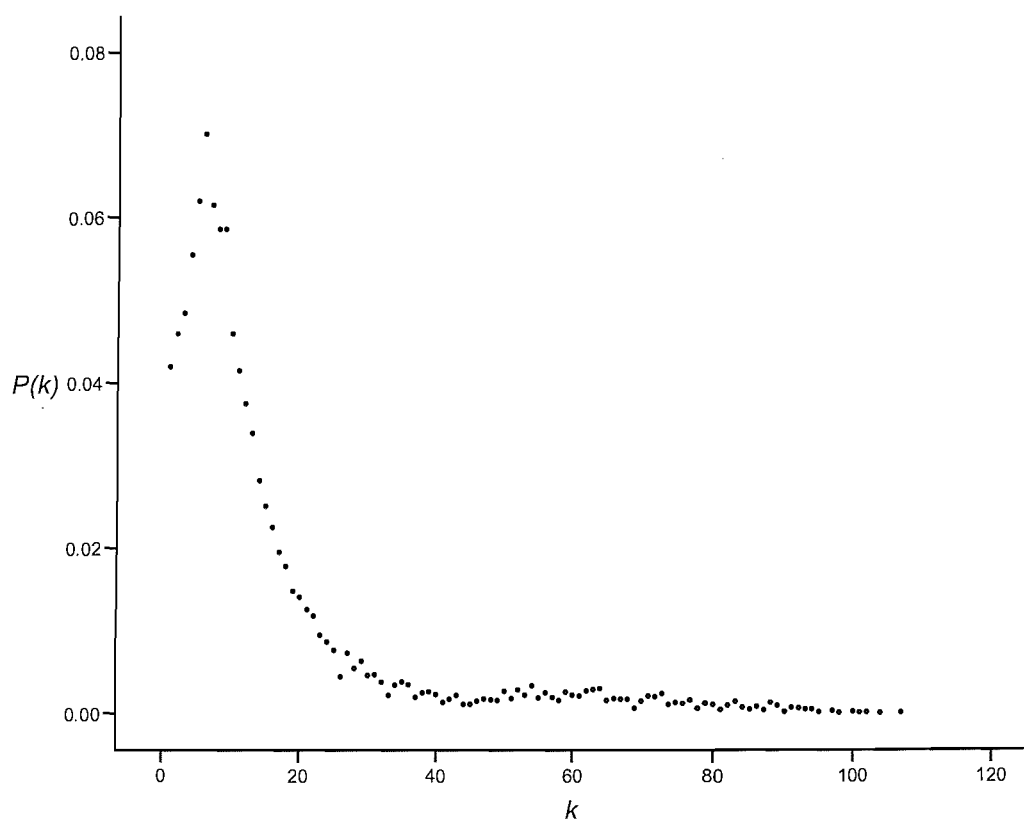


FIGURE 8.4: Distribution of number of contacts for all individuals per day.

8.5.1.2 Calibration of the contact intimacy index

All contacts during the simulation of ISTAM can be recorded. This provides the computation of R_0 according to its definition (i.e., trace the number of new infections from all index cases). After calibration, I was fixed to be 0.03, and at this value the average R_0 was 1.79.

At the I value of 0.03, the simulated SIR plot is shown in Figure 8.5. The simulated epidemic peaked at about 30 days and ended at between 70-80 days, at which time 40% population had been infected. The reason for this mismatch is in reality during the epidemic period, humans will change their behavior according to their own or other people's health state, such as staying at home or go to health care places when symptoms emerge to rest and avoid infecting other people.

8.5.2 Dynamics of population during the whole day

Individuals' daily activities are distributed according to certain patterns both spatially and temporally. Thus, their aggregate patterns (i.e., dynamics of the whole population within the whole city) also exhibit a certain structure during a whole day. This kind of research is especially interesting to the field of transportation, and also has implications for marketing, urban infrastructure and emergency response (Janelle et al. 1998). The word 'rush hour' is coined to express the time during this period when the number of individuals present peaks at locations such as transportation networks, shopping centers and other service facilities. The spatial distribution of all individuals at PC6-level and at a 15-minute interval can be generated from the simulation results. Figure 8.6 shows individuals' spatial distribution at 4 am, 10 am, 4 pm and 10 pm on a given day. It is clear that during 10 am and 4 pm, individuals are more concentrated within a few PC6 where a few special ABs such as schools and industries are located.

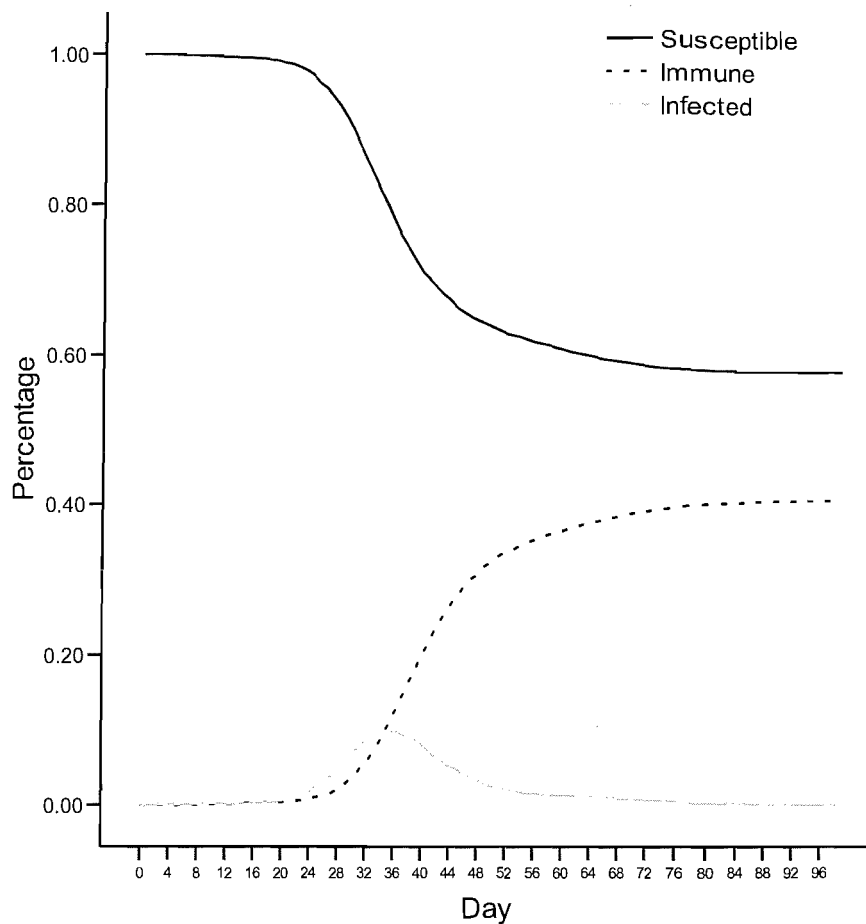


FIGURE 8.5: The changing number of susceptible, infected and immune individuals over time for one simulation of 100 days.

8.5.3 Infection PC6 distribution

Traditional epidemiological studies take the address of a patient's household or the address of a health care facility where the patient is registered as the spatial location where infection occurred. Also, the 'lag' effect which exists between the time of infection, the time of showing symptoms and the time of visiting health care facilities is often not represented. Unfortunately, under most circumstances, these data sources are the main resources for research work (sometimes even such data are not guaranteed to be available).

From the simulation results of ISTAM, for every infection, information such as the

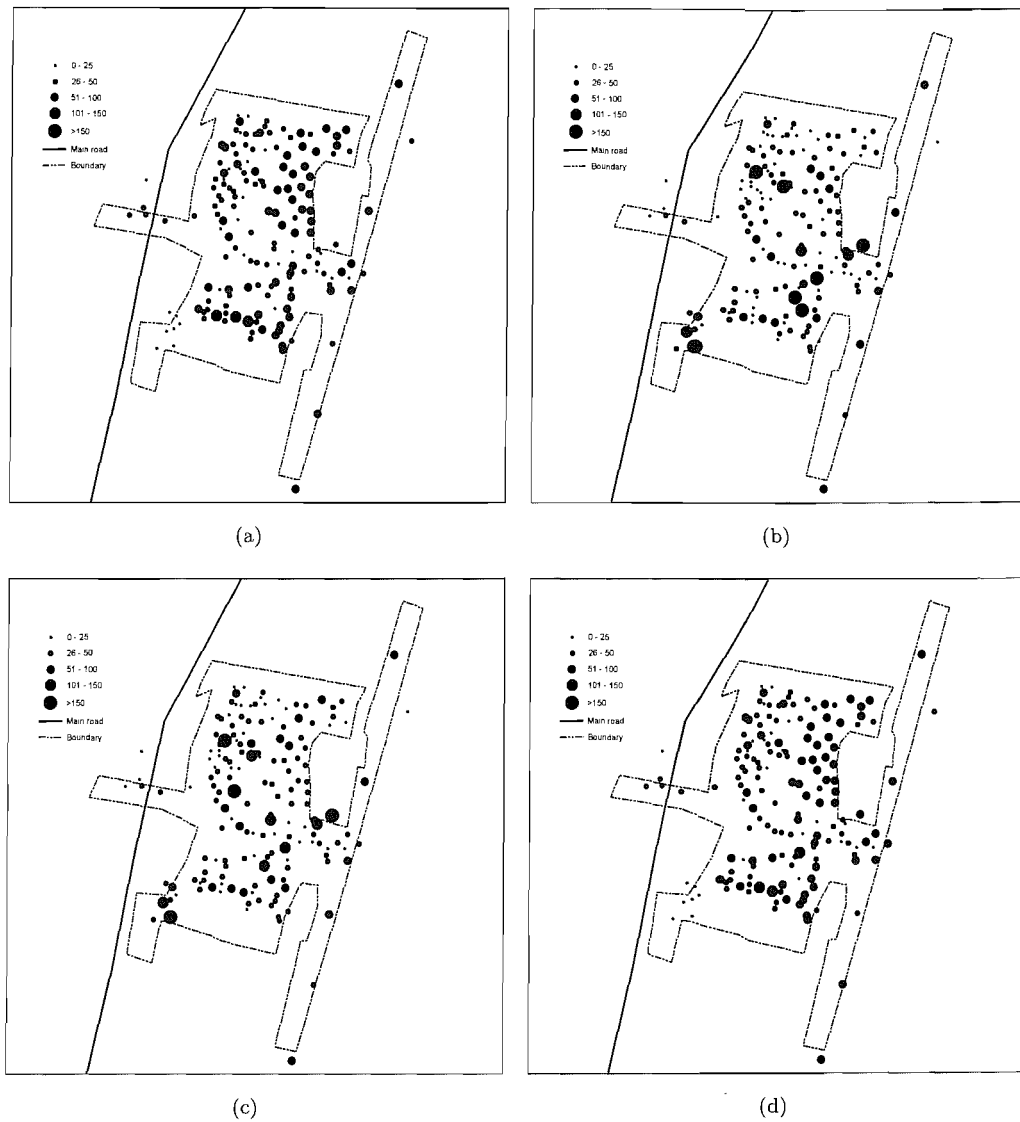


FIGURE 8.6: Individuals' spatial distribution at PC6 levels, at (a) 4 am, (b) 10 am, (c) 4 pm and (d) 10 pm.

attacker, attackee, infection location, infection time and also the change of health state of each patient over time can be accumulated and traced. Figure 8.7 shows the spatial distributions of infections ((a) is for the ABs where the infections really occurred and (b) is for patients' household). It is clear that infections in Figure 8.7(a) concentrate within a smaller number of PC6 than Figure 8.7(b), and there were no PC6 having more than 100 infections in Figure 8.7(b). The reason is that in reality during the daytime, a great number of individuals stay within a relatively small number of ABs such as school, industry and farm so that a great number of infections could occur. Figure 8.8 is an

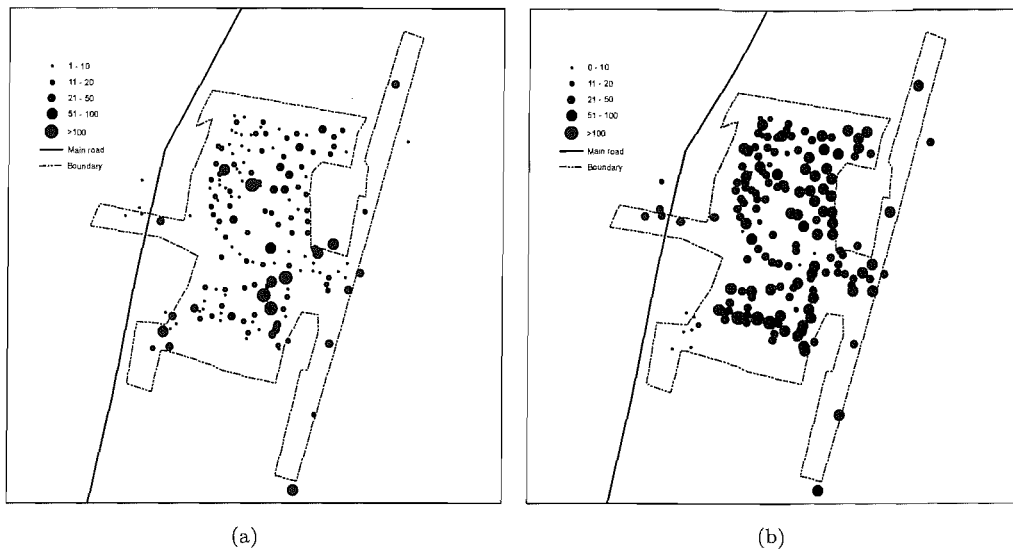


FIGURE 8.7: Spatial distributions of infections at PC6 level: (a) based on the AB where infections occurred; (b) based on the households of infected persons

example to show the number of individuals in the latent stage (i.e., be infected but not infectious), the stage of being infectious without symptoms, the stage of infectious with symptoms and the total number of infected individuals changing over time.

The ‘lag’ effect in time and ‘misplace’ effect in space within the practical epidemic data could be an interesting focus for research. The simulated results from ISTAM provided the possibility of further analysis, and deeper understanding could be gained in terms of the relation between the practical data sources and the real situation.

8.5.4 Infection distribution at different types of AB

In practice, the identification of key ABs for infection is important as this can assist possible infectious disease control measures. If enough information about individuals’ activity patterns and ABs’ physical condition is available, the percentage of infections at different types of AB for one epidemic outbreak can be generated by AB simulation. One example is Longini et al. (2005) mentioned earlier. They simulated that 28% of infections would occur within the family, 20% at household clusters and 21% at

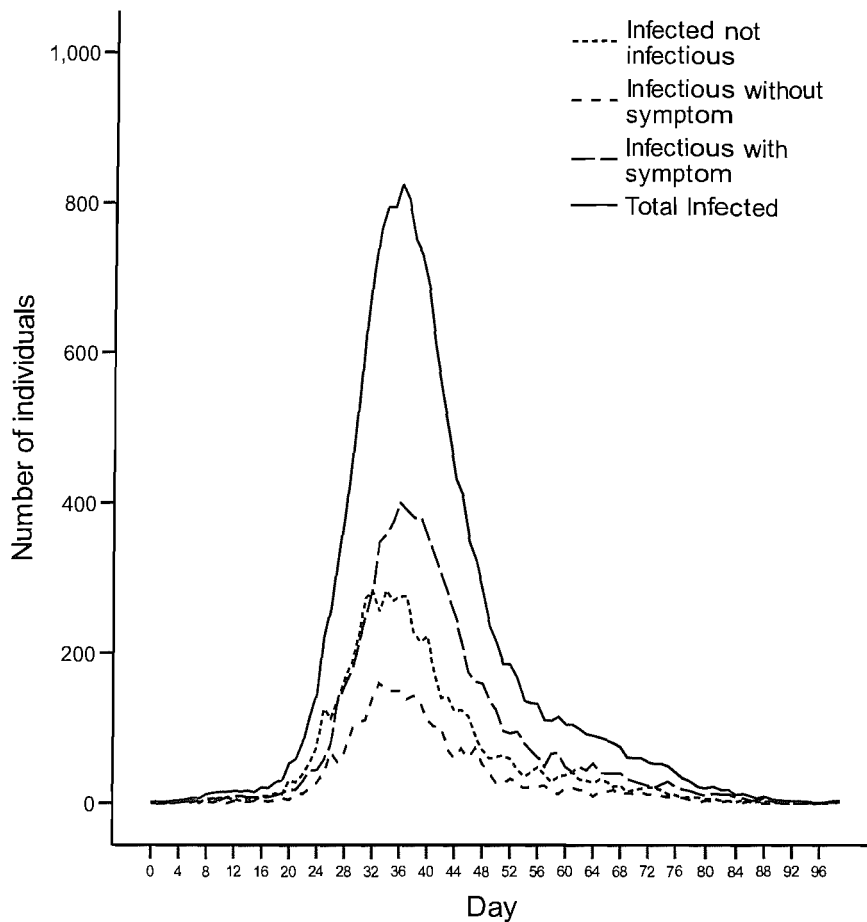


FIGURE 8.8: The changing of number of infected but not infectious, infectious without symptoms, infectious with symptoms and the total infected individuals over time.

school. Another example is Ferguson et al. (2005) who showed through simulation that infection risk comes from three sources with roughly equal proportions: (1) household, (2) workplace and school and (3) random contacts in the community. Table 8.9 is an example showing the percentage of infection at ABs generated by ISTAM. Practical statistical data are needed for comparison with the above simulation results for model calibration and validation purposes.

TABLE 8.9: Percentage of infections at different types of AB.

<i>AB Type</i>	<i>Percentage (%)</i>
Household	50.7
School	19.3
Industry	9.2
Office	8.7
Social place	5.1
Shop	2.8
Cultural place	1.3
Sport place	1.0
Farm	0.8
Post office/bank	0.5
Health care place	0.4
Other service place	0.1

8.5.5 Network analysis

Figure 8.9 is the contact network for one day's duration. Quite a few households exist for which the family numbers did not have contacts with outside persons during that day. Also from Table 8.9, we know that most infections occurred within households and at school. The arithmetic mean of clustering coefficients of this one day contact network is 0.52 with a standard deviation of 0.42. This certified the high number of 'clustered' contacts between individuals.

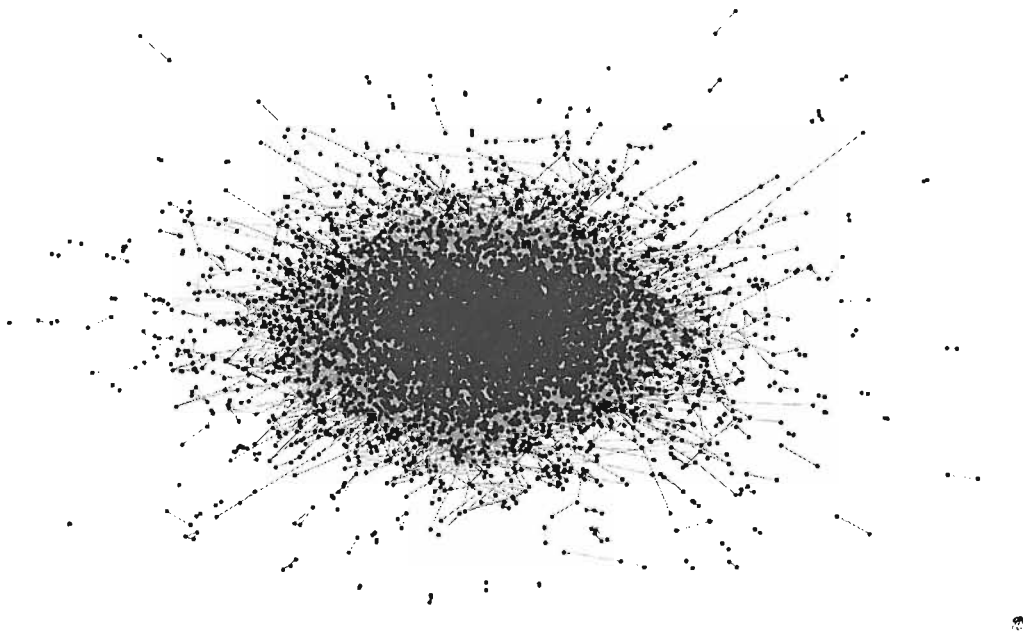


FIGURE 8.9: The contact network between individuals for one day.

8.6 Summary

Cities or urban areas “represent large concentrations of diverse peoples living on relatively small parcels of land” (LaGory 1988). City has been taken as a complex system and complexity theory has been applied to study urban dynamics (Batty 2005*b*). Even within a rather small city such as Eemnes, the human population’s demographic properties, the landscape’s physical condition and individuals’ daily activity patterns can be very complex. Inevitably, problems exist within the application of ISTAM to Eemnes. Due to the shortage of practical data, humans’ activities such as activities related with traffic and outdoor activities were neglected. In reality, humans’ contacts within transportation vehicles, especially the public transportation, always play an important rôle in infectious disease transmission amongst the human population. Some ABs are outdoor, and the weather condition should not be neglected in such a context. To include a climate and weather module is a good choice; an example is the BioWar model (Carley et al. 2004).

Chapter 9

Discussion

9.1 Reflection on two application cases of ISTAM

Although it has been applied to two cases, ISTAM is still a conceptual model. For a real application, the data needed are infectious disease data (individual case), activity data on individuals and spatial data on the interior geometries of buildings. The individual case data problem is not only technological but also legal: often it is hard to acquire real data. However, sometimes, even when data exist, the usability is restricted (Armstrong 2002). To simulate the real world, an individual-based model can cover more detail than a mathematical model. But how much detail is required? There should be a balance (Bian 2004). Too much detail not only makes the model unwieldy but also makes the implementation too time-consuming. Most important of all, the model should be appropriate to the aim of the study. Further simulation within-AB, stochastic properties, intelligent individuals and further definition of effective contact will be the foci of future research.

The emergent space-time pattern of disease in a given region depends on the parameters of both the disease transmission model and the spatial and social network structures in

place in the environment in which transmission takes place. In particular, it is expected that changes in the parameters of the (simulation) model will lead to observable changes in the space-time pattern of disease. Simulation models provide an important means for evaluating the sensitivity of emergent patterns and their space-time character to changes in model parameters.

The above introduction leads to two possible avenues for exploitation of the spatio-temporal information in emergent and possibly aggregated disease patterns: (1) it is hypothesised that statistical models fitted to the space-time patterns of aggregated disease data can be used to infer parameters of the underlying disease transmission process. (2) it is hypothesized that specific changes in the spatial environment and social network structures in which transmission occurs will provide explanations for variations in disease dynamics from place to place (e.g., town to town).

Demonstration of the above hypothesized linkages would have important implications for a range of applications. For example, recent emergent diseases (e.g., SARS, bird flu, biological agents) can pose serious hazards to human health, with little known about their transmission characteristics. In such circumstances, it is extremely important to characterize their transmission properties early in an outbreak in order to plan early warning and containment strategies. To date, little use has been made of spatio-temporal information in this regard.

If the association between particular elements of the environment and social network structure and disease outcomes can be quantified then it should be possible to map the vulnerability of settlements to specific diseases. For example, it is well-known that the behavior characteristics of individuals can be modified to reduce the likelihood of disease transmission. However, spatial elements such as the effects of public versus state school education for children and settlement structure (e.g., out-of-town supermarket versus local shop) are less well studied. Again, such knowledge would be useful in terms of planning containment strategies.

9.2 Calibration and validation of ISTAM

The credibility and reliability of the model results are achieved by model calibration and validation. The process of calibration involves assigning appropriate values to default input parameters according to existing knowledge. Model validation is defined as “substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model” (Schlesinger 1979).

For individual-based models, the model calibration and validation are very important (Ropella et al. 2002) as:

1. The results of an IBM are the emergent properties from the interaction of individuals that exist only in the model; unlike analytical model results, that is, the outcome of an IBM can be reproduced only by exactly reproducing its implementation.
2. The complex outcomes of an IBM make model errors difficult to identify.
3. It is challenging to manage populations of multiple kinds of individuals.

“A model should be developed for a specific purpose (or application) and its validity determined with respect to that purpose. If the purpose of a model is to answer a variety of questions, the validity of the model needs to be determined with respect to each question” (Sargent 2003). As reviewed by Rand et al. (2003), validation approaches include (Sargent 1988, Parker et al. 2003) (1) matching model output to measured variables in the system being modeled, and (2) matching a model’s component structures and processes to structures and processes in the system being modeled. Rand et al. (2003) applied an IBM for land use change and tried to validate their model by matching the macro-level patterns generated by their model to those commonly found in the real

world, since their aim was at the macro level and not the micro level. The models need only be valid for the specific purposes to which they are applied (Koopman 2002).

It is assumed that there are several levels of reality: at a microscopic level, interactions may be described by complicated potentials, but at a macroscopic level, the properties of the system are dominated by the aggregated effect of all microscopic interactions. ISTAM is a bottom-up model, the model is built from the individual (micro) level, the aim is to predict for the population (macro) level, between the micro and macro levels, there are other meso levels which can be aggregated from different dimensions or different angles. Existing knowledge can be applied for model calibration and validation from all these different levels (see Figure 9.1).

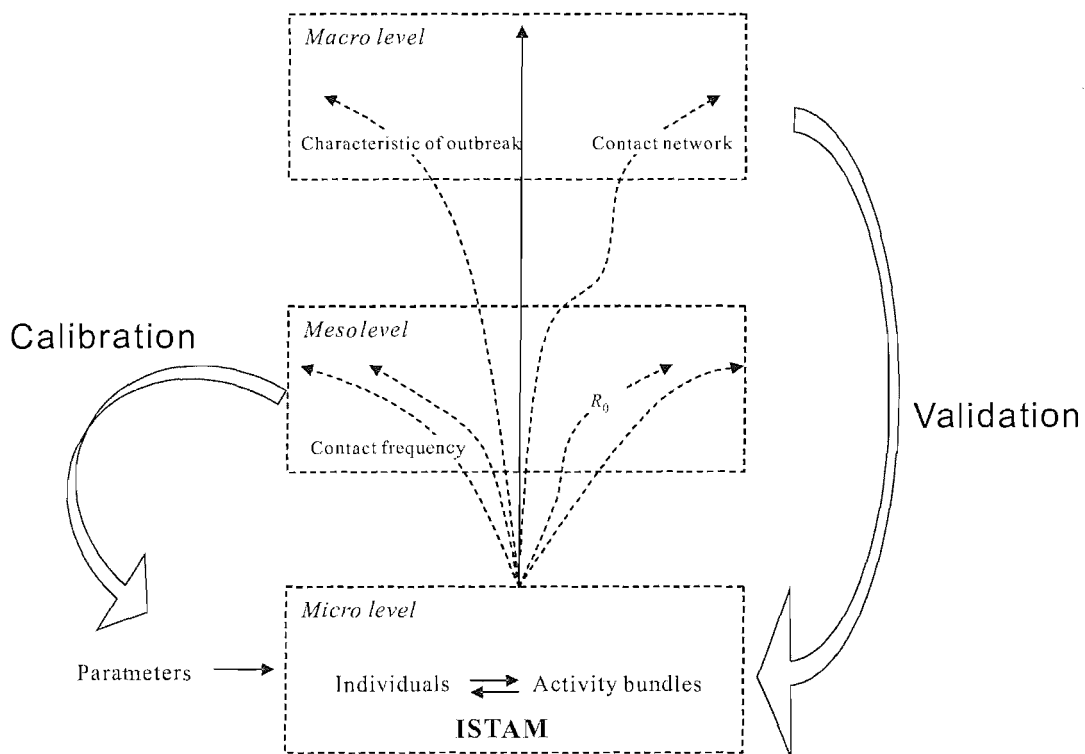


FIGURE 9.1: Process of calibration and validation of ISTAM.

As presented in Chapter 6 and Chapter 8, R_0 and contact frequency were used to calibrate infectivity index, contact frequency index and contact intimacy index. The existing research on contact networks and existing characteristics about influenza outbreaks were used to validate the simulation results. In fact, there are still more aggregated meso results that can be used for this purpose. Examples are: the real data about the changing

numbers of individuals within a certain AB (also the human density at a certain PC6) can be surveyed to calibrate individuals' activity patterns.

9.3 Selection of AB

Eubank (2002) stated that one requirement for the simulation of infectious disease transmission is that the model can identify specific geographic and demographic pathways along which disease spreads. This highlights the importance of the selection of ABs in the design and implementation of a disease transmission model. Selection of ABs varies depending on the study target and data availability.

When data on the space-time activities of individuals are difficult to obtain, the simplest method of AB selection is to include households, workplaces and schools. An example is given by Bian (2004) who simulated an influenza outbreak in 1000 individuals in a metropolitan area, and divided each individual's whole day into daytime and nighttime parts. Then, workplaces and households can be taken as two main types of AB. Another example is provided by the Ferguson et al. (2005) model which explicitly incorporates households, schools and workplaces for the simulation of an influenza pandemic of 85 million people in Southeast Asia. To include other situations, random contacts in the community associated with day-to-day movements and travel were also modelled.

A second method of AB selection is based on the interaction of between and within demographic groups of the target population. The theory behind this method is that people have most contact with people with similar ages and the same gender, and this method is consistent with traditional epidemiology and demography (Edmunds et al. 1997, Eubank et al. 2004). In a study of an influenza epidemic in a small suburban community, Elveback et al. (1976) structured a population of 1000 persons into five age groups (pre-school, grade-school, high-school, young adult and old adult). Sub-group mixing in families, clusters of neighbourhoods, schools, and pre-school playgroups (see

Table 9.1) as well as total community mixing were assumed to occur randomly. In a study to contain bioterrorist smallpox within a community, Halloran et al. (2002) go further as they include more types of AB and define different transmission probabilities for different types of AB (see Table 9.2). For example, they assume that person-to-person transmission probabilities were highest in households; lower in the day care centers, play groups, and schools; and even lower in the neighborhoods and the community at large.

TABLE 9.1: Subgrouping mixing.
(Elveback et al. 1976)

Age group (Mixing group)	Families	Clusters of Neighbourhoods	Playgroups	School	Total community
Preschool	*	*	*		*
School	*	*		*	*
Youth adults	*	*			*
Older adults	*	*			*

TABLE 9.2: Daily smallpox transmission probabilities in a simulated community.
(Halloran et al. 2002)

Contact group	Children					Adults	
	Pre-School		School			Younger	Older
	Small play groups	Large daycare centers	Elementary	Middle	High		
Small play groups	0.09						
Large daycare centers		0.05					
Elementary School			0.006				
Middle School				0.005			
High School					0.002		
Family: child	0.12	0.12	0.12	0.12	0.12	0.05	0.05
Family: adult	0.05	0.05	0.05	0.05	0.05	0.06	0.06
Neighborhood	0.00004	0.00004	0.00005	0.00005	0.00005	0.00014	0.00014
Community	0.00004	0.00004	0.00005	0.00005	0.00005	0.00014	0.00014

A third method of AB selection is based on individuals's daily space-time activities. The ABs where people are present for a long duration and with greater frequency are selected. An exception is given by hospitals and emergency wards. Although the percentage of health care activities for the whole population is of only marginal importance, the crucial roles that the hospital and emergency wards play in the process of transmission mean that they cannot be neglected. Examples are provided by Meyers et al. (2005), Epstein et al. (2002), Brouwers (2005), Longini et al. (2005) and Eubank et al. (2004).

The above methods of AB selection are not mutually exclusive. In fact, they are generated from different viewpoints and reflect the different types of model they serve.

9.4 Within-AB simulation

Chapter 7 provides a detailed analysis of the parameters of raster space AB simulation. However, these parameters are still within the domain of theoretical discussion and far from application for practical use. A few problems existed as follows:

1. The humans' space-time dynamics are too complex to be expressed by a number of parameters. In this research, it was assumed that the space-time dynamics patterns for individuals who are present within the same AB at the same time are the same at within-AB level. In fact, this assumes that the pattern is AB-based. In reality, humans' space-time dynamics depend strongly on the different rôles that they play in their joint activity.
2. The simulations are always based on the patterns that we already discovered or accepted. For example, the four spatial distribution types presented in this thesis may be not suitable or there may be other distribution types which the modeller does not know. The computational complexity is a related problem. For example, complex dynamic types of AB were not implemented within ISTAM (Figure 9.2).

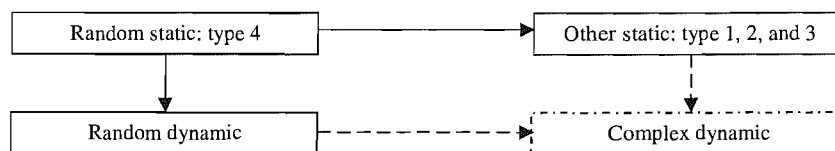


FIGURE 9.2: Relation between different types of static AB and dynamic AB.

3. The scale of space and time of the simulation may be not sufficiently fine to capture the real space-time dynamics of humans. For the spatial scale, raster

space AB simulation uses MinX and MinY to express the minimum distances in two dimensions, while in reality, the space within an AB where humans can move freely is not always a rectangle (it is also necessary to consider the layout of facilities such as furniture). For the temporal scale, real-time patterns of the dynamics of humans are ideal for simulation of the infectious disease transmission, but such patterns are seldom available.

4. Some types of public AB, although very small in number, can play a crucial rôle. Examples are sports stadia and marketplaces. A large number of individuals aggregate together periodically (e.g., weekends during league match season) or aperiodically (e.g., celebrations during holidays). Within some types of AB such as shopping centres or main traffic stations, individuals are highly mobile and the structures of the AB can be complex. It is impossible to choose a space-time scale that is suitable for all types of AB.
5. The process of infection is not clearly understood. Take influenza as an example. Although 2 m is taken as a normal distance for the infection to occur, the precise attack duration and attack distance are not known.

Considering the problems above, vector space AB simulation and rôle-based AB simulation was applied within the framework of ISTAM. Vector space AB simulation takes the space within an AB as continuous. Vector space AB simulation allows representation of individual movements at a finer spatial scale than raster space and the distance between any two individuals within an AB can be represented as a real value (rather than in terms of pixels). Importantly, Hall's (1966) distance rules can be applied directly. An example of vector space AB simulation involving the simulation of influenza transmission amongst customers and sales people inside one shopping mall is developed. This model, although simple, shows the two merits of vector space simulation: (1) the spatial layout can be expressed explicitly by the use of vector data (e.g., within a geographical information system, GIS); (2) the individuals' movement patterns inside the AB can be as detailed as required such as to reflect their activities and rôles. Rôle-based AB

simulation assumes that individuals' physical distances (also their contact frequency and contact duration) between each other are determined mainly by their rôles which they play in joint activities.

To simulate effectively a set of individuals' space-time relations within an AB, different factors need to be considered for each type of AB. Distance rules can be applied to the situation where the physical space provides enough scope for individuals to adjust their distances between each other. In such a situation, the relations between individuals based on their rôles in the joint activity are the primary factor to be considered. Thus, rôle-based simulation can be applied. When the physical conditions of the AB constrain individuals from adjusting their spatial location freely, vector space simulation is a suitable choice. In such a circumstance, mobility and density are the most important factors. For some special ABs, such as lecture rooms or aeroplanes, where individuals' space-time locations are constrained by the layout of the seating, the distances between individuals are in fact the distances between seats and individuals are static most of the time. In these circumstances, raster space simulation is more convenient to apply, and the distribution of the individuals' spatial location is the primary factor to be considered.

It needs to be pointed out that at a fine spatial scale, some factors which can be neglected in the general situation may play an important rôle. For example, during long-haul air travel, air circulation is limited and infectious droplets may persist for longer than normal. Thus, some parameters of an infection model such as the diffusion distance and diffusion duration may increase greatly. Although rôle-based simulation and raster and vector space simulation approach the same problem from different angles, these methods can be applied in combination even for a single AB (e.g., vector space and rôle-based simulation). One important decision when integrating models in this way is the relative contributions made by each component to the final model. Practical data are needed to guide this decision under different circumstances.

For a certain number of people in a certain AB, it is hard (if not impossible) to quantify the probability of infection. However, contact probability is used here as a substitute,

with the assumption of a linear relation between contact and infection. Unfortunately, contact probabilities in different types of ABs are difficult to estimate as well. For random contacts between individuals in different ABs, Meyers et al. (2005) used different probabilities: 1 for households; 0.03 for workplaces and 0.003 for other public places. The authors admitted that these parameters were based on intuition instead of estimated from data. Another example is Brouwers (2005)' MicroPox model for the simulation of smallpox transmission. In MicroPox, ABs were assigned different transmission probabilities (see Table 9.3). It is possible to compare the probability of infection at two ABs if the primary factors which control the infection can be recognized and compared. In practice, the identification of key ABs for infection is important as this may assist control measures. If enough information about individuals' activity patterns and the ABs' physical condition is available, the percentage of infections at different types of AB for one epidemic outbreak can be generated by AB simulation. Figure 6.11 is an example showing the percentage of infection at ABs generated by ISTAM for a hypothetical influenza outbreak in the campus of the University of Southampton. Empirical data are needed for comparison with the above simulation results for model calibration and validation purposes.

TABLE 9.3: Transmission probabilities for different types of AB

AB types	Transmission probabilities
Dwelling	0.25
Kindergarten, School, Hospital	0.15
Office	0.05
Travel	0.15
Emergency ward	0.25
Neighborhood	0.10

9.5 Possibility of further analysis based on the simulated data of ISTAM

The target of this PhD research changed over the course of its execution. The initial aim was to build a relatively simple pilot model for the simulation of the transmission

of infectious disease at the individual level. Then, geostatistical analysis was to be performed based on the simulated data from the above model (this part was supposed to be the focus of this PhD research). During the first two years, the author realized that the more complex the model, the more meaningful the following geostatistical analysis. Thus, the percentage of the modelling research in the whole PhD research increased. The turning point was at the end of the second year, when the author transferred from master of philosophy to PhD. After discussion with the internal examiner of the transfer: Prof. David Martin and PhD supervisor: Prof. Peter Atkinson, the author decided to delete the geostatistical analysis part and focus on the modelling research.

Here, both the pilot model and following geostatistical analysis are presented as an example to show the possibility of further analysis based on the simulated data of ISTAM (of course, ISTAM is more complex and closer to the real situation than the pilot model). The point is: data resulting from ISTAM simulation can be accumulated at micro-level and aggregated to different higher organizational levels. These simulated data can be the base for further analytical research.

9.5.1 Introduction to geostatistics

Geostatistics is a set of tools for the analysis of spatial data treated as a realization of a random function (RF). Geostatistics has been used to explore the statistical relationships in data and to test hypotheses about disease patterns. During this PhD research, a tentative model was built to explore change in the geostatistical variogram as a function of changes in simulation model parameters. Geostatistics is concerned with spatial data. That is, each data value is associated with a location in space and there is at least an implied connection between the location and the data value. The main areas of concern within geostatistics for epidemiology are: disease mapping, ecological analysis, clustering and assessment of putative sources of health hazard.

Geostatistics has its basis in Matheron's (1963) theory of regionalized variables. The most important distinction to make in geostatistics is that between model and data. In most geostatistical analyses, the first step is to compute some function such as the variogram to describe the spatial variation in a region of interest. This function must be obtained from sample data and, thus, it is a statistical function dependent on the data from which it is derived. On the contrary, if we are to use the sample function to infer further information about the region of interest then, generally, we need to adopt some formal model of the variation. Most commonly, the RF model is used. The variogram is then defined as a parameter of this model and may itself be comprised of several parameters. To relate the experimental variogram to the variogram defining the RF model, it is necessary to fit some continuous mathematical function to the observed values. Then, the fitted function estimates the model parameters.

The variogram is defined as half the expected squared difference between the random functions $Z(x)$ and $Z(x + \mathbf{h})$ at a particular lag \mathbf{h} (Matheron,1963). The mathematical expression is as follows:

$$\rho(h) = \frac{1}{2}E[\{Z(x) - Z(x + \mathbf{h})\}^2] \quad (9.1)$$

9.5.2 Pilot model description

To simulate infectious disease transmission, a pilot model was created in which space is essentially discrete (raster) although the model is not a CA in the strict sense. The model was run with a diurnal time-step. The model included three separate sets of parameters: disease transmission, environment and population parameters.

To keep the model simple in the first instance, 100 by 100 residences were distributed evenly over the raster space of 600 by 600 pixels. $M = 25$ schools were distributed using a stratified (5 by 5 cells, in this case) random sampling scheme. N companies were distributed randomly.

Population parameters included both the population structure and individual behavior. The structure of each household was based on published statistical data (see <http://www.statistics.gov.uk>). Two age groups were established (adults, children) with different behaviors. During the day, children attended their nearest school according to non-overlapping catchments as represented by the Thiessen polygon structure. This structure is not dissimilar to the state secondary school system in England. During the day, adults attended either their nearest or a randomly selected company in equal numbers per household. During the evening, some interaction (transmission) between neighbours was allowed. Related figures can be found in Appendix B.

9.5.3 Initial analysis results

From one simulation run the entire space-time data cube was extracted for further analysis. Figure B.4 shows the plot of total number of cases against day number. The information in Figure B.4 is typically all that is used to characterise diseases. However, further spatio-temporal information may be gleaned from geostatistical and similar analyses of the space-time patterns of occurrences. Figure 9.3 shows three variograms obtained for days (a) 65, (b) 149 and (c) 308 (see Figure B.4). It is clear that the character of spatial variation changes through time in subtle ways. In particular, on day 65 there are two scales of spatial variation (3 pixels, 11 pixels; corresponding to local interactions between neighbours and the effect of school catchments). These patterns are evident in adults as well as children. On day 149, the local pattern has been subsumed by the more dominant pattern attributable to school catchments, and by day 308, a larger scale trend has emerged which subsumes the pattern induced by the initial school catchment structure (e.g., day 149).

9.6 Scaling up and scaling down at different scales

Epidemiological data can be distinguished as point data and count data (Elliott et al. 2000). Point data are those that have associated exact spatial and temporal information,

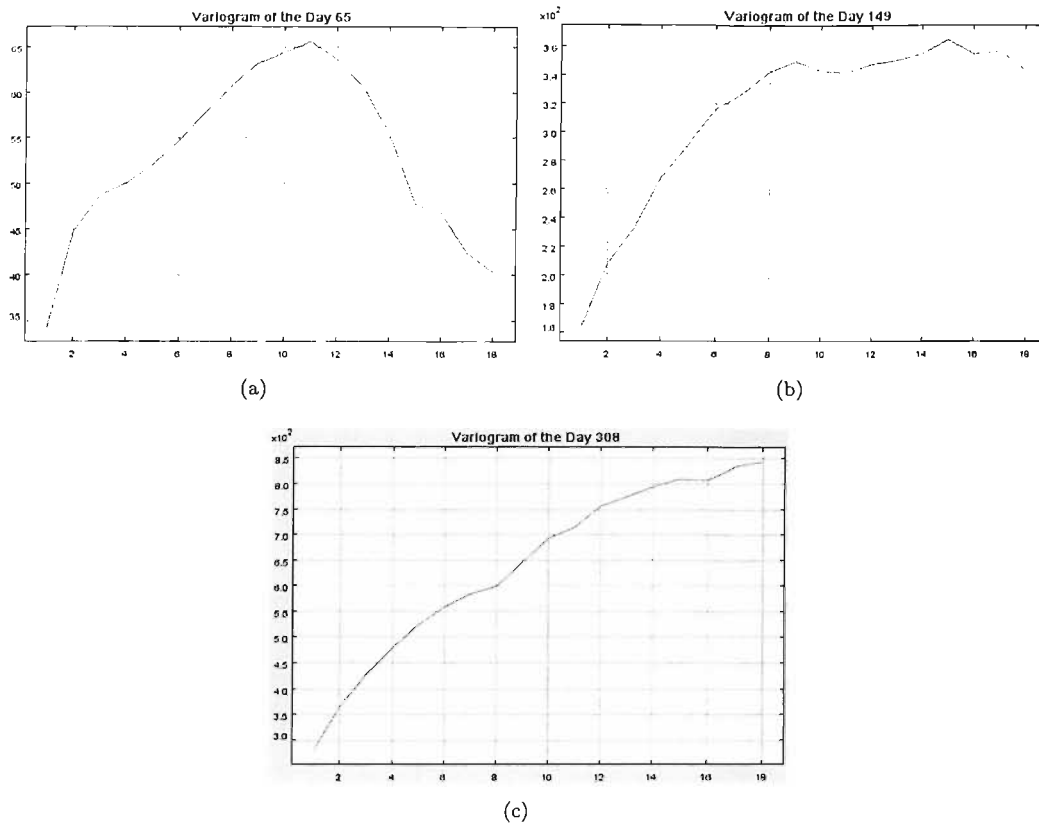


FIGURE 9.3: Variograms for days (a) 65, (b) 149 and (c) 308 showing some of the changes in spatial structure which occurs through simulation sequences.

and count data are those that are available as aggregated summaries. Sometimes health records have been aggregated within a geographical area to a population large enough to ensure non-disclosure and to protect the privacy of the individual. Confidentiality constraints usually prevent data reporting at that level (Armstrong 2002).

Within epidemiological studies, much data, particularly on number of cases, are acquired at health facilities (e.g., hospitals, clinics, dispensaries). Such data are often acquired without reference to geographical coordinates, and even where they are, disclosure and confidentiality restrictions prevent their full use. The geographical reference for a health facility is not a point (x, y) , but a catchment defined using some two-dimensional function. Catchments are often not equally weighted spatially and further may overlap. Often, catchments may be approximated using mathematical functions such that space-time distributions of disease rate can be estimated from health facility data. However,

such estimates are regularized (convolved) by the spatially varying catchments such that it is difficult to make inferences about the underlying processes of transmission.

As discussed above, most of the available epidemiological or public health data are count data, and the areas on which they were aggregated are commonly administrative such as counties or districts. It cannot be denied that political frontiers have effects on the distribution of epidemiological cases but this division brings problems which have been named as the modifiable areal unit problem (MAUP) (Openshaw 1984, Cockings & Martin 2005). For a given set of data, different aggregations or zoning systems will often show apparently different spatial patterns in the data (Openshaw 1984). Relationships between variables which are observed at one level of aggregation may not hold at the individual, or any other level of aggregation (Blalock 1964). This is the so called 'ecology fallacy'.

As pointed out by Grenfell and Harwood (1997), "A fruitful avenue for future work here is the analysis of hierarchical spatial data". It could be an interesting point to compare the individual-based, group-based, household-based or location-based transmission. The transform between these levels can be taken as scale-up and scale-down. Disease incidences have complex patterns of over a range of spatial and temporal scales (Graham et al. 2004).

9.7 Changing of humans' daily activity patterns

People who travel amongst different parts of the world may introduce infectious disease into other parts, and the transportation vehicles also can serve as mechanical vectors for the diffusion of diseases or disease vectors (Mayer 2000). As pointed out by Mayer (2000), "A well-known characteristic of contemporary society is the increasing speed with which individuals and transportation vehicles traverse the earth" (pp. 941). During the outbreak of SARS, the disease was transmitted quickly around the whole world by

airplane travel. Modern transport systems facilitate the spread of epidemics as the movement of potential vectors, disease reservoirs, and other organisms that might be involved in human disease are greatly facilitated by modern transport (Daily & Ehrlich 1996).

Schöfelder and Axhausen (Schöfelder & Axhausen 2002, Schöfelder & Axhausen 2003, Axhausen 2005) studied that change of activity space over long-term duration due to the changing expense of travel and telecommunication. It is a very interesting to study the effect of activity space change on infectious disease transmission over the long term for populations from different regions.

9.8 Problems of current ISTAM and future work

In ISTAM, both time and space are discrete. Important questions are how to choose the time unit of the whole model and how to choose the time unit of the activity pattern of an individual? This should depend mostly on the parameters of the target disease (especially the time scale of disease evolution and the time needed for infection). In RePast, the time behaves as a discrete event simulator whose quantum unit is known as a tick. In ISTAM, 15 or 30 minutes was set as one tick. The problem is: we assume the possibility of getting infection is in direct proportion to the time for the susceptible individual to be within the attack distance from an infectious individual. So given continuous time, $P(2t) = 2 \times P(t)$; given discrete time, $P(2t) = P(t) + (1 - P(t)) \times P(t) = 2P(t) - P(t) \times P(t)$ and $P((n + 1)t) = p(nt) + (1 - p(nt)) \times p(t)$. If only the $P(t) \ll 1$, $P(nt)$ is close to $nP(t)$. The choice of time and space scale can affect the micro-simulation results, especially as the attack distance and attack duration for most infectious diseases are not precisely defined. More factors such as sanitation conditions, temperature and ventilation condition can affect the probability of transmission.

In ISTAM, all ABs are indoor spaces. For outdoor ABs, factors such as weather condition and local natural environment also need to be considered. In reality, humans' activities

inside an AB not only determine the individuals' space-time dynamics, but also affect the probability of effective contact directly. For example, students are silent most of the time during lectures while they talk frequently in a bar. Such differences are important for airborne diseases. All of the above need to be added in further study.

Stochastic properties have been applied widely in ISTAM. For example, the time lines of infectious diseases are drawn stochastically from distributions. However, there are many other components which need to be expressed stochastically such as the infectivity index (Bian 2004). In fact, this value is not constant, it changes over time.

In ISTAM, the intelligent properties are overly simple and there is scope to include the individuals' reactions to the epidemic (Potash & Heinbokel 2003), as well as information transmission between individuals and the ability to change his or her activities under different situations.

For airborne infectious disease, airborne bacteria may stay suspended in the air for an extended period of time. This suggests that an infectious individual can build "an infectious buffer zone" (both space and time) along his or her movement path and also the density of the airborne bacteria in the air will increase if the infectious individual stays longer. To apply a further definition of effective contact will be a challenge for future research.

Chapter 10

Conclusion

In conclusion, AB simulation is an effective method to express small-level spatial effects. This concept bridges the gap between the spatial effect on an individual level and the individual's activity-based simulation. As shown by the application of ISTAM to a hypothetical influenza outbreak amongst a simulated first year undergraduate student population in the University of Southampton, raster space AB simulation could be an effective method to simulate infection at fine space-time scales and for a certain AB the proportion infected can be quantified. This makes the estimation and analysis of the infection risk for a certain AB possible. Given sufficient data (about both the physical structure of ABs and humans' social activities), within-AB simulation and ISTAM can be applied for practical use, for example, control measure testing for epidemics. As shown by the application of ISTAM to Eemnes, rôle-based simulation was demonstrated to be an efficient approach for within-AB simulation especially in the circumstances when detailed spatial structure information for the AB is not available.

ISTAM is a novel model for simulating the transmission of infectious disease. The two-level structure (separating the between-AB and within-AB activities) makes ISTAM flexible such that it can be applied to novel circumstances. The concept of AB plays a

key rôle: both the building of individual activity patterns and simulation within ABs depends on how well the ABs are defined and classified. In conclusion, the merits of simulation by ISTAM are:

1. ISTAM is straight forward and allows a process based representation of the real world.
2. Using ISTAM, it is easy to take into account the important factors, neglect the less important factors and include random factors in the model.
3. Using ISTAM, it is easy to consider human reactions (both active and reactive) and simulate the interaction between humans at fine scales.

As the simulation results are individual-based, AB-based and PC6-based, this provides more aspects for the calibration and validation of ISTAM. Research is currently underway to validate the model by comparison of the simulated results of the number of individuals present at different types of AB over time with related research such as Eubank et al. (2004).

ISTAM (by AB simulation) provides a new way to build a contact network based on the activity patterns of individuals. Most importantly, the contact model is a dynamic model: the contacts between individuals change over time and the order of these contacts also influences disease transmission. Further research, such as how to include the ABs as nodes, is underway.

The application of ISTAM to a hypothetical influenza outbreak amongst the population of Eemnes, a city in the Netherlands, provided an example of the simulation of the individual-based transmission of infectious disease at the city level. This involved the building of a population and spatial and social structure of a city based on limited quality data sources, and the simulation of contact processes from individuals' movements between ABs and interactions within ABs. ISTAM, if fed with activity pattern data,

can simulate effectively individuals' movements at between-AB and within-AB levels. Providing data sources are sufficient, this model can be extended to larger study areas.

Dynamic simulation models of infectious disease provide an important experimental environment for evaluating and quantifying the effects of changes in disease transmission probabilities as well as environmental and social parameters on space-time disease outcomes. ISTAM, although still in development, can be used to test disease control measures. In fact, this is a major advantage of a simulation model: the parameters can be changed to build and study 'What If' scenarios.

Appendix A

Interfaces of the applications of ISTAM



FIGURE A.1: Initial interface of ISTAM.

(The picture shows policemen in Seattle, Washington, December 1918, wearing masks made by the Seattle Chapter of the American National Red Cross during the influenza epidemic, adopted from Grosby (2003))

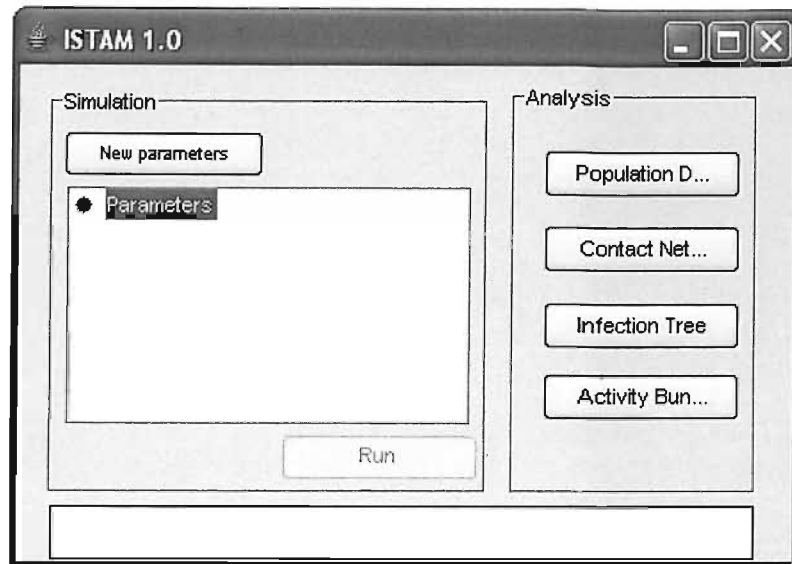


FIGURE A.2: Main interface of ISTAM.

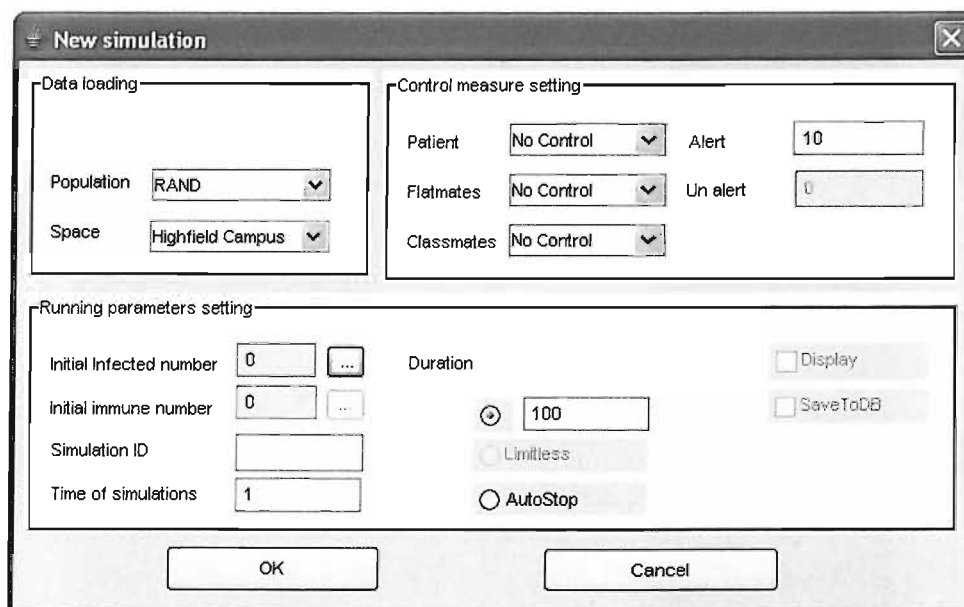


FIGURE A.3: Parameters interface of ISTAM.

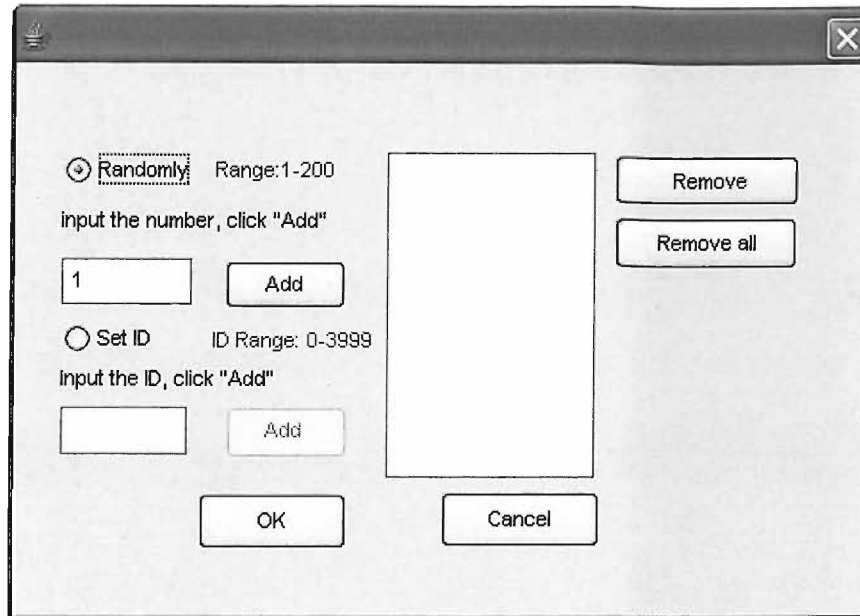


FIGURE A.4: Interface of initial infectious individuals setting.

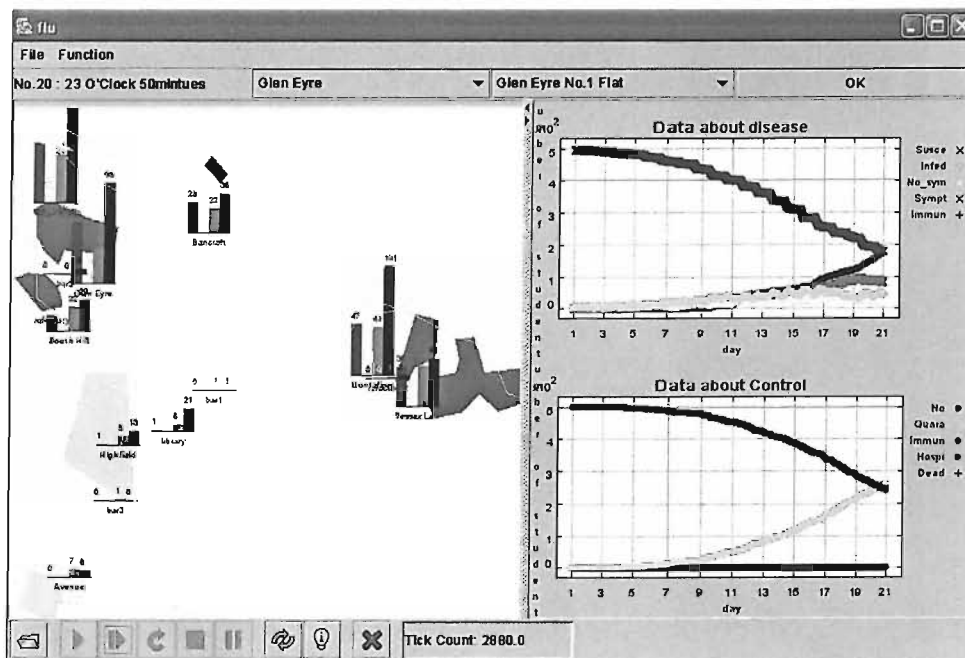


FIGURE A.5: Spatial display of ISTAM application to the campus of University of Southampton.

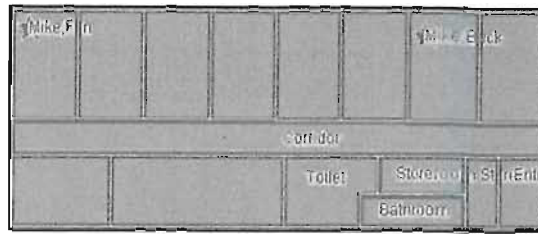


FIGURE A.6: Spatial display within flat of ISTAM application to the campus of University of Southampton.

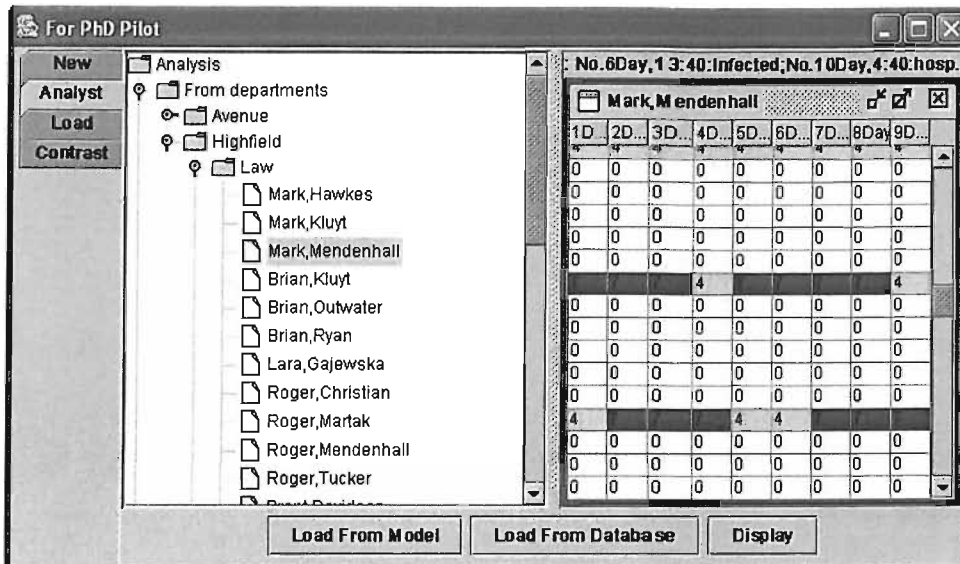


FIGURE A.7: Interface for displaying of individuals' records within ISTAM.

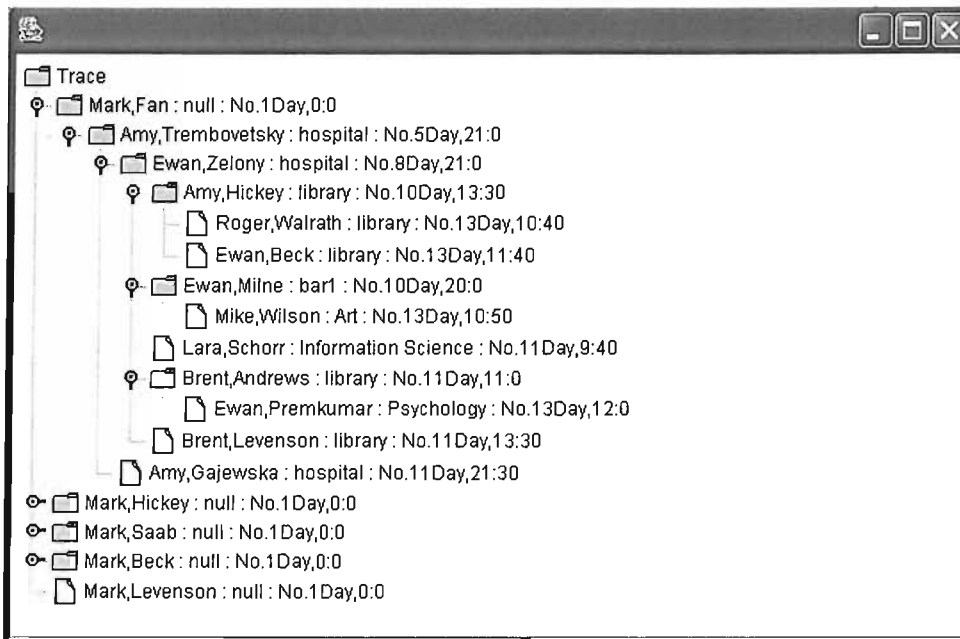


FIGURE A.8: Interface of infection tree within ISTAM.

Appendix B

Interfaces of the pilot project

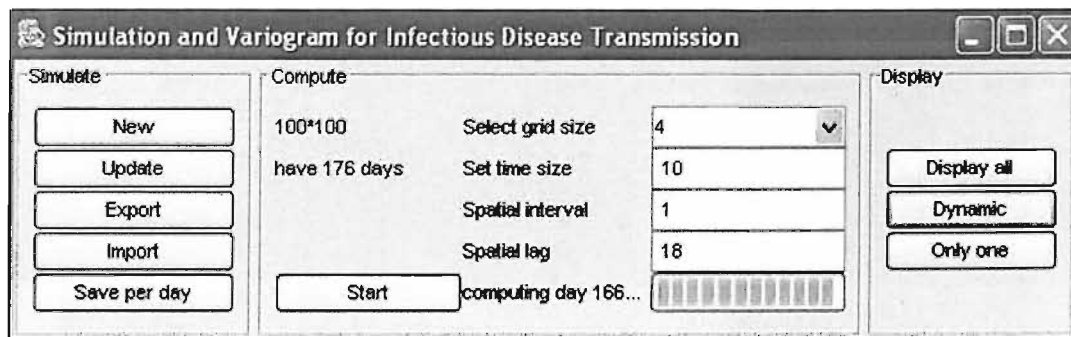


FIGURE B.1: Main interface of the variogram program.

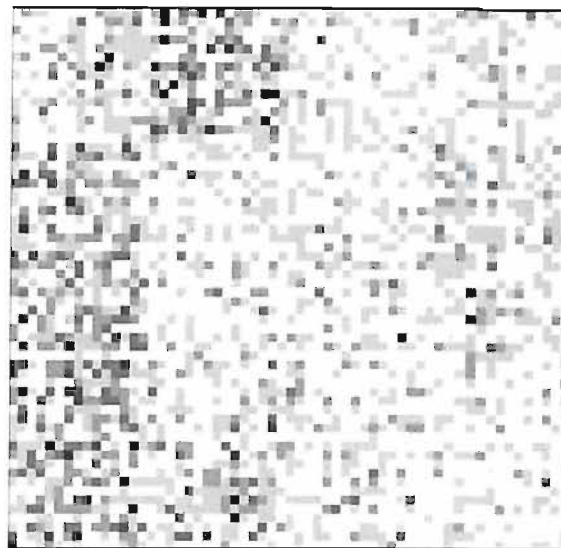


FIGURE B.2: Spatial display of the variogram program (for a cell, the darker the color, the more infected cases).

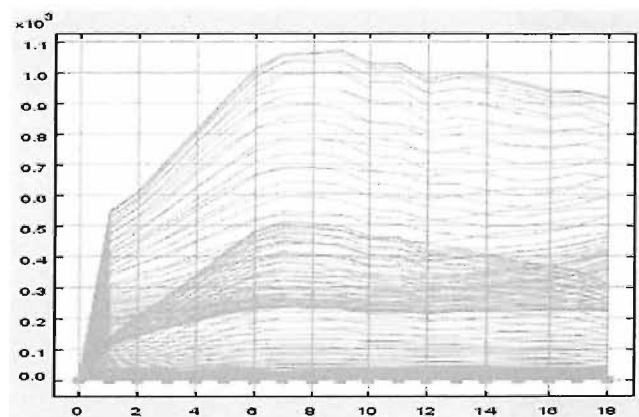


FIGURE B.3: Variograms for many days.

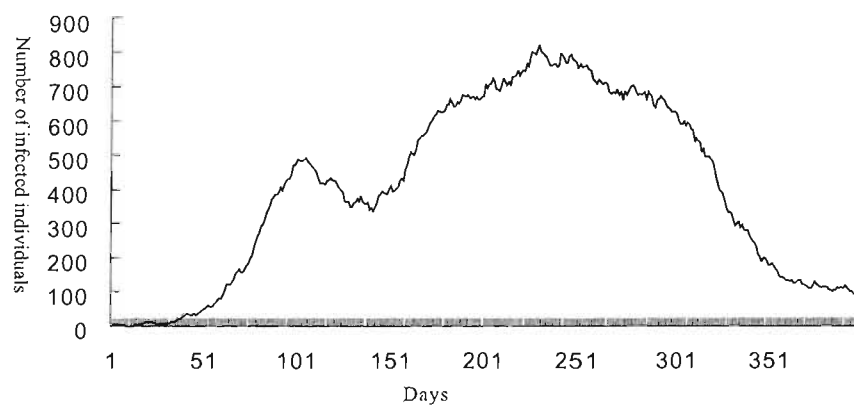


FIGURE B.4: Number of infected individuals by day.

Appendix C

Questionnaire of student activity survey

Yong Yang, PhD student, School of Geography, University of Southampton

Student Activity Survey

This form will be used for PhD study only. Your identity is not recorded, and information on individuals will be kept confidential. If you have any queries or suggestions, please email me at: yy@soton.ac.uk

Please answer the following questions (by circling or specifying the correct answer):

1. Your gender

A. Male B. Female

2. Your hall of residence

A. Glen Eyre Complex	B. Chamberlain	C. South Hill	D. Hartley Grove
E. Aubrey House	F. Romero	G. Gateley Hall	H. Bencraft
I. Wolfe House	J. Connaught	K. Montefiore	L. Other accommodations

3. When do usually you get up?

A. 7:00am B. 7:30 am C. 8:00am D. 8:30am E. 9:00am F. 9:30am G. _____

4. When do you usually leave your room?

A. 8:00am B. 8:30 am C. 9:00am D. 9:30am E. 10:00am F. 10:30am G: _____

5. When do you usually go back to your room in the afternoon or evening (approximately)?

A. 14:00 B. 15:00 C. 16:00 D. 17:00 E. 18:00 F. 19:00 G: _____

6. Where do you usually stay when "between" lectures during the daytime?

A. Library B. Residence C. Student's Union D. _____

7. Where do you usually stay when you have no lectures during the daytime?

A. Library B. Residence C. Student's Union D. _____

8. How many times do you visit the library during the daytime (Monday to Friday)?

_____ times How long is a typical visit? _____ hours

9. How many times do you visit or do the following during the evening (Monday to Friday)?

A. Stay in _____ times
 B. Library _____ times
 C. Gym _____ times
 D. Bar _____ times
 E. _____ _____ times

10. LUNCH:

A. Cooking B. Refectory C. Buying foods D. _____

11. DINNER:

A. Cooking B. Refectory C. Buying foods D. _____

When your lecture has finished, please hand your completed form to the person outside the lecture room. Thank you very much for your help!

FIGURE C.1: Questionnaire of survey on students' activity pattern.

Bibliography

- Abbey, H (1952), 'An examination of the reed frost theory of epidemics', *Human Biology* **24**, 201–233.
- Agarwal, P & R Abrahart (2003), Agent-based simulation of cognitive neighbourhoods in large-scale geographical environments, *in* 'The 7th International Conference on GeoComputation', Southampton, UK.
- Albert, R & A Barabasi (2002), 'Statistical mechanics of complex networks', *Review of Modern Physics* **74**, 47–97.
- Anderson, R M & R M May (1982), 'Directly transmitted infectious diseases: control by vaccination', *Science* **215**, 1053–1060.
- Anderson, R M & R M May (1992), *Infectious diseases of humans: dynamics and control*, Oxford University Press, Oxford, UK.
- Antikainen, H (1999), Spatial aspects of health, individual-level behavior patterns, and public health polity, *in* 'Workshop of Health Research Methods and Data', Turku, Finland.
- Arentze, T & H Timmermans (2007), 'Parametric action decision trees: Incorporating continuous attribute variables into rule-based models of discrete choice', *Transportation research part B: Methodological* **41**(7), 772–783.
- Armstrong, M P (2002), 'Geographic information technologies and their potentially erosive effects on personal privacy', *Studies in the Social Sciences* **27**, 19–28.

- Aschwanden, C (2004), Spatial simulation model for infectious viral diseases with focus on sars and the common flu, *in* 'the 37th Hawaii International Conference on System Sciences', Hawaii, USA.
- Axhausen, K W (2005), Activity spaces, biographies, social networks and their welfare gains and externalities: Some hypotheses and empirical results, Technical report, IVT, ETH Zurich.
- Bagni, R, R Berchi & P Cariello (2002), 'A comparison of simulation models applied to epidemics', *Journal of Artificial Societies and Social Simulation* **5**(3).
- Baldassare, M (1978), 'Human spatial behavior', *Annual Review of Sociology* **4**, 29–56.
- Batty, M (2003), Agent-based pedestrian modelling, *in* P. A Longley & M Batty, eds, 'Advanced Spatial Analysis: The CASA Book of GIS', ESRI Press, Redlands, USA.
- Batty, M (2005a), 'Agents, cells, and cities: new representational models for simulating multiscale urban dynamics', *Environment and Planning A* **37**, 1373–1394.
- Batty, M (2005b), *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*, MIT press, Cambridge, USA.
- Batty, M, J Desyllas & E Duxbury (2003), 'The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades', *International Journal of Geographical Information Science* **17**, 673–697.
- Bauch, C T, J O Lloyd-Smith, M P Coffee & A P Galvani (2005), 'Dynamically modelling sars and other newly emerging respiratory illnesses', *Epidemiology* **16**, 791–801.
- Beggs, C B, C J Noakes, P A Sleight, L A Fletcher & K Siddiqi (2003), 'The transmission of tuberculosis in confined spaces: an analytical review of alternative epidemiological models', *International Journal of Tuberculosis and Lung Disease* **7**(11), 1015–1026.
- Beltran, F S, L Salas & V Quera (2006), 'Spatial behavior in groups: an agent-based approach', *Journal of Artificial Societies and Social Simulation* **9**(3).

- Bian, L (2003), 'The representation of the environment in the context of individual-based modelling', *Ecological Modelling* **159**, 279–296.
- Bian, L (2004), 'A conceptual framework for an individual-based spatially explicit epidemiological model', *Environment and Planning B: Planning and Design* **31**, 381–395.
- Bian, L & D Liebner (2005), Simulating spatially explicit networks for dispersion of infectious diseases, in D Maguire, M Batty & M Goodchild, eds, 'GIS, Spatial Analysis and Modelling', ESRI Press, USA, pp. 245–64.
- Blalock, H (1964), *Causal inferences in nonexperimental research*, University of North Carolina Press, Chapel Hill, USA.
- Bonabeau, E (2002), 'Agent-based modelling: Methods and techniques for simulating human systems', *PNAS* **99**, 7280–7287.
- Bowman, J L & M E Ben-Akiva (2001), 'Activity-based disaggregate travel demand model system with activity schedules', *Transportation research part A: Policy and Practice* **35**(1), 1–28.
- Brouwers, L (2005), Micropox: a large-scale and spatially explicit microsimulation model for smallpox transmission, in 'the 15th International Conference on Health Sciences Simulation', San Diego, USA.
- Brown, D G, R Riolo, D T Robinson, M North & W Rand (2005), 'Spatial process and data models: toward integration of agent-based models and GIS', *Journal of Geographical Systems* **7**, 1–23.
- Carley, K, D Fridsma, E Casman, A Yahja, N Altman, L Chen, B Kaminsky & D Nave (2004), 'Biowar: Scalable agent-based model of bioattacks', *IEEE Transactions on Systems, Man and Cybernetics* **36**, 252–265.
- Cliff, A D, P Haggett, J K Ord & G Versey (1981), *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community*, Cambridge University Press, Cambridge, UK.

- Cockings, S & D Martin (2005), 'Zone design for environment and health studies using pre-aggregated data', *Social Science and Medicine* **60**, 2729–42.
- Collier, N (2001), Repast: An extensible framework for agent simulation, Technical report, Social Science Research Computing, University of Chicago.
- Cox, N J & K Subbarao (2000), 'Global epidemiology of influenza: Past and present', *Annual Review of Medicine* **51**, 407–421.
- Crosby, A W (2003), *America's Forgotten Pandemic: The Influenza of 1918*, Cambridge University Press, Cambridge, UK.
- Daily, G C & P R Ehrlich (1996), 'Global change and human susceptibility to disease', *Annual Review of Energy and Environment* **21**, 125–144.
- DeAngelis, D L & L J Gross (1992), Individual-based models and approaches in ecology, in 'Populations, Communities and Ecosystems', Chapman and Hall, New York, USA.
- Decock, K M & J B McCormick (1988), 'HIV infection in zaire', *New England Journal of Medicine* **319**, 309.
- Demers, M (2000), *Fundamentals of Geographic Information Systems*, John Wiley and Sons, New York, USA.
- Dibble, C & P G Feldman (2004), 'The geograph 3D computational laboratory: Network and terrain landscapes for repast', *Journal of Artificial Societies and Social Simulation* **7**(1).
- Eames, K T D & M J Keeling (2003), 'Contact tracing and disease control', *Proceedings of the Royal Society B* **270**, 2565–2571.
- Edmunds, W J, C J O'Callaghan & D J Nokes (1997), 'Who mixes with whom? a method to determine the contact patterns of adults that may lead to the spread of airborne infections', *Proceedings of the Royal Society B* **264**, 949–957.
- Eichner, M (2003), 'Case isolation and contact tracing can prevent the spread of smallpox', *American Journal of Epidemiology* **158**(2), 118–128.

- Ellegård, K (1999), 'A time-geographical approach to the study of everyday life of individuals a challenge of complexity', *GeoJournal* **48**, 167–175.
- Elliott, P, J C Wakefield, N G Best & D J Briggs (2000), Spatial epidemiology: methods and applications, in P Elliott, J. C Wakefield, N. G Best & D. J Briggs, eds, 'Spatial epidemiology: methods and applications', Oxford University Press, Oxford, UK.
- Elveback, L R, J P Fox & E Ackerman (1975), Stochastic stimulation models for two immunization problems, in D Ludwig & K. L Cooke, eds, 'Proceedings of SIMS Conference on Epidemiology', pp. 90–103.
- Elveback, L R, J P Fox, E Ackerman, A Langworthy, M Boyd & L Gatewood (1976), 'An influenza simulation model for immunization studies', *American Journal of Epidemiology* **103**, 152–165.
- Epstein, J M, D A T Cummings, S Chakravarty, R M Singa & D S Burke (2002), Toward a containment strategy for smallpox bioterror: An individual-based computational approach, Technical report, Center on Social and Economic Dynamics, Johns Hopkins University.
- Epstein, J M & R L Axtell (1996), *Growing artificial Societies: Social science from the bottom up*, MIT Press, Cambridge, Massachusetts, USA.
- Ettema, D (2006), PUMA: Multi-agent modelling of urban systems, in 'Transportation Research Board 85th Annual Meeting', Washington, D.C., USA.
- Eubank, S (2002), Scalable, efficient epidemiological simulation, in 'SAC 2002', Madrid, Spain.
- Eubank, S, H Guclu, V S A Kumar, M V Marathe, A Srinivasan, Z Toroczkai & N Wang (2004), 'Modelling disease outbreaks in realistic urban social networks', *Nature* **429**, 180–184.
- Fenner, F, D A Henderson, I Arita, Z Jezek & I D Ladnyi (1988), *Smallpox and its eradication*, World Health Organization, Geneva, Switzerland.

- Ferguson, N M, D A T Cummings, C Fraser, J C Cajka, P C Cooley & D S Burke (2006), 'Strategies for mitigating an influenza pandemic', *Nature* **442**, 448–452.
- Ferguson, N M, D A T Cummings, S Cauchemez, C Fraser, S Riley, A Meeyai, S Iamsirithaworn & D S Burke (2005), 'Strategies for containing an emerging influenza pandemic in southeast asia', *Nature* **437**(8), 209–214.
- Ferguson, N M, M J Keeling, W J Edmunds, R Gani, B T Grenfell, R M Anderson & S Leach (2003), 'Planning for smallpox outbreaks', *Nature* **425**, 681–685.
- Flamm, M & V Kaufmann (2006), The concept of personal network of usual places as a tool for analysing human activity spaces: A quantitative exploration, in 'STRC 6th Swiss Transport Research Conference', Monte Verit, Switzerland.
- Forer, P (2002), Timelines, environments and issues of risk in health: the practical algebra of (x, y, t, a) , in D. J Briggs, ed., 'GIS for Emergency Preparedness and Health Risk Reduction', Kluwer Academic Publishers.
- Fotheringham, A S & M Wegener (2000), *Spatial models and GIS: new Potential and new models*, Taylor and Francis, London, UK.
- Fuks, H & A Lawniczak (2001), 'Individual-based lattice model for spatial spread of epidemics', *Discrete dynamics in Nature and society* **6**, 191–200.
- Gammaitoni, L & M C Nucci (1997), 'Using a mathematical model to evaluate the efficacy of tb control measures', *Emerging Infectious Diseases* **3**, 335–342.
- Garrett, L (1994), *The Coming Plague: Newly Emerging Diseases in a World Out of Balance*, Farrar Straus Giroux, New York, USA.
- Gatrell, A C (1983), *Distance and Space: A Geographical Perspective*, Oxford University Press, Oxford, UK.
- Gatrell, A C & M L Senior (1999), Health and healthcare applications, in P. A Longley, D. J Maguire, M. F Goodchild & D. W Rhind, eds, 'Geographical Information Systems', John Wiley, Chichester, UK.

- Gimblett, R H (2002), *Integrating Geographic Information Systems and Agent-Based Modelling Techniques for Simulating Social and Ecological Processes*, Oxford University Press, New York, USA.
- Ginot, V, C L Page & S Souissi (2002), 'A multi-agents architecture to enhance end-user individual-based modelling', *Ecological Modelling* **157**, 23–41.
- Given, L M & G J Leckie (2003), "sweeping" the library: Mapping the social activity space of the public library', *Library and Information Science Research* **25**, 365–385.
- Golledge, R G & R J Stimson (1997), *Spatial Behavior: A Geographic Perspective*, The Guilford Press, New York, USA.
- Graham, A J, P M Atkinson & F M Danson (2004), 'Spatial analysis for epidemiology', *Acta Tropica* **91**, 219–25.
- Grenfell, B & J Harwood (1997), '(meta) population dynamics of infectious diseases', *Trends in Ecology and Evolution* **12**(10), 393–9.
- Grimm, C (1999), 'Ten years of individual-based modelling in ecology: what have we learned and what could be learn in future?', *Ecological Modelling* **115**, 129–48.
- Grimm, V & S F Railsback (2005), *Individual-based Modelling and Ecology*, Princeton University Press, Princeton, USA.
- Hägerstrand, T (1975), Space, time and human conditions, *in* A Karlqvist, ed., 'Dynamic allocation of urban space', Saxon House Lexington Book, Lexington.
- Haggett, P (2000), *The Geographical Structure of Epidemics*, Oxford University Press, Oxford, UK.
- Hall, E T (1966), *The Hidden Dimension*, Doubleday, New York, USA.
- Halloran, M E (1998), Concepts of infectious disease epidemiology, *in* K Rothman & S Greenland, eds, 'Modern Epidemiology', Lippincott Raven Publishers, Philadelphia, USA.

- Halloran, M E, I M Longini, A Nizam & Y Yang (2002), 'Containing bioterrorist smallpox', *Science* **298**, 1428–1432.
- Hanski, Ilkka (1998), 'Metapopulation dynamics', *Nature* **396**, 41–49.
- Hay, A, V Gregory, A Douglas & Y Lin (2001), 'The evolution of human influenza viruses', *Philos Trans R Soc Lond B Biol Sci* **356**, 1861–70.
- Hethcote, H W (2000), 'The mathematics of infectious diseases', *Society for Industrial and Applied Mathematics Review* **42**, 599–653.
- Holm, E, U Lindgren & G Malmberg (2000), Dynamic microsimulation, in A. S Fotheringham & M Wegener, eds, 'Spatial models and GIS: New potential and new models', GISDATA 7, Taylor and Francis, London, UK.
- Holmes, E E (1997), Basic epidemiological concepts in a spatial context, in D. T. . P. M Kareiva, ed., 'Spatial ecology: the role of space in population dynamics and interspecific interactions', Monographs in population biology, Princeton University Press, pp. 111–136.
- Huang, C, C Sun, J Hsieh & H Lin (2004), 'Simulating sars: Small-world epidemiological modelling and public health policy assessments', *Journal of Artificial Societies and Social Simulation* **7**(4).
- Huisman, O & P Forer (1998), Computational agents and urban life-spaces: a preliminary realisation of the time geography of student lifestyles, in 'Proceedings of the third International Conference on GeoComputation', Bristol, UK.
- Hutton, M D, W W Stead, G M Cauthen, A B Bloch & W M Ewing (1990), 'Nosocomial transmission of tuberculosis associated with a draining abscess', *Journal of Infectious Disease* **161**, 286–295.
- Jackle, J A, S Brunn & C C Roseman (1976), *Human spatial behavior: A social geography*, Duxbury Press, North Scituate, USA.
- Jacquez, G M (2000), 'Spatial analysis in epidemiology: nascent science or a failure or GIS?', *Journal of Geographical Systems* **2**, 91–7.

- Jacquez, G M, D A Greiling & A M Kaufmann (2005), 'Design and implementation of a space-time intelligence system for disease surveillance', *Journal of Geographical Systems* **7**, 7–23.
- Janelle, D G, B Klinkenberg & M F Goodchild (1998), 'The temporal ordering of urban space and daily activity patterns for population role groups', *Geographical Systems* **5**, 117–137.
- Keeling, M J & K T D Eames (2005), 'Networks and epidemic models', *Journal of the Royal Society Interface* **2**(4), 295–307.
- Kermack, W O & A G McKendrick (1927), 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society A* **115**, 700–721.
- Kleppner, D (1998), 'Physics in 50 years', *Physics Today* **51**, 11–3.
- Klovdahl, A S, E A Gravissb, A Yaganehdooost, M W Ross, A Wanger, G J Adams & J M Musser (2001), 'Networks and tuberculosis an undetected community outbreak involving public places', *Social Science and Medicine* **52**, 681–694.
- Koopman, J S (2002), 'Modelling infection transmission: The pursuit of complexities that matter', *Epidemiology* **13**, 622–624.
- Koopman, J S (2004), 'Modelling infection transmission', *Annual Review of Public Health* **25**, 303–326.
- Koopman, J S (2005), 'Infection transmission science and models', *Journal of infectious disease* **58**, S3–S8.
- Koopman, J S & J W Lynch (1999), 'Individual causal models and population system models in epidemiology', *American Journal of Public Health* **89**(1170-1174).
- Koopman, J S, S E Chick, C S Riolo, A L Adams, M L Wilson & M P Becker (2000), 'Modelling contact networks and infection transmission in geographic and social space using germs', *Sexually Transmitted Diseases* **27**(10), 617–626.
- Kretzschmar, M, S V D Hof, J Wallinga & J V Wijngaarden (2004), 'Ring vaccination and smallpox control', *Emerging Infectious Diseases* **10**(5), 832–41.

- Kuby, M, J Harner & P Gober (2004), *Human Geography in Action*, John Wiley and Sons, New York.
- LaGory, M (1988), 'The organization of space and the character of the urban experience', *Publius: The Journal of Federalism* **18**(Fall), 71–89.
- Last, J A (1988), *A Dictionary of Epidemiology*, Oxford University Press, New York.
- Leautier, F (2006), *Cities in a Globalizing World: Governance, Performance, And Sustainability*, World bank.
- Longini, I M, A Nizam, S Xu, K Ungchusak, W Hanshaworakul, D A T Cummings & M E Halloran (2005), 'Containing pandemic influenza at the source', *Science* **309**, 1083–1087.
- Longini, I M, M E Halloran, A Nizam & Y Yang (2004), 'Containing pandemic influenza with antiviral agents', *American Journal of Epidemiology* **159**, 623–33.
- Maier, J, R Paesler, K Ruppert & F Schaffer (1977), *Sozialgeographie*, Westermann, Braunschweig.
- Mangili, A & M A Gendreau (2005), 'Transmission of infectious diseases during commercial air travel', *Lancet* **365**(9463), 989–96.
- Matheron, G (1963), 'Principles of geostatistics', *Economic Geology* **58**, 1246–66.
- May, Robert M & Alun L Lloyd (2001), 'Infection dynamics on scale-free networks', *Physical Review E* **64**, 066112.
- Mayer, J D (2000), 'Geography, ecology and emerging infectious diseases', *Social Science and Medicine* **50**, 937–952.
- Meade, M, J Florin & W Gesler (1988), *Medical Geography*, The Guilford press, New York, USA.
- Meyers, L A, B Pourbohloul, M E J Newman, D M Skowronski & R C Brunham (2005), 'Network theory and sars: Predicting outbreak diversity', *Journal of Theoretical Biology* **232**, 71–81.

- Miller, E J, J D Hunt, J Abraham & P Salvini (2004), 'Microsimulating urban systems', *Computers, Environment and Urban Systems* **28**, 9–44.
- Miller, H J (2005), 'Necessary space-time conditions for human interaction', *Environment and Planning B: Planning and Design* **32**, 381–401.
- Moore, A B, P Whigham, A Holt, C Aldridge & K Hodge (2003), A time geography approach to the visualisation of sport, in 'the 7th International Conference on GeoComputation', Southampton, UK.
- Moore, C & M E J Newman (2000), 'Epidemics and percolation in small-world networks', *Physical Review E* **61**(5), 5678–5682.
- Morris, M (1994), Epidemiology and social networks: modelling structured diffusion, in S Wasserman & J Galaskiewicz, eds, 'Advances in Social Network Analysis', Sage, USA, pp. 26–52.
- Murray, J D (1989), *Mathematical Biology*, Springer, Berlin, Germany.
- Nagel, K & F Marchal (2003), Computational methods for multi-agent simulations of travel behavior, in '10th International Conference on Travel Behavior Research', Lucerne, Switzerland.
- Newman, M E J (2000), 'Models of the small world: A review', *Journal of Statistical Physics* **101**, 819–841.
- Newman, M E J (2002), 'The spread of epidemic disease on networks', *Physical Review E* **66**, 016128.
- Newman, M E J, I Jensen & R M Ziff (2002), 'Percolation and epidemics in a two-dimensional small world', *Physical Review E* **65**, 021904.
- Noakes, C J, C B Beggs, P A Sleight & K G Kerr (2006), 'Modelling the transmission of airborne infections in enclosed spaces', *Epidemiology and Infection* pp. 1–10.
- Openshaw, S (1984), *The Modifiable Areal Unit Problem*, Geo Books, Norwich, UK.

- Oppong, J R, A R Mikler, P Moonan & S Weis (2006), From medical geography to computational epidemiology dynamics of tuberculosis transmission in enclosed spaces, in 'Innovative Internet Community Systems', Springer Berlin, Heidelberg, pp. 189–197.
- Parker, D C, S M Manson, M A Janssen, M J Hoffmann & P Deadman (2003), 'Multi-agent systems for the simulation of land-use and land-cover change: A review', *Annals of the Association of American Geographers* **93**, 314–37.
- Pastor-Satorras, R & A Vespignani (2001), 'Epidemic spreading in scale-free networks', *Physical Review letters* **86**, 3200–3203.
- Patlak, M (1996), 'Book reopened on infectious diseases', *FDA Consumer* **30**, 19–23.
- Patlolla, P, V Gunupudi, A R Mikler & R T Jacob (2006), Agent-based simulation tools in computational epidemiology, in 'Innovative Internet Community Systems', Springer Berlin, Heidelberg, pp. 212–223.
- Pendyla, R M & R Kitamura (1998), 'Application of an activity-based travel-demand model incorporating a rule-based algorithm', *Environment and planning B* **25**(5), 753–772.
- Peuquet, D J (1999), Time, GIS and human health: Perresentation of space-time dynamics, in 'Workshop of Health Research Methods and Data', Finland.
- Pool, I D S & M Kochen (1978), 'Contacts and influence', *Social Networks* **1**, 5–51.
- Potash, P J & J F Heinbokel (2003), Modelling human behavior as a factor in the dynamics of an outbreak of pneumonic plague, in 'the 21st International System Dynamics Conference', New York, USA.
- Prietula, M. J, K. M Carley & L Gasser, eds (1998), *Simulating organizations: Computational models of institutions and groups*, The MIT Press, Menlo Park, USA.
- Quera, V, A Solanas, L Salafranca, F S Beltran & S Herrando (2000), 'P-SPACE: A program for simulating spatial behavior in small groups', *Behavior Research Methods, Instruments, and Computers* **32**(1), 191–196.

- Railsback, S F, S L Lytinen & S K Jackson (2006), 'Agent-based simulation platforms: Review and development recommendations', *Simulation* **82**(9), 609–23.
- Rand, W, D G Brown, S E Page, R Riolo, L E Fernandez & M Zellner (2003), Statistical validation of spatial patterns in agent-based models, *in* 'Agent Based Simulation', Montpellier, France.
- Read, J M & M J Keeling (2003), 'Disease evolution on networks: the role of contact structure', *Proceedings of the Royal Society B* **270**, 699–708.
- Rhodes, C J & R M Anderson (1997), 'Epidemic thresholds and vaccination in a lattice model of disease spread', *Theoretical Population Biology* **52**, 101–18.
- Riley, E C, G Murphy & R L Riley (1978), 'Airborne spread of measles in a suburban elementary school', *American Journal of Epidemiology* **107**(5), 421–32.
- Riley, R L (1974), 'Airborne infection', *American Journal of Medical Sciences* **57**, 466–475.
- Riley, S & N M Ferguson (2006), 'Smallpox transmission and control: Spatial dynamics in great britain', *Proceedings of the National Academy of Sciences* **103**(33), 12637–42.
- Ropella, G E, S F Railsback & S K Jackson (2002), 'Software engineering consideration for individual-based models', *Natural Resource Modelling* **15**, 5–22.
- Ross, R (1916), 'An application of the theory of probabilities to the study of a priori pathometry', *Proceedings of the Royal Society A* **92**, 204–230.
- Sargent, R G (1988), A tutorial on validation and verification of simulation models, *in* M Abrams, ed., 'Proceedings of Winter Simulation Conference', San Diego, CA, pp. 33–39.
- Sargent, R G (2003), Verification and validation of simulation models, *in* 'Proceedings of Winter Simulation Conference'.
- Schaerstrom, A (1996), *Pathogenic paths? A time geographical approach in medical geography*, Lund University Press, Lund, Sweden.

- Schlesinger, S (1979), 'Terminology for model credibility', *Simulation* **32**(3), 103–4.
- Schöfeller, S & K W Axhausen (2003), On the variability of human activity spaces, Technical report, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zurich.
- Schöfeller, S & K W Axhausen (2002), Measuring the size and structure of human activity spaces - the longitudinal perspective, Technical report, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zurich.
- Scott, S & C J Duncan (2001), *Biology of Plagues*, Cambridge University Press, Cambridge, UK.
- Sirakoulis, G, I Karafyllidis & A Thanailakis (2000), 'A cellular automaton model for the effects of population movement and vaccination on epidemic propagation', *Ecological Modelling* **133**, 209–23.
- Small, M & C K Tse (2005), 'Small world and scale free model of transmission of sars', *International Journal of Bifurcations and Chaos* **15**, 1745–1755.
- Sommer, R (1983), Spatial behavior, in H. H Blumberg, A. P Hare, V Kent & M Davies, eds, 'Small groups and social interaction (Vol. 1)', John Wiley and Sons Ltd, Chichester, UK, pp. 1–15.
- Steinhoff, M (2000), Epidemiology and prevention of influenza, in 'Infectious Disease Epidemiology', Aspen Publishers.
- Tobias, R & C Hofmann (2004), 'Evaluation of free java-libraries for social-scientific agent based simulation', *Journal of artificial societies and social simulation* **7**.
- Tomlinson, J, N. Bullock, P. Dickens, P. Steadman & E. Taylor (1973), 'A model of students' daily activity patterns', *Environment and Planning* **5**, 231–266.
- Vaughn, K M, P Speckman & E I Pas (1997), Generating household activity-travel patterns (hatps) for synthetic populations, in 'the 76th Annual Meeting of the Transportation Research Board', Washington, D.C., USA.

- Wallinga, J, W J Edmunds & M Kretzschmar (1999), 'Perspective: human contact patterns and the spread of airborne infectious diseases', *Trends in Microbiology* **7**, 372–377.
- Watts, D J & S H Strogatz (1998), 'Collective dynamics of 'small-world' networks', *Nature* **393**, 440–442.
- Wells, W F (1955), *Airborne Contagion and Air Hygiene*, Harvard University Press, Cambridge, USA.
- Wong, L H, P Pattison & G Robins (2006), 'A spatial model for social networks', *Physica A* **359**, 793.
- Wooldridge, M (2002), *Introduction To Multi-Agent Systems*, John Wiley and Sons, Chichester, UK.
- Yahja, A (2002), Biowar: Simulating disease outbreaks using social networks, *in* 'Computational Analysis of Social and Organizational Systems Conference', Pittsburgh, USA.
- Zheng, D F, P M Hui, S Trimper & B Zheng (2005), 'Epidemics and dimensionality in hierarchical networks', *Physica A* **352**, 659–668.