

Student thesis series INES nr 429

Investigating the spatial patterns and climate dependency of Tick- Borne Encephalitis in Sweden

Elisabeth Farrington

2017

Department of

Physical Geography and Ecosystem Science

Lund University

Sölvegatan 12



Elisabeth Farrington (2017).

Investigating the spatial patterns and climate dependency of Tick-Borne Encephalitis in Sweden

En undersökning av rumsliga mönster och klimatberoende av fästingburen encefalit (TBE) i Sverige

Master degree thesis, 30 credits in *Geomatics*

Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2017 until *June* 2017

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Investigating the spatial patterns and climate dependency of Tick-Borne Encephalitis in Sweden

Elisabeth Farrington

Master thesis, 30 credits, in *Geomatics*

Supervisor 1: Ali Mansourian

Department of Physical Geography and Ecosystem Science

Supervisor 2: Mohammadreza Rajabi

Department of Physical Geography and Ecosystem Science

Exam committee:

Anna-Maria Jönsson

Department of Physical Geography and Ecosystem Science

Weiming Huang

Department of Physical Geography and Ecosystem Science

Abstract

The increasing prevalence of Tick-Borne Encephalitis (TBE) in Sweden is a cause of concern to both individuals and the public health service. The disease is spread via bites from the vector *Ixodes ricinus* and causes long-term neurological damage in 46% of cases. To date, research in this field has focussed on the relationship between climate and TBE incidence and there is little to no available research regarding the spatial distribution of TBE in Sweden. The aim of this study was to determine whether the relationship between TBE and climate factors is statistically significant and to detect spatial patterns of TBE both nationally and in the endemic Stockholm-Mälaren region in eastern Sweden. The data includes annual TBE incidence from 1986 to 2016. Monthly data was not available.

Multiple linear regressions were used to determine the dependency of TBE on climate variables (temperature and precipitation). Many of the explored years show a statistically significant relationship between TBE and at least one climate variable, though using data at a finer temporal scale would produce results with more years showing statistical significance. Getis-Ord G_i^* and Local Moran's I were used in order to detect spatial autocorrelation between TBE incidence points in the Stockholm-Mälaren region. Significant clustering ($p < 0.1$) was detected in the majority of years in the temporal dataset, with high levels of clustering occurring in southern Stockholm County and along the shores of Lake Mälaren.

Climate change increases the risk of ticks and, as a result, TBE spreading further north in Sweden. This study recommends further spatial analysis and modelling of the spread of TBE in relation to climate factors. This will in turn allow policy makers within the public health sector to make informed decisions regarding preventative schemes and rehabilitation programmes. However, it requires that the reporting of TBE incidents takes place on a monthly basis.

Sammanfattning

Den ökande förekomsten av fästingburen encefalit (TBE) i Sverige är oroväckande för både individer och folkhälsovårdsektorn. Sjukdomen sprids av fästingbett orsakade av vektorn *Ixodes ricinus* och orsakar långvarig neurologisk skada i 46 % av fallen. Forskning inom detta område har hittills fokuserat på förhållandet mellan klimat och TBE-förekomst, men det finns ingen tillgänglig forskning angående den rumsliga fördelningen av TBE i Sverige. Syftet med denna studie var att avgöra om förhållandet mellan TBE och klimatfaktorer är statistiskt signifikant samt att upptäcka rumsliga mönster av TBE i Stockholm-Mälarenområdet i östra Sverige där sjukdomen är endemisk. Data till denna studie består av årliga TBE-infektioner från 1986 till 2016. Månadsdata var ej tillgänglig.

Multipel linjär regression användes för att avgöra TBE-beroendet av klimatvariablerna temperatur och nederbörd. Flera år visar ett statistiskt signifikant förhållande mellan TBE och åtminstone en av klimatvariablerna, dock hade månadsdata antagligen givit bättre signifikans i resultaten. Getis-Ord G_i^* och Local Moran's I användes till att upptäcka rumslig autokorrelation mellan TBE incidenter i Stockholm-Mälarenområdet. De flesta signifikanta TBE fallen är grupperade i södra Stockholms län och runt Mälaren.

Med klimatförändringar finns risk att fästingar och därmed risken för TBE sprider sig norrut i Sverige. Mer forskning inom rumslig analys och modellering av TBE och dess påverkande faktorer rekommenderas därför så att beslutsfattare inom folkhälsovården kan implementera förebyggande system och rehabiliteringsprogram. Det fordrar dock att rapportering av TBE incidenter sker på månatlig basis och inte som nu årsvis.

Acknowledgements

I would like to thank my supervisors, Ali Mansourian and Mohammadreza Rajabi, for their guidance and support throughout all stages of this thesis. I would like to extend my gratitude to Marika Hjertqvist, my contact at the Public Health Agency of Sweden (*Folkhälsomyndigheten*), for providing me with the disease incidence data used in this study. Finally, I would like to thank my friends and family for keeping me in good spirits and inspiring me every day.

Table of Contents

1. Introduction.....	1
1.1 Background.....	1
1.2 Main aim of study.....	2
1.3 Scope of study	2
1.4 Outline	2
2. Literature review	5
2.1 TBE.....	5
2.2 TBE and the influence of environmental conditions.....	7
2.3 Tracking changes in ticks and TBE extent and distribution	7
2.4 Other methods of spatial analysis	9
3. Methodology	11
3.1 Data preparation and sources	11
3.2 Statistical analyses.....	12
3.3 Point pattern analysis.....	14
4. Results	17
4.1 Multiple regressions	17
4.2 Point Pattern Analysis	19
5. Discussion	39
5.1 TBE incidence and climatic factors	39
5.2 Patterns in TBE incidence	40
6. Conclusions.....	45
7. References.....	47

1. Introduction

1.1 Background

Vector borne diseases are infections caused by the bites of infected arthropods. They are responsible for a significant proportion of morbidity and mortality around the world, with 17% of all infectious diseases caused by vectors (World Health Organization 2016). Mosquitos, ticks and fleas are among the most common vectors, with ticks being the predominant vector in regards to the spread of zoonotic pathogens in Europe (Rizzoli et al. 2007). Ticks are vectors for Tick-borne encephalitis virus (TBEV) which causes Tick-borne encephalitis. The prevalence of TBE has risen over the last half century in Europe and is an increasing cause of concern as there is currently no cure available (Haemig et al. 2011).

Geographic Information Systems (GIS) are an important asset in the field of epidemiology. They can be used for visualising the geographical distribution of a disease as well as for more detailed analyses regarding spatial patterns or relationships between pathogenic factors and their geographic environment (Ruankaew 2005). The geographic distribution of ticks cannot be studied using GPS or other tracking devices, due to their small size. GIS provides a simple and cost-efficient means of mapping the distribution of ticks, though this requires suitable spatial data and knowledge of the ticks behavioural patterns (Cromley 2003). This information allows for the production of risk maps emphasizing regions most suited to vector habitation and reproduction as well as predictive maps of where disease may spread in the future.

The prevalence of TBE in Sweden is of great concern to a large number of Swedes, with national news agencies reporting on the disease as recently as May 2017 (Svenska Dagbladet 2017a, b). Many studies concerning the TBE virus have focussed on determining the major factors, particularly biotic and climatic, influencing the transmission of the disease. A handful of studies have attempted to predict how the number of TBE cases may change in the future using variables relating to wildlife and climate in Sweden (Haemig et al. 2011). Despite the overall consensus suggesting that tick populations and TBE cases are on the rise, there is little available research regarding the spatial distribution of the disease in Sweden. It is important to conclude whether the increase in the geographical range of TBE is significant as this will determine which areas are most vulnerable to TBE exposure. Mapping high-risk zones allow for policy makers to allocate resources in a cost effective manner in regards to vaccination programmes or specialised treatment. It also allows for counties and parishes to be one step ahead of the disease. This way they can plan prevention schemes to try firstly to avoid the disease spreading over a greater area and eventually eliminate the disease altogether.

1.2 Main aim of study

The main objectives of this master thesis are:

- 1- To study whether temperature and precipitation are independent variables influencing the prevalence of the tick-borne encephalitis infection
- 2- To perform spatial analysis and point pattern detection of the disease

In order to meet these objectives, the main research questions are:

- 1- Is there a statistically significant relationship between climatic factors and the location of TBE infection?
- 2- Does the spatial distribution of TBE incidence display any evident patterns?

1.3 Scope of study

The geographical scope of this study is limited to Sweden and will examine the spatial relationship between TBE incidence and climate on a national level. As the disease is endemic to Stockholm and the region around Lake Mälaren, TBE cases in the counties of Södermanland, Stockholm, Uppsala and Västmanland will also be studied (Figure 1). It will be referred to as the Stockholm-Mälaren region throughout the study.

1.4 Outline

The thesis is organised in six chapters. Chapter 1 serves as an introduction to the study specifying the aim and providing general background information. Chapter 2 is a literature review characterising the disease and its vector, the relationship between ticks and climate as well as relevant methods used in the modelling of TBE. Chapter 3, the methodology, describes the methods used in data formatting and analysis, the results of which are presented in Chapter 4. Chapter 5 includes a detailed interpretation of the results aided by relevant existing literature. Chapter 6 concludes the thesis by presenting major findings and recommendations for further research.

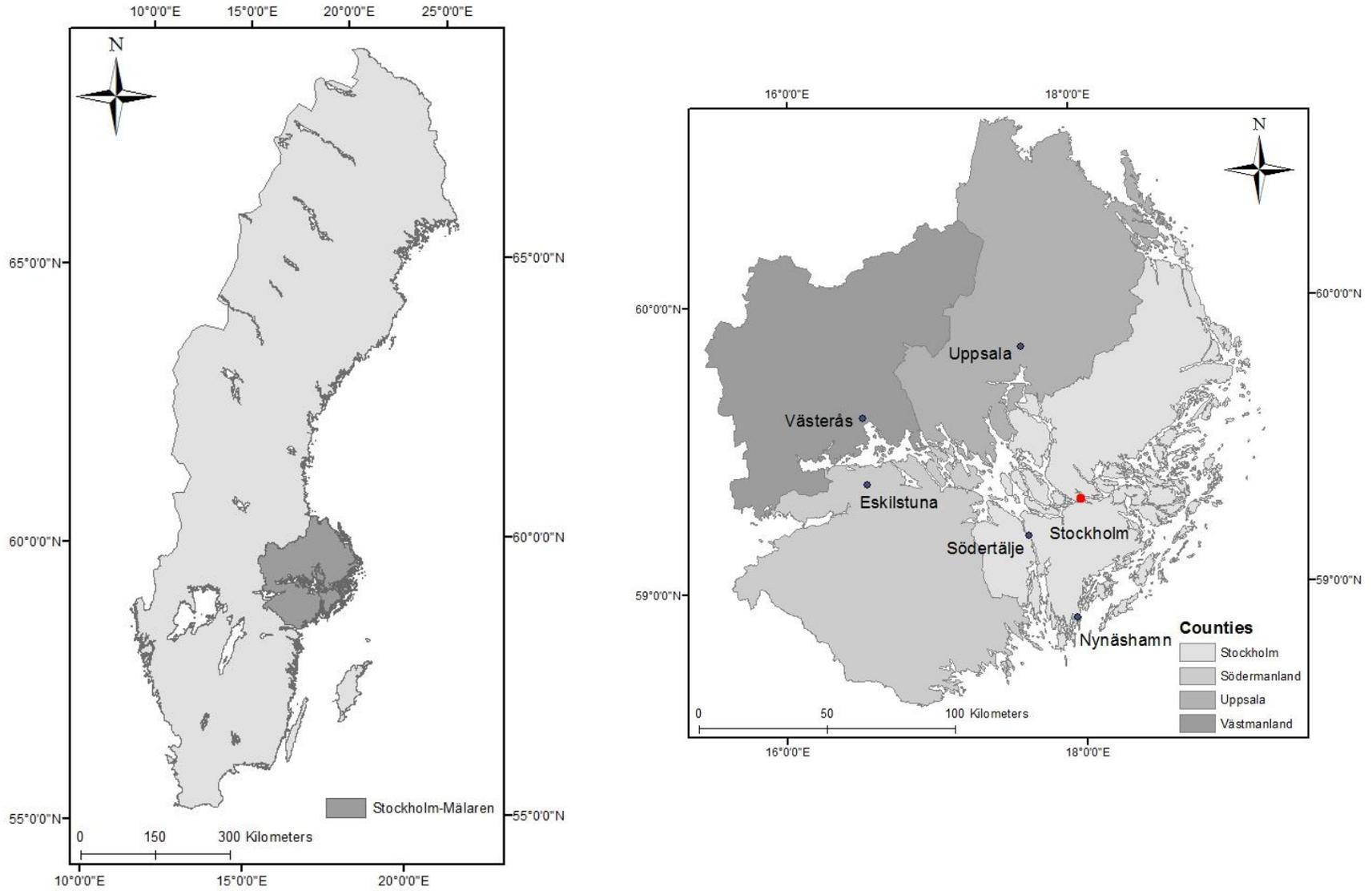


Figure 1 The geographical scope of the study including Sweden and the Stockholm-Mälaren region. Projection: RT90_25_gon_W

2. Literature review

This chapter will review literature pertaining to the disease tick-borne encephalitis and the role of *Ixodes ricinus* in its transmission. It will also examine the influence of environmental factors and review relevant literature discussing modelling techniques.

2.1 TBE

Tick-borne encephalitis, more commonly referred to as TBE, is a vector-borne disease caused by the appropriately named tick-borne encephalitis virus (TBEV). TBEV, of the genus Flavivirus, can present itself as the Russian spring/summer encephalitis (RSSE) or Central European encephalitis (CEE). Whilst the diseases are regarded as fairly similar from a clinical standpoint, the protein structures of the viruses differ slightly, as do the species of tick involved in transmission (Dumpis et al. 1999). This study will focus on the Central European strain of the virus as it is prominent in Sweden.

2.1.1 Transmission

In epidemiology, the term host can be subdivided into transmission hosts and reservoir hosts. Whilst transmission hosts are responsible for passing the disease to other organisms, reservoir hosts promote the survival of a pathogen, allowing for cell growth and reproduction. The main reservoirs of the TBEV include small rodent species such as the yellow-necked mouse (*Apodemus flavicollis*) and the red vole (*Myodes rutilus*) (Lindquist and Vapalahti 2008; Valarcher et al. 2015). Though the host organism carries the virus, it is either immune or naïve in the sense that it does not display any symptoms of the disease (Dumpis et al. 1999; Centers for Disease Control and Prevention 2012).

The hard-bodied tick species *Ixodes ricinus* commonly known as the castor bean tick (in reference to its appearance when engorged) or the sheep tick is the chief vector responsible for the transmission of TBE to larger mammals such as domestic animals and humans. The tick can contract the virus transovarially (from parent to offspring), by mating with an infected tick or by feeding on infected organisms. The tick can contract the virus at any of the three active phases in its life-cycle (larva, nymph and adult) and will remain infected for the duration of its life (Dumpis et al. 1999).

Humans contract the disease as a result of a direct bite from an infected tick whereby the pathogen passes from the saliva of the tick and into the human bloodstream. The majority of tick to human incidences of transmission occurs when the tick is in its nymph stage (Lindgren and Gustafson 2001). The proportion of ticks carrying the virus is low, between 1-4%, and

approximately 1 in 600 tick bites are likely to transmit TBE (Lindgren 1998). The disease can also be transmitted via the consumption of unpasteurised dairy products from sheep, goats and cows, though the risk of infection is minimal (Andreassen et al. 2012).

2.1.2 Symptoms

The illness comprises two phases, though not all patients experience both. The duration of the first phase ranges from 2-10 days with a median of 5 days with onset symptoms including fever, vomiting, fatigue, headaches and muscle pain which manifest more specifically in the neck and shoulders (Gritsun et al. 2003; Lindquist and Vapalahti 2008). The infection can be hard to diagnose in those who only experience the first phase of the illness as mild symptoms can be brushed off as the flu (Dumpis et al. 1999). The two phases are separated by an interval of roughly one week (though it may range from 1-21 days) where there are no persisting symptoms.

For those in whom the disease progresses to the second stage, neurological symptoms manifest as the virus begins to impact the central nervous system (CNS). This can result in meningitis, meningoencephalitis or meningoencephalomyelitis. Meningitis is the inflammation of the meninges membrane surrounding the brain, meningoencephalitis the inflammation of the meninges and the brain, and meningoencephalomyelitis the inflammation of the brain and the spinal cord. Additional second stage symptoms include flaccid paresis whereby muscles in the upper and lower limbs are weakened and unable to contract. Second stage patients may also experience hallucinations and increased sensitivity and discomfort of the eyes from light exposure (photophobia) (Mansfield et al. 2009; Gritsun et al. 2003).

Though the disease has a mortality rate of less than 1% in Sweden (<2% in Europe as a whole), up to 46% of diagnosed cases result in permanent long-term neurological issues as a ramification of the initial TBE infection (Lundkvist et al. 2011; Andersson et al. 2010). There is a vaccine available that is recommended for those living in endemic regions as it provides a >90% protection rate (Folkhälsomyndigheten 2017; Heinz et al. 2013).

2.1.3 Current trend in Sweden

The first clinical case of TBE in Sweden was recorded in 1954 (Berger 2017). The number of cases has since increased considerably, particularly over the last 30 years in a trend resembling that of a second degree polynomial function (Figure 2). In 2004 it became obligatory by Swedish law to report all cases of TBE in accordance with the Communicable Diseases Act (*Smittskydslag*) (2004:168) where it is referred to as viral meningoencephalitis. Diagnosed cases are reported to the medical official for infectious disease (*smittskydsläkaren*) at the county level, as well as to the Public Health Agency of Sweden (*Folkhälsomyndigheten*) (Folkhälsomyndigheten 2017).

While the disease was previously considered endemic to the Stockholm-Mälaren region and the area around Mälaren, since the mid-1990s it has become increasingly prevalent in western Sweden, spreading westwards through Götaland towards the coast (Jaenson et al. 2012a).

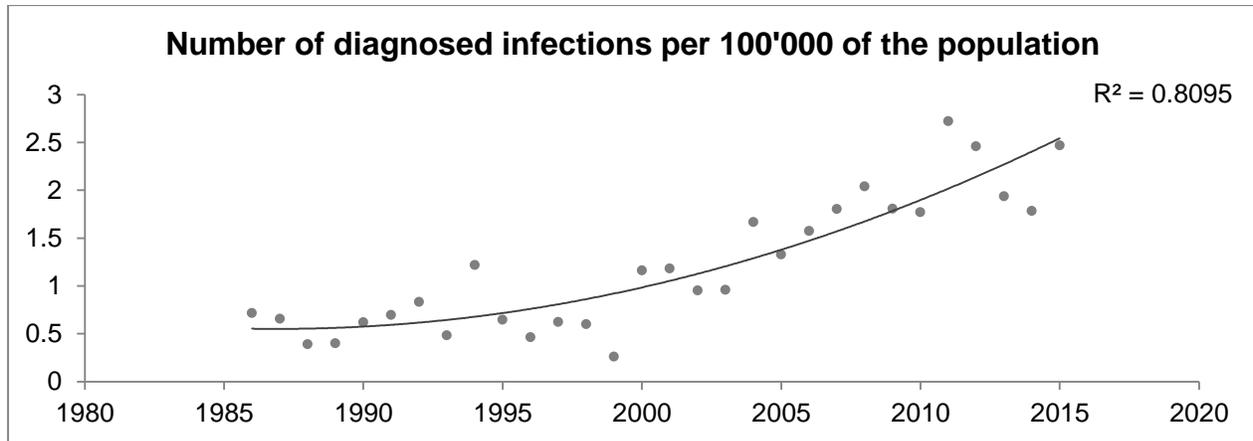


Figure 2 Number of diagnosed cases of TBE in Sweden 1986-2015 per 100'000 of the population. ($P < 0.00$) Population statistics according to (World Bank Group 2017)

2.2 TBE and the influence of environmental conditions

The TBE virus is reliant upon its vector host for survival so any environmental conditions impacting the tick will in turn influence the virus. Ticks are cold-blooded and thus highly susceptible to variations in climate. Ticks become active when temperatures fall between a 5-7°C range (Süss et al. 2008; Lindgren and Gustafson 2001). They are prone to desiccation and need a relative humidity of at least 80%, though preferable over 85% (Gray et al. 2009). Ticks habitats vary though typically include deciduous and coniferous woodland, forests and pastures as these habitats are often shared with roe deer. Roe deer are the preferred tick hosts, though vertebrates such as small rodents, wild boar and livestock are also common hosts.

2.3 Tracking changes in ticks and TBE extent and distribution

Researchers have implemented a variety of methods in order to determine how the spread of ticks and TBE is changing and which variables are pertinent to the survival of the ticks and the virus. The majority of their research focuses on the factors believed to be causing the number of TBE incidences to increase, though there is little research discussing the spatial distribution of the disease in Sweden. The first example considers the geographical distribution of ticks whilst the remainder focus on the relationship between TBE and environmental factors.

As the presence of ticks is essential to the spread of TBE, the research presented by Jaenson et.al (2012) focusses on determining the spatial range of *Ixodes ricinus* in Sweden in order to characterise regions that may become new foci of TBE. Their methodology included asking the general public about their experiences with ticks in and around their residencies in Sweden. In 1994 and 2009, questionnaires were published in local newspapers and free magazines targeting homeowners, dog owners and hunters in particular. They required the participants to answer questions about the occurrence of ticks within a 1km radius of their residence in 2008 and the number found on each family member and pets (Tälleklint and Jaenson 1998). Those taking the 2009 survey were also asked whether they believed that there had been any changes in the abundance of ticks since 1994, with the authors concluding that the vector has indeed spread northward and increased in abundance in central and southern Sweden (Jaenson et al. 2012b). The use of public surveys and opinion-based data in research may increase the risk of bias and subjectivity in the results, however the authors have stated that the cost and time efficiency of this method outweighs these concerns

Swedish TBE research tends to focus on the relationship between TBE and climate or other biological variables, with researchers performing regression analyses to test for significance. Regression analyses are used to determine how well one or more independent variables are able to predict the value of the dependent variable or more simply, how the dependent variable responds to change in the independent variable (Tu 1996; Yan and Su 2009). Regression analyses can highlight causality as well as forecast possible trends in the data (McCarroll 2016). In their TBE research, both Haemig et.al (2011) and Lindgren and Gustafson (2001) used multiple linear regression analysis as in the equation (1).

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots\beta_nx_n + \varepsilon \quad (1)$$

Haemig et.al used climate and wildlife data to model potential increases or decreases in TBE infections in the Stockholm-Mälaren region of Sweden. The wildlife data included populations of red fox, mink and roe deer amongst others and were used with mean monthly precipitation and temperature to perform cross-correlations to see how well they corresponded with the TBE incidences. They later used linear regressions to predict the number of incidences that were likely to occur the following year. Their best model suggested mink population and December precipitation were optimal in predicting any major changes in TBE incidences (Haemig et al. 2011).

Focussing solely on how temperature impacts the incidence of TBE, Lindgren and Gustafson (2001) used multiple linear regression and backwards selection to determine how different temperature ranges impacted TBE incidence in the Stockholm County. Using climate data from

the year of incidence and the year prior, their study included 18 different explanatory variables that accounted for the number of days within specified ranges during winter, spring and autumn. The study concluded that milder winters and an earlier onset spring were significant factors when accounting for TBE incidence.

2.4 Other methods of spatial analysis

There are a wide variety of methods for detecting spatial pattern that have yet to be used in TBE modelling in Sweden. These methods are easy to implement with appropriate software and have been used in several areas of epidemiological research.

2.4.1 Geographically Weighted Regression

When there is apparent spatial dependency within a global model, local models can be used in order to depict how the relationship between the dependent variable and the explanatory variables vary in space. Geographically weighted regression (GWR) is a local method which implies that a regression equation is applied to each feature within a dataset as opposed to simply computing a global mean. GWR, presented in equation (2), applies a moving kernel over the dataset computing the mean value for each spatial window (Miller 2012; Fotheringham et al. 2003). In the equation, the coordinates of point i are denoted as (u_i, v_i) , y and x_{ik} as the dependent and k^{th} independent variable respectively and ε as the error term. The results of the GWR can be depicted visually and coefficient rasters can be produced in order to identify areas where the relationship between the dependent variable and each independent variable are strong. GWR has been successfully implemented in Duque de Caxias municipality in the state of Rio de Janeiro, Brazil where it showed how households with access to running water guarded against new cases of leprosy (Duarte-Cunha et al. 2016). GWR has also suggested a significant relationship between climate variables and the spread of hand-foot-and-mouth disease in Shenzhen in China (Zheng et al. 2014). Whilst GWR is able to determine spatial trends between variables in the model, it is unable to predict how these trends may change.

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (2)$$

2.4.2 Geographically and Temporally Weighted Regression

In environmental processes and human activity, relationships between variables may vary not just in space but also in time. The Geographically and temporally weighted regression (GTWR) is model that is able to account for non-stationarity on a spatial and temporal scale (Fotheringham et al. 2015). Simply an extension of GWR, a temporal element t has been included so as to give a space-time location for feature i as is seen in equation (3) (Huang et al. 2010). GTWR is a new

method of modelling spatiotemporal relations, though it has already been used to track changing house prices in Calgary, Canada as well as modelling the key factors involved in carbon dioxide emissions in China (Huang et al. 2010; Liu et al. 2016). It is also considered to be a suitable method to apply in epidemiological research.

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{kn} \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i \quad (3)$$

2.4.3 Point Pattern Analysis

Whilst clusters and point density can be identified visually in a dataset, there is no certainty that these points are statistically significant. The two main methods for detecting statistically significant spatial autocorrelation on a local scale are Hot Spot Analysis using the Getis Ord G_i^* statistic and Anselin Local Moran's I statistic. Getis Ord G_i^* statistic identifies clusters of high values (hotspots) and low values (cold spots) within a weighted dataset in accordance to equation (4) (Getis and Ord 1992). The equation generates a G_i^* value for each feature i for the weight (w) and distance (d) between i and its neighbour j . A hotspot will have a high value and be surrounded by other high value features.

$$G_i^* = \frac{\sum_j w_{ij(d)}x_j}{\sum_j x_j} \quad (4)$$

Anselin's Local Moran's I is also able to detect clusters of high and low values. An advantage with this method is the ability to identify high and low outliers (features surrounded by features with dissimilar values) in a dataset (Sugumaran et al. 2009). Clusters and outliers are detected using equation (5), where once again, w is the distance weight matrix.

$$I_{i(d)} = \frac{(x_i - \bar{x}) \sum_j w_{ij(d)}(x_j - \bar{x})}{\sum_j (x_j - \bar{x})^2 / n} \quad (5)$$

3. Methodology

The following chapter discusses the various data used in the research as well as the different approaches used in terms of data analysis. The workflow of the study can be summarized by the three major tasks: literature review, data preparation and model implementation and analysis. The literature review provided a deeper insight of the disease and the behaviour of its hosts, whilst data preparation manipulated the data such that it could be implemented in the tools used in the statistical analyses and the point pattern analyses. The original intent of the study was to perform multiple regressions, GWR and GTWR on a national level to see which areas were most susceptible to TBE in relation to precipitation and temperature. It was believed that incidence points located far from the endemic region of Stockholm-Mälaren were being interpreted by the GWR model as outliers, thus warping the output. The geographic scope of the study was reduced to the Stockholm-Mälaren region as this would reduce the suspected outliers and provide a higher number of incidence points per unit area. Reducing the size of the study area meant that there was little to no variation in the climate variables. GWR and GTWR were deemed not suitable for the purpose of this study due to the restrictions in temporal and spatial resolution of the TBE incidence data. Instead new methods including Local Moran's I and Getis-Ord G_i^* statistic were used to determine the location of statistically significant spatial autocorrelation in the Stockholm-Mälaren region and directional distribution was implemented in order to determine whether there was any directional trend in the points of incidence.

3.1 Data preparation and sources

Due to the scope of the study, both geographically and temporally, it would have been impossible to collect sufficient primary data within the available time frame. The secondary data used includes diagnosed TBE incidences and climatic data.

3.1.1 TBE data

The Public Health Agency of Sweden (*Folkhälsomyndigheten*) is responsible for the archives regarding the registered cases of TBE infection. Therefore, it was first necessary to send a request for approval before the location of disease incidences between 1986 and 2016 were made accessible. Despite the law for obligatory notification of TBE being enforced in 2004, it is assumed that all cases prior to this have been accounted for.

It was believed that the TBE data would consist of aggregated data at 100m resolution, however this was not the case and due to time restrictions it was not possible to obtain data at a finer scale. The TBE data is comprised of the names of 959 sites and the years in which a TBE infection had

been recorded there. Of these 959 locations, coordinates were only available for 435 of the infection sites. For the remaining sites, the location names were often clear enough to make an educated guess as to where in the country the infection site was located. The coordinates for these sites were extracted by means of the programme Google Earth. 17 infection sites were removed from the data due to location uncertainty.

3.1.2 Climate data

Data for temperature and precipitation were downloaded from the Swedish Meteorological and Hydrological Institute (SMHI) via their Open Data catalogue (SMHI 2017). Mean monthly temperature and mean monthly precipitation data were downloaded for all stations with continuous records between 1986 and 2016. There were 94 weather stations with continuous temperature records and 201 stations with continuous precipitation records for all of Sweden. The values for each month were imported into ArcGIS and point shapefiles for each month were produced. These data points were then interpolated using the inverse distance weighted (IDW) technique resulting in 24 raster layers depicting mean monthly precipitation and temperature respectively for each month. IDW was used for its simplicity and time efficiency; however local climate conditions may not be properly accounted for.

3.1.3 Data processing

The methods for spatial analysis were implemented using the tools and functions found in ArcMap by ESRI with models being run for each individual year in the time series (ESRI 2011). For each incidence of infection, the monthly climate values were extracted from the previously interpolated rasters. This resulted in a point-shapefile for each year between 1986-2016 with fields for mean monthly temperature and precipitation for each location of disease.

3.2 Statistical analyses

This section describes the various statistical tests performed on the data, including multiple regressions and geographic weighted regressions.

3.2.1 Multiple regressions

A multiple regression test was performed in order to determine whether there was any statistical significance in the relationship between the dependent variable of disease incidence and the two independent variables temperature and precipitation. The TBE data was only available on an annual scale (number of incidences per location per year) and so the time-scale of the climate data had to be changed accordingly from monthly averages to an annual average. Literature suggests that ticks hibernate during the winter and only become active once temperatures have surpassed 5-7 degrees C. The onset of TBE infection in humans occurs approximately one week

after the tick bite has occurred (i.e. The symptoms of the disease cannot be drawn out and appear as a result of a bite from the previous season) so it was deemed appropriate that only the climate values of the months where the ticks were active should be analysed. Literature suggests that tick activity can occur between April and November (Dumpis et al. 1999; Gray et al. 2009). This time frame was reduced by one month (April to October), as temperatures in November were often below the set 5 degree limit. The temperatures in April were all above the 5 degree limit and so it was considered appropriate as the start of the active period. Whilst precipitation is vital to maintain the relative humidity needed by the ticks, temperature is considered to be more important in determining when the ticks become active and thus the same active period determined by the temperature values was used for precipitation too.

The average temperature and precipitation for the active period was calculated for every TBE incidence location. Before the regression was executed, both dependent and independent variables were normalized using the equation (6) below. Normalization ensures that the climate variables are proportional to one another and that one variable does not inadvertently affect the results.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6)$$

Multiple linear regressions follow the equation (1) described in section 2.3. The Analysis Toolbox in Microsoft Excel was used to perform all multiple regressions where the confidence level was set to 95% ($p < 0.05$). Linear regressions were performed with temperature and precipitation as independent variables. These two climate variables were then combined and a multiple linear regression was performed. The implementation of the Analysis Toolbox creates a summary output with sections describing the regression statistics and the regression coefficients where adjusted R^2 value and p-value were of most interest. The adjusted R^2 value was used as it accounts for multiple independent variables and adjusts the proportion of variance accordingly. Multiple regressions were also performed on the TBE incidence points in the Stockholm-Mälaren region after the geographical scope of the study was reduced.

3.2.2 Geographically Weighted Regressions

Geographic Weighted Regression (GWR) is based on the basic linear regression and is used when modelling spatial heterogeneity. GWR is a local model meaning that the regression equation is fit to each feature in the dataset, thus allowing the model to account for any regional variation. Temperature, precipitation and TBE incidences were implemented as input variables in the GWR and the default settings for kernel type and bandwidth method were kept as fixed and

Akaike Information Criterion (AICc) respectively. The output results included information regarding the statistical significance of the regression as well as coefficient rasters for all the exploratory variables. However, due to restrictions in the spatial and temporal resolution of the data, and the fact that many years in the temporal dataset contained less than 50 locations of TBE incidence, it was decided that the input data was too poor to perform a good quality Geographic Weighted Regression so this method was discontinued in the study.

3.3 Point pattern analysis

The study progressed into a point pattern analysis of the disease incidences occurring in the Stockholm-Mälaren region. As stated previously, this area was chosen due to the relatively high concentration of disease incidences over space and time. The analysis was composed of three parts: Analysing patterns, mapping clustering and measuring geographic distributions, all of which were implemented for each year in the time series.

3.3.1 Analysing Patterns and Generating Spatial Weights

For analysing patterns, an initial Average Nearest Neighbour (ANN) test was performed in order to detect whether there was any clustering of TBE cases in the region. The ANN is a ratio where the observed mean distance between each point of TBE infection and its nearest neighbour is compared to the expected mean distance if the pattern of TBE infection was random. If the average distance between the points of infection is less than what is expected if the point pattern were to be random, then the features are considered to be clustered.

In order to continue the analysis and map any possible clustering, the input data required weighting. This was first done using the Collect Events function in the Spatial Statistics Toolbox which grouped all individual incidences of infection that shared the same coordinates into a new point layer (Figure 3). The spatial autocorrelation methods require aggregated data, hence it is important that the Collect events function is utilised.

Spatial weighting is a process whereby the spatial relationships between features are quantified, with nearby features having stronger influence than those further away (Getis and Aldstadt 2010). Two methods of spatial weighting, Fixed Distance Band and K-nearest neighbours, were performed so that the results of each could be compared and contrasted. The optimum distance band for each year was determined by means of incremental spatial autocorrelation. The maximum distance required to ensure each feature has a minimum of one neighbour was input for the option 'Beginning Distance', and the observed nearest neighbour distance as the 'Distance Increment' value. The function tests the spatial autocorrelation at increasing distances, producing a line graph depicting the z-score at each distance increment as an output. A large z-score

signifies intense clustering, so only the years with highlighted peak z-score values were used to determine the fixed band distances. For the K-nearest neighbour weighting, a spatial weights matrix for each year was produced using the eight nearest neighbours of each target feature in the computations. These complex computations essentially work by multiplying the number of TBE incidences by the inverse distance value to the feature point of a features eight nearest neighbours (Fukunaga and Narendra 1975; ESRI 2014)

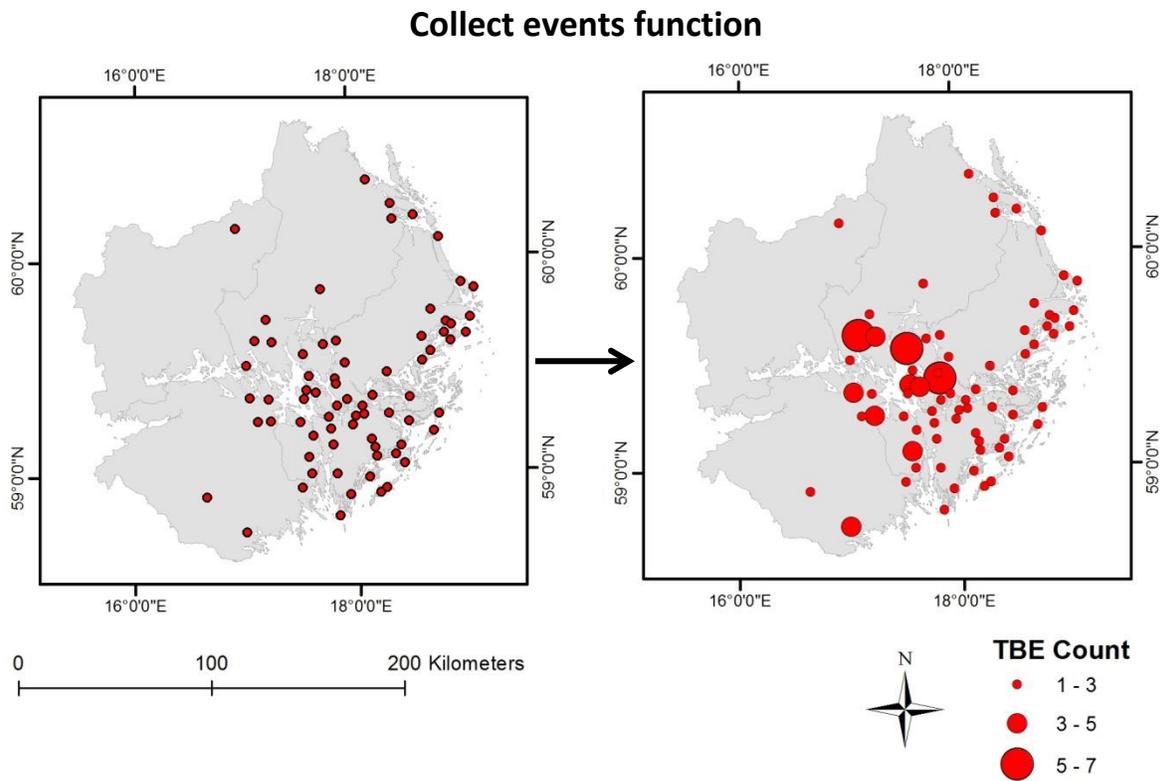


Figure 3 Maps depicting the Collect Events function for 2011. The image on the left shows locations of TBE incidence whilst the image on the right is the result of the Collect Events function and shows locations where multiple cases of TBE have occurred. Projection: RT90_25_gon_W

3.3.2 Mapping Clusters

The Hot Spot Analysis (Getis-Ord G_i^* statistic) function and a Cluster and Outlier analysis (Anselin Local Moran's I statistic) in ArcGIS were two methods used to map clusters in the datasets. Both statistics are used to determine significant positive and negative spatial autocorrelation within a weighted dataset.

The Hot Spot Analysis tool was run for all datasets in the time frame, the first time using the fixed distance band as a method of conceptualising the spatial relationships, and the second time using the spatial weights matrices developed from the K-nearest neighbour analysis. Features

with high values surrounded by neighbour features with high values will resemble hotspots whereas features with low values surrounded by similar neighbours will be cold spots.

The Local Moran's I detects local spatial autocorrelation within a dataset, highlighting clusters of high and low values as well as outliers in the data. The tool, also known as Cluster and Outlier analysis, implemented band-distance weighting as well as k-nearest neighbour weighting, similar to the hotspot analysis method.

3.3.3 Measuring Geographic Distributions

Lastly, a Directional Distribution test was performed as a way of summarizing the changes in spatial distribution of the TBE infection sites during the time frame. Directional distribution tool creates an ellipse based on the standard distance in the x- and y-directions respectively, calculating the standard deviation for X and Y coordinates of each feature. For this study, the ellipse size was set to one standard deviation with stretched or elongated output ellipses suggesting a trend in the direction of the point distribution.

4. Results

The following section describes the results obtained from the methodology including multiple regressions and point pattern analysis over the Stockholm-Mälaren region.

4.1 Multiple regressions

Table 1 and 2 represent years where linear regressions showed any statistical significance when trying to explain TBE incidence. Each column represents the p-value obtained for each independent variable each year as well as the coefficients for the variables, with shading indicating significant values.

When looking at the results of the national dataset, there are nine years in the temporal dataset that showed no significant relationship with TBE incidence. All but two of these years occurred during the nineties. 7 years showed significant relationships with temperature; 18 years showed significant relationships with precipitation and 5 years showed significant relationships with temperature and precipitation combined.

After the study area was limited to the Stockholm-Mälaren region (Figure 1), only 15 years showed statistically significant relationships between the independent variables and the incidence of TBE infection. When compared to the national level, many years show significance for similar variables however, the years 1999 and 2005 show significant temperature relationships which are not found at the national level.

In all but one case, the coefficient signifies a negative trend though the R^2 values remained below 0.05 which shows that the points do not clearly follow the pattern of the regression trend line. For the combined variables, the adjusted R^2 values were often negative meaning that the sample size is too small to determine trends with multiple predictors.

Table 1: Years in the national dataset where the relationship between disease incidence and the independent variables show statistical significance. Significant values are highlighted in grey and Adj R² represents the adjusted R² value.

Year	Temperature			Precipitation			Combined Variables	
	p-value	coefficient	Adj. R ²	p-value	coefficient	Adj. R ²	p-value	Adj. R ²
1986	0.007	-0.379	0.035	0.175			0.044	0.012
1987	0.108			0.021	-0.198	-0.006	0.065	
1988	0.011	-0.373	0.113	0.030	-0.318	0.042	0.012	0.083
1989	0.680			0.042	-0.160	-0.015	0.904	
1990	0.091			0.007	-0.360	0.049	0.022	0.028
1991	0.005	-0.375	0.040	0.334			0.127	
1997	0.012	-0.111	-0.021	0.042	-0.153	-0.020	0.047	-0.032
2000	0.024	-0.127	-0.006	0.232			0.144	
2001	0.016	-0.142	0.016	0.205			0.062	
2002	0.062			0.016	0.071	-0.015	0.161	
2003	0.365			0.003	-0.100	0.018	0.130	
2004	0.390			0.003	-0.050	-0.009	0.333	
2006	0.026	-0.092	-0.005	0.001	-0.130	0.009	0.015	-0.003
2007	0.095			0.000	-0.101	0.001	0.059	
2008	0.966			0.000	-0.145	0.014	0.825	
2009	0.486			0.005	-0.262	0.008	0.244	
2010	0.281			0.001	-0.089	0.002	0.142	
2011	0.484			0.000	-0.055	0.006	0.226	
2013	0.434			0.000	-0.070	0.003	0.218	
2014	0.447			0.000	-0.235	0.019	0.251	
2015	0.229			0.001	-0.072	0.012	0.052	
2016	0.511			0.000	-0.115	0.015	0.145	

Table 2: Years in the Stockholm-Mälaren dataset where the relationship between disease incidence and the independent variables show statistical significance. Significant values are highlighted grey and Adj R² represents the adjusted R² value.

Year	Temperature			Precipitation			Combined Variables	
	p-value	coefficient	Adj. R ²	p-value	coefficient	Adj. R ²	p-value	Adj. R ²
1986	0.006	-0.391	0.039	0.128			0.030	0.015
1987	0.087			0.038	-0.164	-0.013	0.073	
1988	0.012	-0.379	0.117	0.024	-0.372	0.062	0.012	0.091
1990	0.165			0.003	-0.367	0.114	0.003	0.138
1991	0.018	-0.247	0.033	0.077			0.030	0.006
1997	0.019	-0.128	-0.023	0.024	-0.218	-0.012	0.032	-0.018
1999	0.020	-0.577	0.185	0.993			0.096	
2002	0.033	-0.088	-0.020	0.613			1.000	
2005	0.043	-0.119	-0.008	0.065			0.117	
2007	0.025	-0.173	-0.014	0.000	-0.362	0.030	0.003	0.039
2008	0.450			0.009	-0.032	-0.009	0.988	
2009	0.500			0.029	-0.046	-0.005	0.577	
2010	0.046	-0.058	-0.008	0.250			0.220	
2011	0.802			0.037	-0.000	-0.007	0.787	
2014	0.915			0.000	-0.219	0.046	0.128	

4.2 Point Pattern Analysis

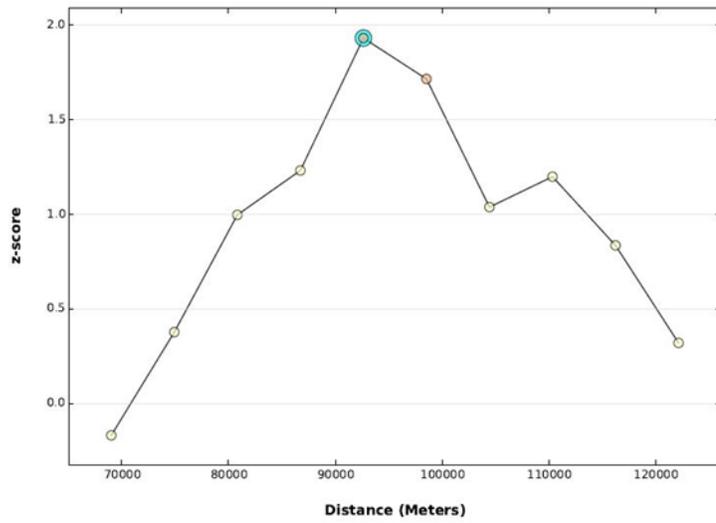
The following section describes the results of the point pattern analysis of TBE over the Stockholm-Mälaren region.

4.2.1 Incremental Spatial Autocorrelation

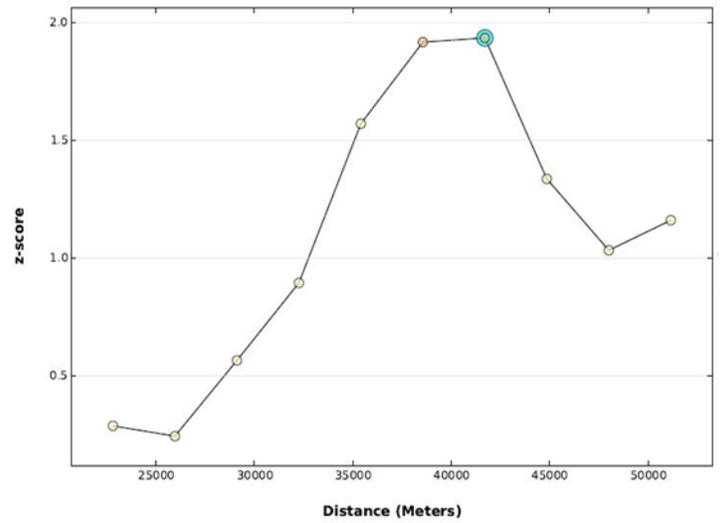
When determining the values for the fixed distance bands, an incremental spatial autocorrelation was performed to determine years with significant peak z-scores (Figures 4-6). For all the years explored, only 10 showed significant peak z-scores with values over 1.65. Majority of the significant band distances fall between 45-60 km. In 2009 and 2013 when two peaks were significant, the peak with the greater distance was chosen to represent the distance band.

Incremental Spatial Autocorrelation

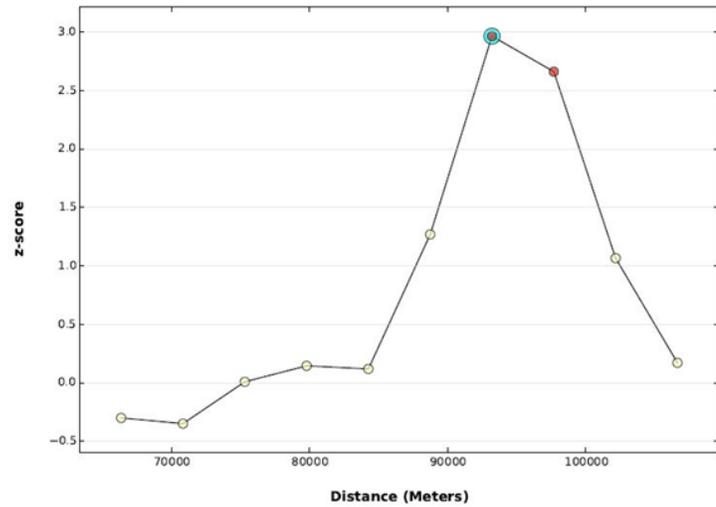
1993



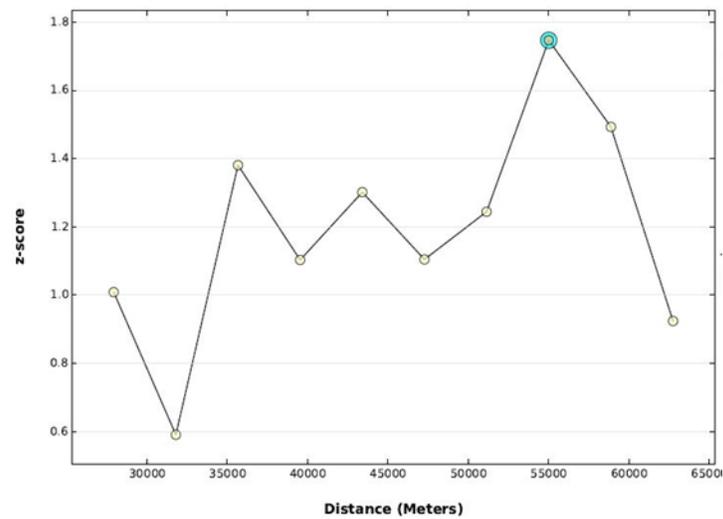
1997



2001



2005



z-score

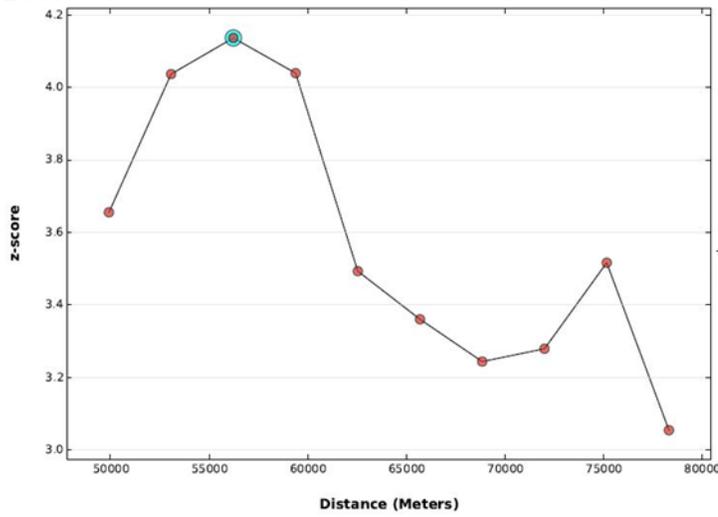
- > 2.58
- 1.96 - 2.58
- 1.65 - 1.96
- -1.65 - 1.65
- -1.96 - -1.65
- -2.58 - -1.96
- < -2.58

● Peaks

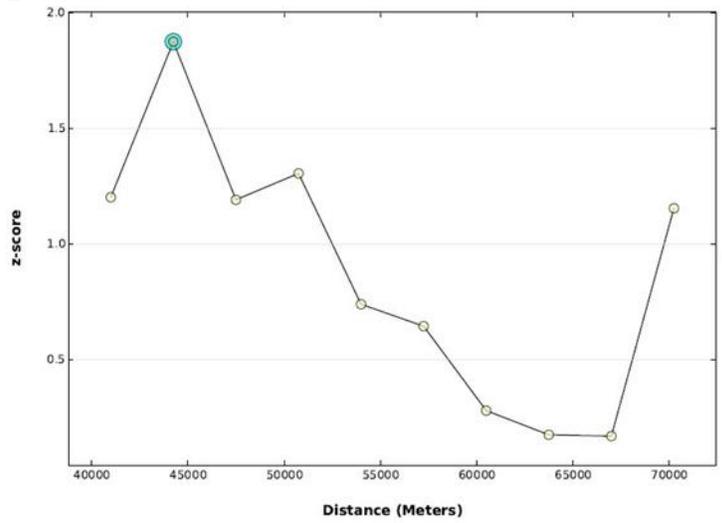
Figure 4 Results of the incremental spatial autocorrelation for years 1993, 1997, 2001 and 2005 where a significant peak z-value was obtained. Z-scores >1.65 indicate a significant degree of clustering in the data thus only these years were used in determining fixed band distance values.

Incremental Spatial Autocorrelation

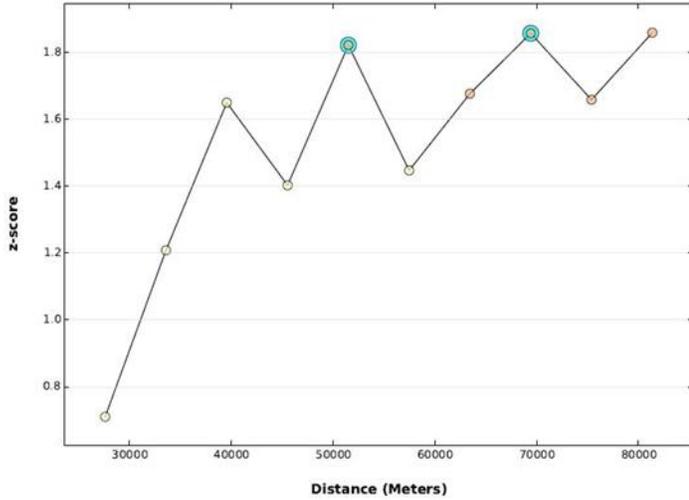
2007



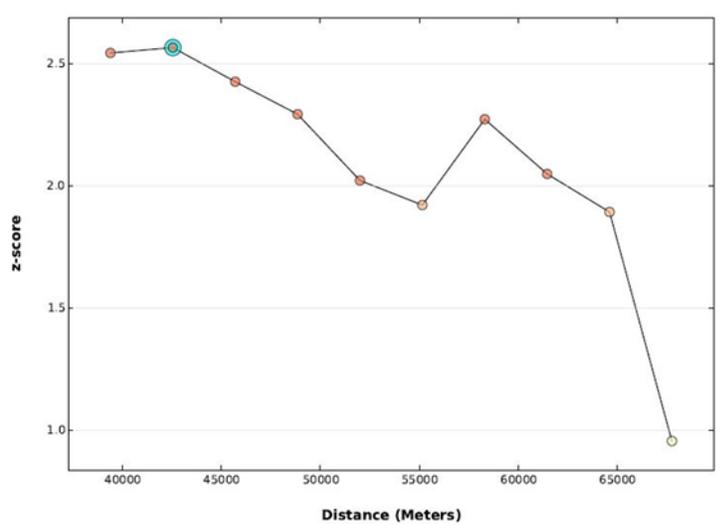
2008



2009



2011



z-score

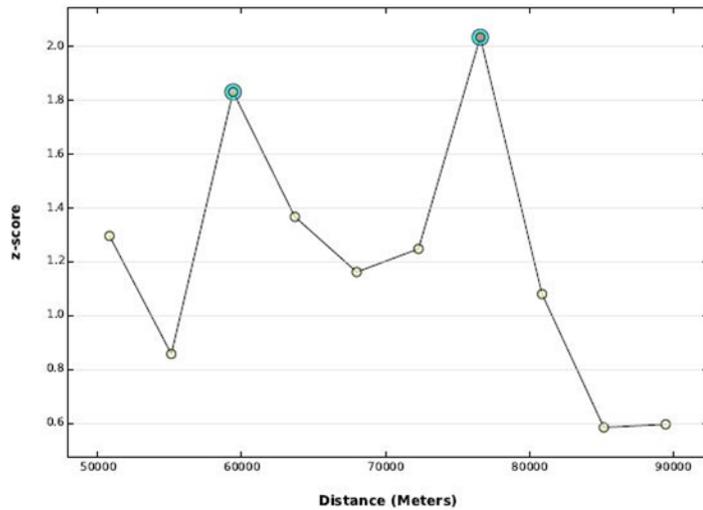
- > 2.58
- 1.96 - 2.58
- 1.65 - 1.96
- -1.65 - 1.65
- -1.96 - -1.65
- -2.58 - -1.96
- < -2.58

○ Peaks

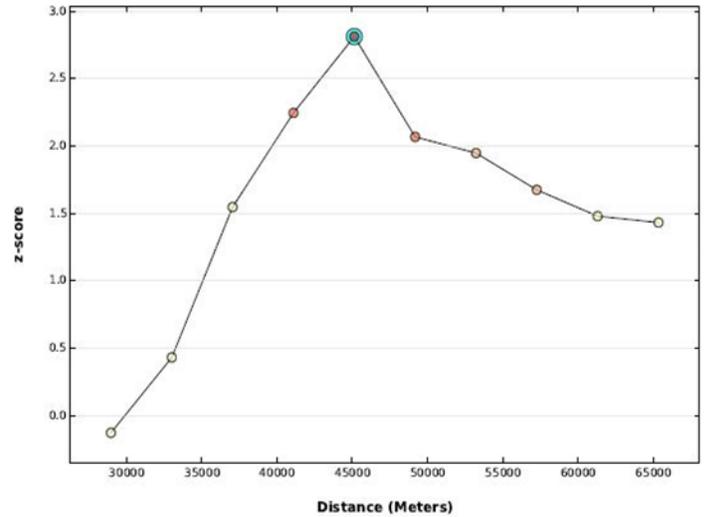
Figure 5 Continued results of the incremental spatial autocorrelation for years 2007, 2008, 2009 and 2011 where a significant peak z-value was obtained. Z-scores >1.65 indicate a significant degree of clustering in the data thus only these years were used in determining fixed band distance values.

Incremental Spatial Autocorrelation

2013



2015



z-score

- > 2.58
- 1.96 - 2.58
- 1.65 - 1.96
- -1.65 - 1.65
- -1.96 - -1.65
- -2.58 - -1.96
- < -2.58

● Peaks

Figure 6 Continued results of the incremental spatial autocorrelation for years 2013 and 2015 where a significant peak z-value was obtained. Z-scores >1.65 indicate a significant degree of clustering in the data thus only these years were used in determining fixed band distance values.

4.2.2 Hotspot Analysis (Getis-Ord G_i^*) for fixed distance band

The incremental spatial autocorrelation performed for the band distance weighting suggested that there were only ten years with a significant peak z-score distance, and thus hot spot analysis with distance band weighting was only performed for these 10 years (Figure 7-8). For the years 2005, 2007, 2011 and 2015 there are a large number of hotspots around the Lake Mälaren with confidence above 95% suggesting that this is an area with a high proportion of TBE infections. In 2005, hotspots were only located in Uppsala county and in 2007 hotspots are found in the two easterly counties, Stockholm and Uppsala, as well as the Stockholm-Södermanland border. It is not until 2011 when the hotspots around Lake Mälaren are present in all four counties and in 2015 the clustering around the lake can be found predominantly in Södermanland and Västmanland. In 2013 there are a number of hotspots ($p < 0.1$) located on the southern coast of Stockholm country. Significant cold spots ($p < 0.1$) are most commonly found along the east coast of Stockholm County in the north of the archipelago. There is an increase in the number of significant clusters over time.

Hot Spot Analysis- Band distance weighting

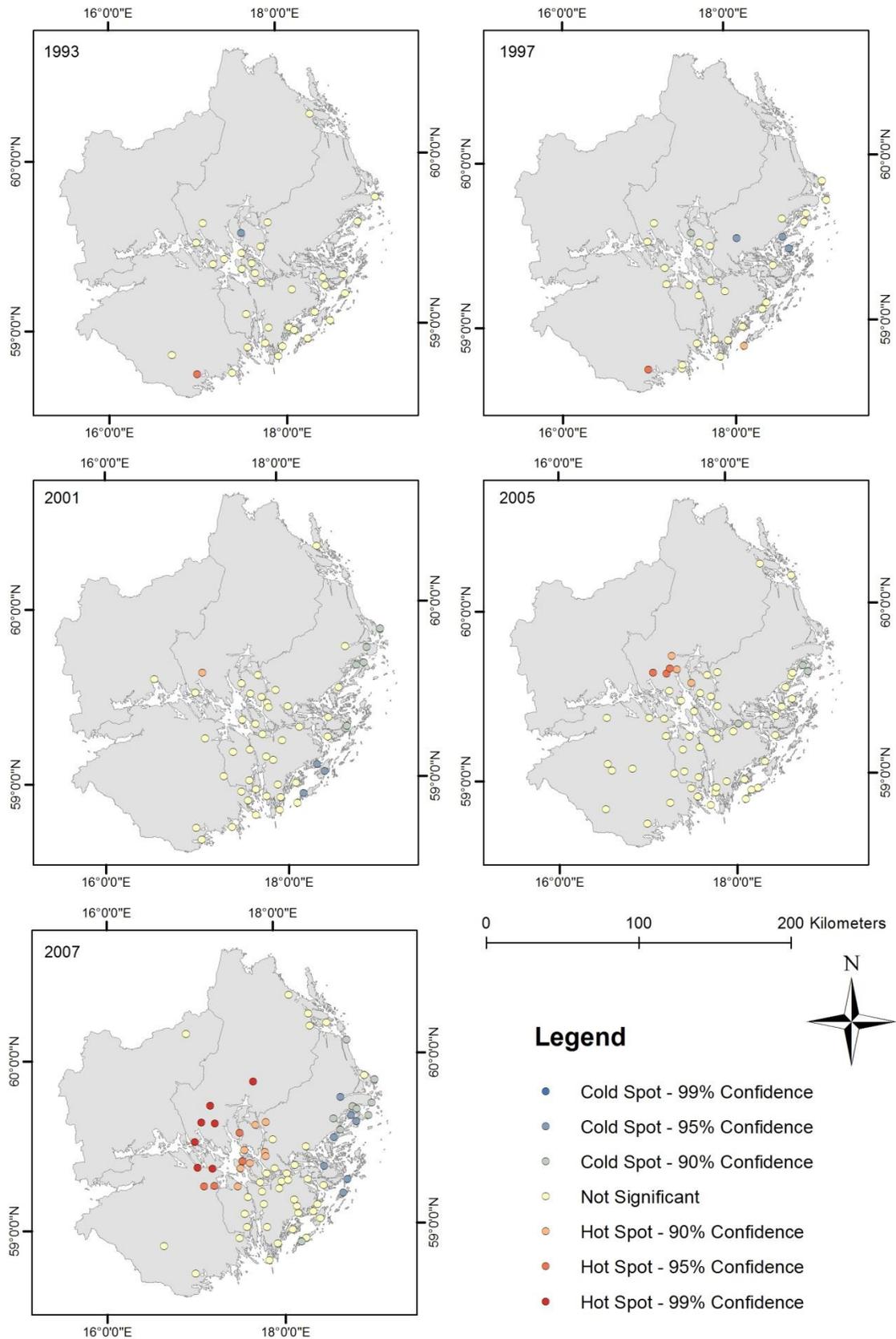


Figure 4 Results of the Hotspot Analysis using Getis-Ord G_i^* with fixed band distance weighting in the Stockholm-Mälaren region. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- Band distance weighting

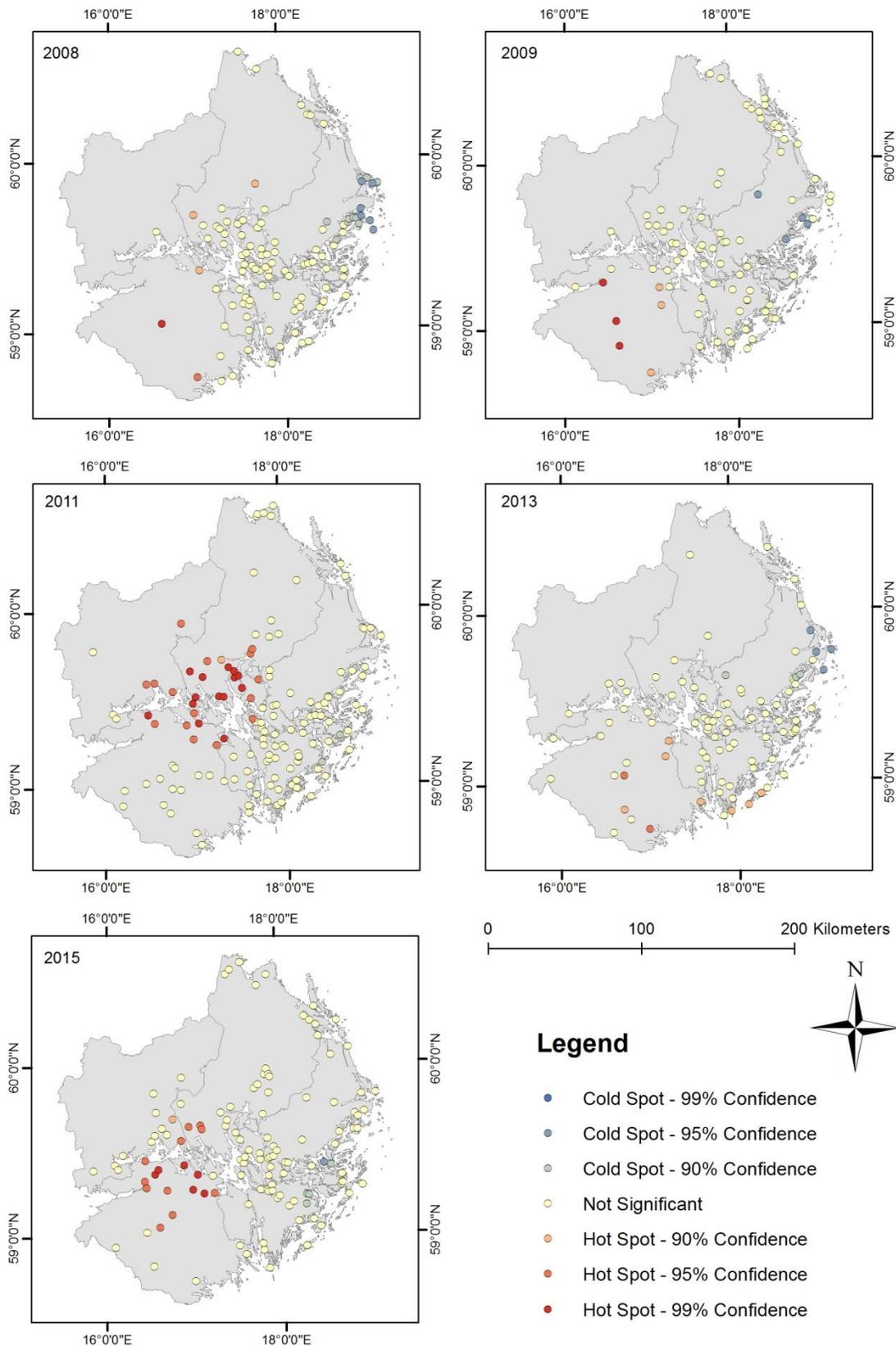


Figure 5 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with fixed band distance weighting in the Stockholm-Mälaren region. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

4.2.3 Hotspot Analysis (Getis-Ord G_i^*) for K-nearest neighbour

Years of significant clustering using K-nearest neighbour weighting are depicted in Figure 9-14. In the 1990's, distinct hotspots are found in the south of Stockholm county near the port Nynäshamn. This was also true for hotspots in 2002 and 2013. After the year 2000, hotspots start becoming increasingly prominent around Lake Mälaren with the first incidence of a 99% confident hotspot occurring in 2004 in this area. In 1987, 1989, 1994, 1996 and 1998, there are also incidences of hotspots occurring around Mälaren, though these are neither to the same extent nor confidence level as those occurring after the millennium. In 1997, 2005 and 2007 there are significant cold spots located in the north-east of Stockholm County. Whilst majority of the TBE incidence points are classed as not significant, in all but 5 years hotspots account for 5-10% of the total number of points. The number of significant hotspots increases over the timeframe, though a significant hotspot does not occur in Västmanland until 2007.

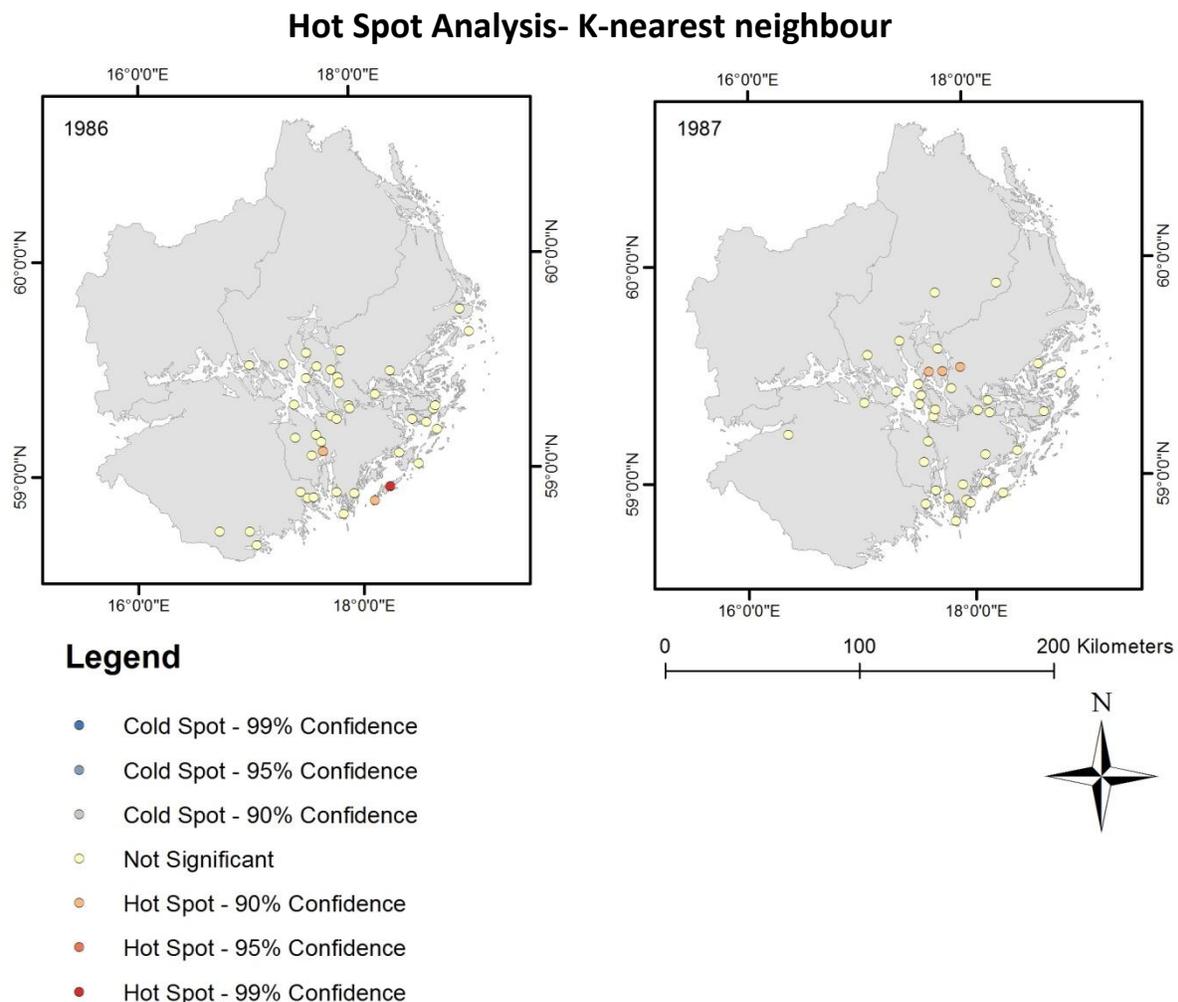


Figure 6 Results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- K-nearest neighbour

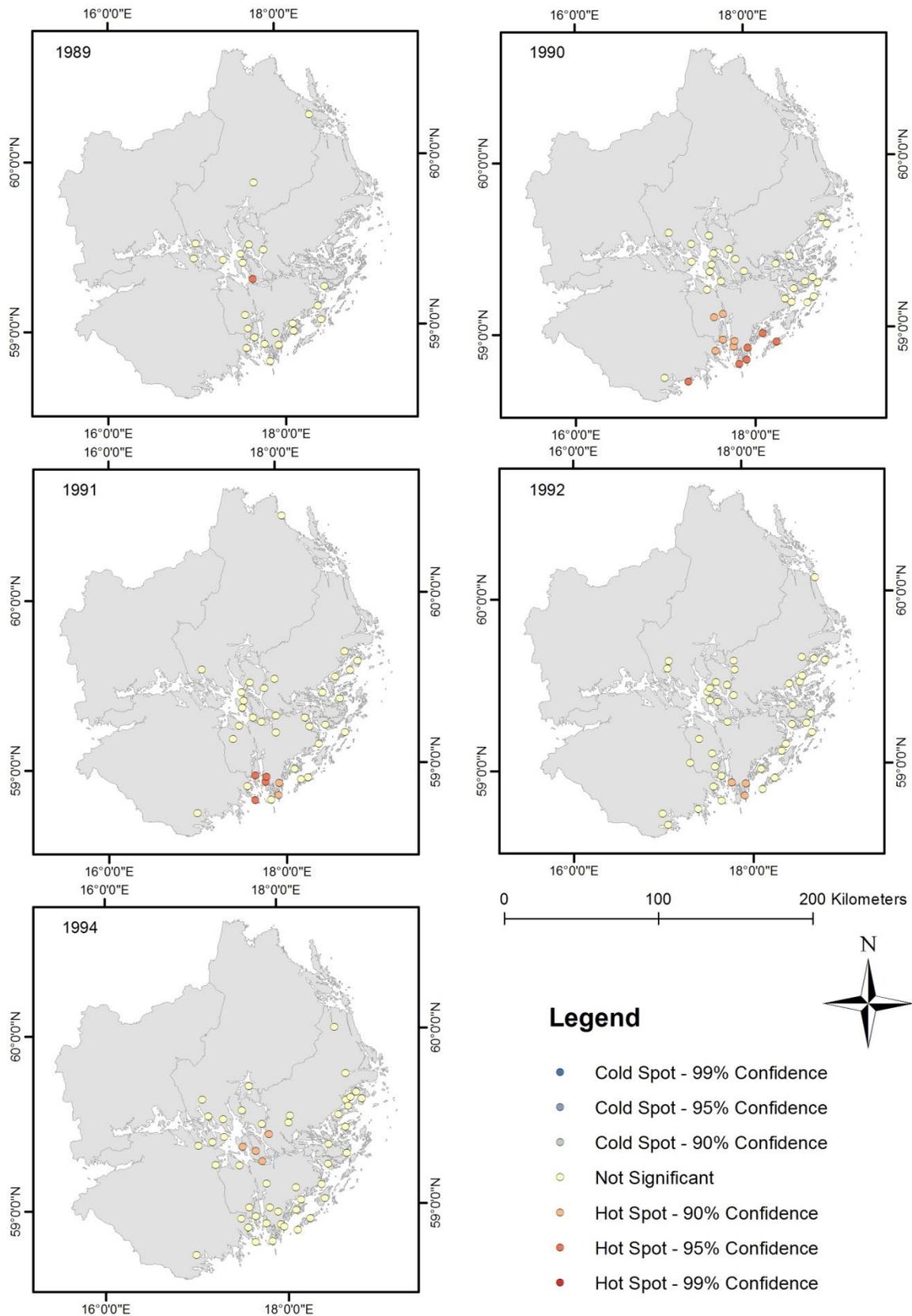


Figure 7 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region for years 1989, 1990, 1991, 1992 and 1994. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- K-nearest neighbour

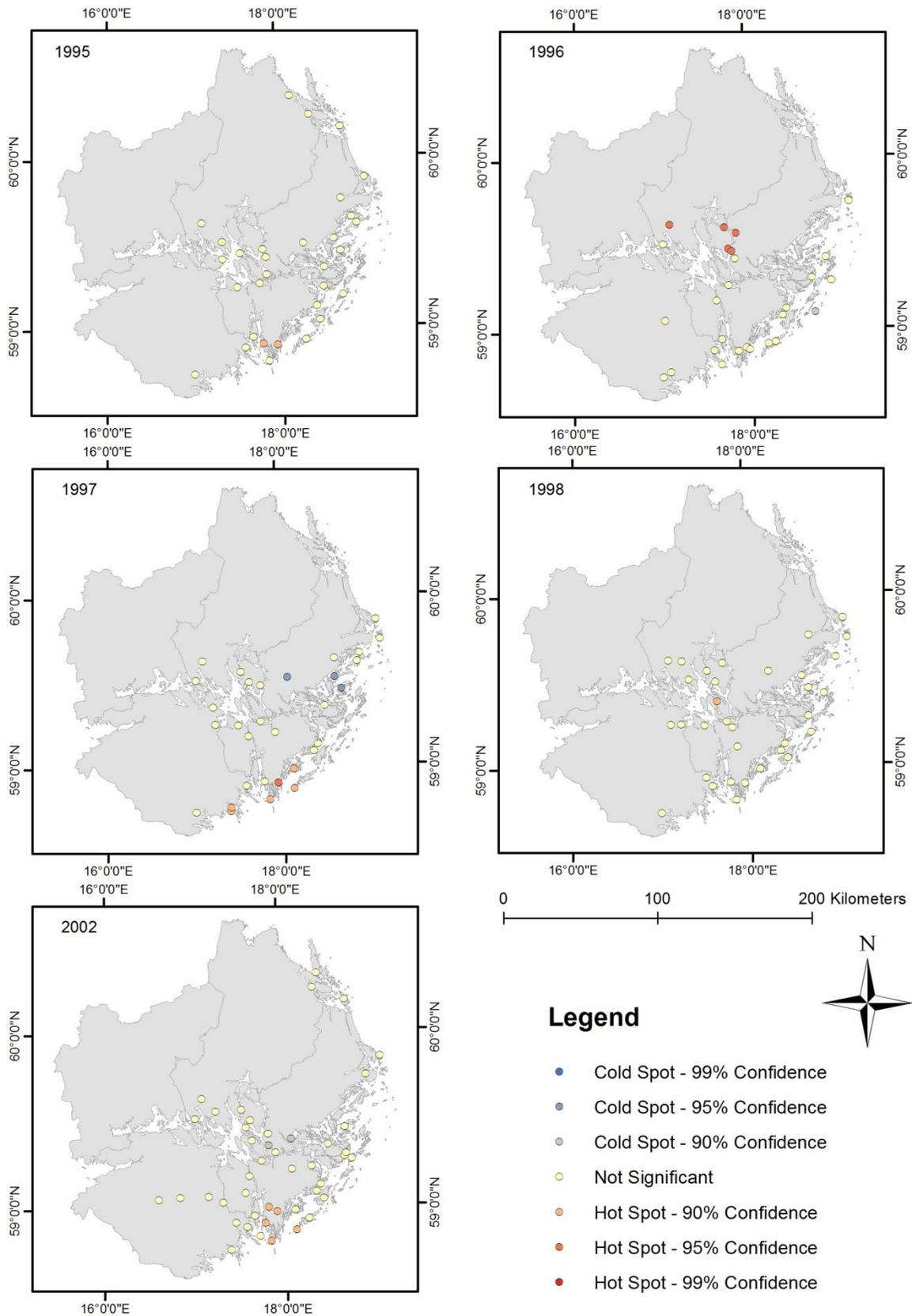


Figure 8 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region for years 1995, 1996, 1997, 1998 and 2002. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- K-nearest neighbour

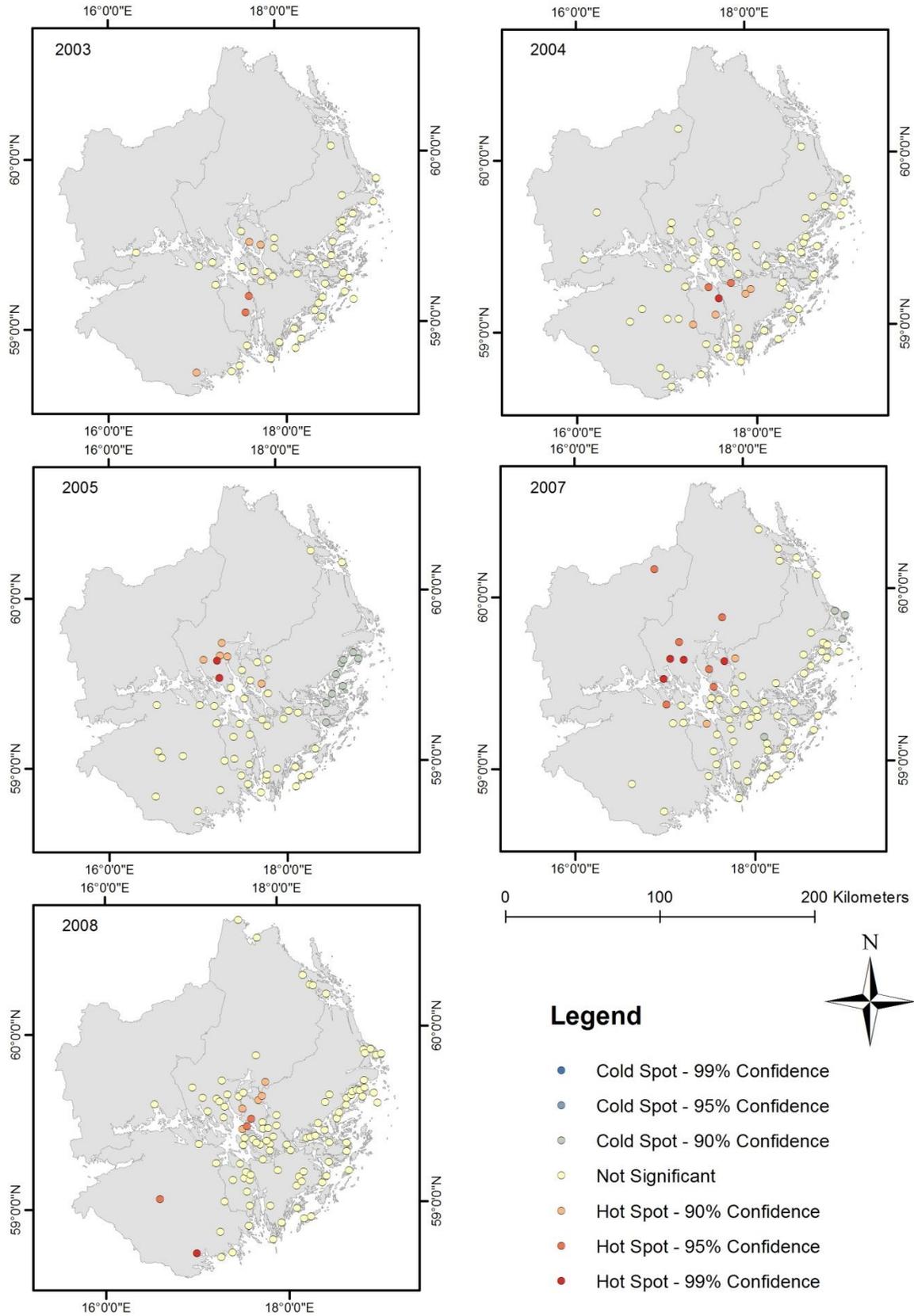


Figure 9 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region for years 2003, 2004, 2005, 2007 and 2008. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- K-nearest neighbour

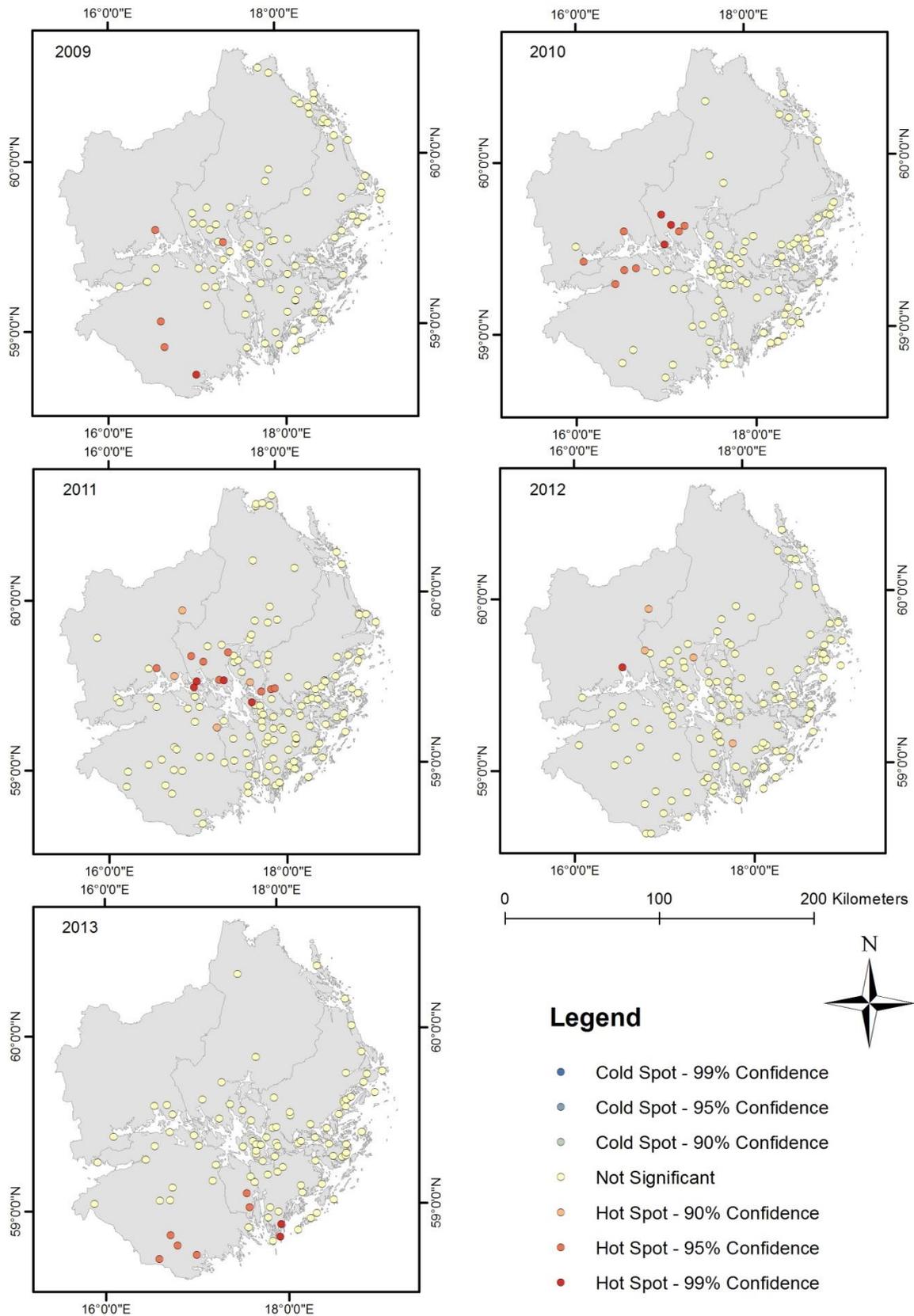


Figure 10 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region for years 2009, 2010, 2011, 2012 and 2013. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

Hot Spot Analysis- K-nearest neighbour

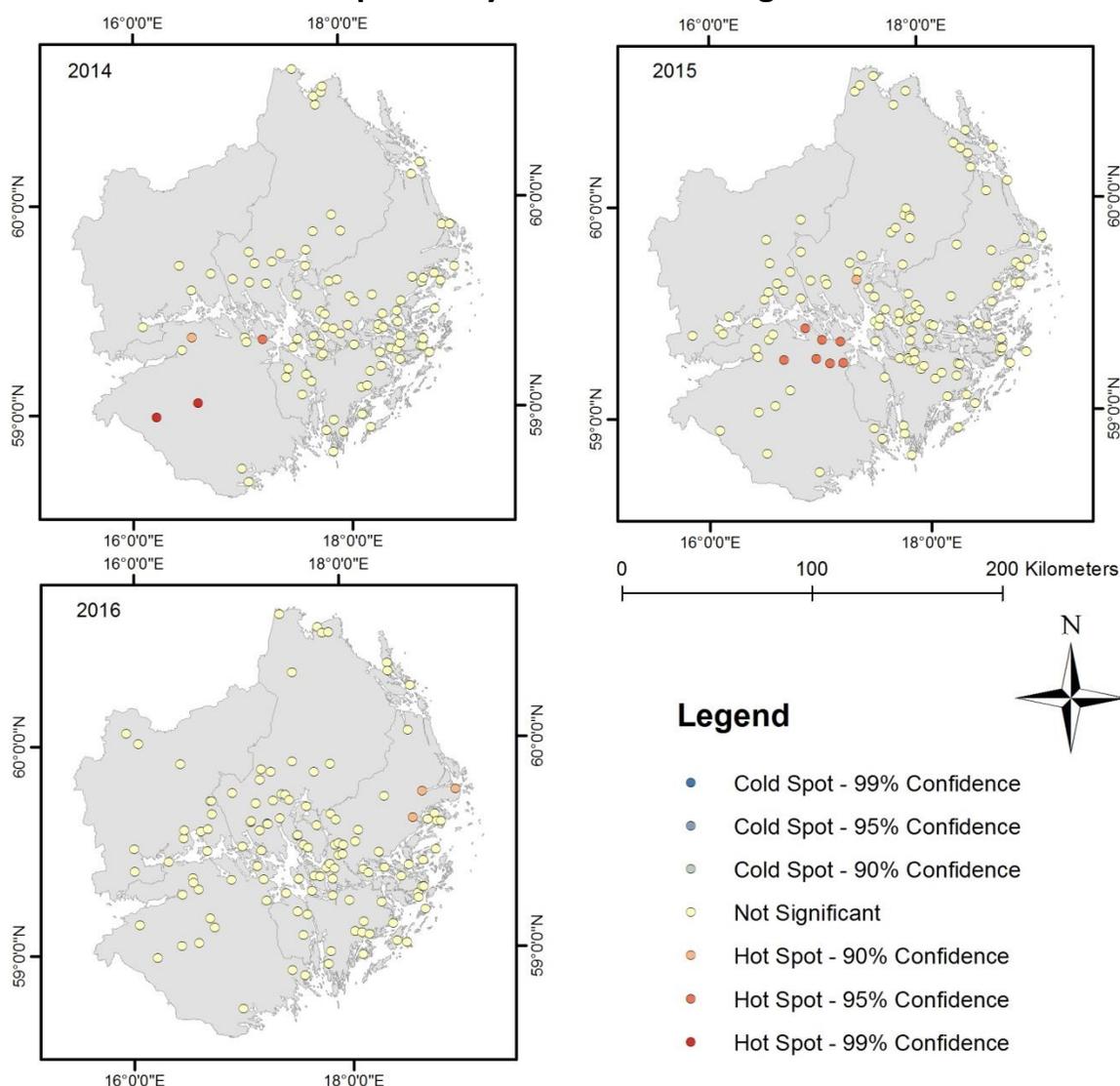


Figure 11 Continued results of the Hotspot Analysis using Getis-Ord G_i^* with KNN weighting in the Stockholm-Mälaren region for years 2014, 2015 and 2016. Orange points show TBE hotspots whilst blue points show TBE cold spots. Projection: RT90_25_gon_W

4.2.4 Local Moran's I for fixed distance band

As mentioned previously, 10 band-distances were obtained through the incremental spatial autocorrelation, however neither 1997 nor 2001 showed any significant clustering when performing the Moran's I local statistic (Figure 15-16). Local Moran's I identifies significant clusters around Lake Mälaren predominantly, though in 2013 significant clusters were found on the south east coast of Stockholm and Södermanland counties. For band-distance weighting, outliers have occurred in the Stockholm archipelago and once in the south of Uppsala County. 2015 is the first year in which a significant cluster occurs in Västmanland.

Local Moran's I- Fixed distance band

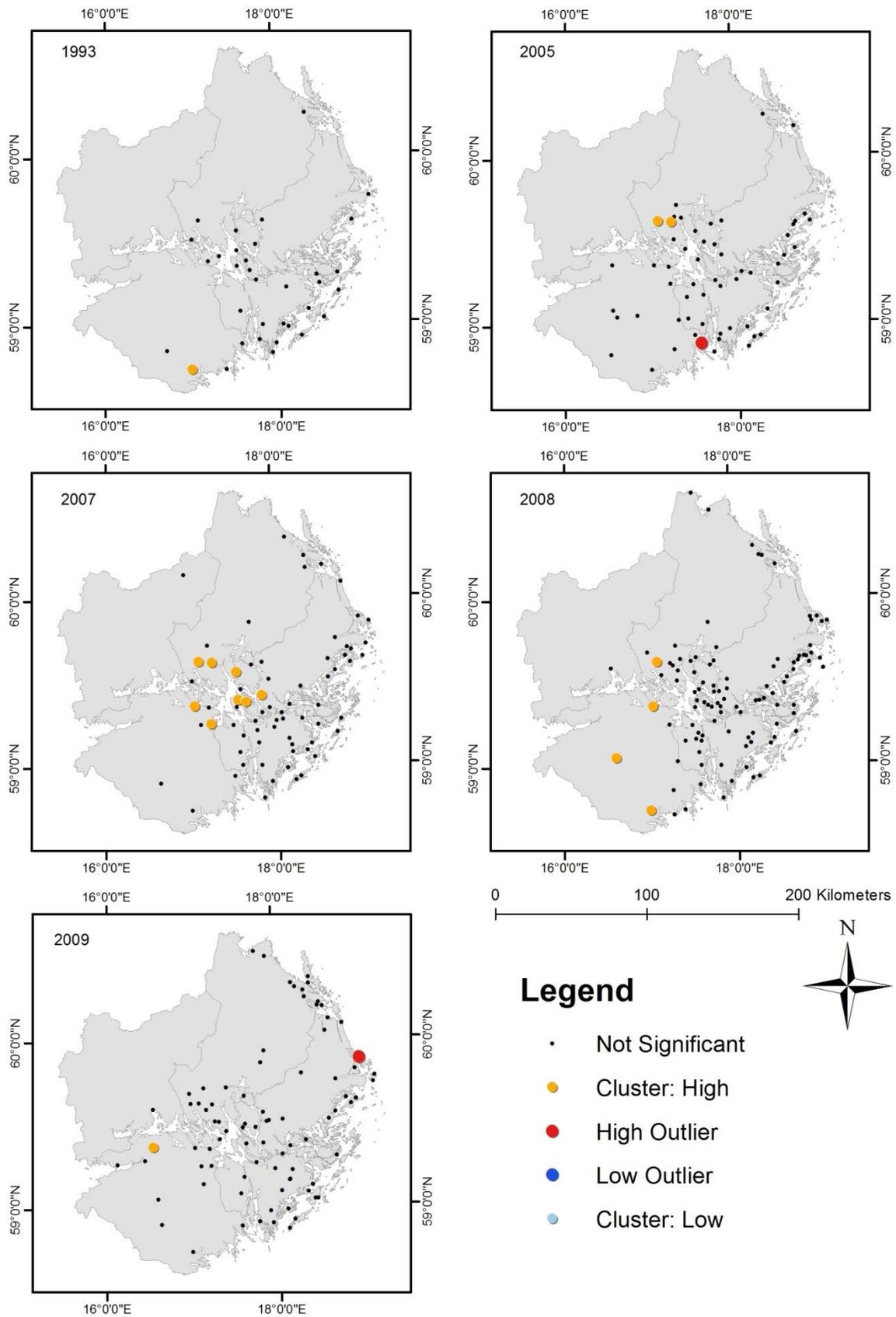


Figure 12 Results of the Local Moran's I with fixed distance band weighting in the Stockholm-Mälaren region for years that showed significant clustering including 1993, 2005, 2007, 2008 and 2009. Projection: RT90_25_gon_W

Local Moran's I- Fixed distance band

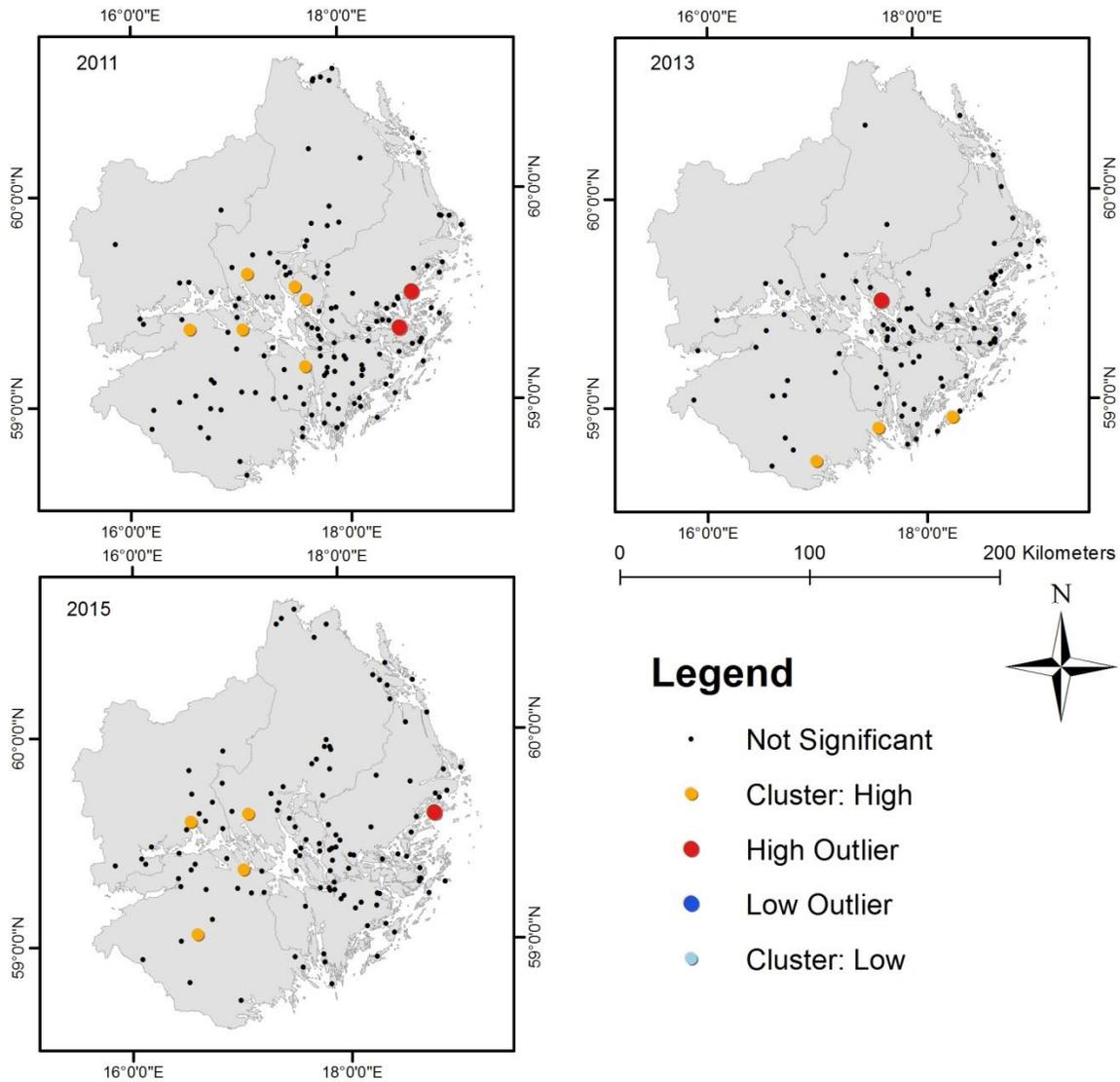


Figure 13 Results of the Local Moran's I with fixed distance band weighting in the Stockholm-Mälaren region for years that showed significant clustering including 2011, 2013 and 2015. Projection: RT90_25_gon_W

4.2.5 Local Moran's I for K-nearest neighbour

The maps below include only the years where clustering occurred (Figure 17-20). Using the K-nearest neighbour weighting, only 1993 and 1999 showed neither clustering nor outliers. There are 12 years, 1989, 1992, 1995-1998, 2000, 2001, 2003, 2006, 2015 and 2016, where no clustering occurred but an outlier was present. The maps depict 12 years where at least one high cluster point was found at Lake Mälaren and 5 years where at least one high cluster point was found in the Nynäshamn area in southern Stockholm County. 2007 is the only year where a low outlier occurs, with the point being found on Lake Mälaren in southern Uppsala County. The first cluster point in Västmanland occurs in 2010. There has been an increase in clustering over time, though Getis-Ord G_i^* detects a larger number of significant clusters in comparison.

Local Moran's I- K-nearest neighbour

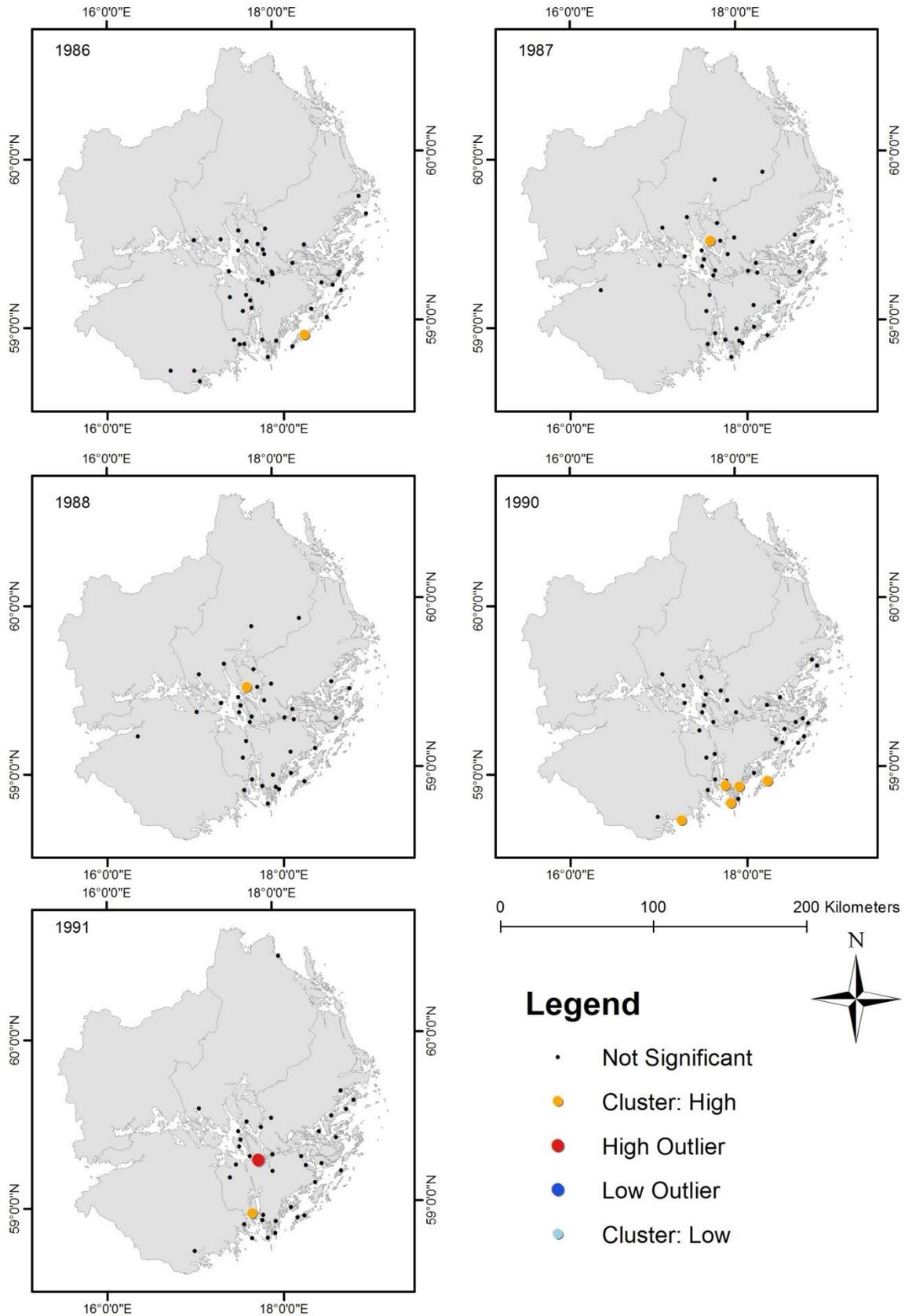


Figure 14 Results of the Local Moran's I with KNN in the Stockholm-Mälaren region for years that showed significant clustering including 1986, 1987, 1988, 1990 and 1991. Projection: RT90_25_gon_W

Local Moran's I- K-nearest neighbour

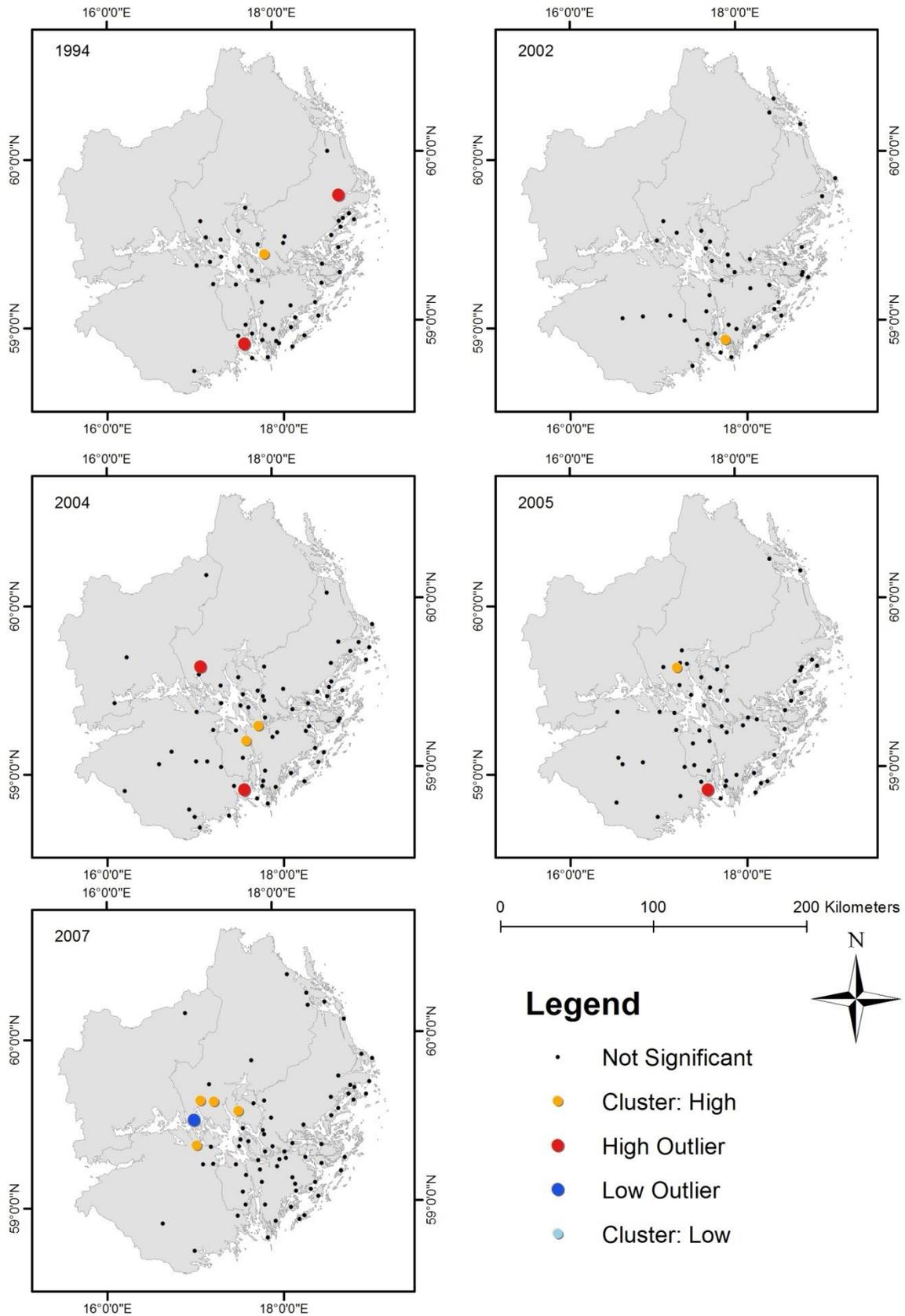


Figure 15 Continued results of the Local Moran's I with KNN in the Stockholm-Mälaren region for years that showed significant clustering including 1994, 2002, 2004, 2005 and 2007. Projection: RT90_25_gon_W

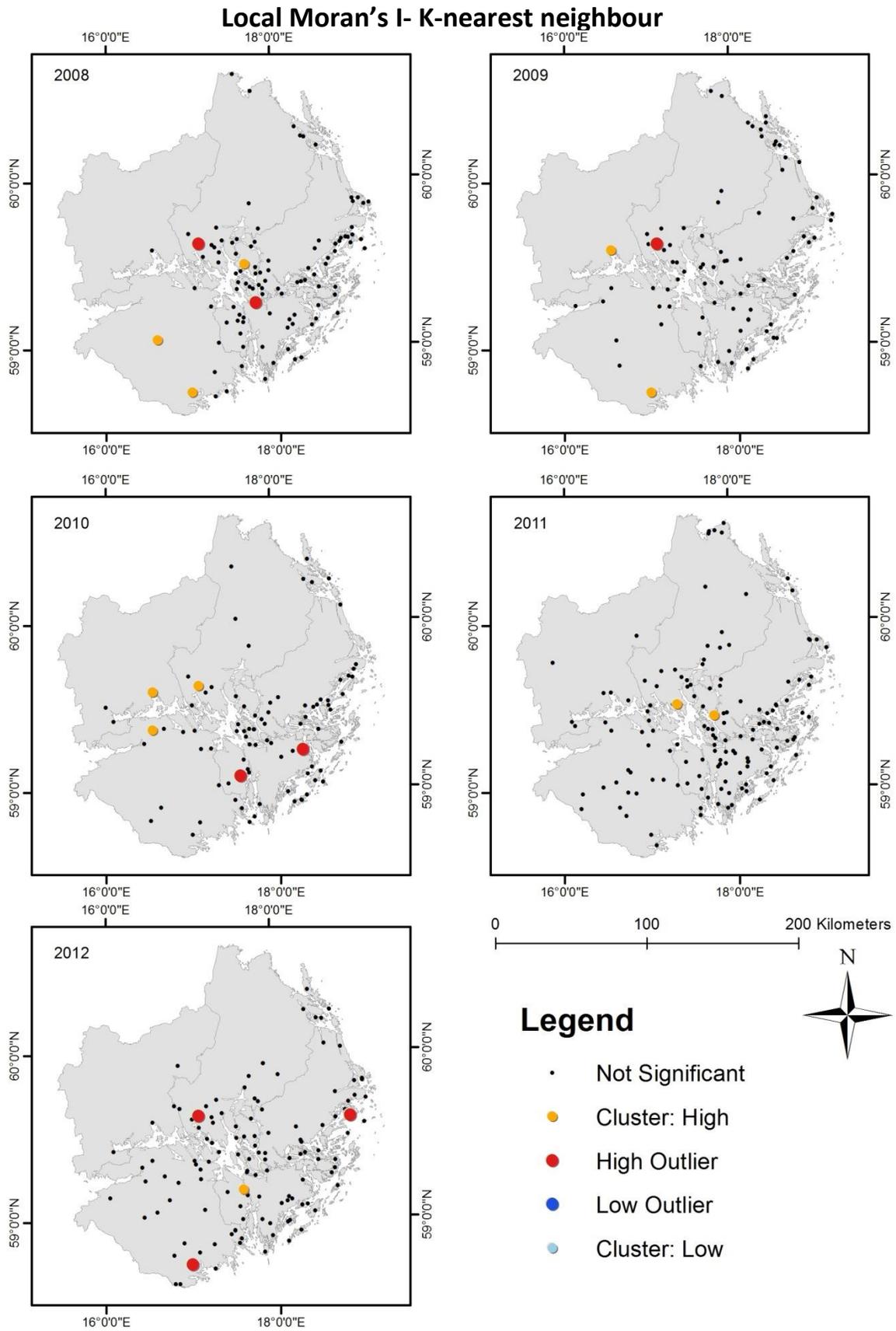


Figure 16 Continued results of the Local Moran's I with KNN in the Stockholm-Mälaren region for years that showed significant clustering including 2008, 2009, 2010, 2011 and 2012. Projection: RT90_25_gon_W

Local Moran's I- K-nearest neighbour

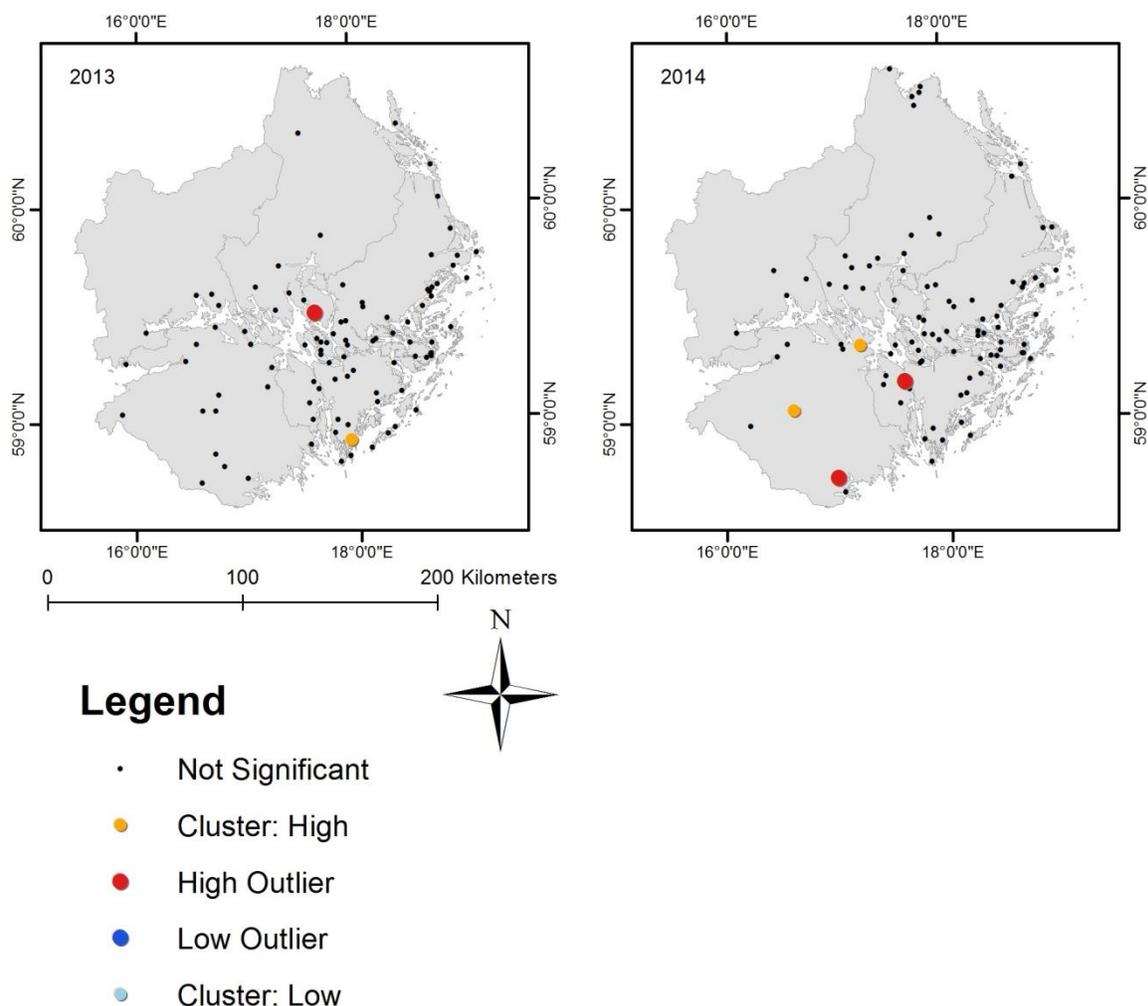


Figure 17 Continued results of the Local Moran's I with KNN in the Stockholm-Mälaren region for years that showed significant clustering including 2013 and 2014. Projection: RT90_25_gon_W

4.2.6 Directional distribution

The standard deviational ellipse was used to determine the trends in orientation for the cases of TBE incidence over the 30 year time frame (Figure 21-22). Long narrow ellipses suggest a stronger trend in the data. The ellipses are centered around the mean center of all the features in the dataset for any given year in the temporal dataset. All of the ellipses are located on the coast of Stockholm County and incorporate part of the Stockholm archipelago. The ellipses increase in size with each year in accordance to the increasing number of disease incidents.

In 1988 and 1989, the directional trend moves in a southeast-northwest direction between southern Stockholm county towards the east of Lake Mälaren. The general directional trend amongst the remaining years is in the northeast-southwest direction and is particularly evident in 1995 and 2009. The most discernable trends can be found between 1993 and 2010 where the

direction of the points is along the coast of Stockholm County. Between 2011 and 2016 there is a less defined trend in the direction of the points distribution which is depicted by the round shape of the ellipses.

Directional distribution with standard deviational ellipses

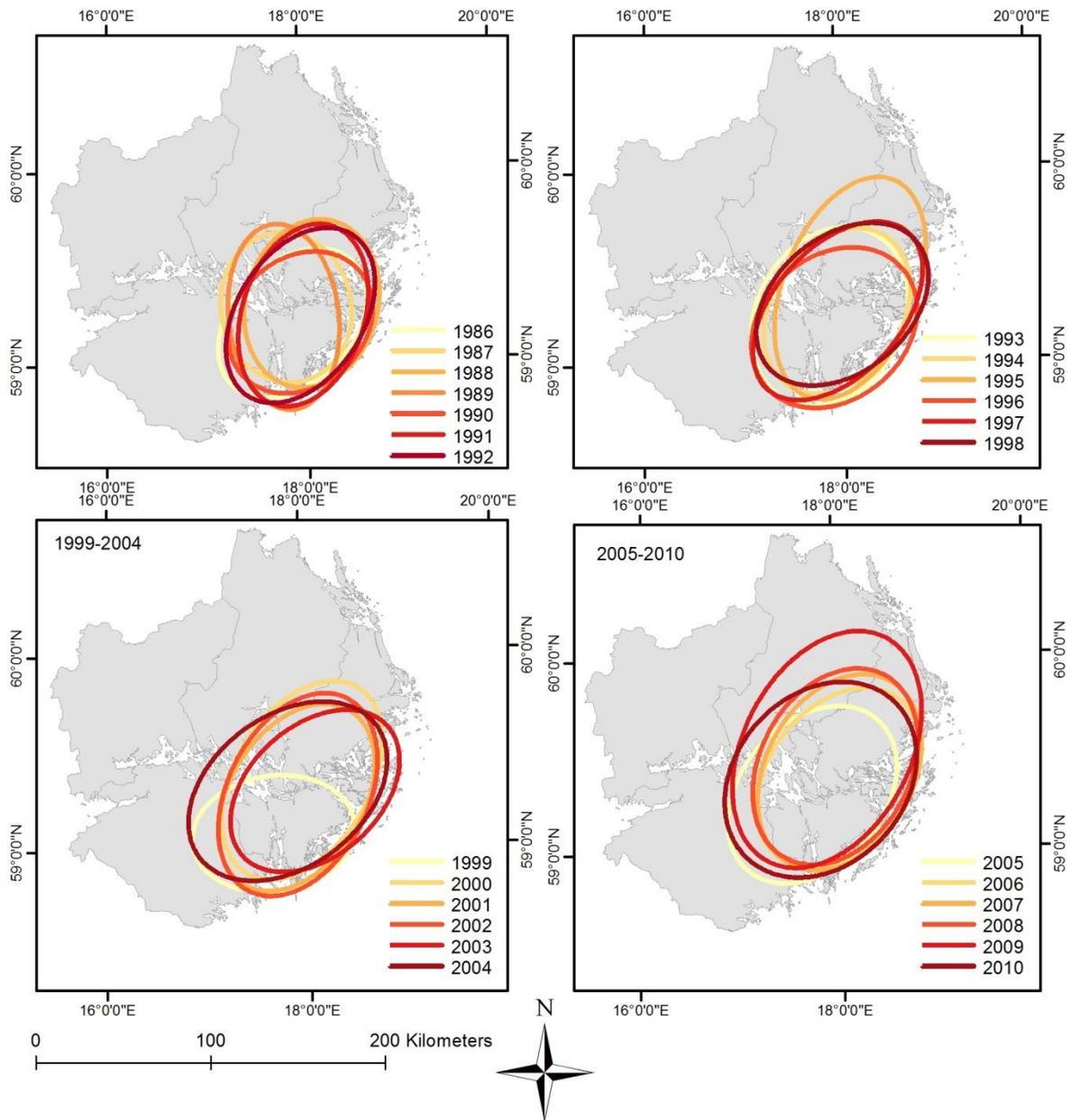


Figure 18 Results of directional distribution with standard deviational ellipses for years 1986-2010. Each ellipse accounts for 1 standard deviation with long narrow ellipses showing more defined directional trend. Projection: RT90_25_gon_W

Directional distribution with standard deviational ellipses

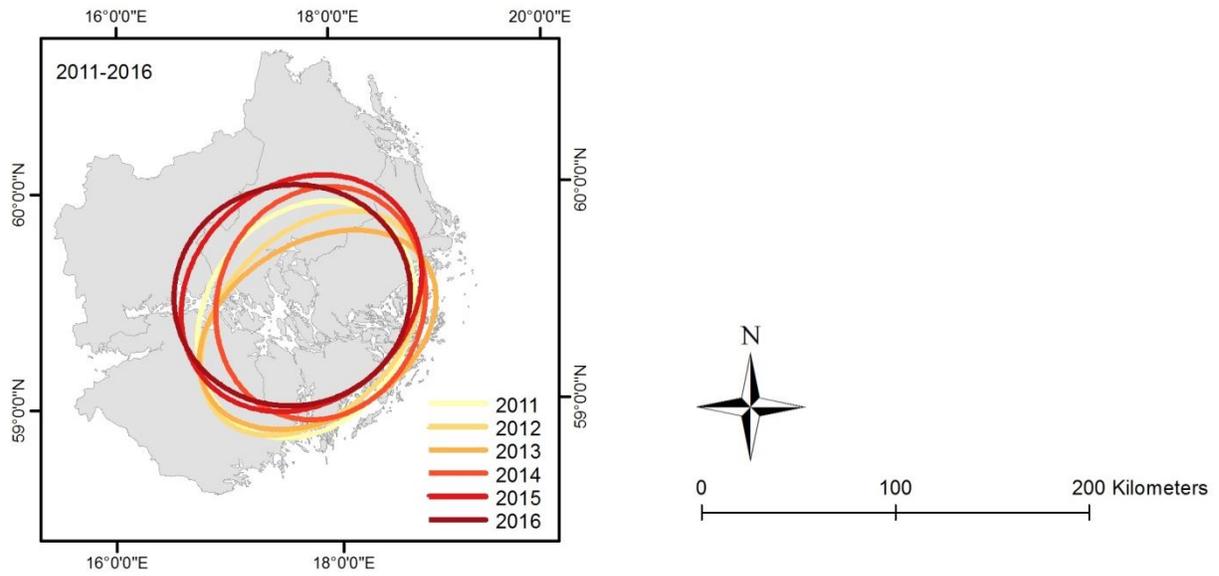


Figure 19 Continued results of directional distribution with standard deviational ellipses for years 2010-2016. Each ellipse accounts for 1 standard deviation with long narrow ellipses showing more defined directional trend.
Projection: RT90_25_gon_W

5. Discussion

The purpose of this study has been firstly to see whether there is a statistically significant relationship between climatic factors and TBE infection, and secondly to see whether the distribution of TBE incidence displays any evident spatial patterns. Few of the explored years showed statistical significance in relation to TBE and the climate variables, though it is believed that with more data at a finer resolution there will be more years depicting statistically significant relationships with climate variables. When investigating spatial patterns, the clustering found in the Stockholm-Mälaren region was statistically significant in the south of Stockholm County as well as along the shores of Lake Mälaren. The clustering around Mälaren has moved westwards towards Västmanland and Södermanland since 2009.

5.1 TBE incidence and climatic factors

At the start of the study it was assumed that the TBE incidence data received would have a finer temporal scale, hence making it possible to investigate monthly variation in disease and climate. On receiving the annual incidence data the temporal scale of the climate data had to be adjusted accordingly, thus disregarding the between-month variation that may in itself have an impact on the intensity of disease incidence. Annually averaged data provided a very generalised overview of the temperature and precipitation conditions in a year and may thus have impacted the results of the statistical analyses. When performing the statistical analyses with climate factors, the winter temperatures were not accounted for as this is a season where ticks are not active and as a result, not able to spread the disease. However, winter temperatures and snow cover have an impact on whether a tick population will survive until the following active season or whether the active season may increase its range (Lindgren 1998).

The multiple linear regression analyses examined the dependency of TBE infection with the variables temperature, precipitation and a combination of the two. In the national dataset, roughly two thirds of the temporal datasets had at least one significant variable (Table 1) whilst on the regional level only half of the temporal datasets showed statistical significance (Table 2). On the national level, there were only five years where the application of two independent variables resulted in statistical significance, and six years on the regional level. Many years showed a significant relationship between TBE incidence and mean precipitation for the active season. This was unexpected as literature suggests precipitation does not have a direct impact on disease transmission, rather on the overall well-being of the tick and its ability to maintain a stable internal water balance (Knulle and Rudolph 1982; Dobson and Carper 1993). This may suggest that the relationship between the two variables is coincidental, though it may be the case that

precipitation and humidity have a greater impact on transmission than previously expected. Further investigation of the relationship between TBE incidences and precipitation suggests a negative relationship between the two variables, suggesting that fewer cases of TBE occur when the average seasonal precipitation is high. This could also be due to the fact that people tend to spend less time outdoors when the weather is bad and it is raining.

Ticks are cold-blooded and are thus expected to be directly affected by changes in their surrounding temperature (Hocking 1971). TBE themed researches have suggested that temperature can impact the degree of TBE infection. It was therefore surprising that so few of the explored years showed a statistically significant relationship with the temperature variable (Lindgren 1998; Lindquist and Vapalahti 2008). The low number of significant relationships could be caused by a variety of factors. Firstly, the analysis was limited by the number of incidence points with many of the explored years having fewer than 50 observed cases. Small sample sizes reduce the ability to detect statistical significance and increase the likelihood of type II error. The coarse temporal resolution using annual data can lead to errors in the statistical analysis as temperature and precipitation data at this scale may not be representative of the conditions on the actual date of infection. It is recommended that further studies include finer data in order to determine more specific relationships.

Increasing temperatures can accelerate the maturation process of the tick, and is thus a huge incentive for research in this field (Dumpis et al. 1999). Years with significant temperature variables displayed a negative trend when plotted with the majority of infections occurring with an average active season temperature between 10.5-12.5 °C. This negative trend could be attributed to ticks being highly susceptible to dehydration, whereby ticks limit movement and questing if there is a drop in relative humidity or if temperatures are high (Ostfeld and Brunner 2015).

5.2 Patterns in TBE incidence

When examining the distribution of TBE infection points in Sweden, there is obvious clustering in the endemic Stockholm-Mälaren region. Investigating how the disease is spatially distributed within the endemic zone can help distinguish significant clustering of the disease, classifying them as high risk zones for infection. Two cluster-mapping methods were used in the study, as well as two spatial-weighting methods in order to see if there were any key similarities or differences as to how clustering and hotspots were detected (Figure 7-20).

5.2.1 Conceptualising spatial relationships

In this study K-nearest neighbour and fixed distance-band were used as spatial weighting methods. Variables are weighted in order to increase the influence of objects in proximity to a feature, thus enforcing Tobler's First Law of Geography whereby "everything is related to everything else, but near things are more related than distant things" (Tobler 1970; Everitt et al. 2012).

The primary difference between the two methods was the number of years that each weighting method was successfully able to discern clusters. KNN produced a spatial weights matrix for each dataset in the time series, making it possible to visualise clusters for a large proportion of the time series. The distance bands were determined through incremental spatial autocorrelation whereby only the datasets where the spatial autocorrelation produced peak z-score depicting statistically significant spatial clustering were implemented in the cluster-mapping tools (Figure 4-6). Peak values occurred in cases where the z-score was greater than 1.65 and thus considered to be clustered.

Both weighting methods have their advantages and disadvantages. KNN ensures consistency in the number of variables (K) that can influence a feature, though this can be problematic if there are a limited number of features in a dataset or if a feature has fewer than K proximal neighbours. Fixed distance creates a sphere of influence with a predetermined distance whereby all values within the sphere of influence are weighted as important. A defined value for proximity limits the influence of distant points though this may also cause the proportion of influential variables to differ from feature to feature. KNN is considered a good choice when working with large numbers of feature points (Qian et al. 2014). Observing the results in years where both weighting methods depicted clustering for Local Moran's I; the fixed band distance weighting produces a larger number of significant clusters in comparison to KNN. For both cluster mapping methods, fixed band distance detects clustering around Lake Mälaren whilst KNN has also detected significant clustering in the south of Stockholm County particularly between 1986 and 2002. The two weighting methods detect cold spots in similar locations when Getis-Ord G_i^* is implemented.

5.2.2 Cluster-mapping

The main difference between the two cluster-mapping methods used in this study is that the Getis-Ord G_i^* statistic is able to determine hot spots and cold spots, whilst Local Moran's I is able to differentiate between high and low clusters as well as high and low outliers (Kies et al. 2009). These outliers are determined by comparing the value of a feature to that of its neighbours as well as the mean of the entire population, in which features that differ from their neighbours

are classified as anomalies. In Getis-Ord G_i^* the value of the feature and its neighbours are used to determine hotspots. This may result in features accidentally being classed as hot spots since features with very high values may increase the local mean (Anselin 1995; Songchitruksa and Zeng 2010). Local Moran's I determines clusters by only looking at a feature's neighbours, thus producing outliers if the value of a feature differs from those around it.

When comparing the results of the two clustering techniques, the Getis-Ord G_i^* (Figure 7-14) produced a larger proportion of significant clusters compared to Local Moran's I (Figure 15-20). This may be due to the fact that points with a 90% confidence level ($p < 0.10$) were also considered to be hotspots. When comparing the results of the spatial weighting methods for Getis-Ord G_i^* for the years where both factors produced hotspots, it is not clear whether one creates more hotspots than the other. When comparing the results of the spatial weighting methods for Local Moran's I for the years where both factors produced clusters, the fixed distance-band produced more clusters whilst the KNN was able to identify more outliers. The outliers presented in Local Moran's I are often located in the same area that Getis-Ord G_i^* detects significant cold spots. The location of the clusters were however fairly similar for both cluster-mapping methods. Many clusters were found by the coast in the south of Stockholm County near the port town Nynäshamn, as well as around Lake Mälaren. Much of the apparent clustering in southern Stockholm County occurred during the 1990s, more specifically in 1990, 1991, 1992, 1995 and 1997, whilst clustering around Mälaren became particularly prominent after 2003. The first clusters around Mälaren typically occurred in Stockholm County near the towns of Sigtuna and Sollentuna, but gradually spread westwards towards Västmanland and along the southerly shores of the lake in Södermanland during the late 2000s. There are a number of reasons why these particular locations are typical hotspots for TBE infection. Firstly, these areas consist mostly of natural landscapes such as woodlands, meadows, and agricultural land, which are particularly suitable habitats for ticks and reservoir hosts such as small mammals. Ticks require humid conditions and the proximity to water bodies ensures this. When visually interpreting the directional distribution of TBE incidence in the Stockholm-Mälaren region, the majority of the points are spread along the coast and archipelago of Stockholm County. In the later years of the dataset when the number of TBE incidence points has increased, the ellipse becomes larger encompassing the shores of Lake Mälaren.

5.3 Further study

There is great potential for further studies which explore the spatial distribution of TBE in Sweden. This study has highlighted the importance of temporal resolution in terms of the data being implemented in statistical and spatial models. Adjusting the resolution of the TBE and

climate data to a finer temporal scale (exact dates of infection or number of incidences per month) would increase the confidence in the results.

Whilst the low number of disease incidence may be positive in terms of the health and well-being of the population, it can result in undetected relationships between variables. A similar study examining the relationships between environmental factors and the occurrence of West Nile Virus in the USA used cases of infected dead birds as an additional resource as they believed that the numbers of human cases were underreported and unreliable (Kala et al. 2017). It is possible for ruminants to become infected with TBE and so serological tests on cattle may allow for the spread of TBE in Sweden to be tracked more efficiently.

TBE infection is not solely influenced by temperature and precipitation variables. A more detailed analysis should also consider land cover, the number of days with snow cover and roe deer population. In Sweden it is common for people to visit a holiday home during the summers, so location of these cottages in regard to infection location may be of interest.

GWR has been successfully implemented in a variety studies in order to understand the spatial distribution of diseases such as cancer, cardiovascular disorders and dengue fever and was thus thought to be an appropriate method for analysing TBE (McKinley et al. 2013; Racloz et al. 2012; Soljak et al. 2011). It is possible that, with additional variables and finer temporal resolution, this may also be a suitable method for monitoring TBE in Sweden.

6. Conclusions

Understanding and limiting the geographical spread and incidents of TBE in light of climate-related factors is of importance in Sweden as TBE is a serious disease with potential long-term consequences for the affected. Vaccines against TBE are available, but there is a lack of consensus regarding the need for a public vaccination strategy. Further, recommendations to vaccinate are based on the number of reported incidents in relation to population.

While some studies have explored the biological requirements of the TBE vector *Ixodes ricinus* and the incidence of TBE cases, few studies have examined the changing spatial distribution of *Ixodes ricinus* in Sweden, and models to forecast the potential increase in the incidence of TBE in view of changing climatic factors. This type of information is essential in planning long-term strategies for cost effective management of TBE in Sweden.

This study has shown that there is significant clustering of TBE in the endemic Stockholm-Mälaren region, with the majority of these clusters located in the south of Stockholm County or along the shores of Lake Mälaren. On a national level, some of the explored years showed statistical significance with climate variables. Further significance will be highlighted if more precise dates of incidences and climate data are made available.

There are already indications that the *Ixodes ricinus*, and with it, the risk of TBE may be spreading further north. Changes to climatic factors such as temperature and precipitation have an effect on the vector and, as a result, on the potential spread of the TBEV virus. This study makes the following key recommendations:

- Further spatial analysis and research around the climatic factors affecting TBEV survival and spread in Sweden is needed. Modelling of the spread of *Ixodes ricinus* in view of climate related factors will contribute to under-pin evidence based public health decision-making and is thus of economic and societal importance.
- Models can provide a forecast of the spread of the vector and virus, and rather than waiting for cases to occur to guide decisions on vaccination or adopting a strategy of mass vaccinations. Modelling of the spread will thus allow for adoption of a *targeted preventative approach*, where vaccinations are offered based on probabilities of occurrence of the TBEV virus.
- To increase the confidence levels of these models at national level more detailed reporting of the incidents of TBE cases is needed from the health sector. This includes at a minimum the date or at least month of the infection, and the area, in order to more

accurately match it to temperature and precipitation and other geographical or climatic factors.

7. References

- Andersson, C. R., S. Vene, M. Insulander, L. Lindquist, Å. Lundkvist, and G. Günther. 2010. Vaccine failures after active immunisation against tick-borne encephalitis. *Vaccine*, 28: 2827-2831.
- Andreassen, A., S. Jore, P. Cuber, S. Dudman, T. Tengs, K. Isaksen, H. O. Hygen, H. Viljugrein, et al. 2012. Prevalence of tick borne encephalitis virus in tick nymphs in relation to climatic factors on the southern coast of Norway. *Parasites & vectors*, 5: 177.
- Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical analysis*, 27: 93-115.
- Berger, S. 2017. *Infectious Diseases of Sweden: 2017 edition*. GIDEON Informatics Inc.
- Centers for Disease Control and Prevention, 2012. Principles of epidemiology in public health practice: An introduction to applied epidemiology and biostatistics. Report. [in Swedish, English summary]
- Cromley, E. K. 2003. GIS and disease. *Annual review of public health*, 24: 7-24.
- Dobson, A., and E. Carper. 1993. Health and climate change: Biodiversity. *Lancet*, 342: 1096-1099.
- Duarte-Cunha, M., A. S. d. Almeida, G. M. d. Cunha, and R. Souza-Santos. 2016. Geographic weighted regression: applicability to epidemiological studies of leprosy. *Revista da Sociedade Brasileira de Medicina Tropical*, 49: 74-82.
- Dumpis, U., D. Crook, and J. Oksi. 1999. Tick-borne encephalitis. *Clinical Infectious Diseases*, 28: 882-890.
- ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.
- ESRI. 2014. Modelling Spatial Relationships. ArcGIS 10.2.2 Help. Retrieved, from
- Everitt, B. S., S. Landau, M. Leese, and D. Stahl. 2012. *Cluster Analysis*. Chichester: Wiley.
- Folkhälsomyndigheten. 2017. Sjukdomsinformation om Tick Borne Encephalitis (TBE). Retrieved 10 April 2017, from. <https://www.folkhalsomyndigheten.se/smittskydd-beredskap/smittsamma-sjukdomar/tick-borne-encephalitis-tbe/>
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Fotheringham, A. S., R. Crespo, and J. Yao. 2015. Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, 47: 431-452.
- Fukunaga, K., and P. M. Narendra. 1975. A branch and bound algorithm for computing k-nearest neighbors. *IEEE transactions on computers*, 100: 750-753.
- Getis, A., and J. Aldstadt. 2010. Constructing the spatial weights matrix using a local statistic. In *Perspectives on spatial data analysis*, 147-163 pp.: Springer.
- Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24: 189-206.

- Gray, J., H. Dautel, A. Estrada-Peña, O. Kahl, and E. Lindgren. 2009. Effects of climate change on ticks and tick-borne diseases in Europe. *Interdisciplinary perspectives on infectious diseases*, 2009.
- Gritsun, T. S., V. A. Lashkevich, and E. A. Gould. 2003. Tick-borne encephalitis. *Antiviral Research*, 57: 129-146. DOI: [http://dx.doi.org/10.1016/S0166-3542\(02\)00206-1](http://dx.doi.org/10.1016/S0166-3542(02)00206-1)
- Haemig, P. D., S. Sjöstedt de Luna, A. Grafström, S. Lithner, Å. Lundkvist, J. Waldenström, J. Kindberg, J. Stedt, et al. 2011. Forecasting risk of tick-borne encephalitis (TBE): Using data from wildlife and climate to predict next year's number of human victims. *Scandinavian journal of infectious diseases*, 43: 366-372.
- Heinz, F. X., K. Stiasny, H. Holzmann, M. Grgic-Vitek, B. Kriz, A. Essl, and M. Kundi. 2013. Vaccination and tick-borne encephalitis, central Europe. *Emerging infectious diseases*, 19: 69.
- Hocking, B. 1971. Blood-sucking behavior of terrestrial arthropods. *Annual Review of Entomology*, 16: 1-28.
- Huang, B., B. Wu, and M. Barry. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24: 383-401.
- Jaenson, T. G., M. Hjertqvist, T. Bergström, and Å. Lundkvist. 2012a. Why is tick-borne encephalitis increasing? A review of the key factors causing the increasing incidence of human TBE in Sweden. *Parasites & vectors*, 5: 184.
- Jaenson, T. G., D. G. Jaenson, L. Eisen, E. Petersson, and E. Lindgren. 2012b. Changes in the geographical distribution and abundance of the tick *Ixodes ricinus* during the past 30 years in Sweden. *Parasites & vectors*, 5: 8.
- Kala, A. K., C. Tiwari, A. R. Mikler, and S. F. Atkinson. 2017. A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. *PeerJ*, 5: e3070.
- Kies, U., T. Mrosek, and A. Schulte. 2009. Spatial analysis of regional industrial clusters in the German forest sector. *International Forestry Review*, 11: 38-51.
- Knulle, W., and D. Rudolph. 1982. Humidity relationships and water balance of ticks. *Physiology of ticks*: 43-70.
- Lindgren, E. 1998. Climate and tickborne encephalitis. *Conservation ecology*, 2.
- Lindgren, E., and R. Gustafson. 2001. Tick-borne encephalitis in Sweden and climate change. *The Lancet*, 358: 16-18.
- Lindquist, L., and O. Vapalahti. 2008. Tick-borne encephalitis. *The Lancet*, 371: 1861-1871. DOI: [http://dx.doi.org/10.1016/S0140-6736\(08\)60800-4](http://dx.doi.org/10.1016/S0140-6736(08)60800-4)
- Liu, Y., Z.-M. Chen, H. Xiao, W. Yang, D. Liu, and B. Chen. 2016. Driving factors of carbon dioxide emissions in China: an empirical study using 2006–2010 provincial data. *Frontiers of Earth Science*: 1-6.
- Lundkvist, A., A. Wallensten, S. Vene, and M. Hjertqvist. 2011. Tick-borne encephalitis increasing in Sweden, 2011. *Euro Surveill*, 16.

- Mansfield, K., N. Johnson, L. Phipps, J. Stephenson, A. Fooks, and T. Solomon. 2009. Tick-borne encephalitis virus—a review of an emerging zoonosis. *Journal of General Virology*, 90: 1781-1794.
- McCarroll, D. 2016. *Simple statistical tests for geography*. Boca Raton ; London ; New York: CRC Press, Taylor & Francis Group.
- McKinley, J. M., U. Ofterdinger, M. Young, A. Barsby, and A. Gavin. 2013. Investigating local relationships between trace elements in soils and cancer data. *Spatial statistics*, 5: 25-41.
- Miller, J. A. 2012. Species distribution models: Spatial autocorrelation and non-stationarity. *Progress in Physical Geography*, 36: 681-692.
- Ostfeld, R. S., and J. L. Brunner. 2015. Climate change and Ixodes tick-borne diseases of humans. *Phil. Trans. R. Soc. B*, 370: 20140051.
- Qian, Y., W. Zhou, J. Yan, W. Li, and L. Han. 2014. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7: 153-168.
- Racloz, V., R. Ramsey, S. Tong, and W. Hu. 2012. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *PLoS Negl Trop Dis*, 6: e1648.
- Rizzoli, A., C. Silaghi, A. Obiegala, I. Rudolf, Z. ˇek Hubálek, G. Földvári, O. Plantard, M. Vayssier-Taussat, et al. 2007. Ixodes ricinus and its transmitted pathogens in urban and peri-urban areas in Europe: new hazards and relevance for public health. *Emerging zoonoses: eco-epidemiology, involved mechanisms and public health implications*: 72.
- Ruankaew, N. 2005. GIS and epidemiology. *Journal of the Medical Association of Thailand= Chotmai het thangphaet*, 88: 1735-1738.
- SMHI. 2017. Observation data and station information. Swedish Meteorological and Hydrological Institution.
- Soljak, M., E. Samarasundera, T. Indulkar, H. Walford, and A. Majeed. 2011. Variations in cardiovascular disease under-diagnosis in England: national cross-sectional spatial analysis. *BMC cardiovascular disorders*, 11: 12.
- Songchitruksa, P., and X. Zeng. 2010. Getis-Ord spatial statistics to identify hot spots by using incident management data. *Transportation Research Record: Journal of the Transportation Research Board*: 42-51.
- Sugumaran, R., S. R. Larson, and J. P. DeGroot. 2009. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *International journal of health geographics*, 8: 43.
- Süss, J., C. Klaus, F. W. Gerstengarbe, and P. C. Werner. 2008. What makes ticks tick? Climate change, ticks, and tick-borne diseases. *Journal of travel medicine*, 15: 39-45.
- Svenska Dagbladet, 2017a. Fästingkarta kan varna för TBE-risk. *Svenska Dagbladet* 17 May 2017.
- Svenska Dagbladet, 2017b. Fruktad liten blodsugare har vaknat. *Svenska Dagbladet* 06 May 2017.

- Tälleklint, L., and T. G. Jaenson. 1998. Increasing geographical distribution and density of *Ixodes ricinus* (Acari: Ixodidae) in central and northern Sweden. *Journal of medical entomology*, 35: 521-526.
- Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46: 234-240.
- Tu, J. V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49: 1225-1231.
- Valarcher, J., S. Hagglund, M. Juremalm, G. Blomqvist, L. Renstrom, S. Zohari, M. Leijon, and J. Chirico. 2015. Tick-borne encephalitis. *Rev Sci Tech OIE*, 34: 453-466.
- World Bank Group. 2017. Population total for Sweden.
- World Health Organization. 2016. Vector Borne Diseases (Fact sheet No. 387). Retrieved 13 April 2017, from. <http://www.who.int/mediacentre/factsheets/fs387/en/>
- Yan, X., and X. Su. 2009. *Linear regression analysis: theory and computing*. World Scientific.
- Zheng, S., C. Cao, J. Cheng, Y. Wu, X. Xie, and M. Xu. 2014. Epidemiological features of hand-foot-and-mouth disease in Shenzhen, China from 2008 to 2010. *Epidemiology and infection*, 142: 1751-1762.

Institutionen för naturgeografi och ekosystemvetenskap, Lunds Universitet.

Student examensarbete (Seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) och via Geobiblioteket (www.geobib.lu.se)

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) and through the Geo-library (www.geobib.lu.se)

- 400 Sofia Sjögren (2016) Effective methods for prediction and visualization of contaminated soil volumes in 3D with GIS
- 401 Jayan Wijesingha (2016) Geometric quality assessment of multi-rotor unmanned aerial vehicle-borne remote sensing products for precision agriculture
- 402 Jenny Ahlstrand (2016) Effects of altered precipitation regimes on bryophyte carbon dynamics in a Peruvian tropical montane cloud forest
- 403 Peter Markus (2016) Design and development of a prototype mobile geographical information system for real-time collection and storage of traffic accident data
- 404 Christos Bountzouklis (2016) Monitoring of Santorini (Greece) volcano during post-unrest period (2014-2016) with interferometric time series of Sentinel-1A
- 405 Gea Hallen (2016) Porous asphalt as a method for reducing urban storm water runoff in Lund, Sweden
- 406 Marcus Rudolf (2016) Spatiotemporal reconstructions of black carbon, organic matter and heavy metals in coastal records of south-west Sweden
- 407 Sophie Rudbäck (2016) The spatial growth pattern and directional properties of *Dryas octopetala* on Spitsbergen, Svalbard
- 408 Julia Schütt (2017) Assessment of forcing mechanisms on net community production and dissolved inorganic carbon dynamics in the Southern Ocean using glider data
- 409 Abdalla Eltayeb A. Mohamed (2016) Mapping tree canopy cover in the semi-arid Sahel using satellite remote sensing and Google Earth imagery
- 410 Ying Zhou (2016) The link between secondary organic aerosol and monoterpenes at a boreal forest site
- 411 Matthew Corney (2016) Preparation and analysis of crowdsourced GPS bicycling data: a study of Skåne, Sweden
- 412 Louise Hannon Bradshaw (2017) Sweden, forests & wind storms: Developing a model to predict storm damage to forests in Kronoberg county
- 413 Joel D. White (2017) Shifts within the carbon cycle in response to the absence of keystone herbivore *Ovibos moschatus* in a high arctic mire
- 414 Kristofer Karlsson (2017) Greenhouse gas flux at a temperate peatland: a comparison of the eddy covariance method and the flux-gradient method

- 415 Md. Monirul Islam (2017) Tracing mangrove forest dynamics of Bangladesh using historical Landsat data
- 416 Bos Brendan Bos (2017) The effects of tropical cyclones on the carbon cycle
- 417 Martynas Cerniauskas (2017) Estimating wildfire-attributed boreal forest burn in Central and Eastern Siberia during summer of 2016
- 426 Ida Pettersson (2017) Ekologisk kompensation och habitatbanker i kommunalt planarbete
- 428 Johanna Andersson (2017) Using geographically weighted regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use : a case study in Lund and Malmö, Sweden
- 429 Elisabeth Farrington (2017) Investigating the spatial patterns and climate dependency of Tick-Borne Encephalitis in Sweden