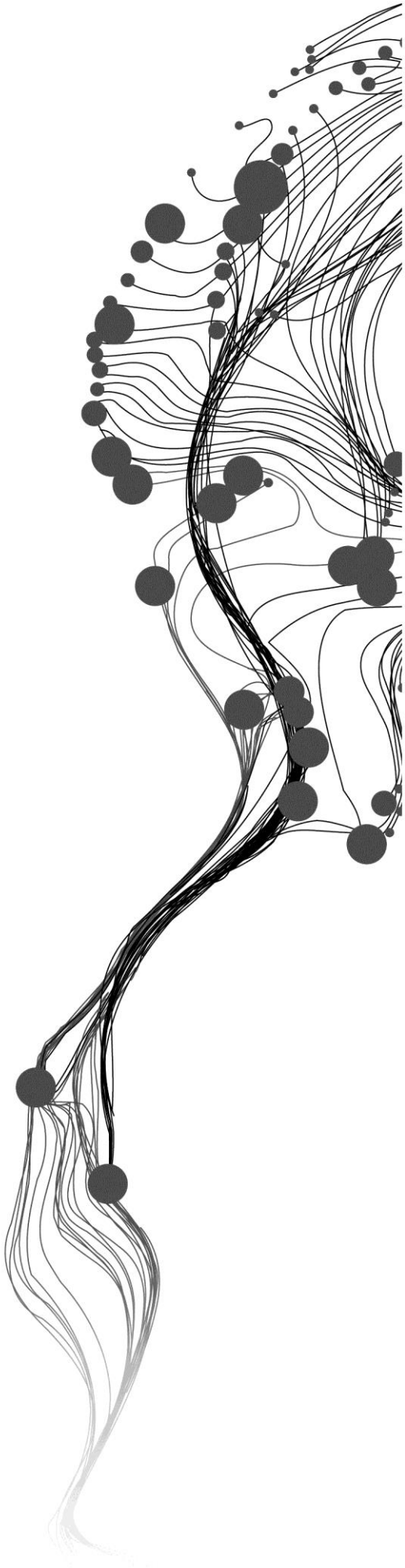
Integrating a hierarchical SOM-based intervention method in an agent-based pertussis model

QUAN ZHOU
February 2018

SUPERVISORS:
Dr. S. Amer
Ir. P.W.M. Augustijn



Integrating a hierarchical SOM-based intervention method in an agent-based pertussis model

QUAN ZHOU
Enschede, The Netherlands, February 2018

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: Urban and Regional Planning and Geo-information Management (PGM)

SUPERVISORS:
Dr. S. Amer
Ir. P.W.M. Augustijn

THESIS ASSESSMENT BOARD:
Dr. D. Reckien
Dr. Gertjan Geerling (External Examiner, Radboud University)]

ABSTRACT

A resurgence of pertussis has been detected in the Netherlands since 1996. Several intervention methods have been implemented but a severe outbreak was detected in 2012. Some authors declared that more efficient alternative interventions should be studied. The main problems of planning a disease intervention are to determine the correct moment in time to start the intervention, the spatial locations in which to conduct the intervention and the target population that should get the intervention. However, previous research always focused either on disease aberration of different age groups for the whole country or the disease aberration in specific geographic regions for all age groups.

The main goal of this research is to construct an intervention model which can draw up intervention planning automatically for different spatial regions based on local disease occurrence to prevent pertussis. Because of time limitation, only vaccination strategies are taken into account as the intervention method in this research. The generated vaccination methods are referred as condition-based vaccination method. The intervention model is agent-based with two entities: GGDs and RIVM. GGDs are local health units which supervise the ongoing disease pattern in their service region and implement interventions. RIVM is the Netherlands national public health and environment institution which collects the disease information from all GGDs and detects disease aberrant change at the national level and orders GGDs to implement intervention in their service area. The behaviors of these two agents were simulated using two sub models: early warning model and intervention selection model. Early-warning model is a hierarchical model which detects disease aberration at both national and regional level. The national aberration determines the time when the interventions should be implemented while the regional detection explores the aberration age groups. A hierarchical self-organizing map (SOM) is used as the aberration detection algorithm which has not been proposed before. The approach is tested on a simulated dataset for Pertussis in the Netherlands and the result can be seen as effectiveness. The aberration age groups will be analyzed in the intervention selection model to determine the target population who should get vaccination priority. Meanwhile, the intervention selection model stipulates the vaccination ratio for each target population groups in different region.

The environment of the intervention model is a disease transmission model which simulates pertussis spread in the Netherlands established by previous research. Several validation experiments were performed at first and the result presents that the model run as expected. The impact of the condition-based vaccination method is assessed regarding the number of infections and the disease diffusion patterns and compared with the current vaccination strategy and maternal vaccination approach. The total number of infections of the condition-based vaccination scenario is obviously less than the number of other two vaccination scenarios and the disease outbreak is restricted in a small area with the condition-based vaccinations.

Keywords:
Intervention model, Agent-based model, SOM, Pertussis, the Farrington method, the Netherlands

ACKNOWLEDGEMENTS

This research would not be possible without the contribution from several people. Therefore, I want to take this chance to extent my sincere gratitude here.

First of all, I want to express my special thanks to my supervisors Sherif Amer and Ellen-Wien Augustijn for their intellectual guidance, support and motivation provided during the whole research process. I also grateful for all UPM stuffs. Thank you for all of your scientific and interesting lectures.

Besides, I incredibly grateful to my girl-friend Christina and all of my friends. Thank you for all of your love and encourage.

Finally, I want to express gratitude to my family for all love and support.

Quan Zhou

Enschede, February 19, 2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1. Background and justification

Pertussis is an infection which is easily transmitted from person to person and especially threatens infants' health. As Figure 1-1 indicates, there has been an increase in the number of pertussis notifications in the Netherlands since 1996. Several interventions have been implemented, like the National Immunization Programme (NIP) was implemented in the Netherlands in 1996 which aims to increase the vaccination coverage of infants and acellular pertussis vaccines (ACV) was administered to four-year-old since 2001 (Van der Maas et al., 2013). However, these interventions have not proven to be very effective as disease cases have not decreased and come with a relatively heavy financial burden (De Melker et al., 2000; De Greeff et al., 2009). Therefore, some authors stated that alternative vaccination strategies should be studied to come to a more efficient approach (Niessen et al., 2008).

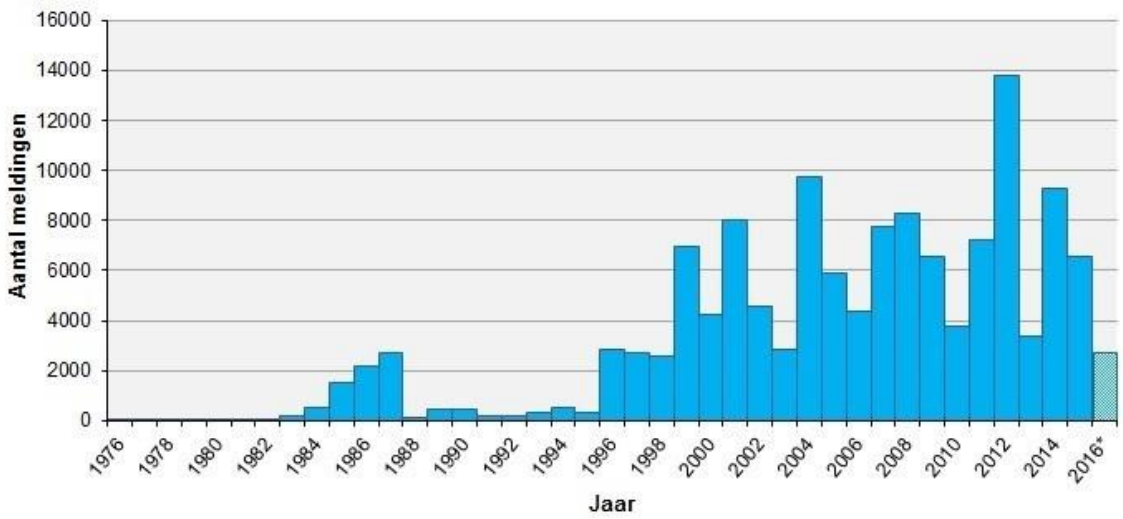


Figure 1-1: Number of pertussis in the Netherlands per year (Source: Kinkhoest - RIVM)

The general problem of planning a new intervention is how to choose target population that should get the intervention priority (Bijkerk.P et al., 2015). Several alternative pertussis intervention strategies have been identified by researchers. Because pertussis is a vaccine-preventable disease (James D. Cherry, 2012), most of the studies focused on the optimal vaccination method. These strategies identified the target population either for particular population groups in the whole Netherland (Rozenbaum et al., 2012; de Vries et al., 2010) or in specific geographic regions for all population groups (Tjalma, 2016). For example, vaccination of pregnant women has already been introduced by a few countries and the Netherlands is also considering this option (Dabrera et al., 2015); Periodic adult vaccination has been proposed for people that have a lot of contact with young children like medical staff, teachers and all family members of new born infant(Lehmann, 2016); Network vaccination has been explored by Tjalma to simulate the efficacy of vaccination to all population in several specific municipalities with the highest daily commuting using certain vaccination coverage (Tjalma, 2016). Although some of these interventions have been verified as effectiveness, they still have their drawbacks. As for the interventions focused on specific population groups, though the measures reduce the number of pertussis notifications of the target groups, they had no impact on the increasing number of infected cases in other population (Rozenbaum et al.,

2012, Van der Maas et al., 2013). Meanwhile, the network vaccination was validated with few effect on the disease diffusion (Tjalma, 2016).

To attempt to overcome these drawbacks and investigate a better intervention strategy, condition-based intervention method is proposed in this research. The assumption is that the ongoing patterns (conditions) of disease occurrence are various for different population groups in different area. The target population groups should be adjusted to fit the aberration population which is identified by regional health units. The expectation is that all population groups can benefit from the condition-based intervention and the disease diffusion will be prevented. Because of the limited time, only vaccination strategies are concerned as the intervention methods in the model. The effectiveness of other kinds of interventions are also worthy of attention but haven't been integrated into the model yet (Keller et al., 1998).

To simulate the behaviors of this condition-based intervention method, one of the most important things is to identify the way in which target population groups are selected. An intervention model is constructed which can supervise the disease occurrence, detects the aberration and identifies the target population groups using Agent-based modeling approach.

Agent-based models (ABMs) are especially suitable for this type of study as they simulate the heterogeneous behaviors of self-organized entities (agents) and the interactions between each other and the environment. The intervention model includes two agents: GGDs and RIVM. GGDs are the local health units in the Netherlands. There are currently 25 GGDs and each has its own service area (Fig. 1-2). The tasks of the GGDs include childhood vaccination, and the implementation of infectious disease surveillance systems for some diseases including pertussis. The primary aim of such surveillance systems is to collect and analyze disease data that can be used to detect outbreaks and to evaluate the impact of interventions (Damm et al., 2016). The regional surveillance data are reported to the national institution for public health and the environment (RIVM) which detects the disease outbreak from the national level and advocates the most appropriate interventions to the GGDs (Hawker et al., 2005).The environment in the model is the disease occurrence in each GGD service region.

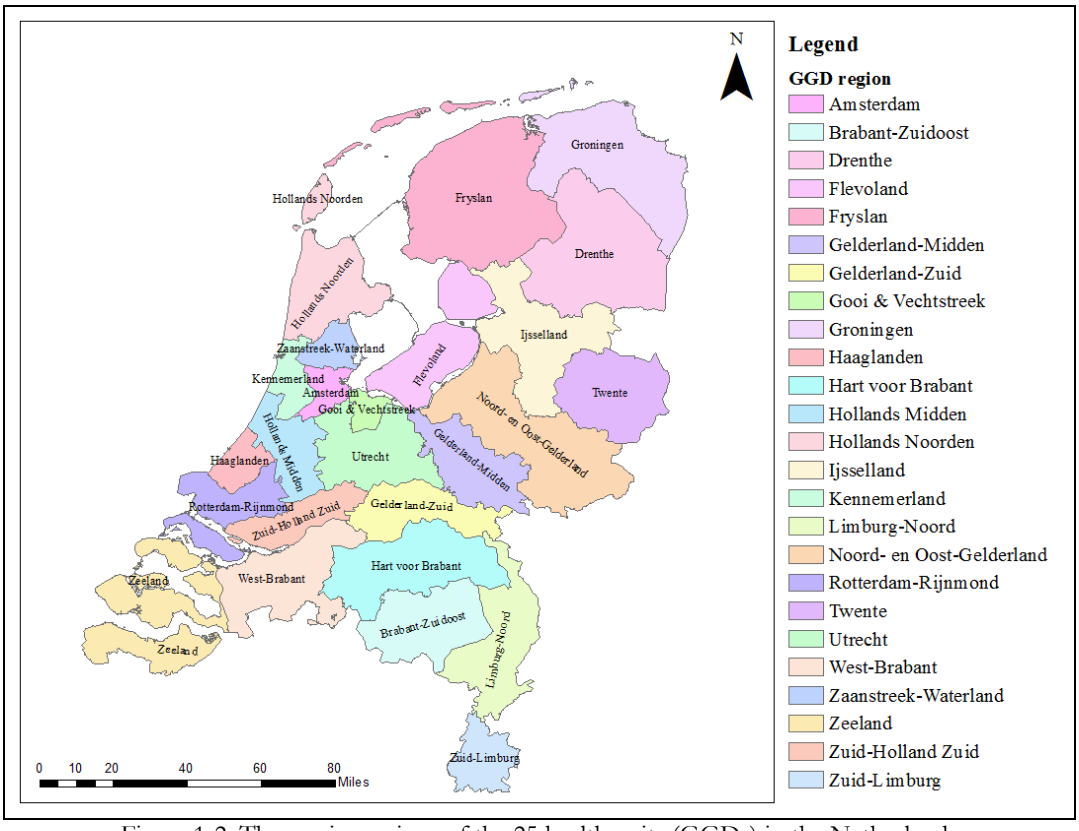


Figure 1-2: The service regions of the 25 health units (GGDs) in the Netherlands

The concept of this model has been proposed by Tjalma but was not yet accomplished (Tjalma, 2016). In the research of Tjalma (2016), the intervention model was used to simulate how the GGDs implemented an intervention strategy (either network vaccination or maternal vaccination) in certain areas when the number of infected people in a given GGD region was higher than a certain percentage of the population. In this research, the intervention (vaccination) strategies that should be implemented in each GGD region are determined by an intervention selection sub model with a rule-based method, and the time that interventions should be executed in each GGD region depends on an early-warning sub model using aberration detection algorithm. There is an attempt that using a hierarchical Self-organizing map (SOM), a classified unsupervised machine learning method, to detect the disease aberration which hasn't been explored before. The performance of using the SOM in disease early-warning detection will be discussed in Section 4.2. In the research of Tjalma (2016), the effectiveness of intervention strategy was tested by integrating with a disease diffusion model. This transmission model is used as the environment of the intervention model in this thesis to simulate the effectiveness of the condition-based intervention strategy.

1.2. Research objective

1.2.1. General objective

The main goal of this research is to simulate the effectiveness of the condition-based intervention method via agent-based modelling and machine learning in order to decrease the spread of pertussis in the Netherlands.

1.2.2. Specific objectives

To be able to achieve this purpose, the following specific objectives are identified:

1. Identify existing alternative pertussis intervention strategies.
2. Construct an early-warning sub model to detect the national disease outbreak and the aberration population groups in each GGD region.
3. Construct an intervention selection sub model using rule-based algorithm.
4. Complete the intervention model with early-warning sub model and intervention selection sub model.
5. Integrate the intervention model and disease transmission model to evaluate the performance of the condition-based intervention strategy.

1.2.3. Research questions

Several questions are identified to achieve specific objectives

Identify existing alternative pertussis intervention strategies:

- Which alternative intervention strategies are proposed by previous researches?
- Which interventions will be investigated in this research?

Construct an early-warning model to determine the intervention strategy should be implemented in which conditions:

- Which early-warning algorithms are proposed by previous researches?
- Which methods are suitable for the pertussis in the Netherlands?
- Whether the SOM can predict the disease outbreak?
- Whether the SOM can detect the aberration population groups in each GGD region?
- Whether the prediction ability of SOM is better than traditional aberration detection method?

Construct an intervention selection sub model using rule-based algorithm:

- Which part of intervention selection in the work of Tjalma (2016) can be used in this research?
- How to design the rule-based graphic to simulate the strategies selection?

Complete the intervention model with early-warning sub model and intervention selection sub model:

- How to integrate the machine learning method into the agent-based model?

Integrate the intervention model and disease transmission model to evaluate the performance of the condition-based intervention strategy:

- How the disease transmission model categorizes the population?
- How to evaluate the performance the condition-based intervention strategy?

1.3. Innovation aim at

Lots of alternative interventions have been proposed in previous research to prevent pertussis in the Netherlands (reference). The target population in these interventions are either a specific age group for the whole country or all population groups in specific geographic regions. In this research, a condition-based intervention strategy is proposed which identifies the target population considering the various infection features of each age group in different regions. An intervention model is constructed to simulate the behaviors of the condition-based intervention method. The intervention model consists of two sub-models: early-warning model and intervention selection model. The algorithm of aberration detection in the early-warning model is the self-organizing map (SOMs) which is first used to detect the disease outbreak in this research.

2. Literature Review

2.1. Introduction to literature review

In this chapter, the outcomes of several research objectives which should be achieved based on the literature review will be fully represented. The first objective is to explore the previous early-warning detection methods. The second one is to identify the current pertussis intervention strategies which have been proposed or implemented in the Netherlands or other countries. The third objective is to discuss the way in which machine learning method can be integrated into the agent-based model.

2.2. The model of Tjalma

The way of how the selected intervention approaches prevention the disease spread will be simulated using pertussis transmission model developed by Tjalma (2016). As shown in Figure 2-1, the Tjalma model consists of four sub models: population model, SEIR model, network model, and intervention model.

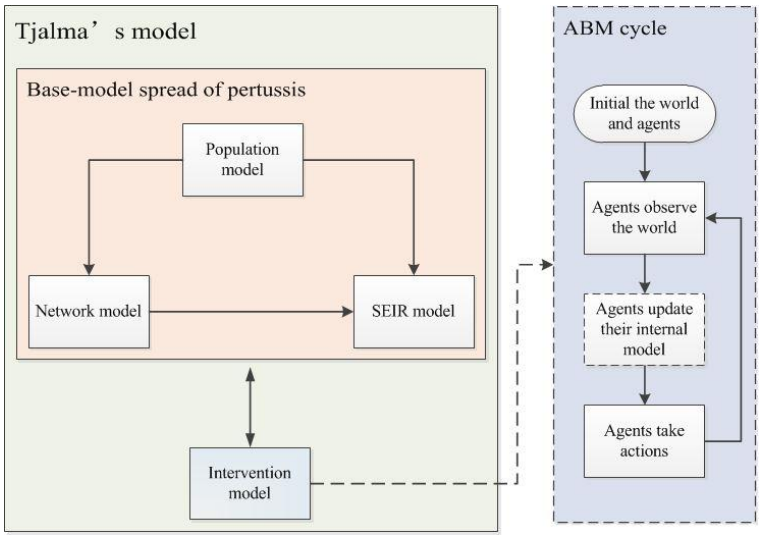


Figure 2-1: The concept framework of Tjalma model (Tjalma, 2016)

The population model creates the initial population. Population is created for all municipalities of the Netherlands and distributes this population in space (Fig. 2-2). The population model controls the population dynamics and provides several categorizes population groups. In the model, population is divided by age, gender (only for 25-35 years' group), commuting state and whether they are part of a households. There are 9 age groups: 0-5 months, 5 months to 5 years, 5 to 12 years, 12 to 17 years, 17 years to 25 years, 25 to 35 years, 35 to 50 years, 50 to 65 years and older than 65 years.

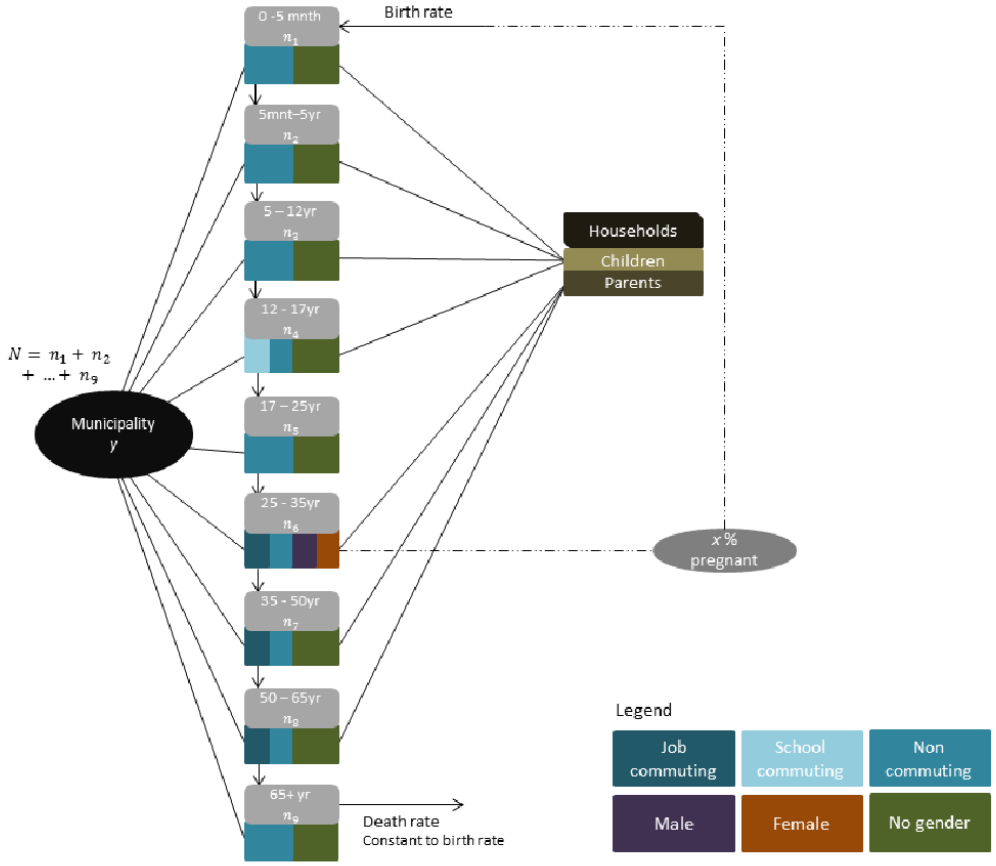


Figure 2-2: The framework of population sub model (Tjalma, 2016)

The SEIR model describes the current pertussis diffusion for each municipality. The model splits the population in four disease groups: susceptible (S), exposed (E), infected (I) and recovered (R). The transition from one group to another model the disease diffusion process in a mathematical form (van den Driessche, 2008):

$$dS/dt = b - \beta SI - pbE - qbI - dS \quad (2.2\text{-}1)$$
$$dE/dt = \beta SI + pbE + qbI - (\varepsilon + d)E$$
$$dI/dt = \varepsilon E - (r + d)I$$
$$dR/dt = \gamma I - dR$$

Where b presents the birth rate, d indicates death rate. The contact rate between susceptible population and infected people is denoted by the bilinear mass action form βSI. The new-borns offspring from the exposed individuals are represented using pbE while the new-borns offspring from the infected individuals are expressed by qbI. The transformation rate from exposed to infected is presented by a constant value ε. The recovered rate of infected individuals is γ (van den Driessche, 2008).

The network model describes the work related adult commuting and the adolescent school commuting through traffic network and contact matrix. The traffic network consists of the municipalities of the Netherlands and their interactions, while the contact matrix which determine whether people get enough time to transmit the disease (Tjalma, 2016). The network model and SEIR model make up the disease model which simulates the spread of pertussis for the whole Netherlands.

The Tjalma model (Tjalma, 2016) was proposed to investigate two intervention strategies, vaccination in certain municipalities and vaccination of pregnant women to identify which one is the better for pertussis

prevention. The intervention model of Tjalma simulates the way of how health units implement these two vaccination methods to the population. The model was designed to be an agent-based model yet, due to time constraints was not implemented in such a way. In the intervention model, GGDs are programmed as the agents. These health units sense the number of infected population in their service region. The intervention strategy will be executed if the infected population is higher than a basic level of disease cases. The intervention method is simulated by increasing the percentage of recovered population and decreasing the ratio of susceptible. How big of the increase percentage is depends on the known children vaccination ratio for each municipality. Two vaccination scenarios are tested in the model, either maternal vaccination or network vaccination. The maternal vaccination increases the recovered population in pregnant women and their infants. The network vaccination enhances the total recovered population for certain municipalities.

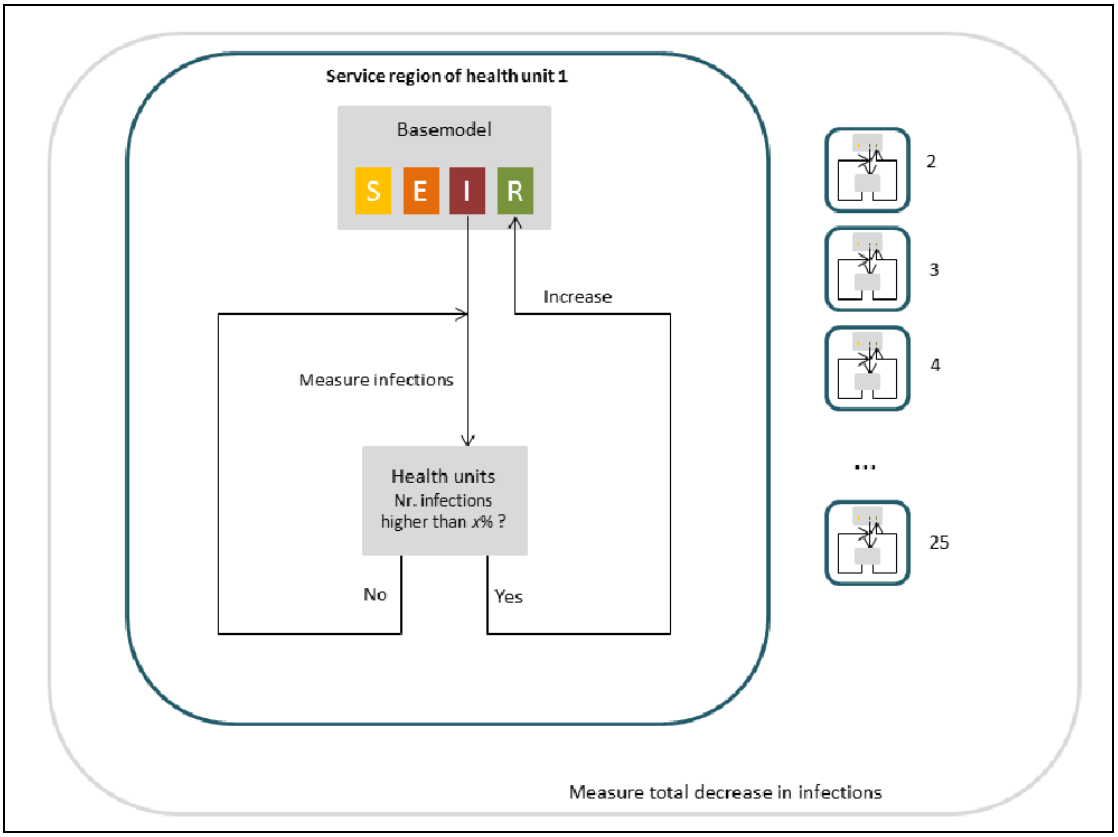


Figure 2-3: Concept framework of interaction between intervention model and disease transmission model (Tjalma, 2016)

2.3. Early warning system

2.3.1. Definition types

The aberration detection for health-related issues was represented as the difference in the distribution or frequency of important health-related events when compared with historical (more than 3 years before) or recent (less than 9 days in the past) data (L. Hutwagner et al., 2003) . The main aberration detection methods can be divided into two categories: case definition methods and pattern recognition methods (L. Hutwagner et al., 2003). Therefore, the aberration detection will be explored from four directions (Tab.2-1):

Table 2-1: Aberration detection methods

Detection direction / Detection methods	Frequency	Spatiotemporal Distribution
Case definition methods	● Historical limits method ● Log-linear regression model (Farrington method) ● Quality control model compound smoothing technique ● Cyclical regression model. ● Time series moving percentile method (MPM) ● 'hhh4'function	● Local indicators of spatial association (LISA) ● Space-time K-function ● 'hhh4'function
Pattern recognition methods	● Self-organizing map(SOM)	● Self-organizing map(SOM) ● Tracking analysis

2.3.2. Frequency based case definition methods

The infectious disease aberration detection methods can be classified based on the length of time that the early warning system will be collecting data. Normally, long-term detection methods expect syndromic surveillance to last more than 30 days. Short-term detection methods which also referenced as drop-in surveillance systems, are used to conduct the supervision for less than 30 days (L. Hutwagner et al., 2003). In this section, three long-term methods will be illustrated: Historical limits method, time series moving percentile method (MPM) and log-linear regression model (Farrington method). Early Aberration Reporting System (EARS) short-term drop-in surveillance method will be introduced as an example for short-term detection methods.

The historical limits method has been applied in the United States to incidence data for nine diseases including pertussis since 1989. The number of current 4-week period reported cases is compared to those of the corresponding historical period in preceding 5 years. The aberration is identified according to the comparison of the ratio of current reports with the historical mean (L. Hutwagner et al., 2003):

$$X_0/\mu \ > 1 + (2 * \sigma_x/\mu) \qquad (2.3\text{-}1)$$

where X_0 is the current total reported cases for a 4-week interval, μ is the mean of 15 totals of 4-week intervals (including the same 4-week period, the preceding 4-week period, and the subsequent 4-week period over the preceding 5 years of historical data), and σ_x is the standard deviation of these 15 historical incidence data values.

China Infectious Disease Automated-alert and Response System (CIDARS) has applied a similar historical limit method, but becomes more simply. The method was modified and referenced as time series moving percentile method based on historical data (Zhang et al., 2014). According to Zhang (2014), the current observation period (7 days) is compared with corresponding historical period in previous 3 years. The corresponding historical period includes the same seven-day period, accompanying with the two preceding seven-day periods and the two following seven-day periods for each of the previous three years, resulting in 15 historical seven-day data blocks. The aberration signal will be determined based on the comparison between the current reported cases and historical median:

$$X_0 > \mu \qquad (2.3\text{-}2)$$

where X_0 is the current total reported cases for a 7 days block, μ is the median of 15 historical seven-day period. If the X_0 is higher than μ, an alert signal will be generated in CIDARS. The current observation period and historical data blocks were dynamically moved forward day by day (W. Yang et al., 2011). Zhang and his team (2014) evaluated the performance of Dengue outbreak detection using the time series

moving percentile method. By using the reported cases and outbreaks during 2009– 2012, CIDARS could detect all the reported outbreaks during the study period, which were currently regarded as almost 100% sensitivity of outbreak detection for CIDARS (Fig. 2-4).

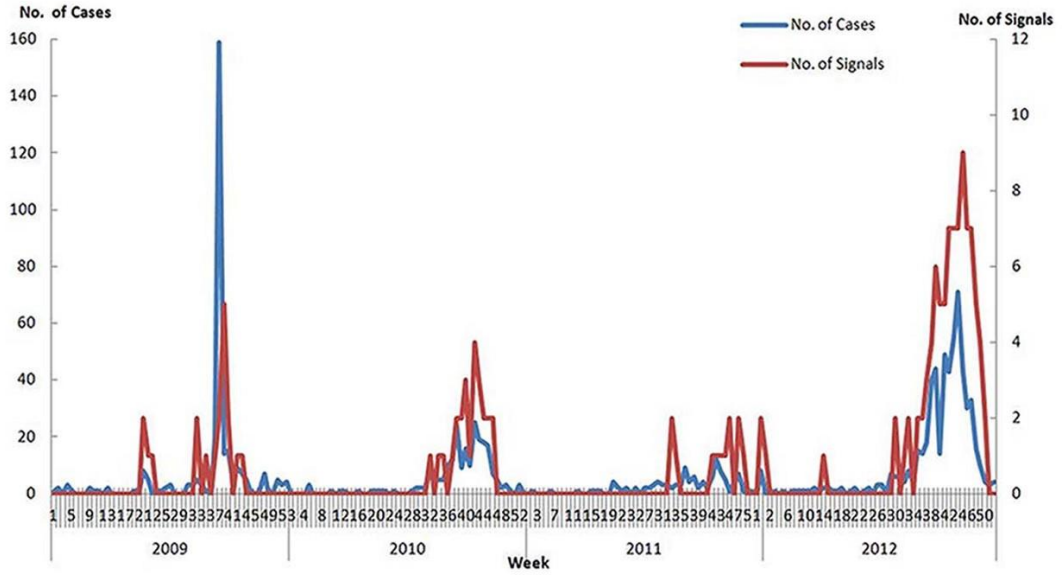


Figure 2-4: The time distribution of dengue fever cases and signals between 2009 and 2012 in China.

A log-linear regression model which was developed by Farrington is used to detect communicable disease outbreak in the United Kingdom (L. Hutwagner et al., 2003), and the method is also implemented at European public health institutes (Maëlle, Dirk, & Michael, 2014). The detection of the current observation week is done by formulating a statistical algorithm for predicting the observed number of counts y_{t0} . The reference data includes the corresponding week in the previous b years together with w weeks before and w weeks later. Therefore, a total of $b * (2w + 1)$ data is extracted as referenced values. An over-dispersed Poisson generalized linear model (GLM) with log-link is fitted to the reference data (Höhle & Mazick, 2010):

$$E(y_t) = \mu_t \ where \ log(\mu_t) = \alpha + \beta t \tag{2.3-3}$$

$\mathrm{Var}(y_t) = \emptyset\mu_t$ while α, β and $\emptyset$ are coefficients to estimated. The expected number of counts μ_0 is anticipated for the current time using GLM. The upper bound U_{to} is determined according to the predicted value and its variance. If the observed number of reported cases is higher than the upper bound U_{to}, the alert will be sounded (Maëlle et al., 2014).

EARS method is one of the short-term surveillance methods which will be recommended while little history data is available. There are three variants C1, C2 and C3. C1 reference to the baseline including seven timepoints before the current event, C2 and C3 are the baseline which covers previous three days to nine days. The aberration is identified while the current reported cases y_{t0} is higher than the threshold U_{to}. The U_{to} is calculated through the formula (for C1):

$$U_{to} = \overline{y_{t0}} + z_{1-a}S_{t_0} \tag{2.3-4}$$

While $\overline{y_{t0}} = \frac{1}{7}\sum_{i=t_0-7}^{t_0-1} y_i$, $S_{t_0}^2 = \frac{1}{7-1} * \sum_{i=t_0-7}^{t_0-1}(y_i - \overline{y_{t0}})^2$ and z_{1-a} is the $(1-a)'th$ quantile of the standard normal distribution (Maëlle et al., 2014).

2.3.3. Spatial distribution-based case definition methods

In this section, local indicators of spatial association (LISA) and Space-time K-function will be introduced. The LISA method was used to detect the hotspot pattern of dengue in Chachoengsao Province (Jeefoo et al., 2011), Thailand. The method detects the location and the significant of the clusters (hotspots). There are two requirements of the method: the LISA for each observation area gives an indication of the extent of significant spatial clustering of similar values around that area; the sum of LISAs for all observations is proportional to a global indicator of spatial association(Anselin, 1995). Local Moran' I and Geary's are two indices of the LISA. Jeefoo et al. (2011) used the local Moran' I to examine the local level of spatial autocorrelation where the value of dengue fever morbidity rate (DFMR) were both extreme and geographically homogeneous. The local Moran' I is defined as:

$$I_i = Z_i \sum_j w_{ij} Z_j \tag{2.3-5}$$

the observations Z_i, Z_j are in deviations from the mean, and the summation over j is such that only neighboring values $j \in J_i$ are included. w_{ij} are the weights and by convention, $w_{ij} = 0$. The Moran scatter plot was constructed with spatial lag of DFMR to represent the result.

Space-time K-function which was developed by Diggle et al. was used to determine the spatiotemporal pattern of H1N1 diffusion in Hong Kong (Wong & Lee, 2011). The method was used to explore the relationship among cases in spatial and temporal scale. The K function is represented as:

$$K(s,t) = K_1(s) * K_2(t) \tag{2.3-6}$$

The right part of the formula describes the K-functions of the spatial and temporal component process respectively. $K_1(s) = \lambda_1^{-1} E[N_d]$ where N_d indicates the number of further events occurring within distance s of an arbitrary event and $K_2(t) = \lambda_2^{-1} E[N_t]$ number of further events occurring within time t of an arbitrary event.

2.3.4. Self-organizing map (SOM) – pattern recognition method

Self-organizing map (SOM) is one of the best known unsupervised neural learning algorithms for large data mining and pattern identifying (Törönen et al., 1999 , Richardson et al., 2003). In the research of Lobo (2009), the author listed 8 application of SOM: clustering, exploring data analysis and visualization, ordering of multidimensional data, supervised data classification, sampling, feature extraction, control or data sensitive processing and data interpolation.

As for the epidemic investigation, the most popular application of SOM is to study multivariate patterns (Basara & Yuan, 2008). In the research of Augustijn (2013), SOM has been used to identify the synchrony between spatial locations, group epidemic waves of Measles in Iceland based on the similarity of spatiotemporal diffusion trajectories. Meanwhile, different severity level of disease occurrence can be detected and mapped in different neurons.

There are three main steps for SOMs algorithm: train SOMs using training data, secondary clustering and data mapping. SOMs are trained follow three steps: determine the size and type of SOM lattice, assign the weights which includes the same dimension with the input data to each neuron and identify the best matching unit (BMU) through the iterative process. The input data vectors which map to the same neuron are synchronized. However, the vectors that map to neighboring neurons might be synchronized as well. The second clustering can group the neurons with similar values. Methods like U-matrix, K-means and hierarchical clustering are recommended to implement the second clustering(Juha Vesanto, 2000). After that, mapping data is used onto the trained SOM to identify the patterns (Augustijn et al., 2013).

2.4. Pertussis Intervention strategies

This section can be divided in two parts: the intervention strategies have been practiced by GGDs in the Netherlands and the intervention strategies have been implemented in other countries or proposed in academic stage.

An accelerated schedule of the National Immunisation Programme (NIP) was implemented in the Netherlands in 1996. The schedule stipulated the first vaccination was given at two months of age. From November 2001 onwards, acellular pertussis vaccines (ACV) was administered to four-year-old, simultaneously with the booster Diphtheria- Tetanus-Inactivated Poliovirus (DT-IPV) vaccine. Finally, in 2005 the DTP-IPV-Haemophilus influenza type b (Hib) combination vaccine including whole cell pertussis vaccines (WCV) was replaced by a combination with ACV at 2, 3, 4 and 11 months of age (Van der Maas et al., 2013).

Except these intervention strategies, there are several intervention measures have been implemented in other countries. The vaccination of pregnant women has been introduced in the United states since 2012 in order to limit the risk of infants (Dabrera et al., 2015). Meanwhile, Tjalma (2016) created a simulation model to evaluate the effectiveness of using the maternal vaccination in the Netherlands. Rozenbaum (2012) proposed implementing the extended vaccination strategies – adolescents vaccination in the Netherlands while the adolescents booster dose has been integrated in the national immunization programmes of USA, Australia and France (Halperin, 2005; Tan et al., 2005)

2.5. Agent based modelling (ABM) meets machine learning (ML)

ABMs are best described as the simulation models in which the unique and independent entities (agents) 'interact with each other and their environment locally (Railsback & Grimm, 2012, p.18). According to the results of the agents' decision and actions, patterns are generated in space and time (Johnston, 2013, p.12). Comparing with other simulation techniques, ABMs emphasize modelling the heterogeneity of agents and concern the emergence of self-organization (Macal & North, 2010). The heterogeneity refers to the different characteristics of agents in size, location, resource reserves and history. The emergence of self-organization can be explained as the agents adjust their behaviors to adapt the current states of themselves to the other agents, and to the environment, in other words, individual agents interact with each other and the environment (Railsback & Grimm, 2012, p.19). The connection between agents are named as ABM topology which depicts 'who' transmits information to 'who' (Macal & North, 2010). Normally, agents are restricted to interact with their neighbours-in geographic space or via network topology (Railsback & Grimm, 2012, p.19).

ABMs are used in many different dimensions, especially in the infectious disease simulation. The most important advantages of ABMs are the 'heterogeneity of both individuals and environment, and also the stochastic essence of infectious disease transmission (Yang et al., 2011). However, one limitation of standard ABMs is the agents can't come up a new behavior of how to take action based on the history (the results of previous actions) (Rand, 2006). Integrating the machine learning with the ABMs can greatly solve this problem.

Machine learning is the field concerning of how to construct computer programs that automatically improve with experience (Mitchell, 1997). Lots of machine learning algorithms have been investigated such as decision tree, neural network, support vector machine (SVM), Bayesian network and so on, however the best method that will be used in this research will be identified during the thesis period.

In the paper of Rand (2006), he briefly introduced the way of how to integrate the machine learning algorithms into the agent-based models. Rand (2006) stated that both ABMs and ML used straightforward and similar structure to control their flow of operation (Fig.2-5). In order to combine these two models, ML cycle could be seen as a model 'refinement engine' for the ABM. Therefore, the integrated cycle

focuses on the ABM and interrupts the third step in its standard progress, 'by sending data to the ML cycle to handle the model refinement' (Rand, 2006) (Fig.2-6).

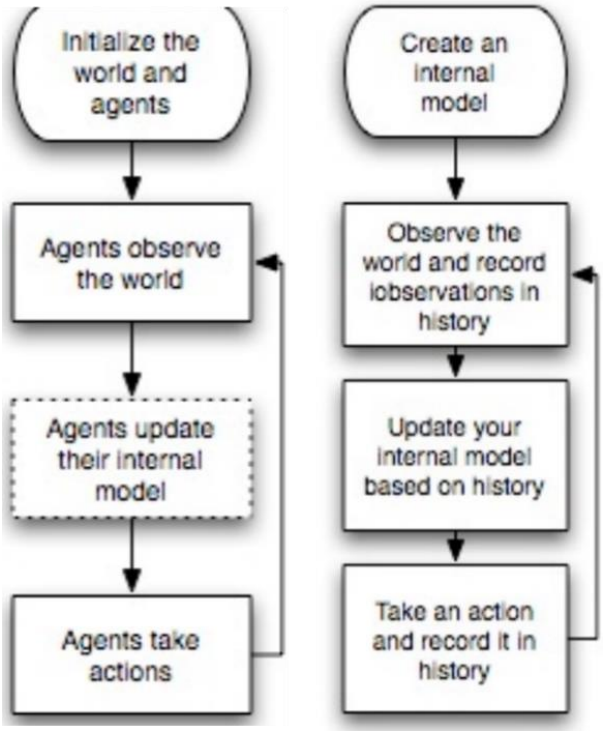


Figure 2-5: Agent-based models cycle (left) and Machine learning cycle (right) (Rand, 2006)

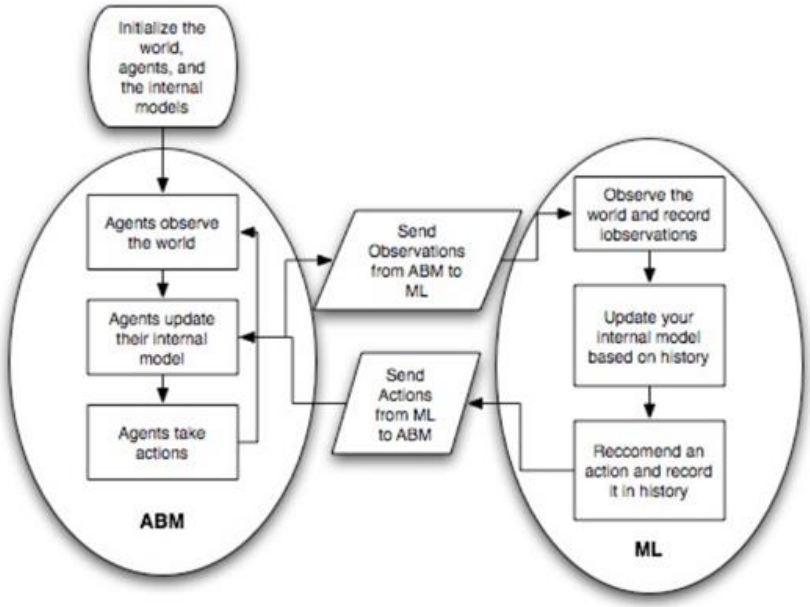


Figure 2-6: Integrated cycle (Rand, 2006)

3. Concept of Methodology

3.1. Introduction to methodology

To simulate the behaviors of condition-based intervention method, an intervention model is constructed. As what has been described in Section 2.2, the intervention model was proposed by Tjalma but has not been implemented. In this chapter, the previous design will be improved to include two sub-models: early-warning model (system) and intervention selection model. The early warning model explores the condition in which the interventions should be implemented (disease outbreak) and the aberration population groups in each GGD region. The intervention selection model determines the intervention strategies that will be implemented and the target population groups for each GGD area. The aberration population groups in each GGD region are sorted out according to the age categories proposed by Tjalma which has been described in Section 2.2. The categories of target population will be fully discussed in Section 3.5. Because of the limited time, only vaccination strategies are introduced as the intervention methods in this research.

3.2. Framework

The intervention model is an agent-based model which consist of the agents, environment and interaction between agents, and environment. Two kinds of entities (agents) are introduced in the model: GGDs and RIVM. GGDs monitor the ongoing patterns (conditions) of disease occurrence in their service region and implement corresponding interventions when a national disease outbreak is detected by the RIVM agent. Attributes and behaviors (interaction) of GGDs and RIVM are presented in the Table 3-1. These behaviors are fully discussed in the following sections.

Table 3-1: Behaviors and attributes of agents in the intervention model

Agent	Behavior	Attributes	Simulated sub model
GGDs	Sense the number of infected cases in their service region and update this value in every time step	- Total infected population in their service region - Infected population for each age group in their service region	Early-warning model
	Compare the current disease occurrence with the history data and detect aberrations using either the Farrington method or SOM method for each age group	- Current infected population for each age group in their service region - Historical infected population for each age group in their service region - Trained SOMs (Section 3.4.2)	
	Identify the targeted population groups for intervention	- Identified target population groups based on the detected aberration age groups in their service region.	Intervention selection model
	Select interventions (only vaccination	- Increasing vaccination coverage ratio for target population groups	

	strategies in this research)	- Time interval between two adjacent interventions	
	Implement interventions	- Increasing the recovered population ratio (vaccinated ratio) for targeted population groups	
RIVM	Sense the number of infected cases and update this value in every time step	- Total national infected population - Infected population for each GGD region	Early-warning model
	Compare the current infected cases with the history data and detecting national aberrations using the Farrignton method and SOM method	- Current total national infected population - Current infected population in each GGD region - Historical national total infected population - Historical infected population in each GGD region	
	Order GGDs to implement interventions	- Alarm signal for each GGD	

The environment of the intervention model is the disease ongoing occurrence of the whole Netherland which is simulated by the disease transmission model constructed by Tjalma (2016). As what has been described in Section 2.2, the disease transmission model consists of three sub models: SEIR (susceptible (S), exposed (E), infected (I) and recovered (R)) model, network model, and population model. The number of infections for different population groups in each GGD region is simulated by the SEIR model. The working flowchart of the intervention model is presented in the Figure 3-1. The solid line in the figure presents the loop of the model running while the dashed line indicates that the intervention selection model will be activated once aberration is detected at the national level. The first step of the intervention model is supervising the disease occurrence data generated by the SEIR model to detect the disease aberrant change in national and regional level using the early-warning model. In this research, two aberration detection algorithms, the Farrington method and the self-organizing map (SOM), were attempted respectively. The performance of these two methods are fully discussed in the chapter 4. The SOM method is finally implemented in the early-warning model as the method performs a better aberration detection ability.

If the aberration is detected at the national level, the intervention selection model will be activated to determine the target population groups and corresponding intervention (vaccination) strategies for each GGD area. The concept of this intervention selection model is described in Section 3.5.

The early-warning model is executed using R while the intervention selection model and disease transmission model are implemented in NetLogo. The interaction between these two programs is implemented using R-Extension in NetLogo.

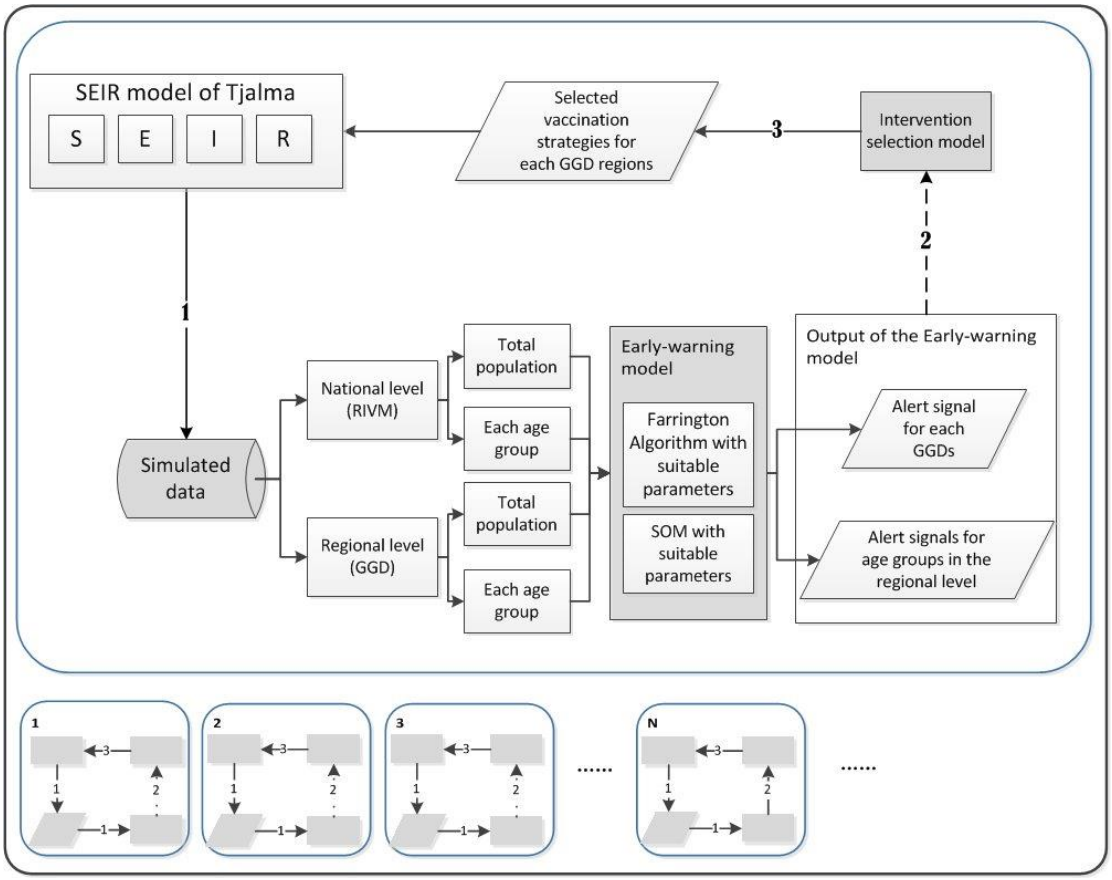


Figure 3-1: Work-flow of intervention model. The dashed line indicates the intervention selection model will be activated once aberration is detected at the national level. The solid line presents the loop of interaction between sub models.

3.3. Data preparation

3.3.1. Simulated data

In this thesis, simulated data is used. There are several reasons for using hypothetical data instead of real data. The first reason is lack of continues long-term historic pertussis information for the whole Netherlands. Furthermore, it is difficult to investigate the predictors of behavior change in the real world (Williams et al., 2015).

The pertussis infection data simulated by the disease transmission model with the current vaccination strategy (infant vaccination) was used to train the early-warning model and role as the benchmark to be compared with the outcome of the model with condition-based intervention (vaccination) method. The data suggests the pertussis condition from 1996 to 2012 of the whole Netherlands. Several parameters in the disease transmission model should be confirmed to be able to simulate the pertussis spread more realistic (Tab.3-2). In this research, these parameters follow the setting provided in Tjalma's thesis. The details will not be discussed further here, and Tjalma (2016) is recommended for the expatiation.

Table 3-2: Parameters requirement of disease transmission model (Tjalma, 2016)

Sub model	**Parameters requirement**
SEIR model	Latent period; Recovery rate; Years of immunity after infection; Infection rate; Vaccinated population
Population model (for each municipality)	Population of each age group; Households; Commuters; Birth/death rate
Network model	Commuting data; Contact matrices

Besides, the disease transmission model simulates the disease spread from three directions in terms of population commuting: only school commuting, only job-commuting, and integrated school and job commuting.

The national pertussis data generated by the model run with current vaccination method considering different commuting conditions are presented in the Figure 3-2. The left graphic presents the data regarding only school commuting while the right one is the data considering both school commuting and job commuting. Because of the similar pattern between only job commuting related data and the data with both school and job commuting, the graphic of previous one is not presented here. The data with only school commuting performs a typical epidemic pattern with disease outbreak every three years which is similar with the real data (Fig. 1-1). However, data including both school commuting and job-commuting suggests an endemic feature after the outbreak in 1997. Meanwhile, the model run with the integrated school commuting and job commuting is quite slower than the model run only introduces the school commuting. The presupposition of the condition-based intervention method and the intervention model is that the disease should presents an epidemic pattern. Thus, only the school commuting is concerned in this research, though this condition is not very realistic. There are six times pertussis outbreaks over the simulated time period. These outbreaks are referenced as 6 epidemic waves (Tab. 3-3).

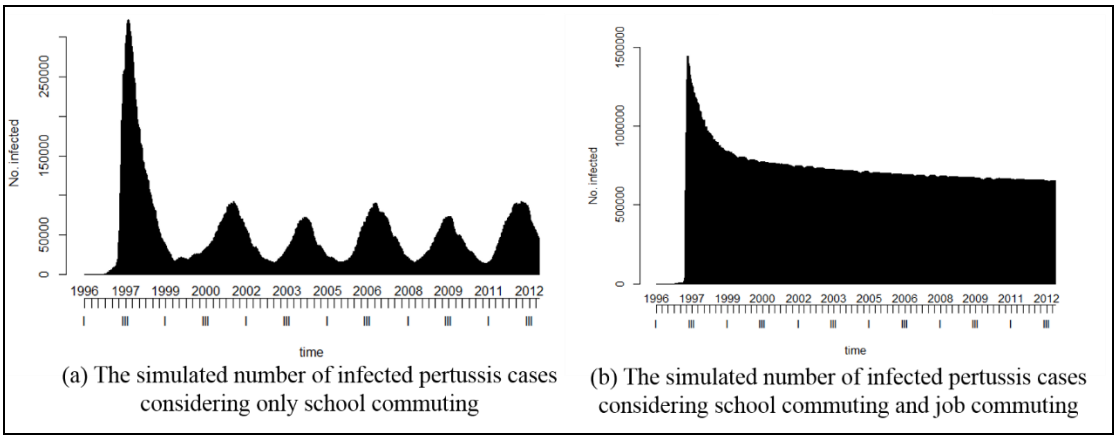


Figure 3-2: The simulated epidemic curves considering different commuting conditions. (a) The total number of infected pertussis cases considering only school commuting (b) The total number of infected pertussis cases considering school commuting and job commuting

Table 3-3: Overall description of first kind of simulated data considering only school commuting

Epidemic wave	Duration (week)	Tick of the peak value
Wave 1	1 – 173	85
Wave 2	174 – 370	289
Wave 3	371 – 494	426
Wave 4	495 – 640	563
Wave 5	641 – 777	705
Wave 6	778 - 881	846

The simulated information which includes infected cases for nine age groups, susceptible population, exposed population, infected population and recovered population are recorded per seven ticks (one week) for each municipality. These data are recorded in a csv file and will be used to as input data for the intervention model.

3.3.2. Disease data for Early-warning model

In order to train the early-warning model, the data generated by the disease transmission model run with current vaccination method over 16 years simulated time period are used. The original input data includes weekly pertussis cases for nine age groups of each municipality and the name of corresponding GGD for each municipality.

The Farrington algorithm in early-warning model is implemented using the Surveillance package in R. The Surveillance package provides a S4 sts data class (Paul & Meyer, 2016). The sts data class consists of several parameters. The first is a slot observed $\boldsymbol{n \times m}$ matrix which includes **n** time series and **m** entities in monitored, e.g. municipalities, age groups (Höhle & Mazick, 2010). The second is epoch which indicates the time index. The index ***freq*** suggests the number of observation per year while 52 is for weekly recorded data and 12 is for the data recorded per month (Maëlle et al., 2014). In this research, the disease occurrence for each municipality are monitored per week. The format of the time series uses the ISO 8601 standard which represents week-numbering for a year from the first Monday of week 1 to the last Sunday before next ISO year. Two types of observed matrixes are introduced to correspond to the RIVM agent and GGD agent. As for the RIVM, the monitored entities**(m)** represent the 25 GGDs. The entities for GGDs agents indicate age groups in their service region.

As for the SOMs method, the input data is log transformed at first to ensure a normal distribution. Then the data is stored in different structures in terms of spatial and temporal perspective. According to Augustijn (2013), there are three types of data organization: space over time **(S × T)**, space over single wave **(S × W)** and time over space **(T × S)**. In this research, GGDs × Time **(S × T)** is used to classify the GGDs with similar disease outbreak pattern; Time × GGDs **(T × S)** explores the national disease diffusion pattern over time. Besides, data organization for the regional level (GGDs) types time over infected cases for nine age groups in certain GGD. This data structure explores the aberration age groups of each GGD in their service region. Complete 16 years training data is used to train the SOM in order to identify all patterns of disease occurrence. This will be fully described in the Section 3.4. Meanwhile, the dimension of the new input dataset should be same with the training dataset if the trained SOM is applied to detect the pattern of real-time data. Besides, the detected real-time data can't be classified directly. If the detected real-time data contains x vectors, the 1^{th} to x^{th} vectors in the training dataset should be replaced by the real-time data resulting in a new dataset. The reason is described in the Section 4.3.

3.3.3. Data input for intervention selection model

The outcomes of the early-warning model will be used as the input data for the intervention selection model. The input data includes the national alarm signal and the tab of the aberration population groups for each GGD.

3.4. Early warning model

The early-warning model detect the aberrations of pertussis infections occurring at an unknown time to warn against a possible outbreak at an early stage. The main problem of the aberration detection is to investigate the spatiotemporal pattern of the disease outbreak. Traditional case-based methods like the Farrington algorithm always bias toward the temporal character of disease using statistic knowledge. In this research, there is an attempt that using self-organizing map (SOM) as the aberration detection algorithm to investigate the spatiotemporal pattern of pertussis outbreak. A hierarchical diffusion pattern has been identified for pertussis spread in previous research (Girmay, 2012). The hierarchical diffusion refers that the disease spread from one place to another in an order of sequence (Cliff et al., 1981). The spread is mostly described by the synchrony between each location (GGD in this research) and the diffusion trajectories which have been investigated using the SOM method (Augustijn et al., 2013). To confirm the early-warning model can be constructed even if the SOM not work as expected, the Farrington method was also used. The outcomes of the Farrington method and the SOM method are discussed in the Chapter 4. These two detection methods observe the disease from both national level (RIVM) and regional level (GGDs). The expect outputs of the Farrington method include:

1. Alarm signal in certain GGDs at time **a.**
2. National alarm signal at time point **b**.
3. Regional or national alarm signal for certain age groups at time point **a** or **b**.

The expect results of SOM method contain:

1. National alarm signal at time point **c**.
2. Regional alarm signal in certain GGDs at time point **d**.
3. Similarity of disease transmission pattern among each GGD.
4. Aberration age groups in national and regional level.

The conceptual design model framework is presented in Figure 3-3. If several regional alarm signals are generated earlier than the national alarm, the intervention selection model will be activated in these GGDs' service area and the region which have similarity disease transmission pattern. If the national alert is sounding, all GGDs should execute interventions. However, only the national alarm and aberration age groups in regional level (GGD) are finally concerned in the model and the rest functions haven't been completed yet because of the limited time.

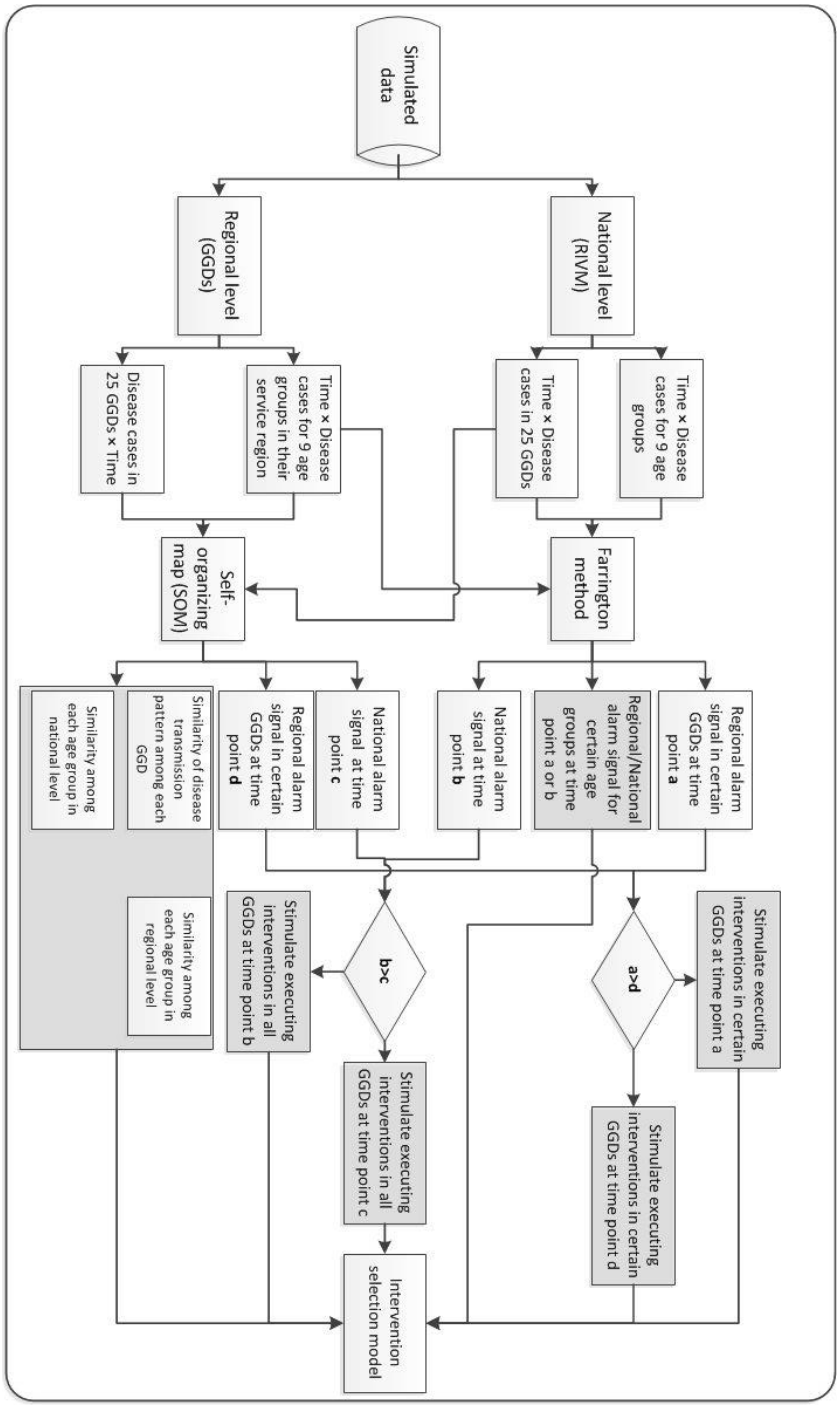


Figure 3-3: The conceptual framework of early-warning model. The shadow polygons indicate the output of the model.

3.4.1. Farrington Algorithm

The purpose of the method was to establish a robust system for the weekly reports on infections at the former Communicable Disease Surveillance Centre. The algorithm is currently implemented at European public health institutes (Hulth et al., 2010). Comparing with other frequency-based case definition algorithms, the Farrington method is readily comprehensive, widespread application and robustness. The method is executed follows these steps:

a) **Determining the reference data (historical data):** The Farrington algorithm performs one-timepoint detection. It predicts the observed number of counts y_{t_0} for the current time point t_0 by a statistical algorithm (Höhle & Mazick, 2010). The prediction only counts from the

corresponding period in the past years. In this research, the current week t_0^{week} of year t_0^{year} is compared with weeks from t_0^{week-3} to t_0^{week+3} of year t_0^{year-4} to t_0^{year-1}, giving a total of 28 base-line reference values. According to Noufaily et al. (2013), it is better to exclude the last 26 weeks before the current time point in order to maintain sensitivity when an outbreak occurred recently.

b) **Fitting the over-dispersed Poisson generalized linear model (GLM) to the reference data:** The current count y_{t_0} and corresponding historical data are assumed to be distributed with mean μ_i and variance $\emptyset\mu_i$ (Farrington et al., 1996). An over-dispersed Poisson generalized linear model (GLM) with log-link is fitted to the reference data y_t:

$$\log(\mu_t) = \alpha + \beta t \tag{3.4-1}$$

Where α and β are coefficients to estimate.

c) **Predicting the expected number of counts μ_{t_0} for the current timepoint t_0:** The original Farrington assumed that the error of the expected counts $y_{t_0} - \hat{\mu}_{t_0}$ is normally distributed, thus:

$$\mathrm{Var}(y_{t_n} - \hat{\mu}_{t_n}) = \mathrm{Var}(\hat{y}_{t_n}) + \mathrm{Var}(\hat{\mu}_{t_n}) = \emptyset\mu_0 + \mathrm{Var}(\hat{\mu}_{t_n}) \tag{3.4-2}$$

where $\mathrm{Var}(\hat{y}_{t_0})$ is the variance of observation and $\mathrm{Var}(\hat{\mu}_{t_0})$ is the variance of predicted value.

d) **Determining the threshold:** The threshold is defined as upper limit of a one-side $(1-a)\cdot 100\%$ confidence interval(Maëlle et al, 2014):

$$U_{t_n} = \hat{\mu}_{t_n} + z_{1-a}\widehat{Var}(y_{t_n} - \hat{\mu}_{t_n}) \tag{3.4-3}$$

with z_{1-a} is the $1-a$ quantile of the standard normal distribution. The assumption of the improved Farrington is that $y_{t_0} \sim NB(\mu_{t_0}, v)$ while $v = \frac{\mu_{t_0}}{\emptyset - 1}$. The upper bound is calculated based on the predicted number of counts μ_{t_0} and its variance for the current time point.

e) **Detecting the disease aberration change:** The observed count y_{t_0} is compared with the threshold U_{t_0}. The time point t_0 will be labelled as "alarm" if the y_{t_0} is higher than U_{t_0}.

These five steps are implemented automatically by the Surveillance R package. Several parameters should be confirmed to improve the method more suitable for the pertussis data (Tab.3-4). The workflow of this method is presented in the figure 3-3.

Table 3-4: Parameters used in the Farrington method in R

Parameter	Description
b	The number of reference year.
w	Weeks to include before and after the current week in each year.
weightsThreshold	The threshold for reweighting past outbreaks
pThresholdTrend	Threshold for deciding whether to keep trend in the model
thresholdMethod	Method to be used to derive the upper-bound.

3.4.2. Self- organizing maps (SOMs)

The self-organizing maps (SOMs) combined with a Sammon's Projection and GIS has been explored to identify the spatiotemporal disease diffusion pattern (Augustijn et al., 2013). However, the applicability of SOMs in detecting the aberration of an epidemic has not been proposed yet. In this research, there are three main outcomes of SOM (combined with Sammon's Projection):

1. Synchrony between GGDs' service regions.
2. Aberration detection of pertussis in national and regional level.
3. Aberration age groups in each GGD region.

SOM is one class of unsupervised artificial neural network which clusters high dimensional data by projecting it into two-dimensional space. This 2-D lattice (space) consists of number of neurons. Neurons are trained to observe the patterns of the input data samples and updated at each iteration. The detected

pattern by certain neuron is represented as a codebook vector (Augustijn et al., 2013). Neurons are assigned in the lattice following an order sequence in which Neurons with similar codebook vector are closer to each other in the lattice than the dissimilar ones (Kohonen, 1998). The input data samples are mapped into the neuron with the most similar extract pattern. The mapped back neuron is also referenced as best matching unit (BMU). In this research, the SOM is trained using Kohonen R package follow these steps:

a) **Normalizing the input data:** As what has been described in the Section 3.3.2, there are three types of input data:
   1. Time × GGDs: Identifying the national disease diffusion trajectories which can be extended used to detect the national aberration change;
   2. GGDs × Time: Exploring the synchrony of disease outbreak characteristic between each GGD;
   3. Time × Age (for each GGD): Detectingt the aberration age groups for each GGD;

b) **Training the SOM:** The training process follows these three steps:
   1. Determining the number and type (the shape of nodes, rectangular or hexagonal) of nodes
   2. Assigning the input data.
   3. Identifying the BMU with default number of iterance and learning rate.

c) **Mapping the clustering result to the original input data:** The input vectors are assigned in a certain neuron after SOM training. This classified result is mapped back to the original input data. One of the advantages of SOM is that the topological properties of input data can be preserved after the model training. This is helpful to explain the characteristics of input data using the result of SOM more easily. As for GGDs × Time SOM, GGDs projected in the same cluster present synchronic on disease transmission pattern. The mapped back results can detect the similarity among GGDs and the aberration age groups for each GGD directly which is fully discussed in the Section 4.2.

d) **Secondary clustering the SOM grid using hierarchical clustering method:** Because of the number of neurons is hard to be determined, a large number always be used at first. The secondary clustering can improve the classified performance of the initial SOM result.

e) **Generating disease diffusion trajectories using Sammon's projection from SOM output:** This step aims to detect the national aberration using the Time × GGDs. Sammon's projection has been introduced integrating with the results of trained Time × GGDs SOM to present the trajectory of disease spread pattern (Augustijn et al., 2013). In this research, codebook vectors in the trained SOM for eight waves (Tab.3-3) are projected in 2-D space respectively using the Sammon's projection. The diffusion trajectory is visualized using the arrow line which connects points (mapped neurons) in the projection space. The way in which using the outcomes of the detect diffusion trajectories to identify the national disease aberration is fully discussed in Section 4.2.

After training SOM follow previous steps, the trained SOM can be used in the early-warning model to interact with the intervention selection model and disease transmission model.

3.5. Intervention selection model

The purpose of the intervention selection model is to simulate the way in which GGDs prevent pertussis when they are alarmed using rule-based method. The main problem is to determine the target groups of population that should get the interventions.

The ongoing pattern of pertussis occurrence is supervised in the early-warning model. If the aberration is detected in the national level (RIVM), all GGDs are ordered to implement interventions. GGDs identify the target groups based on the identified aberration age groups from the early-warning model in their service region. The various disease pattern in each GGD leads different aberration information might be generated for GGDs at the same tick. The various detection results might result in heterogeneous behaviors of GGDs.

As mentioned before, only vaccination is concerned as the intervention method in this research. Similar with the Tjalma model (Tjalma, 2016), vaccinations are simulated by increasing the percentage of recovered population in the target population groups. Several parameters should be confirmed.

The first is the ***target population group*** that should be vaccinated. If the age group 1 (0-5month) is identified as aberrant, the target groups should be the pregnant women. The first reason is the infants younger than 5 to 7 months are still susceptible because the antibodies only take effect after the third injection. The second reason is the antibodies can be passed from the vaccinated pregnant women to their infants. If the age group 2 is presented as aberration, the target population should be age group 5 to 9 (older than 25 years old). The main reason is the infant vaccination has been considered in the original disease transmission model and the effectiveness of the vaccine will maintain for a few years. It is impossible to increase the recovered ratio of young children (less than 5 years old) due to the high vaccination coverage of infants. The second reason is the young children are assumed only contact with their family members. The age groups 5 to 9 are adults might be the parents or grandparents of the child. Vaccination of these adults can be seen as an efficient way to protect the children. If the else age groups are identified as aberrant, they will be seemed as the target groups directly.

The second is the ***increase vaccination ratio*** for the target population (vaccinated) group. In Robin's research (2010) which simulate the performance of adolescents vaccination, the vaccination coverage in the Netherlands of infants was also assumed for adolescents. The same assumption is used in this thesis. Because of the limited time, the immunity is built up immediately for the vaccinated population after injection to shorten the model running time.

The third is the ***minimum implementation interval***. The assumption is only one intervention will be implemented by each GGD in one disease outbreak. Thus, the implementation interval is introduced to constrain GGDs' behavior that the second intervention will not be executed in one disease outbreak and to confirm that GGDs can implement interventions in the following disease outbreak. As Figure 3-2 (a) indicates, the average interval between two disease outbreaks is 3 years and the same interval will be applied for adjacent vaccinations.

4. Construction of the Early-Warning Model

The aberration detection methods described in the previous chapter uses a combination of several parameters to identify the aberration. This chapter describes how the input parameters were confirmed and the comparison between the Farrington method and the SOMs.

4.1. Calibration of the Farrington method

Several parameters should be confirmed both for the Farrington algorithm and SOM to achieve the best prediction ability of pertussis data. Parameters in Farrington method are discussed at first.

The main problem of parameters setting in Farrington algorithm is using whether the original approach proposed by Farrington (1996) or the improved version (Noufaily et al., 2013) which has been discussed in the Section 3.4.1. The selection directly impacts the threshold determined method and corelated parameters *weightsThreshold, pThresholdTrend* and *thresholdMethod*. Parameters setting for these two versions of the Farrington method are presented in the Table 4-1.

Table 4-1 : Parameters for the original and improved Farrington method (Salmon et al., 2016)

Parameter	**The original Farrington method**	**The improved Farrington method**
b	4	4
w	3	3
weightsThreshold	1	2.58
pThresholdTrend	0.05	1
thresholdMethod	nbplugin	delta

Several indexes have been proposed to evaluate the performance of the detection methods, such as in-control run-length which indicates the time interval before the first wrong alarm, out-of-control length which presents average range of time to determine an already occurred change (Höhle et al. , 2010), sensitivity which was defined as the number of detected outbreaks divided the total number of outbreaks, time to detection which presents the average time range between the first correct detection time point and the disease outbreak (Hutwagner et al., 2005) and so on.

However, there is no specific threshold of disease outbreak for the simulated data. In order to evaluate the performance of two Farrington methods, detection before peak value is defined as the average time range between the first correct detection tick and the disease peak tick. This index and the sensitivity are used to overall estimate the outcome of the national simulated data using different version of Farrington method (Tab.4-2). Because the first three years' infected cases are used as the history data, the detection is start from the fourth year. The red triangle above the time shaft indicates the detected aberrations (Fig. 4-1). The first disease outbreak in the figure indicates the wave 2 in the Table 3-3 and the last outbreak is wave 6.

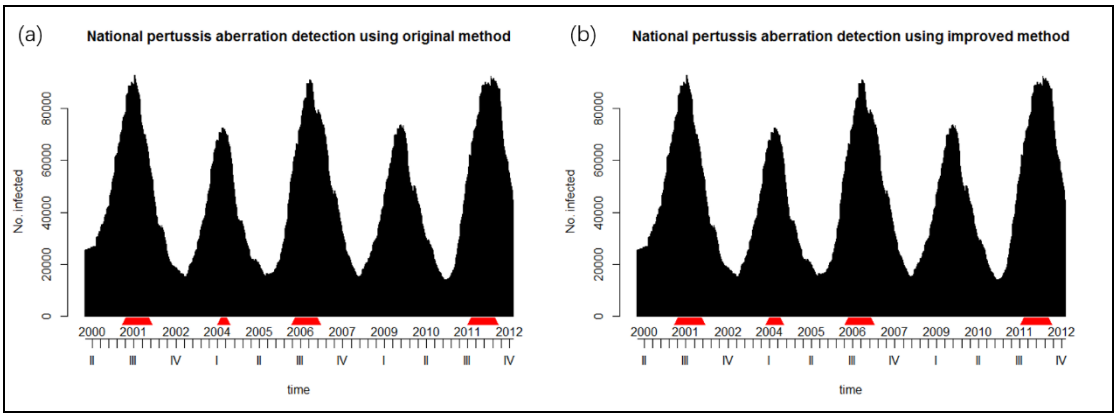


Figure 4-1: National level pertussis aberration detection using (a)original method (left) and (b)improved method(right)

Table 4-2: Overall evaluation of different version of Farrington method for national detection

Algorithm version	**Sensitivity**	**Detection before peak value (weeks)**
Original	0.8	18.5
Improved	0.8	18.75

As table 11 indicates, the detection using two version Farrington methods have similar outcome with 80% sensitivity and around 18.6 detection before peak value. The fourth outbreak is ignored neither by the original method or by the improved method. The same phenomenon also exists in the regional detections which will be discussed later. Detection before peak value of original method is a little bit less than the improved one. The sounding alarm more than 18 weeks before the disease peak value is enough for interventions making impact on disease occurrence.

The detections at regional level for all disease outbreak using two versions are presented in the Figure 4-2 and 4-3 respectively. The red points in the figure indicate the detected aberrations for different GGD region within 5 disease outbreaks. The result presents various detection patterns for different GGDs using whether the original version or the improved method. The outbreaks of wave 2, 4 and 6 are detected by all GGDs using either original approach or improved one. However, almost all GGDs are silent to the wave 5. Meanwhile several GGDs keep quiet to the wave 3 and this phenomenon is various between two versions.

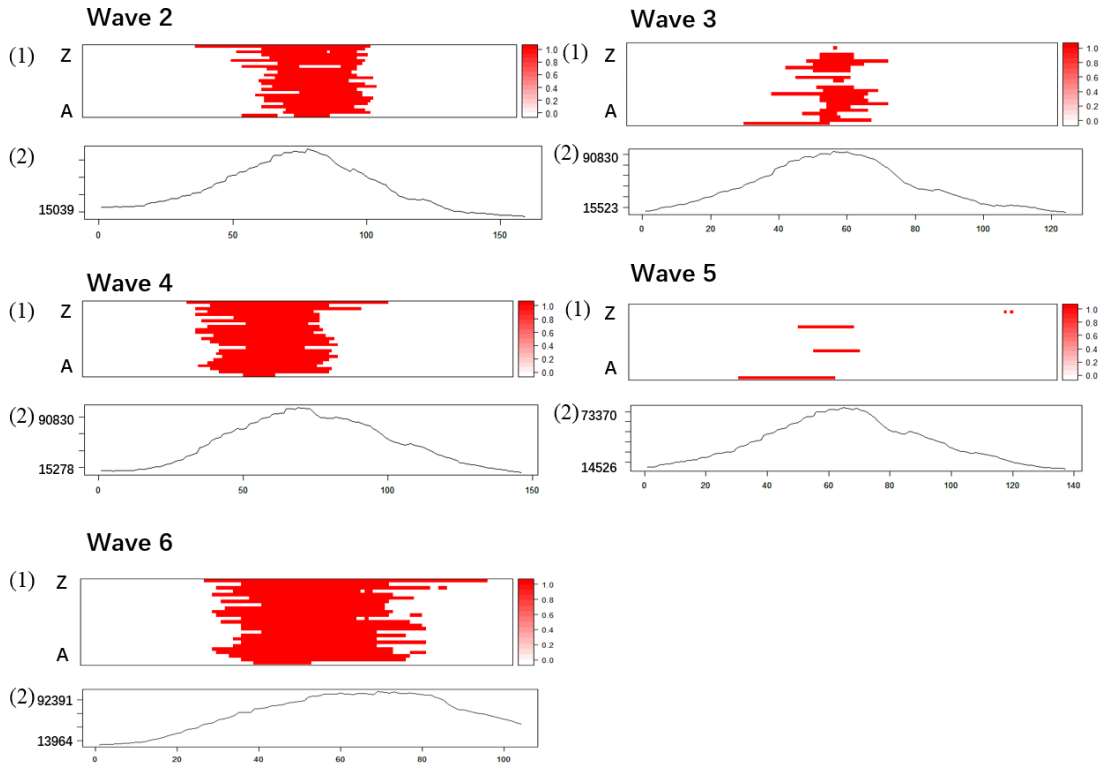


Figure 4-2: Aberration detection for each GGDs using improved Farrington method. (1) Sounding alarms for each disease outbreak wave in 25 GGDs, with GGDs arranged in ascending order of initial. (2) Time series of national total infected cases

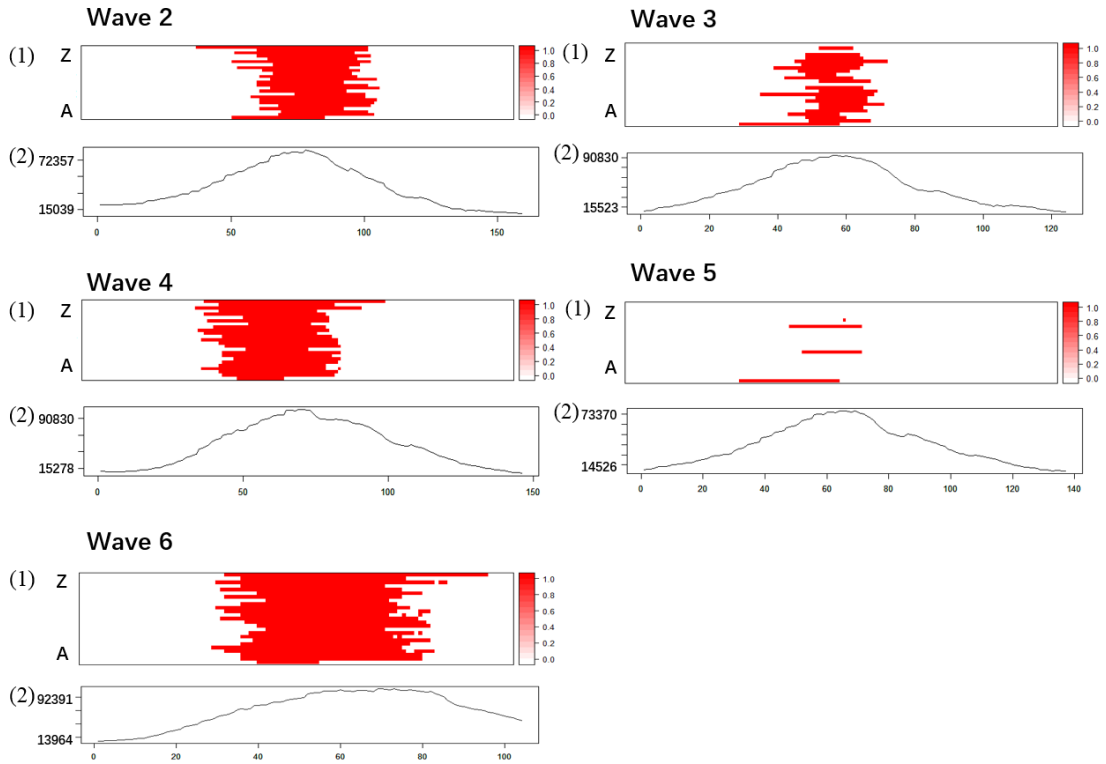


Figure 4-3: Aberration detection for each GGDs using original Farrington method. (1) Sounding alarms for each disease outbreak wave in 25 GGDs, with GGDs arranged in ascending order of initial. (2) Time series of national total infected cases.

In this case, the detection before peak value and sensitivity for each GGD are also used to estimate two versions of Farrington method. As Table 4-3 indicates, both of the sensitivity index and detection before peak value of improved Farrington method is higher than the original one for the regional aberration detection.

Table 4-3: Overall evaluation of different version of Farrington method for regional aberration detection

Algorithm version	Sensitivity	Detection before peak value (weeks)
Original	0.76	20.11
Improved	0.808	20.19

Besides, the aberration changes of nine age groups for either the national level or regional level are also detected in the early-warning model. Same with the previous estimation method, sensitivity and detection before peak value are also used. Table 4-4 presents the sensitivity and detection before peak value of improved Farrington method is higher than the original one for the national age groups aberration detection. As for the comparison between the performance of these two methods for age groups in all GGD regions presents a similar result.

Table 4-4: Overall evaluation of different version of Farrington method for national age groups aberration detection

Algorithm version	Sensitivity	Detection before peak value (weeks)
Original	0.96	16.30
Improved	0.98	16.77

Generally speaking, the improved Farrington method can detect the disease outbreak earlier than the original one, and the improved method is more sensitive than the original approach. Therefore, the improved Farrington is selected to compare with the SOM method.

4.2. Calibration of Self-organizing map (SOM)

The main problem of using the SOM method to detect disease outbreak is how to transform the outcomes of the SOMs to the warning signal. As what has been described in the Section 3.4.2, there are three kinds of data structure: Time × GGDs, GGDs × Time and Time × Age (for each GGD). Several input parameters should be confirmed to train the SOM: *the dimension of the lattice, the topology of SOM lattice, number of iteration updates* and *learning rate*.

The topology of SOM lattice and the learning rate for all input data structures follow the Augustijn research (2013). The hexagonal lattice is used, and the learning rate is declining from 0.05 to 0.01. The distance from neuron's codebook to the input vector is reduced during the iteration process. Ideally, the distance should reach a minimum value. This progress is represented in a curve graphic. If the curve is continually decreasing, more iterations are needed. The number of iteration updates was identified for three input data structures, for Time × GGDs is 1000, 3000 for Time × Age and 5000 for GGDs × Time. As Figure 4-4 indicates, the training process graphic curves of these three data structures reach the minimum plateau and the change is less than 0.05.

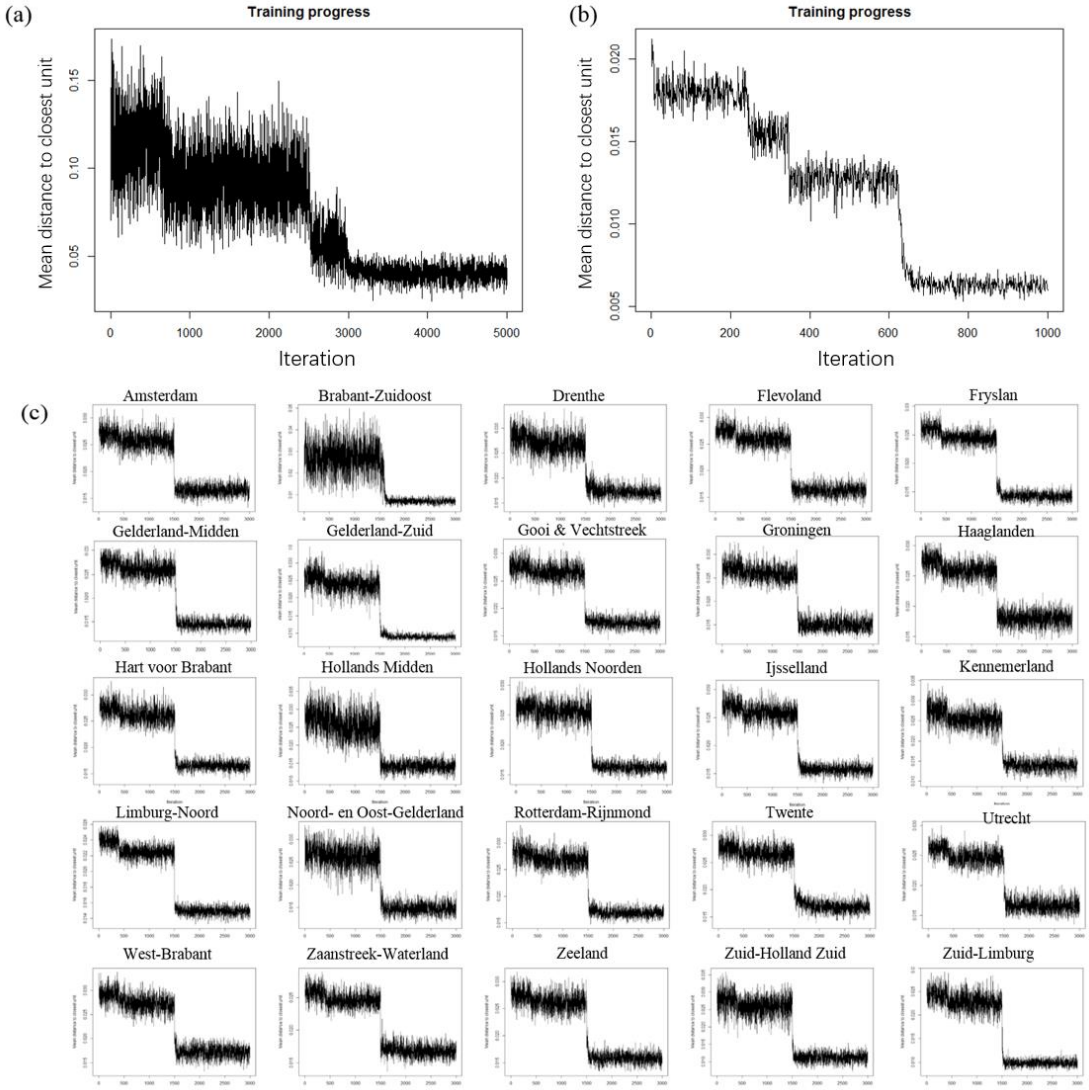


Figure 4-4: Mean distance changes between input vectors and corresponding BMUs during the training process within Time × GGDs SOM (a), GGDs × Time SOM (b) and Time × Age for 25 GGDs (c)

The large lattice dimension always presents more detailed patterns while the small lattice presents more general patterns. Therefore, the selection of the lattice size is a trade-off between more information and manageable few patterns (Liu et al., 2006).

The Time × GGDs SOM is used to detect the national pertussis aberration. The trained SOM requires that all aberrant patterns can be identified correctly using small map size. After training, the dimension of the Time × GGDs lattice is determined as 4 × 5 and the trained lattice is presented in Figure 4-5 (a). Neuron1 is located in the lower left corner and neuron 20 is located in the top right corner. For this type of SOM, the codebook vector is compared with the input vectors which include a serial number of infected people in each GGD at each tick. The input vectors will be mapped to the neuron with the most similar codebook vector. The assumption is that the increase of the number of infected cases in a specific GGDs should be higher than the increase in other GGDs at the beginning of the aberration and the infected cases in all GGDs should be high when warning for an outbreak. In the SOM lattice, the codebook vectors are visualized using a line graph. Therefore, neurons with the codebook vector performing the pattern of previous two assumptions will be identified. The input vectors which are mapped into these neurons are assumed as aberrations. Nine neurons are selected that might indicate an outbreak, neuron 1, 2, 4, 5, 6, 8, 9, 11, 12 and 13. To check the assumption, the input vectors which are classified in these neurons are mapped on the disease curve and visualized using different colors (Fig 4-5. (b)). As Figure 4-5 (b) presents, all the vectors mapped into the selected neurons distribute around the peak value of each disease outbreak.

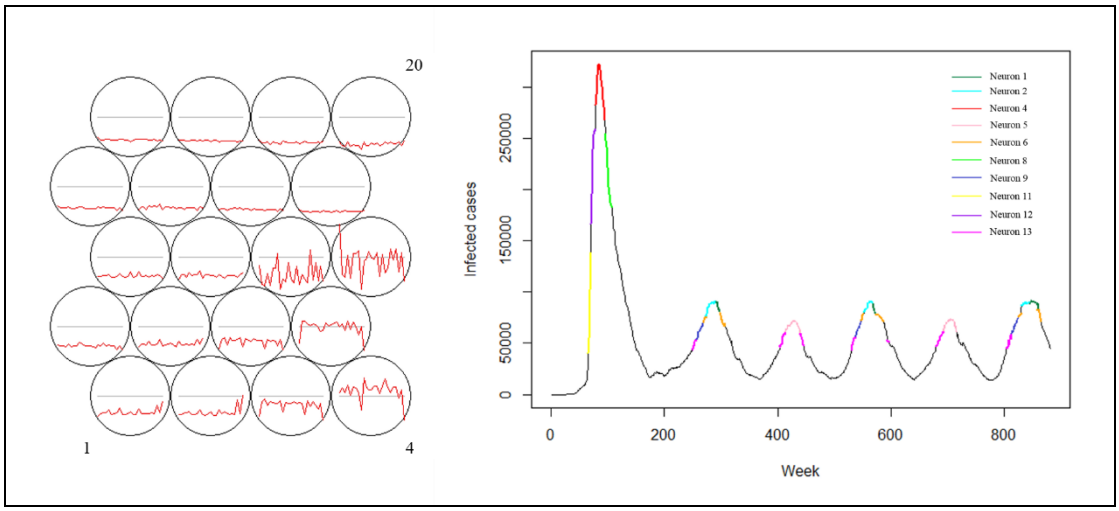


Figure 4-5: Codebook vectors (a) and disease curve with mapped back neuron (b) of Time × GGDs SOM.

The SOM combined with Sammon's projection has been proposed to explore the diffusion trajectory using Time × GGDs data. In this case, this method is introduced to verify further whether the selected neurons can be applied to detect the aberrations and identified which neurons can seem as the beginning of aberrations.

There are 6 epidemic waves in the simulated data (Tab.3-3). As Figure 4-6 and Table 4-5 indicate, the trajectories can be partitioned into 3 groups, wave 1 is group 1, wave 2, 4 and 6 is group 2 and wave 3 and 5 is group 3. The groups 2 and 3 present a symmetrical pattern to a certain level. They always start from neuron 19, go ahead and reach neuron 2 or 5, then go back to the start point. This pattern is distinct similar to the character of the epidemic curve and the feature of codebook vectors (Fig. 4-5). The codebook of neuron 2 and 5 indicates the pattern that the infected cases reach the peak value in wave 2 to 6. The neurons before 2 and 5 in the trajectories present the pattern of increasing infected cases. The codebook of neuron 13 is identified as the pattern which indicates the beginning of disease aberrant change because of several reasons. The first is that the neurons (5 or 9) following the neuron 13 present obviously infected cases increasing pattern. The second is that all waves are sensitive to the neuron 13 except wave 1. The third reason is that there is enough time for GGDs to implement interventions before disease reaching the peak value if the alarm is sounding when the input vector is mapped into the neuron 13.

Neurons with similarity codebook vector are projected closely to each other in the projection space, and the character of vectors should be obviously different when their neuron are far away from each other. As Figure 4-6 indicates, neuron 11, 12, 4, 8, 3 and 7 are scattered in space away from the other neurons. These isolated neurons are all connected in the wave 1. This feature is also similar to the outcome getting from the codebook vectors. According to visual analysis of the outcome of Figure 18 and 19, neuron 4 indicates the pattern that the infected cases reach the peak value in wave 1, and neuron 11 is identified as the beginning of disease aberration in wave 1.

Therefore, the Time × GGDs trained SOM with 4 × 5 lattices can be used to detect the aberration. If the input vector is mapped into the neuron 11 and 13, all GGDs will be alarmed to implement interventions. Meanwhile, the warning state will be maintained unless the pattern of the peak value is identified (neuron 4, 5 and 2).

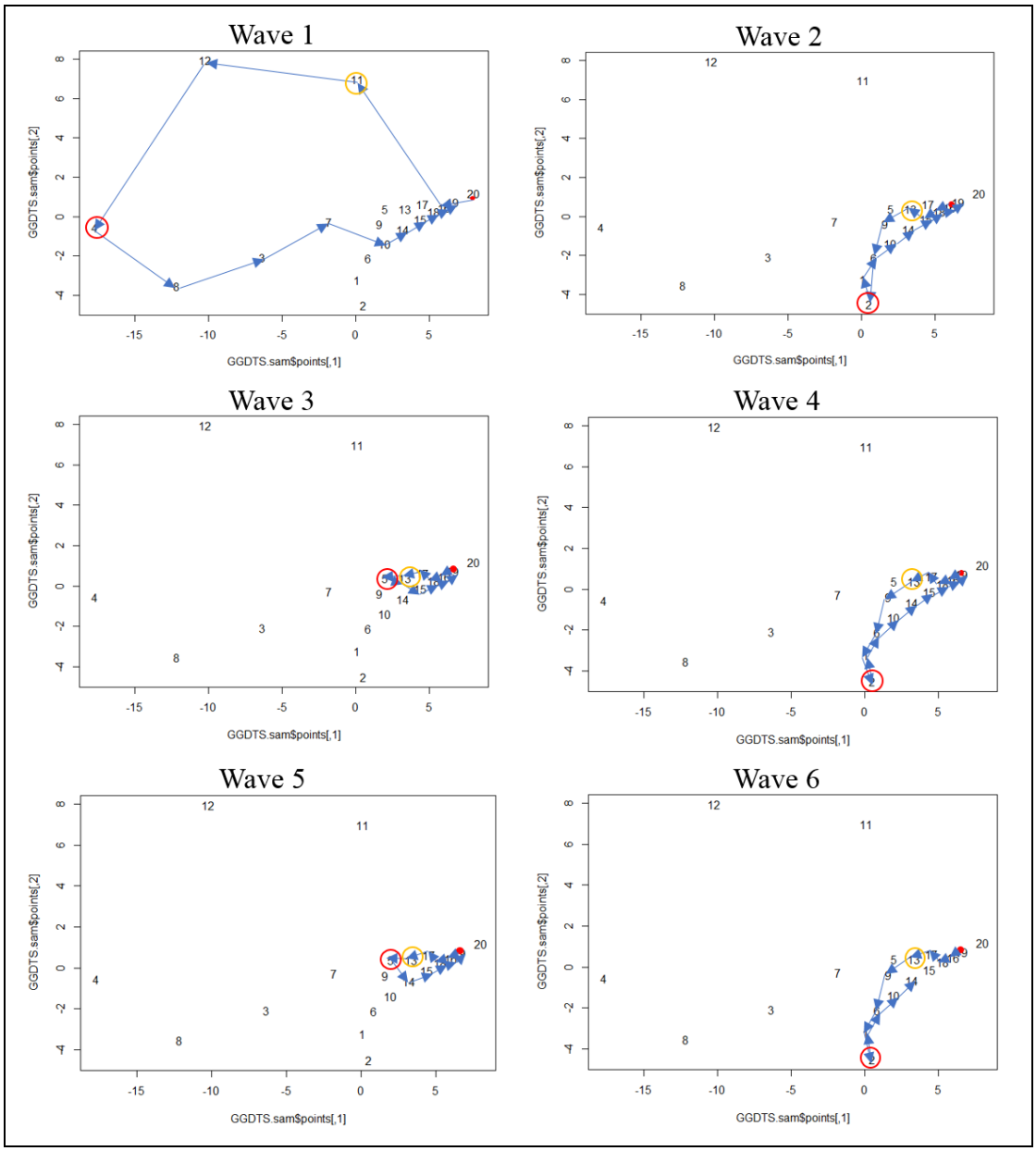


Figure 4-6: Diffusion trajectories on Sammon's Projection. The start neuron for trajectories are highlight using red point. The neurons present the pattern of disease reaching peak value are visualized using red circle and the neurons indicate the pattern of aberration cropping up are represented by orange circle.

Table 4-5: Diffusion trajectory for six waves while the neurons present the pattern of disease reaching peak value are visualized red and the neurons indicate the pattern of aberration cropping up are represented in orange.

Wave	Diffusion trajectory																
Wave 1	20	16	11	12	4	8	3	7	10	14	15	18	16	19			
Wave 2	16	18	15	13	9	6	2	1	6	10	14	15	18	16	19		
Wave 3	19	16	18	17	13	5	13	15	18	16	19						
Wave 4	19	16	18	17	13	9	6	1	2	1	6	10	14	15	18	16	19
Wave 5	19	16	18	17	13	5	14	15	18	16	19						
Wave 6	19	16	18	17	13	9	6	1	2	1	6	10	14				

As mentioned in Section 3.3.2, the detected real-time data should replace corresponding part of the training dataset to generate a new dataset for SOM analysis. The main reason is to confirm the detected data can be scaled to a suitable value. The result of Time × GGDs trained SOM is used as an example. The training dataset indicates the data in the whole simulated period are recorded and the real-time dataset means only the data before week t is known. As the Figure 4-7 indicates, the difference between the observed value at point a and b is clearly presented in their mapper back neuron. The codebook in the neuron 13 is located at the bottom as the scaled value of the data recorded at point b is low compared with the value of the data stored at point a. Though the pointed b and c are recorded at the same time t and store same data vector, the mapped back neuron for them should not be the same. Because the point c is the peak point in the real-time dataset during the observed period which indicates the scaled value for data stored at week t should be high. To confirm the same data input vector can be mapped back into the same neuron, the detected real-time data should replace corresponding part of the training dataset to generate a new dataset at first.

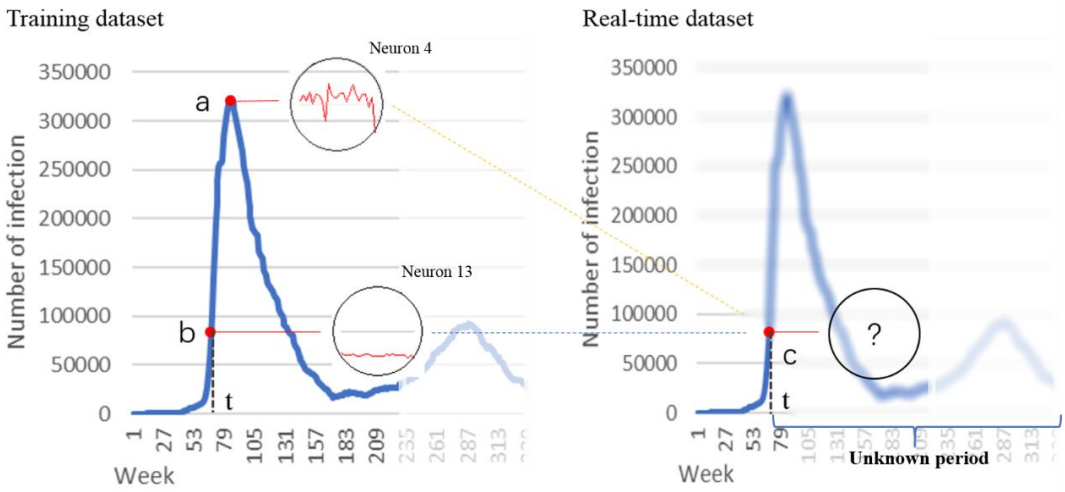


Figure 4-7: The diagram of the reason why the detect real-time data should replace the training dataset to generate a new dataset before SOM analysis.

The SOM using GGDs × Time data explores the synchrony of disease outbreak pattern between GGDs. The dimension of the GGDs × Time SOM lattice using 3 × 4 neurons was tested first. The mapped back results for the GGDs are presented in the Table 4-6. GGDs which are classified in the same neuron have the similarity disease diffusion feature. These correlations can be verified when comparing with the component planes of the Time × GGDs SOM (Fig.4-8). In the experiment, several GGDs with similar pattern are classified in different neurons, i.e. the component planes in neuron 1, 2 and 5. This result indicates the trained lattice (GGDs × Time SOM) is larger than the clusters need.

Table 4-6: Mapped back neuron for each GGD

NEURON	GGD REGION(S)
1	Gelderland-Zuid; Groningen; Noord- en Oost-Gelderland
2	Brabant-Zuidoost; Hollands Midden
3	Utrecht
4	Drenthe; Ijsselland; Twente; West-Brabant
5	Fryslan; Gelderland-Midden; Zuid-Holland Zuid
6	Hart voor Brabant
7	Gooi & Vechtstreek; Kennemerland; Zaanstreek-Waterland
9	Amsterdam
10	Flevoland; Zeeland; Zuid-Limburg
11	Hollands Noorden; Limburg-Noord
12	Haaglanden; Rotterdam-Rijnmond

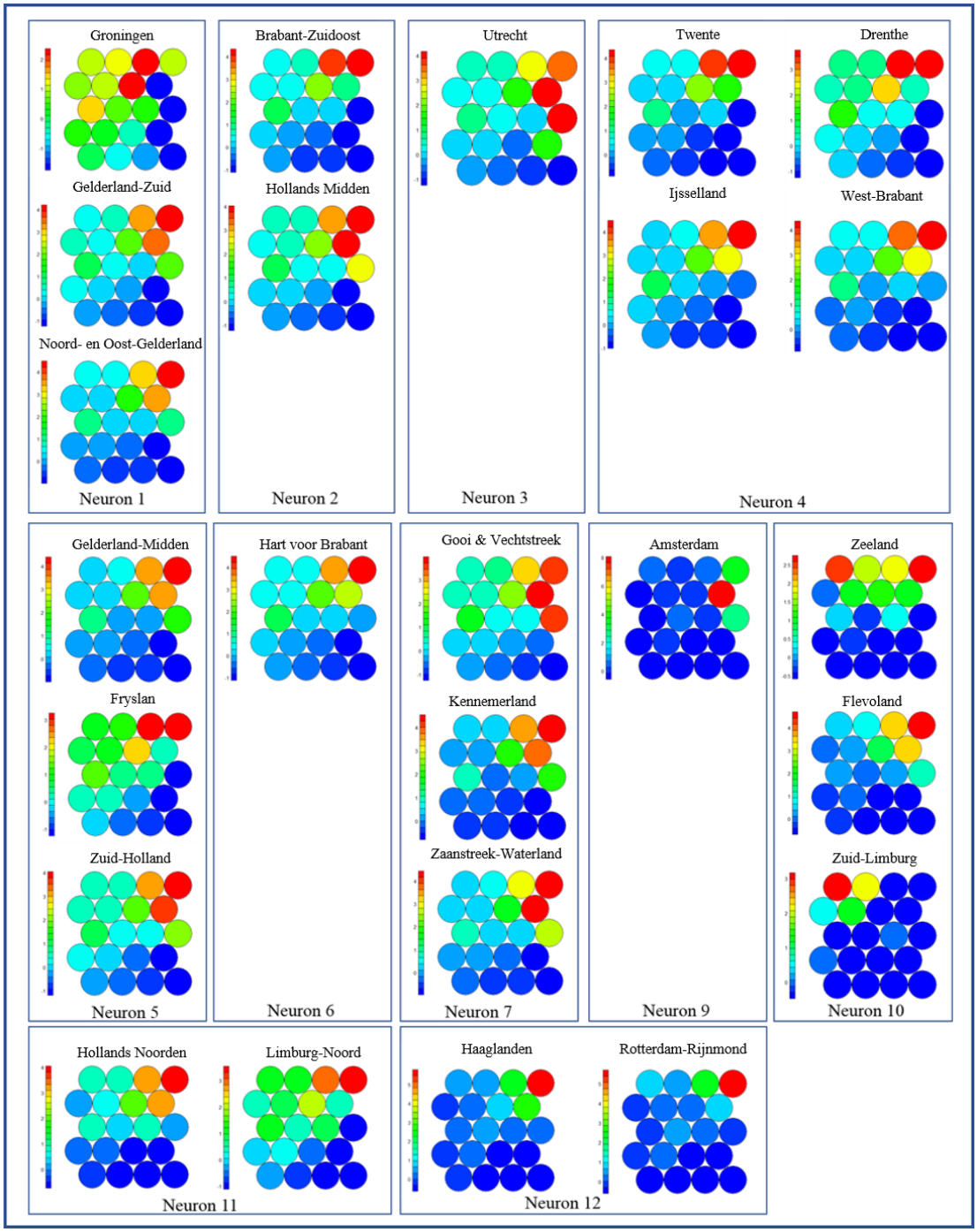


Figure 4-8: Component planes of Time × GGDs SOM.

Besides, the result was also examined by visualizing it in geographic map. As Figure 4-9 shows, GGDs are visualized using different colors based on the neuron they mapped into. Because only school commuting is considered in the model, the commuters only travel short distance (Tjalma, 2016). The assumption is that geographically adjacent GGDs are more likely to be classified in the same neuron. This pattern can be identified in GGDs which are mapped into several neurons like neuron 4 and 12 (Fig. 4-9). However, GGDs which are classified in most neurons present a typical geographical isolation. Therefore, the outcome of GGDs × Time trained SOM using 3 × 4 lattice is not reliable. Because of the limited time, the performance of GGDs × Time SOM has not been improved yet and the synchrony between GGDs are not be integrated in the early-warning model.

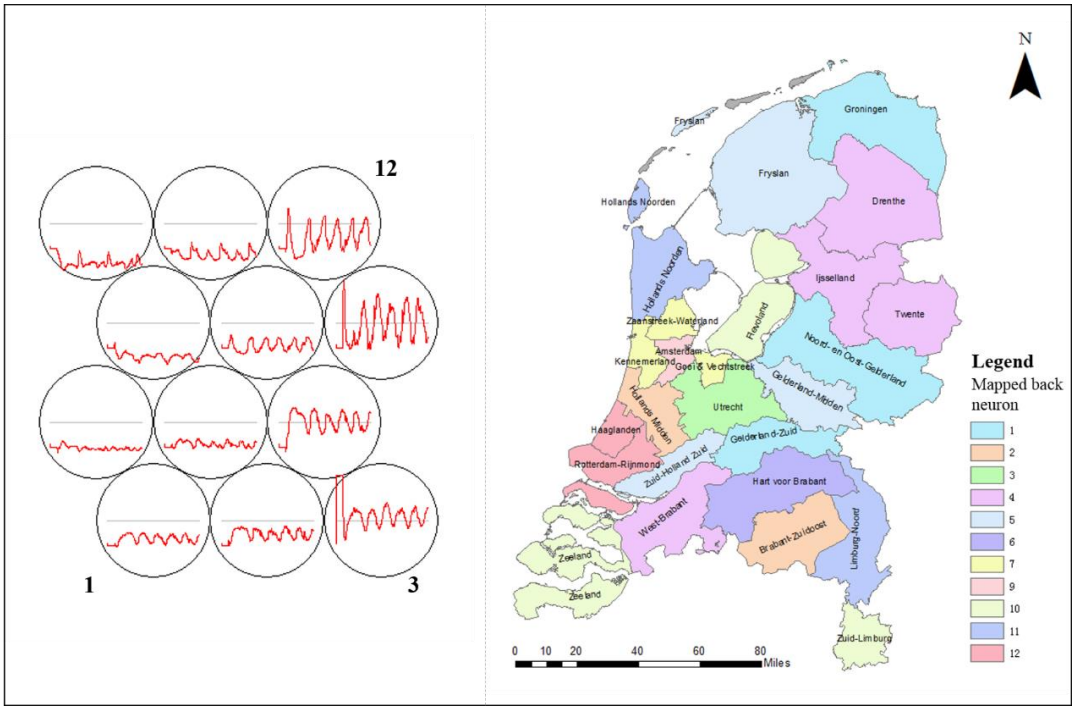


Figure 4-9: Visualization of GGDs based on their mapped back neurons. The left is the codebook vectors for GGDs × Time trained SOM.

The Time × Age data structure for 25 GGDs is applied to detect the aberration age groups. These aberration age groups will be analyzed by the intervention selection model to identify the target age group to implement interventions. The dimension of the Time × GGDs SOM lattice is tested using 3 × 4. The codebook vectors are presented in the Figure 4-10. Neuron 1 is located in the lower left corner and the neuron 12 is located in the top right corner for each GGD lattices. These codebook vectors are visualized in pie charts. The input data includes the infected population for each age groups in a certain GGD over time. The hypothesis is the age groups with higher values in the codebook vector are seemed as the aberration age groups for this neuron. If the input vector of a certain GGD is mapped onto the trained lattice when warning is given, the aberration age groups might be confirmed. What need to be mentioned is that if the neurons with codebook vector which presents puny value for all age groups comparing with other neurons in the same lattice, no aberration age group will be identified in these neurons, i.e. neuron 10 in Time × Age Amsterdam lattice, neuron 1 in Time × Age Brabant-Zuidoost lattice and so on. The codebook vectors are various for different GGD's Time × Age data. Therefore, different aberration age groups might be identified in different GGDs which lead heterogeneous behaviors of GGDs to implement interventions.

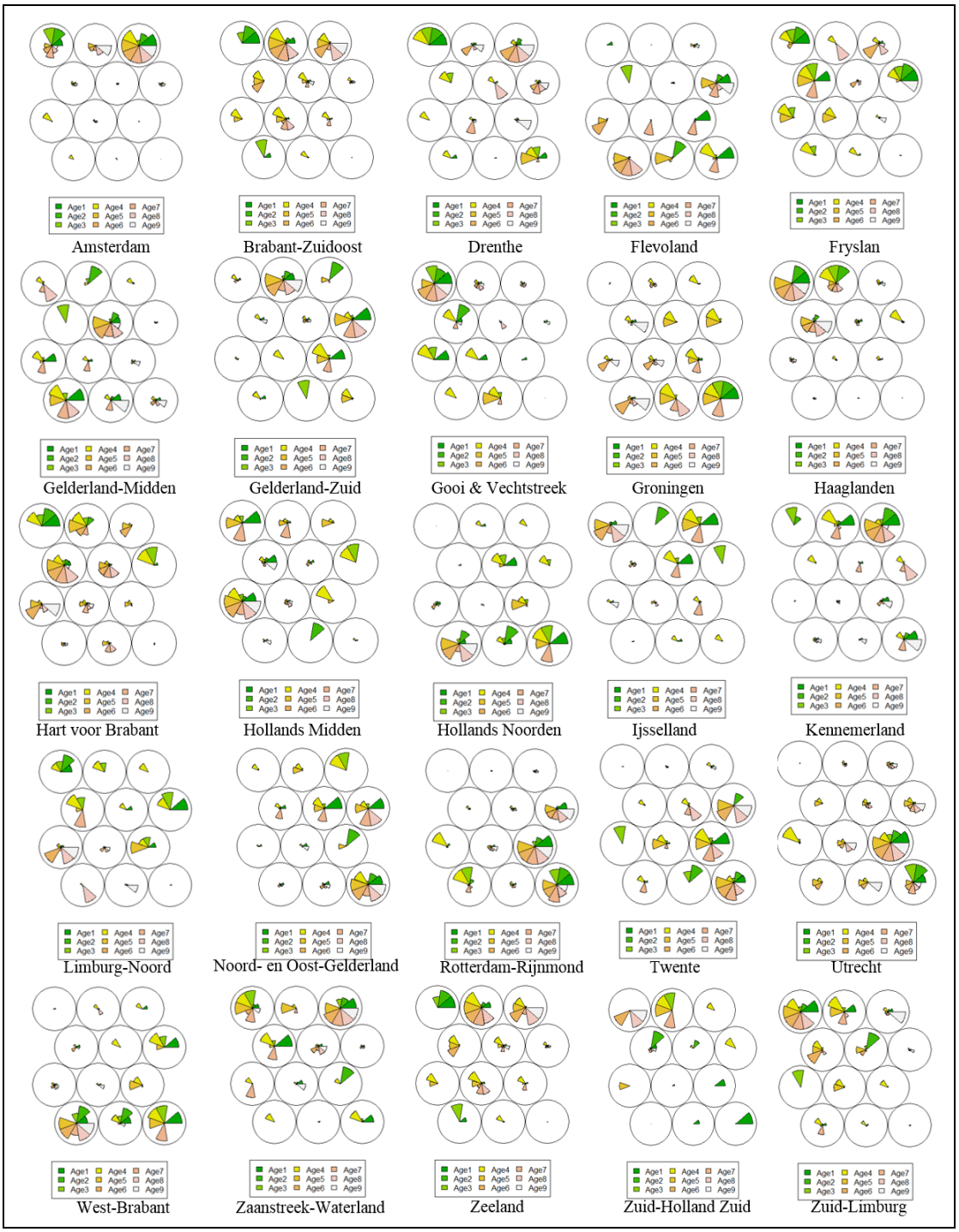


Figure 4-10: Codebook vectors for 25 GGD's Time × Age SOM

The efficacy of the trained SOM using Time × Age data structure for 25 GGDs was verified by examining whether the detected aberration age group with high infected cases, and whether various aberration age groups are identified at the same tick. The percentage stacked chart of infected cases by GGDs using the first kind of simulated data is presented in the Figure 4-11. As the chart shows, several GGDs contributes more to the total infected cases than the others. Amsterdam and Utrecht GGD region are selected as the examples to explore the performance of the trained Time × Age SOMs.

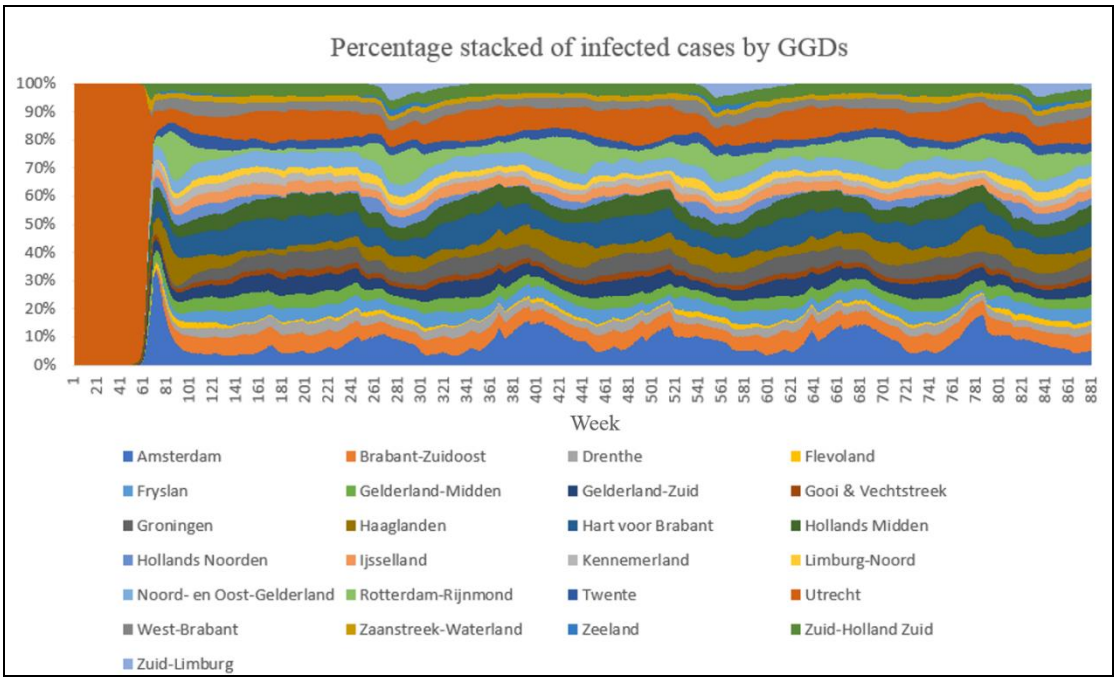


Figure 4-11: Percentage stacked chart of infected cases by GGDs using the simulated data

The six waves are classified in three groups. In this examine, the wave 1 (group 1), 2 (group 2) and 3 (group 3) were chosen to test the Time × Age SOM performance. The mapped back neurons of Time × Age Amsterdam input data and Time × Age Utrecht input data during the three disease outbreak periods are presented in the table 4-7.

Table 4-7: Mapped back neurons of Time × Age Amsterdam input data and Time × Age Utrecht input data during the three disease outbreak periods

Wave	GGD region	Mapped back neuron														
1	Amsterdam	3	6	8	4	10	12	11	9	11	9	8	6	3	6	3
	Utrecht	6	7	6	4	6	4	9	7	4	10	7	1	5	8	
2	Amsterdam	8	5	7	5	2	5	2								
	Utrecht	11	7	11	7	8	7	11	7	8						
3	Amsterdam	9	7	9	8	6										
	Utrecht	8	7	2	7	1	2									

Meanwhile, the infected cases by age groups in Amsterdam GGD region for three disease outbreak periods are presented in the left of Figure 4-12 while the corresponding detected aberration age groups are performed in the right part. As the figure shows, each age group take up similarity proportion of the infected cases when the disease outbreak. This character can also be found in the detected aberration age groups. Almost all age groups are marked as aberrant in the front part of the first disease outbreak period, the middle of the second disease outbreak period and the front part of the third disease outbreak period. Meanwhile, the detection is sensitive to the number change of total infected cases. If the number of total infected cases is less than a certain value, no aberration age group is identified.

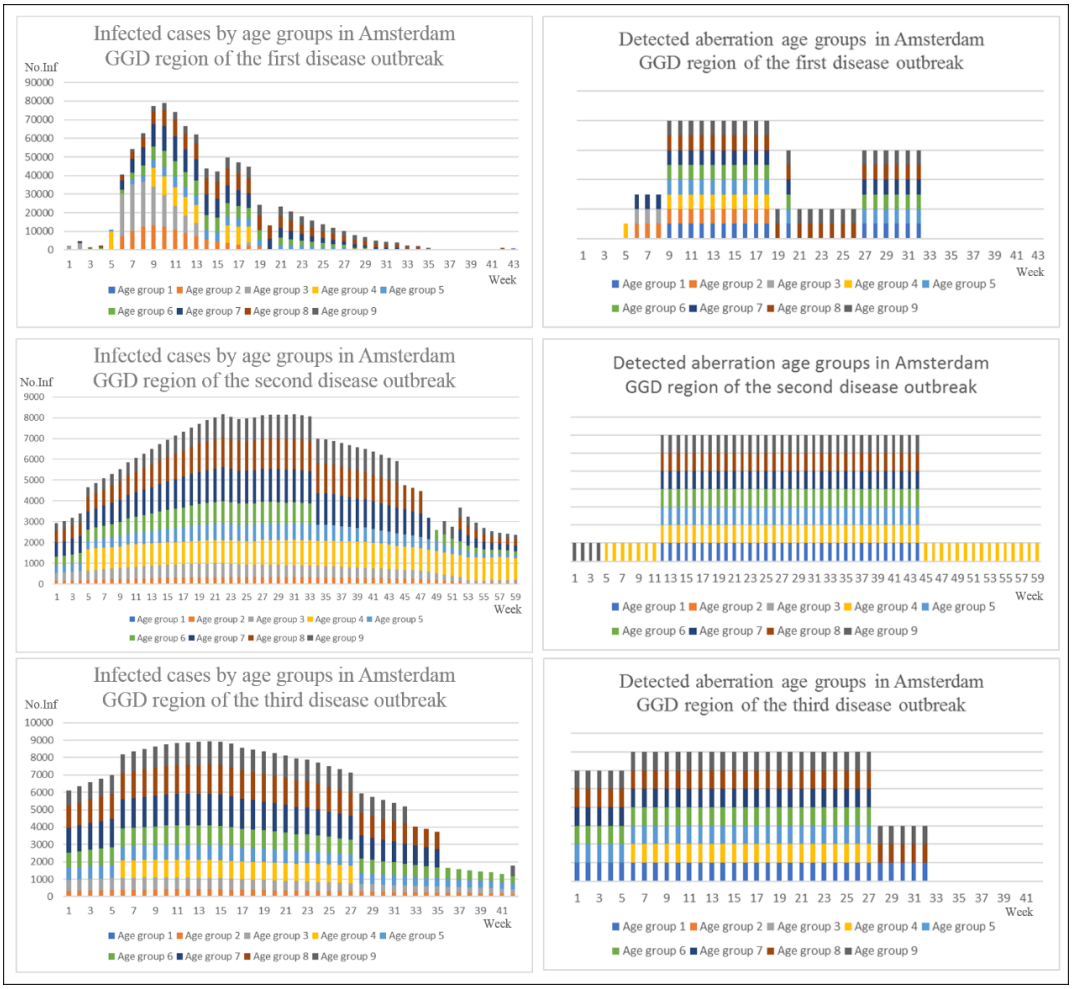


Figure 4-12: Infected cases by age groups in Amsterdam GGD region of three disease outbreaks (left part) and corresponding detected aberration age groups (right part).

The infected cases by age groups in Utrecht GGD region of three disease outbreaks present an obviously different pattern. As Figure 4-13 indicates, age group 4 takes up the highest proportion during the whole period. This feature is captured by the Time × Age SOM. Although the mapped back neurons are various (Tab. 4-7), age group 4 is always seem as aberrant. However, the method seems over sensitive to certain level. Several age groups which take up few proportion are also identified as the aberration, i.e. the front part of in the first disease outbreak.

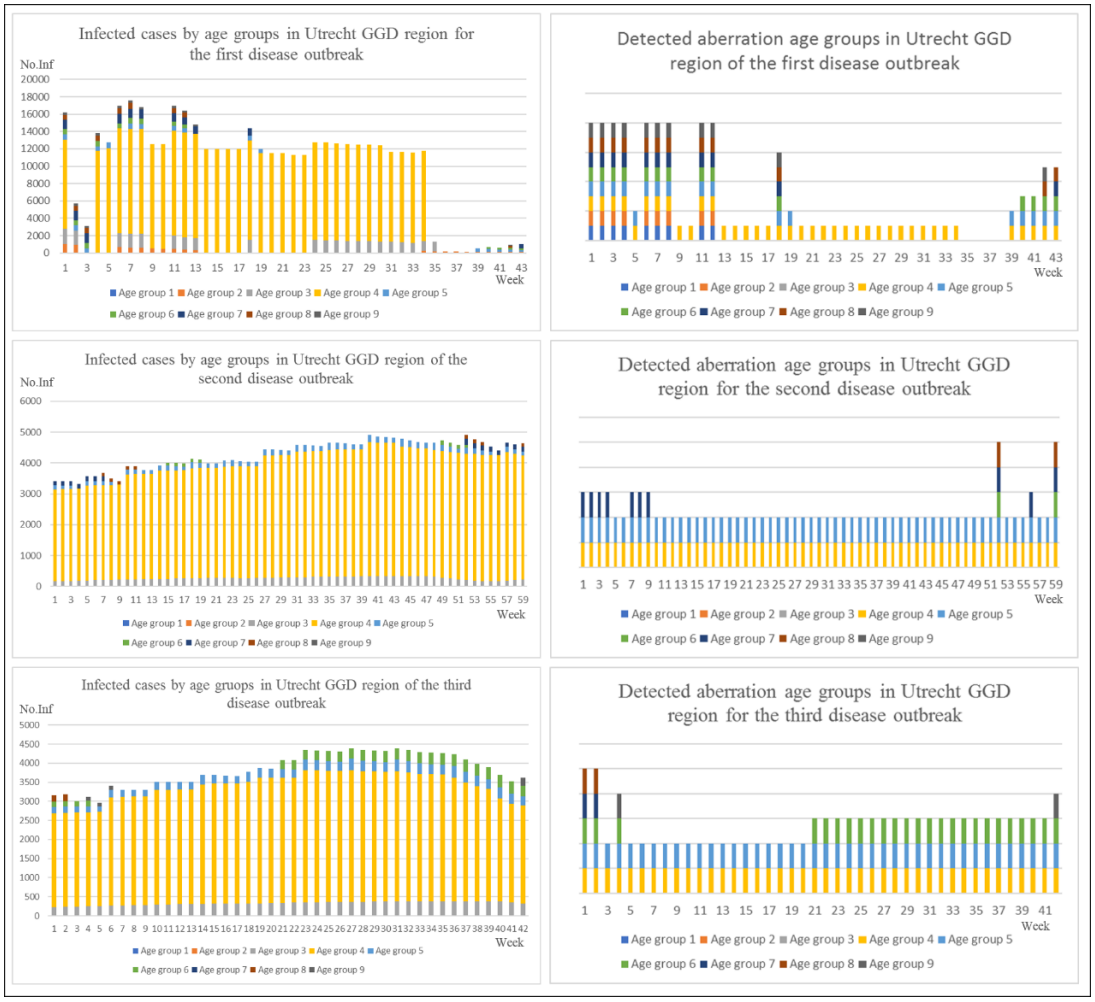


Figure 4-13: Infected cases by age groups in Utrecht GGD region of three disease outbreaks (left part) and corresponding detected aberration age groups (right part)

Generally speaking, the aberration age groups can be detected by the Time × Age trained SOM for 25 GGDs using 3 × 4 lattice. Meanwhile, the pattern infected cases by age groups is various in different GGDs.

4.3. Comparison of the Farrington method and SOMs

Both of the Farrington method and SOMs has been explored to detect the pertussis aberration change in the Netherlands using the first kind of simulated data. In this section, the efficacy of these two methods will be compared.

Similar with the comparison between the original Farrington method and the improved Farrington method, sensitivity and detection before peak value will be used as indexes to compare the prediction ability of SOMs and the (improved) Farrington method. This comparison is only applied for the national level detection as the detection for GGDs hasn't been produced using SOM method. Meanwhile, the wave 0 is not been included in the comparison because the Farrington method can't detect the aberration for the basic historical data (the first three years). As Table 4-8 indicates, the sensitivity and detection before peak value of SOMs are higher than the Farrington method. The SOMs can detect all aberrations and the detected time is earlier than the Farrington method.

Table 4-8: Overall evaluation of the improved Farrington method and SOMs national detection

Algorithm version	**Sensitivity**	**Detection before peak value (weeks)**
Farrington	0.8	18.5
SOMs	1	32.6

Besides, the run-time of SOMs is less than the Farrington method. As for the national level detection, the average run-time for SOMs is 1.41 seconds, and the mean run-time for Farrington method is 8.46 seconds. Therefore, SOMs is selected to be implemented in the early-warning model.

5. Results and Discussion

This chapter consists of two parts, validation of the intervention model to confirm that all agents' behaviors are simulated as expected; a discussion of the impact of the condition-based intervention (vaccination) method in which the outcomes of the method will be compared with the current infant vaccination and maternal vaccination in terms of the infected cases and the disease diffusion pattern.

5.1. Model validation

The intervention model is an agent-based model which includes two entities: GGDs and RIVM. The environment of the model is a disease transmission model which simulates the pertussis spread in the Netherlands constructed by Tjalma (2016). The behaviors of agents are evaluated to validate whether they are simulated as expected. Several examinations have been formulated:

1. Whether the RIVM can supervise the number of infected cases and detect the aberration before the notifications reach a peak value.
2. Whether the RIVM orders GGDs to implement interventions in their service region when an aberration is detected.
3. Whether GGDs can identify the target population groups to implement interventions based on the aberration age groups they detected in their service region.
4. Whether GGDs implement intervention to the target population groups in their service region.

5.1.1. General outcomes of the model

The real-time simulated pertussis data is generated by the disease transmission model. In this thesis, the model only concerns the school-commuting, the start total population recovered (immune) ratio is defined as 70%, and the infection is introduced in the age group 7 (35 – 50 years old) in Utrecht at first. The total simulated period is 640 weeks.

As Figure 5-1 indicates, there are five disease outbreaks. The highest number of infected cases is presented in the first outbreak reaching 190 thousand. The number of notifications decreases dramatically after the peak value in the first outbreak and the second outbreak is identified after 180 weeks. The most extended duration of the outbreak is 151 weeks which occurs in the third outbreak. When looking at the percentage stacked infections over age groups (Fig. 5-1), a high proportion of infections presents in the age group 3 (5 – 12 years old) within the first 70 weeks, followed by a high percentage of infections in the age group 4 (12 – 17 years old) during the remaining period. These outcomes are fully discussed in Section 5.2.2.

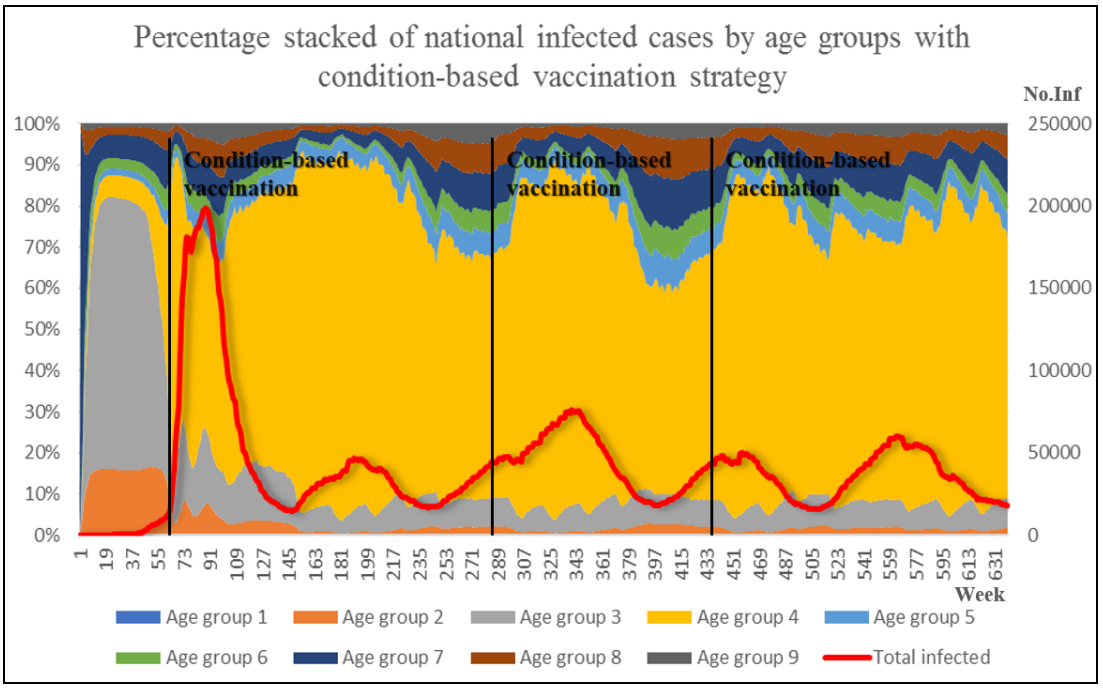


Figure 5-1: General outcomes of the model with condition-based intervention strategy. The red line indicates the total national infected cases. The graphic below is the percentage stacked of infected cases by age groups

5.1.2. Validation of aberration detection

In this section, examination 1 and 2 are tested. The national aberration detection is simulated by the early-warning model using GGDs × Time SOM. There are 5 times detected aberrations, starting at the 65th week, the 178th week, the 283rd week, the 433rd week and the 537th week separately which are corresponding to five disease outbreaks. The average detection before peak value is 28.14 weeks. These outcomes indicate a good aberration detection performance that all aberrations are detected and leave enough time for interventions to take effect. The warning state is activated once aberration is detected and will be maintained until the detected notifications becoming normal.

The expectation is there should be five times interventions can be identified at whether national or regional level. However, an unexpected result is performed. The national population, population in Amsterdam GGD region and Flevoland GGD region by SEIR are presented in the Figure 5-2. The significant increase of the recovered population (vaccinated) indicates vaccinations were implemented at that time. There are three times injections are identified at the national level. These three vaccinations correspond to the first, the third and the fourth disease outbreak. The effect of the interventions at the second and fifth disease outbreak can't be detected at the national level. Meanwhile, the intervention in Amsterdam GGD region and Flevoland GGD region performed different patterns. The Figure 5-2 (middle) presents that there are two apparently vaccinations in the first and the third disease outbreak respectively for the Amsterdam GGD region. Besides, two fluctuations of the recovered population are also detected. As Figure 5-2 (right) indicates, two times vaccinations can be found in the first and the third disease outbreak in the Flevoland GGD region. These various patterns of implemented immunization and the unexpected result can be explained in the following way.

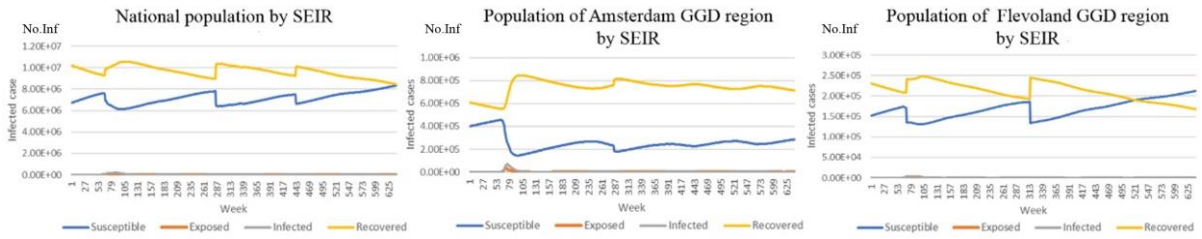


Figure 5-2: The national population (left), population in Amsterdam GGD region (middle) and Flevoland GGD region (right) by SEIR. The yellow line indicates the recovered population, the blue line indicates the susceptible people, the exposed population is expressed

The missing execution in the second and fifth disease outbreak at the national level can be explained by the fact that the period between detected aberration ticks in two adjacent outbreaks might be less than the minimum implement interval (3 years). The interval between the first and the second outbreak is about 102 weeks, and the interval between the fourth and fifth outbreak is 107 weeks. Thus, GGDs which implemented interventions in the first or the fourth disease outbreak period are prohibited to take actions during the second or fifth disease outbreak.

The first slight fluctuation of the recovered population in the third disease outbreak (the 417th week) of Amsterdam GGD region is explained by the fact that there was an intervention implemented at that time. However, the target population groups were already vaccinated in previous executions. Because of the high vaccination coverage, if the target group has been vaccinated before, the effect of the second vaccination should be limited. Therefore, there is not a significant increase in the recovered population. The existing of the second fluctuation is stranger. The assumption is that the total number of infected cases present an endemic feature in Amsterdam GGD region at that time (Fig. 5-3)). More and more infected people becoming recovered and a few outside infected commuters lead to the increase of the total recovered population.

When looking at the pattern of vaccination in Flevoland GGD region, it shows that there is one injection less comparing with the outcome at the national level or the Amsterdam GGD region. A plausible reason is the varying pertussis conditions in each GGD. Several GGDs might present the high number of pertussis infected cases while the aberration might not exist in other GGDs at the same moment. As Figure 5-3 indicates, the patterns of the infected cases in Amsterdam GGD region and Flevoland are significantly different. There are only two detectable outbreaks in the Flevoland GGD region within the first and third national disease outbreak period. This is the reason why there is only two times vaccination in Flevoland.

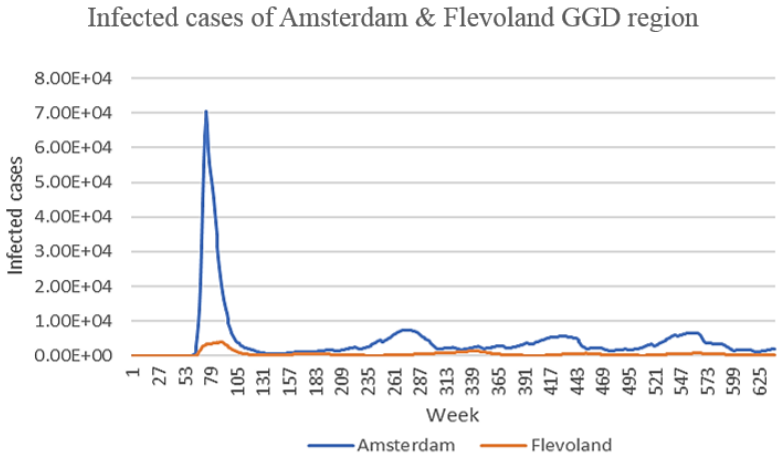


Figure 5-3: Infected cases of Amsterdam & Flevoland GGD region

From what has been discussed above, the RIVM can detect the number of infected cases and detect the aberration before the notifications reaching the peak value in each outbreak. GGDs implement interventions in their service region when they are alarmed as long as the period from last time intervention is higher than the minimum implement interval and the aberration age groups can be detected.

5.1.3. Validation of intervention selection

This section tests the examination 3 and 4. The behavior of GGDs selecting a target age group is simulated by the intervention selection model combined with the early-warning model. At first, the aberration age groups within each GGD are detected by the early-warning model using Time × Age SOM. After that, these detected age groups are analyzed in the intervention selection model to identify the target population groups. The validation of the intervention selection is limited to the Amsterdam GGD region which is chosen as an example. As discussed before, there are three vaccinations implemented by the Amsterdam GGD. The infected cases of the Amsterdam GGD region by age groups at three times vaccinations are shown in Figure 5-4. The instances of infected age group 4 are significantly higher than the other groups at the first times vaccination. The infected instances of age group 7 and 8 are slightly higher than the others at the second times vaccination. The same pattern can also be detected at the third times vaccination. The corresponding mapped back neurons using trained Time × Age SOM of these three times vaccinations are neuron 4, 7 and 5. Thus, the aberration age groups can be identified using the trained codebook vector of Amsterdam GGD's Time × Age SOM (Fig. 4-10 (Amsterdam)). As Table 5-1 indicates, the detected aberration age group(s) for the first and second disease outbreak are similar to the observed age groups which exhibit a high number of infected cases. However, the detection of the third times vaccination is not good enough. The result shows that only age group 4 is identified as the aberration, but the highest number of infections are presented in the age group 7 and 8.

Table 5-1: Detected aberration age group(s) for Amsterdam GGD region at three times vaccinations

	Mapped back neuron	Detected aberration age group(s)
1	4	4
2	7	4 7 8 9
3	5	4

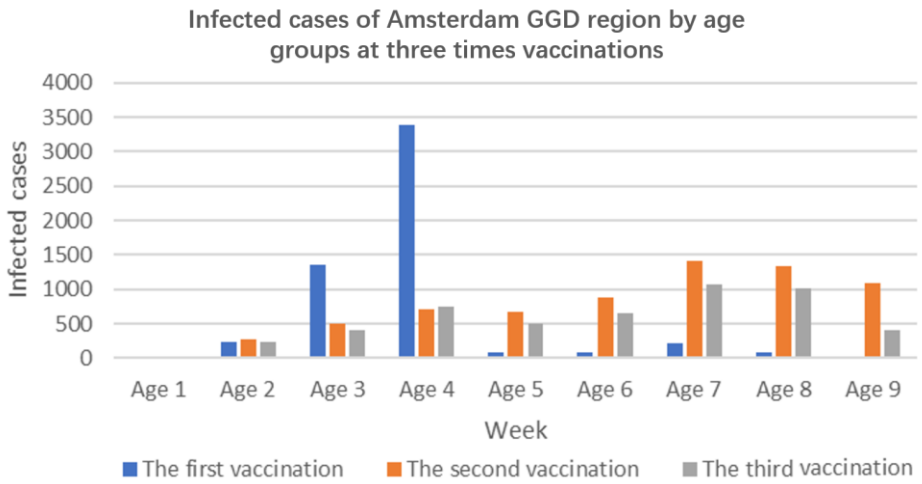


Figure 5-4: Infected cases of Amsterdam GGD region by age groups at three times vaccinations

The target population group is same as the detect aberration age group which is either 4, 7, 8 or 9. The total vaccinated population at each time intervention is 13 302, 49 122 and 1210. In summary, the target

population groups for intervention are identified and vaccinated by GGDs based on the aberration age groups they detected in their service region.

5.2. Model evaluation

In this section, the performance of the condition-based vaccination strategy will be evaluated by comparing with the current (infant) vaccination strategy and maternal vaccination strategy regarding the number of infections and disease diffusion pattern. The outcomes of the model with the current vaccination will be used as the benchmark. Maternal vaccination is one of the most popular alternative pertussis intervention strategies in the Netherlands (Chapelle, 2013) and has been explored in the research of Tjalma (2016) using the disease transmission model. To be able to make the comparison, the maternal vaccination strategy is simulated as that the pregnant women will be vaccinated when national aberration is detected. The vaccination coverage and the minimum implement interval are the same with the condition-based vaccination method.

5.2.1. Condition-based vaccination

One of the main problems of disease intervention is how to select target population groups getting the intervention (Bijkerk P, et al., 2015). Several researchers focused on the characteristics of different pertussis infected age groups. The spatial diffusion feature of epidemic inspired some researchers to detect the target population group based on their geographic information. Condition-based vaccination is proposed to identify the population that should get vaccination considering pertussis infected features of different age groups in different regions. It is assumed as a promising method to prevent pertussis more efficiently.

In this research, behaviors of the condition-based vaccination method are simulated by an intervention model which consist of two sub models, early-warning model combined and intervention selection model. The method implements re-occurring vaccination to the target population in each GGD region when a national disease outbreak is detected, and every times' vaccination percentage is same with the current vaccination ratio in the Netherlands of infants. The assumption is that there should be only once vaccination in each GGD region in each disease outbreak. Thus, the minimum vaccination interval is introduced in the model which defines that the period between two adjacent vaccinations in a specific GGD region should be more extended than a pre-defined threshold. This interval value is determined as 3 years which is same as the average disease outbreak interval in the model with current vaccination which was used to train the intervention model. The average vaccinated population per year of the model with condition-based immunization is 375 273.

5.2.2. Comparison of number infections

The influence of the vaccination is measured by the number of infections per age category during the total simulated period at first. Two kinds of graphics for three vaccination strategies are constructed. The first is the overall national infection per age group curve in percentage which is used to explore how each age groups contributes to the total number of pertussis notifications. The second is the normalized national infected cases by age groups. The number of infected cases for each age group is normalized respectively to the range 0 to 1. The normalization results in the disease pattern for each age group can be compared. The results of three vaccination scenarios are presented below.

The performance of the model with current vaccination strategy is presented in the Figure 5-5. The red line indicates the total number of infected cases.

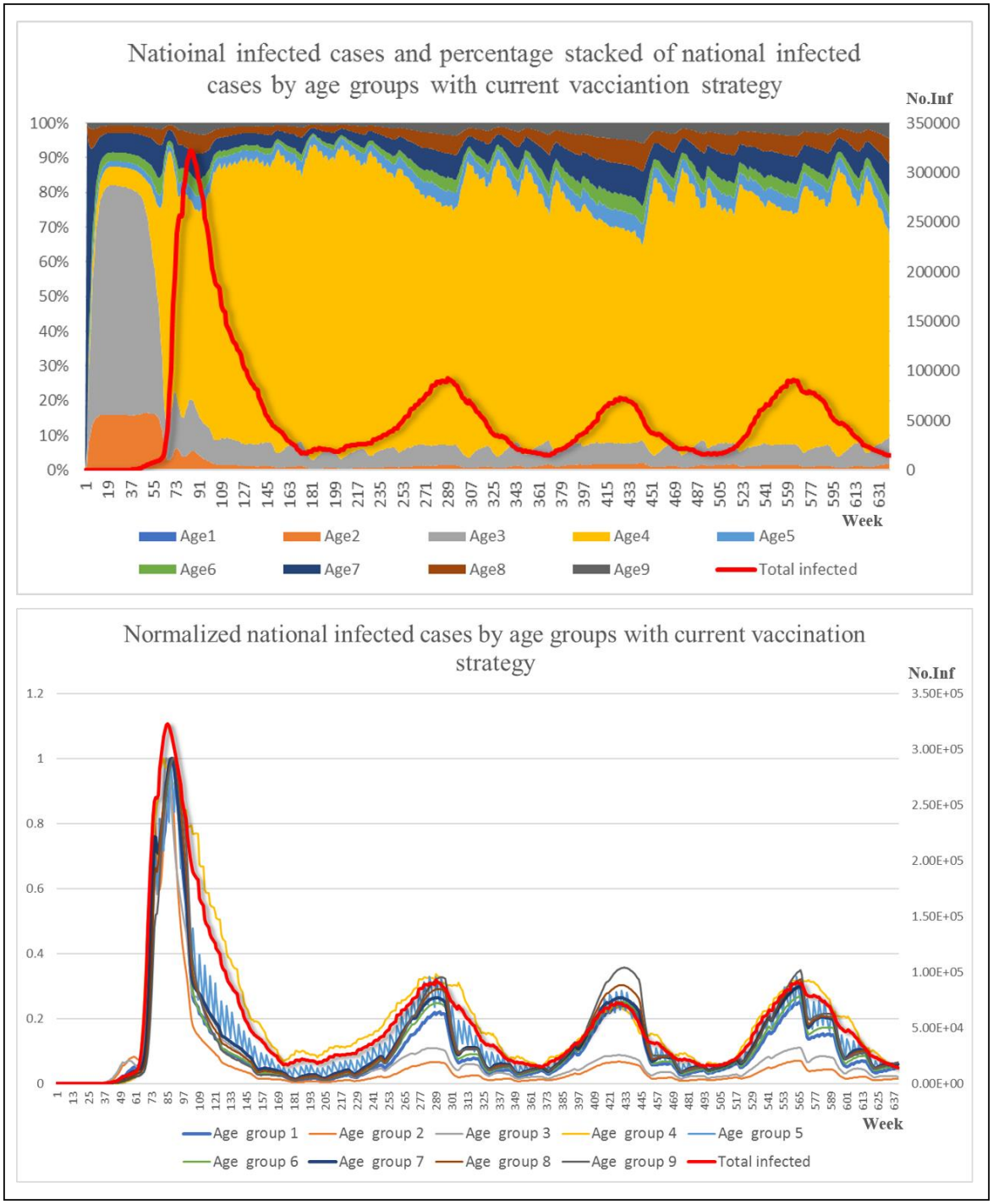


Figure 5-5: National infected cases and percentage stacked of national infected cases by age groups with current vaccination strategy (above) and the normalized national infected cases by age groups with current vaccination strategy. The red line indicates the total infected cases.

The maximum number of infections is 322 511 which is reached in the first disease outbreak (85th week). After this peak, the number of infected cases decreases until the second outbreak is coming. There are three disease outbreaks with the number of pertussis notifications ranging from 15 039 to 90 146 after the first outbreak.

When looking at the infected cases per age group, something interesting can be found. The infection is introduced by the age group 7 (35 -50 years old) at the beginning of the model which is represented as a 100 percent of national infected cases with dark blue in Figure 5-5. After that, a dramatic increase of age

group 3 (5 -12 years old) and age group 2 (5 months to 5 years old). This phenomenon was explained as that people who age from 35 to 50 might be the parents of age group 2 and 3 resulting in a high contact rate (Tjalma, 2016). Afterwards, age group 4 (12 – 17 years old) presents the highest percentage during the remaining period, which can be interpreted by age group 4 is the unique group with commuters while the model only considered school commuting. Commuting leads people who age from 12 – 17 years old are easier being infected than other age groups. It can be noticed that the percentage of age group 4 is always decreasing when the national infected cases reaching the peak value in each disease outbreak accompanying with the increasing percentage of other age groups, though age group 4 still occupies the highest percent. Meanwhile, the percentage of age group 4 constantly reaches a significantly higher value before the decreasing. The explanation is that lots of people are infected by the high number of infected commuting adolescents, and the increased number of infected people leads to the disease outbreak. The normalization graphic (Figure 5-5 (below)) can verify the explanation. All age groups present an obviously increasing trend when the disease outbreak.

The purpose of the maternal vaccination is to protect the infants younger than 5-month old who are the most vulnerable to pertussis by vaccinating to the pregnant women. The result of the model with maternal vaccination is presented in the Figure 5-6. There are three times maternal vaccinations at week 65, week 223 and week 434 which are represented as the black line in the figure. The reason for why only three times vaccination has been described in the Section 5.1.2. The target group is the pregnant women that belong to the age group 6 (25 – 35 years old). It is notable to find that the number of national infected cases (329 910) in the first disease outbreak of the maternal vaccination is even higher than the cases of the current vaccination strategy (322 511). Although vaccination is implemented at the beginning of the first disease outbreak, the effect is negligible as the target population contributes slightly to the explosion. Meanwhile, there is one more detected disease outbreak compared to the current vaccination scenario. The maternal vaccination leads to a shorter outbreak interval (mean 2.2 years) compared with the current scenario (average 3.2 years). Meanwhile, moderately less national infected cases are detected which range from 13191 to 80 129.

When looking at the infected cases per age groups, the pattern during the first disease outbreak is as the current vaccination scenario. Meanwhile, age group 4 also presents the highest percentage of disease cases starting from the 60th week. However, the ratio of the age group 4 maintains a high level when the disease outbreak which is quite different with the current vaccination scenario. Besides, the increments of other age groups' percentages are identified when the number of national infected cases is declining. These two facts might suggest the outbreak is introduced due to the enormous change of infections in age group 4. As the normalization graphic indicates (Figure 5-6 (below)), the disease pattern of the age group 4 is quite similar to the national disease pattern. Although the increments in other age groups can be detected when the disease outbreak, the enhancement is in a small extent. It is notable that the number of infected cases of age group 1 performs a dramatically increase once the disease outbreak at the national level. Three times maternal vaccination effectively decrease the number of infection in the age group 1 (the purple line in the Figure 5-6(below)). However, two times aberrant increments of the normalized infections in age group 1 after the second disease outbreak suggests that the durability of the impact of maternal vaccination is suspicious. Besides, the number of infection of age group 6 performs a significant decline after the third vaccination in the fourth disease outbreak (the green line in the Figure 5-6 (below)).

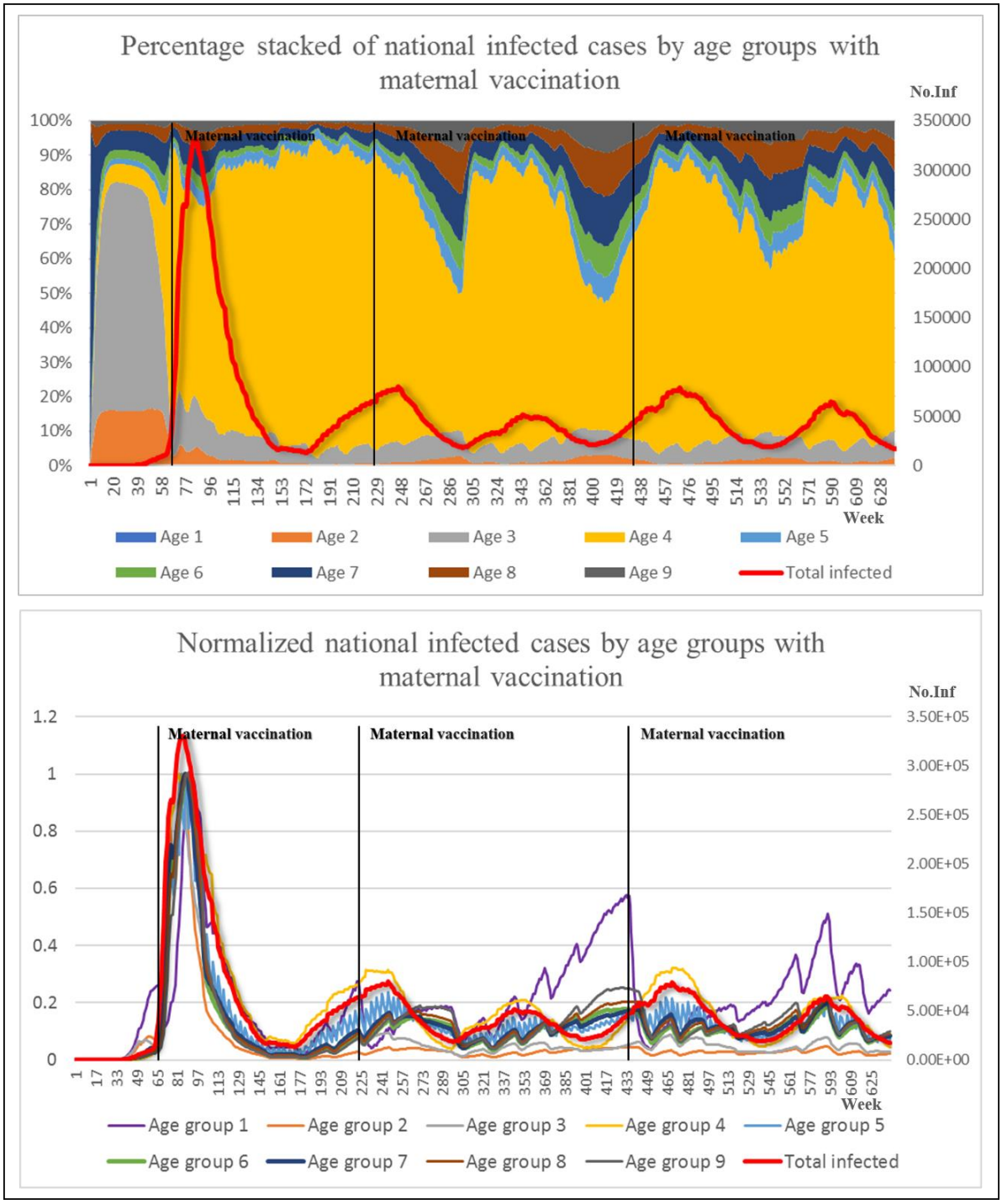


Figure 5-6: National infected cases and percentage stacked of national infected cases by age groups with maternal vaccination strategy (above) and the normalized national infected cases by age groups with maternal vaccination strategy. The red line indicates the total infected cases.

The outcomes of the model run with condition-based vaccination are presented in the Figure 5-7. There are also three times vaccinations, and the details of these vaccinations have been described in Section 5.1. The pattern of the national infected cases is entirely different with the result of other two vaccination scenarios. The maximum number of infection is 198 521 which is significantly less than the model run to current vaccination or maternal vaccination. Similar with the disease curve of maternal vaccination, there are four times moderate disease outbreak after the first fiercely outbreak and the average outbreak interval is also 2.2 years. The second and the fourth outbreak present alike short outbreak period with the number

of infections ranging from 14 623 to 49 865. The third and fifth outbreak performs a relatively extended outbreak period with the number of pertussis notifications ranging from 15 560 to 75 752.

The pattern of the percentage stacked of national infected cases by age groups with condition-based vaccination strongly resembles the figure of the maternal vaccination. However, identified fluctuations of every age groups' percent with the change of national infected cases are slighter comparing with the figure of the maternal vaccination. This feature can be explained as that people within the age groups except for age group 4 are protected more with condition-based vaccination than with the maternal vaccination. The normalization graphic verifies this explanation. It is notable to see that only age groups 4 and 5 present an increase during outbreaks at the national level except the first and fifth outbreak. The number of infections of all age groups except 4 and 5 fluctuates in a small range during the second, third and fourth outbreak. Meanwhile, the second and third vaccination results in a distinct decrease of the number of infections in all age groups except age group 4. Lots of people benefit from the condition-based immunization.

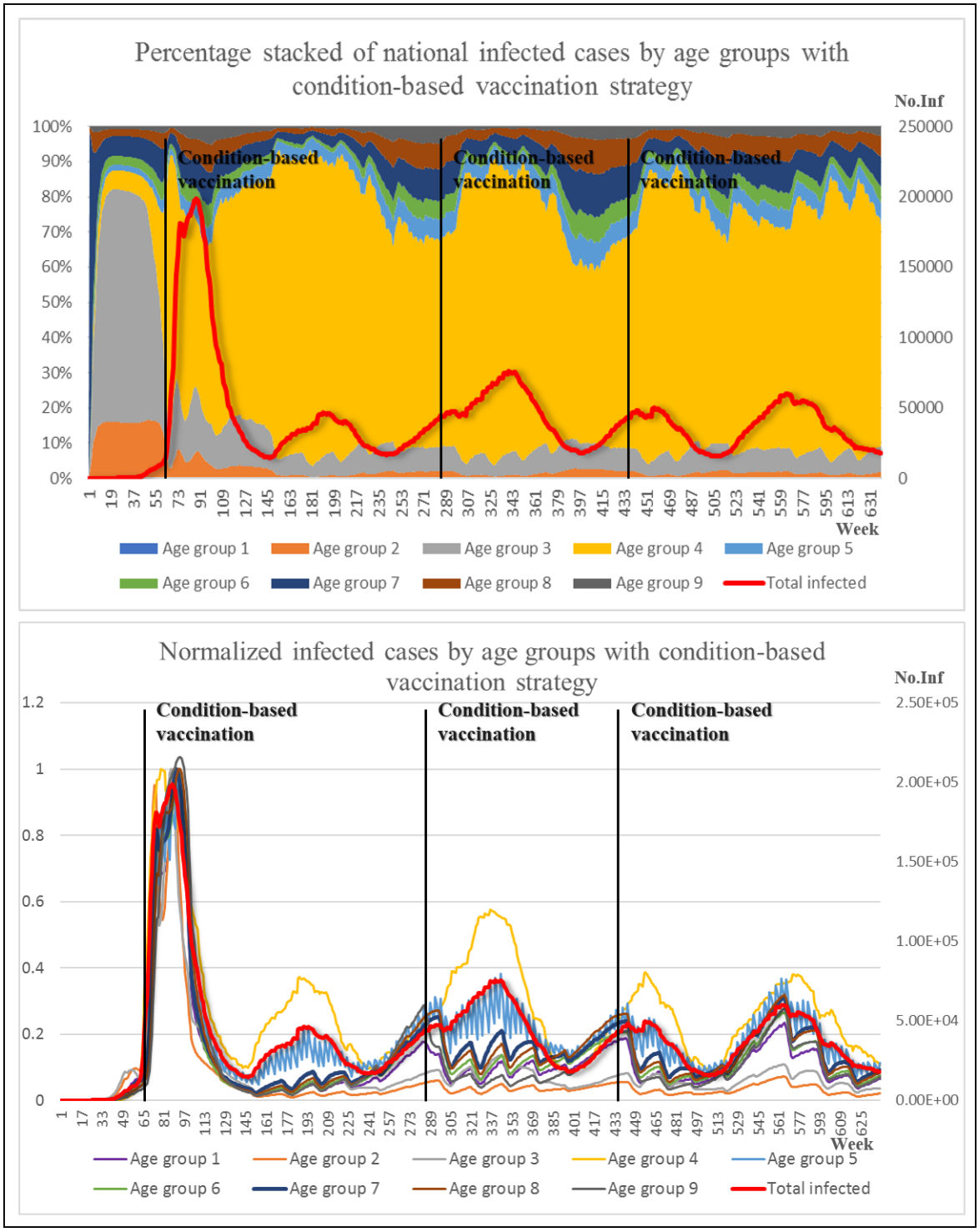


Figure 5-7: National infected cases and percentage stacked of national infected cases by age groups with condition-based vaccination strategy (above) and the normalized national infected cases by age groups with condition-based vaccination strategy. The red line indicates the total infected cases.

To overall compare the performance of these three vaccination scenarios, the term 'sum infections' is defined in this research which describes the sum value of total infected cases at each recorded tick during the whole simulated period. This term should be distinguished by the number of total infected cases over the simulated period as several the infected cases are counted repetitive in the sum infections. The total number of infections of three vaccination scenarios are presented in Table 5-2. The number of vaccinated population in the maternal vaccination and condition-based vaccination are also presented in Table 5-2.

Meanwhile, an index 'vaccination effectiveness' is introduced to simple evaluate the performance of the maternal vaccination and condition-based vaccination. The formula for the vaccination effectiveness is:

$$\mathrm{VE} = (SI_{current} - SI_{m\&c})/vaccination_{m\&c} \quad (5.2\text{-}1)$$

While the $\mathbf{VE}$ indicates the index vaccination effectiveness, SI means the sum infections, $m\&c$ suggests the maternal or condition-based vaccination.

Table 5-2: Indexes for comparison of three vaccination scenarios

Vaccination scenario	Sum infections	Difference with current sum infections	Vaccinated population	Vaccination effectiveness
Current	35 327 921			
Maternal	33 067 925	2 259 996	336 218	6.7
Condition-based	25 999 591	9 328 330	3 436 697	2.7

As the Table 5-2 indicates, the difference number of sum infections between condition-based scenario and the current is 4 times of the value of maternal vaccination. There are more than 9 million people benefit from the condition-based vaccination. However, it is notable that the vaccinated population in the condition-based scenario is even 10 times of the maternal vaccination during the simulated period. The value of the vaccination effectiveness of condition-based scenario suggests that less than 3 people benefit from the vaccination while the value for maternal scenario indicates almost 7 people benefit from the maternal vaccination. It seems that maternal vaccination is more effective than the condition-based method.

5.2.3. Effects on disease diffusion

The condition-based intervention strategy is proposed with the idea that the target population can be identified considering the various infection features of all age groups in different GGD service regions. The assumption is that the most susceptible age groups in each GGD areas can be protected by the condition-based intervention preventing the disease to spread from the original infected age group(s) to the others and the regions beside or have a high commuting ratio with the original disease outbreak area will not be infected. The effects on different age groups have been discussed in the previous section, the impact on the disease spatial diffusion is explored here.

The spatial diffusion pattern of the model run with the current vaccination and maternal vaccination is used as the control group, and the consequence of the current vaccination is seen as the benchmark (base run). Because of the various infection features of these three vaccination scenarios, to be able to make comparison, only the period starting from the week 65 when the condition-based vaccination and the maternal vaccination was first implemented in their scenario, to the week 85 when the number of pertussis notifications is almost reaching the peak value in the first disease outbreak for all scenarios.

As the Figure 5-8 (current vaccination) indicates, the disease is spread to the places adjacent to the Utrecht GGD region first. The number of infections in Amsterdam GGD region reaches 1 – 2 percent of the population which is higher than the other areas at week 70. The disease starts to spread to the further zones and reaches almost the whole Netherlands at the week 85 except the Groningen, Zeeland and Zuid-Limburg GGD areas which locate in the north, southwest and southeast corner of the country. The number of infection in the Amsterdam GGD region begins to decrease in week 80 while the number of infected cases in Haaglanden is increasing and reaches 2 – 5 percentage of the population. The number of infection in Rotterdam GGD region reaches 2 – 5 percent of population as well at week 85.

The diffusion pattern in the maternal vaccination scenario resembles the figure of the current vaccination run performed. However, the disease spread in the maternal vaccination is even faster and broader than the benchmark. The number of infection in Noord- en Oost-Gelderland GGD area, reaches 1 -2 percent of the population at week 70 which is not been detected in the base run at the same time. Meanwhile, the disease diffusion at week 80 with maternal vaccination is similar with outcome of base run presented at week 85. The number of infection in Groningen is within 1 – 2 percent which is less than 1 percent in the based run at week 85.

When looking at the condition-based vaccination scenario, the result can be called good as expected. Most regions are within 0 -1 percent of the population during the whole examination period. There are only 4 GGD regions performed increase of the number of infection at week 70. The infection is only observed in few areas located in the east and south sides of the Netherlands at week 85. Although the number of virus in Rotterdam GGD area presents a moderately increase at week 85, the increment is less than the outcome in other two scenarios. The disease retreats to the regular status (less than 1 percent of population) in Utrecht GGD region which is the original disease outbreak area at week 80.

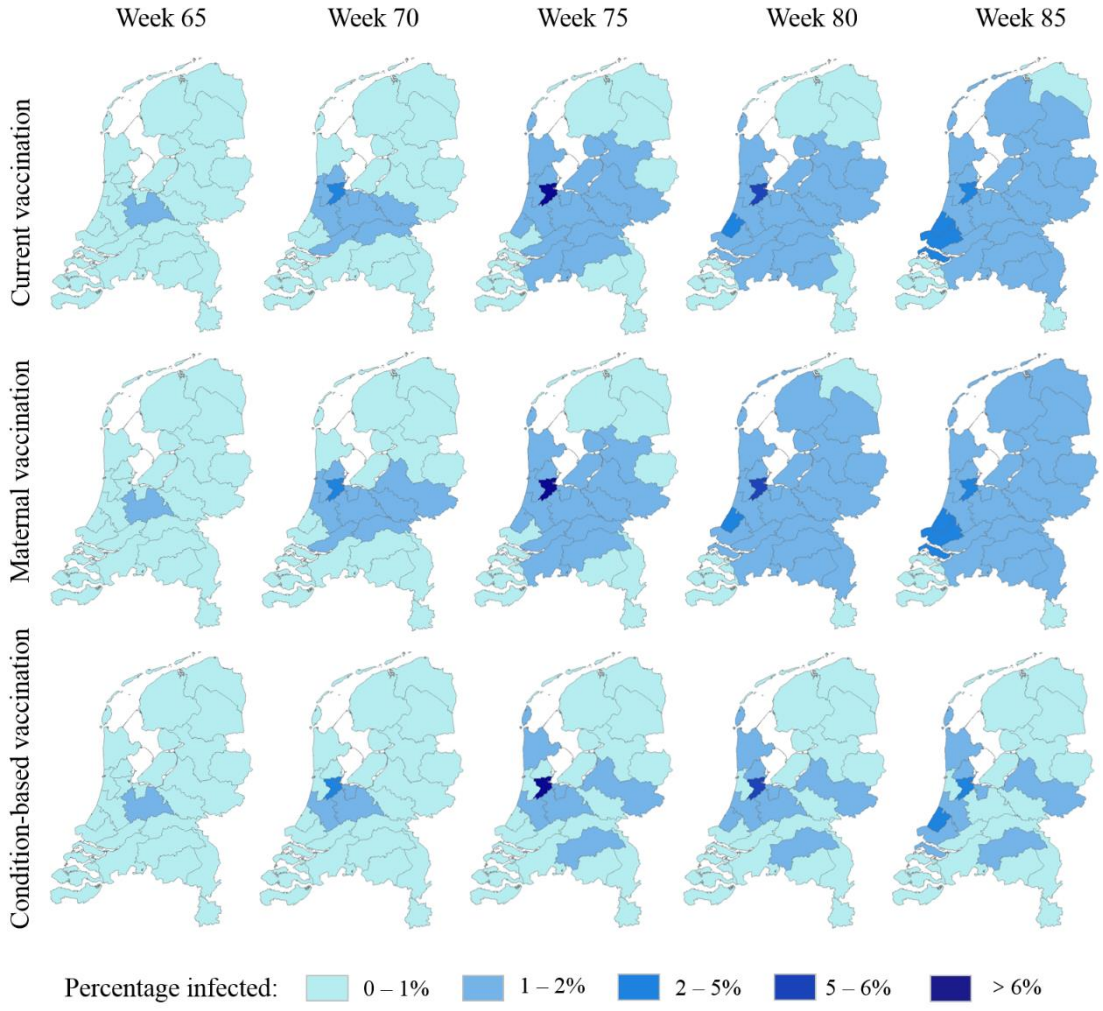


Figure 5-8: Pertussis diffusion of different vaccination strategy.

5.3. Discussion

In this section, the outcomes of the intervention model validation, and the performance of the simulated condition-based intervention strategy will be discussed. As what has been described in Chapter 4 and 5.1, the behaviors of the model performed as expected. As for the early-warning sub model, the SOMs method which is used in this research is more sensitive than the traditional aberration detection method, the Farrington algorithm (Section 4.3). All five times national disease outbreaks within the model run with either the condition-based vaccination or maternal vaccination are detected by the Time × GGDs trained SOM. Meanwhile, the average duration of detected tick before the moment that the number of infections reaches the peak value is more than half years (28.14 weeks) which leaves enough time for GGDs to implement interventions. The identified aberration age groups by the Time × Age trained SOM for Amsterdam GGD region, one of the most severe pertussis infection areas, have been validated with a higher number of infections than other age groups. Meanwhile, the heterogeneous behaviors among GGDs can be simulated according to the various disease occurrence in each GGD region.

The impact of the condition-based intervention method is simulated by the intervention model combined with the disease transmission model which has been constructed by Tjalma. The performance of the method is evaluated in terms of the number infection and effect on disease diffusion pattern. The model with the current vaccination and maternal vaccination are used as the control groups. As what has been mentioned at the beginning of the thesis, the expectation of the condition-based intervention (vaccination) method is that all population groups can benefit from the intervention and the disease diffusion will be prevented.

When looking at the result of the number of infection, the sum infection with the condition-based vaccination is obviously less than other two vaccination scenarios. More than 9 million people benefit from the method comparing with the current vaccination scenario while the value for maternal vaccination method is 2 million. Meanwhile, all age groups present a decrease after the condition-based vaccination except age group 4. However, the total vaccinated population with the condition-based method is quite higher than the maternal method and the result of the vaccination effectiveness can't be called as good. There are less than 3 people benefit from the condition-based vaccination while the number of maternal scenarios is almost 7. Two reasons can explain this. The first is the target population groups might not be the best choice for each GGD region in three times vaccination. The target population groups are determined by a rule-based method which means the algorithm might not be the most suitable for all conditions. The second reason is that the impact of vaccination should be finite. If the total number of vaccinated population is higher than a certain value, the impact of extra vaccination will make no contributes to the number of infected cases. What also needs to be mentioned is the decreasing of the number of infections accompanying with shorter disease outbreak interval. There are five times disease outbreaks in the condition-based vaccination scenario while four times is found in the current scenario. Meanwhile, the number of infection in all age groups except group 4 is decreased once the condition-based vaccination is implemented.

As for the diffusion pattern, the disease is significantly prevented by the condition-based vaccination. There are few GGDs identified with the increment of the number of infections and the disease is locked to a small extent since the disease is outbreak from Utrecht. Summary speaking, the intervention model successfully simulates the condition-based intervention which is also be assessed as a valid pertussis intervention strategy.

6. Conclusion and Recommendation

6.1. Conclusion

The primary object of this research was to construct an intervention model which can draw up intervention planning automatically for different spatial regions based on local disease occurrence to prevent pertussis. The intervention model is agent-based with two entities: GGDs and RIVM. GGDs are local health units which supervise the disease ongoing pattern in their service region and implement interventions (Damm et al., 2016). RIVM is the Netherlands national public health and environment institution which collects the disease information from all GGDs and detects disease aberrant change at the national level and orders GGDs to implement interventions in their service area. The behaviors of these two agents were simulated using two sub models.

The first is an early-warning model in which disease aberrations can be detected at both national and regional level. The national aberration determines the time when the interventions should be implemented. The objects of the regional aberration detection are different age groups which can be used to determine the target population groups will getting interventions priority. There was an attempt that using a hierarchical self-organizing map (SOM) as the aberration detection algorithm which has not been proposed before. The approach was first tested on a simulated dataset for Pertussis in the Netherlands to determine the effectiveness. To confirm the early-warning model can be constructed even if the SOM does not work as expected, the Farrington method which is a traditional and robustness aberration detection approach was also used (Hulth et al., 2010). The performance of these two approaches was evaluated by two indexes: sensitivity and detection before peak value. The results show that the SOM method can detect all 5 national disease outbreaks while the Farrington method only can observe 4 outbreaks. Meanwhile, the SOM method detected the aberrations average 32.6 weeks ahead of the disease reaching peak value and the average period for the Farrington method was 18.5 weeks. The detected aberration age groups in each GGD (health unit) service region using the SOM method were verified with higher infected population than other age groups. The SOM method was examined using the simulated data, and the result might different when using the empirical data, but it has offered a possibility that using the SOM approach to detect the disease aberrant change.

The other sub model is the intervention selection model. Because of time limitation, only vaccination was used as the intervention method in this research. The main goal of intervention selection model is to confirm two factors, who should be vaccinated and how many people will be vaccinated. The people should be vaccinated also refers to the target population group which is identified based on the detected aberration age groups in different GGD regions. The number of people getting the vaccination is measured by an increment in the percentage of the immune population in the target population group. The increased percentage which also refers to vaccination coverage in this research is same with the current Netherlands vaccination ratio for infant vaccination in each municipality. Besides, a minimum implement interval is introduced which defines that the period between two adjacent vaccinations in a specific GGD region should be longer than a pre-defined threshold. The threshold was determined according to the average disease outbreak interval in the simulated dataset for Pertussis.

A disease transmission model which simulates pertussis spread in the Netherlands was introduced as the environment of the intervention model to evaluate the performance of the generated intervention strategies. The early-warning model monitored the simulated real-time disease occurrence in each GGD region and the intervention selection model generated corresponding vaccination strategies. Validation experiments were performed to test whether the model works as expected with the real-time simulated data. The results present that all 6 disease outbreaks were detected by the model over 640 simulated weeks and the identified aberration age groups for each GGD region presented a higher number of infected cases than other age groups. However, an unexpected behavior has been observed that almost all GGDs

only implemented three times vaccinations while there are 6 disease outbreaks. The main reason is the simulated real-time disease outbreak interval shorter than the default minimum implement interval. Despite that, the performance of the intervention model can be seen as good.

The vaccination strategies generated by the intervention model are referred as the condition-based vaccination methods in this research. These condition-based vaccination methods increase the vaccination coverage of the target population groups in different GGD regions. The impact of the condition-based vaccinations was assessed regarding the number of infections and the disease diffusion patterns. Meanwhile, the outcomes of the model running with current infant vaccination method were used as the benchmark and the outcomes of the model with maternal vaccination method were introduced as an additional control group. The total number of infections of the condition-based vaccination scenario is obviously less than the number of other two vaccination scenarios. The number of infections in all age groups except age group 4 present a distinct decrease once the vaccination was applied. Meanwhile, the disease outbreak is restricted in a small area with the condition-based vaccinations. Thus, the condition-based vaccination can be seen as an effective approach to prevent pertussis, though it was only verified in a simulated environment.

6.2. Recommendation

The constructed intervention model and corresponding condition-based intervention methods in this research basically achieve the research objectives, but there still have some drawbacks and limitations of the work. Several of them can be improved or modified to achieve a more effective and reliable intervention model.

Early-warning model

a) The feature of pertussis spread has been identified as a hierarchical diffusion pattern which indicates the disease spread from one place to another in an order of sequence (Girmay, 2012). In this research, the alarm is only generated at the national level by RIVM. If GGDs can determined the time when they should implement intervention, the method will become more effective.

b) Meanwhile, the hierarchical diffusion pattern can be better interpreted by the synchrony among different locations. Though the SOM method has been used to identify the synchrony between GGD regions in this research, the result was unreliable. If the synchrony between GGD regions can be considered in the model, the generated interventions can prevent pertussis more efficient in terms of diffusion pattern.

c) The SOM method has been verified as a good aberration detection approach using the simulated data. It is interesting to examine the performance of the SOM method with empirical data.

Intervention selection model

a) A rule-based method is used in the intervention selection model. However, the rule-based approach only trains the agents to take respondsibility to various conditions. Although the performance of the model is promising, integrating machine learning methods might improve the model to become smarter than the rule-based method. Because machine learning enables the agents to come with a new strategy of how to take action (Rand, 2006). Whether GGDs determine to implement a particular intervention strategy is always a trade-off between several components: the effectiveness of the strategy, number of infected persons in a given time, the financial consequence of strategy and the result of previous practice strategies. The most suitable intervention strategy t is hard to predict. For example, the disease outbreak interval may change when vaccination is implemented which can lead to an increase in the number of disease cases.

b) In this research, only vaccination method was considered. However, other intervention methods like respiratory isolation might also contributes to prevent disease spread (Lechat, 1987).

Intervention model combined with the disease transmission model

a) In this research, only school commuting was considered in the disease transmission model as the disease pattern becomes endemic when other kinds of commuting like job commuting is introduced. But only considered school commuting is not realistic.

b) Although the outcomes of the model are stable within different vaccination scenario, the model only runs one times for each scenario due to limited time which should been run more times to confirm the result's reliability.

LIST OF REFERENCES

Anselin, L. (1995). Local Indicators of Spatial Association - Lisa. *Geographical Analysis*, *27*(2), 93–115.

Augustijn, E.-W., & Zurita-Milla, R. (2013). Self-organizing maps as an approach to exploring spatiotemporal diffusion patterns. *International Journal of Health Geographics*, *12*(60), 1–14. https://doi.org/10.1186/1476-072X-12-60

Basara, H. G., & Yuan, M. (2008). Community health assessment using self-organizing maps and geographic information systems. *International Journal of Health Geographics*, *7*, 1–8. https://doi.org/10.1186/1476-072X-7-67

C. P. Farrington, N. J. Andrews, A. D. B. and M. A. C. (1996). A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *159*(3), 547–563. https://doi.org/10.1890/11-0206.1

Chapelle, C. F. La. (2013). Maternal Vaccination Against Bordetella Pertussis: What Do Pregnant Women Want? *Gynecology & Obstetrics*, *3*(4). https://doi.org/10.4172/2161-0932.1000162

Cliff AD, Haggett P, Ord JK, V. G. (1981). Spatial Diffusion: An Historical Geography of Epidemics in an Island Community. Retrieved February 17, 2018, from https://books.google.nl/books?hl=en&lr=&id=k-88AAAAIAAJ&oi=fnd&pg=PR9&dq=Spatial+Diffusion:+An+Historical+Geography+of+Epidemics+in+an+Island+community&ots=uj314ComyO&sig=sMY4RPPisOHCvMy2pdoPNS-1YWk#v=onepage&q&f=false

Dabrera, G., Amirthalingam, G., Andrews, N., Campbell, H., Ribeiro, S., Kara, E., … Ramsay, M. (2015). A case-control study to estimate the effectiveness of maternal pertussis vaccination in protecting newborn infants in England and Wales, 2012-2013. *Clinical Infectious Diseases*, *60*(3), 333–337. https://doi.org/10.1093/cid/ciu821

Damm, O., Witte, J., Wetzka, S., Prosser, C., Braun, S., Welte, R., & Greiner, W. (2016). Epidemiology and economic burden of measles, mumps, pertussis, and varicella in Germany: a systematic review. *International Journal of Public Health*, *61*(7), 847–860. https://doi.org/10.1007/s00038-016-0842-8

de Greeff, S. C., Lugnér, A. K., van den Heuvel, D. M., Mooi, F. R., & de Melker, H. E. (2009). Economic analysis of pertussis illness in the Dutch population: Implications for current and future vaccination strategies. *Vaccine*, *27*(13), 1932–1937. https://doi.org/10.1016/j.vaccine.2009.01.106

De Melker, H. E., Schellekens, J. F. P., Neppelenbroek, S. E., Mooi, F. R., Rümke, H. C., & Conyn-Van Spaendonck, M. A. E. (2000). Reemergence of pertussis in the highly vaccinated population of the Netherlands: Observations on surveillance data. *Emerging Infectious Diseases*, *6*(4), 348–357. https://doi.org/10.3201/eid0604.000404

de Vries, R., Kretzschmar, M., Schellekens, J. F. P., Versteegh, F. G. A., Westra, T. A., Roord, J. J., & Postma, M. J. (2010). Cost-effectiveness of adolescent pertussis vaccination for the netherlands: Using an individual-based dynamic model. *PLoS ONE*, *5*(10). https://doi.org/10.1371/journal.pone.0013392

Girmay, N. H. (2012). *Performance and Scalability of Geographically-Explicit Agent-Based Disease Diffusion Models* University of Twente.

Halperin, S. A. (2005). Canadian experience with implementation of an acellular pertussis vaccine booster-dose program in adolescents: implications for the United States. *Pediatric Infectious Disease Journal*, *24*(6 Suppl), S141–S146. https://doi.org/10.1097/01.inf.0000166163.02135.8b

Hawker, J., Begg, N., Blair, I., Reintjes, R., & Weinberg, J. (2005). *Communicable Disease Control Handbook*. *Communicable Disease Control Handbook*. Blackwell Pub. Retrieved from https://books.google.nl/books?id=fOyPCll3iZAC&pg=PA361&lpg=PA361&dq=The+relationship+between+GGD+and+RIVM&source=bl&ots=NiyYJBjO_R&sig=KahRJYWFDKsvqtlEOnC_yNvCSu0&hl=en&sa=X&ved=0ahUKEwjqk4eqpK_VAhWDZVAKHTYTDDAQ6AEIXzAH#v=onepage&q&f=false

Höhle, M., & Mazick, A. (2010). Aberration Detection in R Illustrated by Danish Mortality Monitoring. *Biosurveillance: Methods and Case Studies*, *3*, 215–237. https://doi.org/10.1201/b10315-13

Hulth, A., Andrews, N., Ethelberg, S., Dreesman, J., Faensen, D., Pelt, W. Van, & Schnitzler, J. (2010). Practical usage of computer-supported outbreak detection in fi ve European countries.

Hutwagner, L., Browne, T., Seeman, G. M., & Fleischauer, A. T. (2005). Comparing Aberration Detection Methods with Simulated Data. *Emerging Infectious Diseases*, *11*(2), 314–316. https://doi.org/10.3201/eid1102.040587

Hutwagner, L., Thompson, W., Seeman, G. M., & Treadwell, T. (2003). The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *Journal of Urban Health : Bulletin of the New York Academy of Medicine*, *80*(2 Suppl 1), i89–i96. https://doi.org/10.1007/PL00022319

James D. Cherry, M. D. (2012). Epidemic Pertussis in 2012 — The Resurgence of a Vaccine-Preventable Disease. *Perspective*, *367*(1), 1–3. https://doi.org/10.1056/NEJMp1002530

Jeefoo, P., Tripathi, N. K., & Souris, M. (2011). Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao province, Thailand. *International Journal of Environmental Research and Public Health*, *8*(1), 51–74. https://doi.org/10.3390/ijerph8010051

Juha Vesanto, E. A. (2000). Clustering of Self-Organizing Map. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, *11*(May), 586–600.

Keller, L. O., Strohschein, S., Lia-hoagberg, B., Ph, D., Schaffer, M. A., & Ph, D. (1998). Population-Based Public Health Interventions : *Public Health Nursing*, *21*(5), 453–468.

Kevin M.Johnston. (2013). *Agent Analyst: Agent based modeling in ArcGIS*. California: Esri Press.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *21*(1–3), 1–6. https://doi.org/10.1016/S0925-2312(98)00030-7

Lechat, M. (1987). Public health action in emergencies caused by epidemics. *Health Policy*, *8*(1), 141. https://doi.org/10.1016/0168-8510(87)90141-2

Lehmann, B. A. (2016). Artikelen Informatiebehoefte over maternale kinkhoestvaccinatie.

Liu, Y., Weisberg, R. H., & Mooers, C. N. K. (2006). Performance evaluation of the self-organizing map for feature extraction. *Journal of Geophysical Research: Oceans*, *111*(5), 1–14. https://doi.org/10.1029/2005JC003117

Lobo, V. J. A. S. (2009). Application of Self-Organizing Maps to the Maritime Environment. *Information Fusion and Geographic Information Systems*, 19–37. https://doi.org/10.1007/978-3-642-00304-2

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162. https://doi.org/10.1057/jos.2010.3

Maëlle, S., Dirk, S., & Michael, H. (2014). Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance, *70*(10), 1–35. https://doi.org/10.18637/jss.v070.i10

Mitchell, T. M. (1997). *Machine learning*. *McGraw-Hill*. https://doi.org/10.1007/978-3-540-75488-6_2

Niessen, W. J. M., & Schellekens, J. B. J. F. P. (2008). Meldingsplicht voor kinkhoest niet effectief om ongevaccineerde kinderen te beschermen, *152*(2), 86–90.

Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N., & Charlett, A. (2013). An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, *32*(7), 1206–1222. https://doi.org/10.1002/sim.5595

P. Bijkerk, B. de Gier, D.R.E. Nijsten, J.W. Duijster, L.C. Soetens, S. J. M. H. (2015). State of Infectious Diseases in the Netherlands, 2015.

Paul, M., & Meyer, S. (2016). hhh4 : An endemic-epidemic modelling framework for infectious disease counts, (2010), 1–17.

Railsback, S. F., & Grimm, V. (2012). *Agent-based and individual-based modeling : a practical introduction*. Princeton University Press. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=tSI2DkMtoWQC&oi=fnd&pg=PP2&dq=agent+based+and+individual+based+modeling&ots=dQ8ZVw2ESK&sig=GAb8wSkssLEelssyx6BiSKg4vo4#v=onepage&q=agent based and individual based modeling&f=false

Rand, W. (2006). Machine learning meets agent-based modeling: When not to go to a bar. *The Proceedings of Agent*, 9. Retrieved from http://www.ccl.sesp.northwestern.edu/papers/agent2006rand.pdf

Richardson, A. J., Risi En, C., & Shillington, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, *59*(2–3), 223–239. https://doi.org/10.1016/j.pocean.2003.07.006

Rozenbaum, M. H., De Vries, R., Le, H. H., & Postma, M. J. (2012). Modelling the impact of extended vaccination strategies on the epidemiology of pertussis. *Epidemiology and Infection*, *140*(8), 1503–1514. https://doi.org/10.1017/S0950268811002354

Salmon, M., Schumacher, D., & Höhle, M. (2016). Monitoring Count Time Series in *R* : Aberration Detection in Public Health Surveillance. *Journal of Statistical Software*, *70*(10). https://doi.org/10.18637/jss.v070.i10

Tan, T., Trindade, E., & Skowronski, D. (2005). Epidemiology of pertussis. *The Pediatric Infectious Disease Journal*, *24*(5 Suppl), S10-8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15876918

Tjalma, S. (2016a). *An Agent-Based Model To Compare Vaccination Strategies for Pertussis in the Netherlands*. University of Twente.

Törönen, P., Kolehmainen, M., Wong, G., & Castrén, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Letters*, *451*(2), 142–146. https://doi.org/10.1016/S0014-5793(99)00524-4

van den Driessche, P. (2008). Deterministic Compartmental Models: Extensions of Basic Models (pp. 147–157). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78911-6_5

Van der Maas, N. A. T., Mooi, F. R., de Greeff, S. C., Berbers, G. A. M., Spaendonck, M. A. E. C. van, & de Melker, H. E. (2013). Pertussis in the Netherlands, is the current vaccination strategy sufficient to reduce disease burden in young infants? *Vaccine*, *31*(41), 4541–4547. https://doi.org/10.1016/j.vaccine.2013.07.060

Williams, L., Rasmussen, S., Kleczkowski, A., Maharaj, S., & Cairns, N. (2015). Protection motivation theory and social distancing behaviour in response to a simulated infectious disease epidemic. *Psychology, Health & Medicine*, *8506*(May), 1–6. https://doi.org/10.1080/13548506.2015.1028946

Wong, N. S., & Lee, S. S. (2011). The spatiotemporal diffusion of Pandemic Influenza (H1N1)2009 in Hong Kong. *Procedia Environmental Sciences*, *3*, 26–31. https://doi.org/10.1016/j.proenv.2011.02.006

Yang, W., Li, Z., Lan, Y., Wang, J., Ma, J., Jin, L., … Hu, W. (2011). A nationwide web-based automated system for outbreak early detection and rapid response in China. *Western Pacific Surveillance and Response Journal : WPSAR*, *2*(1), 10–15. https://doi.org/10.5365/WPSAR.2010.1.1.009

Yang, Y., Atkinson, P. M., & Ettema, D. (2011). Analysis of CDC social control measures using an agent-based simulation of an influenza epidemic in a city. *BMC Infectious Diseases*, *11*(1), 199. https://doi.org/10.1186/1471-2334-11-199

Zhang, H., Li, Z., Lai, S., Clements, A. C. A., Wang, L., Yin, W., … Yang, W. (2014). Evaluation of the performance of a dengue outbreak detection tool for China. *PLoS ONE*, *9*(8), 1–5. https://doi.org/10.1371/journal.pone.0106144