Student thesis series INES nr 371

Issues of Geographic Context Variable Calculation Methods applied at different Geographic Levels in Spatial Historical Demographic Research

-*A case study over four parishes in Southern Sweden*

Karolina Despoina Pantazatou

2015

Department of

Physical Geography and Ecosystem Science

Lund University

Sölvegatan 12

Karolina D. Pantazatou (2015).
Issues of Geographic Context Variable Calculation Methods applied at different Geographic Levels in Spatial Historical Demographic Research -A case study over four parishes in Southern Sweden

Frågeställningar Beträffande Beräkningsmetoder av Geografiska Kontextvariabler på olika Geografiska Upplösningsnivåer i Rumslig Historisk Demografisk Forskning – en kvalitativ studie över fyra församlingar i Södra Sverige
Master degree thesis, 30 credits in *Geomatics*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *March* 2015 until *October* 2015

Issues of Geographic Context Variable Calculation Methods applied at different Geographic Levels in Spatial Historical Demographic Research

- A case study over four parishes in Southern Sweden

Karolina Despoina Pantazatou

Master thesis, 30 credits, in *Geomatics*

Supervisors:

Lars Harrie, Lund University

Finn Hedefalk, Lund University

Exam committee:
Karin Hall, Lund University
Petter Pilesjö, Lund University

Abstract

Karolina D. Pantazatou

Issues of Geographic Context Variable Calculation Methods applied at different Geographic Levels in Spatial Historical Demographic Research – *A case study over four parishes in Southern Sweden*

Spatial analysis is dependent on the geo-referencing quality of the spatial data, as well as on the definition of the geographic context variables used. However, these facts are rarely taken into consideration in historical demographic research where the geographic factor is considered. An important obstacle in this kind of research is the availability of historical data (spatial and non-spatial), sufficient timeframes, and financial resources to employ qualified scientists to perform the geo-coding and the linking of population to specific geographic levels according to the existent historical sources. Since these are essential issues, it is important to determine if and how much the choice of different geographic context variable definitions calculated over different geographic levels could affect the research outcome. This thesis project attempts to address this problem, by examining how much the results of geographic context variables differ when different definitions of the variables are used or when the variables are calculated over different geographic levels. For this purpose, geographic and demographic data from four rural parishes set in the 19th century southern Sweden is used to define geographic context variables that might affect mortality in historical demographic research (e.g. soil types, proximity to water, proximity to wetlands, proximity to gathering places, and population density). The results show that different definitions of distance might produce contradictory results, depending on the geography of the research location and the shape or size of the geographical units. Similarly, results tend to differ when different geographic levels are used. Though the suitability of what geographic level is chosen is highly dependent upon the research hypothesis, additional research is needed to determine when a geographic level is deemed suitable enough.

Keywords: Geography, Physical Geography, Spatial Longitudinal Demographic Research, Historic Demography, Spatial Demography, Geographic Context Variables, Distance Calculation Methods, Geographic Levels, GIS

Supervisors: Lars Harrie, Finn Hedefalk. Master degree project, 30 credits, in Geomatics. 2015. Department of Physical Geography and Ecosystem Science, Lund University. Student thesis series INES no 371.

Preface

This study is a master thesis in the master program Geomatics at the department of Physical Geography and Ecosystem Science at Lund University. The work was carried out in cooperation with the Research School in Economic Demography at Lund University.

I would like to express my gratitude to supervisor Lars Harrie in the department of Physical Geography and Ecosystem Science at Lund University. His helpful read-throughs, comments and suggestions and overall support have been valuable and productive.

I would also like to thank my supervisor Finn Hedefalk at the department of Physical Geography and Ecosystem Science who is working on his PhD thesis in cooperation with the Centre for Economic Demography at Lund University, for his valuable advice and insight whenever help understanding the structure and theory behind the provided geographic and demographic data was required.

Table of Contents

1 Introduction

1.1 Background

In recent years, *Geographical Information Systems* (GIS) have rapidly evolved and been implemented in a wide scope of application areas including government and public services, business and service planning, logistics and transportation, and environmental management and modeling (Longley et al. 2011). Public and private sector research utilizes GIS to manage geographic information, identify geographical trends and patterns, and to model spatial processes (Harris et al. 2005). Even at an individual level, GIS have influenced many aspects of everyday life. For example, the usage of GPS navigation devices whilst driving is nowadays considered a common practice. Additionally, a variety of GIS application services (e. g. web map services, governmental or institutional geoportals, traffic maps, interactive navigation maps, crime maps, weather maps, statistical maps, etc.) is available on the Internet. These applications are regularly used by the public, which is presented with the opportunity to add or retrieve information with geographical content. But, what is a geographical information system and what are its uses?

In order to better comprehend the definition of a GIS, it is recommended to review the general definition of an information system. An information system is an association of people, machines, data, and procedures working together to collect, manage, and distribute information of importance to individuals or organizations (Worboys and Duckham, 2004). A GIS constitutes a special category of information system whose concept is based on geographically referenced data. More in particular, a geographical information system is a computer based information system that enables collection, modeling, storage, retrieval, sharing, processing, analysis, and presentation of geographically referenced data (Worboys and Duckham, 2004). Depending on the occasion, the output of these processes may be available in the form of digital maps, 3D virtual models, tables or lists. An essential characteristic of GISs is the fact that they support the spatial as well as the temporal dimension (aspect) of geographically referenced data.

This last characteristic makes GIS a valuable tool in the service of demographic research, which is conducted over space and/or time. In science, *demography* is denoted as the statistical study of births, deaths, income, marriages or the incidence of disease, illustrating the changing structure of human populations (Oxford Dictionary 2005). The term is also used to describe the composition of a particular human population. Demography is a very general science, which is why it is divided into several scientific fields. The fields covered in this project are that of historical and spatial demography.

Historical demography is a branch of demography, which is denoted as the quantitative study of past human populations. It deals with matters related to the size and structure of populations as well as with factors that are believed to cause population changes (i.e. fertility, mortality, migration, socioeconomic status, configuration of families, etc.) (Encyclopedia of Public Health, 2002). Historical demographic research is partly based on written records and archaeological findings. Written records often consist of legal documents, taxation records, or ecclesiastical records of baptisms, marriages and funerals. Additionally, if available, historical maps are also deployed to assist the analysis of past populations. Unfortunately, written records are not always available (e.g. prehistoric populations) or complete and with the same rule applying for archaeological findings, historical demographers are often forced to rely on intelligent guesswork and make estimations.

Spatial demography is defined as the spatial analysis of demographic processes (spatial demography org.). It is closely related to urban demography – where the focus is set upon the sociological and ecological factors affecting population behavior –, rural demography – where the focus is set upon migration patterns – and applied demography which focuses upon population estimation research (Voss 2007).

In conclusion, the rapid development of GIS has resulted in an ever-growing list of new application areas. Historical demography is one of the more recent scientific fields to be added to this list. GIS have provided the tools necessary for historical demographers to explore new and old data sources including detailed geographical information. This has allowed them to perform sophisticated methods of spatial analysis and expand their scope of analysis to new dimensions, which was not possible before. As a result, data from national censuses containing geographical identifiers at the scale of the individual or at other levels of aggregation (i.e. municipalities, counties or other administrative areas) as well as detailed historical maps of cities and sites have been digitized and therefore become available for further processing. The integration of contextual geo-spatial information with micro-level socioeconomic and demographic data enables historical demographers to perform a much wider search for demographic patterns in the past, ultimately improving their understanding of these patterns. Spatial modeling of socioeconomic behaviors set in the smaller and more simply-organized communities of the past could contribute to the better understanding of the main drivers behind the first demographic transitions. The utilization of historical GIS also enables further studies regarding the interaction between climatic, environmental, socioeconomic and demographic processes and how they affect different aspects of society such as public health, mortality, fertility and migration.

1.2 Framework & Problem Statement

This study is part of a historical demographic research project, whose aim is to integrate geographic data with longitudinal demographic data in order to discover spatial patterns related to demographic phenomena such as migration, fertility, mortality and health inequalities. This is an interdisciplinary project carried out at Lund University as a collaboration between a group of experts in the fields of demography, epidemiology and GIS. The project covers a study area that includes five parishes (i.e. Halmstad, Hög, Kävlinge, Kågeröd, and Sireköpinge) located in the province of Scania in Southern Sweden. Demographic and geographic data are provided for a time-period that extends from 1800 to 1914. The goal is to research which factors (demographic and/or geographic) play the most significant part in determining the values of the pre-mentioned demographic phenomena across the defined time-period.

The basic methodology of the pre-mentioned project can be summarized in the following steps:

i. Search for appropriate data sources (e.g. historical maps, digitized historical demographic data, etc.).
ii. Digitalization of historical maps and production of geographical units (e.g. parish boundaries, property units, buildings, roads, railways, streams, wetlands, etc.).
iii. Link the demographic, individual-level data from the demographic database to the related geographical units.
iv. Process the demographic data, so that it is in a suitable format, to support different methods of demographic analysis (e.g. COX regression analysis model (Cox, 1972)).
v. Calculate the geographic context variables (soil types, population density index, proximity to water, proximity to wetlands, and proximity to gathering places) for each geographical unit.
vi. Utilize the demographic and geographic variables by inserting them into demographic analysis models and assess the results (e.g. examine if any of the geographic context variables has a higher influence on mortality rates).

The process above raises two important questions. Firstly, how should geographic context variables be defined and, secondly, at what level of aggregation should the linkage between individuals and geographical units be performed? In this case, possible levels of aggregation are for example the geographic units described by buildings, property units and parishes. Linking individuals to small areal units is a cumbersome and time demanding task. The choice of an appropriate level of analysis is directly related to the availability of relevant historical data as well as the financial resources to employ assistant researchers to perform the linking task. The problem at hand also embeds a hidden issue, denoted as the modifiable areal unit problem (MAUP, Openshaw 1984).

MAUP is denoted as the problem that occurs during the spatial analysis of aggregated data, in which the results differ, when the same analysis is applied to the same data, but different aggregation schemes are used. MAUP takes two forms: the scale effect and the zone effect. The scale effect produces different results when the same analysis is applied to the same data, but on different scales of aggregation units. For example, analysis using data aggregated by parishes will differ from analysis using data aggregated by property units. Often this difference in results is valid: each analysis asks a different question because each evaluates the data from a different perspective (different scale). The zone effect is observed when the scale of analysis is fixed, but the shape of the aggregation units is changed. For example, analysis using data aggregated into one-mile grid cells will differ from analysis using one-mile hexagon cells. The zone effect is a problem because it is an analysis, at least in part, of the aggregation scheme rather than the data itself.

1.3 Aim

The aim of this thesis project is to provide guidelines for including geographic factors into historical demographic research. Especially, two things are investigated:

1. The definition and calculation of geographic context variables
2. The variations of produced geographic context variable values when computed over different geographic levels of aggregation.

The reason why these two issues are investigated is to enhance the choice of context variables as well as a suitable geographic level for linking individuals when conducting a spatio-temporal demographic research over historical data.

1.4 Method

The aim of this study is achieved in three steps. Initially, a bibliographic research (literature study) is conducted over the field of historical geodemographic studies. This step is crucial in order to obtain the necessary background knowledge regarding the different analyses methods and challenges most often encountered in this field of studies. The second step is dedicated to using this knowledge for studying definition of the geographic context variables and various methods according to which they can be computed by, as well as the effect of different geographic levels in the analysis. The third step is dedicated to the evaluation of the study and the export of conclusions.

1.5 Limitation

The main limitation of this study is that it does not focus on performing a real demographic **micro-level analysis**. As stated above, the aim is rather to develop a methodology for how to integrate geography into demographic research rather than perform an actual study.

Another limitation is that the data used for evaluating the methodology is limited both temporally and spatially. More in particular, the study area consists of four[1] parishes in southern Sweden, namely: Halmstad, Hög, Kävlinge, and Sireköpinge, while the study period extends from 1800 to 1914 (in most cases, only data from 1910 are used). At this point, it is also important to note that this study is conducted over an exclusively rural type of environment.

Finally, only a limited number of geographic context variables (soil type, population density index, proximity to water, proximity to wetlands and proximity to gathering places) are examined.

1.6 Disposition

The work has been divided into six chapters: introduction, related studies, materials and methods, results, discussion and conclusions.

The first two chapters consist of a general introduction on the specified research field and a documentation of previous studies performed respectively. The introduction contains a set of definitions and summarized descriptions of the main scientific areas (GIS, demography, historical demography and spatial demography) closely linked to the subject of this project. Other topics discussed in the introduction include the problem statement, the aim, the method and the limitations of this project. The second chapter includes a general overview of studies made on the topic of spatial demography.

The section titled “Materials and methods” contains information regarding the study area, data sources, geographic context variables and technical implementation. The data sources are split into two parts (geographic data and demographic data), describing the origin and the main characteristics of each dataset. A separate part is dedicated on providing information regarding the background motivation behind the choice of the specific geographic context variables featured in this project. This is followed by a detailed description of each one of the aforementioned variables. An application is developed to display the results of the computation of all geographic context variables over all geographic levels in the form of a digital map. The results of the computations are also made available in a tabular format through the Graphical User’s Interface of the application. A documentation of the software used to develop the application is included in the technical implementation.

The results are being presented in the form of tables and diagrams. The information presented in the results is being evaluated in the discussion section. Quality issues and future developments are also included in the previous section. A summarized description of the main findings of this project are presented in the conclusions.

[1] Kågeröd parish was excluded from this work, since the quality of the geographic data there was somewhat worse than for the other parishes.

Appendix A contains a vocabulary of important terms frequently used in the study. To enhance readability, terms found in the vocabulary are marked in bold in the text. Finally, Appendix B includes some important code snippets (e.g. SQL query examples).

2 Related Studies

This part focuses upon studies previously conducted in the broader research area of spatial demography. It specifically highlights the difficulties and challenges of conducting a longitudinal demographic research on historical data with the apposition of related articles. To enhance readability, the related articles are divided in four sections. The first section is dedicated to the problem of geocoding historical populations. The second section contains information regarding ongoing projects that attempt to bridge the gap between historical and modern spatial demography by building databases storing historical and contemporary data. The third section presents some challenges (e.g. Modifiable Areal Unit Problem & Uncertain Geographic Context Problem) that are usually connected with the choice of demographic analysis method in spatial demographic research and provides some examples of how to overcome them. The last part discusses the use of geographic context variables in spatial demographic research.

2.1 Historical Demography Studies on Population Geocoding

In the present day, the number of **micro-level** longitudinal studies including geographic factors is limited. This is attributed to the fact that large historical datasets in which individuals are linked to detailed physical locations are sparse, mainly due to the lack of complete records and other ancillary information. Another important reason is the high financial cost of employing personnel to perform the complex and time-consuming cross-reference tasks of assigning individuals to detailed physical locations. This has resulted in, with a few exceptions (e.g. Hedefalk 2014; Villarreal et al. 2014), of individuals that have being linked to geographic units of a higher aggregation level such as parishes/counties or other aggregated regions (Gutmann et al. 2005). Another consequence of the pre-mentioned situation is that in cases where more detailed geographic data has been analyzed (e.g., Ekamper 2010), only a few years or a short time period has been studied. The same applies when larger areas (e.g. countries) are being studied (Sarkar and Hall 2011). Consequently, a significant number of spatio-temporal, historical, demographic studies are being conducted over smaller areas (DeBats 2008). Therefore, the effects of the spatial micro-level factors, within aggregated regions, have rarely been studied for historical data.

The results of longitudinal historical studies are largely based upon the geographic contexts on the aggregated level. Gutmann et al. (2005) researched the factors (weather conditions, environmental amenities, occupation, and population characteristics) that influence net migration rates in the USA Great Plains between 1930 and 1990. The data set used as well as the corresponding results were available on a county level. The chosen level of aggregation was well-suited for the purposes of the study, since migration is defined as a residential change that crosses a county boundary (Long, 1988). Nevertheless, environmental parameters such as temperature or precipitation do not conform to administrative boundaries (e.g. county boundaries). Consequently, the county values of the aforementioned variables will most probably exhibit some level of correlation, violating the conditions of a regression analysis. This problem was dealt with by the implementation of a standard error function, which allowed spatially correlated errors (panel-corrected standard errors).

DeBats (2008) studied the spatial distribution of wealth in two mid-19th century U.S. cities (namely; Alexandria (Virginia) and Newport (Kentucky)) of commercial and industrial character. He urged the importance of selecting a suitable study area, representative of the examinable period that; on one hand has a population that is large enough to demonstrate all the typical urban characteristics (economic, social, political, etc.) and to segregate the effects of industrialization from urbanization, and, on the other hand, is small enough to be studied in a holistic manner. Mapping was done on an individual level. With the help of tax records, national censuses, plat maps, poll lists and city directories, approximately 80% of the inhabitants were successfully relocated to their respective residences and associated –as individuals, families as well as economic units– with all available social, economic and political information.

Regarding the choice of mapping level in historical data sets, DeBats (2008) supported the idea of mapping a whole community and not just a population sample on an individual level. He based his arguments on the same opinion previously expressed by Ruggles and Menard (1995) who argued that, although processing national samples may help locate “individual behavior in temporal and spatial context”, only the study of whole communities could place individuals in the context of their immediate neighbors, which is the factor most likely to have influenced their behavior. He also expressed concerns of using higher levels of aggregation except to the one expressed by the residential unit for performing analysis on. It was his belief that higher levels of aggregation could provide erogenous results by obscuring the versatility of results sometimes only apparent at the lower levels of aggregation (e.g. averaging data on higher levels of aggregation).

Villarreal et al. (2014) studied the historical health and environmental conditions in seven major U.S. cities from 1830 to 1930. They developed the Historical Urban Ecological data set (HUE) that contained digitized crime, disease, demographic, property, land and tax information from annual municipal reports on a *ward*-level (see vocabulary). **Wards** were chosen as the most appropriate areal unit, because most statistics at the time (national and federal censuses and municipal reports) included information on that level. By including accurate, digital representations of historical city street centerlines, they provided future researchers with the necessary tools to link the ward-level data set to smaller areal units (e.g. city blocks, residences), thus enabling statistical calculations to be performed over several levels of spatial granularity (multilevel analysis). Furthermore, it should be mentioned, that they also attempted to link individuals (war veterans) to their respective residences, but decided that this was unnecessary, and linked them to the equivalent blocks instead, by spatially geo-coding each one of them at the address-level. All addresses that could not be geocoded or placed with certainty, due to incomplete records or sporadic availability of ancillary historical sources, were ultimately excluded from the sample.

2.2 Databases Bridging the Gap Between Historical & Contemporary Demographic Studies

Assigning individuals to detailed physical locations is relatively easier when conducting modern longitudinal studies, where adequate records are present assigning individuals to physical locations belonging to a low aggregation level, with the help of for example zip codes. However, efforts are being made to create large indexes, allowing the conduct of full-scale analyses on data with large areal and temporal extent. The ultimate objective is to attempt bridging the gap between the past and the present following the course of population evolution. The North Atlantic Population Project (North Atlantic Population Project 2008) and Terra Populus (Terra Populus 2013) are indicative examples of this attempt.

The North Atlantic Population Project (NAPP) is a machine-readable database, containing complete censuses of several Northern Atlantic countries (e.g. Canada, Great Britain, Norway, Sweden, the United States of America, and Iceland) from the mid-19th century and onwards (Hall, 2011; North Atlantic Population Project 2008). More in particular, NAPP contains individual- and household-level data that is available at several levels of aggregation (e.g. state, county, municipality, district, sub-district, province, parish, and town). As the choice of administrative boundaries varies between countries, so does the availability of the aforementioned data at different levels of aggregation. It is important to mention that the NAPP database has enabled linking of individuals between census years to support longitudinal analysis. NAPP only contains complete censuses for a limited number of years per country. Therefore, in order to support cross-temporal analyses, NAPP has also included samples of national census data for a different selection of years. The aim of NAPP is to, eventually; reconstruct the population of the North Atlantic world from the mid-19th century up to the present.

Terra Populus (Terra Pop) is a collaborative project, whose goal is to provide an organizational and technical framework to preserve, integrate, disseminate, and analyze global-scale spatiotemporal data describing population and the environment (Terra Populus 2013). TerraPop incorporates the following kind of data: census and survey microdata (including information on an individual level); aggregate census and survey data (including aggregate population, land use, and land cover characteristics); data describing land cover and other environmental characteristics (remote sensing); climate data (e.g. temperature, precipitation, etc.); and Vector-based GIS data representing administrative and census unit boundaries. The temporal dimension of the data expands over several decades, with some sources reaching back to the 19th century. Climate data was acquired from multiple weather stations (WorldClim) and interpolated to create continuous grids. Land/Use information is interpreted from satellite images (Global Land Cover 2000) and sometimes supplemented by information from surveys and governmental records (Global Landscapes Initiative). The distributable data is available in three formats: census microdata with attached characteristics describing land use, land cover, and climate for local areas; aggregate data for administrative districts with tabulated population data and environmental characteristics; and gridded data with population and environmental characteristics.

2.3 Demographic Analysis Methods

When attempting to collect, organize, store and distribute data in this scale, or at any scale, it is important to have performed a thorough analysis, in order to avoid MAUP related issues. Ekamper (2010) and Sarkar & Hall (2011) highlight the effects of MAUP, by demonstrating how performing the same analysis, over the same data but in different aggregation levels, can produce different results.

Ekamper (2010) used cadastral maps in urban historical demographic research in the Dutch city of Leeuwarden in the mid-19th century. He performed different kinds of analysis over population density, spatial distribution of wealth, spatial distribution of religion and infant mortality rates in the study area. The areal units used in the analysis and the corresponding visualized results were property-units and dwellings (i.e. residences). The temporal extent of each result was usually one or two years. Additionally, Ekamper (2010) demonstrated how cadastral maps of, for instance population density, can provide different information than other types of maps aggregated at the level of urban districts.

Though the importance of space in regression methods appears to be generally recognized in the social sciences, the choice of areal unit (Modifiable Areal Unit Problem - MAUP) remains an open issue (Openshaw 1984). Geographically Weighted Regression (GWR) and multilevel modeling are two methods that have been developed to address scale issues. More in particular, both the GWR and multilevel modeling approaches employ different techniques in their attempt to link macro-level to micro-level analyses (Entwisle 2007; Voss 2007).

Geographically Weighted Regression is defined as an exploratory statistical technique that permits variations in relationships over space to be measured within a single modeling framework (Fotheringham et al. 2002). GWR has been used in recent demographic (Işık and Pınarcıoğlu 2006; Johnson et al. 2005) and health research (Yang et al. 2009; Chen et al. 2010). A disadvantage of this method is that it is excessively hard to implement in the case of large datasets (Matthews et al. 2013).

Multi-level modeling has widely been used in demography and social epidemiology research (Entwisle 2007). This methodology is applicable either when examining the consequences of contextual factors on social behavior at a lower level, or when estimating the extent at which individual behaviors and demographic and health outputs are effected by the characteristics of a single individual as well as by the attributes of the larger geographic area (e.g. neighborhood, district, city, county, etc.). Two rare studies combining spatial and multilevel modeling in a demographic framework were conducted by Chaix and colleagues (Chaix et al. 2005; Chaix, Merlo, and Chauvin 2005).

Another attempt to work around the MAUP was made by Spielman and Logan (2013), who tried to employ the social composition of areas at different scales in order to identify neighborhoods and specify what defines them. By using an eco-centric perspective, they explored how segregation (by socioeconomic indicators or race/ethnicity) can vary by spatial scale. The study was carried out over 19th-century individual-level U.S. census data. The usage of individual-level data offers researchers the flexibility to study intergroup relations from a spatial perspective, unconstrained by administrative boundaries and scale.

Except MAUP, there are is another problem denoted as the Uncertain Geographic Context Problem - UGCoP (UGCoP, Kwan (a) 2012). Kwan ((b) 2012) defines this problem as: "*UGCoP refers to the problem that findings about the effects of area-based contextual variables on individual behaviors or outcomes may be affected by how contextual units (e.g., neighborhoods) are geographically delineated and the extent to which these areal units deviate from the true geographic context. It is a significant methodological problem because it means that analytical results can differ for different delineations of contextual units even if everything else is the same*". This issue is hard to address in longitudinal historical demographic studies, because detailed movement records of the individuals are not available. However, other factors might mitigate the magnitude of the problem. For example, providing information related to the type of community (e.g. rural), the transportation infrastructure (e.g. availability of developed road or railroad network, etc.) and the typical activities (indicating movements) of the inhabitants of the study area during a specific period can help to build a profile of the average, typical citizen of that period in time.

2.4 Geographic Context Variables in Demographic Research

From previous sections, it is evident that the choice of at what aggregation level individuals should be linked to physical locations to is crucial when conducting a longitudinal historical demographic research. Another important factor is the choice of variables. Except for the socioeconomic and demographic variables, demographers also deploy geographic context variables.

Mennis and Liu (2005) studied the relationship between socioeconomic change and urban growth by trying to identify patterns within a database containing socioeconomic and land cover change data. The geographic context variable in this case was the land cover data. A time series of land cover and socioeconomic data was produced and both data layers were overlayed for different periods in time. Specific information regarding the justification of the land cover classification was provided in the documentation. The result of the overlay process was used to extract dependencies between land cover change and the socioeconomic status.

Population density is a commonly used context variable in epidemiologic and demographic research. Seng et al. (2005) found correlation between dengue fever cases and population density in Malaysia. Except from epidemiology, the pre-mentioned variable has also been used in other demographic research. De Beer and Deerenberg (2007) studied fertility changes in the Netherlands in association to amongst other factors population density. Population density in this case was calculated on a regional as well as on a municipality level. More in particular, it was defined as the total number of people within a geographical unit defined by the choice of geographic level the study was conducted over.

Altitude and land use are also geographic context variables that are used in epidemiology (Gouri et al. 2010; Pezeshki et al. 2012). Fuster and Colantonio (2004) conducted a study of consanguinity over several municipalities or groups of municipalities in Spain for the time periods: pre-1889, 1890–1929, 1930–1959, and 1960–1979. They studied the occurrence of marriages between people with close blood relations in association to a set of demographic and geographic factors. The geographic variables introduced in their study were population density index, altitude and geographic characteristics. The geographic characteristics consisted of a spatial layer of a land cover classification (e.g. mountain, valley, island and coast, plain, etc.). The exact criteria on which the classification was based on or how the geographic computation was performed are not mentioned. The same applies for the population density, where even though it is defined as the division of population by localities, these localities are not specified.

Proximity to specific places is also classified as a geographic context variable. Pezeshki et al. (2012) studied the spread of cholera in villages belonging to the Chabahar district of Iran by monitoring the climate conditions (e.g. temperature, precipitation and humidity), altitude and the distance to local or regional health centers, rivers, subterranean canals as well as the distance to the eastern borders of Iran. The distance was calculated by using the straight-line distance ESRI ArcGIS Spatial Analyst function that calculates the Euclidean distance in a straight line from every cell in a raster dataset to the nearest source. Villages constituted the geographic level the study was conducted over.

Moreover, historical demographers have pointed out that regional differences in mortality were often large in the past, and often much larger than socioeconomic differences. Possible responsible factors behind this are population density, communication networks, sanitation and access to safe water, organization of poor relief and health care, breast-feeding practices or differences in agricultural productivity (Smith, 1983; Reid, 1997; Woods et al., 1993; van Poppel et al., 2005; Garrett et al., 2001). Potential geographic context variables that can be derived from this analysis and be used in mortality studies are population density, proximity to communication networks, proximity to clear water, proximity to health centers and hospitals, and soil types that affect agricultural productivity.

It is important to draw attention to the fact that a number of spatial demographic studies do not include detailed information regarding the exact geographic context variable computation methods applied. The potential issues caused by a non-representative choice of geographic level whilst attempting to present aggregated data in the geographic space, are documented in the definition of the Modifiable Areal Unit Problem (MAUP). The choice of a suitable geographic context variable may be equally important. The value of a geographic context variable is directly dependent upon the geography within or near the geographic unit for which it is being calculated. Consequently, the result is dependent upon the spatial distribution of the geographic object (e.g. streams) examined by the geographic context variable and the general characteristics (e.g. shape, size) of the geographic unit (e.g. village, county), which are to a large extent dictated by the chosen geographic level. All the above factors in conjunction with the research objective need to be taken into serious consideration before deciding upon which geographic context variable calculation method best serves the objective. Moreover, every method has its advantages and its disadvantages and not including any argumentation over why a method is chosen may become a liability problem when examining the results of the study. The extent on which this may or may not affect the outcome of the study is hard to determine, but nevertheless, should constitute a good enough reason to include a short but descriptive explanation over the main principles behind the implemented calculation method in the documentation.

3 Materials & Methods

3.1 Study area

The study area includes four parishes located in the province of Skåne in Southern Sweden, namely: Halmstad, Hög, Kävlinge and Sireköpinge (Figure 3.1). These parishes were rural parishes with an economy mainly based upon agricultural activities. However, from the 1870s and onwards, Kävlinge’s economy was also based upon some industrial activity (e.g. potato, wool and sugar factories). The area of each parish was estimated at: Halmstad 27.8 km^2, Hög 6.8 km^2, Kävlinge 8 km^2, and Sireköpinge 22.5 km^2. Added together the five parishes covered a total area of approximately 65 km^2. The fertile plains of Hög and Kävlinge were situated at between 10 and 15 meters above sea level, while the plains of Halmstad and Sireköpinge were situated between 60 and 70 meters above sea level. On average the climate of the study period is characterized by mild winters and not very hot summers, with the temperature from 1766 to 1860 ranging between -3 to 15 degrees Celsius (Bengtsson & Broström, 2010). The study area mainly consisted of open farmland with a minor exception of some northern parts of Halmstad which were covered by woods. According to Bengtsson and Broström (2010), land was the most important source of income for the population living in this area. They also mention how the land in Halmstad and Sireköpinge was primarily owned by the nobility, though the land in Kävlinge and Hög was predominantly owned by the crown or by free holders.

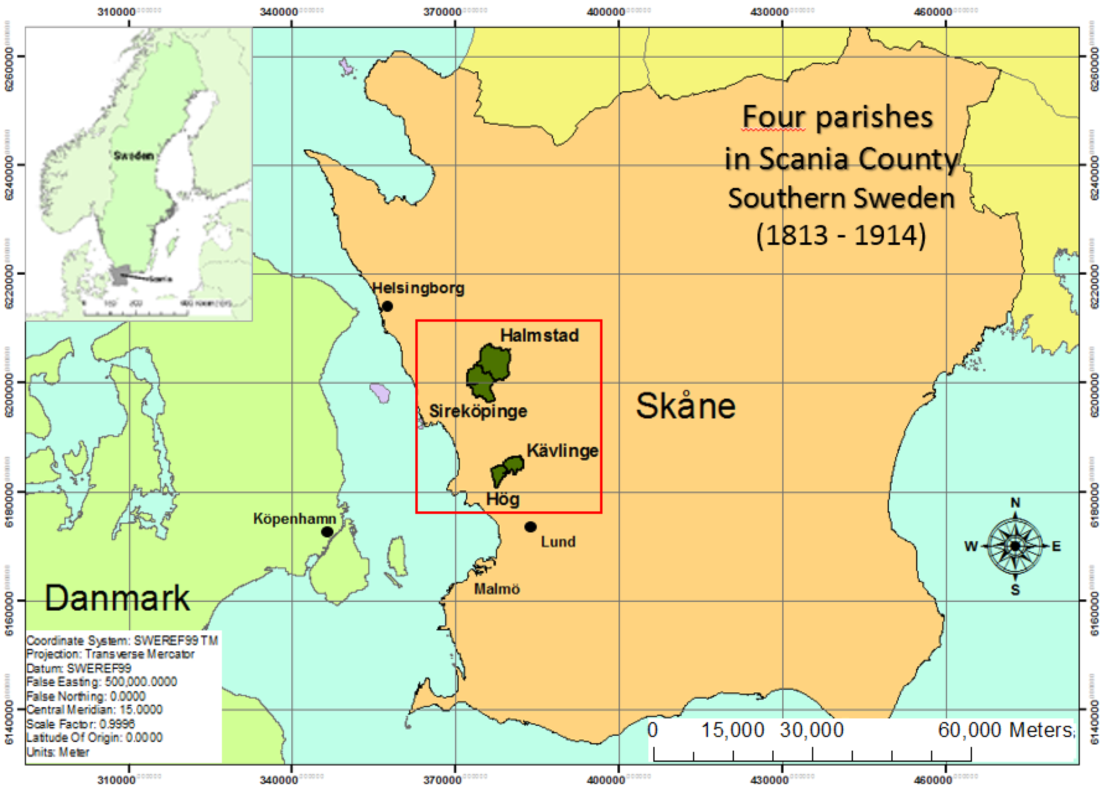


Figure 3.1: *Study area covering 4 parishes in Skåne (Sweden) and their surrounding areas. ©Lantmäteriet[I2014/00579]*

Up until the beginning of the 19th century, farmers were living in villages and cultivating parcels in the surrounding area (Dribe & Olsson 2003). The parish land was divided in several smaller parcels and each farmer could own more than one parcel. The idea was that all farmers should have an equal opportunity to cultivate crops upon fertile and less fertile soil. Nevertheless, in the early 19th century, several farmers expressed their will to proceed with a land reform that would rearrange the distribution of land parcels, so that each farmer would get one parcel equal to the added area of all the previous ones he owned. This request was discussed in parliament and an enactment took effect, stating that a land reform would be granted for the inhabitants of a village if, and only if, as much as a single farmer so requested (Johanesson, 1984). This enactment, which in Swedish is denoted as “Enskiftet”, took effect for several villages in Scania during the early 1800s. It should be noted that not all villages implemented it. Those villages that eventually did implement it did not necessarily do so at the exact same time. However, in 1827 a new land reform (Laga Skifte) was introduced, making the implementation of the previous land reform (Enskiftet) mandatory for all villages. The concentration of smaller not adjacent parcels into one larger property unit allowed people to move out from the villages, build farms and explore more efficient ways of agricultural cultivation. For instance, it was documented that the period following the land reform was characterized by the cultivation of previously uncultivated land, the drainage of lakes and wetlands as well as the implementation of crop rotation (Lewan, 1989). All of the previously mentioned innovations led to an increased productivity of agricultural products over the following years (Skansjö, 1997).

The property units created during the land reforms are denoted as *initial property units*. As time elapsed, only a few initial property units remained unchanged. Most of them were subdivided or partitioned to smaller property units, or aggregated with other property units to form larger properties. This was a direct consequence of people buying or selling land, or splitting their property amongst their children as inheritance. The merge of adjacent property units, originating from different initial property units, was a quite usual phenomenon for this study period. However, the initial property units continued to have an important role in the taxation of families and therefore the only address provided in the poll tax register was on an initial property unit level.

Sweden experienced a rapid growth in population during the given study period, which was mainly attributed to the end of hostilities with Norway, the introduction of vaccines and better health care, improved personal hygiene as well as the implementation of enclosure reforms and more efficient cultivation methods (Dribe & Olsson 2003). As a result, it could indicatively be mentioned that Sireköpinge counted 1250 and Halmstad about 900 inhabitants respectively in 1870 (Dribe & Olsson 2003), while the smaller parish of Hög had 479 inhabitants in 1899 and Kävlinge, the most populated parish in the study area, counted 1728 inhabitants in 1899 (Kävlinge Municipality).

3.2 Data

Datasets used for this project include geographic and demographic data. The demographic data are stored in Microsoft SQL Server 2012 Standard Edition. The geographic data are either stored in SQL Server tables or in ESRI shape-files. The following subsections contain a more detailed description of all different datasets used in the study.

3.2.1 Geographic Data

The geographic data are both modern and historic. The modern data are topographic data from National Land Survey of Sweden (Lantmäteriet, http://www.lantmateriet.se/en/) and soil data from the Geological Survey of Sweden (SGU, http://www.sgu.se/en/). The historical geographic data are based on around 60 historical maps (land survey maps - LSMs 1757-1863; military topographical survey maps – MTSMs 1812 – 1820; topographic maps – TMs 1860-1865; and economic maps – EMs 1910-1915) from National Land Survey of Sweden. Some indicative samples of the aforementioned maps are shown in Figure 3.2. All maps were scanned, geo-referenced and digitized. About 900 property units, streams, lakes and wetlands have been digitized.

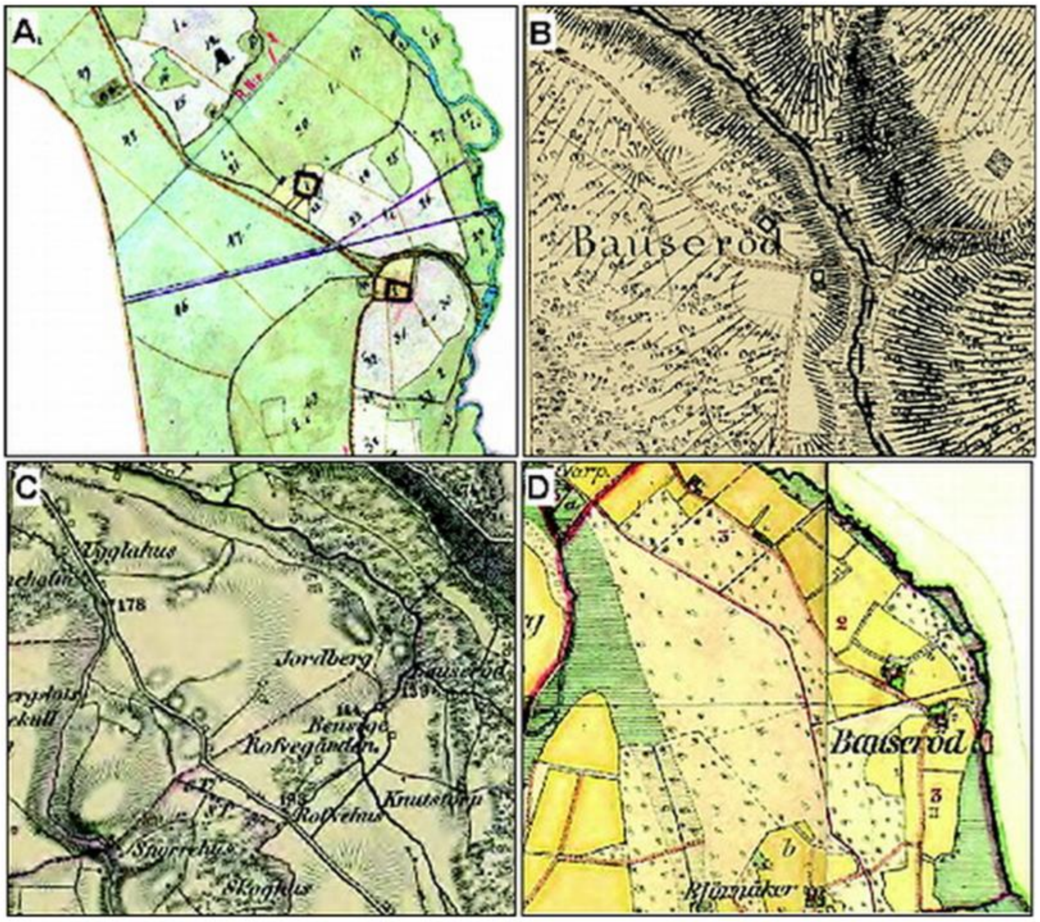


Figure 3.2: *Four historical map series depicting areas of the same part of Kågeröd parish. A) Land survey map from 1831. B) Military topographic survey map for the time period 1812-1820. C) Topographic map from 1860. D) Economic map for the time period 1910-1915 ©Lantmäteriet[I2014/00579] (Hedefalk, 2014)*

As mentioned earlier, soil type data was retrieved from the Swedish Geological Survey (SGU). Soil type data over the study area is represented by polygons in vector format and stored in shape files.

Geographic information regarding the streams and lakes was derived from the military topographical survey map series (1812 – 1820) and provided by the Swedish County Administrative Board of Scania. Streams are represented by lines whereas lakes are represented by polygons. The data is stored in vector shape files. It should be mentioned that no metadata regarding the specific data collection year is included in the attribute table for every stream or lake.

Wetland data was derived from the military topographic survey map series (1812 – 1820), which in turn was made available by the Swedish County Administrative Board of Scania. Wetlands are represented by polygons, in vector format, and stored in shape files.

Data regarding the gathering places and more in particular buildings were derived from the economic map series (1910 – 1915) provided by the National Land Survey of Sweden. Buildings are represented by point geometries, in vector format and are stored in shape files.

All the geographic data, modern as well as historical, are defined in the geodetic reference system SWEREF 99 TM (Swedish Realization of ETRS 89 in UTM 33 projection).

3.2.2 Definition and Data of the Geographic Levels

This study implemented the use of four different geographic levels: property units, initial property units, addresses and parishes. In general, a property unit represents an area that is smaller or equal to an area covered by an initial property unit or address, which in turn is smaller than the area covered by a parish. Initial property units and addresses represent two different ways of expressing ownership.

The property unit data has been transformed from snap shot representation (i.e., based on a single map) to an object lifeline representation. This implies the existence of related data regarding the year each property unit was created, changed and possibly ceased to exist. To create this object lifeline representation, textual data (from e.g. maps and poll-tax registers) as well as cadastral dossiers were used (Hedefalk et al., 2015). Property units represent the geometry of a property (Figure 3.3). Property units do not contain any information regarding their owner. The size of property units in the study area ranged from 0.001 – 5 km^2. Notably, the largest property units are located in Halmstad and Sireköpinge, which could be linked to the fact that the land in the aforementioned parishes was predominantly owned by the nobility. A few very large property units are also found in Kävlinge. This suggests the existence of greater differences in the size of property units within these parishes than in Hög.

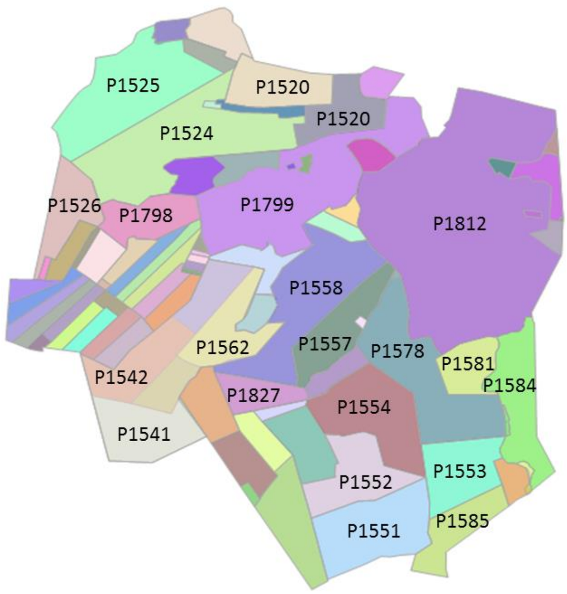


Figure 3.3: *Halmstad Property Units*

As mentioned before, initial property units were formed as a result of the land reforms and were constituted of the sum of the total property unit area belonging to a certain individual (Figure 3.4). Because not all villages implemented the land reforms during the same time, additional work was carried out to create an initial-property-unit layer with the initial property units of all villages for all parishes in the study area. The information on which the creation of the initial property unit geometries were based upon was extracted from the Land Surveyor Map Series 1757-1863. The size of initial property units ranged from 0.002 – 8 km^2 approximately. Whilst smaller initial property units exist in all parishes, the largest initial property units are mainly found in Sireköpinge and Halmstad.

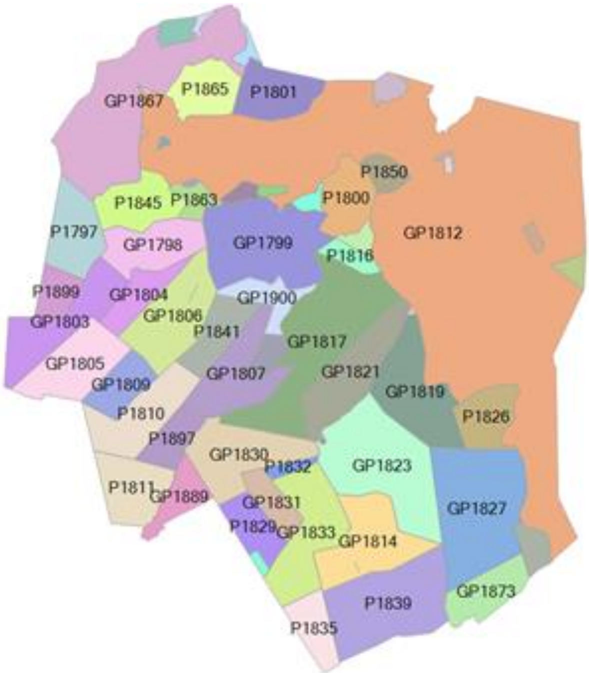


Figure 3.4: *Initial Property Units of Halmstad Parish*

Address units are areal expressions of properties derived from taxation documents. Since these documents are primarily concerned with individual property information, an address unit reflects the owner of the property. Because of this, it is possible for several property units to have the same address without being adjacent (Figure 3.5.). The size of an address unit could range from 0.002 – 8 km^2 approximately. It is important to mention that the smallest and largest address units are found in Halmstad, Sireköpinge and Kävlinge, suggesting a less homogenous in size set of address units in these parishes. Hög is the only parish where all property units that belong to the same address, constitute a single-part address unit. On the contrary, all other parishes have occurences of address units consisting of two or more, not adjacent parts. The extent to which the parts of an address unit can be spread over the parish area varies. However, Kävlinge can be noted as the parish that contains the address units whose parts are most widely spread.

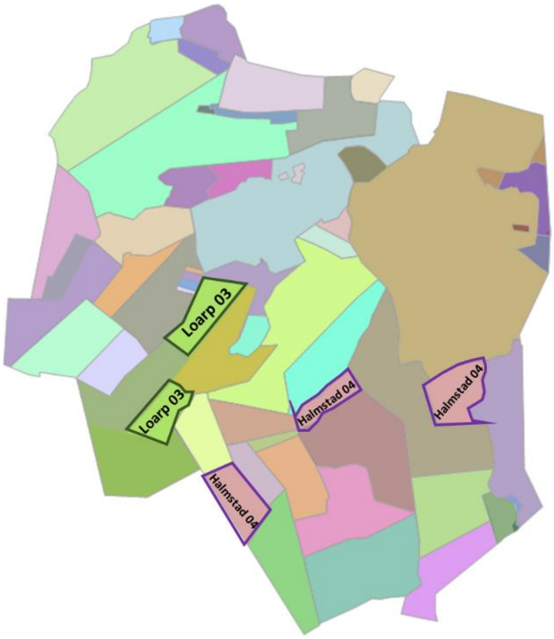


Figure 3.5: *Address Units of Halmstad Parish. The two addresses Loarp 03 and Halmstad 04 are highlighted to demonstrate how an address unit can be constituted of property units that are not adjacent.*

A parish unit is represented by a geometry containing all the addresses, initial property units and property units belonging to the same parish (Figure 3.6). Parishes were also used as administrative boundaries during the given time period.

Figure 3.6: *Parish Unit of Halmstad Parish*

3.2.3 Demographic Data

Demographic data was derived from SEDD (Scanian Economic-Demographic Database), which in turn acquired the relevant information from vital registers, catechetical examination registers/parish registers and annual poll-tax registers. In this study, demographic data is restricted to the number of individuals inhabiting a certain property unit during a specific year in time.

SEDD contains information from the parishes Halmstad, Hög, Kågeröd, Kävlinge and Sireköpinge (Figure 3.1). However, for this study we have excluded data from Kågeröd since the quality of the geographic data there is somewhat worse than for the other parishes.

Additional work has been carried out to link individuals to their place of residence on an initial-property-unit level. Information regarding the property units and their assigned residents was made available on an annual basis, by the cross-reference of annual poll-tax registers, parish registers and vital registers (Hedefalk et al., 2015). The initial property units used for the georeferencing of individuals have coordinates in the geodetic reference system SWEREF 99 TM.

3.3 Method

The objective of this project is achieved by implementing the following. Firstly, the database data was thoroughly inspected in order to obtain a good knowledge of the provided information. Secondly, the definition of the geographic context variables was mainly conducted as a theoretical investigation of effects of different definitions. Simulations were performed over the provided data set, in order to evaluate these definitions. The results of the aforementioned simulations are described in greater detail in chapter 4.

Before any of the aforementioned simulations could be performed, the database data had to be tested for possible faults. Attempts were made to clean the data from spatially and temporally overlapping information. Duplicates were also removed. Property units that were categorized to belong to one parish but spatially overlapped with another were also removed. Polygons representing buildings in the property unit data set were excluded from the study. New tables were created to store information regarding the parishes and initial property units. Additional work was done to create links between property units and the initial property units they belonged to. The centroid of all polygon geometries was computed and stored in a separate column for reasons of optimization. Minor corrections were made in misspellings of address names, to ensure the accuracy of the results.

The investigation of suitable geographic level is performed as a quantitative study. Values for the geographic context variables are computed for every geographic unit level. These values are later used to statistically compare the results on different geographic unit levels.

Simpler analyses are performed using standard tools such as Microsoft SQL Server 2012 – Developer Edition and Microsoft Office Excel 2013. Some analyses are also performed with the help of a specific application, which was designed to visualize the results of the geographic context variables.

3.4 Software Used

The application code was written in C# using the Integrated Development Environment **Microsoft Visual Studio 2012 Ultimate Edition. Microsoft SQL Server 2012 Developer Edition** was the relational database management system used to store the geographic data. Some figures depicting specific spatial results were also derived from Microsoft's SQL Server Management Studio tab "Spatial Results", which provides a visual result of spatial queries. All geographic context variable computations were performed by spatial queries written in the database structured querying language **SQL**. Some parts of the statistical analysis were also done using SQL code. The linkage between the application form and the spatial database was achieved with the help of **DotSpatial**, which is a geographic information system library that allows developers to incorporate spatial data, analysis and mapping functionality into their applications (DotSpatial, 2015). **ESRI ArcGIS 10.2.2 ArcMap** was used to store the computed results of the geographic context variables in the form of shape files, as well as to produce some of the maps. Finally, **Microsoft Office Excel 2013** was used to perform the overall statistical analysis.

3.5 Geographic Context Variable Application

An application was developed in order to provide users with the opportunity to view the results of the computed geographic context variables. The application presents the results in a tabular output as well as in the form of a digital map. Users have the choice to select the results of which geographic context variable they want to view. Additionally, they have the option to choose the geographic level (i.e. property unit level, initial property unit level, address unit level or parish unit level) and the areal extent of the results (i.e. they can choose between viewing the results of one specific parish or the results of the entire study area). The application was built in order to enhance the interpretation of the results by visualizing them in the form of digital maps. It also allows the user to further process the results. Tabular outputs can be exported as excel files and be imported to statistical packages for further processing. A screenshot of the application Graphical User's Interface (GUI) is shown in Figure 3.7.

The application's main window (Figure 3.7) consists of seven parts:

1. Menu-Bar
2. Toolbar
3. Selection-Tab (Includes the application form)
4. Legend-Tab
5. Progress-Bar
6. Map Frame
7. Coordinates

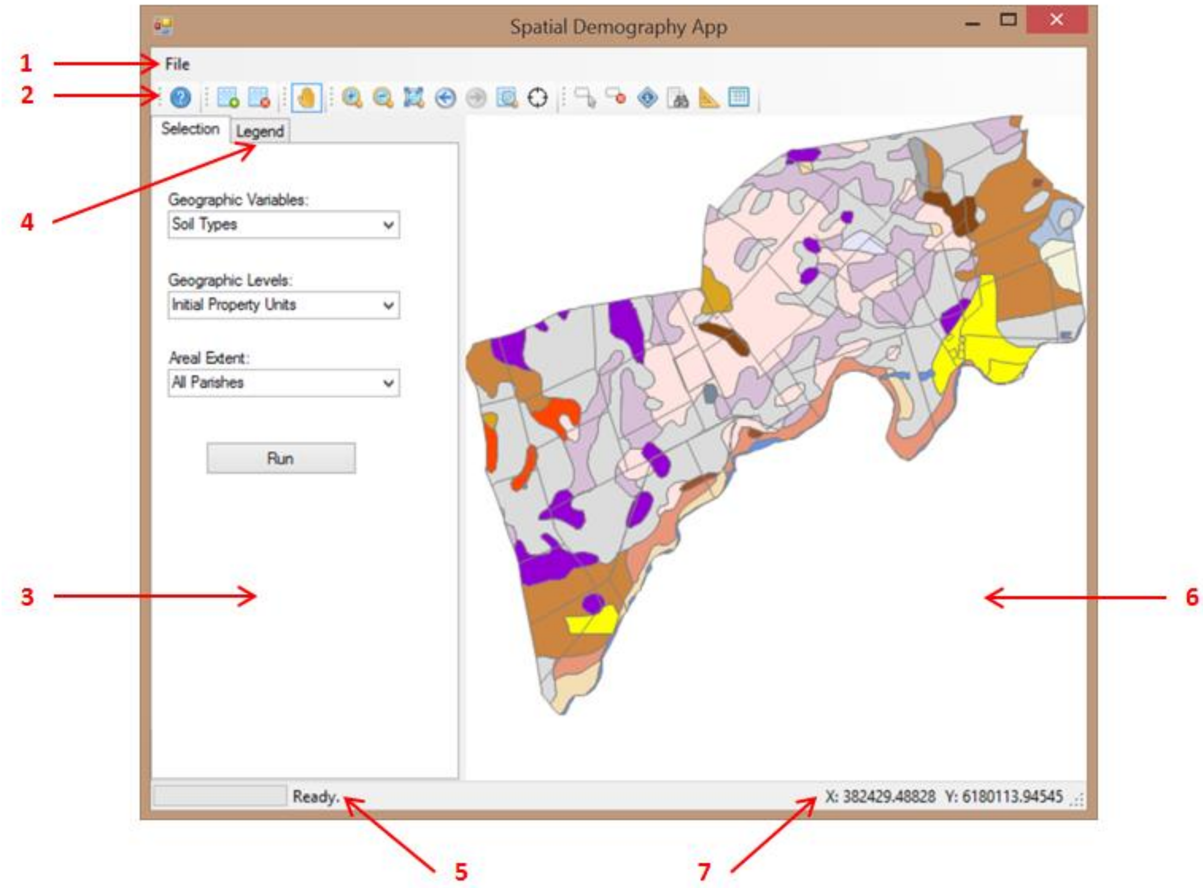


Figure 3.7: *Application GUI – Selection tab*

From the menu bar the user is able to open a new main window, browse to import shapefiles, save the project, print the current map composition in a pictorial format, restore the window to its original exstent, exit the application or access additional information regarding the application.

The tools in the toolbar allow the user to further process the contents of the map area. Every toolbar-icon corresponds to a specific process. A brief explanation of the function of every tool is presented below:

 Help – tool (provides info regarding the function of the toolbar buttons)

 Add Layer – tool (adds a layer to the map composition frame)

 Remove Layer – tool (removes the selected layer from the map composition frame)

 Pan – tool (allows the user to move objects around in the map frame by using the mouse cursor)

 Zoom In – tool (zooms in to a selected area)

 Zoom Out – tool (zooms out from aselected area)

 Zoom to Full Extent – tool (zooms out to the original extent of the map)

 Zoom to Previous – tool (zooms back to previous scale extent)

 Zoom to Next – tool (zooms back to the following extent)

 Zoom to Layer – tool (zooms in to the selected layer from the data frame)

 Zoom to Coordinates – tool (asks the user to provide some coordinates and zooms in to that area)

 Select – tool (highlights the selected spatial object in the map)

 Unselect All – tool (Unselects all selected spatial objects in the map)

 Identify – tool (Shows all information related to the selected spatial object)

 Find – tool (Highlights all spatial objects that comply to the conditions of a specific sql query)

 Measure – tool (Measures the distance between different points)

 View Attribute Table (Opens the attribute table of the selected data layer)

The selection-tab contains a form which allows the user to configure certain parameters related to the visual presentation of the geographic context variable results. As shown in Figure 3.8, the legend-tab contains information related to every map layer. The user fills the form and presses the execution button. This results in the formation of a request which is sent to the database to retrieve the specified data. The progress of this procedure is shown in the progress-bar. The output of the aforementioned procedure is displayed in the map frame. When the mouse cursor is in the map frame, the coordinates of its exact location are displayed in the bottom left of the window.

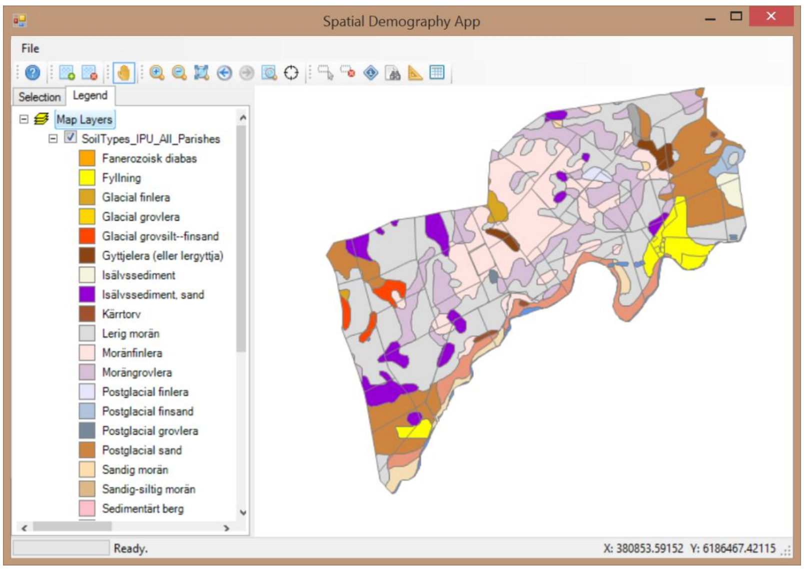


Figure 3.8: *Application GUI – Legend tab*

3.5.1 System Design

The system consists of the application form and an ESRI ArcGIS GeoDatabase. The user communicates with the system through the application form. Once the user has filled the form and submitted it, the system creates an instance call to the database requesting the specified data. The database retrieves and returns this information to the application. The results are being communicated to the user as digital maps and tabular data. The data flow within the system is illustrated in Figure 3.9.

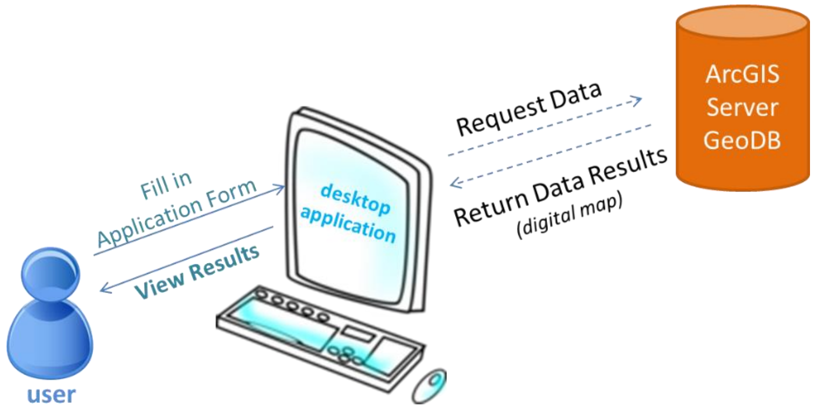


Figure 3.9: *Application Deployment Diagram*

The user communicates with the system through the application's GUI. This is achieved with the help of several controls (i.e. toolbars, menubars, checkboxes, drop-down lists and buttons) that control the user's inputs. Figure 3.10 presents all the options the user is presented with in the menubar.

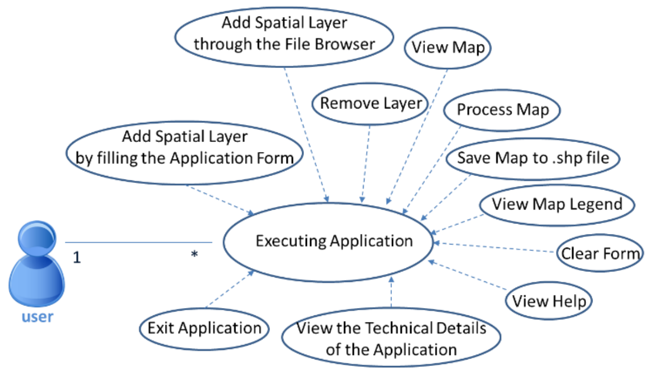


Figure 3.10: *Use Case Diagram showing options for further processing of the spatial data in the map frame.*

Through the form in the selection-tab, the user can choose the geographic context variable, the geographic level, and the areal extent at which he/she wishes the geographic context variable to be calculated over (Figure 3.11).

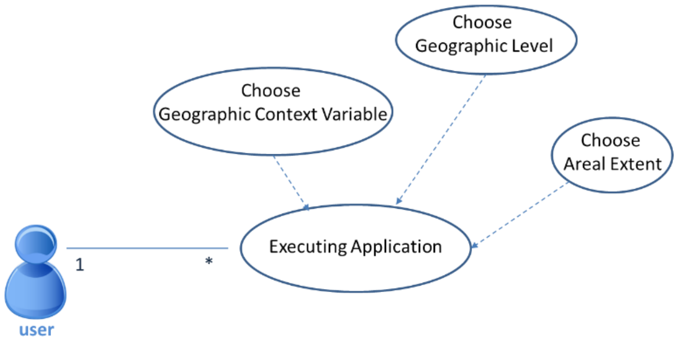


Figure 3.11: *Use Case Diagram depicting the application form options*

The use case diagram in Figure 3.12 depicts additional options to further process the spatial data in the map frame.

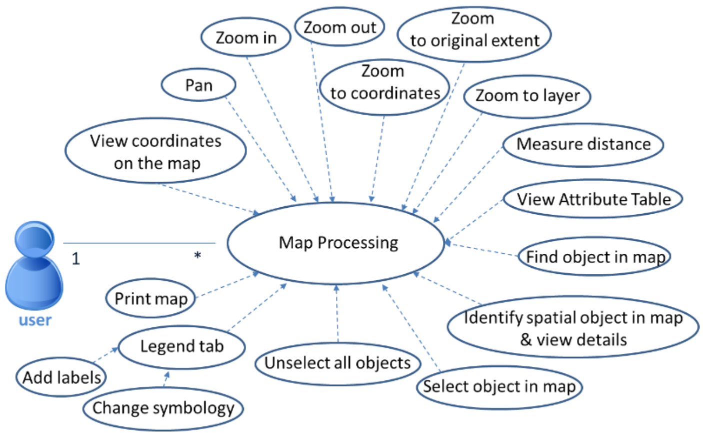


Figure 3.12: *Use Case Diagram showing the options of the user whilst processing a digital map*

More in particular, the screenshots in Figure 3.13 display how the distance-measuring tool works. With the help of this tool, the user is able to calculate distances between areas of interest in different units (i.e. meters, kilometers, miles, etc.). Distances can be measured directly through a single line segment or indirectly through several line segments. In the second case, the tool's window shows the distance of the last line segment as well as the total distance of all drawn line segments.

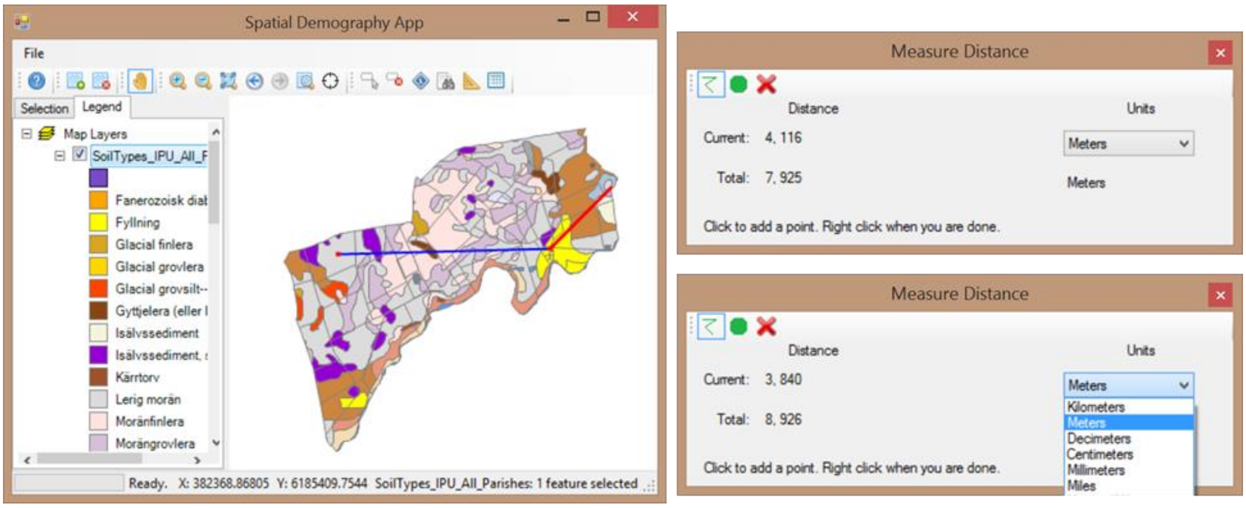


Figure 3.13: *Measuring-Distance Tool*

The "identification" tool lets the user select a spatial object in the map and then prompts a window with summarized information about that object (Figure 3.14).

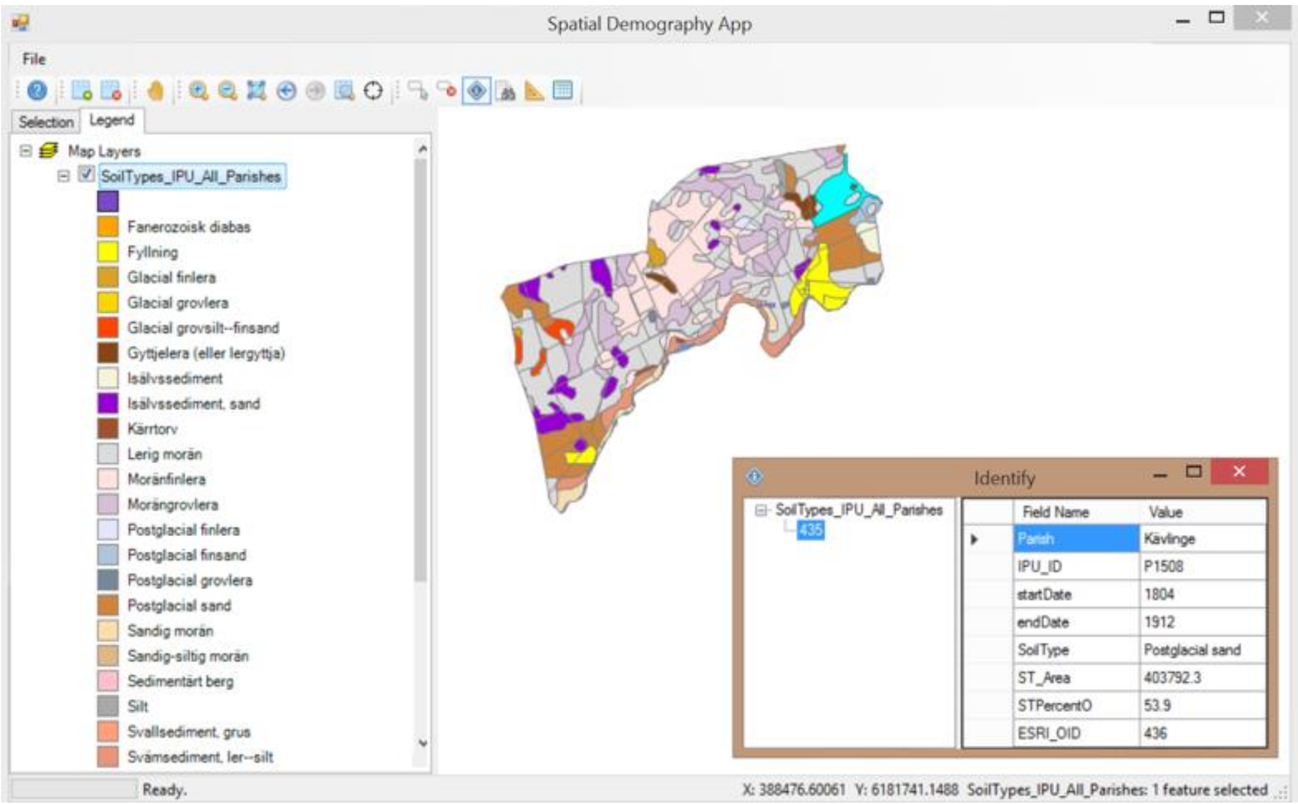


Figure 3.14: *Identify Tool*

If the user is specifically interested in a particular area, then it is possible to zoom in to that area by entering the corresponding coordinates in the “zoom-to-coordinates” tool, as shown in Figure 3.15.

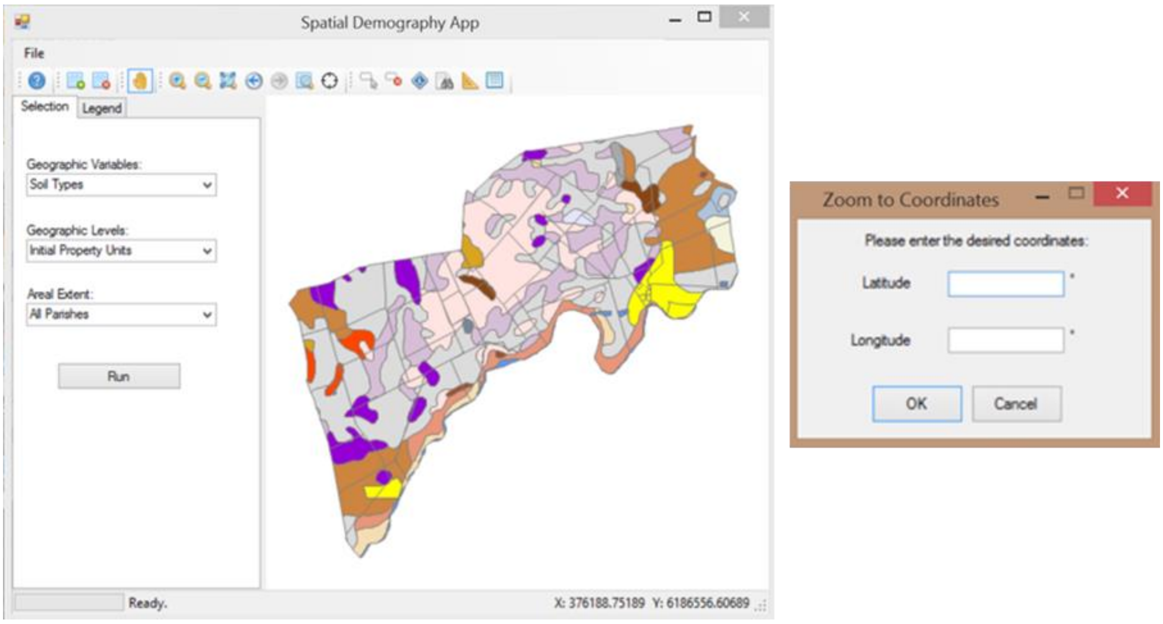


Figure 3.15: *Zoom-to-Coordinates Tool*

As mentioned before the results of a user’s request are also shown in tabular format. This is achieved through the attribute table of every selected spatial data layer. The application allows the user to further process the tabular output. Figure 3.16 presents all the available options, whilst the screenshots in Figures 3.17, 3.18, 3.19, 3.20, and 3.22 depict the attribute table editor and its different tools and editing, viewing, selecting and options correspondingly.

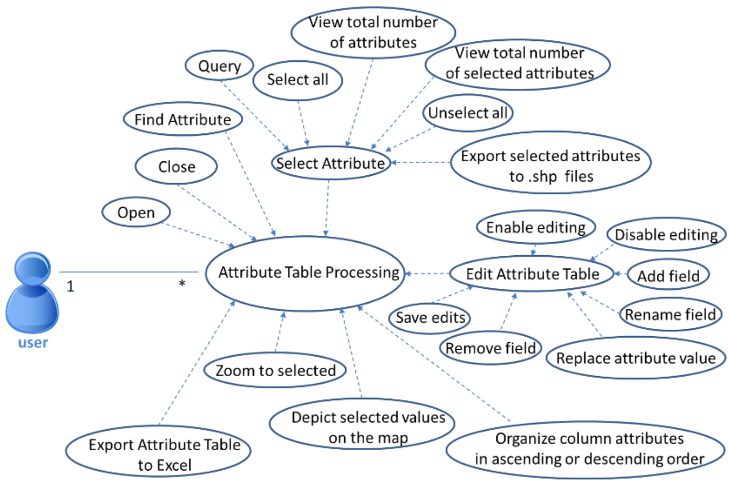
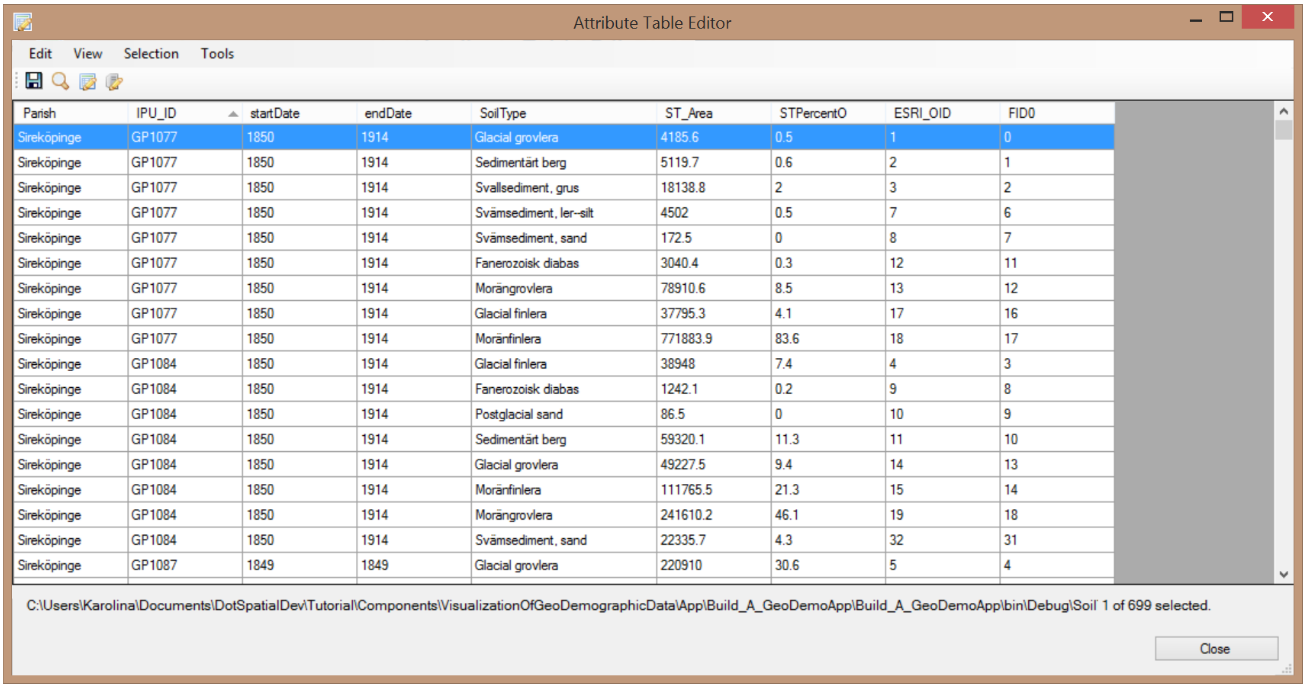
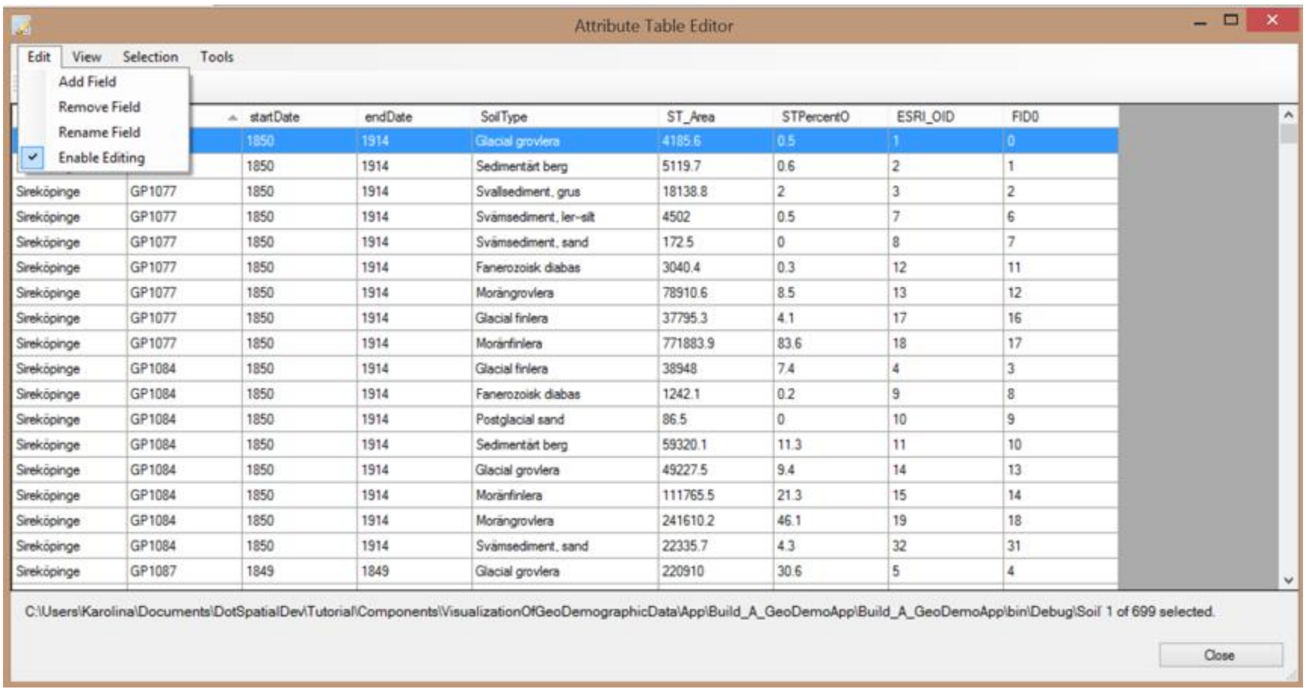


Figure 3.16: *Use Case Diagram showing the options of processing the attribute table of a spatial data layer*



Parish	IPU_ID	startDate	endDate	SoilType	ST_Area	STPercentO	ESRI_OID	FID0
Sireköpinge	GP1077	1850	1914	Glacial grovlera	4185.6	0.5	1	0
Sireköpinge	GP1077	1850	1914	Sedimentärt berg	5119.7	0.6	2	1
Sireköpinge	GP1077	1850	1914	Svallsediment, grus	18138.8	2	3	2
Sireköpinge	GP1077	1850	1914	Svämsediment, ler–silt	4502	0.5	7	6
Sireköpinge	GP1077	1850	1914	Svämsediment, sand	172.5	0	8	7
Sireköpinge	GP1077	1850	1914	Fanerozoisk diabas	3040.4	0.3	12	11
Sireköpinge	GP1077	1850	1914	Morängrovlera	78910.6	8.5	13	12
Sireköpinge	GP1077	1850	1914	Glacial finlera	37795.3	4.1	17	16
Sireköpinge	GP1077	1850	1914	Moränfinlera	771883.9	83.6	18	17
Sireköpinge	GP1084	1850	1914	Glacial finlera	38948	7.4	4	3
Sireköpinge	GP1084	1850	1914	Fanerozoisk diabas	1242.1	0.2	9	8
Sireköpinge	GP1084	1850	1914	Postglacial sand	86.5	0	10	9
Sireköpinge	GP1084	1850	1914	Sedimentärt berg	59320.1	11.3	11	10
Sireköpinge	GP1084	1850	1914	Glacial grovlera	49227.5	9.4	14	13
Sireköpinge	GP1084	1850	1914	Moränfinlera	111765.5	21.3	15	14
Sireköpinge	GP1084	1850	1914	Morängrovlera	241610.2	46.1	19	18
Sireköpinge	GP1084	1850	1914	Svämsediment, sand	22335.7	4.3	32	31
Sireköpinge	GP1087	1849	1849	Glacial grovlera	220910	30.6	5	4

Figure 3.17: *Application – Attribute Table Editor (Tabular Output)*



		startDate	endDate	SoilType	ST_Area	STPercentO	ESRI_OID	FID0
		1850	1914	Glacial grovlera	4185.6	0.5	1	0
		1850	1914	Sedimentärt berg	5119.7	0.6	2	1
Sireköpinge	GP1077	1850	1914	Svallsediment, grus	18138.8	2	3	2
Sireköpinge	GP1077	1850	1914	Svämsediment, ler–silt	4502	0.5	7	6
Sireköpinge	GP1077	1850	1914	Svämsediment, sand	172.5	0	8	7
Sireköpinge	GP1077	1850	1914	Fanerozoisk diabas	3040.4	0.3	12	11
Sireköpinge	GP1077	1850	1914	Morängrovlera	78910.6	8.5	13	12
Sireköpinge	GP1077	1850	1914	Glacial finlera	37795.3	4.1	17	16
Sireköpinge	GP1077	1850	1914	Moränfinlera	771883.9	83.6	18	17
Sireköpinge	GP1084	1850	1914	Glacial finlera	38948	7.4	4	3
Sireköpinge	GP1084	1850	1914	Fanerozoisk diabas	1242.1	0.2	9	8
Sireköpinge	GP1084	1850	1914	Postglacial sand	86.5	0	10	9
Sireköpinge	GP1084	1850	1914	Sedimentärt berg	59320.1	11.3	11	10
Sireköpinge	GP1084	1850	1914	Glacial grovlera	49227.5	9.4	14	13
Sireköpinge	GP1084	1850	1914	Moränfinlera	111765.5	21.3	15	14
Sireköpinge	GP1084	1850	1914	Morängrovlera	241610.2	46.1	19	18
Sireköpinge	GP1084	1850	1914	Svämsediment, sand	22335.7	4.3	32	31
Sireköpinge	GP1087	1849	1849	Glacial grovlera	220910	30.6	5	4

Figure 3.18: *Application – Attribute Table Editor (Edit Options)*

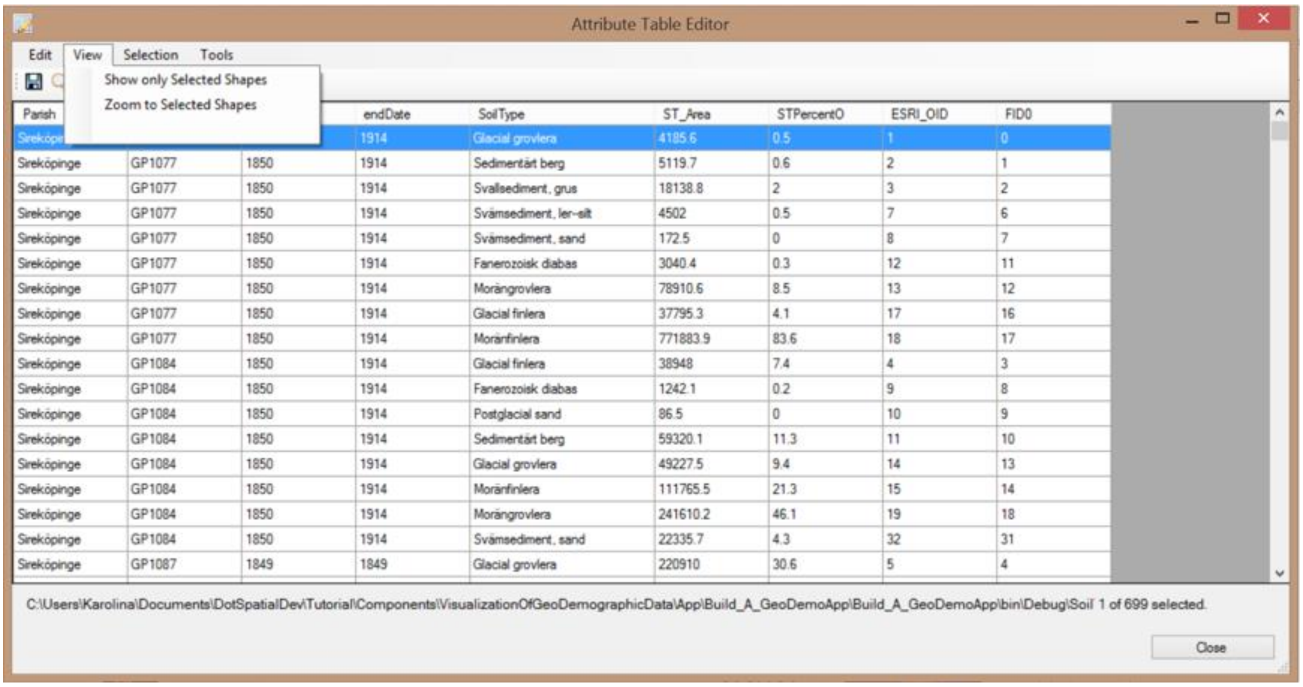


Figure 3.19: *Application – Attribute Table Editor (View Options)*

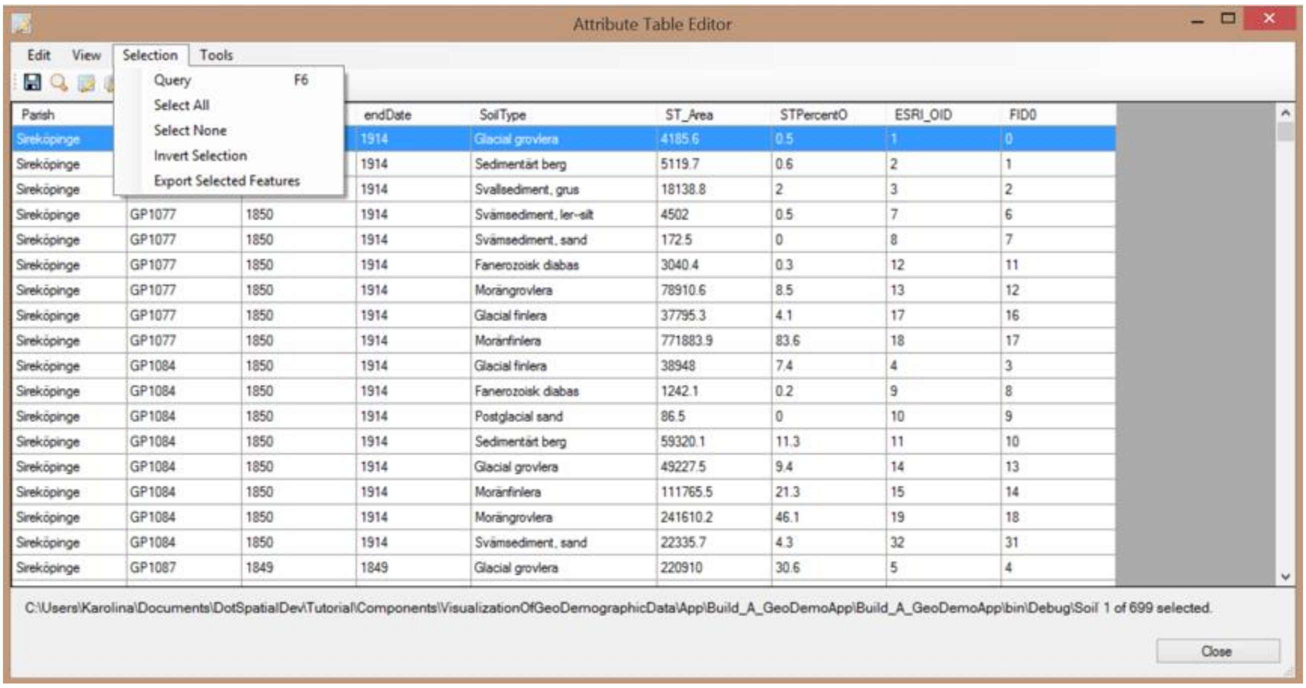


Figure 3.20: *Application – Attribute Table Editor (Selection Options)*

The “Selection” tab includes the “Query” option. This options prompts a window (Figure 3.21), called the Expression Editor, which allows the user to select only the spatial objects that have the desirable attribute values.

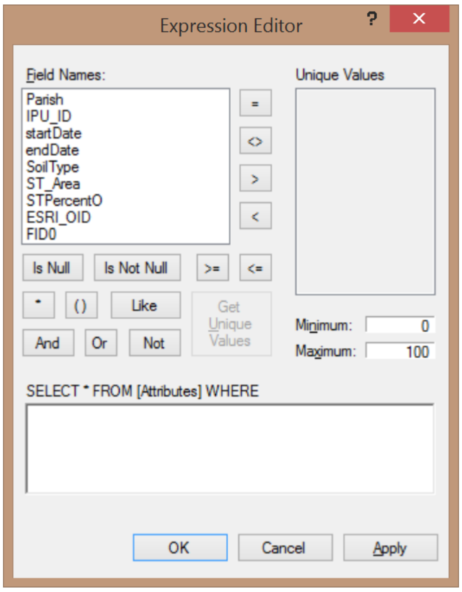


Figure 3.21: *Application – Attribute Table Editor (Find Option – Expression Editor)*

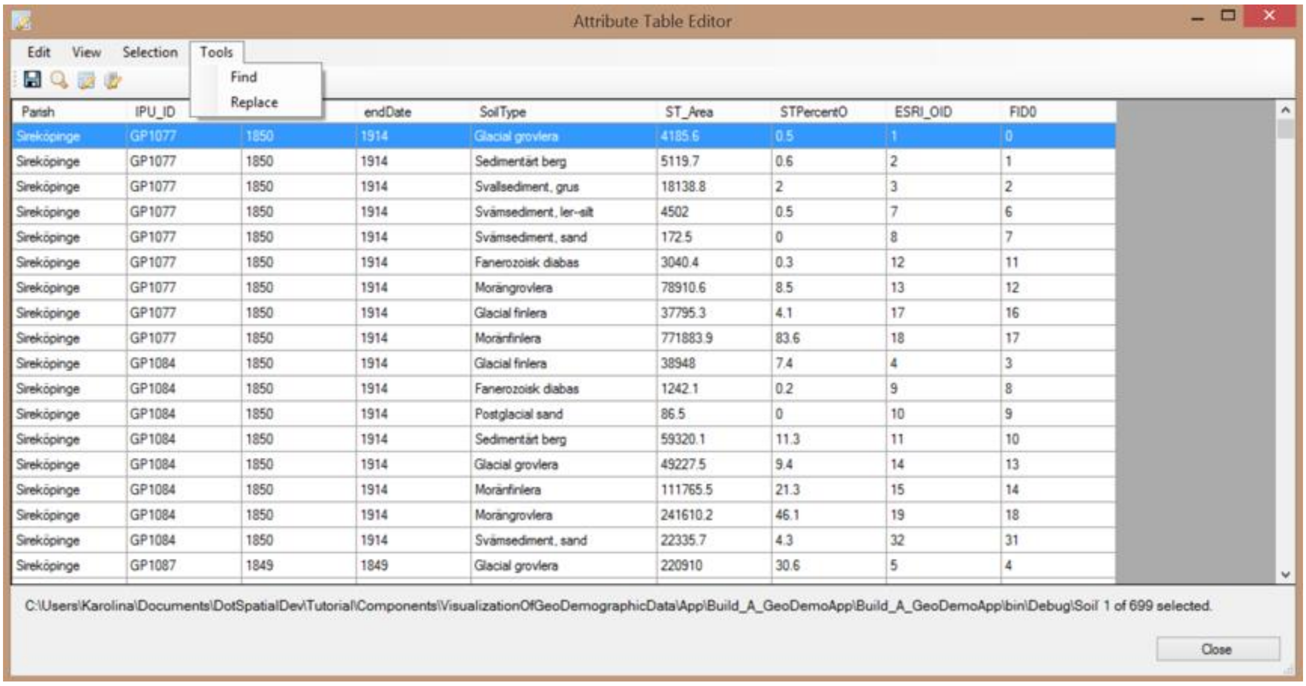


Figure 3.22: *Application – Attribute Table Editor (Tools)*

Figure 3.23 contains a use case diagram wich illustrates all the available formatting options for printing a map.

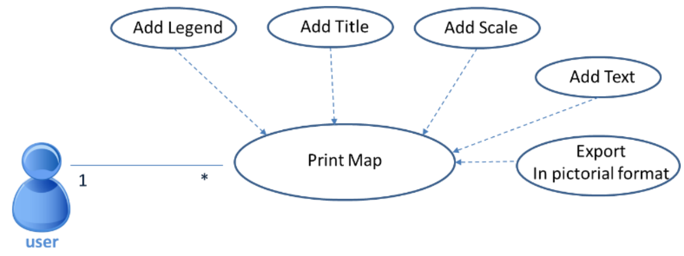


Figure 3.23: *Use Case Diagram showing the available formatting options for printing a map*

The screenshots in Figures 3.24, 3.25, 3.26, and 3.27 depict the different processing options that are available through the Print Layout tool.

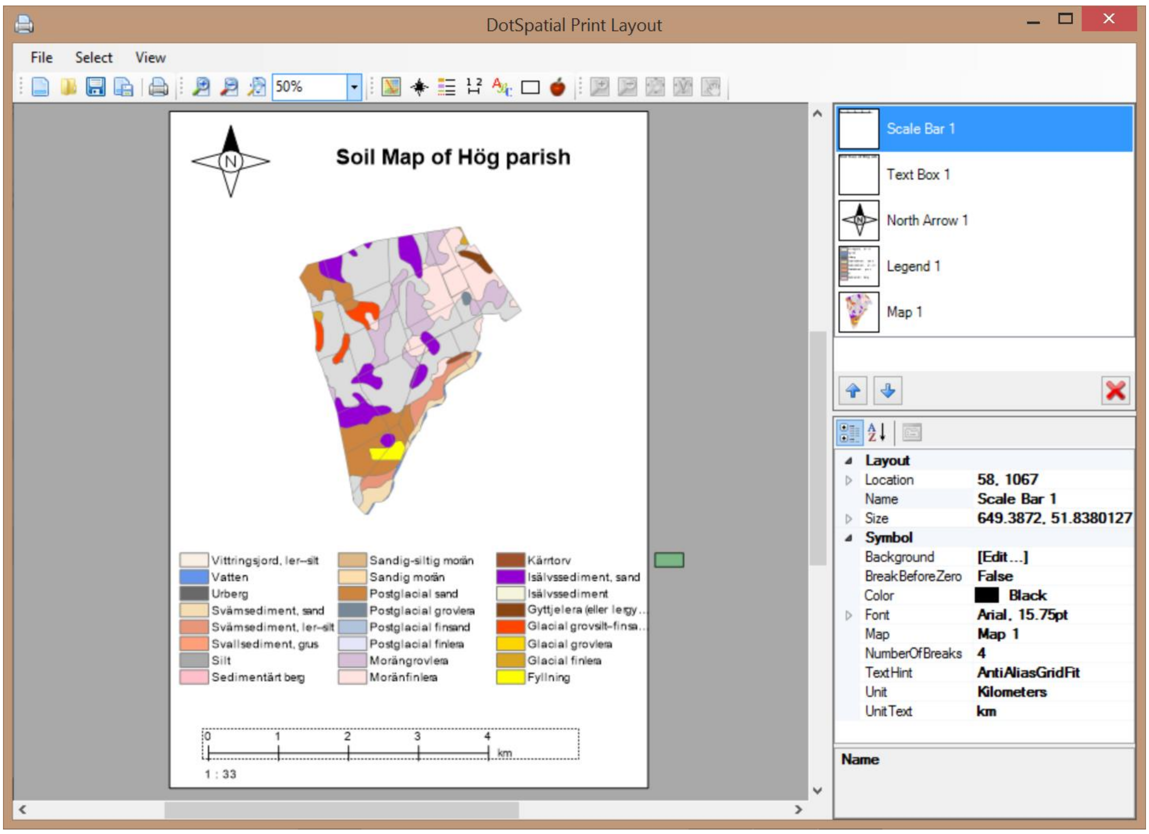


Figure 3.24: *Screenshot of the Print Map Layout Tool*

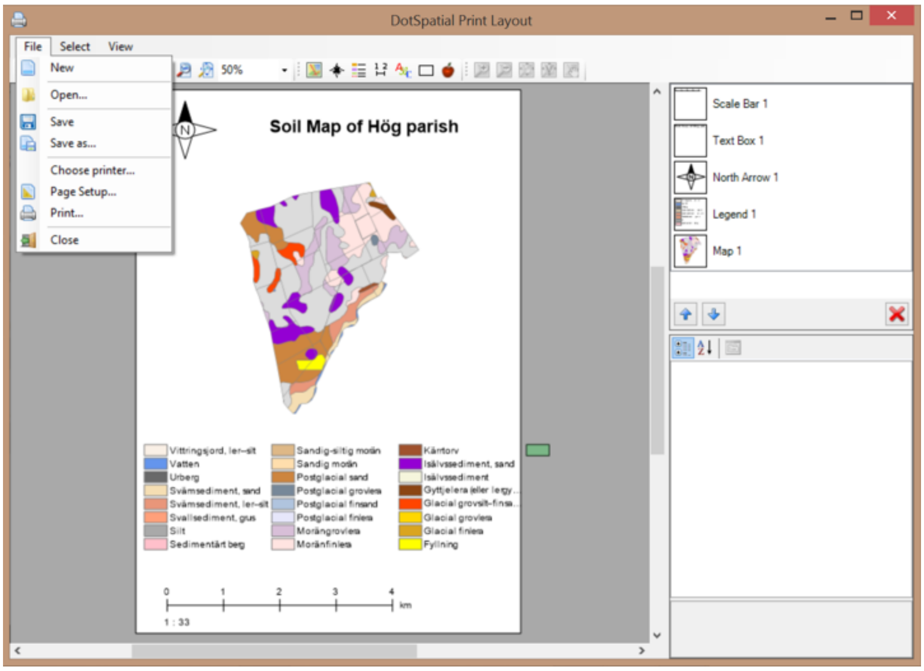


Figure 3.25: *Screenshot of the Print Map Layout Tool – File Options*

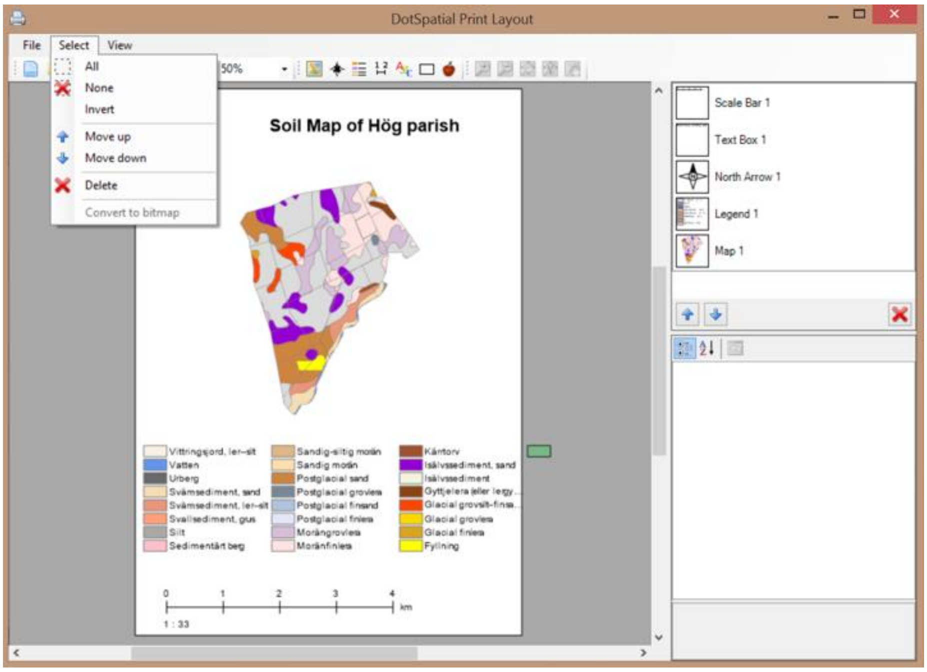


Figure 3.26: *Screenshot of the Print Map Layout Tool – Select Options*

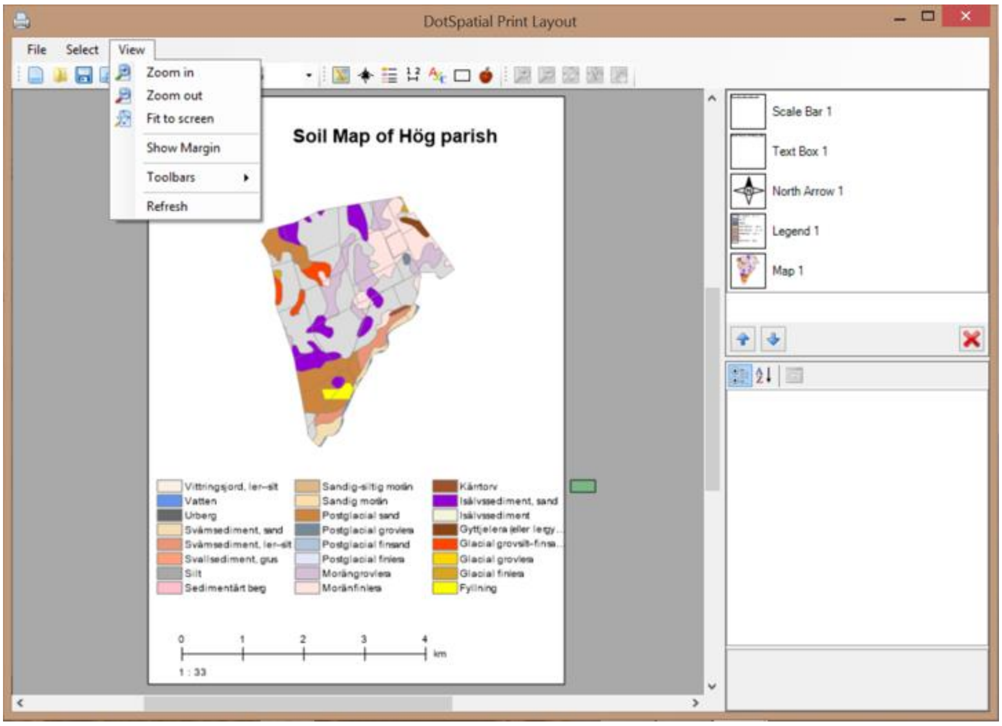


Figure 3.27: *Screenshot of the Print Map Layout Tool – View Options*

The application consists of two main classes; the application form and the map-selection class, and three enumerations (holding the geographic variables, the geographic levels, and the parishes). An instance of the program can support multiple forms, whilst a form can support multiple map selections. This information is depicted in the Application Class Diagram in Figure 3.28.

The schema of the database used to compute the various geographic context variables over the different geographic levels is included in Figure 3.29. There are nine tables in total in the database, of which six are related to the geographic context variables (e.g. population, soil type, wetlands, streams, lakes, and buildings) and three are related to the geographic levels (e.g. property units, initial property units, and parishes). Information regarding the address level is included as an attribute in the property unit table. All tables contain a geometry attribute (usually denoted as *Shape*) except from the population table. This attribute contains the spatial information and uniquely identifies every entry in the table. For practical reasons, a primary key field (different from the geometry field) has been assigned to every table.

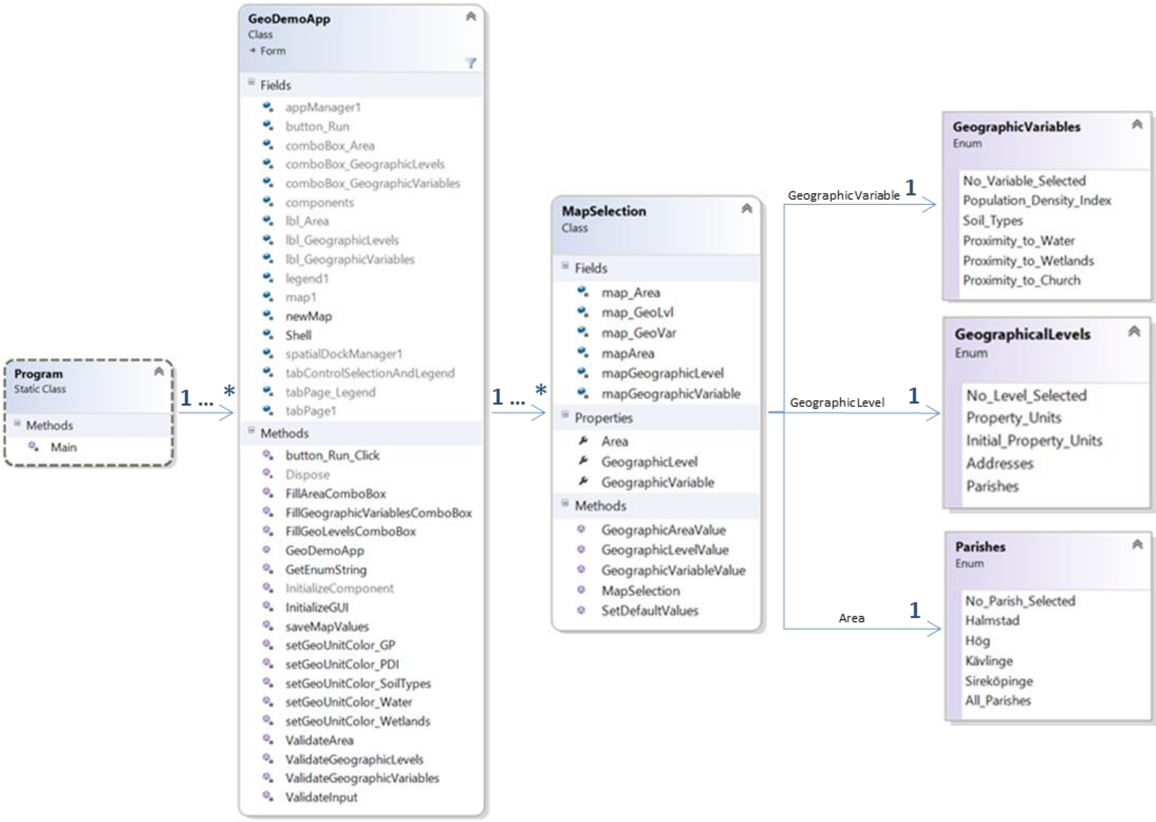


Figure 3.28: *Application Class Diagram*

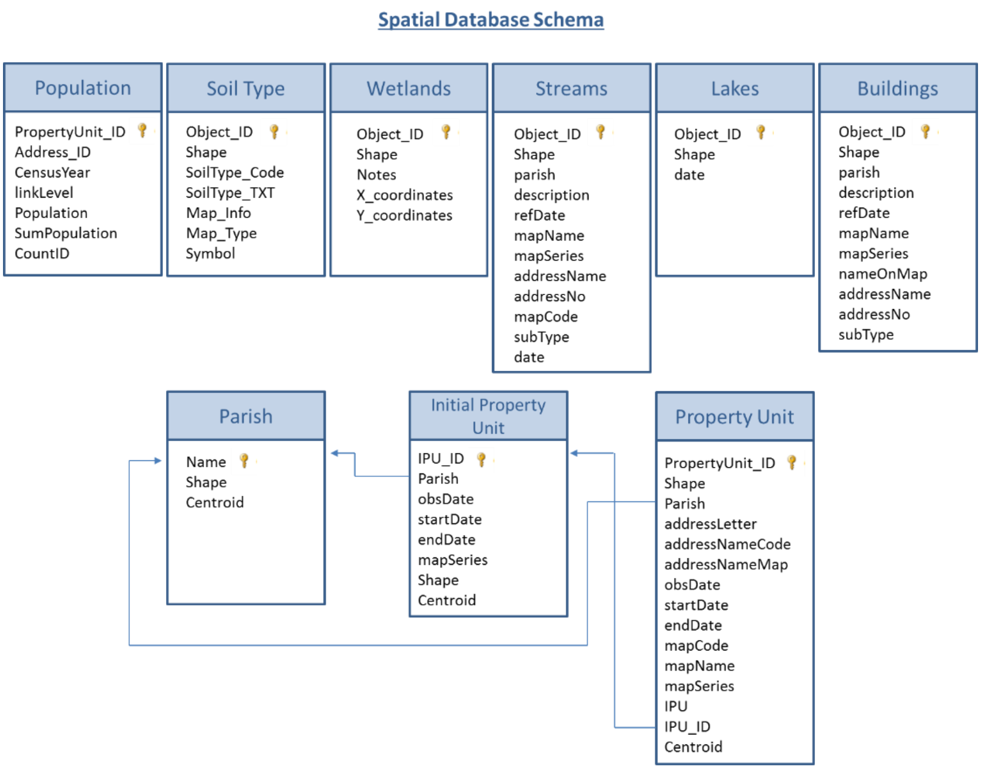


Figure 3.29: *Application Database Schema*

4 Definition of Geographic Context Variables

4.1 Introduction

In recent years, demographic research has increasingly come to utilize geographic context variables in its studies. Demography, which is inextricably linked or closely intertwined with the concept of time and space, has experienced a rapid evolution due to the development and wide dissemination of GIS. GIS contribute a technological platform to perform complex operations between spatial and non-spatial data including the temporal dimension. These tools enabled demographers to study and compare much larger areas, for a much longer period in time than ever before.

The notion or term of geographic context variables has emerged as a result of the integration of geographic and non-geographic data in the context of a GIS. Any variable that has a geographic extent or can be represented in the geographical space is denoted as a geographic context variable. Common geographic context variables are various types of distances, indexes (e.g. population density index, income index, vegetation index, etc.) or physical attributes of geographical locations (e.g. elevation, slope) or other attributes related to geographical locations such as soil type and climate conditions (e.g. temperature, precipitation, humidity, etc.).

4.2 Selection of Geographic Context Variables

Bengtsson and Dribe (2011) mention that, though population mortality varied amongst the five parishes during the time period of this study, socioeconomic factors were not responsible for this varying pattern. It is revealed how individuals with low income or low social status display similar mortality rates as individuals with higher income or more elevated social status. Therefore, the responsible factors need to be searched for elsewhere. Famine, poor sanitation knowledge and relevant infrastructure as well as and epidemics were significant problems during the period in question. Therefore, there are several geographic context variables that could be regarded as possible mortality factors (Bengtsson & Borström, 2011). For instance, soil types (quality of food) and population density (spreading of infectious diseases) are generally included in mortality-rate researches. Proximity to gathering locations (e.g. churches, schools, markets, train stations) was also contemplated as a possible risk factor in terms of contamination is more likely to occur when more people from different locations are gathered. Since malaria was a real threat at the time, proximity to wetlands, lakes and streams (natural habitat of mosquitos – known carriers of the malaria parasitic protozoans) was also proposed as a possible affecting factor.

For the purpose of this project, data is acquired from two separate databases containing historical geographic data (i.e. buildings, property units etc.) and modern geographic data (i.e. soil types). A third database storing demographic data (SEDD) is being used to calculate the population density index for every geographic unit (property units, initial property units, addreses and parishes) at a given time. Data from all three databases is integrated and processed to produce the desired results.

4.3 General Issues Regarding the Definition of Geographic Context Variables

Before we continue with the definition of the chosen geographic context variables, it is important to reflect how a significant, fundamental concept such as distance is defined. All of the geographical context variables in this study, with the sole exception of soil types, are dependent upon this measure. There are many methods of estimating distance. Euclidean, spherical (lat/long) and network distance are amongst the most widely used methods. Because the study area is too small for the great-circle distance over the Earth's surface to differ significantly to the planar Euclidean distance, lat/long distance was not preferred.

Network distance was not implemented because of the notion that people living in rural, agricultural communities of the past, often used footpaths rather than roads to move from one place to another. A possible explanation could be that road networks were not always that widely developed and if the ground geography (e.g. planes) permitted it, footpaths allowed for a much faster way of transportation. Unfortunately, all of these paths or footpaths were usually not recorded in the maps of the time.

Another reason why network distance was not selected was that, when studying epidemiology, scientists are not only interested in how people moved but also in their exposure to certain elemments. For instance, since malaria was a quite common disease during the mid 19th century in the study area, epidemiologists might be interested in the exposure to wetland or water covered areas (natural habitat of the disease-carrying mosquitos) to estimate the probability of people being exposed to the infected mosquitos. Since exposure is not only expressed in terms of how near mosquito habitat areas people moved around during their daily activities, but also in terms of how big an area the infected mosquitos could affect, Euclidean distance would be a more representative method to implement compared to network distance.

Because of all the previously stated reasons, Euclidean distance was chosen as the most appropriate method for the needs of this study and is implemented for all distance calculations performed.

In Spatial DBMSs (Data Base Management Systems), there are several ways of calculating the shortest Euclidean distance between a geographic unit that is described as a polygon and other geographic units (e.g. polygons, lines or points). SQL Server's built in function STDistance() calculates the shortest Euclidean distance between the geometries of two separate geographic units. Similar functions exist in other Spatial DBMSs. It is possible to separate the different ways of computing the shortest Euclidean distance between geographical units -where one of them is always a polygon- in three categories:

I. **Polygon – Point**
II. **Polygon – Line**
III. **Polygon – Polygon**

By examining the first category, two different methods of calculating the shortest Euclidean distance between the geographic units involved can be suggested. The first method estimates distance as the shortest distance between the borders of a polygon PU_i and a point (Figure 4.1).

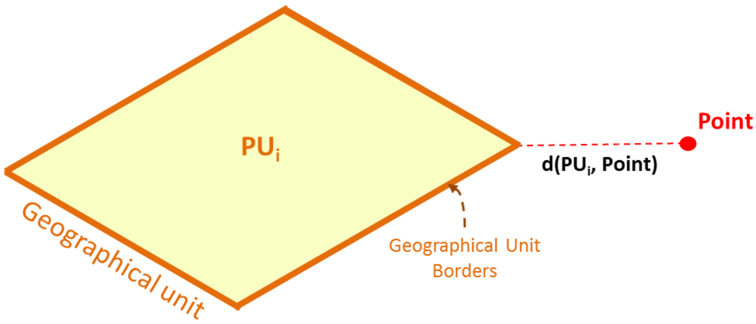


Figure 4.1: *Calculate the shortest Euclidean Distance between the borders of a polygon and a point.*

According to the second method, the shortest distance between a polygon PU_j and a point can be computed as the shortest line segment *d(PU_j-centroid, Point)* between the polygon's centroid and the point (Figure 4.2).

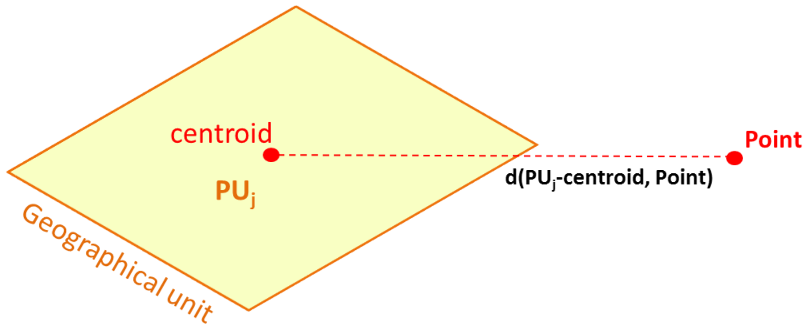


Figure 4.2: *Calculate the shortest Euclidean Distance between a polygon centroid and a point.*

At this point, it should be mentioned that the *border-to-point* distance calculation method yields a zero-distance result if the point is located within or at the border of the polygon in question. On the contrary, the *centroid-to-point* distance calculation method produces a zero-distance result, if and only if, the point coincides with the polygon's centroid. In any other case, both methods return distance values greater than zero.

The second category, like the first, also suggests two methods of performing the distance calculations. As before, the distance is either calculated from the borders or the centroid of the polygon. The shortest Euclidean distance between a polygon PU_i and a line are computed as the shortest line segment $d(PU_i, Line)$ between the polygon borders and the line (Figure 4.3).

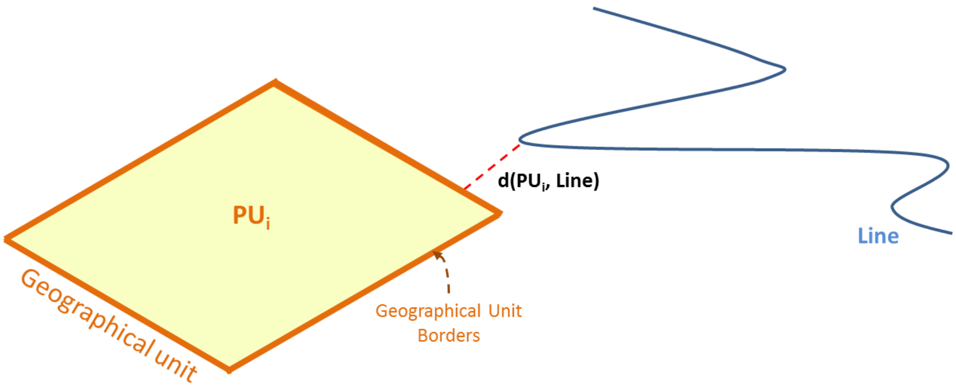


Figure 4.3: *Calculate the shortest Euclidean Distance between the borders of a polygon and a line.*

According to the second method, the shortest Euclidean distance between a polygon PU_j and a line are computed as the shortest line segment $d(PU_j\text{-}centroid, Line)$ between the polygon's centroid and the line (Figure 4.4).

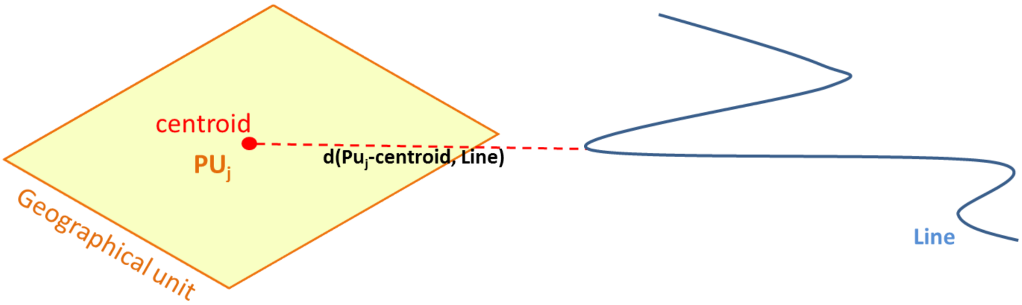


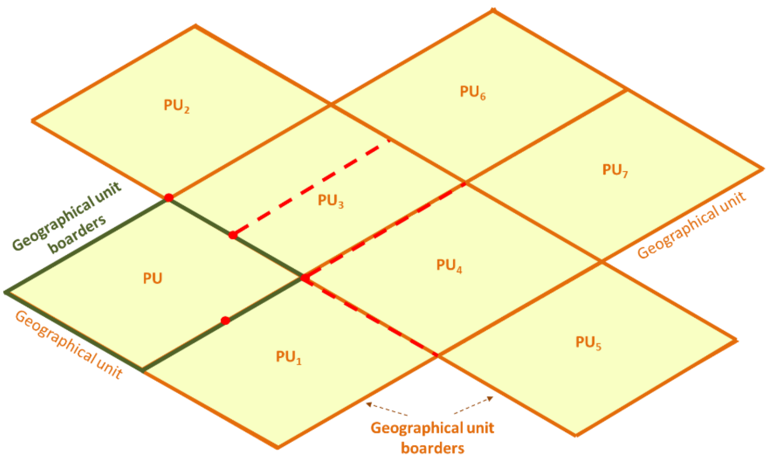
Figure 4.4: *Calculate the shortest Euclidean Distance between the centroid of a polygon and a line.*

Similarly to the *Polygon-to-Point* category, this category yields zero-distance results if parts of the line are included within or touch the polygon borders, if the border-to-line method is implemented. If the centroid-to-line method is used, a distance result is equal to zero only if the polygon centroid overlaps with the line. Any other instance produces a distance result higher than zero for both methods.

By assessing the third and final category, four different ways could be found to calculate the shortest Euclidean distance between two geographical units expressed as polygons:

i. Border to Border
ii. Border to Centroid
iii. Centroid to Border
iv. Centroid to Centroid

According to the first distance calculation method, the distance between two geographical units is expressed as the shortest Euclidean distance between their borders (Figure 4.5). The second calculation method calculates the shortest Euclidean distance between the borders of the geographical unit -whose PDI value is being computed- and the centroid of all the geographical units within the buffer area (Figure 4.6). The "Centroid to Border" method estimates the shortest Euclidean distance from the centroid of the geographical unit -whose PDI value is being computed- to the border of another geographical unit (Figure 4.7). Finally, the last calculation method computes the shortest Euclidean distance between the centroids of two geographical units (Figure 4.8).



$dPU_1 = dPU_2 = dPU_3 = dPU_4 = \mathbf{0}$

$dPU_5 \neq \mathbf{0}$, $dPU_6 \neq \mathbf{0}$, $dPU_7 \neq \mathbf{0}$

Figure 4.5: *Calculate the shortest Euclidean Distance with the 'Border to Border' distance calculation method.*

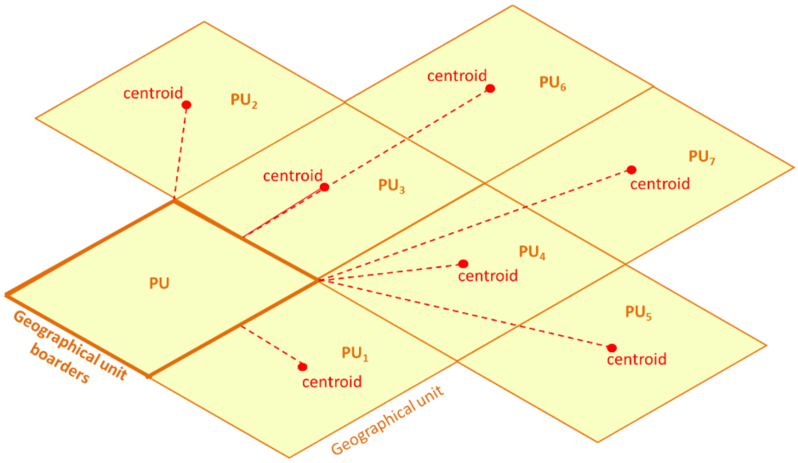


Figure 4.6: *Calculate the shortest Euclidean Distance with the 'Border to Centroid' distance calculation method.*

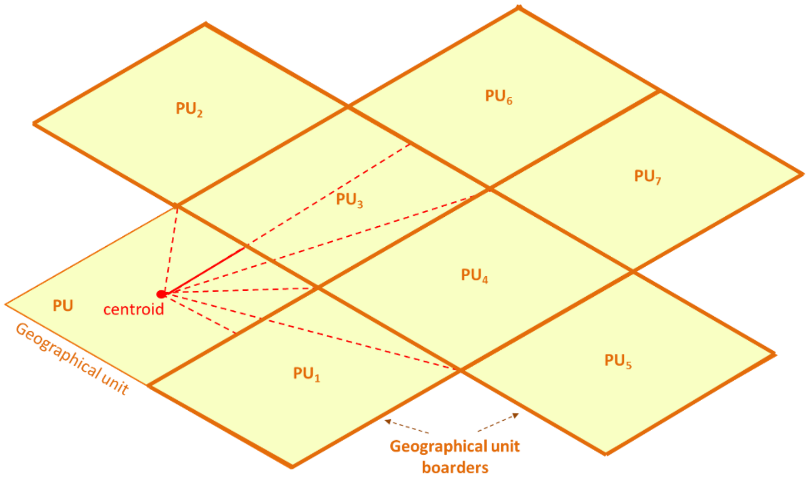


Figure 4.7: *Calculate the shortest Euclidean Distance with the 'Centroid to Border' distance calculation method.*

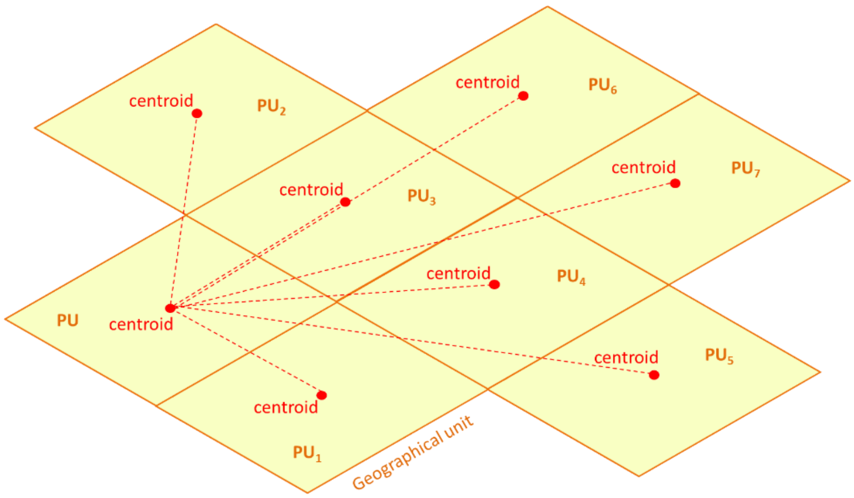


Figure 4.8: *Calculate the shortest Euclidean Distance with the 'Centroid to Centroid' distance calculation method*

The distance calculation method "Border-to-Centroid" was not implemented in this study as it was deemed irrelevant to this specific implementation. Assuming that people in old, rural, agricultural communities are mainly occupied with agricultural activities near their residential area as well as that this residential area was often located near the center of their property, it makes more sense to perform distance calculations from the centroid rather from the border of their property. Of course this is just an educated assumption, and if the uncertainty of in which part of their property unit people spent the largest part of their day, then estimating proximity from the property unit's borders might have been a good approximation. Nevertheless, when contemplating the distance to wetland area which is represented by a polygon, the distance to its border rather than its centroid is of more interest because the focus is set upon where the typical wetland conditions start to appear. The same argument applies for the distance computation to lakes, which are also represented by polygons. For these reasons, the "Border-to-Centroid" method was excluded from the study.

4.4 Definition & Computation of Geographic Context Variables

This part is dedicated to presenting the results of the computation of the geographic context variables. It is subdivided in sections, one for every geographic context variable. Every section includes a short definition of the geographic context variable, a detailed description of the computation method that was implemented, the results and a short discussion. The implementation section contains arguments of the reason why a certain method is chosen with regards to the specific requirements of the study. Examples of problematic situations that might occur as a consequence of implementing a certain computation method are also presented. The results are presented in the form of diagrams. Every diagram shows the results of a certain geographic context variable, computed by a specific method, at a given geographic level for one parish at a time. In cases where more than one computation methods have been implemented, a statistical comparison of the results is provided. The statistical comparison is performed by computing the Root Mean Square Deviation (RMSD) between the results of the two compared methods. RMSD values are produced for every parish separately as well as cumulatively for the whole study area.

The content of this part is meant to highlight the importance of selecting the most suitable calculation method that fulfils the specific requirements of the respective demographic research, as well as the importance of being aware of the vulnerabilities the chosen method entails. In cases where a demographer cooperates with a GIS expert and the computation of the geographic context variables is performed by the latter, it is essential for the demographer to be thoroughly informed about the particulars of the implemented methods. That way the demographer, in close collaboration with the GIS expert, is in better position to determine which calculation method better suits the data of the study area and at the same time better serves the objective of the research hypothesis. If not, there is a high risk of misinterpreting the meaning of the produced geographic context variable values, which may further affect the outcome of the demographic analysis. Therefore, it is very important for every demographic study including a geographic context to enclose a short detailed description of the implemented computation methods as well as the reason why they were preferred. A list of their vulnerabilities and to what extent they may come to affect the result should also be included. Good practices, such as the aforementioned, may also enhance the comparison between demographic studies that deal with the same topic.

4.4.1 Population Density Index

4.4.1.1 Definition

Population Density Index (PDI) is computed at all four geographic levels (property units, initial property units, addresses and parishes) according to the following formula (Fei Lu, 2015):

$$\mathbf{PDI} = \frac{\mathrm{M}_0}{50} + \sum_{i=1}^{n} \frac{M_i}{d_i} \qquad (1)$$

Where $\boldsymbol{M_o}$ stands for the number of people living in the geographical unit for which PDI is estimated, $\boldsymbol{M_i}$ represents the total number of people living in geographical unit ***i***, $\boldsymbol{d_i}$ is the Euclidean distance in meters between geographical unit ***o*** and geographical unit ***i***, and ***n*** is the total number of geographical units in the buffer area. The buffer area in this case has been decided to be big enough to cover all geographic sub-units within a parish. The unit Population Density Index is measured at is m^{-1}. PDI is a dynamic variable since it depends upon variables that change over time, such as the number of individuals inhabiting the area within a certain geographic unit. However, for the purpose of this study, as the linking of population information to geographical units is still a working progress, the PDI will only be calculated for the year of 1910.

This formula was preferred over the approach of just calculating the total population per geographic unit, as it considers the impact of the surrounding population as well. Ultimately, the PDI formula is capable of producing results that better capture the population patterns in the study area in relation to the demographic research objective, which focuses on mortality. Mortality caused by infectious diseases that are transmittable from human to human, should not only be dependent upon the fact of how many people live in the same geographic unit, but also on how many people live in geographic units within the wider range of their daily activities. Because most people of the time were involved in agricultural activities it is likely that, especially after the implementation of the land reforms, they spent most of their time near their property. This means that in terms of risking contracting a human-to-human transmittable infection they would be most vulnerable to the people they lived together with in the same property unit as well as to their closest neighbors. People living further away might still have constituted a risk but not one that was as direct. As the risk of exposure to human-transmittable infectious diseases seems to decline with distance, it is important to account for that when defining a geographic context variable that attempts to measure the exposure of humans to humans.

4.4.1.2 Implementation

When calculating the PDI according to the aforementioned formula, issues might rise depending on which distance calculation method is applied. Therefore, it is deemed necessary to determine how distance is to be calculated between the geographical objects.

The PDI formula does not allow the distance between two geographical units to be equal to zero. This constraint is imposed by the fact that distance is a denominator in the formula. However, if the distance between two geographical units is estimated as zero, a “division by zero” error will occur, terminating the calculation process. In theory, zero distances between geographical units can be found whilst implementing all of the previous calculation methods. Nevertheless, some of them are more prone to producing these results than others. It should also be noted that the specific characteristics (e.g. shape) of the geographical units in the study area also affect the probability of zero-distance results to be found between geographical units.

If the first calculation method was to be examined, namely the computation of the shortest Euclidean distance between the borders of two geographical units (e.g. property units), the problem becomes obvious whilst attempting to estimate the distance between a property unit and its neighboring property units. Because neighboring property units are sharing common borders, the shortest distance between their borders is equal to zero (Figure 4.9).

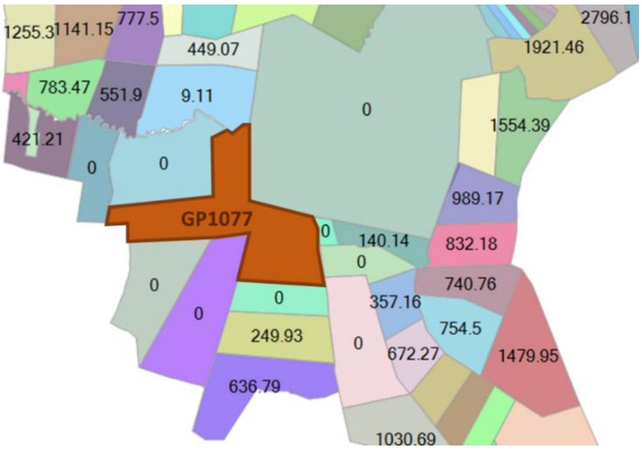


Figure 4.9: *Example of zero-distance results being produced for neighboring property units to property unit 'GP1077' whilst implementing the 'Border-To-Border' Euclidean Distance Calculation method in Sireköpinge parish*

PDI cannot be calculated for property units if the distance between them is zero. Because of this constraint, the neighboring property units to each property unit in the study area would either have to be excluded from the calculation or, alternatively, a default distance value would have to be decided for all neighboring property units. Both suggested solutions would produce a distorted representation of reality. More in particular, in the first case the populations of the neighbouring to a property unit property units are not to be included in the calculations of its PDI, and in the second case the distances between the neighbouring to a property unit property units are given a default value which will ultimately affect the PDI value of the property unit in question. If the default distance value between adjacent geographic units is too small (e.g. less than a meter), then the populations of all adjacent to the geographic unit -whose PDI is being calculated- geographic units will significantly determine the PDI result. These cases, depending on the total amount of population assigned to an adjacent geographic unit, are prone to produce extremely high PDI values. This is because the fraction which divides the population of an geographic unit by the distance between that geographic unit and the geographic unit whose PDI is being computed, has a remarkably smaller denominator compared to the denominators of other geographic units that are not adjacent to that geographic unit.

When Euclidean distances from a property unit's centroid to another property unit's border are being computed, a distance could be equal to zero if the centroid would be located outside of the property unit boundaries. The case where a geographical unit's centroid is located outside of its boundaries is denoted as the "horseshoe problem". An example of this issue in the data set is shown in Figure 4.10:

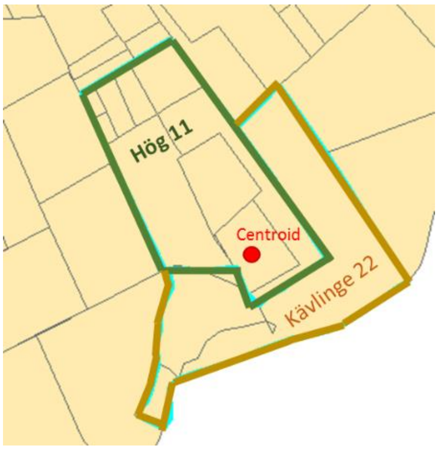


Figure 4.10: *Initial Property Unit "Kävlinge 22" has its centroid located within another property unit "Hög 11"*

In the figure above the centroid of initial property unit "Kävlinge 22" is located outside of its boundaries and inside the boundaries of initial property unit "Hög 11". Because, SQL Server's STDistance() function estimates the Euclidean distance between a point (Kävlinge 22- centroid) contained by a polygon (Hög 11 initial property unit boundaries) as zero, the distance between the two aforementioned initial property units according to the "Centroid-to-Border" distance calculation method is zero.

It is important to know the extent of the "horseshoe problem" in the dataset, before deciding upon a suitable distance calculation method. By executing the SQL-query in Figure 4.11 (which selects the parish, id, area and geometry of all property units whose centroid is located outside the property unit area), it is possible to find the exact number of property units for which this problem occurs in the dataset:

```
SELECT  Parish, uniqueIdObj AS PropertyUnit_ID,
        Shape.STArea() AS Property_Unit_Area,
        Shape AS PropertyUnit_Geometry

FROM    PropertyUnits_All


WHERE   Centroid.STIntersects(Shape)=0
```

Figure 4.11: *SQL-query returning the property units whose centroid is located outside of their borders*

There are four in total property units in the study area that exhibit this problem. They are located in Halmstad or Kävlinge parish and their respective area ranges from 12,000 m^2 to 350,000 m^2 approximately. By executing similar SQL-queries for the remaining geographic levels, it is revealed that two initial property units in Kävlinge, and four address-units in Kävlinge and Halmstad parishes are affected by the "horseshoe problem". The areal extent of these initial property units or address-units is similar to the areal extent of the aforementioned property units dealing with the same issue. None of the parish units have this problem. Subsequently, it is relatively safe to say that the extent of the "horseshoe problem" is limited and therefore does not have a significant effect upon the overall distance calculation results.

Another instance where the distance between two different geographical units may be equal to zero according to this calculation method, is when a geographical unit and its centroid are contained by another geographical unit (Figure 4.12). The reason why the distance is equal to zero is the same as the previously mentioned one.

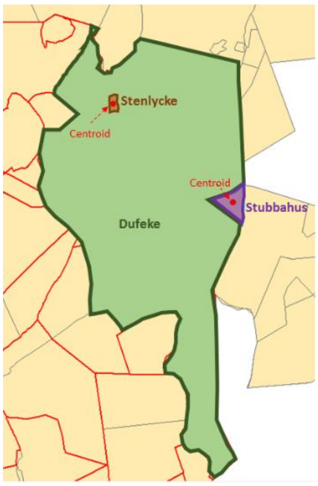


Figure 4.12: *Initial Property Unit "Stenlycke" has its centroid located within another initial property unit "Dufeke" in Halmstad parish*

The issues spotted in the "Centroid-To-Border" shortest Euclidean distance calculation method would also apply for the "Border-to-Centroid" distance calculation method.

In the context of this data set, no "division by zero" errors occurred whilst the "Centroid-to-Centroid" distance calculation method was implemented. In theory, such an error could potentially occur, if a situation as shown in Figure 4.13 would occur.

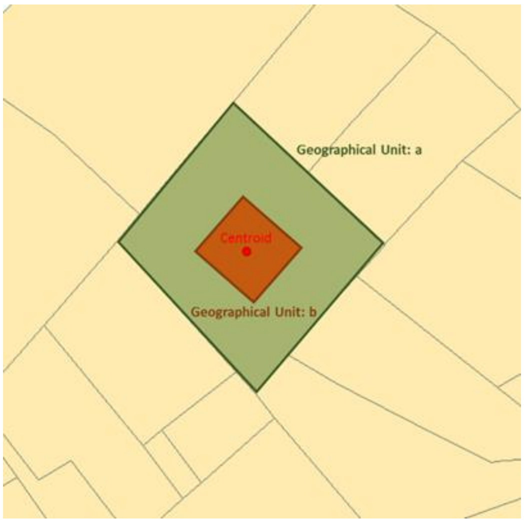


Figure 4.13 *Geographical Units sharing the same centroid*

In this case, two geographical units share the same centroid. This can happen if one geographical unit is contained by another geographical unit. The distance between the centroid of geographical unit "**a**" and the centroid of geographical unit "**b**" would be equal to zero.

Because of the previously stated reasons, for the purpose of this study, the "Centroid-To-Centroid" shortest Euclidean distance calculation method (Figure 4.8) was deemed most suitable to implement for the PDI computation.

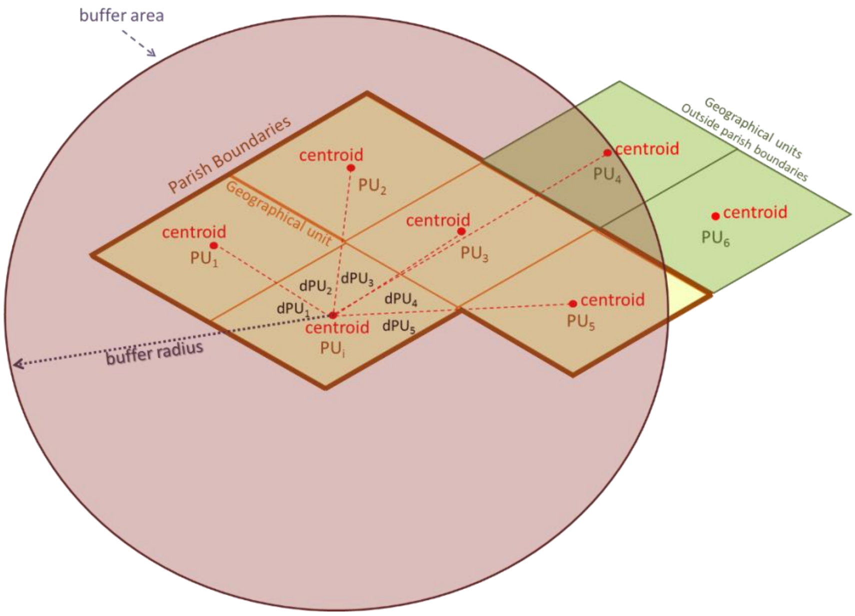


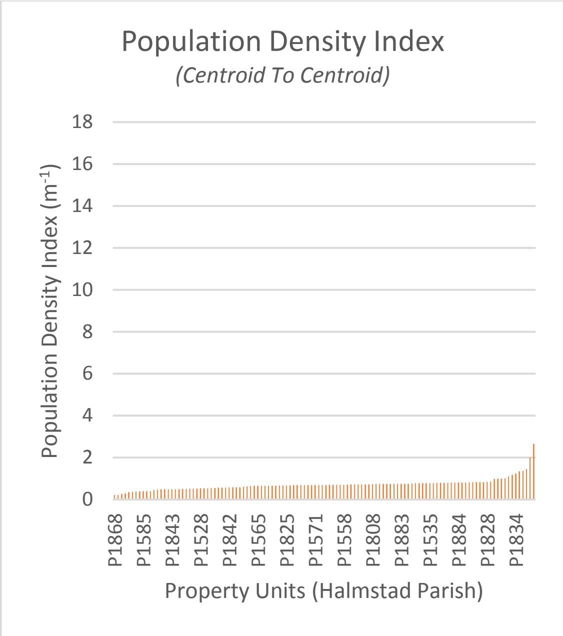
Figure 4.14: *Example of calculating the PDI of property unit PU_i*

Figure 4.14 illustrates how the population density index is calculated for property unit PU_i. PDI is separately calculated for every property unit in the study area. PU_i's PDI value is affected by the populations of the property units whose centroid is located within the circular buffer area the center of which is PU_i's centroid. In the figure below, this means that the property units PU_1, PU_2, PU_3, PU_4 and PU_5 are included in the calculation of PU_i property unit's PDI. Property unit PU_6 is excluded from the procedure, as its centroid is not located within the buffer area. All the property units belonging to the same parish are included in the calculations. In the results presented here, the buffer area is assigned with a radius of 4,000 m. This radius was chosen to ensure that all geographical units within a parish should be included in the estimation of the PDI value for each independent geographical unit within the parish. The choice of a radius also potentially permits for other geographical units -outside the parishes of the study area (e.g. PU_4)- to be included in the PDI calculation of the geographical units.

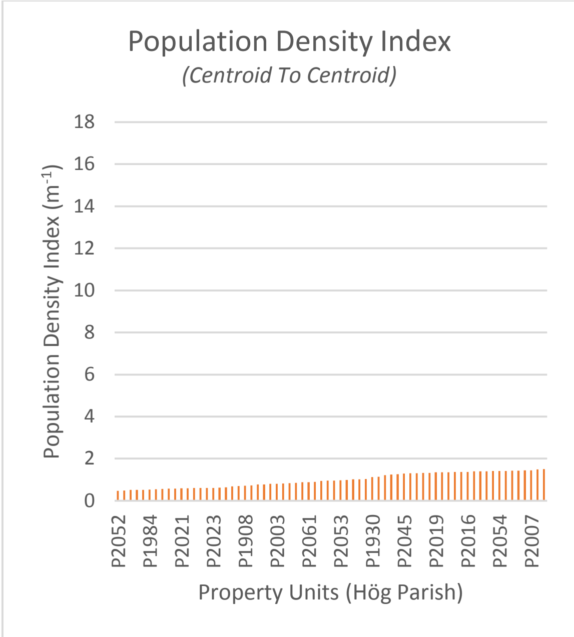
However, for the purpose of this study, as the linking of population information to geographical units is still a working progress, the PDI will only be calculated for the year of 1910. No results on parish level are computed as it would not make sense to do so for this geographic context variable. The reason why this is the case, is that it would have been required to have a number of adjacent parishes to create a parish level, suitable for implementing this particular PDI formula. Because only Kävlinge is adjacent to Hög and Halmstad to Sireköpinge there are not enough data to permit a valid execution of the formula. The results would be too biased.

4.4.1.3 Results

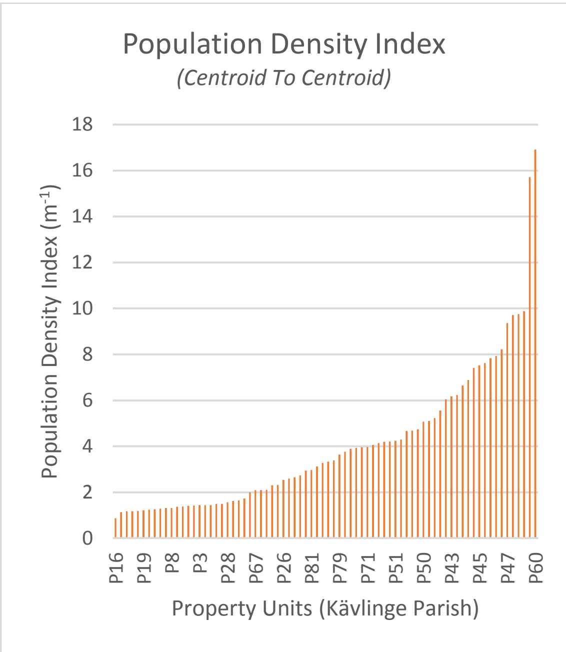
Diagrams in Figure 4.15 illustrate the PDI-values produced for every property unit in the parishes of Halmstad, Hög, Kävlinge and Sireköpinge respectively. The results are presented in an ascending order.



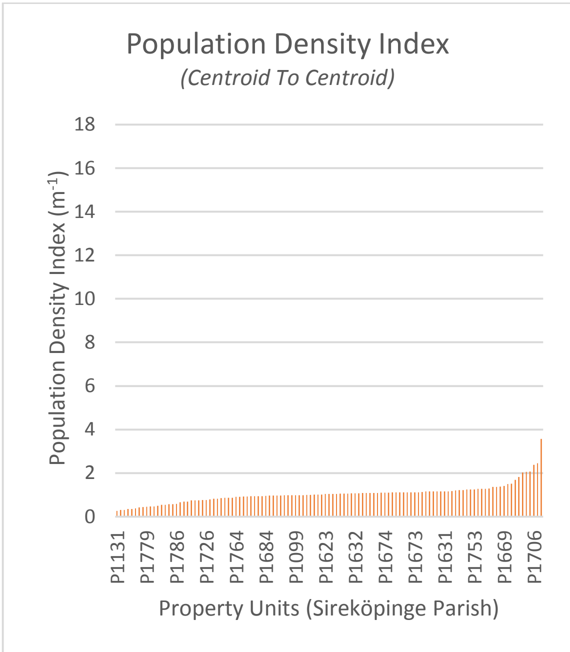
(i)



(ii)



(iii)



(iv)

Figure 4.15: *Diagrams depicting the Population Density Index of property-units on parish level. (i) PDI-values of property units in Halmstad parish. (ii) PDI-values of property units in Hög parish. (iii) PDI-values of property units in Kävlinge parish. (iv) PDI-values of property units in Sireköpinge parish.*

The diagrams in Figure 4.16 present the PDI values for initial property units located in the parishes of Halmstad, Hög, Kävlinge and Sireköpinge correspondingly.

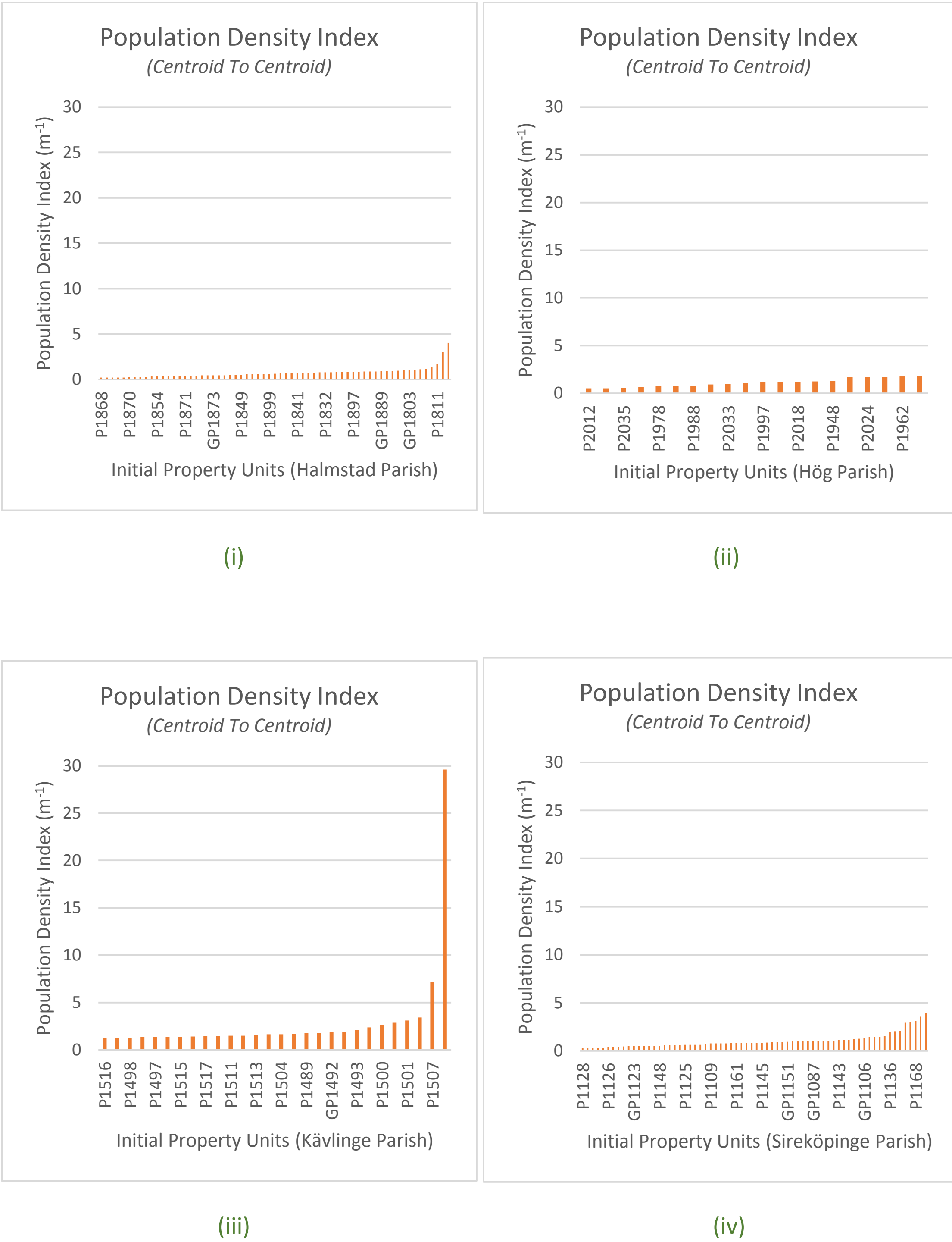
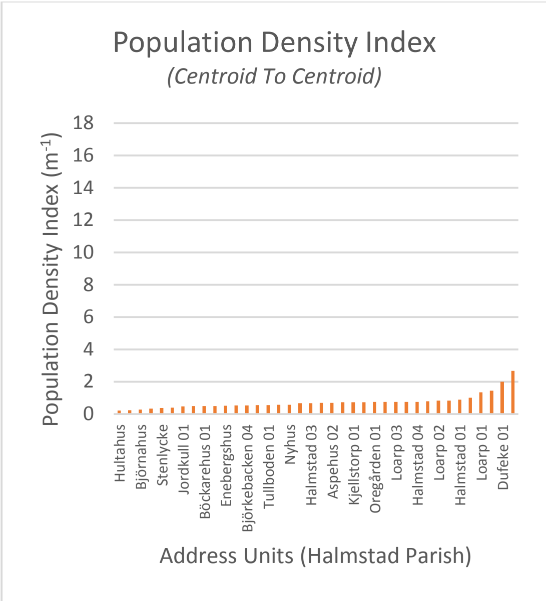
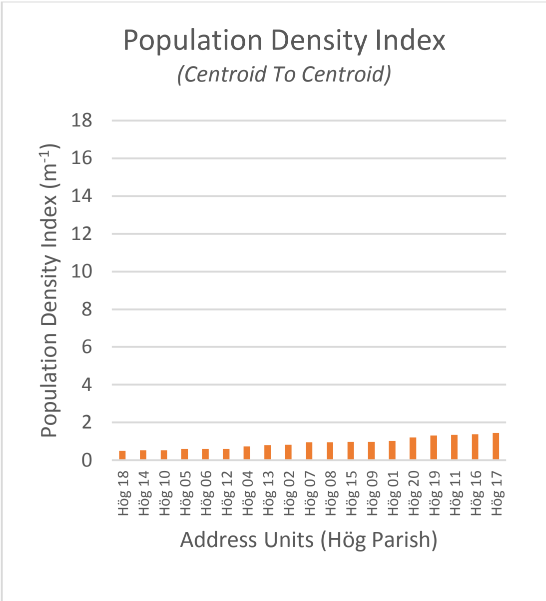


Figure 4.16: *Diagrams depicting the Population Density Index of initial-property-units on parish level. (i) PDI-values of initial property units in Halmstad parish. (ii) PDI-values of initial property units in Hög parish. (iii) PDI-values of initial property units in Kävlinge parish. (iv) PDI-values of initial property units in Sireköpinge parish.*

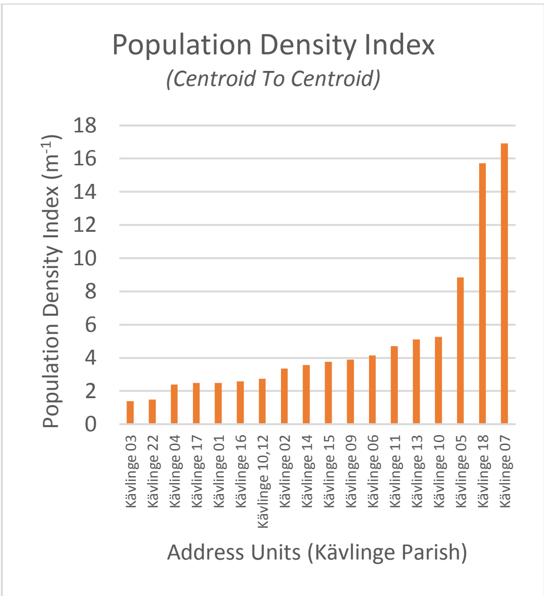
The population density index of address units of Halmstad, Hög, Kävlinge and Sireköpinge is presented in the corresponding diagrams of Figure 4.17.

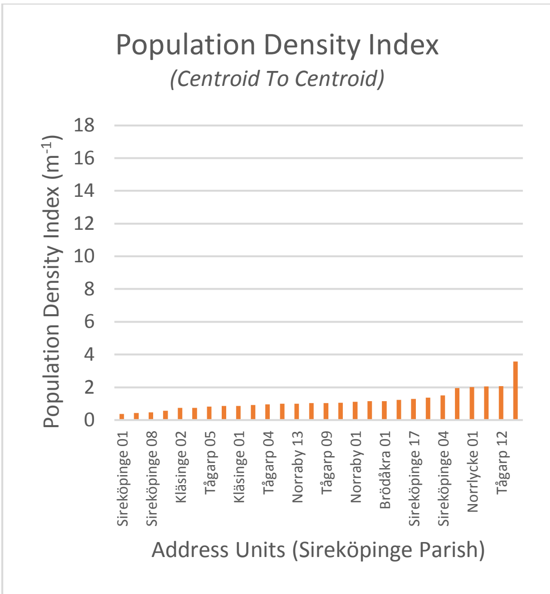




(i)

(ii)





(iii)

(iv)

Figure 4.17: *Diagrams showing the Population Density Index of address-units on parish level. (i) PDI-values of address units in Halmstad parish. (ii) PDI-values of address units in Hög parish. (iii) PDI-values of address units in Kävlinge parish. (iv) PDI-values of address units in Sireköpinge parish.*

4.4.1.4 Discussion

From the knowledge obtained from the historical sources mentioned in the study area introduction regarding the total population of every parish and the general occupation patterns of the inhabitants of each parish in the study area, it seems that the PDI formula succeeds in reproducing a faithful representation of reality. Kävlige which was documented as the parish with the largest population in the historical records, is shown in diagram (iii) Figure 4.14, diagram (iii) Figure 4.15 and diagram (iii) Figure 4.16 to have produced the highest overall PDI-values per geographic unit. The PDI results verify that Kävlinge was the most highly populated and densely populated of parishes in the study area. It is important to note that Kävlinge is the parish that displays the highest dispersion in PDI-values, indicating that the population is not evenly distributed over the parish area. This can partially be explained by the fact that Kävlinge since the early 1870s was affected by industrialization (e.g. establishment of factories and railway communications) and as a result attracted more people to live close to their working place. The effect of industrialization in this case was counteracting the effects of the land reforms which urged people to move out from the village. However, the parts of Kävlinge that were not involved in any industrial activity were affected by the land reforms. This might very well be the underlying reason behind the large dispersion of PDI-values in Kävlinge parish.

Hög is the parish whose population seems to be more evenly distributed over the parish area, which is evident by the fact that the PDI-values do not seem to vary significantly over the parish area regardless the geographic level of resolution. This could be attributed to the fact that the parish of Hög includes only one village and was amongst the first parishes in this area to have implemented the land reforms in 1804.

Halmstad and Sireköpinge exhibit somewhat different patterns than Kävlinge or Hög. A possible explanation could be that they did not have any significant industrial activity, the land reforms were implemented later in time compared to Hög, and in contrast to Hög they included more than one village within their parish boundaries.

The PDI formula is affected by distance between geographic units and the amount of population linked to every geographic unit. In general, high PDI values are produced for geographic units that have themselves large population or are located in close proximity to other geographic units with relatively high population. The distance between geographic unit centroids is affected by size (the larger the units the longer the distance between their centroids) and shape. Narrow or non-rectangular geographic units might have their centroids located in closer proximity to other geographic unit centroids than wider rectangular or circular geographic units, whose centroid's are more likely to be located closer to the center of their geometries and, therefore, leaving an equal amount of distance to any of their borders. Size and shape (to an extent) are affected by the geographic level. Overall, property units tend to be smaller, narrower and more rectangular than addresses or initial property units. But because they are smaller they have less population linked to them, so even though the distance between geographic units gets smaller so does the total amount of population. Initial property units are usually bigger than property units and accommodate more population. The uniformity-homogeneity of the size and/or geographic units belonging to the same geographic level affects the PDI value. This is important to consider while trying to interpret the PDI values of a single geographic level as well as when comparing PDI results between different geographic levels. Do all geographic units of the same geographic level have similar size or similar shapes? If there is a lack of homogeneity as to the size of the units of one level then that could partly explain the differences in PDI, when the population factor has been accounted for.

4.4.2 Proximity to Water

Waterbodies in our study area consist of streams and lakes (Figure 4.18). The largest lakes are found in the northern parts of Halmstad parish and some smaller ones can be found in the western parts of the same parish as well as in the southern parts of Sireköpinge. Though several streams flow through Halmstad and Sireköpinge, this is not the case for Hög and Kävlinge whose closest waterbody is a river which constitutes the southern natural border of their parishes. No lakes exist near Hög or Kävlinge. Overall, Halmstad could be considered as the parish containing the largest number of waterbodies. The neighboring parish of Sireköpinge also has a quite dense stream network.

4.4.2.1 Definition

Proximity to water is defined as the shortest Euclidean distance between a geographic unit and a lake or stream in the study area. Generally, streams and lakes change over time. These changes can be periodic and repeat themselves annually (e.g. lakes and streams often contain less water during the warmer months of the year due to evaporation and/or limited rainfall, and more water during the rainy periods of the year or just after the snow has melt). Other changes can occur as a result of the ever going bedrock erosion under a lake or stream. The water erodes the bedrock, and, as time elapses, this could ultimately cause changes along the course of a stream, and, potentially, even alter the extent of a lake. Water bodies might also be affected by volcanic activity, earthquakes, climate change or human intervention (e.g. draining of lakes, altering the course of a river). However, as the historic data source only provides information regarding the waterbodies for a limited time period (1812-1820), it is assumed, for the purpose of this study, that the shape and size of water bodies remains constant throughout the study period. Consequently, the proximity to water is considered a static variable.

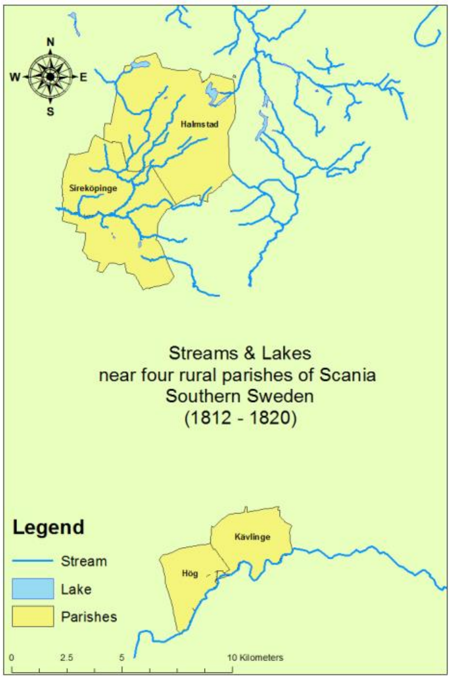


Figure 4.18: *Map over the streams & lakes near the parishes of Halmstad, Sireköpinge, Hög and Kävlinge ©Lantmäteriet[I2014/00579]*

4.4.2.2 Implementation

Every geographic unit's distance to water (e.g. lakes or streams) is calculated as the shortest Euclidean distance (in meters) between the:

I. centroid of a geographic unit and
   a. a point on the border of a lake (Figure 4.8), or
   b. a point on a stream (Figure 4.5).

II. border of a geographic unit and
   a. a point on the border of a lake (Figure 4.6), or
   b. a point on a stream (Figure 4.4).

*Method **I*** (Centroid-to-Border) calculates the shortest Euclidean distance between the centroid of a geographic unit and a point on the border of a lake. It also performs a similar procedure to determine the shortest distance between the centroid of the same geographic unit and a point on the nearest stream line. In the end, the method selects the shortest distance of the two to be the geographic unit's minimum distance to water.

*Method **II*** (Border-to-Border) computes the shortest distance between the border of a geographic unit and a point on the border of a lake. It repeats a similar procedure to estimate the shortest distance between the borders of the same geographic unit and a point on the closest stream line. Then it compares the two values and selects the lowest one to represent that geographic unit's minimum distance to water.

This process is applied to all geographic levels but one. The address level differs from the other geographic units because it is the only one that consists of more than one, often not adjacent, property units (Figure 3.4). More in particular, as an address unit consists of one or more property units, the calculation of the shortest distance between the whole address unit and water cannot be completed in one step.

If *method I* was implemented for an address unit as an entity, then the centroid of that entity might be located outside of the address unit producing a false shortest distance estimation to water (Figure 4.19). Figure 4.19 illustrates how the centroid for the property units belonging to address unit 'Halmstad 04' is located outside the address unit boundaries. The value of the shortest distance d_c between the address unit centroid and the stream is much lower than the corresponding shortest distance values between the respective property units and the stream ($d_{c_1}, d_{c_2}, d_{c_3} > d_c$). Therefore, the value of d_c is not representative for this particular address unit. Problems such as the aforementioned occur in cases where an address unit consists of more than one property units which are not adjacent.

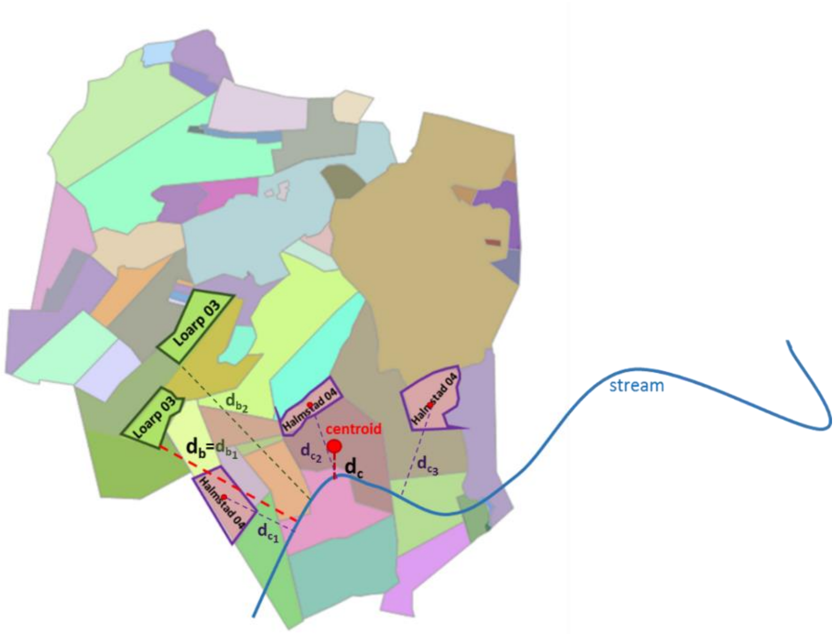


Figure 4.19: *Examples of minimum distance to water calculations for Address units Loarp 03 and Halmstad 04*

Similarly, if *method II* was used to calculate the shortest distance to water from the borders of an address unit that consists of several property units which are not adjacent to each other, then the address unit would be assigned with a shortest distance that would represent the shortest distance from the borders of the property unit that was located in the closest proximity to a stream or lake. An illustrating example of this case is included in Figure 4.19 regarding the address unit "Loarp 03". According to *method II*, the shortest distance d_b to water for this address unit is equal to the shortest distance to water d_{b_1} from the borders of the property unit that is in closest proximity to water. However, the difference between the shortest distances between the property units of "Loarp 03" d_{b_1}, d_{b_2} and water is not negligible. The problem becomes even greater if the property units of an address unit are more widely spread, then that estimation would not be representative for the rest of the property units included in the address unit. To overcome the abovementioned issues, the shortest distance of an address unit to water is computed as the average of the shortest distance to water of all property units sharing the same address.

4.4.2.3 Result

Figure 4.20 illustrates the minimum distance to water results by property unit, computed by the two aforementioned methods, in the form of diagrams. The results computed according to the centroid-to-border distance calculation method are shown in green color and are sorted in an ascending order. The results of the border-to-border distance calculation method are depicted in blue and are sorted to match the same property unit order as the one followed by the centroid-to-border method. The choice of presenting the results grouped by parish and not cumulatively was made because the spatial distribution of waterbodies differs from parish to parish, and in this way it is possible to examine the behavior of the calculation methods whilst implemented in environments with different characteristics.

Figure 4.20: *Diagrams depicting the Minimum Distance to Water from property units on a parish level. (i) Proximity to Water results for property units in Halmstad parish. (ii) Proximity to Water results for property units in Hög parish. (iii) Proximity to Water results for property units in Kävlinge parish. (iv) Proximity to Water results for property units in Sireköpinge parish.*

The difference between the distance calculation results of the two aforementioned methods, can be estimated with the help of the Root Mean Square Deviation, which is calculated according to the following formula:

$$\boldsymbol{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,1} - x_{i,2})^2}{n}} \qquad \textbf{(2)}$$

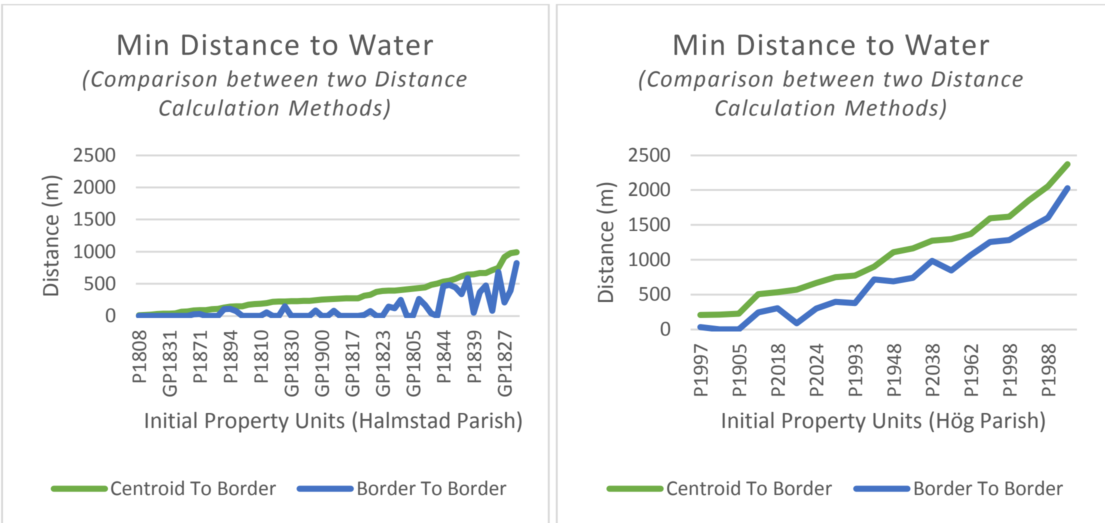
With $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every property unit i, and n the total number of property units in the study area, the

RMSD is equal to 229 meters approximately. The RMSD results for every parish separately and for all parishes can be found in table 4.1:

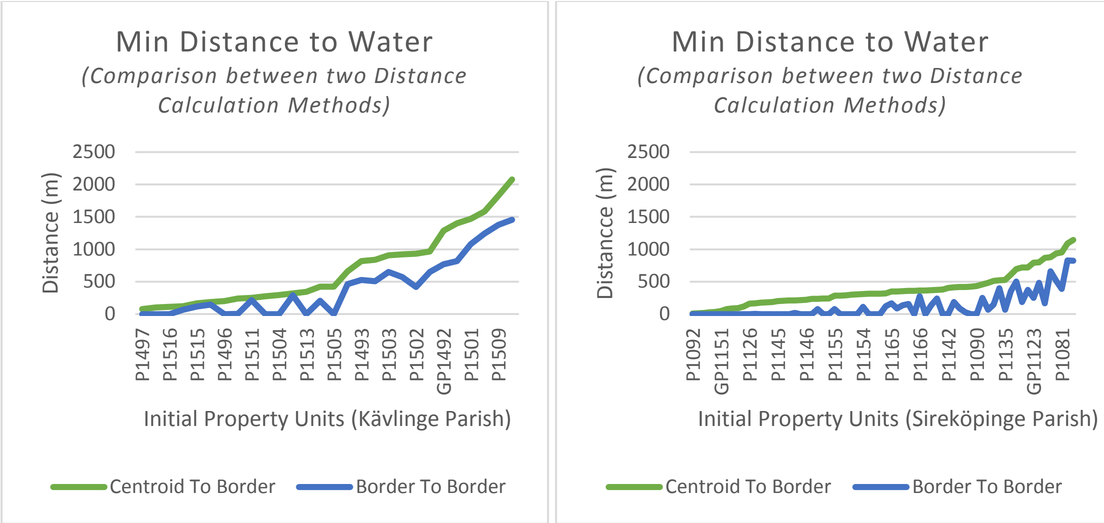
Parish	RMSD (m) *Property Unit Level*
Halmstad	249.7
Hög	214.4
Kävlinge	219.5
Sireköpinge	228.1
All Parishes	**229**

Table 4.1: *"Root Mean Square Differences between the 'Centroid-to-Border' and the 'Border-to-Border' distance calculation methods on Property Unit Level by Parish"*

Figure 4.21 illustrates the minimum distance to water results by initial property unit, computed by the two aforementioned methods, in the form of diagrams.



(i) (ii)



(iii) (iv)

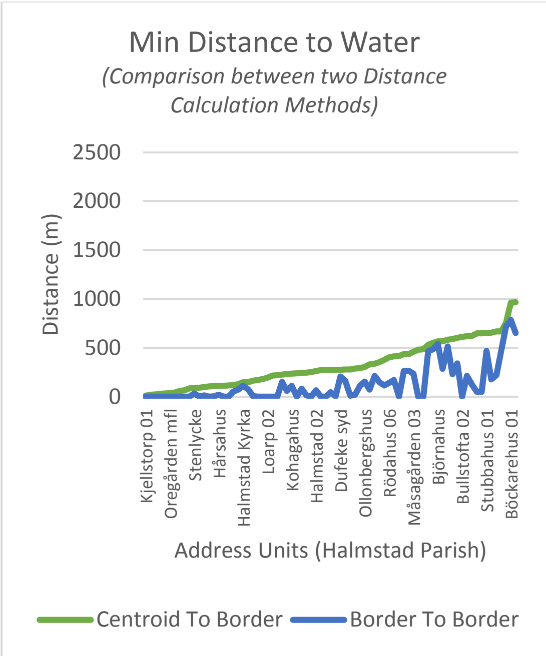
Figure 4.21: *Diagrams depicting the Minimum Distance to Water from initial property units on a parish level. (i) Proximity to Water results for initial property units in Halmstad parish. (ii) Proximity to Water results for initial property units in Hög parish. (iii) Proximity to Water results for initial property units in Kävlinge parish. (iv) Proximity to Water results for initial property units in Sireköpinge parish.*

With $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every initial property unit *i*, and *n* the total number of initial property units in the study area (all parishes), the RMSD from formula 2 is equal to 292.1 meters approximately. The RMSD results for every parish separately and for all parishes can be found in table 4.2:

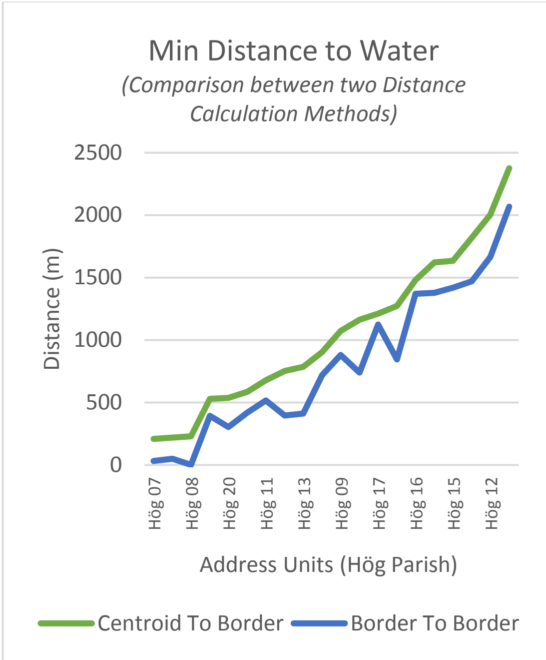
Parish	RMSD (m) *Initial Property Unit Level*
Halmstad	264.1
Hög	344.7
Kävlinge	321.2
Sireköpinge	287.2
All Parishes	**292.1**

Table 4.2: *"Root Mean Square Differences between the 'Centroid-to-Border' and the 'Border-to-Border' distance calculation methods on Initial Property Unit Level by Parish"*

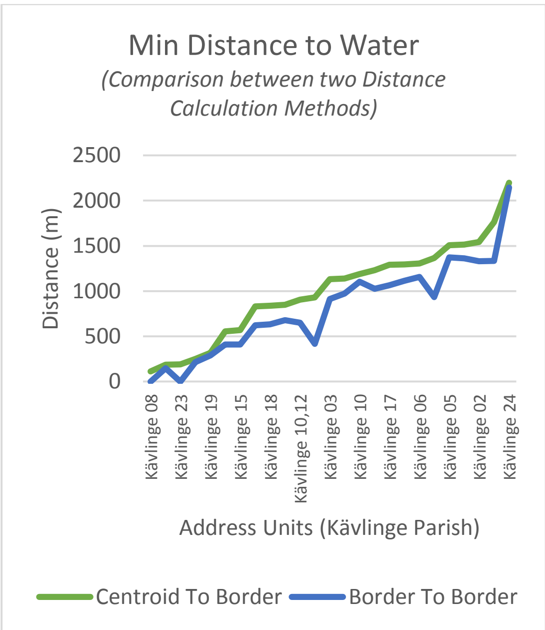
Figure 4.22 illustrates the minimum distance to water results by address unit, computed by the two aforementioned methods, in the form of diagrams.



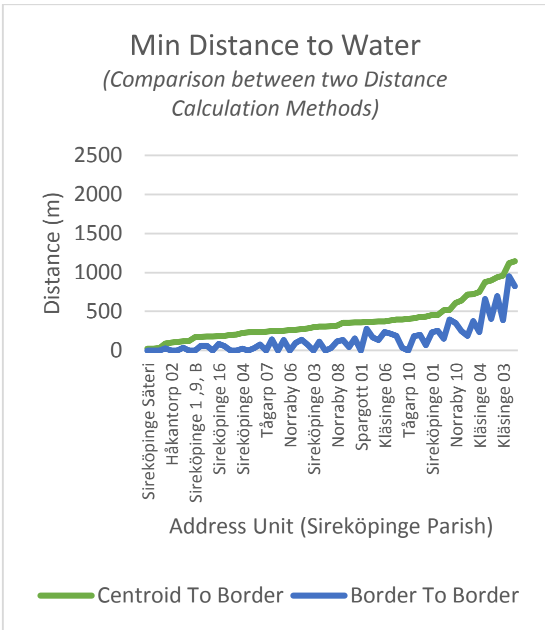
(i)



(ii)



(iii)



(iv)

Figure 4.22: *Diagrams depicting the Minimum Distance to Water from address units on a parish level. (i) Proximity to Water results for address units in Halmstad parish. (ii) Proximity to Water results for address units in Hög parish. (iii) Proximity to Water results for address units in Kävlinge parish. (iv) Proximity to Water results for address units in Sireköpinge parish.*

With $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every address unit *i* in each parish, and *n* the total number of address units in the study area (all parishes), the RMSD from formula 2 is equal to 246.5 m approximately. The RMSD results for every parish separately and for all parishes can be found in table 4.4:

Parish	RMSD (m) *Address Unit Level*
Halmstad	244.6
Hög	264.2
Kävlinge	221.5
Sireköpinge	252.3
All Parishes	**246.5**

Table 4.3: *"Root Mean Square Differences between the 'Centroid-to-Border' and the 'Border-to-Border' distance calculation methods on Address Unit Level by Parish"*

Figure 4.23 shows a diagram of the minimum distance to water results by parish unit, computed by the two implemented calculation methods.

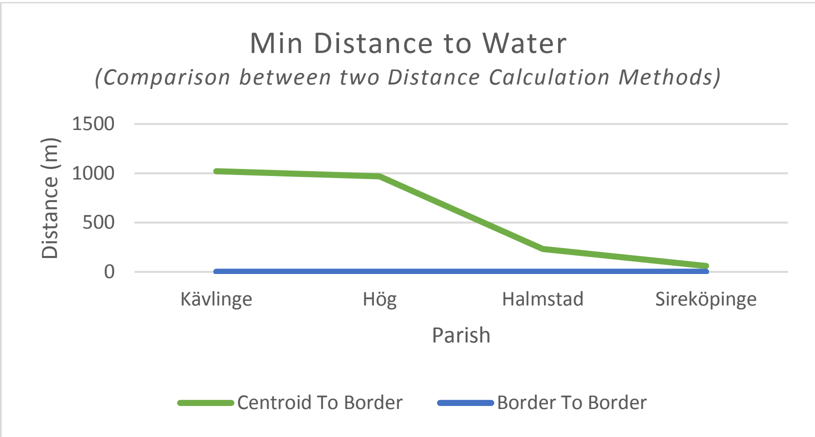


Figure 4.23: *Minimum Distance from Parish-Unit Polygon Centroid or Border to Lake-Polygon Boundary or Stream-Polyline Point*

With $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every parish unit *i*, and *n* the total number of parish units in the study area, the RMSD from formula 2 is equal to 714,2 m approximately. The RMSD results for every parish separately and for all parishes can be found in table 4.5:

Parish	RMSD (m) *Parish Unit Level*
Halmstad	232.4
Hög	969.5
Kävlinge	1021.3
Sireköpinge	59.3
All Parishes	**714.2**

Table 4.4: *"Root Mean Square Differences between the 'Centroid-to-Border' and the 'Border-to-Border' distance calculation methods on Parish Unit Level by Parish"*

4.4.2.4 Discussion

The centroid-to-border distance calculation method yields higher distance results for all parishes in all levels of geographic resolution. The border-to-border distance calculation method is more prone to produce zero-distance results. Particularly for the parishes of Halmstad and Sireköpinge, where the stream network is denser and smaller or bigger lakes exist, the line representing the border-to-border method in diagrams (i) and (iv) in figures 4.20, 4.21 and 4.22 respectively shows a lot of fluctuations towards zero. This implies that the method is sensitive to the spatial distribution of waterbodies, in a way that the denser the distribution of rivers and lakes in the study area the higher the probability of producing zero-distance results. In theory, this is also true for the centroid-to-border method, but as the distance in that case is calculated between two points and not between a polygon and a point, the risk of producing zero-distance results is significantly lower. Diagrams (ii) and (iii) in figures 4.20, 4.21 and 4.22 respectively, illustrate how zero-distance results are not that likely to occur in areas which are not defined by a dense distribution of waterbodies.

Big differences between the results of the two methods whilst implemented for the same geographic unit can be attributed to the shape and/or size of the geographic unit. For instance, a very narrow but long property unit might have some part of its borders in a very close proximity to water, though its centroid could be located quite further away.

From tables 4.1, 4.2, 4.3 and 4.4 it can be derived that the bigger the geographic units get, the bigger the difference between the results produced by the two distance calculation methods. There are two facts that attribute to this result. Firstly, the fact that if any segment of a stream or part of a lake, however small, overlaps with a geographic unit, then the proximity to water for that geographic unit, according to the border-to-border distance calculation method, will be equal to zero. Secondly, the fact that the bigger the size of a geographic unit the greater the likelihood of a stream or lake overlapping it. These facts in conjunction with the results depicted in Figure 4.23, where the border-to-border distance calculation method yields zero-distance results for all parish units, suggest that the aforementioned method should not be used with geographic levels of coarse resolution.

Consider the case of Hög, where the only close waterbody is a river in the southern borders of the parish. The river flows along the parish borders and crosses it at some parts, consequently the distance of all the parish to water is set to zero. Properties in the northern part of the parish can be located as far as 4km away from the river. Thus, if the water was contaminated the population of these properties were not that likely to be affected in contrast to those living near the river. Now, if this contamination affected the mortality rates of Hög parish and a study was conducted over proximity to water and mortality, implementing the border-to-border distance calculation method on a parish level, the association between close proximity to contaminated water and mortality may not have been spotted. A possible explanation for this could be that the assigned value representing the parish’s shortest distance to water would be equal to zero and, because mortality is computed over the entire parish as a whole, the overall mortality rate might only have changed a little or not at all.

4.4.3 Proximity to Wetlands

4.4.3.1 Definition

Proximity to wetlands is defined as the shortest Euclidean distance between a geographic unit and a wetland in the study area. In general, the total area covered by wetlands may change with time. These changes can be result of natural processes, climate change or human intervention. Capturing these changes is essential in providing good quality of wetland data. Though scientists, who perform contemporary research, are well equipped with tools to extract this kind of information on a regular basis with the help of for example satellites, **LiDAR** (NOAA 2015) or other sources, this is not the case for historic research. Historic research on wetlands is dependent upon the existence of well preserved, good quality historical maps. There is information stating that significant efforts to dry out wetland area were carried out during the second half of the 19th century, to produce more arable land. At the time this study was conducted no such data was available in digital format. This is why for the purpose of this study the geographic context variable of wetlands is considered as a static variable.

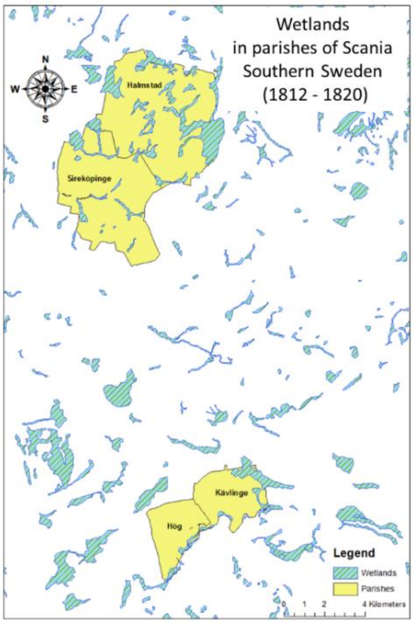


Figure 4.24: *Map over the wetlands near the parishes of Halmstad, Sireköpinge, Hög and Kävlinge ©Lantmäteriet[I2014/00579]*

4.4.3.2 Implementation

Every geographic unit's distance to wetland is calculated as the shortest Euclidean distance (in meters) between:

i. the centroid (point) of a polygon (e.g. property unit or parish) and a point in the border of a polygon (i.e. wetland).

ii. the border of a polygon (e.g. property unit or parish) and a point in the border of a polygon (i.e. wetland).

The distances were decided to be calculated only to the borders of the wetland areas as there was no point of including distance calculations to their centroid. This is due to the fact that the study is more interested in the minimum distance from locations which are more likely to have served as natural habitats for malaria infected mosquitos. Therefore, estimating the distance from wetland borders is deemed as a more suitable choice.

The implementation of two different methods of distance estimation was carried out in order to stress the importance of critically examining the methods used to produce a distance value rather than just concentrating on the magnitude of the produced value. Distance is defined as the distance between a starting point and an end point. The choice of distance calculation method as well as what is considered to be a starting point and what an end point affects the value of the result. In the implementation of any distance calculation method, it is important to understand what is considered to be the concept of distance. The following figures provide proof as to why this is necessary.

Figure 4.25 shows how the same numerical result from the two calculation methods represent two totally different situations in reality. In the first case, whilst calculating the distance from the geographical unit's centroid to the wetland border, the wetland area is actually located within the geographical unit (e.g. property unit) boundaries, where people are more likely to be present. On the contrary, in the second case, whilst calculating the distance from the geographical unit's border to the wetland border, the wetland area is found outside the geographical unit boundaries. In theory, exposure to wetland conditions is higher in the first case than the second, but this fact is not evident from the comparison of the numerical results provided by the two different calculation methods.

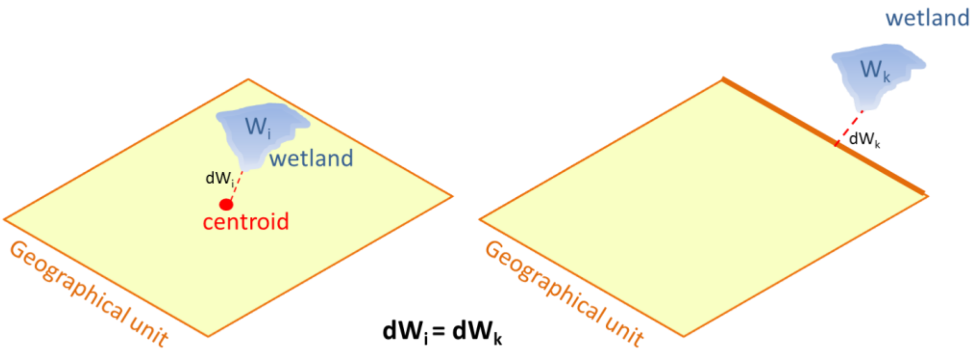


Figure 4.25: *Same distance results between a property unit and a wetland, calculated by two different methods*

Another instance where misleading results might be produced, is shown in Figure 4.26. This figure shows how differently both methods treat the case where a wetland area is located within the geographical unit boundaries. According to the "centroid-to-border" method, the distance to wetland area is only equal to zero if the geographical unit's centroid is inside wetland area. This is fine as long as the centroid is located within the geographical unit's boundaries. A problematic situation would occur if the centroid happened to be located within wetland area but outside the boundaries of the geographical unit (horseshoe-problem). On the contrary, according to the "border-to-border" calculation method, the distance between a geographical unit and a wetland area is equal to zero as long as both object intersect. The second case is more prone to provide zero-distance results than the first one.

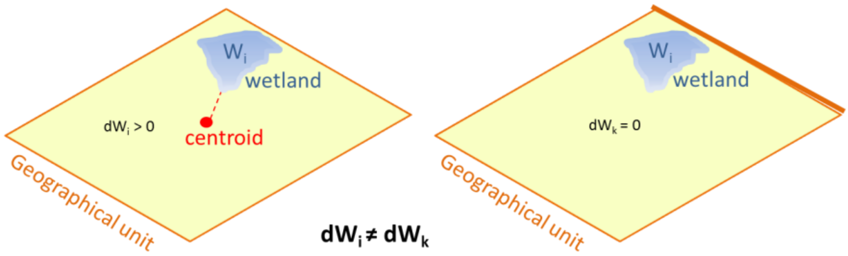


Figure 4.26: *Wetland within geographical unit produces different distance results whilst calculated by two different methods*

Figure 4.27 shows three different cases where the "border-to-border" calculation method would yield a zero-distance result. In the first case, only a small part of the wetland area is included within the geographical unit boundaries. The second example shows a situation where wetland area and geographical unit area barely abut, while in the third example the geographical unit contains the whole wetland area. In terms of exposure to wetland conditions, these three examples show three very different situations despite the fact they share the same value regarding their distance from wetland area.

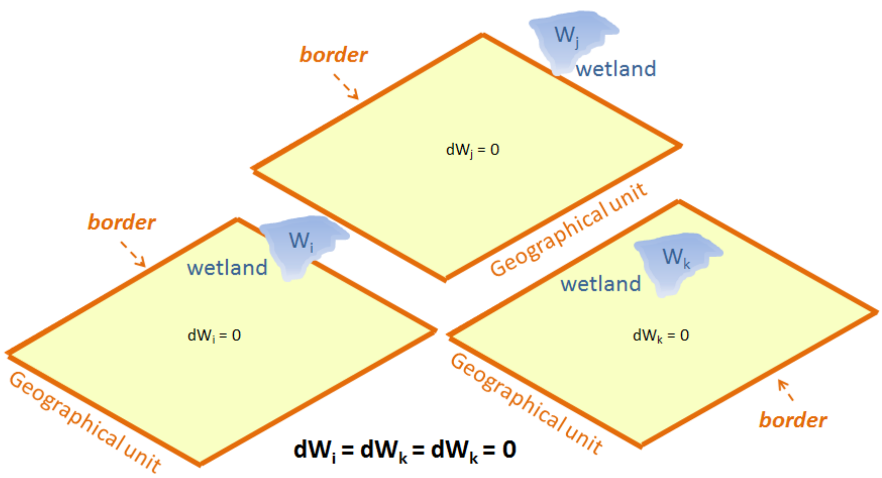


Figure 4.27: *Cases where the distance to wetlands is equal to zero with the border-distance calculation method*

On the contrary, Figure 4.28 illustrates an example of two geographical units that both contain wetland area but have very different distance-from-wetland values produced by the same calculation method (centroid-to-border). Here the exposure to wetland conditions is very similar for both cases, but the distance values are different.

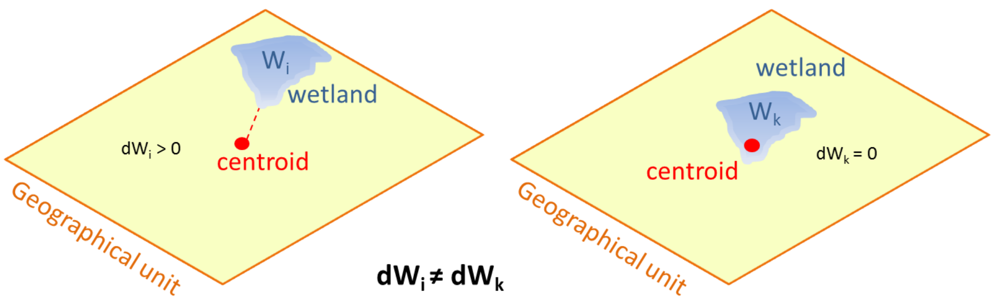


Figure 4.28: *"Wetland within geographical unit produces different distance results whilst calculated between the geographical unit polygon centroid and the wetland polygon border"*

When examining conditions of exposure to wetland conditions, just relying on distance numerical values and distance calculations methods is not enough. The shape and size of wetlands and geographical units is also important. Figure 4.29 shows how two similar geographical units share the same distance (calculated with the "centroid-to-border" method) to wetland borders but the exposure to wetland conditions is higher in the case where the wetland area is larger. Again this is not detectable from just the distance value. The same situation would have occurred if the distance was calculated with the "border-to-border" method.

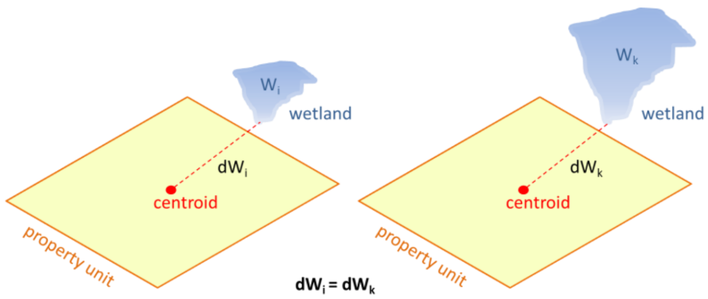


Figure 4.29: *The distance between a property unit and the nearest wetland is the same but the area of the wetlands differs*

Additionally, exposure to wetland conditions is affected by the total amount of wetland area near a certain location. As depicted in Figure 4.30, the comparison of exposure to wetlands between two geographical units cannot only be estimated based on which geographical unit has the shortest distance to wetland area. If a geographical unit is within a closer proximity to a small wetland area, then its level of exposure to those conditions might be lower than the case where a geographical unit is surrounded by scattered parts of larger wetland areas in slighter longer proximity.

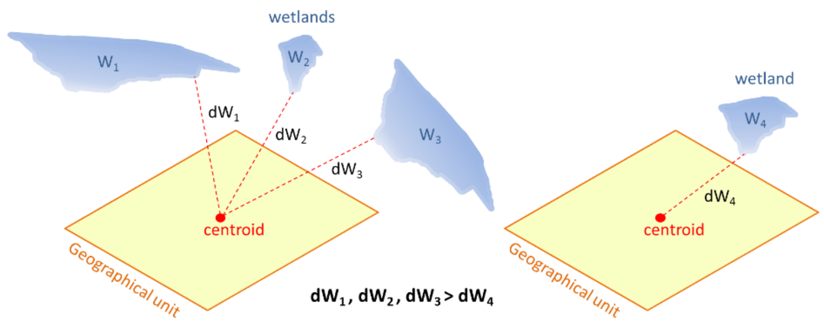
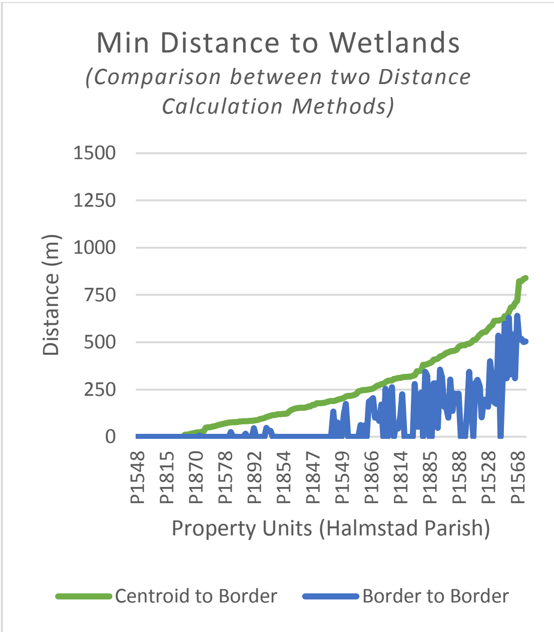


Figure 4.30: *The distance between a property unit and the nearest wetland in (b) is smaller than in (a) but in the latter the total exposure to wetlands is higher*

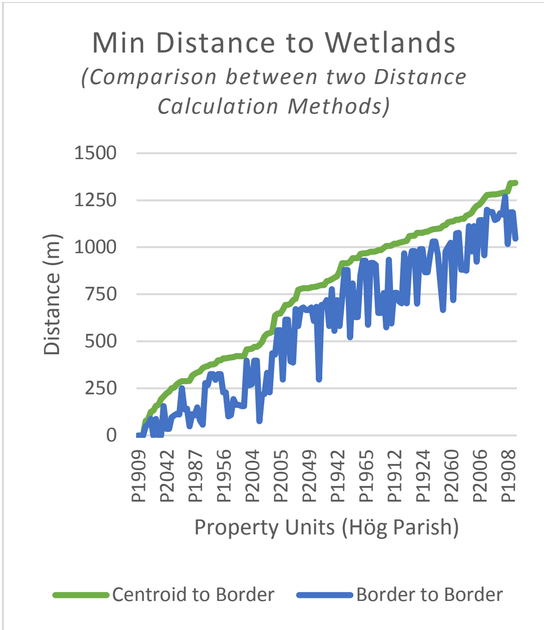
All of the above instances show why it is very important to provide detailed information about the exact details of how distance calculations are being performed between spatial objects, when attempting to link geographic and demographic statistical results. If this issue is not treated carefully, misleading conclusions might be drawn, providing a scewed or false view of reality.

4.4.3.3 Result

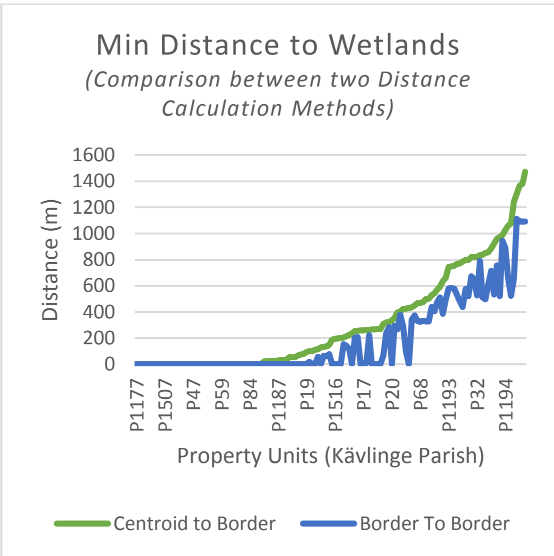
Figure 4.31 illustrates the minimum distance to wetlands results by property unit, computed by the two aforementioned methods, in the form of diagrams. The results computed according to the centroid-to-border distance calculation method are shown in green color and are sorted in an ascending order. The results of the border-to-border distance calculation method are depicted in blue and are sorted to match the same property unit order as the one followed by the centroid-to-border method.



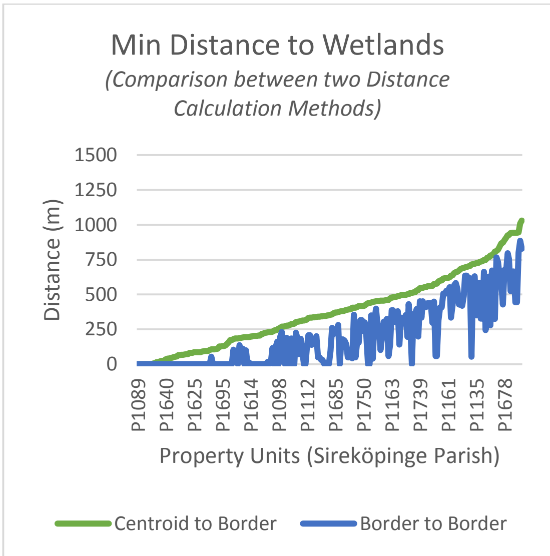
(i)



(ii)



(iii)



(iv)

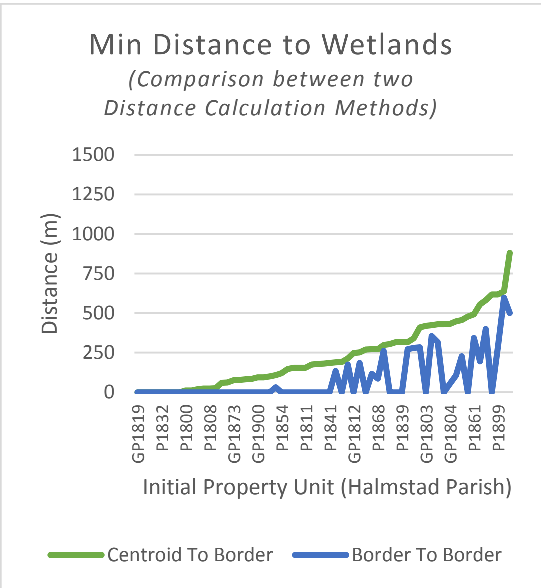
Figure 4.31: *Diagrams depicting the Minimum Distance to Wetlands from property units on a parish level. (i) Proximity to Wetlands results for property units in Halmstad parish. (ii) Proximity to Wetlands results for property units in Hög parish. (iii) Proximity to Wetlands results for property units in Kävlinge parish. (iv) Proximity to Wetlands results for property units in Sireköpinge parish.*

Table 4.5 depicts the RMSD between the results of the two distance computation methods whilst implemented for property units and wetlands. The results are shown on a parish level as well as for the entire study area. With $x_{i,1}$ representing the centroid-to-boarder distance calculation method and $x_{i,2}$ the border-to-border calculation method for every property unit *i* in each parish, and *n* the total number of property units in the study area (*n=159*), the RMSD from formula 2 is equal to 202,1 m approximately.

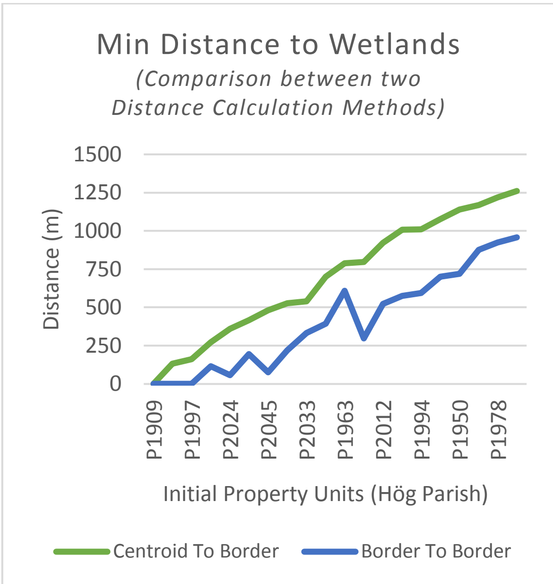
Parish	RMSD (m) *Property Unit Level*
Halmstad	204.6
Hög	206.0
Kävlinge	172.6
Sireköpinge	215.5
All Parishes	**202.1**

Table 4.5: *"Root Mean Square Differences between the two distance-calculation methods on Property Unit Level by Parish"*

Figure 4.32 illustrates the minimum distance to wetlands results by initial property unit, computed by the two aforementioned methods, in the form of diagrams.



(i)



(ii)

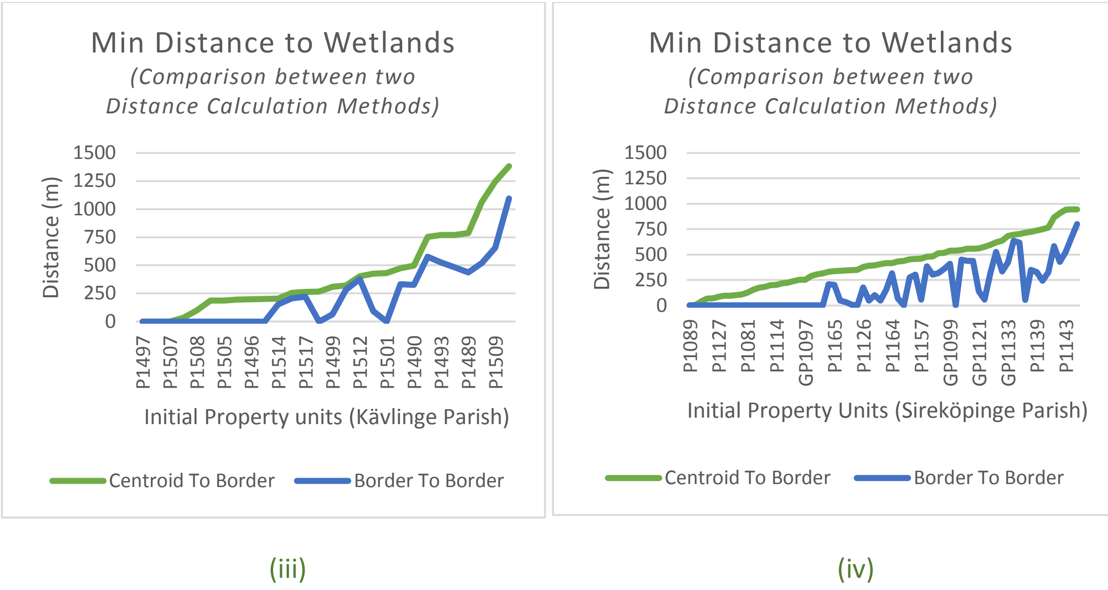


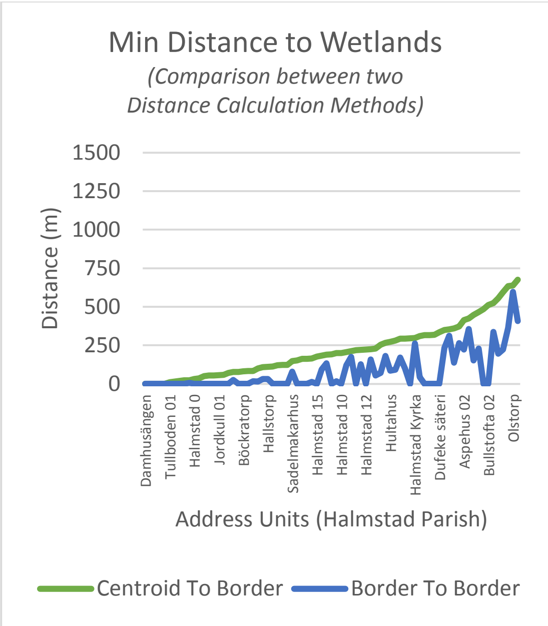
Figure 4.32: *Diagrams depicting the Minimum Distance to Wetlands from initial property units on a parish level. (i) Proximity to Wetlands results for initial property units in Halmstad parish. (ii) Proximity to Wetlands results for initial property units in Hög parish. (iii) Proximity to Wetlands results for initial property units in Kävlinge parish. (iv) Proximity to Wetlands results for initial property units in Sireköpinge parish.*

The Root Mean Square Deviation from formula 2, with $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every initial property unit *i* in each parish, and *n* the total number of initial property units in the parish area, produces the results shown in table 4.6:

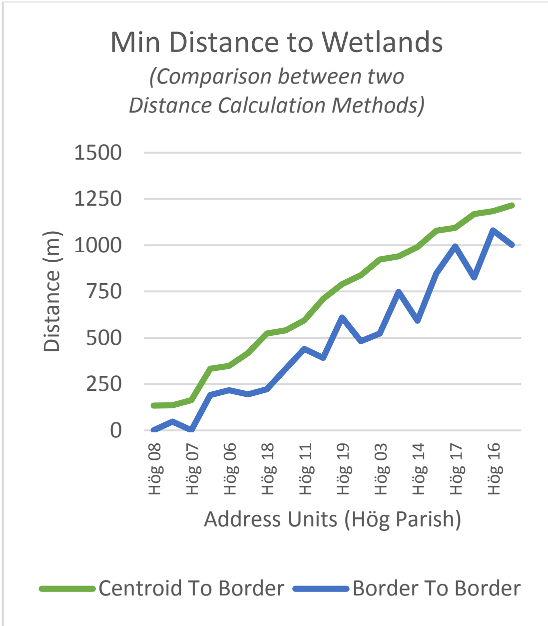
Parish	RMSD (m) *Initial Property Unit Level*
Halmstad	203.3
Hög	315.0
Kävlinge	246.6
Sireköpinge	275.9
All Parishes	**253.1**

Table 4.6: *"Root Mean Square Differences between the two distance-calculation methods on Initial Property Unit Level by Parish"*

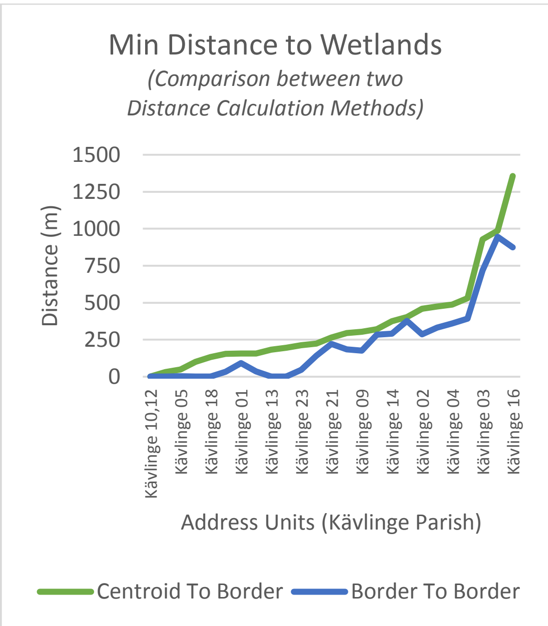
Figure 4.33 presents the minimum distance to wetlands results by address unit, computed by the two aforementioned methods, in the form of diagrams.



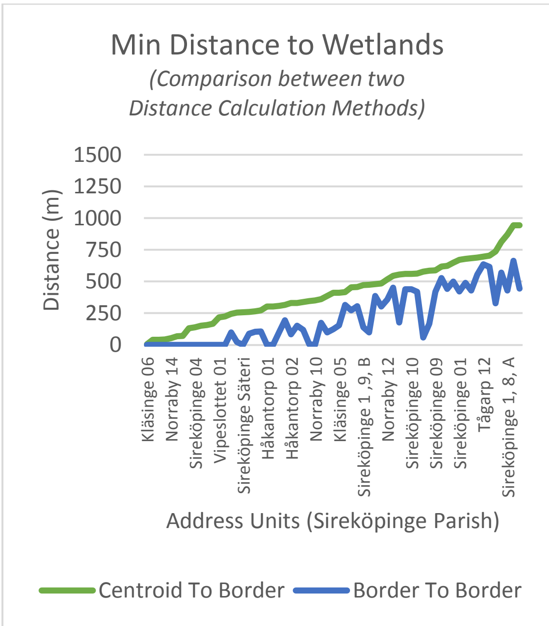
(i)



(ii)



(iii)



(iv)

Figure 4.33: *Diagrams depicting the Minimum Distance to Wetlands from address units on a parish level. (i) Proximity to Wetlands results for address units in Halmstad parish. (ii) Proximity to Wetlands results for address units in Hög parish. (iii) Proximity to Wetlands results for address units in Kävlinge parish. (iv) Proximity to Wetlands results for address units in Sireköpinge parish.*

The Root Mean Square Deviation from formula 2, with $x_{i,1}$ representing the centroid-to-border distance calculation method and $x_{i,2}$ the border-to-border calculation method for every address unit *i* in each parish, and n the total number of address units in the study area (n=23), produces a result equal to 202.7 m (Table 4.7).

Parish	RMSD (m) *Address Unit Level*
Halmstad	179.4
Hög	239.5
Kävlinge	151.5
Sireköpinge	232.5
All Parishes	**202.7**

Table 4.7: *"Root Mean Square Differences between the two distance-calculation methods on Address Unit Level by Parish"*

Figure 4.34 shows the minimum distance to wetlands results by parish unit, computed by the two implemented methods, in the form of diagrams.

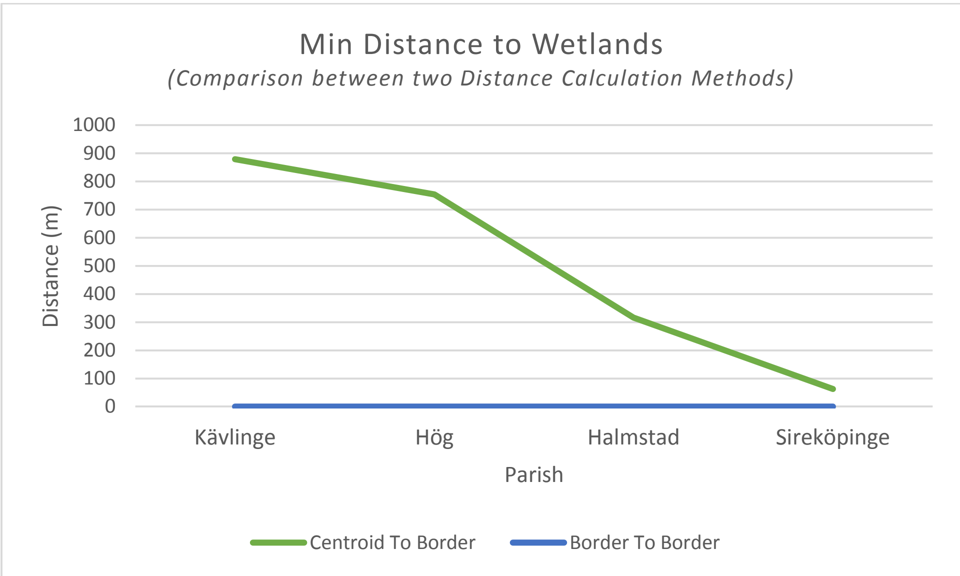


Figure 4.34: *Comparison of two Minimum Distance Calculation Methods from Parish Polygon to Wetland Polygon*

The Root Mean Square Deviation from formula 2, with $x_{i,1}$ representing the centroid-to-border distance calculation method, $x_{i,2}$ the border-to-border calculation method for every parish *i* in the study area and *n* the total number of parishes (*n*=4), produces the following result: 601.1 meters (table 4.8).

Parish	RMSD (m) *Parish Unit Level*
Halmstad	316.3
Hög	753.6
Kävlinge	879.6
Sireköpinge	62.2
All Parishes (RMSD)	**601.1**

Table 4.8: *"Root Mean Square Differences between the two distance-calculation methods on Parish Unit Level by Parish"*

4.4.3.4 Discussion

Similarly to the proximity to water results, the border-to-border distance calculation method continues to yield lower distance values than the centroid-to-border method. Once again it is obvious from the diagrams that the first named method favours zero-distance results. This is shown by the many fluctuations towards zero of the border-to-border method line in the diagrams. The coarser the geographic unit level, the lower the border-to-border distance values are produced and the greater the differences in distance values between the methods.

4.4.4 Proximity to Gathering Places

4.4.4.1 Definition

Proximity to gathering places is, in this case, defined as the shortest Euclidean distance between a geographic unit and the church. Every location where people tend to assemble can be considered as a gathering place. Other suggestions for possible gathering places for people belonging to this study period are markets, schools, hospitals, train stations, etc. The reason why determining gathering locations and estimating the distance from them is that it is important to estimate the level of exposure people have to other people. This might give an indication as to how infectious diseases spread and which populations where at higher risk of contracting a disease depending on how close they lived to locations or sites people used to assemble and thus were in higher risk of infecting each other with air borne diseases or any other kind of infectious diseases. In this study only the distance to church is examined. However, similar techniques can be followed to estimate proximity to any other gathering location.

4.4.4.2 Implementation

Every geographic unit's distance to a gathering place (e.g. church) is calculated as the shortest Euclidean distance (in meters) between the centroid (point) of a geographical unit (polygon) and a point representing the gathering place (Figure 4.2) or, alternatively, between the border of a geographical unit and the point of a gathering place (Figure 4.1).

4.4.4.3 Result

Figure 4.35 presents the minimum distance to church results by property unit, computed by the two aforementioned methods, in the form of diagrams. The results computed according to the centroid-to-border distance calculation method are shown in green color and are sorted in an ascending order. The results of the border-to-border distance calculation method are depicted in blue and are sorted to match the same property unit order as the one followed by the centroid-to-border method.

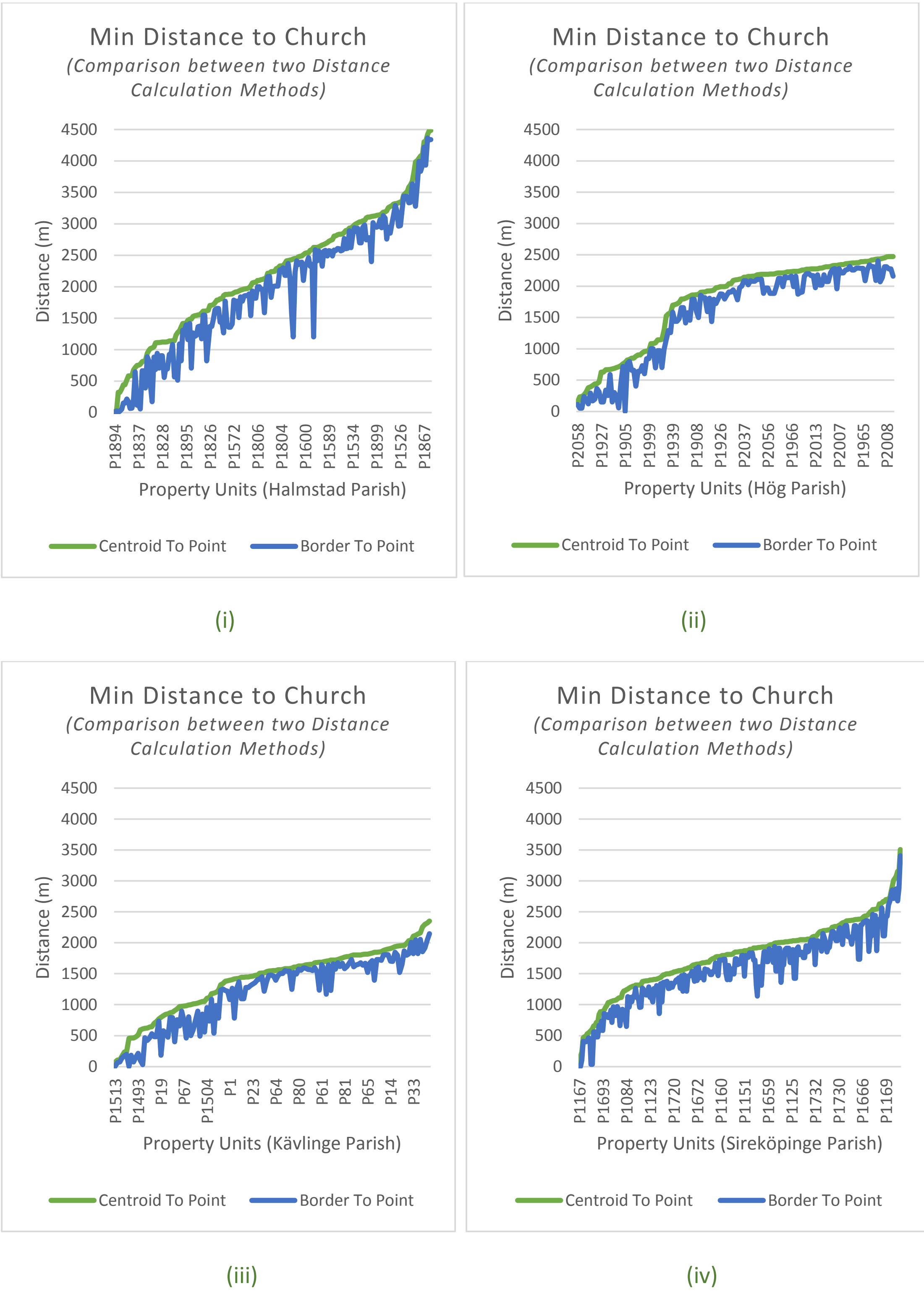


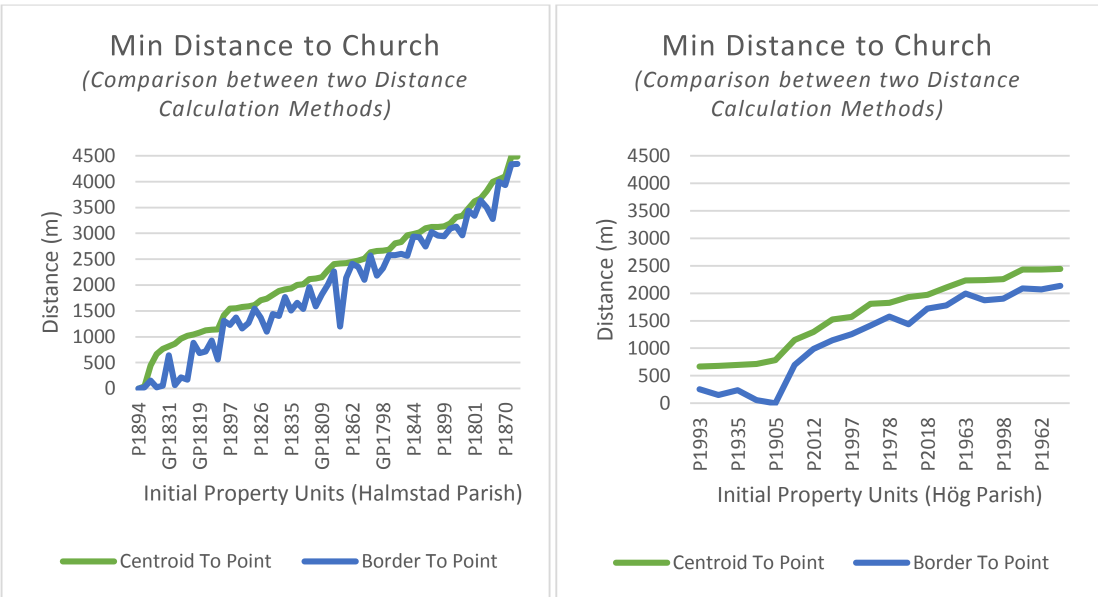
Figure 4.35: *Diagrams depicting the Minimum Distance to Church from property units on a parish level. (i) Proximity to Church results for property units in Halmstad parish. (ii) Proximity to Church results for property units in Hög parish. (iii) Proximity to Church results for property units in Kävlinge parish. (iv) Church to Wetlands results for property units in Sireköpinge parish.*

Table 4.9 contains the Root Mean Square Deviation from formula 2, where $x_{i,1}$ represents the centroid-to-border distance calculation method, $x_{i,2}$ the border-to-border calculation method for every property unit i in the study area and n the total number of property units.

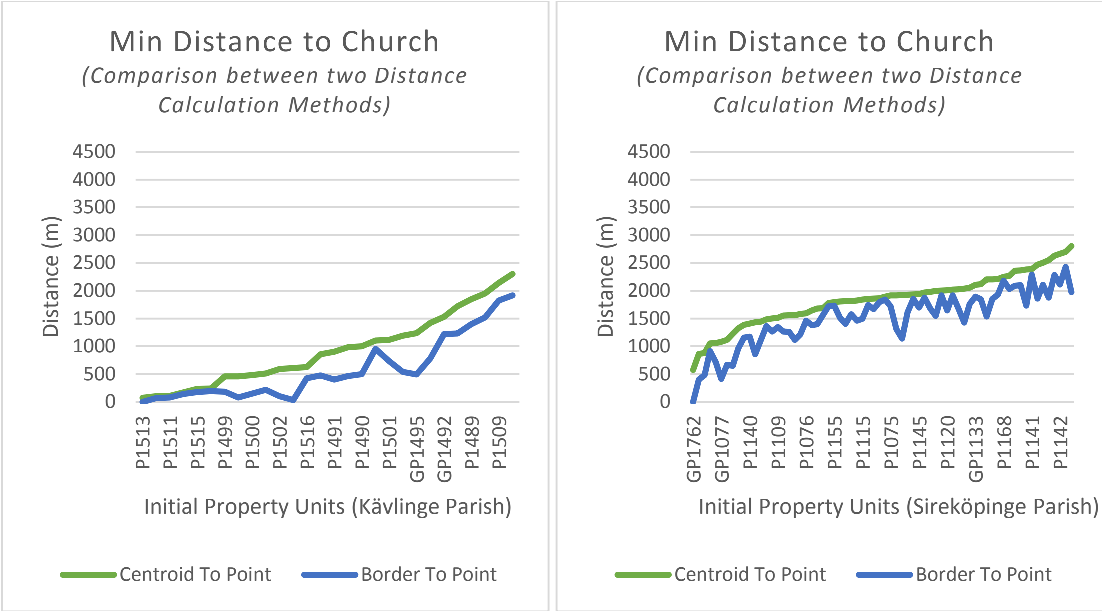
Parish	RMSD (m) *Property Unit Level*
Halmstad	342
Hög	240.3
Kävlinge	251.4
Sireköpinge	260.6
All Parishes	**278.8**

Table 4.9: *"Root Mean Square Differences between the two distance-calculation methods on Property Unit Level by Parish"*

Figure 4.36 presents the minimum distance to church results by initial property unit, computed by the two aforementioned methods, with the help of diagrams.



(i) (ii)



(iii) (iv)

Figure 4.36: *Diagrams depicting the Minimum Distance to Church from initial property units on a parish level. (i) Proximity to Church results for initial property units in Halmstad parish. (ii) Proximity to Church results for initial property units in Hög parish. (iii) Proximity to Church results for initial property units in Kävlinge parish. (iv) Church to Wetlands results for initial property units in Sireköpinge parish.*

Table 4.10 contains the Root Mean Square Deviation from formula 2, where $x_{i,1}$ represents the centroid-to-border distance calculation method, $x_{i,2}$ the border-to-border calculation method for every initial property unit i in the study area and n the total number of initial property units.

Parish	RMSD (m) *Initial Property Unit Level*
Halmstad	390.2
Hög	420.6
Kävlinge	400.4
Sireköpinge	376.6
All Parishes	**390.3**

Table 4.10: *"Root Mean Square Differences between the two distance-calculation methods on Initial Property Unit Level by Parish"*

Figure 4.37 presents the minimum distance to church results by address unit, computed by the two centroid-to-point and border-to-point methods, with the help of diagrams.

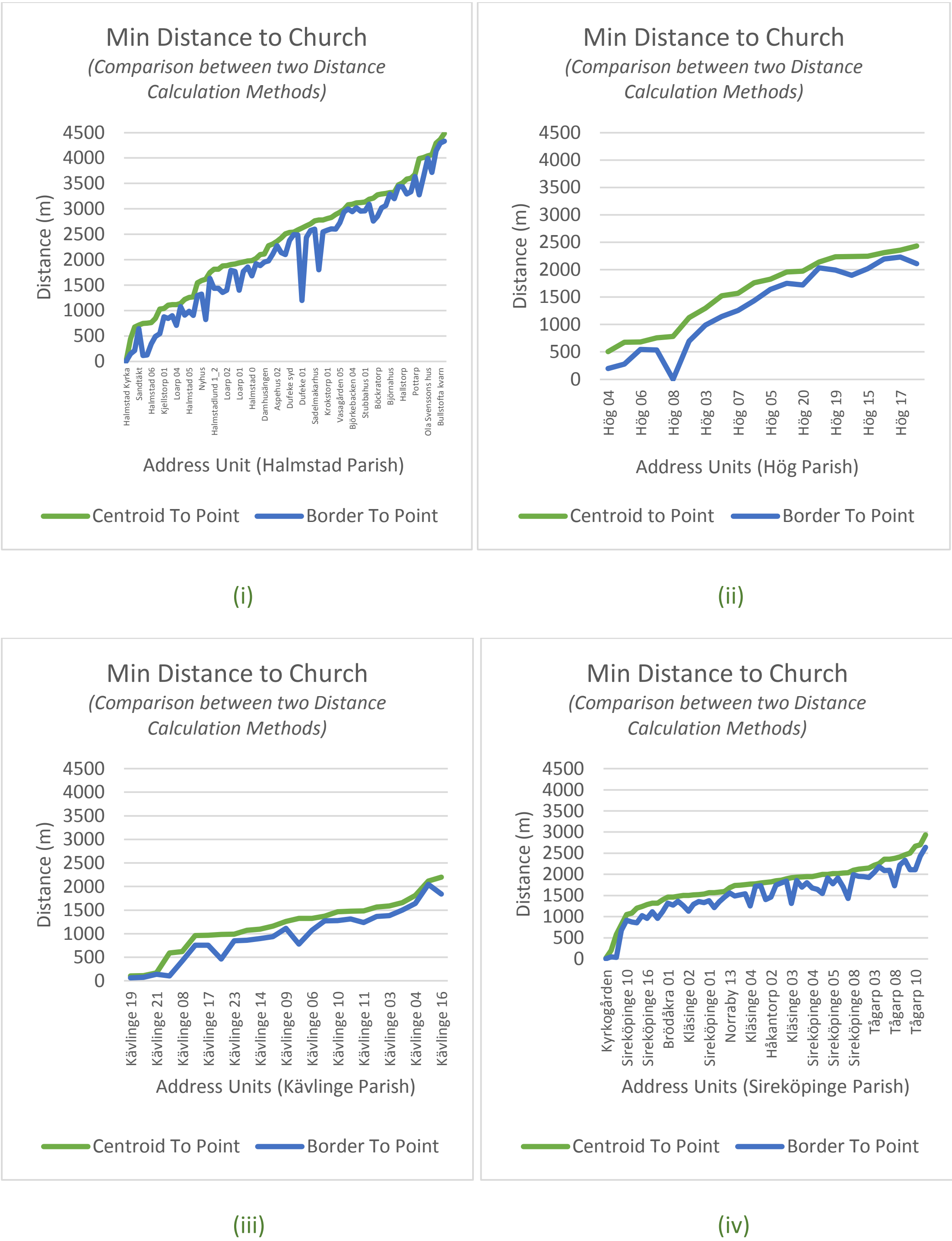


Figure 4.37: *Diagrams depicting the Minimum Distance to Church from address units on a parish level. (i) Proximity to Church results for address units in Halmstad parish. (ii) Proximity to Church results for address units in Hög parish. (iii) Proximity to Church results for address units in Kävlinge parish. (iv) Church to Wetlands results for address units in Sireköpinge parish.*

Table 4.11 presents the Root Mean Square Deviation from formula 2 for all parishes, where $x_{i,1}$ represents the centroid-to-border distance calculation method, $x_{i,2}$ the border-to-border calculation method for every address unit *i* in the study area and *n* the total number of address units.

Parish	RMSD (m) *Address Unit Level*
Halmstad	355.8
Hög	322.2
Kävlinge	250.6
Sireköpinge	284.6
All Parishes	**316.2**

Table 4.11: *"Root Mean Square Differences between the two distance-calculation methods on Address Unit Level by Parish"*

Figure 4.38 presents the minimum distance to church results by parish unit, computed by the two aforementioned methods, with the help of diagrams.

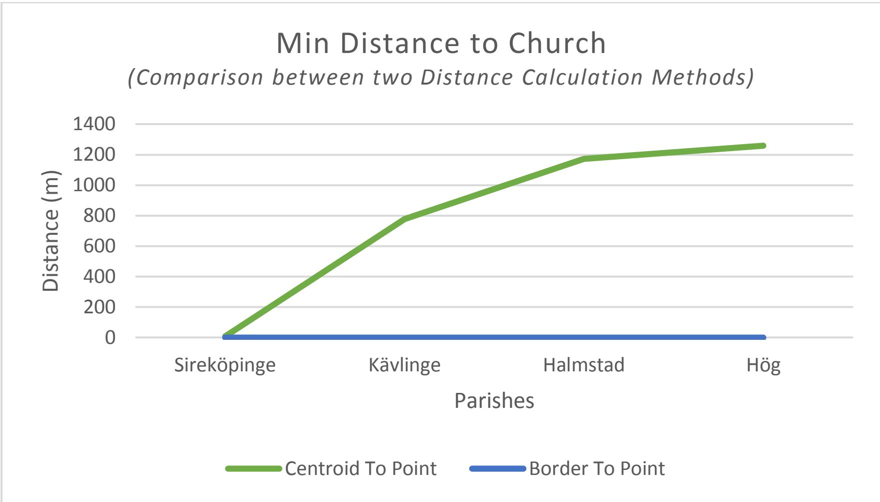


Figure 4.38: *Comparison of two Minimum Distance Calculation Methods from Parish-Unit Polygon to Church Point*

Finally, table 4.12 contains the Root Mean Square Deviation from formula 2, where $x_{i,1}$ represents the centroid-to-border distance calculation method, $x_{i,2}$ the border-to-border calculation method for every parish unit *i* in the study area and *n* the total number of parish units.

Parish	RMSD (m) *Parish Unit Level*
Halmstad	1173.6
Hög	1258.2
Kävlinge	778.2
Sireköpinge	6.5
All Parishes	**944.2**

Table 4.12: *"Root Mean Square Differences between the two distance-calculation methods on Parish Unit Level by Parish"*

4.4.4.4 Discussion

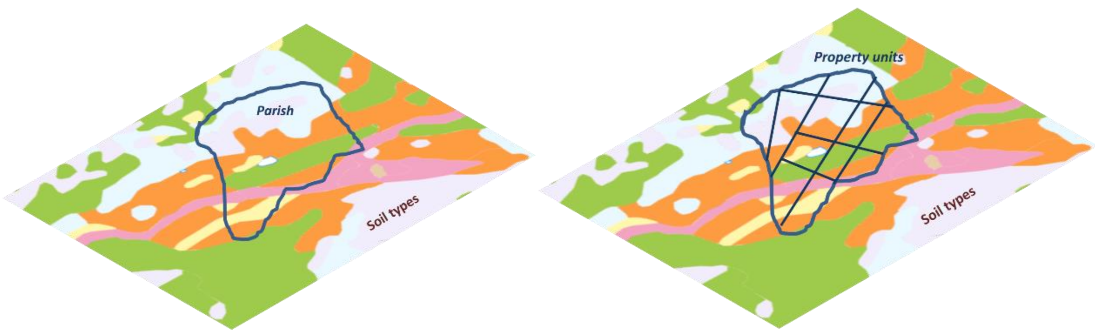
Whilst estimating the proximity to church for every geographical unit using the two pre-described calculation methods, it does yet once more become evident that the centroid-to-border method yields higher distance values than the border-to-border method. What is different this time is that though the difference between the two methods' results is considerable, the diagrams show less fluctuated curves. This can be attributed to the fact that the shortest distance to one point is measured. Since points don't cover as large an area as other geometries such as polygons and polylines and given the assumption that the point-layer is not dense (the church point-layer is not a particularly dense layer, as there is usually only one church in every parish), then the probability of a polygon including that point is small. Subsequently, the border-to-border distance calculation method is not likely to frequently produce zero-distance results. Even the difference in values between the two methods is greater in this case, it seems that the border-to-border diagram line follows the patterns of the centroid-to-border line more than in the case where proximity to polygon or line objects was computed. The differences between the two calculations methods in this instance are more affected by the size of the geographic units belonging to the same geographic level, rather than the shape.

4.4.5 Soil Types

4.4.5.1 Definition

Soil type information in the study is contemporary. The reason why contemporary soil type data was used in the study, is, firstly, because such detailed information was not available from historical sources and, secondly, because soil types covering certain areas don't undergo significant changes over short periods of time, such as the duration of the study period. Because of this, soil types can be described as a **static variable**.

Soil type information of each geographic unit is being stored in a table for each geographic level. More in particular, property units can be assigned to more than one soil types (Figure 4.39), as they are represented by polygons. The property-unit-table includes additional information regarding the area of each soil type (in m^2) as well as the percentage (%) of the total property unit area it covers. The same format is implemented for all geographic levels.



(a) (b)

Figure 4.39: *Overlay example between Soil types and Parish boundaries (a) or property units (b).*

4.4.5.2 Implementation

As a result of natural distribution of soil types or the overlay of soil types and geographic units, a geographical unit might include more than one soil type objects belonging to the same category (Figure 4.40). For example, in order to calculate correctly the percentage of soil type *i* covering the parish area in Figure 4.40, it is necessary to sum up the areas belonging to the soil type objects: i_1, i_2, i_3 and i_4. This problem is solved by performing a union function in SQL based on the soil type category for all geographic levels. The area of each soil type per geographical unit is then calculated as the sum of the unified areas belonging to the same soil type.

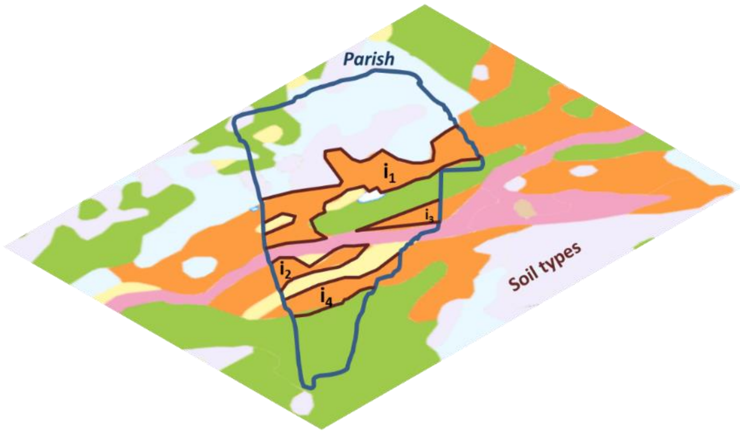


Figure 4.40: *Example of Soil Type Calculation Challenges.*

4.4.5.3 Results

Figures 4.41 – 4.44 demonstrate the frequency of the soil types in the four parishes.

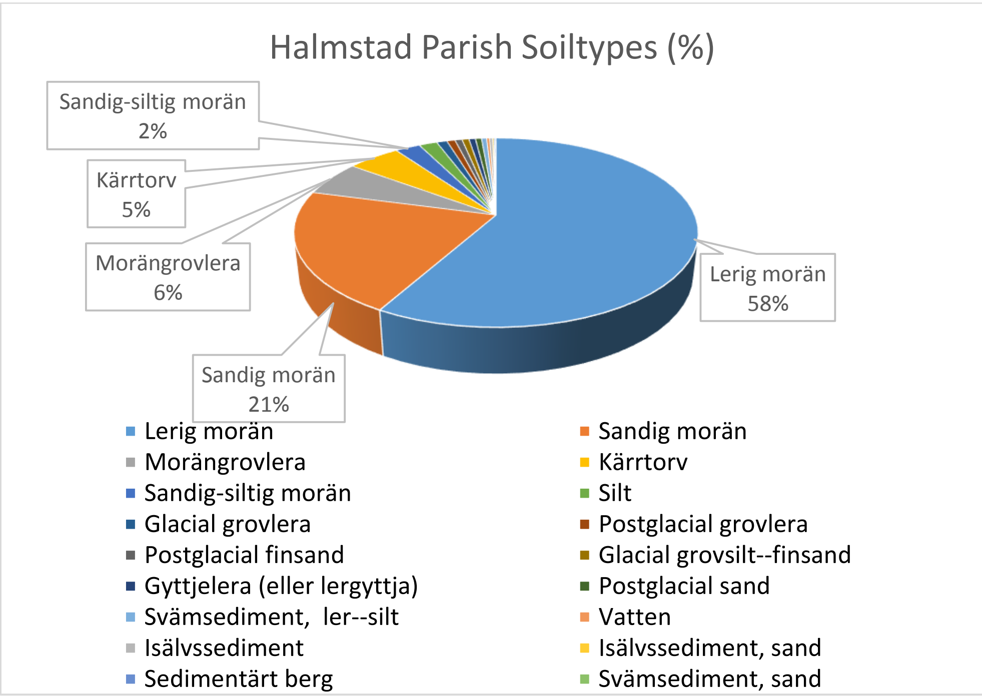


Figure 4.41: *Soil Types of Halmstad Parish*

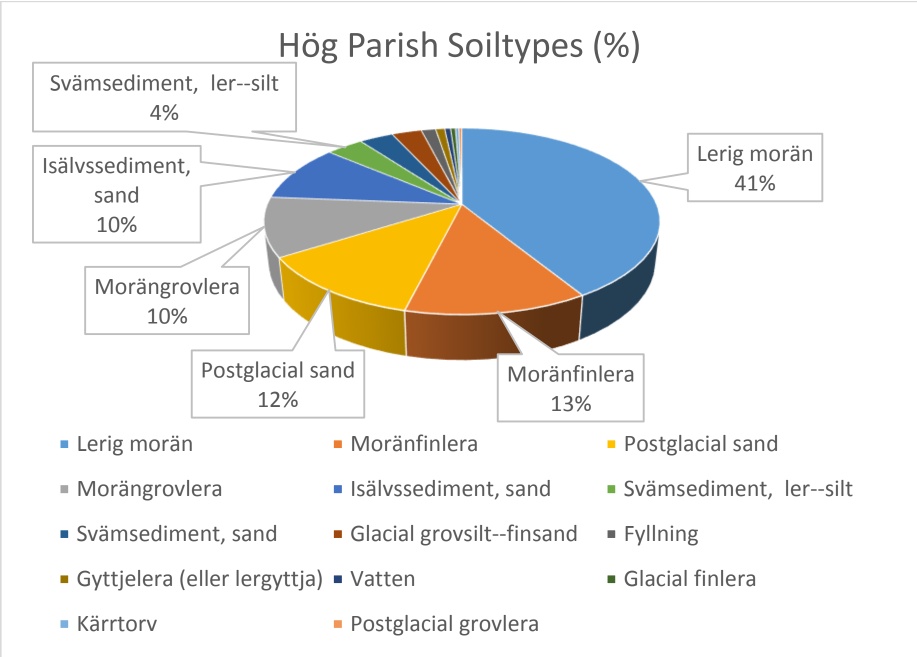


Figure 4.42: *Soil Types of Hög Parish*

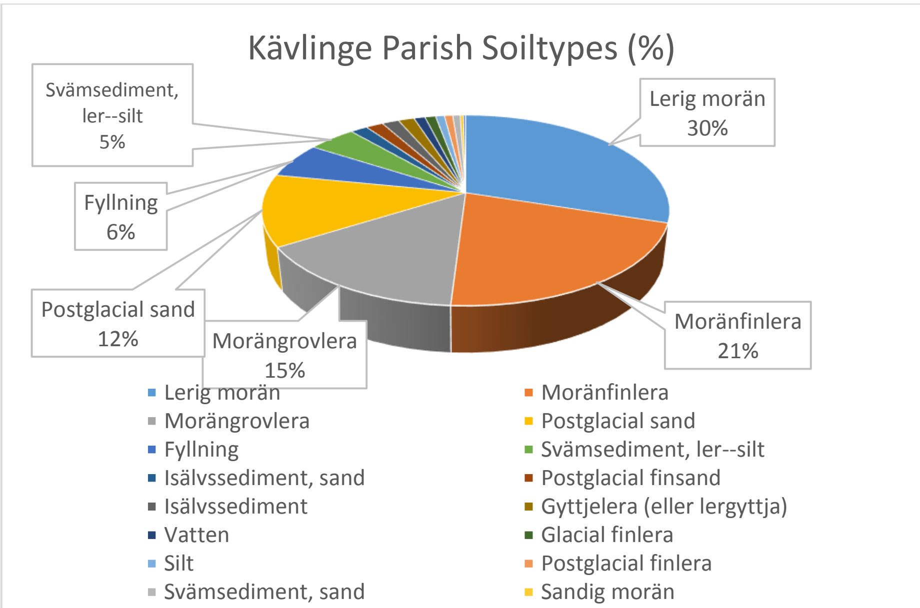


Figure 4.43: *Soil Types of Kävlinge Parish*

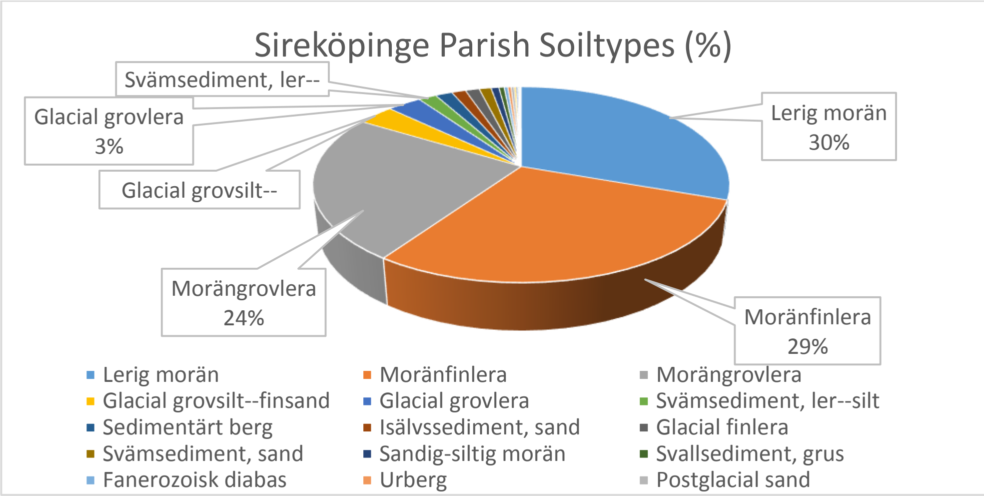


Figure 4.44: *Soil Types of Sireköpinge Parish*

4.4.5.4 Discussion

As illustrated in all soil type diagrams, it is evident that clay till (lerig morän) is the most dominant soil type in all four parishes. The largest concentration of clay till is found in Halmstad parish where it covers more than half of the parish area (58%). Hög also has a large portion of clay till where a total of 41% of the parish is covered by this soil type. Similar to Hög, Kävlinge and Sireköpinge exhibit a more varying soil type pattern where clay till is included amongst the more dominant soil types, covering about 30% of the parish area in both cases. The second most dominant soil type for Halmstad parish is sand till, covering a total 21% of the parish. Other soil types covering 13% to 10% of Hög parish include fine-clay till, postglacial sand, coarse clay till and fluvial sediment sand. Kävlinge also largely covered by some of the aforementioned soil types. More in particular it is 21% covered by fine clay till, 15% by coarse clay till and 12% by postglacial sand. Similarly, Sireköpinge is 29% covered by fine clay till and 24% covered by coarse clay till. In brief, it is evident that all parishes generally contain the same prevalent soil types.

The method applied is suitable for listing all soil types inside a geographic unit and does not leave room for error, as it is based on a simple overlay function followed by a union function than merges all areas sharing the same soil type within the same geographical unit.

5 Study of Geographic Levels

5.1 Introduction

One of the objectives of this study is to provide guidelines to which geographic level the individuals should be linked to in a historical spatial demographic research. In general, the values for the geographic context variables will be more representative the higher the geographic resolution level chosen (exceptions exist). Unfortunately, performing a detailed geocoding of historical individuals is a very time-consuming and expensive task. Therefore, it is interesting to examine if and how much the results of geographic context variables differ whilst computed over other geographic levels with coarser resolution. This part is dedicated to perform such comparisons with the help of Chebyshev's inequality and the concept of dominant soil types. For all non-nominal geographic context variables, a table containing the maximum, minimum, median and average variable values for each geographic level is included. This is done to provide a general overview of the range and frequency of values per geographic level. Then the values of geographic levels are compared by implementing Chebyshev's inequality. More details regarding the exact methodology followed are included in following sections. At this point, it should be mentioned that the distances for all geographic context variables including distance calculations (population density index, proximity to water, proximity to wetlands and proximity to gathering places) were computed with the centroid-to-centroid method. The implementation of this method was deemed more suitable for comparing results from different geographic levels, as it wasn't prone to producing misleading zero-distance results for the geographic levels with coarse resolution.

The list below includes the pairs of geographic levels whose geographic context variable values were statistically compared to each other.

- Initial property unit – property unit
- Address – property unit
- Parish – property unit

5.2 Geographic Context Variable Computation Results per Geographic Level

5.2.1 Population Density Index

Table 5.1 displays the minimum, maximum, median and average population density index value on different geographic unit levels (i.e. property unit level, initial property unit level, address level and parish level) for the entire study area:

Statistical Values	Property Unit (m^{-1})	Initial Property Unit (m^{-1})	Address Unit (m^{-1})
Min	**0.2**	**0.2**	**0.2**
Max	**16.9**	**29.6**	**16.9**
Median	**1.1**	**0.9**	**1.0**
Average	**1.5**	**1.2**	**1.9**

Table 5.1: *Population Density Index results per geographic unit level computed with the "Centroid-to-Centroid" distance calculation method*

A method was used to estimate how much the geographic context variable results differ between geographic levels. For clarity reasons, the steps of the implemented method will be explained with the help of an example. In this example, the population density index results computed on an initial-property-unit level are compared to the corresponding results computed on a property-unit level. Firstly, the absolute difference between every property unit's PDI-value and its corresponding initial property unit PDI-value is calculated (Formula 3).

$$\textbf{Absolute Difference } (PDI_{IPU_{(i)}}, PDI_{PU_{(i,m)}}) = |\boldsymbol{PDI_{IPU_{(i)}} - PDI_{PU_{(i,m)}}}| \quad \textbf{(3)}$$

In Formula 3, $\boldsymbol{PDI}_{\boldsymbol{IPU}_{(i)}}$ represents the PDI-value of initial property unit ***i*** and $\boldsymbol{PDI}_{\boldsymbol{PU}_{(i,m)}}$ represents the PDI-value of property unit ***m*** which belongs to initial property unit ***i***. This difference is calculated between all initial property units and their respective property units in the data set.

In the next step, the mean ***μ*** of all calculated absolute differences is computed as well as the standard deviation ***σ***. Because the values of the absolute differences do not necessarily follow a normal (Gaussian) distribution, the rule that applies for the percentage of values that lie within a bound around the mean in a normal distribution with a width of one, two and three standard deviations respectively, cannot be implemented. Instead, Chebyshev's Inequality (Formula 4), which can be applied to arbitrary distributions, is used to estimate the probability of how many absolute-difference values lie within ***k*** standard deviations of the mean.

$$\boldsymbol{Pr}(|\ \mathbf{X}-\boldsymbol{\mu}\ | \geq \boldsymbol{k\sigma}) \leq \frac{1}{k^2} \qquad , k>0 \qquad \textbf{(4)}$$

X in Chebyshev's Inequality represents a randomly chosen absolute difference value from the computed data set, ***μ*** is the mean of all absolute-difference values, ***σ*** is the standard deviation and ***k*** is any real number higher than zero. Only the case where ***k* >1** is useful, because if k≤1, then $\frac{1}{k^2}$ ≥1 and therefore the inequality is trivial as all probabilities are less or equal to one. For example, if *k = 2*, then according to Chebyshev's inequality, the probability that values lie outside the interval (μ − 2σ, μ + 2σ) does not exceed ¼. In other words, the probability of a value being within the aforementioned interval is equal to 75%.

The comparison between other pairs of geographic levels is performed by implementing the same method. Table 5.2 shows the average absolute difference between PDI-values of the initial property unit level and the property unit level as well as between the address unit level and the property unit level. The interval in table 5.2 is computed for two separate values of ***k*** (k=$\sqrt{2}$ & k=2). According to Chebyshev's Inequality, the probability of the absolute difference between a property unit's PDI-value and its corresponding initial property unit's PDI-value to lie within the interval of (μ-$\sqrt{2}$σ, μ+$\sqrt{2}$σ) or (μ-2σ, μ+2σ) is 50% or 75% respectively.

Area	Initial Property Unit - Property Unit (*Absolute difference in m^{-1}*)				Address Unit - Property Unit (*Absolute difference in m^{-1}*)			
	Mean	**St. Dev.**	**Interval (μ-$\sqrt{2}$σ, μ+$\sqrt{2}$σ)**	**Interval (μ-2σ, μ+2σ)**	**Mean**	**St. Dev.**	**Interval (μ-$\sqrt{2}$σ, μ+$\sqrt{2}$σ)**	**Interval (μ-2σ, μ+2σ)**
Halmstad	0.4	0.7	0 – 1.4	0 – 1.8	0	0.1	0 – 0.2	0 – 0.2
Hög	0.4	0.3	0 – 0.8	0 – 1	0.1	0.1	0 – 0.2	0 – 0.3
Kävlinge	8.4	10.2	0 – 22.9	0 – 28.8	1.2	1.1	0 – 2.8	0 – 3.4
Sireköpinge	0.8	0.7	0 – 1.8	0 – 2.2	0.1	0.1	0 – 0.2	0 – 0.3
All Parishes	**2.1**	**5.6**	**0 – 10.1**	**0 – 13.2**	**0.3**	**0.7**	**0 – 1.3**	**0 – 1.7**

Table 5.2: *Population Density Index result comparison between geographic unit levels*

According to the contents of table 5.2, the PDI-value of a property unit has a 75% probability to differ 0 to 13.2 m^{-1} from the PDI-value of its corresponding address unit and a 75% probability to only differ 0 to 1.7 m^{-1} from its corresponding initial property unit. This does not necessarily mean that initial-property-unit PDI-values and their corresponding property unit PDI-values generally differ much more than the address unit PDI-values do from their corresponding property unit PDI-values, but rather that the maximum absolute difference found between the PDI-value of an initial-property-unit and that of its corresponding property unit is bigger than the maximum absolute difference found between the PDI-value of an address-unit and that of its corresponding property unit. Table 5.1 shows that property units and address units have similar minimum, maximum, average and median values. On the contrary, though initial property units have a somewhat lower median and average value, they have a much greater maximum value, indicating that there is a much wider range of PDI-values produced at this geographic level. Consequently, the mean and standard deviation of the absolute differences are observably higher whilst comparing property units to initial property units, rather than property units to address units. This fact combined with the smaller absolute difference intervals in the second case, could be indicative to the fact that PDI-values of property units generally tend to differ less from their respective address unit's PDI-value than from their initial property unit's PDI-value.

Aggregating population on initial property units and computing the PDI on that geographic level, could have three possible consequences. The PDI-value of an initial property unit could be lower, higher or equal to the PDI-value of its corresponding property units. In general, distances between initial property unit centroids are equal or bigger than distances between property unit centroids. If the increase in population does not counteract the increase in distance, the PDI-value of an initial property unit will become lower than the PDI-values of its property units. Alternatively, if the increase in population is higher than the increase in distance between the initial property unit centroids, then the PDI-value will become higher than the PDI-value of the respective property units. If the aggregated population value counteracts the difference in distance, then the new PDI-value will be similar to the one estimated in the property unit level. Regarding initial property units, it seems that the aggregation of population to larger units has counteracted any increase between the distances between their centroids, enough to produce much higher PDI-values. This is proved by the increase in maximum produced PDI-value for initial property units (Table 1). Kävlinge is the parish where this effect is the strongest, as the mean and standard deviation of absolute differences is remarkably higher, and at the same time close to the documented in table 5.1 maximum PDI-values of the two levels. The fact that initial property units retain the same minimum PDI-value with the property units in conjunction with the slight decrease in median and average PDI-value, suggests that there is a significant enough number of initial property units whose PDI-value has dropped compared to the PDI-value of their related property units. The extent to which the aforementioned situations occur will be decisive in the choice of if the initial property unit geographic level is representative enough to replace the property unit level.

A possible reason why the PDI-value of an address unit tends to differ less from the ones of its corresponding property units may be related to the way the PDI is calculated on the address level. As property units that constitute an address unit are not always located in the same part of the parish, the PDI-value for the overall unit is estimated as the average of the individual PDI-values of those property units. This would smoothen out any big differences in PDI-values between the property units of an address unit, ultimately underestimating or overestimating the true population exposure to population. Because of this, questions might rise regarding how representative the implementation of this PDI formula actually is to estimate the population density on the address geographical level. The absolute difference in PDI-values between property units and their corresponding address unit might be small, but there are issues of representativeness.

5.2.2 Proximity to Water

When estimating the distance to water in this study, we do so in order to define the probability of people living in closer proximity to water showing higher mortality rates. To be able to make accurate associations between proximity to water and mortality, it is essential to estimate as accurately as possible the real distance between where people actually lived and water. Because the study area is a rural area and the main occupations of its population are related to agriculture, it is assumed that the people spent most of their time within their property. So using the geographic level of property units, which has the highest resolution of all other geographic levels in this study, would be a good approximation for estimating results that come closer to the real distance between people and water. However, an important question is whether geographic levels with coarser resolution still manage to produce relatively representative distance estimations. In order to decide that, it is good to estimate the average difference between the distance to water from a geographic unit belonging to one geographic level and the distance to water of its corresponding geographic unit at another geographic level. A general idea of the distance to water measured at different geographic levels can be obtained by examining the contents of table 5.3.

Statistical Values	Property Unit (m)	Initial Property Unit (m)	Address Unit (m)	Parish Unit (m)
Min	**0.3**	**7.9**	**9.7**	**59.3**
Max	**2433.9**	**2371.1**	**2375.7**	**1021.3**
Median	**420**	**336.0**	**357.9**	**600.9**
Average	**633.3**	**481.2**	**512.5**	**570.6**

Table 5.3: *Proximity to Water result comparison per geographic unit level computed with the "Centroid-to-Border" distance calculation method.*

By viewing the maximum distance values in table 5.3, it is observed that the coarser the resolution of a geographic level the lower the maximum value of its computed shortest distance to water. On the contrary, the minimum distance values increase as the resolution of a geographic level gets coarser. This suggest that the size of geographic units, which is largely determined by the choice of geographic level, might be affecting the shortest distance to water calculations. However, the median and average distance to water values do not follow a similar pattern, as they generally neither decrease or increase with the decrease in resolution of geographic levels. This could be an indication that the size of a geographic unit is not the only factor affecting calculations of the shortest distance to the nearest waterbody. As waterbodies are not equally dispersed over the study area (Figure 4.18), another possible factor affecting the difference in distance averages between different geographic levels, could also be the spatial distribution of the streams and lakes.

The comparison between proximity-to-water values computed over different geographic levels is conducted with the help of Chebyshev's Inequality. Table 5.4 shows the interval ($\mu-\sqrt{2}\sigma$, $\mu+\sqrt{2}\sigma$) or ($\mu-2\sigma$, $\mu+2\sigma$) within which the absolute difference between the proximity-to-water values of two geographic units belonging to different geographic levels might lie in, with a probability of 50% or 75% respectively.

Area	Initial Property Unit - Property Unit *(Absolute difference in m)*				Address Unit - Property Unit *(Absolute difference in m)*				Parish Unit - Property Unit *(Absolute difference in m)*			
	Mean	**St. Dev.**	**Interval (μ-$\sqrt{2}\sigma$, μ+$\sqrt{2}\sigma$)**	**Interval (μ-2σ, μ+2σ)**	**Mean**	**St. Dev.**	**Interval (μ-$\sqrt{2}\sigma$, μ+$\sqrt{2}\sigma$)**	**Interval (μ-2σ, μ+2σ)**	**Mean**	**St. Dev.**	**Interval (μ-$\sqrt{2}\sigma$, μ+$\sqrt{2}\sigma$)**	**Interval (μ-2σ, μ+2σ)**
Halmstad	90.3	112.8	0 – 249.8	0 – 315.9	87.1	108.3	0 – 240.3	0 – 303.7	206	188.3	0 – 472.3	0 – 582.6
Hög	122.3	105	0 – 270.7	0 – 332.3	114.1	88.2	0 – 238.9	0 – 290.5	443.6	291.3	31.7 – 855	0 – 1026
Kävlinge	136.8	134.5	0 – 327	0 – 405.8	468.3	310	29.9 – 907	0 – 1088	607.7	319.3	156 – 1059	0 – 1246
Sireköpinge	79.9	103.6	0 – 226.4	0 – 287.1	93.5	103.3	0 – 239.6	0 – 300.1	286.6	262	0 – 657.1	0 – 810.6
All Parishes	**102.2**	**115**	**0 – 264.8**	**0 – 332.2**	**166**	**217.5**	**0 – 473.6**	**0 – 601**	**359.2**	**301.5**	**0 – 785.6**	**0 – 962.2**

Table 5.4: *Proximity to Water result comparison between geographic unit levels – Results calculated with the Centroid-To-Border distance calculation method.*

According to table 5.4, there is a 50% probability of the absolute difference between a property unit's shortest distance to water and the shortest distance to water of its respective initial property unit, to be within the range of 0 to 264.8 meters. Whilst comparing the differences in produced distance values between property units and address units, it is shown that there is a 75% probability for the absolute difference in distance to be included within the interval of 0 to 601 meters. The respective interval within which the absolute difference in distance to water between property units and parishes is included in, is 0 to 962.2 meters. Consequently, it seems that the coarser the resolution of the geographic level to which the property unit level is being compared to, the larger the absolute difference in distance to water. Consequently, the coarser the resolution of a geographic level the less accurate the estimation of minimum distance to water for that level. The largest values of absolute difference averages, standard deviations and ranges of intervals are observed in parishes that do not have a dense waterbody network (e.g. Kävlinge and Hög). On the contrary, the smallest absolute differences are found in the comparison between geographic levels of parishes that have a much denser waterbody network, like Halmstad and Sireköpinge. This indicates that the spatial distribution –of the related to the geographic context variable– geographic objects (e.g. lake-polygons and stream-lines), is affecting the value of the absolute difference in shortest distance to water between units of different geographic levels.

5.2.3 Proximity to Wetlands

Table 5.5 shows the minimum, maximum, median and average distance to wetlands value on different geographic unit levels (i.e. property unit level, initial property unit level, address level and parish level) for the entire study area:

Statistical Values	Property Unit (m)	Initial Property Unit (m)	Address Unit (m)	Parish Unit (m)
Min	**0**	**0**	**0**	**62.2**
Max	**1471.2**	**1382.4**	**1357.4**	**879.6**
Median	**333.6**	**320.8**	**298.0**	**535.0**
Average	**410.5**	**384.5**	**356.1**	**502.9**

Table 5.5: *Proximity to Wetlands results per geographic unit level computed with the "Centroid-to-Border" distance calculation method.*

The contents of table 5.5 follow similar patterns to those observed in table 5.3, suggesting that the size of geographic units, which are largely determined by the choice of geographic level, as well as the spatial distribution of wetlands are affecting the results of the computation of the shortest distance to wetlands.

The comparison between proximity-to-wetlands values computed over different geographic levels is conducted with the help of Chebyshev's Inequality. Table 5.6 shows the interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$ or $(\mu-2\sigma, \mu+2\sigma)$ within which the absolute difference between the proximity-to-wetland values of two geographic units belonging to different geographic levels might lie in, with a probability of 50% or 75% respectively.

Area	Initial Property Unit - Property Unit *(Absolute difference in m)*				Address Unit - Property Unit *(Absolute difference in m)*				Parish Unit - Property Unit *(Absolute difference in m)*			
	Mean	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**	**Mean**	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**	**Mean**	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**
Halmstad	80.9	97.4	0 – 218.6	0 – 275.7	81.5	98.3	0 – 220.5	0 – 278.1	194	111.5	36.3 – 351.7	0 - 417
Hög	110	100.4	0 – 251.9	0 – 310.8	103.4	78.4	0 – 214.3	0 – 260.2	325.3	175.8	76.7 – 573.9	0 – 676.9
Kävlinge	100.8	129.3	0 – 283.6	0 – 359.4	228.9	172.5	0 – 472.8	0 – 573.9	613.5	291.8	201 – 1026.1	0 – 1197
Sireköpinge	85.1	120	0 – 254.8	0 – 325.1	100.4	110.1	0 – 256.1	0 – 320.6	327.2	248.6	0 – 678.8	0 – 824.4
All Parishes	**92.1**	**113.2**	**0 – 252.2**	**0 – 318.5**	**120.2**	**127.8**	**0 – 300.9**	**0 – 375.8**	**347.1**	**259**	**0 – 713.4**	**0 – 865.1**

Table 5.6: *Proximity to Wetlands result comparison between geographic unit levels – Results calculated with the Centroid-To-Border distance calculation method*

The information provided in table 5.6 largely resembles the information extracted from table 5.4. Once more, the coarser the resolution of a geographic level the less accurate the estimation of minimum distance to wetlands for the geographic units of that level. As wetlands are more evenly distributed in the study area compared to lakes and streams, it is harder to find strong patterns that reveal differences between parishes.

5.2.4 Proximity to Gathering Places (Church)

Table 5.7 shows the minimum, maximum, median and average distance to gathering places value on different geographic unit levels (i.e. property unit level, initial property unit level, address level and parish level) for the entire study area:

Statistical Values	Property Unit (m)	Initial Property Unit (m)	Address Unit (m)	Parish Unit (m)
Min	**0.7**	**0.7**	**0.7**	**6.5**
Max	**4487.7**	**4487.7**	**4483.8**	**1258.2**
Median	**1855.8**	**1847.8**	**1826.8**	**975.9**
Average	**1824**	**1820.4**	**1922.1**	**804.1**

Table 5.7: *Proximity to Church results per geographic unit level computed with the "Centroid-to-Border" distance calculation method*

The contents of table 5.7 show that the minimum, maximum, median and average distance-to-church results of property units, initial property units and address units are quite similar. On the contrary, parish units have a higher minimum distance-to-church value and observably lower maximum, median and average values than the previous geographic levels. This could be indicative as to the location of every parish church being closer to the centroid of its parish unit. The large difference between minimum and maximum values in conjunction with the lower average values suggests that the shortest distance to church varies significantly between the geographic units of the property unit, initial property unit and address unit geographic levels.

The comparison between proximity-to-church values computed over different geographic levels is conducted with the help of Chebyshev's Inequality. Table 5.4 shows the interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$ or $(\mu-2\sigma, \mu+2\sigma)$ within which the absolute difference between the proximity-to-church values of two geographic units belonging to different geographic levels might lie in, with a probability of 50% or 75% respectively.

Area	Initial Property Unit - Property Unit *(Absolute difference in m)*				Address Unit - Property Unit *(Absolute difference in m)*				Parish Unit - Property Unit *(Absolute difference in m)*			
	Mean	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**	**Mean**	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**	**Mean**	**St. Dev.**	**Interval $(\mu-\sqrt{2}\sigma, \mu+\sqrt{2}\sigma)$**	**Interval $(\mu-2\sigma, \mu+2\sigma)$**
Halmstad	137.9	196.6	0 – 416	0 – 531.1	143.5	192.8	0 – 416.1	0 – 529.1	1192.4	802.9	56.9 – 2328	0 – 2798
Hög	134.5	124.7	0 – 310.9	0 – 383.9	127	107.8	0 – 279.5	0 – 342.6	797	294.6	380 – 1214	208 – 1386
Kävlinge	132.6	117.8	0 – 299.2	0 – 368.2	301.8	256.2	0 – 664.1	0 – 814.2	717.1	386.8	170 – 1264	0 – 1490.7
Sireköpinge	129.6	162	0 – 358.7	0 – 453.6	172.6	182.4	0 – 430.5	0 – 537.4	1788.4	581	967 – 2609	626 – 2950
All Parishes	**133.3**	**157.9**	**0 – 356.6**	**0 – 449.1**	**179.7**	**198.8**	**0 – 460.8**	**0 – 577.3**	**1225.8**	**725.4**	**200 – 2251**	**0 – 2676**

Table 5.8: *Proximity to Gathering Places (Church) result comparison between geographic unit levels – Results calculated with the Centroid-To-Border distance calculation method*

The information included in table 5.8 shows that the coarser the resolution of a geographic level the less accurate the estimation of minimum distance to church for the geographic units of that level. This is proved by the large range of absolute difference intervals whilst comparing the absolute difference of distance values produced by property units and their corresponding initial property unit, address unit or parish unit. No large differences between the mean absolute differences of the distance-to-church values are observed whilst comparing property units to initial property units or addresses. The only exception is Kävlinge, where whilst comparing address units to their respective property units the difference between their values is larger. This could be attributed to the fact that Kävlinge contains addresses whose property units are located far away from each other, causing the average shortest distance to church for the address unit as an entity to be larger or smaller than the actual distance to church of each one of its property units. With the absolute difference between a property units' minimum distance to church having 75% probability to be within the range of 0 to 2.676 meters from its corresponding parish unit, it is doubtful as to how representative the parish level is to depict the population's actual distance to church.

5.2.5 Soil Types

The method used to calculate the total area of each soil type within a geographic unit was described in chapter four. In this section, the difference between soil-type results produced for geographic units of different geographic levels will be analyzed. Because every geographic unit, depending on its size and location, may have from one to several soil types included within its boundaries, it would be too complicated to compare the results between two geographic units of different geographic levels for all the soil types in their area. Therefore, only the most dominant soil type of every geographic unit will be taken into account for the purpose of this analysis. The most dominant soil type is expressed as the soil type that covers the largest proportion of the geographic unit area. Another reason why only the most dominant soil type was used to conduct this analysis was that, in theory, this is also the soil type which most affects the individuals within the same geographic unit. In practice, the difference between the most dominant and the second most dominant soil type could be very small, especially if the study area is characterized by the existence of a widely dispersed, large variety of soil types. Nevertheless, considering the diagrams in figures 4.42 – 4.45, it can be derived that three to four soil types are prevailing over the study area, whereas the rest of them only cover small proportions.

More in particular, the comparison between two geographic levels (e.g. initial property units and property units) was done by computing the total percentage of geographic units belonging to the finer resolution geographic level (property units) that had the same dominant soil type with their corresponding geographic unit from the coarser geographic level (initial property units). The same logic is applied to the rest of the geographic unit layer comparisons. Although this approach is not definite, it can be considered as a good and indicative first attempt of controlling how much the results between two different geographic levels differ. Table 5.9 shows the variations in top soil type from one geographic unit layer to another.

Area	Initial Property Unit - Property Unit	Address Unit - Property Unit	Parish Unit - Property Unit
Halmstad	78.9%	80.9%	62.2%
Hög	73.5%	73.5%	32.7%
Kävlinge	59.6%	51.5%	36.8%
Sireköpinge	86.2%	80.1%	41.9%
All Parishes	**80.4%**	**73.6%**	**44.3%**

Table 5.9: *Soil type result comparison between geographic unit levels*

From table 5.9 it is becoming clear that the coarser the resolution the smaller percentage of property units sharing the same dominant soil type with their respective geographic unit from the coarser geographic level. It should be noted, that when the two geographic levels (i.e. property units and parishes) which have the largest resolution difference are being compared, less than 45% of the property units share the same dominant soil type with their corresponding parish unit. This can be attributed to the fact that their geographic units have a significant difference in size, and therefore the probability of them sharing the same dominant soil type becomes increasingly lower. However, the difference in size being the only reason can be questioned by considering the cases of Kävlinge and Hög. Even though they are the smallest parishes in the study area, only 36.8% and 32.7% of their property units share the same dominant soil type with their parish unit. The diagrams in figures 4.43 and 4.44 show that these parishes have a larger number of soil types covering larger proportions of their parish than Halmstad or Sireköpinge which are mainly dominated by two or three soil types respectively. The same point can be proven by the fact that even though Halmstad is the largest parish in the area and therefore could be expected to have a smaller number of property units sharing the same dominant soil type, this is not the case. On the contrary, it is the only parish which has a relatively to the others high percentage of property units sharing the same dominant soil type. Figure 4.42 provides the answer. Nearly 80% of the parish area is covered by only two different soil types. Consequently, the number of prevailing soil types in an area as well as their spatial distribution can play a significant part in the decision of, if selecting a coarser geographic level to link individuals to, is a representative enough choice.

Initial property units and address units show small differences in dominant soil types compared to their corresponding property units. Overall, initial property units seem to have slightly smaller differences. A possible reason for this could be that they consist of one concrete unit, whereas address units might consist of several property units that are located in further away proximity from each other. Now, depending on the number and spatial distribution patterns of the soil types in every parish, widely dispersed property units belonging to the same address unit have a higher probability of not sharing the same dominant soil type with their corresponding address unit. Kävlinge, which is the parish with the most remotely located property units belonging to the same address unit is a typical example of this, since on average only 51.5% of the property units share the same dominant soil type with their respective address unit. However, remembering that this is also the parish with the largest number of dominant soil types in the study area, it is not surprising that it yields the lowest percentages, even whilst comparing property units to initial property units or address units. This is further proof of how the total number of dominant soil types affects the difference between geographic levels.

6 Discussion

As mentioned before, choosing a representative distance calculation method can be very difficult. In this study two such methods were implemented. The border-to-border method was found to be partial at producing zero-distance results depending on the spatial distribution of the geographic object to which the proximity to was being estimated and the choice of geographic level (Figure 6.3). Examples over how the specific location of the church can be affecting the results produced by the border-to-point method are shown in figures 6.1 and 6.2. Examples over how the choice of geographic level can affect the results produced by the border-to-border method are shown in figure 6.3.

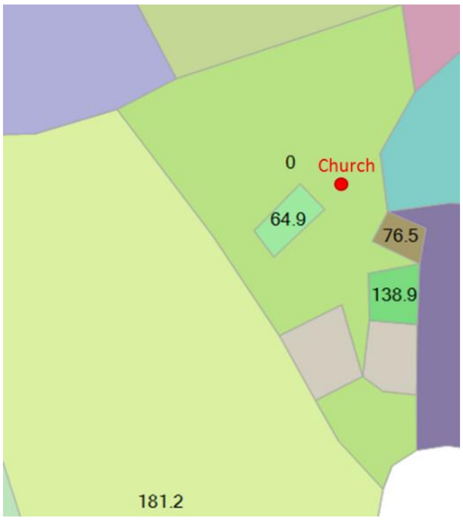
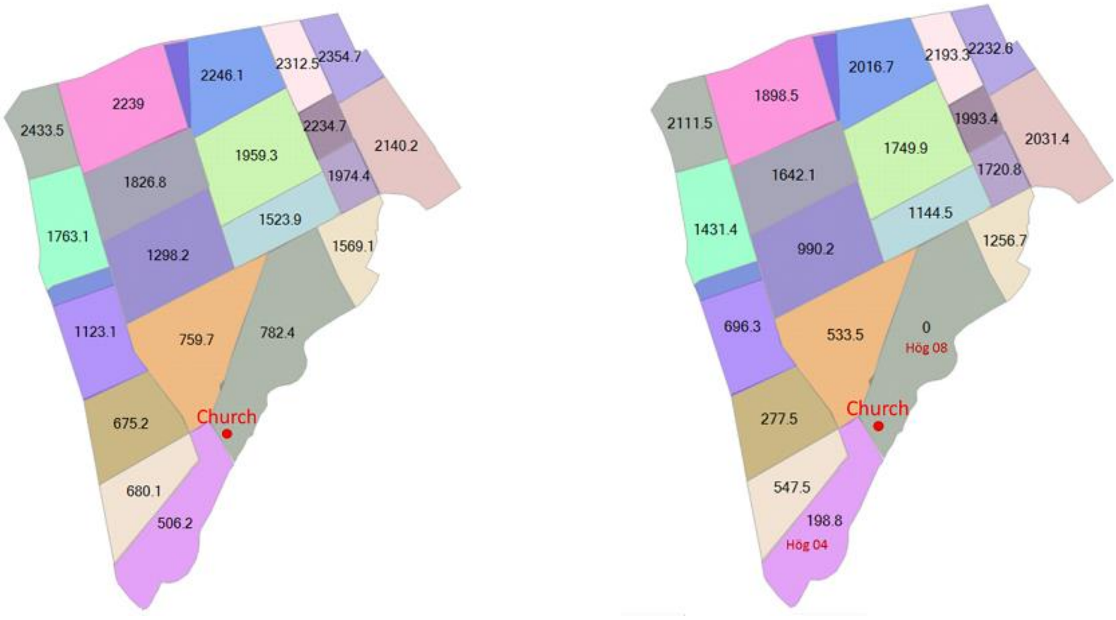


Figure 6.1: *Kävlinge parish property unit containing the church point geometry as well as other property units. Distances from the containing property units to the church are larger than 0m even if they are closer to the church than the average distance from the larger containing polygon's borders.*

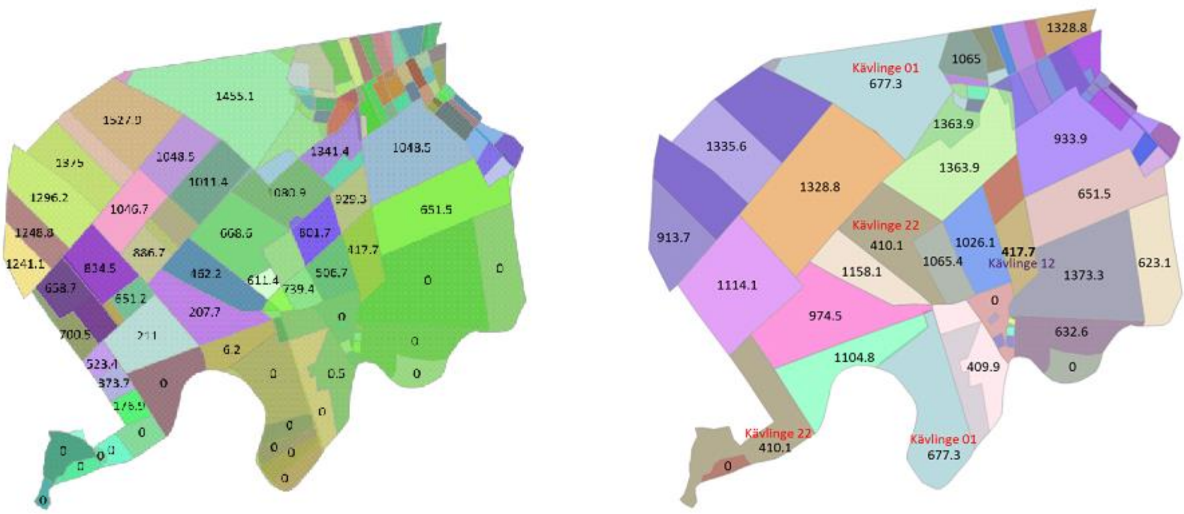
If the distance between a polygon (e.g. property unit) and a point (e.g. building-church) is estimated as the shortest Euclidean Distance between the border of the property-unit polygon and the building point, then the distance will be equal to zero if the point is contained by the polygon. However, if the property unit that contains the church also contains other property units, then their distance to the church will not be equal to zero (Figure 6.1). This produces a skewed representation of reality where the average distance from any point within the larger containing polygon (green polygon) to the church is much greater than 0 meters.



(a) (b)

Figure 6.2: *(a) Distance to church computed with the centroid to point distance calculation method on initial property unit level for Hög parish. (b) Distance to church computed with the border to point distance calculation method on initial property unit level for Hög parish.*

According to the border-to-point, method initial property unit “Hög 08” is closer to the church than initial property unit “Hög 04” (Figure 6.2 (b)). This result is the complete opposite compared to the result of the centroid-to-point method in figure 6.2 (a).



(a) (b)

Figure 6.3: *(a) Distance to water computed with the border-to-border distance calculation method on property unit level for Kävlinge parish. (b) Distance to water computed with the border-to-border distance calculation method on address level for Kävlinge parish.*

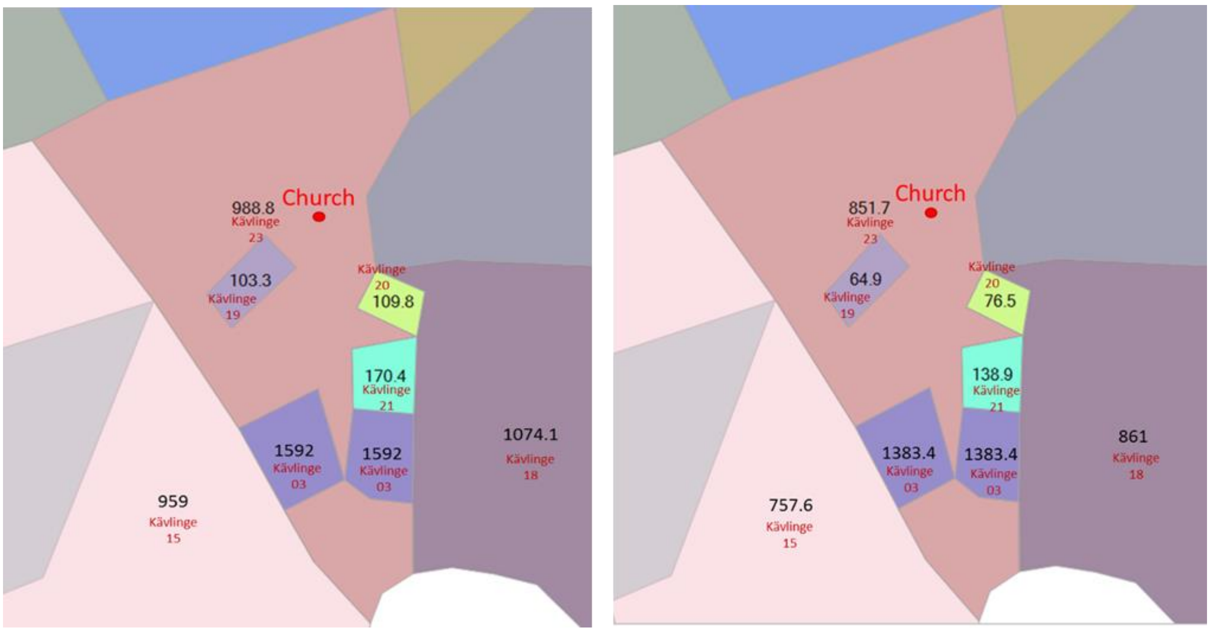
In figure 6.3 (b), the property unit with address “Kävlinge 01” in the south was zero when distance was calculated on the property unit level (figure 6.3 (a)). Because there is a property unit with the same address in the north of the parish, the distance has changed from 0 to 677.3 meters. The property unit with address “Kävlinge 22” in the southwestern part of Kävlinge is located within 410.1 meters to water according to figure 6.3 (b). However, this produces a false representation of reality since most property units of that address are located in the southwestern part of Kävlinge near the stream, having a distance from 0 to 500 meters from it (Figure 6.3 (a)). Only the property unit of the same address located in the north has a distance of 1450 meters from the stream. The results shown in figure 6.3 (b) are very misleading, because property units whose distance to the stream was equal to zero before now have been assigned with a distance of 410.1 meters and the northern property unit of the same address has been assigned with a much lower value that its own distance value to the stream. Subsequently the 410.1 meters are not representative to any of the property units of this address unit.

In addition to the above, it was also found that the border-to-border method constantly produced lower distance-results than the centroid-to-border method. The coarser level of geographic resolution implemented the greater the probability of the border-to-border method producing a zero-distance result (e.g. parish level). Depending on the topography and the population distribution within a geographic unit of the coarsest level, the results produced by the aforementioned method might be unrepresentative. Therefore, the centroid-to-border method was considered as a more appropriate method to implement. However, given its vulnerability to the “horseshoe-problem” and other size and shape problems causing a centroid to be placed at parts within the geographical unit, which may or may not be at all representative of the location where people linked to that geographical unit were actually present at, stresses the need for considering possible improvements.

The choice of a suitable geographic level to link demographic and geographic information is crucial to the outcome of a historical spatial demographic research. In this study, the results of a number of geographic context variables computed over different geographic levels were statistically compared to show if and how much results computed over different geographic levels could differ. Results of geographic context variables computed over the property unit level are assumed as the most accurate, since this is the most detailed level at which demographic information has been linked to, and because of the assumption that people in the rural communities of the 19th century Scania were mainly involved with agricultural activities and therefore probably spent most of their time inside their properties. Generally, results calculated over the property unit level are closer to those calculated over the initial property unit level or address level, rather than those computed over the parish level.

More in particular, property units produce more similar results to their respective initial property units when calculating the proximity to water, the proximity to gathering places and the dominant soil type. Property units display results that are more similar to those produced by their corresponding address units when calculating the population density index. Regarding the proximity to wetlands, property units tend to differ relatively equally as much from the results of their corresponding initial property units or address units. The reason why these differences occur is closely related to the general characteristics of the geographic units in every geographic level (size, shape, level of homogeneity in size or shape), the geography of the study area, the spatial distribution of the geographic feature (e.g. lake-polygons) the geographic context variable is concerned of, and the implemented geographic context variable calculation method. Population Density Index calculations are affected by the choice of geographic level (shape, size and level of shape and size homogeneity within the geographic level) and population patterns. All geographic context variables involving the concept of distance to specific geographic features (e.g. streams, lakes, wetlands and churches) are aslo affected by the choice of geographic level as well as the spatial distribution of the geographic feature. Finally, soil types are affected by the choice of geographic level and the apportionment and total number of soil types in the study area.

Initial property units can differ from property units in size, shape and potentially in level of homogeneity regarding size or shape. The same fact applies for the difference between property units and parish units. The size of an initial property units was related to the total area of land a person owned. The level of homogeneity regarding the size of initial property units is closely related to how the land was divided amongst the population. If few wealthy land owners existed then the initial property units would be covering large areas, and consequently –depending on the geography of the study area– the difference between computing geographic context variables over this geographic level and the property unit level would be greater. On the contrary, if the land was more evenly distributed amongst the people, the size of an initial property unit might not vary much amongst initial property units or from the size of property units, and therefore the results of their geographic context variable computations would be more similar. Hög parish could be regarded as such an example. However, if the land is not equally distributed amongst the people there would be greater differences between the sizes of initial property units, and, therefore, the difference between the computation of geographic context variables over the initial property unit level and the property unit level would yield from similar to very different results. For instance, this was the case for the parishes of Halmstad and Sireköpinge, where the land was predominantly owned by the nobility and the rest of it was divided to smaller land parcels. The extent to which the aforementioned characteristics differ across the study area, in conjunction with the geography and spatial distribution of geographic features related to the examined geographic context variable, would be considered as decisive factors as to if the initial property unit level could be used to link individuals to instead of the property unit level.



(a) (b)

Figure 6.4: *(a) Distance to church computed with the centroid to point distance calculation method on address level for Kävlinge parish. (b) Distance to church computed with the border to point distance calculation method on address level for Kävlinge parish.*

Address units and property units might yield more similar or different results, depending on the size and shape of their units as well as on the number of address units consisting of property units which are located in longer distance from each other. The last characteristic of an address unit could be problematic and produce unrepresentative results, if individuals were to be linked to this layer. Typical examples of this problem are shown in figures 6.3 and 6.4. In Figure 6.4 (a), the property unit where the church is located has a longer distance to church than other property units in the area. This is caused by the fact that he property unit containing the church has the same address with another property unit located further away, near the southwestern border of Kävlinge. The distance to church for this address unit is calculated as the average distance to church for all property units belonging to this address. Consequently, instead of having a zero distance to church, the property unit where the church is located has been assigned with a distance to church which is equal to 988.8 meters (Figure 6.4 (a)). Correspondingly, the other property unit of the same address unit will be assigned with a lower distance value than its actual one. This problem persists even whilst implementing the border-to-point distance calculation method (Figure 6.4 (b)). As shown in both cases, calculating proximity on address level by averaging the distances of all property units having the same address may produce misleading results. People living in different property units of the same address, but in a long distance from each other, are exposed to totally different conditions and should therefore not be examined as one entity. Important information regarding the characteristics of the environment people are living in as well as their response to it would be lost, if people were linked to addresses and geographic context variables were computed over their geographic units. For all the above reasons, this geographic level should most preferably not be used to link individuals to.

The geographic level of parishes, according to the results in chapter five, has a too coarse resolution to adequately represent the true conditions individuals are exposed to and triggers the scale effect of MAUP. For the computation of most geographic context variables the spatial distribution of geographic features such as lakes, streams, wetlands and soil types was too varying to be represented by one value.

6.1 Quality Issues

This section presents several factors that might affect to some extent the results of this study. Some of these factors are listed below:

- Incomplete data for all four parishes. The most complete information, at this point in time, is provided for the parish of Hög.
- Incomplete information regarding the assignment of individuals to property units, since individuals are only linked to the initial property units through the annual poll-tax registers. Associations are finally made based upon scientific research and intelligent guesswork.
- Information about the location of wetlands is only available through historical maps. A historical map is only a snapshot of the real world in time. There is no source that would provide information regarding the existence and/or the extent of the depicted wetlands during frequent enough time intervals. This is important since it is known that many wetlands were dried out to produce more space for cultivated land in the past. It is a difficult task to find the exact period in time when these desiccations took place as well as to know how climatic factors influenced the extent of the wetlands. The quality of the historical maps and the absence of detailed periodical information of the wetlands raise accuracy concerns.

- Calculating distances only by implementing the Euclidean method. In some cases, it might have been useful to calculate distances by implementing the network method instead (e.g. distance from gathering places). It might have been useful to implement both distance-calculation methods and study if and how much their results actually varied as well as how much that could potentially affect the outcome of the study.

  Factors that affect accuracy of distance calculations:
  - Accuracy of methods used to produce the historical maps
  - Quality and condition of the paper the maps were drawn on (distortion due to dampness or dried out maps or torn maps, etc.)
  - Accuracy of the digitization process of the historical map.
  - Modern projection the map used (risk for distortion).
  - Choice of geographic (e.g. lat/long) or geometric distance (e.g. Euclidean, Network, etc.).
  - How we define a shortest distance between two geographic objects (e.g. distance between the centroid of one polygon and a point in the border of another polygon or the distance between a point in the border of one polygon and a point in the border in another polygon, etc.).
  - Research question (e.g. are we interested in merely the distance or the exposure as well; e.g. distance from nearest wetland or distance from nearest wetland combined with nearest wetland area or nearest wetland areas within a defined buffer-zone)
  - Mobility (movement) patterns of individuals. How individuals tend to move in the study area as it is described by the chosen geographic level. In this case the choice of geographic level affects the distance calculation as it affects the distance between the centroid or border of a geographic unit and a geographic feature (e.g. wetlands).

Moreover, there are some additional factors which affect the Population Density Index results. In the property unit's index for the parishes of Kävlinge and Halmstad, there are cases where property units cover parts or the whole area of more than two initial property units. These very big property units could not be linked to any initial-property-unit and were therefore excluded from the calculations regarding all context variables. However, as some of them might have population linked to them for 1910, their exclusion from the computations of the Population Density Index, might have resulted in the presentation of a false estimation regarding the true population density of the areas they covered as well as their surrounding areas within the 4 km radius for 1910.

There are cases where the same property unit is represented twice in the dataset for 1910. One of the copies belongs to the land survey map series and the other to the economic map series. The size and or shape of the property units are not always identical in the map. This is partly due to some minor distortion caused by the fact that they originate from different maps and some distortion is expected during the georeferencing and digitization process. Both property units have been assigned with population for the year of 1910. One has directly been assigned population and the other has only a portion of the population assigned to the address unit it belongs to. In theory, these two property units should never coexist in time because they overlap. The fact that both have been assigned with population and therefore are included in the PDI calculation process is responsible for the emergence of some extremely high PDI values in the study area. This is because the distance between their centroids is extremely small, producing an extremely low denominator in the PDI formula and depending on whether or not they have been assigned with a large population the final PDI value can be hugged and totally unrealistic.

Because using data from only one map series would not leave enough data covering the whole study area to calculate the PDI, data from both map series were kept. To mitigate the problem, a buffer was introduced, which only permitted the PDI formula to be calculated for property units whose distance between their centroids is longer than 40 meters. The value of the previously mentioned threshold was determined after thorough examination of the problematic property units. All property units whose distance between their centroids was shorter than 40 meters were exhibiting this problem. The distance of the first pair of property units which did not have this issue but were in close proximity to each other was over 40 meters. This does not mean that the problem ceased to exist for the rest of the dataset, but it solved the problem of the extreme PDI values.

6.2 Future Developments

In the future, a new way of calculating geographic context variables on the property unit level might be introduced. This method can be applied on the already existent data in the following manner. Information about the buildings of the study area is available for the study period in different datasets. This information was obtained from geocoded historical maps. Some datasets include buildings in the form of polygons and others in the form of points. Most datasets cover different time periods, but some overlap to a certain extent. In the latter case, a decision needs to be made as to which dataset should be used to depict what period. Since the temporally overlapping datasets do not usually agree, a suggestion could be to choose the most recent dataset to cover the most recent period in time. It is more likely that the datasets produced to cover a later period in time, provide a more accurate account of the study area in that particular point in time. Once this has been done, the convex hull of all buildings present at a certain property unit or initial property unit at a specific period in time can be calculated. Next, the geographical context variables can be calculated for all the convex hulls. For instance, distance calculations can be performed by calculating the distance from the convex hull centroid or borders.

Provided that the building data is adequately accurate, this method does not entail any other limitations or assumptions that would affect the accuracy of the calculations of geographic context variables on that aggregation level. If the buildings are not too scattered within the enclosing property unit, in most cases, the convex hull will cover a significantly smaller area. Subsequently, individuals can be placed to specific locations in a more accurate way. This is important in historical longitudinal studies of rural areas where individuals are believed to move more frequently in locations closer to their area of residence and work (in this case coinciding since we deal with a primarily agricultural type of society). Moreover, this method is easy and fast to implement and does not stress the availability of resources (e.g. time, finances, historical material).

A possible solution for the issues shown in Figure 4.29 and Figure 4.30 could be provided by introducing a buffer area (Figure 6.5). The distance from a geographical unit to the nearest wetland would be calculated in the same way as before, with the exception that only the wetlands that intersect the buffer-area would be included in the process. Moreover, to measure the extent of exposure to wetlands, the results could also include the area of the nearest wetland, the area of the nearest wetland that intersects the buffer-area and the total sum of wetland area that is included within the buffer area. By changing the radius of the buffer, the user is able to obtain a better understanding of the magnitude of exposure to wetland area (e.g. important in cases of malaria studies, where wetlands are the natural habitat of the disease-carrying mosquitos). This method could be applied when calculating the shortest distance between the centroid of a geographical unit polygon and the border of a wetland polygon as well as when calculating the shortest distance between the border of a geographical unit polygon and the border of a wetland polygon. It should be noted that even if it were to be implemented, the problems caused by the differentiations between the two distance calculation methods would still persist.

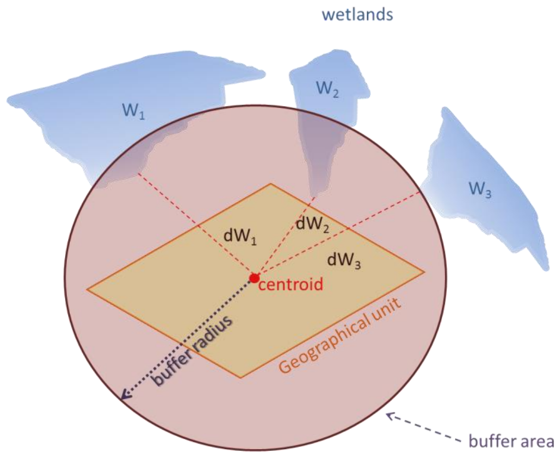


Figure 6.5: *Example of applying a buffer area in the problem of estimating exposure to wetlands alongside with calculating the shortest distance to them*

Addressing the vulnerabilities of the “centroid-to-border” distance calculation method, a plausible alternative could be to randomly place a certain number of points inside every geographical unit and calculate the shortest Euclidean distance to e.g. wetlands for each one of them. Then the shortest distance to wetlands for the entire geographical unit could be estimated as the computed average of all these partial distances. This could possibly provide a more representative result regarding the actual overall shortest distance between the geographical unit and the nearest wetland, and, at the same time, eliminate the probability of the “horseshoe problem” to occur and affect the estimation of the distance result.

7 Conclusions

After reflection, the key conclusions of this thesis project can be summarized in the following points:

- The results of geographic context variables involving distance calculations should be interpreted with causion, as their results vary depending on:
  - the definition of the geographic context variable (e.g. measuring distance from polygon centroid to polygon border, or polygon border to polygon border).
  - the implemented distance calculation method (e.g. Euclidean, spherical – lat/long, network, etc.).
  - geographic level (size and shape of geographic units, homogeneity in shape and size of geographic levels).

- From the methods implemented in this study, the centroid-to-border distance calculation method is considered as the most suitable. Although it is sensitive to the shape of geographic units (horseshoe problem), it is not prone to produce unrepresentative zero-distance results like the border-to-border method. The centroid-to-border method always yields higher distance-values compared to the border-to-border method.

- In spatial historical demography, the choice of geographic level is directly related to the research hypothesis, the characteristics of the study area, data availability and financial resources.

- From the geographic levels in this study, the property unit is considered as the most representative geographic level. This is based on the research hypothesis which concerns mortality patterns of individuals, the rural character of the study area and the notion that people involved in agricultural activities of the past did tend to spend most of their time in the property unit they cultivated.

- Further research is needed to conclude if initial property units could produce representative enough results for this type of study.

- The geographic level of addresses is not representative for this study. Cases where address units consist of more than one, non-adjacent property units, depending on the variability of the study area and the distance between the property units, increase the risk of associating individuals with conditions they were not exposed to.

- Geographic units belonging to the geographic level of parishes are too big to efectively represent the true conditions every individual was exposed to and face the scale effect of MAUP.

References

Bengtsson, T., & Dribe, M. (2011). "The late emergence of socioeconomic mortality differentials: A micro-level study of adult mortality in southern Sweden 1815-1968," Explorations in Economic History, Elsevier, vol. 48(3), pages 389-400, July.

Bengtsson, T., & Broström, G. (2010). Mortality Crises in Rural Southern Sweden 1766–1860. Demographic Responses to Economic and Environmental Crises, 1-16.

Bengtsson, T., & van Poppel, F. (2011). Socioeconomic inequalities in death from past to present: An introduction. Explorations in Economic History 48 (2011) 343-356.

Bengtsson, T., Dribe, M., Quaranta, L., and Svensson, P. (2014). The Scanian Economic Demographic Database, Version 4.0 (Machine-readable database).

Chaix, B., Merlo, J., and Chauvin, J. (2005). Comparison of a spatial approach with the multilevel approach for investigating place effects on health: The example of healthcare utilization in France. Journal of Epidemiology and Community Health 59: 517-526. doi:10.1136/jech.2004.025478.

Chaix, B., Merlo, J., Subramanian, S.V., Lynch, J., and Chauvin, P. (2005). Comparison of a spatial perspective with a multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. American Journal of Epidemiology 162(2): 171-182. doi:10.1093/aje/kwi175.

Chen, V.Y-J., Wu, P-C., Yang, T-C., and Su H-J. (2010). Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. Science of the Total Environment 408(9): 2042–2049. doi:10.1016/ j.scitotenv.2009.11.044.

Cox, D. R. (1972). Regression Models and Life Tables. Journal of the Royal Statistical Society. B, 34, 187-200.

DeBats, D.A. (2008). A tale of two cities: Using tax records to develop GIS files for mapping and understanding nineteenth-century U.S. cities. Historical Methods, 41(1), 17-38.

Demography, n.” Oxford Dictionaries Online. December 2014. Oxford University Press. 2 March 2015 < http://www.oxforddictionaries.com/definition/english/demography />.

DotSpatial. https://dotspatial.codeplex.com/, assessed September 15, 2015.

Dribe, M., & Olsson, M. (2003). Spelade äganderätten någon roll? Om friköp av frälsejord i Skåne under 1800-talet. In Swedish Economic History Conference, Lund, October.

Fei L. (2015). “Compute a Crowdedness Index on Historical GIS Data- A Case Study of Hög Parish, Sweden, 1812-1920”. Master Thesis 2015. Lund University. Sweden.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. (2002). Geographically Weighted Regression: The analysis of spatially varying relationships. Chichester, UK: John Wiley & Sons.

Garrett, E., Reid, A., Schürer, K., Szreter, S. (2001). Changing Family Size in England and Wales: Place, Class and Demography, 1891–1911. Cambridge UniversityPress, Cambridge.

Gutmann, M., Deane, G., Lauster, N., Peri, A. (2005). Population and Environment, Vol. 27, No. 2, Two Population-Environment Regimes in the Great Plains of the United States, 1930-1990, November 2005 ©2006 Springer Science+Business Media, Inc. DOI: 10.1007/s11111-006-0016-3.

Hall, P. (2011). "Big Data: Large-Scale Historical Infrastructure from the Minnesota Population Center," Historical Methods Special Issue, Part 1 (Vol. 44, No. 1), and Part 2 (Vol. 44, No. 2).

Harris, R., Sleight, P., Webber, R. (2005). Geodemographics, GIS and Neighbourhood Targeting, e-book, accessed 27 February 2015, <http://LUND.eblib.com/patron/FullRecord.aspx?p=242963>.

Hedefalk, F., (2014). Life histories across time and space: methods for including geographic factors on the micro-level in longitudinal demographic research, In Dissertations from Department of Physical Geography and Ecosystem Science, University of Lund. (link: http://lup.lub.lu.se/record/4732650, visited 27-4-2015).

Hedefalk, F., Harrie, L., Svensson, P. (2015). Methods to Create a Longitudinal Integrated Demographic and Geographic Database on the Micro-Level: A Case Study of Five Swedish Rural Parishes, 1813–1914. Historical Methods: A Journal of Quantitative and Interdisciplinary History 48(3): 153-173, Routledge.

Historical Demography. Answers.com. Encyclopedia of Public Health, The Gale Group, Inc, 2002. http://www.answers.com/topic/historical-demography, accessed March 02, 2015.

Işik, O., and Pinarcioğlu, M.M. (2006). Geographies of a silent transition: A geographically weighted regression approach to regional fertility differences in Turkey. European Journal of Population/Revue européenne de Démographie 22(4): 399-421.

Jiang, B., Yao, X., and Wheeler, D.C. (2010). Visualizing and diagnosing coefficients from geographically weighted regression models. In: Sui, D.Z., Tietze, W., Claval, P., Gradus, Y., Park, S.O., and Wusten, H. (eds.). Geospatial analysis and modelling of urban structure and dynamics. Netherlands: Springer: 415-436. doi:10.1007/978-90-481-8572-6.

Johannesson, G. (1984). Skåne, Halland och Blekinge. Om Skånelandskapens historia. Stockholm. Sweden.

Johnson, K.M., Voss, P.R., Hammer, R.B., Fuguitt, G.V., and Mcniven, S. (2005). Temporal and spatial variation in age-specific net migration in the United States. Demography 42(4): 791-812. doi:10.1353/dem.2005.0033.

Kwan, M.-P. (2012a) The uncertain geographic context problem. Annals of the Association of American Geographers, 102(5): 958-968.

Kwan, M.-P. (2012b) How GIS can help address the uncertain geographic context problem in social science research. Annals of GIS, 18(4): 245-255.

Kävlinge Municipality. http://www.kavlinge.se/kommunochpolitik/kommunfakta/historia/fran13smasocknartillenkommun.4.3c1e3a3e13cce78e3f44b2e.html, assessed September 18, 2015.

Lewan, N 1989, *Bebyggelse Och Landskap : En Översikt*, n.p.: Lund : Inst. för kulturgeografi och ekonomisk geografi,Lunds univ., 1989 ; (Lund : Lunds univ. repro), Library catalogue (Lovisa), EBSCO*host*, viewed 22 July 2015.

Long, L. H. (1988). Migration and residential mobility in the United States. New York: Russell Sage Foundation.

Longley, Paul; Goodchild, Michael; Maguire, David; Rhind, David 2011, Geographic Information Systems & Science 3rd ed., ISBN 978-0-470-72144-5, John Wiley & Sons, Inc.

Matthews, S. A., & Parker, D. M. (2013). Progress in Spatial Demography. Demographic Research Volume 28, Article 10, p. 271-312.

Mennis, J., & Liu, J. W. (2005). Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change. Transactions in GIS, 9(1), 5-17.

Microdata. Minnesota Population Center. . IPUMS International (Integrated Public Use Microdata Series, International). University of Minnesota. Retrieved April 16, 2015, from https://international.ipums.org/international-action/faq#ques7

Minnesota Population Center. North Atlantic Population Project: Complete Count Microdata. 2008, Minneapolis: University of Minnesota. Retrieved May 4, 2015, from https://www.nappdata.org/napp/intro.shtml

Minnesota Population Center. Terra Populus. 2013, Minneapolis: University of Minnesota. Retrieved May 4, 2015, from http://www.terrapop.org/

NOAA - National Oceanic and Atmospheric Administration of the United States department of Commerce. 'What is Lidar'. Retrieved May 30th, 2015 from http://oceanservice.noaa.gov/facts/lidar.html

Openshaw, S. (1984). The Modifiable Areal Unit Problem. Norwich, UK: Geobooks.

Pezeshki, Z., Tafazzoli-Shadpour, M., Mansourian, A., Eshrati, B., Omidi, E., & Nejadqoli, I. (2012). Model of cholera dissemination using geographic information systems and fuzzy clustering means: Case study, Chabahar, Iran. Public Health (Elsevier), 126(10), 881-887 7p. doi:10.1016/j.puhe.2012.07.002.

Reid, A. (1997). Locality or class? Spatial and social differentials in infant and child mortality in England and Wales, 1895–1911. In: Corsini, C., Viazzo, P.-P. (Eds.), The Decline of Infant and Child Mortality: The European Experience. Martinus Nijhoff Publishers, Dordrecht, pp. 129–154.

Richards S.J. (2008). Applying Survival Models to Pensioner Mortality Data. British Actuarial Journal, 14, pp 257-303. DOI: 10.1017/S1357321700001720.

Ruggles, S., and Menard, R. (1995). Introduction to the Minnesota historical census projects. Historical Methods 28 (1): 6–10.

Sarkar, S., and Hall, P., K. (2011). "Mapping NAPP: Examples of Spatial Analysis of Nineteenth-Century Migration in the North Atlantic Countries Using NAPP Data." Historical Methods: A Journal of Quantitative and Interdisciplinary History 44.1: 25-36.

Seng S.S. Chong A.K. Moore A. (2005). Geostatistical modelling, analysis and mapping of epidemiology of dengue fever in Johor State, Malaysia. The 17th annual colloquium of the spatial information research centre. Dunedin, New Zealand: University of Otago. p. 15.

Skansjö, S. (1997). Skånes historia. Lund. Sweden.

Smith, D.S., 1983. Differential mortality in the United States before 1900. Journal of Interdisciplinary History 13, 735–759.

Sobek, M., Cleveland, L., Flood, S., Kelly Hall, P., King, M. L., Ruggles, S., & Schroeder, M. (2011). Big data: large-scale historical infrastructure from the Minnesota Population Center. Historical methods, 44(2), 61-68.

Spatial Demography. http://spatialdemography.org/, assessed March 02, 2015.

Spielman, S., E., Logan, J., R. (2013). Using high-resolution population data to identify neighborhoods and establish their boundaries. Annals of the Association of American Geographers 103: 67–84 doi: 10.1080/00045608.2012.68504923279975.

Van Poppel, F., Jonker, M., Mandemakers, K. (2005). Differential infant and child mortality in three Dutch regions, 1812–1909. Economic History Review 58,272–309.

Villarreal, C., Bettenhausen, B., Hanss, E., Hersh, J. (2014). Historical Health Conditions in Major U.S. Cities, Historical Methods: A Journal of Quantitative and Interdisciplinary History, 47:2, 67:80, DOI: 10.1080/01615440.2013.874005.

Voss, P. R. (2007). Demography as a spatial social science. Population Research and Policy Review, 26(5-6), 457-476.

Ward (electoral subdivision). Wikipedia. Retrieved March 12, 2015 from https://en.wikipedia.org/wiki/Ward_(electoral_subdivision)

Woods, R., Williams, N., Galley, C. (1993). Infant mortality in England 1550–1950. Problems in identification of long-term trends and geographical and social variations. In: Corsini, C.A., Viazzo, P.-P. (Eds.), The Decline of Infant Mortality in Europe 1800–1950. Four National Case Studies. UNICEF and Instituto degli Innocenti di Firenze, Florence, pp. 35–50.

Worboys, M.., Duckham, M. (2004). GIS: A Computing Perspective 2nd ed., ISBN 0-415-28375-2, CRC Press LLC.

Appendix A

Vocabulary

ArcMap is the main component of Esri's ArcGIS suite of geospatial processing programs, and is used primarily to view, edit, create, and analyze geospatial data. ArcMap allows the user to explore data within a data set, symbolize features accordingly, and create maps.

DotSpatial DotSpatial is a geographic information system library written for the .NET 4 Framework. It allows developers to incorporate spatial data, analysis and mapping functionality into their applications or to contribute GIS extensions to the community (DotSpatial, 2015).

Dynamic Variable a variable whose value changes as time passes.

LiDAR a remote sensing technology used to examine the surface of the earth by measuring distance by illuminating a target with a laser and analyzing the reflected light. This technology is used to make high-resolution maps in a wide range of application areas such as geomatics, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing, atmospheric physics, laser altimetry, and contour mapping (NOAA 2015).

Macro Level Analysis Analysis performed at an aggregated level (e.g. building, property unit or parish).

Microdata In the study of survey and census data, microdata is information at the level of individual respondents. For instance, a national census might collect age, home address, educational level, employment status, and many other variables, recorded separately for every person who responds; this is microdata (IPUMS International, 2015).

Micro Level Analysis Analysis performed at the lowest level of integration (e.g. individual level).

Microsoft SQL Server is a Relational DataBase Management System (RDBMS) developed by Microsoft. As a RDBMS, its primary function is to store and retrieve data as requested by other software applications which may run either on the same computer or on another computer across a network.

Network Distance Network distance generally refers to the distance between a node A (e.g. building A) and a node B (e.g. building B) which are directly or indirectly connected within a network (e.g. road network).

Static variable a variable whose value remains the same (static) as time passes.

SQL is an acronym for Structured Query Language. SQL is a special-purpose programming language, based on the principles of relational algebra, which was specifically designed to manage data in a Relational Database Management System.

Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for the Microsoft Windows operative system, as well as for web sites, web applications and web services.

Ward a subdivision of a local authority area, typically used for electoral purposes. Wards are usually named after neighborhoods, thoroughfares, parishes, landmarks, geographical features and in some cases historical figures connected to the area. It is common in the United States for wards to simple be numbered. In Australia, Canada, Monaco, New Zealand, South Africa, United Kingdom, United States of America and Zimbabwe they are electoral districts within a district or municipality, used in local government elections. In the United States, wards are usually subdivided into precincts for polling purposes (Wikipedia, 2015).

B. Code Snippets

B.1. SQL Queries

This section of the appendix presents the SQL code used to compute the project context variables. It is divided in two parts. The first part is dedicated to the SQL code used to declare the geometries later used in the spatial computation queries. The second part contains all the SQL queries deployed for the computation of the geographic context variables.

B.1.1. Declare Geometries

```
DECLARE @geoSoilTypes geometry;
DECLARE @geoPropertyUnits geometry;
DECLARE @geoSoilTypes_Hog geometry;

SELECT @geoPropertyUnits=Shape FROM dbo.PropertyUnit_Hog;
SELECT @geoSoilTypes=Shape FROM dbo.SOILTYPES;
SELECT @geoSoilTypes_Hog=Shape FROM dbo.SOILTYPES_HOG;
```

B.1.2. SQL Queries – Calculation of the Soil-Type Context Variable

This part includes the SQL code used to obtain soil type information of physical locations on different aggregation levels (e.g. property unit, initial property unit, address and parish). The calculation process is performed in three logical steps. In the first step, the physical location units are intersected with the soil types. In the second step, soil type areas belonging to the same soil type and included in the same physical location unit boundaries are being unified. The total area (in m^2) of each soil type within the boundaries of every physical location unit are also being calculated in this step. The results of the final step show the physical location unit with its corresponding soil types, the area covered by each soil type (in m^2) as well as the proportion (%) of each physical location unit area covered by every soil type.

B.1.2.1. Calculation of the Soil-Type Context Variable (Property Unit Level)

```
-----------------------------------------------------------------------------------------------------
--                                                                                                 --
------Calculate the Soil-Type Geographic Context Variable for all parishes (property-unit level)-----
--                                                                                                 --
-----------------------------------------------------------------------------------------------------

--
--Perform an Intersection over the property-unit layer and the soil-type layer:
--
WITH PU_Interesection_ST (PropertyUnit_ID, PropertyUnit_Geom, SoilType, PropertyUnit_SoilType_Area,
PropertyUnit_SoilType_Geom) AS
(
        SELECT  dbo.PropertyUnits_All.uniqueIdObj AS PropertyUnit_ID,
                dbo.PropertyUnits_All.Shape AS PropertyUnit_Geom,
                dbo.SOILTYPES.JG2_TX AS SoilType,
                PropertyUnits_All.Shape.STIntersection(SOILTYPES.Shape).STArea() AS
                PropertyUnit_SoilType_Area,
                PropertyUnits_All.Shape.STIntersection(SOILTYPES.Shape) AS PropertyUnit_SoilType_Geom

                FROM    dbo.PropertyUnits_All, dbo.SOILTYPES

                WHERE   PropertyUnits_All.Parish!='Kågeröd' AND
                        PropertyUnits_All.Initial_Property_Unit IS NOT NULL AND
                        PropertyUnits_All.Shape.STIntersects(SOILTYPES.Shape)=1
)

--
--Perform a union over the various soil-type geometries produced from the intersection of the
--soil-type layer with the property-unit layer, so that all soil-type geometries having
--the same soil-type, within an property-unit, are merged.
--
, ST_UNION_PER_PU (PropertyUnit_ID, SoilType, Soil_Area, Total_SoilType_Geom) AS
(
        SELECT  PropertyUnit_ID, SoilType, ROUND(SUM(PropertyUnit_SoilType_Area),0) AS Soil_Area,
                geometry::UnionAggregate(PropertyUnit_SoilType_Geom) AS Total_soil_type_geom

        FROM            PU_Interesection_ST

        GROUP BY        PropertyUnit_ID, SoilType
)

--
--Select the parish, property-unit id, start date, end date, soil type, soil type area and
--soil type percentage for every property unit in the study area:
--
SELECT  Parish, PropertyUnit_ID, startDate, endDate,
        SoilType, Soil_Area AS SoilType_Area,
        ROUND(Soil_Area*100/Shape.STArea(), 2) AS ST_PercentageOfPU,
        Total_SoilType_Geom, PropertyUnits_All.Shape AS PU_Geom

FROM    dbo.PropertyUnits_All, ST_UNION_PER_PU

WHERE   dbo.PropertyUnits_All.uniqueIdObj=ST_UNION_PER_PU.PropertyUnit_ID

ORDER BYParish, PropertyUnit_ID, Soil_Area DESC;
```

B.1.2.2 Calculation of the Soil-Type Context Variable (Initial Property Unit Level)

```
-----------------------------------------------------------------------------------------------------
--                                                                                                 --
---Calculate Soil-Types Geographic Context Variable for all parishes (initial-property-unit level)---
--                                                                                                 --
-----------------------------------------------------------------------------------------------------

--
--Perform an Intersection over the initial-property-unit layer and the soil-type layer:
--
WITH IPU_Interesection_ST (InitialPropertyUnit_ID, InitialPropertyUnit_Geom, SoilType,
InitialPropertyUnit_SoilType_Area, InitialPropertyUnit_SoilType_Geom)     AS
(
SELECT  dbo.Initial_Property_Unit.uniqueIdObj AS InitialPropertyUnit_ID,
        dbo.Initial_Property_Unit.Shape AS InitialPropertyUnit_Geom,
        dbo.SOILTYPES.JG2_TX AS SoilType,
        Initial_Property_Unit.Shape.STIntersection(SOILTYPES.Shape).STArea() AS
        InitialPropertyUnit_SoilType_Area,
        Initial_Property_Unit.Shape.STIntersection(SOILTYPES.Shape) AS
        InitialPropertyUnit_SoilType_Geom

FROM    dbo.Initial_Property_Unit, dbo.SOILTYPES

WHERE   Initial_Property_Unit.Shape.STIntersects(SOILTYPES.Shape)=1
)

--
--Perform a union over the various soil-type geometries produced from the intersection of the
--soil-type layer with the initial-property-unit layer, so that all soil-type geometries having
--the same soil-type, within an initial-property-unit, are merged.
--
, ST_UNION_PER_PU (InitialPropertyUnit_ID, SoilType, Soil_Area, Total_SoilType_Geom) AS
(
SELECT  InitialPropertyUnit_ID, SoilType,
        ROUND(SUM(InitialPropertyUnit_SoilType_Area),1) AS Soil_Area,
        geometry::UnionAggregate(InitialPropertyUnit_SoilType_Geom) AS Total_soil_type_geom

FROM            IPU_Interesection_ST
GROUP BY        InitialPropertyUnit_ID, SoilType
)

--
--Select the parish, initial-property-unit id, start date, end date, soil type, soil type area and
--soil type percentage for every initial property unit in the study area:
--
SELECT  Parish, InitialPropertyUnit_ID, startDate, endDate,
        SoilType, Soil_Area AS SoilType_Area,
        ROUND(Soil_Area*100/Shape.STArea(), 1) AS ST_PercentageOfPU,
        Total_SoilType_Geom, Initial_Property_Unit.Shape AS IPU_Geom

FROM    dbo.Initial_Property_Unit, ST_UNION_PER_PU

WHERE   dbo.Initial_Property_Unit.uniqueIdObj=ST_UNION_PER_PU.InitialPropertyUnit_ID

ORDER BYParish, InitialPropertyUnit_ID, Soil_Area DESC;
```

B.1.2.3 Calculation of the Soil-Type Context Variable (Address Level)

```
-----------------------------------------------------------------------------------------------------
--                                                                                               --
---------Calculate the Soil-Type Geographic Context Variable for all parishes (address level)--------
--                                       --All Parishes--                                       --
-----------------------------------------------------------------------------------------------------

--
--Perform a geometric aggregation of the property units on an address level:
--
WITH AddressInfo (Parish, ID, Geom) AS
(
        SELECT  Parish, addressNameCode, geometry::UnionAggregate(Shape) AS Aggr_IPU_Geom

        FROM    dbo.PropertyUnits_All

        --Except property units from Kågeröd(not completed) and property units with no registered
        --value for the addressNameCode-field from the calculation.
        WHERE   Parish!='Kågeröd' AND
                addressNameCode!='<Null>' AND
                addressNameCode!='' AND addressNameCode IS NOT NULL

        GROUP BY Parish, addressNameCode
)

--
--Perform an Intersection over the address layer and the soil-type layer:
--
, AddrGeom_Interesection_ST (Parish, Address_ID, Address_Geom, SoilType, Address_SoilType_Area,
Address_SoilType_Geom)   AS
(
        SELECT  Parish, AddressInfo.ID AS Address_ID,
                AddressInfo.Geom AS Address_Geom,
                dbo.SOILTYPES.JG2_TX AS SoilType,
                AddressInfo.Geom.STIntersection(SOILTYPES.Shape).STArea() AS Address_SoilType_Area,
                AddressInfo.Geom.STIntersection(SOILTYPES.Shape) AS Address_SoilType_Geom

        FROM            AddressInfo, dbo.SOILTYPES
        WHERE           AddressInfo.Geom.STIntersects(SOILTYPES.Shape)=1
)

--
--Perform a union over the various soil-type geometries produced from the intersection of the
--soil-type layer with the aggregated-address layer, so that all soil-type geometries having
--the same soil-type, within the same address, are merged.
--
, ST_UNION_PER_Address (Parish, Address_ID, SoilType, Soil_Area, Total_soil_type_geom) AS
(
        SELECT  Parish, Address_ID, SoilType, SUM(Address_SoilType_Area) AS Soil_Area,
                geometry::UnionAggregate(Address_SoilType_Geom) AS Total_soil_type_geom
        FROM            AddrGeom_Interesection_ST
        GROUP BY        Parish, Address_ID, SoilType
)

--
--Select the parish, address id, soil type, soil type area and soil type percentage for every
--address in the study area:
--
SELECT  ST_UNION_PER_Address.Parish, AddressInfo.ID AS AddressName, SoilType, Soil_Area AS SoilType_Area,
                ROUND(Soil_Area*100/AddressInfo.Geom.STArea(), 2) AS ST_PercentageOfAddress,
                AddressInfo.Geom AS Address_Geom,
                ST_UNION_PER_Address.Total_soil_type_geom AS Total_SoilType_Geom

--INTO  SoilTypes_Address_All_Parishes

FROM    AddressInfo INNER JOIN ST_UNION_PER_Address
                ON      AddressInfo.ID=ST_UNION_PER_Address.Address_ID AND
                AddressInfo.Parish=ST_UNION_PER_Address.Parish

ORDER BYST_UNION_PER_Address.Parish, AddressInfo.ID, Soil_Area DESC;
```

B.1.2.4 Calculation of the Soil-Type Context Variable (Parish Level)

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
------------------------Calculate Soil Types Context Variable (Parish Level)-------------------
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Perform an Intersection between the Soil Type layer and the Parish boundaries:
--
WITH  ST_Intersection_P (Parish, Parish_Geom, SoilType, Parish_SoilType_Area,
Parish_SoilType_Geom) AS
(
       SELECT Name AS Parish, Parish.Geom AS Parish_Geom, SOILTYPES.JG2_TX AS SoilType,
              Parish.Geom.STIntersection(SOILTYPES.Shape).STArea() AS Parish_SoilType_Area,
              Parish.Geom.STIntersection(SOILTYPES.Shape) AS Parish_SoilType_Geom

       FROM   Parish, SOILTYPES

       WHERE  Parish.Name!='Kågeröd' AND Parish.Geom.STIntersects(SOILTYPES.Shape)=1
)

--
--Compute the union of the soil type geometries within the parish boundaries,that belong to the
same soil type:
--
, ST_Union_Per_P (Parish, SoilType, SoilType_TotalArea, SoilType_TotalGeom) AS
(
       SELECT Parish, SoilType, SUM(Parish_SoilType_Area) AS SoilType_TotalArea,
              geometry::UnionAggregate(Parish_SoilType_Geom) AS SoilType_TotalGeom

       FROM          ST_Intersection_P
       GROUP BY      Parish, SoilType
)

--
--List the soil types, their area and % percentage of the parish area they cover:
--
SELECT        Parish.Name, SoilType, ROUND(SoilType_TotalArea, 1) AS SoilType_Area,
              ROUND(SoilType_TotalArea*100/Parish.Geom.STArea(), 2) AS ST_PercentageOfP,
              SoilType_TotalGeom, Parish.Geom AS Parish_Geom

FROM          Parish INNER JOIN ST_UNION_PER_P ON Parish.Name=ST_UNION_PER_P.Parish

ORDER BY      Parish.Name, ST_PercentageOfP DESC;
```

B.1.3 SQL Queries – Calculation of the Proximity-To-Water Context Variable

This section of the appendix contains the SQL code used to calculate the distance between physical locations (e.g. property units, initial property units, addresses and parishes) and water sources. The proximity to water is ultimately expressed as the shortest distance between a geographical unit and a lake or stream. More in particular, the proximity to water is calculated as the shortest Euclidean distance between the centroid of the geographical unit polygon and the nearest point in the border of the nearest lake polygon or the nearest point on the nearest river polyline. The proximity to water is also calculated as the shortest Euclidean distance between the boarder of the geographical unit polygon and the nearest point in the boarder of the nearest lake polygon or the nearest point on the nearest river polyline.The results include the shortest distance to a water source in meters per geographical unit.

B.1.3.1.a Calculation of the Proximity-To-Water Context Variable (Property Unit Level) with the "Centroid to Border" Calculation Method

```
-----------------------------------------------------------------------------------------------------------
--                                                                                                       --
--------------Calculate the "proximity-to-Water" context variable (property-unit level) ---------------
--                                       (Centroid To Border)                                            --
--                                                                                                       --
--                                         --All Parishes--                                              --
--                                                                                                       --
-----------------------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the property-unit-polygon centroid and
--the points of the stream-polyline:
--
WITH MinDistanceToStream (Parish, PropertyUnit_ID, Distance_To_Stream) AS
(
        SELECT   PropertyUnits_All.Parish, uniqueIdObj AS PropertyUnit_ID,
                 ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Distance_To_Stream

        FROM     PropertyUnits_All, dbo.MTSM_1812_1820_STREAM

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND
                 PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the property-unit-polygon centroid and the
--points of the lake-polygon border:
--
, MinDistanceToLake (Parish, PropertyUnit_ID, Distance_To_Lake) AS
(
        SELECT  PropertyUnits_All.Parish, uniqueIdObj AS PropertyUnit_ID,
                ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                AS Distance_To_Lake

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_LAKE

        WHERE   PropertyUnits_All.Parish!='Kågeröd' AND
                PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between a property unit and
--a lake or stream:
```

```
--
, MinDistanceToWater (Parish, PU_ID, MinDistanceToWater) AS
(
        SELECT  MinDistanceToStream.Parish, MinDistanceToStream.PropertyUnit_ID,
        CASE
                WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                        THEN MIN(Distance_To_Stream)
                ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

        FROM    MinDistanceToStream, MinDistanceToLake
        WHERE   MinDistanceToStream.PropertyUnit_ID=MinDistanceToLake.PropertyUnit_ID

        GROUP BY MinDistanceToStream.Parish, MinDistanceToStream.PropertyUnit_ID
)

--
--Select info regarding the parish, id, geometry and min distance to water for every
--property unit:
--
SELECT  MinDistanceToWater.Parish AS Parish,
        MinDistanceToWater.PU_ID AS PropertyUnit_ID,
        MinDistanceToWater, PropertyUnits_All.Shape AS PU_Shape

FROM    PropertyUnits_All INNER JOIN MinDistanceToWater
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToWater.PU_ID

ORDER BY MinDistanceToWater.Parish, MinDistanceToWater.PU_ID;
```

B.1.3.1.b Calculation of the Proximity-To-Water Context Variable (Property Unit Level) with the "Border to Border" Calculation Method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
-----------Calculate the "proximity-to-Water" context variable (property-unit level)----------
--                                     (Border to Border)                                    --
--                                                                                           --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the property-unit-polygon border and
--the points of the stream-polyline:
--
WITH MinDistanceToStream (PropertyUnit_ID, Distance_To_Stream) AS
(
        SELECT   uniqueIdObj AS PropertyUnit_ID,
                 ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Distance_To_Stream

        FROM     PropertyUnits_All, dbo.MTSM_1812_1820_STREAM

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND
                 PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the property-unit-polygon border and the
--points of the lake-polygon border:
--
, MinDistanceToLake (PropertyUnit_ID, Distance_To_Lake) AS
(
        SELECT  uniqueIdObj AS PropertyUnit_ID,
                ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                AS Distance_To_Lake

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_LAKE

        WHERE   PropertyUnits_All.Parish!='Kågeröd' AND
                PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between a property unit and
--a lake or stream:
--
, MinDistanceToWater (PU_ID, MinDistanceToWater) AS
(
        SELECT  MinDistanceToLake.PropertyUnit_ID AS PropertyUnit_ID,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                        ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

        FROM    MinDistanceToStream INNER JOIN MinDistanceToLake
                ON MinDistanceToStream.PropertyUnit_ID=MinDistanceToLake.PropertyUnit_ID

        GROUP BY MinDistanceToLake.PropertyUnit_ID
)

--
--Select info regarding the parish, id, geometry and min distance to water for every
--property unit:
--
SELECT  PropertyUnits_All.Parish AS Parish,
        PropertyUnits_All.uniqueIdObj AS PropertyUnit_ID,
```

```
        MinDistanceToWater,
        PropertyUnits_All.Shape AS PropertyUnit_Geom

FROM    PropertyUnits_All INNER JOIN MinDistanceToWater
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToWater.PU_ID

ORDER BY PropertyUnits_All.Parish, PropertyUnits_All.uniqueIdObj;
```

B.1.3.2.a Calculation of the Proximity-To-Water Context Variable (Initial Property Unit Level) with the "Centroid to Border" calculation method

```
-------------------------------------------------------------------------------------------
--                                                                                       --
------Calculate the "proximity-to-Water" context variable (initial-property-unit level):------
--                                  (Centroid to Border)                                 --
--                                     --All Parishes--                                  --
--                                                                                       --
-------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the initial-property-unit-polygon centroid
--and the points of the stream-polyline:
--
WITH MinDistanceToStream (IPU_ID, Distance_To_Stream) AS
(
        SELECT  Initial_Property_Unit.uniqueIdObj AS IPU_ID,
                ROUND(MIN(Initial_Property_Unit.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                AS Distance_To_Stream

        FROM     dbo.Initial_Property_Unit, dbo.MTSM_1812_1820_STREAM

        GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the initial-property-unit-polygon centroid and
--the points of the lake-polygon border:
--
, MinDistanceToLake (IPU_ID, Distance_To_Lake) AS
(
        SELECT   Initial_Property_Unit.uniqueIdObj AS IPU_ID,
                 ROUND(MIN(Initial_Property_Unit.Centroid.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                 AS Distance_To_Lake

        FROM     dbo.Initial_Property_Unit, dbo.MTSM_1812_1820_LAKE

        GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between an initial-property-unit
--and a lake or stream:
--
, MinDistanceToWater (Parish, IPU_ID, MinDistanceToWater) AS
(
        SELECT  Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                        ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

        FROM    dbo.Initial_Property_Unit INNER JOIN MinDistanceToStream
                ON Initial_Property_Unit.uniqueIdObj=MinDistanceToStream.IPU_ID
                INNER JOIN MinDistanceToLake
                ON Initial_Property_Unit.uniqueIdObj=MinDistanceToLake.IPU_ID

        GROUP BY Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj
)
--
--Select parish, initial-property-unit id, min distance to water & geometry for every
--initial property unit:
--
SELECT  Initial_Property_Unit.Parish AS Parish,
        Initial_Property_Unit.uniqueIdObj AS IPU_ID,
        MinDistanceToWater, Initial_Property_Unit.Shape AS IPU_Geom

FROM    Initial_Property_Unit INNER JOIN MinDistanceToWater
        ON Initial_Property_Unit.uniqueIdObj=MinDistanceToWater.IPU_ID

ORDER BY Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj;
```

B.1.3.2.b Calculation of the Proximity-To-Water Context Variable (Initial Property Unit Level) with the "Border to Border" calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
------Calculate the "proximity-to-Water" context variable (initial-property-unit level):------
--                                   (Border to Border)                                      --
--                                     --All Parishes--                                      --
--                                                                                           --
-----------------------------------------------------------------------------------------------
--
--Calculate the shortest distance (geometry) between the initial-property-unit-polygon border
--and the points of the stream-polyline:
--
WITH MinDistanceToStream (IPU_ID, Distance_To_Stream) AS
(
        SELECT  Initial_Property_Unit.uniqueIdObj,
                ROUND(MIN(Initial_Property_Unit.Shape.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                AS Distance_To_Stream

        FROM    Initial_Property_Unit, dbo.MTSM_1812_1820_STREAM

        GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the initial-property-unit-polygon border and
--the points of the lake-polygon border:
--
, MinDistanceToLake (IPU_ID, Distance_To_Lake) AS
(
        SELECT   Initial_Property_Unit.uniqueIdObj,
                 ROUND(MIN(Initial_Property_Unit.Shape.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                 AS Distance_To_Lake

        FROM     Initial_Property_Unit, dbo.MTSM_1812_1820_LAKE

        GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between an initial-property-unit
--polygon boarder and a lake or stream:
--
, MinDistanceToWater (Parish, IPU_ID, MinDistanceToWater) AS
(
        SELECT  Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                        ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

        FROM    Initial_Property_Unit INNER JOIN MinDistanceToStream
                ON Initial_Property_Unit.uniqueIdObj=MinDistanceToStream.IPU_ID
                INNER JOIN MinDistanceToLake
                ON Initial_Property_Unit.uniqueIdObj=MinDistanceToLake.IPU_ID

        GROUP BY Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj
)

--
--Select parish, initial-property-unit id, min distance to water & geometry for every
--initial property unit:
--
SELECT  Initial_Property_Unit.Parish AS Parish,
        Initial_Property_Unit.uniqueIdObj AS IPU_ID,
        MinDistanceToWater, Initial_Property_Unit.Shape AS IPU_Geom

FROM    Initial_Property_Unit INNER JOIN MinDistanceToWater
                ON Initial_Property_Unit.uniqueIdObj=MinDistanceToWater.IPU_ID

ORDER BY Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj;
```

B.1.3.3.a Calculation of the Proximity-To-Water Context Variable (Address Level) with the "Centroid to Border" calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
-------------Calculate the "proximity-to-Water" context variable (address level):-------------
--                                   (Centroid to Border)                           --
--                                      --All Parishes--                                    --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the property-unit-polygon centroid and
--the points of the stream-polyline:
--
WITH MinDistanceToStream (Parish, PropertyUnit_ID, Distance_To_Stream, Max_Distance_To_Stream) AS
(
        SELECT   PropertyUnits_All.Parish, uniqueIdObj AS PropertyUnit_ID,
                 ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Distance_To_Stream,
                 ROUND(MAX(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Max_Distance_To_Stream

        FROM     PropertyUnits_All, dbo.MTSM_1812_1820_STREAM

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the property-unit-polygon centroid and the
--points of the lake-polygon periphery:
--
, MinDistanceToLake (Parish, PropertyUnit_ID, Distance_To_Lake) AS
(
        SELECT  PropertyUnits_All.Parish, uniqueIdObj AS PropertyUnit_ID,
                ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                AS Distance_To_Lake

        FROM   PropertyUnits_All, dbo.MTSM_1812_1820_LAKE

        WHERE PropertyUnits_All.Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between a property unit and
--a lake or stream:
--
, MinDistanceToWater (Parish, PU_ID, MinDistanceToWater) AS
(
        SELECT  MinDistanceToStream.Parish, MinDistanceToStream.PropertyUnit_ID,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                        ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

FROM    MinDistanceToStream, MinDistanceToLake
WHERE   MinDistanceToStream.PropertyUnit_ID=MinDistanceToLake.PropertyUnit_ID

GROUP BY MinDistanceToStream.Parish, MinDistanceToStream.PropertyUnit_ID
)

--
--Compute the min distance to water for every address as the average min distance
--to water of the property units they consist of.
--Display the results by showing the parish, address name,
--min distance to water (m) and address geometry for every address:
--
```

```sql
SELECT  MinDistanceToWater.Parish AS Parish, AddressNameCode AS Address_ID,
        ROUND(AVG(MinDistanceToWater), 1) AS MinDistanceToWater,
        geometry::UnionAggregate(PropertyUnits_All.Shape) AS Address_Shape

FROM    PropertyUnits_All INNER JOIN MinDistanceToWater
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToWater.PU_ID

WHERE   addressNameCode!='' AND addressNameCode!='<Null>' AND
        addressNameCode IS NOT NULL

GROUP BY MinDistanceToWater.Parish, AddressNameCode

ORDER BY MinDistanceToWater.Parish, AddressNameCode;
```

B.1.3.3.b Calculation of the Proximity-To-Water Context Variable (Address Level) with the "Border to Border" calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
-------------Calculate the "proximity-to-Water" context variable (address level):-------------
--                                    (Border to Border)                                    --
--                                      --All Parishes--                                    --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the property-unit-polygon border and
--the points of the stream-polyline:
--
WITH MinDistanceToStream (PropertyUnit_ID, Distance_To_Stream, Max_Distance_To_Stream) AS
(
        SELECT   uniqueIdObj AS PropertyUnit_ID,
                 ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Distance_To_Stream,
                 ROUND(MAX(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Max_Distance_To_Stream

        FROM     PropertyUnits_All, dbo.MTSM_1812_1820_STREAM

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY uniqueIdObj
)

--
--Calculate the shortest distance (geometry) between the property-unit-polygon border and the
--points of the lake-polygon border:
--
, MinDistanceToLake (PropertyUnit_ID, Distance_To_Lake) AS
(
        SELECT  uniqueIdObj AS PropertyUnit_ID,
                ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                AS Distance_To_Lake

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_LAKE

        WHERE   PropertyUnits_All.Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY uniqueIdObj
)

--
--Select the shortest distance to water, as the shortest distance between a property unit and
--a lake or stream:
--
, MinDistanceToWater (PU_ID, MinDistanceToWater) AS
(
SELECT  MinDistanceToLake.PropertyUnit_ID AS PropertyUnit_ID,
        CASE
                WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                        THEN MIN(Distance_To_Stream)
                ELSE MIN(Distance_To_Lake)
                END AS Min_DistanceToWater

FROM    MinDistanceToStream INNER JOIN MinDistanceToLake
        ON MinDistanceToStream.PropertyUnit_ID=MinDistanceToLake.PropertyUnit_ID

GROUP BY MinDistanceToLake.PropertyUnit_ID
)

--
--Select info regarding the parish, id, geometry and min distance to water for every
--property unit:
--
SELECT  PropertyUnits_All.Parish AS Parish,
        PropertyUnits_All.addressNameCode AS Address_ID,
        ROUND(AVG(MinDistanceToWater), 1) AS MinDistanceToWater,
```

```
        geometry::UnionAggregate(PropertyUnits_All.Shape) AS Address_Shape

FROM    PropertyUnits_All INNER JOIN MinDistanceToWater
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToWater.PU_ID

WHERE   addressNameCode!='' AND addressNameCode!='<Null>' AND
        addressNameCode IS NOT NULL

GROUP BY PropertyUnits_All.Parish, PropertyUnits_All.addressNameCode

ORDER BY PropertyUnits_All.Parish, PropertyUnits_All.AddressNameCode;
```

B.1.3.4.a Calculation of the Proximity-To-Water Context Variable (Parish Level) with the "Centroid to Border" calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
----Calculate the "proximity-to-Water" context variable (parish level - CentroidToBorder):---
--                                        --All Parishes--                                   --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the shortest distance (geometry) between the parish-polygon centroid and the points
--of the stream-polyline:
--
WITH MinDistanceToStream (P_ID, Distance_To_Stream, Max_Distance_To_Stream) AS
(
        SELECT   Parish.Name AS Parish_ID,
                 ROUND(MIN(Parish.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 2)
                 AS Distance_To_Stream,
                 ROUND(MAX(Parish.Centroid.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 2)
                 AS Max_Distance_To_Stream

        FROM     dbo.Parish, dbo.MTSM_1812_1820_STREAM

        WHERE   Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Calculate the shortest distance (geometry) between the parish-polygon centroid and the points
--of the lake-polygon border:
--
, MinDistanceToLake (P_ID, Distance_To_Lake) AS
(
        SELECT   Parish.Name AS Parish_ID,
                 ROUND(MIN(Parish.Centroid.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 2)
                 AS Distance_To_Lake

        FROM     dbo.Parish, dbo.MTSM_1812_1820_LAKE

        WHERE   Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Select the shortest distance to water, as the shortest distance between the parish-centroid
--and a lake or stream geometry:
--
, MinDistanceToWater (Parish_ID, Min_DistanceToWater) AS
(
        SELECT Parish.Name,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                         ELSE MIN(Distance_To_Lake)
                         END AS Min_DistanceToWater

        FROM    dbo.Parish, MinDistanceToStream, MinDistanceToLake

        WHERE   Parish.Name=MinDistanceToStream.P_ID AND
                Parish.Name=MinDistanceToLake.P_ID AND
                MinDistanceToStream.P_ID=MinDistanceToLake.P_ID

        GROUP BY Parish.Name
)

--
--Select the name, min distance to water and geometry of every parish:
--
SELECT  Parish_ID AS Parish, Min_DistanceToWater,
```

```
        Parish.Geom AS Parish_Geom

FROM    MinDistanceToWater INNER JOIN dbo.Parish
        ON MinDistanceToWater.Parish_ID=Parish.Name

ORDER BY Min_DistanceToWater;
```

B.1.3.4.b Calculation of the Proximity-To-Water Context Variable (Parish Level) with the “Border to Border” calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
-------------Calculate the "proximity-to-Water" context variable (parish level)---------------
--                                    (Border to Border)                                     --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the parish-polygon boarder and the points
--of the stream-polyline:
--
WITH MinDistanceToStream (P_ID, Distance_To_Stream, Max_Distance_To_Stream) AS
(
        SELECT   Parish.Name,
                 ROUND(MIN(Parish.Geom.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Distance_To_Stream,
                 ROUND(MAX(Parish.Geom.STDistance(dbo.MTSM_1812_1820_STREAM.Shape)), 1)
                 AS Max_Distance_To_Stream

        FROM     Parish, dbo.MTSM_1812_1820_STREAM

        WHERE    Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Calculate the shortest distance (geometry) between the parish-polygon border and the points
--of the lake-polygon border:
--
, MinDistanceToLake (P_ID, Distance_To_Lake) AS
(
        SELECT   Parish.Name,
                 ROUND(MIN(Parish.Geom.STDistance(dbo.MTSM_1812_1820_LAKE.Shape)), 1)
                 AS Distance_To_Lake

        FROM     Parish, dbo.MTSM_1812_1820_LAKE

        WHERE    Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Select the shortest distance to water, as the shortest distance between the parish-polygon
--border and a lake or stream geometry:
--
, MinDistanceToWater (Parish_ID, Min_DistanceToWater) AS
(
        SELECT Parish.Name,
                CASE
                        WHEN MIN(Distance_To_Stream)<MIN(Distance_To_Lake)
                                THEN MIN(Distance_To_Stream)
                        ELSE MIN(Distance_To_Lake)
                        END AS Min_DistanceToWater

        FROM    Parish, MinDistanceToStream, MinDistanceToLake

        WHERE   Parish.Name=MinDistanceToStream.P_ID AND
                Parish.Name=MinDistanceToLake.P_ID AND
                MinDistanceToStream.P_ID=MinDistanceToLake.P_ID

        GROUP BY Parish.Name
)

--
--Select the name, min distance to water and geometry of every parish:
--
SELECT   Parish_ID AS Parish, Min_DistanceToWater,
```

```
        Parish.Geom AS Parish_Geom

FROM    MinDistanceToWater INNER JOIN Parish
        ON MinDistanceToWater.Parish_ID=Parish.Name

ORDER BY Min_DistanceToWater;
```

B.1.4 SQL Queries – Calculation of the Proximity-To-Wetlands Context Variable

This part includes the SQL queries used to calculate the distance between physical location units of different aggregation levels (e.g. property units, initial property units, addresses and parishes) and wetlands. The proximity to wetlands is calculated as the shortest Euclidean distance between the centroid of the physical location unit polygon and the nearest point in the periphery of the nearest wetland polygon.

B.1.4.1.a Calculation of the Proximity-To-Wetlands Context Variable (Property Unit Level) with the "Centroid to Border" distance calculation method

```
---------------------------------------------------------------------------------------------
--                                                                                         --
---------Calculate the "proximity-to-wetlands" context variable (property-unit level):--------
--                                        (Centroid-To-Border)                           --
--                                          --All Parishes--                               --
--                                                                                         --
---------------------------------------------------------------------------------------------

--
--Calculate the min shortest distance (geometry) between the property-unit polygon border
--and the wetland-polygon border:
--
WITH PU_Distance_Wetlands (Parish, PropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT    PropertyUnits_All.Parish,
                  PropertyUnits_All.uniqueIdObj AS PropertyUnit_ID,
                  ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 1)
                  AS Min_Distance_To_Wetlands

        FROM      PropertyUnits_All, dbo.MTSM_1812_1820_WETLANDS

        WHERE     Parish!='Kågeröd'

        GROUP BY PropertyUnits_All.Parish, PropertyUnits_All.uniqueIdObj
)

--
--Select the parish, id, min shortest distance to wetlands and geometry of every
--property unit:
--
SELECT  PU_Distance_Wetlands.Parish AS Parish,
        PU_Distance_Wetlands.PropertyUnit_ID AS PropertyUnit_ID,
        Min_Distance_To_Wetlands,
        PropertyUnits_All.Shape AS PropertyUnit_Geom

FROM    PU_Distance_Wetlands INNER JOIN PropertyUnits_All
        ON PU_Distance_Wetlands.PropertyUnit_ID=PropertyUnits_All.uniqueIdObj

ORDER BY PU_Distance_Wetlands.Parish, PU_Distance_Wetlands.PropertyUnit_ID;
```

B.1.4.1.b Calculation of the Proximity-To- Wetlands Context Variable (Property Unit Level) with the "Border to Border" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
--------Calculate the "proximity-to-wetlands" context variable (property-unit level):---------
--                                    (Border-To-Border)                                     --
--                                                                                           --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the min shortest distance between the property-unit polygon border and the
--wetland-polygon border:
--
WITH PU_Distance_Wetlands (Parish, PropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT  Parish, PropertyUnits_All.uniqueIdObj AS PropertyUnit_ID,
                ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 1)
                AS Min_Distance_To_Wetlands

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_WETLANDS

        WHERE   Parish!='Kågeröd'

        GROUP BY Parish, PropertyUnits_All.uniqueIdObj
)

--
--Select the parish, id, min shortest distance to wetlands and geometry of every
--property unit:
--
SELECT  PU_Distance_Wetlands.Parish AS Parish,
        PU_Distance_Wetlands.PropertyUnit_ID AS PropertyUnit_ID,
        Min_Distance_To_Wetlands,
        PropertyUnits_All.Shape AS Property_Unit_Geom

FROM    PU_Distance_Wetlands INNER JOIN PropertyUnits_All
        ON PU_Distance_Wetlands.PropertyUnit_ID=PropertyUnits_All.uniqueIdObj

ORDER BY PU_Distance_Wetlands.Parish, PU_Distance_Wetlands.PropertyUnit_ID;
```

B.1.4.2.a Calculation of the Proximity-To- Wetlands Context Variable (Initial Property Unit Level) with the "Centroid to Border" distance calculation method

```
---------------------------------------------------------------------------------------------
--                                                                                         --
----Calculate the "proximity-to-wetlands" context variable (initial-property-unit level):-----
--                                                                                         --
--           (initial-property-unit-polygon Centroid To wetland-polygon Border)            --
--                                                                                         --
--                                     --All Parishes--                                    --
--                                                                                         --
---------------------------------------------------------------------------------------------


--
--Calculate the min shortest distance (geometry) between the initial-property-unit polygon centroid
--and the wetland-polygon border:
--
WITH IPU_Distance_Wetlands (Parish, InitialPropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT   Initial_Property_Unit.Parish AS IPU_Parish,
                 Initial_Property_Unit.uniqueIdObj AS InitialPropertyUnit_ID,
                 ROUND(MIN(Initial_Property_Unit.Centroid.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 2)
                 AS Min_Distance_To_Wetlands

        FROM     Initial_Property_Unit, dbo.MTSM_1812_1820_WETLANDS

        GROUP BY Initial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj
)

--
--Select the parish, id, min shortest distance to wetlands and geometry of every
-- initial-property-unit:
--
SELECT  IPU_Distance_Wetlands.Parish AS Parish,
        IPU_Distance_Wetlands.InitialPropertyUnit_ID AS InitialPropertyUnit_ID,
        Min_Distance_To_Wetlands,
        Initial_Property_Unit.Shape AS IPU_Geom

FROM    IPU_Distance_Wetlands INNER JOIN Initial_Property_Unit
        ON IPU_Distance_Wetlands.InitialPropertyUnit_ID=Initial_Property_Unit.uniqueIdObj

ORDER BY IPU_Distance_Wetlands.Parish, IPU_Distance_Wetlands.InitialPropertyUnit_ID;
```

B.1.4.2.b Calculation of the Proximity-To- Wetlands Context Variable (Initial Property Unit Level) with the "Border to Border" distance calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
-----Calculate the "proximity-to-wetlands" context variable (initial-property-unit level):----
--                                      (Border-To-Border)                                  --
--                                                                                          --
--                                        --All Parishes--                                  --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the min shortest distance (geometry) between the initial-property-unit polygon border
--and the wetland-polygon border:
--
WITH IPU_Distance_Wetlands (Parish, InitialPropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT   Parish, Initial_Property_Unit.uniqueIdObj AS InitialPropertyUnit_ID,
                 ROUND(MIN(Initial_Property_Unit.Shape.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 1)
                 AS Min_Distance_To_Wetlands

        FROM     Initial_Property_Unit, dbo.MTSM_1812_1820_WETLANDS

        GROUP BY Parish, Initial_Property_Unit.uniqueIdObj
)

--
--Select the min, max and avg shortest distance to wetlands, showing if the initial-property-unit
--polygon intersects with the wetland polygon:
--
SELECT  IPU_Distance_Wetlands.Parish AS Parish,
        IPU_Distance_Wetlands.InitialPropertyUnit_ID AS InitialPropertyUnit_ID,
        Min_Distance_To_Wetlands,
        Initial_Property_Unit.Shape AS IPU_Geom

FROM    IPU_Distance_Wetlands INNER JOIN Initial_Property_Unit
        ON IPU_Distance_Wetlands.InitialPropertyUnit_ID=Initial_Property_Unit.uniqueIdObj

ORDER BY IPU_Distance_Wetlands.Parish, IPU_Distance_Wetlands.InitialPropertyUnit_ID;
```

B.1.4.3.a Calculation of the Proximity-To- Wetlands Context Variable (Address Level) with the “Centroid to Border” distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
-------------Calculate the "proximity-to-wetlands" context variable (address level):----------
--                                  (Centroid-To-Border)                                   --
--                                    --All Parishes--                                       --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the min shortest distance (geometry) between the property-unit polygon centroid
--and the wetland-polygon border:
--
WITH PU_Distance_Wetlands (Parish, PropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT  Parish, uniqueIdObj AS PropertyUnit_ID,
                MIN(PropertyUnits_All.Centroid.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape))
                AS Min_Distance_To_Wetlands

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_WETLANDS

        WHERE   Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY Parish, uniqueIdObj
)

--
--Select the parish, name, min distance to wetlands and shape of every address:
--
SELECT  PropertyUnits_All.Parish, addressNameCode AS Address_ID,
        ROUND(AVG(Min_Distance_To_Wetlands), 1) AS Min_Distance_To_Wetlands,
        geometry::UnionAggregate(Shape) AS Address_Shape

FROM    PropertyUnits_All INNER JOIN PU_Distance_Wetlands
        ON PropertyUnits_All.uniqueIdObj = PU_Distance_Wetlands.PropertyUnit_ID

WHERE   addressNameCode!='' AND addressNameCode!='<Null>' AND
        addressNameCode IS NOT NULL

GROUP BY PropertyUnits_All.Parish, addressNameCode

ORDER BY PropertyUnits_All.Parish, addressNameCode;
```

B.1.4.3.b Calculation of the Proximity-To- Wetlands Context Variable (Address Level) with the “Border to Border” distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
-----------Calculate the "proximity-to-wetlands" context variable (address level):------------
--                                    (Border-To-Border)                                   --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the min shortest distance (geometry) between the property-unit polygon border
--and the wetland-polygon border:
--
WITH PU_Distance_Wetlands (Parish, PropertyUnit_ID, Min_Distance_To_Wetlands) AS
(
        SELECT  Parish, uniqueIdObj AS PropertyUnit_ID,
                MIN(PropertyUnits_All.Shape.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape))
                AS Min_Distance_To_Wetlands

        FROM    PropertyUnits_All, dbo.MTSM_1812_1820_WETLANDS

        WHERE   Parish!='Kågeröd' AND Initial_Property_Unit_ID IS NOT NULL

        GROUP BY Parish, uniqueIdObj
)

--
--Select the parish, name, min distance to wetlands and shape of every address:
--
SELECT  PropertyUnits_All.Parish, addressNameCode AS Address_ID,
        ROUND(AVG(Min_Distance_To_Wetlands), 1) AS Min_Distance_To_Wetlands,
        geometry::UnionAggregate(Shape) AS Address_Shape

FROM    PropertyUnits_All INNER JOIN PU_Distance_Wetlands
        ON PropertyUnits_All.uniqueIdObj = PU_Distance_Wetlands.PropertyUnit_ID

WHERE   addressNameCode!='' AND addressNameCode!='<Null>' AND
        addressNameCode IS NOT NULL

GROUP BY PropertyUnits_All.Parish, addressNameCode

ORDER BY PropertyUnits_All.Parish, addressNameCode;
```

B.1.4.4.a Calculation of the Proximity-To- Wetlands Context Variable (Parish Level) with the "Centroid to Border" distance calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
------------Calculate the "proximity-to-wetlands" context variable (parish level)------------
--                                  (Centroid to Border)                                    --
--                                                                                          --
--                                     --All Parishes--                                     --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the min shortest distance (geometry) between the parish polygon centroid and
--the wetland-polygon border:
--
WITH P_Distance_Wetlands (Parish, Min_Distance_To_Wetlands) AS
(
        SELECT  Parish.Name AS Parish_ID,
                ROUND(MIN(Parish.Centroid.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 2)
                AS Min_Distance_To_Wetlands

        FROM    dbo.Parish, dbo.MTSM_1812_1820_WETLANDS

        WHERE   dbo.Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Select the parish, min shortest distance to wetlands and geometry of every parish:
--
SELECT  P_Distance_Wetlands.Parish AS Parish,
        Min_Distance_To_Wetlands,
        Parish.Geom AS Parish_Geom

FROM    P_Distance_Wetlands INNER JOIN dbo.Parish
        ON P_Distance_Wetlands.Parish=Parish.Name

ORDER BY P_Distance_Wetlands.Parish;
```

B.1.4.4.b Calculation of the Proximity-To- Wetlands Context Variable (Parish Level) with the "Border to Border" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
------------Calculate the "proximity-to-wetlands" context variable (parish level)-------------
--                                (Border to Border)                                         --
--                                  --All Parishes--                                         --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the min shortest distance (geometry) between the parish polygon border and the
--wetland-polygon periphery:
--
WITH P_Distance_Wetlands (Parish, Min_Distance_To_Wetlands) AS
(
        SELECT   Parish.Name, ROUND(MIN(Parish.Geom.STDistance(dbo.MTSM_1812_1820_WETLANDS.Shape)), 1)
                 AS Min_Distance_To_Wetlands

        FROM     Parish, dbo.MTSM_1812_1820_WETLANDS

        WHERE    Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)
--
--Select the min shortest distance to wetlands, showing if the parish polygon intersects with
--the wetland and include the total wetland area within each parish boundaries:
--
SELECT  P_Distance_Wetlands.Parish AS Parish,
        Min_Distance_To_Wetlands,
        Parish.Geom AS Parish_Geom

FROM    P_Distance_Wetlands INNER JOIN Parish ON P_Distance_Wetlands.Parish=Parish.Name

ORDER BY P_Distance_Wetlands.Parish;
```

B.1.5 SQL Queries – Calculation of the Proximity-To-Gathering-Places Context Variable

This part includes the SQL queries used to calculate the distance between physical location units of different aggregation levels (e.g. property units, initial property units, addresses and parishes) and gathering places (e.g. churches). The proximity to gathering places is calculated as the shortest Euclidean distance (in meters) between the centroid of the physical location unit polygon and the nearest point representing a gathering place. Additionally, the proximity to gathering places is calculated as the shortest Euclidean distance (in meters) between the border of the physical location unit polygon and the nearest point representing a gathering place.

B.1.5.1.a Calculation of the Proximity-To- Gathering-Places Context Variable (Property Unit Level) with the "Centroid to Point" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
----------Calculate the "proximity-to-Church" context variable (property-unit level):---------
--                                     (Centroid to Point)                                   --
--                                                                                           --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between an property-unit polygon centroid
--and the church point:
--
WITH MinDistanceToGP (PU_ID, DistanceToChurch) AS
(
        SELECT  PropertyUnits_All.uniqueIdObj AS Property_Unit_ID,
                ROUND(MIN(PropertyUnits_All.Centroid.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape)), 1)
                AS Distance_To_Church

        FROM     dbo.PropertyUnits_All, dbo.EM_1910_1915_HOUSE_EXTENDED

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND
                 (descriptio LIKE 'Church%' OR descriptio LIKE '%Church') AND
                 PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY PropertyUnits_All.uniqueIdObj
)

--
--Select the parish, id, min distance to church and geometry of every
--property unit:
--
SELECT  PropertyUnits_All.Parish AS Parish, PU_ID, DistanceToChurch,
                PropertyUnits_All.Shape AS PU_Geom

FROM    dbo.PropertyUnits_All INNER JOIN MinDistanceToGP
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToGP.PU_ID

ORDER BYPropertyUnits_All.Parish, PropertyUnits_All.uniqueIdObj;
```

B.1.5.1.b Calculation of the Proximity-To- Gathering-Places Context Variable (Property Unit Level) with the “Border to Point” distance calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
---------Calculate the "proximity-to-Church" context variable (property-unit level):----------
--                                   (Border to Point)                                      --
--                                                                                          --
--                                    --All Parishes--                                      --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between an property-unit polygon border
--and the church point:
--
WITH MinDistanceToGP (PU_ID, DistanceToChurch) AS
(
        SELECT   PropertyUnits_All.uniqueIdObj AS Property_Unit_ID,
                 ROUND(MIN(PropertyUnits_All.Shape.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape)), 1)
                 AS Distance_To_Church

        FROM     PropertyUnits_All, dbo.EM_1910_1915_HOUSE_EXTENDED

        WHERE    PropertyUnits_All.Parish!='Kågeröd' AND
                (descriptio LIKE 'Church%' OR descriptio LIKE '%Church') AND
                 PropertyUnits_All.Initial_Property_Unit IS NOT NULL

        GROUP BY PropertyUnits_All.uniqueIdObj
)

--
--Select the parish, property unit ID, min distance to church from property unit border
-- and geometry for every property unit:
--
SELECT  PropertyUnits_All.Parish AS Parish, PU_ID, DistanceToChurch,
        PropertyUnits_All.Shape AS PU_Geom

FROM    PropertyUnits_All INNER JOIN MinDistanceToGP
        ON PropertyUnits_All.uniqueIdObj=MinDistanceToGP.PU_ID

ORDER BYPropertyUnits_All.Parish, PropertyUnits_All.uniqueIdObj;
```

B.1.5.2.a Calculation of the Proximity-To- Gathering-Places Context Variable (Initial Property Unit Level) with the "Centroid to Point" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
------Calculate the "proximity-to-Church" context variable (initial-property-unit level):-----
--                                     (Centroid to Point)                                   --
--                                                                                           --
--                                       --All Parishes--                                    --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the shortest distance (geometry) between an initial-property-unit polygon centroid
--and the church point:
--
WITH MinDistanceToGP (IPU_ID, DistanceToChurch) AS
(
SELECT  Initial_Property_Unit.uniqueIdObj AS Initial_Property_Unit_ID,
        ROUND(MIN(Initial_Property_Unit.Centroid.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape)), 1)
                         AS Distance_To_Church

FROM     dbo.Initial_Property_Unit, dbo.EM_1910_1915_HOUSE_EXTENDED

WHERE    Initial_Property_Unit.Parish!='Kågeröd' AND
         (descriptio LIKE 'Church%' OR descriptio LIKE '%Church')

GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Select the parish, id, min distance to church and geometry of every
--initial property unit:
--
SELECT  Initial_Property_Unit.Parish AS Parish, IPU_ID, DistanceToChurch,
        Initial_Property_Unit.Shape AS IPU_Geom

FROM    Initial_Property_Unit INNER JOIN MinDistanceToGP
        ON Initial_Property_Unit.uniqueIdObj=MinDistanceToGP.IPU_ID

ORDER BYInitial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj;
```

B.1.5.2.b Calculation of the Proximity-To- Gathering-Places Context Variable (Initial Property Unit Level) with the "Border to Point" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
------Calculate the "proximity-to-Church" context variable (initial-property-unit level):-----
--                                   (Border to Point)                                       --
--                                                                                           --
--                                    --All Parishes--                                       --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the shortest distance (geometry) between an initial-property-unit polygon border
--and the church point:
--
WITH MinDistanceToGP (IPU_ID, DistanceToChurch) AS
(
SELECT   Initial_Property_Unit.uniqueIdObj AS Initial_Property_Unit_ID,
         ROUND(MIN(Initial_Property_Unit.Shape.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape)), 1)
         AS Distance_To_Church

FROM     Initial_Property_Unit, dbo.EM_1910_1915_HOUSE_EXTENDED

WHERE    Initial_Property_Unit.Parish!='Kågeröd' AND
                          (descriptio LIKE 'Church%' OR descriptio LIKE '%Church')

GROUP BY Initial_Property_Unit.uniqueIdObj
)

--
--Select the parish, initial property unit id, min distance to gathering location and
--geometry for every initial property unit:
--
SELECT   Initial_Property_Unit.Parish AS Parish, IPU_ID, DistanceToChurch,
         Initial_Property_Unit.Shape AS IPU_Geom

FROM     Initial_Property_Unit INNER JOIN MinDistanceToGP
         ON Initial_Property_Unit.uniqueIdObj=MinDistanceToGP.IPU_ID

ORDER BYInitial_Property_Unit.Parish, Initial_Property_Unit.uniqueIdObj;
```

B.1.5.3.a Calculation of the Proximity-To- Gathering-Places Context Variable (Address Level) with the "Centroid to Point" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
-------------Calculate the "proximity-to-Church" context variable (address level):------------
--                                    (Centroid to Point)                                    --
--                                                                                           --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between an property-unit polygon centroid
--and the church point:
--
WITH PU_MinDistanceToGP (Parish, PU_ID, DistanceToChurch) AS
(
        SELECT  PropertyUnits_All.Parish AS Parish, uniqueIdObj AS Property_Unit_ID,
                MIN(PropertyUnits_All.Centroid.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape))
                AS Distance_To_Church

        FROM    PropertyUnits_All, dbo.EM_1910_1915_HOUSE_EXTENDED

        WHERE   PropertyUnits_All.Parish!='Kågeröd' AND
                (descriptio LIKE 'Church%' OR descriptio LIKE '%Church') AND
                PropertyUnits_All.Initial_Property_Unit_ID IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Select the parish, id, min distance to church and geometry of every address unit:
--
SELECT  PU_MinDistanceToGP.Parish, AddressNameCode AS Address_ID,
        ROUND(AVG(DistanceToChurch), 1) AS DistanceToChurch,
        geometry::UnionAggregate(Shape) AS Address_Shape

FROM    PU_MinDistanceToGP INNER JOIN PropertyUnits_All
        ON PU_MinDistanceToGP.PU_ID=PropertyUnits_All.uniqueIdObj

WHERE   addressNameCode!='<Null>' AND addressNameCode!='' AND
        addressNameCode IS NOT NULL

GROUP BY PU_MinDistanceToGP.Parish, AddressNameCode

ORDER BY PU_MinDistanceToGP.Parish, AddressNameCode;
```

B.1.5.3.b Calculation of the Proximity-To- Gathering-Places Context Variable (Address Level) with the “Border to Point” distance calculation method

```
----------------------------------------------------------------------------------------------
--                                                                                          --
-------------Calculate the "proximity-to-Church" context variable (address level):------------
--                                    (Border to Point)                                     --
--                                                                                          --
--                                     --All Parishes--                                     --
--                                                                                          --
----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between an property-unit polygon border
--and the church point:
--
WITH PU_MinDistanceToGP (Parish, PU_ID, DistanceToChurch) AS
(
        SELECT  PropertyUnits_All.Parish AS Parish, uniqueIdObj AS Property_Unit_ID,
                MIN(PropertyUnits_All.Shape.STDistance(dbo.EM_1910_1915_HOUSE_EXTENDED.Shape))
                AS Distance_To_Church

        FROM    PropertyUnits_All, dbo.EM_1910_1915_HOUSE_EXTENDED

        WHERE   PropertyUnits_All.Parish!='Kågeröd' AND
                (descriptio LIKE 'Church%' OR descriptio LIKE '%Church') AND
                PropertyUnits_All.Initial_Property_Unit_ID IS NOT NULL

        GROUP BY PropertyUnits_All.Parish, uniqueIdObj
)

--
--Select the parish, id, min distance to church and geometry of every address unit:
--
SELECT  PU_MinDistanceToGP.Parish, addressNameCode AS Address_ID,
        ROUND(AVG(DistanceToChurch), 1) AS MinDistanceToChurch,
        geometry::UnionAggregate(Shape) AS Address_Shape

FROM    PU_MinDistanceToGP INNER JOIN PropertyUnits_All
        ON PU_MinDistanceToGP.PU_ID=PropertyUnits_All.uniqueIdObj

WHERE   addressNameCode!='<Null>' AND addressNameCode!='' AND
        addressNameCode IS NOT NULL

GROUP BY PU_MinDistanceToGP.Parish, addressNameCode

ORDER BY PU_MinDistanceToGP.Parish, addressNameCode;
```

B.1.5.5.a Calculation of the Proximity-To- Gathering-Places Context Variable (Parish Level) with the "Centroid to Point" distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
--------------Calculate the "proximity-to-Church" context variable (parish level):------------
--                                    (Centroid to Point)                                    --
--                                                                                           --
--                                      --All Parishes--                                     --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Calculate the shortest distance (geometry) between the parish polygon centroid and the church point:
--
WITH MinDistanceToGatheringPlace (Parish, Min_Distance_To_GP) AS
(
        SELECT   Parish.Name AS Parish,
                 ROUND(MIN(Parish.Centroid.STDistance(dbo.LSM_1757_1863_HOUSE.Shape)), 1)
                 AS Min_Distance_To_Church

        FROM     dbo.Parish INNER JOIN LSM_1757_1863_HOUSE
                 ON Parish.Name=LSM_1757_1863_HOUSE.parish

        WHERE    subType='Church' AND Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Select name, min distance to church and geometry of every parish:
--
SELECT  Parish.Name AS Parish,
        Min_Distance_To_GP AS DistanceToChurch,
        Parish.Geom AS Parish_Geom

FROM    dbo.Parish INNER JOIN MinDistanceToGatheringPlace
        ON Parish.Name=MinDistanceToGatheringPlace.Parish;
```

B.1.5.5.b Calculation of the Proximity-To- Gathering-Places Context Variable (Parish Level) with the “Border to Point” distance calculation method

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
--------------Calculate the "proximity-to-Church" context variable (parish level):------------
--                                   (Border to Point)                                       --
--                                                                                           --
--                                    --All Parishes--                                       --
--                                                                                           --
-----------------------------------------------------------------------------------------------

--
--Calculate the shortest distance (geometry) between the parish polygon border and the church point:
--
WITH MinDistanceToGatheringPlace (Parish, Min_Distance_To_GP) AS
(
        SELECT   Parish.Name AS Parish,
                 ROUND(MIN(Parish.Geom.STDistance(dbo.LSM_1757_1863_HOUSE.Shape)), 1)
                 AS Min_Distance_To_Church

        FROM     Parish INNER JOIN LSM_1757_1863_HOUSE
                 ON Parish.Name=LSM_1757_1863_HOUSE.parish

        WHERE    subType='Church' AND Parish.Name!='Kågeröd'

        GROUP BY Parish.Name
)

--
--Select the name, geometry and min distance to gathering place for every parish:
--
SELECT   Parish.Name AS Parish,
         Min_Distance_To_GP AS MinDistanceToChurch,
         Parish.Geom AS Parish_Geom

FROM     Parish INNER JOIN MinDistanceToGatheringPlace
         ON Parish.Name=MinDistanceToGatheringPlace.Parish;
```

B.1.6 SQL Queries – Calculation of the Population-Density-Index Context Variable

This part includes the SQL queries used to calculate the Population Density Index on different aggregation levels (e.g. property units, initial property units, and addresses. The Population Density Index of each geographical unit is being calculated according to the formula previously explained in the implementation section of this study. The distance between the geographical units is computed as the distance between their centroids. The population census year is 1910. The buffer area in this case is set to 4000 meters.

B.1.6.1 Calculation of the Population-Density-Index Context Variable (Property Unit Level)

```
-----------------------------------------------------------------------------------------------
--                                                                                           --
----Calculating Population-Density-Index Geographic Context Variable (property-unit level)------
--                                 (Centroid to Centroid)                                    --
--                                                                                           --
--                                   --All Parishes--                                        --
--                                      Year 1910                                            --
--                                                                                           --
-----------------------------------------------------------------------------------------------


--
--Select the id and the population of the property units that existed in year 1910.
--The population information is available either directly on a property unit level
--(linkLevel 1) or on an address level (linkLevel=2).
--In some cases population information has successfully been linked to a property-
--unit directly. In other cases, population information is only available on an
--address level. According to this approach, the population of one address is divided
--equally between the property units related to that address. This means that there
--are property units that have population information linked to them from both
--linkLevel 1 and linkLevel 2. This information is summed up to produce a single
--population value for every property-unit in the data set for year 1910:
--
WITH PU_Population (PU_ID, PU_Population) AS
(
        SELECT  uniqueIdObj, SUM(Population) AS Population

        FROM    Population_PropertyUnits

        WHERE   CensusYear='1910'

        GROUP BY uniqueIdObj
)


--
--Select the parish, id, population information and geometry of every property
--unit.
--Exclude the large overlapping property units of Hög & Kävlinge parishes from the
--calculations.
--
, PU_Population_Geom (Parish, PU_ID, PU_Population, Geom, Centroid) AS
(
        SELECT  Parish, uniqueIdObj,
                ISNULL(PU_Population.PU_Population, 0) AS PU_Population,
                Shape AS Geom,
                Centroid

        FROM    PU_Population LEFT JOIN PropertyUnits_All
                ON PU_Population.PU_ID=PropertyUnits_All.uniqueIdObj

        --Exclude the large containing PUs of Hög & Kävlinge:
        WHERE   Initial_Property_Unit_ID IS NOT NULL
)


--
--Calculate the PDI of every property-unit on a 4000m radius from it's
--centroid to the other property-units' centroid:
--
```

```sql
, Population_Density_Index (PU_ID, PDI) AS
(
        SELECT  a.PU_ID AS PU_ID,
                ROUND(a.PU_Population/50+SUM(b.PU_Population/a.Centroid.STDistance(b.Centroid)), 2)
                AS Population_Density_Index

        FROM    PU_Population_Geom a, PU_Population_Geom b

        WHERE   a.PU_ID!=b.PU_ID AND

                --Measure to limit the EM, LSM map series property-unit overlapping problem.
                --Exclude PU-centroids that are in closer distance than 40m from the calculations.
                a.Centroid.STDistance(b.Centroid)>40 AND
                a.Centroid.STDistance(b.Centroid)<=4000 --buffer radius set to 4km

        GROUP BY a.PU_ID, a.PU_Population
)

--
--Select the ID, PDI and geometry of each property unit:
--
SELECT  Parish, Population_Density_Index.PU_ID AS PU_ID,
        PDI AS Population_Density_Index,
        PU_Population_Geom.Geom AS PU_Geom

FROM    Population_Density_Index INNER JOIN PU_Population_Geom
        ON Population_Density_Index.PU_ID=PU_Population_Geom.PU_ID

ORDER BY Parish, PU_ID;
```

B.1.6.2 Calculation of the Population-Density-Index Context Variable (Initial Property Unit Level)

```
---------------------------------------------------------------------------------------------
--                                                                                         --
-----------------Calculating Population-Density-Index Geographic Context Variable ------------
----------------------------------(initial-property-unit level)------------------------------
--                                                                                         --
--                                   (Centroid to Centroid)                                --
--                                                                                         --
--                                      --All Parishes--                                   --
--                                         Year 1910                                       --
--                                                                                         --
---------------------------------------------------------------------------------------------


--
--Select the id and the population of the property units that existed in year 1910.
--The population information is available at an address level (linklevel=2).
--Therefore, according to this approach, the population of one address is divided
--equally between the property units related to that address.
--In some cases, the population has been linked directly to a property unit (linkLevel 2).
--
--As the population of some property units has both been calculated based on their
--address and property-unit information (double uniqueIdObj entries), in this case,
--we choose to sum the population information of both entries.
--
WITH IPU_Population (IPU_ID, IPU_Population) AS
(
        SELECT  uniqueIdObj, SUM(Population) AS Population

        FROM    Population_PropertyUnits

        WHERE   CensusYear='1910'

        GROUP BY uniqueIdObj
)


--
--As many initial property units did no longer exist in 1910, the calculation of the
--Population Density Index on the initial property unit level will be computed as the
--sum of the population of the property units included in the former initial-property-unit
--area. Sum the population of the property units that are included in the same initial-
--property-unit. Furthermore, as time passed, initial-property units where divided to smaller
--property units, which where in turn divided to even smaller property units or merged with
--smaller property-units belonging to neighbouring initial-property units.
--The summed area of all property units linked to a specific initial property unit is
--not always identical to the total area of the original initial-property unit.
--Here we will only use the area of the original initial-property-unit.
--
, IPU_Population_SUM(Parish, IPU_ID, IPU_Population) AS
(
        SELECT  Parish, Initial_Property_Unit_ID AS IPU_ID,
                ISNULL(SUM(IPU_Population.IPU_Population), 0) AS IPU_Population

        FROM    IPU_Population RIGHT JOIN PropertyUnits_All
                ON IPU_Population.IPU_ID=PropertyUnits_All.uniqueIdObj

        --Perform the calculation only for Hög parish.
        --Exclude the large PUs that contain more than one IPUs & are therefore not linked to
        --any initial-property-unit:
        WHERE   Initial_Property_Unit_ID IS NOT NULL

        --Sum the populations of the property-units that have been merged to form
        --an initial-property-unit:
        GROUP BY Parish, Initial_Property_Unit_ID
)


--
--Select the pariah, id, population and geometry of every initial property unit:
--
, IPU_Population_Geom (Parish, IPU_ID, Population, Geom) AS
```

```sql
(
	SELECT	IPU_Population_SUM.Parish, IPU_Population_SUM.IPU_ID,
		IPU_Population_SUM.IPU_Population AS Population,
		Initial_Property_Unit.Shape AS Geom

	FROM	IPU_Population_SUM LEFT JOIN Initial_Property_Unit
		ON IPU_Population_SUM.IPU_ID=Initial_Property_Unit.uniqueIdObj
)


--
--Calculate the PDI of every initial property-unit on a 4000m radius from it's
--centroid to the other initial-property-units' centroid:
--
, Population_Density_Index (IPU_ID, PDI) AS
(
	SELECT	a.IPU_ID AS IPU_ID_a,

	ROUND(a.Population/50+SUM(b.Population/a.Geom.STCentroid().STDistance(b.Geom.STCentroid())), 2)
		AS Population_Density_Index

	FROM	IPU_Population_Geom a, IPU_Population_Geom b

	--Exclude the examined initial property unit's distance from itself from the calculations.
	--Exclude all the initial-property-units whose distance from the examined initial
	--property unit's centroid is larger than 4000m.
	WHERE	a.IPU_ID!=b.IPU_ID AND
		a.Geom.STCentroid().STDistance(b.Geom.STCentroid())<=4000

	GROUP BY a.IPU_ID, a.Population
)


--
--Select the ID, PDI and geometry of each initial property unit:
--
SELECT	Parish, Population_Density_Index.IPU_ID AS IPU_ID,
	PDI AS Population_Density_Index,
	IPU_Population_Geom.Geom AS IPU_Geom

FROM	Population_Density_Index INNER JOIN IPU_Population_Geom
	ON Population_Density_Index.IPU_ID=IPU_Population_Geom.IPU_ID

ORDER BY Parish, IPU_ID
```

B.1.6.3 Calculation of the Population-Density-Index Context Variable (Address Level)

```
----------------------------------------------------------------------------------------------
--                                                                                          --
-----Calculating Population-Density-Index Geographic Context Variable (address-unit level)----
--                                                                                          --
--                                    (Centroid to Centroid)                                --
--                                                                                          --
--                                       --All Parishes--                                   --
--                                          Year 1910                                       --
--                                                                                          --
----------------------------------------------------------------------------------------------


--
--Select the id and the population of the property units that existed in year 1910.
--The population information is available either directly on a property unit level
--(linkLevel 1) or on an address level (linkLevel=2).
--In some cases population information has successfully been linked to a property-
--unit directly. In other cases, population information is only available on an
--address level. According to this approach, the population of one address is divided
--equally between the property units related to that address. This means that there
--are property units that have population information linked to them from both
--linkLevel 1 and linkLevel 2. This information is summed up to produce a single
--population value for every property-unit in the data set for year 1910:
--
WITH PU_Population (PU_ID, PU_Population) AS
(
        SELECT  uniqueIdObj, SUM(Population) AS Population

        FROM    Population_PropertyUnits

        WHERE   CensusYear='1910'

        GROUP BY uniqueIdObj
)


--
--Select the parish, id, population information and geometry of every property
--unit.
--Exclude the large overlapping property units of Hög & Kävlinge parishes from the
--calculations.
--
, PU_Population_Geom (Parish, PU_ID, PU_Population, Geom, Centroid) AS
(
        SELECT  Parish, uniqueIdObj,
                ISNULL(PU_Population.PU_Population, 0) AS PU_Population,
                Shape AS Geom,
                Centroid

        FROM    PU_Population LEFT JOIN PropertyUnits_All
                ON PU_Population.PU_ID=PropertyUnits_All.uniqueIdObj

        --Exclude the PUs that did not exist in 1910 and the large containing PUs of Hög & Kävlinge:
        WHERE   Initial_Property_Unit_ID IS NOT NULL
)


--
--Calculate the PDI of every property-unit on a 4000m radius from it's
--centroid to the other property-units' centroid:
--
, Population_Density_Index (PU_ID, PDI) AS
(
        SELECT  a.PU_ID AS PU_ID,
                ROUND(a.PU_Population/50+SUM(b.PU_Population/a.Centroid.STDistance(b.Centroid)), 2)
                AS Population_Density_Index

        FROM    PU_Population_Geom a, PU_Population_Geom b

        WHERE   a.PU_ID!=b.PU_ID AND
```

```sql
                --Measure to limit the EM, LSM map series property-unit overlapping problem.
                --Exclude PU-centroids that are in closer distance than 40m from the calculations.
                a.Centroid.STDistance(b.Centroid)>40 AND
                a.Centroid.STDistance(b.Centroid)<=4000

        GROUP BY a.PU_ID, a.PU_Population
)

--
--Select average PDI per address-unit:
--

SELECT  Parish, addressNameCode AS Address_ID,
        ROUND(AVG(PDI), 2) AS Population_Density_Index,
        geometry::UnionAggregate(Shape) AS Address_Geom

FROM    Population_Density_Index INNER JOIN PropertyUnits_All
        ON Population_Density_Index.PU_ID=PropertyUnits_All.uniqueIdObj

GROUP BY Parish, addressNameCode

ORDER BY Parish, addressNameCode;
```

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)

335 Fei Lu (2015) Compute a Crowdedness Index on Historical GIS Data- A Case Study of Hög Parish, Sweden, 1812-1920
336 Lina Allesson (2015) Impact of photo-chemical processing of dissolved organic carbon on the bacterial respiratory quotient in aquatic ecosystems
337 Andreas Kiik (2015) Cartographic design of thematic polygons: a comparison using eye-movement metrics analysis
338 Iain Lednor (2015) Testing the robustness of the Plant Phenology Index to changes in temperature
339 Louise Bradshaw (2015) Submerged Landscapes - Locating Mesolithic settlements in Blekinge, Sweden
340 Elisabeth Maria Farrington (2015) The water crisis in Gaborone: Investigating the underlying factors resulting in the 'failure' of the Gaborone Dam, Botswana
341 Annie Forssblad (2015) Utvärdering av miljöersättning för odlingslandskapets värdefulla träd
342 Iris Behrens, Linn Gardell (2015) Water quality in Apac-, Mbale- & Lira district, Uganda - A field study evaluating problems and suitable solutions
343 Linnéa Larsson (2015) Analys av framtida översvämningsrisker i Malmö - En fallstudie av Castellums fastigheter
344 Ida Pettersson (2015) Comparing Ips Typographus and Dendroctonus ponderosas response to climate change with the use of phenology models
345 Frida Ulfves (2015) Classifying and Localizing Areas of Forest at Risk of Storm Damage in Kronoberg County
346 Alexander Nordström (2015) Förslag på dammar och skyddsområde med hjälp av GIS: En studie om löv- och klockgroda i Ystad kommun, Skåne
347 Samanah Seyedi-Shandiz (2015) Automatic Creation of Schematic Maps - A Case Study of the Railway Network at the Swedish Transport Administration
348 Johanna Andersson (2015) Heat Waves and their Impacts on Outdoor Workers – A Case Study in Northern and Eastern Uganda
349 Jimmie Carpman (2015) Spatially varying parameters in observed new particle formation events

350 Mihaela – Mariana Tudoran (2015) Occurrences of insect outbreaks in Sweden in relation to climatic parameters since 1850

351 Maria Gatzouras (2015) Assessment of trampling impact in Icelandic natural areas in experimental plots with focus on image analysis of digital photographs

352 Gustav Wallner (2015) Estimating and evaluating GPP in the Sahel using MSG/SEVIRI and MODIS satellite data

353 Luisa Teixeira (2015) Exploring the relationships between biodiversity and benthic habitat in the Primeiras and Segundas Protected Area, Mozambique

354 Iris Behrens & Linn Gardell (2015) Water quality in Apac-, Mbale- & Lira district, Uganda - A field study evaluating problems and suitable solutions

355 Viktoria Björklund (2015) Water quality in rivers affected by urbanization: A Case Study in Minas Gerais, Brazil

356 Tara Mellquist (2015) Hållbar dagvattenhantering i Stockholms stad - En riskhanteringsanalys med avseende på långsiktig hållbarhet av Stockholms stads dagvattenhantering i urban miljö

357 Jenny Hansson (2015) Trafikrelaterade luftföroreningar vid förskolor – En studie om kvävedioxidhalter vid förskolor i Malmö

358 Laura Reinelt (2015) Modelling vegetation dynamics and carbon fluxes in a high Arctic mire

359 Emelie Linnéa Graham (2015) Atmospheric reactivity of cyclic ethers of relevance to biofuel combustion

360 Filippo Gualla (2015) Sun position and PV panels: a model to determine the best orientation

361 Joakim Lindberg (2015) Locating potential flood areas in an urban environment using remote sensing and GIS, case study Lund, Sweden

362 Georgios-Konstantinos Lagkas (2015) Analysis of NDVI variation and snowmelt around Zackenberg station, Greenland with comparison of ground data and remote sensing.

363 Carlos Arellano (2015) Production and Biodegradability of Dissolved Organic Carbon from Different Litter Sources

364 Sofia Valentin (2015) Do-It-Yourself Helium Balloon Aerial Photography - Developing a method in an agroforestry plantation, Lao PDR

365 Shirin Danehpash (2015) Evaluation of Standards and Techniques for Retrieval of Geospatial Raster Data - A study for the ICOS Carbon Portal

366 Linnea Jonsson (2015) Evaluation of pixel based and object based classification methods for land cover mapping with high spatial resolution satellite imagery, in the Amazonas, Brazil.

367 Johan Westin (2015) Quantification of a continuous-cover forest in Sweden using remote sensing techniques

368 Dahlia Mudzaffar Ali (2015) Quantifying Terrain Factor Using GIS Applications for Real Estate Property Valuation

369 Ulrika Belsing (2015)The survival of moth larvae feeding on different plant species in northern Fennoscandia

370 Isabella Grönfeldt (2015) Snow and sea ice temperature profiles from satellite data and ice mass balance buoys