

PERFORMANCE AND SCALABILITY OF GEOGRAPHICALLY-EXPLICIT AGENT-BASED DISEASE DIFFUSION MODELS

NETSANET HAGOS GIRMAY
February, 2012

SUPERVISORS:

Ir. P.W.M. Ellen-Wien Augustijn
Dr.Ir. R.A. (Rolf) de By



PERFORMANCE AND SCALABILITY OF GEOGRAPHICALLY-EXPLICIT AGENT-BASED DISEASE DIFFUSION MODELS

NETSANET HAGOS GIRMAY

Enschede, The Netherlands, February, 2012

Thesis submitted to the Faculty of Geo-information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.

Specialization: Geoinformatics (GFM)

SUPERVISORS:

Ir. P.W.M. Ellen-Wien Augustijn

Dr.Ir. R.A. (Rolf) de By

THESIS ASSESSMENT BOARD:

Dr.Ir. R.A. (Rolf) de By (chair)

Dr. T. Filatova, University of Twente

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Due to its direct consequence on physical and psychological impact of human beings, an epidemic is considered as one of the main threats of any country. Given its unlimited geographical boundary and rapid transmission rate, efficient methods are required to understand temporal and spatial aspects of disease diffusion, individual contact processes as well as their behavior that greatly affects the course of an epidemic outbreak. Computer simulation models have been used to accurately and efficiently depict human interactions, emergent behaviors and commuting that cover large areas which, however, require large computational resources. In this study an already existing Agent-based model for the simulation of Pertussis in Enschede is used. However, enabling the model to run for larger number of agents and spatial extent requires implementation of commuting of agents, and scaling temporal extent of the model requires recovery of infected agents as well.

This study proposed innovative methods for scaling agent-based models toward simulating a large spatial and temporal extent while concurrently improving its performances. The study first examines, test and identifies factors influencing performance of the existing pertussis model. It then develops a conceptual model to adjust and improve the identified limitations that enable to scale up the model spatially and temporally. Subsequently to implement the developed conceptual model Netherlands was selected as a case study area since the existing agent based simulation model simulates the spread of infectious disease pertussis/ whooping cough for the area of Enschede.

The concept of hierarchy based hybrid modelling approach which uses advantages of agent based modelling, metapopulation modeling, and hierarchical modeling approaches was utilized to develop a conceptual model toward simulating the process of pertussis diffusion for the whole Netherlands. The concept of hierarchical modeling technique was used to break down the spatial extent (Netherlands) into three levels (urban system level, municipality level, individual based level) and the temporal extent of one year into 30 minutes, 12 hours and twice a week. Similarly, the concept of age structured metapopulation modeling technique was utilized for computing disease diffusion and for creating conceptual models for both municipality and urban system levels.

The existing model was adjusted and its scalability tested spatially and temporally. The analysis of the tests revealed that the model was scaled up from the first 47 neighborhoods to the complete Enschede area as well as from 67,000 to around 155,000 agents. To implement the prototype two sub-models (commuting model and disease model) for the municipality level were implemented and tested individually. Apart from these developed two separate models, attempt has also made to join these two models to examine the pattern and direction of disease spread. Generally, the developed conceptual model and employed approaches in this study has proven to be promising for scaling of the Pertussis model without increasing the number of agents significantly, flow of information between municipalities and providing insights toward simulating a large spatial and temporal extent while concurrently improving its performances.

Keywords

Pertussis, epidemic modelling, agent based model, hierarchical modelling approaches, metapopulation modelling approach, hybrid modelling, Enschede, Netherlands

ACKNOWLEDGEMENTS

I thought writing acknowledgments is simple, but it is not. It is even harder than writing abstract. "A lot of supports, a lot of love which one to mention?"

In the course of carrying out this study, I benefited the support from many individuals to whom I would like to extend my sincere gratitude. My utmost thanks go to my first supervisor "My Mom", Ms. Ellen-Wien Augustijn. She has been there for me in every way, even when I couldn't believe in myself. She gave me courage, strength and confidence. She has been a source of intellectual inspiration and guidance. THANK YOU SO MUCH.

I would also like to thank my second supervisor Dr.Ir. R.A. (Rolf) de By, for his support, encouragement and brilliant ideas.

Many thanks to Mr. Bas Retsios (Repast Expert) who is the smartest, experienced, but also kind, generous and cooperative person I had ever seen. Thank you so much. With all their constructive comments, enthusiasm, continuous guidance and feedback throughout the thesis process made this study possible.

I am greatly indebted to The Netherlands Fellowship Programme-NFP who generously granted the funding for my study. Without the financial support this study would not have been made.

I would like to thank the academic staff members of GFM including instructors, secretaries and other supporting staff for their guidance through this study and their inspiring lectures, in-depth knowledge and practical exercises and research skills during my study.

Special thanks go to the Library staff members of ITC for supporting me especially when I was in dire need of books and resources. I owe special thanks to all persons who helped me.

Time of my studies at ITC is unforgettable as it gave me an opportunity to meet, learn and share experience with students from across the globe with different background and learning process. I have learned much from each of you throughout the year. Now I can proudly say I have friends across the globe.

I would like to take this opportunity to owe my heart felt sincerest appreciation to my best friend Negera Dinsa, for his continuous support and encouragement through the course of this study. "Without you nothing could be possible".

I would also like to thank my best friend Kiflom Desta who filled the scholarship for me. Thank you so much you are the reason I am here.

I am grateful to my parents, my brothers, sister and friends for their love and support. This work is dedicated to all of you and without whom I would not be where I am today. Since it is difficult to mention all who assisted me, I take this opportunity to say, GOD BLESS YOU ALL.

Netsanet Hagos Girmay

Enschede,

The Netherlands, February 2012

TABLE OF CONTENTS

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Motivation And Problem Statement	1
1.2 Research Identification	3
1.2.1 Research Objectives	3
1.2.2 Research Questions	3
1.2.3 Innovation Aimed At	4
1.2.4 Related Work	4
1.3 Project Set-Up	5
1.3.1 Method Adopted	5
1.4 Overview of the following chapters	7
2 Background Information	9
2.1 Characteristics Of Disease Diffusion (Epidemics)	9
2.1.1 What Are Epidemics?	9
2.1.2 Diffusion process of epidemics	9
2.1.3 Factors facilitating disease diffusion	11
2.1.4 Characteristics Of Pertussis	12
2.1.5 Pertussis in the Netherlands	14
2.2 Modelling Disease Diffusion	17
2.2.1 Agent Based (Individual Based) Modelling	17
2.2.2 Metapopulation modelling	20
2.3 Modelling approaches for Scaling epidemic models	21
2.3.1 Grid based/ distributed modelling	21
2.3.2 Hierarchical modelling	22
2.3.3 Hybrid modelling	26
2.4 Summary	27
3 Analysis of the existing model	29
3.1 The existing pertussis model	29
3.2 Limitations of the existing model	32
3.2.1 Applications to large spatial extent	33
3.2.2 Time extent and recovery	33
3.2.3 Immunity levels	33
3.2.4 Commuting/ movement	33
3.2.5 Hardware requirement issue	34
3.3 Results analysis of the existing model	34
3.3.1 Testing spatial scalability of the model	34
3.3.2 Hardware requirement issue (memory size requirement))	37
3.3.3 General property and performance check	39

3.4	Summary	42
4	Conceptual model	45
4.1	The hierarchical model	45
4.1.1	Spatial and temporal hierarchies	46
4.1.2	The respective hierarchical models	49
4.1.3	Urban system level model	50
4.1.4	Municipality level model	54
4.1.5	Commuting model	58
4.1.6	Disease model	60
4.2	Adjustments to the existing model	62
5	Simulation data preparation	65
5.1	Simulation data preparation	65
5.1.1	Introduction	67
5.1.2	Simulation data	70
6	Model implementation, results and discussion	75
6.1	Implementation tools and language	75
6.2	Implementation stages	76
6.2.1	Stage one: Activating population	76
6.2.2	Stage two: Implementation of the commuting model	79
6.2.3	Stage three: Disease model	80
7	Conclusion and Recommendation	83

LIST OF FIGURES

1.1	Research Work Flow Diagram	6
2.1	spatial diffusion pattern	10
2.2	Potential hierarchical disease diffusion in the Netherlands [14]	11
2.3	Pertussis Analysis in the world	13
2.4	Vaccination coverage in the Netherlands (%), Adopted from: [59]	14
2.5	Diffussion of pertussis in the Netherlands	15
2.6	Map of the Netherlands showing areas of the daily urban system commuting: Adopted from: [62]	16
2.7	Schematic representation of a metapopulation model, Adopted from: [26] . . .	21
2.8	Concept of distributed modeling approach	22
2.9	Concept of Hierarchy theory, Adoupted from: [70]	24
2.10	Types of hierarchy i.e. top-down and bottom-up, Adopted from: [70]	26
2.11	Concept of hybrid modeling approach	27
3.1	The class diagram of the existing pertussis model [1]	30
3.2	Demonstration of individuals attending an activity [1]	31
3.3	Types of Social interactions	31
3.4	Total number of neighborhoods (63 neighborhoods) and their agents (around 70000)	36
3.5	Complete run times (in minutes) of one month and two months	36
3.6	Number of infections of one month and two months	37
3.7	Correlation Between Memory And Number of Population	38
3.8	Number of infections and complete run time (in minutes) of each run for the 63 neighborhoods in five test runs	40
3.9	Average complete run time (in minutes) versus total number of agents	41
3.10	Complete run time (in minutes) versus total number of infections of the individual runs	42
4.1	Spatial and temporal hierarchies and flow of information	46
4.2	Netherlands and its hierarchical levels	48
4.3	Flow of information of the hierarchical	50
4.4	Spatial, Temporal, and Agent characteristics of the hierarchical model	51
4.5	General work flow of the urban system model	53
4.6	Structure of the modeled population	56
4.7	Flow of information between the hierarchies	58
4.8	Concepts for adjustment of the existing disease model	63
4.9	Daily activities of the commutes in the individual level model	64
5.1	Abstract diagram of the model	66
5.2	Twente Population (2009, 2010) per municipality	68
5.3	Number of jobs per municipality	69
5.4	Vaccination data of Twente Region per 5 years	72
5.5	GIS environment used for visualization	73
5.6	Twente Region Centroids (places used as living places for commuter and noncom- muters agents)	74

5.7	Agents of Twente Region per age group and commuter and noncommuters . . .	74
6.1	Stages of the implementation process	76
6.2	UML diagram of the modeled agent groups	77
6.3	Types of agent groups	78
6.4	Result of the model after the agents are activated	79
6.5	Result of the model after implementation of the commuting model	80
6.6	Work flow of the disease model	81

LIST OF TABLES

3.1	Neighborhoods with error on their data and their total number of population. .	35
3.2	Population and the needed amount of memory	38
3.3	Model summary of the performed memory test	38
3.4	Model parameters and the significance values of the performed memory test . .	39
3.5	Total number of created and infected agents in each run for the 63 neighborhoods and their initialization, loading, and complete run time	41
5.1	Twente Population (2009, 2010) per municipality	67
5.2	Number of jobs per municipality	69
5.3	Population per municipality and age group	71

Chapter 1

Introduction

1.1 MOTIVATION AND PROBLEM STATEMENT

Due to its direct consequence on physical and psychological impact of human beings, health is regarded as a very important issue in all nations. One of the main tasks of any government throughout the world is to keep their population healthy. Some of the common actions undertaken to protect spread of disease are control on the import of goods such as seeds and animal products, restrictions on travel of humans and checks on human beings for disease such as tuberculosis and yellow fever. For example, regular x-ray check on the immigrants for tuberculosis disease is carried out by The Netherlands. Despite these efforts, there are numerous reports on epidemics or outbreaks of diseases around the world.

Disease is not limited to geographical boundaries and does not distinguish one society from another. For this reason, diseases are considered as one of the main threats in any country. For instance, easily preventable infectious diseases such as Cholera, Malaria, and Typhoid are the major threats in the developing countries. These societies are suffering from significant socio-economical influence as well as adverse impact of illness on their way towards development. Parts of Central, Eastern and South Eastern Europe, North Africa, and South Asia are the most active areas of pandemic influenza transmission [17]. Generally, epidemics have been accountable for main losses of life and they are still a cause of concern today and in the future [52].

Interventions can reduce the devastating impact of the disease diffusion in society. To perform an effective, efficient and informed intervention preparedness measures are important [54]. Such measures include: raising awareness of the population, creating local stocks of supplies and equipment for diagnosis, strengthening health surveillance systems and practicing protocols for managing information on certain diseases (e.g. developing a computer model for the outbreaks) [54]. Among the above measures, computer modeling or simulation provides theory and understanding of the spread by increasing the level of information in understanding speed of the spread, mechanisms of the spread, and identification of high risk areas [38]. Then, through proper knowledge and concerted action the spread of an epidemic can be controlled/ eliminated [58]. Generally, representing the spread of diseases geographically and analyzing it using different modeling tools and techniques have a great importance [57].

The process of simulating disease infection is complex as it is the result of the cumulative outcomes of human interaction depending on the biophysical mechanisms of the disease spreading over a range of scales in space and time. Human interaction usually includes a variety of movements such as trips to work or shopping or schools, retailing activities, or migration which is difficult to simulate. Accordingly, the spatial spread of disease follows different forms such as: expansion, relocation or combination of both [21]. Expansion refers to the process whereby the disease remains and intensifies in the original region and also expands to the nearby areas over time. In contrast, relocation refers to movement of infected individual from area of origin to new areas [21]. Generally, the transmission of disease is a dynamic process that involves the interaction between individuals both spatially and temporally [72]. Thus, simulating spatial infectious

diseases spread requires detailed and accurate depiction of human interaction and the nature of the disease. One way to deal with such phenomena is to use agent-based modelling (ABM) techniques which is the most widely used modeling approach [10, 29].

Agent-based models are capable of representing very complicated systems that are irreducibly heterogeneous, irreducibly random and contain irreducible interactions [10]. The principle of ABM is to represent explicitly the heterogeneous parts of a system in a computer model, rather than attempting to "coarse grain" [3]. This method enables us to take into consideration temporal and spatial aspects of disease diffusion, individual contact processes as well as their behavior that greatly affects the course of an epidemic outbreak. Furthermore, it represents components of the real system and keeps track of individual behaviors over time [3]. Due to this, agent-based models are widely used to simulate disease diffusion [68, 72].

A number of studies have been done using agent-based modelling to simulate disease diffusion. For instance, O'Neil and Sattenspiel [51] developed an agent-based model to study the 1981-1919 flu epidemics in three small communities in central Manitoba. Yang and Atkinson [72] developed a model called ISTAM for the simulation of airborne infectious-disease transmission by integrating the contact patterns of individuals, an infectious disease process model, and a stochastic infection model. Carpenter [15] also developed a multi-agent-based model to study the relation between Spanish flu and seasonal population movement. Similarly, Perez and Dragicevic [57] developed a multi-agent model to simulate the spread of disease in an urban environment using measles outbreak as a case study. Recently, agent-based simulation for pertussis (whooping cough) has been developed by Abdulkareem [1], for the area of Enschede (Twente region, Netherlands). However, except for the models by Yang and Atkinson [72], Perez and Dragicevic [57], and Abdulkareem [1] the other models are not spatially explicit. Even those models are limited to capture the spatial expansion of disease, since they are individual based, and building complex and large scale models using individual based approach is difficult and computationally expensive. In recent years large-scale computational models known metapopulation models have been used for the simulation of epidemic outbreaks.

Metapopulation theory is based on the idea of a Population of local populations. The approach addresses particular characteristics of populations in patchy environments. Specifically the approach is based on the assumptions that a landscape consists of discrete patches of habitat which interact at some level. This method creates a better insight in understanding the spatial and temporal pattern of infectious disease dynamics in a global scale.

For its advantages many modelers have been using metapopulation modeling to model different phenomena's such as: Colizza et al. [22] investigated the role of the large-scale properties of the airline transportation network in determining the global propagation pattern of emerging diseases. Using the model the authors were able to assess the reliability of numerical forecast with respect to the intrinsic stochastic nature of the disease transmission and travel flows. Bollker and Grenfell [9] use metapopulation modeling to explore the relations between persistence and dynamics in measles epidemics. Summary measures of persistence and triennially achieved from the model allow to compare epidemics in England, New York and Copenhagen. Even though, those models depicts spatiotemporal aspects of disease diffusion, in both models individual contact processes as well as their behavior that greatly affects the course of an epidemic outbreak is not considered. Furthermore, although the metapopulation approach is useful, models should be developed to describe levels of local populations rather to coarse grain events. Hence, it calls for scaling approach which is able to detect the pattern of disease diffusion and elements facilitating it to the required level.

Nicol [48] describes scalability as "How the performance of a certain application behaves as the application problem size increases and the parallel architecture executing it increases."

Pawlaszczyk and Strassburger [56] also demonstrated scalability as a process that considers the architectural as well as the algorithmic side of a simulation system. According to Rana and Stout [60] the issue of scalability arises when there is an increase on the total number of agents involved, size of the data (rules) the agents are operating on, diversity of agents or spatial extent of the model, and the phenomenon which model depicts. To assess whether a given agent-based system scales successfully, certain metrics for measuring scalability can be identified [60]. For instance, a scaling strategy can be evaluated from spatial extent of the model, behavior of the model, or generally from the temporal resolution, spatial resolution, and capability of the model in depicting the modeled phenomena. According to the authors, these metrics can be generalized by the term performance of the simulation.

Performance could take into account the effects of the various subsystems such as processor, memory, disk, software efficiency, and algorithms [46]. This is a high communication-to-computation ratio resulting from interaction between agents participating, maintaining coherence and causality of events, granularity of the application code executed in simulation degrades the model performance. Generally, the case of scalability management is directly related to methodologies for agent analysis and design [60].

Thus, the focus of this study is to build a method for designing scalable agent-based systems with a large number of agents operating concurrently and provide a framework for improving the performance of such systems. The applicability of this method will be tested by implementing it on the agent-based disease diffusion model of Abdulkareem [1]. The study will also investigate the consequences of the change on the conceptual model and the architecture to ensure an adequate performance of the model.

1.2 RESEARCH IDENTIFICATION

This section explains the identification of the research objectives and the corresponding research questions that need to be answered. Also it includes, the aimed innovation and related works of other researchers in tackling similar problems.

1.2.1 Research Objectives

The main objective of this research is to develop a method for scaling agent-based simulations that can be implemented on a pertussis model and improving its performance.

In order to achieve the main objectives the following sub objectives will be addressed:

- To examine how the existing model is functioning
- To identify factors influencing performance of the existing model
- To develop a method for improving the performance of the existing model
- To evaluate the performance and scalability of the extended model

1.2.2 Research Questions

Research questions for the respective specific objectives

1. To examine how the existing model is functioning
 - 1.1 What are the general characteristics of the existing ABM?
 - 1.2 What is the emerging property of the simulated model(output)?

2. To identify factors influencing performance of the existing model
 - 2.1. How can ABM be tested for performance?
 - 2.2. What are the factors influencing the performance of the agent-based models?
3. To develop a method for improving the performance of the existing model
 - 3.1. What are the appropriate approaches to scale the current ABM to enable this model to run for a large number of agents while improving its performance?
 - 3.2. How to create a conceptual model using the selected method (in question 3.1)?
 - 3.3. How to adjust the structure of the code to implement the created conceptual model?
4. To evaluate the performance and scalability of extended model
 - 4.1. How to verify the extended model?
 - 4.2. What are the strengths and weaknesses of the extended model?
 - 4.3. What are the possible recommendations to improve the extended model?

1.2.3 Innovation Aimed At

The main theme and contribution of this research is to develop a method for scaling agent-based simulations of diffusing phenomena. This method will encompass a conceptual framework that shows how to increase the scalability and improve the performance of an agent-based model of a diffusing phenomenon. Generally, the method attempts to explain how to make a model scalable without affecting the performance but efficiently using the resources at hand (without additional resources like memory, CPU, hard disk.)

The developed model will be implemented and tested in an existing pertussis model of Abdulkareem [1].

1.2.4 Related Work

Agent-based models have the advantage that they explicitly represent epidemic processes at the local level, however, they may impose a heavy parametric and computational resources [8]. That means, simulating large number of interacting agents requires tracking and scheduling the interactions which leads to serious computational requirements and analytical challenges in ABMs. To overcome such limitations, many researchers used different types of approaches. For instance: Parker [55] create an agent based model (ABM) of disease transmission in a distributed systems using JAVA programming language. Similarly, Šišlák, et al. [39] (2009) simulate a huge amount of airplanes called AgentFly air-traffic testbed using spatially distributed simulation, partitioning and dynamic load balancing. Also, Xu and McGinnis [71] use hierarchical distributed simulation for 300mm wafer fab (Wafer fabrication) and demonstrated that hierarchical distributed simulation is promising in achieving significantly speedup via massive distributed simulation. However, this approach requires extensive computational experiments.

To reduce the high computational requirement of the distributed approaches researchers have been using alternative methods. For example, Bobashev, et al. [8] introduced a hybrid model that can dramatically save computational times and resource by combining advantages of agent-based and equation-based approaches. More importantly, the output of the authors allow for the mathematical analysis of emerging structures generated by the ABM. Similarly, Ajelli et al. [2] provide a side-by-side comparison between a structured metapopulation model and an agent-based model to assess the epidemic predictions. The results of Ajelli et al. [2] showed that the

possibility of hybrid models by combining the agent-based and the metapopulation approaches according to the available data and computational resources.

Generally, scalability research in multi-agent systems is important in order to have an appropriate methodology for designing complex systems with large number of agents operating concurrently. The designed methodology will provide a design for improving performance of such systems. Consequently, the method will envisage applications employing a large number of agents, such as the design of a commercial computing related to electronic commerce and financial trading, Manufacturing, Financial management, Supply chain management, Disease diffusion modelling where performance cannot be easily predicted in advance.

1.3 PROJECT SET-UP

This section briefly describes how the study attempts to address the formulated research questions. It includes the description of the proposed approach and methods to be used, the work flow that leads from the initial research questions to the final results and conclusions, and the research work plan that gives an overview of the sequence of the research with respect to the allocated time.

1.3.1 Method Adopted

To achieve the predefined objective and to answer the corresponding research questions the study carried out the following phases: Knowledge building, implementation, evaluation, and conclusion phases. Each of them will be briefly explained in the subsequent sections.

Knowledge Building

At this stage relevant literatures are reviewed to get knowledge and understanding of:

- Definition, Nature and characteristics of Epidemics/ disease diffusion (e.g. Pertussis)
- Nature, characteristics and methods for modelling disease diffusion
- Types of modelling approaches (e.g. ABM)
- Approaches for increasing scalability and performance of ABM (e.g. metapopulation modeling, Grid/ distributed, Hybrid, Hierarchical modelling approaches)
- Approaches for testing scalability and performance of an agent based model

From the literature review possible methods used by other researchers for testing scalability and performance, and approaches for scaling and improving performance of agent based models are obtained.

Using the acquired methods for testing the scalability and performance of the existing agent based simulation of pertussis is studied to understand how the model works, emerging properties (output) and factors influencing scaling and performance. In the study of the existing model, knowledge of Java based modeling tool which is Repast Symphony is also required. Furthermore, data of population, vaccination, and commuting were used for testing the model. Generally, this stage by acquiring the relevant knowledge from the literatures, it is optimizing and preparing the model for scaling.

Implementation Phase

Possible approach for scaling and improving the performance were selected from literature review, expert consultation, and then a conceptual design is developed to solve the identified problem. The developed conceptual design describes the possible steps followed in scaling and performance improvement of an agent-based model.

For the implementation of the conceptual model, data is prepared based on:

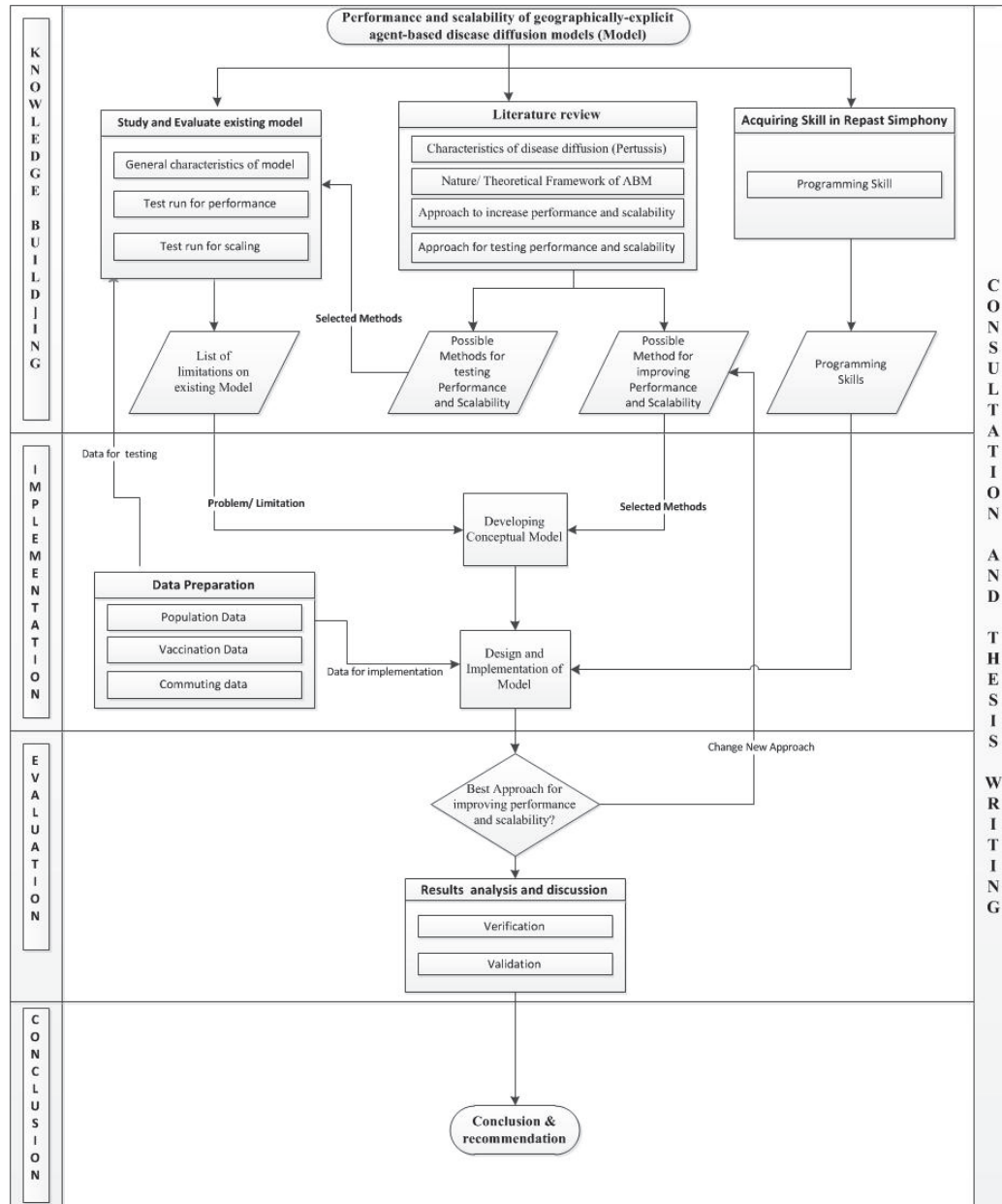


Figure 1.1: Research Work Flow Diagram

- Literature review and other sources to extract the appropriate data for the simulation
- Analyzing the vaccination data
- Analyzing the commuting pattern of the society

Finally, according to the developed conceptual design, the existing pertussis model is adjusted, which helps in analyzing the applicability of the developed conceptual design. For the implementation input data (population, vaccination, and commuting data), the acquired programming skill, and the developed conceptual design were utilized.

Evaluation and Conclusions Phase

Performance of the developed approach is evaluated by identifying the merit and demerits of the new extended model. Then, adjustments were made in the areas where the method was not functioning as intended. Also, validation and verification of the model output is assessed. The flow chart below (Figure 1.1) shows the steps that are followed to develop the method.

1.4 OVERVIEW OF THE FOLLOWING CHAPTERS

From initial description of the concepts and methods to further conclusion and recommendation made, this document consists of seven chapters which are listed below:

Note: chapter one of the research is explained in the above section, therefore, here the description starts from chapter two.

Chapter two addresses the used theories and background of the research with a brief and detailed description of the concepts. Specifically, detailed description of spatial diffusion according to characteristics of a disease (specifically pertussis), different methods/ modeling approaches, and modeling languages are made. Beside to description of the modeling approach in achieving the objectives and sub objectives of the study, a comparison of different modelling approaches such as distributed, hybrid, and hierarchical is included. Moreover using the grasped concepts, some of the research questions are answered and described.

Chapter three encompasses description and analysis of the existing agent based pertussis model which is built to show the spread pattern of pertussis in Enschede, Netherlands. The analysis mainly focuses in checking the performance and scaling bounds of the model. Results of the analysis are later used as guidelines for the creation of the conceptual model for adjustments, performance improvements and scalability of agent based models.

Chapter four provides a detail description of the used conceptual design for the adjustments of the existing pertussis model and creation of the hierarchical model. Specifically, this section contains different conceptual diagrams and mathematical equations which are used in the creation of the existing model. The used concepts, equations and approaches are based on the modeling approach selected in chapter two, results of the analysis in chapter 3 and with general aim of creating a methodology for scaling and increasing the performance of the existing agent-based pertussis model.

Chapter five consists, the data preparation and implementation of the extended model. The data preparation section includes the required data, source and characteristics of the acquired data, steps used in the data preparation and explanation and usage of the created data. On the other hand, the implementation section discusses the implementation actions of the created conceptual model; particularly, the steps and stages of the implementation and the step by step implementation results and visualization of the extended model.

Chapter six presents result, analysis and discussion of the extended or new model. In detail, the outcome of the implementation, correctness of the implemented model, challenges faced dur-

ing/ after implementation and the effects/ influence on the implemented model and validity of the used approached are analyzed and assessed.

Chapter seven is the concentrated description of conclusion and recommendation the implemented hierarchical model. The conclusion presented on this chapter is based on results of the analysis in chapter six. Then based on the conclusion reached recommendations for further works are proposed.

Chapter 2

Background Information

This chapter describes the nature/ characteristics of disease diffusion (Epidemics) which includes concepts such as what are epidemics, how epidemics diffuse, elements/ factors of epidemics diffusion, what is Pertussis (Nature/ characteristics of Pertussis), and nature and characteristics of pertussis in the Netherlands. In addition this concept includes description on modeling disease diffusion; types of modeling approaches (individual based and meta-population) and java based modeling toolkit (Repast Symphony), and methods for scaling increasing performance in an agent-based model.

2.1 CHARACTERISTICS OF DISEASE DIFFUSION (EPIDEMICS)

2.1.1 What Are Epidemics?

An epidemic (outbreak) occurs when an infectious disease spreads rapidly to many people. This diffusion may occur in one community or even extend to several countries. It can last from days to years. Sometimes a single case of a contagious disease is considered an outbreak. This may be true if it is an unknown disease, is new to a community, or has been absent from a population for a long time. For example, in 2003, the severe acute respiratory syndrome (SARS) epidemic took the lives of nearly 800 people worldwide. According to Colizza et al. [24] original source of the disease, social interaction between infected and susceptible individuals, individual mobility, and social interaction and transportation (mobility) are the origin/ factors of disease diffusion.

According to Colizza et al. [24] disease spread depends on occurrences of the disease which is "the place where disease cases are found (i.e. infected individuals) and the locations where the necessary circumstances for causing illness are prevalent". Disease Spread has a temporal and spatial component, which means it spreads from one place to another (spatially) through time (temporally).

Season, time and duration of the diffusion process are considered as the temporal components of disease diffusion. The spatial component of diffusion considers the dispersal and the structure dimension of disease diffusion. The dispersal aspect considers "the routes of commuting" while the structure aspect "refer to the reciprocal relationships between the locations along the route". The dispersal and structural component describe the different disease diffusion process.

The disease diffusion process, is the process by which an infectious disease is carried through space over time [21, 36]. The dispersal component describes either contagious or hierarchical diffusions and structural component describe either expansion or relocation processes [21]. This content is described in section 2.1.2.

2.1.2 Diffusion process of epidemics

According to Cliff [21] disease spread from one place to another in different forms like in expansion, relocation or combinations of both. Relocation diffusion refers to movement of infected individuals from area of origin to new areas [21]. In contrast, expansion refers to the process

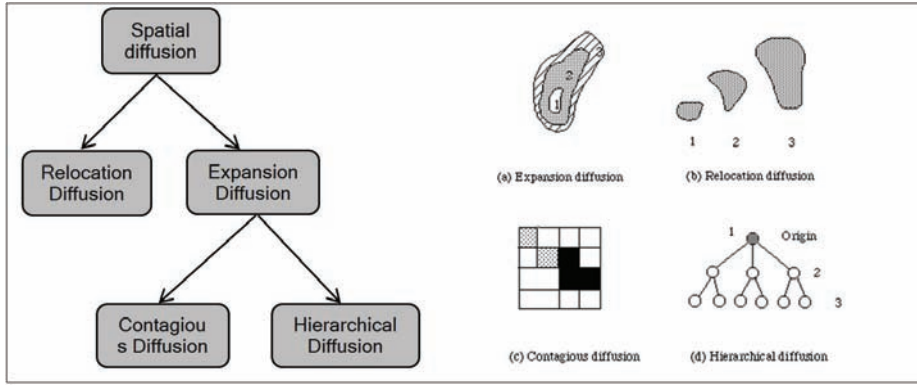


Figure 2.1: Spatial Diffusion Pattern, adapted from: [21]

whereby the disease remains and intensifies in the original region and also expands to the nearby areas over time; mostly the spread of disease is in this form. Moreover, expansion diffusion occurs in contagious and hierarchical manner. Figure 2.1 describes the different forms of spatial spread.

As illustrated in the above figure (figure 2.1(c)) contagious expansion diffusion depends on direct contact between people, the phenomenon spreads out through a uniform medium (centrifugal manner) and it is subject to distance decay that means it is strongly influenced by distance because nearby individuals or regions have a much higher probability of contact than remote individuals or regions; therefore, the further you get from the point of origin the less you are to be infected. Thus in contagious spread diffusion process is slow.

Hierarchical expansion diffusion (figure 2.1(d)) involves transmission through an ordered sequence of classes or places, for example, from large metropolitan centers to remote villages, within socially structured populations, innovations may be adopted first on the upper level of the social hierarchy and then trickle down to the lower levels [21]. With this type of spread, the source of the disease remains and intensifies within the originated area but spreads out to a new location at a later period and fades away faster. Distance and movement are the factors which influence hierarchical spread [20].

According to Cliff et al [20] and Viboud et al [67] the rate of spread is faster if there is a good road network and distance between cities and their hinterlands is small. Hierarchical spread exhibit cascading diffusion. Cascade diffusion is a term reserved for processes that are always assumed to be downwards from larger to smaller centers [20]. When mapped out hierarchical diffusion often produces a series of non-contiguous spots which reflect the spatial expression of a structure.

Example: Pertussis is an air borne infectious disease. This disease spreads from one person to the other and depicts a hierarchical spread pattern. Broutin et al. and Grenfell et al. [13, 14, 35] analyze that the first arrival cases of pertussis starts from the urban areas and then spreads out to the hinterlands. Since, these cities contain large population density as well as spatial extent and social interactions. For instance, in the Netherlands a disease may start from Amsterdam, because it is a big city and every day many travelers travel in and out of the city. Figure 2.2 below shows the possible diffusion pattern.

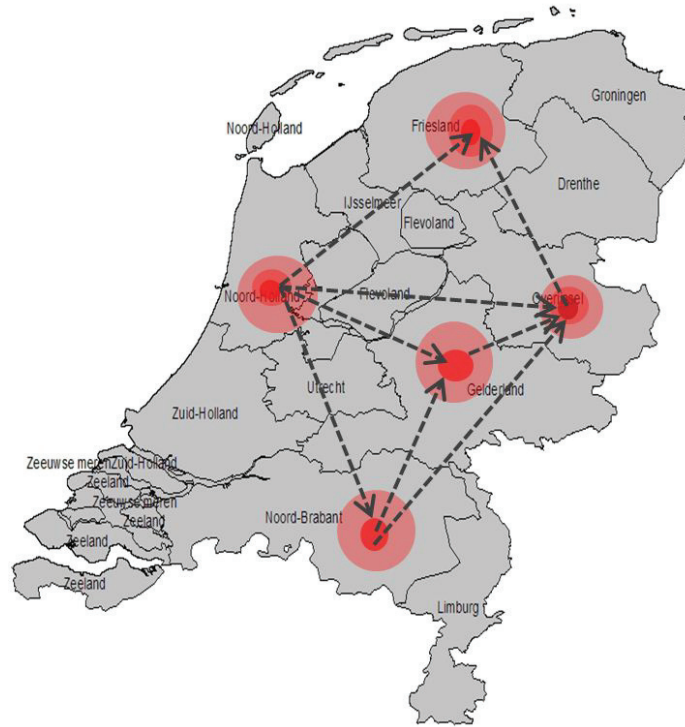


Figure 2.2: Potential hierarchical disease diffusion in the Netherlands [14]

2.1.3 Factors facilitating disease diffusion

2.1.3.1 Commuting

Due to population growth and expansion of cities, urbanization and long distance commuting has increased. This has caused rapid growth of the transportation network leading to an increase in the establishment of businesses, industries and educational institutions. As more attention is drawn to economic infrastructure, less housing facilities are available in the urban areas causing a lot of people to commute daily to work or to school. Regular or repeated traveling/ movements between specific locations such as home, work or school by means of transportation is called commuting. Commuting occurs within space and time.

Time, distance and purpose are the factors that affect the commuting behavior of people. For instance, people become outgoing commuters when they are commuting out of their locality and incoming commuters when they are commuting within their locality (e.g. city or neighborhood). These factors aid to determine the means and the types of commuting. Means of commuting could be through private and public transportation devices; and type of commuting could be regular or random commuting. Regular commuting occurs when the same travel pattern is repeated each day between fixed pair of location, such as For example commuting from home to work or verse vice. Random or irregular commuting occur when the travel pattern is not repeated to a specific location [32].

Commuting is one of the main factors of disease diffusion. This is because commuting creates an opportunity for the interaction of people who live in different areas/ places. This interaction could be between an infected and susceptible people which will then be the reason for disease diffusion [20, 32, 61, 54]. The interaction of people could be for different reasons, for example, to work or socialize.

2.1.3.2 Social Interaction

One of the basic elements which facilitate disease diffusion is social interaction [24]. Social interaction or social network is described as the links or relationship among a group of individuals. Social interaction is based on individual activities which differ according the purpose, population size, time and space. These activities could create individuals within a group or between groups to have casual (face-to-face or conversational) or close (skin-to-skin or physical) contact.

The type and frequency of interaction between individuals is affected by many things such as: culture, religion, age, intimacy and social context, and gender. For example: individual interactions of family members have different format and frequency than interaction between friends. Moreover these interactions are constrained by time (duration) and space (distance between the individuals). When the number of interacting individuals and their interaction increases, it is called Social Networks. Generally, these contacts facilitate the spread of disease [72].

To diffuse different diseases require different levels of interaction. For instance, smallpox requires very close (physical) contact of infected and susceptible individuals. Similarly, infectious sexual disease such as HIV/AIDS requires sexual (physical) contact. In the other hand diseases like pertussis, influenza, and measles could require only a conversational proximity between people since these diseases diffuse through airborne droplets. However, it has been argued that defining social interaction types, which are causes for the diffusion of airborne diseases are harder than that of sexually transmitted diseases. Because, according to the analysis of Eames et al. [33], per day individuals have approximately seven times more conversational contact than physical contact.

2.1.4 Characteristics Of Pertussis

Pertussis, or whooping cough, is an air borne and respiratory tract disease caused by bacteria called *Bordetella* that live in the mouth, nose, and throat. It was isolated in 1906 by Jules Bordet and Octave Gengou in pure culture. A person who is infected by pertussis suffers from coughing spells that last 4-8 weeks [69]. This whoop sound followed by coughing gave the disease the popular name of "whooping cough". The disease is most dangerous in infants and young children, and is easily transmitted from person to person, mainly through droplets produced by coughing or sneezing.

According to world Health Organization [69], pertussis has different stages of symptoms and can be difficult to diagnosis since the early symptoms of whooping cough are similar to the common cold. First symptoms (first stage) of pertussis which is also called catarrhal stage appear 7-10 days after infection. This stage includes mild fever, watery eyes, runny nose, and cough, which in typical cases gradually develop into a paroxysmal cough (second stage) followed by whooping, vomiting, or both. An infected person struggles to breathe after the high-pitched coughing and the person is most contagious during this stage (second stage). Fits may occur on their own or triggered by yawning, stretching, eating, laughing, or yelling; they usually occur in groups, with multiple episodes every hour around the clock. The second stage usually lasts two to eight weeks, or sometimes longer. The proximal stage is then followed by the convalescent stage (third stage). The convalescent stage usually lasts one to two weeks and is marked by decrease in paroxysms of coughing, both in severity and frequency, and a cessation of vomiting.

According to Hethcote and Rie [37], in most countries, the highest incidence rate occurs in young infants. Most of the morbidity, hospitalization and mortality occur in infants. Young infants often present with atypical symptoms, which makes early diagnosis and effective management difficult [37]. Of all infants diagnosed with pertussis and reported to surveillance systems, two-thirds are admitted to a hospital, and some die from pertussis.

The main prevention of pertussis is through DTP3 vaccination which began in 1974 and cur-

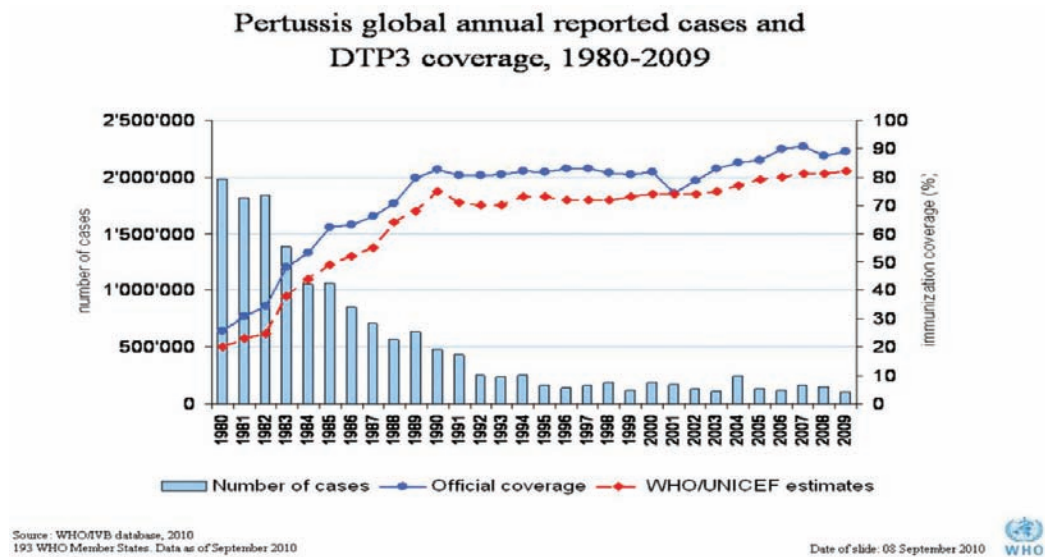


Figure 2.3: Pertussis Analysis in the world

rently spread throughout the world. Furthermore, immunity against pertussis infection can be gained by natural infection or through vaccination but does not last lifelong. Thus, pertussis can occur in previously immunized and infected individuals, but immunization and prior infection decrease the clinical picture. In highly immunized populations most morbidity and nearly all mortality occur in infants under one year, whereas, in unimmunized populations morbidity and mortality also occur in older children and infants. Adults and adolescents could become infected with pertussis but with milder disease (i.e. less than 3 weeks of cough), but play greater role in diffusing the infection to infants and small children [49]. Generally, the clinical severity varies among individuals with different age groups and vaccinated individuals, for example, adults rarely have the classic "whoop sound" [69]. Experts have suggested that young infants will continue to be at risk until adolescents and adults are immunized, thereby reducing the reservoir of pertussis [37].

Prior to the introduction of pertussis vaccination (DTP3), most mortality cases of children occurred in young children [12]. After introduction of the vaccination, the illness and death are sharply decreased. Now the highest incidence, morbidity and mortality are in infants who are too young to be vaccinated. However, more cases have recently been occurring in adolescents and adults. This change in the epidemiology of pertussis is due to the waning immunity that occurs after both disease and vaccination, and to a reduction in natural boosting.

According to world Health Organization [69] thirty per cent of adults with a cough lasting longer than 2 weeks may have pertussis; most infants and young children who contract pertussis are infected by a family member. Worldwide, 30-50 million pertussis cases and about 300,000 deaths occur per year [27, 53]. Figure 2.3 shows world wide spread of pertussis and vaccination starting from the year 1980-2009.

"Recently the level of circulation of pertussis in the adult population is also becoming quite high. For instance, in the yearly incidence of infection in persons over 15 year in Finland was estimated at 248 per 100 000, and in the incidence of infection in German adults was estimated at 133 per 100 000. Furthermore, in Finland it was found that not less than 26% of college students followed during a 2 year period had been infected at least once [12]. This is a potential problem since infected adults may transmit the pathogen to very young children that have not yet been

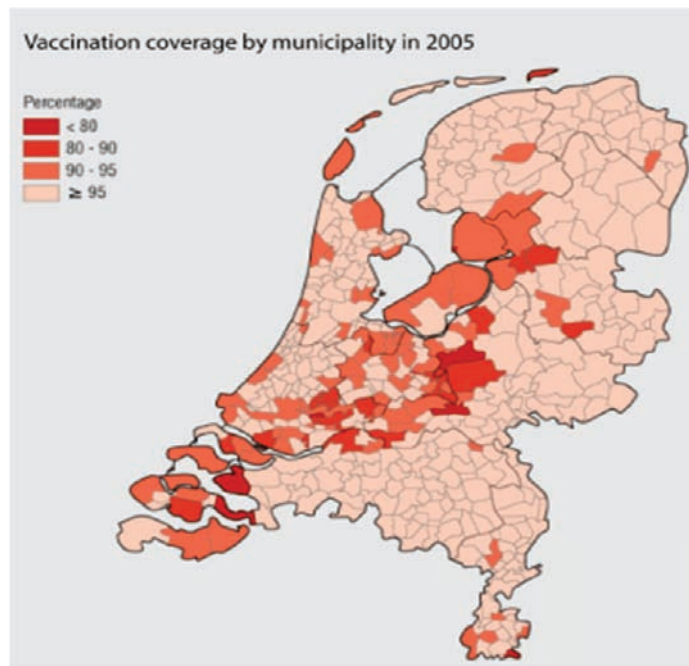


Figure 2.4: Vaccination coverage in the Netherlands (%), Adopted from: [59]

vaccinated, and that are most vulnerable to severe complications" [12].

Diffusion Pattern in Pertussis

Pertussis is a disease that depicts hierarchical diffusion pattern. According to Broutin et.al and Grenfell et al [13, 35] cases of pertussis starts from the urban areas and then spreads out to the hinterlands. To analyze the hierarchical spread observed in the empirical data that occurs during pertussis epidemic years, geographic distance, and population size and density were the parameters that are used. The results depicted that large population size and density showed large cluster and longer stay of the disease which is due to lots of social interaction and commuting within large population size. Whereas less social interaction and commuting occur within small population size causes the disease to fadeout faster.

2.1.5 Pertussis in the Netherlands

In 1996, 2,771 cases of pertussis were reported in the Netherlands (population 15 million), compared with 319 cases in 1995 [12]. Due to the sudden increase in pertussis the incidence was reported as outbreak in the Netherlands. This epidemic outbreak took place until 2004 [30]. The exact factors that contributed to the increased number of cases are not clearly identified [31], but it is stated that the waning immunity among adolescent and adults was one of the factors.

In the Netherlands, vaccination has been introduced since 1953, and currently its coverage has been high (approximately 96%) [59]. Vaccination against pertussis is given in the first year of life starting at 3, 4, 5, and 11 months of age. The intrinsic quality of the Dutch vaccine is constantly measured for each vaccine batch, and satisfies international standards. Nevertheless, the incidence of pertussis in vaccinated infants has increased significantly since 1994. As a result, it is estimated that vaccine efficacy estimated with the screening method shows a decreasing trend [59]. Figure 2.4 show the vaccination coverage of the Netherlands in 2005 [59]. Recently, in the Netherlands the incidence based on notifications in persons over 7 years of age in the vaccinated and unvaccinated populations is comparable. This indicates that protection against infection after

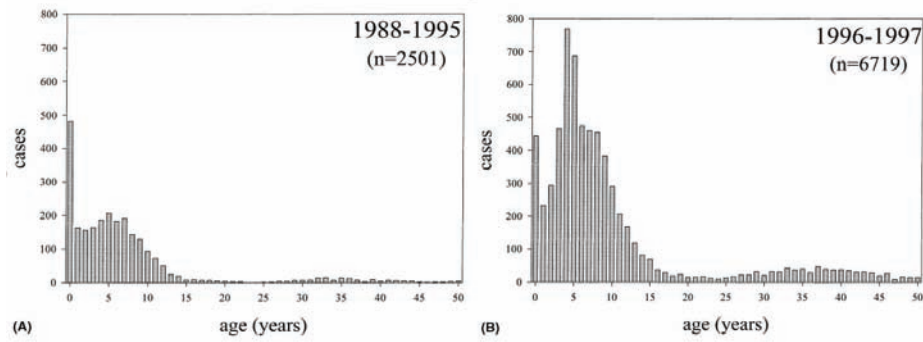


Figure 2.5: Diffusion of pertussis in the Netherlands

vaccination may be rather short-lived, in the order of 5-10 years. Previously, most reported cases of pertussis were in children younger than 15 years of age. To illustrate this, Figure 2.5 shows the observed number of reported cases stratified by age in the 1988-1995 periods, and during the 1996-1997 epidemics. The number of cases is high in (unvaccinated) infants less than 1 year old, it is much lower in (predominantly vaccinated) infants of 1-2 year of age, and it reaches a second peak between 4 and 7 year. Since the age distribution of the Dutch population is relatively uniform (approximately 200000 individuals per cohort of a year), the figure is indicative for the incidence based on notification in The Netherlands. In fact, they early incidence in children younger than 10 year was less than 10 per 100 000 in the pre-epidemic period, and exceeded 100 cases per 100 000 during the 1996-1997 epidemic.

Figure 2.5 shows: Age-specific numbers of pertussis cases in The Netherlands. A shows the distribution from 1988 up to 1995 (before the epidemic), while B shows the age distribution between 1996 and 1997 (during the epidemic) [12].

As described in section 2.1.3 diffusion of disease is dependent in elements/ factor such as commuting behavior of the people. With this in mind, the commuting behavior of the Netherlands is explained below.

2.1.5.1. Commuting Behavior In The Netherlands

Over the past, de-concentration of population and employment from the core of the city to the periphery have been increasing due to several reasons like the use of information and communication technologies, globalization and deregulation. In the process, many urban regions became polycentric [62]. For example the Netherlands has been divided into various provinces, regions and municipalities. Some areas are more urbanized compared to others.

The changes in urban forms reverse the strong concentration of jobs in the city center into urban regions. As a result, daily travel pattern became tangential instead of radial in many metropolitan areas [62]. For example, in the late 1980s, 57% of all the inter-municipal trips taken by the inhabitants of Randstad Holland were between suburban communities. Travel between the central cities and the suburban communities of the Randstad were less frequent: 41% of the inhabitants' trips [28]. In addition to this, due to the small geographical space of the Netherlands outgoing and incoming commuting behavior within urban areas are feasible.

The Netherlands is made up of 16.3 million inhabitants out of which, 5 million people commute daily to work by various means of transport such as private car, bicycle and public transport among others. Over 60% of the Dutch inhabitants commute by private cars, 25% by bicycle, nearly 10% by public transport and the remaining percentage by walking [47]. The commonest means of commuting for less than 3km are by walking and cycling [62].

According to Schwanen, et al. [62], distance and the time of commuting determine the

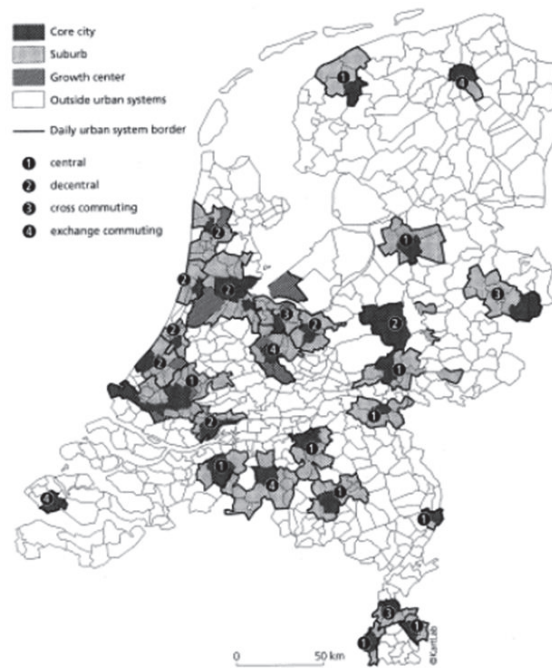


Figure 2.6: Map of the Netherlands showing areas of the daily urban system commuting: Adopted from: [62]

types of commuters that commute. In the Netherlands, young workers between the ages of 25 and 34 commute longer distances and are outgoing commuters. From these young commuters males are the majority as compared to female commuters. In the contrary, older aged workers commute shorter distances because they prefer to work within their locality. On the other hand, higher income and highly educated workers are long distance commuters. The highly educated commuters prefer other means of commuting than private cars [47]. Generally, the age, distance, time and individual activity pattern will determine the commuting behavior of commuters.

The individual activity pattern such as activities at home or at work determines the time a person leaves or arrives at home [62]. According to this author, the daily activity pattern of a Dutch worker contains five periods. Within these periods two levels of activity patterns occurred. The first level of activities represent bundles of activities that occur in the home, at the work site and the time window which is "the block of time that start when a person leaves either his/her home or work place and arrive at the same or another base location". The second level differentiates other activities and the travel period within the time window. From Schwann's [62] analysis 70% of Dutch population, are full-time workers, who spend on the average 8.5 hours at work and 30% are part time workers. Schwann's [62] analysis is illustrated in figure 2.6.

According to Schwanen et al [62], more women commute by public transport as compared to men, since women are short distance commuters due to their responsibility to keep their homes. However, as compared to men, women interact more with other people such as family members and friends. This implies that women are more infectious and susceptible to infectious diseases during commuting and socializing, as the main sources of infections are commuting and social interaction.

2.2 MODELLING DISEASE DIFFUSION

Imitation of some real processes, state of affairs, or thing in the real world is called modeling [10]. Modelling generally entails representing certain key characteristics or behaviors of a selected physical or abstract system. It is a powerful method for understanding and predicting of both simple and complex phenomenon. Models can be used to get answers for questions, such as 'What is the cause of a certain phenomenon?', 'Which areas or population are affected by the phenomenon?', 'How to control that phenomenon?'. Generally models are built to inform the population and/or policy makers in the provision and understanding of a certain phenomenon. Today sophisticated computer models are used [3].

A simulation refers to a computerized version of a model which is run over time to study the implications of the defined interactions. Simulations are generally iterative in their development, and they have an advantage in representing and understanding the complex social structure and the heterogeneous patterns in the contact networks of real-world populations determining the transmission dynamics. One develops a model, simulates it, learns from the simulation, revises the model, and continues the iterations until an adequate level of understanding is developed. Simulation models have become an important information tools to study real world phenomenon, such as epidemiology, physics, engineering, education, industry, video games.

Over the years, many epidemiological scientists developed epidemiologic models since epidemics have direct impact on human beings. Epidemiological models provide tools and principles to represent better understanding and perspective to such issues. Moreover, the better understood the diffusion of disease the more efficient the reaction becomes [19]. These models are complex because to efficiently model the outbreak they need to involve contents such as: ever increasing level of interconnectedness, globalization, society with a high level of diversity and heterogeneity. As a result, a modeling approach that is capable of tracing such complexity and detail human and environmental structure is requires.

The geotemporal spreading pattern found by the different modeling approaches may differ, but to select the best modeling approach for a phenomenon it all depends on the different approximations and assumptions used ([2]. Over the years, epidemics have been modeled using different modeling approaches such as: very detailed agent-based models, meta-population models. Meta-population model consistently yields a larger incidence than the agent-based model, as expected due to the differences in their structure [10].

2.2.1 Agent Based (Individual Based) Modelling

Agent-based modeling is a powerful modeling approach that has been applied in a number of applications in the last few years, including complex applications like real-world business problems, or epidemics [10]. Agent-based models are capable of representing very complicated systems that are irreducibly heterogeneous, irreducibly random and contain irreducible interactions [10]. The principle of ABM is to represent explicitly the heterogeneous parts of a system in a computer model, rather than attempting to "coarse grain" [3].

This method enables us to take into consideration temporal and spatial aspects of disease diffusion, individual contact processes as well as their behavior that greatly affects the course of an epidemic outbreak. Furthermore, it represents components of the real system and keeps track of individual behaviors over time [3]. According to Barnes and Chu; Castle et al. [3, 16] agent-based modeling have the following advantages: (i) ABM captures emergent phenomena; (ii) ABM provides natural description of system and (iii) ABM is flexible.

1. ABM captures emergent phenomena: Interactions of individual entities generate growing phenomena. By definition, they cannot be reduced to the system's parts: the whole is more

than the sum of its parts because of the interactions between the parts. For example, the spread pattern of a disease is the interaction of the infected individual with other people who may be susceptible to contract the disease. The interactions of individual enhance the spread pattern of the disease. One may want to use ABM when there is potential for emergent phenomena, i.e., when:

- Individual behavior is nonlinear and can be characterized by thresholds, if-then rules, or nonlinear coupling. Describing discontinuity in individual behavior is difficult with differential equations.
 - Individual behavior exhibits memory, path-dependence, and hysteresis (There is no way to predict the system's output without looking at the history of the input), or temporal correlations, including learning and adaptation.
 - Agent interactions are heterogeneous and can generate network effects.
 - Averages will not work. Aggregate differential equations tend to smooth out fluctuations, not ABM, which is important because under certain conditions, fluctuations can be amplified: the system is linearly stable but unstable to larger perturbations.
2. ABM provides a natural description of a system: Based on a set of rules ABM has the ability to simulate individual behavior, makes it possible to mimic natural systems, and makes the model seem closer to reality. For example, it is more natural to describe how people move and interact in a place than to come up with the equations that govern the dynamics of the density of the individuals. One may want to use ABM when describing the system from the perspective of its constituent units' activities is more natural, i.e., when:
- The behavior of individuals cannot be clearly defined through aggregate transition rates.
 - Individual behavior is complex. Everything can be done with equations, in principle, but the complexity of differential equations increases exponentially as the complexity of behavior increases. Describing complex individual behavior with equations becomes intractable.
 - Activities are a more natural way of describing the system than processes.
 - Stochasticity applies to the agents' behavior. With ABM, sources of randomness are applied to the right places as opposed to a noise term added more or less arbitrarily to an aggregate equation.
3. ABS is flexible to work with - ABS has the ability to tune the dynamic nature of individuals' behavior in relation to changes in reality by changing the characteristics of individuals in. It is easy to add more agents to an agent-based model. Another dimension of flexibility is the ability to change levels of description and aggregation: one can easily create aggregate agents, subgroups of agents, and single agents, with different levels of description coexisting in a given model. One may want to use ABM when the appropriate level of description or complexity is not known ahead of time and finding it requires some tinkering.

The components of an ABM are set of agents, environment, time, the interactions between agents, and the interaction between agents and their environment.

Agents

An agent is "a system situated within and a part of an environment that senses the environment and acts on it over a period of time, in the pursuit of its own agenda and so as to affect what

it senses in the future". Agents are the fundamental part of any ABM, representing the entities that act in the world being modeled. These agents are the central units of the model and their aggregate behavior will determine the outcome of the model. Agent behavior is based on a set of rules. In ABM a system contains collection of autonomous decision making entities called agents. According to [16] agents have many properties such as:

- **Autonomy:** are self-governing which are driven to achieve certain goal. Therefore, there is no top down control over these agents.
- **Goal-directed** - agents have set of goals to be achieved.
- **Agents interact with their environment as well as with other agents.** An agent is situated, in the sense that its behavior is situationally dependent, which means that its behavior is based on the current state of its interactions with other agents and with the environment.
- **Reactivity:** agents are capable of discovering their environment and responding to it.
- **Social ability/ Interactivity:** agents use computer language rather than natural language to communicate with other agents. In addition, they have the ability to interact with each other.
- **Pro-activity:** in a goal directed behaviors, agents are capable of taking a lead.
- **Heterogeneity:** the concept of mean-individuals is redundant: agents permit the development of autonomous individuals, but also groups of agents can exist. However, they are spawned in bottom-up, collections of similar autonomous individuals.
- **Mobility:** this behavior of agents indicates that agents can move from one place to the other which is a very important feature of agents.
- **Learning:** agents can be designed to learn from their environment at individual level or at a population level (collection of agents). They also have motive, belief, desires and motions.
- **Explicit space:** events occur in explicit/ specific space, which could be dynamic social network, or GIS.

Environment

Agents must be embedded in some type of environment, that is, a space in which they exist. The choice of the environment can have important effects on the results of the simulation runs, but also on the computational requirements of the model; how to represent the environment and how much detail to include will always be a case specific issue that requires a lot of pragmatism. Environment can be geographical explicit which makes the agent have a geographical location or spatially implicit which make the location of the agent not relevant in the simulation model. An environment can be either static or dynamic. Within a static environment, no changes occur when agents interact with the environment while as the dynamic environment changes occur even if the agents do not interact with the environment. Multiple environments can be modeled which may lead to interaction between environments which may influence the behavior of agents within dose environments. In ABS the environment is the world where agents act [3].

Time

Agent-based simulation models are driven by time. Time can be continuous, discrete and discrete event. Continuous time changes occur continuously across time. Discrete time is split into regular time step while discrete event checks the process when change will occur [3].

Interactions

Interaction of agents is the fundamental concept of ABS. Interaction in ABS is expressed by behavioral rules which consist of actions and conditions. The specific action (behavior) describes what should be done and when the specific condition actions should be fired. Interactions can be continuous or discrete. Interaction can occur among agent-agent, agent-environment and environment-environment [3].

The key to developing useful agent-based model are identifying agents, accurately specifying their behaviors, and clearly assigning their environment [10]. One begins developing an agent-based model by identifying the agent types (classes) along with their attributes. Agents are generally the decision-makers in a system. Once the agents are defined, agent behaviors are specified based on our theory of agent behavior for modeling the agent behavior. Finally, appropriately representing agent interactions is also basic feature. Agents have repetitive interactions with each other.

The interaction between agents is complex, and requires methods that control which agents interact, when they interact, and how they interact. These interactions relay on the power of computers to explore the dynamics. Even the simplest ABM consists of a system of agents and their relationships, that is, a simple agent-based model can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system that it emulates. In addition, agents may be capable of evolving, allowing unanticipated behaviors to emerge. A synonym of ABM would be microscopic modeling or individual based modeling. Due to this, agent-based models are widely used to simulate disease diffusion [68, 72].

A number of studies have been done using agent-based modelling to simulate disease diffusion. For instance, O'Neil and Sattenspiel [51] developed an agent-based model to study the 1981-1919 flu epidemics in three small communities in central Manitoba. Yang and Atkinson [72] developed a model called ISTAM for the simulation of airborne infectious-disease transmission by integrating the contact patterns of individuals, an infectious disease process model, and a stochastic infection model. Carpenter [15] also developed a multi-agent-based model to study the relation between Spanish flu and seasonal population movement. Similarly, [57] developed a multi-agent model to simulate the spread of disease in an urban environment using measles outbreak as a case study. Recently, agent-based simulation for pertussis (whooping cough) has been developed by Abdulkareem [1], for the area of Enschede (Twente region, Netherlands) . However, except for the models by Yang and Atkinson [72], Perez and Dragicevic [57], and Abdulkareem [1] the other models are not spatially explicit, also those models are limited to capture the spatial diffusion pattern of an outbreak. Thus, a modelling approach that is capable of capturing the complex interaction is required.

2.2.2 Metapopulation modelling

The metapopulation modeling approach is an essential theoretical framework in the understanding of the large scale spreading of epidemics. Metapopulation models rely on the basic assumption that all individuals form a single, homogeneously mixed population in which the population is structured and localized in relatively isolated discrete patches or sub-populations. This modeling approach explicitly model the spatial structured population as a system of loosely coupled sub-populations (or patches), whereby homogeneous mixing is assumed to be a reasonable approximation within sub-populations and well defined patches connected through individual mobility [23, 25, 26, 4].

Agents within each subpopulation are distributed into classes such as susceptible, infected, recovered etc. representing their state with respect to the modeled disease and possibility that individuals in the same location may get into contact and change their state according to the

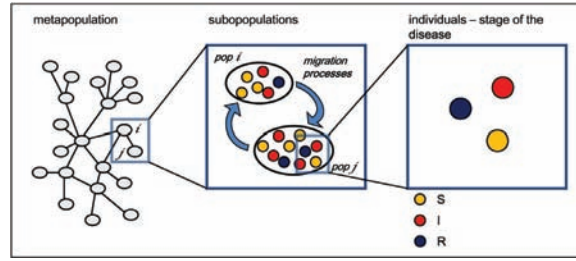


Figure 2.7: Schematic representation of a metapopulation model, Adopted from: [26]

infection dynamics. Basically, the main issue in metapopulation modeling approach is to what extent the commuting pattern of agents is described accurately.

Figure 2.7 illustrates a collection of subpopulation or patches connected by individual commuting. Each patch contains a population of agents who are categorized with respect to their stage of the disease (S=susceptible, I=infected, R=recovered). Agents commute from a subpopulation to other subpopulations.

As compared to agent-based models, structured meta-population models are fairly scalable and can be conveniently used to provide large scale scenarios and diffusion patterns. While on the one hand, the level of information that can be extracted using this method is less detailed than those of agent based models, the spatial and temporal ranges and the number of realizations that can be computationally analyzed is much larger. Moreover, the amount of data to be integrated is less massive than in agent-based frameworks (see section 4.3.2 for detail description). However, the choice of a method depends on the amount of detailed needed for the phenomenon to be modeled. For this reason, there is a need for a modelling approach that is capable of capturing the complex interaction in a more detailed manner.

2.3 MODELLING APPROACHES FOR SCALING EPIDEMIC MODELS

Epidemic modeling contains large number of interacting agents also known as multi-agent systems because they contain large (short and long distance) and complex interactions of different components. To simulate such applications require huge resource and a sophisticated modeling approach, and the study of epidemics could be facilitated, if there exists a method that efficiently and effectively model epidemics/ outbreaks [68, 72]. Therefore, to support the execution of such applications without affecting the performance and scaling of the model several modeling approaches/ methods have been emerged and used. For example, grid based /distributed modeling, hybrid modeling and hierarchical modeling. The application, advantage, disadvantage, and the method that is selected for this research will be discussed in the proceeding sections.

2.3.1 Grid based/ distributed modelling

Simulating large scale epidemics requires the availability of sufficient computing power and a method which is capable of utilizing these resources. A Grid based modelling offers a robust distributed computing infrastructure needed for simulations of these type [46]. According to Mengistu, et al. [46] Grid based modeling can offer the proper execution environment for multi agent based system (MABS) applications for the following reasons.

1. As explained in section: 1.1. the computational resources needed for epidemic simulations are so large thus they cannot run in small systems effectively. These simulations require



Figure 2.8: Concept of distributed modeling approach

agents (thousands or even millions of agents) to perform highly in complex and data intensive tasks.

2. Since multi agent based systems grew out of distributed artificial intelligence, a distributed computing platform like the Grid based modeling serves as a natural environment to plan, develop and implement agent based simulations (see figure 2.8).

If application of the multi-agent system is installed on a stand-alone machine, all tasks (threads) run on that same machine, and then communication is essentially data movement within the same physical memory. However, in Grid based modelling, the communication involves transfer of data over the network, because the threads are launched on separate machines. Furthermore, this modeling approaches involves partitioning the simulation into equivalent tasks, with each task to be launched on a node. The relations in the physical entities involve transferring data within the same machine (inbound messaging) or between nodes (outbound messaging). The terms inbound and outbound refer to the destination of the messages with respect to the location of the sender. Thus, communication is of primary interest and thus should be well defined in the Grid based model [46].

Generally, the Grid is an architecture that is based on data distribution and computation, contains several computers that are networked, divides the task into equivalent tasks that will later be launched on a node, message transferring is inbound or outbound. It has the advantage of solving larger problems and executing applications that are distributed in nature, scalability, enables coordinated resource sharing, and allows resource negotiation and dynamical allocation [18]. Thus Grid environments can be used both for computer intensive tasks and data intensive applications as they offer resources, services, and data access mechanisms [73].

However, this approach is challenged by the complexity of dividing a big problem in to small separate tasks, characterization, discovery, monitoring of resources and services, and computations. This challenge is due to the considerable diversity, large numbers, dynamic behavior; and geographical distribution of the entities.

2.3.2 Hierarchical modelling

The involvement or combination of large number of human activities, interactions, behaviors, and different components and elements make epidemics modeling complex. Hence, requires an

effective and efficient method to depict and breakdown such complexity. Hierarchical modeling approach is among those methods which are used in complex systems. Thus the section below describes the theory and concept behind hierarchical modeling approach.

2.3.2.1. Hierarchy theory

Epidemic modelling is complex since it contains a large number of human interactions, emergent behavior and activities, and movements that cover large areas as explained in section: 1.1. Emergence, thus, can be understood as the property of complex systems where large interactive systems naturally evolve toward a self-organized critical state in which a minor event can lead to a cascading catastrophe. Thus, emergence is the central point of the theory of complexity. In addition, complexity often results from the nonlinear interactions among a large number of system components which frequently lead to emergent properties, unexpected dynamics, and characteristics of self-organization [42]. Levin [42] argued that heterogeneity, nonlinearity, hierarchical organization, and flows are four key elements of complex systems.

In addition, complexity takes the form of hierarchy that means complex systems are usually hierarchically organized [42, 63, 70]. According to the authors, the construction of a complex system using a hierarchical approach is likely to be more successful. Simon [63] also argued that, hierarchic systems have some common properties that are independent of their specific content. In other words, hierarchy is a central structural scheme of the architecture of complexity, and often manifests itself in the form of modularity in nature.

According to hierarchy theory, complex systems have both a vertical structure that is composed of levels and a horizontal structure that consists of holons [70]. Higher levels are characterized by larger entities whereas lower levels are composed of smaller entities. Generally, the relationship between two adjacent levels is asymmetric: the upper level exerts constraints (e.g. as boundary conditions) to the lower level, whereas the lower provides initiating conditions to the upper. On the other hand, the relationship between subsystems (holons) at each level is relatively symmetric in that they interact in both directions [63, 70]. Furthermore, at each level, holons are defined in such a manner that the interactions between their components are significantly more frequent within than between holons. For example, an epidemic spreading in a country can be described by the spread of the disease between the regions at the higher level and the spread of the disease from individual to individual at the lower level, with stronger interactions between individuals within the same city than between individuals found in different cities, and so on.

Hierarchy does not always refer to a system that contains top-down interaction. Certainly, hierarchy theory emphasizes both top-down and bottom-up perspectives approach [70].

The central idea of the hierarchy theory is to derive the hierarchical organization from differences in temporal and spatial scales between the phenomena of interest. And provide a framework that allows for handling complex systems (see figure 2.9).

2.3.2.2. How To Determine Hierarchies

Over the years, increasingly many researchers' biologists, social scientists, and physical scientists raise the word "complexity" to describe their orientation to the problems on which they work [40, 63, 70]. According to these authors, there are four different kinds of frameworks for handling complexity and organizing complex system in to smaller and manageable parts. These frameworks are: order hierarchy, inclusion hierarchy, control hierarchy and level hierarchy.

1. Order Hierarchy

Order hierarchy considers a hierarchy to be equivalent to an ordering by the values of a variable defined on some set of element. According to Batty and Lane [40, 5] hierarchy is

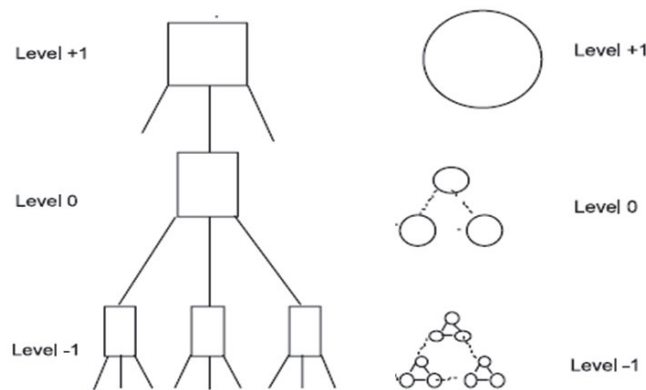


Figure 2.9: Concept of Hierarchy theory, Adoupted from: [70]

a natural ordering that is initially based on size but size can be measured in many different ways. For example, in cities, size is typically based on the number of individuals or households or workers. In populations it may also be based on the area over which such location occurs or energy is used, or the field of influence over which individuals in the hierarchy have control.

To the other types of hierarchy discussed below, order hierarchy does not refer to relationships and interactions among the entities that comprise the hierarchy, much less give any role to hierarchy in conditioning entity relationships and interaction structures [40, 5]. Even though order is essential to hierarchy, order alone is not what makes hierarchy important for complex systems.

2. Inclusion Hierarchy

Inclusion hierarchy is used to refer to a recursive organization of entities in which an entity is a container, and what it contains are other entities [40]. According to the author, this notion is just a special case of an order hierarchy, where the ordering variable is the number of entities one opens before arriving at the particular place of interest. For example, in an inclusion hierarchy an urban system is a collection of cities, which in turn contain firms, households, institutions and organizations.

Simon [63] also explains inclusion hierarchy as a set of Chinese boxes of a particular kind. A set of Chinese boxes usually consists of a box enclosing a second box, which, in turn, encloses a third box. The Chinese boxes called 'hierarchies' are a variant of that pattern. Opening any given box in a hierarchy discloses not just one new box within, but a whole small set of boxes; and opening any one of these component boxes discloses a new set in turn.

3. Control Hierarchy

In reference to social organization, hierarchy refers to a control system in which every entity has an assigned rank, and all power is concentrated in the (usually single) entity with the highest rank. Entities with a specified rank are entitled to give orders to entities with lower ranks, while they must obey orders received from elements with a higher rank [40]. In control hierarchy, orders flow rank-downwards; while information and requests rank upwards. For example, a church, political party, or army can be organized as a control

hierarchy. According to Lane [40], entities that comprise a control hierarchy do not form an inclusion hierarchy.

4. Level Hierarchy

Many authors view hierarchy as entities with a particular kind of ontological organization, in which entities are posited to exist at different "levels". Each level is characterized by a particular spatio-temporal scale for its associated entities and for the processes through which the entities at this level interact with one another. The higher the level, the more extended the associated spatio-temporal scale [40, 63].

According to those authors, entities at a given level may construct and maintain entities at higher levels through their interaction, and higher-level entities may be, at least in part, composed of lower level entities: these are often described by the term upward causation. Through upward causations, this type of hierarchy may form inclusion hierarchies. However, level hierarchies are also characterized by downward causation. Incorporation into a higher level entity can change the properties and interaction modalities of lower-level entities.

For the rest of this paper, level hierarchy is selected for creating a spatiotemporal hierarchy of disease diffusion. This is because from the above mentioned types of hierarchies, level hierarchy is based on structure, process and interaction of a system. In addition, it includes the structures of order and inclusion hierarchies.

To have a complete understanding of the complex structure of epidemics, its way of transmission, its spatial extent, and prevention mechanisms, a hierarchical modeling approach is required. This approach is capable of reasonably and accurately depicting the system, its hierarchies and its interactions.

2.3.2.3. Hierarchical modelling approach

Hierarchical modeling approach is a way of discovering or rendering order and breaking down complexity. This modeling method emphasizes both top-down and bottom-up perspectives [70]. The created hierarchies are based on the theory of level hierarchy. In addition, the hierarchical model has a set of functional positions which describes a hierarchical structure (for example, higher level which is country level activity will not be individual level). This structure is used to coordinate the activities of the agents, communication pattern etc. These hierarchical organizations could be of different types temporal, spatial or agent characteristics (behavior) (see figure 2.10).

The first step in the hierarchical approach is to derive the hierarchical organization from differences in temporal and spatial scales between the phenomena of interest. This means, to identify the appropriate spatial and temporal hierarchies, or to decompose the complex system in to smaller parts. This decomposition could be both in top-down, bottom-up or both schemes. A top-down approach identifies levels and holons by progressively partitioning the entire system downscale, whereas a bottom-up scheme involves successively aggregating or grouping similar entities upscale.

For example, to analyze how pertussis spreads in a country, different spatial hierarchies could be created. These are hierarchies based on the size and number of population in an area. Such as, from high scale (urban systems) to its equal scale (urban system to urban systems), from high scale (urban systems) to low scale (from urban systems to their sub-populations (cities)), or vice versa [43].

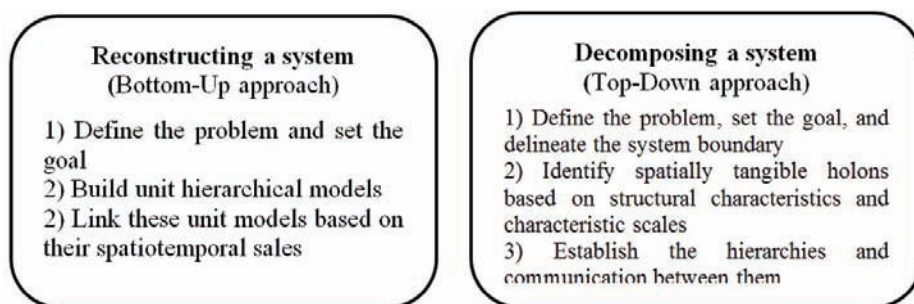


Figure 2.10: Types of hierarchy i.e. top-down and bottom-up, Adopted from: [70]

Once an appropriate hierarchy is established, processes can be studied at focal levels (corresponding to characteristic domains of scale and time), by properly choosing size (interval or spatial resolution) and extent (study duration or area). A process-relevant spatial/ temporal hierarchy effectively groups these factors into relatively separate regions according to their characteristic scales in space and time. Since, the phenomena of interest are only observable at the appropriate scale of observation. Generally, the focuses of this stage is on making observations and developing models at focal levels [70].

As defined above, a level organization is composed of interacting components. Some of the components interact weakly and others strongly. For this reason, after the hierarchical levels and their organization are created, a communication between the interacting components is constructed from spatial and temporal aspects; for example, levels which are closer to each other interact frequently than levels that are far away from each other. The communication/ interaction between levels could be flow of information or process, and the output of one level could be an input to the other level. In general, an appropriate hierarchical, dynamic structure provides opportunities for diversity, flexibility, and creativity. Also it provides higher efficiency and stability that are difficult to obtain in non-hierarchical complex systems. Furthermore, hierarchical approach have the advantage of minimizing execution and searching time, keep tasks in ordered format, gives conceptual structure of how agents are inserted in to the tree, reducing in model execution time. Contrary to its advantages, this modeling approach requires detailed knowledge of the model hierarchy and knowledge of appropriate type of communication between the hierarchies.

However, as compared to the other modeling approaches (for example, Grid/ Distributed based) hierarchical architecture tends to evolve faster, allow for more stability [70]. Thus it is selected as a modeling approach for decomposing the complexity of the study and creating hierarchies for this research.

2.3.3 Hybrid modelling

Hybrid modeling is a modeling approach that combines the advantages of two modeling paradigms and introduces a hybrid model that starts as one modeling approach and switches to the other approach after a certain condition is reached [8]. Because, in complex systems (multi-agent systems) the joint behavior of the agents can be very complex and tracking the behavior requires a disciplined approach. Also inadequate representation of the detailed population structure can lead to spurious results, especially when the simulation process is beginning and individual variation is critical [8]. For example, the model of Bobashev, et al. [8] start as agent-based and switches to equation-based after the number of infected individuals is large enough to support a population-averaged approach.

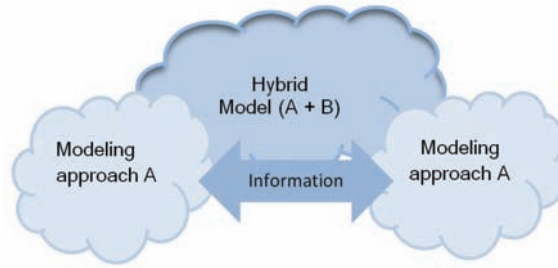


Figure 2.11: Concept of hybrid modeling approach

According to Bobashev, et al. [8] "The concept of the hybrid threshold model originates from the premise that when the number of active agents is large, the law of large numbers and central limit theorem could be applicable to the observations. Thus, it should be possible to aggregate the behavior of similar agents and to model their behavior through mean-field approximations".

Hybrid modeling approach is advantageous because it enables modelers to avoid at some defined point the disadvantages of one modeling approach and use the advantages of another approach, to dramatically save computational times. For instance, the model of Bobashev, et al. [8] have the advantage of saving computational time and, more fundamentally, allows for the mathematical analysis of emerging structures generated by the ABM.

However, according to Lättilä et al. [41] hybrid modeling approach could be disadvantageous because implementation of two modelling approaches in to one problem is complex. This is because, it is important to know the limitations and strengths of both methods, using output of one modeling approach that is not validated as an input to the other approach may lead in to a fault model result. Furthermore, translating the mental model to a simple logic might be extremely difficult as there are many different factors impacting the actual decisions, and there may not be specialized program for automating the work thus it could require relatively heavy investments in time and skills.

Yet, despite its disadvantages a proper hybrid modeling approach provides opportunities for diversity and flexibility. Furthermore, it could make the model more robust, flexible and scalable. Therefore, in this research a hierarchy based hybrid modeling approach which combines the advantages of agent based, metapopulation and hierarchical modeling approaches is selected and used. For the implementation java based modeling approach called Repast J is used (see figure 2.11).

2.4 SUMMARY

This chapter highlights the used theories, backgrounds and concepts such as meaning and diffusion process of epidemics specifically spatial disease diffusion according to the characteristics of diseases (specifically pertussis), elements which facilitate disease diffusion, and ways of modeling disease diffusion.

Furthermore, the chapter provides in-depth knowledge on the effective approaches for modeling disease diffusion in relation to computer simulation modeling. Computer simulation modeling delivers tools and principles to represent better understanding and perspective of such phenomena. However, simulating disease diffusion requires an effective modeling approach which is capable of depicting the modeled phenomenon to the required level.

Agent-based modeling techniques belong to the category of individual based approaches; and therefore are very suitable for simulating smaller geographic areas with relatively few agents.

Metapopulation modeling is a mathematical modeling approach which assumes homogeneous mixing or distribution of population and is capable of modeling very large areas with large number of populations. However, have limitations in not including detailed structure and behavior of population which could influence characteristics of the disease diffusion. Hierarchical modeling approach groups the population according to different assumptions such as age, behavior, geographical location, and socioeconomic position. The approach mainly focuses on discovering or rendering order and breaking down complexity. Yet it requires highly detailed knowledge of the structure of the phenomena to be modeled.

With this in mind for this research a hierarchy based hybrid modeling approach which combines advantages of agent based, metapopulation, and hierarchical modeling approaches is selected. For the implementation java based repast modeling approach is selected since it is powerful modeling approach. However, before introducing and describing the prepared conceptual model (chapter four), chapter three reviews and analyzes the existing agent based pertussis model which is used as a bases for the preparation and implementation of the conceptual model.

Chapter 3

Analysis of the existing model

The section contains description and analysis of an agent based pertussis model. The model is built to show the spread pattern of pertussis in Enschede, Netherlands. The analysis is mainly directed in checking the needed adjustments, performance and scalability of this model. Result of the analysis will later be used in creating a conceptual model for scaling and improving performance of agent based models.

3.1 THE EXISTING PERTUSSIS MODEL

An agent-based model of pertussis in Enschede area have been developed by Abdulkareem [1]. The model is based on the Individual Space-Time Activity-based Model (ISTAM) which is an agent-based disease model developed by Yang and Atkinson [72]. ISTAM is a model which takes into account the individual activities among human activities bundles. Activity bundle (AB) simulation is a method for obtaining a specific contact network of individuals from the space time dynamics of individuals constrained both by their social activity and by the physical condition of the space [72]. Using AB for modeling disease diffusion makes a model to be flexible. This is because activities bundles capture each individual's space time dynamics at fine scale of space and time.

The model by Abdulkareem [1] simulates the social interaction of individuals that causes the spread of pertussis in Enschede which is a city in the Netherlands. The synthetic population of the model has been generated by using synthetic reproduction approach and Enschede census data. To create and assign agents, the Enschede census data which is used by the synthetic reconstruction approach contains data of families and their children. In the simulation the first element that is created is the Household heads per neighborhood. Subsequently, partners of the household heads if exist and then children are created. As agents are created their attributes are assigned. Individuals/ agents differ from each other Because of some of their attributes (for example, by gender, work status, age etc.). The attribute assist to investigate how individual contract and spread the disease. The class diagram in figure 3.1 shows the attributes of an agent. The city environment (which includes GIS layers) and agents (which have home address and activity location) are loaded and created as a fourth stage. Then these agents perform their activities based on the activities listed on the activity table. Additionally, agents look around to find their friends and the group where they fit in. For modeling different tasks of agents the existing pertussis model consists of three main models which are the daily activity, social interaction and disease models.

Daily Activity Model

The daily activity model, models both the individual and group activities. Each individual performs an activity within a specific time of the simulation (figure 3.2 shows example of an agent/ individual attending an activity). Agents with similar activities form a group. Groups are formed based on various purposes. For example, class mate group are formed because they go to the same school, at the same class, and they are in the similar/ same age category. Group activities are formed when two or more individuals together perform the same activities. The individual

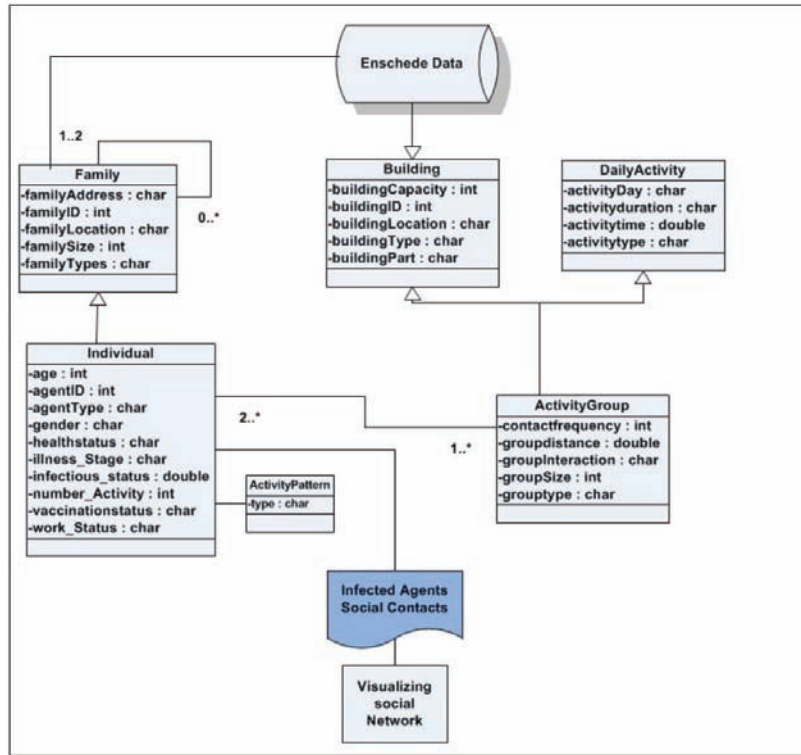


Figure 3.1: The class diagram of the existing pertussis model [1]

has an activity schedule which is used to determine what activity an individual should perform.

The activities are placed per age group, work status, and they may or may not be the same every day of the week. The activity schedule assists the individual/agent to change his/her activity after completing an activity. Changing an activity could cause the individual to be involved in another group which led to an interaction with other individuals and a linked interaction between two groups.

The model contains two types of activities. These activities are considered as static and dynamic activities. Static activities are activities such as going to school or work while dynamic activities are social activities which vary from day to day like visiting a friend. Figure 3.2 shows an example of an individual attending activity.

Social Interaction Model

The social interaction model, models the full and partial interactions that occur within groups or between groups, as social interaction is one of the key factors to disease diffusion. According to Abdulkareem [1] full interaction occurs when all members within a group interact. For example within a family group, parents interact with all members in the family. On the other hand, partial interaction occur when some of the members interact within group, for instance, a staff agent can only interact with staff in the same department or at the office but not all staff in the institution (see figure 3.3).

Figure 3.3 shows the full and partial interaction types. Individuals are presented by the faces, the group boundary is represented by the circle around the individuals and straight line between two individuals represents the contacts that are occurred. (Adopted from [1] from [7].

Activities are constrained by both physical and social constraints, also basic rules for social interaction such as the size of a group, the distribution of individuals within a group and the distance between individuals are considered. The distance within which an individual will interact

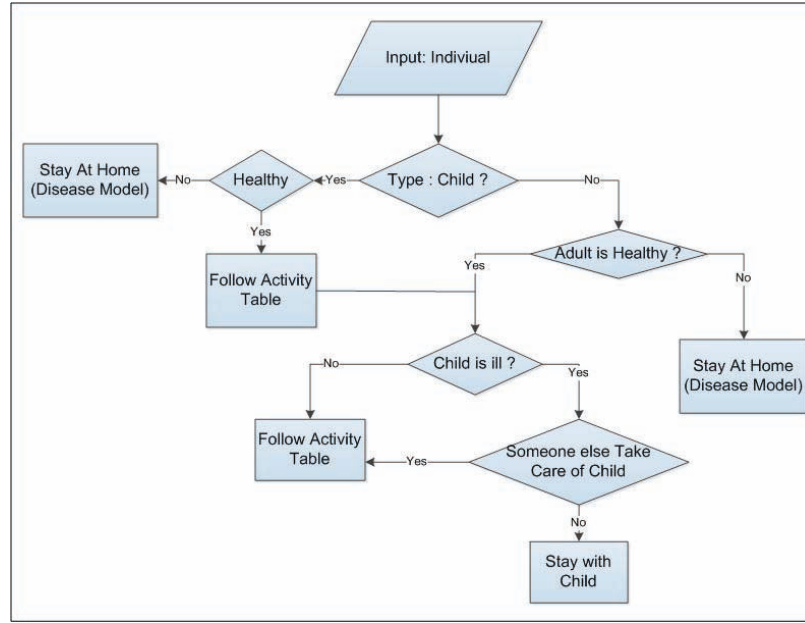


Figure 3.2: Demonstration of individuals attending an activity [1]



Figure 3.3: Types of Social interactions

was used to determine if an agent is susceptible or not. During group activities social interactions occur and enhance the disease diffusion.

Disease Model

In the first stage of the simulation individuals/ agents are created and assigned appropriate properties (attributes). Vaccination level is among the most important attributes of an agent. Vaccination level determines whether the agent is immune and will not be infected or is susceptible/ not immune and can be infected at any moment.

An agent is considered susceptible if s/he is not vaccinated against pertussis, has uncompleted vaccination or waning of vaccination. Even if the agent had been vaccinated, the vaccine might not give the agent the required level of immunity, which could be due to the effectiveness of the vaccine [45]. However, because of lack of data the vaccination type is not considered. These, immunity levels of agents are assigned using the available data.

In the model immunity levels of agents depend on their age, this means:

- Infants younger than the age of six months are considered as under vaccination,
- Children under the age of four but older than six months are considered to be 96% vaccinated with partial vaccination,

- Children between the ages of four to nine are 96% vaccinated with full vaccination, and their immunity is not yet deteriorated because of waning, which makes them the best-protected groups.
- Children between ages ten to eighteen are vaccinated but the vaccine is called serum (the old serum, not the new booster). The immunity of these children is waned by unknown level.
- Adults who are older than 18 years old are assumed to have waned immunity. It is assumed that 10 to 30% of them have had pertussis or still protected by vaccination.

The model uses the above assumptions to determine the vaccination level of the created agents. Thus, based on the above explained assumptions and the group size of created agents, vaccinated ratio of the group is created then assigned to individual agents. However, for adults above the age of eighteen years, around 30% of them by random ratio of the total created adults agents are considered to be susceptible. Finally, for modeling the spread of the disease there is a model called disease model.

The elements of the disease model are the individuals, individual vaccination, and individual's social interaction. The disease model and activity behavior of agents are related to each other.

The disease model comprises the infection model and illness evolution. The infection model evaluates the health status of individuals during interaction. Disease diffusion occurs when these constraints are met, which are; (a) distance between individuals during interaction should be less than or equal to one meter that means there should be a close contact interaction, and (b) within a group there should be an infectious and susceptible individual.

For disease to be transferred the disease model calculates the contact duration or the frequency of contact per day. That is in order for the disease to be transmitted the calculated contact duration should be more than or equal to an hour or the calculated frequency of contact per day between two individuals should be more than or equal to one hour for daily activity.

Illness evolution starts when an individual is infected. It models the characteristic stages of pertussis which are the incubation period (up to 7 days), catarrhal period (1-2 weeks), paroxysmal period (1-6 weeks) and convalescent period (4 weeks). An infected individual follows these stages. The health status of the infected individual changes when s/he is in the paroxysmal period and stops every activity and stays at home. In the model after infection agents do not recover.

Initial infected case is selected at every simulation run randomly, out of the loaded neighborhoods of Enschede. Immediately after infection an agent is not infectious, which indicates the agent is in an incubation period for three days. Starting from the fourth to the seventh day the individual becomes infectious but still follows her/ his daily activity. For example, moving from place to place, joining and changing groups. Thus, if an infected agent joins a group after two hours the probability of infecting the susceptible agents within the group increases. After seven days the infected agent stops activity and stays at home. Using the discussed process the simulation runs for one month with a time interval of 30 minutes (which is one tick).

3.2 LIMITATIONS OF THE EXISTING MODEL

To address the limitations emanating from the existing model, factors influencing the existing model should be identified and discussed. These potential limitations will therefore be considered as general benchmarks towards increasing performance of the existing model. For this reason, in the proceeding sections some of the limitations of the existing model will be addressed.

3.2.1 Applications to large spatial extent

The pertussis model currently models the individual person and the household. The total number of agents is equal to the population which have children (parents and their children) of Enschede and its map extent is restricted to the Enschede area [1]. Enschede consists of 70 spatial blocks (neighborhoods). These 70 neighborhoods are included in implementation of the existing model of Abdulkareem [1].

However, in reality, the spread of a disease is not spatially restricted [61] which means that it spatially spread to the large parts of the country and nearby neighbors. Since people move from place to place for many reasons (e.g. to work, visit, etc.).

Moreover, even though the model contains data of the 70 neighborhoods simulating disease diffusion over the 70 neighborhoods is not possible. Thus, currently, the model is able to simulate disease diffusion pattern over 46 neighborhoods with maximum population of 55345 [1, 50, 65]. When the neighborhood or population increases the model creates an error and stops running.

3.2.2 Time extent and recovery

The existing simulation model simulates the spread pattern of pertussis within one month that is 31 days. In the model, 31 days is represented by 1488 ticks, 1 day is represented by 48 ticks, and 1 tick represents 30 minutes of real time. However, as described in the above paragraph, pertussis is a disease that could last up to two months. Therefore, to show the natural disease pattern and the epidemic process of the Netherlands, the simulation's time extent should be increased in to two months.

In addition, as explained above the current model doesn't include recovery of infected agents/ individuals. This is because with time extent of one month recovery of infected individuals is not expected (according to the characteristics of pertussis stages). In the current model when an individual is infected he/she stops his/her daily activity and stays at home (see section 3.1). However, if the time extent is increased in to two months infected individuals should be recovered and go back to his/ her daily activity.

3.2.3 Immunity levels

The spread pattern and speed of any disease is determined by the immunity level of the people [6, 66, 54]. The stronger the immunity level of a society the less this society will be affected by disease, and vice versa. The immunity level of a society can be increased/ improved by different mechanisms, for example, in the case of pertussis the society's immunity level may be increased by vaccination or one may gain immunity naturally after infection [?, 37].

In an epidemics model adding immunity level of agents is relevant because it determines the pattern of spread, number of infections, speed of the outbreak etc. Immunity level is an important element in the pertussis model; this means it determines if the model output reflects reality or not. This is because the immunity level of agents determines the number of agents that are susceptible.

3.2.4 Commuting/ movement

Quite often, people move from place to place for various reasons to the nearby or distant cities. For example, trip to work, trip to school, trip to shopping, trip to recreation, etc. These indicate that spatial interaction occurs over a range of scales in space and time within a city, between a city and across regions. The commuting behavior in the Netherlands has been discussed in the preceding section 2.1.

As explained in section 3.1, the existing model depicts the spread pattern of pertussis within Enschede city. This indicates that the city is modeled as a closed city, which means, there are no incoming or outgoing commuters. Only effect of the behavior of individuals who commute within the city is modeled. As explained in the preceding sections one of the main focuses of this study is to consider the movement of people to the neighboring cities and regions.

3.2.5 Hardware requirement issue

Epidemic modelling contains large numbers of human interactions, emergent behavior and activities, and movements that cover large areas (see section 1.1). These properties make modeling disease diffusion complex. Running these models (converting them in to computer programs) and making these models behave the way they would in real life requires computers that are capable of executing the instructions and data of the simulations.

The capacity of a computer determines the performance of a model. This is because, performance takes into account the effects of the various subsystems such as processor, memory, disk, software efficiency, and algorithms [46]. Therefore, to check and upgrade the performance of a model its hardware and software requirements should be tested.

The existing pertussis model is built using the Java based application Repast Symphony 2.0 beta (see section 2.4.4). Currently, the default hardware allocation of Repast are used to run the model, which is 250 MB of memory. To store the model, 403 MB of Hard Disk is needed. However, these resources could be creating a problem on the performance of the model. Therefore, a test is needed to see and find out the appropriate computer (hardware) requirements.

3.3 RESULTS ANALYSIS OF THE EXISTING MODEL

As explained in the above sections the model developed by Abdulkareem [1] requires modification, as it has limitations in showing the realistic pertussis diffusion pattern. To gain better understanding and check performance and overcome limitation of the model a test is conducted. This test also aims at getting an insight for creating a conceptual model, which will be used for developing a method for scaling and improving performance of an agent based systems.

3.3.1 Testing spatial scalability of the model

As explained in the above sections the spatial extent of the existing pertussis model is the boundary of Enschede. Thus, in the model, data for the 70 neighborhoods is found in comma separated file called City_data.CSV. The city data is loaded to the model according the order in the CSV file (for example starting from row 1 up to 70). Even though data of the existing model includes all neighborhoods of Enschede, loading all areas in the model is not possible. For this reason, currently the numbers of spatial areas that are loaded to the model are 36, but, this number can be increased to 46 [1, 50, 65].

For the loaded 46 spatial areas the total number of population is 55345. In such cases, the model runs effectively without creating an error. However, when the number of spatial areas (number of population) is increased, the model returns an error message. This error is created during loading of the GIS environment, and after the model successfully created agents, neighborhoods and report the status of infection.

To identify the cause of the error, several tests are done including checking correctness of the city data. In the test, firstly, the data have been loaded by a group of 5 spatial areas (Neighborhoods) which becomes a total test of 14 groups (see Appendix 1). Each group is tested 5 times (total of 70 tests). As a result, seven spatial areas have been detected containing error in their data.

Table 3.1 Neighborhoods with error on their data and their total number of population.

No	Name
1	MARSSTEDEN
2	DORP BOEKELO
3	LONNEKER-WEST
4	BROEKHEURNE
5	USSELO
6	GOORSEVELD
7	TWEKKELO

To distinguish the source of the data error several tests has been done. The result showed the Shapefile of the model called City_Neighborhood_Buildings.shp containing an error.

The City_Neighborhood_Buildings is a file that contains points which represent the city buildings. Among the attributes of the file, Building_T is an attribute with a string [50] data type. Building_T contains the type of buildings (for example, type of a point could be a House, WorkPlace, or School etc.). With this in mind, it has been identified that for the seven neighborhoods this attribute is sometimes empty. This means that in these neighborhoods there are some buildings whose type is not described which is creating an error. Table 3.1 shows the seven spatial areas which contain error in their data and their number of agents (population).

To see the effect of these data's in the performance of the model, a test has been done by deleting their data from the model, and performing several tests. Yet, the error continued to appear at each run, which indicates that the source of the error is not only these data's. Furthermore, the order in which the data is loaded is reversed to see its effect on the model. The test then proved the order in which the data is loaded doesn't have effect on the model which indicates the error is occurred due to repast settings.

By default repast settings are allocated with 256 MB of memory. This memory allocation can be increased depending on the capacity of the computer to 512 MB or 1 GB of memory. The setting is also dependent on the disk size (needed hard disk) size of the simulation data which is 403 MB in the Pertussis model. Therefore, to enable the model to run for a large dataset, the memory limitation problem has been modified and allocated with 1 GB memory.

After the repast setting is fixed the remaining 63 neighborhoods were loaded and tested. The test then proved the ability of the model to simulate spread of the disease in all the 63 neighborhoods. For the 63 neighborhoods the number of agents is around 67000. In the test the number of infection, model initialization and loading of the GIS environment time, and complete run times has increased. Figure 3.4 below shows the output of the model, which is the number of neighborhoods and their total number of created agents, after the repast settings is fixed.

To see the applicability of the model in simulating large number of agents than the 67000, several tests have been conducted by increasing the number of agents found in an area. The test then showed the model is capable of simulating up to 155000 numbers of agents. However, this number depends on the capacity of the computer. This means, when the capacity of the computer increases the simulated number of agents is expected to increase and vice versa.

Test on increasing the simulation temporal extent

Part 3.2.2 explains that the pertussis model is currently simulating the spread pattern of pertussis within one month which is represented by 1488 ticks; while the natural stages of pertussis have

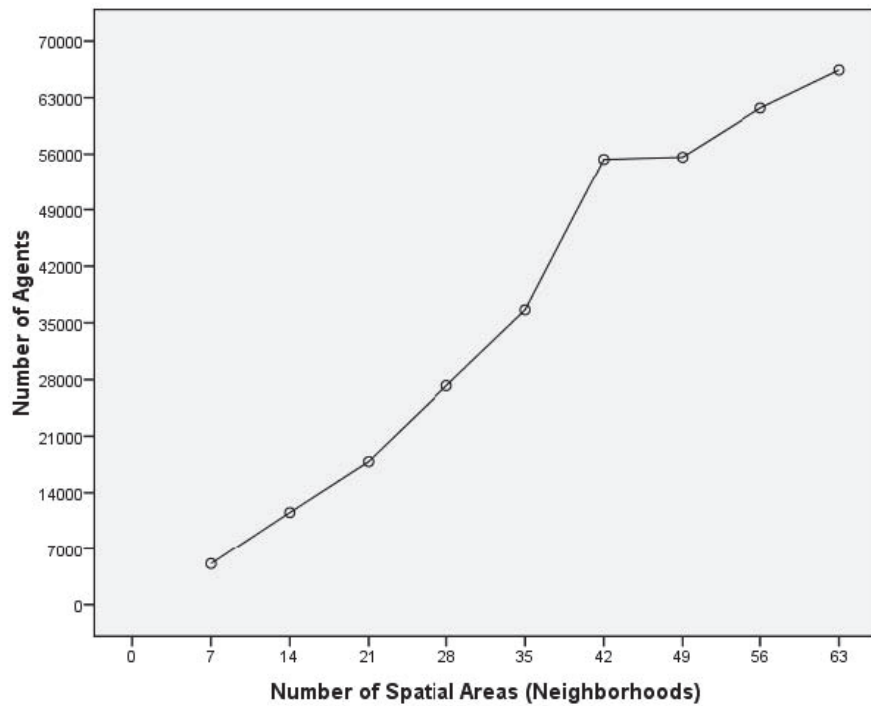


Figure 3.4: Total number of neighborhoods (63 neighborhoods) and their agents (around 70000)

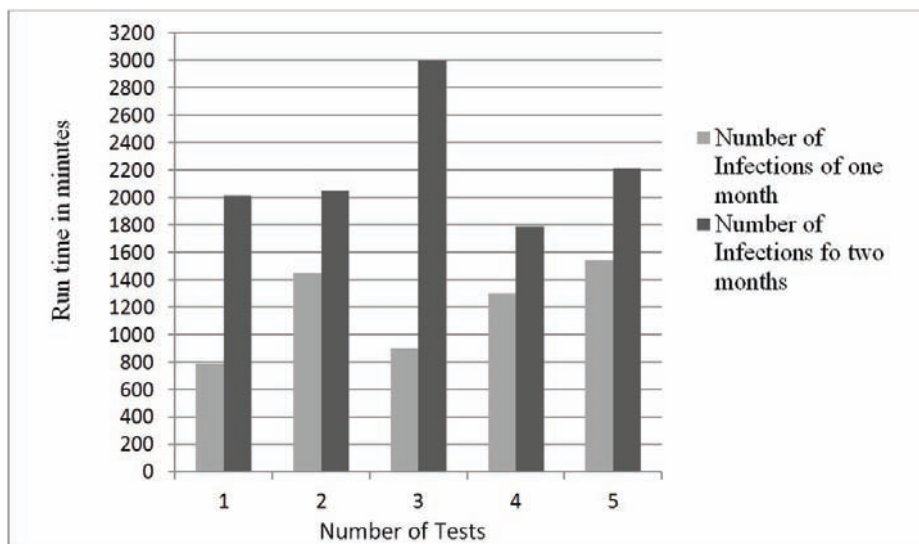


Figure 3.5: Complete run times (in minutes) of one month and two months

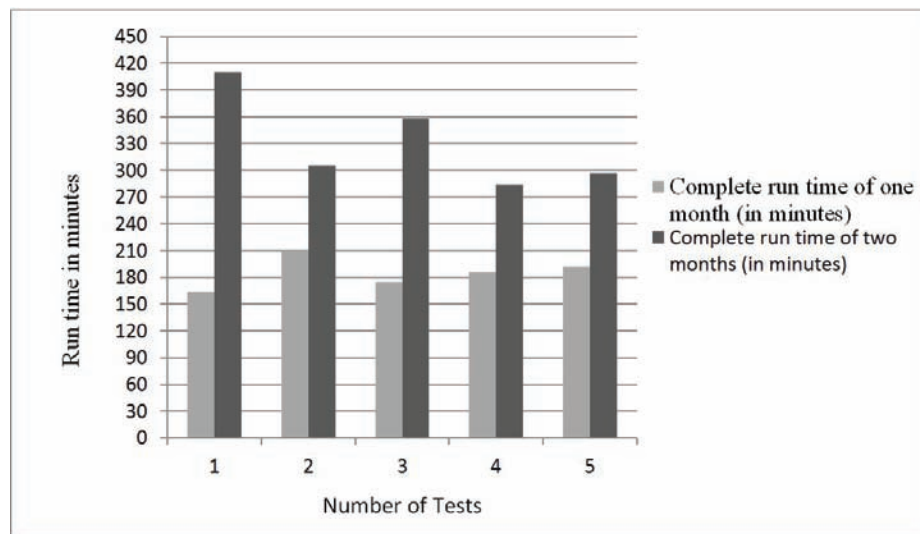


Figure 3.6: Number of infections of one month and two months

a life time of around two months. In addition, as explained in section 2.1 pertussis epidemics occurs in the Netherlands every 3-5 years. For this reason, by increasing the temporal extent of the pertussis model to two months (which is 2976 ticks) a test is conducted. Result of the test implied that the model becomes slower (the complete run time takes longer time than before). In addition, the number of infected agents increased.

The above figure 3.5 and 3.6 illustrate the impact of elongating the simulation temporal extent according to the performed analysis. The analysis is made by testing the model 5 times for a time extent of one month, two months. For the test all the 63 neighborhoods are used. The result then indicated the initialization and GIS environment loading time appear to be the same for all tests and temporal extents. However, the number of infected agents and complete run time has almost doubled for the temporal extent of two months and one year.

The GIS environment is used for visualization and as an environment for the agents interactions and daily activities. Therefore, even though temporal extent of the model is changed the GIS environment loading time is not expected to change because its direct relationship is with the number of agents and spatial extent not with the temporal extent of the model.

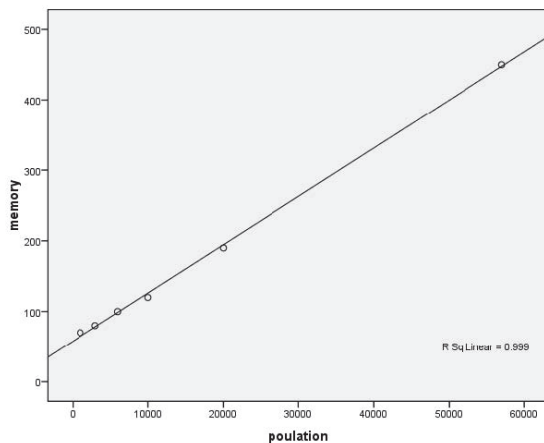
Increase of the complete run time and number of infections is the result of increasing THE simulation temporal extent to two months. In addition, lack of the existing model in recovering infected agents is also causing an increase in the number of infections as the time increases. The increase in the number of infections creates a negative impact on the model output because in reality the time is not the only factor which facilitates disease diffusion. Furthermore, people infected from pertussis should be cured at least one and half month after infection. To represent the natural disease diffusion pattern all agents whose time since infection is six weeks should be cured and get back to their normal daily activity. Therefore, to use the model with larger temporal extent recovery of the infected agents should be implemented (see chapter 4).

3.3.2 Hardware requirement issue (memory size requirement))

To determine the appropriate memory requirement of the model several tests have been done by increasing and decreasing the memory of the repast run time settings as well as the population size. Table 3.2 shows the data which was found on the testing of the model and which will be used for the analysis found in this section.

Table 3.2 Population and the needed amount of memory

Population	Memory (in MB)
1000	70
3000	80
6000	100
10000	120
20000	190
57000	450



(a)

Correlations			
		population	memory
population	Pearson Correlation	1.000	1.000**
	Sig. (2-tailed)		.000
	N	6.000	6
memory	Pearson Correlation	1.000**	1.000
	Sig. (2-tailed)	.000	
	N	6	6.000

** . Correlation is significant at the 0.01 level (2-tailed).

(b)

Figure 3.7: Correlation Between Memory And Number of Population

This data has been analyzed in SPSS and the results are discussed below.

The scattergram and the pearson product-moment correlation coefficient indicate that the number of population is positively related to the memory space. That means the increase in population needs the increment of memory space to run the model. The significance value indicate that a genuine relationship between population and memory space.

Table 3.3 shows both the value of R2 and adjusted R2 is 0.999, which indicate that population account for increase in memory space. This means that nearly 100% of the increase in memory space can be explained by population alone. Therefore, memory space is strongly correlated to the number of population of the simulation model.

Table 3.3 Model summary of the performed memory test

Mode	R	R Square	Adjusted R Square	std.Error of the Estimation
1	1.000	.999	.999	4.946

Table 3.4 Model parameters and the significance values of the performed memory test

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. error	Beta		
1 (Constant)	57.752		2.636	21.908	.000
population	.007	.000	1.000	65.255	.000

Table 3.4 provides details of the model parameters and the significance of these values. From the table b is 57.752, this can be interpreted as when the number of population is zero, and the model predicts that 57.752 memory spaces will be needed. The value of m is 0.007 which represents the gradient of the regression line. It represents the change in outcome associated with a unit change in the predictor. This means if the number of population increased by one unit, then the model predicts that 0.007 extra values of memory space is needed. From the table it is possible to estimate regression equation assuming the relation between population and memory space is linear.

$$Y = mx + b, \text{ where } m = 0.007 \text{ and } b = 57.752$$

$$Y = 0.007x + 57.752$$

$$\text{Memoryspace} = 0.007(\text{population}) + 57.752$$

Given the calculated linear regression equation of above, it is possible to make a prediction about memory space needed to run a model for specific number of population. Generally, from the above interpretation it should be noted that, in reality memory space yields a range, not a point. That means it is challenging to get the exact memory space needed for specific population size. However, it is possible to estimate the approximate of memory space needed to run the model. This indicates that, memory test that has been done in this section is toward serving mostly as bases for estimation of memory space to run the model.

To prove the applicability or correctness of the created equation memory size for 154280 agents have been calculated using the equation:

$$\text{Population} = 154280$$

$$\text{Memoryspace} = 0.007(\text{population}) + 57.752 = 1137.712$$

The calculated amount of memory has been assigned to repast settings and 154280 populations were loaded to the model. The model was tested and the correct amount of memory for the loaded population is analyzed. The output of the model analyses indicated that to run the given population the model requires 1130 MB of memory; which proves that the formula is able to estimate the approximate amount of memory requirement only with error of only 7 MB memory.

3.3.3 General property and performance check

Besides the above assumptions and performance checks the existing agent based model is tested several times to evaluate and understand its general characteristics. First, to study the temporal pattern of the model, the 63 neighborhoods are loaded and tested by running the model five times. The output of the test indicated that, in all tests the time required for initialization and loading of the GIS environment are the same.

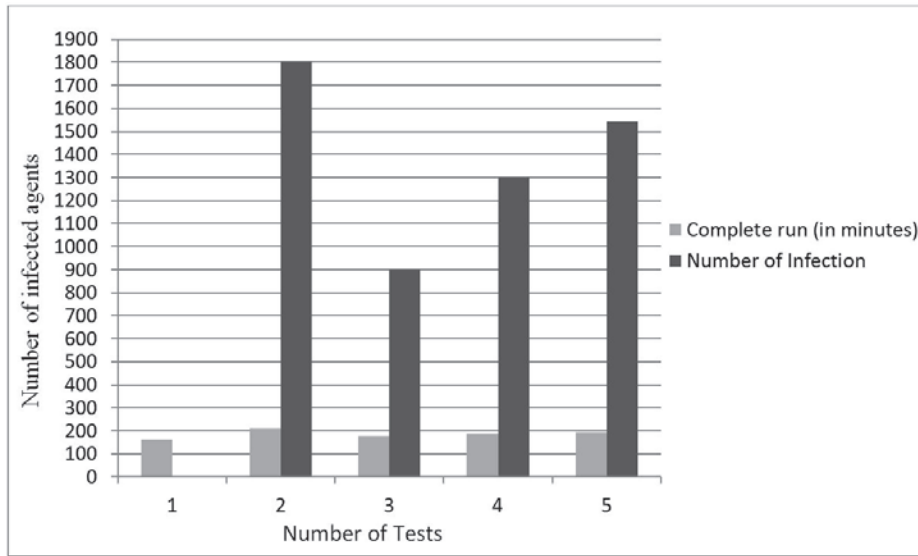


Figure 3.8: Number of infections and complete run time (in minutes) of each run for the 63 neighborhoods in five test runs

For this particular test the model runs five times with the same number of neighborhoods (which is 63). In the test initialization time is 22 minutes and GIS environment loading time is 9 minutes. For all test runs the loading and initialization time is the same. The initialization time is larger than loading of the GIS environment. This is because at initialization the model is creating all agents. In addition, it assigns a house, work/ school, activity, friend, family, vaccination level etc. for the agents. While the loading only loads the GIS environments for visualization.

From the test result it is clear that the run time and number of infections (See: figures 3.8). Moreover, the number of infections doesn't follow a normal distribution pattern. This means in some tests the number of infections increases rapidly and in others it reduces in to a single infection. For instance, in figure 3.8, test number one have only one infected agent in contrast to the others tests. This variation could be due to the randomness of the model, this means, in the model the first infection is assigned randomly. And the first infected agent could be, for example, surrounded by people whose immunity level is high enough to prevent them from being infected then clearly the infection will stop at one person only; if he/she is an infant and is only child of a family. In addition to this, if his/her family is fully vaccinated then the transmission will stop; on the contrary, if the first infected agent is surrounded by people whose immunity levels are low then the infection rate will increase.

In the test the total number of agents is around 67000. And the average total simulation time is 217 minutes (see Table 3.5).

To understand more about model behavior in running the total number of areas (the 63 neighborhoods) a second group of test is conducted. This test is done by grouping the spatial areas in two 9 groups in which each group contain 7 neighborhoods. Each category is tested 5 times. From this test the following key points are conducted.

From the test it was observed that as the number of spatial areas increase the number of agents increases and reaches a maximum of around 67000. For each spatial area the number of created agents is the same (at each run). In addition, enabling and disabling the disease model doesn't

Table 3.5 Total number of created and infected agents in each run for the 63 neighborhoods and their initialization, loading, and complete run time

No Test	Initialization Time(In Minutes)	Loading GIS Environment(In Minutes)	Complete run(In Minutes)	Number of Infections	Total Number of Agents
1	22	9	194	1	66435
2	22	9	241	1805	66435
3	22	9	206	900	66435
4	22	9	217	1300	66435
5	22	9	223	1544	66435

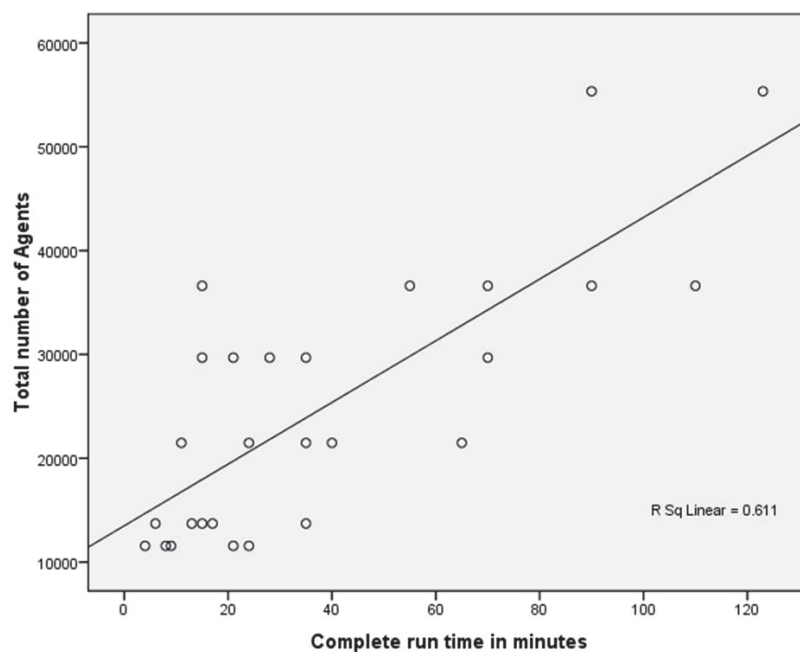


Figure 3.9: Average complete run time (in minutes) versus total number of agents

affect the created number of agents, but it creates a variation on the complete run time of the model. Which means, if the disease model is not enabled the complete run time is constant, in the other hand, the complete run time shows a difference from one run to the other (sometimes it increase and in the others it decrease) (see in appendix 2).

Note: (To show the impact of the infection model on the number of created agents 1 sample for each test run is taken for this analysis because the results of all tests are the same)

Furthermore, the test proved that currently the model is able to simulate for 63 neighborhoods with around 67000 populations. Generally, the result proves the number of agents and spatial extent are highly correlated with R value 0.99 and adjusted R2 value of 0.98.

From the test it can be understood that as number of agents increase the time for the complete run increases. This is because, to create the agents assign their houses, family activity, work, load the GIS environment etc. the model require a lot of time. This increase doesn't result in time only but also in memory (see section 3.3 number 3).

Generally, figure 3.9 shows that the number of agents and the complete run time are correlated

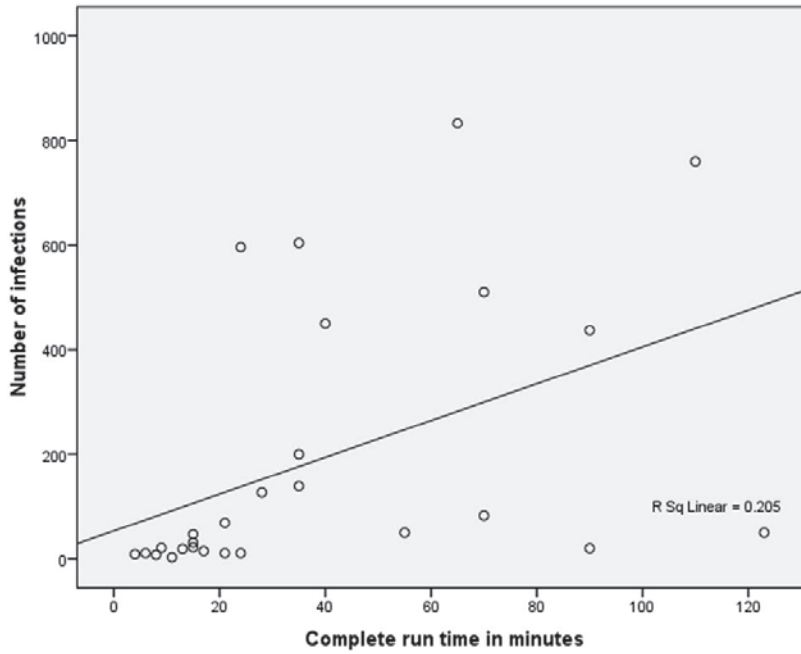


Figure 3.10: Complete run time (in minutes) versus total number of infections of the individual runs

with R value 0.611, this indicates the complete run time of the simulation is highly dependent on the number of agents. This means, as the number of agents increase the complete run time increases and vice versa.

In the model the number of infections doesn't affect the simulation time. This means time required for a complete run, is not correlated with the total number of infections ($R=0.205$). This can be seen in figure 3.10. The figure indicates that when the time increases the number of infections doesn't grow proportionally rather the number of infection sometimes increase and in the other decrease. For example, at complete run time of 75 minutes the number of infection approaches to 200, 600 and 800 at some point.

Therefore, complete run is not correlated with the number of infections. This could be due to the randomness of the model as described in the above sections (see the beginning of section 6).

3.4 SUMMARY

In this chapter, initially, description of the conceptual model of the existing agent based pertussis model of Enschede is discussed. Specifically the three main models which are contained to build the model are explained in detail. Finally, the model is tested and analyzed.

The analysis mainly focuses in checking performance and scaling bounds of the model. This is because the model was functioning only for 46 neighborhoods of Enschede even though data was available for the whole 70 neighborhoods of Enschede. In addition, the model doesn't incorporate recovery of infected agents and waning immunity of agents of all age groups which limits the model scaling temporally. Furthermore, the model was built as a closed city which means commuting/ movement of agents was not included which greatly affect the realistic diffusion process of disease.

Results of the analysis indicated that spatial scaling was affected by data errors which were used for the model and setup of repast settings (repast was using limited memory space). Hence,

removing or fixing these data (corrupted data) and repast settings maximizes the spatial extent and number of agents (around 150000) of the model.

Enabling the model to run for larger number of agents and spatial extent requires implementation of commuting of agents. Moreover, for scaling temporal extent of the model the recovery of infected agents should be included to the model as explained in the above paragraph. The next chapter (chapter four) discusses the adjustments and new concepts which are included to scale and improve performance of the pertussis model.

Chapter 4

Conceptual model

This section describes the conceptual design for the adjustments of the existing pertussis model and the design of the hierarchical model. The conceptual model is prepared based on results of the analysis in chapter 3 and with the general aim of creating a methodology for scaling and increasing the performance of the existing agent-based pertussis model.

To achieve the intended objectives a conceptual model which consists of several parts is created. Including the following:

- General conceptual model of the spatial and temporal hierarchy
- Conceptual model of the three hierarchical levels (disease and commuting models)
- Conceptual model for the adjustments of the existing pertussis model

Understanding of the conceptual model, limitations and results of testing of the existing model, which was described in section 3.1, 3.2, and 3.3, are the basis to understand the conceptual model of the adjusted model.

4.1 THE HIERARCHICAL MODEL

As explained in section 2.3.2 a hierarchical modeling approach is selected for scaling the existing pertussis model. This is because epidemic modelling is complex and it contains large number of human interactions, emergent behavior and activities, and movements that cover large numbers of areas. According to Levin, Simon, and Wu and David [42, 63, 70], complexity takes the form of hierarchy that means complex systems are usually hierarchically organized. In addition, these authors stated that the construction of a complex system using a hierarchical approach is likely to be more successful.

The hierarchical modeling approach is based on the central ideas of hierarchy theory, i.e., (i) the system is decomposed vertically into levels and horizontally into holons (elements at one level, which are complete systems when seen from the lower level), and (ii) higher levels are characterized by slower processes than lower levels. The first step in the hierarchical modeling approach is to create different hierarchies such as spatial, temporal etc. based on different kinds of frameworks for handling complexity and organizing complex system into smaller and manageable parts.

Hierarchies in a hierarchical modeling approach could be created based on order hierarchy, inclusion hierarchy, control hierarchy or level hierarchy (see section 2.3.2). For this study, level hierarchy is selected because it is based on structure, process and interaction of a system. In addition, it includes the structures of order and inclusion hierarchies.

Therefore, to implement the scaling process, the study area (Netherlands) is divided into three-level spatial hierarchies namely the urban systems level, the regional level, and municipality level. In addition, the temporal resolution is divided into twice a week, twice a day and every thirty minutes. This choice was made in order to maintain the natural ordering of municipalities and

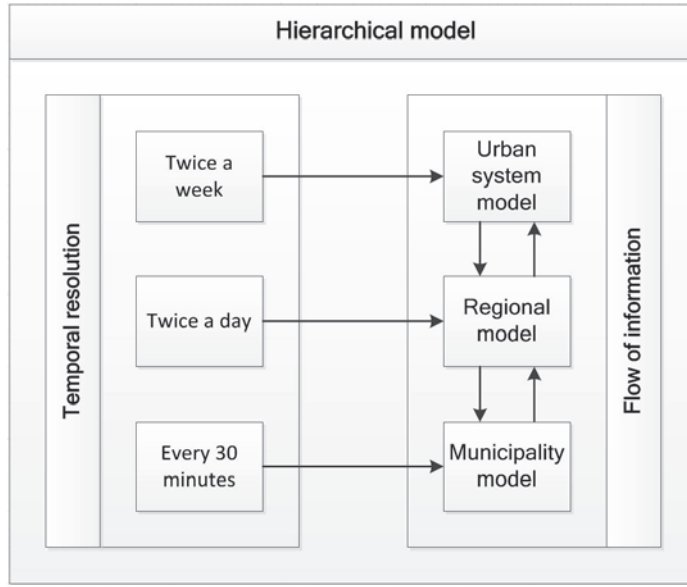


Figure 4.1: Spatial and temporal hierarchies and flow of information

regions in the Netherlands and for keeping the simulations reasonably fast. Consecutively, after the hierarchies are created the communication and flow of information between them is created. The hierarchies and their temporal description are illustrated in the figure 4.1 and sections below.

4.1.1 Spatial and temporal hierarchies

Spatial hierarchy

The Netherlands is developing into a network society and a network economy. On the one hand, individualization continues to progress; on the other hand, all those individuals are increasingly closely interconnected in numerous networks. This development also has major consequences for spatial planning. There is more and more coherence between the various cities and urban areas. The government applauds this development towards urban networks. Partnerships between such networks expand the support base of public facilities and services and open up opportunities for optimal use of the scarce space. To respond to this trend, the national government has designated 6 national urban networks [64]. A national urban network is defined as an entity of larger and smaller cities, including the open spaces in between [64]. The cities and centres that comprise these networks complement and reinforce each other's strengths, so that they have more to offer together than they do as individual cities. These national urban networks are: Randstad Holland, Brabantstad, Southern Limburg, Twente, Arnhem-Nijmegen and Groningen-Assen. Thirteen economic core areas are largely found within these networks (National Spatial Strategy, 2010).

Commensurate with the National Spatial Strategy, in this study, the above mentioned national urban networks/ urban systems are considered as a strategic division for upper level hierarchy. The assumption is that these six urban networks are enough to capture the interaction and mobility of the Dutch people. In addition, the first arrival cases of pertussis starts from the urban areas and then spreads out to the hinterlands; because of the large population density as well as spatial extent and social interactions of these cities [13, 35]. Although the assumption considers only the six urban systems, if required, the model should be capable to include (model) for the remaining cities and villages of the Netherlands.

In the Netherlands more intense interaction is observed between urban areas over the years; however, the intense interaction is often confined to urban areas that are geographically close to one another [44]. According to those authors, particularly the areas located within a region/urban system have more interaction than the areas which are found in different regions/urban system. This suggests that physical distance plays an important role in the social, economic activities and commuting behavior of the people. This leads to the conclusion that studying the pattern of disease diffusion within an urban system/region is reasonable. For this reason the middle level/second level in the hierarchy municipalities that are found within an urban system (as shown in figure 4.2).

Since the main objective of this study is to increase the scalability and performance of the existing model of Abdulkareem [1] the case of Enschede city, Twente region is considered as a strategic area for middle level hierarchy among the six urban systems.

The Twente Region is situated in the Eastern part of the Netherlands. The number of inhabitants of the region is approximately 625,000 (4% of the Dutch population). Twente is located on the axis that runs from Amsterdam via Berlin to Moscow. There is an international intercity connection by train from Amsterdam to Berlin that stops in Hengelo and direct connections to the European rail nets (Netherlands: Thalys, Germany: ICE, France: TGV). According to Self-Evaluation Report of Twente (2005), in Twente, there are a number of large companies with an international reputation, such as: Texas Instruments (microelectronics), Urenco (uranium enrichment), Siemens, Philips, Stork, Eaton-Holec (all in the metalelectrical sector), Vredestein (tyres), Grolsch (beer), Ten Cate (industrial textiles), Polaroid (photographic films), Cannondale (bikes), Bolletje (industrial bakery), Uniq (salads), Zwanenberg (fine meat products), Akzo (salt production), SASOL (chemical additives), Thales (military radar equipment). Furthermore, there are two main cultural institutions located in Twente: the Nationale Reisopera (National Travel Opera) and the Orkest van het Oosten (The Netherlands Symphony Orchestra).

The region comprises of 14 municipalities situated in the Province of Overijssel. Among those municipalities, Enschede is the spatial extent of the lower level model and it contains the highest number of population.

The third and final level in the hierarchy is the settlement within a city. As explained above it is selected due to the availability of the existing pertussis model as well as its advantage of capturing the spread of pertussis in an individual based model.

In the hierarchical model, there are three types of holons: urban system holon, municipality holon and agent holon. The urban system holon contains the 6 urban systems/regions. The municipality holon contain all municipalities that are found within a single urban system. And the agent holon is a collection of all individuals (agents) and their children with in a municipality. Each holons have different spatial extent and granularity.

The extent of the urban system holon is the Netherlands and its granularity are the 6 urban systems. Agents of an urban system holon move from one urban system to the other. The spatial extent of the municipality holon is an urban system while its granularity is a municipality. All agents of the municipality holon live in the different municipalities. And they are allowed to move from one municipality to the other which is found within the same urban system twice a day. Unlike the others, the agent holon has a detailed and smaller granularities since it is individual based. Thus, agent holon contains individual houses, shopping centers, offices, sport centers, churches, and schools as its granularity and a city as its spatial extent or boundary (see figure 4.2).

Temporal hierarchy

The developed conceptual model for epidemic modeling contains both spatial and temporal hierarchies as mentioned in the previous sections. The lower level in the hierarchy which is the individuals in a municipality are computed based on the model prepared by Abdulkareem [1]

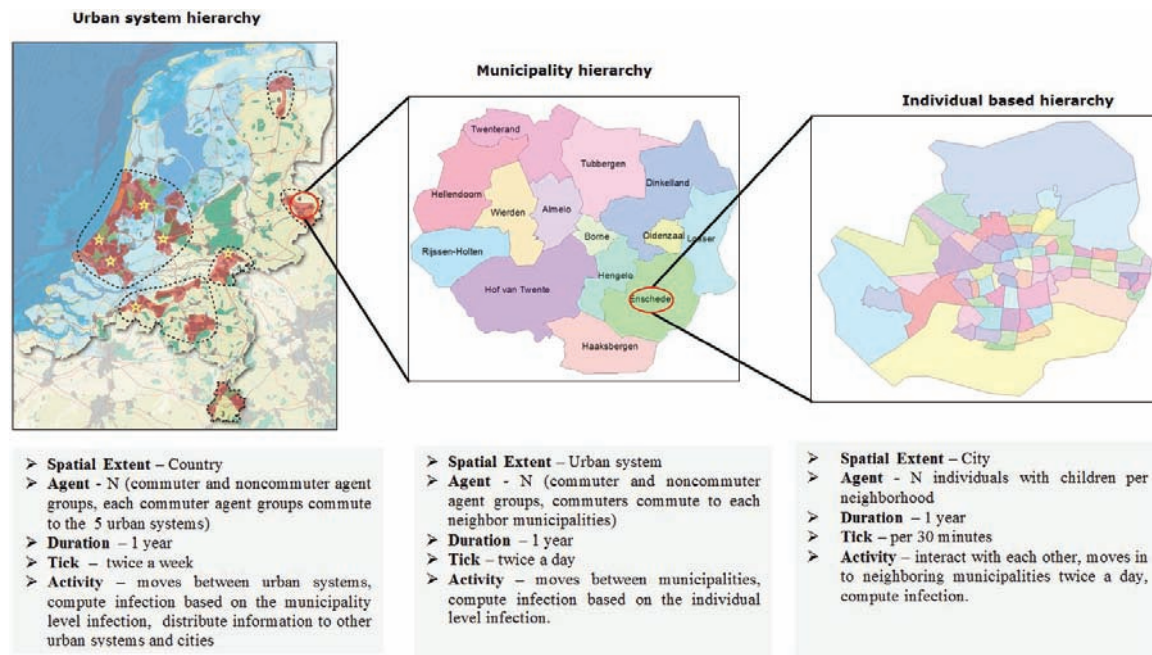


Figure 4.2: Netherlands and its hierarchical levels

with some modifications (for example, increasing the spatial and temporal extent, addition of recovery and waning immunity level to all age groups (see chapter 3)).

The model by Abdulkareem [1] schedules tasks of agents and updates information per 30 minutes of real time and one tick in simulation time; that means, in her model one day is represented by 48 ticks. Duration of the simulation was one month but now it is extended to one year, because, simulation of pertussis diffusion pattern within one month is not long enough to show the pattern of pertussis outbreak in the Netherlands. Therefore, the duration of this simulation is updated from 1488 (48 ticks x 31 days) to 17520 (48 ticks x 365 days). For the lower level in the hierarchy the temporal resolution of Abdulkareem [1] is adopted.

However, for the middle level in the hierarchy/ municipality level, adopting this temporal resolution (30 minutes) is not relevant/ realistic. Because, as explained by Okla [50] due to distance between the municipalities, people commute from one municipality to the other once or twice a day. For example, in the morning students may travel from their home municipality to other municipality for school and return back in the afternoon. For this reason, in weekdays, the temporal resolution of the middle level model is changed in twice a day. This means information is updated twice a day (every 24 ticks) which is one tick into simulation time. On weekends the commuting activity of people is often changed; for example, people may stay at their home municipality or commute long distance for social visits (e.g shopping, family visit). Therefore, in the municipality model the commuters stay at their home municipality.

On weekends the commuting activity of people is often changed; for example, people may stay at their home municipality or commute long distance for social visits (e.g shopping, family visit). Therefore, in the municipality model the regular commuters (week day commuters) stay at their home municipality while the weekend commuters, commute to other urban systems. The weekend commuters are assumed to commute on Friday afternoon and return back on Monday morning; which leads to the assumption of the weekend temporal extent. The temporal extent is: every 216 ticks commuters commute from their source urban system to the destination urban system and return back every 337 ticks; which is twice a week.

Finally, the upper level in the hierarchy which is the urban systems level updates information twice a week. This means two ticks in simulation time is 1 week in real time. Therefore, duration of the simulation is one year which is 96 ticks (2 ticks x 48 weeks). This temporal resolution is selected by taking into consideration the distance between the urban systems, and it is assumed that most people commute from one urban system to the other twice a week. For example, people may commute to visit friends or relatives to a municipality of other urban system on Friday afternoon and could return back Monday morning.

4.1.2 The respective hierarchical models

Large scale epidemic estimation is dependent upon the accurate and realistic modeling where the movement of individuals at various levels is taken into account [26]. To model such complex systems a modeling approach which is able to handle the complexity is required. Metapopulation approaches explicitly include spatial structures and consist of multiple subpopulations coupled by commuting/ movement of people. For this reason, the urban system model and the municipality models are built based on theories and assumptions of the metapopulation modeling approach (see section 2.2.2 for detail description). For the individual based model all the previous assumptions are adapted with some modification.

Quite often, families with children have the highest probability to be infected with pertussis compared to individuals with no children. Therefore, in this study, families (who have children) aged between 0 to 55 years were considered while individuals aged between 19 to 25 and individuals older than 55 years were excluded. In addition, these assumptions were in accordance with the individual based Abdulkareem [1] model.

In all levels of the model the population or number of people is constant, i.e. no births or deaths are included. In addition, population or people's behavior is constant, i.e. no people will lose their jobs or start to work different hours, no people will change their municipality or commuting destination.

In addition, the number of commuters from one municipality/ urban system to the other is different. This difference could be due to the economic strength, economic activity, and job attraction of the areas. Which means, the extent to which municipalities attract people living elsewhere but commuting into the municipality depends on factors, such as the kind, number and level of jobs the municipality offers, the municipality's accessibility by car and/or public transport, the number and kind of jobs in the neighboring municipalities, etc.

The urban system and municipality level models contain an agent called Group-agent (also known as agent-groups). Group-Agent represents a group of people/ individuals that are in the same age category and are grouped together. Therefore, the basic assumption of the group agent is that all individuals found within similar age categories are structured. The assumption is based on the basic idea of age structured metapopulation modeling.

The idea of group-agent allows giving a group behavior, activity and reasonably permits to model behavior and activity of relatively big number of individuals/ people in a compacted manner. The assumption was adapted because in these models the number of people to be modeled is large and modeling behavior, interaction and pattern of large amount of people in an individual bases is computationally expensive.

However, in the hierarchical models the concept of individual behavior and activity is not completely disregarded. Yet, it was included when the middle level municipality model is connected to the individual level model. This means agents from the municipality model are sent as groups, but later when they enter the individual based model they are decomposed and become individuals (see figure 4.3).

Figure 4.3 illustrates the concept of the spread of information between the hierarchical levels.

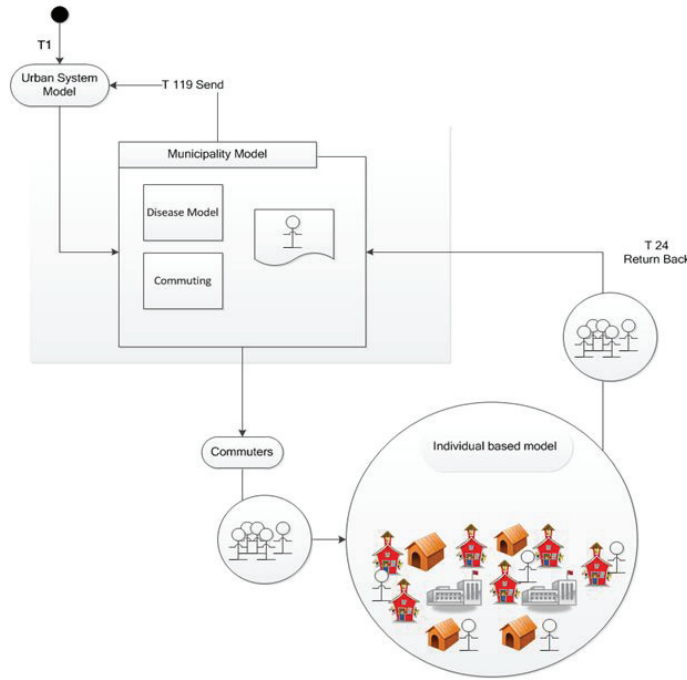


Figure 4.3: Flow of information of the hierarchical

That means in the urban system and municipality level models the concept of group agents is used. Therefore, the commuter agents are group-agents. However, the lower level model is individual based, thus when municipality level commuters commute to the model they are decomposed and become individuals. The over all concept of the hierarchical model is illustrated in figure 4.4.

The proceeding sections (sections 4.1.3 and 4.1.4) present a detailed description of the two hierarchical models for the two corresponding hierarchical levels which are discussed in section 4.1.1 and the adjustments that are made to the existing model. The concepts are mainly focused on adjusting the identified and discussed limitations of the existing model in its effort to enhance the scalability and performance concurrently.

4.1.3 Urban system level model

Spatial extent of the urban system model is the Netherlands and its six urban systems. In the urban system model, an urban system consists of an agent and agent-groups. An agent-group consists of populations which are grouped by age into infants, children, teenagers, and adults. These groups could be commuters or noncommuters.

The individual agent represents the urban system (government), in addition there are four "agent-groups" based on four population categories (Infants, children, Teenager, and adults). These agent-groups are noncommuter agent-groups. Furthermore, an urban system consists of commuter agent-groups which involve only teenager and adults (see municipality model for detail). In the model all urban systems have N numbers of group-agents which involve the mentioned two categories. That means, an urban system have 4 noncommuter plus 12 commute agent-groups.

Noncommuter agent-groups consider the population who live and work within their urban systems. Therefore these agent-groups are not allowed to move from their home urban system to other urban systems. On the other hand, commuter agent-groups are allowed to move from one

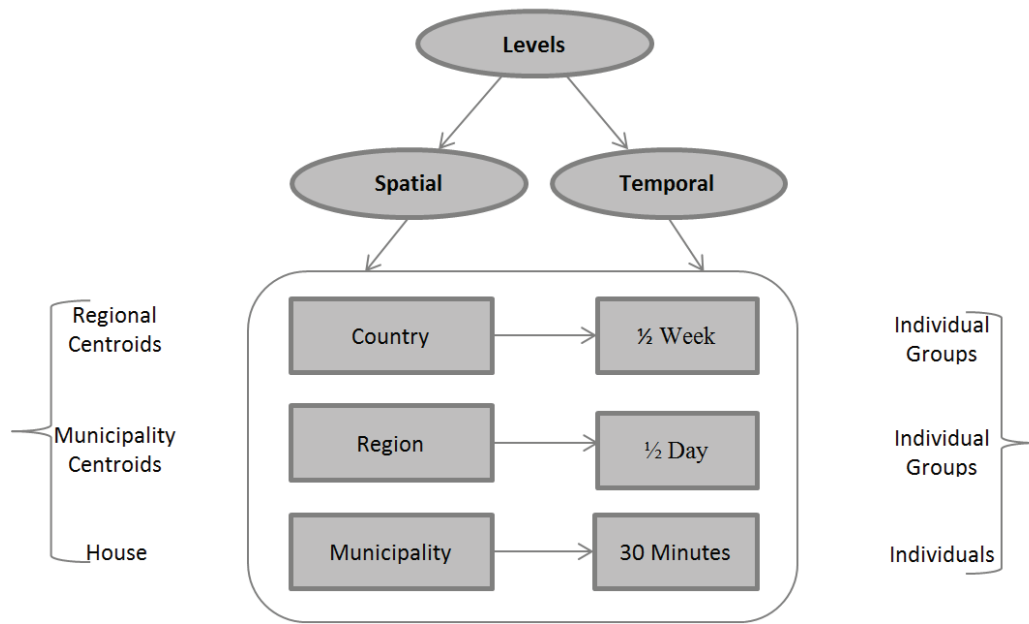


Figure 4.4: Spatial, Temporal, and Agent characteristics of the hierarchical model

urban system to the other for different reasons. Having commuter agent-groups enables an urban system to have interaction with all urban systems and allows to detect the spatial diffusion process of the disease between the urban systems.

Commuter agents-groups represent people who commute from one urban system to the other. These groups could be regular commuters such as students and full time workers who travel every weekend from school/ work urban system to home urban system or irregular (random or occasional) commuters who commute for social visits.

Update of information is 124 times within the simulation time extent of 365 days. This means, the model updates information twice a week, which is on Monday morning (every 337 ticks) and on Friday evening (every 216 tick). The model keeps track of information update time by detecting the current tick of the simulation time. The transferred information includes: total number of incoming/ outgoing commuters, address of incoming/ outgoing commuters, and status of incoming and outgoing commuters (is infected or not) per age group.

When infection is detected within an urban system, the model sends information to the other urban systems, which used to warn the other urban systems about the infection; and it sends information to all its cities and villages, to advise them to vaccinate their people. As a result, the other urban systems increase the vaccination status of their agents, and close schools.

Assumption is made that all municipalities that are part of the urban system are actually running in the middle level (they are part of the municipality level model), but the highest level can calculate the number of diseased based on data collected from the municipalities.

Activities and behavior of agents

As explained in the above section, the urban system model consists of two types of agent (an urban system agent and a group-agent). The groups have their own behavior and activities, each behavior and activity of an agent-group is constant.

- " Activity: throughout the simulation period noncommuter agent-groups stay within their home municipality. If the group is commuter, in weekends they commute to their destination urban system and Monday morning they return back to their home municipality until

the next commuting time.

- Behavior: there is no increase or decrease of population therefore the size of the population stays constant. If the group is commuter, no people losses job or change commuting or home municipality, no commuter changes age group, if the group consists of agents who have been infected for 7 days, on the next commuting the group have to eliminate these agents in their home municipality.

When infection occurs within the urban system, the information is spread to the other urban systems. Then these urban systems take an action that will increase the vaccination level of their population.

The urban system model consists of computing model sending people to other urban systems (based on commuting data) and a disease model which counts the number of infected commuters and calculates the urban systems infection based on the municipality level infections.

Work flow of the model

The main purpose of the urban system model is to send commuters and receive information from municipality model. The information could be infection status of the region including the number of infections, infected commuters, and total number of commuter per each group and each municipality. Finally, based on the information acquired from the municipality model (e.g. status), the model spread information to the urban systems.

On initial tick the commuting model starts to collect commuters and send them to their commuting location. The commuters are distributed over the municipalities, and in case these municipalities are linked to the lowest level they will be transferred to the municipality level model.

Then the disease model checks if there are infections. When an infection has occurred, this model sends information to all urban systems. As a result these urban system change their behavior for example, they increase vaccination level of their population.

During the temporal extent commuters come from the municipality level model twice a week therefore the commuting model of the urban system model collects these commuters and distributes them to their destination location. Concurrently, the disease model checks and calculates the number of infected agent-groups and distributes information. Figure 4.5 illustrates the activities of the disease and commuting models of the urban system level model.

Daily activity model

Unlike the disease model of the municipality level model, disease model of the urban system model is simplified and it performs simple arithmetic operations. That is, it check if any infected commuter is coming from the municipality model or if there exists any infection on the urban systems. When infection is detected from the municipality level model, this model distributes information to the urban systems which notifies them to change their behavior and increase their population's vaccination level or close schools or stop commuters from entering the urban system.

The model simply counts the infections, update number of infection, and distribute the infection to the urban systems.

Commuting model

Commuter agent-groups of the urban system model are Adults; which could be workers going to work or coming back from work, shoppers going for shopping or returning back from shopping, visitors going for family (friend, holiday) visits, or all.

The commuter agent-groups are allowed to commute from their source urban system to the destination urban systems twice a week. This indicates the commuting behavior of the urban system level model is twice a week. This is because, due to the distance between the urban systems

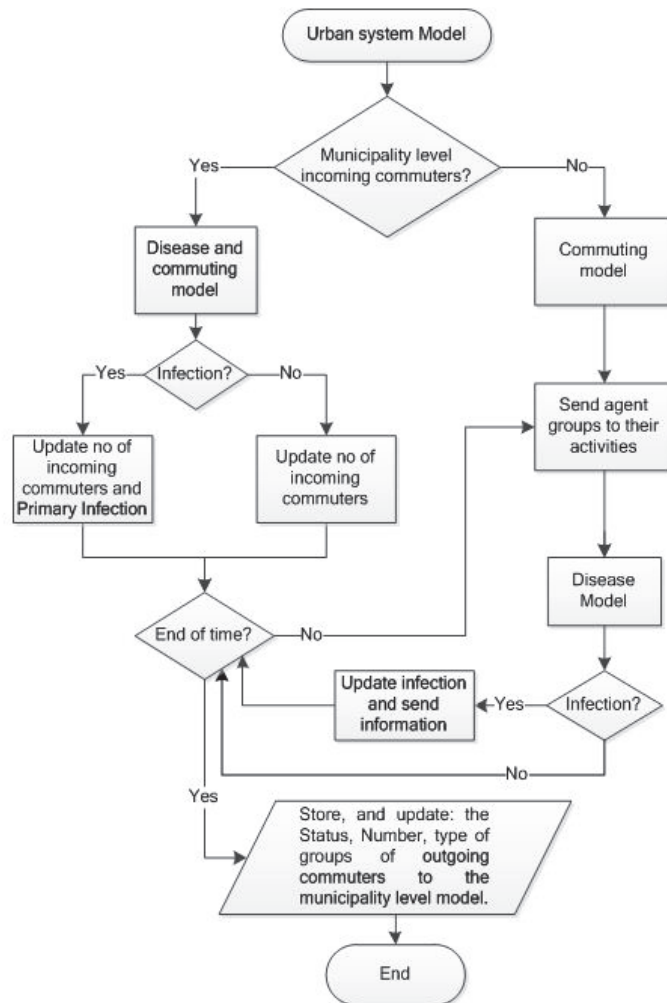


Figure 4.5: General work flow of the urban system model

it is assumed that people go to work, school, etc. on Monday morning and return back on Friday afternoon.

Commuting occurs on weekend that is Friday afternoon and Monday morning. Therefore, on Friday afternoon agent-groups commute from their home urban system to other urban systems and return back on Monday morning; that means every Monday morning commuters return back to their destination urban system. The reason of their commuting could differ from one group to the other, for example, it could be for social visits or shopping on weekends.

In the model there are two types of commuting which are regular and irregular (random or occasional) commuting. Regular commuting occurs when agent-groups commute to the current urban system for work and they return to their homes in the weekends. On the other hand, occasional/ irregular coming is occurred when the agent-groups commute to other urban systems to visit friends, relatives, or shopping on the weekends. Therefore, agent-groups of this model could be among those categories.

Regular commuters are expected to be the same agent-groups every week while occasional commuters could be the same or different groups. Information and list about regular commuters is based on data. However, the irregular or occasional commuters are generated randomly since they could differ from week to week.

As for the commuters of the municipality level model, commuters of the urban system model are aware of the day and time of commuting, source and destination address, age group, and type of activity. Every 216 ticks commuters commute from their source urban system to the destination urban system and return back every 337 ticks. When commuters are infected they change their activity and stays at home; which creates a decrease in the number of commuters.

In the model transportation or path in which commuters follow to commute from one urban system to the other is not considered. The simulation tracks only activities of the commuters after they reach a destination municipality. Commuter agents can be infected at their home municipality or at the place they commute and vice versa. On the other hand they can infect noncommuter agents.

With the commuter their status such as health status, age group, source and destination location, etc. is sent to the destination municipality at the initial tick. Once they reach to their destination, they interact with noncommuter population of the municipality. Consequently the disease model calculates if there are new secondary infections.

The commuting model of the urban system model is responsible to send the commuters according their destination address and collects them when the commuting time is completed. In addition, this model collects commuters from the municipality level model and distributes them according their destination address. At the end of their commuting time the model collects these commuters and sends them to the municipality level commuting model.

4.1.4 Municipality level model

At the moment the spatial extent of the municipality model is Twente region and its 14 municipalities. But the model could be used for other areas by changing the used data (the used simulation data of the Twente region). The choice was motivated due to the availability of the individual based model of Enschede.

In the model each municipality consists of an age structured population/ agents. That means, N number of agents per age group represents each municipality, where N is the number of groups plus number of neighbors of a city. This concept enables a city to have interaction with all its neighbors at a time. There are two types of agents-groups: commuter and noncommuter agent-groups. Noncommuter agent-groups are not allowed to move from their home municipality to other municipalities. These agent-groups represent population of a municipality who works and

lives within their home municipality. On the other hand, commuter agent-groups are those agents who commute in to other municipalities within the urban system for different reasons. For example: an agent-group can commute from their home municipality to work/ school and back from work/ school to their home.

Agents are grouped according to their age groups into infant 0 to 3, child 4 to 12, Teenager 13 to 18, and adult 25 to 55. Further, these agent-groups are named as:

- TeenagerStudent agents are between the ages of 13 to 18, and
- Agents who are adults and works full time are called WorkerFull agents, WorkerHalf if they are half time workers, and NonWorkers if they are jobless.

Each municipality consists of 6 agent-groups, in which four groups are noncommuters and the 2 are commuters, for example, if the total numbers of population of a municipality in the age category 0 to 3 is 500, then in the model the municipality have an infant group with size 500.

Each agent-group has associated attributes such as: age groups, health status, type (commuter or noncommuter), vaccination level, municipality code. Initially the health status attribute of agents could be susceptible or infected, hence; this attribute could be changed in the simulation period. This means, each group has collection of susceptible, infected and recovered agent which could vary through time. In addition, agent-groups have daily activities which are situated in space and time. The daily activities of each agent-groups are constant, but differ between groups; for e.g., the activity of commuter is different from noncommuter.

As explained above noncommuter agent-groups are groups who work and live within their home municipality. These agent-groups are includes infants, children, teenagers, and adults.

During the simulation's temporal extent these noncommuters are activated and perform daily activities within their municipalities. The activities could be going to work, schools, interacting with one another, etc. However, the daily activities of these agent-groups are not considered in the model. Therefore, the agents are created and based on the metapopulation approach the infection between them is calculated.

Commuter agent-groups are assumed to be students and full time workers who travel every day to school/ work. The agent-groups are between the ages 13 to 18 for student and 25 to 55 for full time worker adults. However, infants, children, WorkerHalf, and NonWorkers are not considered as commuters.

Fig. 4.5 Illustrates the schematic representation of the structured agent-group in municipality (big circles) and the interaction between the municipalities (broken arrows). The agent-groups inside the municipality and the individual agents (Small circles and smallest circles inside) and the interaction between them is the (solid arrow). The interaction also indicates that some members of a group go out of the group to a group of another municipality

In the model the spatial environment consists of centroids of the municipalities. This means that for each municipality, agent-groups are positioned in those municipality centers. This is because the intension of the municipality model is to calculate the disease spread within the structured populations without detecting the individual interactions and the causes of these interactions.

In the simulation model the agent-groups are represented by a single point symbols. Each municipality consists of 30 points symbols, i.e. 4 points represent the noncommuter population (infant, child, teenager, and adult) and the other 26 represent the commuter agents which commute to each municipality (teenager and adult).

Activities and behavior of agents

As explained in the above section, the municipality model consists of two types of agent-groups (commuter and noncommuter). The groups have their own behavior and activities, each

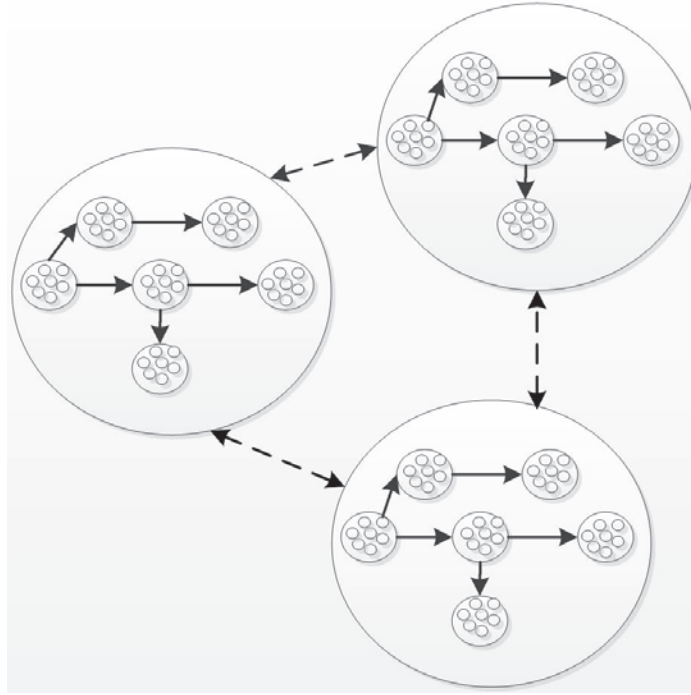


Figure 4.6: Structure of the modeled population

behavior and activity of agent-groups is constant.

- Activity: throughout the simulation period noncommuter agent-groups stay within their home municipality. If the group is commuter, in day time commuters commute to destination municipality and in night time they return back to home (destination) municipality and stay at the municipality until the next commuting time.
- Behavior: there is no increase or decrease of population therefore size of the population stays constant. If the group is commuter, no people losses job or change commuting or home municipality, no commuter changes age group, if the group consists of agents who have been infected for 7 days, on the next commuting day the group have to eliminate these agents in their home municipality.

Work flow of the model

For the simulation an age structured population, commuting and vaccination data of each municipality is required. The population data should include information of the total number of population, workers, nonworkers, and students of each municipality per age group. Similarly, the commuting data needs to include the exact information of the total number of commuters and their commuting destination per age group and per municipality.

The first step of the simulation is to calculate initial number of infection within the agent-group/ groups are calculated (see the discussion of the disease model below). Then within the commuting model (see the discussion of the commuting model below) collects the commuters of each municipality and distributes them to their destination locations. The outgoing and incoming commuters will cause an increase or decrease of the municipality's population. The following point discusses the underlined assumptions of the population configuration:

- Let i represent the municipalities of Region Twente

$$(i = 1, 2, 3, \dots, 14)$$

- Let P_i represent the total population of each municipality in the region,

$$(i = 1, 2, 3, \dots, 14).$$

- Based on the commuting data commuting population of each municipality are identified including adult commuters (x) and teenager commutes (y).

$$Commuteradults(w) = X(P_i) \quad (4.1)$$

$$Commuterteenager(s) = Y(P_i)$$

- The commuters of each municipality are composed of each of the two agent-groups (teenager and adult). Total commuters

$$(c) = w + s \quad (4.2)$$

- P_s is simulated population per agent-group

$$P_s = \sum (100\% \text{ of } P_i(\text{final})) \quad (4.3)$$

- Where, $P_i(\text{final})$ indicated the populations per agent-group of a municipality after the outgoing commuters are deducted and incoming commuters are added.

$$P_i(\text{final}) = P_i(\text{original}) + C \quad (4.4)$$

$$P_i(\text{original}) + C(\text{incoming}) \text{ or}$$

$$P_i(\text{original}) - C(\text{outgoing})$$

The second step of the simulation could be done in two ways: (i) check if any commuter is infected initially, when infected commuter is detected the disease model calculates secondary infection for the destination municipality. (ii) Without checking if any commuter is infected, calculate infection of all age groups.

The result of both computations is the same that is because if there is no infection in commuters the calculated infection value of the second method becomes zero; which makes the model output the same. For this reason the second choice is selected. Therefore, after calculating the initial infection and distributing commuters the disease model calculates change in number of infection of all age groups and all municipalities. Along with the change in infection change in susceptible and recovered agents is calculated. The process is listed in the disease model below.

The processes of the disease and commuting model continue until infection is detected in Enschede and the individual based model should be activated or it is the end of the week and information should be sent to the urban system model (see commuting model below).

In the model except Enschede the other municipalities are not connected to the individual based model. This is because the existing individual based model is calculating the diffusion pattern of pertussis in Enschede. Therefore, the disease model of the municipality model calculates

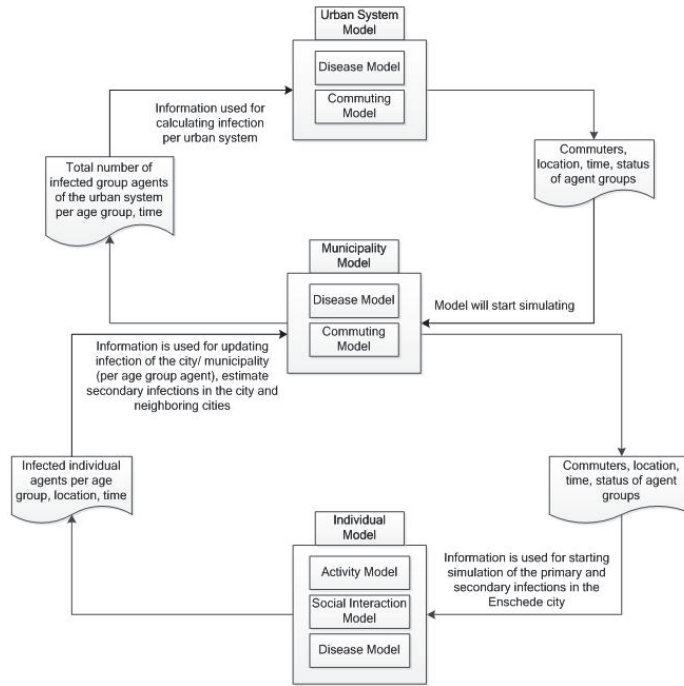


Figure 4.7: Flow of information between the hierarchies

the diffusion of disease of all municipalities but not for Enschede. Total flow of information between the hierarchical levels is shown in the figure 4.7.

The municipality model comprises of both the commuting model and disease model. The commuting model is responsible for detecting the activities of the commuters while the disease model is responsible for calculating the spread of the disease. Detailed description of these models is discussed in the sections below.

4.1.5 Commuting model

Commuting is regular travel between one's place of residence and place of work or school. It sometimes refers to any regular or often repeated traveling between locations. Commuting is among the main factors which facilitate disease diffusion. This is because commuting creates an opportunity for the interaction of people who live in different areas/ places. This interaction could be between an infected and susceptible people which will then be the reason for disease diffusion [20, 32, 61, 53]. The interaction of people could be for different reasons, for example, to work or socialize. For this reason commuting activity of people is included in the model.

In the model, the activity of these commuters during commuting is determined in space and time. It is assumed that all commuters follow the same commuting activity to reach to their destination municipality. Every day all commuting activities might be the same or different, for example, on weekends the commuter agents stay at their home municipalities.

Commuters have two types of activities namely: commute to work or school which is found in other municipality and return back to destination municipality. The activities are followed unless it is weekend or the commuters are infected and should stay at their home municipality. Monday to Friday commuters have two types of activities (going to school/ work or returning back from school/ work). On weekends the commuters doesn't commute so they stay at their home municipality. The commuting activity is done from Monday to Friday; since these days are

the official working/ school days.

Unlike commuting between urban systems, municipality commuting is shorter distance and frequent. Thus, commuter agents are allowed to commute from their home municipality to other municipalities (holon municipalities) twice a day. This means, every morning agent-groups travel from one municipality to the other to work, and to school. In the afternoon the commuters return back to their home. Therefore, every 24 ticks or twice a day, commuters travel from source to the destination municipality. Which indicates that update of information of the municipality level model is twice a day (per 24 ticks).

Friday afternoon some agents commute to the upper level urban system model. The reason of their commuting could be because, these agents are the once who come to the current municipality for work and they return to their homes in the weekends (regular commuters); or they are agents who are traveling to other municipalities to visit friends, relative, or shopping (occasional commuters). These commuters then return back to their school/ working municipalities on Monday morning. Regular commuters are expected to be the same group of agents every week while occasional commuters could be the same or different groups every week. As for the commuters of the municipality level model, the commuters of the urban system model are aware of the day and time of commuting, source and destination address, age group, and type of activity. Every 216 ticks commuters commute from the municipality model to the urban system model and return back every 337 ticks. When an individual is infected it changes its activity and it stays at home; which creates a decrease in the number of commuters.

In the model the transportation or path in which the commuters follow to commute from one municipality to the other is not considered. Therefore, in the simulation, only activities of the commuters after they reach a destination municipality are modeled. Commuter agents can be infected at their home municipality or at the place they commute and vice versa. On the other hand they can infect noncommuter agents.

With the commuter their status such as health status, age group, source and destination location, etc. is sent to the destination municipality at the initial tick. Once they reach to their destination, they interact with noncommuter population of the municipality. Consequently the disease model calculates if there are new secondary infections.

Types of agent groups

As explained commuting contains large numbers of movement of people which can the cause for disease diffusion. To store this information an accurate and efficient method is required. The method should be capable of tracking the changes occurred in the agent-groups.

The issue becomes difficult since individual information is not included in the municipality level model. This means, during commuting agent-groups can be infected or infect others, and at the aggregated level losing information is easy because it there is no information about individual agents. Hence, the information has to be tracked and kept. This is because:

1. To compute disease diffusion between municipalities and groups, the disease model required updated information on the number of susceptible, infected, and recovered individuals. In addition, the model this information to recover infected agent-groups.
2. Commuting model requires accurate information on status of the agent-groups, because this information will be used to collect all agent-groups who are allowed to commute between the municipalities. This means, all infected agents have to stay at their home municipality if it is more than three days since their infection.

There are two possibilities to create agent-groups. The first choice is to create a single agent-group of each category and update the values whenever information needs to be updated. The second one is to create two types of agent groups of each category and represent them as day

time group agent and night time group agent. Day time group agents represent the population of a municipality after commuting is occurred and night time population represents the original population of the municipality (when there is no commuting).

The first choice is easily implementable however during the commuting process the updated information could be lost and detecting the change could become challenging. The second option allows the model to track the changes easily so there will be no loss of information. But the second approach could be challenging because a strong way of detecting the appropriate agent groups for updating is required.

For this study the second approach is selected. Therefore, the model incorporates two type of agent-groups which are night time and day time agent groups.

Day time agent groups are activated when commuting model starts. These agent-groups function of stay alive for 24 ticks, and are disabled after commuters return back to their municipality. All the necessary information of the commuters such as, infection status, number of infections, and moment of infections are transferred to the night time agent-groups when they are activated, because this information will be used by the disease model.

4.1.6 Disease model

The disease model calculates the transfer of disease from one group to the other. This is done by calculating the change in susceptible, infected, and recovered (S, I, R) for every group of every municipality. The computation is done at every tick, which means the three values are calculated at every step and the values are updated.

A major assumption of SIR model is that encounters between infected and susceptible agent-groups happen at a rate comparative to their corresponding numbers in the population. The rate of new infections is also assumed to be constant. Similarly, infected agent-groups are assumed to recover with a constant likelihood at time t . Based on these assumptions it is possible to drive formula for the disease model as discussed in preceding section. In this case, there are two conditions, primary/ initial and secondary, that should be defined in order to have epidemic disease outbreak.

In the case of initial condition there should be at least single infected agent group that is placed into a fully or partially susceptible population. Once initial infection is introduced the disease spreads and new infections start to appear at the defined disease transmission rate. It is a situation in which the number of infected agent-groups increases from the initial value. That means the total number of infections produced must exceed the value of initial infection. This total number of agent groups infected during an epidemic is called secondary infection contrary to the primary initial infection. The step by step detailed theory of the disease model is discussed in the proceeding sections.

A model of disease spread where agent-groups were susceptible to a disease, potentially contracted the disease, and then becomes recovered or immune to future infections (from the disease) is called SIR model. In this disease model there are three groups, namely susceptible $S(t)$, infectious $I(t)$, and recovered $R(t)$. As it has been discussed in section 2.2 all susceptible are equally at risk of infection (homogenously mixing) and total population was assumed to be constant over the course of infectious period (no deaths and births. Where,

$$N = S(t) + I(t) + R(t) \quad (4.5)$$

Let,

$$St = \text{thenumberof susceptible agent groups in the population at time } t \quad (4.6)$$

I_t = the number of infected agent groups in the population at time t

R_t = the number of recovered agent groups in the population at time t

N = the population size.

Correspondingly if the three groups are divided by the total population size N , fraction of population at time t is computed.

$$st = St/N \text{ (the susceptible fraction of the population at time } t) \quad (4.7)$$

$it = I_t / N$ (the infected fraction of the population at time t) $rt = R_t / N$ (the recovered fraction of the population at time t)

It should be noted that each agent-groups in the population is in one of the three groups, i.e.

$$St + It + Rt = N \text{ and } st + it + rt = 1 \quad (4.8)$$

Here the assumption was that all agent-groups potentially move from the susceptible group to the infected group and then to the recovered. Assume that each infected agent-group contacts (γ) agent groups in each period of time on average. Each contact may not result in transmission of the disease. Possibly only (α) percent of the contacts result in transmission. Consequently the potential number of transmissions may be at most $(\gamma) * (\alpha) = (\beta)$. (β) is the average number of transmissions possible from a given infected person in each period. (β) is the probability of susceptible agent-groups to become infective between the time t and $t+1$. Since only st percent of the population is susceptible each infected agent-groups generates only $(\beta) * st$ new infections each period. Each infected agent-groups recovers at some rate (μ) .

Given the current state of the population in period t described by St , It and R_t the following equation can be derived to describe the SIR disease model.

From susceptible population on average $((\beta) * (st) * It)$ will be lost. Hence in period $t + 1$,

$$St + 1 = St - (\beta)stIt \quad (4.9)$$

$$R_t + 1 = R_t + (\mu)It$$

$$It + 1 = It + ((\beta)stIt - (\mu)It) = It((\beta)st - (\mu))$$

Similarly each of these variables can be written in terms of the population fractions (change rate)

$$st + 1 = st - (\beta)stit \quad (4.10)$$

$$rt + 1 = rt + (\mu)it$$

$$it + 1 = it(1 + (\beta)st - (\mu))$$

These two equations should equal to 1, i.e.

$$st + 1 + it + 1 + rt + 1 = st + it + rt = 1 \quad (4.11)$$

In SIR model there has to be two conditions (initial and final condition). In the initial there has to be an initial infected agent-groups ($it > 0$) in order for there to be more infected agent-groups. In other words, there are no infected agent-groups if $it = 0$ consequently there is no possibility of anyone else being infected. Therefore, in a model initial infection could be generated manually by

assigning a certain value to a specific groups or making the model to choose a random group and assign initial infection (as explained above).

In most cases it will either be increasing or decreasing. From equation $it+1 = it (1 + (\beta)st - (\mu))$,

let $pt = 1 + (\beta)st - (\mu)$, where pt is the threshold for the SIR model with a constant population. If pt is greater than 1, $it+1 > it$ where as $it+1 < it$ if pt is less than, increasing and decreasing the number of infected agent groups. Obviously $pt > 1$ or $pt < 1$ it is not at a steady state.

In equation $pt = 1 + (\beta)st - (\mu)$, there are constants 1, (β) , and (μ) and as well as st variable which is decreasing whenever $it > 0$ and hence $pt+1 < pt$. Therefore, if it is increasing or decreasing it is doing so at a decreasing rate or increasing rate respectively. This indicates that the infected population will always vanish in the long run if the population is at steady state (constant population).

It has been discussed that st is decreasing function for $it > 0$ and rt is increasing function and both st and rt will reach a steady state value since it goes to 0 at t goes to infinity.

Usually pertussis has three stages catarrhal, proximal and convalescent. In reality, the second stage which is about 1 to 6 weeks s is assumed to be period of infectiousness. The length of time that it takes to become unable to pass on the disease assumed to be 3 ($(\mu) = 1/3$). The transmission rate (β) is perhaps a little challenging to estimate because how changing the value of (β) (0.1, 0.2, 0.3) affects the course of the epidemic. For example, if initial conditions for $(\mu) = 1/3$, $(\beta) = 0.1$ or 0.2 or 0.3, $I_0 = 1$, essentially nothing will happens because It will only grow if $1 + (\beta)st - (\mu) > 1$ and if $(\beta) < (\mu)$ then an epidemic will not occur.

If (μ) is smaller (people recover from the epidemic faster) the epidemic is smaller i.e. a shorter period of infectiousness implies that an epidemic is of smaller magnitude and lasts for a shorter period of time.

4.2 ADJUSTMENTS TO THE EXISTING MODEL

Disease model

Naturally or through some kind of medicine, the immunity levels of individuals within a society can differ from one another. But, usually individual within the same or similar age groups have somewhat similar immunity to a certain disease. As described in chapter two individuals which have similar age have similar immunity to pertussis unless there are other special cases to the individual.

For this reason in the existing pertussis model the immunity level is assigned per age groups. Ages from 0 to 3, 4 to 9, and 10 to 18 are the three age groups which have vaccination level assigned in the current model. The vaccination level status is one of the attributes of an agent. Figure 4.8 shows the adjusted conceptual model of immunity levels and disease infections.

One of the limitations of the existing model is the lack of the recovery in infected agents as explained in chapter 3. Therefore, to include this idea different assumptions are taken. All agents whose time since infection is 6 weeks or 2016 ticks they should recover. This means they should start to follow their activity pattern, their immunity level should get back to normal and become fully immune, the number of infections in the model should be reduced. To apply this concept a change is required on the conceptual model of the disease model. However, due to the limitations of time the immunity level of the individual based model is not included in the individual based model. However, recovery of agent recovery is included in the middle level.

Activity model

As explained in the above sections the existing model is connected with the municipality level

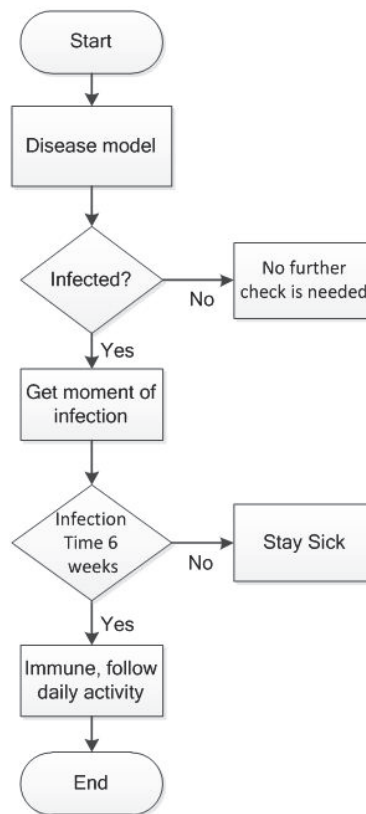


Figure 4.8: Concepts for adjustment of the existing disease model

model by its commuters. At the time of commuting the agent groups are sent to the individual based model. Yet, this model doesn't consider aggregated groups. Therefore, when they enter the city they are decomposed and become individuals.

On their commuting location these commuters perform activities which are determined in space and time. The activity of the commutes is managed by activity model of the individual based model. This means, during arrival these commuters are assigned to schools or work places. As the individual level model the activities of the commutes is stored in a daily activity table. The activity table specifically states the time and commuting location of these commuters (see figure 4.9). The activity is always constant that is if commuter is a student commute to school or commute to work if the commuter is worker.

By reading from the table, the activity model assigns to the agent to the appropriate location. But, still this model requires the source address of these commutes to assigns the agent to the appropriate location. This indicates the commuting location of these commuters is based on distance.

After the commuting location is allocated, the commuters are treated as the population of the city and all assumptions of the individuals of the individual based model are applied on them. Such as, these commuters can be involved in a group activity, can interact with people, and be infected and infect others.

However, these commuters are not thought to perform activities out of their school or work. This means, when the commuting time is over these commutes are collected and return back to their source location.

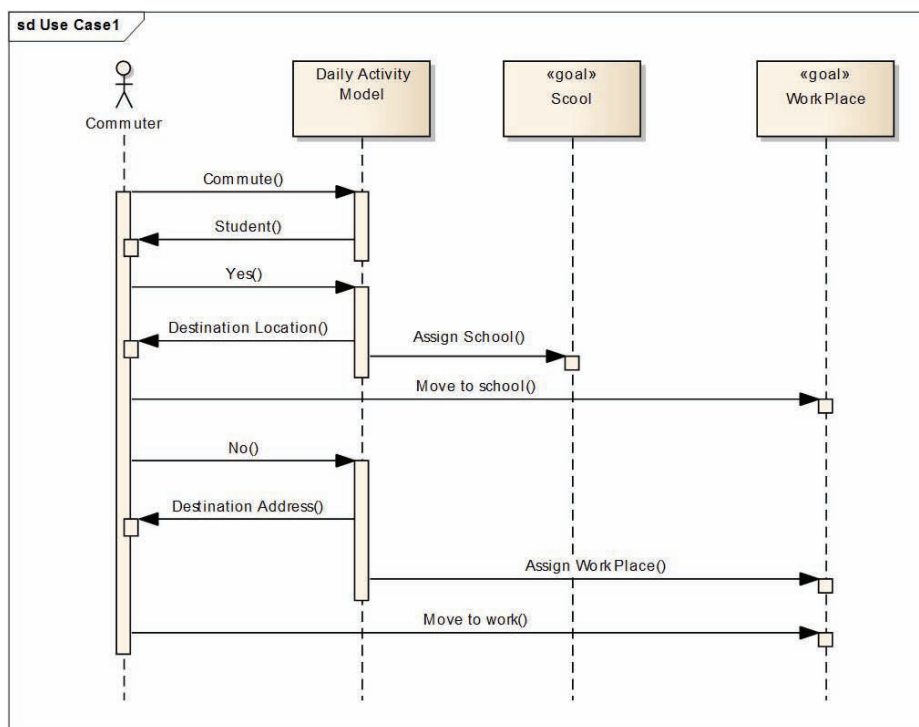


Figure 4.9: Daily activities of the commutes in the individual level model

Chapter 5

Simulation data preparation

Chapter four contains descriptions of the conceptual model which was created based on the analysis in chapter 3 and literature review in chapter two. The general aim of the created conceptual model is to create a methodology for scaling and increasing the performance of the pertussis model. The created conceptual models include conceptual model for the hierarchical models and for adjustments of the existing model. The proceeding section is a description of the data preparation and implementation of the conceptual model.

For the simulation population, commuting and vaccination data is required. For this reason from different sources the data have been collected and prepared. The data preparation was done according to the conceptual model prepared in chapter four, thus trying to understand the current data selection and preparation requires understanding of the objective of this study as well as the prepared conceptual model. Subsequent to the data preparation process the municipality level model is implemented.

Due to the availability of lower level individual based model for the Enschede municipality, the implementation of the middle level/ municipality model is done by using datasets of Region Twente. However, by changing the used datasets the implemented models can be used for all municipalities and urban systems/ regions of the Netherlands.

5.1 SIMULATION DATA PREPARATION

For the simulation population, commuting, GIS environment and vaccination data are required. For this reason, the downloaded population data (CBS population), the GIS environment and the vaccination data gained from ITC are used. The usage of the prepared simulation data is illustrated in figure 5.1.

The required population data should include age structured population data for all municipalities. For the commuting model, commuting data which includes the commuters commuting information per municipality and per age group is required. The commuting data should include the number of commuters, their age as well as their commuting location. To use the data in the model the commuting location and number of commuters per age group is crucial. Furthermore, the model requires vaccination levels per age group.

For this reason, four types of data are prepared which comprises:

- Twente population data set
- Twente commuting data set
- Twente vaccination data set
- Twente spatial dataset

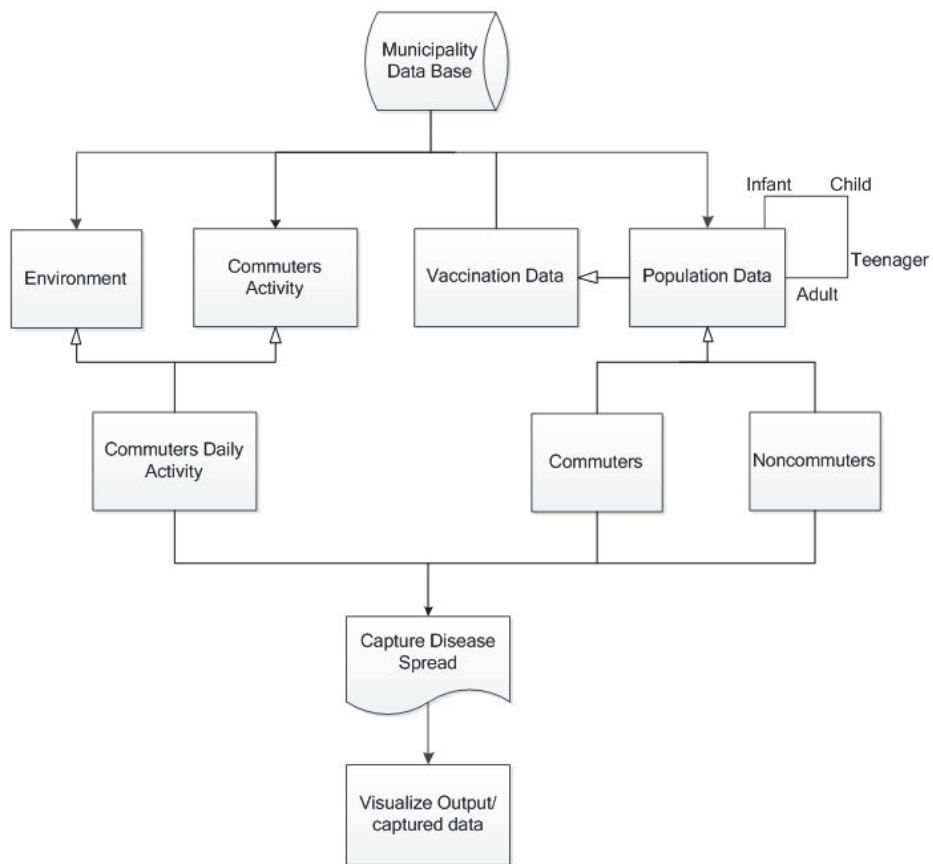


Figure 5.1: Abstract diagram of the model

Table 5.1 Twente Population (2009, 2010) per municipality

Gemnr	Regions	2009	2010
141	Almelo	72428	72602
147	Borne	21106	21330
1774	Dinkelland	26066	26058
153	Enschede	56071	157052
158	Haaksbergen	24495	24486
163	Hellendoorn	35846	35791
164	Hengelo (O.)	80925	80772
1735	Hof van Twente	35151	35468
168	Losser	22589	22647
173	Oldenzaal	31764	31974
1742	Rijssen-Holten	36787	37080
183	Tubbergen	20992	21145
1700	Twenterand	33605	33580
189	Wierden	23467	23447

5.1.1 Introduction

The census dataset of Twente region is downloaded from the official website of the Central bureau of statistics of the Netherlands. The dataset is prepared in Dutch, thus for the translation Google translator have been used. The census data is prepared for the years 1960 up to 2010 per 10 years period. The derived dataset is in a tabular form containing population growth of Twente Region by birth, death and migration by sex, and population data per municipality which contain all age groups ranging from 0 - 5 to 95 and above. In addition, number of jobs per municipality is included in dataset.

In the dataset each municipality has a code. The municipal code indicates the numerical designation of municipalities. The code is adopted by the CBS in consultation with the Ministry of the Interior and Kingdom Relations (BZK). This four digit code is associated with the name of the municipality. Thus if rename a municipality, change the code.

The population per municipality is derived from the Structure Count GBA (GBA) and organized by CBS. For most municipalities data is included for all years. However, the dataset doesn't encompass data for the municipalities of Dinkelland, Hof van Twente, Rijssen-Holten, and Twenterand from the years 1960 up to 2000. For this reason, for the simulation population data which is available in the dataset will be used.

The simulation data is prepared by using data of the year 2009, because the data for the same year was used in the simulation of the individual based Abdulkareem [1](2010). Moreover, comparison of the population dynamics with in these two years (the year 2009 and 2010) indicated, during these years the increase of the population is minimal which is 2140 (see table 5.1 and figure 5.2).

In 2009 the total population of Twente region is approximately 621292. Half of the population of 0.6 million live in urban areas, predominantly in the three major municipalities of Enschede, Hengelo and Almelo. As Enschede is the main municipality of the region, it has almost 160,000 inhabitants (As explained in section 4.3.2. and shown in figure 5.2 and table 5.1).

Twente have a number of large companies with an international reputation. The current state of the Twente economy was strongly influenced by its economic history as an old industrial re-

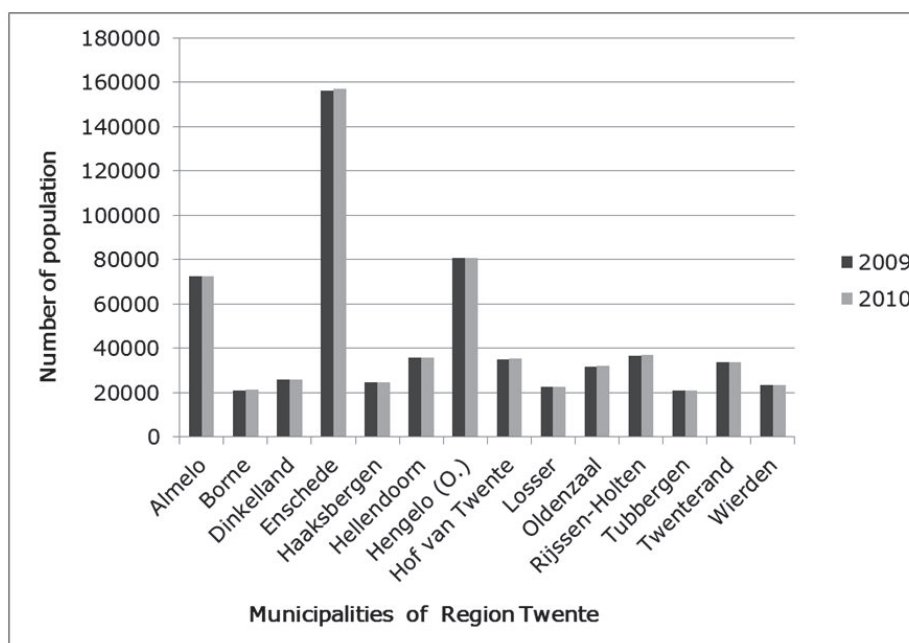


Figure 5.2: Twente Population (2009, 2010) per municipality

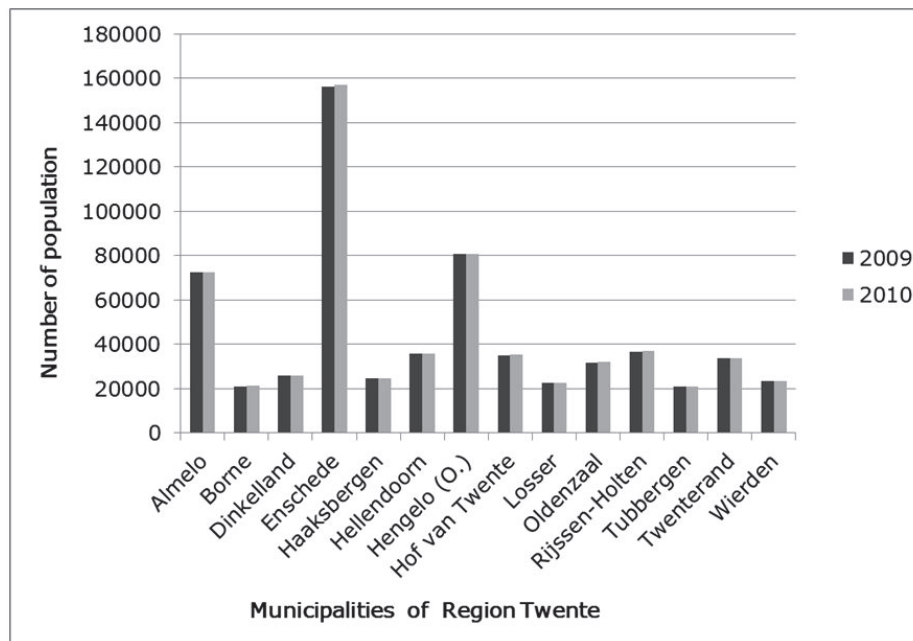
gion. From the 1830s, a textile processing complex grew up across the urban areas, with Hengelo specializing in textiles machinery and other towns (Enschede, Almelo) in weaving and tailoring. The textiles agglomeration began a period of secular decline from the 1940s onwards, however, industrial analysts believed that the decline was not terminal. The machinery sector also played a big roll and accounts for around 10% of regional employment. Generally, Twente to this day remains a far more industrial (manufacturing) region than the rest of the Netherlands; the industrial sector of the region accounts for a large employment regionally and nationally [34].

Economic activities can be an indicator of the types of jobs in the city and can provide general information on the local workforce. Employment opportunities can also affect the growth rate of cities and allow people to settle in the city. If residents cannot find work in the city, then they will most probably move to the nearby cities or move elsewhere for jobs. Activities have spatial locations and mobility relates to which activities are taking place where. Some are routine activities, because they occur regularly such as commuting and shopping and each type of activities has its own specific mobility requirements. They involve commuting to work, for example residential to industrial or commercial or administrative or schools, and shopping, for example residential to retailing. Generally, there is diversity of urban activities in a diversity of urban contexts. With this in mind, Twente economic data have been considered and illustrated in figure 5.3. As can be seen from the figure Enschede, Almelo and Hengelo, relatively have better job opportunities than the other municipalities (Table 5.2 and Figure 5.3).

In the data, the proportion of persons employed outside the municipality is expressed as a fractional percentage of total employment from 18 to 65 years. This data is derived from the Social Statistical Database (SSB). The percentage is listed in more than 10 employees who work in another municipality. In addition the data set includes the proportion of students, who commute every day to school which is found in other municipalities than their home municipality. The data of commuter students is also acquired from the same source (SSB). And as the other data's the student commuting data also is found in a general manner. The data contains the total number of commuters per age group and per municipality.

Table 5.2 Number of jobs per municipality

Gemnr	Name	Jobs
141	Almelo	33755
147	Borne	4668
1774	Dinkelland	8870
153	Enschede	46398
158	Haaksbergen	8138
163	Hellendoorn	10573
164	Hengelo (O.)	31271
1735	Hof van Twente	12094
168	Losser	5372
173	Oldenzaal	16302
1742	Rijssen-Holten	15796
183	Tubbergen	7125
1700	Twenterand	9195
189	Wierden	6618

**Figure 5.3:** Number of jobs per municipality

In Netherlands the proportion of people who commute from a smaller municipality to large municipalities is much bigger than the corresponding figure for other municipalities. In 2009 in Twente region a total number of around 9626 commuters per municipality commute daily to school/ work. The commuting pattern of the inhabitants of region Twente is illustrated in appendix.

As explained above the data in this study is a subset from CBS, a database including the total number of individuals who commute to their working destination and return back to their origin (home municipality). The simulation model has age structured agent-groups which divide the population into 0-3, 4-12, 13-18, and 25-55. As it has been discussed in preceding section adult groups age 55+ and age 19 to 25 are excluded in this study. Thus, some assumptions are necessary to estimate the average incoming and outgoing commutes of each municipality to match with group-agent structuring of the model. In addition, CBS data has no description about the destination of outgoing commuters except their total number per each municipality per age-group at 5 years intervals. For example, 0 to 5 years, 5 to 10years, 10 to 15 years, 15 to 20 years, 20 to 25 years, 25 to 30 years, 30 to 35 years, 35 to 40 years, 40 to 45 years, 45 to 50 years, 50 to 55 years and 55 years older.

On the basis of these divisions, total number of incoming and outgoing commuters were estimated and allocated randomly for each municipality as per the defined age structured agent-groups in the simulation model. The assumption was that the number of commuters in CBS data per age was assumed to be same though in reality can vary. On the basis of this assumption, for example, the total number of commutes for age-group between 0 to 3 is three-fifth of the total number of commutes for age-group between 0 to 5. Suppose the total number of commuters in CBS data between 0 to 5 age-groups is 100. Hence, the estimated total number of commuters for age-group between 0 to 3 is 60 ($3/5 \times 100 = 60$). The same can be applied to the other age-groups to estimate the total number of commuters as the defined age structured agent-groups. The estimated total number of commuters for the incoming and outgoing commuters utilized in this study is illustrated in Appendix.

5.1.2 Simulation data

2.1.2.1. Population of Twente (Agents)

The simulation requires population per municipality data. As explained in the above sections the Twente census data is used for the simulation. During data preparation a subset of data is generated from the downloaded population information based on the purpose of the model, which is simulating the pertussis diffusion pattern in the region and using the combination of hierarchical modeling, metapopulation and agent based modeling for scaling agent based disease diffusion models.

The subset data will present the population of the each municipality, which are structured in to ages groups, commuters and noncommuters. These categories of the population are grouped according to their age groups. Even though in the data all age groups ranging from 0 - 5 to 95 and above are included for the simulation, the groups which were included in the individual level model of Abdulkareem (2010) are selected and used. The following table shows the prepared population data per municipality:

Thus the data is grouped into age groups which are 0 to 3 infant, 4 to 12 child, 13 to 18 teenager, and 25 to 55 adult. These groups have a list of common attributes (for example, age group, immunity level, health status, municipality name, municipality code, population size, etc.). Therefore, here individuals are grouped in to infant, child, teenager, and adult according to their age.

Furthermore, the population groups consist of commuter and noncommuter individual/ agents. The agent groups are the most important parts of the simulation. Noncommuter agents

Table 5.3 Population per municipality and age group

Municipality Name	Muni Code	Age 0 to 3	Age 4 to 12	Age 13 to 18	Age 19 to 55
Almelo	141	2626	8209	4436	46085
Borne	147	842	2291	1408	13168
Haaksbergen	158	800	2747	1527	15004
Hellendoorn	163	1264	4092	2168	22251
Hengelo	164	2814	8938	4848	51024
Losser	168	710	2431	1339	14240
Oldenzaal	173	1126	3541	1985	19848
Tubbergen	183	751	2794	1555	12957
Wierden	189	827	2744	1536	14528
Twenterand	1700	1319	4230	2149	21108
Hof van Twente	1735	1108	3929	2182	21271
Rijssen-Holten	1742	1584	4895	2667	22255
Dinkelland	1774	878	3234	1854	15613
Enschede	153	3193	11620	6302	66331

are the once that don't commute to other cities and commuter are the once who move from one city to the other for work or school (see section 4.1.2). As explained in section 4.2.1 the commuter agents commute twice a day and all commuters perform the same type of activity to reach their destination location; which is in the morning go to work or to school in other municipality then in the afternoon return back home. As a result, these agents could be infected and infect other commuter or noncommuter agents.

The daily commuting activities is tracked by the population who commute every day to other municipalities. These commuters are generated in groups, and as explained in section 4.1.2 for this level the groups of commutes are considered to be workers and students. In addition, it is assumed that all commuter teenagers are student and all commuter adults are workers; this indicates the age groups that are allowed to commute are 13 up to 18 which are students and 25 to 55 who are workers. Each municipality could have certain number of commuters per age group (adult and teenager). Therefore, the total groups of commuters (total teenager and adult groups) which commute between municipalities are 392.

Daily commuter agents have two types of activities, going to school/ work and being at home municipality. Monday to Friday these agents commute to the destination city. These commuters go to work/ school at 9: 00am and return back to their home at 7: 00pm; this activity continues every day except for weekends in which agents are supposed to stay at their home municipality (see table 4.1).

To assign the attribute values such as population size, age group, destination location agents use the census data and other datasets that are used (such as vaccination level, commuting activity). These agents differ in terms of attributes that effect disease diffusion including physical (susceptible, infected, immune), social (age, commuter, noncommuter) and economic (workers, students, non-workers).

It has been observed that in the Netherlands the residents are not forced to be vaccinated unlike other countries. That means vaccination is based on the willingness of the residents, for example, as it has been illustrated in figure 5.4, rijssen-holten seems very religious municipality because the vaccination level is relatively low. Other municipality in general has relatively higher

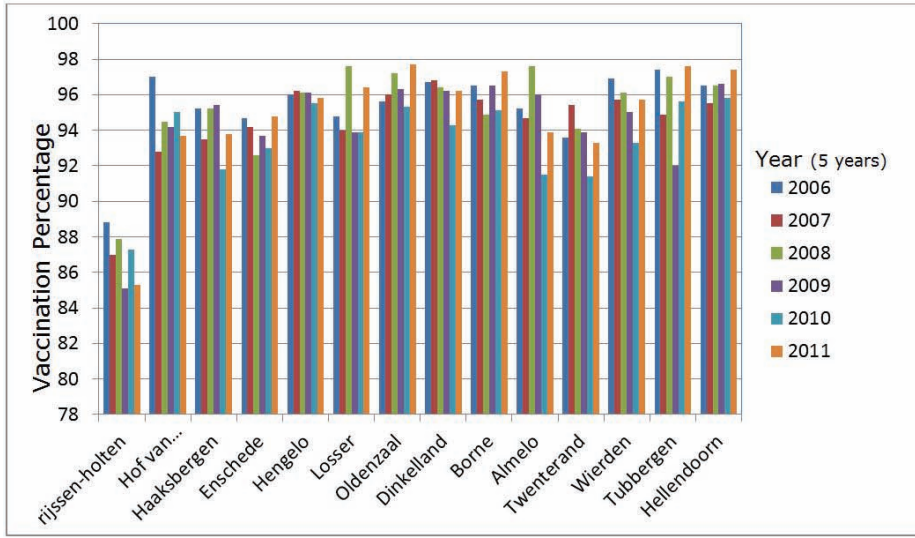


Figure 5.4: Vaccination data of Twente Region per 5 years

vaccination level.

Generally, for the simulation three types of data's are prepared and stored in a CSV file.

- Population data
- Commuters data
- Vaccination data

2.1.2.2. Twente GIS environment

As explained in the above section from the census data of Twente region population per age groups, commuter per age groups have been created. In addition as those agents require an environment to live in this section describes the preparation steps and the prepared GIS environment data. To apply the behavior of agent during the simulation an environment is needed which assists the agents to interact.

The spatial data was provided by ITC. The dataset contains of all the topographic features of Netherlands in a vector format at a scale of 1:10,000; for the year in 2004. From the data, necessary attributes are used and the missing ones are created and modified. The created GIS environment is two types (for visualization and for actual simulation). The spatial dataset used for visualization consists of the administrative boundary of the municipalities (see figure 5.5). The spatial data created for visualization purpose is in order for users to get the idea of the shape of the region and its municipalities.

The second group of spatial areas is used to give agents the geographic address (coordinates) of places of their activity and living environment.

In this region a municipality sometimes consists of three or four settlements that are scattered in space. The centroid of the municipality does not coincide with any of these settlements. Therefore, to create the geographic locations there are two options: centroids and municipality center.

The first option is to create the centroids at each central location (settlement) of a municipality. When there is one or more centroids in a municipality; which could make the model mode realistic, since the actual centroid locations are preserved.

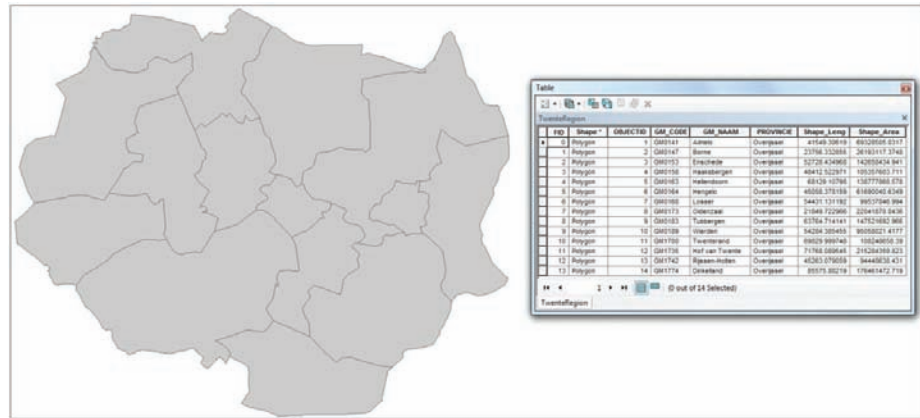


Figure 5.5: GIS environment used for visualization

The second option is to put the agents in one centroid location which is created at center of a municipality. But the second option doesn't preserve the realistic location of cities. Yet, since this idea supports the assumption of the model in putting all agent-groups in one location and using option one leads to the violation of the model assumption, for this study the second option is selected and used.

Therefore, in this model the agents are placed in centroids which are created in the center of each municipality using ArcMap. Each municipality contains one centroid and it is used to place all agents (see figure 5.6).

Centroids have the following attributes Location (x and y coordinates), OBJECTID, Shape (which is point), PROVINCE (which indicates where in the Netherlands Twente region is found), Muni_Name (which is the municipality name), GM_Code (which is the municipality code), Pop_Type (which indicates if the population who live in the centroid is commuter or not), Pop_Size (which is the size of the population who live in the centroid per age group).

Each municipality has four noncommuter agent groups (infant, child, teenager, and adult centroids) and 26 (2 x 13) commuter agent-groups (TeenagerStudent and WorkFull). Therefore, the total created agent-groups are 420 (see figure 5.7).

At the beginning of the simulation the created agents are placed in the centroids according to the age groups. Each agent group has a Location (source and destination x and y coordinates for commuters), OBJECTID, Shape (which is point), PROVINCE (which indicates where in the Netherlands Twente region is found), Muni_Name (which is the municipality name), GM_Code (which is the municipality code), Group_type (which indicates if the agent group is commuter or not), Agent_Type (infant, child, teenager, and adult), Group_Size (which is the size of the population who live in the centroid per age group), and Health status.

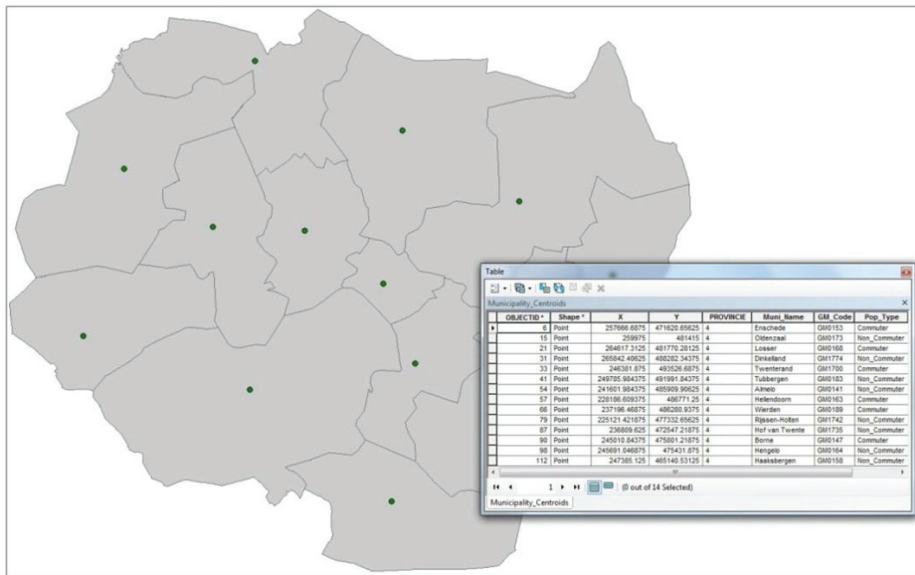


Figure 5.6: Twente Region Centroids (places used as living places for commuter and noncommuters agents)

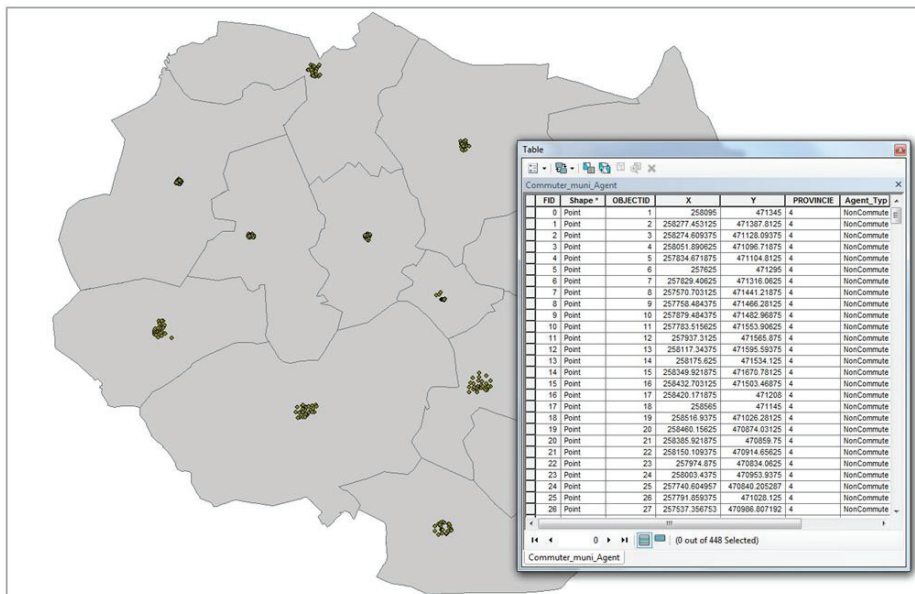


Figure 5.7: Agents of Twente Region per age group and commuter and noncommuters

Chapter 6

Model implementation, results and discussion

In chapter four, a conceptual model for scaling agent based systems have been created. The created approach is based on the analysis in chapter three and the analysis of scaling methods selected in chapter two. For the implementation of the model population, commuting and vaccination data's are required. For this reason, different data were studied and preparation of these data's has been explained in chapter four.

Overall the proceeding chapter consists of descriptions of the implementation platform and programming language, and the stages that are followed to implement the model (group agent creation, commuting model and disease models).

6.1 IMPLEMENTATION TOOLS AND LANGUAGE

Numerous simulation/modelling systems are available for creating agent-based models which support the direct integration of geospatial data, for example Swarm, Repast, OBEUS, NetLogo and StarLogo (see Castle and Crooks, 2006 for a recent review).

Repast is one of the few simulation/modelling software systems that supports the integration of geospatial data both raster and vector datasets (especially that of vector-based geometries) into an agent-based model. Furthermore, it contains Repast J, which uses Java programming language.

Object-Oriented Programming (OOP) Language is a paradigm which is build based on the decomposition principle of hierarchy theory [11]. OOP includes C, C++, visual basic, and Java. To improve the quality of a program or software OOP is the best programming language [11, 70]. Also, this programming language is the best way to build complex, but stable and efficient simulation models [11, 70].

A number of research and development investments use OOP language and produced several ABM software environments that are now freely available. These include Repast (Recursive Porous Agent Simulation Toolkit), Swarm, NetLogo, and MASON (Multi agent simulation toolkit), as well as many others.

Recursive Porous Agent Simulation Toolkit was originally developed by David Sallach and other researchers at the University of Chicago and the Argonne National Laboratory in 2003, it is now managed by the non-profit volunteer organization ROAD (Repast Organization for Architecture and Development). This software is developed as a pure Java implementation. Repast is consists of a set of tools which was originally based on Swarm. This simulation toolkit is specifically designed for social science applications. Repast permits the systematic study of complex system behaviors through controlled and replicable computational experiments. According [11, 70] Repast is the best software documentation that is built for the implementation of agent based model.

Generally, using Java based Repast for creating a model is the best choice for more complex models as well as GIS models. Furthermore, models created with java based programming language are plate form independent. Therefore, because of its advantages Repast J is used for implementation phase of this study.

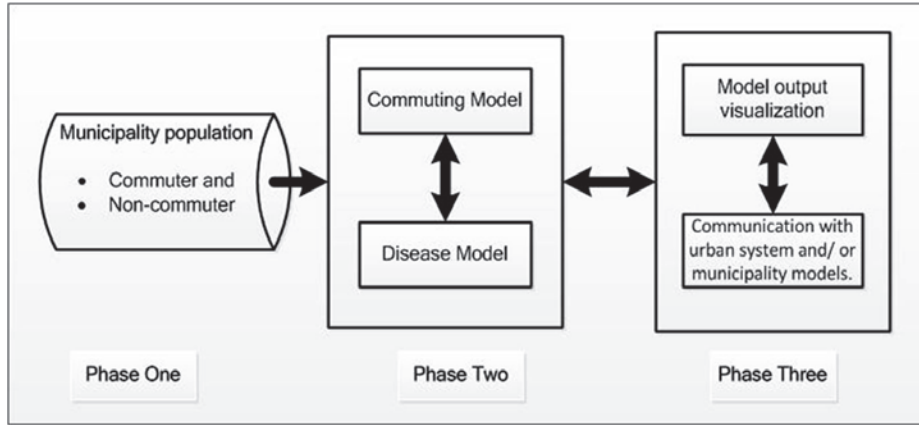


Figure 6.1: Stages of the implementation process

6.2 IMPLEMENTATION STAGES

As explained in the above sections the developed conceptual model consists of three hierarchies including the individual based model with some adjustments. However, in the implementation process due to the time limitations an attempt was done only to implement the middle level municipality model. The choice was supported due to the availability of the existing model and because this model is the mediator between the two models (urban system model and individual based model).

Implementation of these models is done in three stages. These are: stage one agent creation, stage two implementations of the sub models, and stage three visualization of the model output.

6.2.1 Stage one: Activating population

As explained in the conceptual model the municipality level is built based on metapopulation modeling approach. In the model the population is grouped in to different groups, namely: infant, child, teenager, and adults. Infants consists of population that are found from ages 0 to 3, children are all populations between the ages 4 to 12, teenagers are between the ages 13 to 18, and adults are between the ages 19 to 55. As explained in the above sections the population that is between the ages 19 to 25 is excluded from the model (see section four for detail).

Agent of the municipality level model is a group agent. A group agent is a collection of population that are structured by age. Municipality group agent could be commuters and non-commuters. The information of these group agents is read from a csv file which was prepared in the data preparation process.

Each group agent consists of different attributes such as address (municipality code), population size, age group, group type (commuter or noncommuter), vaccination percentage, number of susceptible, number of infected, and number of recovered. Therefore when a group agent is created it contains the listed information.

Municipality source or destination code, population size, age group, group type, and vaccination percentage are static variables which are acquired from the stored csv file. However the number of susceptible, number of infected, and number of recovered are dynamic variables which can change their values when the disease model is activated.

In the implementation process agent groups can be created in two ways. One way is to create the agent groups in ArcGIS and load them at the beginning of the simulation. The second way is to store information of these agent groups in a csv file and activate them at the beginning

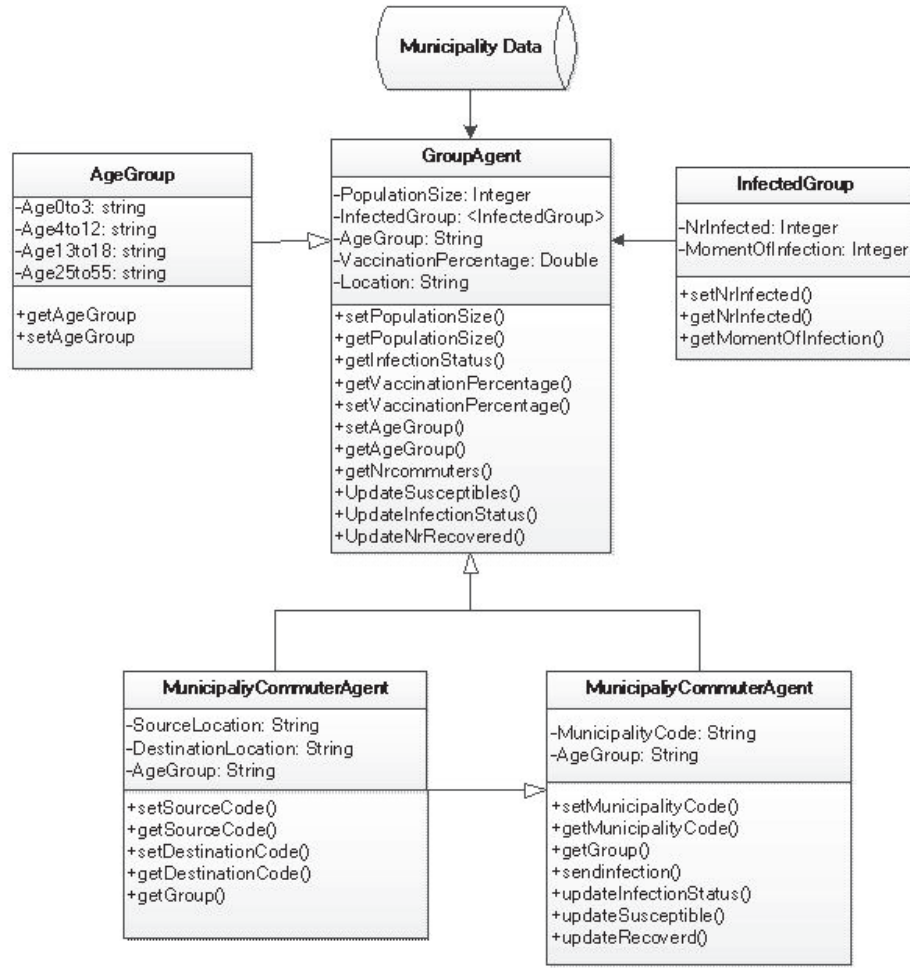


Figure 6.2: UML diagram of the modeled agent groups

of the simulation. The first method is computationally expensive as compared to the second method. In addition, the first method is appropriate when social interaction of simulated agent needs to be studied. For this study the second method is selected because the modeled entities are in aggregated level, and concept of the social interaction or individual contact is not considered.

At the beginning of the simulation agents groups are loaded from the stored csv file and assigned to their locations. However, in the model GIS context is not implemented so the agent groups store the values but they are not assigned spatially. The class diagram illustrates the property of the created group agents (see figure 6.2).

To create the group agents four classes are created. The first one is the main class which stores the overall information about the group. The second and biggest class is the MunicipalityAgent class which contains every information of the municipalities agent groups including information of the commuters. The class inherits properties about the group agents from GroupAgent class. The getter and setter functions store the information of the group agents to the class.

In the model there were two choices for creating the group agents. One is to create a single group agent of each category and update the values whenever information needs to be updated. The second one is to create two types of agent groups of each category and represent them as day time group agent and night time group agent. Day time group agents represent the population

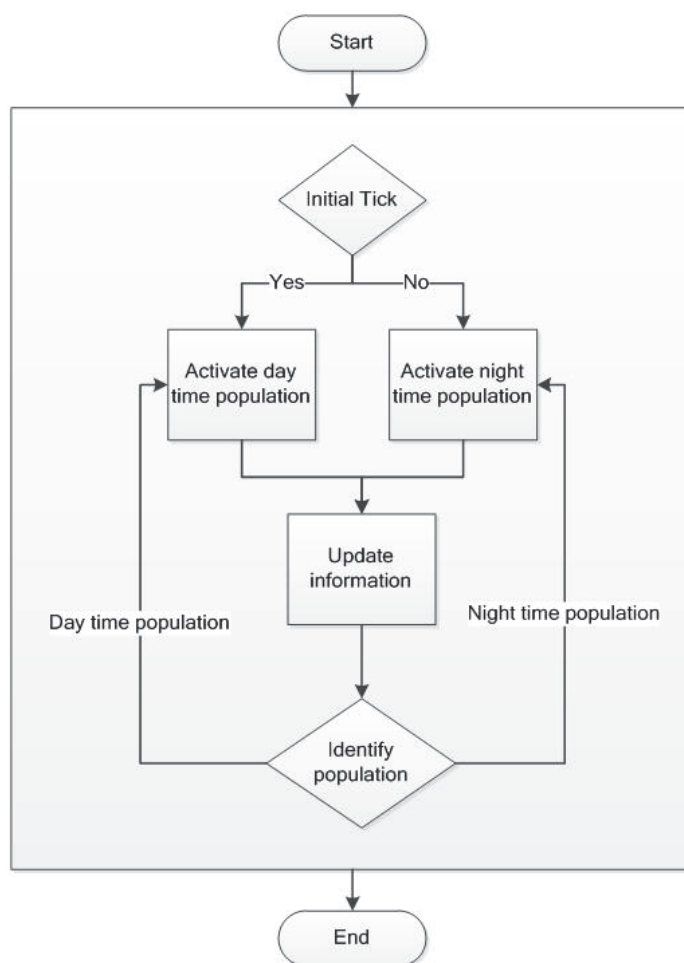


Figure 6.3: Types of agent groups

of a municipality after commuting is occurred and night time population represents the original population of the municipality (when there is no commuting).

The first choice is easily implementable however during the commuting process the updated information could be lost and detecting the change could become challenging. The second option allows the model to track the changes easily so there will be no loss of information. But the second approach could be challenging because a strong way of detecting the appropriate agent groups for updating is required. Due to its advantages the created group agents are based on the second approach (see figure 6.3).

In the model, two types of agent groups, for each age group are created. At the beginning of the simulation information is stored to the day time population and after commuters return back to their home municipality, the night time agent groups are activated. The difference between the two categories is the number of populations which are contained within the group agents. This means, the day time and night time population differ from one another by the number of incoming and outgoing commuters.

In the simulation the behavior of the agent groups is constant, that is if they are noncommuters they stay at their home municipality otherwise they commute from their home municipality to the destination municipality every 24 ticks. Other than this the activities, interactions, and tasks these agent groups perform is not modeled. Figure 6.4 illustrates the implementation

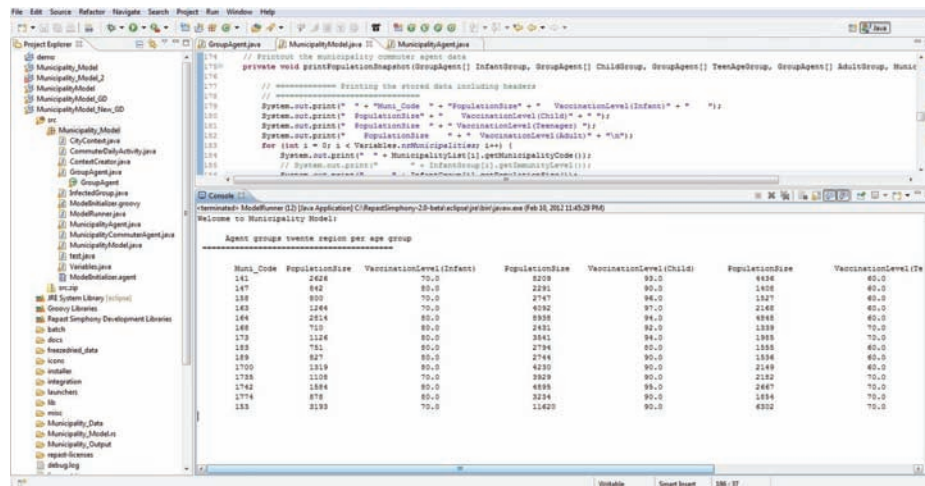


Figure 6.4: Result of the model after the agents are activated

process's output after the group agents are created and stored.

After the agents are created the second step is to create the commuting model.

6.2.2 Stage two: Implementation of the commuting model

In a large scale models commuting behavior of the population should be included. This is because commuting creates an opportunity for the interaction of people who live in different areas/ places. The interaction could be between infected and susceptible people which will then be the reason for disease diffusion [20, 32, 61, 54]. For this reason the in the approach commuting models are considered.

In the model, the activity of these commuters during commuting is determined in space and time. All commuters follow the same commuting activity to reach their destination municipality. Therefore, the commuting model is accountable for sending and collecting the commuters from their source location to their destination location and vice versa.

The commuting model is a submodel of the disease model. The main task of this model is to send and receive commuters from their source location to their destination location according the commuting g time.

There are two types of commuters, which are: students and workers. However, this model is built based on aggregated behavior, activity and interaction therefore; the model only collects the commuters from their source location and sends them to their destination location and vice versa. This means the activities or interactions which occurred during commuting or after commuting is not modeled.

At each commuting time (every 24 ticks) the number of commuters of a municipality could decrease or turn to zero. Because when infection is occurred the model reduces them from their respective group agents. Then the reduced agents stays at their home municipality.

The commuting model is not only responsible for the commuters of the municipality level model. Yet this model also sends/ receives incoming and outgoing commuters to/ from commuters of the urban system model every (sends them every 216 and receives them every 337 ticks). However, because the implementation at this point is only for the municipality model, currently, the model is only responsible for the municipality level commuters. Figure 6.5 illustrates the implementation outputs of the commuting model.

Figure 6.5 illustrates the activity of the day time and night time commuter agent-group. At day

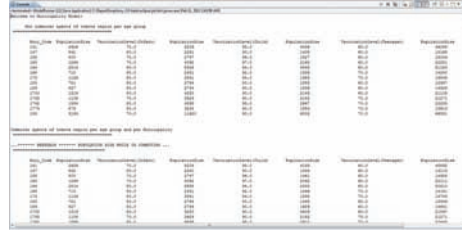


Figure 6.5: Result of the model after implementation of the commuting model

time the number of day time population increase by the same amount of incoming commuters while decreasing by amount of outgoing commutes. This indicates the commuting of people from one place to the other which are the driving forces of disease diffusion.

At night, the night time agent-groups are activated. And start their activity. Whenever commuters start their activity the night time population is disabled. Yet, at their return all changes occurred during commuting which could affect the night agent-groups are updated and stored to those groups.

From the result of the displayed commuting model it is possible to see the increase and decrease of those two type of agent groups. In addition, it can clearly show that the two agent-groups categories share the same attributes.

Generally, the main aim of the commuting model is to send commuters in the morning and return those distributed commuters in the afternoon. The model capable of collecting all commuters who commute to/ from every municipality.

6.2.3 Stage three: Disease model

The disease model is one of the sub models of the disease model. The main task of the model is to calculate disease diffusion based on the metapopulation modeling approach.

The initial task of the disease model is to distinguish the susceptible from the immune population. The basic element of this process is the vaccination level of each municipality and each age group; which is one of the attributes of a group. Vaccination percentage indicates how much individuals within the group are susceptible and immune. The following formulas are used by the disease model in calculating the Susceptible, Infected, and Recovered agent groups.

Time = 0

$S(t+1) = \text{population} * \text{vaccination}$ $I(t+1)=0$ $R(t+1)=0$

Time = 1

$$S(t+1) = St - St * It * \beta * \Delta t$$

$$I(t+1) = It + St * It * \beta * \Delta t - \mu * It * \Delta t$$

$$R(t+1) = Rt + \mu * It * \Delta t$$

After computing the susceptible the model calculates the initial infections. The disease model calculated the process of the disease diffusion in the following steps:

- > Start initial infection, in the model randomly creates a number of infections to a randomly selected municipality and group.
- > Commuters commute to their destination municipalities

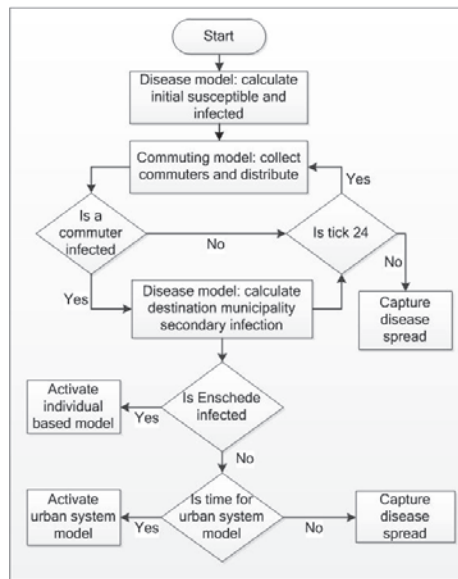


Figure 6.6: Work flow of the disease model

- > For each group disease model calculates change in Infection and recovery
- > Let's, there is infection in a group. That means change in infection is > 0 . Thus, if change in infection of a group is > 0 , change of recovery of the infected group will be calculated (b/c if 1 is infected from a group recovery could also happen and I is calculated in the above step so it is not needed to calculate for the previously infected group). In order to update the change in I and in R the model stores these values in two variables.
- > Then a destination for change in I is selected, that means a random group is picked and then infection (I change) is transferred to the picked group. At this point the infection is transferred to a random group because when one is infected the infection can be transferred to individuals in the group or to individuals found in another group.
- > Finally, the model calculates the total number of infected (I), recovered (R), and susceptible (S). Every 24 ticks the above three processes are repeated commuting model collects the distributed commuters and returns them to their destination municipality. Finally the model calculates the total number of susceptible infected and recovered. Figure 6.6 illustrates the process of municipality model.

When an infection is occurred in Enschede the third or lower level in the hierarchy which is the individual based model is connected to calculate infections of the municipality. In the individual based model all the concepts of the model by Abdulkareem (2010) with some modifications (e.g. recovery of infected agents, includes commuter activity in its daily activity model) are applied as described in 3.1. Furthermore, section 4.3.2 contains a detail description of flow of information between the hierarchical levels.

Chapter 7

Conclusion and Recommendation

The main objective of this research was to develop a method for scaling agent-based simulations that can be implemented on an existing pertussis model and improving its performance. To achieve the predefined objective several questions have been addressed and studied. The specific objectives and their answerd will be discussed below.

To develop the method the first objective was to examine, test and identifies factors influencing performance of the existing pertussis model. The analysis mainly focuses in checking performance and scaling bounds of the model. This is because the model was functioning only for 46 neighborhoods of Enschede even though data was available for the whole 70 neighborhoods of Enschede. In addition, the model doesn't incorporate recovery of infected agents and waning immunity of agents of all age groups which limits scaling the model temporally. Furthermore, the model was built as a closed city which means commuting/ movement of agents was not included which greatly affect the realistic diffusion process of disease.

Findings of the test which has been depicted in chapter three, revealed that the model can be scaled up from 47 to 70 neighborhoods as well as from 67, 000 to around 155, 000 agents. This was mainly achieved by fixing a repast setting to allocate more running memory for the complete Enschede area instead of the first 47 neighborhoods and by fixing the error in the input data. Even though the model is capable to simulate the total population of Enschede, pertussis is an airborne disease which is not restricted in one city. The diffusion could be due to many reasons among which commuting is one the main factor facilitating the diffusion. For this reason, enabling the model to run for larger number of agents and spatial extent requires implementation of commuting of agents. For scaling temporal extent of the model recovery of infected agents should be included. Moreover, agent base modeling techniques belong to the category of individual based approaches; and therefore are very suitable for simulating smaller geographic areas with relatively few agents. Because the approach studies the individual persons, activities and contact processes and interactions between people which may lead to the transmission of the disease. However, when modelling large numbers of agents, for example for the complete country it this is computationally demanding.

Three approaches (distributed, hybrid and hierarchical modeling) for scaling the model have been evaluated. Grid based modeling approach divides complex systems in smaller parts and distributes these parts across several platforms. At a certain point these parts interact with one another and update information. A Grid based modelling offers a robust distributed computing infrastructure needed for simulations of complex systems by offering proper execution environment for multi agent based system (MABS). Yet the approach is computationally expensive and it doesn't provide a tool to break the complexity of a system. Hybrid modeling approach combines the advantages of two modeling paradigms and introduces a hybrid model that starts as one modeling approach and switches to the other approach after a certain condition is reached. The approach is provides opportunities for diversity and flexibility. Furthermore, it could make the model more robust, flexible and scalable. However, hybrid modeling approach could be disadvantageous because implementation of two modelling approaches in to one problem is complex..

Hierarchical modeling approach is a way of discovering or rendering order and breaking down complexity. The approach emphasizes in solving a big system by dividing it into smaller parts and creating an interaction between the parts. The approach provides different methods such as order, inclusion, control and level hierarchies for decomposing complex system in to smaller parts. From the analysis these approaches have been selected to prepare a method for scale agent based systems, since their limitations are compensated by one another. Therefore, in this study a hierarchy based hybrid model approach has been developed. And for the implementation process the hierarchical models are supposed to be created in a distributed manner. The developed model computes the disease diffusion process in the Netherlands. In addition to the selected scaling modeling approaches the method also combines advantages of agent based modeling and metapopulation modeling approaches. To implement the prototype two sub-models for the middle tire have been implemented (commuting model and the disease model) and have been tested individually. This has proven the general concepts of the approach to develop a method for scaling agent-based models toward simulating a large spatial and temporal extent while concurrently improving its performances. Apart from the developed two separate models, commuting and disease models, based on the created and discussed conceptual frameworks in chapter four, attempts has also made to join these two models to examine the pattern and direction of disease spread in region Twente. Basically, it has been demonstrated that both models run separately. The main challenges faced in the execution combining process were the lack of the flow of information between the models which might be arise either due to efficiency of the written implementation Programme or due to incompatibility of Repast step function and model steps function. This indicates that though the model step (tick) is increasing, Repast step function may not detect the process that leads to missing of updated flow of information. Such more integrated modeling process would have brought up new insight that could therefore facilitate targeted discussions around spread of disease and other related factors in the model. These issues together with other key issues might be further research areas which will be the subjects of discussion in the recommendation section.

Recommendations:

The developed conceptual model and employed approaches in this study enabled a thorough investigation of the first three research objectives. The challenge face in joining the two models, however, only allowed limited investigation around the last research question from the third sub objectives aiming to join the two models. Obviously further research implementing the model and further to evaluate the performance and scalability of the extended model needs to investigate and identify other important factors apart from the issues that have been discussed.

The hierarchical approach has proven to be promising for the following reasons:

- > It allows scaling of the Pertussis model without increasing the number of agents significantly
- > Behavior can be given to the municipality and the urban center which could lead to activities and behavioral changes to control and infection. For example, when infection is occurred in an urban system, information could be spread to the other urban systems, to warn them about the infection as a result the urban systems could change their behavior by increasing vaccination or closing schools.
- > Communication between municipalities could be implemented; this could lead to the realistic detection of the spread behavior.
- > Captures the complexity of the system could be captured by studying the interactions between the systems

However, to continue with the current implementation, calibration of the disease model could be made by checking on the parameters such as R (reproductive number), the number of com-

muters could be calibrated, the disease transmission, recovery rate, and other related parameters could be calibrated. The created models could be combined with the lowest level individual based model. Furthermore, to implement the created prototype the implementation could be extended by creating a link between the sub models (disease and commuting models), implementing the top level model and creating a communication or link with the lower level models, multiple lower level models could be connected with the municipalities and the output could be calibrated, and distributed implementation of these sub models could be facilitated.

Generally further research area looking at scaling agent-based simulations could be done by continuing with the current implementation, focusing on the following related issues:

Calibration of the disease model check on the R

-> Calibration of the number of commuters

-> Making the link with the lowest level model

This could be enabled by extending

-> Top level model

-> Check if multiple lowest level models can be attached to the models

-> Strengthening the distributed implementation

Overall, modeling could be a useful tool to identify a reliable set of intervention options and to assess their relative effectiveness under certain conditions.

The results provide insights into the application of the model to calculate ratios of susceptible/infected in specific time frames and urban environments, due to its ability to depict the disease progression based on individuals' interactions. It is demonstrated that the dynamic spatial interactions within the population lead to high numbers of exposed individuals who perform stationary activities in areas after they have finished commuting. As a result, the sick individuals are concentrated in geographical locations like schools and universities.

LIST OF REFERENCES

- [1] S.A. Abdulkareem. Simulating the spread of pertussis in enschede region using agent - based modelling, 2010.
- [2] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J.J. Ramasco, S. Merler, and A. Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1-13):190, 2010.
- [3] D.J. Barnes and D. Chu. *Agent-Based Modeling*, pages 21–77. Springer London, 2010.
- [4] M. Barthelemy, C. Godreche, and J. Luck. Fluctuation effects in metapopulation models: Percolation and pandemic threshold. 2010.
- [5] M. Batty. *Hierarchy in Cities and City Systems: Hierarchy in Natural and Social Sciences*, volume 3, pages 143–168. Springer Netherlands, 2006.
- [6] A. D. Bentsi-Enchill, S. A. Halperin, J. Scott, K. MacIsaac, and P. Duclos. Estimates of the effectiveness of a whole-cell pertussis vaccine from an outbreak in an immunized population. *Vaccine*, 15(3):301–306, 1997.
- [7] L. Bian. A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B: Planning and Design*, 31(3):381–395, 2004.
- [8] G. V. Bobashev, M. D. Goedecke, Yu Feng, and J. M. Epstein. A hybrid epidemic model: Combining the advantages of agent-based and equation-based approaches. In *Simulation Conference, 2007 Winter*, pages 1532–1537, 2007.
- [9] B. Bollker and B. Grenfell. Space, persistence and dynamics of measles epidemics. 1995.
- [10] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. 99, 2002.
- [11] G. Booch. *Object-Oriented Analysis and Design(2nd Ed.)*. 1994.
- [12] M.V. Boven, H.E. Melker, J. F. P. Schellekens, and M. Kretzschmar. Waning immunity and sub-clinical infection in an epidemic model: implications for pertussis in the netherlands. *Mathematical Biosciences*, 164(2):161–182, 2000.
- [13] H. Broutin, E. Elguero, F. Simondon, and J. Guégan. Spatial dynamics of pertussis in a small region of senegal. In *Proceedings of the Royal Society of*, 2004.
- [14] H. Broutin, Francois S. Elguero, E. and, and J.F. Guegan. Spatial dynamics of pertussis in a small region of senegal. 1997.
- [15] C. CARPENTER. *Agent-Based Modeling Of Seasonal Population Movement and the Spread of the 1918-1919 FLU: The Effect on a Small Community*. PhD thesis, 2004.
- [16] C.J.E. Castle, A.T. Crooks, . Centre for Advanced Spatial, and . Analysis (University College London): Working Paper, 110. Principles and concepts of agent-based modelling for developing geospatial simulations. 2006.

- [17] M. Chan. World health organization(who), May 25 2010.
- [18] D. Chen, G.K Theodoropoulou, S.J. Turner, W. Caib, and R.M. Yi Zhanga. Large scale agent-based simulation on the grid. 2008.
- [19] B. Claude, D. Perrin, and H. J. Ruskin. Considerations for a social and geographical framework for agent-based epidemics. In *Computational Aspects of Social Networks, 2009. CASON '09. International Conference on*, pages 149–154, 2009.
- [20] A. Cliff and p. Haggett. Time, travel and infection. british medical bulletin. *Department of Geography, University of Cambridge, Cambridge and Institute for Advanced Studies, University of Bristol, Bristol, UK*, 2004.
- [21] A.D. Cliff. *Spatial Diffusion: An Historical Geography of Epidemics in an Island community*. Press Syndicate of the University of Cambridge, New Yourk, 1981.
- [22] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. 2006.
- [23] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. 2006.
- [24] V. Colizza, M. Barthélemy, A. Barrat, and A. Vespignani. Epidemic modeling in complex realities. *Comptes Rendus Biologies*, 330(4):364–374, 2007.
- [25] V. Colizza, M. Barthélemy, A. Barrat, and A. Vespignani. Epidemic modeling in complex realities. *Comptes Rendus Biologies*, 330(4):364–374, 2007.
- [26] V. Colizza and A. Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. 2008.
- [27] Center For Disease Control and Prevention. Pertussis (whooping cough), Page last updated: August 26, 2010 2010.
- [28] C. Cortie, M. Dijst, and W.I.M. Ostendorf. The randstad a metropolis? *Tijdschrift voor economische en sociale geografie*, 83(4):278–288, 1992.
- [29] A. Crooks, C. Castle, and M. Batty. Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, 32(6):417–430, 2008.
- [30] S.C. de Greeff, J.F.P. Schellekens, F. R. Mooi, and H.E. de Melker. Pertussis in the netherlands, 2001-2002. 2003.
- [31] H.E. de Melker, M.A.E. Conyn-van Spaendonck, H.C. Rijkman, J.K. van Wijngaarden, F.R. Mooi, and J.F.P. Schellekens. Pertussis in the netherlands: an outbreak despite high levels of immunization with whole-cell vaccine. 3, 1997.
- [32] R.E. Dickinson. The geography of commuting: The netherlands and belgium. *Geographical Review*, 47(4):521–538, 1957.
- [33] K.T.D. Eames, Jonathan M. Read, and W. John Edmunds. Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70–76, 2009.
- [34] S. Garlick, P. Benneworth, J. Puukka, and P. Vaessen. Supporting the contribution of higher education institutions to regional development. Technical report, 2006.

- [35] B. Grenfell and J. Harwood. (meta)population dynamics of infectious diseases. *Trends in Ecology and Evolution*, 12(10):395–399, 1997.
- [36] Hagerstrand. Spatial diffusion: an historical geography of epidemics in an island community door andrew david cliff. 1952.
- [37] H.W. Hethcote and A.V. Rie. Adolescent and adult pertussis vaccination: computer simulations of five new strategies. 2004.
- [38] J. L. Hsieh, C. T. Sun, G. Y. M. Kao, and C. Y. Huang. Teaching through simulation: Epidemic dynamics and public health policies. *Simulation*, 82(11):731–759, 2006.
- [39] D. ÅiÅlÅk, P. Volf, and M. Pechoucek. Large-scale agent-based simulation of air-traffic. 2009.
- [40] D Lane. Hierarchy, complexity, society. *University of Modena and Reggio Emilia*, 2008.
- [41] L. Lattila, P. Hilletoft, and B. Lin. Hybrid simulation models Å when, why, how? *Expert Systems with Applications*, 37(12):7969–7975, 2010.
- [42] S.A. Levin. *Fragile Dominions: Complexity and the Commons*. Perseus Publishing, USA, 1999.
- [43] Z. Li and G. Shao. Halting infectious disease spread in social network. pages 305–308, 2009.
- [44] N. Limtanakool, T. Schwanen, and M. Dijst. Development in hierarchy in the dutch urban system on the basis of flows. *Utrecht University*, 2005.
- [45] Schellekens J.F.P. Neppelenbroek S.E. Mooi F.R. RÅijmke H.C. Melker, H.E.D. and M.A.E.C. Spaendonck. Reemergence of pertussis in the highly vaccinated population of the netherlands: Observations on surveillance data. 2000.
- [46] M. Mengistu, L. Lundberg, and P. Davidsson. Performance prediction of multi-agent based simulation applications on the grid. 2007.
- [47] Statistics Netherlands. The dutch virtual census of 2001 analysis and methodology editors, 2004.
- [48] D. M. Nicol. Scalability, locality, partitioning and synchronization in pdes. pages 4–11, 1998.
- [49] American Academy of Pediatrics. 11. pertussis introduced in 1952/3 (dtp) dtap introduced 1996 notifiable. Technical report, 2006.
- [50] R. Okla. *Simulating the spatial hierarchichal spread pattern of pertussis in the twente region using Agent-Based Modelling*. PhD thesis, 2011.
- [51] C.A. O’Neil and L. Sattenspiel. Agent-based modeling of the spread of the 1918Å1919 flu in three canadian fur trading communities. *American Journal of Human Biology*, 22(6):757–767, 2010.
- [52] World Health Organization. World health assembly opens with appeals for more ambitious approach to health problems, 2005.
- [53] World Health Organization. Immunization surveillance, assessment and monitoring, 2011.

- [54] World Health Organization. Manual 11. control of communicable diseases and prevention of epidemics. 2011.
- [55] J. Parker. A flexible, large-scale, distributed agent based epidemic model. 52, 2007.
- [56] D. Pawlaszczyk and S. Strassburger. Scalability in distributed simulations of agent-based models. In *Winter Simulation Conference (WSC), Proceedings of the 2009*, pages 1189–1200, 2009.
- [57] L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, 8, 2009.
- [58] National Center For Chronic Diseases Prevention and Health Promotion. How to investigate an outbreak, 2005.
- [59] RIVM Immunisation Programme. The national immunisation programme in the netherlands, 2010.
- [60] O.F. Rana and K. Stout. What is scalability in multi-agent systems?, 2000.
- [61] ResearchPennState. How disease spreads, 2009.
- [62] T. Schwanen, F.M. Dieleman, and M. Dijst. Travel behaviour in dutch monocentric and policentric urban systems. *Journal of Transport Geography*, 9(3):173–186, 2001.
- [63] H.A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962.
- [64] National Spatial Strategy. Summary national spatial strategy. creating space for development, 2010.
- [65] F. Sugi. Analysis of an existing agent - based simulation model for the spread of pertussis, disease : whooping cough, in the enschede area. Technical report, University of Twente Faculty of Geo-Information and Earth Observation ITC, 2010.
- [66] A. Van Rie and H.W. Hethcote. Adolescent and adult pertussis vaccination: computer simulations of five new strategies. *Vaccine*, 22(23-24):3154–3165, 2004.
- [67] C. Viboud, O. N. Björnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. 2006.
- [68] J. Wang, J. Xiong, K. Yang, S. Peng, and Q Xu. Use of gis and agent-based modeling to simulate the spread of influenza. *Geoinformatics, 2010 18th International Conference on*, pages 1–6, 2010.
- [69] world Health Organization. Pertussis, 2010.
- [70] J. Wu and J. David. A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, 153(1-2):7–26, 2002.
- [71] Sheng Xu and Leon F. McGinnis. Hierarchical distributed simulation for 300mm wafer fab. pages 1774–1780, 2007.
- [72] Y. Yang and P. M. Atkinson. Individual space - time activity-based model: a model for the simulation of airborne infectious-disease transmission by activity-bundle simulation. *Environment and Planning B: Planning and Design*, 35(1):80–99, 2008.

- [73] S. Zhu, Z. Du, and X. Chai. Gdsa: A grid-based distributed simulation architecture. In *Cluster Computing and the Grid Workshops, 2006. Sixth IEEE International Symposium on*, volume 2, pages 66–66, 2006.

Appendix

Note: The following tables are some of the data's that have been used in the study. The first four are results of testing of the existing model. these tests are done by changing different parameters of the model. The first table indicates was conducted when testing the model by disabling the disease model; this is done mainly to get understanding of temporal process of the model. The second Table result that was obtained by enabling the disease model; it was done to see the out put of the disease model. The third table shows the result of the existing model before it was scaled to run for the complete area of Enschede.

The final two tables are the commuting computed commuting data, which was used for the municipality level commuting model. The first table is commuter students table while the second table is Adult commuters table. In both tables Y axis indicates the in coming commuters of a municipality and X axis indicates outgoing commuters of a municipality.

No	Number of Spatial Area	Initialization and Loading of the GIS environment time (in minutes)	Complete run time (in minutes)	Infection	Total Number of Agents
1	1 up to 5	Less than 1	2	0	3304
1	1 up to 5	Less than 1	2	0	3304
1	1 up to 5	Less than 1	2	0	3304
1	1 up to 5	Less than 1	2	0	3304
1	1 up to 5	Less than 1	2	0	3304
2	6 up to 10	1	3	0	4575
2	6 up to 10	1	3	0	4575
2	6 up to 10	1	3	0	4575
2	6 up to 10	1	3	0	4575
2	6 up to 10	1	3	0	4575
3	11 up to 15	1	4	0	3690
3	11 up to 15	1	4	0	3690
3	11 up to 15	1	4	0	3690
3	11 up to 15	1	4	0	3690
3	11 up to 15	1	4	0	3690
4	16 up to 20	1	4	0	5660
4	16 up to 20	1	4	0	5660
4	16 up to 20	1	4	0	5660
4	16 up to 20	1	4	0	5660
4	16 up to 20	1	4	0	5660
5	21 up to 25	1	4	0	5394
5	21 up to 25	1	4	0	5394
5	21 up to 25	1	4	0	5394
5	21 up to 25	1	4	0	5394
5	21 up to 25	1	4	0	5394
6	26 up to 30	around 2 minutes	5	0	7067
6	26 up to 30	around 2 minutes	5	0	7067
6	26 up to 30	around 2 minutes	5	0	7067
6	26 up to 30	around 2 minutes	5	0	7067
6	26 up to 30	around 2 minutes	5	0	7067

Number of Spatial Area	Initialization (time in minutes)	Loading of the GIS environment time (in minutes)	Complete run time (in minutes)	Number of infections	Total Number of Agents
5	Less than 1	Less than 1	2	0	3304
5	Less than 1	Less than 1	2	0	3304
5	Less than 1	Less than 1	2	0	3304
5	Less than 1	Less than 1	2	0	3304
5	Less than 1	Less than 1	2	0	3304
7	Less than 1	Less than 1	2	0	3304
7	Less than 1	Less than 1	2	0	3304
7	Less than 1	Less than 1	2	0	3304
7	Less than 1	Less than 1	2	0	3304
7	Less than 1	Less than 1	2	0	3304
14	1	1	9	21	11564
14	1	1	8	8	11564
14	1	1	24	596	11564
14	1	1	4	9	11564
14	1	1	21	11	11564
17	1	seconds around 2 minutes	15	22	13715
17	1	seconds around 2 minutes	35	604	13715
17	1	seconds around 2 minutes	6	11	13715
17	1	seconds around 2 minutes	13	19	13715
17	1	seconds around 2 minutes	17	15	13715
24	2	1	65	833	21486
24	2	1	35	200	21486
24	2	1	24	11	21486
24	2	1	40	450	21486
24	2	1	11	3	21486
30	around 1 minutes and 30 seconds	around 2 minutes and 30 seconds	15	31	29690
30	around 1 minutes and 30 seconds	around 2 minutes and 30 seconds	70	510	29690
30	around 1 minutes and 30 seconds	around 2 minutes and 30 seconds	28	127	29690
30	around 1 minutes and 30 seconds	around 2 minutes and 30 seconds	35	139	29690
30	around 1 minutes and 30 seconds	around 2 minutes and 30 seconds	21	69	29690

Number of Spatial Area	Initialization time (in minutes)	Loading of the GIS environment time (in minutes)	Complete run time (in minutes)	Infection	Total Number of Agents
37	3	2	55	50	36610
37	3	2	15	47	36610
37	3	2	90	437	36610
37	3	2	110	760	36610
37	3	2	70	83	36610
46	7	5	123	50	55345
46	7	5	90	20	55345
46	7	5	Error	Error	55345
46	7	5	141	87	55345
46	7	5	Error	Error	55345
57	9	6	Error	Error	61742
57	9	6	Error	Error	61742
57	9	6	Error	Error	61742
57	9	6	Error	Error	61742
57	9	6	Error	Error	61742
67	11	9	Error	Error	68736
67	11	9	Error	Error	68736
67	11	9	Error	Error	68736
67	11	9	Error	Error	68736
67	11	9	Error	Error	68736
70	13	10	Error	Error	69348
70	13	10	Error	Error	69348
70	13	10	Error	Error	69348
70	13	10	Error	Error	69348
70	13	10	Error	Error	69348

Number of Spatial Area	Initialization time (in minutes)	Loading of the GIS environment time (in minutes)	Complete run time (in minutes)	Total Number of Agents
14	1	1	8	11564
17	around 1 minute	around 2 minutes and 30 seconds	9	13715
24	2	1	11	21486
30	around 1 minute	around 2 minutes and 30 seconds	30	29690
37	2	2	40	36610
46	7	5	57	55345
57	9	6	THE MODEL STOPPED RUNNING (Error)	61742
67	11	9	THE MODEL STOPPED RUNNING (Error)	68736
70	13	10	THE MODEL STOPPED RUNNING (Error)	69348

GM-CODE	141	147	153	158	163	164	168	173	183	189	1700	1735	1742	1774
141		11	10	15	11	13	15	13	12	13	12	11	12	13
147	21		24	20	23	24	23	25	17	21	23	17	22	18
153	10	11		9	11	11	10	10	10	11	11	9	10	8
158	18	17	16		17	16	16	16	18	18	19	18	19	18
163	16	14	13	16		16	17	18	17	15	16	18	18	16
164	14	14	15	12	14		14	14	12	15	14	14	15	14
168	19	20	19	21	14	16		19	20	19	19	16	19	20
173	14	15	15	16	14	15	17		14	19	18	18	19	16
183	19	16	14	19	20	16	21	22		18	17	18	15	16
189	24	20	19	23	22	24	26	25	18		19	19	15	20
1700	19	17	17	15	19	21	22	15	14	12		14	19	15
1735	17	17	16	18	18	18	18	16	16	12	13		15	15
1742	14	15	13	15	13	14	15	14	14	8	12	15		13
1774	18	18	16	19	18	18	19	18	21	13	15	12	16	

GM-CODE	141	147	153	158	163	164	168	173	183	189	1700	1735	1742	1774
141		23	19	31	21	25	29	27	24	27	23	23	25	26
147	44		47	39	46	49	47	50	34	42	47	34	43	37
158	36	34	32		33	33	33	32	36	36	38	37	37	35
163	31	29	26	32		33	34	35	35	29	32	37	36	32
164	28	29	30	25	28		29	28	24	29	29	29	31	28
168	39	39	38	41	27	32		37	40	37	39	33	39	41
173	29	31	30	33	28	31	35		28	37	36	37	39	31
183	39	31	29	38	41	33	42	45		37	33	36	30	32
189	47	39	37	47	45	49	51	50	37		39	38	30	40
1700	37	33	35	29	38	41	43	31	29	24		29	39	31
1735	33	34	31	35	36	35	37	31	32	23	25		29	31
1742	28	31	26	30	27	28	30	27	28	16	25	29		25
1774	36	37	32	38	35	37	39	36	43	27	30	25	31	