Epidemiological cellular automata: A case study involving AIDS

A thesis submitted to the Auckland University of Technology,
in fulfilment of the requirements for the degree of
Master of Philosophy (MPhil)

Ofosuhene Okofrobour Apenteng

School of Computing and Mathematical Sciences

June 2013

Supervisor: Prof. Ajit Narayanan

Table of Contents

Table of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted to the award of any other degree or diploma of a university or other institution of higher learning.

During the course of this research, the following papers have appeared:

In December 2012 I presented a paper (Apenteng, Narayanan et al. 2012) in migratory reinfection using cellular automata as a modelling and simulation tool (Chapter 4). I am also drafting a paper on the application of SEIR models to infectious disease with delay and have prepared some results using simulated data (Chapter 5). This paper will appear in Journal of Computer Science, Technology and Application (ISSN 2155-7969), USA.

Signed: 

Date: 21 May 2014

Acknowledgements

I would like to sincerely thank my supervisors and Professor Ajit Narayanan for their regular discussions which contributed to the production of this thesis. Their insightful comments and rigorous analysis of my arguments, as well as their willingness to offer practical advice, were very helpful.

I am indebted to my family who offered encouragement and bore with me the inevitable sacrifices at all stages of the production of this project. I am also grateful to all colleagues and friends, especially Dr. Waseem Ahmad who in no small way was truly supportive. You are the best.

I dedicate this:

To Jennifer,
my lovely wife, who makes me happy
when skies are gray for her massive support.

Abstract

The spread of disease is a major health concern in many parts of the world. In the absence of vaccines and treatments, the only method to stop the spread of disease is to control population movements. Human mobility is one of the causes of the geographical spread of emergent human infectious diseases and plays a key role in human-mediated bio-invasion, the dominant factor in the global biodiversity crisis. One of the most serious emergent infectious diseases in the last 30 years or so is AIDS (acquired immunodeficiency syndrome), where multiple pathogen species infect a human body. HIV/AIDS is now considered much more commonplace than previously thought. AIDS leads to interaction effects between the pathogens that may alter previously understood patterns of disease spread. There has been longstanding interest in how to model population movements in order to find optimal control strategies for a particular disease. The simulation models proposed here use cellular automata based on sound mathematical principles and epidemiological theory to model HIV/AIDS to provide a suitable framework to study the spatial spread of disease in different scenarios. This work investigates how probabilistic parameters affect the model in terms of time, location, gender, age and subgroups of the population. The cellular automaton modelling approach is used to forecast numbers of cases in different subgroups. An approach using wavelet transforms analysis is illustrated to understand the impact of delay on the spread of infectious disease. The results confirm that the higher the frequency, then the slower the spread of disease and vice versa. The thesis concludes with showing how co-infection can be modelled in future work on a theoretical base.

Keywords: SIR, SEIA Models, Reinfection Rates, Migration Scheme, Delay, Wavelets, Latent Period, Epidemics Models, Cellular Automata

Chapter 1
Introduction

The continuum of infectious disease is changing rapidly in conjunction with dramatic social and environmental changes. Worldwide, there is a fast increase in population growth with consequent increases in urban and international migration, leading to increased risk of exposure to infectious agents. The human immunodeficiency virus (HIV) infection, which leads to acquired immunodeficiency syndrome (AIDS), has become an important infectious disease in both the developed and developing countries. AIDS is one of the most serious, deadly diseases in human history and is caused by the human immunodeficiency virus (HIV) (Hogan, Zaslavsky et al. 2010). In many African countries, AIDS is already a major cause of death and it is predicted by experts that it will soon become so in Asian countries having larger scale populations. It is well known that the HIV virus has a long incubation and infectious period. In the absence of antiretroviral therapy, the median time of progression from HIV infection to AIDS is 10 years and the median survival time after developing AIDS is only 9 months. Moreover, the rate of clinical disease progression varies widely between individuals, from two weeks up to 20 years. The amount of time it takes for symptoms of AIDS to appear varies from person to person.

In Ghana, AIDS has a constant measurable incidence, both of cases and of potential transmission over a number of years. In 2004, in sub-Saharan Africa, 2 million people died of AIDS, 33.4 million were infected and 2.7 million people were newly infected with an estimated 440,000 in Ghana (see report on AIDS Epidemic Updates in (UNAIDS 2010)). Sub-Saharan Africa is more heavily affected by AIDS than any

other region of the world. In 2008 around 1.4 million people died from AIDS in sub-Saharan Africa. There is a growing need to model the effects of environmental factors, including migration patterns, on the spread of AIDS to gain an increased understanding of how AIDS spreads. Therefore, possible ways to combat that spread, not just through drugs but also through enhanced environmental control can be proposed. Such models can lead to novel hypotheses and predictions that can be grounded in the data available to health organisations responsible for modelling or controlling the spread of AIDS.

1.1 **Scientific Background**

AIDS causes mortality, morbidity of millions of people and expenditure of enormous amounts of money in public health care and disease control. AIDS was first reported in the USA on June 5, 1981, by the Center for Disease Control (CDC) which recorded a cluster of *Pneumocystis carinii* (a form of rare pneumonia) in five homosexual men in Los Angeles (Gottlieb, Schanker et al. 1981). In the initial stage, the CDC did not have an official name for the disease, often referring to it by way of the diseases that were associated with it, for example, lymphadenopathy - disease of the lymph nodes, and the disease after which the discoverers of HIV originally named the virus (Control 1982; Barre-Sinoussi 2004). The CDC later used the name 'Kaposi's Sarcoma' (a form of skin cancer) and 'Opportunistic Infections'. By 1986 there was sufficient understanding of the retrovirus to lead to a new name, 'HIV', or human immunodeficiency virus. It is now accepted that AIDS is a disease of the human immune system caused by the human immunodeficiency virus (HIV) (Weiss 1993; Sepkowitz 2001). The HIV condition progressively reduces the effectiveness of the immune system and leaves individuals susceptible to opportunistic infections and tumours.

Lack of access to health care and the existence of coexisting infections such as tuberculosis and hepatitis C may also accelerate disease progression. A person with HIV becomes sick with AIDS when his or her immune system is seriously compromised. HIV is transmitted through direct contact of a mucous membrane or the bloodstream with a bodily fluid containing HIV, such as blood, semen, vaginal fluid, pre-seminal fluid, and breast milk (AvertingHIV/AIDS 2013). The transmission can involve anal, vaginal or oral sex, blood transfusion, contaminated hypodermic needles, exchange between mother and baby during pregnancy, childbirth, breastfeeding or other exposure to one of the above bodily fluids. When HIV enters the body, it infects white blood cells known as T-cells, commonly termed as CD4+ cells, but also infects other cells such as dendrite cells (Srivastava, Banerjee et al. 2009). The HIV virus attacks CD4+ cells causing them to die while at the same time releasing billions of HIV back into the blood stream (Guide4Living 2013). The HIV virus then infects new cells and so the lethal cycle continues, as shown in Figure 1. There are several stages that come before one can progress from HIV to AIDS. The 'incubation' or induction period' is the 'latency period', which is the time between the onset of infection and seroconversion (May 1987; Donnelly and Cox 2001; Kaddar, Abta et al. 2011). The end of the disease process is death, since death from the disease is unfortunately the current prognosis for the overwhelming majority of AIDS cases (Keeling and Rohani 2008).

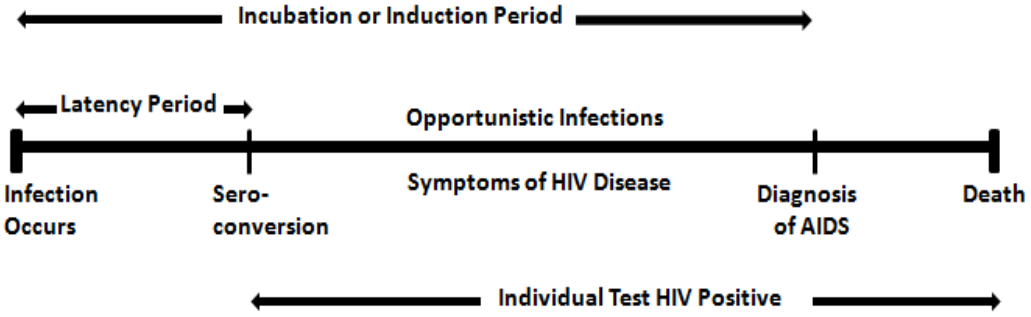

Figure 1: Basic picture of the natural progression of the AIDS disease (Keeling and Rohani 2008)

1.2 **Motivation**

Research on epidemic models that incorporate time dependent biological and environmental parameters, disease related death, varying total population and time delay is becoming one of the important areas in the mathematical theory of epidemiology. Cellular automata (CA) have been used to model the spread of a disease, with the aim of testing mathematical models of how a disease spreads by generating predictions of infection and/or identifying parameter values of the model that lead to actual real world data being fitted by the outputs of the CA model. There is little understanding of how movement of individuals from one population level to another plays a significant role in modelling the spread of AIDS using CA and mathematical models, as seen in (Sections 2.2.1, 2.3 and 3.1.1). Many models assume a homogeneous spatial geometry, with no distinctions being made between rates of spread in, for example, urban areas and rural areas, or their interaction. The aim of this research is to extend our understanding of how cellular automata can be used to model AIDS to help us understand, in the long term, the spread of disease in heterogeneous, spatially variable areas where population movements (e.g. from rural to urban) can play a critical role. The beneficiaries of this research will not just be AIDS healthcare professionals but other medical areas where there is a high degree of co-infection.

1.3 **Objectives**

Statistical and mathematical models have important roles in the study of the AIDS epidemic. Statistical models can be used for short-term forecasting of AIDS cases and mathematical models contribute to the identification of those epidemiological factors that affect the spread of the HIV virus. There is very little understanding of how the major diseases are related in mathematical terms. Without such models, we cannot make effective predictions of spread. So, we cannot currently identify the best preventive strategies. The key objectives of the study are:

1. To explain the natural course of HIV/AIDS – the relationship between the susceptible exposed infected and either immune or (removal) populations.
2. To improve the methodological basis for modelling the HIV/AIDS epidemics in terms of spatial geometry to represent the effects of population migration on HIV transmission.
3. To understand through simulation how population control strategies may have an effect on policy makers.
4. To introduce and explore principles of co-infection that may be useful for future extensions of the work reported here.

The outcome will provide some guidelines for understanding and interpreting the potential implication of current prospective changes in behaviour of population due to the spread of AIDS. The use of artificial intelligence and mathematical theories will be adopted to model the spread of AIDS. In addition, novel mathematical models will be constructed for future research in co-infection.

1.4 **Detail of the Thesis**

The thesis consists of the following sections. In the first, we look at past literature to help us to identify research gaps and frame research questions. This literature survey

will focus on mathematical models of susceptible, infected, recovery (SIR) and susceptible, exposed, infected, recovery (SEIR), respectively. The second section will focus on the methodological approach needed to design and construct cellular automata that can be used for the mathematical modelling of the HIV/AIDS disease. The third section will deal with experimental simulations and results that demonstrate the applicability of cellular automata for modelling. Finally, the fourth section of this thesis consists of an analysis of the results, followed by summary and concluding remarks concerning possible future research directions.

Chapter 2

Literature Review

In order to rationalise the research design, some existing research will be outlined and existing knowledge will be explained to justify how the research benefits this field.

It has now been established that the main cause of the AIDS disease is HIV infection, although in some very few cases it is known that not all HIV lead to AIDS (AvertingHIV/AIDS 2013). It is not clear where or how HIV/AIDS originated and there are many hypotheses (AvertingHIV/AIDS 2013).

Figure 2 shows that the total number of people living with HIV and AIDS increased from around 8 million in 1990 to 33.3 million at the end of 2009. This growth trend of the epidemic has stabilised in recent years and this is due to the significant increase in people receiving antiretroviral therapy; AIDS-related death cases have also declined (AvertingHIV/AIDS 2013).

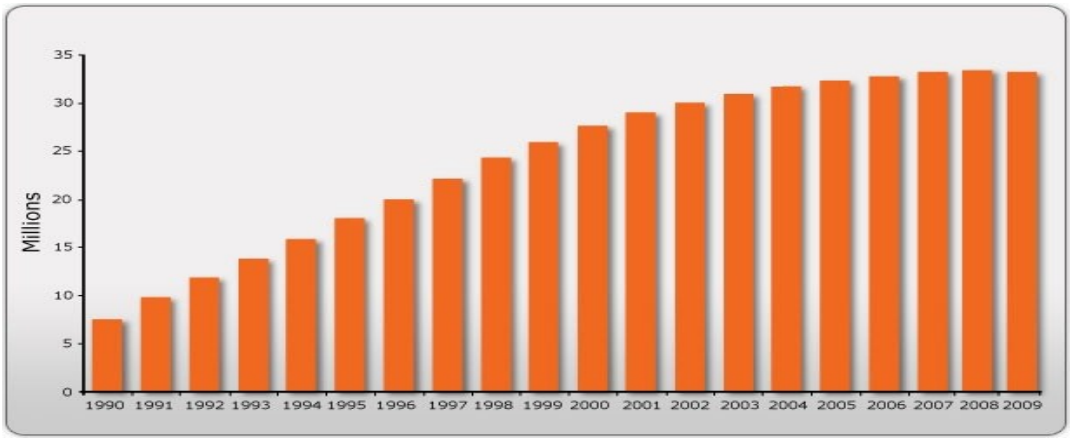

Figure 2: Numbers of HIV/AIDS cases in the world 1990-2009 (AvertingHIV/AIDS 2013)

However, global statistical data have shown that the number of AIDS cases had increased tremendously from 2009 to 2010, especially the number of children infected from 2.5 to 3.4 million (Table 1).

Table 1: Numbers of HIV/AIDS cases in the world 2009 and 2010 (Avert 2011)

	2009	2010	% increase
People newly infected with HIV	33.3 million	34.0 million	2.1
Children infected with HIV/AIDS	2.5 million	3.4 million	36.0
People newly infected with HIV	2.6 million	2.7 million	3.9

Table 2 shows that about 67.6 percent of all people living with HIV are from sub-Saharan Africa, which indicates that the region carries the greatest burden of the epidemic in the world. Epidemics in Asia have remained relatively stable and are still largely concentrated among high-risk groups. Likewise in Eastern Europe and Central Asia the number of people living with HIV has almost tripled since 2000 (AvertingHIV/AIDS 2013).

Table 2: Summary of World HIV/AIDS

Region	Adults & children living with HIV/AIDS	Adults & children newly infected	Adult prevalence*	AIDS-related deaths in adults & children
Sub-Saharan Africa	22.9 million	1.9 million	5.0%	1.2 million
North Africa & Middle East	470,000	59,000	0.2%	35,000
South & South-East Asia	4 million	270,000	0.3%	250,000
East Asia	790,000	88,000	0.1%	56,000
Oceania	54,000	3,300	0.3%	1,600
Latin America	1.5 million	100,000	0.5%	67,000
Caribbean	200,000	12,000	0.9%	9,000
Eastern Europe & Central Asia	1.5 million	160,000	0.8%	90,000
North America	1.3 million	58,000	0.6%	20,000
Western & Central Europe	840,000	30,000	0.2%	9,900
Global Total	**34 million**	**2.7 million**	**0.80%**	**1.8 million**

**** Proportion of adults aged 15-49 who are living with HIV/AIDS***

For instance, the first AIDS case diagnosed in Ghana was in March 1986 (Aboagye-Sarfo, Cross et al. 2010), five years after HIV was first found globally by the United States Centre for Disease Control and Prevention. By the end of 2004 a total of 440,000 people were found to be infected with the virus in Ghana. The majority of these cases were in the eastern region, which has the highest prevalence when compared to the northern region, which has the lowest rate of incidence (Aboagye-Sarfo, Cross et al. 2010).

Since there is no vaccine currently available for AIDS, it is essential to prevent AIDS from spreading (Kibona, Mahera et al. 2011). Many immunologists, epidemiologists, mathematicians and statisticians have been able to model the

progression of AIDS and how it affects different populations (Wang, Heather et al. 2010). Isham (1988) reviewed the context of the transmission of HIV and AIDS and Naresh et al. (2011) studied the different stages during which an HIV positive patient is not aware of the disease prior to developing AIDS. The period at which a person is declared to have AIDS varies from one country to another (Guide4Living 2013). The HIV/AIDS epidemic has continued to increase and cause havoc in sub-Sahara Africa for the past three decades. It has eroded the continent's health care system (Joshi, Lenhart et al. 2008) and is poorly controlled in Africa (Bhunu, Mushayabasa et al. 2011).

Migration is usually defined as the movement of people from one place to another temporarily, seasonally or permanently, for a host of voluntary or involuntary reasons. This includes refugees and internally displaced persons. To distinguish among categories of migrants, the word migrant is usually restricted to those who move for voluntary reasons (internally or internationally), while refugees and internally displaced persons are those who move involuntarily (Allen and Burgin 2000). Migration may impact HIV progression by connecting geographically separate epidemics and by changing sexual behaviour (Coffee, Lurie et al. 2007). Migrants are at higher risk for HIV than non-migrants. Migrant labour has played a significant role in the initial spread of HIV in the southern part of the African region (Brummer 2002). Migration may be an important driver of HIV in southern and western India. Also in south east Asia due to high numbers of migrant workers who periodically return home. See for example Overseas Filipino Workers. A study conducted by Deering, Vickerman et al. (2008) suggests that migration could play a significant role in the HIV epidemic spread in southern and western India.

2.1 Mathematical Modelling of Epidemics

The first mathematical epidemiological model was formulated in 1760 by Daniel Bernoulli (Choisy, Guégan et al. 2006) with the aim of evaluating the impact of variation of human life expectancy. In the early 1920s, Kermack and McKendrick made crucial suggestions concerning a compartment model to study the spread of infectious diseases. Their suggestions formed the foundations of mathematical epidemic modelling. The epidemic of any population at any time consists of different proportions (compartments) of susceptible, infective and recovered individuals (Kermack and McKendrick 1927). This model is discussed in more detail in section 2.2. Greenwood (1931) also formulated a different chain binomial model independently in which the number of incidences was proportional to the number of susceptibles alone. In the chain model, infections are categorized by generations and, due to this, time is considered as discrete. Subsequently, the chain-binomial models were applied to diseases such as measles and the common cold (Brimblecombe, Cruickshank et al. 1958; Heasman 1961). Anderson, May and McLean (1988) introduced the first mathematical models of the transmission dynamics of HIV to assess the potential impact of HIV and AIDS on demographic and epidemiological processes. AIDS is capable of changing the population growth rate structure in some developing countries from positive to negative values during a set timescale. En'ko formulated the progression of an epidemic: the incidence is proportional to the number of susceptibles as well as the number of infected (En'ko 1989). Normally, models identify behaviours that are unclear in experimental data, because the data are non-reproducible and the number of data points is limited and subject to errors in measurement. For instance, the fundamental result in mathematical epidemiology is that most mathematical epidemic models, including those that have high degree of heterogeneity, usually show threshold behaviours which in epidemiological terms

can be stated as: If the average number of secondary infections caused by an average infective is less than one, then a disease will die out, while if it exceeds one there will be an epidemic (Hethcote 2000). This principle, consistent with observations and quantified via epidemiological models, has been used to estimate the effectiveness of vaccination policies and the likelihood that a disease may be eradicated or removed. Thus, even if is not possible to check hypotheses accurately, agreement with hypotheses of a qualitative nature is always available (Brauer and Castillo-Chávez 2001). Mathematical modelling in epidemiology provides an understanding of the underlying mechanisms that influence the spread of disease and, in the process, it suggests control strategies (Culshaw 2006; Pietro G 2007).

2.2 The Basic SIR Model

Kermack and McKendrick (1927) formulated a classical epidemic model in which they considered a closed population with only three compartments, susceptible *S(t)*, infected $I(t)$, and removed $R(t)$, as seen in Figure 3. The susceptible $S(t)$ represents the individuals who have not yet been infected with the disease at time t. The $I(t)$ denotes the individuals who have been infected with the disease and are able to spread the disease. Finally, the removed $R(t)$ represents the individuals who have been infected and then recovered with immunity or isolated or died from the disease.

Figure 3: The SIR model

Mathematically, the changes can be described as follows:

$$\frac{dS}{dt} = -\beta SI \quad (1)$$

$$\frac{dI}{dt} = \beta SI - \alpha I \quad (2)$$

$$\frac{dR}{dt} = \alpha I \quad (3)$$

where β is the transmission rate and α is the removal rate.

Model (1-3) is based on the following three basic assumptions:

1. Migration, birth and death rates are not included in the model. So,

$S(t) + I(t) + R(t) = N$, where N is a constant (the closed population size).

2. An average infective makes contact (sufficient) to transmit infection with βN others per unit time. The law of mass action is applied to epidemics, which states that: *the rate of new infection of the disease is proportional to both the number of susceptible and the number of infectives*: $\beta N * S/N * I = \beta SI$

3. Individuals are removed from the infected class at a rate α. They either recover with immunity, become isolated or die.

The proof of *Survival and Total Size, and Second Threshold Theorem* of the SIR model is found in Appendix A. These proofs were meant to determine:

- If $\Re \leq 1$, then the solution to (1-3) is to approach the disease-free equilibrium $\lim_{t\to\infty} I(t) = 0, \ \lim_{t\to\infty} S(t) = N$
- If $\Re > 1$, then the solution to (1) is to approach a unique positive endemic equilibrium $\lim_{t\to\infty} I(t) > 0, \ \lim_{t\to\infty} S(t) > 0$

  where $\Re$ is the basic reproduction number.

Changing the assumptions, several modifications of the SIR model were developed over the years including those with two (SIS), four (SEIS, SEIR, MSIR), five (MSEIR) and six (MSRIRS) compartments (Li, Graef et al. 1999; d'Onofrio 2002; Roberts, Baker et al. 2007; Vynnycky and White 2010; Griffiths 2011), where $M(t)$ represents passively immune infants and $E(t)$ represents the exposed class in the latent period.

2.3 Some modifications of SIR model

A deterministic model is one whose behaviour is entirely no randomness. The system is perfectly understood, and then it is possible to predict precisely what will happen. A stochastic model is one whose behaviour cannot be entirely predicted (Allen and Burgin 2000).

Isham (1988) formulated a simple stochastic epidemic model that gives a good understanding of the spread of HIV infection at the initial stage of the AIDS epidemic. The merit of using the simple stochastic epidemic model is that its properties are much easier to obtain by an algebraic approach. The SIR model was appropriately based on the assumption that those who have been infected with HIV cease to transmit the infection after they have been diagnosed as being in the AIDS class. While the model was a simple stochastic epidemic, in the context of AIDS models it is well known that the number of death cases related to AIDS is important and was not included in the model. In the model, there were no opportunities for the susceptible class of the individuals to become infected by infected partners. The individuals in the population were not stratified by their age since it is well-known that the length of the incubation period tends to decrease with increasing age in adults as compared to young children who have short incubation periods. Mukandavire and Das et al. (2011) pointed out that deterministic models have

difficulty in predicting future system states accurately while stochastic models are used to find the volatility in variables to forecast the models accurately in the future. The introduction of the random volatility was as an additive white noise fluctuation to model a system with a discrete time delay. Their results showed that a change in delay and the introduction of white noise do not have an effect on the dynamic behaviour of the system. There was no comparison between the additive and multiplicative noise to show that the additive noise was beneficial.

Bhunu and Mushayabasa et al. (2011) present deterministic models to investigate how counselling and testing could modify sexual activity that will result in a decrease in the HIV epidemic in resource-limited communities. The assumption was that the HIV epidemic in Africa can be partially reduced through education efforts based on different risk activities. Their models showed that effective counselling and testing could have a positive impact in reducing the HIV epidemic. The disadvantage of this model is that there is no incubation period parameter in the model and hence no ability to represent the duration of time (in years) for the individual to develop AIDS.

A continuous model is one where the state variables change in a continuous way, and not abruptly from one state to another. Kim (2009) formulated a simple continuous model for HIV infection by using the well-known SIR model. Although exact solutions for the model cannot be found, numerical simulations assist in understanding the model. The following assumptions were made:

- A particular population, which is reasonably restricted, is at high-risk to HIV by sexual contact only. There is no vertical infection (also known as mother-to-child transmission), no infection by blood transfusion or injection-drug use.

- The population is uniformly mixed, so the probability of acquiring HIV is the same for every single individual within the community.
- Once infectious individuals are classified as AIDS patients, they are no longer engaged in the infection.
- There are no subtractions of the population except for disease-induced death.

The removed class (R) was replaced by AIDS patients class (A) (Kim (2009)), to give Model (4-6) as follows.

$$\frac{dS}{dt} = K - \alpha SI \quad (4)$$

$$\frac{dI}{dt} = \alpha SI - \beta I \quad (5)$$

$$\frac{dA}{dt} = \beta I - \delta A \quad (6)$$

where:

K : Recruitment rate to population

α : Infection rate

δ : Conversion rate (from infection to AIDS)

S : Susceptible

I : Infectious (infected) class

A : AIDS patients' class.

From the simulations results, it was observed that the infection obeys exponential growth i.e. $I = I(0)e^{\alpha Nt}$. This implies that $\frac{dI}{dt} = \alpha NI$, where $N \approx S$ and $-\beta I \approx 0$.

The disadvantage of this model was that there were unknown constant migration rates to the susceptible class. Additionally, there was no incubation period in the model. Hence, the time required for HIV infection to develop into AIDS was not included in the model.

A differential equation is linear if the unknown function and its derivatives appear to the power 1 (products are not allowed) and nonlinear otherwise. Naresh and Tripathi et al. (2011) introduced a nonlinear mathematical model to analyse the effect of contact tracing on reducing the spread of HIV/AIDS in homogeneous populations with constant migration to the susceptible class. Their results show that contact tracing might be of immense help in reducing the spread of AIDS epidemic in a population. Also, infection was reduced when infective class members know their status of infection and modify their sexual activities. Furthermore, it was observed that the individuals who know their status of HIV infection, detected by screening and contact tracing, do not contribute to the spread of the disease. In the absence of screening and contact tracing, the infected individuals continue to spread the disease without any precaution due to lack of awareness of their status.

Colijn and Cohen et al. (2009) demonstrated why SEIR models can be used to model latent co-infection of tuberculosis (TB) with different strains. The model showed that tuberculosis in people may experience latent co-infection with drug sensitive and drug resistant strains. Roeger and Feng et al.(2009) formulated simplified deterministic models of co-infection between TB and HIV. Different independent reproduction numbers were used to represent TB and HIV, respectively. The overall reproduction number for the model was the maximum reproduction number between TB and HIV. They observed that an increase in HIV prevalence also increased the level of infection for TB. The presence of TB may have a significant impact on HIV dynamics. The drawback of this model was that it did not take into account sexual-transmission, which is the cause of HIV and TB co-infection.

As can be seen from the above, a number of epidemiological models have been proposed that make assumptions about the domain that simplify epidemiological aspects. All models need to make simplifying assumptions, including our own (to be described later). Simplifying assumptions allow us to make incremental steps towards ever better and more complete models, or to model subset-problems that are analytically or numerically tractable. The question being tackled in this thesis is whether it is possible, with the use of cellular automata as the modelling tool, to incorporate more accurate assumptions that still produce realistic simulations but have increased expressive power because of the inclusion of more information.

In mathematical modelling, a deterministic model is viewed as a useful approximation of reality that is easier to build and interpret than a stochastic model, and a stochastic model only determined random process. In a continuous model, a variable changes continuously over a period of time, while in a discrete model the state of the variable changes at a discrete set of points in time intervals. Another important concept in modelling is linearity. A linear model uses parameters that are constant and do not vary throughout a simulation. For example, solutions to linear equations can be expressed in terms of a general solution and also linear models equations have explicitly defined solutions. A non-linear model introduces dependent parameters that are allowed to vary throughout the course of a simulation run, and its use becomes necessary where interdependencies between parameters cannot be considered insignificant. For example, nonlinear model typically do not have general solutions, and also nonlinear models may or may not have implicitly defined solutions. The choice between using a linear and a non-linear model is dependent upon how significantly the values of any of the parameters involved vary in relation to any of the other parameters.

2.4 **Cellular Automata Models**

Mathematical modelling is the only way to analyse the effectiveness of different disease control strategies, as stated by (Ferguson, Keeling et al. 2003). Simulation is the obvious way, since real life experimentation is impossible. Cellular automata have become a powerful tool to model the spread of disease. A cellular automaton (CA) is a mathematical 'machine' or 'organism' that lends itself to some very remarkable and elegant ideas. One of the most exciting aspects of this area of mathematics is that cellular automata arise from very basic mathematical principles which, when applied repeatedly, produce complex outcomes. More details on cellular automata are presented in the next Chapter.

Situngkir (2004) presented the use of cellular automata simulations in spatial epidemiology, with a focus on analysing avian influenza disease in Indonesia. The result of the simulation shows the spreading-rate of influenza in a simple way that describes possible preventive action through isolation of the infected as a major step of preventing a pandemic. Lichtenegger (2005) investigated the spread of epidemic disease with cellular automata models. Some of the models used were complicated due to the high dimensional configuration space. The introduction of re-growth and evolution provided new ways to interpret the host and the disease existing in dynamic equilibrium points. White (2007) also used CA for modelling epidemics. The main purpose was to understand the time course of the disease with the aim of controlling the spread of epidemic diseases. White suggested that most of the existing mathematical models failed to explain the following: 1) the individual contact processes, 2) the effects of individual behaviour, 3) the spatial aspects of the epidemic spreading, and 4) the effects of mixing patterns of the individuals. The mathematical deterministic model was used to simulate the epidemic spread of the

disease. The model was realised in a two-dimensional cellular automata, where the population was divided into three compartments: susceptible, infected and recovered (SIR).

Fresnadillo (2010) introduced a new mathematical epidemiological model based on cellular automata of graphs. The model was SEIR (susceptible-exposed-infected-recovered), where each node of the cellular automaton was assigned for only one individual. Mikler (2005) concentrated on modelling infectious disease using global stochastic cellular automata (GSCA) and also used the basic SIR model for the study of infectious disease. One of the disadvantages of using just SIR models is, as pointed out earlier, that the models do not take into account the geographical or the spatial dimensions of the region. Zorzenon dos Santos and Coutinho (2001) investigated the dynamics of HIV infection at the beginning of AIDS by using the CA approach. The model was based on three features: the global immune response to the pathogen, mutation rate of the HIV, and amount of the geographical localization of the disease. The drawback of the model was that the parameter for generating special geographical structures which play an important role in HIV spread was not fully specified. Liu and Wang et al. (2008) formulated a classic SEIR model based on ordinary differential equations. They explored the spatial behaviour of epidemic diseases that were showing seasonal trend. They implemented a model called 'dependent on the neighbourhood' to simulate the spatial-temporal movement related with the different waves of an epidemic. Alimadad (2011) modelled the spread of HIV between individuals by taking into consideration numerous features, such as type of sexual behaviour, impact of HIV testing, explicit sexual contact with different kinds of frequencies depending on their sexual behaviour, infectivity depending on the stage of the disease, sexual activity period of the disease and social

influences. Dascalus and Stafen et al. (2011) presented a multilevel simulator using hierarchical cellular automata to demonstrate how the spread of infectious disease could be modelled. The disadvantage of this model was that no quarantine and vaccination strategies were included in the model. Shih and Milne (2004) outlined how to capture the behaviour of disease spread in a tractable model with cellular automata. There were basically two scenarios: 1) where they presented an epidemic model which was used to explain the realistic pattern of disease spread with no defence or barriers to prevent the spread of the disease in a landscape distribution, and 2) where they introduced barriers to prevent the spread of disease, and showed the differences.

More detailed models using the CA approach have been proposed. The discretized quality of CA models allows individual cells and their life history to be examined and is thus ideal for small, heterogeneous populations that cannot be described accurately with ordinary or partial differential equations. However, traditional CA models have the disadvantage of not including continuous time-dependency in some of the models described in previous sections. For this reason, CA models and mathematical models will be modified and developed in which the AIDS model is described by an $n \times n$ CA lattice in which each cell corresponds to an individual cell. The reasons for using CA in this research are:

- the CA have significant role in epidemic modelling because each individual, or cell, or small region of space 'updates' itself independently in parallel
- This allows for the concurrent development of several epidemic spatial clusters, defining its new state based on the current state of its surrounding cells (locating) and on some shared rules of change.

The major drawback of CA modelling is the lack of "analytical" methods, similar to those that are useful in dealing with differential equations.

The key research problems and issues investigated in this research are:

- Previous work does not always deal with population migration – one of the key factors in the spread of HIV – although individual-to-individual transmission and population to population transmission (or compartment to compartment) is modelled. The question arises as to whether adding population migrations explicitly to a model provide any benefit over individual-to-individual transmission and compartment to compartment transfer.
- Previous research has, by and large, focused on modelling an individual disease spread. The question arises as to whether modelling two strongly-associated diseases, where the occurrence of one can lead to the subsequent occurrence of the other, adds to our understanding of HIV spread.

Chapter 3

Design of the Study

The mathematical modelling of an epidemic in this work is concerned with the infection processes mainly from person-to-person contact within populations (Isham 1988; Rao 2003; Roberts, Baker et al. 2007). As seen in the previous chapter, epidemic models have been used to study patterns of health, illness and associated factors at the population level in epidemiology (Roberts and Heesterbeek 1993). Epidemiological modelling is a base method in public health research and, in addition to modelling disease spread, helps inform evidence-based medicine for identifying risk factors for disease and determining optimal treatment approaches to clinical practice and for preventative medicine (Hove-Musekwa, Nyabadza et al. 2011; Nyabadza and Mukandavire 2011).

This research aims to develop cellular automata to model the spread of the AIDS disease taking into account geographical (spatial) distributions and mathematical models based on the literature discussed in more detail in sections 2.2.1 and 2.3. Several considerations need to be taken to achieve the objectives of this study.

3.1 **Methodology**

Formulation, validation, verification, forecasting and prediction have a pivotal role in quantitative modelling of AIDS. In quantitative models, the inputs, assumptions and the logical structure are set out clearly and therefore adaptable over time. In recent years, quantitative methods to predict future AIDS epidemics have become more sophisticated and widely used (Chen 2011; Nyabadza and Mukandavire 2011). However, an epidemiological model is only useful if it either fits existing data or is derived from existing data.

3.1.1 Mathematical modelling

As defined by Eykhoff (1974), a mathematical model is a representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in usable form. We often describe a particular phenomenon mathematically by means of a function or an equation. Such a mathematical model is an idealization of the real world phenomenon and never a completely accurate representation. The process of developing a mathematical model is termed mathematical modelling. The process involves a four-stage process: formulation, solution, interpretation and underpinning decision support. Mathematical models can take many forms, including but not limited to: dynamical systems, statistical models, differential equations, or game theoretic models. These and other types of models can overlap, with a given model involving a variety of abstract structures. Even though any model has its limitations, a good one can provide valuable results. In many cases, the quality of a scientific field depends on how well the mathematical models developed on the theoretical side agree with results of repeatable experiments. Lack of agreement between theoretical mathematical models and experimental measurements often leads to important advances as better theories are developed. For example, in this thesis, an infectious disease such as AIDS will be used. For infectious diseases, mathematical modelling uses mathematical language to describe the disease's progress to show the likely outcome of an epidemic and help inform public health interventions. There are now a number of such models in the literature. For example, as seen in South Africa (Brummer 2002; Coffee, Lurie et al. 2007) and South-West India (Deering, Vickerman et al. 2008) migration modelling of HIV have played a significant role in the spread of AIDS. Modelling the spread of AIDS has being formulated in various ways, for example, as seen in (Merli, Hertog et al. 2006; Kim 2009; Naresh, Tripathi

et al. 2011; Nyabadza, Mukandavire et al. 2011). CA has also been used to model the spread of AIDS as was seen in (Zorzenon dos Santos and Coutinho 2001; Alimadad, Dabbaghian et al. 2011). The ideas from these research programmes of migration movement of an individual from one population to the other will be used to demonstrate the significant role of the spread of AIDS. Models use some basic assumptions and mathematics to find parameters for various infectious diseases and use those parameters to calculate the effects of possible interventions, like mass vaccination programmes.

In general, model complexity can occur, for instance, when attempting to use a model given by a system of nonlinear algebraic equations. The problem maybe so large in terms of the number of factors involved that it might be impossible to account for all the necessary information in a single mathematical model. In some cases we may attempt to replicate the behaviour directly by experimental trials.

As shown in Figure 4, the present model would take into account model construction (formulation of the models) and simulation.

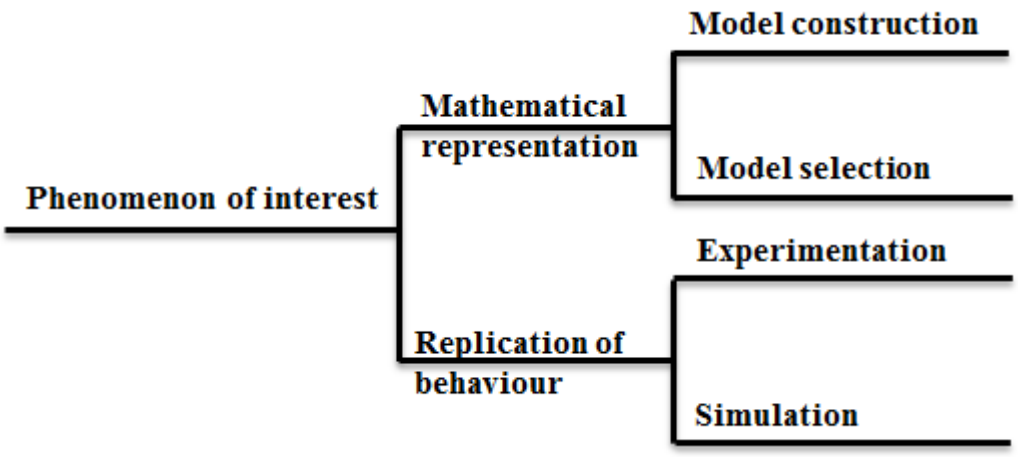

Figure 4: General nature of a model (Giordano, Fox et al. 2009)

The attempt to simulate the spread of disease within spatial is based on a simple, two-dimensional CA model. The simulation procedure has been divided into two stages, namely the construction and the implementation stages, which are discussed in the following sub-sections.

As shown in Figure 5, our methodology includes a model is intended to be an experimental and predictive model, with the aim of providing possibly new insight into how population migration may affect disease spread and co-infection.

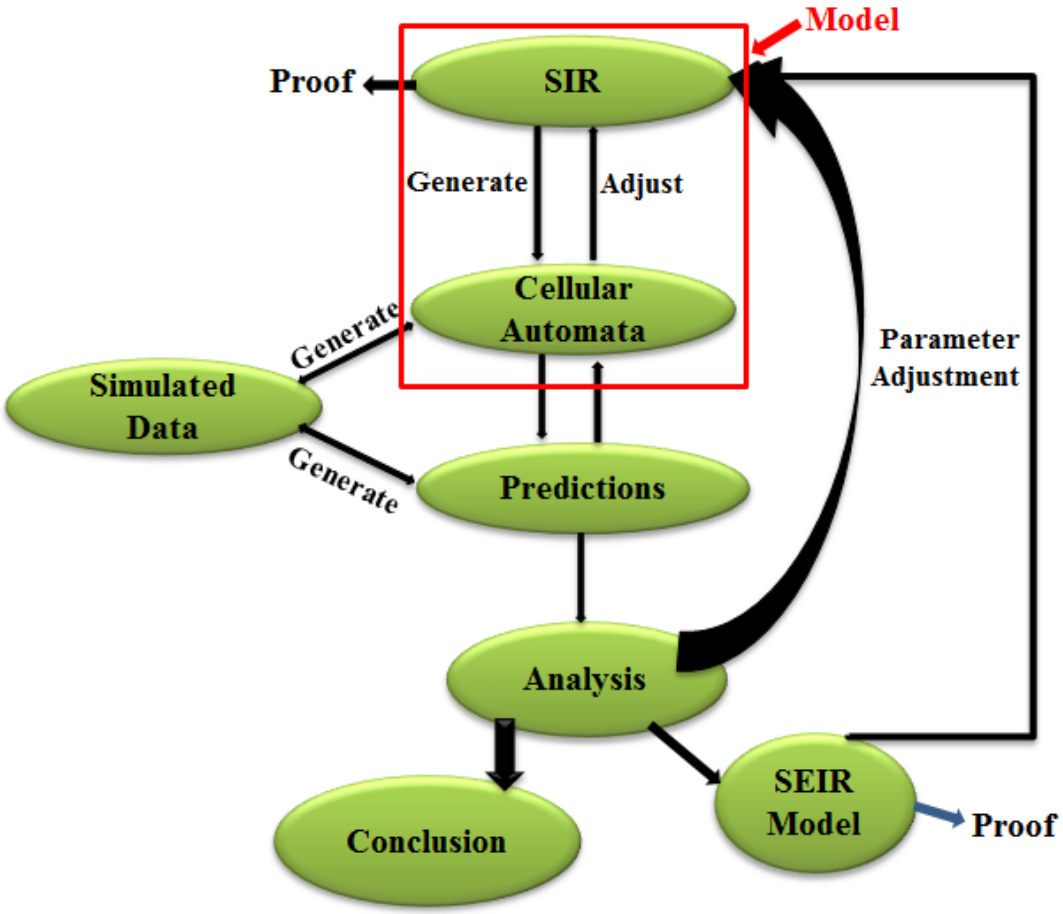

Figure 5: Generic Methodology

Our model is not meant to fit existing, real-world data, since the data is unlikely to contain the geospatial and co-infection information being modelled. It will therefore not be possible to demonstrate validity in terms of comparison of model outcomes against real-world observations/data. Instead, the approach adopted in this thesis is to generate artificial HIV datasets according to a set of geospatial parameters that are

gross simplifications of how a disease has spread, both in terms of population migration and in terms of co-infection. These rules, and hence the artificial data produced by them, are representations of a fictional HIV world. The task of the model is to generate artificial data. Parameter settings used for generating the best-fit models will then be analysed in the final part of the thesis for real-world plausibility and implications. If the model is initially shown to be mathematically sound through proof, trust can be placed in the analysis of real-world plausibility. That is, any interesting aspects found during analysis can be identified as model implications rather than the result of unsound mathematics.

First, a mathematical model will be formulated by extending the SIR model to an SEIR model. The introduction of *E* is meant to include the latency or delay aspects of HIV. This SEIR will be analysed for mathematical soundness. Artificial HIV data will be generated using CA with parameters derived from the literature concerning what is known about the effects of population migration on HIV disease spread. Where there is not sufficient information, intuition will be used to derive the rules. However, a number of simplifying assumptions will be made to demonstrate the feasibility of incorporating population migration into disease spread models. The formulated model will be used to fit the simulated data by using CA to show the geographical distribution of the spread of the HIV/AIDS disease. By adjusting and modifying the formulated model we will try to produce outcomes that can be analysed with respect to what we can learn about future SEIR models involving a geospatial dimension as well as co-infection. Based on the analysis, a number of conclusions will be drawn on what we can expect if and when real-world data involving geospatial location and co-infection becomes available.

3.1.2 Description of Cellular Automata Models

The basic element of a CA is the cell. A cell is a kind of a memory element and stores ‘states’. The model demonstrate CA, each cell can have a value of 1 or 0. That is, each cell represents an occupied cell (1) or a non-occupied cell (0). ‘Occupied’ here means by an individual member of the population being modelled. In a two-dimensional (2D) approach such as the one adopted here, the cells represent a topological map in which individuals are located in a geospatial relationship to each other, with the default interpretation of ‘north’ being located at the top of the 2D topology. In more complex simulations, the cells can have more states.

The cells arranged in this spatial map form a lattice. These cells arranged in a lattice represent a static state of the system as a whole. To introduce dynamics into the system, we add rules. The "job" of these rules is to define the state of the cells for the next time step. In CA, a rule defines the state of a cell in dependence of the neighbourhood of the cell.

Different definitions of neighbourhoods are possible. Considering a two dimensional lattice, the following definitions are common.

1. Von Neumann Neighbourhood: four cells. The cell above and below, right and left from each cell are called the Von Neumann neighbourhood of this cell. The total number of neighbour cells including itself is 5 cells (Figure 6).

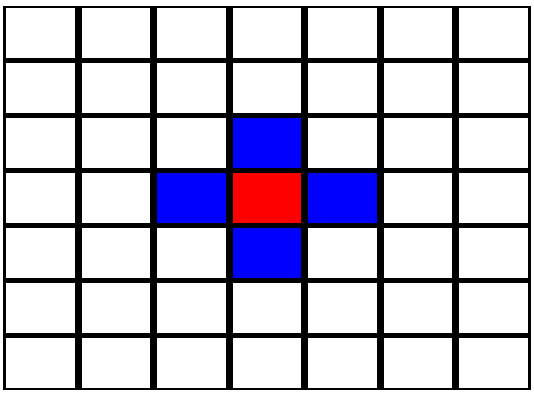

Figure 6: Von Neumann Neighbourhood

2. Moore Neighbourhood: eight cells. The Moore neighbourhood is an enlargement of the von Neumann neighbourhood containing the diagonal cells too. The total number of neighbour cells including itself is 9 cells (Figure 7).

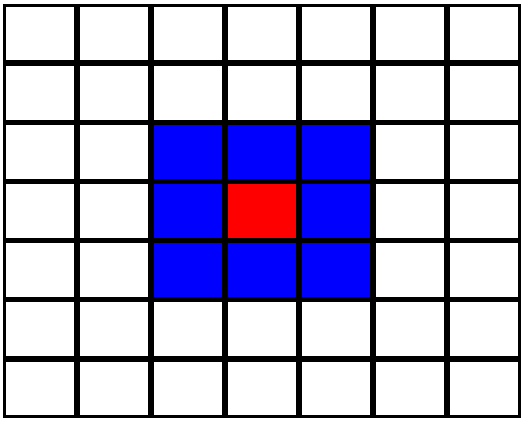

Figure 7: Moore Neighbourhood

3. Extended Moore Neighbourhood, equivalent to the description of the Moore Neighbourhood above. However, the neighbourhood reaches over the distance of the next adjacent cells. The total number of neighbour cells including itself is 25 cells (Figure 8).

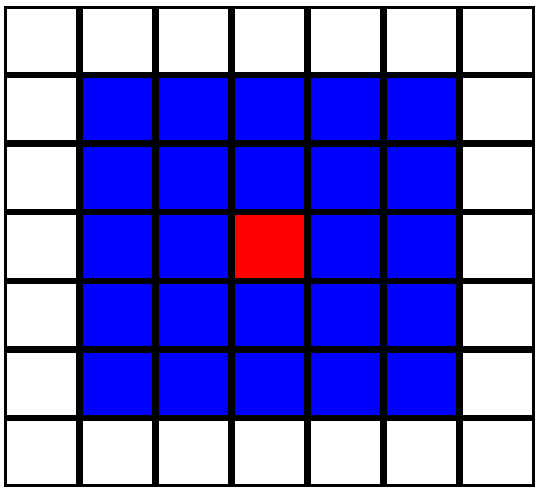

Figure 8: Extend Moore Neighbourhood

The Moore Neighbourhood Model will be used in this research to represent possible infection by neighbouring individuals of an individual at the centre of the cell. The red cell is the centre cell, the blue cells are the neighbourhood cells. The states of these cells are used to calculate the next state of the (red) centre cell according to the defined rule.

3.2 **Why cellular automata?**

Many existing epidemic models use differential equations which do not take into consideration spatial factors such as variable population density and population dynamics. They tend to assume that populations are closed and well-mixed – that is, host numbers are constant and individuals are free to move wherever they wish. CA capture the probabilistic nature of disease transmission and are characterized by their discretization of space and time. As the CA evolves, the update function, which takes into account a cell's current state and those of its neighbours, determines how microscopic interactions can influence the macroscopic behaviour of the system.

Also, CA models of disease spread all share one property: that the virtual world in which they run is an idealized and approximate one. There is therefore a trade-off between a model's degree of abstraction and its usefulness.

3.3 **Why Wavelet Analysis?**

The wavelet transform is a function which is an improved technique of the Fourier transform. In time-frequency analysis in a set of data, the classical Fourier transform analysis is inadequate due to a lack of any local information contained in the data.

Let a be a certain class of functions and $f_0, f_1, f_2, ..., f_n$ be a frequency of simple functions such that each $f(x) = \sum_{n=0}^{\infty} a_n f_n(x)$ for some coefficients a_n. We will consider a function in Laplace transform $L_2(R)$ in order to demonstrate the wavelet function, i.e.

$$L_2(R) = \left[\left\{ f : R \to C / \int_R |f(x)|^2 < \infty \right] \right. \quad (7).$$

The function becomes period when $f = 0$.

In that case we consider a wavelet function ψ such that

$$f(x) = \sum_{j \in Z} \sum_{k \in Z} d_{j,k} \psi_{j,k}(x) \quad (8)$$

where $d_{j,k}$ are wavelet coefficients and $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ are the translated and scaled version of wavelet ψ.

Let s represent each compartment and ψ be the wavelet. Then the wavelet coefficient of s at scale a and position b is defined by:

$$C_{a,b} = \int_R s(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a} \right) dt \quad (9)$$

since $s(t)$ is discrete, we will use a piecewise constant interpolation of the $s(k)$ values, $k = 1$ to *length*(s). The algorithms for this wavelet and wavelet transform are in Appendix D.

The reasons why wavelet analysis is used are:

a) to analyse the frequency of the spread of the disease; and

b) to understand the impact of delay in the spread of AIDS disease.

These will be achieved by analysing the frequency and interpreting exposed as delay.

The following research questions will be addressed:

1. How can the spread of HIV/AIDS disease be modelled to reflect geographical (spatial) distributions of AIDS by using CA and mathematical models?
2. What are the implications of our model for future co-infection models?

The following statement will be investigated:

- The higher the frequency, then the slower the spread of disease and vice versa.

Chapter 4

Experimental and Results

As noted in the preceding chapters, this chapter will cover the experimental work of the research. We will formulate a model based on Figure 5, sections 5.2, 5.3.2 and 5.3.3. The experiments were done based on the methodological approach discussed in section 3.1. The following sections first provide an overview of the experiments.

4.1 **SIR Model: CA**

A CA model for epidemic spreading was defined by White, del Rey et al. (2007) as a set of (L, Q, V, f).

Where:

- L is the space where the automata operates. It is represented as a lattice of cells (or elements).
- Q is the space of states of cells. Each cell assumes a state and the space of states represent the diverse, finite and integer possible states.
- V is the set of neighbour cells of a given cell - the neighbourhood – the environment where the automata time-space evolution takes place.
- f is the local transition rule. When applied to a cell it determines the state assumed at the next time step.

4.1.1An SIR Model

The model was with non-lethal disease with permanent immunity due to lack of vaccine. The natural birth and natural death rate of these two parameters control the birth and death rate in the population. These two parameter rates are discarded. The

infection procedure takes place when susceptible individuals are infected through contact to infected individuals.

4.1.2 Design of the Cellular Automata

A 2D lattice is used in order to define the dimensionality of the CA model. Each of the cells in the automaton has a unique index, and is stored using a one dimensional array. However, the coordinates are easier to handle by defining them as x and y coordinates. Moore and Von Neumann neighbourhood are used during runtime.

4.2 **Simulation setup**

There are many implementations of CA using the MATLAB programming environment (Athanassopoulos, Kaklamanis et al. 2012). We will offer simple processes coupled with power. CA codes were written to deal with the simulations below. There are two models; the first model is a three state model: susceptible, infected and recovery. The second model is an extension of the first model by adding an exposed state and making it a four state model.

4.3 **Simulation scenario**

In this research, r is the radius used for the spatial neighbourhood of an individual. The 1st order Moore neighbourhood is defined as the 8 nearest neighbours using a 3x3 neighbourhood, with the individual at the centre, i.e. $r = 1$. The degree of infectiousness in this spatial dimension, the value of the infected parameter is used to calculate the probability ρ that the current cell is infected in the next time step depending on the values of neighbours:

$$p = \frac{2 * (SUM\_DIS)}{(2r+1)^2 - 1} \qquad (10)$$

where the denominator $(2r+1)^2 - 1$ is the cardinality of the neighbouring cell and SUM_DIS is infected parameter.

The original SIR model formulated by Kermack and McKendrick (1927) was used without any modification to provide a benchmark for the basis of the research work. However, one addition to their SIR model is the use of a small number of initial, geographically dispersed populations. No migration patterns were specified for this initial SIR model. Hence, population spread was based purely on infection spread.

4.3.1 Experiment and Results

Experiments are performed on populations with domain size of 50x50, 60by60, 70x70 & 90x90 representing the four geographically dispersed populations P_1-P_4, respectively. Ten infected individuals were introduced into the four different populations. Table 3 summarises the parameter values used.

Table 3: Parameters of simulation

Pop 1	Pop 2	Pop 3	Pop 4	δ	β	α
50x50	60x60	70x70	90x90	14	0.25	7

4.3.2 Description of $(SIR)^n$ Model

In this section, the CA model for $(SIR)^n$ using multiple populations and a predefined migration pattern (which can vary) are introduced with the aim of generating within-population and between-population patterns and trends, where *n* represents different number of population sizes. This is to distinguish between individuals who were infected and recovered but who subsequently (at suitably identified time lags) become susceptible to the same or other disease. There is growing interest in how population movements affect infection and re-infection (Aagaard-Hansen, Nombela et al. 2010). The approach proposed here, as described in section 3.1, uses CA to

model re-infection under one possible migration scheme and demonstrates how population movements from one population to another population (i.e. towns, districts, cities and regions) can lead to different patterns of re-infection among four populations of different sizes.

4.3.3 Rules for Disease Spread

The rules described hereafter determine the state transitions of individual cells in the CA for the (SIR)^n model (Figure 9):

- A cell changes its state from *susceptible* to *infected* ($S \rightarrow I$) when it comes in contact with an infected cell in its defined neighbourhood and the mode of transmission rate is β.
- The state of the cell changes from *infected* to *recovered* ($I \rightarrow R$) after being in state I for a given α. In state I, the cells are capable of passing on the infection to neighbouring cells.
- The state of a cell changes from recovery to susceptible $(R \rightarrow S)$ after being in state R for a given time δ which takes an individual to move from recovery state back to susceptible state.

Blue, red and orange squares correspond to 0, 1 and 0.1 with susceptible individuals, infected individuals and recovery individuals, respectively, in Figure 9. The light blue colour represents the R to S recurrence. 365 iterations were simulated, with one iteration representing a day.

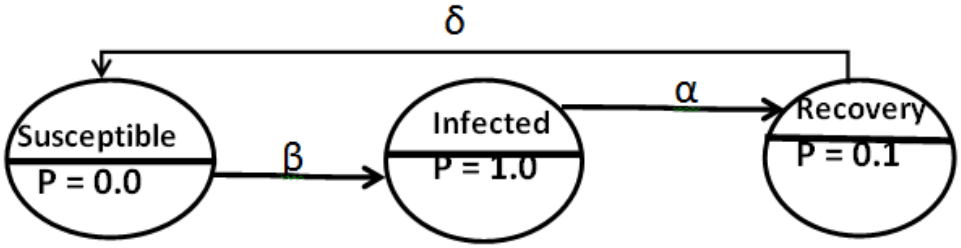

Figure 9: $(SIR)^n$ model

As shown Figure 9, the addition of δ in the thesis is an extension of the SIR model done by Kermack and McKendrick (1927).

$$\frac{dS}{dt} = \delta R - \beta SI \qquad (11)$$

$$\frac{dI}{dt} = \beta SI - \alpha I \qquad (12)$$

$$\frac{dR}{dt} = \alpha I - \delta R \qquad (13)$$

The analysis of this $(SIR)^n$ model is presented in Appendix B. The proof demonstrates the steady state, the equilibrium and the basic reproduction number.

4.3.4 Simulation Setup

Four spatially separated populations (P_1-$P_{4)}$ were created with a migratory schema described in Figure 10. Each arrow represents the movement of one randomly chosen infected individual per time step making a move to another population (five such movements in Figure 10).

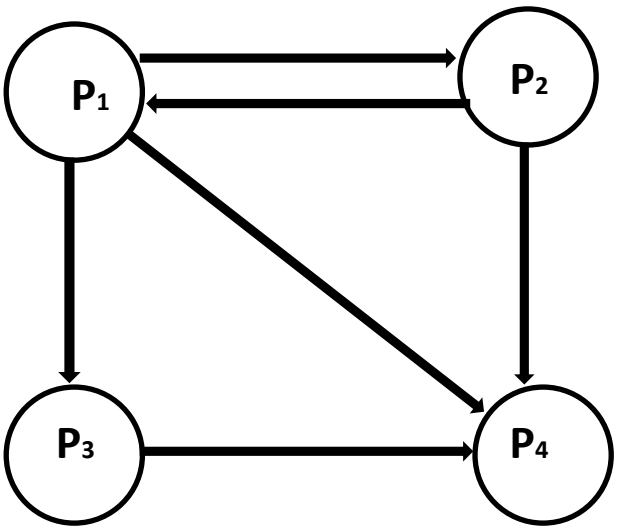

Figure 10: Migration scheme model

Figure 10 represents how the migration rules are applied. The movement of an infected individual depends on both transition probabilities and are deterministic. In P_1 and P_2, the movement of the infected individual is probabilistic. For example, with P_1, the migrant rule is either from (P_1 to P_2), (P_1 to P_3) or (P_1 to P_4) and the decision for any these three routes is selected by random probability. With the migrants rule applied to P_2, the movement of an infected individual is either from (P_2 to P_1) or (P_2 to P_4) and the decision for these two routes to occur is random probability. The migrant rule in P_3 is deterministic, while no movement occurs in P_4. These are based on the CA written codes.

Figure 11 shows the simulations of model (a-3) over 365 iterations, showing plots of population 1 to population 4 for grids of 50x50, 60x60, 70x70 and 90x90. Pop.1, Pop.2, Pop.3 and Pop.4 respectively indicates that no migration of individuals infected with HIV migrated from one population to other.

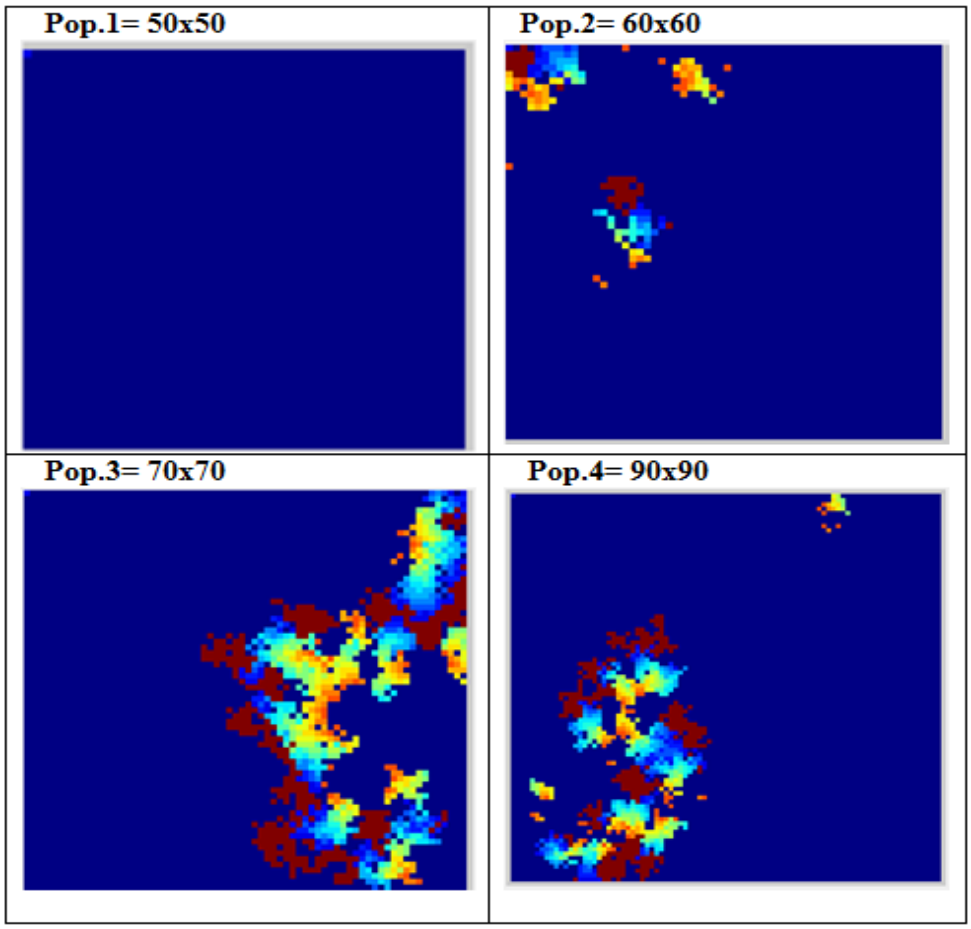

Figure 11: A dynamic spread process without migration

In each of the four populations, 20 HIV individuals were introduced in each of the populations. The colours appears as red (HIV), light green (AIDS), orange (re-infection) and blue (Susceptible). In the Pop.1 model, the shown lattice contains 50x50 cells, with no visible period after 365 steps. It can be seen that the 20 introduced HIV individual populations die as compared to Pop.2, Pop.3 and Pop.4 that did not have migration movement.

Figure 12 Simulation of model (3) (365 iterations) showing plots of the population 1 to population 4 for grid of 50x50, 60x60, 70x70 and 90x90 indicating that migration of individuals infected with HIV does not depend on the area size of the population. The following colours appear as red (HIV), light green (AIDS), orange (re-infection).

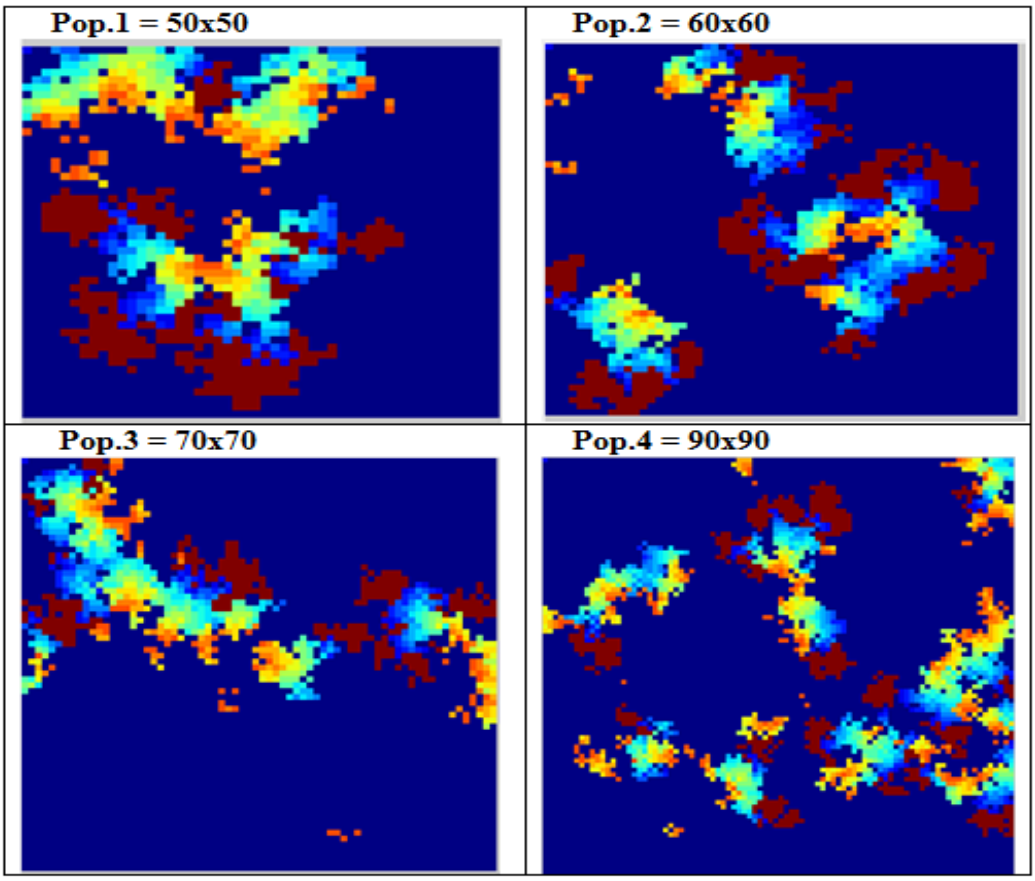

Figure 12: A dynamic spread process with migration

It can be seen that the 20 introduced HIV individual populations in each of the four populations with migration movement continue to exist as compared to Figure 11 for Pop.1.

The model indicates that there is increasing volatility in P_4 after 200 days due to incoming but no outgoing migrants (Figure 13).

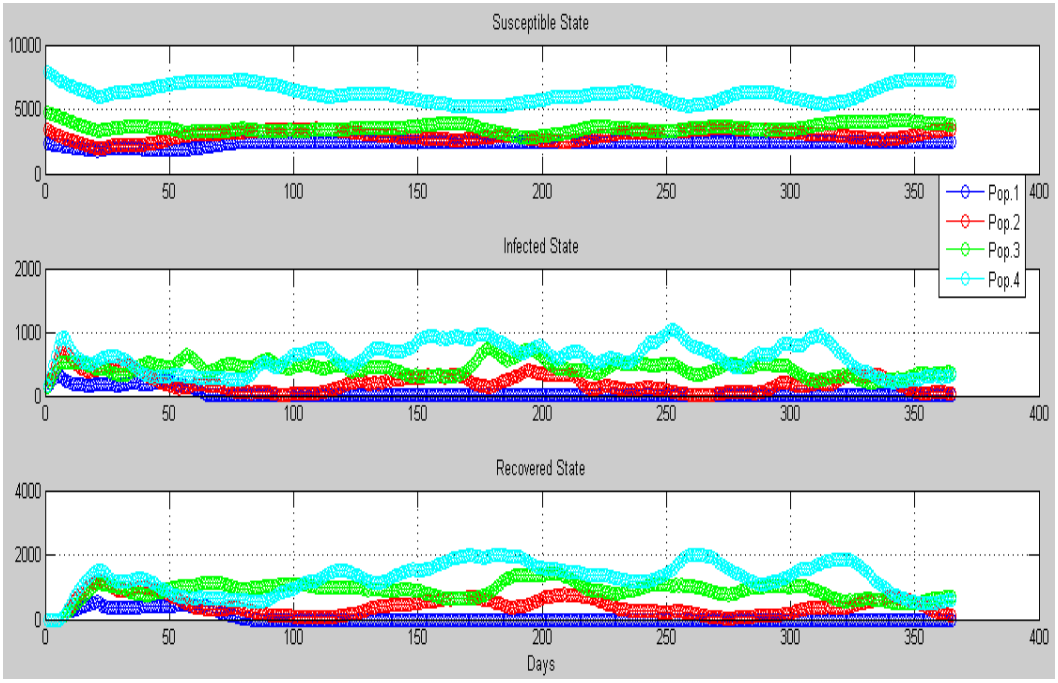

Figure 13: Size of the four populations (y-axis) without migration over time (x-axis)

Disease in P_2 disappeared at 90 days but reappeared after 320 days due to non-migration. As seen in Pop.1, all the 20 HIV infected individuals introduced die as indicated in the infected state and recovery state respectively.

The model indicates that there is increasing volatility in P_4 after 200 days due to incoming but no outgoing migrants (Figure 14).

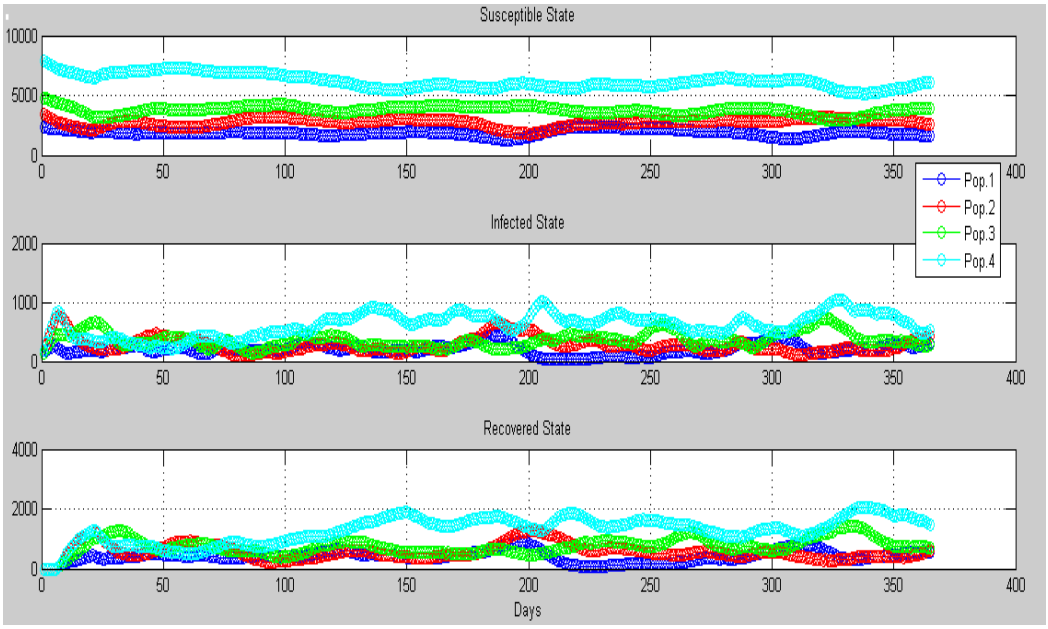

Figure 14: Size of the four populations (y-axis) with migration over time (x-axis)

Disease in P_1 disappeared at 201 days but reappeared after 250 days due to migration. It was noted that, in many cases, disease spread was at the leading edge of movements within a population (Figures 11, 12), resulting in the disease moving backward and forwards in waves as infection spread within the confines of populations and infected migrants entered the population from another population.

4.4 **Summary of Simulations**

These simulations have demonstrated that it is possible to create four distinct populations with a predefined migration schema that allows for the modelling of the $(SIR)^n$ epidemic model. The spread of the disease does not depend on the size of population but on the migrant direction of the infected individual in Figure 12. Interestingly, the number of infected converged on the 365^{th} iteration for the first three populations (Figure10). The design of CA and the model analysed here do not incorporate the effect of delay or a latent period, birth and mortality in Figure 9. The migration is significant in disease modelling as seen with spread of infected HIV individuals persisting in Figure 12 seen in (Deering, Vickerman et al. 2008) as compared to Figure 11 (without migration movement). The results obtained will be used as a benchmark to compare the impact of introducing exposed and latencies parameters in the next chapter.

Chapter 5

Extended SIR Model to SEIR Model

The incubation time for AIDS is anything from a few months to years after the patient has been shown to have antibodies to the HIV. We can, for example, incorporate this in mathematical model as a delay effect, or by introducing a new class, $R(t)$ say (removal stage), in which members of the susceptible population remain for a given length of time before moving into the infective class. Such models give rise to differential equation formulations and they can exhibit oscillatory behaviour as shown in Figure 15, as might be expected from the inclusion of delays.

A delay in many population models considers that the transmission dynamic behaviour of the disease at time *t* can destabilise the equilibrium. There are two types of such delay: a discrete or fixed delay, and a continuous or distributed delay. In models with discrete or fixed delay, the dynamic behaviour of the disease depends on the state at time $t-\tau$, where τ is a fixed constant (Kaddar, Abta et al. 2011). For instance, the number of new-born at time t depends on the state of population at time $t-\tau$, where τ the period of pregnancy. In the case of continuous or distributed delay, the dynamic behaviour of the model at time t depends on the states during the entire period prior to time t.

We let individuals i in patch1 j in patch2 at time t be represented by S_{ij} for the number of susceptibles in the population and E_{ij} for susceptible individuals becoming exposed, that is, infected but not yet infective. The individual remains in the exposed class for a certain latent period before becoming infective (Özalp and

Demirci 2011). Let I_{ij} represent the number infected and R_{ij} the number recovered. $\beta_{ij} \geq 0$ is the effective contact-rate per individual per unit of time, which introduces the Law of Mass Action assumption: the rate of β_{ij} from susceptible to exposed the natural death rate is assumed to be proportional to the population number in each state, with the rate constant μ. The assumption is that τ (delay time in days) is constant. By using similar work by Yan and Liu (2005), the probability that an individual survives the latent period to infected at $[t-\tau, t]$ is $e^{-\mu\tau}$. Since, the number of susceptible individuals that become exposed at time $t-\tau$ is $\beta_{ij} S_{ij}(t-\tau) I_{ij}(t-\tau)/N_{ij}(t-\tau)$, this leads to a deterministic model of $\beta_{ij} \frac{S_{ij}(t-\tau) I_{ij}(t-\tau)}{N_{ij}(t-\tau)} e^{-\mu\tau}$.

5.1 **Rules for Disease Spread**

The rules described below determine the state transitions of individual cells in the CA for the SEIR model which will incorporate other probabilistic parameters as shown in Table 4 (Apenteng, Narayanan et al. 2012).

- A cell changes its state from *susceptible* to *exposed* ($S \rightarrow E$) when it comes in contact with an infected cell in its defined neighbourhood.
- A cell changes its state from *exposed* to *infected* ($E \rightarrow I$) after being in the E state for a given ε which is the transition time.
- The state of the cell changes from *infected* to *recovery* ($I \rightarrow R$) after being in state I for a given k. In state I, the cells are capable of passing on the infection to neighbouring cells.
- The cell remains in state R.

Let $C_{(i,j)}$ denote the state of each cell. We defined (10) as the probability that an infective individual removed. Let p_r and p_d be the probability that infected individual become removal state and disease-induced mortality cell respectively. If $C_{(i,j)}(t)=0$, then $C_{(i,j)}(t+1)=1$ with the probability P, i.e. if $rand < P$, then the state of the individual in the $C_{(i,j)}$ change from susceptible to exposed at $(t+1)$ time, otherwise it remains in the susceptible, .i.e. $C_{(i,j)}(t+1)=0$, where $rand$ refers to random number between 0 and 1.

The following outline the detail of the CA model how works including the cell state, rules and parameters used in the CA for the exposed, infection and removal respectively.

1. If $C_{(i,j)}(t) \in I$, then $C_{(i,j)}(t+1)=-1$ (death cell with probability p_d), or $C_{(i,j)}(t) \in R$ with probability p_r;
2. If $C_{(i,j)}(t) \in R$, then $C_{(i,j)}(t+1)=0$ with probability (10);
3. If $C_{(i,j)}(t) \in E$, then $C_{(i,j)}(t+1)=C_{(i,j)}(t)+1$, and after t time steps the cell becomes infected.

Figure 15 shows the summary of the rules for the disease spread.

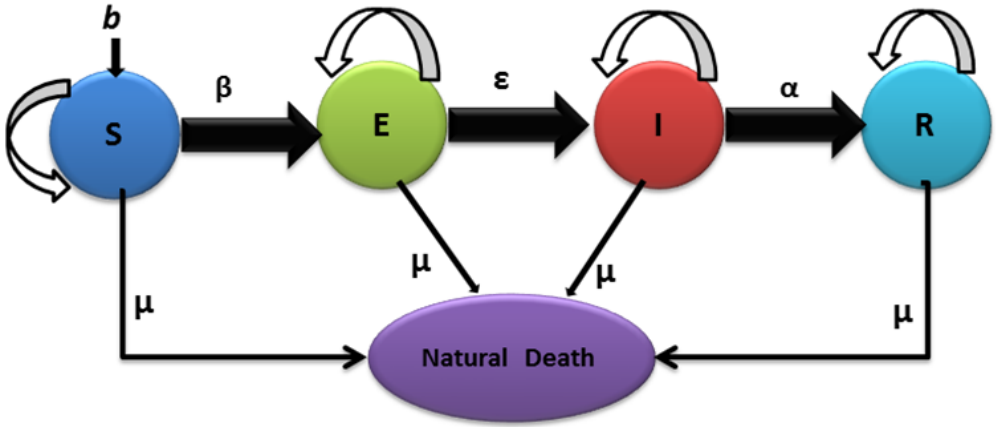

Figure 15: SEIR Model

Based on the aforementioned assumptions the CA rules, the ideas behind the classic SEIR models are based on differential equations. The set of ordinary differential equations corresponding to the CA model is

$$S_{ij}^{'}(t) = b - \beta_{ij}\frac{S_{ij}(t)I_{ij}(t)}{N_{ij}(t)} + \delta A_{ij}(t) \tag{14}$$

$$E_{ij}^{'}(t) = \beta_{ij}\frac{S_{ij}(t)I_{ij}(t)}{N_{ij}(t)} - \beta_{ij}k\frac{S_{ij}(t-\tau)I_{ij}(t-\tau)}{N_{ij}(t-\tau)} - (\mu+\varepsilon)E_{ij}(t) \tag{15}$$

$$I_{ij}^{'}(t) = \beta_{ij}k\frac{S_{ij}(t-\tau)I_{ij}(t-\tau)}{N_{ij}(t-\tau)} - (\mu+\alpha)I_{ij}(t) + \varepsilon E_{ij}(t) \tag{16}$$

$$A_{ij}^{'}(t) = \alpha I_{ij}(t) - (\mu+\delta)A_{ij}(t) \tag{17}$$

where $k = e^{-\mu\tau}$

This models a population of a constant size $N_{ij}(t) = S_{ij}(t) + E_{ij}(t) + I_{ij}(t) + A_{ij}(t)$.

Table 4: The meaning of the parameters

Parameter	Meaning
β	The parameter controlling how often a susceptible-infected contact results in a new exposure
ε	The rate at which an exposed person becomes infective
μ	The natural mortality rate (this is unrelated to disease).
α	The rate an infected (HIV) that moves into the AIDS.
τ	Days in delay
δ	The death rate related to AIDS
S	The number of susceptible individuals at the beginning of the model run
E	The number of exposed individuals at the beginning of the model run
I	The number of infected (HIV) individuals at the beginning of the model run
A	The number of AIDS individuals at the beginning of the model run
Days	Controls how long the model will run

5.2 **Stability Analysis of SEIR Model**

An important equilibrium point for any stochastic or deterministic disease model is the disease free equilibrium (DFE). The stability of the DFE is especially important since it determines whether or not a virus is capable of attacking a population. The

reproduction number R_0 is a threshold value or number which determines the stability of the DFE. The reproduction number is the expected number of secondary cases produced by typical infection in completely susceptible populations (van den Driessche and Watmough 2005). If $R_0 > 1$, an epidemic occurs and if $R_0 < 1$, an epidemic does not occur. R_0 can be calculated as $R_0 = \rho(FV^{-1})$ (where F and V are found below). The spectral radius is (ρ) and is called next generation operator and represents the rate at which individuals in compartment j generate new infections in compartment i.

The (i, j) is the number of entries of matrix FV^{-1} which represents the number of new infections in compartment i due to an infected individual being introduced into compartment j (van den Driessche and Watmough 2005; Arino, Jordan et al. 2007).

Let $F_i(x)$ be the rate (transmitted rate) at which new infected enter compartment i.

Let $V_i = V_i^- - V_i^+$ denote the transfer of individuals out of (V_i^-) and into (V_i^+) i^{th} compartment. Then:

$$F_{ij} = \left.\frac{\partial F_i}{\partial x_j}\right|_{x0} \qquad V_{ij} = \left.\frac{\partial V_i}{\partial x_j}\right|_{x0} \quad \text{For } i, j = 1,2,...,m$$

Note: x_0 is the disease free equilibrium (DFE).

The reproduction number R_0 is given by dominant eigenvalues (or spectral radius) of FV^{-1}.

5.3 **Experiment and results**

We implemented in MATLAB the cellular automata model based on the rules described in Section 5.1. Experiments are performed on populations occupying a

(100x100) cell space. Twenty infected individuals were introduced into the four different populations. Tables 4 and Figure 12 summarise the parameter values used, whereas Table 5 summarises the simulation protocol. Blue, light green, red and light blue squares correspond to 0, 0.5, 1 and 0.1 with susceptible individuals, exposed individuals, infected individuals and recovery individuals, respectively. 365 iterations were simulated with one iteration representing a day.

Table 5: Simulation Protocol

Events	Model Prediction	CA Model
Simulator	matlab	matlab
Susceptible	80 people	80 people
Exposed	nil	From model prediction
Infected	20 people	20 people
Removal	nil	From model prediction
Natural death	$0.0005yr^{-1}$	$0.0005yr^{-1}$
Incubation period	$7\ yr^{-1}$	$7\ yr^{-1}$
No. Of time step	365 iterations	365 iterations
Infection rate	0.25	0.25

Figure16 depicts how the propagation of disease spread clusters in geographical distribution with 40 simulations.

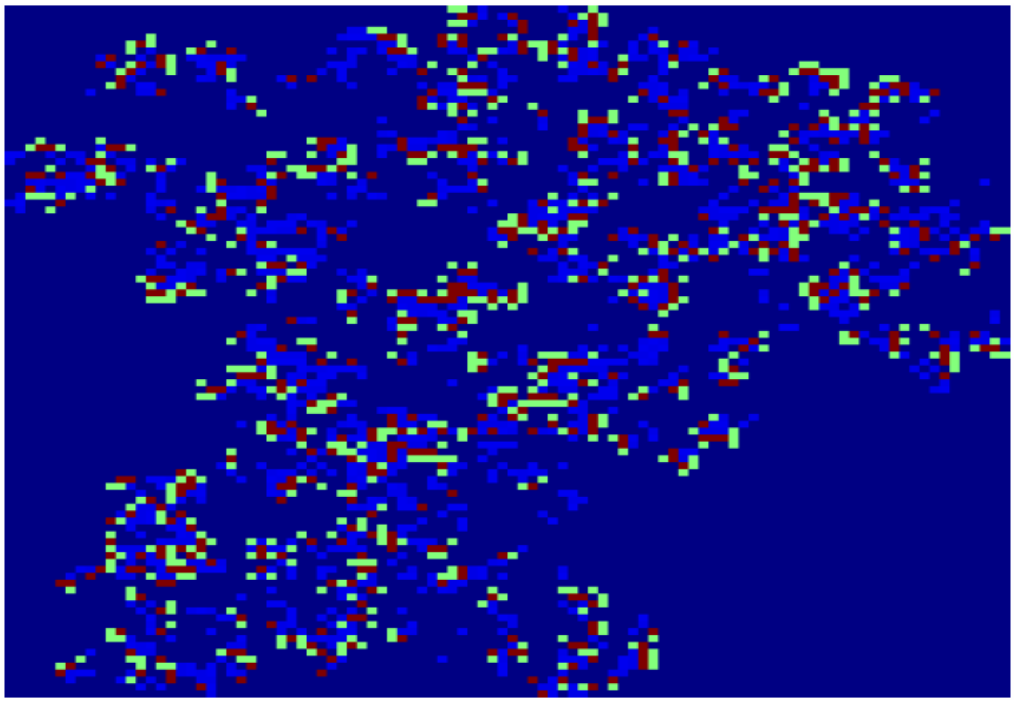

Figure 16: A dynamic spread process

The colours: dark blue, green, red and light blue represent healthy individuals (susceptible), infected individuals (exposed), infectious individuals HIV and AIDS patients (recovery replaced by AIDS), respectively. We can see the breakdown of an initially homogenous spread-of-disease pattern. As the phase separation takes place, a persistent compact spread of disease is formed.

Figure 17 depicts how the propagation of disease spread clusters in geographical distribution with 65 simulations.

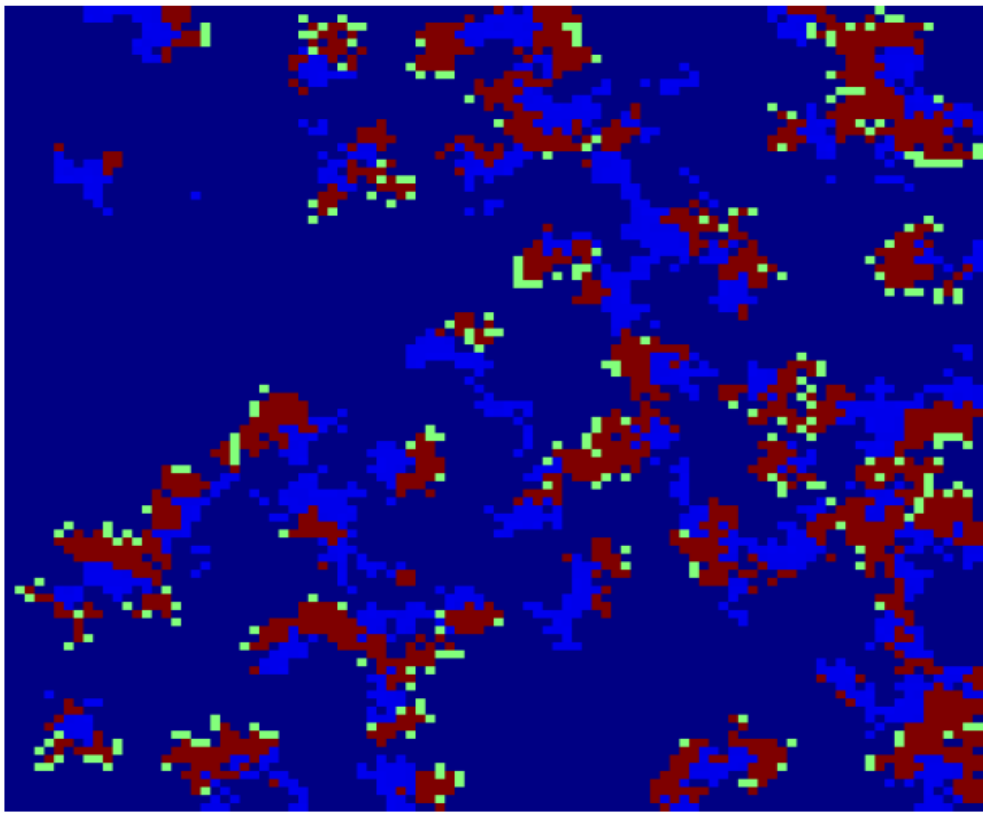

Figure 17: A dynamic spread process

The colours: dark blue, green, red and light blue represent healthy individuals (susceptible), infected individuals (exposed), infectious individuals HIV and AIDS patients (recovery replaced by AIDS), respectively. As the phase separation takes place, a persistent compact spread of disease is formed (HIV), surrounded by free exposed individual populations.

Figure 18 depicts how the propagation of disease spread clusters in geographical distribution with 365 simulations.

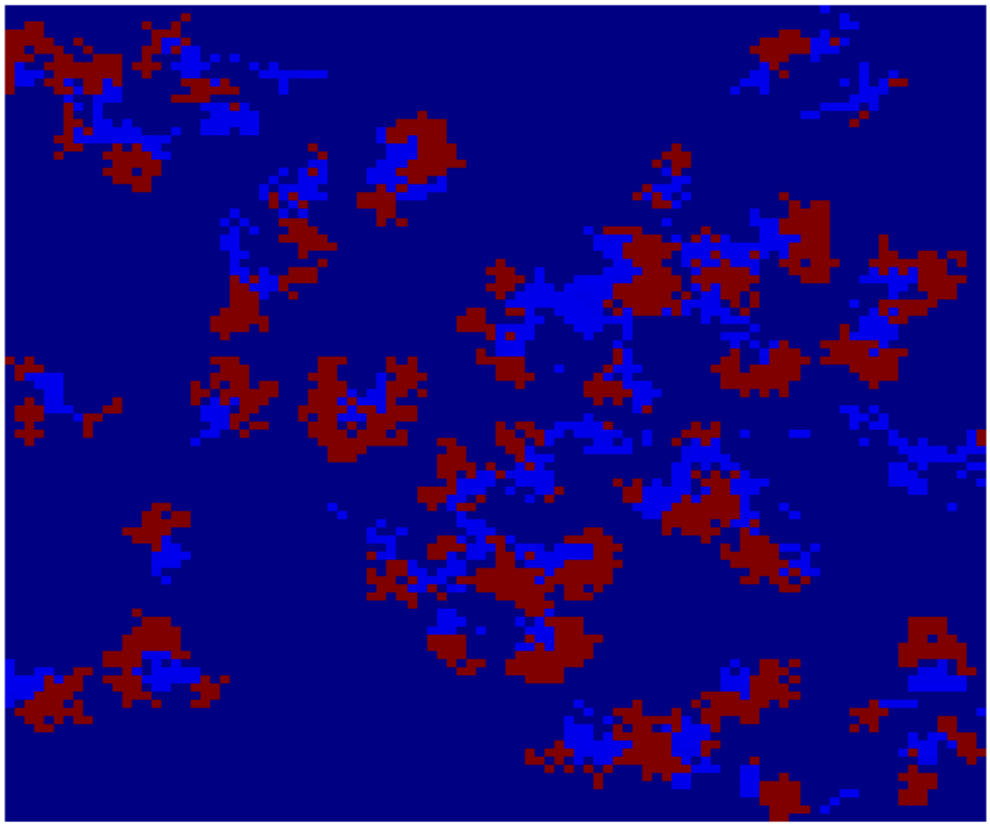

Figure 18: A dynamic spread process

The colours: dark blue, red and light blue represent healthy individuals (susceptible), infectious individuals (HIV) and AIDS patients (recovery replaced by AIDS), respectively. As the phase separation takes place, a persistent compact spread of disease is formed (HIV and AIDS) within the population.

The model indicates that there is increasing volatility in the susceptible population after 240 days due to outgoing moves from the susceptible to the exposed (Figure 19A).

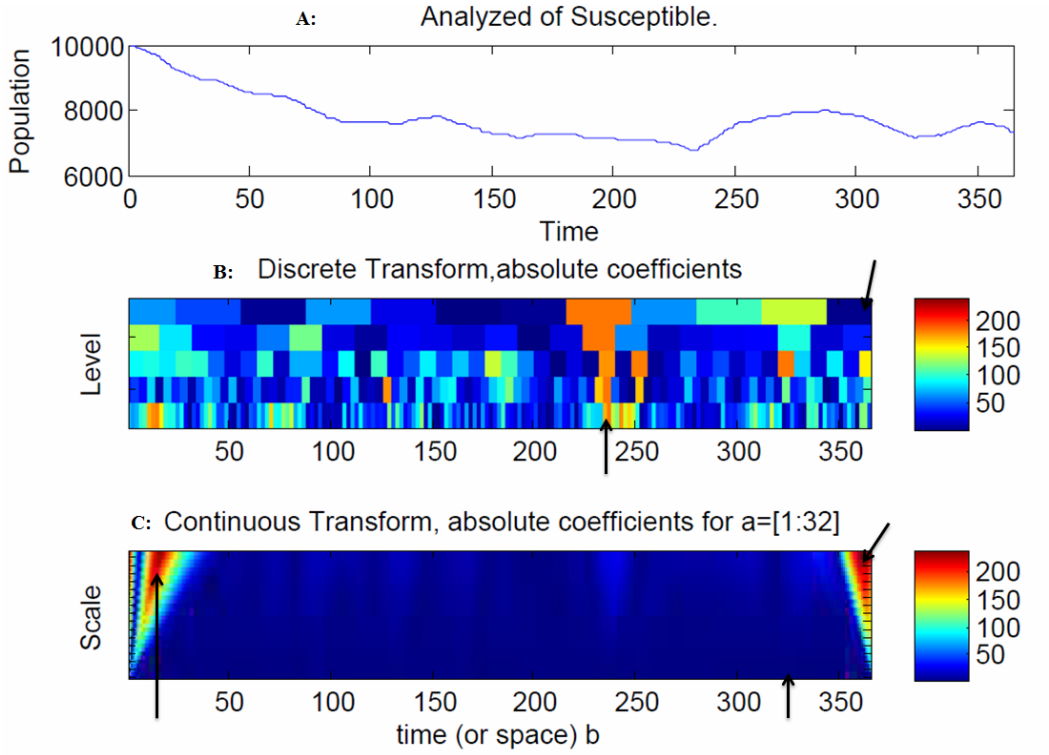

Figure 19: Snapshot of Susceptible

The spread of the disease in discrete transform were at a peak at 230 and 360 days (Figure 19B) as indicated by the arrows. There were slower and faster frequency at the initial stages as well as the final stages at 10 and 360 days indicated by the arrows. The blue regions denote the probability densities. For example, at 330 days there was an indication of slower frequency which resulted in faster spread of the disease (Figure 19C) due to large coefficients in continuous wavelet transform.

The model indicates that there is increasing volatility in the exposed after 200 days due to outgoing moves from exposed to infectious (Figure 20A).

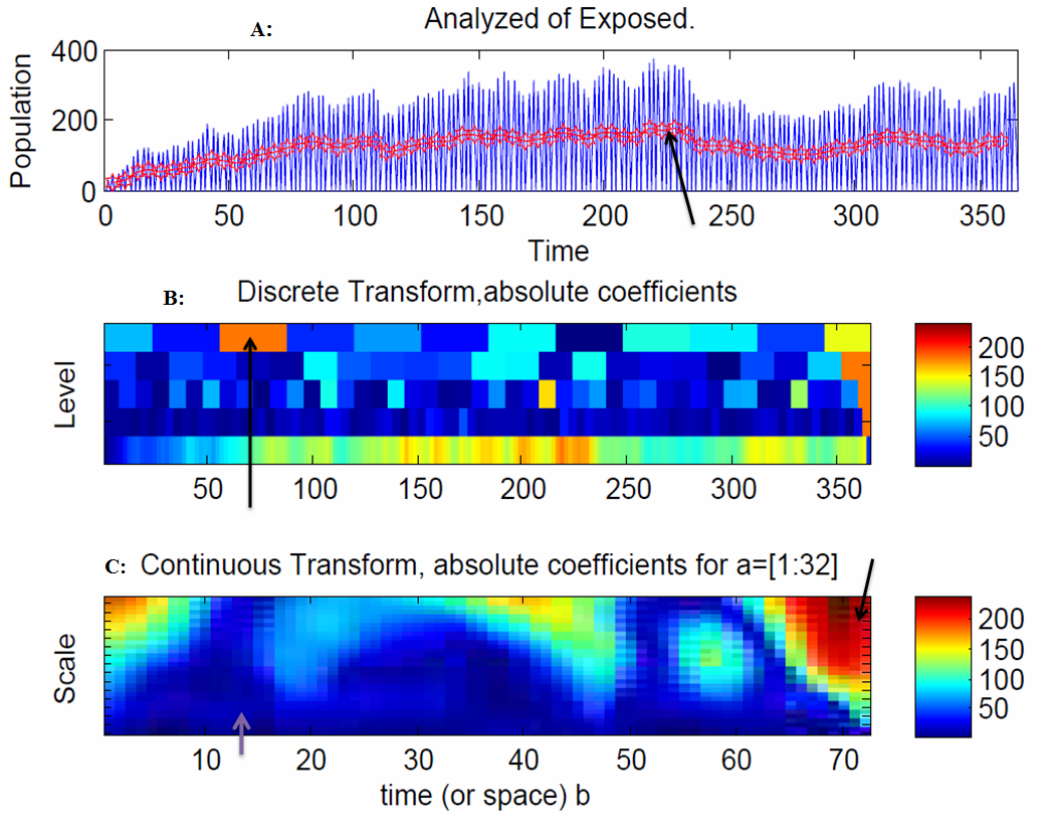

Figure 20: Snapshot of Exposed

The spread of the disease in discrete transform were at a peak at 70 days due to large coefficients (Figure 20B). In the continuous wavelet transform, at 72 days, there were no exposed individuals again (Figure 20C). They moved to the infectious stage.

The model indicates that there is increasing volatility in the infectious after 240 days due to outgoing moves from infectious to recovery (Figure 21A).

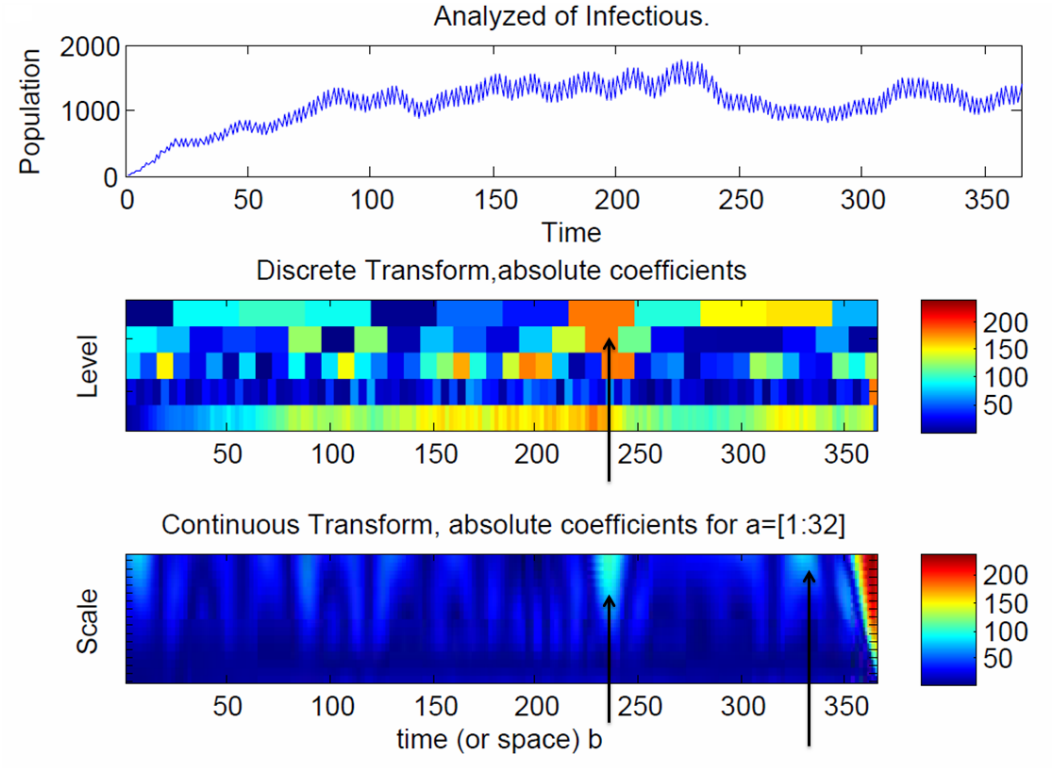

Figure 21: Snapshot of Infected

The spread of the disease in discrete transform were at a peak at 240 days (Figure B), and 240 and 280 days due to large coefficients (Figure C). The range of frequencies used in averaging is indicated by the arrow at 240 day which the peak of the disease spread (Figure 21C).

The model indicates that there is increasing volatility in recovery after 240 days due to outgoing moves from recovery to natural death (Figure 22A).

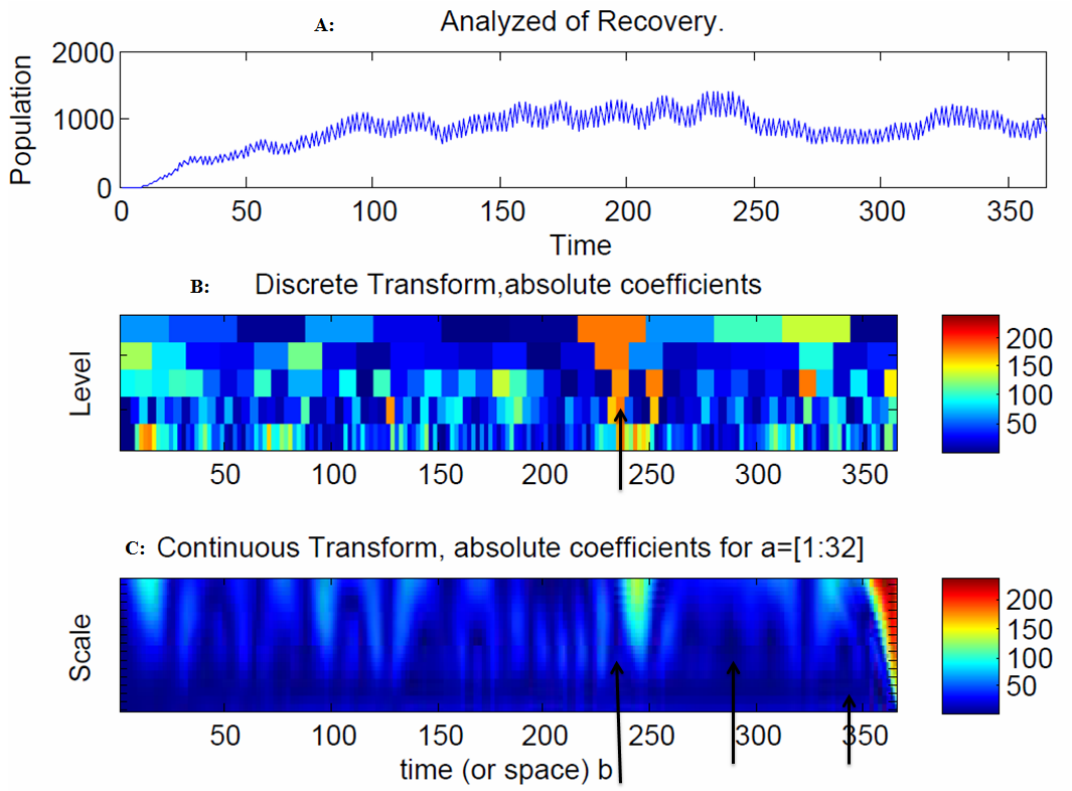

Figure 22: Snapshot of Recovery

The model indicates that there is increasing volatility in recovery after 240 days due to outgoing moves from recovery to natural death (Figure 22A). The spread of the disease in discrete transform were at a peak at 240 days (Figure 22B), and 230, 250 and 340 days due to large coefficients (Figure 22C). The range of frequencies used in averaging is indicated by the arrow at 250 and 362 days which are the peak of the disease spread (Figure 21C).

The model indicates that there is increasing volatility in natural death after 50 days due to outgoing moves from infectious and recovery to natural death (Figure 23A).

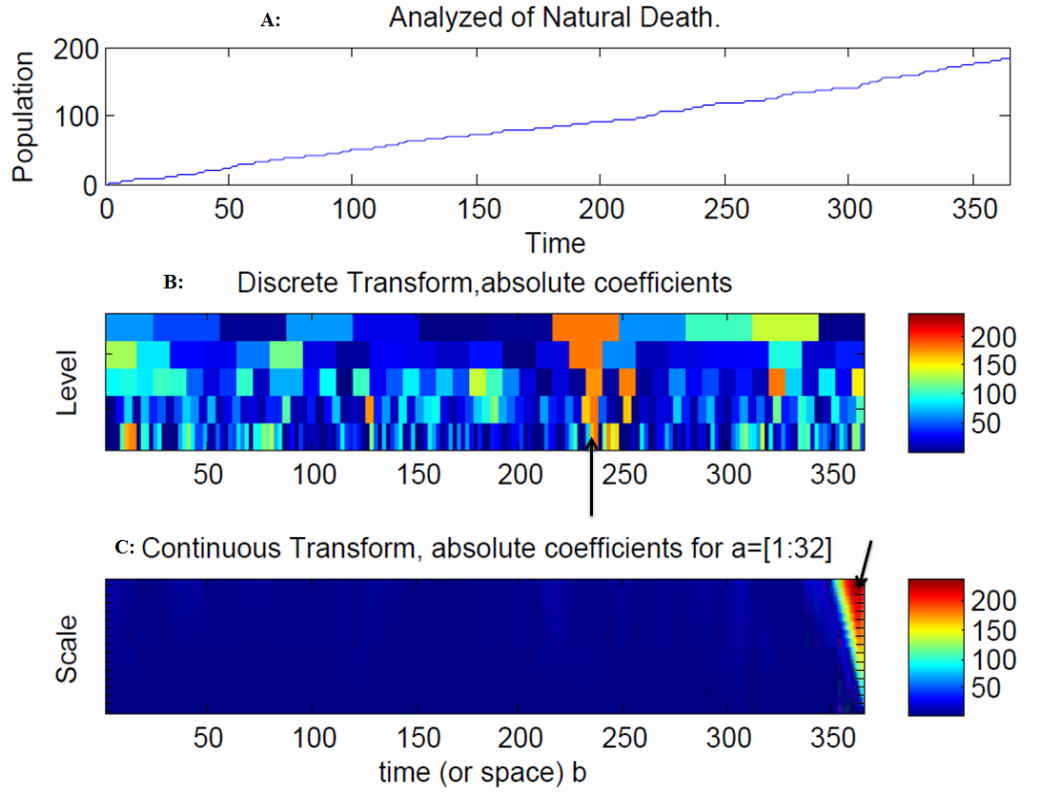

Figure 23: Snapshot of Natural Death

The spread of the disease in discrete transform were at a peak at 240 day (Figure 22B). The range of frequencies used in averaging is indicated by the arrow at 362 days which are the peak of the disease spread (Figure 21C).

As shown in Figure 24, the CWT coefficients are large in scales (which represent the absolute coefficients).

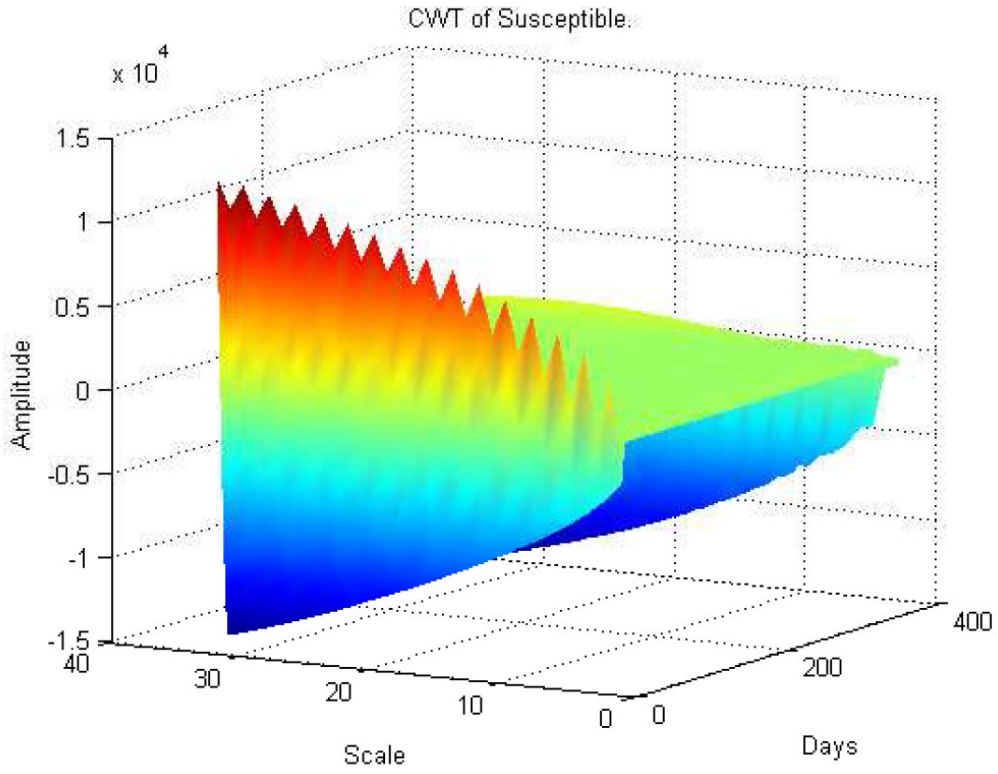

Figure 24: Sinusoidal of Susceptible

The frequencies of the susceptible waves depict the sinusoidal pattern in the continuous wavelet transform (CWT) coefficients at these scales. The amplitude is the power of the spread of the disease. The sinusoidal wave amplitude is the height of the crest and frequency is the number of oscillations per second. Hence, amplitude remains same for any change in frequency. The maximum and minimum is 1.5×10^4 Hz which demonstrates that the spread of the disease in the susceptible was high.

Figure 25 plots the same transform from a different angle for better visualization.

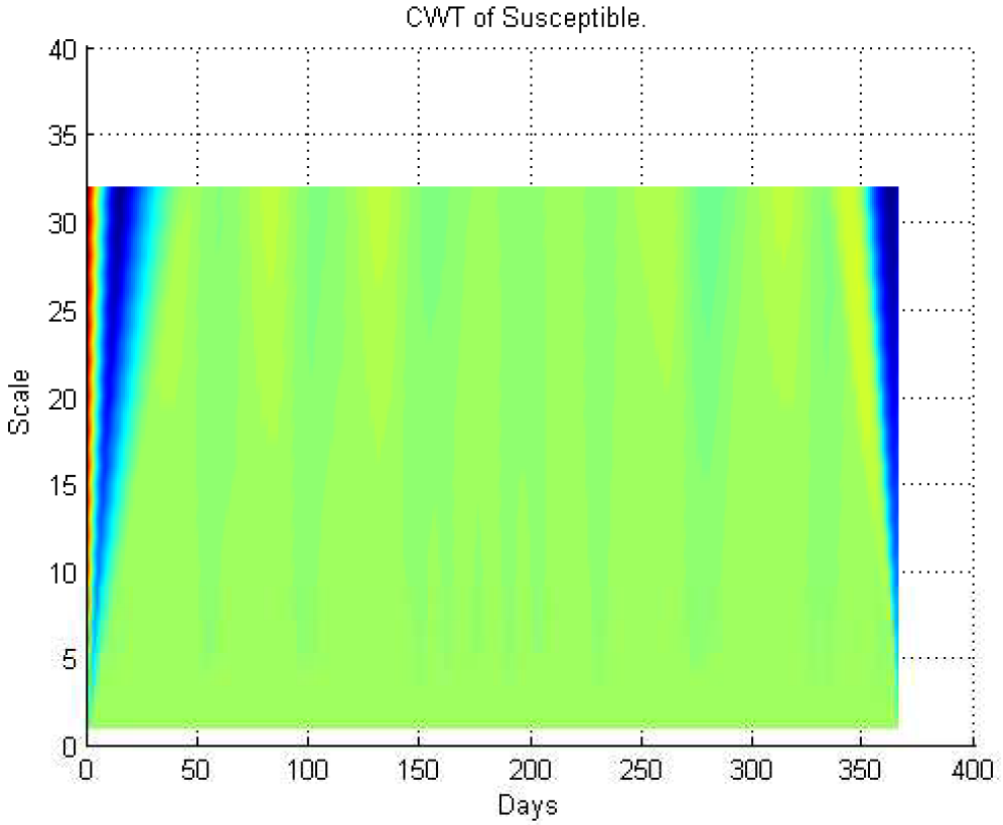

Figure 25: Sinusoidal of Susceptible

Figure 25 is a typical time-space diagram of the CA-model (14-17). The closed-loop lattices contain 100x100 cells, with a visible period of 365 time steps. The strong interaction between the (*x, y*) coordinates is 0.026Hz. As seen in Figure 19C, the arrow at 330 days indicated slower frequency which resulted in faster spread of the disease.

As shown in Figure 26, the continuous wavelet transform (CWT) coefficients are large at scales near the frequencies of the exposed waves and clearly depict the sinusoidal pattern in the CWT coefficients at these scales for the exposed.

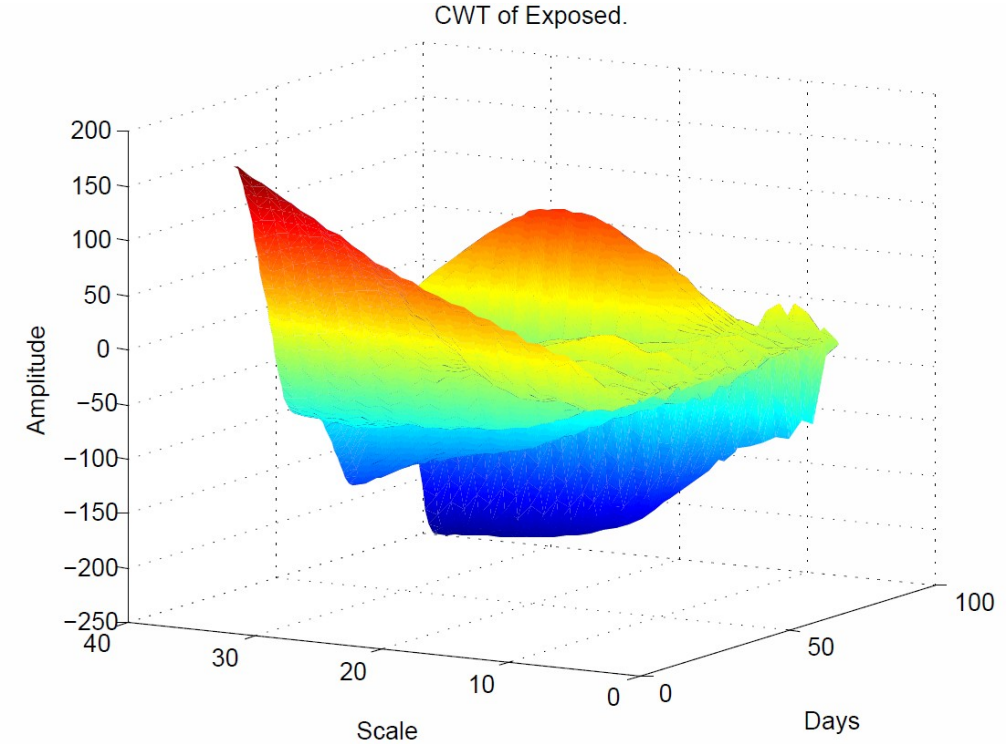

Figure 26: Sinusoidal of Exposed

The sinusoidal wave amplitude is the height of the crest and frequency is the number of oscillations per second. Hence, amplitude remains same for any change in frequency. The maximum and minimum is 225 Hz which demonstrates that the spread of the disease in the exposed was at the peak state.

Figure 27 plots the same transform from a different angle for better visualization.

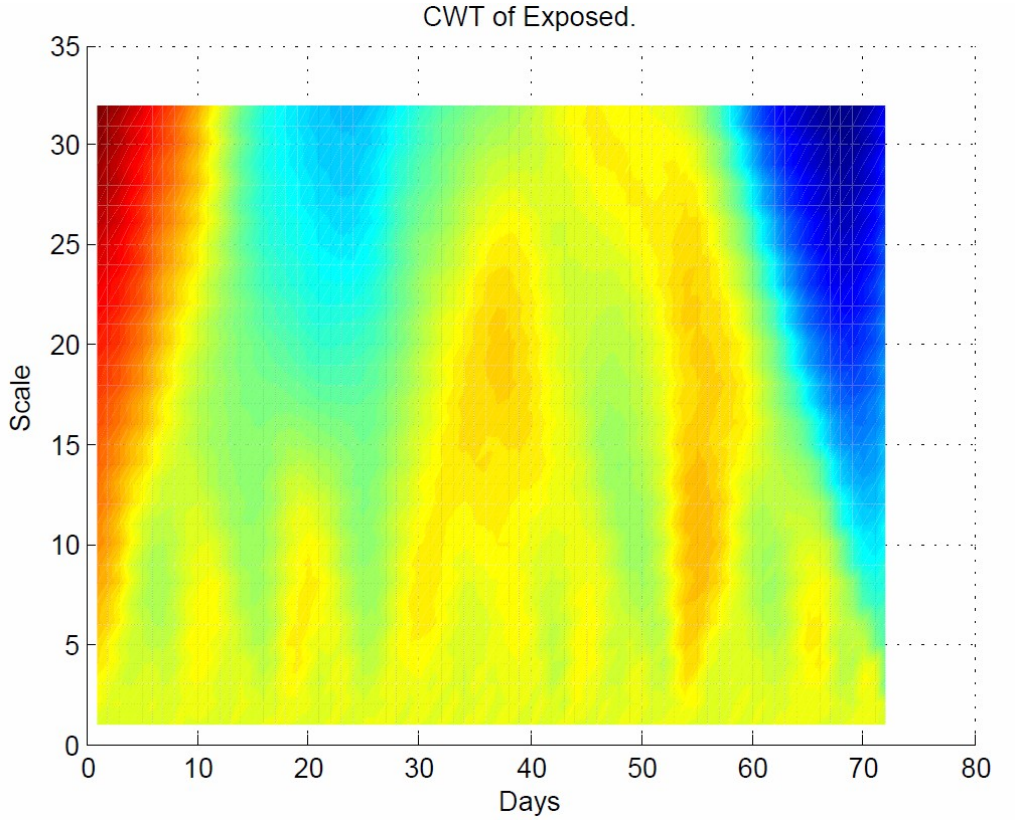

Figure 27: Sinusoidal of Exposed

In this plot, the value of each (x, y) coordinate represents the strength of spread of the disease between coordinates. The strong interaction between the (*x, y*) coordinates is 0.026Hz. As seen in Figure 20C, the arrow at 70 days indicated faster frequency which resulted in slower spread of the disease.

From Figure 28, the continuous wavelet transform (CWT) coefficients are large at scales near the frequencies of the infectious waves and clearly depict the sinusoidal pattern in the CWT coefficients at these scales for the infectious.

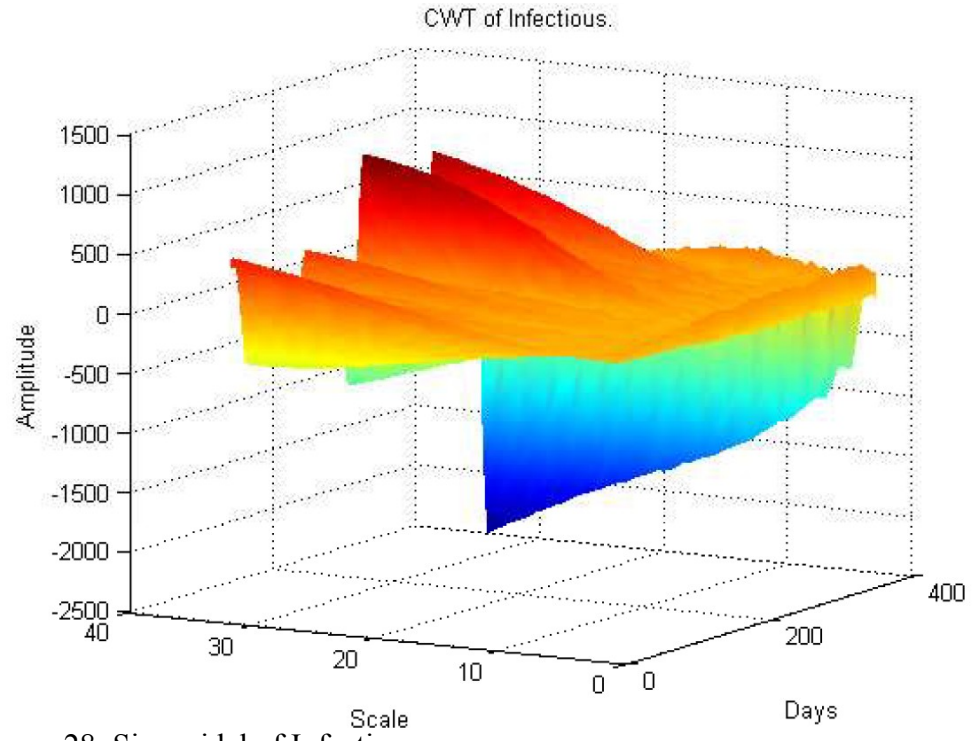

Figure 28: Sinusoidal of Infectious

The greater the amplitude of a wave, the more energy it carries. The sinusoidal wave amplitude is the height of the crest and frequency is the number of oscillations per second. Hence, amplitude remains same for any change in frequency. The maximum and minimum is 2000Hz which demonstrates that the spread of the disease in the infectious was at peak state.

Figure 29 plots the same transform from a different angle for better visualization.

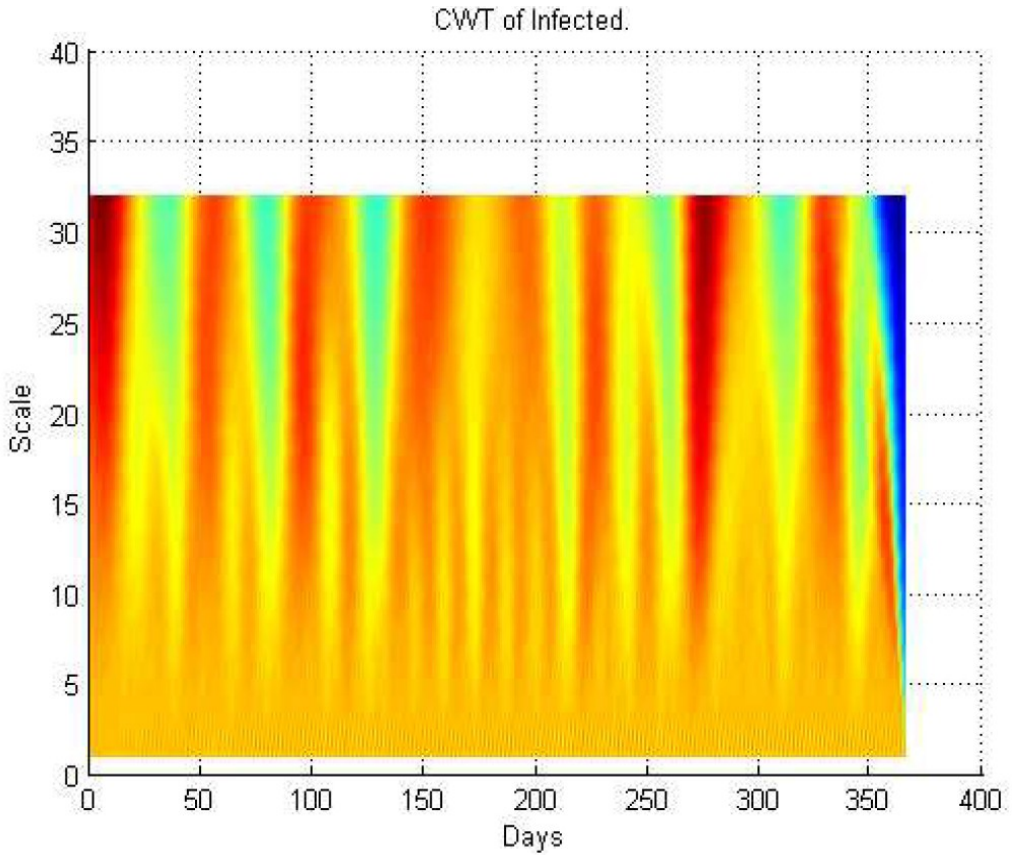

Figure 29: Sinusoidal of Infected

Figure 29 is a typical time-space diagram of the CA-model (14-17). The closed-loop lattices contain 100x100 cells, with a visible period of 365 time steps. The strong interaction between the (*x, y*) coordinates is 0.0056Hz. As seen in Figure 21C, the arrows at 240 and 320 days indicated slower frequency which resulted in faster spread of the disease.

In Figure 30, the CWT coefficient are large at scales near the frequencies of the recovery waves and also shows the sinusoidal pattern in the CWT coefficients at these scales for recovery.

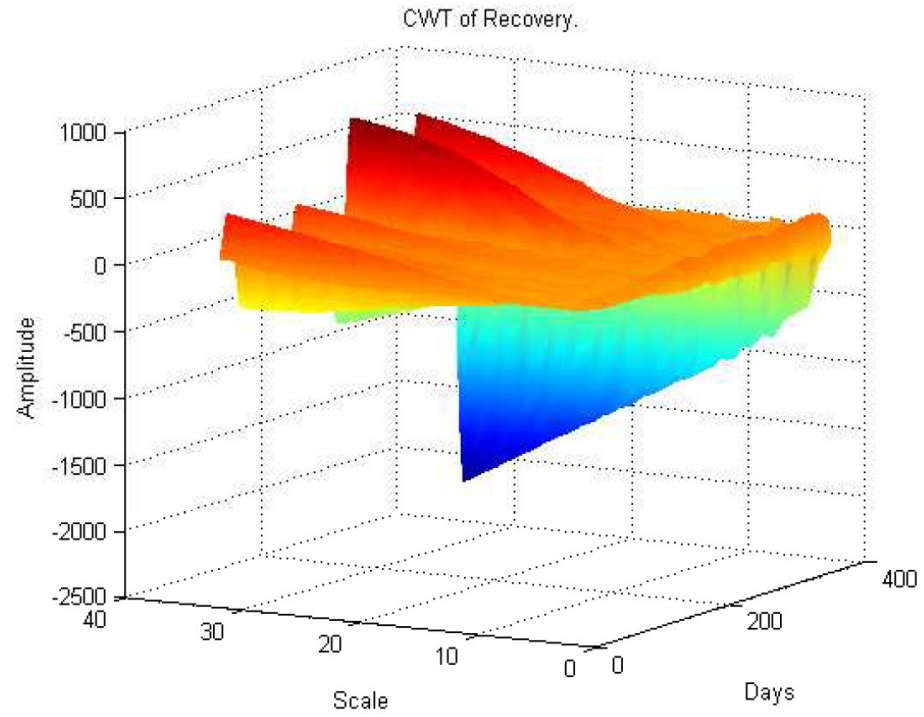

Figure 30: Sinusoidal of Recovery

The greater the amplitude of a wave, the more energy it carries. The sinusoidal wave amplitude is the height of the crest and frequency is the number of oscillations per second. Hence, amplitude remains the same for any change in frequency. The maximum and minimum is 1750Hz which demonstrates that the spread of the disease in recovery was at peak state.

Figure 31 plots the same transform from a different angle for better visualization.

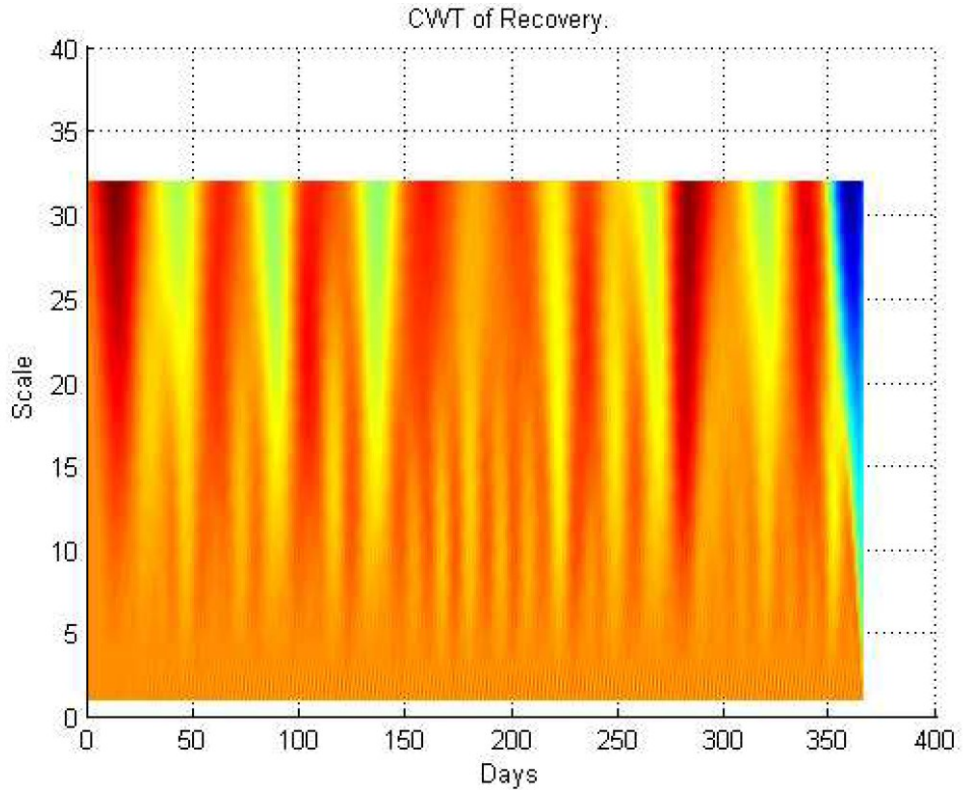

Figure 31: Sinusoidal of Recovery

Figure 31 is a typical time-space diagram of the CA-model (14-17). The closed-loop lattices contain 100x100 cells, with a visible period of 365 time steps. The strong interaction between the (*x, y*) coordinates is 0.006Hz. As seen in Figure 22C, the arrows at 240, 280 and 320 days indicated slower frequency which resulted in faster spread of the disease.

The CWT coefficients are large at scales near the frequencies of the death waves and clearly depict the sinusoidal pattern in the CWT coefficients at these scales for death.

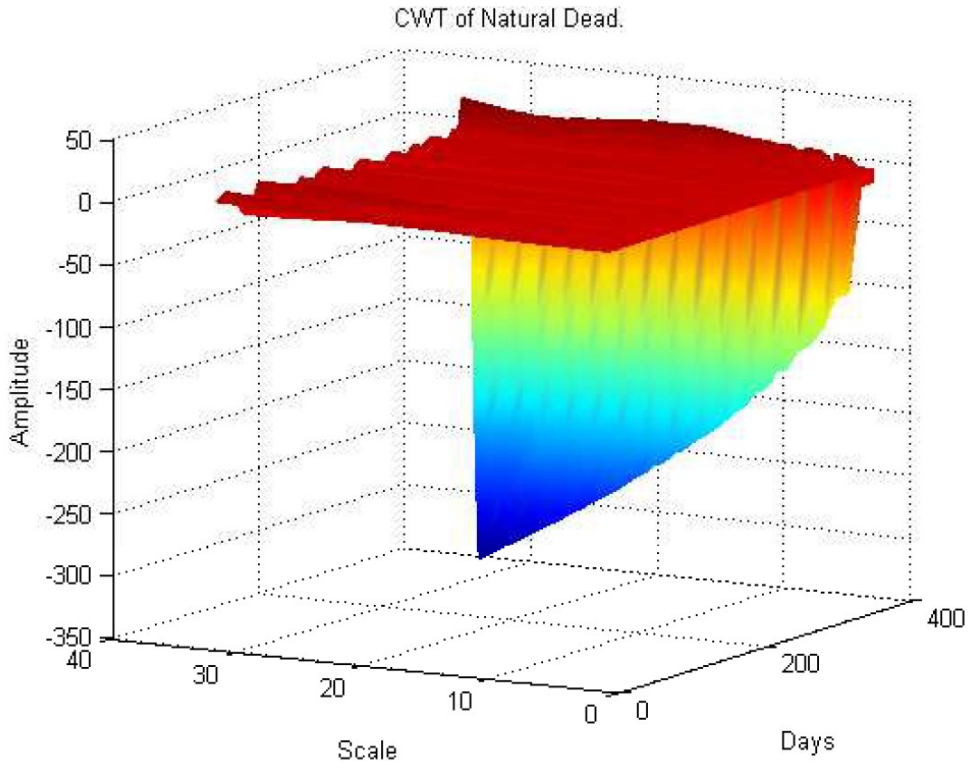

Figure 32: Sinusoidal of Natural Death

The sinusoidal wave amplitude is the height of the crest and frequency is the number of oscillations per second. Hence, amplitude remains same for any change in frequency. The maximum and minimum is 200Hz which demonstrates that the spread of the disease in natural death was at peak state.

Figure 33 plots the same transform from a different angle for better visualization.

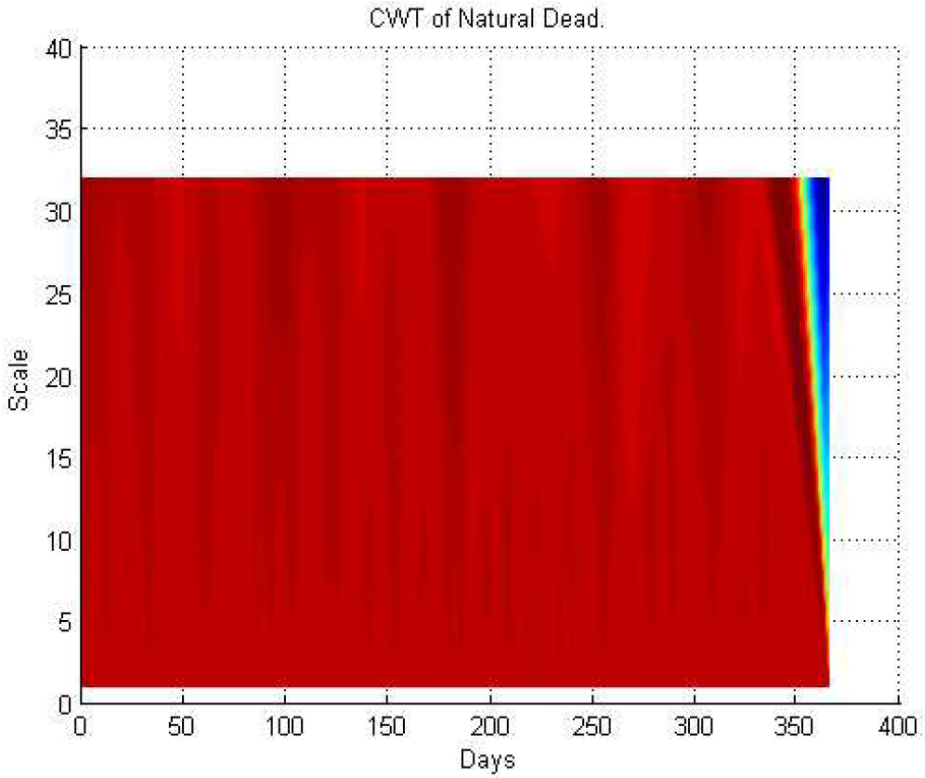

Figure 33: Sinusoidal of Natural death

Figure 33 is a typical time-space diagram of the CA-model (14-17). The closed-loop lattices contain 100x100 cells, with a visible period of 365 time steps. The strong interaction between the (*x, y*) coordinates is 0.0027Hz. As seen in Figure 23C, the arrow at 362 days indicated faster frequency which resulted in slower spread of the disease.

5.4 **Summary of Simulations**

In summary, we extended the SIR model to an SEIR model into four compartments. The disease stopped spreading in the exposed state at 72 days but continue to spreading in susceptible, infectious and recovery after 240 days. The spread of AIDS continue to persist in Figures 19, 21, 22 and 23. This has demonstrated that the slower the frequency the higher the spread of the disease in (Figure 25, 29, 31 and 33), as stated in section 3.3. The model was used to generate artificial data to describe the spread of AIDS in terms of different populations over time.

Chapter 6
Discussion

This thesis gave a review of both CA and mathematical models of AIDS. And overview of CA and mathematical models, their background and simulation setup are given. These techniques are applied to this work to model the spread of AIDS.

During the experiments there was much evidence of different interactions between infected individuals and others. The simulations show the competition between the three different individual populations which differ in infection period (Figures 11 and 12). The simulation which shows the competition between the three different individual populations which differ in infection period correspond to Figure 16. HIV period 1 (red) resulted in higher outbreak frequency compared to a period of 0 (blue) for susceptible, 0.5 (green) for exposed and recovery (AIDS) 0.1 (light blue). For example, 166 were infected with HIV as compared to 67 exposed, 15 got AIDS in 10 days. In addition 373 got HIV, 127 were exposed and 210 got AIDS in 20 days. The interactions between infectious and susceptibles leads to the exposed stage of the disease. As shown in Figure 16, the individuals who were exposed to the disease were much more at the beginning of the spread of the disease, which agrees with study done by Deering, Vickerman et al. (2008) concerning the impact of out-migrants and out-migration on the HIV/AIDS epidemic in India. More individuals became exposed after 40 days. The simulation which shows the competition between the three different individual populations which differ in infection period correspond to Figure 17. HIV period 1 (red) resulted in higher outbreak frequency compared to a period of 0 (blue) for susceptible, 0.5 (green) for exposed and recovery (AIDS) 0.1 (light blue). For example, 700 were infected with HIV as compared to 172 exposed

and 663 got AIDS in 60 days. In addition 945 got HIV, 274 were exposed and 859 got AIDS in 83 days. The interactions were between the infected and the susceptible which leads to the exposed stage of the disease, for example, see recent work by Özalp and Demirci (2011). There were 7886 of susceptible individual who became exposed to the disease (Figure 17) in 83 days. The exposed individuals move to infectious stage (Figure 21).

The simulation which shows the competition between the four different individuals which differ in infection period correspond to Figure 18. HIV period 1 (red) was wining against a period of 0 (blue) for susceptible, 0.5 (green) for exposed and recovery (AIDS) 0.1 (light blue), for example 1180 were infected with HIV as compared to 701 AIDS in 140 days. The interactions between the infected and the susceptible lead to an exposed stage of the disease, which agrees with global dynamics behaviours for a new delay SEIR epidemic disease model studied by Meng and Chen et al. (2007). As shown in Figure 14, the individuals who were infectious to the disease were much more at the beginning of the spread of the disease. The simulation which shows the competition between the three different individuals which differ in infection period correspond to Figure 17. HIV period 1 (red) resulted in higher outbreak frequency compared to period of 0 (blue) for susceptible, 0.5 (green) for exposed and recovery (AIDS) 0.1 (light blue). For example, 1457 were infected with HIV as compared to 911 AIDS in 170 days. Also, 1850 got HIV and 1098 had AIDS in 202 days. The interactions were between the infected and the susceptible which leads to the exposed stage of the disease. There were more exposed individuals who became infectious to the disease (Figure 18). The exposed individuals move to the infectious stage (Figure 23 and 24) with 365

simulations. For example, 1221 were infected with HIV, 848 AIDS and 7459 were susceptible.

The $(SIR)^n$ model experiment has demonstrated that it is possible to create four distinct populations with a predefined migration schema that allows for the modelling of the $(SIR)^n$ epidemic model. Our simulations show that the spread of the disease does not depend on the size of population but on the migrant direction of the infected individual. Interestingly, the number of infected converged on the 365^{th} iteration for the first three populations (Figure 14). It was noted that, in many cases, disease spread was at the leading edge of movements within a population (Figures 13, 14), resulting in the disease moving backward and forwards in waves as infection spread within the confines of populations and infected migrants entered the population from another population, see, for example recent work by Aagaard-Hansen and Nombela et al. (2010).

Wavelets are very important for detecting abrupt changes in the distribution of disease spread (Bjørnstad 2005). These abrupt changes occur from one compartment to another (susceptible, exposed, infectious, and recovery). It produced relatively large wavelet coefficients centred around the discontinuity at all scales. The location of the discontinuity based on the CWT coefficients is obtained at the smallest scales. This selection for shorter infection period was also reported in (Apenteng, Narayanan et al. 2012). It should be noted that, due to the stochastic nature of the model parameters, the introduction of the dead period is only a transient phenomenon. An infection will only cause a new infection if, at the time, the individual is still exposed. Alimadad and Dabbaghian et al. (2011) also reported that

an individual become exposed to an infection when a randomly chosen neighbour individual is susceptible to it.

The smooth oscillations in Figures 19 to 23 produced relatively large wavelet coefficients at scales where the oscillation in the wavelet correlates fit best with the SEIR model, where wavelet analysis was taken into account. This is in contrast to Meng, Chen et al. (2007) who focused on delay functional and impulsive differential equation. The CWT detects both the abrupt transitions and oscillations in the susceptible, exposed, infectious and recovery states. The abrupt transitions affect the CWT coefficients a t all scales and clearly separate themselves from the smooth in the four compartments at small scales.

For example, in Figure 25 the precise location of the high frequency (discontinuity part) is explicit in the case of different time intervals; 10 and 365 days corresponding to frequencies 0.1Hz and 0.0027Hz respectively. The average of these frequencies is 0.051Hz. Whereas, in Figure 27, the precise location of the high frequency (discontinuity part) is explicit at the time interval of 38 days corresponding to a frequency of 0.026Hz, the simulation was stopped in Figure 27 at 72 days, indicating that those individuals who had been exposed have moved to the infectious states. In Figure 17, at step 65, the individuals who have been exposed were fewer than the infectious and recovery because they were all moving to the infectious states. See, for example, previous work by Coffee and Lurie et al. (2007).

On the other hand, in Figure 29, the precise location of the high frequency (discontinuity part) is explicit in the case of different time intervals 50, 100, 150, 200, 270, and 320 days corresponding to frequencies 0.020Hz, 0.010Hz, 0.007Hz,

0.005Hz, 0.004Hz and 0.003Hz respectively, these frequencies show the spread of the disease (Figure 30). The average of these frequencies is 0.008Hz. Figure 33 has the location of high frequency at 360 days which corresponds with 0.0027Hz due to the low disease spread (Figure 19). The dominant order of the frequencies shows that delay in the transition between exposed to infectious states takes place after the HIV and AIDS are at peak points in the different time intervals, for example see studies by Huang and Shen et al. (1998). The detection of these discontinuities (delays) was associated with the spread of disease and is detectable in the frequency and phase of the CWT. The results have also shown how the contact rates dispersed from one compartment to other in Figures 23-31, with the sinusoidal representing the delay in the exposed state. Figures 17, 19 and 20 show how the sinusoidals are affected by the population movement.

Chapter 7

Further Theoretical Work

In this chapter, we explore the implications of extending our model to deal with co-infection from a theoretical perspective.

It has recently been discovered that HIV patients also commonly acquire HCV (Hepatitis C) through shared routes of transmission (Askarieh, Alsiö et al. 2010). While it is known that there is a route from HIV to HCV, the rate of infection from HCV to HIV is unknown and is the subject of controversy (Falconer, Sandberg et al. 2009). Also, relationships between HIV and tuberculosis (TB) are also being reported (Sharma, Mohan et al. 2005). While co-infection models are relatively well known in the influenza literature (Colijn, Cohen et al. 2009) and other commonly co-occurring diseases (Keeling 1999), there is still uncertainty as to how such co-infection models can be applied to HIV and related infections. Finally, while there has been previous work on the use of probabilistic CA to model co-infection, such work does not take into account the population characteristics of the infected populations nor their possible migration routes. The aim of this final piece of research will be used to extend our knowledge of how CA can be used to model co-infection involving HIV and other associated diseases as well as to identify appropriate strategies for constructing CA models that best fit any real world data for further research work. These models may then be useful for extracting parameter values for exposure, susceptibility and infection that can be used in future preventative strategies.

As seen earlier, there have been many mathematical models of single infection, such as HIV/AIDS infection, based on susceptibility (*S*), infection (*I*) and recovery (*R*). However, there has been much less research on co-infection (simultaneous infection by two or more pathogens). There are a growing number of articles that demonstrate that HIV and Hepatitis C (HCV) share co-infection. It appears that HIV-positive patients are commonly co-infected with HCV, probably due to shared routes of transmission. Below, we describe the possibility of modelling HIV and HCV co-infection through our SEIR mathematical model introduced in Chapter 5. For example: Colijn and Cohen et al. (2009) demonstrated how SEIR models can be used to model latent co-infection of TB with different strains. Roeger and Feng et al. (2009) formulated simplified deterministic models of co-infection between TB and HIV.

Based on the research problems and issues that are outlined, the following two research issues are explored further in this chapter:

- Neither model above deals with population migration – one of the key factors in the spread of HIV and HCV as well as other diseases – although individual-to-individual transmission is modelled.
- There is very little understanding of how HIV and HCV, and other major co-infection diseases, are related in mathematical terms.

This research will extend the theoretical framework to allow for interactions in co-infected diseases.

7.1 **SEIR Model Enhancement of Co-infection**

When a person infected with HIV acquires another major infection like HCV, this is called co-infection. HIV and HCV are transmitted by exposure to infected blood.

One-quarter of the people infected with HIV also have HCV (Control 2002). To deal with this problem, our SEIR model will be modified in order to make it more effective for modelling co-infection. The SEIR model will be converted to an SEIA (susceptible-exposed-infectious- AIDS) model by replacing the recovery state to an AIDS state. This is because the final state of HIV and co-infection involving HIV will be AIDS.

7.2 **The Future Model**

The transmission and evolution of HIV/AIDS and HCV are dynamic processes. The model classifies the sexually active population into four states with respect to HIV and HCV. These are susceptible, exposed state for HIV, exposed state for HCV, infected state for HIV, infected state for HCV and finally AID cases. Population numbers in each state are denoted as functions of time by $S(t)$, $E_1(t)$, $E_2(t)$, $I_1(t)$, $I_2(t)$ and $A(t)$, respectively. Individuals enter the susceptible state at constant rate $\wedge$. The natural death rate is assumed to be proportional to the population number in each state, with constant μ. There is an AIDS related death constant δ in the AIDS state.

The model assumes that the spread of the disease will be present in future times and changes to the trend only come with human actions at given times during the disease. These activities might include behaviour change and increasing the number of people under HCV. The movement of individuals from one compartment to another is as described in Figure 34.

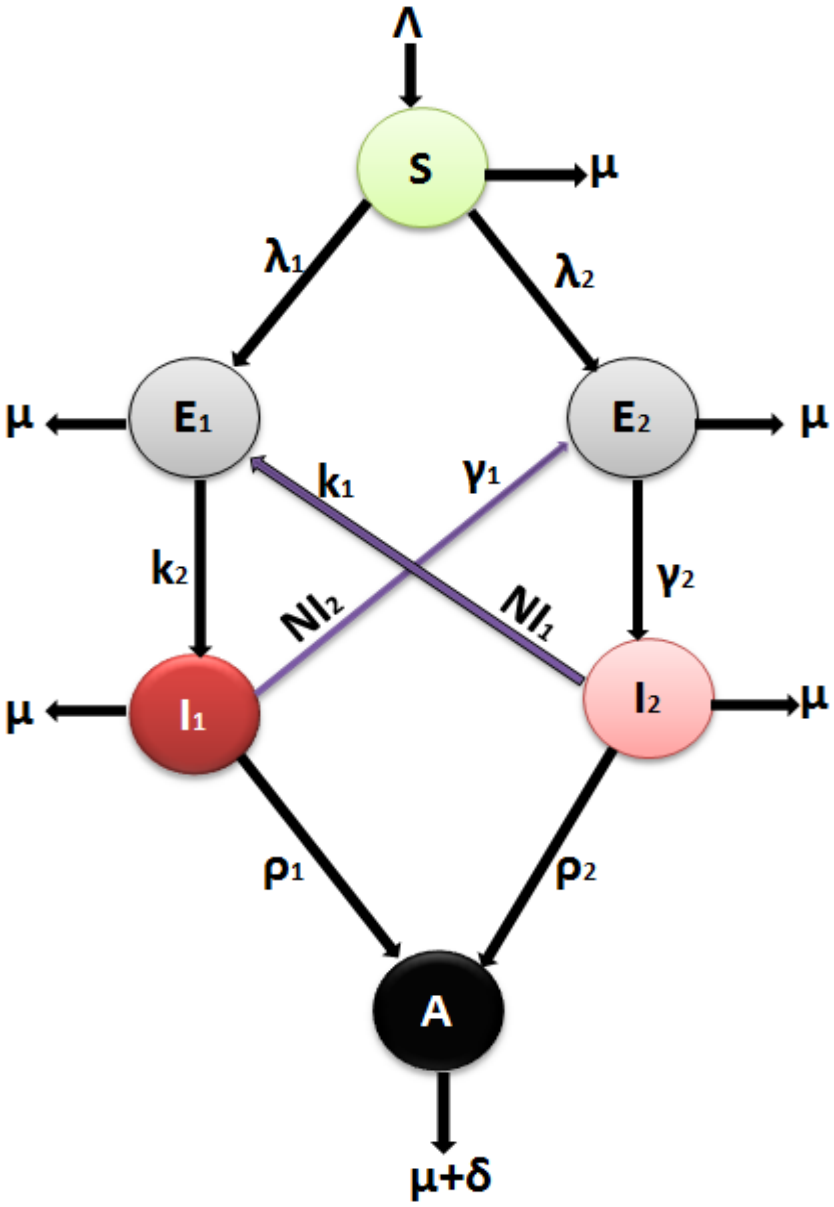

Figure 34: The proposed model for co-infection of HIV/HCV

The following models will be formulated based on Figure 34.

1. SE_1I_1A
2. $SE_1I_1E_2I_2A$
3. SE_2I_2A
4. $SE_2I_2E_1I_1A$

The equations of the SE_1I_1A model are thus given by

$$S^{'}(t) = \Lambda - (\lambda_1 + \lambda_2)S - \mu S \quad (18)$$

$$E_1^{'}(t) = \lambda_1 S + k_1 I_2 - (k_2 + \mu)E_1 \quad (19)$$

$$I_1^{'}(t) = k_2 E_1 - (\gamma_1 + \rho_1 + \mu)I_1 \quad (20)$$

$$A^{'}(t) = \rho_1 I_1 + \rho_2 I_2 - (\mu + \delta)A \quad (21)$$

The total states involved in (18-21) are $N(t) = S(t) + E_1(t) + I_1(t) + A(t)$, where λ_1 and λ_2 are the rates of transmission from the susceptible state to the exposed state of HIV and the exposed state of HCV, respectively. k_2 and γ_2 represent the two exposed rates to the infected state of HIV and the infected state of HCV, respectively. γ_1 and k_1 represent the two infected rates to the exposed state of HIV and the exposed state of HCV, respectively, representing NI_1 and NI_2 (the *N* means Not).Finally, ρ_1 and ρ_2 represent the two infected rates to the AIDS state. Model 2, 3 and 4 will be formulated later.

7.2.1Invariant Region

To check the changes in the population, the variables and the parameters are assumed to be positive for al $t \geq 0$. The model (18-21) will therefore be analysed in a suitable feasible region Φ of preventive interest. The following Lemma1 on the region of the model (4a) is restricted to.

Lemma1

The feasible region Φ defined by

$$\Phi = \left\{ (S, E_1, I_1, A) \in \Re_+^4 \middle| N \leq \frac{\Lambda}{\mu} \right\} \quad (22)$$

with initial conditions $S(0) \geq 0$, $E_1(0) \geq 0$, $I_1 \geq 0$, $A(0) \geq 0$ is positively invariant and attracting with respect to model (4) for $t > 0$.

***Proof*:** Adding the model (18-21) given by

$$\frac{dN}{dt} = \Lambda - \mu N - \delta A \qquad (22)$$
$$< \Lambda - \mu N$$

The solution $N(t)$ of the differential equation (22) has the following property

$$0 \le N(t) \le N(0)e^{-\mu t} + \frac{\Lambda}{\mu}\left[1 - e^{\mu t}\right] \qquad (23)$$

where $N(0)$ represents the sum of initial values of the variables. Note that $N \to \frac{\Lambda}{\mu}$ as $t \to \infty$. If $N(0) \le \frac{\Lambda}{\mu}$, then $\frac{\Lambda}{\mu}$ is the upper bound of N. On the other hand if $N(0) > \frac{\Lambda}{\mu}$, then N will decrease to $\frac{\Lambda}{\mu}$. This means that if $N(0) > \frac{\Lambda}{\mu}$, then the solutions $(S(t), E_1(t), I_1(t), A(t))$ approaches it asymptotically. Hence it is positively invariant under the flow induced by model (18-21).

7.2.2 Positivity of solutions

For model (18-21), it is important to prove that all the state variables remain non-negative so that the solutions of the system with positive initial conditions will be the main positive for all $t > 0$. This is shown in the following lemma 2.

Lemma 2

Given that the initial conditions for model (4) are $S(0) > 0$, $E_1(0) > 0$, $I_1(0) > 0$ and $A(0) > 0$, the solutions (St), $E_1(t)$, $I_1(t)$, and $A(t)$ are non-negative for all $t > 0$.

***Proof*:** Assume that

$\bar{t} = \sup\{t > 0 : S > 0, E_1 > 0, I_1, A > 0\} \in [0, t]$.

Hence $\bar{t} > 0$ and it follows from the first equation of model (22) that

$$\frac{dS}{dt} = \Lambda - (\mu + \lambda_1 + \lambda_2)S. \tag{24}$$

Thus,

$$\frac{d}{dt}\left[S(t)\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\right\}\right]$$

$$= S^{'}(t)\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds + S(t)\exp\left(\mu t + \int_0^t (\lambda_1 + \lambda_2)(S)ds \right)(\mu + (\lambda_1 + \lambda_2)(t))\right\} \tag{25}$$

$$= \exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\left[S^{'}(t) + S(t)(\mu + (\lambda_1 + \lambda_2)(t))\right]\right\} \tag{26}$$

$$= \exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\left[\Lambda - (\mu + (\lambda_1 + \lambda_2)S(t) + S(t)(\mu + (\lambda_1 + \lambda_2)(t))\right]\right\} \text{since} \begin{vmatrix} 0 \le S(t) \le t \\ \lambda(s) \le \lambda(t) \end{vmatrix} \tag{27}$$

Therefore,

$$\frac{d}{dt}\left[S(t)\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\right\}\right] \ge \Lambda\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\right\} \tag{28}$$

Hence,

$$\int_0^{\bar{t}} \frac{d}{dt}\left[S(t)\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\right\}\right] \ge \int_0^{\bar{t}} \Lambda\exp\left\{\mu t + \int_0^t (\lambda_1 + \lambda_2)(s)ds\right\} \tag{29}$$

$$S(\bar{t})\exp\left[\mu t + \int_0^{\bar{t}} (\lambda_1 + \lambda_2)(s)ds \right] - \mathrm{S}(0) \ge \int_0^{\bar{t}} \Lambda\exp\left\{\mu \hat{t} + \int_0^t (\lambda_1 + \lambda_2)(w)dw\right\} d\hat{t} \tag{30}$$

$$S(\bar{t})\exp\left[\mu t + \int_0^{\bar{t}} (\lambda_1 + \lambda_2)(s)ds \right] \ge \mathrm{S}(0) + \int_0^{\bar{t}} \Lambda\exp\left\{\mu \hat{t} + \int_0^t (\lambda_1 + \lambda_2)(w)dw\right\} d\hat{t} \tag{31}$$

so that

$$S(\bar{t}) \ge \mathrm{S}(0)\ \exp\left[-\left(\mu t + \int_0^{\bar{t}} (\lambda_1 + \lambda_2)(s)ds \right)\right] + \exp\left[-\left(\mu t + \int_0^{\bar{t}} (\lambda_1 + \lambda_2)(s)ds \right)\right]\left[\int_0^{\bar{t}} \Lambda\exp\left\{\mu \hat{t} + \int_0^t (\lambda_1 + \lambda_2)(w)dw\right\} d\hat{t} \right] > 0.$$

From the second equation of the model (18-21), we have

$E_1^{'} = \lambda_1 S + k_1 I_2 - (k_2 + \mu) E_1$,

$\Rightarrow E_1(t) \geq E_1(0) \exp[-(k_2 + \mu)(t)] > 0.$

Similarly, it can be shown that, $I_1 > 0$ and $A(t) > 0$ for all $t > 0$ and this completes this proof.

As shown in (11-13) and (14-17), the model and extended version will allow the proposed model to act as death rate parameters for AIDS and there will be additional probabilistic parameters to possibly reduce the spread of the co-infection. Intensive investigation and analysis of this proposed model will be conducted in order to evaluate its accuracy and the capability of the CA model.

It has not been possible in the time available to produce CA simulations of this theoretical model. However, there is also the possibility of real-world data from Ghana being made available to evaluate both this theoretical model and the simulations reported in Chapter 5.

It is natural for any research to have some slight adjustment in the research direction based on the new input gained during the period of research work. The co-infection will be determined by a deterministic model.

Chapter 8

Conclusion

HIV/AIDS has slow transition stages in terms of population migration, therefore the long term effects of HIV/AIDS cannot be observed by investigation of infected individuals. Mathematical models and CA of HIV/AIDS help us understand the dynamics of HIV transmission at the population level of migration. In this thesis, a number of models from the literature are modified to capture different types of HIV/AIDS infection, where different treatment strategies and drug resistance could be applied to minimise the spread of HIV/AIDS infections in Figure 14. We also presented models which focus on how multiple populations can be modelled with four different geographically distributed populations showing how the infected individuals are dominant at different population-levels in the susceptible-infected-recovery $(SIR)^n$ model experiment. The spread of the disease does not depend on the size of the population but on the migratory direction, as shown in Figure 10. Preventative strategies can now be modelled through adjusting the migration schema (Figure 9).

The mathematical disease propagation model of $(SIR)^n$ has been successfully extended to a four compartmental epidemiological model with SEIA which is the spread of infectious disease in a population with delay. The dominant order of the frequencies shows that delay in the transition between exposed to infectious states takes place after peak points in the different time intervals. The detection of these discontinuities (delays) was associated with the speed of spread of disease and is detectable in the frequency and phase of the CWT. This implies that: a) the higher the frequency, the lower the spread of the disease, and b) the faster the spread of the

disease, the slower the frequency. The discrete and continuous transforms demonstrate where and how the propagation of disease was dominated in the population.

In conclusion, the findings showed that migration could play a significant role in the epidemic modelling of HIV. Understanding how the spread of AIDS change in areas with migration is crucial for understanding the overall impact of migration, HIV prevention design as well as for migration destination (Figure 10).

Chapter 9

Summary

The transmission models for directly transmitted HIV/AIDS have been introduced and analysed. The models consist of two different systems of three and four coupled non-linear ordinary differential equations which do not possess a formula solution. However, simple tools from CA allow extracting a great deal of information about the solutions.

The discrete and continuous transform methods were used to evaluate the intensities of the frequencies of the population in sinusoidal patterns. The experimental simulations of the $(SIR)^n$ and SEIR delay model were carried out to assess the difference in the infected individual behaviour in each model, see (Kaddar, Abta et al. 2011). The results show that changes in delay of days have an effect on the distribution of the infected individuals, see, for example (Sen, Agarwal et al. 2010). Although some assumptions are made in epidemiological models, mathematical models still capture the dynamics of HIV/AIDS to an extent. They anticipate the course of an HIV/AIDS epidemic, which is usually longer than the average lifetime of an individual. These estimates direct control efforts for HIV/AIDS and improve success for eradication of the disease. Models identify certain migrant HIV groups at risk at particular time intervals from one population to other in Figure 10. Therefore, epidemiological models and mathematical models with the supplement of CA give an overall view of HIV/AIDS dynamics and direct efforts for surveillance and reduction of AIDS.

Along the way, it was illustrated how the simple $(SIR)^n$ model is possible to create four distinct populations with a predefined migration schema that allows for the

modelled geographical (spatial) distributions of HIV/AIDS by using CA. The spread of the disease does not depend on the size of population but on the migrant direction of the infected individual. Interestingly, the number of infected converged on the 365^{th} iteration for the first three populations (Figure10).

The implications of our models have helped to lay a theoretical foundation for future co-infection (as described in Chapter 7) for public health interventions and how several cornerstones of public health required such a model to illuminate future work.

Finally, the proposed model needs to be tested against real world data to test for fitness in order to evaluate its accuracy and the capability of the CA model, as stated in section 1.3. Also, the proposed model needs to verify the higher the frequency, then the slower the spread of disease and vice versa.

References

Aagaard-Hansen, J., N. Nombela, et al. (2010). "Population movement: a key factor in the epidemiology of neglected tropical diseases." Trop Med Int Health **15**(11): 1281-1288.

Aboagye-Sarfo, P., J. Cross, et al. (2010). "Trend analysis and short-term forecast of incident HIV infection in Ghana." African journal of AIDS research **9**(2): 165-173.

Alimadad, A., V. Dabbaghian, et al. (2011). Modeling HIV spread through sexual contact using a cellular automaton. Evolutionary Computation (CEC), 2011 IEEE Congress on.

Allen, L. J. S. and A. M. Burgin (2000). "Comparison of deterministic and stochastic SIS and SIR models in discrete time." Mathematical Biosciences **163**(1): 1-33.

Anderson, R. M., R. M. May, et al. (1988). "Possible demographic consequences of AIDS in developing countries." Nature **332**(6161): 228-234.

Apenteng, O. O., A. Narayanan, et al. (2012). The Use of Cellular Automata for Modelling (SIR)^N Diseases with Migratory Reinfection. The 2012 3rd International Conference on Computer and Computational Intelligence, Bali Island, Indonesia., ASME

Arino, J., R. Jordan, et al. (2007). "Quarantine in a multi-species epidemic model with spatial dynamics." Mathematical Biosciences **206**(1): 46-60.

Askarieh, G., Å. Alsiö, et al. (2010). "Systemic and intrahepatic interferon-gamma-inducible protein 10 kDa predicts the first-phase decline in hepatitis C virus RNA and overall viral response to therapy in chronic hepatitis C." Hepatology **51**(5): 1523-1530.

Athanassopoulos, S., C. Kaklamanis, et al. (2012). Cellular Automata: Simulations Using Matlab. CDS 2012, The Sixth International Conference on Digital Society Valencia, Spain**:** 63-68.

AvertingHIV/AIDS. (2013). "Worldwide HIV & AIDS Statistics. Regional statistics for HIV and AIDS, end of 2010." Retrieved 21st June, 2013, from http://www.avert.org/worldstats.htm.

Barre-Sinoussi, F. (2004). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Revista de investigacion clinica **56**(2): 126.

Bhunu, C., S. Mushayabasa, et al. (2011). "Mathematical Analysis of an HIV/AIDS Model: Impact of Educational Programs and Abstinence in Sub-Saharan Africa." Journal of Mathematical Modelling and Algorithms **10**(1): 31-55.

Bjørnstad, O. (2005). "SEIR models."

Brauer, F. and C. Castillo-Chávez (2001). Mathematical models in population biology and epidemiology. New York, Springer Verlag.

Brimblecombe, F. S. W., R. Cruickshank, et al. (1958). "Family Studies of Respiratory Infections." BMJ **1**(5063): 119-128.

Brummer, D. (2002). Labour Migration and HIV/AIDS in Southern Africa, International Organisation for Migration Regional Office for Southern Africa**:** 1-26.

Chen, C.-C. (2011). "Quantitative Methodology: Appropriate use in Research for Blind Baseball Ergonomics and Safety Design." The Journal of Human Resource and Adult Learning **9**(4): 6.

Choisy, M., J. F. Guégan, et al. (2006). Mathematical Modeling of Infectious Diseases Dynamics. Encyclopedia of Infectious Diseases, John Wiley & Sons, Inc.**:** 379-404.

Coffee, M., M. N. Lurie, et al. (2007). "Modelling the impact of migration on the HIV epidemic in South Africa." AIDS **21**(3): 343-350 310.1097/QAD.1090b1013e328011dac328019.

Colijn, C., T. Cohen, et al. (2009). "Latent Coinfection and the Maintenance of Strain Diversity." Bulletin of Mathematical Biology(71): 247–263.

Control, C. f. D. (1982). "Opportunistic infections and Kaposi's sarcoma among Haitians in the United States." MMWR Morb Mortal Wkly Rep. **31**(26): 353-354, 360-351.

Control, C. f. D. (2002). Hepatitis C virus and HIV coinfection. Retrieved from http://www.cdc.gov/idu/hepatitis/hepc_and_hiv_co.pdf**:** 1-3.

Cowpertwait, P. S. P. and A. V. Metcalfe (2009). Introductory Time Series with R. London New York, Springer.

Culshaw, R. V. (2006). "Mathematical Modelling of AIDS Progression: Limitations, Expectations, and Future Directions." Journal of American Physicians and Surgeon **11**: 4.

d'Onofrio, A. (2002). "Stability properties of pulse vaccination strategy in SEIR epidemic model." Mathematical Biosciences **179**(1): 57-72.

Dascalus, M., G. Stafen, et al. (2011). "Applications of multilevel cellular automata in epidemiology." 439-444.

Deering, K. N., P. Vickerman, et al. (2008). "The impact of out-migrants and out-migration on the HIV/AIDS epidemic: a case study from south-west India." AIDS **22**: S165-S181 110.1097/1001.aids.0000343774.0000359776.0000343795.

Donnelly, C. and D. Cox (2001). "Mathematical biology and medical statistics: contributions to the understanding of AIDS epidemiology." Statistical Methods in Medical Research **10**(2): 141-154.

En'ko, P. D. (1989). "On the course of epidemics of some infectious diseases." International journal of epidemiology **18**(4): 749-755.

Eykhoff, P. (1974). System Identification: Parameter and State Estimation, Wiley-Interscience.

Falconer, K., J. K. Sandberg, et al. (2009). "HCV/HIV co-infection at a large HIV outpatient clinic in Sweden: Feasibility and results of hepatitis C treatment." Scandinavian Journal of Infectious Diseases **41**(11-12): 881-885.

Ferguson, N. M., M. J. Keeling, et al. (2003). "Planning for smallpox outbreaks." Nature **425**(6959): 681-685.

Giordano, F. R., W. P. Fox, et al. (2009). Mathematical Modelling. United State, Charlie Van Wagner.

Gottlieb, M., H. Schanker, et al. (1981). Epidemiologic Notes and Reports Pneumocystis Pneumonia Los Angeles, Centers for Disease Control and Prevention. **30:** 1-3.

Greenwood, M. (1931). "On the statistical measure of infectiousness." J Hyg (Lond) **31**(3): 336-351.

Griffiths, M. (2011). "The M in STEM via the m in epidemiology." Teaching Mathematics and its Applications **30**(3): 151-164.

Guide4Living. (2013). "What is AIDS? - Understanding the Epidemic." Retrieved 22nd June, 2013, from http://www.guide4living.com/hiv-aids/aids.htm.

Heasman, M. A. (1961). "Theory and observation in family epidemics of common cold." Journal of epidemiology and community health (1979) **15**(1): 12.

Hethcote, H. W. (2000). "The Mathematics of Infectious Diseases." SIAM Review **42**(4): 599-653.

Hogan, D. R., A. M. Zaslavsky, et al. (2010). "Flexible epidemiological model for estimates and short-term projections in generalised HIV/AIDS epidemics." Sexually Transmitted Infections **86**(Suppl 2): ii84-ii92.

Hove-Musekwa, S., F. Nyabadza, et al. (2011). "Modelling Hospitalization, Home-Based Care, and Individual Withdrawal for People Living with HIV/AIDS in High Prevalence Settings." Bulletin of Mathematical Biology **73**(12): 2888-2915.

Huang, N. E., Z. Shen, et al. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**(1971): 903-995.

Isham, V. (1988). "Mathematical Modelling of the Transmission Dynamics of HIV Infection and AIDS: A Review." Journal of the Royal Statistical Society. Series A (Statistics in Society) **151**(1): 5-49.

Joshi, H., S. Lenhart, et al. (2008). "Modelling the effect of information campaigns on the HIV epidemic in Uganda." Math Biosci Eng **5**(4): 757-770.

Kaddar, A., A. Abta, et al. (2011). "A comparison of delayed SIR and SEIR epidemic models." **16**(2).

Keeling, M. J. (1999). "The effects of local spatial structure on epidemiological invasions." Proceedings of the Royal Society of London B **266**: 859–867.

Keeling, M. J. and P. Rohani (2008). "Modeling Infectious Diseases in Humans and Animals." 1-366.

Kermack, W. O. and A. G. McKendrick (1927). "A Contribution to the Mathematical Theory of Epidemics." Proceedings of the Royal Society of London. Series A **115**(772): 700-721.

Kibona, I., W. Mahera, et al. (2011). "A deterministic model of HIV/AIDS with vertical transmission in the presence of infected immigrants." International Journal of the Physical Sciences **6**(23): 5383–5398.

Kim, J.-H. (2009). Dynamic partnerships and HIV transmissions by stage. United States -- Michigan, University of Michigan. **Ph.D.**

Li, M. Y., J. R. Graef, et al. (1999). "Global dynamics of a SEIR model with varying total population size." Mathematical Biosciences **160**(2): 191-213.

Lichtenegger, K. (2005). "Stochastic Cellular Automata Models in Disease Spreading and Ecology." 1-106.

Liu, Q.-X., R.-H. Wang, et al. (2008). "Persistence, extinction and spatio-temporal synchronization of SIRS cellular automata models." J. Stat. Mech.: 1-12.

M.J, F., E. García, et al. (2010). Epidemiological Modeling based on Dynamic Neighbourhoods. Proceedings of the Seventh International Conference on Engineering Computational Technology, Civil-Comp Press, Stirlingshire, UK, Paper 98.

May, R. M. A., Roy M. (1987). "Transmission dynamics of HIV infection." **326**(6109).

Meng, X.-z., L.-s. Chen, et al. (2007). "Global dynamics behaviors for new delay SEIR epidemic disease model with vertical transmission and pulse vaccination." Applied Mathematics and Mechanics **28**(9): 1259-1271.

Merli, M. G., S. Hertog, et al. (2006). "Modelling the Spread of HIV/AIDS in China: The Role of Sexual Transmission." Population Studies **60**(1): 1-22.

Mikler, A. R., S. Venkatachalam, et al. (2005). "Modelling infectious diseases using global stochastic cellular automata." Journal of Biological Systems **13**(04): 421-439.

Milne, G. J. and S. C. Fu (2004). "A Flexible Automata Model for Disease Simulation." Lecture Notes in Computer Science **3305**: 642-649.

Mukandavire, Z., P. Das, et al. (2011). "HIV/AIDS Model with Delay and the Effects of Stochasticity." Journal of Mathematical Modelling and Algorithms **10**(2): 181-191.

Naresh, R., A. Tripathi, et al. (2011). "A nonlinear HIV/AIDS model with contact tracing." Applied Mathematics and Computation **217**(23): 9575-9591.

Nyabadza, F. and Z. Mukandavire (2011). "Modelling HIV/AIDS in the presence of an HIV testing and screening campaign." Journal of Theoretical Biology **280**(1): 167-179.

Nyabadza, F., Z. Mukandavire, et al. (2011). "Modelling the HIV/AIDS epidemic trends in South Africa: Insights from a simple mathematical model." Nonlinear Analysis: Real World Applications **12**(4): 2091-2104.

Özalp, N. and E. Demirci (2011). "A fractional order SEIR model with vertical transmission." Mathematical and Computer Modelling **54**(1-2): 1-6.

Pietro G, C. (2007). "How mathematical models have helped to improve understanding the epidemiology of infection." Early Human Development **83**(3): 141-148.

Rao, A. S. R. S. (2003). "Mathematical modelling of AIDS epidemic in India." **84** (9): 1192-1197.

Roberts, M. and H. Heesterbeek (1993). "Bluff your way in epidemic models." Trends in Microbiology **1**(9): 343-348.

Roberts, M. G., M. Baker, et al. (2007). "A model for the spread and control of pandemic influenza in an isolated geographical region." J R Soc Interface **4**(13): J R Soc Interface.

Roeger, L.-I. W., Z. Feng, et al. (2009). "Modelling TB and HIV co-infections." Math Biosci Eng **6**(4): 815–837.

Sen, M. D. l., R. P. Agarwal, et al. (2010). "On a Generalized Time-Varying SEIR Epidemic Model with Mixed Point and Distributed Time-Varying Delays and Combined Regular and Impulsive Vaccination Controls." Advances in Difference Equations **2010**: 42.

Sepkowitz, K. A. (2001). "AIDS — The First 20 Years." New England Journal of Medicine **344**(23): 1764-1772.

Sharma, S. K., A. Mohan, et al. (2005). "HIV-TB co-infection: epidemiology, diagnosis & management." Indian J Med Res. **121**(4): 550-567.

Shuai, Z. and P. V. D. Driessche (2011). "Global dynamics of a disease model including latency with distributed delays." Canadian Applied Mathematics Quarterly **19**(3): 235-253.

Situngkir, H. (2004). Epidemiology Through Cellular Automata: Case of Study Avian Influenza in Indonesia**:** 1-10.

Srivastava, P., M. Banerjee, et al. (2009). " Modeling the drug therapy for HIV infection." Journal of Biological Systems **17**(2): 213-223.

UNAIDS. (2010). "AIDS Epidemic Updates."

van den Driessche, P. and J. Watmough (2005). "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission." Mathematical Biosciences **180**(1-2): 29-48.

Vynnycky, E. and R. G. White (2010). "An Introduction to Infectious Disease Modelling." 368.

Wang, J.-j., R. K. Heather, et al. (2010). "Dynamic mathematical models of HIV/AIDS transmission in China." Chinese medical journal **123**(15): 2120-2127.

Weiss, R. (1993). "How does HIV cause AIDS?" Science **260**(5112): 1273-1279.

White, S. H., A. M. del Rey, et al. (2007). "Modeling epidemics using cellular automata." Applied Mathematics and Computation **186**(1): 193-202.
Yan, P. and S. Liu (2005). "SEIR epidemic model with delay." ANZIAM J. **48(2006)**: 119-134.
Zorzenon dos Santos, R. M. and S. Coutinho (2001). "Dynamics of HIV Infection: A Cellular Automata Approach." Physical Review Letters **87**(16): 168102.

Appendix A

In the SIR model, R can be determined if and only if S and I are known. By dropping R from the (1-3) from the SIR model, we have

$$\frac{dS}{dt} = -\beta SI \tag{A1}$$

$$\frac{dI}{dt} = (\beta S - \alpha)I \tag{A2}$$

These become very difficult to solve analytically but we can solve these equations following this qualitative approach. We can observe that $\frac{dS}{dt} < 0$ for all t and $\frac{dI}{dt} > 0$ if and only if $S > \alpha/\beta$. With this assumptions it is given that (A1) and (A2) can be solve analytically as

$$\frac{dI}{dS} = \frac{(\beta S - \alpha)I}{\beta SI} \tag{A3}$$

$$\frac{dI}{dS} = -1 + \frac{\alpha}{\beta S} \tag{A4}$$

$$I = -S + \frac{\alpha}{\beta}\ln S + K \tag{A5}$$

This implies that

$$K = I_0 + S_0 - \frac{\alpha}{\beta}\ln S_o \text{, where } K \text{ is constant.} \tag{A6}$$

On the condition that at $t = 0$, $I = I_0$ and $S = S_0 = N - I_0$, at $R_0 = 0$ for all $t > 0$ $0 \le S + I < N$. We have

$$I = -S + \frac{\alpha}{\beta}\ln S + I_0 + S_0 - \frac{\alpha}{\beta}\ln S_o \tag{A7}$$

$$I = N - S + \frac{\alpha}{\beta}\ln\frac{S}{S_0} \tag{A8}$$

For maximum, $I_{\max}$ occurs at $S = \alpha/\beta$ where $\frac{dI}{dt} = 0$. From (11) with $S = \alpha/\beta$

$$I_{\max} = \frac{\alpha}{\beta}\ln\frac{\alpha}{\beta} + I_0 + S_0 - \frac{\alpha}{\beta}\ln S_o \quad \text{(A9)}$$

$$I_{\max} = N - \frac{\alpha}{\beta} + \frac{\alpha}{\beta}\ln\left[\frac{(\alpha/\beta)}{S_0}\right] \quad \text{(A10)}$$

Hence, I increases as long as $S > \alpha/\beta$ but since S decreases for all t, I basically decreases and approaches zero. If $S(0) < \alpha/\beta$, I decreases to zero there is no epidemic. On other hand, if I increases to maximum when $S = \alpha/\beta$ and then decreases to zero, there is an epidemic). The quantity $\beta S(0)/\alpha$ is a threshold quantity, called the basic reproduction number and is denoted by R_0, which demonstrates whether there is an epidemic or not. If $R_0 < 1$ then infection dies out, while if $R_0 > 1$ there is an epidemic.

Threshold Theorem of Epidemiology

From the epidemic model we can observe in terms of phase plane that the course of the epidemic at t runs t_0 to t to ∞. This implies that

$$\frac{dS}{dt}(t) < 0 \quad \text{(A11)}$$

that $S(t)$ is a decreasing function and

$$\lim_{t\to\infty} = S_\infty \quad \text{(A12)}$$

The unique root of (1) can obtain S_0 for $t \to \infty$:

$$I_0 + S_0 - S_\infty + \rho\ln\frac{S_\infty}{S_0} = 0, \text{ with } \rho = \frac{\alpha}{\beta} \quad \text{(A13)}$$

S_∞ depicts the number of susceptible who were infected before. From (1-3) the number of susceptible can be calculated at any time t

$$\frac{dS}{dR} = \frac{1}{\rho} S(t) \tag{A14}$$

$$\frac{dS}{S} = \frac{1}{\rho} S(t)\, dR \tag{A15}$$

$$\left[\ln S(t)\right]_0^t = \left[-\frac{R}{\rho}\right]_0^t \tag{A16}$$

$$\ln S(t) - \ln S_0 = \frac{R_0}{\rho} - \frac{R(t)}{\rho} \tag{A17}$$

$$S(t) = S_0 \exp\left[-\frac{\left(R_0 - R(t)\right)}{\rho}\right] \geq S_0 e^{-\frac{N}{\rho}} > 0 \tag{A18}$$

This will always show a positive value, thus, there always remains some susceptible who have never infected.

Theorem 1

Kermack-Mckendrick: A general epidemic evolves according to the differential equation (1) from initial values $(S_0, I_0, 0)$; where $S_0 + I_0 = N$.

(i) (*Survival and Total Size*) When infection ultimately ceases spreading, a positive number S_∞ of susceptible remains uninfected, and the total number of individuals ultimately infected and removed equals to $S_0 + I_0 - S_\infty$. It is the unique root of the equation

$$N - R_\infty = S_0 + I_0 - R_\infty = S_0 e^{-\frac{R_\infty}{\rho}}, \tag{A19}$$

where $I_0 < R_0 < S_0 + I_0$, $\rho = \alpha/\beta$ being the relative removal rate.

(ii) (*Second Threshold Theorem*) If S_0 exceeds ρ by a small quantity v, and if the initial number of infectious I_0 is small relative to v, then the final number of susceptible left in the population is approximately $\rho - v$, and $R_\infty \approx 2v$. In other words, the level of susceptible is reduced to appoint as far below the threshold as it originally was above it.

Proof:

(i) From (1-3), we can say that

$$\frac{1}{S}\frac{dS}{dt} = \frac{\beta}{\alpha}\frac{dR}{dt} = \frac{1}{\rho}\frac{dR}{dt} \tag{A20}$$

from this we get

$$S(t) = S_0 e^{\frac{R_0 - R(t)}{\rho}} \tag{A21}$$

Thus,

$$S(t) = S_0 e^{-\frac{R(t)}{\rho}} \tag{A22}$$

we can assume that R_0 equals zero. In (A13) and (A18), for $t \to 0$ as already shown, we get

$$S_\infty = S_0 e^{-\frac{R_\infty}{\rho}} > 0. \tag{A23}$$

$N - R_\infty = S_0 + I_0 - R_\infty = N - R_\infty = I_\infty + S_\infty$, where $I_\infty = 0$. Therefore:

$N - R_\infty = S_0 e^{-\frac{R_\infty}{\rho}}$ □

(ii) Using the relations between (A1) and (A17) with constraints on the population size we obtain

$$\frac{dR}{dt} = \alpha\left(N - R(t) - S_0 e^{-\frac{R_\infty}{\rho}}\right) \tag{A24}$$

The equation is very different to have an explicit solution for R in terms of *t*. By using the Taylor's expansion of the exponential term $\left(\text{i.e. } e^{-\frac{R(t)}{\rho}}\right)$ according to the formula $e^{-x} = 1 - x + \frac{1}{2}x^2 + Q(x^3)$ and by neglecting the last term. We get

$$\frac{dR}{dt} \approx \alpha\left[N - S_0 + R\left(\frac{S_0}{\rho} - 1\right) - \frac{R^2 S_0}{2\rho^2}\right] \tag{A25}$$

The expression of the right-hand side is as follows:

$$\frac{dR}{dt} \approx \frac{\rho^2 \alpha}{2S_0}\left[(N-S_0)\frac{2S_0}{\rho^2}+\left(\frac{S_0}{\rho}-1\right)^2-\left(\frac{S_0}{\rho^2}\left[R-\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)\right]\right)^2\right]. \quad \text{(A26)}$$

Let

$$\gamma = \left[(N-S_0)\frac{2S_0}{\rho^2}+\left(\frac{S_0}{\rho}-1\right)^2\right]^{\frac{1}{2}} \quad \text{(A27)}$$

We get

$$\frac{dR}{dt} \approx \frac{\rho^2 \alpha}{2S_0}\left[\gamma^2-\left(\frac{S_0}{\rho^2}\left[R-\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)\right]\right)^2\right] \quad \text{(A28)}$$

By substituting

$$\gamma \tanh\upsilon = \frac{S_0}{\rho^2}\left[R-\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)\right]. \quad \text{(A29)}$$

So we get:

$$\frac{dR}{dt} \approx \frac{\rho^2 \alpha}{2S_0}\left(\gamma^2-\gamma^2\tanh^2\upsilon\right) \quad \text{(A30)}$$

From the relation for γ in (A29) we also have:

$$R = \frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)+\frac{\rho^2\alpha}{S_0}\tanh\upsilon \quad \text{(A31)}$$

at $t = 0 \Rightarrow 0 \Rightarrow R_0 = 0 \Rightarrow$

$$\upsilon_0 = \tanh^{-1}\left(-\frac{1}{\gamma}\left[\frac{S_0}{\rho}-1\right]\right) \quad \text{(A32)}$$

Hence, (A29) and (A30) gives:

$$\frac{dR}{dt} \approx \frac{\rho^2 \alpha}{2S_0}\left(\gamma^2-\gamma^2\tanh^2\upsilon\right) = \frac{\rho^2}{S_0}\gamma\sec h^2\upsilon\frac{d\upsilon}{dt} \quad \text{(A33)}$$

$$\frac{\rho^2\alpha\gamma^2}{2S_0}\left(1-\tanh^2\upsilon\right)=\frac{\rho^2}{S_0}\gamma\sec h^2\upsilon\frac{d\upsilon}{dt} \tag{A34}$$

Hence,

$$\frac{d\upsilon}{dt}=\frac{1}{2}\alpha\gamma \tag{A35}$$

$$\upsilon=\frac{1}{2}\alpha\gamma t+\upsilon_0 \tag{A36}$$

and we can now obtain the solution for $R(t)$ by solving the differential equation

$$\frac{dR}{dt}=\frac{\rho^2\gamma}{S_0}\sec h^2\upsilon\frac{d\upsilon}{dt} \tag{A37}$$

$$\int_0^t dR=\int_0^t\frac{\rho^2}{S_0}\gamma\sec h^2\upsilon\,d\upsilon \tag{A38}$$

$$\int_0^t dR=\frac{\rho^2}{S_0}\gamma\int_0^t\sec h^2\upsilon\,d\upsilon \tag{A39}$$

$$R(t)=\frac{\rho^2}{S_0}\gamma\left[\tanh\upsilon\right]_0^t \tag{A40}$$

$$R(t)=\frac{\rho^2}{S_0}\gamma\left[\tanh\upsilon-\tanh\upsilon_0\right] \tag{A41}$$

where we use the formula (A36) to get

$$R(t)=\frac{\rho^2}{S_0}\gamma\left[\tanh\left(\frac{1}{2}\upsilon\alpha t+\upsilon_0\right)-\tanh\upsilon_0\right] \tag{A42}$$

$$R(t)=\frac{\rho^2}{S_0}\gamma\left[\tanh\left(\frac{1}{2}\upsilon\alpha t+\upsilon_0\right)\right]+\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right) \tag{A43}$$

$$R(t)=\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)+\frac{\gamma\rho^2}{S_0}\left[\tanh\frac{\gamma\alpha t}{2}+\upsilon_0\right] \tag{A44}$$

and finally, using (A32) we have:

$$R(t)=\frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1\right)+\frac{\gamma\rho^2}{S_0}\tanh\left(\frac{\gamma\alpha t}{2}-\omega\right) \tag{A45}$$

$$\omega = \tanh^{-1}\left[\frac{1}{\gamma}\left(\frac{S_0}{\rho}-1\right)\right]. \qquad \text{(A46)}$$

Equations (A45) and (A46) show a symmetric bell shaped curve in Figure 34. It is also known as the epidemic curve of the disease. The epidemic curve is widely used due to its significant comparison between the results predicted by models with the data from public health statistics in various models.

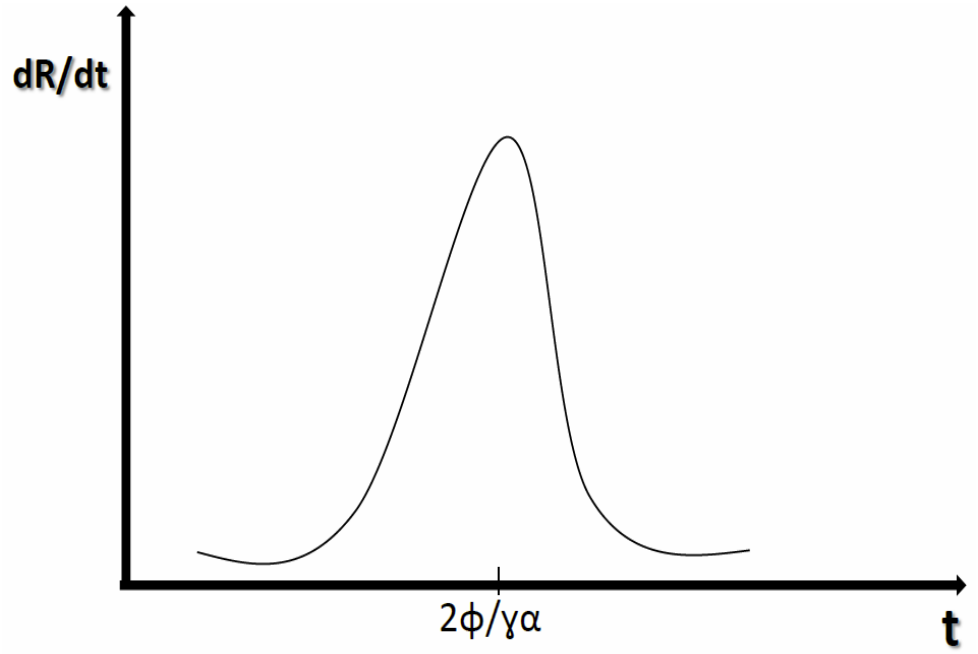

Figure 35: Epidemic curve

From (A46), $R_\infty = \lim_{t\to\infty} R(t)$ is given by

$$R_\infty \approx \frac{\rho^2}{S_0}\left(\frac{S_0}{\rho}-1+\gamma\right) \qquad \text{(A47)}$$

Taking (A22) into consideration, when $2S_0(N-S_0) << (S_0-\rho^2)$ and $S_0 > \rho$,

$$R_\infty = 2\rho\left(1-\frac{\rho}{S_0}\right), \qquad \text{(A48)}$$

From (A48) we get

$R_\infty \approx 2v$ □

when taking into consideration that $S_0 = \rho + v$ for some $v > 0$.

Equivalently, $S_\infty \approx \rho + v - 2v = \rho - v$.

Appendix B

$$\frac{dS}{dt} = \delta R - \beta SI \quad \text{(B1)}$$

$$\frac{dI}{dt} = \beta SI - \alpha I \quad \text{(B2)}$$

$$\frac{dR}{dt} = \alpha I - \delta R \quad \text{(B3)}$$

Theorem 2

Let the initial data be $S(\theta) = S_0(\theta) \geq 0$, $I(\theta) = I_0(\theta) \geq 0$, and $R(\theta) = R_0(\theta) \geq 0$ for all $\theta \in [-\tau, 0)$, with $S_0(0) > 0$, $I_0(0) > 0$, and $R_0(0) > 0$. Then solutions $S(t)$, $I(t)$ and $R(t)$ of system (3) are positive for all $t \geq 0$. For the model system (B1-B3), the region $\Re$ is positively invariant and all solutions starting in $\Re_{+0}$ or $\Re_{+}$ approach, enter or stay in $\Re$.

Proof:

To look for steady states we solve the system:

$$0 = \delta R - \beta SI \quad \text{(B4)}$$

$$0 = \beta SI - \alpha I \quad \text{(B5)}$$

$$0 = \alpha I - \delta R \quad \text{(B6)}$$

From (B6) we get $R = \frac{\alpha}{\delta} I$ (B7)

From (B5) we get either $I = 0$ or $S = \frac{\alpha}{\beta}$ (B8)

If $I = 0$, $R = 0$. Consequently, we get a disease free equilibrium given by $E_0 = (1,0,0)$

The basic reproduction number of system (B5) is given by $R_0 = \frac{\beta}{\alpha}$.

To find the endemic equilibrium, we substitute $S = \frac{\alpha}{\beta}$ and $R = \frac{\alpha}{\delta} I$ in

$$S_e + I_e + R_e = 1 \qquad \text{(B9)}$$

Thus,

$$\frac{\alpha}{\beta} + I_e + \left(\frac{\alpha}{\delta}\right) I_e = 1 \qquad \text{(B10)}$$

$$I_e \left(\frac{\delta + \alpha}{\delta}\right) = \frac{\beta - \alpha}{\beta} \qquad \text{(B11)}$$

$$I_e = \frac{\delta}{\beta}\left(\frac{\beta - \alpha}{\delta + \alpha}\right) \qquad \text{(B12)}$$

Therefore

$$R_e = \frac{\beta}{\alpha}\left(\frac{\beta - \alpha}{\delta + \alpha}\right) \qquad \text{(B13)}$$

This leads to a unique endemic equilibrium given by $E_e = \left(\frac{\alpha}{\beta}, \frac{\delta(\beta - \alpha)}{\beta(\delta + \alpha)}, \frac{\beta(\beta - \alpha)}{\alpha(\delta + \alpha)}\right)$.

Theorem 3

The disease-free equilibrium is locally and globally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.

Theorem 4

The endemic equilibrium is locally asymptotically stable if $R_0 > 1$ and unstable otherwise.

Appendix C

Let the initial data be $S(\theta) = S_0(\theta) \geq 0$, $E(\theta) = E_0(\theta) \geq 0$, $I(\theta) = I_0(\theta) \geq 0$, and $R(\theta) = R_0(\theta) \geq 0$ for all $\theta \in [-\tau, 0)$, with $S_0(0) > 0$, $E_0(0) > 0$, $I_0(0) > 0$, and $R_0(0) > 0$. Then solutions $S(t)$, $E(t)$, $I(t)$, and $R(t)$ of the SEIR model are positive for $t \geq 0$.

Proof:

We use matrix F to represent the rate of new infections in different compartments, differentiated with respect to E and I and then evaluated at the disease-free equilibrium.

$$F = \begin{pmatrix} \frac{\partial F}{\partial E} & \frac{\partial F}{\partial I} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{ij} - \beta_{ij}k \\ 0 & 0 \end{pmatrix} \tag{C1}$$

We use V to represent the rate of transfer of infected from one compartment to another $(V_i = V_i^- - V_i^+)$:

$$V = \begin{pmatrix} \frac{\partial V}{\partial E} & 0 \\ -\varepsilon & \frac{\partial V}{\partial I} \end{pmatrix} = \begin{pmatrix} \mu + \varepsilon & 0 \\ -\varepsilon & \mu + \alpha \end{pmatrix} \tag{C2}$$

$$V = \frac{1}{(\mu + \alpha)(\mu + \varepsilon)} \begin{pmatrix} \mu + \alpha & 0 \\ \varepsilon & \mu + \varepsilon \end{pmatrix} \tag{C3}$$

$$FV^{-1} = \begin{pmatrix} 0 & \beta_{ij} - \beta_{ij}k \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\mu + \varepsilon & 0 \\ \varepsilon/(\mu + \varepsilon)(\mu + \alpha) & 1/\mu + \alpha \end{pmatrix} \tag{C4}$$

$$= \begin{vmatrix} \frac{(\beta_{ij} - \beta_{ij}k)\varepsilon}{(\mu + \varepsilon)(\mu + \alpha)} - \lambda & \frac{\beta_{ij} - \beta_{ij}k)}{\mu + \alpha} \\ 0 & -\lambda \end{vmatrix} = 0 \tag{C5}$$

Thus,

$$R_0 = \frac{(\beta_{ij} - \beta_{ij}k)\varepsilon}{(\mu+\varepsilon)(\mu+\alpha)} = \frac{\beta_{ij}\varepsilon(1-e^{-\mu\tau})}{(\mu+\varepsilon)(\mu+\alpha)} \quad \square$$

$$R_0 = \rho(FV^{-1})$$

where ρ denotes the spectral radius. If $R_0 < 1$ then the DFE is globally asymptotically stable, and if $R_0 > 1$ then the DFE is unstable (Shuai and Driessche 2011).

Appendix D

Algorithm for Wavelet

For simplicity, we let $s(k)$ represent susceptible $S(k)$, exposed $E(k)$, infected $R(k)$, and recovery $R(k)$ in each cases respectively.

$$C_{a,b} = \int_R s(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \qquad \text{(D1)}$$

$$C_{c,b} = \sum_k \int_k^{k+1} s(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \qquad \text{(D2)}$$

since $s(t) = s(k)$, if $t \in [k, k+1]$

then

$$C_{a,b} = \frac{1}{\sqrt{a}}\sum_k s(k)\int_k^{k+1}\psi\left(\frac{t-b}{a}\right)dt \qquad \text{(D3)}$$

$$C_{a,b} = \frac{1}{\sqrt{a}}\sum_k s(k)\left[\int_{-\infty}^{k+1}\psi\left(\frac{t-b}{a}\right)dt - \int_{-\infty}^{k}\psi\left(\frac{t-b}{a}\right)dt\right] \qquad \text{(D4)}$$

so at any scale *a*, the wavelet coefficients $C_{a,b}$ for $b = 1$ to *length* (*s*) can be obtained by convolving the signal *s* and a dilated and translated version of the integrals of the form

$\int_{-\infty}^{k}\psi(t)dt$.