Spatial Temporal Dynamics in disease data

Editha Didas Temu
March, 2010

Spatial Temporal dynamics in disease data

by

Editha Didas Temu

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geo-informatics

Thesis Assessment Board

Chair:	Ms Dr.Ir. W. Bijker
External examiner:	Dr. P.F.D. Scheelbeek
Supervisor:	Ms Ir. P.W.M. Augustijn
Second supervisor:	Prof.Dr. A. Stein

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

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

Abstract

Identification of a disease patterns is an important aspect in epidemiology studies. The study of disease patterns comprises of spatial, temporal and spatial-temporal patterns of a disease. Several methods exist to uncover disease patterns. These methods can be classified into retrospective and prospective studies. In retrospective studies one can study about the history of a disease. This study can be grouped into visualization studies, clustering analysis and correlation studies to find the explanatory variables influencing the detected pattern. Prospective analysis is performed for an earliest detection of a future outbreak particularly for infectious disease. This can be performed by either simulation models such as agent based simulation models or cellular automata. The major challenge of using these models is the validation.

This research aims to find whether spatial temporal patterns that exist in empirical data can be used for validation of an agent based simulation model by using pertussis dataset. The spatial layers of interest of this research include, European countries, the Netherlands at municipality level, Twente regions at the municipality and postcode level and Enschede at the postcode level. Temporal extent was year 1993 to year 2004.

Methods used in this research are the mathematical formula which was used to define epidemic and endemic years of the European counties. At European level, the analysis of synchrony and travelling waves were performed. Week rank method was used to identify hierarchical pattern of spread of pertussis at in the Netherlands at the municipality level, Twente region at municipality and postcodes and Enschede at postcodes. Space-time scan statistic was used to detect spatial temporal clusters which was further categorised to see if there is hierarchical spread pattern that can be compared to week rank.

At the European countries, we observe synchrony and travelling waves of epidemic for some of the countries. No global patterns observed for this. Week rank method identify the hierarchical pattern In the Netherlands at the municipality level but this pattern could not observed in Twente region or Enschede city. Moreover space-time statistic detects statistical significant clusters of an epidemic and endemic years. These clusters were analysed and resulted to 7 clusters which originated in urban areas against 23 clusters which originates in rural areas during epidemic year while during endemic year 6 clusters originates in urban areas and 16 clusters originated in rural areas.

This findings fails to validate the agent based simulation models since for validation to take place multiple patterns should be observed in different hierarchical level. The observed pattern should be used as a guidance to validate agent based simulation models. With this patterns in the simulated data should match with patterns in the empirical data. Further research is recommended to use different methods to identify patterns in a disease data.

Key words: Synchrony, travelling waves, week rank, space-time scan statistic

Acknowledgements

I would like to take this opportunity to express my appreciation to all people who assisted me to Accomplish this research.

First I thank to the Almighty God and my guide who gave me the strength and knowledge to execute this work and who makes me to reach this far. Second I would like to give thanks to my late father Didas Alphonce Temu who assists me until I am here now. May God rest his Soul in peace Amen.

I would like to give also my sincere gratitude to ITC who offered me the sponsorship and the opportunity to broaden my knowledge.

I would like to forward my deep gratitude to my first supervisor, Ms Ir. P.W.M. Augustijn for her support in advice, material wise and her valuable comments to make sure that the final output should be in a good quality. She used most of her time to read my report and made comments that helped me To have a readable thesis in a consistent manner. I would like to say thank you very much and I appreciated for what you did, God may bless you. In other hand, I would like to thank you my second Supervisor, Prof.Dr. A. Stein for his suggestion and comments. My thesis will never reach this far without his comments. God may bless you.

Further more, I appreciate for all my classmates of Geoinformatics for the support and good time we had during our stay and studies. Also I thank you the Tanzanian community at ITC who enabled me to feel like I was at home.
Last but not least, my special thanks go to my mother Yovitha P. Temu and my sisters Jennifer Didas Temu, Rose Didas Temu, Winifrida Shayo, Diana Didas Temu and Joyce Didas Temu for their prayers and spiritual support during my hard time here in ITC. God may bless you all.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Motivation and problem statement

Identification of disease patterns has been of great importance in the study of infectious disease. Epidemiology can be defined as the science concerned with the study of disease behaviour in both space and time and their causes in a human population. Study of disease patterns will guide to identify high risk areas and seasons as well as the prediction of the coming outbreak so as to assist in the control and preventive measures of the disease at an earliest possible time[1]. Disease patterns can be studied differently depending on the type of the disease for instance many infectious disease such as influenza, measles, cholera, tuberculosis and pertussis have various characteristics in common so they can be grouped together [2]. Disease patterns can be classified as spatial patterns, temporal patterns and spatial-temporal patterns.

Spatial disease patterns identify physical location at which a disease is to be observed. During the investigation of spatial disease patterns of an infectious disease on the appropriate scale of analysis is required so as to identify the existing pattern properly[3]. The identification of spatial patterns may change as the spatial scale of analysis changes due to the phenomenon called Modifiable Areal Unit Problem (MAUP).To minimize the effect of MAUP it is required to investigate spatial patterns of a disease in more than one areal unit. The appropriate pattern will be observed in the areal unit where MAUP effect is regarded as minimal.

Temporal patterns are important in identifying seasonal influences as well as periodic patterns of a disease outbreak [4]. In order to get a clear idea of seasonality of transmission and periodic of a disease it is necessary to compare more than one temporal oscillations to see if they oscillate simultaneous, with rising and falling in the same time period [5]. For instance, the periodic pattern before vaccination can be different from the pattern after vaccination.

A spatiotemporal pattern is important for generating disease spread hypothesis which may in turn inform the disease control and prevention strategies [6]. In the study of spatial temporal patterns we can consider epidemic and endemic year of and infectious disease. Epidemic periods can be defined as phases when new cases of a certain disease, in a given human population, and during a given period, substantially exceed what is expected and endemic phases can be defined as time periods in which the disease occurs at a constant but relatively low rate in the population.

Furthermore, study of disease spread includes disease rates in the neighbour's areas since geographically close areas tends to have similar rates of disease than distant areas [3].

Several existing methods have been carried out to study infectious diseases with the use of Geographic Information System (GIS). These methods are categorised into retrospective studies and predictive studies. Retrospective studies look back at an outbreak that already took place to

understand their spread mechanism. Moreover, the retrospective studies can be grouped into visualization studies, cluster analysis (to find unusual high number of occurrences), wavelet analysis to find the repeated periodic waves. Predictive studies are for the detection of future outbreaks of the disease whereby several statistical methods such as spatial scan statistics.

Furthermore, different simulation models such as agent based simulation models and cellular automata have been developed to assist in predictive studies. A major challenge of using these Agent based simulation models is the validation of these models. In ecology they used the strategy called pattern oriented modelling (POM) to validate these simulations. A key idea of this POM is to use patterns observed in a real data. These patterns have to be identified at different hierarchical level. Then based on the observed pattern in areal system, a model structure of an agent based simulation is designed [7].

This research aims to find whether spatial temporal patterns that exist in a disease data can also be used in the validation of an agent-based pertussis simulation model

(Figure 1-1).

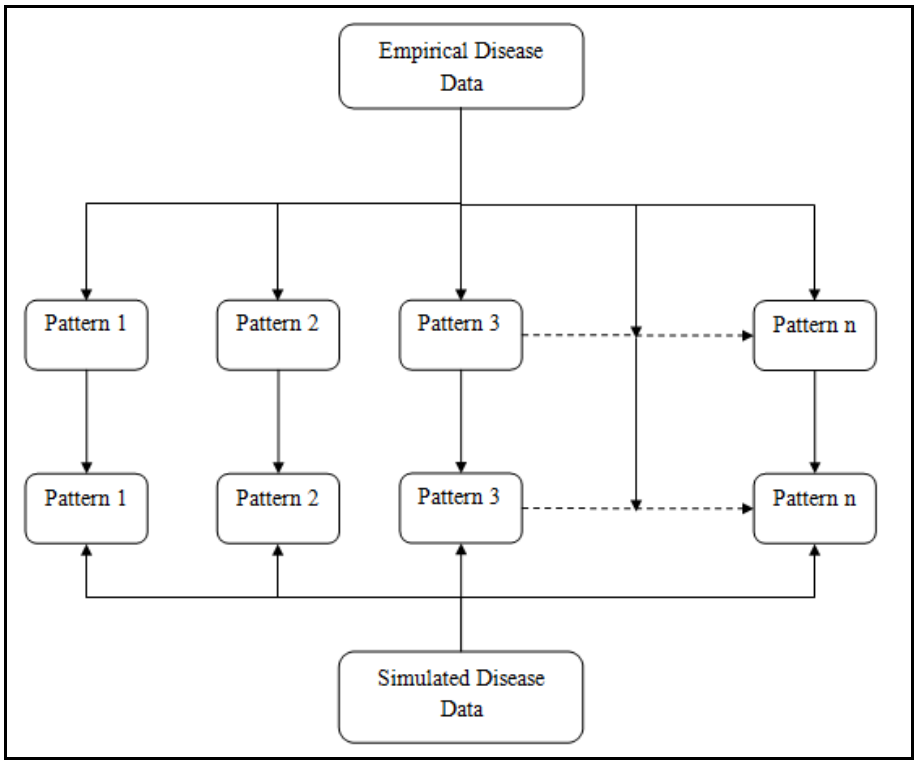


Figure 1-1 Patterns in empirical data and patterns in simulated data

1.2. Research objective

The main objective of this research is to find spatial-temporal patterns by comparing endemic and epidemic of disease data.

1.2.1. Specific objectives

1. To identify hierarchical spread patterns of a disease data
2. To identify synchrony and travelling waves of a disease data
3. To detect clusters of a disease data
4. To use the detected patterns in the validation of an agent based simulation model

1.3. Research Questions

1. Can we observe hierarchical spread patterns of a disease data?
2. Can we identify synchrony and travelling waves of a disease data?
3. Can clusters be identified according to their spatial temporal characteristics and can this lead to identification of different patterns of an epidemic and endemic year?
4. Can the detected patterns be used to validate agent based simulation model?

1.4. Overview of the remaining chapters

Chapters in this research are arranged as follows;

Chapter 2- Literature review

This chapter discuss the concept of pattern oriented modelling and disease patterns. In addition ,methods for identifying spatial temporal patterns art different scales are also explained.

Chapter 3- Data description and Study area

In this chapter the data used are described. Source and the origin of the data will also be described. The chapter will also give the quality of the dataset used and criteria used for selecting appropriate data suitable in this research.

Chapter 4- Research methods

Methods used to identify disease patterns are described .These methods includes week rank which used to determine the hierarchical spread patterns and Satscan which used to detect clusters in a disease data. The Chapter will also provide the description of method used to define endemic and epidemic years for European countries.

Chapter 5- Results of the analysis

The results obtained via the methods will be examined in this chapter

Chapter 6- Discussion conclusion and recommendations

The results obtained via the method will be discussed. In addition the conclusion and recommendation for further research in areas which needs further research will be provided.

2. Literature review

2.1. Introduction

This chapter gives a general overview of patterns oriented modelling. It gives description of disease patterns including spatial, temporal and Spatial-temporal patterns and methods that exist to identify those patterns. Moreover the chapter gives an overview of pertussis and patterns that exists for pertussis. The Chapter will end by the summary of the chapter.

2.2. Pattern oriented modelling

Pattern-Oriented Modeling (POM) is an approach to bottom-up complex system analysis which was developed in ecology for agent-based complex system models. POM explicitly follows the science research strategies which must begin with the explanations of patterns observed in real system. A pattern refers to characteristics, clearly identified structure of nature itself or data extracted from nature. Patterns contain hidden information about the process generating them of which purpose of POM is to uncover this information[8].
In pattern oriented modeling, the observed pattern can be used in three aspects: to optimize the model structure, test theory and reduce the parameter uncertainty of an agent based modeling [7, 9].

2.2.1. Optimize the model structure

Patterns in a designed model must be able to capture important features of real systems at different spatial and temporal scales and different hierarchical levels[9]. If the model is too simple it will not be able to capture all important feature of a real system. Again if the model is too complex it will contain many degrees of freedom to fit the empirical data. Therefore, a designed model should be structured in a way that it balances the fitness and the complexity of the model.

The main objective of pattern oriented modeling is to use multiple patterns observed in real system which guides in the design of the model structure[10]. Focusing on multiple patterns on the model design, the patterns observed at different spatial and temporal scales and different hierarchical level are needed at the same time. This is to produce the simulation results that are close to all patterns from the real system. To do so, one can include in the designed model structure, characteristic of the observed pattern and its dynamics together with the variables and the process causing the pattern to emerge.

The designed models are more likely to be structurally realistic model if there is larger number of different patterns at different scales and hierarchical levels that are matched [7-9]. Additionally, structurally realistic model that matches patterns at different scales of the real system facilitate a direct comparison of the model in relation with observation. It should be complex as is required to reproduce the patterns and to fulfill the objectives.

2.2.2. Test theories

After the designing of the model structure, different alternative theories of the individual agent's behavior can be used in testing how well the system reproduces the characteristic patterns. This approach of using alternative theories is referred to as "strong inferences" [10-11]. These alternative theories are then implemented in agent-based system model to test which alternative best reproduces the variety of patterns that characterizes the real system. The theories that reproduce patterns that are closer to the observations are maintained and the theories that reproduce poor predictions are rejected.

2.2.3. Reducing parameter Uncertainty

To define how well the predicted patterns match with the patterns in the real system, one can set the uncertainty boundaries of the model parameters. All patterns observed in the model need to be met for the parameter setting to be accepted. If the parameter setting fails to produce the pattern at that range of acceptance level then the designed model have to be improved. Additionally, if the designed model is within, the uncertainty boundary is more likely to be structurally realistic as well as making the interaction parameters to interact in a way similar to interactions in the real system.

2.3. Disease patterns

Detecting patterns in disease data is the most important aspect in epidemiology studies. In disease studies, patterns identifies disease behaviour in space, time or both space and time in a defined human population [5]. Disease patterns can be observed differently depending on the type of a disease to be studied. For instance, infectious disease such as pertussis, influenza and measles have various characteristics in common like; the spread of disease is contributed by personal contact, treated by vaccination and they are endemic disease with epidemic peaks which occur at a certain period of time hence they can have similar patterns. Study of disease patterns can be categorised into spatial patterns, temporal patterns and spatiotemporal patterns.

2.3.1. Spatial Patterns

This study of spatial patterns of a disease is of a particular interest in epidemiology since it identifies high risk areas. When performing the analysis of spatial disease patterns, one must consider the appropriate spatial scale (areal unit) to use in order to identify the existing spatial patterns and the best disease measure [3].The identification of spatial pattern of a disease may change as the spatial scale of analysis changes due to the phenomenon called Modifiable Areal Unit Problem (MAUP).To minimize the effect of MAUP it is required to investigate spatial patterns of a disease in more than one areal unit.

The smallest areal unit are spatially accurate but they are heterogeneous and the largest areal unit have less accuracy and more homogenous. Spatial distribution arising from areal unit of low population have higher variances and are therefore more unstable compared to areas with higher population [3]. Spatial patterns can be described in terms of clusters where unusual high number of incidence cases than expected is detected. These clusters may be observed at a variety of scales or spatial units

ranging from world wide, regional, local to household scales [12-13]. Spatial unit of analysis should match the scale at which the clustering occurs so as to have a clear overview for high risk areas to enable further investigation. Appropriate detection of disease clusters may contribute to better understanding of disease spreading mechanisms.

2.3.2. Temporal Patterns

The study of temporal patterns is important in identifying seasonal environments as well as periodic patterns of disease outbreaks [4]. In order to get clear idea of seasonality of transmission and periodic of a disease it is necessary to compare more than one temporal oscillations to see if they oscillate simultaneous, with rising and falling in the same time period [5]. Clusters may also occur in time with which high rates and low rates of a disease in a particular location can be detected. Moreover, in relation to best temporal scale, appropriate timing (the temporal window) for the identification of disease rates must be defined as accurate as possible. Temporal clusters may be defined within a range of years, months, weekly up to single day.

For an infectious disease improvement in detection timeliness for even one day might be of particular significance in public health to react quickly at the beginning of an outbreak [14-16]. However, if the temporal window is too small it can reduce the power of rapid detection of a disease and if the temporal window is very large it might result to late detection of an outbreak whereby public health officials will not benefit from these [17]. Therefore, the temporal scale of analysis must be appropriately selected to enable early detection of a disease.

2.3.3. Spatiotemporal patterns

Spatiotemporal patterns of a disease data in an epidemiology, describes study of the spread of diseases in both space and time [18]. This can lead to knowledge of when and where outbreak occurs so as to have better understanding of the underlying causes of a disease and future prediction of an outbreak.

Spatial temporal patterns of a disease can be studied in terms of disease diffusion, disease persistence and disease fadeout. Disease spreading or diffusion occurs when a disease is transmitted to a new location leaving out its origin or source at a given period of time [19]. Example of a disease spreading pattern is illustrated in Figure 2-1. In the figure it showed different direction of spread of a disease as the time changes. This is a kind of hierarchical spread of a disease where by the disease is spreading to the new location leaving out its origin.

Disease diffusion patterns of an outbreak must considers geographical distance for a disease to spread. In relation to this, it has been reported that disease spread is much related to nearest neighbour areas due to the fact that geographically close areas tend to have similar patterns of disease spread than distant areas [3]. This pattern of neighbourhood can also be explained by the travelling waves of a disease. Travelling waves can be explained by the movement of a disease especially if a disease is infectious from one location to another. This waves might change due to the effect of spatial heterogeneity in population density or population size and temporal changes in parameters (e.g. the introduction of vaccination against a disease), in disease dynamics [4, 20-22]. The study of travelling waves is of particular important for understanding mechanism of spread of a disease particularly

infectious one. In most of the cases the travelling waves might not be detected due to lack of spatial-temporal data at an appropriate scale level. These travelling waves may also spread in terms of hierarchies whereby the waves move from the area of high population to areas where population is low. This hierarchical spread of a disease has proved to occur in most of the infectious diseases.

Different methods exist to identify travelling waves and hierarchical spread of a disease. For instance the hierarchical spread and travelling waves pattern of measles has been studied by B. T. Grenfell, et al [4] by using wavelet analysis. From this study they concluded that the disease spreads from large cities to small cities. The study hierarchical patterns of pertussis was also studied by Hélène Broutin, et al., [20] by using week rank method. They also reported that the disease spread starts in urban before it moved to the surrounding rural areas.

Furthermore, disease persistent occurs when there is a minimal population size below which a disease can not maintain itself without external input [20, 23-25]. The disease persistence may have large increase in number of cases and longer period of time but little increase in the infected area. Disease fadeout occurs when there is a gradual decline of a disease until it dies out in a given location at a given period of time.

The spatial temporal synchrony can also be explained by the identification of spatial synchrony of a disease. For an infectious disease synchrony can be identified if epidemic peaks of a disease in all populations occur simultaneous in space. It is easier to control a disease if an outbreak is synchronous in space (e.g. if an outbreak peaks in all population simultaneously) than when an outbreak is desynchronized in space. If an outbreak is desynchronized in space, disease fadeout in one location might be only for a short period of time since the disease outbreak can be reintroduced again through new cases from the neighbouring locations [5]. For this reason, detection of synchrony in disease spread is vital in epidemiology for controlling and to access the whole picture of a disease dynamics.

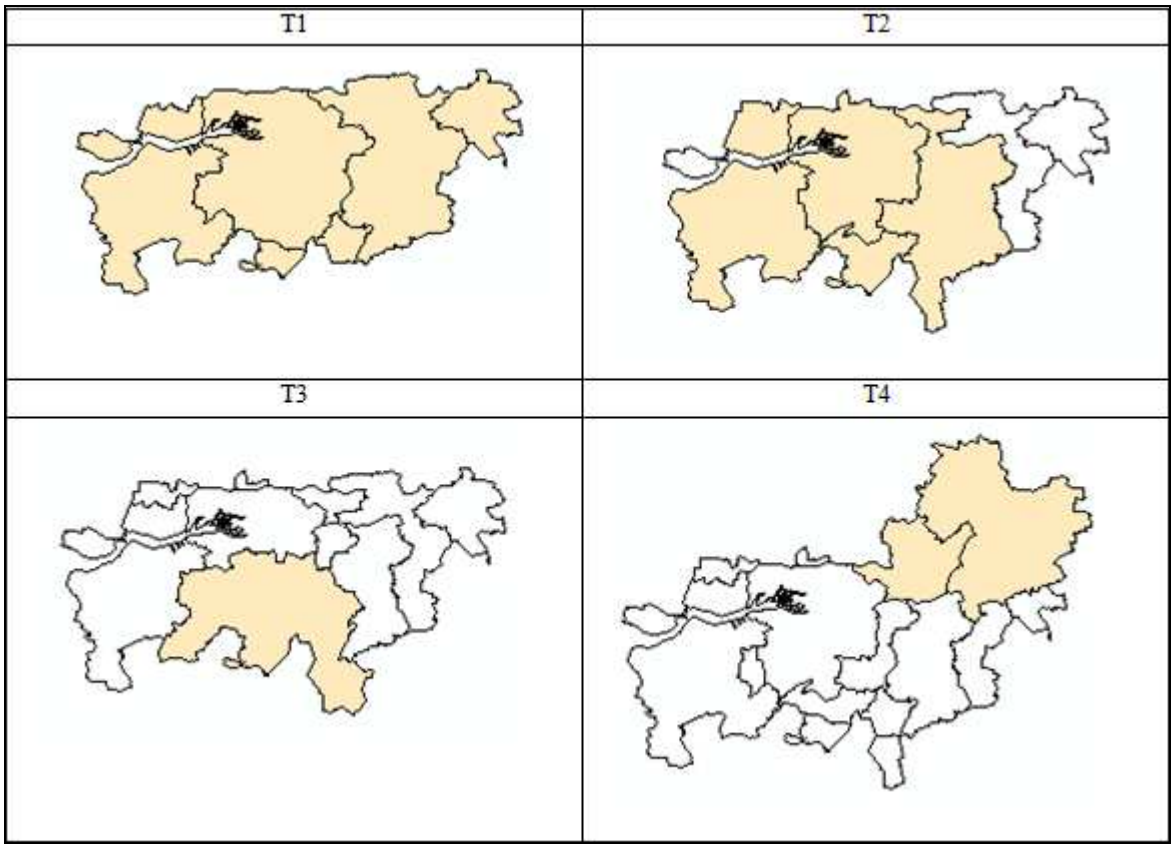


Figure 2-1 : Example of spatial temporal dynamic pattern of a disease. The coloured polygon is the active one. T1, T2, T3 and T4 indicate time periods of a cluster.

2.4. Overview of Pertussis disease

Pertussis, also known as the whooping cough, is a highly infectious disease caused by the bacterium *Bordetella pertussis* and is transmitted via airborne droplets [26]. Pertussis derived its name from the "whoop" sound made from the inspiration of air after a cough. The disease illness is characterised by a prolonged paroxysmal cough often accompanied by an aspiratory whoop. Disease presentation varies with age and history of previous exposure or vaccination.

This disease has a worldwide prevalence and occurs in all age groups. However, it is most severe in young infants who are not fully vaccinated [27]. The introduction of whole-cell pertussis vaccine from 1940s to the 1960s in many countries resulted in dramatic decrease in illness and deaths from pertussis but still the disease remains a considerable to public health problem worldwide. During 1990s an increase of incidence cases was observed in several countries despite high level of vaccination coverage [28]. In many countries despite high vaccination coverage, every year there is an estimation of 40 million cases with approximately 360,000 deaths which occur globally, with which 90% of the occurrence is from developing countries [29]. Pertussis is an epidemic disease with regular epidemic outbreak occurring every 3 to 4 years [20]. In Europe, Australia, Canada and United states pertussis is an endemic in a vaccinated population with an epidemic periodic interval of 3 to 5 years [30].

One of the countries in which this disease is re-emerging despite high vaccination coverage is the Netherlands. After the National Immunisation Programme introduced the mass vaccination, with a whole cell pertussis vaccine in 1990s, the incidence and mortality of pertussis decreased significantly [31]. In view of the fact that, pertussis is most severe to young infants, children were immunized at ages of 3, 4, 5 and 11 months and the vaccine coverage for pertussis, for more than three immunizations, is 96% at the age of 12 months [32]. But, still in 1996, 2771 cases of pertussis were reported in the Inspectorate of the Netherland which was unexpected when compared with 319 cases in 1995 [32]. Since 1996 onwards, this disease has been re-emerging with epidemic peaks after two to three years. Periodic of an epidemic pattern differs from place to place due to the strength of vaccination coverage.

The disease is characterized by the incubation period, a catarrhal phase, paroxysmal phase and convalescent phase. The incubation period is between seven to ten days during which no symptoms are observed. In the catarrhal phase the disease is highly contagious with symptoms like sneezing, low grade fever, bursts of coughing during a single exhalation followed by whooping sound [26, 33]. This phase takes approximately one to two weeks. The paroxysmal phase takes approximately one to six weeks with symptoms such as rapid coughs followed by a long aspiratory effort often associated with a high pitched whoop. The convalescent phase also takes weeks to months with the gradual decrease in frequency and severity of coughing episodes and then disappear.

Several patterns of pertussis in the Netherlands and other countries have been studied. De GREEFF et al. [34] studied long term periodicity and seasonality of pertussis in the Netherlands and reported a slightly increase in long term trend with highest rates seen in 1996, 1999,2001 and 2004 and annual peak incidence in August. However, in some studies they reported pertussis to have seasonality which is not consistent in space and time [35-36], while other studies have shown no seasonality during the period of high introduction of vaccination strategy [34]. Broutin et al [5] performed a comparative analysis of pertussis time series in 12 countries to detect periodicity and synchrony and reported that there were no global synchrony observed for European countries. Another pattern of pertussis is urban-rural hierarchy studied by Broutin et al [20] , who used Pertussis in a small region of Senegal and reported to have urban rural hierarchy pattern during epidemic years.

3. Study area and data description

3.1. Introduction

This chapter describes the datasets used in this study. Section 3.2 describes worldwide data while section 3.3 describes Netherlands Data. The quality of the dataset is described in section 3.4. and their sources and an overview of the study areas. Further, this section will give short description on the quality of the dataset used and problems in the collection of some of the dataset. Lastly, this chapter will be finalised by summary of the chapter.

3.2. Study area

Originally this study aimed at analysing spatial-temporal patterns in pertussis datasets collected worldwide. However, after data quality evaluation we found that African countries had incomplete datasets. Section 3.5 gives a detailed description of data quality. This section describes geographical location of European countries. In the Netherlands, the analysis considers Municipality level, Twente at municipality and postcodes and Enschede at postcode level.

Europe is the world's second-smallest continent by surface area, covering approximately 107 km^2 or approximately 7% of its land area. Of Europe's 45 states, Russia is the largest in area and population, whereas Vatican City is the smallest. Europe is the third most populous continent after Asia and Africa, with a population of 731 million or approximately 11% of the world's population. According to the United Nations (medium estimate), Europe's share may fall to about 7% by 2050. In 1900, Europe's share of the world's population was 25% Figure 3-1.

The Netherlands is located in North Western Europe. West and North-West borders are the coastline of the country with the North Sea. It borders Belgium to the South and Germany to the East. The country is located within Geographic coordinates 520 30’N and 50 45’E (Figure 3-1). This country covers an area of 41,526 km^2, a large part being covered with water. The Netherlands is divided into twelve administrative regions called provinces and each province is divided into municipalities, which in turn are divided into postcodes. The country has approximate population of 16,491,851 with a total number of 395 inhabitants per km^2. It is the most densely populated country in Europe and is the 25th most densely populated country in the world.

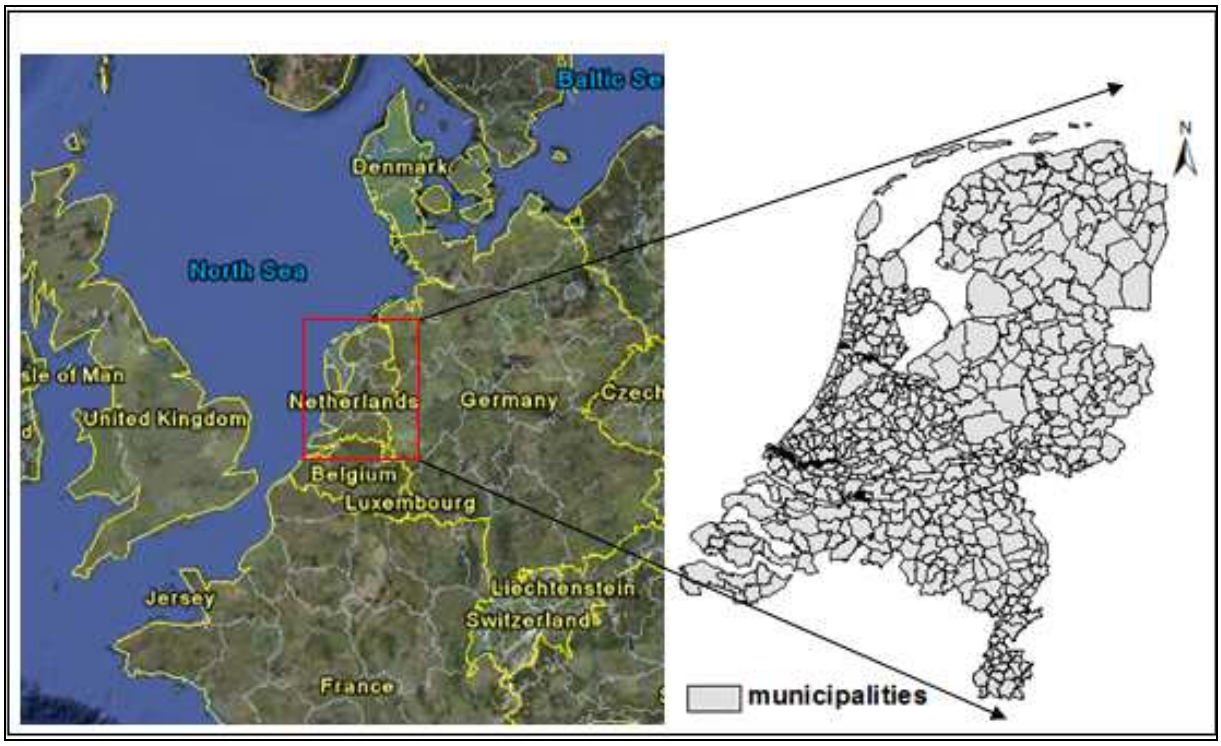


Figure 3-1: Map of the Netherlands. Left in the box is its location in Europe right side is detailed area

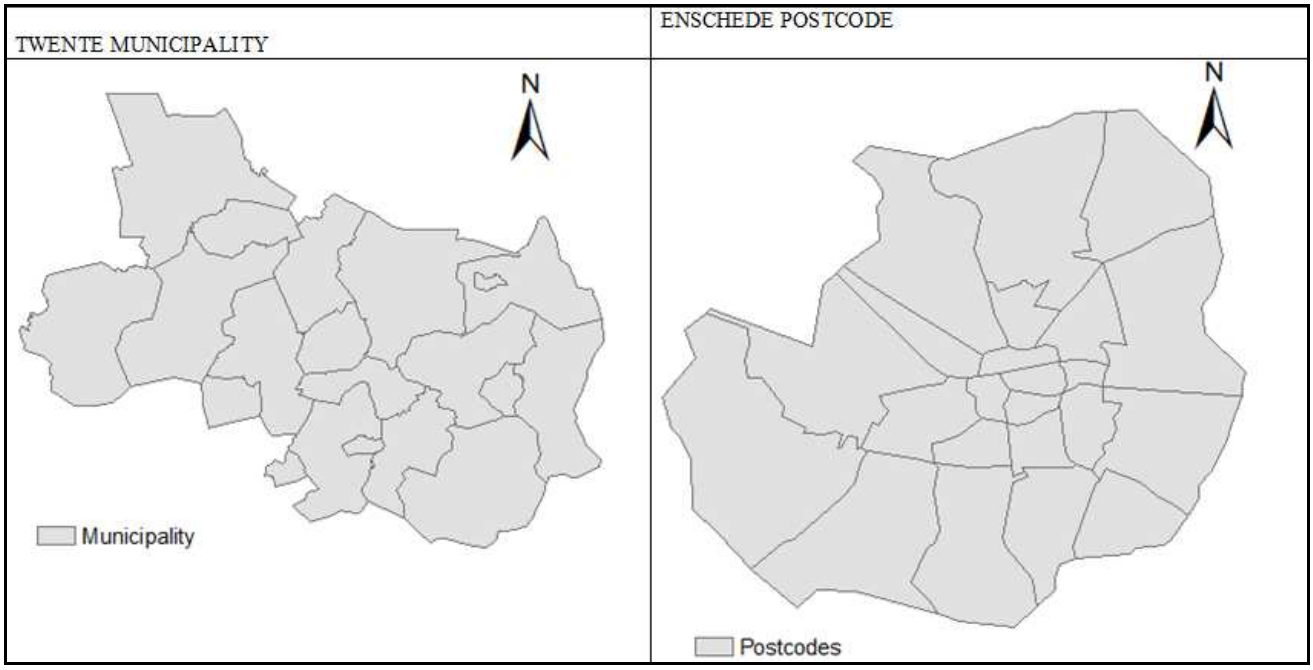


Figure 3-2 : Twente and Municipality postcodes from the Netherlands

3.3. European Countries data

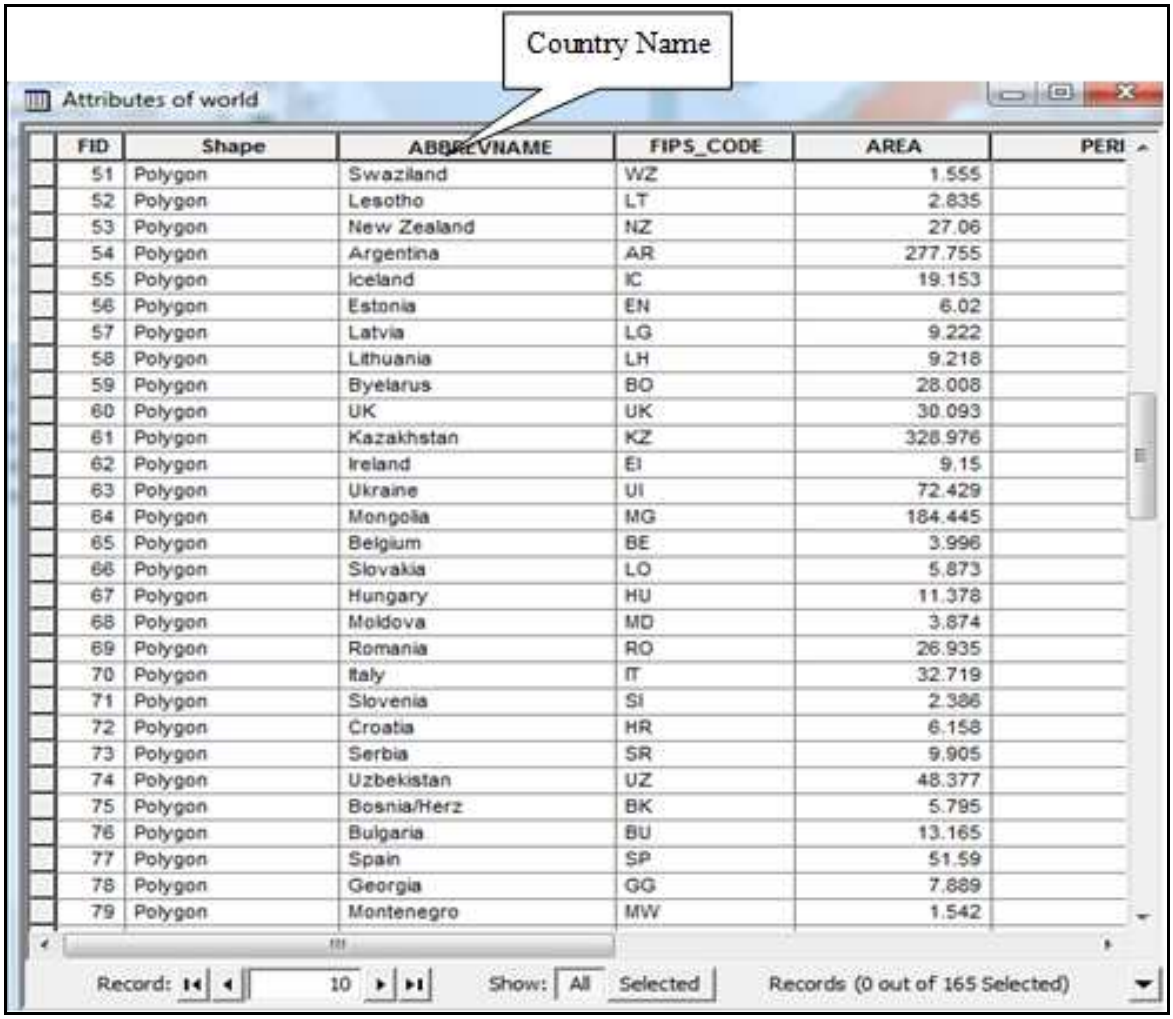
3.3.1. Disease data

The first dataset used was downloaded from [37]. These data contains world wide Microsoft excels yearly pertussis cases. This worldwide data is the officially reported figures of pertussis incidence for individual countries. The database for this data was last updated on 14th October 2009. The data contains the total number of 15432392 pertussis cases starting from 1980 to 2008. The spatial layers of this database are at the country level, the continental level and the global level. It includes incidences for a total of 193 countries. Temporal layers include yearly cases with a total of 29 years.

3.3.2. Spatial data

The country vector map of the world contains attribute values from 252 countries worldwide (Table 3-1) and uses a country shape files downloaded from [38] .This map shows the disease status at the country level and is considered suitable for this thesis, serving as the basis of analysis at the country level of the pertussis spatial temporal patterns.

Table 3-1 Country attribute table showing its contents



3.3.3. Population data

Population data for the worldwide data were downloaded from [39]. The database contains population at the global level. Temporal extent of the population data starts from 1990 to 2008. From this website it is indicated that the data source of these population is International Monetary Fund (IMF).

3.3.4. Data pre-processing

As described in section 3.3.1 the dataset downloaded consists of worldwide pertussis incidence cases. From this we started checking the quality of the dataset for the applicability of these data in this research. After checking the quality of the dataset as described in section 3.7 the European countries were found to be appropriate in this research. Therefore, from the worldwide dataset we selected pertussis cases for all European countries by using Microsoft access via the query;

```
SELECT * into Europe
FROM   world
WHERE c_code=EUR
```

This expression retrieves a selection of all (represented by the *) attributes from the table world that have the country code EUR and put them into table Europe. With this pertussis cases for 53 European countries were selected. These European data consist of 1076988 pertussis incidence cases starting from 1990 to 2008. The result for selected countries is presented in Table 3-2.

Table 3-2 : Pertussis cases for European countries

data2 : Table

C_CODE	CNAME	2008	2007	2006	2005	2004	200
EUR	Armenia	3	1	2	6	7	
EUR	Austria	183	133	72	130	130	
EUR	Azerbaija	5	12	60	2	2	
EUR	Belarus	125	156	76	80	80	
EUR	Belgium	260	293	196	231	96	
EUR	Bosnia an	41	46	42	39	24	
EUR	Bulgaria	193	269	335	313	222	
EUR	Croatia	102	123	80	124	197	
EUR	Cyprus	3	9	8	6	15	
EUR	Czech Rep	767	186	234	412	373	
EUR	Denmark	105	81	55	129	228	
EUR	Estonia	485	409	153	63	455	
EUR	Finland	511	480	535	552	1631	
EUR	France			246			
EUR	Georgia	129	63	135	165	207	
EUR	Germany						
EUR	Greece	22	29		24	23	
EUR	Hungary	33	48	17	22	31	
EUR	Iceland	2	4	3	6	1	
EUR	Ireland	102	78	62	84	90	
EUR	Israel	2169	2635	1495	357	1653	
EUR	Italy	174	474	545	403	763	
EUR	Kazakhsta	45	69	120	222	57	
EUR	Kyrgyzsta	45	71	56	91	9	

Record: 3 of 53

Furthermore, from the worldwide vector map we selected all European countries and export it to EuropeMap via ArcGIS 9.3. With this a total number of 48 European countries were selected.

3.4. The Netherlands Data

3.4.1. Disease data

The second dataset is a database on pertussis incidence from the Netherlands. It contains a total number of 44202 cases observed during 3534 days of pertussis diagnosed between January 1, 1993 and October 16, 2004 at the postcode level. Spatial layers include the regional level, the municipality level and the 4-digit postcode level. The postcode vector map contains information of 3973 postcodes. The postcode layer aggregates to 538 municipalities and 38 regions. Temporal layers include daily, monthly and yearly (with a total of 12 years). Addresses of patients are not provided and there is no information about further medical records of the patient.
The procedure of reporting is as follows. General practitioners are responsible to make home calls to find out the health status of people in different postcode areas. After every two weeks these practitioners have to report the number of pertussis cases to the Public Health Centers (GGD).Then the GGD forwards the reported cases to a central database of Institute of Public Health and environment (RIVM). It is not indicated whether this routine is either followed or not [40].

3.4.2. Population Data

Population data for the municipality and postcode level in the Netherlands from 1993 to 2004 were downloaded from the Central Bureau of Statistic (CBS) Netherlands [41]. The procedure for data collection in the Netherlands falls under the law of the Municipal basic Administration [41]. This law requires the CBS to receive municipality data from all inhabitants as recorded in the Municipal basic Administration at the beginning of every year. Illegal immigrants are excluded from the population count. Also tourists are not included as their number can be temporary and locally large.

3.5. Data Quality issues

The worldwide database contains the absolute number of incidences, spatially aggregated at the country level and temporally aggregated to the year. The level of aggregation serves as the starting point in the identification of space or space-time patterns of a disease data. When performing analysis of patterns in disease data, one must consider the appropriate scale to use in order to identify existing patterns. Aggregation of a disease data at a very large scale for instance at the Country level, might result into inaccurate detection of a disease patterns in terms of location, since the area is too homogenous. Also, if the temporal scale of the patterns is too large, e.g. a year, it might result to late detection of a disease outbreak in terms of controlling outbreak of infectious disease. Moreover, the temporal duration for the reported cases should be such that, it allows us to identify preventive measures of a disease at the earliest possible time. This report is lacking in the worldwide dataset on how long it takes for pertussis incidence cases to be reported. At this level of scale, it was possible only for comparative analysis at the country level. Therefore, we used these worldwide to determine spatial, temporal and spatial temporal patterns of an epidemic and endemic of pertussis data.

This worldwide dataset consists of 3813985 pertussis cases from 1990 to 2008 and it includes incidences for 193 countries. From these countries, there were incidence cases for 46 African

countries and 53 European countries. Starting from 1990 to 2008, there were a total number of 727450 cases from African countries and 1076988 pertussis cases from European countries. Moreover, in this dataset, there were blank spaces and 0 cases which were reported as no cases reported and zero cases reported respectively. Based on this information, in African countries there were a total number of 355 blank spaces and 173 zero cases while in European countries, there were 142 blank spaces and 39 zero cases. From this observation, developing countries were not appropriate for this research since they had a lot of missing data. Ideally, it was in our plan to perform a comparative analysis between European countries and Developing countries by comparing endemic and epidemic years but due to lack of appropriate dataset from developing countries, this analysis was impossible.

Since European countries data were found to be relatively reliable, this research used European countries dataset for the analysis of disease patterns at the country level. It was used to compare endemic and epidemic years and to identify the periodic pattern and synchrony of epidemic and endemic years of pertussis incidence cases. With this, 18 countries with pertussis incidence cases starting from 1993 to 2004 were selected. These countries were selected due to completeness of the data. Countries which were selected includes Austria, Bulgaria, Irelands, Netherlands, Norway, Sweden, Poland, United kingdom, Russia ,Estonia, Latvia, Croatia, Italy, Finland, Ukraine, Romania, Slovakia and Czech Republic which sum to 18 countries. In Norway and Sweden there was no data in 1998 and 1999, in Ireland there was no data in 2000, in Italy there was no data in 2001 and in Finland there was no data in 1999 and 2000, therefore we excluded these years in the analysis. Table 1-2 present the information of counties used for the comparative analysis in this research. Although we used these dataset for the analysis we are not sure if these reported incidence cases reflects a true increase or decrease of a disease or not.

Table 3-3 : A Table showing blank space for some of the countries

data4 : Table

COUNTRY/YEAR	YEAR_1990	YEAR_1991	YEAR_1992	YEAR_1993	YEAR_1994	YEAR_1995	YE
Bulgaria	26	13	82	74	22	77	
Netherlands	471	164	169	294	536	319	
Russia	24960	30876	24004	39439	48614	20626	
Estonia	87	222	200	279	161	50	
Latvia	129	208	106	115	137	58	
Croatia	304	392	313	666	310	252	
Poland	292	302	590	314	697	549	
Ireland	803	843	860	869	353	436	
Norway	208	140	124	17	107	71	
Sweden	10697	11432	9782	11542	13187	10655	
United Kingdom	16605	6282	2750	4718	4837	2399	
Italy	16992	19356	7055	4246	13360	14359	

Record: 1 of 12

The Netherlands dataset covers the entire country. This dataset was appropriate for detection of disease patterns at different spatial and temporal scale.

3.6. Summary

This chapter gives the description of the study area and the quality of the dataset. Next chapter will deal with explaining the method used to attain the main objective of this research.

4. Research methods

4.1. Introduction

This research study aims to detect spatial temporal patterns in an empirical pertussis data .To accomplish this objective, we identified spatial, spatial-temporal and temporal patterns by comparing epidemic and endemic years of pertussis data whereas the hierarchical spread, synchrony and travelling waves were determined.

Initially we started by detecting synchrony and travelling waves for European countries. Next, detection of hierarchical spread of pertussis was determined on the scale of the Netherlands by using week rank method and space-time scan statistic. In space-time scan statistics, spatial temporal clustering analysis was detected. Further the criteria for analysis of spatial temporal clusters detected by space-time span statistics were explained.

4.2. Synchrony and traveling waves

The spatial, temporal and spatial-temporal synchrony of epidemic and endemic years for pertussis in European countries were identified so as to observe if there is inter epidemic and endemic periods for European countries. Initial plan was to compare epidemic and endemic periods between developing countries and European countries but due to lack of reliable data from developing countries we decided to concentrate only in European countries. Although, the data was also incomplete, countries with reliable data were selected. For this, Austria, Bulgaria, Irelands, Netherlands, Norway, Sweden, Poland, United Kingdom, Russia, Estonia, Latvia, Croatia, Italy, Finland, Ukraine, Romania, Slovakia and Czech Republic were countries of interest.

We started our analysis by classifying European countries if the year was epidemic or endemic. This was done because only count data existed for European data but there were no existing information on whether the years were endemic or epidemic. To determine this, we used Netherlands data to define epidemic and endemic years for other European countries. The Netherland data was selected due to the fact that, epidemic and endemic years in the Netherlands are already recognised.

In the Netherlands years 1996, 1999, 2001 and 2004 were epidemic years and years 1993, 1994, 1995, 1997, 1998, 2000, 2002 and 2003 were endemic years. Therefore, we used the Netherland data with referenced epidemic year to generate mathematical formula for the definition of epidemic and endemic years for other European countries. With this, temporal extent of analysis was 12 years starting from 1993 to 2004.To get spatial distribution (or standard rate) of the disease for all countries, we divided the number of incidence cases of each country and for every year by 100,000. Thereafter, expected rates and increment rate of a disease was determined so as to set the threshold value which will be used to define epidemic and endemic years of the other countries. Expected rate (E_R) was calculated by finding the average incidence rates for every year and for each country. The expected

increment (S_C) was calculated by finding the sum of difference between the current years (Y_N) and the previous year ($_{YN-1}$) divided by the total number of their differences ($_{N-1}$). Expected and increment rates were all combined via the formula;

$=IF (OR (AND (Y_N>_{YN-1}, Y_N>E_R, Y_N<Y_{N+1}), Y_N-Y_{N-1}>S_C),"1"," 0")$

Where;
S_C = SUM $(Y_N-Y_{N-1})/N-1$
Y_{N+1}= Following year
E_R=Average incidence rate for all years
1= Epidemic
0= Endemic

Finally, epidemic and endemic years in the Netherlands were compared with epidemic and endemic years of other European countries. All cartographic manipulations and displays were performed using ArcGIS 9.3 software. Lastly, we visualise the data to detect the synchrony or the spread of a disease for epidemic and endemic years.

4.3. Hierarchical spread pattern

4.3.1. Netherlands data aggregation process

Since our aim is to identify hierarchical spread of a disease at different scale levels, the Netherlands dataset was employed in this analysis. This data was from daily reported cases from 1st January 1993 to October 2004 and was collected at the postcode level. Based on this, we started our analysis by aggregating daily cases of every postcode and for each year into weekly reported cases. Then, weekly reported cases at postcode level were joined with the Netherlands postcode vector map. The Idea for this was to get the municipality name for each postcodes and weekly reported cases at municipality level. "Select by attributes tool" in ArcGIS 9.3, was then used to select those postcodes belonging to Enschede municipality and Twente region. In Enschede municipality we got a total number of 22 postcodes and in Twente region we got a total number of 119 postcodes. Afterwards, ArcGIS "dissolve tool" was used to aggregate postcode level to municipality level by summing up the weekly cases of the year and for each year. At this stage a total number of 538 municipalities were obtained in the Netherlands. The resulting table (Table 2-1) indicates an example of weekly cases of pertussis at the Municipality level for the year 1996 in the Netherlands.

Table 4-1 : Weekly cases at municipality level of the year 1996

Attributes of 1996

OID	GEMNM	WEEK1	WEEK2	WEEK3	WEEK4	WEEK6	WEEK7	WEEK8	WEEK9	V
21	Amsterdam	3	0	2	1	1	0	1	2	
371	Ridderkerk	2	0	0	0	0	0	0	0	
1	Aalburg	1	0	0	0	0	0	0	0	
13	Alphen aan den Rijn	1	0	0	0	0	0	0	0	
37	Bathmen	1	0	0	0	0	0	0	0	
107	Deventer	1	0	0	0	0	0	0	0	
141	Enschede	1	0	0	0	0	0	0	1	
178	Haarlem	1	0	0	0	0	0	0	1	
183	Hardenberg	1	0	0	0	0	0	0	1	
199	Heerlen	1	0	0	0	0	0	0	0	
206	Helmond	1	0	0	0	0	0	0	0	
323	Nijmegen	1	0	0	0	0	0	0	0	
335	Oegstgeest	1	0	0	0	0	0	0	0	
351	Oss	1	0	0	0	0	0	0	0	
381	Rotterdam	1	0	1	0	0	0	0	0	
519	Zaltbommel	1	0	0	0	0	0	0	2	
523	Zeewolde	1	0	0	0	0	0	0	0	
0	Aa en Hunze	0	0	0	0	0	0	0	0	
2	Aalsmeer	0	0	0	0	0	0	0	0	
3	Aalten	0	0	0	0	0	0	0	0	
4	Abcoude	0	0	0	0	0	0	0	0	
5	Achtkarspelen	0	0	0	0	0	0	0	0	
6	Akersloot	0	0	0	0	0	0	0	0	
7	Alblasserdam	0	1	0	0	0	0	0	0	
8	Albrandswaard	0	0	0	0	0	0	0	0	
9	Alkemade	0	0	0	0	0	0	0	0	

Record: 0 Show: All Selected Records (0 out of 538 Selected) Options

4.3.2. Definition of urban and rural areas

For detecting an urban-rural hierarchy, a classification (urban-rural) was needed for all administrative units. Rural areas were defined as those locations with low number of people and urban areas as location with dense population. This is because in the Netherlands there is no official definition of what are Urban and Rural areas. With this, two variables were determined as relevant parameters; absolute population and population density. The method used was to define urban and rural areas based on only absolute population, only population density and the combination of absolute population and population density. For the absolute population we started by defining urban areas as those municipalities with population size greater than 50000 and then those municipalities with population size greater than 100000. If the population was less than the defined values then the area was regarded as rural. Figure 2-2 shows the result of definition of urban and rural areas based on the absolute population. The shortcoming of this approach is that large municipalities in area with low population density were still defined as urban area which is incorrect.

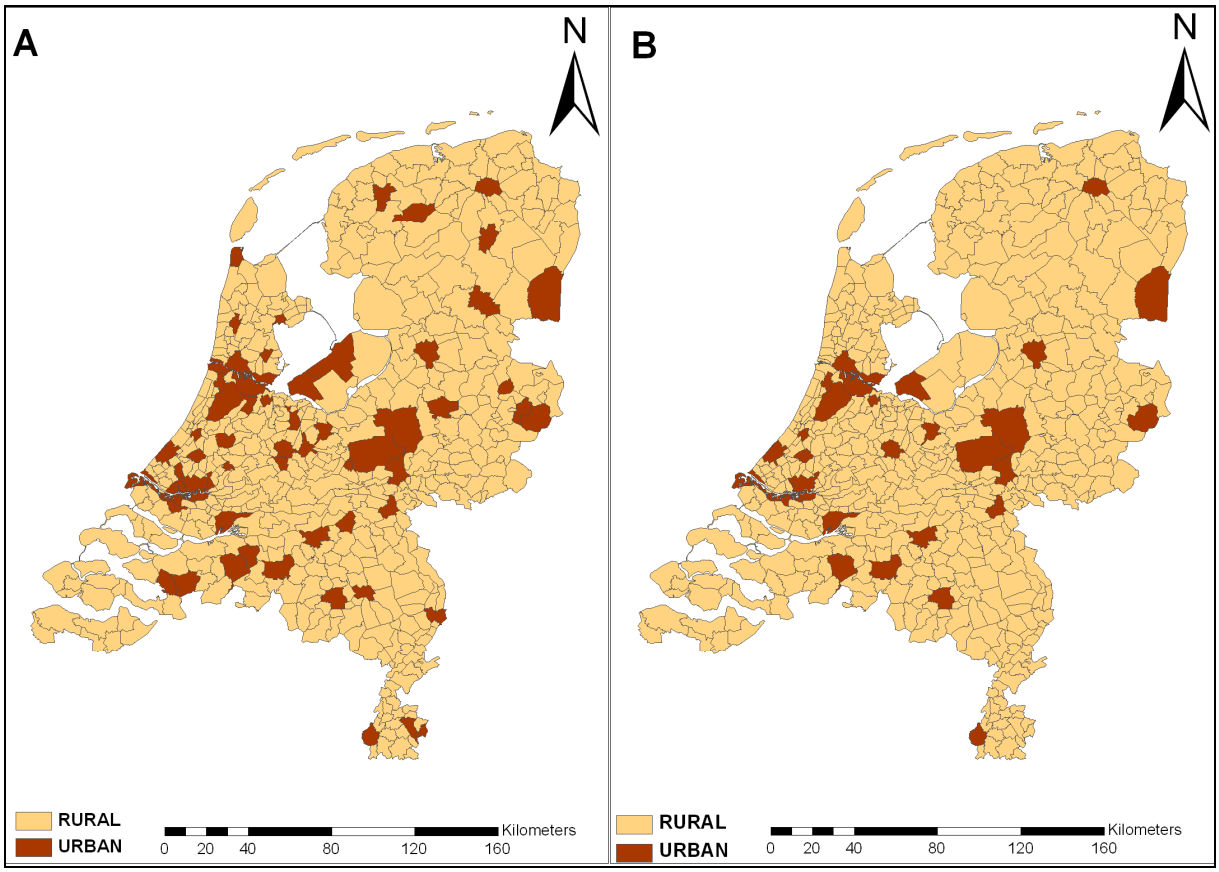


Figure 4-1 : Definition of urban and rural areas in the Netherlands. Map A indicates urban areas for the population greater than 50000. Map B urban areas for population greater than 100000.

Secondly we defined urban and rural areas based on the population density, urban areas were defined as those municipalities with population density greater than 1000 km^{-2}. If population density was less than this threshold value it was considered as rural areas. The result for this is presented in Figure 2-3. This result was also not very good since small area with high population was also regarded as urban areas.

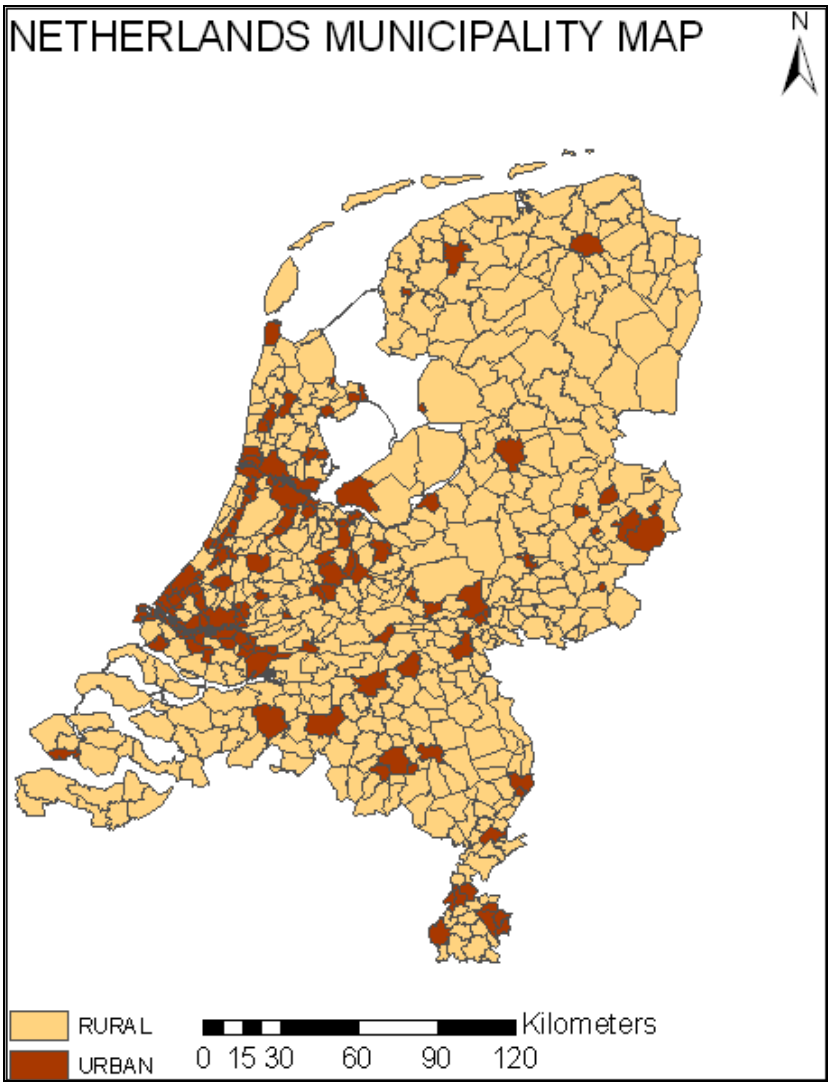


Figure 4-2 : Result of definition of urban areas based on population density

The last approach was to define urban and rural areas based on the combination of population density and absolute population. From this, urban areas were defined as those municipalities with population size greater than 75000 and population density greater than 750 km^{-2}otherwise the area was regarded as rural. With this, a total number of 31 municipalities were classified as urban areas and 507 municipalities as rural areas. The result for this is presented in Figure 2-4. This result was found to be appropriate since the urban area classification have large population size which is proportional to the size of the area.

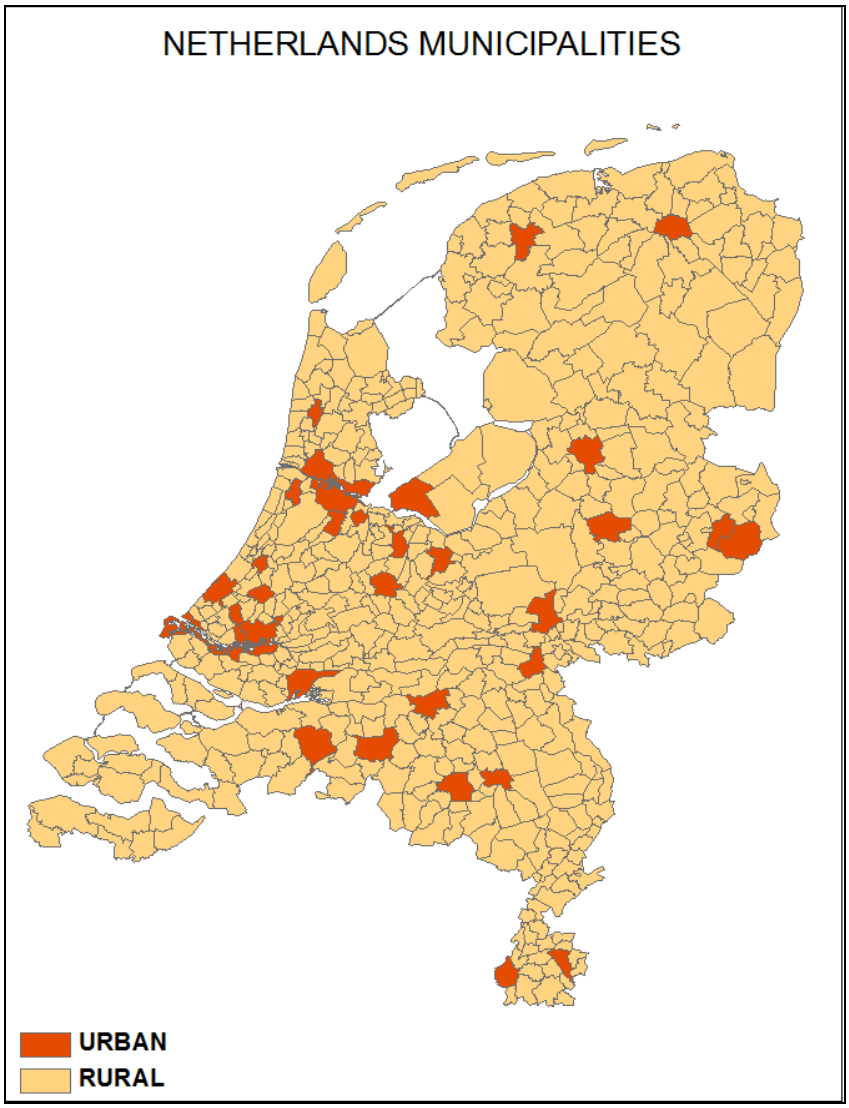


Figure 4-3 : Definition of urban and rural areas in the Netherlands Municipalities

We used similar approach to define urban and rural areas in Twente region and Enschede municipality. Starting with Twente, we defined urban areas as those municipalities with population size greater than 20000 and population density greater than 1000 km^{-2}. With this, a total number of 5 municipalities in Twente region were categorized as urban areas and 17 municipalities as rural areas.

At the level of postcode, we defined urban and rural areas based on population density since the areal unit is small. We defined urban areas in Twente postcodes and Enschede postcodes as those postcodes with population density greater than 2000 km^{-2}. Of which, a total number of 28 postcodes in Twente region were classified as urban and 91 postcodes as rural areas. Consequently, in Enschede municipality a total number of 10 postcodes were classified as urban and 12 postcodes as rural areas. Results for this classification is presented in Figure 2-5 and Figure 2-6 respectively.

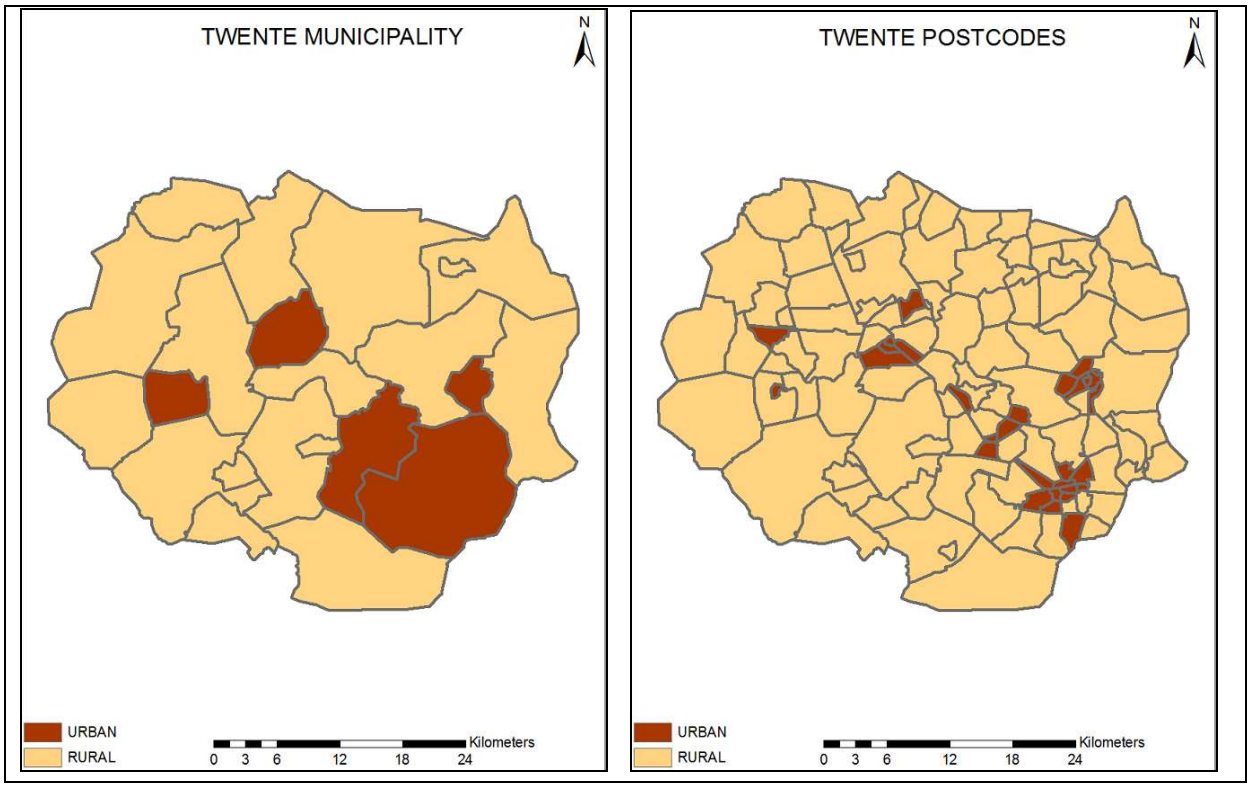


Figure 4-4 : Urban and rural areas in Twente region. Left side is Municipality and right side is postcode

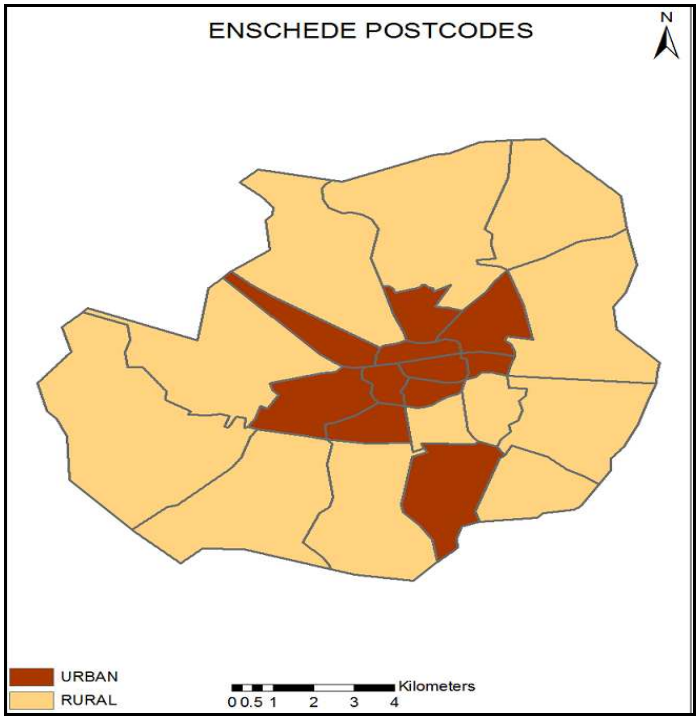


Figure 4-5 : Urban an rural areas in Enschede

4.3.3. Week rank Method

From the work of Broutin et al [20] ,who used week rank method to detect hierarchical pattern of pertussis in small area of Senegal, showed a clear indication that pertussis disease starts in urban areas before it spreads to the surrounding rural areas. Based on this method, we detected hierarchical pattern of pertussis in the Netherlands. Since this research aimed to identify patterns at different scales, week rank method was used to identify hierarchical patterns in the Netherlands at Municipality, in Twente region at the level of Municipality and postcodes and in Enschede at postcode level. The multi scale analysis was also important so as to test if the hierarchical pattern of spread of disease will be synchronous in all levels of analysis starting from the Netherlands to postcode level. With this, few steps have to be followed;

Based on the definition of urban and rural areas defined earlier for Enschede municipality, Twente region and the Netherlands, weeks of the first cases of the year for each urban and rural area were determined since the beginning of the year and for every year. Means for the weeks of first cases for all urban areas and all rural areas were then calculated for both endemic and epidemic years. For endemic years, the combination of 8 years was used while for epidemic years a combination of 3 years was used. Endemic years were, 1993, 1994, 1995, 1997, 1998, 2000, 2002 and 2003 and epidemic years were 1996, 1999 and 2001. However, year 2004 was also an epidemic year in the Netherlands, but it was not included in this analysis since it had daily cases up to 16th October 2004 and for the remaining months, data was reported on monthly basis hence the data could not be used on weekly basis. Furthermore, to determine the persistence of a disease in a location, weeks of the maximum cases of the year for each urban and rural area were determined since the beginning of the year and for every year. We then calculated means for the weeks of maximum cases for all urban areas and all rural areas for both endemic and epidemic years. The dates of maximum cases and the first cases in weeks for every locality were determined. Finally, we compared means for the weeks of first cases and maximum cases of pertussis for both endemic and epidemic years by using non-parametric Kruskal-Wallis test statistics to see if there is a significant difference in means between urban and rural areas in both endemic and epidemic years, or if there is a significant difference of means in urban areas during endemic and epidemic years, or if there is a significant difference of means in rural areas during endemic and epidemic years.

Kruskal-Wallis test is a non parametric (distribution free) test, which can be used to compare two or more groups of sample data. It can be used to test the hypothesis that the number of unpaired sample originate from the same population. If the null hypothesis, being the hypothesis that two means are similar is rejected (When $P < 0.05$) then the conclusion is that there is a statistically significant difference between two means. This test is recommended if the data is not normally distributed. From this analysis, weeks of first cases and maximum cases in all urban areas and in all rural areas were not normally distributed since number of rural areas were large compared to number of urban areas at all scale levels, starting from the Netherlands up to Enschede postcodes. Additionally, for the comparison of the combination of weeks of first cases and maximum cases for all urban areas and all rural areas between endemic and epidemic years were also not normally distributed. This is due to the fact that, the number of endemic years was larger compared to the number of epidemic years (i.e. 8 years against 3 years). Hence Kruskal-Wallis test was appropriate for this analysis to test if the calculated means are of statistically significant difference.

4.3.4. Space-time scan statistic method

SatScan Software was developed to analyze spatial, temporal and space-time data using Spatial, temporal and Space-time scan statistic [42]. Space-time scan statistics implemented using SatScan program is the most widely used approach for the detection of spatial-temporal clusters together with identifying their approximate location and timing [43-44]. Additionally, it can be used as a tool to identify clusters which require further investigation and the one which are most likely to occur [17, 45]. Space-time clusters are determined when unusual high occurrences of disease events are observed within a spatial and temporal range defined by a user [43, 46].

Moreover, space-time scan statistics uses cylindrical window to scan for areas with high rates of cases, low rates of cases or both. This cylindrical window is in three dimensions with which a circular (or elliptic) base of a cylinder represents geographic space and height of the cylinder is representing time [17, 42-43, 47].The key idea of the scanning window is to move in both space and time, whereby it visits each possible locations and size and each time periods as specified by a user. At each spatial location and size of the window used, the number of observed cases and expected cases is counted of which the location with high rate of disease than expected is reported as clusters of cases [42, 48-49].

Space-time data in space-time scan statistic can be analyzed in either prospective or retrospective manner [17, 42]. In retrospective, analysis is performed only once in a fixed geographic location. This analysis scans both live clusters that reach the end of the study period and historic clusters that die out before the study period end date. Prospective analysis is used only in detection of clusters that reach the end of the study period. This analysis is repeated every day, week, month or year which is suitable for an earliest detection of an outbreak.

Usually, space-time scan statistic required either the uniform population at risk, control group data or data that provides the explanation about location and temporal distribution of the underlying population at risk [48-49]. Whereas, based on the information of the population at risk, the expected number of cases is accurately estimated, this becomes less reliable when the catchment area such as emergence department visits, and pharmacy sales is not well defined [48-49]. In this study, the catchment area would correspond to the place of origin of any individual that might happen to have a particular disease with which this is impossible, since we only have reported pertussis cases with no other record about the patient. The space-time permutation scan statistic is the only solution to this since it only requires the actual number of cases [49].

In this study, we used a prospective space-time permutation scan statistic that requires the actual number of cases to scan for areas with high rates of disease throughout the study area. Space-time permutation models produce accurate results when the study period is less than a year otherwise there might be a risk of population shift bias [42]. To meet the space time permutation requirement, we used only short periods of time for the analysis with which number of pertussis cases for each municipality for each separate year starting from 1996 to 1999 was used as a case file. This data was arranged on a weekly basis in order to detect significant clusters for every week since the beginning of the year and for every year (i.e. years, 1996, 1997, 1998 and 1999). The spatial scale of interest was Municipality level in the Netherlands. The latitude and longitude of centroids of each municipality

were used as the coordinate file. Prospective analysis was performed for an earliest detection of an outbreak. Also this analysis was performed in order to know the spatial location of clusters, temporal and spatial-temporal duration of clusters.

Since space-time permutation scan statistic does not require the population-at-risk data, the expected number of cases is calculated by using only the number of cases. Therefore, the expected cases were calculated as follows:

The number of cases for a particular for a particular location (in this case it is municipality or postcodes) p over a given period of time (week in this case) is defied as C_{pt}. Then the total number of observed cases C is found as:

$$C = \sum_{p} \sum_{t} C_{pt} \qquad \textbf{4-1}$$

The expected number of cases at location p and time t, is calculated as follows:

$$\mu_{pt} = \frac{1}{C}\left(\sum_{p} C_{pt}\right)\left(\sum_{t} C_{pt}\right) \qquad \textbf{4-2}$$

This equation indicates that the expected number of cases (μ_{pt}), corresponds to the proportion of all cases occurring in a particular location *p*, times the total number of cases that occurred during time period *t*. A cylinder is defined as the set of *p*-*t* pairs that fall within a particular circular area and occurs within the same time interval. The expected number of cases in that particular cylinder *A* is then calculated as:

$$\mu_{A} = \sum_{(p,t) \in A} \mu_{pt} \qquad \textbf{4-3}$$

This calculation of expected number of cases takes an assumption that, the probability of a case being in location p that observed in a time period *t* is the same to all time periods *t*. If C_A is the observed number of cases for that particular cylinder, there must be evidence that the particular cylinder contains an outbreak (or cluster). Therefore, observed outbreak is tested by using by using Poisson generalized likelihood ratio (GLR) which is given by the formula:

$$GLR = \left(\frac{C_A}{\mu_A}\right)^{C_A}\left(\frac{C - C_A}{C - C_A}\right)^{(C - C_A)} \qquad \textbf{4-4}$$

The cylinder with the highest *GLR* is considered to be the most likely cluster [48-49]. To test the significance of the cluster detected, Monte Carlo simulations was used. In this case, null hypothesis was high number of cases within an area as compared to outside areas. Hence, *p*-values of these tests were estimated by comparing the rank of the maximum likelihood from the real data sets, with the maximum likelihood from the random dataset. If this rank is *R*, then *p*= *R*/ (1 + number of simulations) [42]. In this case we used 9999 Monte Carlo replications to estimate the significance levels of observed clusters.

For all analysis the most likely clusters and secondary clusters with statistical significance of $p < 0.01$ and recurrence interval of 1 year were reported. For prospective analysis, recurrence interval is used as an alternative to the *p*-value. This measure reflects how often the observed or maximum likelihood ratio will be observed by chance, by assuming analysis is repeated within a specified time interval length.

4.3.5. Parameter setting

According to SatScan user guide, the maximum spatial scanning window should not exceed 50% of population at risk [42]. This threshold appear to be unrealistic as such window might detect clusters which are composed of not only the high risk location which are of interest but also many low risk locations which are not of interest [46]. Therefore, to avoid these inconveniences, the maximum spatial scanning window was chosen to detect clusters within an area covering 6% of population at risk. Depending on the size and the shape of the study area, this maximum spatial window was selected to avoid scanning outside the study area and to avoid inclusion of low risk areas with which we are not interested with. Although, also at this spatial window the low risk areas was also detected but at the minimal level compared to 50 %.

The default value of maximum temporal window was also set to be at most 50% of the study period [42]. This study determine the maximum temporal window based on temporal risk window, which means that an infected cluster remains infectious and the disease could be spread to other clusters. Therefore, we selected the maximum temporal scanning window (temporal risk window) to be 14 days with which it will include the incubation period of pertussis which takes approximately 7 to 10 days and contagious period which takes approximately 1 to 2 weeks. As the maximum temporal window was already there, the time aggregated was set to 1 day.

4.3.6. Criteria for analysing spatial temporal results

Spatial temporal clustering patterns were identified to determine the catchment areas (location where the first infection occurred), diffusion (dispersion) and the persistence of disease clusters in both epidemic and endemic years. With this, two checks will be performed:

- Can hierarchical spread be determined of which we will compare this result with the one obtained by week rank method?
- Can clusters be identified according to their spatial temporal characteristics and can this lead to identification of different patterns for epidemic and endemic years?

The spatial locations of clusters were considered as urban and rural areas. For instance, in case of origin of a disease, locations with which the disease occurred both for the first time were examined. Afterwards, check was made to see if the observed location is an urban or rural area by comparing first occurrence to urban and rural map as defined earlier in this chapter.

Then, we classify the origin of a disease as urban or rural area and present results in a table (Figure 4-7). Lastly, based on results presented on the table the difference between endemic an epidemic year was made.

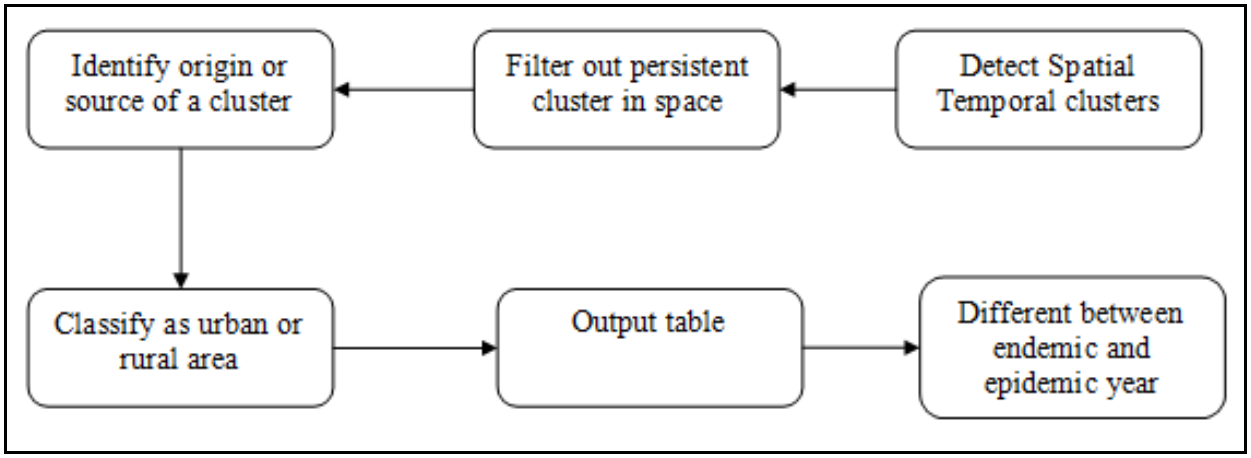


Figure 4-6 : Spatial temporal analysis flow chart

From the work of Broutin et al. [20] who used pertussis incidence cases in a small area of Senegal, there was a clear indication that during epidemic year first cases of pertussis arrived sooner in urban areas than in rural areas and more fadeouts in rural areas than in urban areas. Based on this, we hypothesize to have more diffusion in urban than in rural areas and to have more fadeouts in rural areas than in urban areas.

4.3.7. Classification of clusters as diffusion, persistence or fadeout

For the detecting the persistent, spreading and fadeout period of clusters we formulated the criteria for the relationship in both space and time. As a starting point we defined a spreading cluster as a cluster that is close, overlap or meet with each other in both space and time. We considered two clusters to be close spatially if there is no more than one spatial unit without cases in location. Similarly, spatially overlapping clusters were regarded as those clusters which their interior intersects with one another and the interior of one cluster intersects with the exterior of another cluster. Two or more clusters were considered to meet in space when each cluster touches the boundary of another cluster. Examples of meet overlap and persistence clusters are presented in Figure 5-14.

Consequently, two clusters were considered to be close time if the separation in time detection is not more than three weeks. When separation time of detection between clusters are more than three weeks we grouped it as fadeout cluster. This definition of fadeout period was based on the study of Broutin et al. [20, 24] who defined fadeout period of pertussis to be three weeks without cases at location. The clusters were considered to meet in time if their detection time is the same. Lastly, we regard two clusters to overlap in time if one cluster touches time boundary of another cluster and vice versa. For instance, if cluster A occurs within time interval 1996/2/29 to 1996/3/13 and cluster B occurs within time interval 1996/3/10 to 1996/3/23 we consider it as time overlap.

We defined persistent cluster when in a given spatial location two or more clusters overlaps or come on top of other cluster in a continuous period of time.

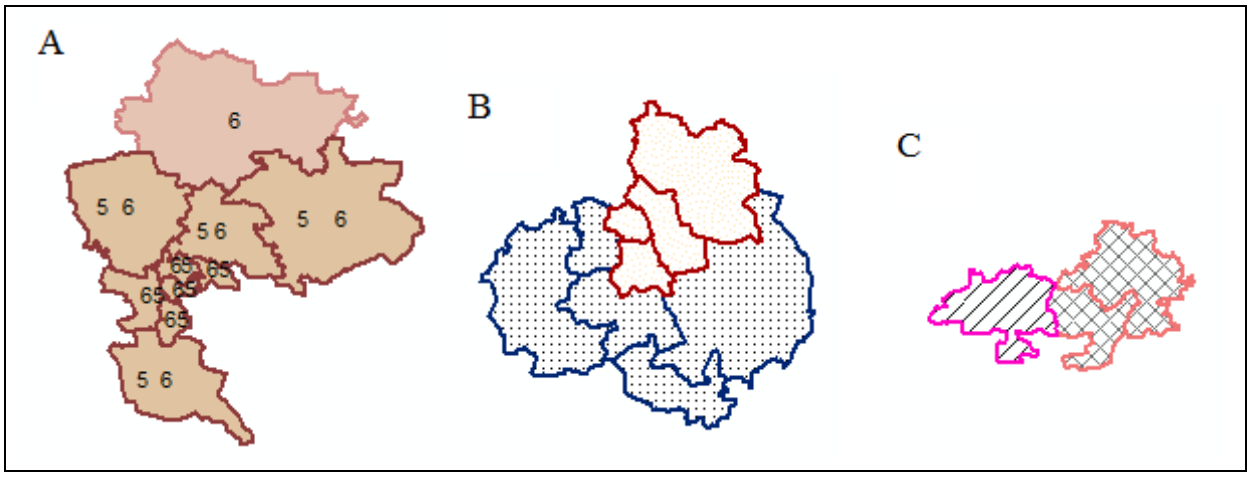


Figure 4-7 : Examples of clusters. A is persistent, B is overlap and C is clusters that meets or touches

4.4. Summary

In this chapter the general sequence and the techniques used in this research were explained. Week rank methods for identifying the hierarchical patterns of a disease were explained in detail and Space-time scan statistic. Further, the criteria for analyzing the spatial temporal clusters detected by space-time permutation model were described. In addition the chapter describes the formula used to describe epidemic and endemic years in the Netherlands. Next chapter present the result of the method used in this chapter and it interprets the resulted data.

5. Result of the analysis

5.1. Introduction

This Chapter presents results of analysis of the procedures described in chapter 4. It analyzes the urban-rural patterns of an epidemic and endemic years as well as setting criteria for analysing spatial temporal patterns of an epidemic and endemic data detected by space-time scan statistic. Then it gives the result of spatial synchrony and travelling waves of an epidemic.

5.2. Result of spatial synchrony and traveling waves

5.2.1. Result of spatial synchrony

Spatial synchrony was checked to find if epidemic for all European countries occurs simultaneously. Spatial synchrony was checked for 18 European countries. Including Austria, Bulgaria, Irelands, Netherlands, Norway, Sweden, Poland, United Kingdom, Russia, Estonia, Latvia, Croatia, Italy, Finland, Ukraine, Romania, Slovakia and Czech Republic.

The results of endemic and epidemic synchrony are illustrated in Figures 5-1 and 5-2 respectively. Figure 5-1 shows the patterns of a disease from 1993 to 1998 and figure 5-2 shows the patterns of a disease from year 1999 to 2004.

During 1993 epidemic synchrony was observed between Sweden, United Kingdom, Russia, Croatia, Ireland, Ukraine, Romania and Slovakia. During 1994 epidemic synchrony was observed between Austria, Sweden, United Kingdom, Russia, Latvia, Italy and Finland.

During 1995 synchrony was observed only in Bulgaria and Italy. During 1996 epidemic synchrony was observed between the Netherlands, Croatia, Romania and Czech.

During 1997 epidemic synchrony was observed between Poland, Russia, Estonia, Latvia, and Ireland. 1998 epidemic synchrony was observed between Poland and Finland.

During 1999, epidemic synchrony was observed between the Netherlands, Bulgaria and Slovakia. During 2000, epidemic synchrony was observed between Poland, Russia, Estonia and Latvia.

During 2001, epidemic synchrony was observed between Austria, the Netherlands and Latvia. During 2002 no synchrony observed.

During 2003, epidemic synchrony was observed between Bulgaria, Poland, Estonia, Finland and Czech Republic. During 2004, epidemic synchrony was observed between the Netherlands, Bulgaria, Norway, Finland, Ukraine and Czech Republic.

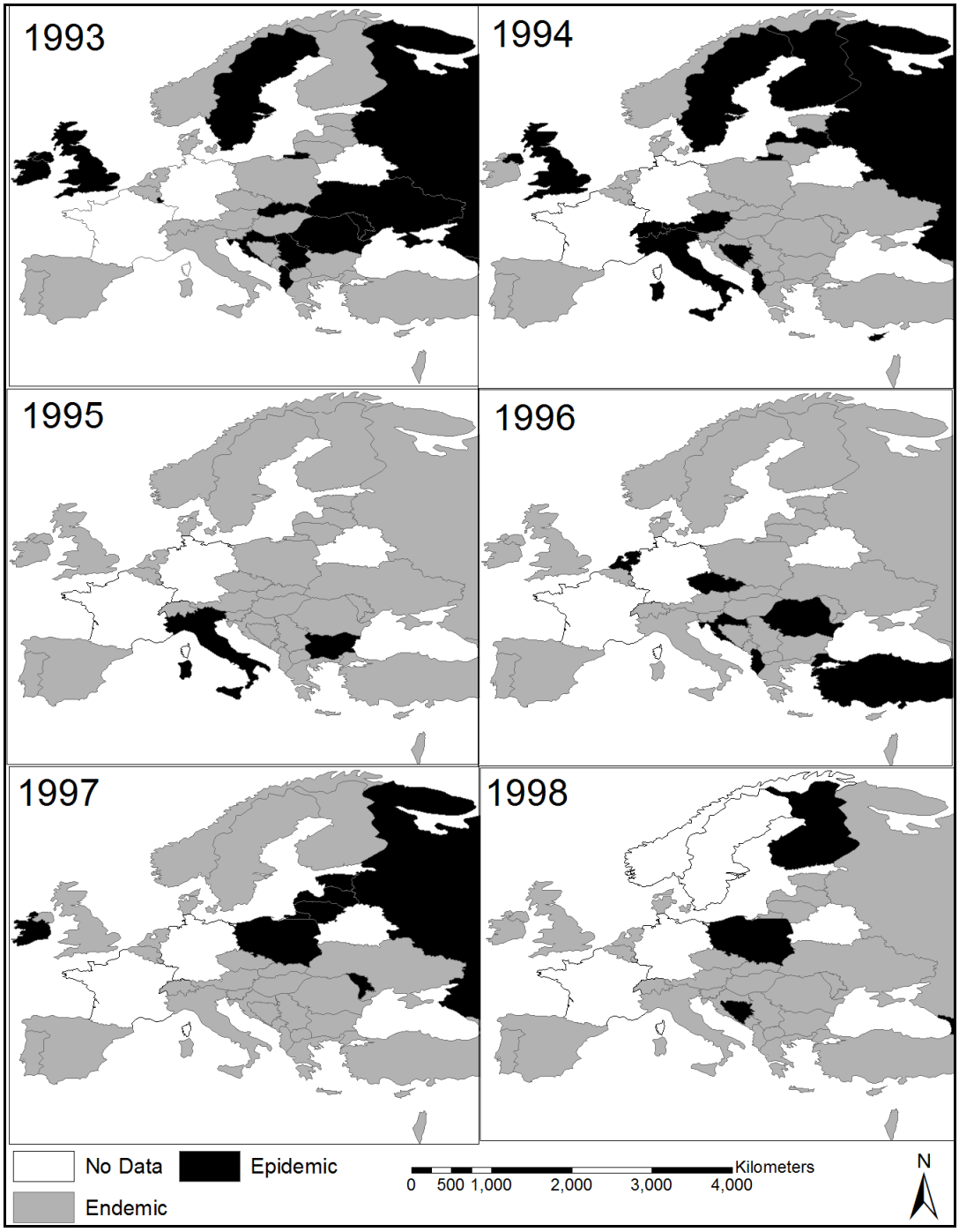


Figure 5-1 : Epidemic patterns in the Netherlands compared other European countries

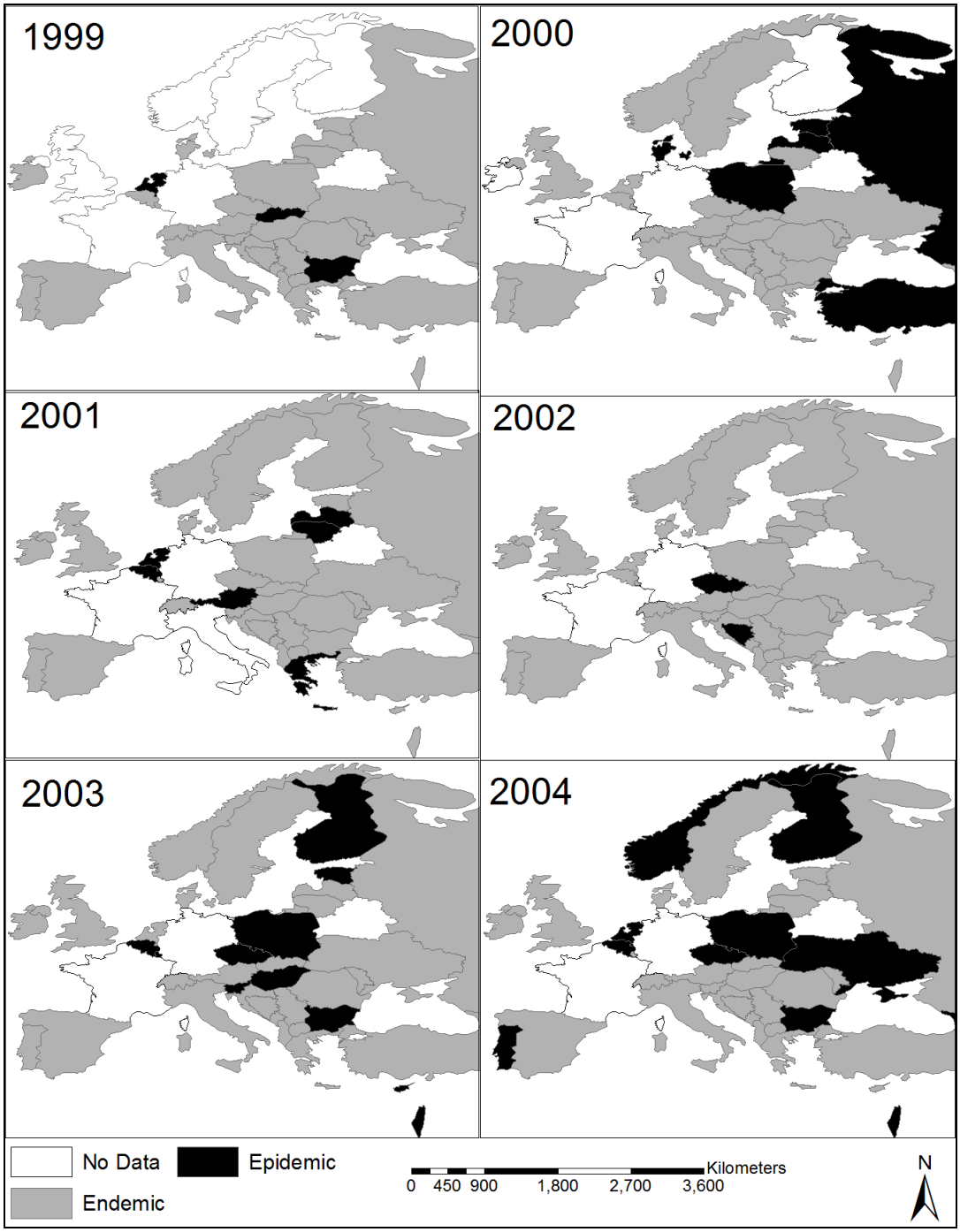


Figure 5-2 : Endemic patterns in the Netherlands compared to other European countries

5.2.2. Result of travelling waves of epidemic

Travelling was checked to find if epidemic is spreading to the nearby country. The travelling wave of epidemic was checked for 18 countries. Including Austria, Bulgaria, Irelands, Netherlands, Norway, Sweden, Poland, United Kingdom, Russia, Estonia, Latvia, Croatia, Italy, Finland, Ukraine, Romania, Slovakia and Czech Republic.

During 1993 epidemic was in Sweden and Russia in 1994 the epidemic was in Finland (Figure 5-3).

Furthermore, during 1995 epidemic was in Bulgaria but during 1996 the epidemic showed to travel to Romania (Figure 5-4). During 1997 Russia was an epidemic which showed to travel to Finland during 1998(Figure 5-5). During 1999 Slovakia showed epidemic but 2000 it seems the disease travel to Poland. During 2001 epidemic was in Austria, 2002 epidemic travel to Czech Republic a, 2003 from Czech Republic to Poland and 2004 epidemic from Poland to Ukraine (Figure 5-7)

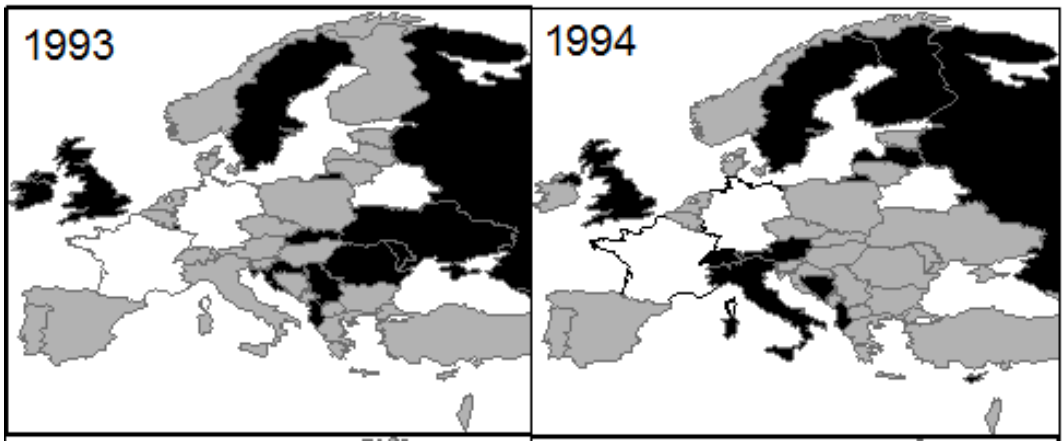


Figure 5-3 Travelling epidemic from Russia and Sweden during 1993 to Finland 1994

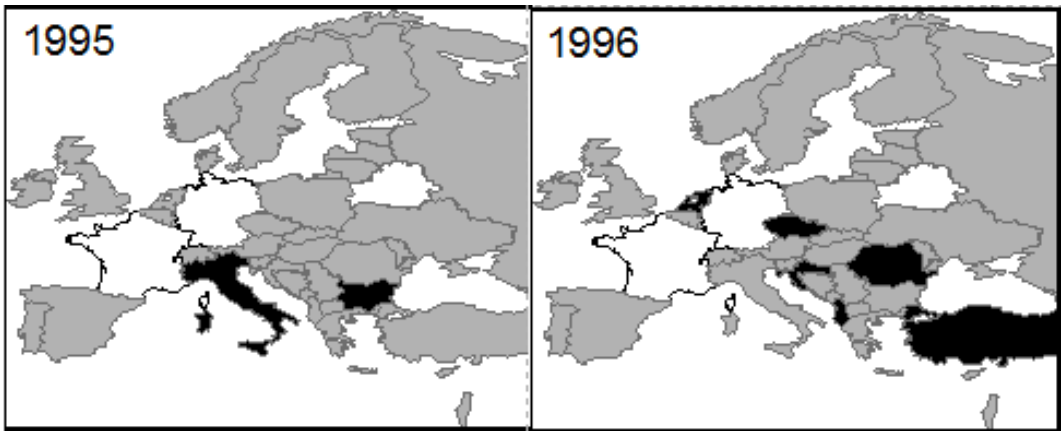


Figure 5-4 Travelling epidemic from Bulgaria during 1995 to Romania 1996

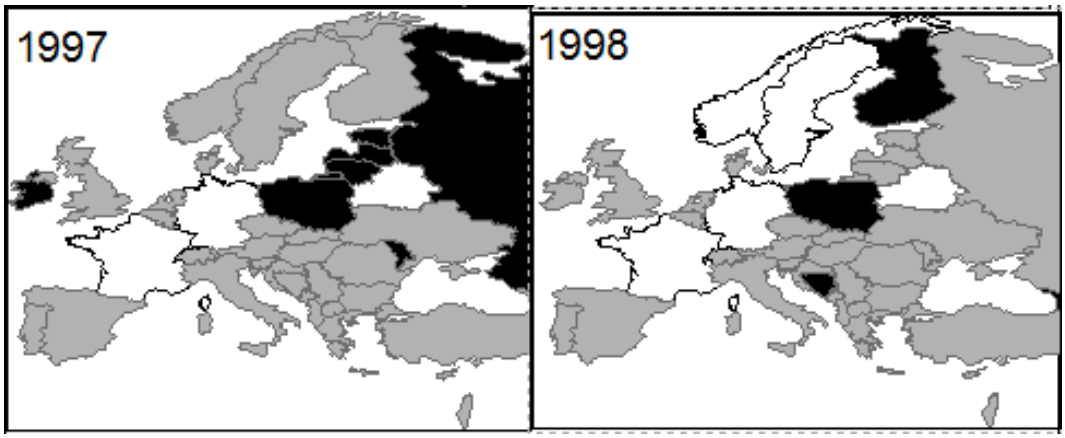


Figure 5-5 Travelling epidemic from Russia during 1997 to Finland 1998

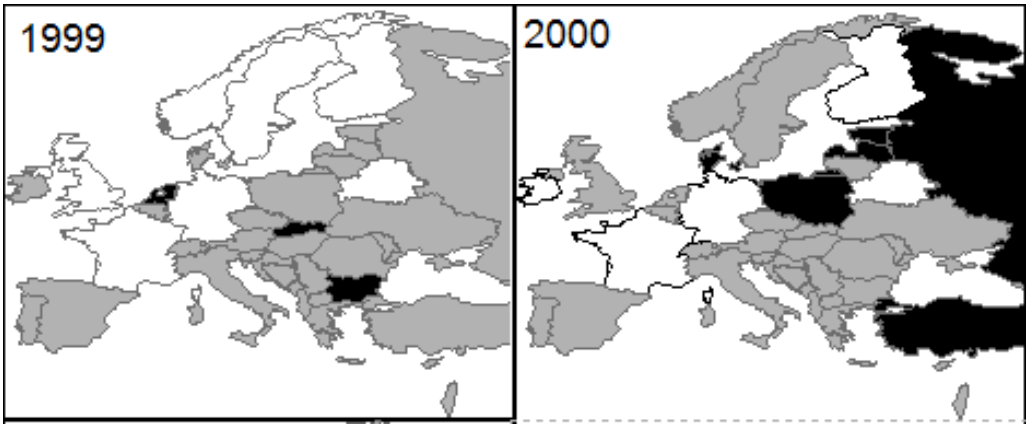


Figure 5-6 Travelling epidemic from Slovakia during 1999 to Poland 2000

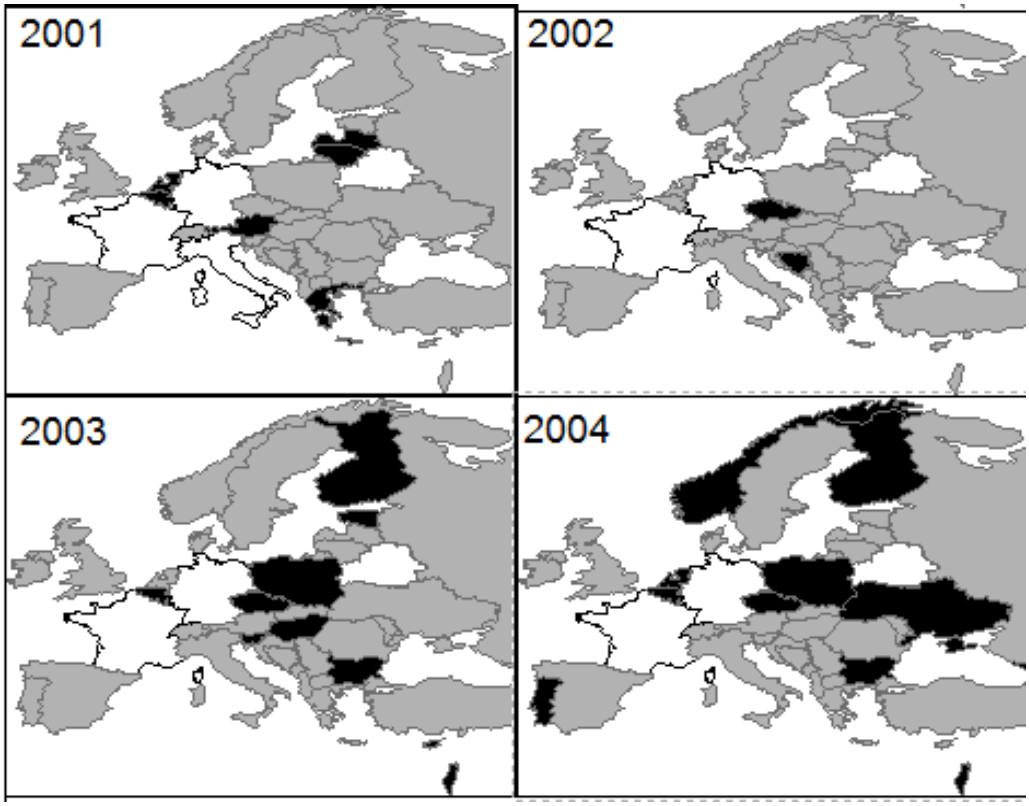


Figure 5-7 travelling waves from Austria 2001 to Czech Republic 2002 to Poland 2003 to Ukraine 2004

5.3. Result of detection of hierarchical spread pattern

5.3.1. Result of the Netherlands Municipality

The week rank method (described in section 4.3.3) has been applied at the scale of the Netherlands, using municipalities for the years 1993 to 2003. Two tests have been performed me, week of the first cases during a year and week for the maximum cases of pertussis during a year.

Results of week rank means for the first cases of pertussis for both endemic and epidemic years in The Netherlands are illustrated in Figure 5-3. Week of first cases in rural areas during both endemic and epidemic year were 19.57 (se=0.33), and 19.88 (se= 0.41) respectively. Week ranks of the first cases in Urban aggregate for both endemic and epidemic years were 12.30 (se= 0.92) and 19.88 (se= 0.41)

respectively (Table 5-1). Rural areas showed similar dates for the first cases during both endemic and epidemic years (H= 1.125, d.f= 1, p= 0.29), while in urban areas the disease arrive earlier during epidemic year than during endemic year (H= 7.1894, d.f= 1, p =0.007). During both endemic (H= 58.2504, d.f = 1, p= 2.31 x $10^{-14)}$ and epidemic (H= 68.534, d.f = 2.2 x 10^{-16}) years, the disease arrived earlier in urban areas than in rural areas (Table 5-2).

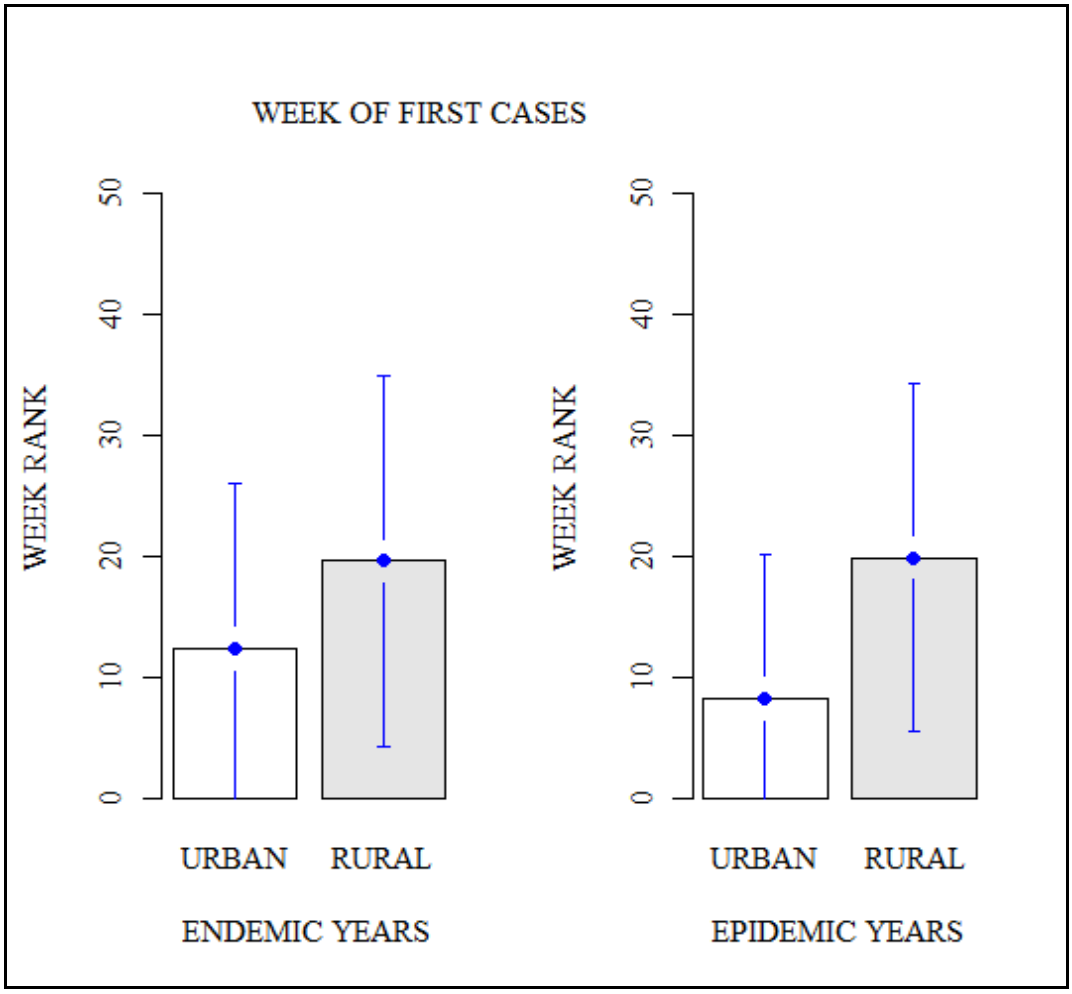


Figure 5-8 : Week rank means of first cases of pertussis in the Netherlands Municipality

Table 5-1 : Means and standard error of first cases during a year in the Netherlands Municipalities

Years	***Location***	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	12.30	0.92
	Rural	19.57	0.33
Epidemic	Urban	8.21	1.25
	Rural	19.88	0.41

Table 5-2 : comparison of means for the first cases of pertussis in the Netherlands

YEARS/ LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	58.2504	1	2.31×10^{-14}
EPIDEMIC	RURAL VS URBAN	68.534	1	2.2×10^{-16}
RURAL	ENDEMIC VS EPIDEMIC	1.125	1	0.29
URBAN	ENDEMIC VS EPIDEMIC	7.1894	1	0.007

Week rank means of pertussis for the maximum number of cases for pertussis in both endemic and epidemic years in the Netherlands are illustrated in Figure 5-4. Week rank of a maximum number of cases in rural areas for both endemic and epidemic years were, 23.75 (se= 0.32) and 27.92 (se= 0.39) respectively. Week rank means of a maximum number of cases in urban areas for both endemic and epidemic years were 24.26 (se= 1.00) and 29.97 (se= 1.48) respectively. The result week rank means of a maximum number of cases are presented in Table 5-3. In rural areas, during endemic year, the disease dies out earlier then during epidemic years (H= 59.9751, d.f=1, P= 9.61×10^{-15}). Similar pattern was also observed in urban areas were the disease dies out sooner during endemic years than during epidemic years (H= 10.3096, d.f= 1, P= 0.001323). During endemic year, there was no significant difference in urban and rural areas (H= 0.1682, d.f =1, P=0.68). Likewise, during epidemic years week of a maximum cases was similar in both urban and rural areas (H= 2.2658, d.f=1, P= 0.1323). Summary for the comparison are presented in Table 5-4.

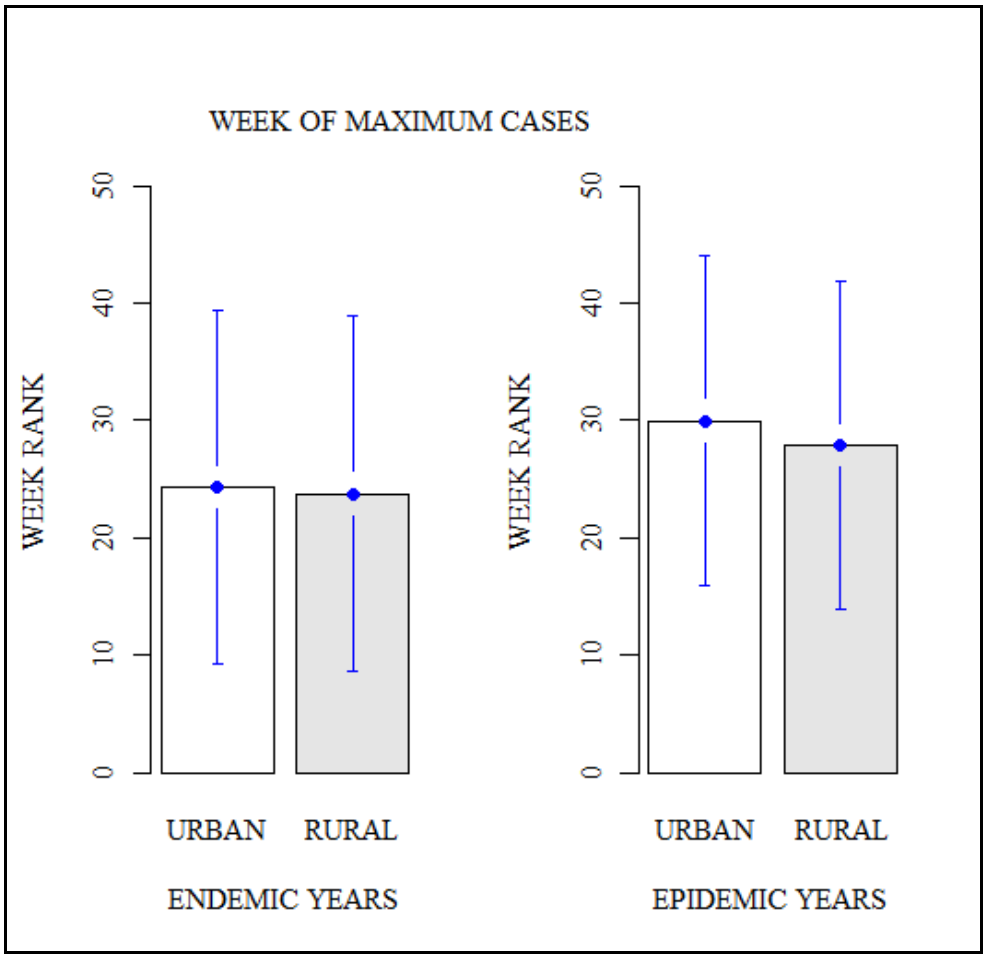


Figure 5-9 : Week rank means of the maximum cases in the Netherlands Municipality

Table 5-3 : Means and standard errors for the maximum cases of pertussis in the Netherlands

Years	***Location***	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	24.26	1.00
	Rural	23.75	0.32
Epidemic	Urban	29.97	1.48
	Rural	27.92	0.39

Table 5-4 : Comparison of means for the maximum cases of pertussis in the Netherlands

YEARS/ LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	0.1682	1	0.68
EPIDEMIC	RURAL VS URBAN	2.2658	1	0.1323
RURAL	ENDEMIC VS EPIDEMIC	59.9751	1	9.61×10^{-15}
URBAN	ENDEMIC VS EPIDEMIC	10.3096	1	0.001323

5.3.2. Result of Twente Municipalities

The same analysis as were performed for the Netherlands were repeated for the Twente at two spatial aggregation levels: municipality and postal codes. Definition of urban and rural areas was performed as described in section 4.3.2. In this analysis means of the first cases and means of the maximum cases were determined. Results were generated for the years 1993 to 2003. The years were aggregated for epidemic and endemic years.

Result of week rank means of first cases for Twente Municipality in both endemic and epidemic years are illustrated in Figure 5-5. Weeks of first cases in rural areas during both endemic and epidemic years were, 22.53 (se=1.79) and 30.18 (se= 2.60) respectively. Weeks of first cases in urban areas during both endemic and epidemic years were, 18.76 (se=2.58) and 22.23 (se=4.89) respectively. These results are clearly presented in Table 5-5. The disease seems to persist in rural areas for a longer period of time during epidemic years than during endemic years (H=6.453, d.f= 1, P=0.01107). There were no significant differences during endemic years between urban and rural areas (H=1.7933, d.f=1,P=0.1805) ,during epidemic years between urban and rural areas (H=1.8552, d.f=1, P=0.1732) and in urban areas between epidemic and endemic years (H=0.0819,d.f=1, P=0.7748) Table 5-6.

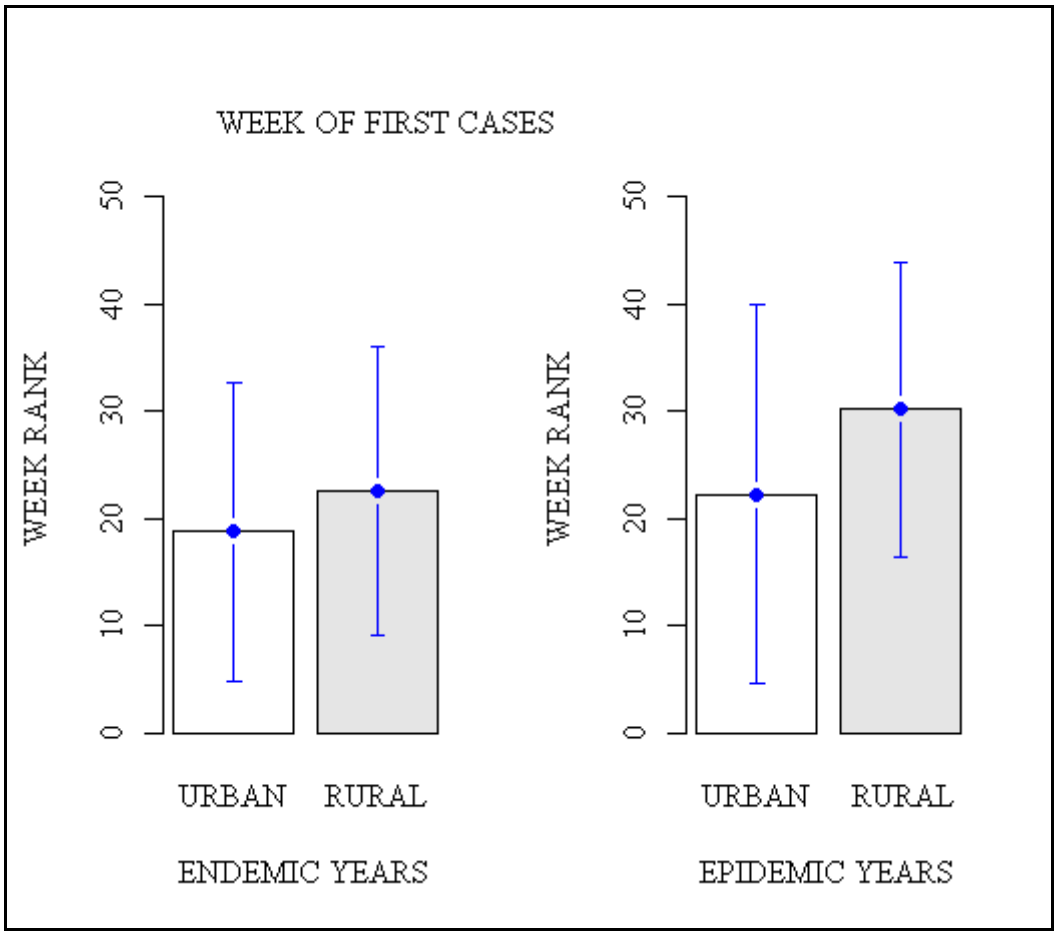


Figure 5-10 : Week rank means for the first cases of pertussis in Twente Municipality

Table 5-5 : Means and standard error of means for the first cases of pertussis in Twente Municipality

Years	***Location***	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	18.76	2.58
	Rural	22.53	1.79
Epidemic	Urban	22.23	4.89
	Rural	30.18	2.60

Table 5-6 : Comparison of the means for the first cases of pertussis in Twente Municipality

YEARS/LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	1.7933	1	0.1805
EPIDEMIC	RURAL VS URBAN	1.8552	1	0.1732
RURAL	ENDEMIC VS EPIDEMIC	6.4543	1	0.01107
URBAN	ENDEMIC VS EPIDEMIC	0.0819	1	0.7748

Results of week rank means of maximum cases for Twente Municipalities are illustrated in Figure 5-6. Week for a maximum cases in rural areas during both endemic and epidemic years were, 24.98 (se= 1.63) and 33 (se=2.40) respectively. Weeks of a maximum cases in urban areas during both endemic and epidemic period were, 28.3 (se=2.590 and 27.69 (se=4.93) respectively. Results for week rank means are presented in Table 5-7. There were no significant difference during epidemic years (H= 0.5745, d.f=1, P=0.4485) and during endemic years (H=0.993, d.f=1, P=0.319) between urban and rural areas. More so, week of a maximum cases in urban areas during both endemic and epidemic years were not different (H=0.015, d.f=1, P=0.9024). In rural areas, during epidemic year there is maximum persistence of the disease than during endemic years (H=8.7674, d.f=1, P=0.003067) Table 5-8.

Table 5-7 : Means and standard error of means for the maximum cases in Twente Municipality

Years	*Location*	*Mean*	*Standard error of the mean (se)*
Endemic	Urban	28.03	2.59
	Rural	24.98	1.63
Epidemic	Urban	27.69	4.93
	Rural	33.0	2.40

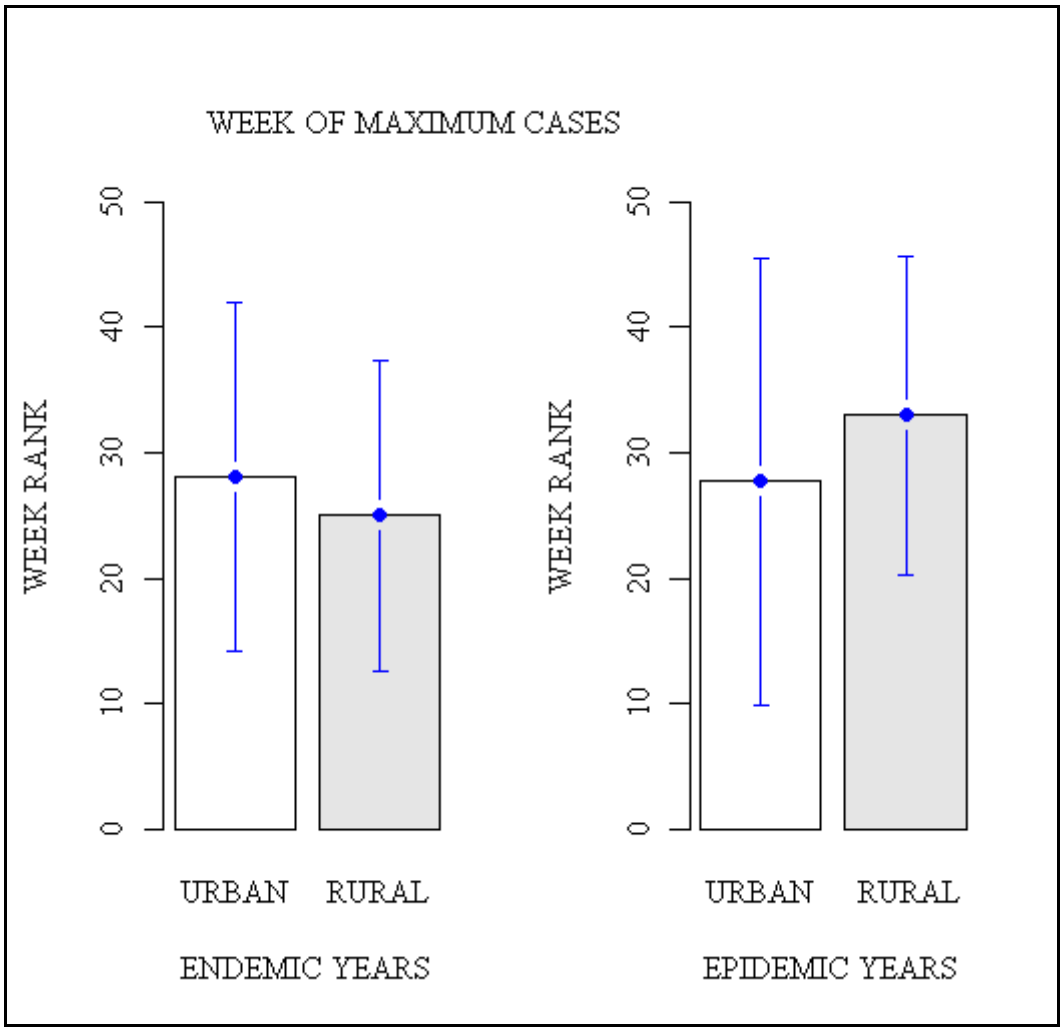


Figure 5-11 : Week rank means for the maximum cases of pertussis in Twente Municipality

Table 5-8 : Comparison of means for the maximum cases in Twente Municipality

YEARS/ LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	0.993	1	0.319
EPIDEMIC	RURAL VS URBAN	0.5745	1	0.4485
RURAL	ENDEMIC VS EPIDEMIC	8.7674	1	0.003067
URBAN	ENDEMIC VS EPIDEMIC	0.015	1	0.9024

5.3.3. Twente postcodes

Result of week rank means of first cases for Twente postcodes during both endemic and epidemic years are illustrated in Figure 5-7. Weeks of first cases in rural areas during both endemic and epidemic years were, 25.28 (se= 1.10) and 31.45(se=1.54) respectively. Weeks of first cases in urban areas during both endemic and epidemic years were, 25.31 (se=1.73), 28.57 (se=2.45) respectively Table 5-9. The rank means of first cases were homogenous for urban areas between endemic and epidemic years (H=1.4565, d.f=1, P=0.2275), during endemic years between urban and rural areas (H=0.0128, d.f=1, P=0.91) and during epidemic years between urban and rural areas (H=1.5551, d.f=1, P=0.21). In rural areas there is the possibility of the disease to arrive earlier persist during epidemic years than during endemic years (H=12.7621, d.f=1, P=0.0003537) Table 5-10.

Table 5-9 : Means and standard error of the means for the first cases in Twente postcodes

Years	***Location***	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	25.31	1.73
	Rural	25.28	1.10
Epidemic	Urban	28.57	2.45
	Rural	31.45	1.54

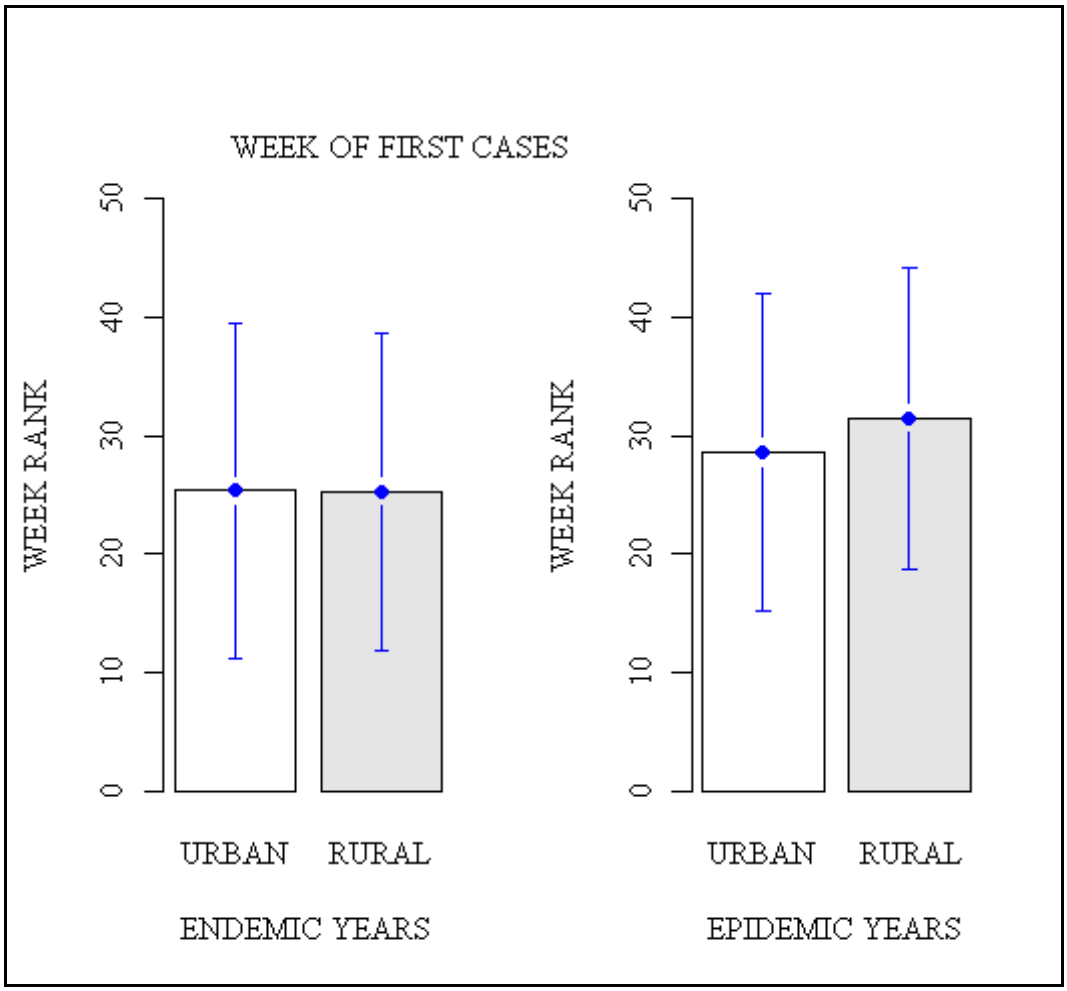


Figure 5-12 : Week rank means for the first cases of pertussis during a year in Twente postcodes

Table 5-10 : Comparison of means for the first cases of pertussis in Twente postcodes

YEARS/LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	0.0128	1	0.91
EPIDEMIC	RURAL VS URBAN	1.5551	1	0.21
RURAL	ENDEMIC VS EPIDEMIC	12.7621	1	0.0003537
URBAN	ENDEMIC VS EPIDEMIC	1.4565	1	0.2275

The result for the maximum number of cases during both endemic and epidemic years for Twente Postcodes is shown in Figure 5-8. Weeks for the maximum cases in rural areas during endemic, epidemic years and for urban areas during both endemic and epidemic years were, 25.82 (se=1.09),

32.59 (se=1.59), 25.94 (se=1.70) and 31.90(se=2.11) respectively Table 5-11. There were no differences for weeks of a maximum number of cases during endemic year between urban and rural areas (H=0.02, d.f=1, P=0.9636) and during epidemic years between urban and rural areas (H=0.6204, d.f=1, P=0.4309). It seems to have more fadeouts in rural areas during endemic years than during epidemic years (H=14.8175, d.f=1 P=0.0001184). More so, the disease take a short period of time for its persistence in urban areas during endemic years than during epidemic years (H=4.80 d.f=1, P=0.02843). Results for the comparisons of means are presented in Table 5-12.

Table 5-11 : Means and standard error of the means for the maximum cases in Twente postcodes

	Location	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	25.94	1.70
	Rural	25.82	1.09
Epidemic	Urban	31.90	2.11
	Rural	32.59	1.59

Table 5-12 : Comparison of means for maximum cases of pertussis in Twente postcodes

YEARS/ LOCATION	***COMPARISON***	***KRUSKALL-WALLIS TEST (H)***	***d.f***	***P- VALUE***
ENDEMIC	RURAL VS URBAN	0.002	1	0.9639
EPIDEMIC	RURAL VS URBAN	0.6204	1	0.4309
RURAL	ENDEMIC VS EPIDEMIC	14.8175	1	0.0001184
URBAN	ENDEMIC VS EPIDEMIC	4.8015	1	0.02843

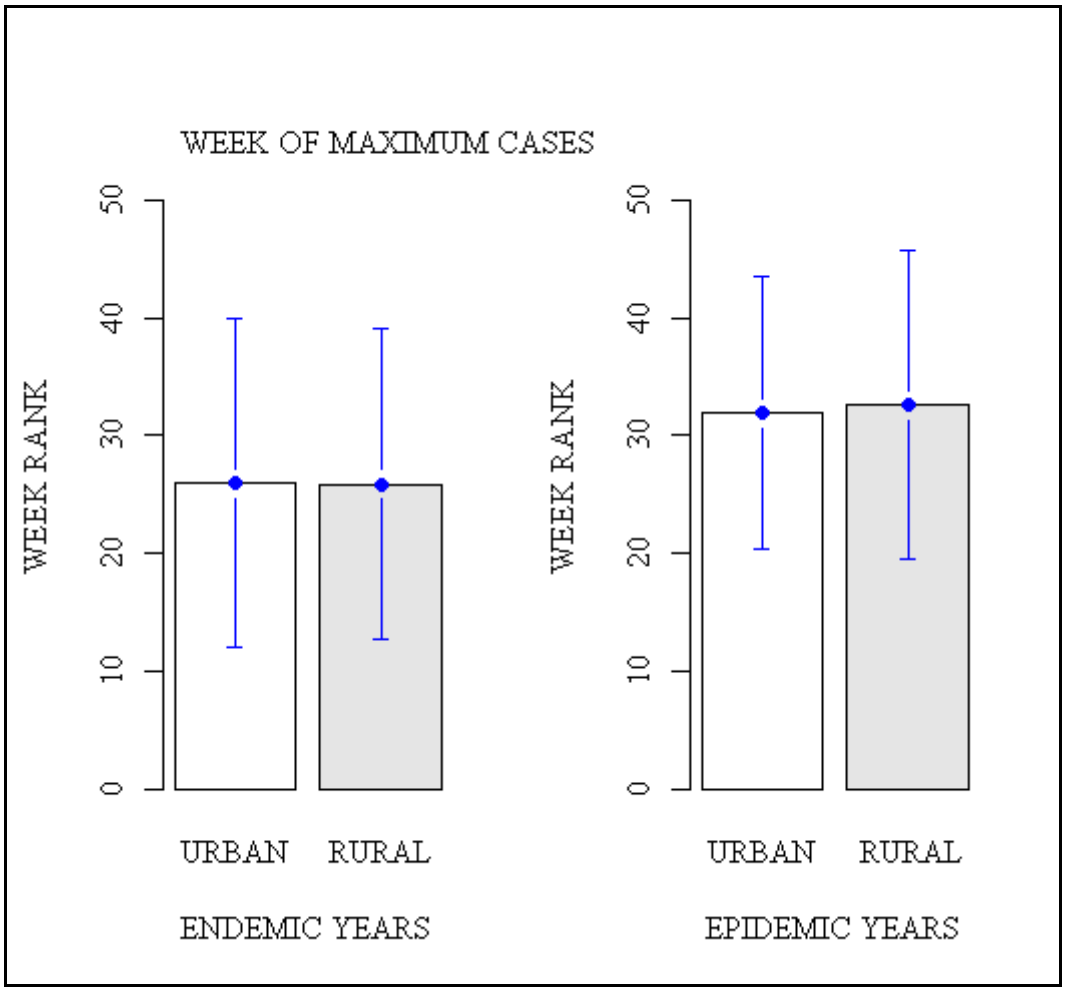


Figure 5-13 : Week rank means for the maximum cases of pertussis in Twente postcodes

5.3.4. Result of Enschede postcodes

The analysis was performed for the Netherlands were repeated for Enschede at postcode level Definition of urban and rural areas was performed as described in section 4.3.2. In this analysis means of the first cases and means of the maximum cases were determined. Results were generated for the years 1993 to 2003. The years were aggregated for epidemic and endemic years.

Results of week rank means of first cases in Enschede area shown in Figure 9-19. Week of first cases in rural areas in both endemic and epidemic years were, 28.14 (se= 2.71) and 25.12 (se= 6.85) respectively. Week of first cases in urban areas during endemic and epidemic years were, 24.96 (se= 2.59), 30.61 (se=3.09) respectively. This result is presented in Table 5-13. Rural areas showed similar week of first cases in both endemic and epidemic year (H= 0.0138, d.f=1, P=0.9065). Similarly, Urban areas there was no significant difference between urban and rural areas (H= 2.1034, d.f=1, P= 0.1470). Moreover, during epidemic years the difference of the mean was not statistically significant. (H= 0.0845, d.f=1, P= 0.7713). Also, dates of first cases in urban and rural areas during endemic year were not statistically significant (H= 0.4814, d.f= 1, P=0.48). The results for the comparison are presented in Table 5-14.

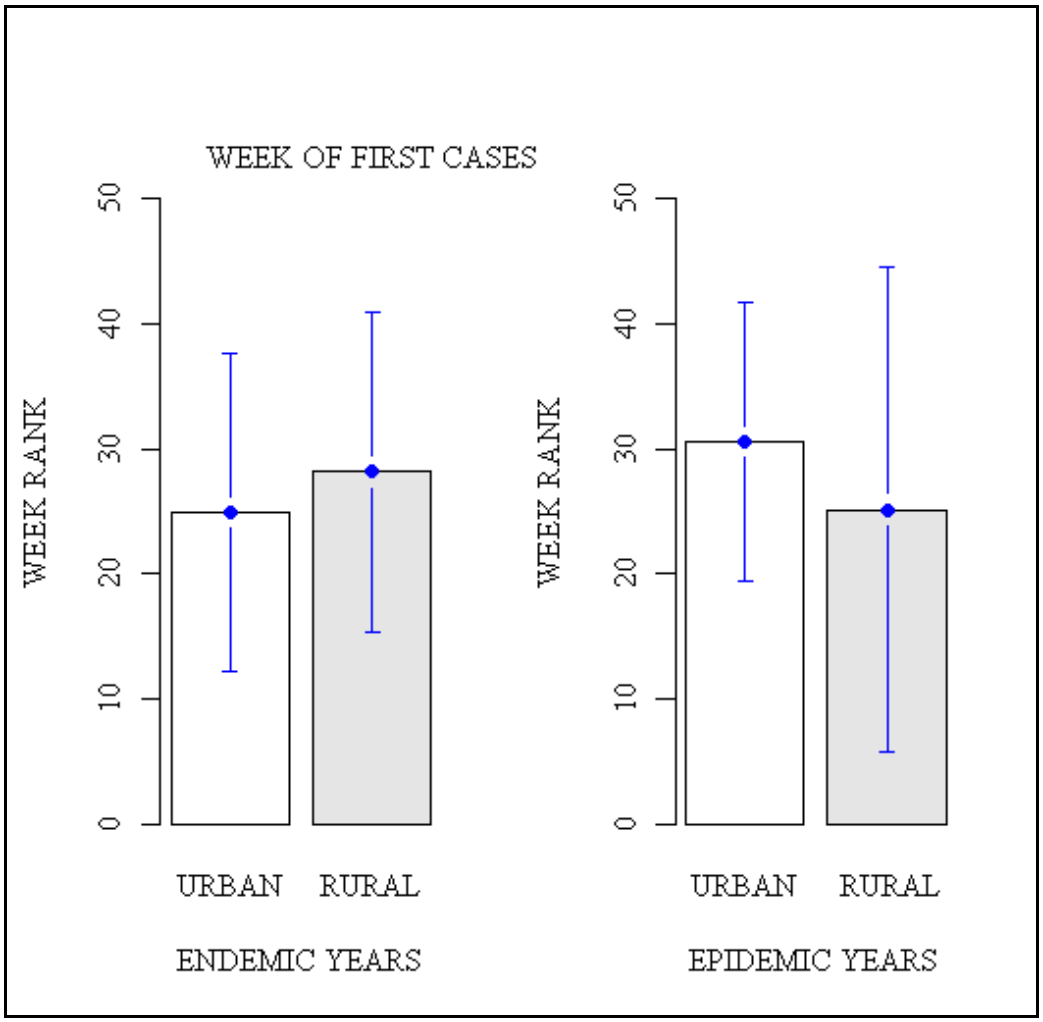


Figure 5-14 : Week rank means for the first cases during the year in Enschede postcodes

Table 5-13: Means and standard errors of means for the first cases in Enschede

Years	***Location***	***Mean***	***Standard error of the mean (se)***
Endemic	Urban	24.96	2.59
	Rural	28.14	2.71
Epidemic	Urban	30.61	3.09
	Rural	25.12	6.85

Table 5-14 : Comparison of means for first cases of pertussis in Enschede

YEARS/LOCATION	*COMPARISON*	*KRUSKALL-WALLIS TEST (H)*	*d.f*	*P- VALUE*
ENDEMIC	RURAL VS URBAN	0.4814	1	0.4878
EPIDEMIC	RURAL VS URBAN	0.0845	1	0.7713
RURAL	ENDEMIC VS EPIDEMIC	0.0138	1	0.9065
URBAN	ENDEMIC VS EPIDEMIC	2.1034	1	0.1470

Week rank means for the maximum cases of pertussis in Enschede postcodes are illustrated in Figure 5-10. Week of a maximum cases in rural areas during both endemic and epidemic years were, 28.14 (se= 2.71) and 25.12 (se= 6.85) respectively. Furthermore, week of a maximum cases in urban areas for both endemic and epidemic years were, 24.96 (se= 2.59) and 34.46 (se=1.77). Table 5-15 shows the results for week of first cases. During endemic year the disease reaches its peaks earlier than during epidemic year (H=6.2616, d.f=1, P=0.01234). There were no difference in rural areas between both epidemic and endemic years (H= 0.0138, d.f=1, P=0.9065). Additionally, week for the maximum number of cases were homogeneous during both endemic and epidemic years between urban and rural areas. That is, during endemic years between urban and rural areas (H= 0.4814, d.f=1, P=0.4878) and during epidemic years between urban and rural areas (H= 0.2975, d.f=1, P=0.5854). Comparison for the means is presented in Table 5-16.

Table 5-15 : Means and standard error of the means for the maximum cases of pertussis in Enschede

Years	*Location*	*Mean*	*Standard error of the mean (se)*
Endemic	Urban	24.96	2.59
	Rural	28.14	2.71
Epidemic	Urban	34.46	1.77
	Rural	25.12	6.85

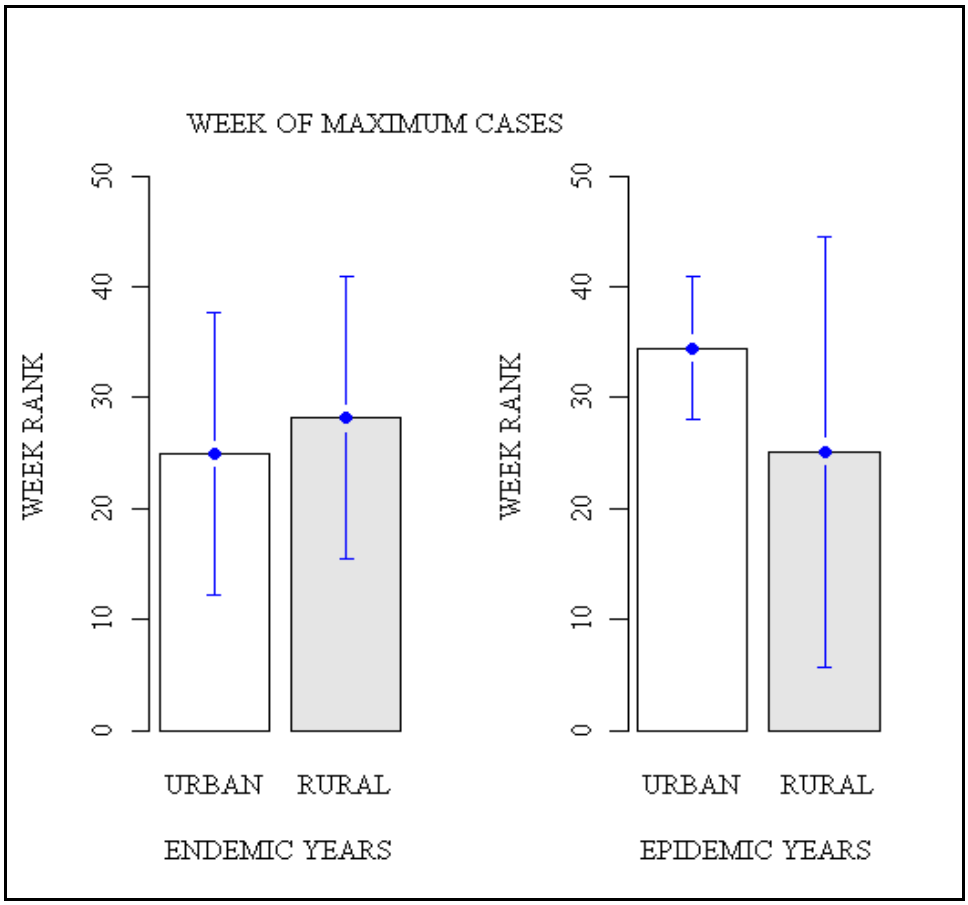


Figure 5-15 : Week rank means for the maximum cases during the year in Enschede postcodes

Table 5-16 : Comparison of means for the maximum cases of pertussis in Enschede

YEARS/ LOCATION	***COMPARISON***	***KRUSKALL-WALLIS TEST (H)***	***d.f***	***P- VALUE***
ENDEMIC	RURAL VS URBAN	0.4814	1	0.4878
EPIDEMIC	RURAL VS URBAN	0.2975	1	0.5854
RURAL	ENDEMIC VS EPIDEMIC	0.0138	1	0.9065
URBAN	ENDEMIC VS EPIDEMIC	6.2616	1	0.01234

5.4. Result for Detection Spatial-temporal clustering patterns

5.4.1. Results of analysing spatial temporal clustering patterns

Analysis in this part was done in the Netherlands at the municipality level over a period of 4 years starting from 1996 to 1999. Ideal plan was to perform the analysis at different scale level starting from the Netherlands up to Enschede postcode levels from 1993 to 2004. But due to time constraints the analysis was performed only in the Netherlands at the level of Municipality. The clusters were detected weekly throughout the year that is around 52 to 53 clusters per annum. From these 52 to 53 clusters we selected statistically significant clusters ($P<0.01$) and numbered them in a chronological order. For instance during 1996 epidemic year, the first statistically significant cluster was observed in week 9 and labelled as cluster number 1, whereas the 2^{nd} and 3^{rd} statistically significant clusters were observed in week 10 and 15 respectively.

For the analysis purpose, if a given statistically significant cluster occurs concurrently at different locations within the same period of study, we labelled each area with additional alphabetical letter. For instance, during 1996 epidemic year statistically significant cluster number 8 that occurs in week 24 was found to affect two different locations, therefore, we labelled them as cluster 8A and 8B to simplify the analysis. Additionally, a detected significant clusters were analysed based on the criteria define in chapter 4 section 4.3.6 and the defined urban and rural map in section 4.3.2. Finally, we describe in detail the most persistence cluster during a year.

During 1996 epidemic year a total number of 18 statistically significant clusters were detected by space time permutation scan statistic as illustrated in Figure 5-16. The result showed that the earliest cluster was detected in week 9 and it is originating from Hengelo urban area. The most persistent cluster starts in week 38 up to week 44**.** The clusters which leads to another cluster or the clusters which occurs in the same location at different time interval include: cluster 1 which starts in week 9, cluster 3 which occurred in week 15, cluster5 occurred in week 20, cluster 7 identified from week 22 to week 23, cluster 8A identified from week 24 to week 25 and cluster 11 and 13 which occurred in week 33 and week 38 respectively (Figure 5-17). The clusters which were detected within a single study period and dies out is cluster 15 identified from week 39 to week 40. Cluster 8 detected in week 24 occurred at different location and was classified as cluster 8A and cluster 8B. Also cluster 9 identified in week 24 to week 25 occurred in different location and it was classified as cluster 9A and 9B.

Cluster 1 and cluster 2 , cluster 5 and cluster 6, cluster 7 and cluster 8B and cluster 9B, cluster 8A and 9A cluster 11 and cluster 12, cluster 11 and cluster 14 and cluster 16 were detected in the same spatial location but different time period respectively. Cluster 3 touches the boundary of cluster 4 and cluster 9A which detected in week 24 to 25 is close in location and time with cluster 10 which was detected in week 27 to 28. The properties and description of the classified clusters are presented in Appendix 1.1 and 1.2 respectively.

Clusters which are persist in location for more than 3 weeks include: cluster 7 which originated in rural areas and it persists in a location for 4 week starting from week 22 to week 25. This cluster was detected with 12 numbers of observed cases. Cluster 8A was also persistence since it exists for 5

weeks in location starting from week 24 to week 28 of the year. This cluster originates in rural area and it was detected with 34 observed cases. The last persistence cluster is cluster 13 which persists in a location for 7 weeks. This cluster was also originated in rural area and it was detected with 13 observed cases. Moreover, fadeout clusters were cluster 1, cluster5 and cluster 11 which exists for only 2 weeks, cluster 3 which takes 3 weeks and cluster 15 and 18 which takes only one week in a location. Leave

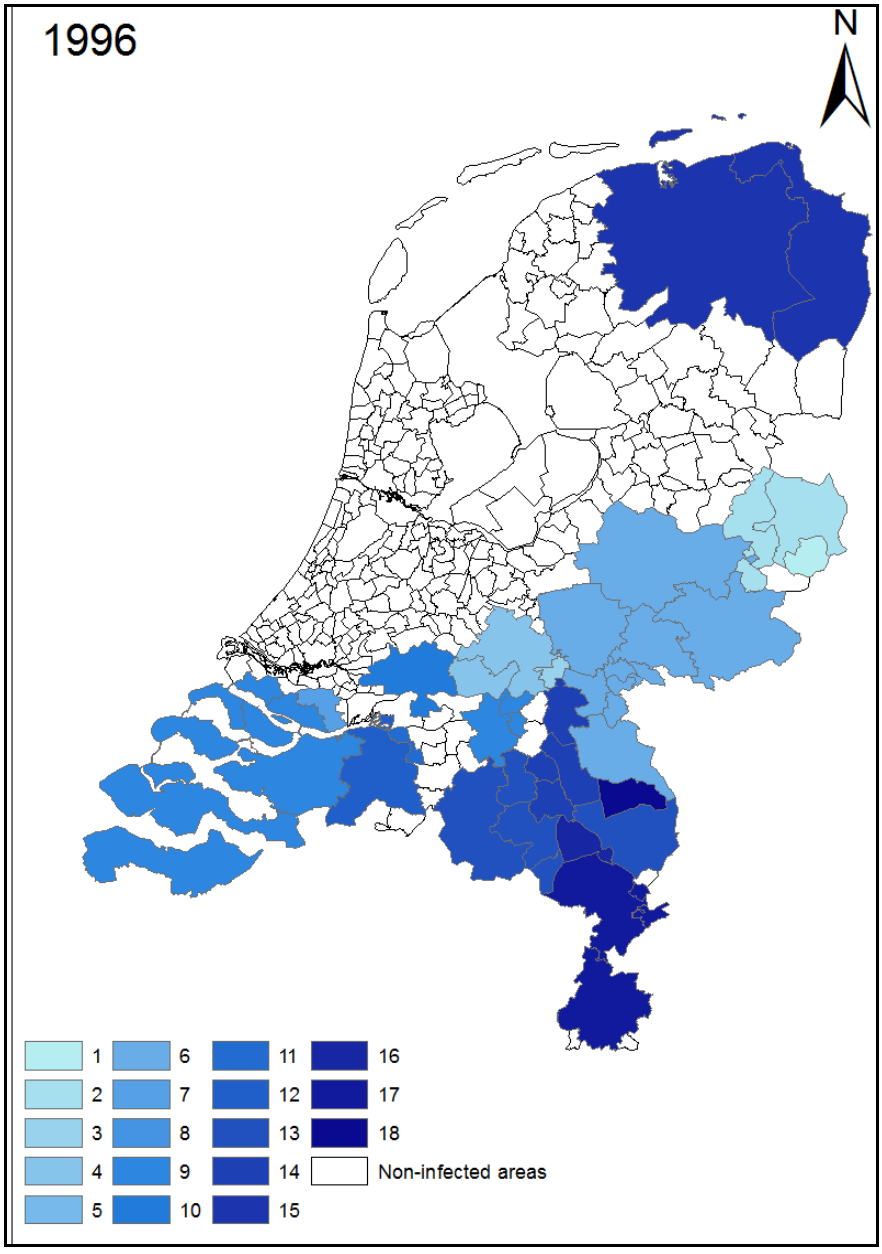


Figure 5-16 : Spatial temporal clusters during 1996 epidemic year. The number in the legend refers to the number of the clusters and it follows a sequence of time in order of its appearance

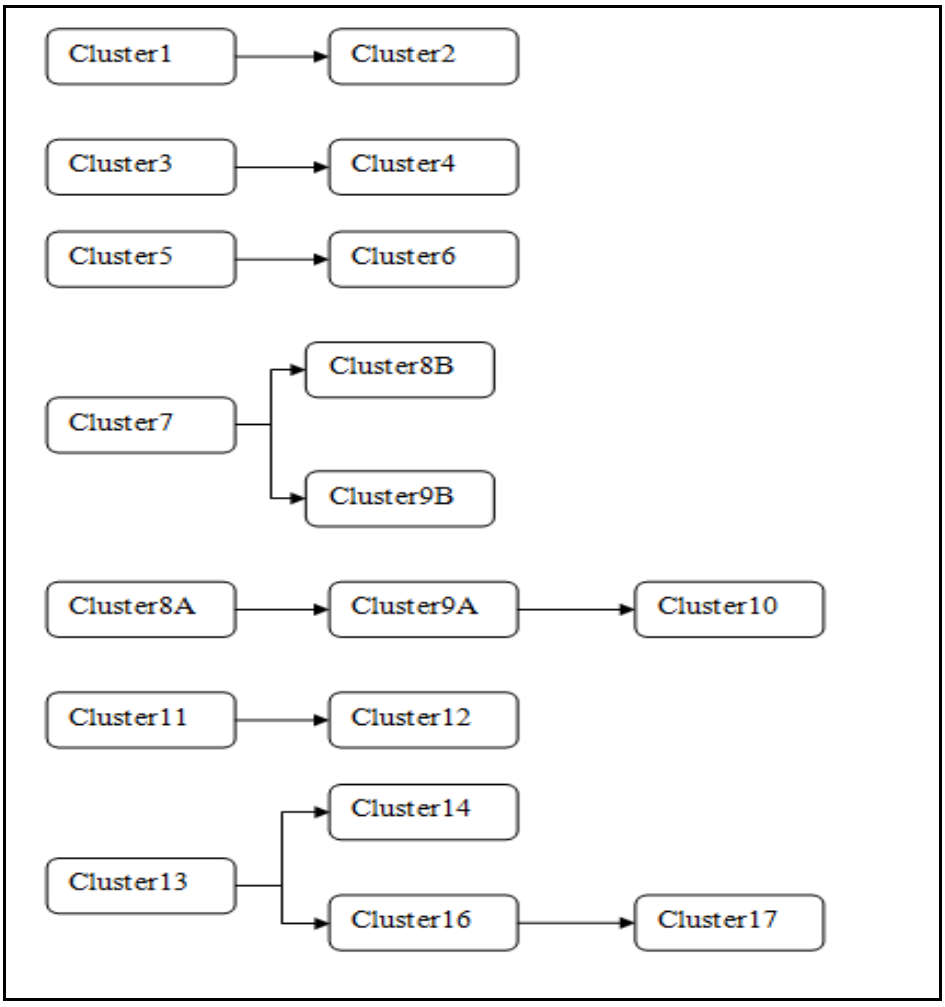


Figure 5-17 : Fadeout and persistence clusters during 1996 epidemic year.

The detail description of cluster 13 is illustrated in Figure 5-18. This cluster originated in rural area was detected in week 38 with 119 cases observed. Between weeks 38 to 39 the number of observed cases decreased to 65 and the coverage area also decreases. From week 40 to 43 the cluster was also observed to another nearby location with a total number of 22 cases. In week 43-44 the number of observed cases increased to 34 cases.

During week 47 the cluster was also observed again in the same location but different time period hence this cluster was classified as fadeout cluster.

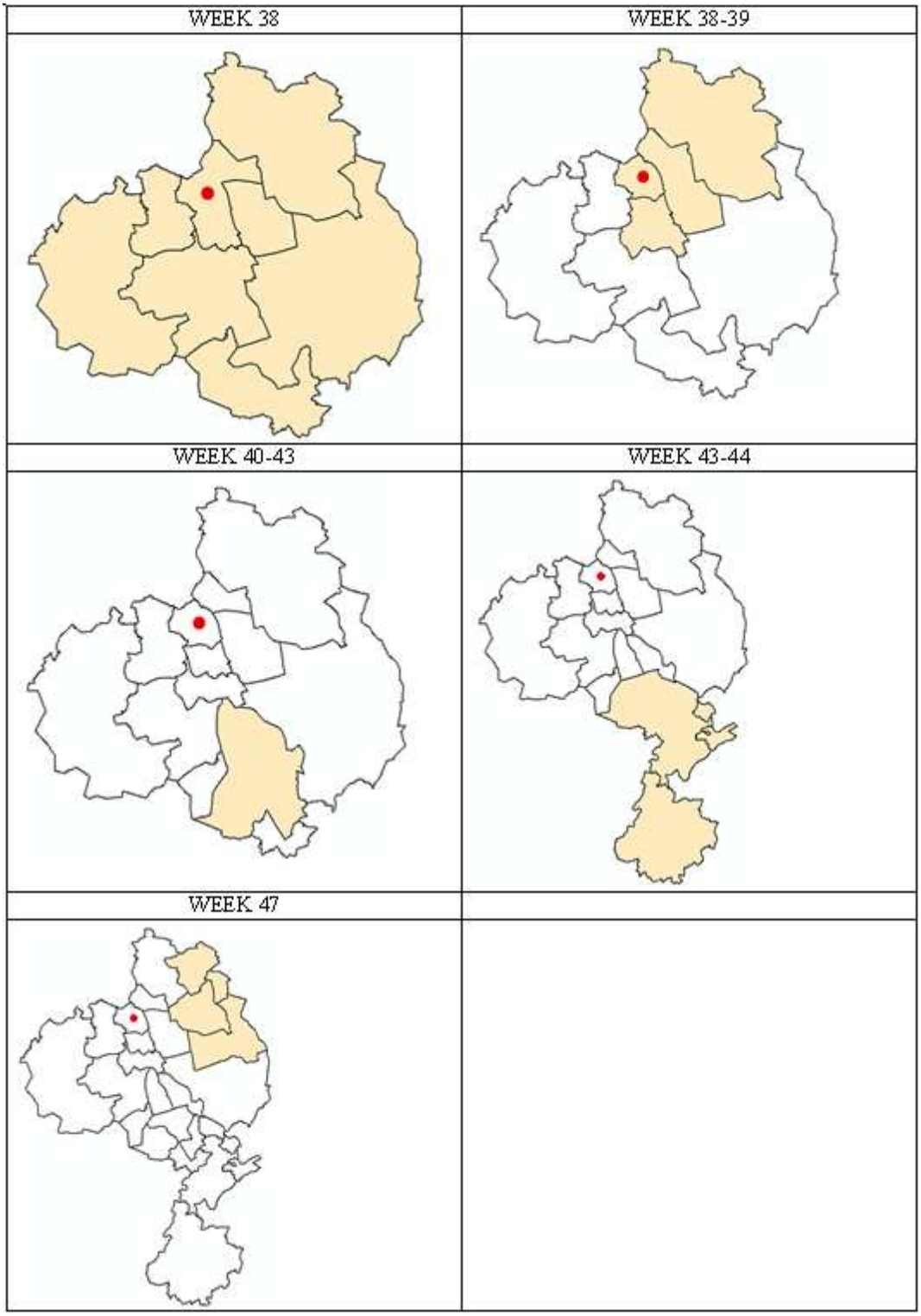


Figure 5-18 : Details of persistent cluster 13 during 1996. Red round is origin of a disease

Result of clusters detected during 1997 endemic year is illustrated in Figure 5-19. A total number of 21 significant clusters were detected. The first cluster was detected from week 53 1996 to week 1 1997. The clusters which occurs in different location but within the same time period include cluster 1 identified from week 53 1996 to week 1 1997. This cluster was classified as cluster 1A and cluster 1B, cluster 7 detected in week 16 and classified as 7A and 7B. More so cluster 11 was detected from

week 18 from week 18 to week 20 and classified as 11A and 11B, cluster 13 was identified in week 23 and classified as cluster 13A and cluster 13B and cluster 14 was identified from week 22 to week 24 and classified as 14A and 14B. Clusters which leads to another cluster include cluster 1A which occurred from week 53 to week 1, Cluster 4 which was identified in week 5, cluster 9 identified in week 16 to week 18 and cluster 12 which was detected from week 21 to week 22. The illustration of these clusters is shown in Figure 5-20. Clusters which were identified for only a single period of time and disappear include Cluster 3, cluster 20 and cluster 21 detected in week 3-4, week46 and week 42-43 respectively.

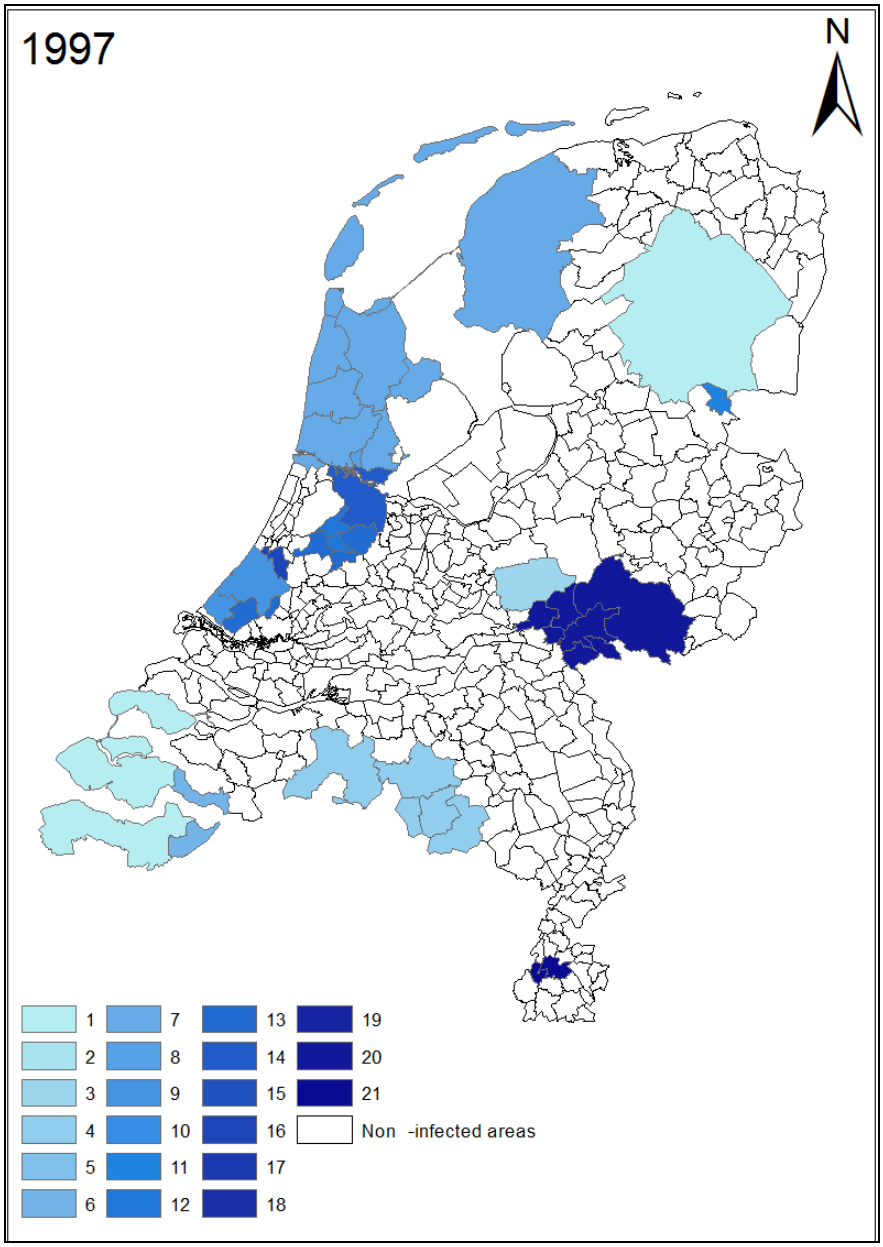


Figure 5-19 : Spatial temporal clusters during 1997 endemic year. The number in the legend refers to the number of the clusters and it follows a sequence of time in order of its appearance

Furthermore, the clusters which were identified in the same spatial location includes cluster 1A, 2 and 6, cluster , cluster 1B and cluster 11A ,7A, 8 and 19, cluster 9 , 10, 13B, 14B, 16, 17 and 18 , cluster 12, 13A, 14A and 15 and lastly is cluster 7B and 11B . Clusters which shows the persistence

in space and time includes cluster 9, 10, 13B, 14, 16, 17 and cluster 18. The second persistence cluster includes cluster 12, 13A, 14A and cluster 15. The properties and description of the classified clusters are presented in Appendix 2.1 and 2.2 respectively. The most persistence cluster is cluster number 9. Clusters which originated from urban areas include cluster 7A, cluster 9 and cluster 19. Clusters which originate from rural areas include clusters 1A, and 1B, 3, 4, 6, 7B, 11A, 11B, 12, 20 and cluster 21.

From this result there was 2 persistence clusters. The first diffusion cluster is cluster 9 which was identified in week 16 to week 18 and it originates in urban areas. This cluster exists in a location for 17 weeks. Another persistence cluster is cluster 12 which was identified in rural area and it persists in a location for 5 weeks. From this result the most persistence cluster is cluster 9.

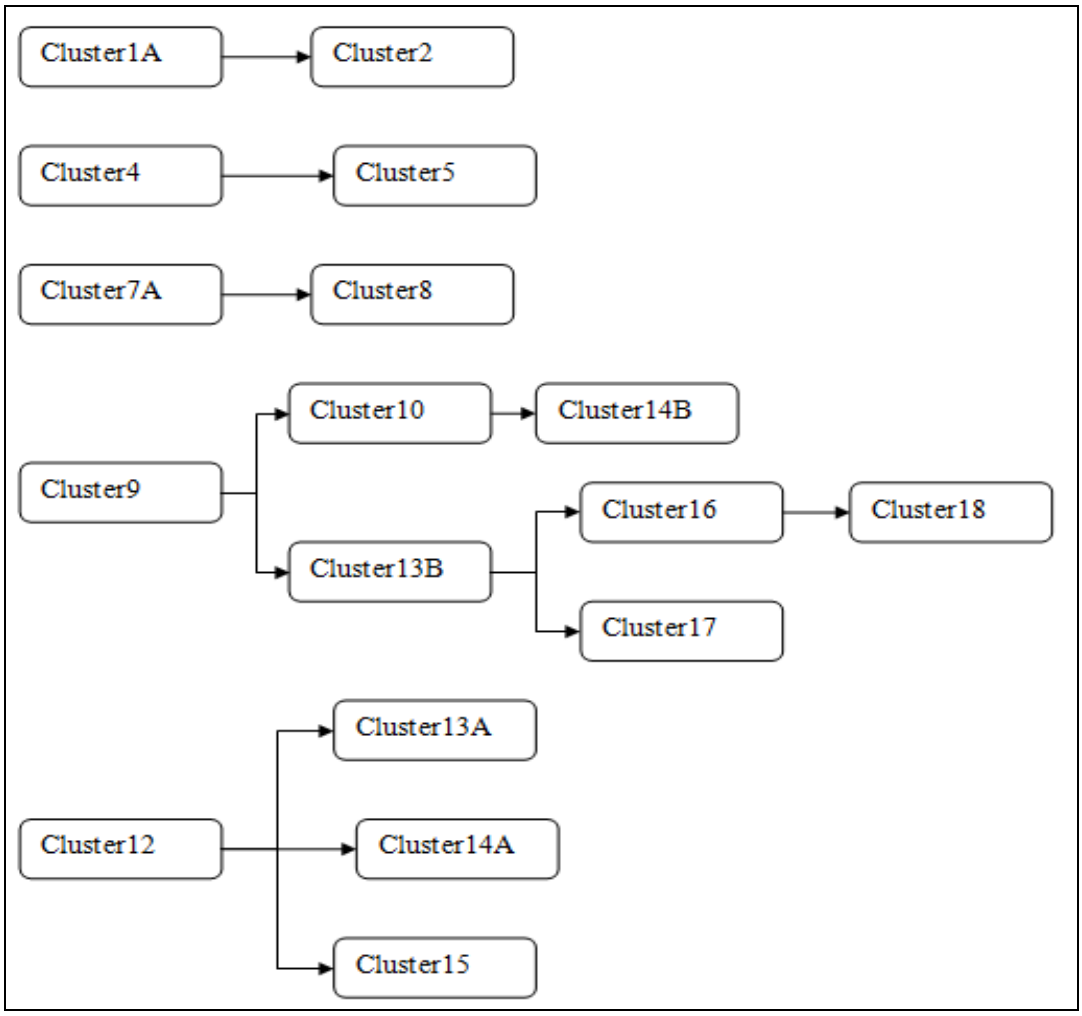


Figure 5-20: Clusters which overlaps in both space and time during 1997 endemic year

Figure 5-21 illustrate how cluster 9 was persisting in a location. This cluster was first detected from week 16 to week 18 with 32 observed cases. In the same location the cluster was again detected from week 17 to week 19 with a total number of 34 observed cases. In week 20 nothing was detected in the location but from week 21 to week 23 the cluster was also detected in the same location leaving out few areas with a total number of 18 cases. The cluster showed to persist in the location with 38 observed cases within week 22 to week 24. Further, from week 25 to week 27 the area size of the infection and the observed cases decreased to 22 cases. From week 26 to 28 some of the area which

was not infected in the previous week showed infection other areas showed uninfected. The total number of cases in this location was maintained to 22 observed cases. No cases observed in this area in week 29 but again there were 26 observed cases between weeks 30 to 32. Starting from this date up to the end of the year no any cases was observed in this location.

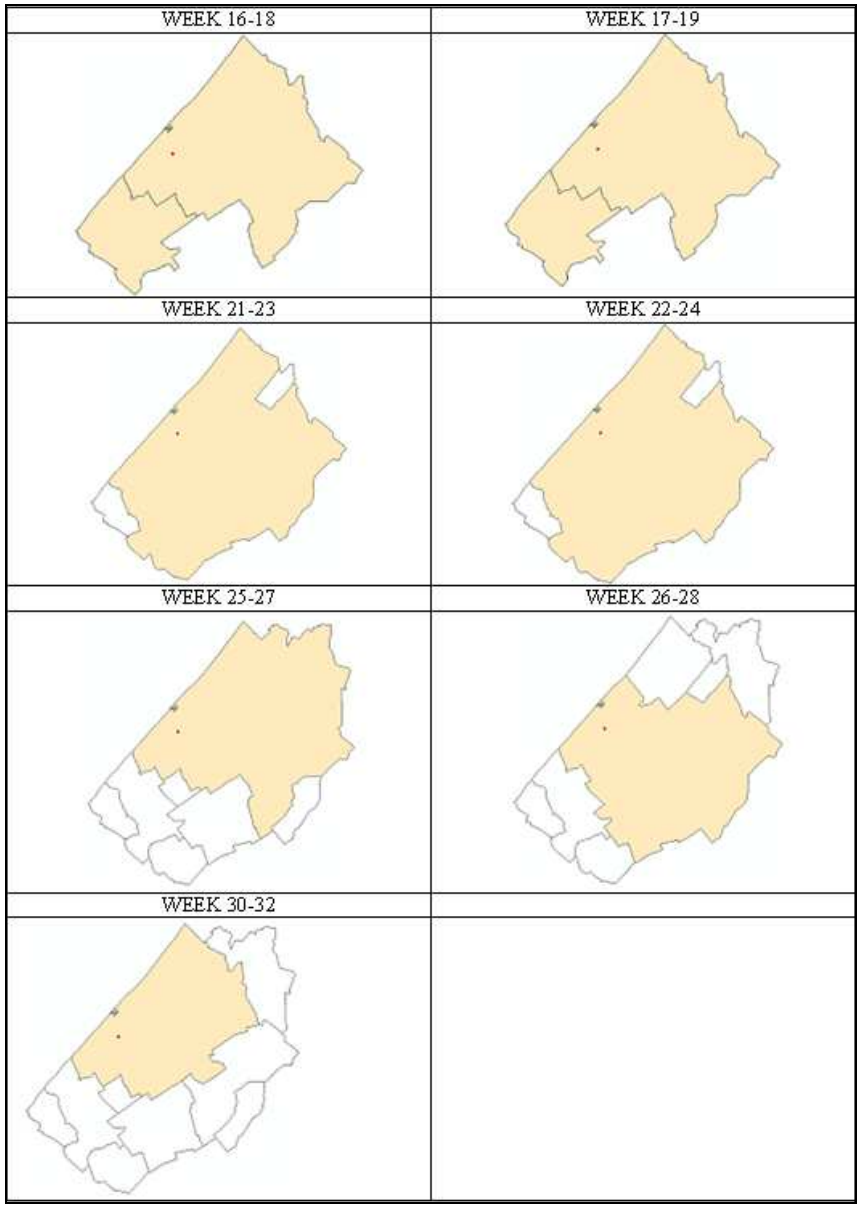


Figure 5-21 : Detailed description of cluster 9 during 1997 endemic year. Red dot is the origin of the disease.

The result of clusters detected during 1998 is presented in Figure 5-22. This year consists of a total number of 13 clusters. The earliest cluster was detected in week 5. Clusters which were identified within the same study period but different location include cluster 3 identified between week 25 to week 27 and classified as cluster 3A and cluster 3B. Cluster 5 detected between week 34 to 36 and classified as cluster 5A and 5B. cluster 10 which was identified between week 39 to 41 and classified as cluster 10A and 10B. Clusters which leads to other clusters include cluster 3B, cluster 5A and cluster 10A Figure 5-23.

Clusters detected in one location for a different period of time include cluster 1, 5A, 6, 7, 8, 9, 10B, 12, and 13A, cluster 3 and 4, cluster, cluster 11 and 10A. Clusters which was identified only once throughout a year includes cluster 2 and cluster 5B. All cluster properties and cluster description is presented in Appendix 3.1 and Appendix 3.2 respectively, 3 clusters were found to originate in urban areas and 5 clusters originated in rural areas. Clusters which originate from rural areas include clusters 1, 2, 3A, 3B and cluster 13A. Clusters which seem originate from urban areas include cluster, 5A, 5B and cluster 10B. Of these clusters cluster 1, 2, 3A, 3B, 5B, 10A and 13B resulted into fadeout due to their short period of existence. Cluster 5A is the most persistence cluster it takes a long time period in a location it starts from week 34 up to week 45 and it originated in urban area.

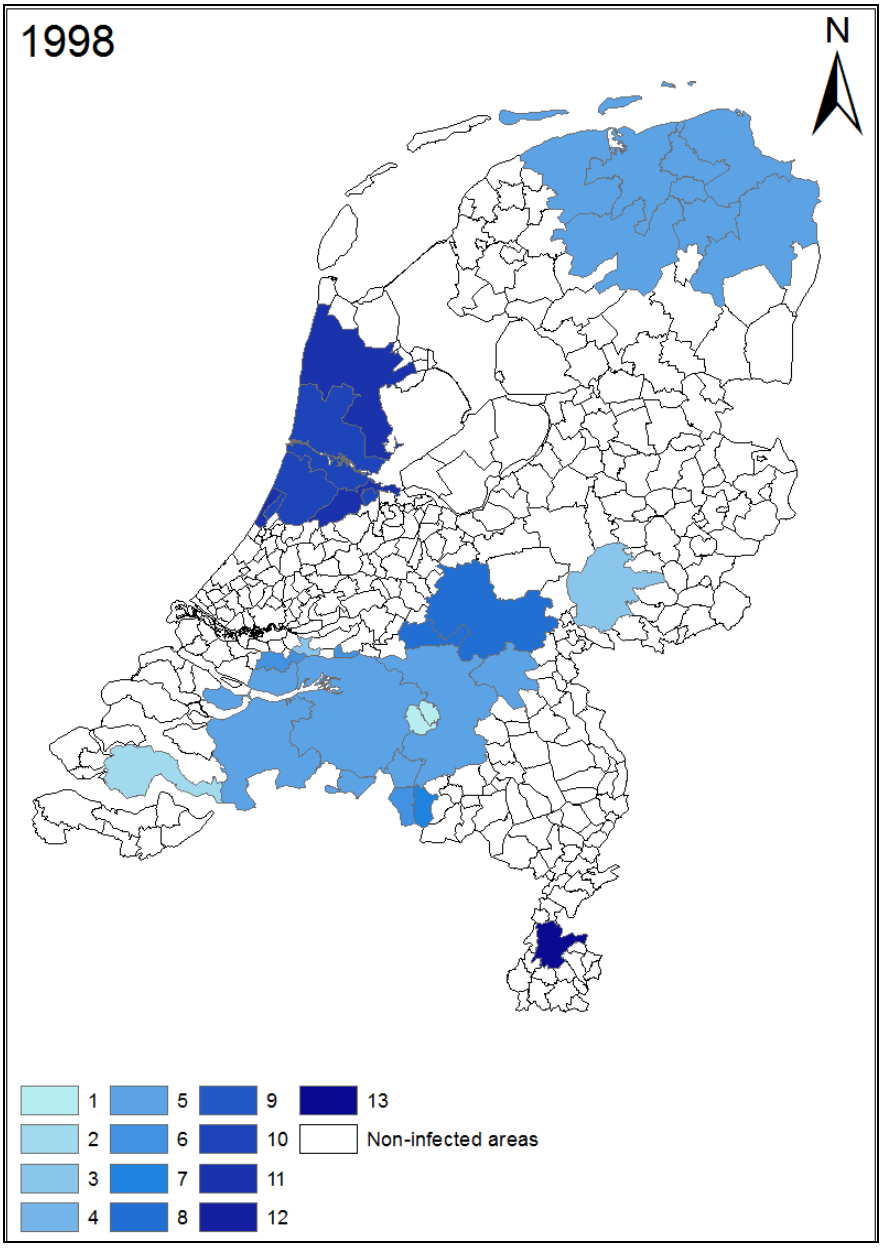


Figure 5-22: Spatial temporal clusters during 1998 endemic year. The number in the legend refers to the number of the clusters and it follows a sequence of time in order of its appearance

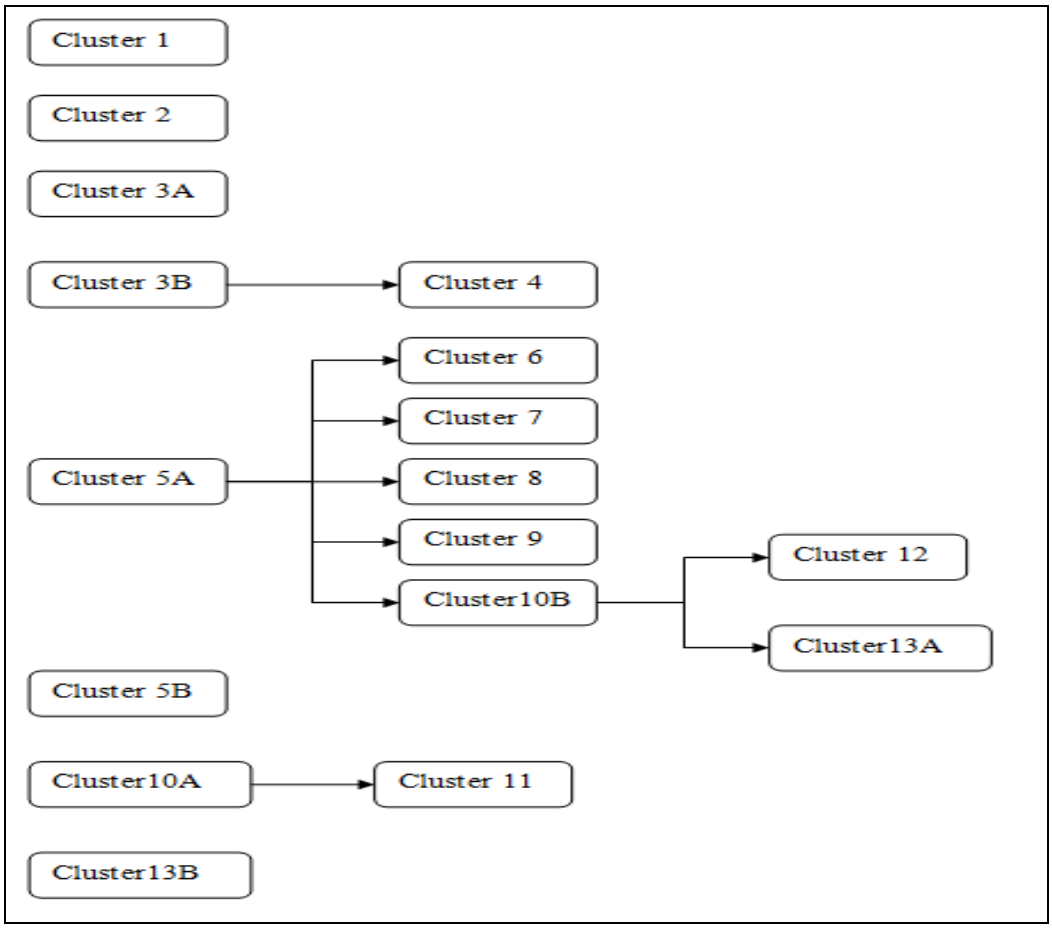


Figure 5-23 : Relationship of clusters during 1998 endemic year.

Figure 5-24 showed the detailed persistence of cluster 5A starting from its origin. Cluster 5 firstly detected in week 36 with 36 numbers of observed cases and 4.23 relative risks at the radius of 21.51 I. Afterwards, in week 37 the number of observed cases were 34 and relative risk of 4.92 with radius of 15.48. Then in week 38 the number of observed cases were 19 and relative risk of 4.47 within the radius of 24.99. In week 39 still this cluster was there with 10 cases observed and relative risk of 11.68 within a radius of 9.81. Moreover, in week 40, number of cases observed in that cluster was 23 with relative risk of 3.77 within a radius of 23.45. In week 41 numbers of cases was 28 a radius of 23.45. Again in week 45 the cluster was also detected with 21 number of cases and relative risk of 3.09 within and radius of 20.35. This cluster exits until week 46 then it fadeout.

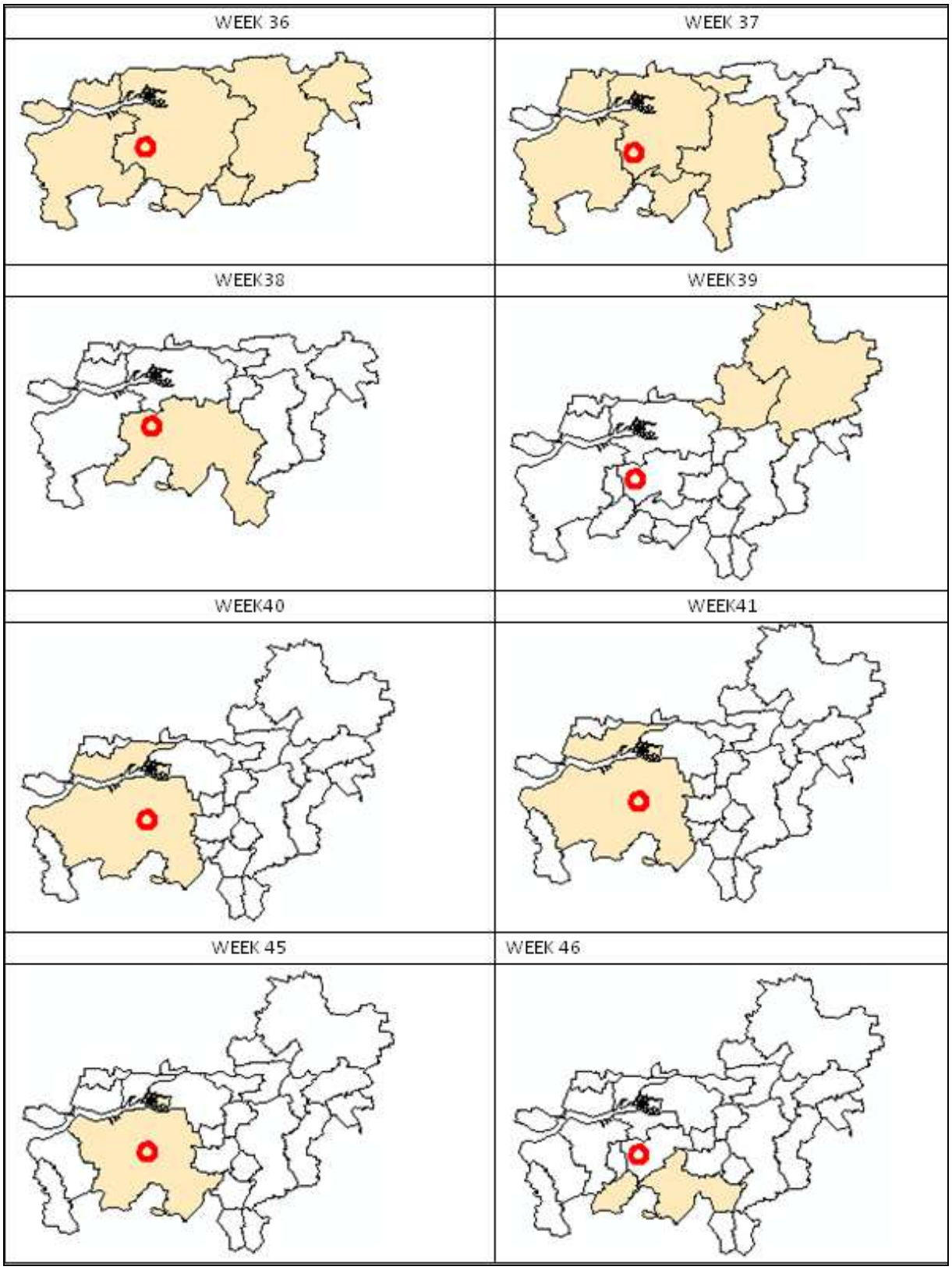


Figure 5-24: Detailed description of cluster 5A during 1998 endemic year. Red circle is the origin of the disease.

The result of detected spatial temporal clusters during 1999 epidemic year is illustrated in Figure 5-25. In this year a total number of 35 statistically significant clusters were detected. The first earliest cluster was identified from week 2 to week 3. Clusters which were detected at different location but

within the same period of time include cluster 15 which occurred at two different spatial location detected from week 26 to 27. This cluster was classified as cluster 15A and 15B. Cluster 18 detected from week 30 to 31 was also identified at two different locations and it was classified as cluster 18A and 18B. Cluster 19 detected from week 30 and week 32 and it was classified as cluster 19A and cluster 19B. Cluster 22was detected from week 34 to week 35 in three different spatial locations and it was classified as cluster 22A, 22B, and 22C. Cluster 23 was detected from week 35 to week 36 and it was classified as cluster 23A and 23B. Cluster 24 was detected from week 36 to week 37 and it was classified as cluster 24A and 24B. Cluster 25 was detected from week 38 and it was classified as 25A and 25B. Lastly cluster 27 was also detected in two different spatial locations from week 40 to 41 and it was classified as cluster 27A and cluster 27B. Properties of cluster and their description are presented in Appendix 4.1 and Appendix 4.2 respectively.

Clusters which was detected in the same spatial location in a different period of time include, clusters2, 3, 5, 15B, 20, 22C, 23B, 24B and cluster 25A which were detected in the Northern part of the Netherlands. Clusters 15A, 14, 17, 27B, 19A, 23A, 28, 18A, 24A, 22A and 25B were detected in the Southern part of the Netherlands. Additionally, clusters 6, 13, 12 which appear to occur in the same spatial location but different time periods were detected in the Western Southern part of the Netherlands. Clusters 33, 22B, 34, 33, 19B, 19C, 26, 30, 8, 21, 10, 11, 7, 18B and cluster 9 were detected in the same location but different time period in the Western part of the Netherlands. Also cluster 29 and 35 overlap in location but different time period were detected in the Eastern part of the Netherlands and clusters 31, 32 and 16 were detected in the Northern East part of the Netherlands. Cluster 4 was detected in week 9 and disappears for the whole period of time. Cluster 29 was detected in week 43 and reoccurs again in week 51.

Clusters which lead to another cluster include, cluster 1, cluster7, cluster 10, cluster12, cluster 14, cluster 18B, cluster 20 and cluster 26. Clusters which originate from urban area include, cluster 1, cluster2, cluster 4, cluster 7 and cluster 18B.

From these clusters, Cluster 14 and cluster 26 resulted to persistence cluster. Cluster 14 which originated in rural area was initially detected in week 25 with a total number of 110 numbers of cases and it exists until week 37. The duration of this cluster is 13 weeks. Cluster 26 which also originated from rural areas was initially detected in week 40 with a total number of 42 cases and it exists until week 49 and counted the duration period of 10 weeks.

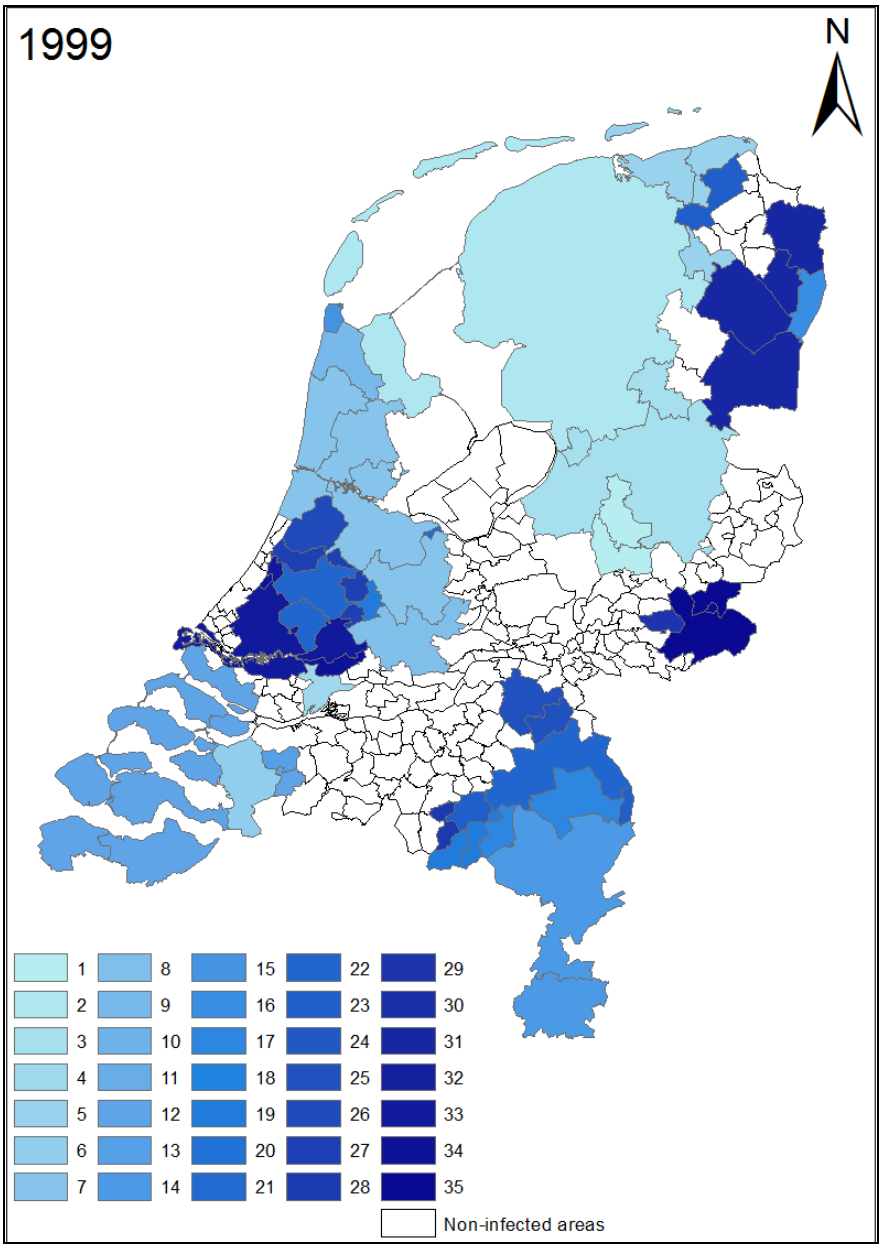


Figure 5-25 : Spatial temporal clustering patterns during 1999. The number in the legend refers to the number of the clusters and it follows a sequence of time in order of its appearance

Further, persistence cluster were classified as those clusters that takes not more than 8 weeks in the location without fadeout. Therefore, cluster 1, cluster 7, cluster 18B and cluster 20 resulted to persistent clusters. Cluster 1, and cluster 7exists in a location for 5 weeks whereas cluster 18B and cluster 20 takes 7 weeks in a location. The remaining clusters were classified as fadeout since they appear in a location for only short period of time (less than 3 weeks) weeks and dies out. Figure 5-26 illustrate how persistent and diffusion clusters are connected

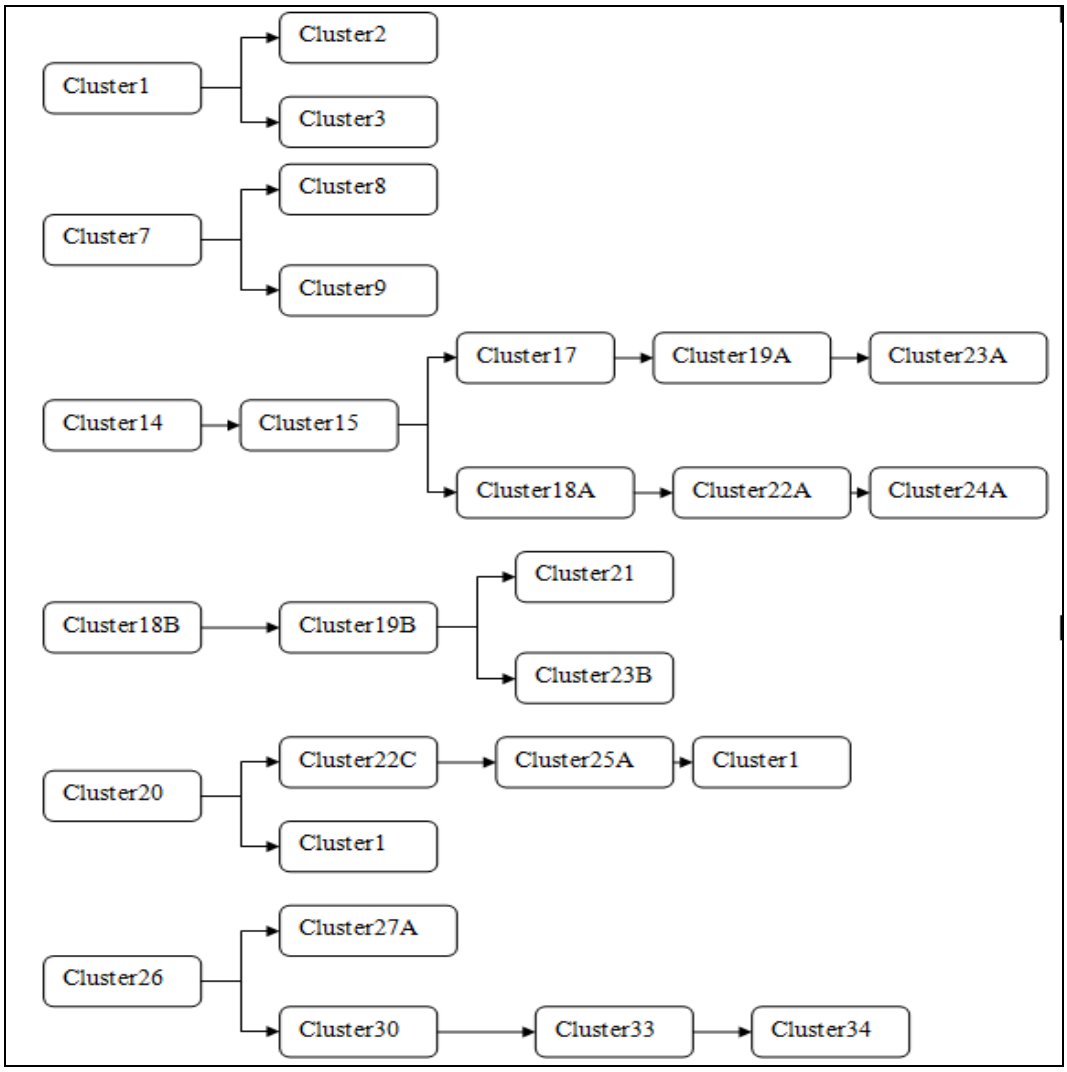


Figure 5-26 : persistent clusters during 1999 epidemic year

The persistent cluster is illustrated in Figure 5-27, 5- 28 and 5-29. This cluster was originated from urban area and detected between week 14 – 15 with a total of 72 numbers of observed cases. Between week 15- 16 the cluster was again detected in a location with 118 number of observed cases. Then between week 17 – 18 in the same location the cluster was also detected with a total number of 13 observed cases and it fades out. In week 31 the cluster was again detected in the same location with 18 number of observed cases. This cluster persists in a location until week 36 and then fadeouts. In week 40 the cluster occurs again in the same location with 42 cases. This cluster persist up to week 49 and it dies out.

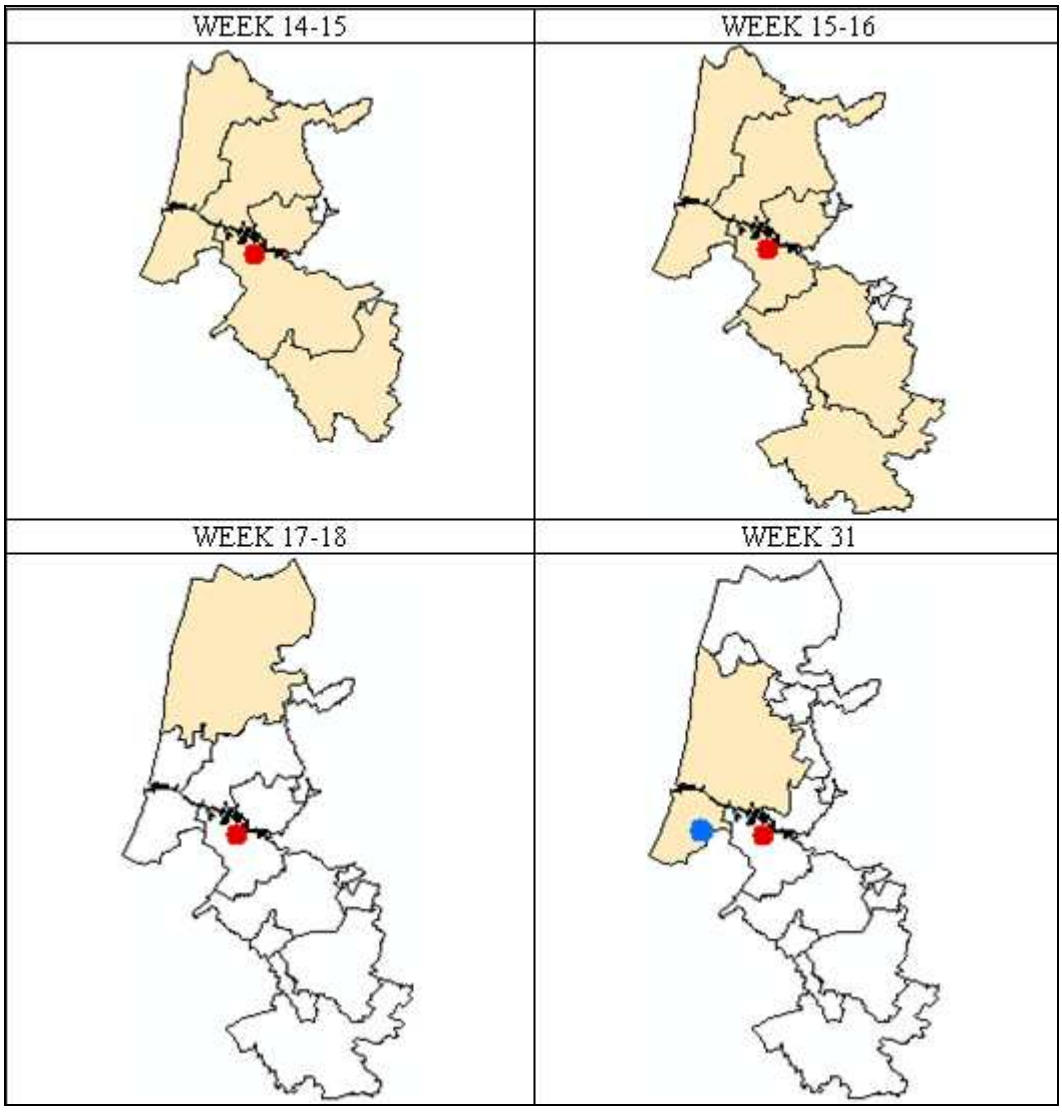


Figure 5-27 : Detailed description of cluster 7, 8, 9 and cluster 18B. Red dot is the first origin of the disease and blue dot is the second origin.

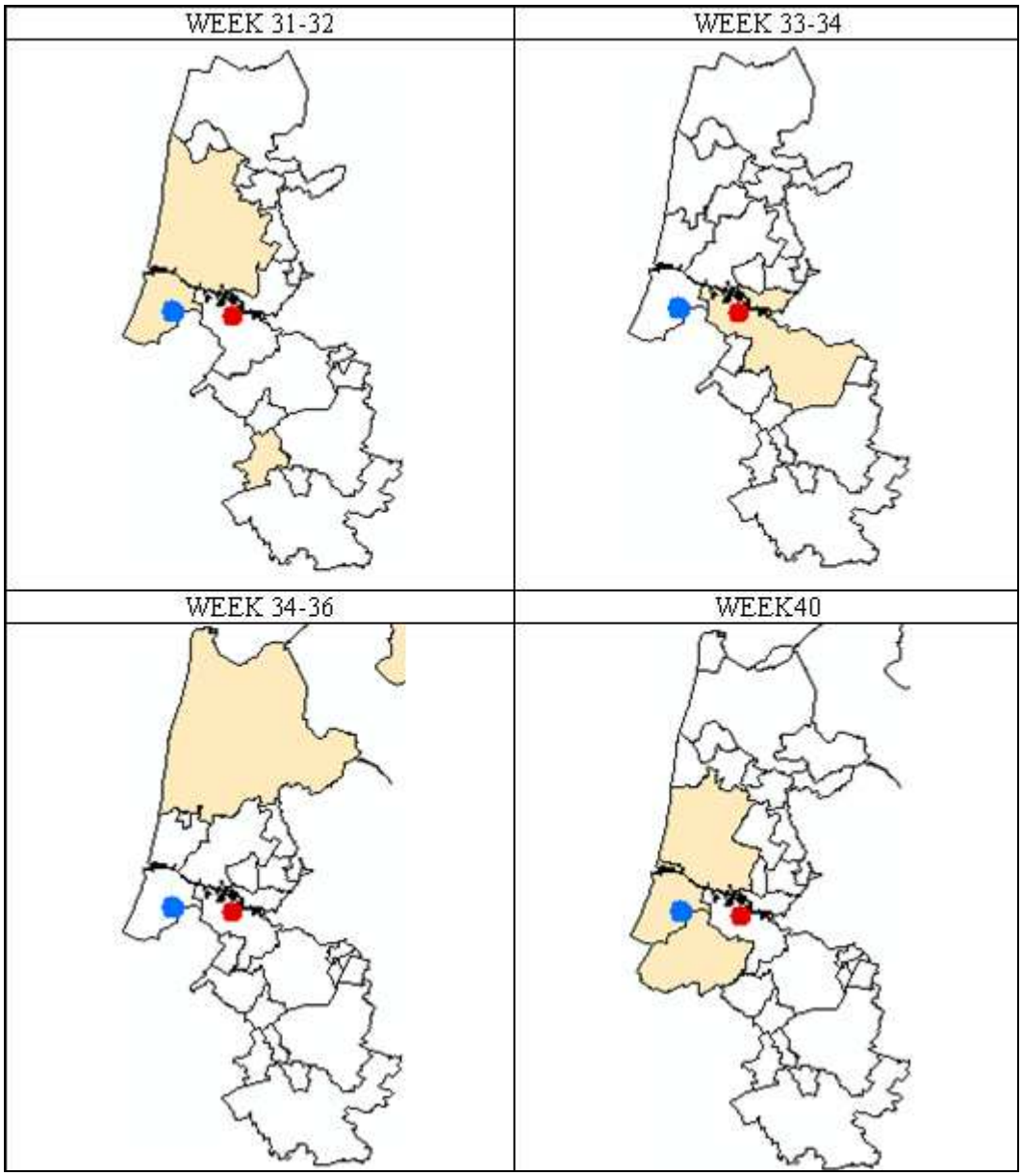


Figure 5-28 : Detailed description of cluster 19B, 21,23B and 26. Red dot is the first origin of the disease and blue dot is the second origin.

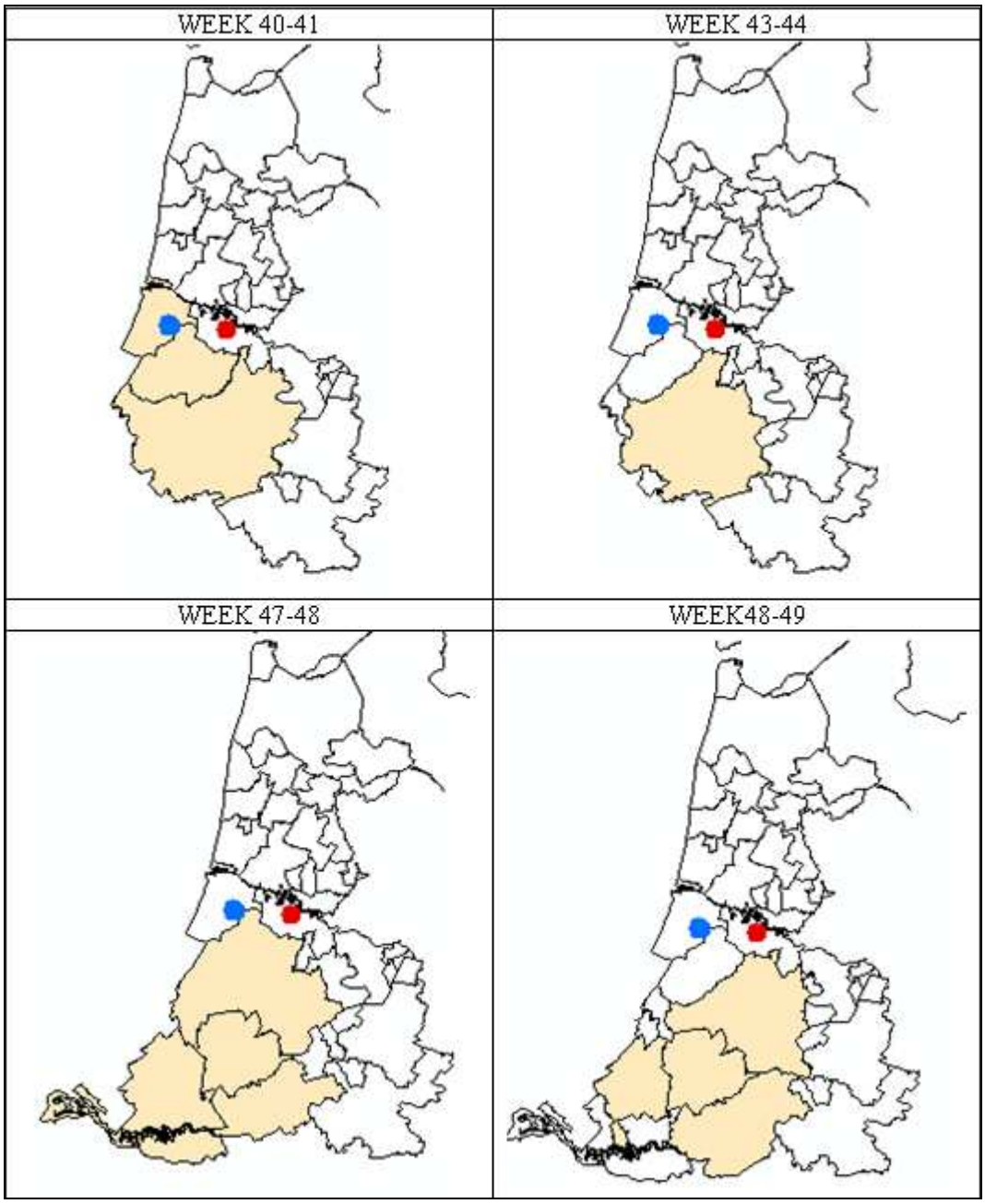


Figure 5-29: Detailed description of cluster 27A, 30, 33 and 34. Red dot is the first origin of the disease and blue dot is the second origin.

5.5. Summary

In this chapter the result of the analysis was presented. Synchrony and travelling waves for the European countries was analysed and found that there was no global synchrony. Week rank method identifies the hierarchical spread of a disease in the Netherlands at the municipality. Moreover, the result of the cluster detected by the space-time permutation statistic showed the persistence of disease more in rural areas than in urban areas.

6. Discussions, conclusion and Recommendations

6.1. Introduction

The objective of this research was to detect patterns in empirical pertussis data that can be used to validate agent-based simulations using a pattern-oriented approach. Based on literature research, different types of spatial, spatial-temporal and temporal patterns were identified that occur in infectious diseases like pertussis. Patterns that were identified included “hierarchical spread”, “synchrony” and “travelling waves”. The relation between the spatial and temporal scale on which these patterns can be observed and the statistical methods available to detect them has been made.
Methodological approach applied was to analyze at different spatial scales (Europe at country level, Netherlands at municipality level, Region (Twente) at municipality and postal code level, and Enschede at postal code level.
The following checks have been performed:

- A check on “synchrony” at the European level (do European countries experience epidemic outbreaks during the same years?)
- A check on “travelling waves” at the European level (does the spread of pertussis show a wave like spread from one side of Europe to another?)
- A check of “Hierarchical diffusion” on the spatial scale of the Netherlands using the week-rank methods
- A check on “Hierarchical diffusion” on the spatial scale of the Netherlands using a combination of spatial scan statistics with a further analysis of the observed clusters at the spatial scale of the cluster

Besides the methods as described above, throughout the analysis difference between epidemic years and endemic years were identified in order to determine if differences in patterns between these groups of years can be detected.

6.2. Synchrony and traveling waves

We performed a check in European countries to find if all countries have epidemic simultaneously and if epidemic is usually travelling to the nearby countries. The synchrony was identified but not for all European countries. Also the observed synchrony was not always for every year. Similar pattern was observed by Helene Broutin, et al [5] who performed a comparative analysis of pertussis time series in 12 countries.

Furthermore, travelling waves of epidemic was observed only for some of the countries not all. This could be due to difference vaccination strategies among the countries. The reason for this could also explained by the method used. It is possible that some of the countries were regarded as endemic

while they are epidemic. Incompleteness of data for some of the countries could also be a reason to these results. For instance, countries close to the Netherlands France and Germany there were no data so we can not say anything about this.

6.3. Week rank method

Week rank method was used to detect hierarchical diffusion pattern in the Netherlands, Twente region and Enschede city. In the Netherlands analysis was performed in 538 municipalities, in Twente region the analysis was performed in 22 municipalities and 119 postcodes and In Enschede the analysis was performed in 22 postcodes.

In the Netherlands at municipality level of analysis the method identify clearly the hierarchical diffusion between endemic and epidemic year. During endemic years the disease arrive sooner in urban areas than in rural areas i.e. approximately week 12 against week 19 as well during epidemic years pertussis seems to arrive sooner in urban areas than in rural areas i.e. approximately week 8 against week 19. This pattern may be explained by population density or population size of the area. Pertussis is transmitted by contact so the area with dense population is more likely to be affected by the disease than the area with sparse population density or population size. A similar pattern was also reported by Broutin et al [20] who used pertussis data in a small region of Senegal.

In Twente region at the level of municipality there was a difference in urban and rural areas between both epidemic and endemic years but this difference was not statistically significant. In addition to that, in Twente region at the postcode level during epidemic year we also observed a difference but not statistically significant. This result might be explained by the fact that this method requires a lot of data in order to produce appropriate results since the test statistic used to test for the significant difference is unable to detect small differences in the result [20] . The result at this level of Twente region could be improved by doing the analysis using only one year instead of using multiple years for both endemic and endemic. Since we used only small areas, it is not obvious that this area is affected by the disease every year. In addition, pertussis is a vaccine preventable disease so it is possible that in some of the year vaccination of that area is high so no disease affects that area. In Enschede at the level of the municipality the method does not detect any difference between urban and rural areas. The Area is not large enough to detect hierarchical diffusion. The spatial extent of the region is more suitable for this type of analysis.

Furthermore, the Kruskall –Wallis test used to compare means for the significant difference produces a very small P value for instance, in the comparison of means for the first cases of pertussis in the Netherlands Table 5-2. the comparison of Urban and rural areas during endemic and epidemic year showed P value of 2.31 X 10 -14 and 2.2 X 10 -16 respectively. Although difference of urban and rural area was significant the values of p was very small.

6.4. Space-time scan statistic method

The space-time permutation scan statistic was used to detect significant spatial temporal clusters for the data aggregated at the municipality level in the Netherlands. The technique for analysing detected significant clusters was defined whereby the spreading clusters were defined as those clusters that

meet, are close in space and time and overlap in both space and time. Clusters were considered to be close in space if they were disjoint in location for not more than one spatial location while close in time if they disjoint in time for not more than three weeks. Clusters were considered to meet in space and time if each cluster touches the boundary of another cluster in terms of space and two clusters collide in time i.e. if the start date of the present cluster is the end date for the previous cluster. The clusters were considered as overlapping clusters if their interiors intersect each other. Problem of this technique is that it fails to identify isolated clusters although they were close in location See the illustration in Figure 6-1.

Figure 6-1: Example of isolated clusters

Furthermore, clusters detected by space-time permutation statistic seem to persist in the southern part and west southern part of the Netherlands during epidemic years. The reason for this might be low vaccination coverage in these areas. In addition to that these clusters could be explained by the fact that the areas affected are close to the borders of the Belgium the movement over there is high compared to the centres of the Netherlands which makes us to be unsure with the vaccination status of the people over there.

Some detected covers the very large area but not all areas contain the number of cases. This is the major problem of the SatScan since it detect both clusters of high risk areas of which we are interested with and clusters with low risk areas which we are not interested with [46].

Moreover, SatScan method does not have cartographic support for understanding the identified clusters. It only provide the information about the identified clusters such as centre location, the radius of the cluster detected and data that describe each cluster via the text format. This output should be processed and exported to ArcMap which is a time consuming process and makes the interpretation of the result more difficulty. For instance, for the clusters which were detected within the same time period it is difficult to classify them manually. Additionally, if there is an overlapping cluster in the same location it is difficult to visualize it in map this might hinder the clusters that require further investigation. Because of this limitation it makes the interpretation of SatScan clusters relative to the original dataset very difficult. For instance, in case of identifying urban and rural areas, some clusters might be classified as they originate in rural areas while they are originating from Urban as SatScan can not test the significant of the origin of the cluster. Also this could be possible reason for the failure to identify the hierarchical spread of the disease during epidemic year and endemic year.

6.5. Comparison of epidemic and endemic years

At the European (countries of Europe) scale where the analysis was performed to observe if there is synchrony and travelling waves of an epidemic, the results indicate that there was no global epidemic synchrony observed. Also some countries in Europe showed clear travelling waves of epidemic but other countries did not.

At the Netherlands scale week rank method was used to identify the hierarchical spread of a disease. At the level of Municipality the result showed that there was a hierarchical spread in both endemic and epidemic years but during epidemic years the disease seems to arrive sooner in urban areas than in rural areas that is week 8 against 19. Additionally, the method showed that during epidemic years the disease appeared to persist more in urban areas than in rural areas. At Twente region and Enschede postcodes no difference was observed between urban and rural areas in both endemic and epidemic period.

Result of space-time scan statistic showed that during both endemic and epidemic years, the number of clusters in rural areas is higher than the number of clusters obtained in urban area. During epidemic years the number of clusters that originate in rural areas was 23 compared to 7 clusters that originated in urban areas. During endemic years, the number of clusters that originate in rural areas was 16 compared to 6 which originate in urban areas. Because of low number of clusters observed in urban areas it is difficult to conclude on the hierarchical spread. Therefore we can not differentiate between epidemic and endemic year. Table 6-1 and 6-2 respectively

Table 6-1 Epidemic years

Location	***Pattern***	***Number of clusters***
Urban	Fadeout	3
Urban	Persistence	3
Urban	Diffusion	1
Rural	Fadeout	18
Rural	Persistence	4
Rural	Diffusion	1

Table 6-2 Endemic years

Location	***Pattern***	***Number of clusters***
Urban	Fadeout	4
Urban	Persistence	-
Urban	Diffusion	2
Rural	Fadeout	15
Rural	Persistence	1
Rural	Diffusion	-

6.6. Validation of Agent based models

The main objective of this research is to identify multi-spatial temporal patterns that can be used to validate agent based simulation models by comparing endemic and epidemic years. Initial plan was to run simulated data to observe if patterns observed in empirical data can be observed as accurate as possible via the simulated data but due to time constraints we decided to reflect on this.

We first identified the patterns in a disease data at different hierarchical scale level so as to assist in the validation of an agent based simulation model. To do so, patterns observed in a disease data should be used as guidance in designing a model structure of an agent based simulation model. The model should include all the characteristics of the patterns (i.e. the model should include variables which makes that pattern to emerge). The designed model should be close or should match with the patterns observed in empirical data [50].

In this study we identified patterns at the country level of which we identify the synchrony and travelling waves for pertussis, we identified hierarchical spread patterns in the Netherlands at the level of municipality to Enschede at the level of postcodes. Then we identified hierarchical spread patterns in the Netherlands at the municipality level by using space-time permutation statistics.

European countries showed the travelling waves but not for all countries. Also the spatial synchrony observed was not always for all counties since each country have their own vaccination strategy. By the week rank method in the Netherlands at the level of municipality we identified hierarchical spread patterns but at in Twente region and Enschede postcode this pattern could not be observed.

Moreover the pattern was not clear observed in the Netherlands at municipality level by using space-time scan statistic. This makes difficult to conclude even the results of week rank method. For agent based simulation model it is required to have patterns at different hierarchical level and one must be sure that that pattern exist and that pattern can also emerge in the model output [10]. Additionally, patterns in a designed model must be able to capture important features in the empirical dataset at different spatial and temporal scales and at different hierarchical level [50]. In this analysis we have only one pattern observed in the Netherlands at the municipality level When using only one pattern in an empirical dataset the model could not capture all important features in that dataset. Due to these findings validation of an agent based simulation model is not possible.

6.7. Conclusion

The main objective of this research is to find multi-scale spatial temporal patterns in disease data that can be used to validate agent-based simulations models by comparing endemic and epidemic of disease data.

Based on this objective we conclude that;

Patterns that are useful are “synchrony”, “travelling waves” and “Hierarchical spread”. Synchrony was checked at the spatial level of European countries but did not detect useful results. Travelling waves were checked both for European Countries and the Netherlands but the methods used for

detecting this type of spread did not lead to good results. Hierarchical spread was only detected using the week rank method for the Netherlands at the municipality. No clear differences were detected between epidemic and endemic years. Some differences were found using the week ranks at the Netherlands level. In epidemic years the disease arrives earlier in urban areas compared to rural areas.

Limitations in the data had clear influences on the results. At the European scale missing data for many countries made it difficult to identify diffusion patterns. At the level of the Netherlands it is difficult to determine the "direction of spread" as the origin of a cluster can never be completely detected. Differences exist between different regions of the country related to the awareness of pertussis. This results in differences in detection rates. In general only a fraction of the true pertussis case is registered (15% for primary infection and 1% for repeat infection) [51].

The methods applied were week rank, space-time permutation model and mathematical formula for defining epidemic and endemic years

6.8. Recommendation

Further study is recommended that;

1. To investigate the hierarchical spread patterns in Twente region and Enschede by using different method, and to repeat the regional check for other regions of the Netherlands
2. To check for synchrony between larger cities in the Netherlands as this information will provide a link between the hierarchical spread information of the Netherlands
3. Find a different method to detect travelling waves.

6.9. Summary

In this chapter, the result and implication of the analysis carried out in this thesis was discussed. The conclusion and recommendation was also provided.

References

1. Martin, S.W., P. Michel, J. Holt, D. Middleton, and J. Wilson, *Geographical and temporal distribution of human giardiasis in Ontario, Canada.* International Journal of Health Geographics, 2003. **2**(1): p. 5.
2. Jia, Z.W., X.W. Li, W. Wang, and S.M. Cheng, *Promising of spatial-temporal model in public health.* Chinese Medical Journal, 2009. **122**(3): p. 349-350.
3. Odoi, A., S.W. Martin, P. Michel, J. Holt, D. Middleton, and J. Wilson, *Geographical and temporal distribution of human giardiasis in Ontario, Canada.* International Journal of Health Geographics, 2003. **2**(1): p. 5.
4. Grenfell, B.T., O.N. Bjornstad, and J. Kappey, *Travelling waves and spatial hierarchies in measles epidemics.* Nature, 2001. **414**(6865): p. 716-723.
5. Broutin, H., J.-F. Guegan, E. Elguero, F. Simondon, and B. Cazelles, *Large-Scale Comparative Analysis of Pertussis Population Dynamics: Periodicity, Synchrony, and Impact of Vaccination.* Am. J. Epidemiol., 2005. **161**(12): p. 1159-1167.
6. Farnsworth, M.L. and M.P. Ward, *Identifying spatio-temporal patterns of transboundary disease spread: examples using avian influenza H5N1 outbreaks.* Veterinary Research, 2009. **40**(3).
7. Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W.M. Mooij, S.F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D.L. DeAngelis, *Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology.* Science, 2005. **310**(5750): p. 987-991.
8. Rossmanith, E., N. Blaum, V. Grimm, and F. Jeltsch, *Pattern-oriented modelling for estimating unknown pre-breeding survival rates: The case of the Lesser Spotted Woodpecker (Picoides minor).* Biological Conservation, 2007. **135**(4): p. 555-564.
9. Janssen, M.A., N.P. Radtke, and A. Lee, *Pattern-Oriented Modeling of Commons Dilemma Experiments.* Adaptive Behavior, 2009: p. 1059712309342488.
10. Grimm, V. and S. Railsback, *Agent-Based Models in Ecology: Patterns and Alternative Theories of Adaptive Behaviour*. 2006. p. 139-152.
11. Piou, C., U. Berger, and V. Grimm, *Proposing an information criterion for individual-based models developed in a pattern-oriented modelling framework.* Ecological Modelling, 2009. **220**(17): p. 1957-1967.
12. Marshall, R.J., *A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease.* Journal of the Royal Statistical Society. Series A (Statistics in Society), 1991. **154**(3): p. 421-441.
13. Glass, G.E., *Update: Spatial Aspects of Epidemiology: The Interface with Medical Geography.* Epidemiol Rev, 2000. **22**(1): p. 136-139.
14. Goldenberg, A., G. Shmueli, R.A. Caruana, and S.E. Fienberg, *Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(8): p. 5237-5240.
15. Lazarus, R., K. Kleinman, I. Dashevsky, A. DeMaria, and R. Platt, *Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection.* BMC Public Health, 2001. **1**(1): p. 9.
16. Hashimoto, S., Y. Murakami, K. Taniguchi, and M. Nagai, *Detection of epidemics in their early stage through infectious disease surveillance.* Int. J. Epidemiol., 2000. **29**(5): p. 905-910.
17. Onozuka, D. and A. Hagihara, *Spatial and temporal dynamics of influenza outbreaks.* In: Epidemiology, 19(2008)6, pp. 824-828, 2008.
18. Erwig, M., *Toward Spatiotemporal Patterns,* in *Spatio-Temporal Databases,* D. Caluwe, Editor. 2004, Springer. p. 29-54.

19. Cromley, E.K. and S. McLafferty, *GIS and public health*. 2002, New York: Guilford Press.
20. Broutin, H., E. Elguero, F. Simondon, and J.-F. Guégan, *Spatial dynamics of pertussis in a small region of Senegal.* Proceedings of the royal society B, 2004. **271**(2091-2098).
21. Viboud, C., O.N. Bjornstad, D.L. Smith, L. Simonsen, M.A. Miller, and B.T. Grenfell, *Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza.* Science, 2006. **312**(5772): p. 447-451.
22. Grenfell and Bolker, *Cities and villages: infection hierarchies in a measles metapopulation.* Ecology Letters, 1998. **1**(1): p. 63-70.
23. Lagorio, C., M.V. Migueles, L.A. Braunstein, E. López, and P.A. Macri, *Effects of epidemic threshold definition on disease spread statistics.* Physica A: Statistical Mechanics and its Applications, 2009. **388**(5): p. 755-763.
24. Broutin, H., F. Simondon, and J.F. Guegan, *Whooping cough metapopulation dynamics in tropical conditions: disease persistence and impact of vaccination.* Proceedings of the Royal Society of London Series B-Biological Sciences, 2004. **271**: p. S302-S305.
25. Broutin, H., F. Simondon, and J. Guégan, *Whooping cough metapopulation dynamics in tropical conditions: disease persistence and impact of vaccination.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2004. **271**(Suppl 5): p. S302-S305.
26. Gregory, D.S., *Pertussis: A Disease affecting all ages.* Amercian Familiy Physician, 2006. **74**(3): p. 420-426.
27. Versteegh, F.G.A., *Pertussis: new insights in diagnosis, incidence and clinical manifestations.* . 2005, Enschede: Febodruk bv.
28. Schouls, L.M., H.G.J.v.d. heide, L. Vauterin, P. Vauterin, and F.R. Mooi, *Multiple-Locus Variable-Number Tandem Repeat Analysis of Dutch Bordetella pertussis Strains Reveals Genetic Changes with Clonal Expansion during the Late 1990s.* Journal of Bacteriology, 2004. **186**(16): p. 5496-5505.
29. Preziosi, M.-P., A. Yam, S.G.F. Wassilak, L. Chabirand, A. Simaga, M. Ndiaye, M. Dia, F. Dabis, and F. Simondon, *Epidemiology of Pertussis in a West African Community before and after Introduction of a Widespread Vaccination Program.* Am. J. Epidemiol., 2002. **155**(10): p. 891-896.
30. Tanaka, M., C.R. Vitek, F.B. Pascual, K.M. Bisgard, J.E. Tate, and T.V. Murphy, *Trends in Pertussis Among Infants in the United States, 1980-1999.* JAMA, 2003. **290**(22): p. 2968-2975.
31. Greef, S.C.d., J.F.P. Schellekens, and F.R. Mooi, *Pertussis in the Netherlands 2001-2002,*. 2003: Bilthoven, The Netherlands.
32. Melker, H.E.d., M.A.E. ConynvanSpaendonck, and J.F.P. Schellekens, *Pertussis in the Netherlands: an Outbreak Despite High Levels of Immunization with Whole-Cell Vaccine.* Emerging Infectious Diseases 1997. **3**(2): p. 175-178.
33. Mattoo, S. and J.D. Cherry, *Molecular Pathogenesis, Epidemiology, and Clinical Manifestations of Respiratory Infections Due to Bordetella pertussis and Other Bordetella Subspecies.* Clin. Microbiol. Rev., 2005. **18**(2): p. 326-382.
34. De GREEFF, S.C., A.L.M. DEKKERS, P. TEUNIS, J.C. RAHAMAT-LANGENDOEN, F.R. MOOI, and H.E. De MELKER, *Seasonal patterns in time series of pertussis.* Epidemiology and Infection, 2009. **137**(10): p. 1388-1395.
35. FINE, P.E.M. and J.A. CLARKSON, *Seasonal Influences on Pertussis.* Int. J. Epidemiol., 1986. **15**(2): p. 237-247.
36. Skowronski, D.M., S. Gaston De, D. MacDonald, W. Wu, C. Shaw, J. Macnabb, S. Champagne, D.M. Patrick, and S.A. Halperin, *The Changing Age and Seasonal Profile of Pertussis in Canada.* The Journal of Infectious Diseases, 2002. **185**(10): p. 1448-1453.
37. WHO. *Immunization surveillance, assessment and monitoring*. 2009 14 /October /2009 [cited 2009 19 /October /2009]; Available from: http://www.who.int/immunization_monitoring/data/data_subject/en/.
38. CDC, *Shapefiles for Epi Info*. 2009.

39. URL *EconomyWatch*. 2009 [cited 2009 October, 15]; Available from: http://www.economywatch.com/economic-statistics/Albania/Population/.
40. Asiedu-Dartey, J.A.F., *Modelling of disease through space and time : a case study of pertussis in the Netherlands*. 2005, ITC: Enschede. p. 86.
41. CBS. *Population in the Netherlands (CBS)*. 2010 [cited 2009 1-12-2009]; Available from: http://www.swov.nl/uk/research/kennisbank/inhoud/90_gegevensbronnen/inhoud/bevolking.htm.
42. Kulldorff, M., *SaTScan User Guide*. SatScan, 2009. **Version 8.0**: p. 102.
43. Si, Y., A.K. Skidmore, T.J. Wang, W.F. de Boer, P. Debba, A.G. Toxopeus, and ... *Spatio - temporal dynamics of global H5N1 outbreaks match bird migration patterns.* Geospatial health : health applications in geospatial science, 2009. **4**(1).
44. Nunes, C., *Tuberculosis incidence in Portugal: spatiotemporal clustering.* International Journal of Health Geographics, 2007. **6**(1): p. 30.
45. Kulldorff, M., W.F. Athas, E.J. Feurer, B.A. Miller, and C.R. Key, *Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico.* Am J Public Health, 1998. **88**(9): p. 1377-1380.
46. Chen, J., R. Roth, A. Naito, E. Lengerich, and A. MacEachren, *Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality.* International Journal of Health Geographics, 2008. **7**(1): p. 57.
47. Kulldorff, M., *Prospective time periodic geographical disease surveillance using a scan statistic.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2001. **164**(1): p. 61-72.
48. Jasso, H., C. Baru, T. Fountain, W. Hodgkiss, D. Reich, and K. Warner, *Using 9-1-1 call data and the space-time permutation scan statistic for emergency event detection*, in *Proceedings of the 2008 international conference on Digital government research*. 2008, Digital Government Society of North America: Montreal, Canada. p. 92-98.
49. Kulldorff, M., R. Heffernan, J. Hartman, R. Assuncao, and F. Mostashari, *A space-time permutation scan statistic for disease outbreak detection.* Plos Medicine, 2005. **2**(3): p. 216-224.
50. Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W.M. Mooij, S.F. Railsback, H.H. Thulke, J. Weiner, T. Wiegand, and D.L. DeAngelis, *Pattern-oriented modeling of agent-based complex systems: Lessons from ecology.* Science, 2005. **310**(5750): p. 987-991.
51. Wearing, H.J. and P. Rohani, *Estimating the Duration of Pertussis Immunity Using Epidemiological Signatures.* PLoS Pathog, 2009. **5**(10): p. e1000647.

Appendices

From appendix 1.1, 1.2, 2.1, 2.2, 3.1, 3.2, 4.1 and 4.2 the cluster type M indicates most likely clusters and S is the secondary clusters. For SatScan results within a given study period the most likely clusters is numbered as 1 followed by secondary clusters starting from number 2 in the list. Hence the number of S is ordered according to the number obtained in the list.

Time frame of cluster is the range of time on which the particular cluster was observed.

Appendix 1 .1 : Properties of clusters during 1996 epidemic year

Cluster Number	Cluster type	Time frame of cluster (weeks)	Origin of cluster	Observed cases	Relative Risk	Duration of cluster (weeks)	Radius (Km)
Cluster1	M	9	Urban	12	18.09	1	10.96
	S2	"	"	12	17.30	1	6.15
	S3	"	"	12	17.30	1	13.42
	S4	"	"	12	17.30	1	12.76
	S5	"	"	10	14.42	1	7.98
Cluster2	M	9-10	From cluster1	14	5.25	2	10.96
	S2	"	"	14	5.04	"	13.42
	S3	"	"	13	4.88	"	12.76
Cluster3	M	14-15	Rural	6	12.13	2	7.68
Cluster4	M	16	From cluster3	4	39.67	1	9.71
	S2	"	"	4	30.86	1	13.66
Cluster5	M	20	Rural	6	41.31	1	0
	S2	"	"	6	41.31	1	3.56
	S3	"	"	6	41.31	1	7.04
	S4	"	"	6	41.31	1	5.74
	S5	"	"	6	41.31	1	12.82
	S6	"	"	6	30.98	1	29.75
	S7	"	"	6	20.65	1	7.75
	S8	"	"	6	20.65	1	17.41
	S9	"	"	6	19.07	1	16.72
Cluster6	M	20-21	From cluster5	10	16.76	2	0.00
	S2	"	"	10	16.76	2	3.56
	S3	"	"	10	16.76	2	7.04
	S4	"	"	10	16.76	2	5.74
	S5	"	"	10	16.76	2	12.82
	S6	"	"	10	13.96	2	29.75

	S7	“	“	10	10.47	2	7.75
	S8	“	“	10	10.47	2	17.41
	S9	“	“	10	9.86	2	16.72
	S10	“	“	11	6.36	2	31.27
Cluster7	M	22-23	Rural	12	4.79	2	9.33
Cluster8	M	24	Rural	34	12.96	1	0.00
	S2	“	“	34	12.61	1	6.89
	S3	24	From cluster7	5	20.42	1	19.70
	S4	“	“	5	18.15	1	51.04
Cluster9	M	24-25	From cluster 8A	34	8.06	2	0.00
	S2	“	“	34	7.84	2	6.89
	S3	“	“	37	6.78	2	8.66
	S4	24-25	From cluster 7	12	6.77	2	19.70
	S5	“	“	12	6.00	2	43.82
Cluster10	M	27-28	From cluster 9A	7	10.51		11.41
Cluster11	33	33	Urban	10	6.60	1	15.63
Cluster12	M	33-34	From cluster 11	16	4.81	2	14.26
Cluster13	M	38	Rural	13	12.48	1	8.17
	S2	“	“	17	7.45	1	8.43
	S3	“	“	21	5.72	1	16.44
	S4	“	“	12	9.30	1	8.32
	S5	“	“	17	5.44	1	15.74
	S6	“	“	19	3.91	1	27.52
	S7	“	“	20	3.70	1	34.58
Cluster14	M	38-39	From cluster 13	12	7.96	2	0.00
	S2	“	“	13	6.38	2	8.09
	S3	“	“	18	4.32	2	8.43
	S4	“	“	22	3.40	2	16.44
Cluster15	M	39-40	Rural	52	2.51	2	31.13
Cluster16	M	41-43	From cluster13	22	3.67	2	10.98
Cluster17	M	43-44	From cluster 16	34	2.61	2	14.93
	S2	“	“	54	2.02	2	28.37
Cluster18	M	47	Rural	4	230.31	1	0.00
	S2	“	“	4	83.75		9.03
	S3	“	“	4	70.87		13.99
	S4	“	“	4	51.18		8.05

Appendix 1 .2 : Description of classified clusters during 1996 epidemic year

Cluster number	Cluster type	Cluster new number
Cluster 8	M,S2	8A
	S3,S4	8B
Cluster 9	M, S2,S3	9A
	S4,S5	9B

Appendix 2 .1 : Properties of clusters during 1997 endemic year

Cluster Number	Cluster type	Time frame of cluster (weeks)	Origin of cluster	Observed cases	Relative Risk	Duration of cluster (weeks)	Radius (Km)
Cluster1	M	53-1	Rural	25	5.47	2	33.68
	S2	"	Rural	25	3.36	2	26.29
Cluster2	M	53-2	From cluster1A	27	4.60	2	33.68
Cluster3	M	3-4		7	19.01	2	0.00
Cluster4	M	5	Rural	6	163.46	1	0.00
	S2	5	"	6	118.05	1	6.34
	S3	"	"	6	96.59	1	11.30
	S4	"	"	6	75.89	1	15.36
	S5	"	"	6	27.24	1	18.81
Cluster5	M	5-6	From cluster 4	6	24.18	2	0.00
Cluster6	M	12	Rural	6	23.45	1	31.01
Cluster7	M	16	Urban	7	186.07	1	7.55
	S2	"	"	7	114.93	1	12.24
	S3	"	"	7	86.83	1	15.81
	S4	"	"	7	52.10	1	17.13
	S5	"	"	7	38.69	1	31.61
	S6	"	"	7	24.42	1	25.92
	S7	"	Rural	7	16.84	1	70.38
Cluster8	M	16-17	From cluster7A	7	64.45	2	7.55
	S2	"	"	8	30.33	2	15.62
	S3	"	"	7	39.81	2	12.24
	S4	"	"	8	20.35	2	17.13
	S5	"	"	9	10.74	2	25.92
	S6	"	"	8	12.47	2	32.31
Cluster9	M	16-18	Urban	19	5.26	3	8.78

	S2	“	“	13	5.92	3	7.51
Cluster10	M	17-19	From cluster9	20	4.87	3	8.78
	S2	“	“	14	5.55		7.51
Cluster11	M	18-20	Rural	4	91.68	3	0.00
	S2	“	“	6	26.19	3	15.04
	S3	“	“	4	73.34	3	0.00
	S4	“	“	4	73.34	3	8.58
Cluster12	M	21-22	Rural	4	62.45	2	0.00
	S2	“	“	4	52.04	2	4.79
Cluster13	M	21-23	From cluster12	9	29.30	3	0.00
	S2	“	“	9	25.86	3	4.79
	S3	“	“	9	18.32	3	5.08
	S4	“	“	9	17.58	3	5.02
	S5	“	“	9	15.70	3	7.47
	S6	21-23	From cluster9	18	5.32	3	11.56
Cluster14	M	22-24	From cluster12	9	42.41	3	0.00
	S2	“	“	9	38.56	3	4.79
	S3	“	“	9	29.25	3	5.08
	S4	“	“	9	28.28	3	5.02
	S5	“	“	9	25.71	3	7.47
	S6	“	From cluster 13B	22	4.86	3	9.26
	S7	“	“	16	5.69	3	7.51
	S8	“	From cluster12	10	8.98	3	11.36
Cluster15	M	23-25	From cluster12	7	16.02	3	7.15
Cluster16	M	25-27	From cluster 13B	22	4.00	3	9.81
Cluster17	M	26-28	From cluster 14B	22	4.28	3	9.26
Cluster18	M	30-32	From cluster16	26	3.43	3	7.44
Cluster19	M	35	Urban	5	95.62	1	6.40
	S2	“	“	5	83.14	1	9.59
	S3	“	“	5	79.68	1	7.55
	S4	“	“	5	73.55	1	12.53
	S5	“	“	4	80.52	1	10.50
Cluster20	M	42-43	Rural	6	23.31	2	0.00
	S2	“	“	6	21.89	2	7.48
	S3	“	“	7	14.80	2	8.50
	S4	“	“	10	7.95	2	20.80
Cluster21	M	46	Rural	3	300.86	1	0.00
	S2	“	“	3	188.04	1	4.92

Appendix 2 .2 : Description of classified clusters during 1997 endemic year

Cluster number	Cluster type	Cluster new number
Cluster 1	M	1A
	S2	1B
Cluster 7	M, S2,S3,S4,S5,S6	7A
	S7	7B
Cluster 11	M,S2	11A
	S3,S4	11B
Cluster 13	M,S2,S3,S4,S5	13A
	S6	13B
Cluster 14	M,S2,S3,S4,S5,S8	14A
	S6,S7	14B

Appendix 3 .1 : Properties of clusters during 1998 endemic year

Cluster Number	Cluster type	Time frame of cluster (weeks)	Origin of cluster	Observed cases	Relative Risk	Duration of cluster (weeks)	Radius (Km)
Cluster 1	M	5- 7	Rural	5	28.09	3	0.00
	S2	5 – 7	Rural	5	28.09	3	4.72
Cluster 2	M	10 –12	Rural	5	39.83	3	12.30
Cluster 3	M	26 – 27	Rural	7	12.89	2	4.49
	S2	25 – 27	Rural	10	7.13	3	13.51
Cluster 4	M	26 – 28	From cluster 3B	11	6.70	3	13.51
Cluster 5	M	34- 36	Urban	36	4.23	2	21.51
	S 2	35 – 36	Urban	7	26.50	2	0.00
	S 3	35 – 36	Rural	7	22.36	2	8.32
	S 4	34 – 36	Rural	31	3.12	2	35.99
	S 5	35- 36	Rural	7	18.35	2	13.79
	S6	35- 36	Rural	7	17.89	2	17.06
	S7	35- 36	Rural	5	36.51	2	8.80
	S8	35 –36	Rural	7	16.26	2	8.78
	S9	35 – 36	Rural	23	3.48	2	24.16
	S10	35 – 36	Rural	7	14.60	2	37.65

	S11	35 – 36	Rural	7	12.13	2	20.84
	S12	35- 36	Rural	7	11.93	2	21.84
Cluster 6	M	35 – 37	From cluster 5A	34	4.92	3	15.48
	S2	35 – 37	“	37	3.77	3	28.32
	S3	35 – 37	“	32	3.05	3	25.12
Cluster 7	M	36- 38	From cluster5A	19	4.47	3	24.99
Cluster 8	M	37 – 39	“	10	11.68	3	9.81
	S2	“	“	12	5.90	3	21.59
Cluster 9	M	39 – 40	“	23	3.77	2	23.45
Cluster10	M	40- 41	Urban	11	18.66	2	8.14
	S2	40- 41	Urban	11	11.98	2	22.26
	S3	39- 41	From cluster 5A	28	3.24	3	23.45
Cluster11	M	40 – 42	From cluster 10 A	24	3.59	3	33.32
Cluster12	M	43- 45	From cluster 5A	21	3.09	3	20.35
Cluster13	M	44- 46	From cluster 5A	10	10.20	3	18.02
	S 2	45-46	Rural	5	31.61	2	6.18

Appendix 3 .2 : Description of classified clusters during 1998 endemic year

Cluster number	Cluster type	Cluster new number
Cluster 3	M	3A
	S2	3B
Cluster 5	M, S4,S7,S9	5A
	S2,S3,S5,S6,S8,S10,S11,S12	5B
Cluster 10	M,S2	10A
	S3	10B
Cluster 13	M	13A
	S2	13B

Appendix 4 .1 : Properties of clusters during 1999 epidemic year

Cluster Number	Cluster type	Time frame of cluster (weeks)	Origin of cluster	Observed cases	Relative Risk	Duration of cluster (weeks)	Radius (Km)
Cluster1	M	2-3	Urban	13	5.82	2	8.49
	S	2-3	"	13	5.81	2	12.48
Cluster2	M	5	Urban	19	3.83	1	58.54
Cluster3	M	5-6	From cluster2	20	4.14	2	48.03
	S	5-6	"	26	3.33	2	31.47
Cluster4	M	9	Urban	22	3.33	1	5.38
Cluster5	M	11	Rural	4	149.24	1	0.00
	S	11	Rural	4	130.58	1	6.44
	S	11	"	4	130.58	1	8.16
	S	11	"	4	130.58	1	16.21
	S	11	"	4	116.07	1	12.21
	S	11	"	4	74.62	1	37.55
	S	11	"	6	18.88	1	12.94
	S	11	"	4	47.48	1	21.37
Cluster 6	M	13-13	Rural	11	8.95	2	11.86
Cluster7	M	15	Urban	13	9.26	1	17.60
	S	15		14	5.51	1	14.46
	S	15		10	7.54	1	19.11
	S	14-15		18	4.00	2	27.28
Cluster8	M	15-16	From cluster7	17	9.06	2	12.39
	S	15-16	"	18	6.79	2	13.68
	S	"	"	18	3.04	2	19.11
	S	"	"	24	3.89	2	27.28
	S	"		19	4.28	2	11.04
	S	"	"	22	3.55	2	16.28
Cluster9	M	17-18	"	13	6.07	2	19.47
Cluster10	M	21-22	Rural	18	4.06	2	12.99
Cluster11	M	22-23	From cluster10	18	4.08	2	12.99
Cluster12	M	24	Rural	7	24.25	1	32.03
	S	24	"	7	17.09	1	81.53
Cluster13	M	24-25	From cluster 12	24	3.78	2	34.95
	S 1	24-25	"	22	3.61	2	81.53
Cluster14	M	25-26	Rural	37	3.67	2	27.87
	S2	"	"	39	3.21	2	58.72
	S3	"	"	34	2.43	2	43.41

Cluster15	M	26-27	From cluster14	37	3.71	2	27.87
	S2	“	“	39	3.29	2	58.72
	S 3	“	Rural	28	3.36	2	54.88
Cluster16	M	29	Rural	4	73.13	1	0.00
Cluster17	M	30	From cluster15A	11	7.29	1	21.75
Cluster18	M	30-31	From cluster15A	62	3.17	2	32.24
	S 2	30-31	“	64	3.02	2	68.45
	S 3	31	Urban	18	6.03	1	22.62
Cluster19	M	31-32	From cluster 17	61	2.72	2	32.24
	S 2	“	“	63	2.63	2	68.45
	S3	“	From cluster 18B	29	4.04	2	22.62
	S4	“	Rural	7	15.67	2	5.41
Cluster20	M	32-33	Rural	21	3.67	2	32.16
Cluster21	M	33-34	From cluster19B	34	2.60	2	13.89
Cluster22	M	34-35	From cluster18A	22	4.61	2	14.18
	S2	“	“	34	3.04	2	29.05
	S3	“	Rural	19	4.56	2	8.62
	S4		From cluster20	7	16.33	2	0.00
	S5	35	Rural	18	4.37	1	12.15
	S6	34-35	From cluster18A	29	3.00	2	25.05
	S7	34-35	From cluster20	18	4.19	2	27.36
	S8	35	From cluster18A	36	2.57	1	65.45
	S9	34-35	From cluster20	7	13.07	2	8.68
Cluster23	M	35-36	From cluster19A	33	4.60	2	14.78
	S2	“	“	48	2.87	2	32.31
	S3	“	“	49	2.54	2	56.61
	S4	35-36	From cluster 19B	18	5.20	2	18.86
	S5	36	“	14	6.31		42.66
	S6	35-36	From cluster19A	26	3.38	2	16.58
	S7	35-36	From cluster 19B	34	2.62	2	75.17
Cluster24	M	36-37	From cluster20	26	5.19	2	14.78

	S2	36-37	“	26	3.73	2	15.54
	S3	“	“	29	3.32	2	32.93
	S4	36-37	From cluster22A	25	3.41	2	55.64
	S5	“	From cluster20	20	3.87	2	16.58
Cluster25	M	38	From cluster 22C	8	25.80	1	21.15
	S2	38	“	10	12.44	1	39.08
	S3	38	Rural	7	21.45	1	6.77
	S4	38	From cluster 22C	10	8.93	1	73.93
	S5	38	Rural	9	8.52	1	8.80
Cluster26	M	40	Urban	14	12.06		8.87
	S2	40	“	14	9.86	1	11.88
	S3	40	“	14	7.48	1	18.53
Cluster27	M	40-41	From cluster 26	18	6.53	2	8.87
	S2	40-41	“	18	5.48	2	11.88
	S3	“	“	29	3.29	2	20.54
	S4	41	Rural	13	5.84	1	21.60
Cluster28	M	41-42	From cluster 27B	26	3.52	2	23.60
Cluster29	M	43	Rural	3	203.49	1	0.00
Cluster30	M	43-44	Rural	28	3.21	2	16.09
Cluster31	M	46	Rural	11	33.21		0.00
	S2	46	“	11	18.98	1	8.74
	S3	46	“	11	15.94	1	15.47
	S4	46	“	11	13.07	1	17.64
	S5	46	“	11	9.27	1	25.61
Cluster32	M	46-47	From cluster 31	11	19.40	2	0.00
	S2	“	“	11	11.08	2	8.74
	S3	“	“	11	9.31	2	15.47
	S4	“	“	12	8.19	2	17.64
	S5	“	“	13	6.32	2	25.61
Cluster33	M	48	Rural	28	4.72	1	9.30
	S2	“	“	23	4.58	1	11.84
	S3	“	“	32	2.91	1	17.93
	S4	“	“	34	2.80	1	15.54
Cluster34	M	48-49	From cluster 33	28	4.69	2	9.30
	S2	“	“	32	3.50	2	14.62
	S3	“	“	33	3.11	2	17.45
	S4	“	“	32	2.69	2	14.62
Cluster35	M	51	Rural	2	1329.00	1	0.00
	S2	“	“	2	1329.00	1	4.14
	S3	“	“	3	137.48	1	13.16

Appendix 4 .2 : Description of classified clusters during 1999 epidemic year

Cluster number	Cluster type	Cluster new number
Cluster 15	M,S2	15A
	S3	15B
Cluster 18	M, S2	18A
	S3	18B
Cluster 19	M,S2	19A
	S3	19B
	S4	19C
Cluster 22	M,S2,S6,S8	22A
	S3,S5	22B
	S4,S7,S9	22C
Cluster 23	M,S2,S3,S6	23A
	S4,S5,S7	23B
Cluster24	M,S2,S3,S5	24A
	S4	24B
Cluster25	M,S2,S4	25A
	S3, S5	25B
Cluster27	M,S2,S3	27A
	S4	27B