Master's Thesis
Geography
Geoinformatics

CLUSTERING OF IMMIGRATION POPULATION IN HELSINKI METROPOLITAN AREA, FINLAND: A COMPARATIVE STUDY OF EXPLORATORY SPATIAL DATA ANALYSIS METHODS

Vladimir Kekez

2015

Supervisor:
Senior lecturer Mika Siljander

UNIVERSITY OF HELSINKI
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
DIVISION OF GEOGRAPHY

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section	Laitos – Institution – Department
Faculty of Science	Department of geosciences and geography

Tekijä – Författare – Author
Vladimir Kekez

Työn nimi – Arbetets titel – Title
Computing Capabilities of Exploratory Spatial Data Analysis Methods in different GIS software: Comparative study on the Case of Immigrant Population in Capital Region of Helsinki (Pääkaupunkiseutu)

Oppiaine – Läroämne – Subject
Geography (Geoinformatics)

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's Thesis	January 2015	109

Tiivistelmä – Referat – Abstract

In the world of globalization immigration processes represent consequence of the search for better life. Every year more immigrants are coming to stay and live in Finland. Understanding patterns of living, spatial locations and clustering of this specific population becomes important and integral step towards integration of immigration population in society.

Studies of immigration population conducted in Finland and Helsinki Metropolitan Area are mostly done with descriptive statistical methods mostly employed for describing social patterns and participation of immigrant population within the whole population. Employment of inferential statistical methods, spatial statistical methods, precisely Exploratory Spatial Data Analysis (ESDA methods), specifically Global and Local Moran's Index is becoming extremely important because of the quantitative and qualitative results which can be gained.

This thesis is consisted of analysis of immigrant population patterns, conducted by Global and Local Moran's Index used by ArcGIS and GeoDa software. ArcGIS is a market leading, commercial GIS package for computation, analysis and production of different sorts of GIS analysis and results. Spatial statistic toolbox, as integral part of ArcGIS software package is used for interpretation of spatial statistics results (maps, graphs, reports etc.), which can be obtained, by use of several different methods. GeoDa is non-commercial software, relatively new in GIS practice in Finland, focusing specifically in spatial statistics analysis. It is used for manipulation and operationalization of spatial data analysis, designed for implementation of different and unique (Bivariate Moran’s I, etc.) ESDA techniques. Both software are computing comparable but different results, quantitatively and visually. For global measurements of spatial autocorrelation and presence of clustering within analyzed area Global Moran’s Index is employed. Local measurements and for mapping of possible cluster and outlier occurrences (Anselin Local Moran’s Index) is being used. Employment of weight matrix produced in ArcGIS and GeoDa is allowing creation of conceptualization of spatial weight matrix on the same principles in ArcGIS and GeoDa. Conceptualization of weight matrix in the case of lattice data with shared border is contiguity concept. Contiguity concept is using queen concept for defining neighbors, because it allows bigger analyzing capacity.

Both software are using same statistical equations but outcome results are showing variety of differences, because of the differences in computing, presenting and visual displaying of the results. GeoDa is producing more significant statistical and visual results. The task is to test and compare computational, visual and analytical capabilities and possibilities of both software and analyze quality of outcome results (maps, diagrams, box plots, etc.) Data on immigration population is provided by HSY (Helsingin Seudun Ympäristö) with the lattice grid level size (1x1km, 500x500m, 250x250m). Purpose of my thesis is also to analyze lattice data with new square grid sizes (50x50m), which are inputting more specific local area inputs for location of local spatial autocorrelation and hot spot activities. Creation of new lattice size is motivated by conceptualizing of the data which is aggregated on the building level (Pks_vaki). Main motive is to try to detect new trends in development of clustering and clusters of immigrant population in Greater Helsinki, formulate and impose scale and area size from a different perspective.

Results are informing about undetected process of clustering in the central areas of Helsinki not noticed in the previous studies of immigration population. They are offering different perspective on the problem of clustering of immigration population in Helsinki Metropolitan area.

Avainsanat – Nyckelord – Keywords
spatial statistics, ArcGIS, GeoDa, Global and Local Moran's Index, immigration

Säilytyspaikka – Förvaringställe – Where deposited
Kumpula campus library

Muita tietoja – Övriga uppgifter – Additional information

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CO type	Cluster or outlier type
CRA	Capital Region Area of Helsinki
CSR	Complete Spatial Randomness
EDA	Exploratory Data Analysis
ESDA	Exploratory Spatial Data Analysis
EU	European Union
GAL	Geographical Algorithms Library
GIS	Geographical Information Systems
GMI	Global Moran's Index
HH	High-High cluster value
HL	High-Low outlier value
HMA	Helsinki Metropolitan Area
HSY	Helsingin Seudun Ympäristöpalvelut-kuntayhtymä
KKJ	Mapping coordinate system (Kartastokoordinaattijärjestelmä)
lagged_ULKOKANS	lagged version of observed immigrant population
LH	Low-High outlier value
LISA	Local Index of Spatial Autocorrelation
LISA_CL	Cluster, outlier or Not Significant type
LISA_I	Local Moran's Index
LISA_P	p-value of Local Moran's Index
LL	Low-Low cluster value
LMI	Local Moran's Index
MAUP	Modifiable areal unit problem
OAPEC	Organization of Arab Petroleum Exporting Countries
OECD	Organization for Economic Co-operation and Development

PKS_VAKI	File consisting information about different type of population language
PySAL	Python Spatial Analysis Library
RUUDUT	Preprocessed grids
SUM_ULKOKANS	Term in the table of context (PKS_VAKI) for immigrant population
TOC	Table of Contents
UK	United Kingdom
ULKOKANS	observed immigrant population
USSR	The Union of Soviet Socialist Republics

SYMBOLS

I	Global Moran's Index value
n	Total number of spatial units indexed by i and j
i and j	Spatial units, neighbors
z_i	Deviation of an attribute for feature i from its mean $(x_i - \overline{X})$
x_i	Variable of interest
$\overline{X}$	Mean of x_i
$w_{i,j}$	Spatial weight between feature i and j
S_o	Aggregate of all the spatial weight
$E[I]$	Expected Moran's Index value
$V[I]$	Variance from Moran's Index value
W	Weight matrix
W_{ij}	Normalization of weigh matrix
I_i	Local Moran's Index value

x_i	Attribute for feature i
$\overline{X}$	Mean of corresponding attribute
$w_{i,j}$	Spatial weight between feature i and j and
zI_i	Z-score for Local Moran's Index value
$E[I_i]$	Expected Local Moran's Index value
$V[I_i]$	Variance of Local Moran's Index value
R^2	R square measure of the regression
const a	representation of regression lagged analysis
std-err a	positive or negative error value
t-stat a	value of t-statistic which is the product of Std-err a
p-value a	representation of the statistical significance of a.
slope b	parameter b of the regression lagged analysis
std-err b	positive or negative error value
t-stat b	value of t-statistic which is the product of Std-err b.
p-value b	representation of the significance of the Moran index

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In the world of globalization immigration represents consequence of the search for better life. Globalization has brought upon us global security concerns, humanitarian crises and skill shortages of migration and immigration which have rooted themselves as a central concept of economic, political and social debates at the beginning of this century (Samers, 2010). Contemporary worldwide trends of immigration are consequence of neoliberal order of global economy. Huge income differences all around the world are the consequence of extreme differences in mean incomes of the countries. From the beginning of 1980's huge amount of countries all around the world, especially the poorest ones were experiencing a systematic growth failure (Milanovic, 2006). Neoliberal economic system put up in 1980's by governments of Thatcher in UK and Reagan in USA has reshaped worldwide economy, affecting immigration and migration trends in the 1990's all around the world. "Developed world" represented by Western and Northern European economies due to demographic shrinking (ageing population), economic growth (need for more labor force) and different social factors has experienced need for new wave of immigration. New immigration wave was represented by various groups like refugees, asylum-seekers, highly skilled personnel, manual workers and family members (Castles, 2011).

New trends from 2000's are giving us clear inputs. By the year 2005, workers with foreign background were creating a quarter of the labor force in Australia and Switzerland. In Canada it was 20%, in the USA, New Zealand, Austria and Germany 15%. Following up this trend Western European countries were having around 12% of this kind of population (OECD, 2007).

Present immigration processes and trends in Finland, Helsinki Metropolitan area and especially Capital Region area of Helsinki (CRA) (Pääkaupunkiseutu) are reflecting the same pattern. Immigration is a relatively new process in Finland, in comparison to Great Britain, Germany, Sweden, USA, New Zealand and Australia. Traditionally Finland was emigrational country (Koivukangas, 2003). After the economic crisis at the beginning of 1990's, Finland became stable and fast growing economy, one of the most prosperous in Europe. Some of the reasons for this could be joining European Union as well as fast development of IT industries (e.g. NOKIA) which led to arrival of highly skilled migrants. This period is marked by increasingly huge immigration in country where the number of foreign arrivals has increased for five times throughout the decade, where 20 % population were refugees and huge number

of Ingrian Finns due to the collapse of USSR (Koivukangas, 2003). HMA is the biggest immigration hub (Heikkilä & Peltonen, 2002) in Finland. As an administrative, cultural, educational and economically most prosperous region with the biggest amount of jobs, HMA represents most desirable final destination for majority of the immigrant population.

Migration and immigration have been studied extensively by social sciences. Small number of published scientific studies has been dealing with notion of migration and immigration through spatial concepts. Geography, as a science in its core is dealing with spatial concepts as one of the center problems. Disproportionately low number of studies in geography is dealing with the phenomenon of migration and immigration through exploration of spatial concepts (Samers, 2010). Most of the previous studies dealing with immigration population in HMA have been using sociological, socio-economic and descriptive statistical approach to describe spread, concentration and spatial location of immigrant population (Vilkama, 2007, 2011; Vilkama & Dhalmann, 2009). Some studies have explored social phenomenon of immigration from specific geographical perspective – in terms of "space", "place" and "scale" (Samers, 2010). As Vaattovaara (2001) is briefly acknowledging, "*Pattern of migration from foreign countries (Former Soviet Union and Africa) has also revealed a spatially clustering pattern"*, "space", "place" and "scale" have started to matter in immigration studies in Finland. Vaattovaara's work, marks the beginning of studying social aspects and spatial patterns of immigrant population with more extensive use of GIS methods.

According to Fotheringham et al. (2000), the main task of geographical research is to create understanding about processes which are affecting creation of spatial patterns on the surface of the Earth. However, the notion of spatial location and correlation among them as a core of geographical investigation has not been widely used. Spatial data is special and different from any other data and therefore spatial data needs to be treated differently than other types of data (Anselin, 1989; Anselin & Getis, 1992).

One of the first studies with advanced GIS approach that used spatial location analysis in Finland, was PhD dissertation by Vaattovaara (1998). Method of factorial analysis was implemented in analysis of certain economic aspects of life and their spatial manifestation in terms of spatial location of population of HMA. However, only few of the previous population studies in Finland (Vasanen, 2009; Lehtonen & Tykkyläinen, 2010) have been employing concept of spatial autocorrelation, conceptualized with spatial statistics approach,

implementing use of exploratory spatial data analysis (ESDA) methods as an analysis tool. Motivated by studies conducted by Vaattovaara (1998, 2001 and 2002) and Vilkama (2007 and 2011), dealing with immigration population, this thesis is going to implement concept of spatial autocorrelation as a center question of investigation of the processes of spatial clustering and formation of specific spatial clusters in HMA area among immigration population. Spatial autocorrelation represents reliance between values of variable in neighboring or contiguous locations (Griffith, 2009). According to Anselin (2008), clustering represents pattern as a whole and cluster is a specific location. Detection of clustering as a global process and occurrence of clusters in specific locations of immigrant population in HMA represents main goal of this thesis. Vaattovaara (1998, 2001 and 2002) and Vilkama (2007 and 2011) provided meaningful input for possible detection of specific locations and occurrences of specific local clusters in different locations across the study area.

This thesis represents comparative study of computing capabilities of ESDA methods (***global and local Moran's Index***) performed in two GIS software packages (ArcGIS and GeoDa). Quality and accuracy of the results (maps, statistical values, etc.) are going to be tested and presented. ArcGIS is a market leading, commercial GIS package for computation, analysis and production of different sorts of GIS analysis and results. Spatial statistic toolbox, as integral part of ArcGIS software package is used for interpretation of spatial statistics results (maps, graphs, reports etc.), which can be obtained, by use of several different methods. GeoDa is non-commercial software, relatively new in GIS practice in Finland, focusing specifically in spatial statistics analysis. It is used for manipulation and operationalization of spatial data analysis, designed for implementation of different and unique (Bivariate Moran's *I*, etc.) ESDA techniques. Both software are computing comparable but different results, quantitatively and visually. Correct measurement of spatial autocorrelation is required but at the same time it is *"open to a wide variety of subtle variation"* (O'Kelly, 1994). Using spatial autocorrelation, scientist is responsible for appropriate statistical use of the gained results and comprehends key role of the units and scale of analysis in the process of gaining final results (Chou, 1991). This conceptualization is opening up question of scale and Modifiable Areal Unit Problem (MAUP) being analyzed, as well as the results of clustering and their interpretation in certain scales.

Theory lying beneath proper statistical functions running the operation of different spatial autocorrelation processes being produced, needs to be completely understood by final user (Haining, 1978). Outcome is a comparison of computing capabilities, analyzing patterns and

production of maps, graphs and other results conducted by ArcGIS and GeoDa. Performing spatial autocorrelation is meaningless without adequate use of gained information, performed by trained user (Getis, 1991). Interpretation, question of accuracy and meaningfulness of produced results is a core question of this thesis. Insights of development of certain patterns of spread of immigrant population and possible discovery of new trends, which were, not present or noticed before is the main challenge.

1.1 Motives of the study

Geographic information systems are conducting four elementary functions on space data: input, storage, analysis and output (Goodchild, 1987). Spatial analysis has a wide range of different techniques, from basic description all the way up to complex modelling based on inferential statistical methods (Anselin & Getis, 1992). Previous immigration studies conducted in Finland and HMA area (Heikkilä & Peltonen, 2002; Vaattovaara, 2001, 2002; Vilkama, 2011), have been using less advanced methods. However, these studies used visual representations rather than spatial statistical methods thus they can be therefore categorized as descriptive statistical studies. GIS software like MapInfo and ArcView or statistical packages like SPSS and spreadsheet like Excel, were used for a descriptive representation of exploration of certain spatial socio-economical investigations, mostly conducted by employment of qualitative methods or non-spatial methods.

ESDA methods are exploratory data analysis methods, which are specifically paying attention to dimension of space of certain data being analyzed (Anselin, 1996). Used by GIS oriented computing software (ArcGIS and GeoDa), ESDA methods are going to be employed to analyze patterns of spatial distribution through the concept of spatial autocorrelation of immigrant population in HMA.

Main motive of the study is improvement of quality of analyzing methods and techniques for analysis of immigration patterns in HMA, which are based in quantitative geography and spatial statistics methods. ESDA methods have played important role in the concept of integration of spatial analysis and GIS (Anselin & Getis, 1992; Goodchild et al., 1992). Employing the usage of quantitative methods for measuring of clustering and clusters of immigrant population in this case, represents a second phase. If we would say that a first phase in exploration of clustering would be use of descriptive statistical methods (e.g. percentage or percentile map) for detection and location of immigrant population (Vilkama, 2011) (Figure 1), then the second phase would be use of ESDA methods allowing us more

precise research and conceptualization on the patterns of spatial autocorrelation of immigrant population. ESDA methods are using inferential statistical approach in defining certain undiscovered patterns, which cannot be determined and confirmed, by the use of descriptive spatial statistics methods. Thesis is conceptualizing and focusing on use of inferential statistical methods, more accurately explanation of spatial aspect of distribution of immigration population. It will try to introduce a new perspective and methods, complementary but different from the previous studies dealing with immigration population and HMA.

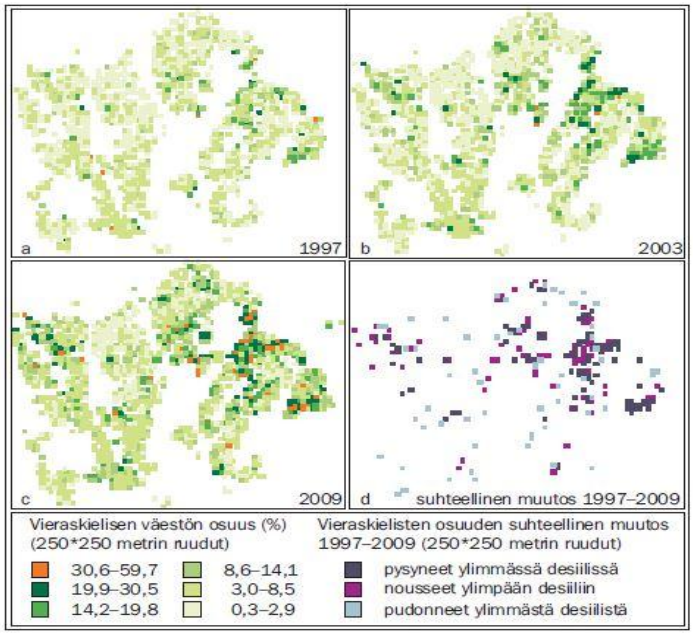


Figure 1. Descriptive statistic map (Vilkama, 2011).

Explanation of functionality of specific ESDA methods for measurement of spatial autocorrelation, determination of possibility to compute and compare statistically significant and precise results, computed by ArcGIS and GeoDa is a main aim of this study. Question of proper lattice level size, analyzing area, production of spatial statistically meaningful results and introduction of new methodological approach in measuring spatial autocorrelation of immigrant population in Helsinki Metropolitan Area is the main task.

1.2 Aims and key questions of the study

The aim of this thesis is employment of spatial statistical methods, based on inferential statistical (ESDA methods) and comparison of their functional capabilities and performances. Testing performing capabilities, computing results and visual representation is going to provide useful information about potential use and performance of this methods. Main idea,

besides explaining the process of making displayable and visually precise representation of clustering of immigration population is to introduce inferential statistical methods in research of immigration population in HMA. Introduction of these methods can help us to better understand underlying processes going on in the area. At the same time it will try to present specific spatial clusters and hot spots of immigration population in certain areas. Contiguity, as a concept lying beneath spatial autocorrelation hasn't been used in the previous studies (Vaattovaara, 2001 and 2002; Dhalmann & Yousfi, 2010*;* Vilkama, 2011) of immigration population of HMA. This thesis is trying to implement a new approach by utilizing spatial statistics methodology in the research field of immigration population concentration in Helsinki Metropolitan Area (HMA). The main research questions of this thesis are:

1) *What are the differences in computational capabilities of ESDA methods performed in ArcGIS, leading commercial desktop GIS software and GeoDa, free, open source, cross-platform GIS software?*

2) *How is lattice (cell size) affecting spatial distribution of specific clustering values and its spatial distribution in analyzed area?*

3) *How is the scale (MAUP) influencing specific clusters results?*

4) *Is there some specific clustering of immigrant population not noticed beforehand in other similar studies?*

5) *Is the concept of spatial autocorrelation providing us with a different visual and quantitative explanation of specific clustering of immigration population in HMA area?*

The main hypothesis of this thesis is that use of advanced ESDA methods in discovering and mapping potential clusters of immigration population of HMA is providing new visual, statistical and presentational capabilities which are changing, improving and providing more precise information on the level of clustering and its physical distribution throughout certain specific areas and HMA as a whole.

2. EMPIRICAL, THEORETICAL AND METHODOLOGICAL FRAMEWORK

2.1. Immigration

Emigration vs. Immigration

" A person who changes residence from one country to another one is considered an emigrant relative to the country of origin and immigrant relative to the country of destination." (Peters & Larkin, 1999)

Immigration is the process in which one person lives a country of origin to come and live permanently in a foreign country (Oxford Dictionaries, 2014). Due to the frequent misinterpretation of the terms emigration and immigration, need for precise explanation is acquired. To emigrate is to leave domicile town, region or country to settle in another. It doesn't necessarily means to displace from a country of origin, but for certain it means migrating from a place of birth and moving to another region or country. On the contrary, immigration means moving and residing in a country which is not ones native country on a permanent basis. Distinct difference in both terms is a country of origin and dependent on the destination country, person can be determined as immigrant or emigrant (Diffen, 2014), see also Figure 2.

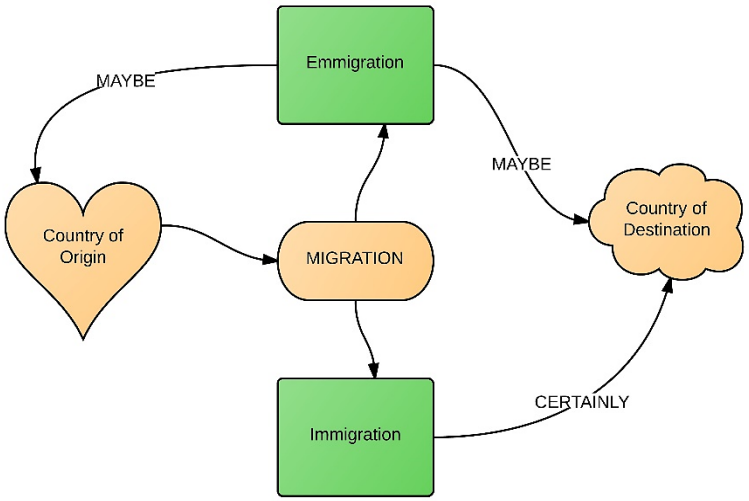


Figure 2. Schematic display of emigration vs. immigration (Kekez, 2014).

2.1.1. Migration trends and policies in Europe

Migration research is not homogenous field; it uses a wide range of theories to explain migration. Classical approaches are based on economic factors which are explaining migration processes on the global level or decisions to migrate on the local level. Changes in migration processes since the 1990s cannot be explained by classical theories. New approaches, explaining contemporary migration trends are focusing on the *"meso-level of migration through exchange processes between social spaces"* (Kepsu et al., 2009). Migration flows across Western Europe at the beginning of the twenty-first century are facing us with complex and confusing picture. These flows can be categorized in four different trends, or concepts of entering the country (Stalker, 2003):

1. *Labor migration* - long- and short-term immigrants and seasonal workers
2. *Family reunification* - attempts of close relatives to join the family which already has long-term settlement rights
3. *Undocumented workers or "illegal immigrants"* - are people who entered the country illegally or have entered on tourist visas and have overstayed, most of the time in order to work
4. *Refugees* - asylum seekers have been granted asylum

For following up all this different flows and for setting up the context for explanation of modern migration, most precise starting point is the end of World War II. Four different phases can be noticed:

Late 1940s and early-1950s – Refugee movements

After Second World War, fifteen million people moved from one country to another, with a great number of them forced to relocate themselves as a consequence of changes of borders, especially between Germany, Poland, and the former Czechoslovakia. Refugees made up thirty percent of the whole population of West Germany till 1950 (Borrie, 1970). These trends started to slow down by the mid-1950s, but still they were present all the way till the Berlin wall was not raised in 1961.

Early-1950s to 1973 – Labor force migration

Revival and reconstruction of the Europe resulted in economic boom. OECD countries average annual growth rate of the economies was around five percent between 1950 and 1973. Major economic forces Germany, France and UK were under the process of revival and

reconstruction of its own economies and reconstruction which caused a big demand for labor force. They started to experience shortage of workers, which was replaced among the population of displaced during the war which was still not enough. Migration from less urbanized and industrialized parts of south-European countries started to occur, predominantly from Italy, Portugal and Spain. Processes of urbanization and industrialization in these countries started to develop rapidly, which caused big shift in the migration process, from emigration countries, they became migration desired countries. In the new prospect, development of migration flows of France and UK has shifted. Migration policy was focused more on the old colonial countries which in the case of France was countries of North Africa and in the case of UK, Caribbean and Indian subcontinent. Germany, which was less dominant colonial force in the past focused more on the policy of short-term contract workers from former Yugoslavia and Turkey. *"Net immigration for Western Europe reached around 10 million (compared with net outflows of 4 million for the period 1914 to 1949)"*(Stalker, 1994).

1974 to mid-1980s – Restrictive politic of migration

Immigration policy has become stricter and with more limitation towards future immigrants already in the 1960s, which for example in the UK caused lowering down the number of people who could have a possibility to emigrate from British Commonwealth. Recession and oil crises in 1973 caused by OAPEC oil embargo, affected migration policies directly by imposing further restrictions in labor immigration and expectance that previous immigration should return to country of origin. Most of the governments allowed previous immigration to stay and allowed family members of existing immigrants to join. Even earlier, but especially in these new circumstances immigration has started to shift from West Europe to South Europe to countries which have now developed themselves to become strong and respectable economies, which was the case with Italy. By joining EU other South European countries got much needed economic "injection" which made them attractive destination for immigrant population.

Mid-1980s to 2001 – New trends (asylum seekers, refugees, and illegal immigrants)

This period was represented by turbulent and rapid political change, especially in the East Europe marking the end of the communism and beginning of shifting to neoliberal capitalism. Opening up of the "Iron Curtain" marked a new period of immigration from East Europe which was already traditionally known as emigration hub (huge immigrations caused by

famine, terrible living conditions, armed conflicts and similar reasons have demographically marked beginning of 20th century) joining already existing immigration trends in West Europe. This phenomenon had been evident as far back as 1980 when some 108,000 Turkish citizens applied for asylum in West Germany. From 1989–1998, more than 4 million people applied for asylum in Europe, 43 per cent of whom came from elsewhere in Europe, 35 per cent from Asia, and 19 per cent from Africa (Salt, 2000). Under the pressure of constant growth of the population of asylum seekers governments of West European countries have started to sharpen even more policy of asylum seeking. That raised up number of illegal immigrants, which were traveling either by themselves or through different modes of human trafficking.

Understanding flows and phases of migration is important for clarifying processes of international migration and understanding different perspectives for exploring the phenomenon. Explaining the current situation in Finland cannot be done without combinations of approaches which are offering more fitting perspective. The best way to describe the process of international immigration to Finland is probably, taking it to account as a part of Nordic migration system which is integral part of a bigger European migration system. Theory of migration systems is defining migration as just one of many intensive exchanging processes (information, goods, ideas, capital, persons etc.) among specific countries which for the end result has creation of the stable system. Conceptualization can be done through connection of several countries of emigration to the one region of immigration or by different approach marking one country as an emigration, but migration spread through many different regions of immigration. Main role in putting up this kind of system is played by *"social and ethnic networks, multinational firms, educational institutions or other corporations - as mediators between macrostructures and individuals as well as between the different countries"* (Kepsu et al., 2009). This theory is focusing on different aspects of migration systems (political, economic, social, demographic and historic), but without dealing with the problem of genesis of migration systems. Spatial proximity is not dealt with, but the main focus is on the influences caused by political and economic relations considering migration systems (Fawcett, 1989; Kepsu et al., 2009).

Taking into account EU as developing migration system, beneath which Nordic migration system comes as integral part, Finland and Helsinki are representing just a micro level in which theory of migration systems can describe processes of migration. Finland started to

gain more immigration population, around the time country entered EU integrating itself within a bigger migration system, which is constantly redefying and posting new forms of interdependencies and transfers within itself and towards outside as well. Beside Nordic migration subsystem, there are also other migration subsystems operating and functioning within EU, like subsystems in Central and South Europe. Considering all this complexity of the different levels of systems and their mutual correspondence and interoperability, there is strong indices that there is a huge number of transnational migrants moving between Nordic countries due to proximity and similarity in culture and within EU in general as a part of bigger system. As a part of this system, Finland is strongly connected to the north-west axis because of the long-lasting Nordic cooperation and migration tradition, and to the east axis along the Baltic countries, especially Estonia. Finland is also attracting a growing number of migrants from other parts of the EU and outside the EU migration system (Kepsu et al., 2009).

2.1.2. Immigration in Nordic region

Most of the migration in Nordic region is happening between countries of the region. There is many reasons and treaties signed between countries which have created that kind of situation: free labor market, languages which are related (exception of Finnish and Sami) and favorable rules allowing studying anywhere in the region. All these things are making moving in between countries easy.

Immigration to the region is marked by two types of immigrants. One type is represented by citizens returning to their home country and other one are citizens of foreign countries who have been granted residence permit. Analyzing situation country by country, proportion, percentage and type of immigrant population is varying.

Sweden is country with a largest proportion of immigrant population, with 13% of its all population being represented by people born in other country than Sweden. *The amount of immigrants and refugees in the country of about 8.8 million inhabitants has risen to over one million; several hundreds of thousands are from countries outside the so-called western world* (Hannikainen, 1996). Iceland is quite close to Sweden with 10 % of population, Denmark and Norway are having slightly smaller number of same population (8%), while Finland leads with the smallest number (less than 4%). The number of immigrants has increased in the whole region due to the different global reasons and global migration causes.

Citizens of Iraq are representing the largest population group from one country. There are 64,000 of them spread out through the region, from which half of them is living in Sweden. Region data from 2007 is marking 46,000 Polish citizens and 35,000 citizens from Baltic countries spread out through the region. In the region there is 45,000 Turkish citizens living, from whom most of them live in Denmark. Russian citizens are marking 44,000 people spread out through the region, of whom more than a half is living in Finland and majority of the rest of this population is living in Norway and Sweden. (Norden, 2013). Finland and HMA, as the largest immigration hub (Heikkilä & Peltonen, 2002) are special in comparison with other European and Nordic capitals by having short immigration history and peripheral location. Tradition of immigration is short and the city's position on the hierarchy of world cities is still relatively marginal. Earlier in the European history, the migration patterns in Finland were marked by emigration to other countries and can be stated by the classical macro-economic explanations for labor migration.

2.1.3. Immigration in Finland

In Finland immigrants are defined (Statistic Finland) by their nationality, country of origin or mother tongue. This thesis is going to use Finnish statistics population data. Collected data determines nationality of one person based on a citizen's mother tongue. Based on data collected on this principle determination of immigration background of citizen is acquired. Native population is constituted from citizens which mother tongue is Finnish, Swedish or Sami. Because of small number of foreigners residing in Finland before 1990's most of population having some other than Swedish or Finnish language for a mother tongue belongs to a group of recent migrants. Language statistic is based on a personal declaration of the mother tongue claimed by each individual.

Finland traditionally represented country of emigrants. Historical development of population in Finland was hugely affected by emigration of population from the middle of 19th century all the way up to beginning of 21st century. For the country which population is 5 472 421 inhabitants (Statistics Finland, 2014) at present, number of 1.3 million Finns which have emigrated since 1860's represents a huge population. Since II World War 755 000 people have emigrated from Finland, with a peak at the beginning of 1970's (Koivukangas, 2003). Since 1970's emigration of Finns have started to decline, but the immigration to Finland started to increase. Next decade was marked by complete change in trends of emigration from

and immigrating in Finland. This period was marked by receiving more immigrants than, emigrants leaving Finland (Heikkilä & Peltonen, 2002) (Figure 3).

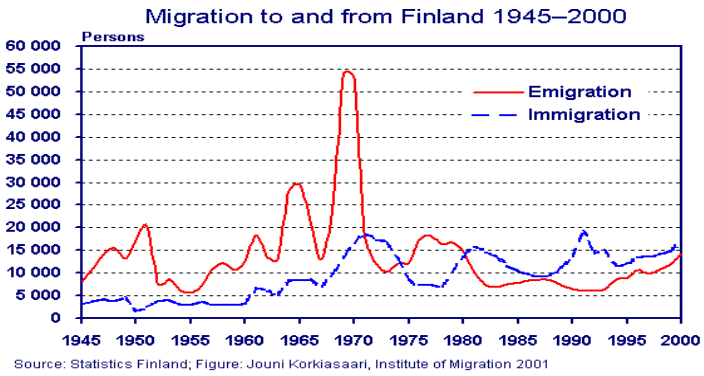


Figure 3. Emigration and immigration in Finland 1945–2000 (Heikkilä & Peltonen 2002).

At the beginning of 1990's number of foreigners was around 21000 people, approximately 0.4% of the total population of Finland. In next twelve years situation has rapidly change, so in 2002 there was around 100000 foreigners living in Finland and at that point representing 1.9% of the total population (Figure 4). That was one of the lowest percentages in the EU and Europe in general. In a comparison, if Finland would have proportionally same amount of immigrants like e.g. Germany, this population would count half a million (Koivukangas, 2003).

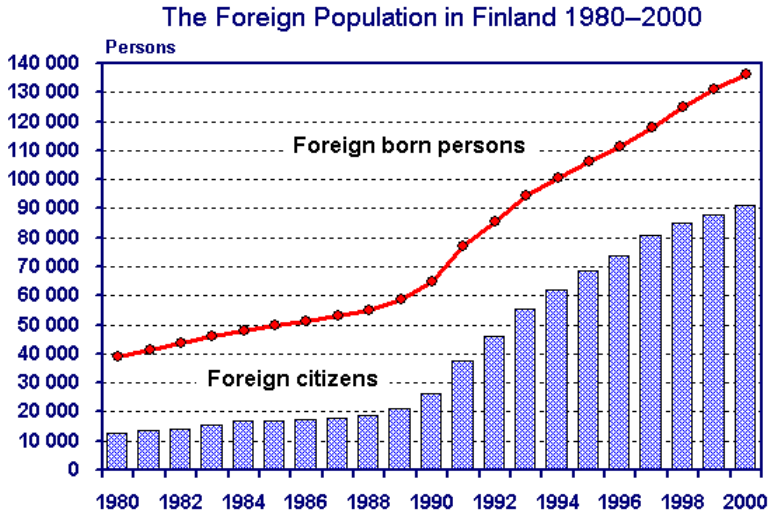


Figure 4. The foreign population in Finland 1980–2000 (Heikkilä & Peltonen 2002).

Global changes which have been happening in world politics (end of the Cold War, influence of neoliberal capitalism, brake of USSR and Yugoslavia, Falling of the Berlin wall, civil war

in Somalia, Iraq, Yugoslavia etc.), also have affected market of labor force and had a strong influence in recent decades on a huge increase of immigration in Finland. Proportion of permanent residents with immigrant background born outside of Finland is 4.4 % in 2009, which is still quite low in comparison with other Nordic countries (Dhalmann & Yousfi, 2010).

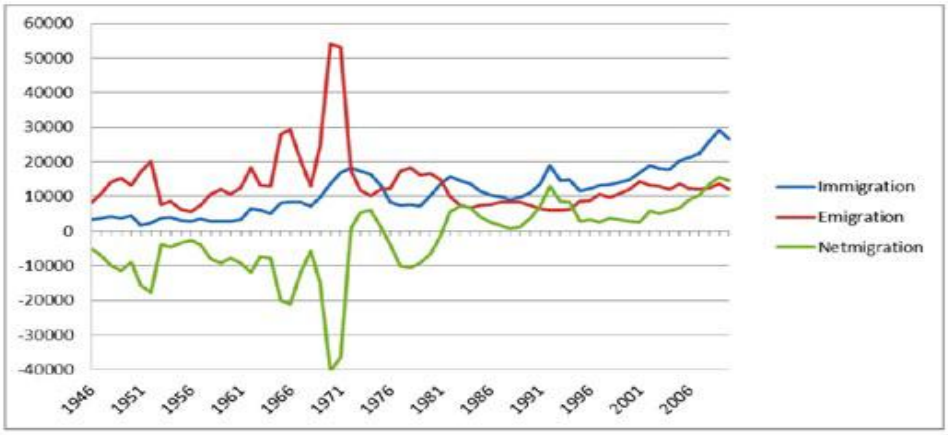


Figure 5. The migration balance (Dhalmann &Yousfi, 2010).

Recent trends show increase in immigrant population, especially in 2000s. Immigrant population at the beginning of 1990s represented 1.3% of total population, in 2000s that was 2.6% and in 2009 it reached 4.4% of total population (Figure 5). Majority of this population has settled in Helsinki region and other highly urbanized areas, where proportion is higher than average (Dhalmann & Yousfi, 2010).

Official data for 2014 is informing that there is 195.511 foreign born people residing in Finland, which corresponds to 3.77% of complete population. If we add Russian speaking population which consist 1.15% we are getting around 260000 inhabitants or 4.92% which are not having Finnish, Swedish or Sami as their mother tongue language (Statistics Finland, 2013).

2.1.4. Immigration in Helsinki Metropolitan Area

Immigration processes in Helsinki Metropolitan Area (HMA) area can be explained by a pattern different from previous immigration movements. New types and forms of migration have appeared in majority of EU countries, including Finland (Chapain et al., 2010). The densest concentration of immigrant population is in South Finland, specifically in Helsinki Metropolitan area (Koivukangas, 2003). In this part of the country lives 50 % of all foreign

citizens (Forsander, 2003). Out of total immigrant population 40% is living in Helsinki conurbation. Finnish government tried to settle immigrant population in less inhabited and sparsely populated areas of Finland (especially refugee population) due to different reasons and motives. Remigration within Finland has happened and majority of this population has settled in Helsinki metropolitan area. HMA municipalities (Helsinki, Espoo and Vantaa) and Municipality of Turku are the only municipalities with immigrant population bigger than 5000 people (Forsander, 2001; Heikkilä & Peltonen, 2002; Kokko, 2002). In 2009 out of total population, 6.7% of population was population with a foreign background (Dhalmann & Yousfi, 2010).

Elaborated facts are determining HMA area as a space with a biggest concentration of immigrant population, which is leading to the fact that the biggest possibility of spatial concentration and distribution of this specific population is emerging in this area. Spatial distribution of clustering and clusters of immigrant population needs to be determined in this specific area, by the use of ESDA methods provided by different GIS software.

2.2. Quantitative geography and spatial statistics

2.2.1. Quantitative geography

Quantitative geography is conducting one or more actions such as: analysis of numerical spatial data, development of spatial theory and construction and testing of mathematical models of spatial processes, aiming at understanding spatial processes better. Main objective of quantitative research is to optimize output of spatial processes with minimizing error percentage. Its specificity is in dealing with spatial data (Fotheringham et al., 2000).

Classification is a basic human mental process (Milligan & Cooper, 1987). Grouping of entities can help us to better understand going on processes, allow us better predictions in assessing certain phenomena and help us develop possible theoretical constructions.

From the philosophical point of view, two major directions in quantitative geography can be recognized and defined as naturalist and anti-naturalist (Graham, 1997). Naturalist movement tended to set up human geography as a spatial science with general laws, which can be particularly seen in the works on migration research. Researchers used methods adopted from physics (gravity model-explaining and predicting movement of people) and tended to treat processes in geography conceptualizing them as the laws in physics (Graham, 1997). This

approach is searching for general (global) "laws" and general (global) relationships (Fotheringham et al., 2000). Anti-naturalist concept is completely opposite, disapproving naturalist concept. Their philosophy is conceptualized on analyzing variations of the relationships over space by the use of so called "local" forms of analyses (Fotheringham, 1998; Fotheringham & Brunsdon 1999). Conceptualizing these approaches is the center of methodological research in analyzing space and location of certain phenomenon and interactions in between, qualifying quantitative geography as a logical path in describing phenomenon of spatial autocorrelation of immigrant population in the HMA area.

Quantitative geography is prevailingly focusing on spatial data, which makes clear distinction in comparison with econometrics or quantitative sociology. *"Spatial data are those which combine attribute information with locational information" (Fotheringham et al., 2000).*

2.2.2. Spatial statistic and spatial autocorrelation

Spatial data analysis is a statistical study of certain phenomenon manifested in space (Anselin, 1996). Special techniques and methods are developed for classification of objects which have topological, geometric and geographic properties. All together these techniques are called spatial analysis or spatial statistics techniques and mostly they are used in the analysis of different geographic data and its spatial dispersal.

With advanced development of computers many automatic spatial techniques algorithms have been created or re-introduced from field of statistics for measuring different sorts of spatial dispersal (Mantels test, Pearson's correlation test, Moran's *I*, Geary's *C*, Getis-Ord General *G*, etc.). Probably the best term describing this process is "geocomputation". Geocomputation represents quantitative analysis conducted by computer in which computer is having a key role (Fotheringham, 1998).

Spatial statistics comprises a set of techniques for describing and modeling spatial data. In many ways they extend what the mind and eyes do, intuitively, to assess spatial patterns, distributions, trends, processes and relationships. Unlike traditional (non-spatial) statistical techniques, *spatial* statistical techniques actually use space – area, length, proximity, orientation, or spatial relationships – directly in their mathematics (Scott & Getis, 2008; Scott & Janikas, 2001). There are many different types of spatial statistics: descriptive, inferential, exploratory, geostatistical and econometric statistics are just some of the most widely used (ESRI, 2013a).

Inferential statistic is trying to reach conclusions that extend beyond immediate data alone and it's opposite to descriptive statistics, which is organizing and describing already existing data (Rice, 2003). Methods used in this thesis belong to inferential statistic. Inferential statistical techniques are using statistical tests, which are gathering accurate probabilistic inferences from data set (Taylor, 1977).

"Spatial autocorrelation is a measure of the degree to which a set of spatial features and their associated data values tend to be clustered together in space (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation)." (ESRI, 2013b)

Spatial autocorrelation is trying to understand the degree of similarity between objects or activities on one spot of Earth's surface and location nearby. ***"First law of geography"*** defined by Tobler (1970)*: "Everything is related to everything else, but near things are more related than distant things"* has described spatial autocorrelation in the most precise manner (Goodchild, 1987). If we have certain variable *Z*, which we are observing on certain spatial location *s* which is determined by certain coordinates *x* and *y* then we can explain spatial autocorrelation as a correlation between $Z(s_i)$ and $Z(s_j)$. Autocorrelation is the correlation of variable with itself, but spatial autocorrelation is correlation of variable with itself on different spatial locations (Schabenberger & Gotway, 2005).

Spatial autocorrelation modeling started to develop more at the end of 1940's and throughout 1950's. At the end of that decade Moran (1948) revealed Moran's Index. Some year afterwards Geary (1954) has implemented same but slightly different concept, by presenting Geary's C. The work of Whittle (1954) was important additional contribution to the field. Based on these works following example of older colleagues, Cliff & Ord (1969, 1970) are employing revolutionary concept of spatial autocorrelation.

Further development, especially visual representations of the gained results of the inferential statistics were developed by John Tukey (1977). His work was extremely important for the development of what today we know as Exploratory Spatial Data Analysis (Anselin, 1996, 1999; Messner et al., 1999) with his concept of Exploratory Data Analysis (EDA). It marked a huge discovery at that time and it opened up new horizons and possibilities, for further development.

In following years, spatial autocorrelation analysis has been used increasingly for making inferences concerning the factors that underlie observed patterns of spatial variation in processes like human and animal migration (Sokal et al., 1988). Contemporary analysis is

marked with Exploratory Spatial Data Analysis (ESDA) conceptualized by Anselin (1996), following up the path and tradition of Tukey.

"*Exploratory spatial data analysis (ESDA) techniques are used for specific analysis of spatial characteristics of data, their spatial association or heterogeneity. Their task is to focus on describing spatial patterns of association (spatial clustering), spatial regimes and identifying spatial outliers" (Anselin, 1996).*

Statistical equations used for the calculations by these methods are the same in ArcGIS and GeoDa (Anselin & Rey, 2010). Final outcome of their results is interesting for comparison and further analysis. Theoretical and methodological approach of global and local methods of spatial autocorrelation in ArcGIS and GeoDa is almost the same, but visual representation of the gained results is slightly different. Therefore, certain prerequisites are needed to be taken into account before global and local methods of spatial autocorrelation are implemented.

2.2.3. Global method of spatial autocorrelation

Notion of spatial autocorrelation exploited in GIS, through use of spatial statistics methods has its own strict structure and limitations. When global methods of spatial autocorrelation come into account Concept of null hypothesis has to be tested so the validation of the result can be obtained.

2.2.3.1. Concept of null hypothesis in spatial statistics

Gained results of inferential statistical methods are always interpreted within null hypothesis. The null hypothesis is claiming that in the certain area there is Complete Spatial Randomness (CSR) of certain phenomenon, features themselves or of the values that are associated with that certain feature. When results of z-score and p-value are obtained by calculations of the tool, null hypothesis can be either accepted or rejected.

The p-value represents probability that observed spatial pattern was created by some random process. When p-value has a small value, it tells us that observed spatial pattern is created by some random process, so the null hypothesis can be rejected. Values of z-score are representations of standard deviations. Both of these values are connected with standard normal distribution represented in Table 1.

Visually displayed in Figure 6 very high or very low values of the z-scores which are connected with a very small p-values can be found at the ends of diagram. When the pattern being observed is run by the tool and as a result in return outcome is small p-values and either very high or very low z-score it informs us that results of the spatial pattern being analyzed is not representation of theoretical background behind null hypothesis of Complete Spatial Randomness (CSR). For the rejection of the null hypothesis decision have to be made about confidence level. Most common confidence levels are 90, 95 or 99 percent. On the basis of confidence levels and different combinations of values of p-value and z-score we can determine can we reject or accept null hypothesis

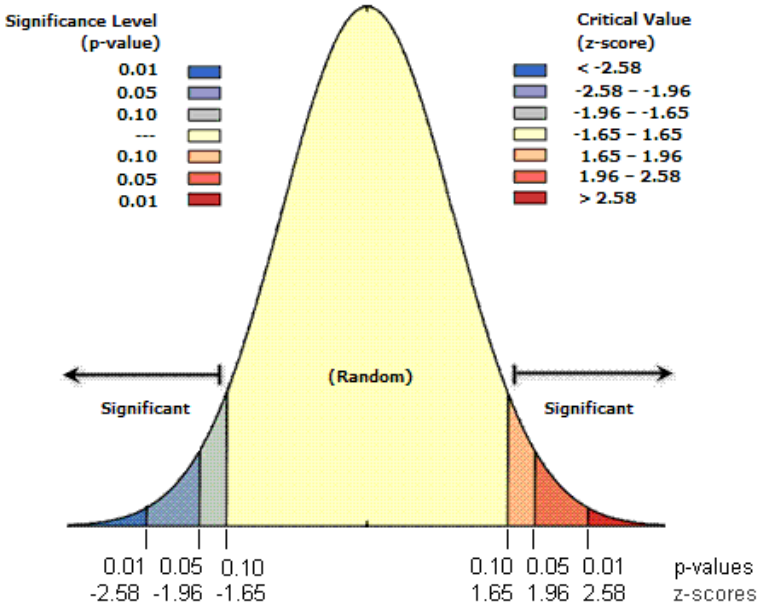


Figure 6. Visual interpretation of distribution of Significance Level (p-values) and z-score in ArcGIS (ESRI, 2013c).

If the values are within the range than it is acceptable to reject the null hypothesis, it is a common practice then to analyze what is causing statistically significant structure in a spatial dataset. Values in the middle of the distribution in Figure 6 are representing expected outcome and they are telling us that nothing unusual is happening with data, but if z-score values have a large value and p-values are small (located at the ends of normal distribution), than it means that there is something statistically significant happening with data (clustering or dispersion).

Table 1. Critical values of p-value and z-score for different confidence levels (ESRI, 2013c).

z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%

2.2.3.2. Global Moran's Index (GMI)

First measure of spatial autocorrelation was presented by Moran (Moran, 1948, 1950). He was studying random or nonrandom distribution of certain phenomena in space in one or two dimensions. It is used to calculate the strength of correlation between observations as a function of the distance separating them (Oliveau & Guilmoto, 2005).

Moran's Index is calculating spatial autocorrelation, similarity between certain features, which is based on a feature location and values for that certain feature simultaneously and at the same time multi-directionally. It compares neighboring areal units over complete study area, and informs us about positive spatial autocorrelation (clustering) if the neighboring units have similar values. If the values of the neighboring units are dissimilar it indicates negative spatial autocorrelation (dispersal) (ESRI, 2013d). Dispersion with geographic data is less common than clustering, but might be seen with some kind of competitive or territorial spatial process, where similar features try to be as far away from each other as possible.

Mathematics

Global Moran's Index is defined as (Getis and Ord, 1992)

The Moran's *I* statistic for spatial autocorrelation is defined as:

$$\boldsymbol{I = \frac{n}{S_o} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}}$$

(1)

$\boldsymbol{n}$ – is total number of spatial units indexed by $\boldsymbol{i}$ and $\boldsymbol{j}$

$\boldsymbol{i}$ and $\boldsymbol{j}$ – are spatial units

$\boldsymbol{z_i}$ – is deviation of an attribute for feature i from its mean $\left(\boldsymbol{x_i} - \overline{\boldsymbol{X}}\right)$

$\boldsymbol{x_i}$ – is variable of interest

$\overline{\boldsymbol{X}}$ – mean of x_i

$\boldsymbol{w_{i,j}}$ – is the spatial weight between feature i and j

$\boldsymbol{S_o}$ – is the aggregate of all the spatial weight

$$\boldsymbol{S_o} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{w_{i,j}}$$

(2)

The z_I score for the statistic is computed as:

$$\boldsymbol{z_I} = \frac{\boldsymbol{I} - \boldsymbol{E}[\boldsymbol{I}]}{\sqrt{\boldsymbol{V} - [\boldsymbol{I}]}}$$

(3)

which is based on:

$$\boldsymbol{E}[\boldsymbol{I}] = \ -\boldsymbol{1} \ / \ (\boldsymbol{n} - \boldsymbol{1})$$

(4)

$$\boldsymbol{V}[\boldsymbol{I}] = \ \boldsymbol{E}[\boldsymbol{I^2}] - \ \boldsymbol{E}[\boldsymbol{I}]^2$$

(5)

Additional calculations for Moran's Index:

$$\boldsymbol{E}[\boldsymbol{I^2}] = \frac{\boldsymbol{A} - \boldsymbol{B}}{\boldsymbol{C}}$$

(6)

$$\boldsymbol{A} = \boldsymbol{n}\ [(\boldsymbol{n^2} - \boldsymbol{3n} + \ \boldsymbol{3})\ \boldsymbol{S_1} - \boldsymbol{nS_2} + \boldsymbol{3S_o^2}]$$

(7)

$$\boldsymbol{B} = \boldsymbol{D}\ [(\boldsymbol{n^2} - \boldsymbol{n})\ \boldsymbol{S_1} - \boldsymbol{2nS_2} + \boldsymbol{6S_o^2}]$$

(8)

$$\boldsymbol{C = (n-1)(n-2)(n-3)S_o^2}$$

(9)

$$\boldsymbol{D = \frac{\sum_{i=1}^{n} z_i^4}{(\sum_{i=1}^{n} z_i^2)^2}}$$

(10)

$$\boldsymbol{S_1 = (1/2)\ \sum_{i=1}^{n}\sum_{j=1}^{n}(w_{i,j} + w_{j,i})^2}$$

(11)

$$\boldsymbol{S_2 = \sum_{i=1}^{n}(\sum_{j=1}^{n} w_{i,j} + \sum_{j=1}^{n} w_{j,i})^2}$$

(12)

Mathematics behind equation is computing mean and variance for a certain attribute from a data set which is being evaluated (i or j). For every feature value, mean is calculated, by creation of *deviation from the mean (zi or zj)*. Deviation of the mean is calculated, by calculation of difference between each value in data set and mean. After that deviation values for all the neighboring features (neighboring grid cells in this case) are multiplied together to form a *cross-product*. The cross-products of the deviations from the mean are then summed for all pairs of areal units as long as they are neighbors.

Cross-product's results can vary, dependent on the feature values, value of mean and deviations in data. Because of this summed cross-product is always used in this version of Global Moran's Index equation. If both neighboring values are above the mean, the product is a positive number. Product is negative, if both neighboring values are below the mean (product of two negative numbers). So the bigger value of deviation from the mean is, the higher cross-product result is.

When values in dataset have intention to cluster spatially (high value clusters close to other high value clusters and low value clusters close to other low value clusters) Global Moran's Index is positive, which reflects the presence of positive spatial autocorrelation, where similar values are next to each other.

But if the value of one areal unit is above the mean and the value of the neighboring unit is below the mean, which are at the same time neighboring units, the product of the two mean deviations will be negative, indicating the presence of negative spatial autocorrelation and a negative value of Global Moran's Index.

The final result which can occur is that positive and negative cross-product values are in balance, which would lead to that Global Moran's Index value would be zero. Global Moran's Index values are ranging between -1 and +1.

The denominator of Moran's *I* is essentially the sum of the squared deviations scaled by the total weight of the matrix.

Explanation

Because GMI belongs to inferential spatial statistical methods, results are always interpreted within the context of null hypothesis. Null hypothesis for GMI is informing that attribute that is analyzed is appearing as randomly spatially distributed process in the area.

Patterns which can occur for a certain set of features are having following possible outcomes: clustered, dispersed or random phenomenon.

If the p-value has a statistically significant figure, null hypothesis can be rejected. Values of p-value can help us interpret processes:

Interpretation of the p-value

1. *If the p-value is not statistically significant*

Null hypothesis cannot be rejected, most probably spatial distribution of that certain feature being analyzed is the result of random spatial process. Spatial pattern, representing spatial randomness is just one possible version of complete spatial randomness (CSR)

2. *If the p-value is not statistically significant and Z-score is positive*

The null hypothesis can be rejected, because the spatial distribution of high and/or low values is indicating underlying spatial clustering process.

3. *If the p-value is statistically significant and Z-score is negative*

Rejection of null hypothesis is expected. In this case spatial distribution of high and low values produces a dispersed spatial pattern.

Interpretation of the z-score

Statistically significant POSITIVE z-score:

Similar values cluster spatially - high values are found closer together, and low values are found closer together, than we would expect from an underlying random spatial process.

Statistically significant NEGATIVE z-score:

Similar values are spatially dispersed - high values are found far away from other high values, and low values are found far away from other low values, and this dispersion is more pronounced than we would expect from an underlying random spatial process.

2.2.4. Local method of spatial autocorrelation

If there is assumption that spatial autocorrelation is not consistent throughout the region (*spatial homogeneity*), but varies on the basis of the location of the certain feature, there is a need for applying different and modified set of methods. Very often level of spatial autocorrelation is high in certain subregions and low in other subregions of the area being analyzed. One of the possible outcomes can be that in one certain subregion there is a positive and in another there is a negative autocorrelation. This phenomenon is called *spatial heterogeneity*.

To be capable to measure spatial heterogeneity of spatial autocorrelation, specific set of inferential statistical methods have to be used. Measures which are modified to observe spatial autocorrelation on local scale are Local Moran's Index, Local Geary's *C* and Getis-Ord *Gi**. They are based on their doublets which are measuring global magnitude of spatial autocorrelation (Lee & Wong, 2001).

Focus of further explanation will be on method used by ArcGIS and GeoDa, which is Local Moran's Index. For gaining accurate results certain conceptualization needs to be applied before running local spatial autocorrelation. Prior to creation of local spatial autocorrelation tests the creation of weight matrix has to be performed.

2.2.4.1. Weight Matrix

Spatial statistics is combining set of distinct spatial aspects (area, distance, length, proximity, connections, etc.) and spatial relationships of certain phenomena being analyzed. Spatial relationship between phenomena is defined by certain set of values represented through spatial weights matrix (ESRI, 2013e). Weight matrix is used to create neighborhood structure for certain data set and to demonstrate extent of similarity between locations and values, which is going to be further developed through concepts of spatial autocorrelation (GeoDa, 2014). Beginning of usage of spatial weight matrix starts with works of Moran (1950) and Geary (1954) and their binary weights matrix concepts (Cliff & Ord, 2009).

Spatial weight matrix file is quantifying spatial relationships between features in dataset, which is based on the conceptualization of the relationship among features. Spatial weight matrix is represented by binary weight matrix or variable weight matrix. There are many different concepts of conceptualization among features of certain dataset. Binary matrix conceptualization is represented by methods of K-nearest neighbors, fixed distance, space-time window and contiguity of spatial relationships. Binary weight matrix values are either 1 or 0. Variable weight matrix is represented through method of inverse distance methods or inverse time spatial method. Variable weights values are occurring between 0 and 1, conceptualizing spatial autocorrelation in a way that near neighbors are having larger value for the weight than neighbors that are more distant.

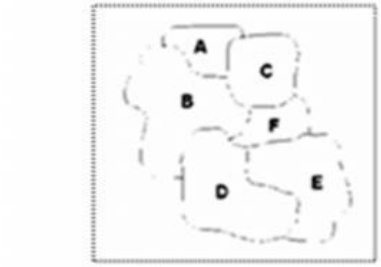
Lattice data, in this case should conceptualize weight matrix on contiguity basis, which is representing binary matrix. Conceptualization of weight matrix is a table set with one row and one column for every feature in the set, row-standardized matrix. Row-standardized matrix is one in which values of itch of its rows sum to one. This is conceptualized that every neighbor weight is divided by the sum of all neighbor weights of certain feature being analyzed.
The location at the center of its neighbors is not included in the definition of its neighbors and because of that is set to a zero (GeoDa, 2014). It is a location from which neighbors are conceptualized. Weight is cell value for any row/column combination, which is explaining quantitatively spatial relationship between row and column values. Neighbors of certain features are defined binary (values 0 and 1), where features which have value 1 are representing neighbors and features which have value 0 are representing non-neighbors and location itself. Option of defining higher order of contiguity exist by defining neighbors of neighbors. Higher order of contiguity also includes option to exclude or include lower orders in calculation.

According to, ESRI and GeoDa glossary terms definitions, conceptualization of weight matrix in the case of lattice data with shared border is contiguity concept. Contiguity concept within itself includes two different principles of defining neighbors. First one which is using plain north-south east-west concept is rook principle. Rook concept, on the basis of contiguity is defining neighbors as entities with which same border is shared. In the case of lattice data, which is used in this study neighbors are represented by cells North-South and West-East from the basic cell (4 immediate ones following the principle). Other concept which is used is a Queen contiguity concept. This concept is, besides incorporating rook concept within itself

also using vertices. Vertices are nodal points which are defining boundary corners of a certain polygon in this case cells because of the nature of lattice data. Principle of locating neighbors is including beside North-South and West-East also Northwest, Northeast, Southwest and Southeast. For example, in this case neighbors some certain cell A would be all cells which are sharing boundary with cell A in any direction (Higazi et al., 2013). Analyzed cells are doubled by this method and they represent all possible neighbors with whom border is shared (Figure 7).

Generally, contiguity weight matrix is producing value "0" or "1" as follows (Higazi et al., 2013):

$$W = \begin{cases} \mathbf{1} & \boldsymbol{i}\ \mathbf{neighbor}\ \boldsymbol{j} \\ \mathbf{0} & \mathbf{othervise} \end{cases}$$



	A	B	C	D	E	F
A		1	1	0	0	0
B	1		1	1	0	1
C	1	1		0	0	1
D	0	1	0		1	1
E	0	0	0	1		1
F	0	1	1	1	1	

Figure 7. Conceptualization of the weight matrix in the case of shown neighboring units (Haining, 2003).

Written as weight matrix file it would be (Higazi et al., 2013):

$$W = \begin{bmatrix} w_{11} & w_{12} \dots & w_{16} \\ w_{21} & w_{22} \dots & w_{26} \\ & \dots\dots & \\ & \dots\dots & \\ & \dots & \\ w_{61} & w_{62} \dots & w_{66} \end{bmatrix} = \begin{bmatrix} 0\,1\,1\,0\,0\,0 \\ 1\,0\,1\,1\,0\,1 \\ 1\,1\,0\,0\,0\,1 \\ 0\,1\,0\,0\,1\,1 \\ 0\,0\,0\,1\,0\,1 \\ 0\,1\,1\,1\,1\,0 \end{bmatrix}$$

Normalization of the weight matrix is then conducted in the following manner (Higazi et al., 2013):

$$W_{ij} = \frac{w_{ij}}{\Sigma_i w_{ij}} \qquad 0 \;\; w_{ij} \leq 1$$

Tests for spatial autocorrelation for a single variable in cross-sectional data set are based on the quantitative capacity of an indicator, which combines the value observed at each location with the average value at neighboring locations (so called spatial lags) (Higazi et al., 2013).

They are representing a measurement of similarity among certain features with a specific focus on the values of association (covariance, correlation or difference) and their association in space (contiguity). Spatial autocorrelation is considered to be significant when the spatial autocorrelation statistic takes on an extreme value, compared to what would be expected under the null hypothesis of no spatial autocorrelation (Anselin, 1992).

2.2.4.2. Local Moran's Index of Spatial Autocorrelation (LISA)

Different than GBI which is measuring spatial autocorrelation of the area being analyzed on global scale, Anselin Local Moran's Index (LMI) is identifying clustering on local scale, with more precise calculating capabilities and outcome results. If there is a certain set of features (input feature class) and analysis field (Input Field) LMI is identifying spatial clusters with high (High-High) (HH) or low value (Low-Low) (LL) and at the same time it is identifying spatial outliers (High-Low) (HL) and (Low-High) (LH).

To be possible to produce this LMI is calculating local Moran's Index value, z-score, p-value and a code representing one of the four code types (HH, LL, HL, and LH). (ESRI, 2013f).

Mathematics

Anselin Local Moran's Index is defined as (Anselin, 1995)

$$\boldsymbol{I_i = \frac{x_i - \overline{X}}{S_i^2} \sum_{j=1, i \neq i}^{n} w_{i,j} \left(x_i - \overline{X}\right)}$$

(1)

Where unknowns are:

$\boldsymbol{x_i}$ – attribute for feature $\boldsymbol{i}$

$\overline{\boldsymbol{X}}$ – mean of corresponding attribute

$\boldsymbol{w_{i,j}}$ – spatial weight between feature $\boldsymbol{i}$ and $\boldsymbol{j}$

$$\boldsymbol{S_i^2 = \frac{\sum_{j=1, i \neq i}^{n} (x_i - \overline{X})^2}{n-1} - \overline{X}^2}$$

(2)

$\boldsymbol{n}$ - defining total number of features

The $\boldsymbol{zI_i}$ – score for the statistics is calculated:

$$\boldsymbol{zI_i = \frac{I_i - E[I_i]}{\sqrt{V - [I_i]}}}$$

(3)

Where:

$$\boldsymbol{E[I_i] = -\frac{\sum_{j=1, i \neq i}^{n} w_{i,j}}{n-1}}$$

(4)

$$\boldsymbol{V[I_i] = E[I_i{}^2] - E[I_i]^2}$$

(5)

Additional mathematics for LMI is:

$$\boldsymbol{E[I^2] = A - B}$$

(6)

$$\boldsymbol{A = \frac{(n - b2_i) \sum_{j=1, i \neq i}^{n} w^2{}_{i,j}}{n-1}}$$

(7)

$$\boldsymbol{B = \frac{(2b2_i - n) \sum_{k=1, k \neq i}^{n} \sum_{h=1, h \neq i}^{n} w_{i,k} w_{i,h}}{(n-1)(n-2)}}$$

(8)

$$\boldsymbol{b2_i = \frac{\sum_{i=1, i \neq j}^{n} (x_i - \overline{X})^4}{\left(\sum_{i=1, i \neq j}^{n} (x_i - \overline{X})^2\right)^2}}$$

(9)

Explanation

Occurrence of the positive value for LMI is pointing that a feature has a neighboring features with a similarly high or low values of the certain phenomena being analyzed, and that at the same time it indicates that these features are belonging to a certain type of cluster, meaning they are clustering. If the values of LMI are negative it indicates that a certain feature has neighboring features with a dissimilar values, showing that the feature is and outlier, infect that it does not correlate with other neighboring features, meaning they are not clustering.

In both of the cases p-value has to be small enough (< 0.05) so the cluster or outlier has to be considered as statistically significant. Features that have non-significant statistical value are marked as Not Significant. They occur if the p-value which represents probability has values greater than (> 0.05).

LMI is a relative measure and to interpret it z-score and p-value have to be computed. With accurate assessment of all three values LMI can be interpreted within Null hypothesis, as one of the inferential statistics methods.

The outcome fields, branded with different cluster or outlier types (CO type) are four possible solutions:

1. *High-High (HH)*

Statistically significant, p-values are lower than 0.05 representing cluster of high values

2. *Low-Low(LL)*

Statistically significant, p-values are lower than 0.05 representing cluster of low values

3. *High-Low(HL)*

Statistically significant, p-values are higher than 0.05 representing outlier in which high value is surrounded by low values

4. *Low-High(LH)*

Statistically significant, p-values are higher than 0.05 representing outlier in which low value is surrounded by high values.

3. STUDY AREA

3.1. Helsinki Metropolitan Area (Pääkaupunkiseutu)

Finland is situated in northern Europe and borders to Sweden and Norway in the west and north-west, and to Russia in the east. The Gulf of Bothnia, the Baltic Sea and the Gulf of Finland lies in the west and the south.

Country is divided into state provinces (*alue* or *läänit* in Finnish), regions (*maakunta* in Finnish), sub-regions (*seutukunta* in Finnish) and municipalities (*kunta* in Finnish). Municipalities are divided into smaller organizing units like large area (*suuripiiri),* basic area *(peruspiiri),* section *(osa-alue),* small area *(pienalue),*block *(kortteli)* and property *(kiinteistö)* (Vilkama, 2011).

Study area belongs to Etelä-Suomi province, Uusimaa region, subregion of Greater Helsinki (*Helsingin seutu* in Finnish), and territory of Helsinki Metropolitan Area *(Pääkaupunkiseutu)*. Helsinki Metropolitan Area territory is formed by municipalities of Helsinki, Vantaa, Espoo and Kauniainen (Inkinen & Vaattovaara, 2007). Kauniainen is the smallest municipality with total population of 9,039 inhabitants (Statistics Finland, (2014) Center of Finland, 2014) and it is the only municipality in the whole Finland which borders are surrounded by only one municipality, Espoo. It is geographically part of the study area, but because data is not provided (by HSY) and knowing that most of the inhabitants are Finnish-Swedish speaking minority and that the number of immigrants living there is small (4%) it is going to be excluded from further analysis. Helsinki Metropolitan Area (Pääkaupunkiseutu) is surrounded with municipalities of: Kirkkonummi and Vihti in West, Nurmijärvi, Hyvinkää, Tuusula, Järvenpää and Kerava in North, Sipoo in East and Baltic Sea in South.

HMA as an integral core part of Greater Helsinki (Metropolitan Area) represents the only metropolis area in Finland. By the latest data, provided by Statistics Finland at the end of 2013 population of Helsinki Metropolitan Area was consisted of 1081515 inhabitants (Population Register Center of Finland, 2014). Around 19 % of the country's population lives in just 0.2 % of Finland's surface area. However, housing density of the HMA is high by Finnish standards: 34.2 m^2/person in comparison with, 39.4 m^2/person in Finland (Urban Facts, 2013) Helsinki Metropolitan Area has a high concentration of employment: approximately 580 000 jobs.

3.2. Location, area and demographics

Helsinki Metropolitan Area is located on the shore of Gulf of Finland at the Baltic See some 80 km north of Tallinn, Estonia, 400 km east of Stockholm, Sweden, and 300 km west of Saint Petersburg, Russia.

Metropolitan area occupies the space of 770.26 km². Eurostat is trying to standardize the concept of metropolitan area. Defined by Eurostat, European Union project for standardization of metropolitan area of Helsinki is made of kernel consisted of: Helsinki, Espoo, Vantaa and Kauniainen (Urban Audit, 2006) see also Figure 8.

Given the total population of Helsinki Metropolitan Area by census data from the end of 2013 which is 1081515 inhabitants and taking into account total area which is 770.26 km², density of population is 1404.09 inhabitants per km²

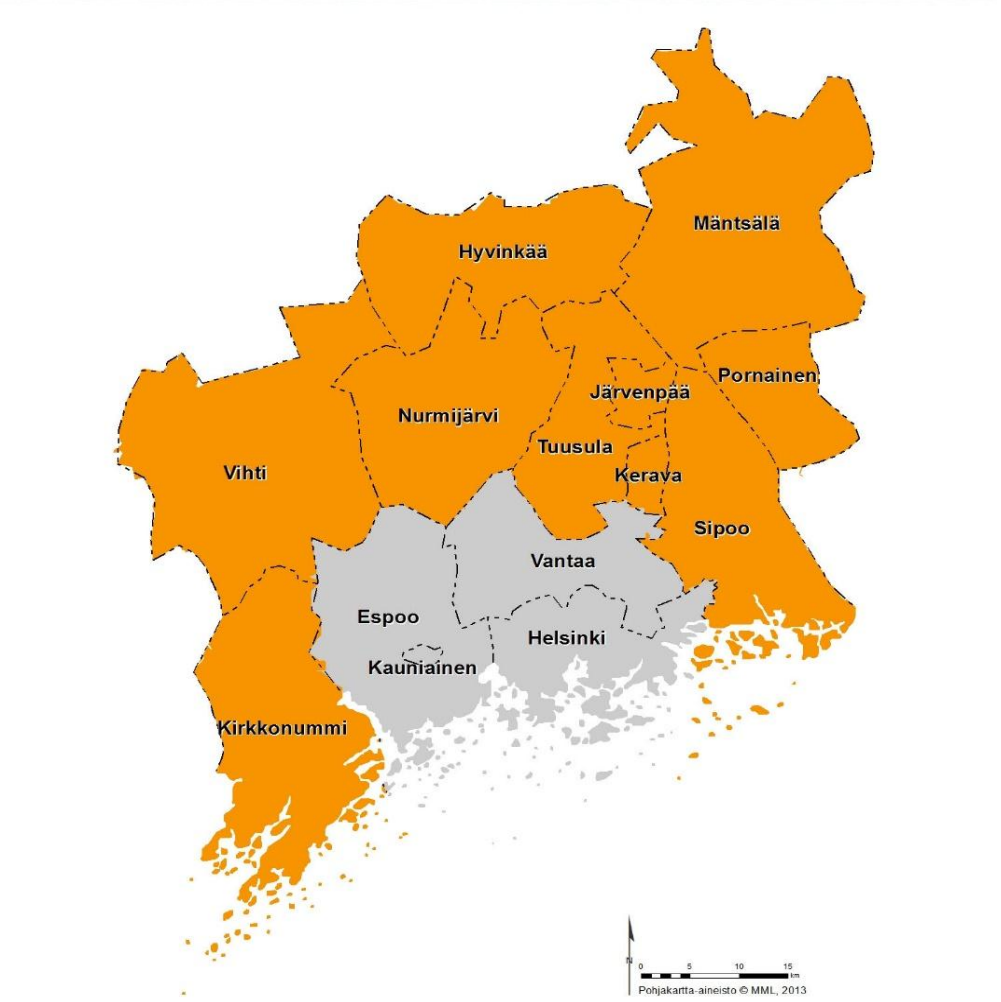


Figure 8. Helsinki Metropolitan Area and surrounding municipalities (Kuuma, 2013).

4. DATA

4.1. About HSY Data

Geographical Information, maps and SeutuCD

Data is provided by HSY (Helsingin Seudun Ympäristöpalvelut-kuntayhtymä). It represents basic geographical information data. HSY is producing comprehensive registration and map data which is supporting planning, research and policy making conducted by municipalities or researchers interested in Helsinki Metropolitan Area. Map materials data is gained from several different producers, mainly to be used by regional public authorities. Regional map materials, used in this work are produced are produced in the regional SeutuCD compact disk (Register data: Source SeutuCD'08). SeutuCD represents a data package integrated as a cross-section of the SePe (seudullisen perusrekisterin) Helsinki Metropolitan Area register data covering buildings, real estate, zoning plans and planning units. It includes several maps of Helsinki Metropolitan Area population at a building level and a geographic dataset of establishments produced by Statistics Finland (HSY, 2014).

4.2. Basic SeutuCD Data

Basic data is provided on the compact CD for the purpose of research conducted by individual researcher. For gaining access to certain data researcher has to apply personally and provide information on explicit type of research that is going to be conducted so he would be provided with specific data. Metadata is divided into two folders one consisting of digital maps and another one consisting of data for research. Researcher is provided with detailed information and structure about the nature of data in Finnish and how data is organized. Each data type has detailed explanations about production, format, form, spatial type, scale, coordinate system, regional coverage and other specific information on the nature and production of data. There is a dictionary of variables explaining meaning of certain terms in legend.

5. METHODS

First step in the spatial pattern analysis of immigrant population in the study area is the study of possible spatial autocorrelation, based on the two features: location and values of immigrant population simultaneously. It is done by the use of several inferential spatial statistics methods. ArcGIS and GeoDa are using same global (Global Moran's Index) and local (Local Moran's Index) methods. In this thesis these two methods are going to be presented.

5.1. Preprocessing data

Because the major purpose of SeutuCD data is a creation of maps, data is mostly use by planning offices of major public authorities (municipalities) and for their convenience it is produced in MapInfo format. All explanations and notifications done about data are produced in Finnish, so understanding of the nature of data can be challenging and difficult for non-native speakers.

Data is produced for use in MapInfo software. Specific MapInfo file formats cannot be used in ArcGIS and GeoDa. Data needs to be transformed to readable ArcGIS and GeoDa format. Transformation was made with Quantum GIS software. During the transformation of the files, specific coordinate system KKJ (Kartastokoordinaattijärjestelmä) which is particularly used in Finland has to be saved in the same manner how it was created for MapInfo. This has to be done due to avoiding distortion of the coordinate system and the whole data in general.

Population data is aggregated on the level of buildings and delivered as point pattern PKS_VAKI file consisting information about different type of population (coordinates of the points, men, women, different age groups, different language groups, foreigners, etc.). Information about immigrant population is based on personal statement of the native language, which is collected from information provided by individuals. Abbreviation used in the table of context dealing with population is SUM_ULKOKANS. Small percentage of this population are the Finnish citizens which don't have immigrant status in a legal sense but are due to inconsistency in grouping of data, marked as foreigners.

Due to the nature of data (point pattern data) and its spatial capability of processing preprocessed grids (RUUDUT) are delivered with grid cell size of different levels: 250m, 500m, 1km and 2km. Lattice is allowing us greater spatial analytical capabilities with specific locational information of the certain phenomenon being analyzed. Nevertheless, due to the

specific interest in local spatial autocorrelation creation of new grid level size of 50 m was executed. Almost all previous studies were done by lattice cell level size 250 m which represents correct size in certain areas of built environment which are located in suburban areas (Puotinharju, Hertoniemi, Vuosari and most of the parts of Vantaa and Espoo municipalities). Central parts of Helsinki municipality (Töölö, Kallio, Sörnainen, Punavuori) are compactly packed built environment areas where 250 m cells represent block level size area. So due to the precise and more accurate measurements of local spatial autocorrelation new grid level size of 50 m is imposed for analysis.

Using the spatial joining option in ArcGIS, point pattern data resembling population of single living unit (apartment building, house, attached house, etc.) in HMA area was aggregated and connected with different lattice level sizes i.e. 1000m, 500m, 250m and 50m, respectively. When produced on lattice data are easier to be comprehended following that the size of each unit is the same and concept of contiguity can be adopted in terms of comparison of the units of the same size. Preprocessing phases can be seen from Figure 9. Final data sets where used for assessing phenomenon of spatial autocorrelation throughout the HMA area with a use of ESDA methods.

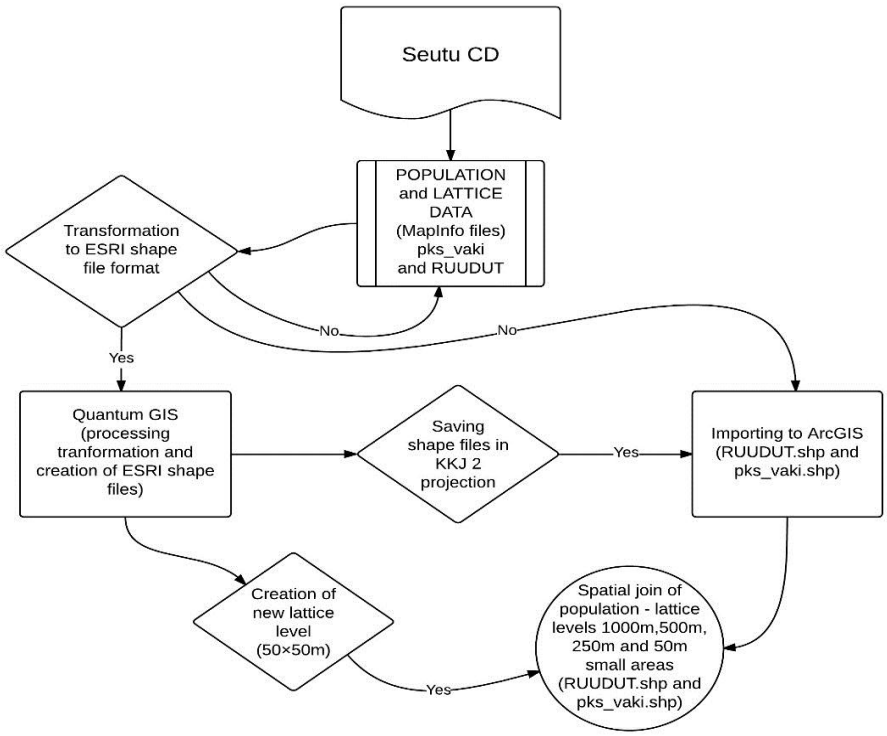


Figure 9. Visual presentation of reading and preprocessing data (Kekez, 2014).

5.2. Global method of spatial autocorrelation

Method implemented in this thesis is Global Moran's Index (GMI) which is used by both software, but produced results are resembled in a different manner.

5.2.1. Global Moran's Index

Output in ArcGIS

In ArcGIS, Global Moran's Index (GMI) produces five different values: Global Moran's Index, expected Index, variance, z-score and p-value.

Values of GMI and Expected Index, being produced at the same time are compared. Z-score and p-value are being produced on the basis of number of features in dataset and variance for data set overall. Variance value is representing how far the values are lying from the mean (expected value), or how far set of values is spread out. After getting p-value and z-score determination of statistical significance is assembled and further on interpreted within the context of null hypothesis (Figure 10).

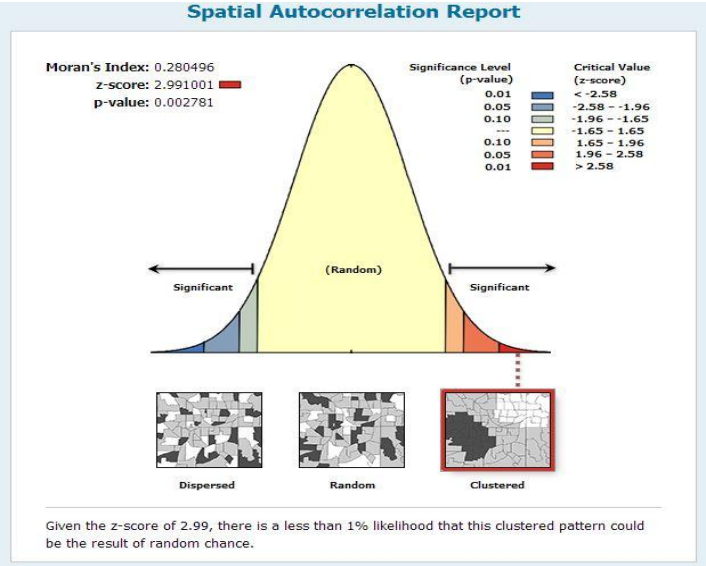


Figure 10. Report of the GMI in ArcGIS (Kekez, 2014).

Positive Global Moran's Index values are indicating occurrence of clustering in specific area and negative results are informing us about the dispersion. The values can be positive or negative for Moran's Index and depended on its value we can talk about, positive or negative spatial autocorrelation. Values are ranging from -1 which indicates perfect dispersion of data in the area, till +1 which indicates perfect correlation. Negative (positive) values indicate negative (positive) spatial autocorrelation. Values range from −1 (indicating perfect

dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern and confirmation of null hypothesis, meaning that certain phenomena being analyzed is randomly distributed around the area (Lee & Wong, 2001).

Output in Geoda

In Geoda, Global Moran's Index (GMI) is producing scatterplot graph and ten different statistic values: #obs, R^2, const a, std-err a, t-stat a, p-value a, slope b, std-err b, t-stat b, p-value b.

Variable #obs is displaying the number of observations and other statistic values are representing the results of simple linear regression, which is the least squares estimator of linear regression model with a single explanatory variable which is in this case Sum_ULKOKA.

Scatterplot graph is a result of a simple linear regression, which is producing least squares regression analysis (Figure 11).

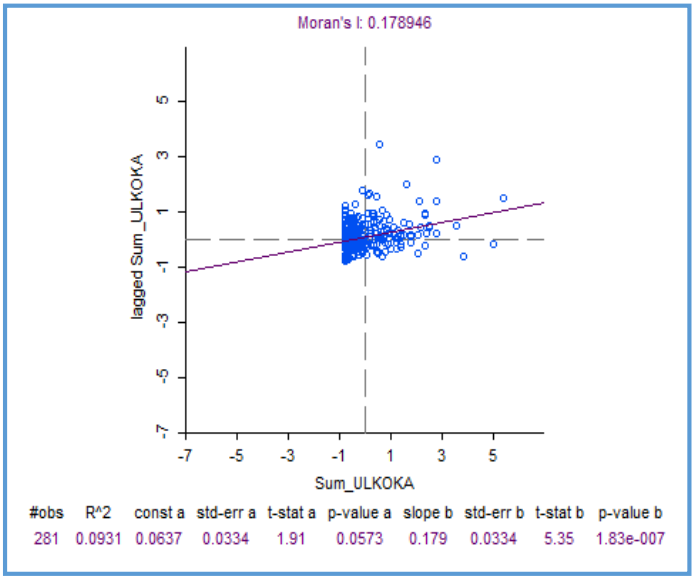


Figure 11. GMI Scatter Plot Graph in Geoda with additional statistic (Kekez, 2014).

Graph produces best fitted line through the set of *n* points making the sum of square residuals of the model (vertical distances between points of the data set and fitted line) as small as possible. The slope of the fitted line is equal to correlation between Sum_ ULKOKA and lagged Sum_ULKOKA corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass (*Sum_ULKOKA, lagged Sum_ULKOKA*) of the data points.

Interpretation of gained statistics is:

R^2 - Represents R square measure of the regression lagged Sum_ULKOKA = a + b*Sum_ULKOKA. It provides information to what extent Sum_ULKOKA explains lagged Sum_ULKOKA. Value range is from 0 to 1.

const a - The parameter **a** is the representation of regression lagged analysis, in form of equation: Sum_ULKOKA = a + b*Sum_ULKOKA.

std-err a – is representing positive or negative error value which is estimated to be in the range (a - std-err up to the a + std-err)

t-stat a - is the value of t-statistic which is the product of Std-err a. It is the ratio of the coefficient to its standard error. In this case it would be:

t-stat a = (slope b) / (std-err a) = (Moran's *I*) / (std-err a)

It explains how much different is the estimated value and the Std-err a (uncertainty) of estimation.

p-value a – is the representation of the statistical significance of a. In general a cut-off value of 0.05 is applied for determination of statistical significance. If the value is less than 0.05 it is statistically significant and if the value is bigger than 0.05 statistical significance doesn't exist.

slope b – represents parameter b of the regression lagged analysis, in form of equation: Sum_ULKOKA = a + b*Sum_ULKOKA. It is the value of Moran's *I* index.

std-err b – is representing positive or negative error value which is estimated to be in the range (b - std-err up to the b + std-err)

t-stat b - is the value of t-statistic which is the product of Std-err b. It is the ratio of the coefficient to its standard error. In this case it would be:

t-stat b = (slope b) / (std-err b) = (Moran's *I*) / (std-err b) - is explaining difference in the estimated value and the Std-err b (uncertainty) of estimation.

p-value b – is the representation of the significance of the Moran index. In general a cut-off value of 0.05 is applied for determination of statistical significance. If the value is less than 0.05 it is statistically significant and if the value is bigger than 0.05 statistical significance doesn't exist.

Additional information about test is omitted by performing permutation test for the GMI. Permutation test represents a numerical approach for testing the significance of statistic performed by GMI. In this case it is used for improvement of the result of the approximate normal test and gaining information about sampling distribution under spatial randomness. In each replication the observed values of variable are randomly assigned to the regions. In this way random map patterns of the spatial distribution of a variable are obtained. For each random pattern, the Moran coefficient is computed. The observed value of Moran's *I* is compared to simulated sample distribution. The observed Moran's *I* value has a low probability to stem from a spatial random distribution of the variable, if it is found in the tails of the sample distribution. In particular the null hypothesis of spatial randomness has to be rejected, if the pseudo p-value of Moran's *I* is lower than the significance level set by the user. Minimum number of permutations is 9 and maximum is 99999 which can be manually set.

Number of permutations will provide different p-value dependent on the number set up. Different number of permutation is going to provide different p-value. With an increasing number of replications the approximation of the generated sample distribution sample is improving. If the spatial autocorrelation is statistically significant (at the 95% margin), the p-value will always be smaller than 0.05 (i.e. it may be 0.001 or 0.000001 – the number of permutations influences the fraction). The choice of final digits of maximum 99999 for the number of permutations in GeoDa is motivated by the maximum computation capabilities and highest certainty in produced results (bigger the number of permutations, higher certainty of the results). Permutation test is performed by right-clicking inside scatter-plot and choosing from pop-up menu option Randomization test (Figure 12).

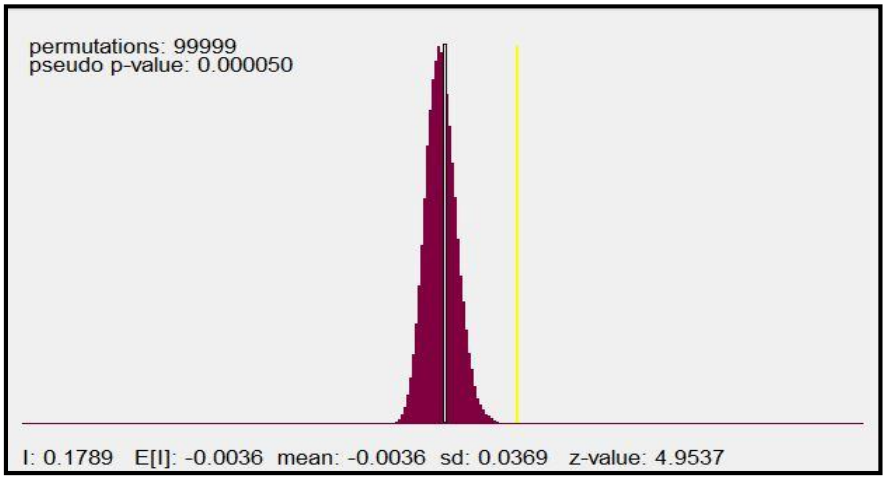


Figure 12. Permutation test of GMI scatterplot graph in Geoda (Kekez, 2014).

Graphical representation as seen in Figure 12 provides us important statistical parameters such as:

permutations – number of permutations performed in the test

pseudo p-value – pseudo significance level

Moran's I (I) – statistics of the Moran's Index

Expected Moran's I (E[I]) – expected value of Moran's *I* for the study area

mean – average of Moran's *I* for the simulated distributions

standard deviations – standard error of Moran's *I* computed from its simulated distribution (99999 in this case)

histogram – random distribution of the value of I (red piles) and the value of the actual data (yellow line)

Moran's *I* points to possible positive spatial autocorrelation. However, it can only be concluded from a significance test whether the measured spatial dependence is a characteristic feature of the variable in the population or due to sampling errors.

Moran scatterplot graphs are used to represent the results of the analysis. They are considered as crucial outcome of the exploration of spatial patterns (Anselin, 1996; Anselin & Bao, 1997). A scatterplot graph is conceptualizing visual statistic derived from the results of GMI. The slope of regression line is indicating level of global association. Statistic is reassembled into four different types of association (Leitner & Brecht, 2007). Lower left and upper right quadrant are indicating positive spatial autocorrelation. Lower one is indicating the presence of similar low values and upper one presence of high values in neighboring locations. Other two quadrants (upper left and lower right) are indicating negative spatial autocorrelation, indicating presence of dissimilar values in neighbor locations (Frank, 2003).

5.3. Local method of spatial autocorrelation

Conceptualization of weight matrix in ArcGIS and GeoDa

Creation of weight matrix in ArcGIS and GeoDa is conducted on the similar principles. Results which are gained are comparable which makes final results of spatial autocorrelation tests relevant.

"Most relevant contemporary software provides for the storage of spatial weights once computed, although establishment of a standard format would be of great help, especially if its use became prevalent, as this would permit comparison without the risk that observed differences were due to the weights being handled differently" (Bivand, 2009).

Overcoming this difficulties and making comparable weight matrix is done by the use of PySAL toolbox produced by GeoDa team. PySAL (Python Spatial Analysis Library) is an independent toolbox (developed specifically for ArcGIS) for the creation and conceptualization of spatial weight matrix on the same principles for ArcGIS and GeoDa. Toolbox within itself is allowing creation of same spatial weight matrix, but with a choice of different file extensions; .swm for ArcGIS and .gal and .gwt for GeoDa (Anselin & Rey, 2010). Unfortunately PySAL toolbox is not operational at the moment, but in the next edition of ArcGIS it is expected to be fully operational. Format used for weight matrix in GeoDa comes from the Geographical Algorithms Library (GAL, university of Newcastle, during 1980's) (Anselin, Syabri & Kho, 2006). GAL format sets up neighbor set membership, no matter if the conceptualization of the matrix is done with a principal of contiguity or some other criteria (Bivand, 2009).

Validation of the conceptualization of weight matrix and p-values in ArcGIS and GeoDa

Due to the differences in computation of the gained results specific attention has to be put on the analyses of the produced results in ArcGIS and GeoDa.*"GeoDa and ArcGIS use permutations to generate the reference distribution. This produces different results not in the values of the Ii, but in their statistical significance"* (Monasterio, 2006).

Different results of statistical significance can perform completely different visual presentations of clustering values in the maps produced by local methods. That's why it is out of high importance to gain closest probability results which can be later compared.

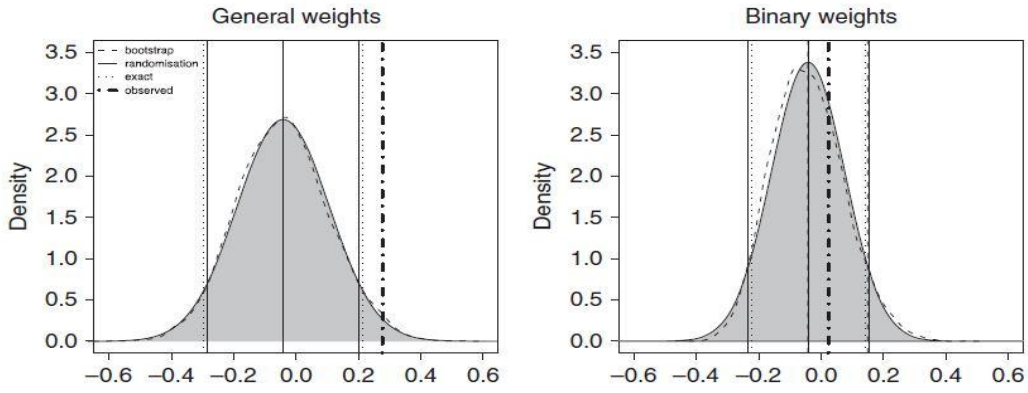


Figure 13. Permutation bootstrap displays for variable five, unstandardized general and binary weights; vertical lines show values of the observed statistic, its expectation, and $\alpha = 0.05$ two-sided ($\alpha = 0.1$ one-sided) ''confidence interval'' lines (Bivand, 2009).

Bivand (2009) is giving superb visual interpretation of the distribution of "confidence intervals" under the different levels of permutation values applied in local statistical methods in both software (Figure 13). Results gained by visual representation can help us understand principles behind statistical methods, more similar or different once the same weigh matrix is applied. Permutation levels can be manipulated in GeoDa giving providing specific results.

5.3.1. Local Moran's Index

Output in ArcGIS

Beside local Moran's *I* index, z-score and p-value, LMI is creating a new output feature class for each feature in the input table feature class. New feature is called COtype and it can be (HH, LL, HL, and LH). In the case if the feature is statistically non-significant it is marked as Not Significant and it is not labeled anyhow. COtype and Not Significant features are added as a new output feature class to the table of contents (TOC) together with local Moran's *I* index, z-score and p-value features for the feature class from input TOC.

Output GeoDa

Output in Geoda is allowing us to choose what kind of the results we would want to produce and save concerning LMI. As how it was the case with GMI in GeoDa we can produce Moran Scatter Plot, but also Cluster Map and Significance Map. The results produced by the Moran Scatter Plot are identical like in GMI which is previously explained. Results of Cluster Map and Significance Map have to be saved and added as a new variable to already existing table of values if they are going to be used afterwards. The outcome is the same, but saved in a table of context. Local Moran's *I* index will be saved automatically as LISA_I, CO type as LISA_CL and p-value as LISA_P.

Complete processes of calculating Global Moran Index and Local Moran Index results produced by both software, as well as employment of all the steps of the processes is explained and presented in Figure 14.

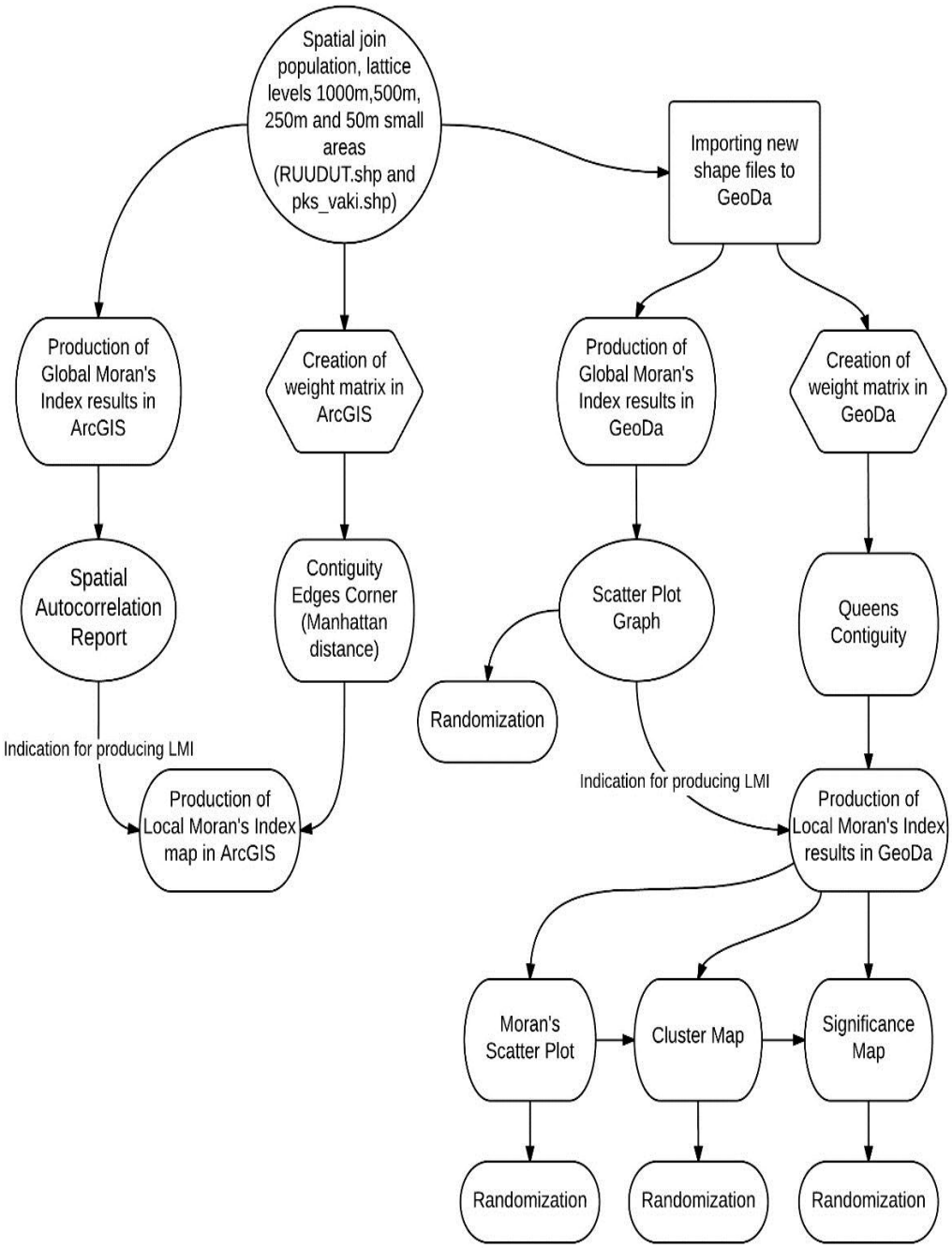


Figure 14. Creation of GMI and LMI spatial statistical results in ArcGIS and GeoDa (Kekez, 2014)

6. RESULTS

Based on the methods and data gained formulation of the results is conceptualized through creation of different maps and statistical results, gained by the employment of different spatial statistical methods performed in ArcGIS and GeoDa. Small areas (pienalue) are mostly used as the spatial level size in addressing the problem of spatial spread of immigration population. ESDA methods performed in ArcGIS and Geoda are going to be employed in creation of the maps representing clusters and outliers of immigration population in HMA.

For more structuralized analysis of data (buildings containing population information), for the purpose of conceptualization of spatial correlation, data is aggregated to lattice. Lattice cells have the size of 1000×1000m, 500×500m, 250×250m which are provided by SeutuCD. Rethinking about conceptualization of the data provided by SeutuCD, which is on the smallest level (250×250m), conclusion was made that most of the buildings do not have that size and that size corresponds to certain extent to a size of small block area. Population point pattern data (PKS_VAKI) is conceptualized like a point data, representing living units. New lattice grid size of 50×50m is made, being more realistic representation of living units in space. Imposing new lattice cell level size, with value of 50×50m, which is more realistically correlating with *in situ* situation analytical capabilities of local ESDA methods (LISA) should improve performance and compute completely new results. Results are going to be presented for five differently conceptualized areas, where basic units are: small areas *(pienalue),* 1000×1000m, 500×500m, 250×250m and 50×50m lattice grid cell size.

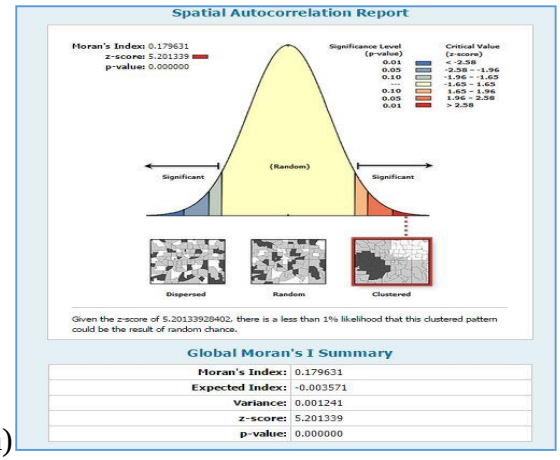
6.1. Global method of spatial autocorrelation

Results resembling GMI performed in ArcGIS and GeoDa are going to be interpreted within null hypothesis, indicating presence of clustering as a global phenomenon presented in the area of HMA. Statistical results represented are exhibiting the same values. Visual results are represented in different manner. Computed statistics is much more detailed in GeoDa and it allows manipulations (randomization levels) and further testing (possibility of manipulating with different p-significance levels), while output in ArcGIS has static form.

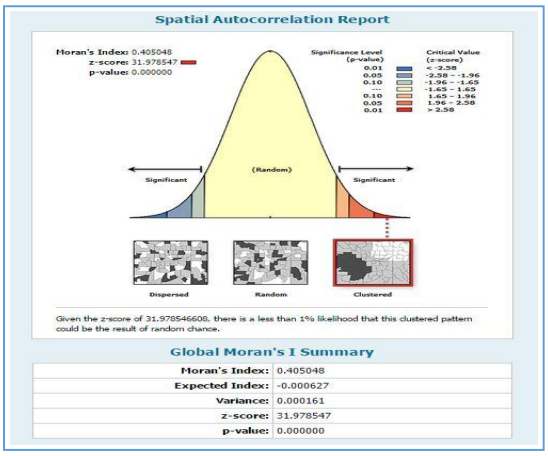
6.1.1. Global Moran's Index results in ArcGIS

All the reports of GBI are indicating presence of high level of clustering in the area (small areas, 1000×1000m, 500×500m, 250×250m and 50×50m) which is visually noticeable from Spatial Autocorrelation reports presented in Figure 15.

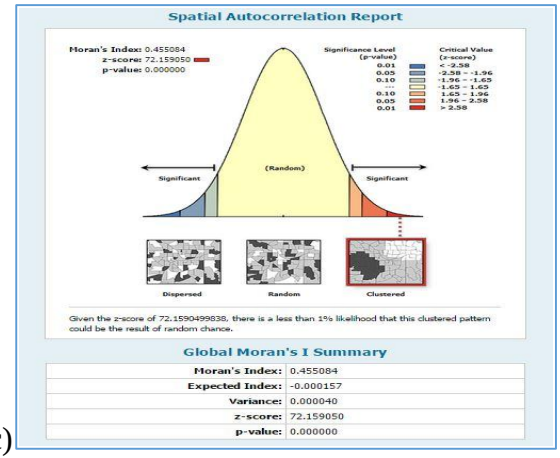
a)



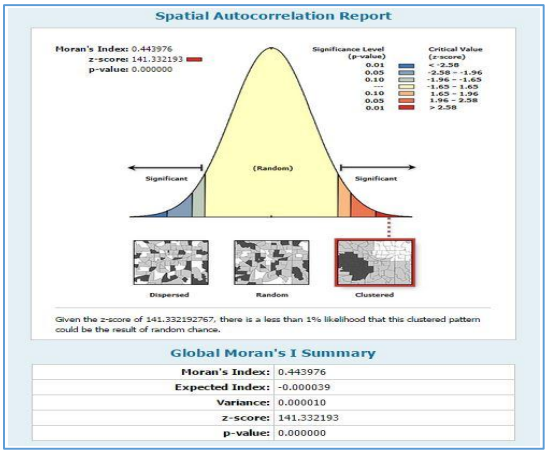
b)



c)



d)



e)

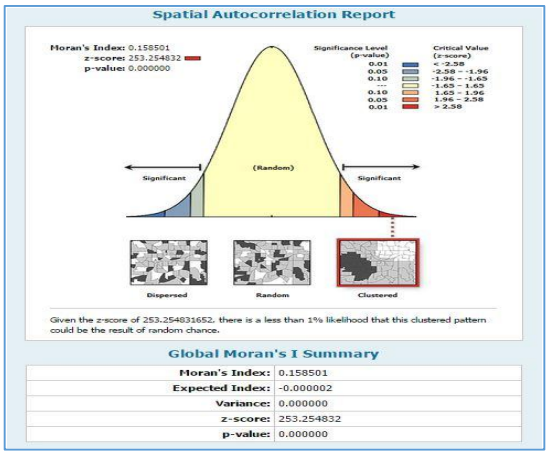


Figure 15. Spatial autocorrelation reports produced by ArcGIS with additional GMI statistics for: a) small areas (pienalue), b) 1000×1000m, c) 500×500m, d) 250×250m and e) 50×50m lattice level size (Kekez, 2014)

Statistical results presented in Table 2 are indicating strong spatial autocorrelation for immigration population in the HMA area. Moran's Index values are indicating positive spatial auto correlation and their distribution indicates best fit for the lattice level of 250×250m.

Expected Index value results are interpreted together with Moran's Index values indicating presence of clustering. Variance for the data values overall is produced indicating distance from the mean (expected value), confirming the best fit of values in lattice level of 250×250m. Significance of the results (p-values) and critical values (z-score) are extremely high, indicating presence of clustering in all levels (small areas, 1000×1000m, 500×500m, 250×250m and 50×50m) of analyzed HMA area. All the analyzed areas are represented with p-values (> 0.1) and z-scores (> 2.58). Z-scores are exhibiting exponential growth of the values, from small areas towards 50×50m lattice level size units. Best fit of the results is represented in 50×50m lattice level size indicating that results represented in that lattice level are most accurate and statistically significant. Concept of CSR can be reject all together with Null hypothesis. Confirmation of clustering processes going on in the HMA area can be accepted, but locations of specific spatial clusters and their formation needs to be investigated by employment of LMI method.

Table 2. Global Moran's Index statistics in ArcGIS (Kekez, 2014).

	Small areas	1000×1000m	500×500m	250×250m	50×50m
Moran's Index	0.179631	0.405048	0.455084	0.443976	0.158501
Expected Index	-0.003571	-0.000627	-0.000157	-0.000039	-0.000002
Variance	0.001241	0.000161	0.000040	0.000010	0
z-score	5.201339	31.978547	72.159050	141.332193	253.254832
p-values	0	0	0	0	0

6.1.2. Global Moran's Index results in GeoDa

Results of GMI performed in GeoDa are also indicating presence of clustering for immigration population for all the studied scales i.e. in small areas, 1000×1000m, 500×500m, 250×250m and 50×50m, respectively. This can be also confirmed from visual representations of Global Moran's Index scatterplots presented in Figure 12. Spatial spread of points (resembling observed values) are indicating strong spatial positive autocorrelation. Moran's Index and Expected Index are having exactly the same values like in ArcGIS indicating that conception of spatial relationship is exactly the same allowing us comparison of the results presented in Table 3.

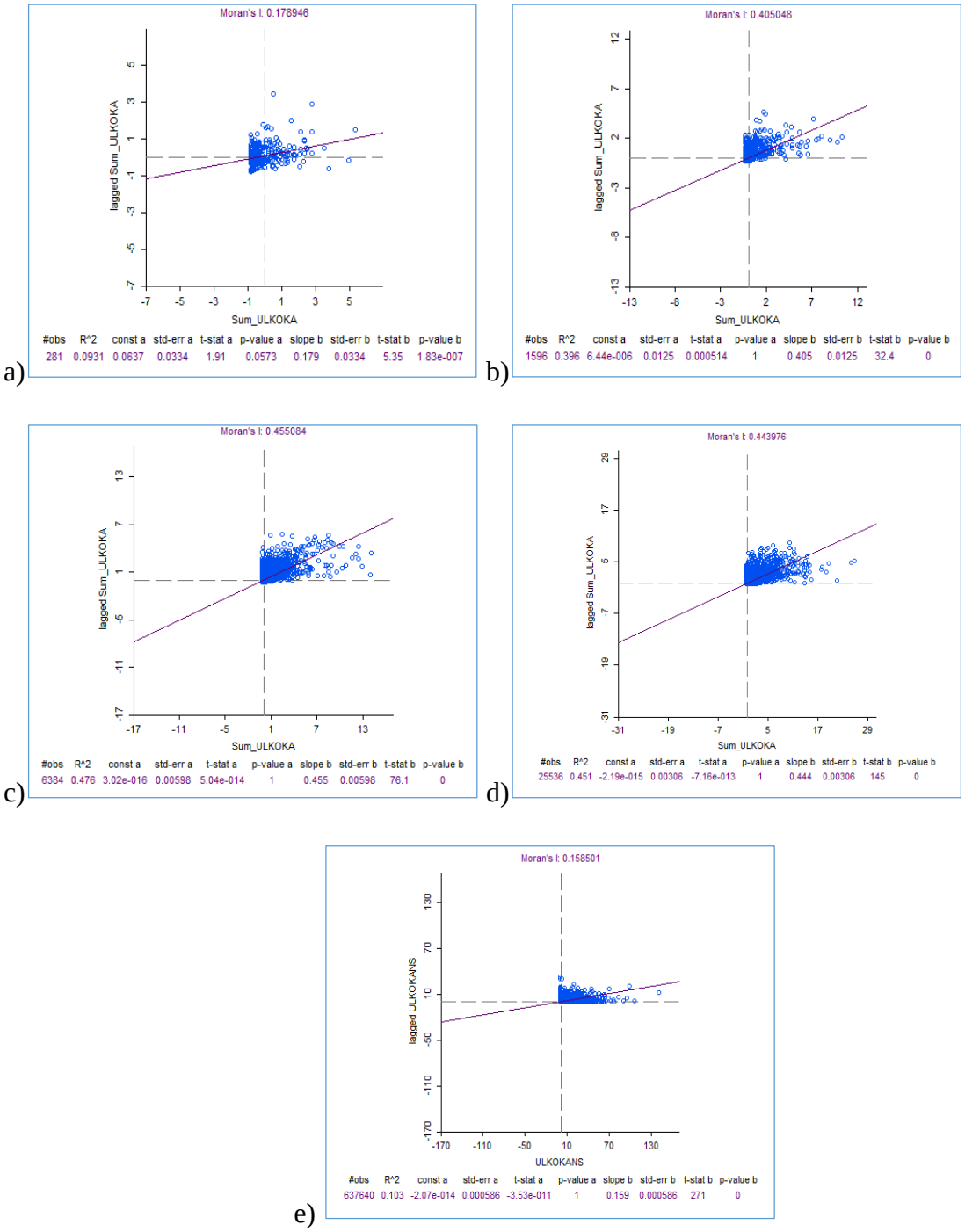


Figure 16. Global Moran's Index scatterplots for small areas: a) small areas (pienalue), b) 1000×1000m, c) 500×500m, d) 250×250m and e) 50×50m lattice level size (Kekez, 2014).

Table 3. Global Moran's Index statistics in GeoDa* (*values of Global Moran's Index, other statistical values of Moran's scatter plot and additional statistics gained from results of randomization levels)(Kekez, 2014).

Global Moran's Index statistics in GeoDa

	Small areas	1000×1000m	500×500m	250×250m	50×50m
Moran's Index	0.178946	0.405048	0.455084	0.451	0.158501
Expected Index	-0.0036	-0.0006	-0.0002	-0.0000	-0.0000
#obs	281	1596	6384	25536	637340
R^2	0.0931	0.396	0.476	0.451	0.103
const a	0.0637	6.44e-006	3.02e-016	-2.19e-015	-2.07e-014
std-err a	0.0334	0.0125	0.00598	0.00306	0.000586
t-stat a	1.91	0.000514	5.04e-014	-7.16e-013	-3.53e-011
p-value a	0.0573	1	1	1	1
slope b	0.0179	0.405	0.455	0.444	0.159
std-err b	0.0334	0.0125	0.00598	0.00306	0.000586
t-stat b	5.35	32.4	76.1	145	271
p-value b	1.83e-007	0	0	0	0
permutations	99999	99999	99999	99999	99999
pseudo p-value	0.000050	0.000010	0.000010	0.000010	0.00010
mean	-0.0036	-0.0006	-0.0002	-0.0000	-0.0000
st. deviation	0.0368	0.0127	0.0063	0.0031	0.0015
z-value	4.9569	31.9644	72.2444	141.3849	255.7654

In Table 3 values of R^2, const a, std-err a and t-stat a, are indicating statistical values for the relationship between observed immigrant population (ULKOKANS) and lagged version (lagged_ULKOKANS). More important are the results of slope b, std-err b, std-err b and t-stat b which are indicating result of immigrant population (ULKOKANS). All the results are confirming presence of positive spatial autocorrelation in the area. Most important result is p-

value b representing significance of Moran's Index of immigrant population indicating very high statistical significance with its values being much smaller than general cut-off value of 0.05.Additional statistics gained from permutation tests is confirming previous results. Moran's Index have the same values like in a scatterplot graphs and permutation test in ArcGIS, which is also the case with Expected Index and z-values. Pseudo p-values are testing one more time significance, with even bigger confidence levels (0.000050 and 0.000010) confirming that the results are highly significant. With values lower than 0.05, we can acknowledge presence of immigration population clustering in the area, reject concept of CSR and null hypothesis and continue with implementation of LMI and mapping of possible clusters.

Results provided by ArcGIS and GeoDa are comparable. However, they are slightly, but statistically non-significantly different. It can be therefore stated that these results are providing enough proof that output results of analysis conducted in ArcGIS and GeoDa are comparable, allowing comparison of the results, gained by Global Moran's Index method. Results produced by LMI (maps computed in ArcGIS and GeoDa) are allowing visual interpretation of clusters located in HMA area.

6.2. Local method of spatial autocorrelation

On the basis of the results produced by GBI which indicated presence of clustering in HMA area we are certain that produced results of LMI and there p-value is are going to be statistically significant, so the clusters or outliers created by LMI, produced in ArcGIS and GeoDa can be compared.

Features that have non statistical significance are marked as Not Significant. Comparable results produced in ArcGIS and GeoDa are maps producing cluster and outlier values. Maps are comparable having two different clustering outputs (High-High and Low-Low), as well as two different outlier values (High-Low and Low-High).

6.2.1. Mapping immigrant population clusters of small areas (pienalue) using ArcGIS

Clusters and outliers of immigration population of small areas being analyzed in this set up are exhibiting specific spatial behaving in certain parts of HMA (Figure 17). Presence of High-High clusters is noticeable, but there is absence of Low-Low clusters. Results presented in Figure 17 are exhibiting presence of High-High clusters in municipality of Helsinki and

Vantaa (administrative cross-border cluster). Eastern part of Helsinki is characterized by presence of massive High-High cluster, indicating huge concentration of immigrant population. Second cluster is located in Western part of Helsinki and Southern part of Vantaa, forming cross- border area. Downtown and East of Helsinki are exhibiting High-High value (three small areas). Presence of outliers is depicted and marked by High-Low outliers presented in Vantaa (two small areas). In Helsinki area noticeable is the presence of Low-High outlier, indicating low values of immigrant in specific area which is surrounded by high amount of immigrants in surrounding area. It is a consequence of spatially significant High-High cluster presented in the area.

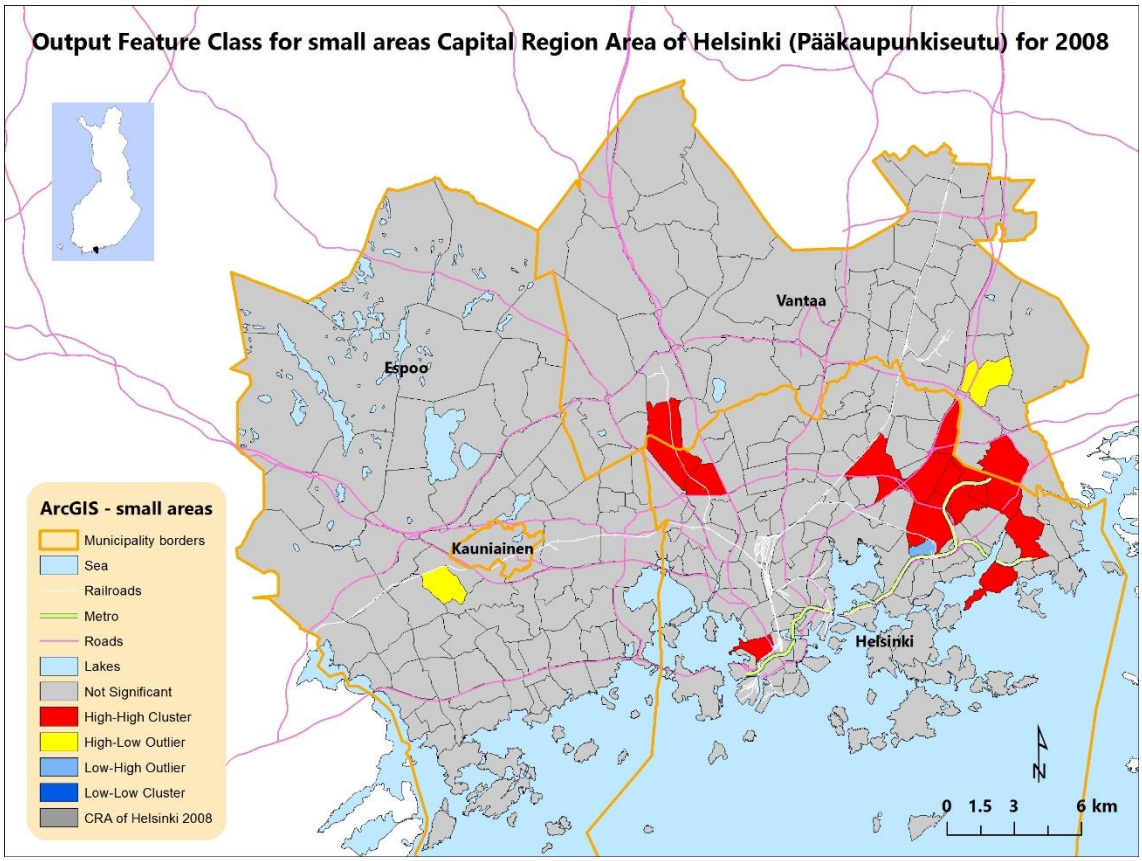


Figure 17. Clustering values and spatial behavior of immigrant population for small areas (pienalue) ArcGIS LMI map (Kekez, 2014).

6.2.2. Mapping immigrant population clusters of small areas (pienalue) using GeoDa

In GeoDa, presence of clusters of High-High (red) and Low-Low (blue) values in HMA small areas (Figure 18) is marked with two significant clusters of each type, occupying significant spatial territories of analyzed area. Cluster of High-High values located in East of Helsinki

and second cluster is located in a triangle between administrative borders of municipalities of Helsinki, Espoo and Vantaa. In Helsinki, there are two more small areas (pienalue) with High-High values not belonging to any clusters, one located in South and other in East. Low-Low clusters are located in North of Espoo and Vantaa and West part of Helsinki next to administrative border of municipality of Vantaa. Outskirts of municipality of Helsinki are exhibiting Low-Low values in certain areas without forming noticeable cluster. Outliers are presented throughout the analyzed area of HMA. There is two High-Low (yellow) small areas (pienalue), located in Vantaa. One is located in the South-West part, connecting two clusters of Low-Low areas. Other one is located in South-east part of Vantaa. There is five small areas (pienalue) which are marked as Low-High (light blue) outliers, four of them being present in municipality of Helsinki and one located in municipality of Espoo.

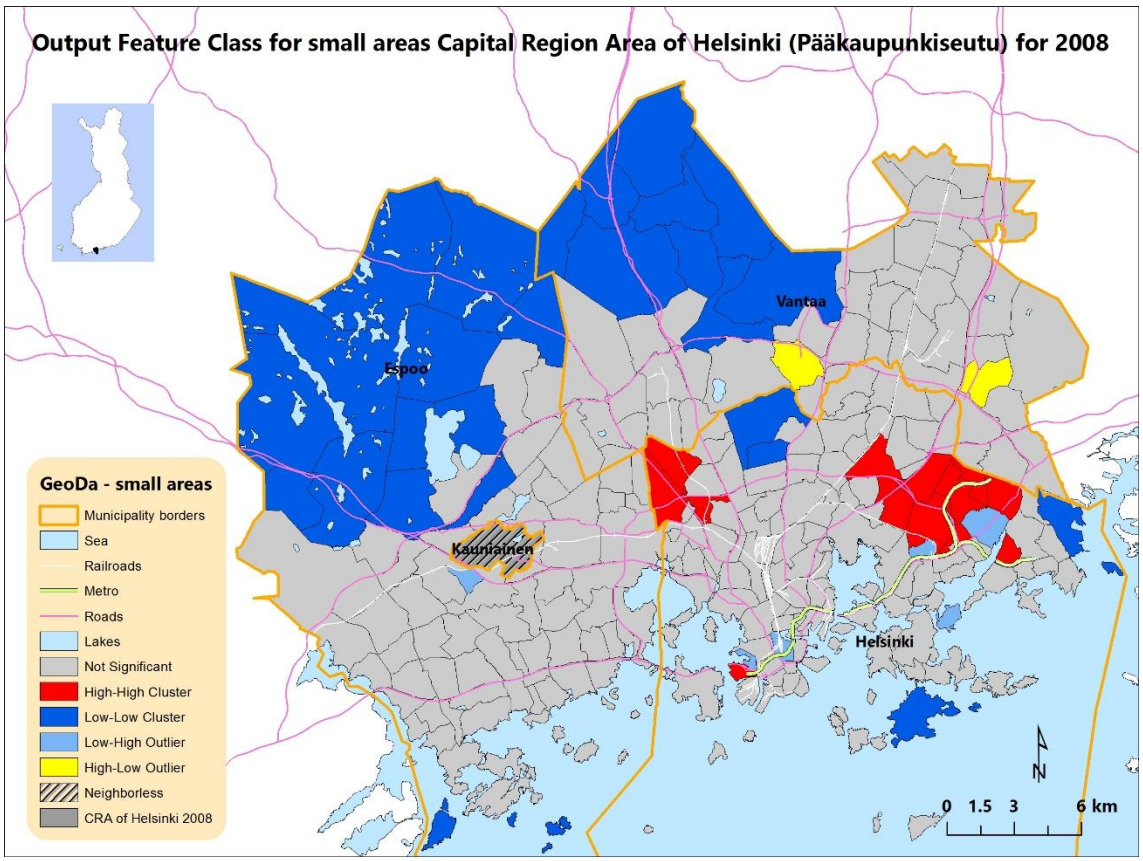


Figure 18. Clustering values and spatial behavior of immigrant population for small areas (pienalue) GeoDa LMI map (Kekez, 2014)

6.2.3. Mapping immigrant population clusters with grid cell size of 1000×1000m using ArcGIS

Map of lattice cell level 1000×1000m size, produced in ArcGIS (Figure 19) is providing representation of clusters of High-High values spread out throughout HMA. There is one dominant spatial cluster crossing over administrative borders of Helsinki, Vantaa and Espoo (the majority is concentrated in Helsinki municipality, connecting it with same value grid cells in Western part of Espoo municipality, South-East and South-West part of Vantaa municipality). Following the concept of contiguity of spatial autocorrelation of similar values, two smaller clusters of High-High values are formed in North of Vantaa and three are formed in Espoo (one in central part and two in the South). Presence of outliers is marked by one Low-High value cell, presented attached to the dominant cluster of High-High values not forming significantly spatial area, but notifying us about certain specific behaving on that territory. In this case it is represented by a one square kilometer area concentrated in the East part of Helsinki municipality. Dark grey color areas are representing non-significant areas.

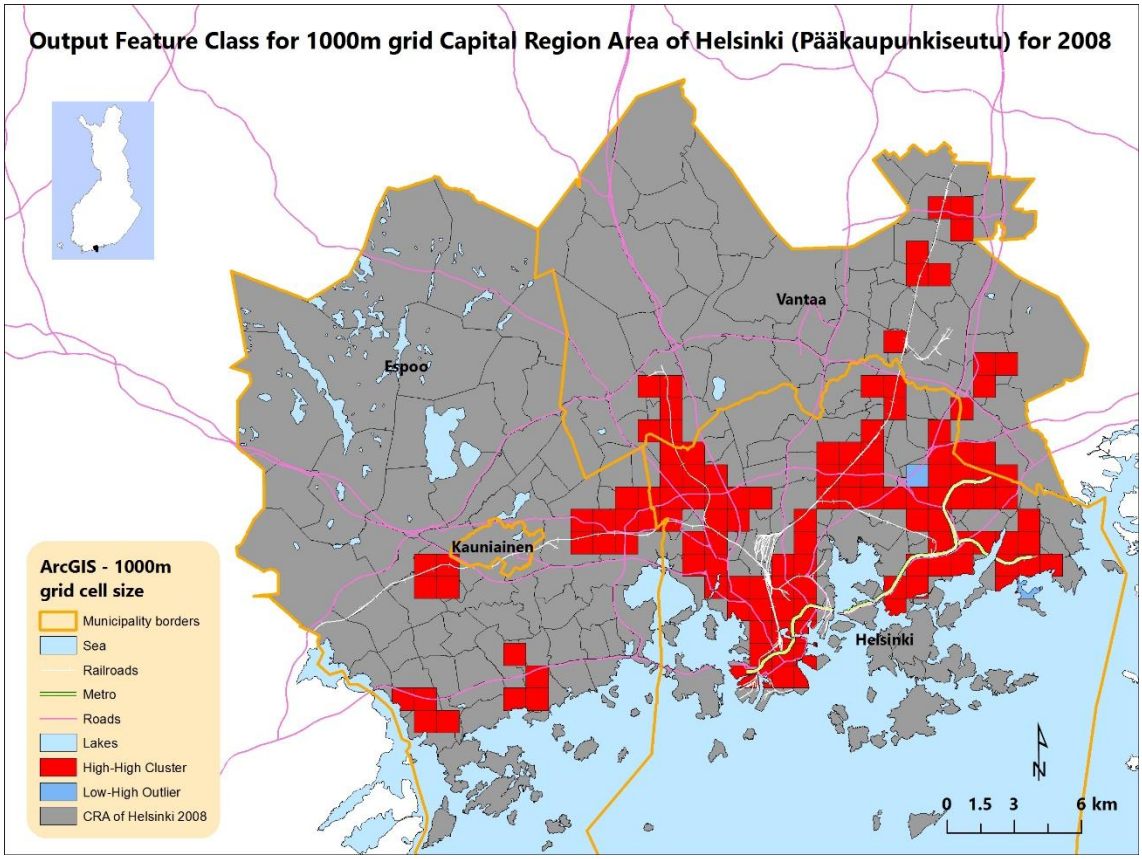


Figure 19. Clustering values and spatial behavior of immigrant population for lattice grid level size 1000×1000m ArcGIS LMI map (Kekez, 2014)

6.2.4. Mapping immigrant population clusters with grid cell size of 1000×1000m using GeoDa

Presence of clusters of High-High and Low-Low values is more distinct and location specific in GeoDa's map of lattice level 1000×1000m grid cell size (Figure 20). The best indication is the formation of the two spatially significant clusters of High-High values instead of one produced in ArcGIS. One is located in the Western part of Helsinki sharing border with Espoo and Vantaa taking into account certain areas of these municipalities and other one is located in East part of Helsinki and South of Vantaa forming a cross-border cluster between these two municipalities. Majority of the area of these spatially significant clusters is the same like in ArcGIS (same level map) with exception of presence of Low-High value outliers located attached to the outer border of the both High-High clusters. In the "East cluster", belonging to Helsinki and Vantaa there is one spatially specific Low-High outlier of three square kilometers formed within cluster of High-High values. In Espoo three spatially smaller outliers of Low-High values are occurring, located within almost the same areas presented in ArcGIS map.

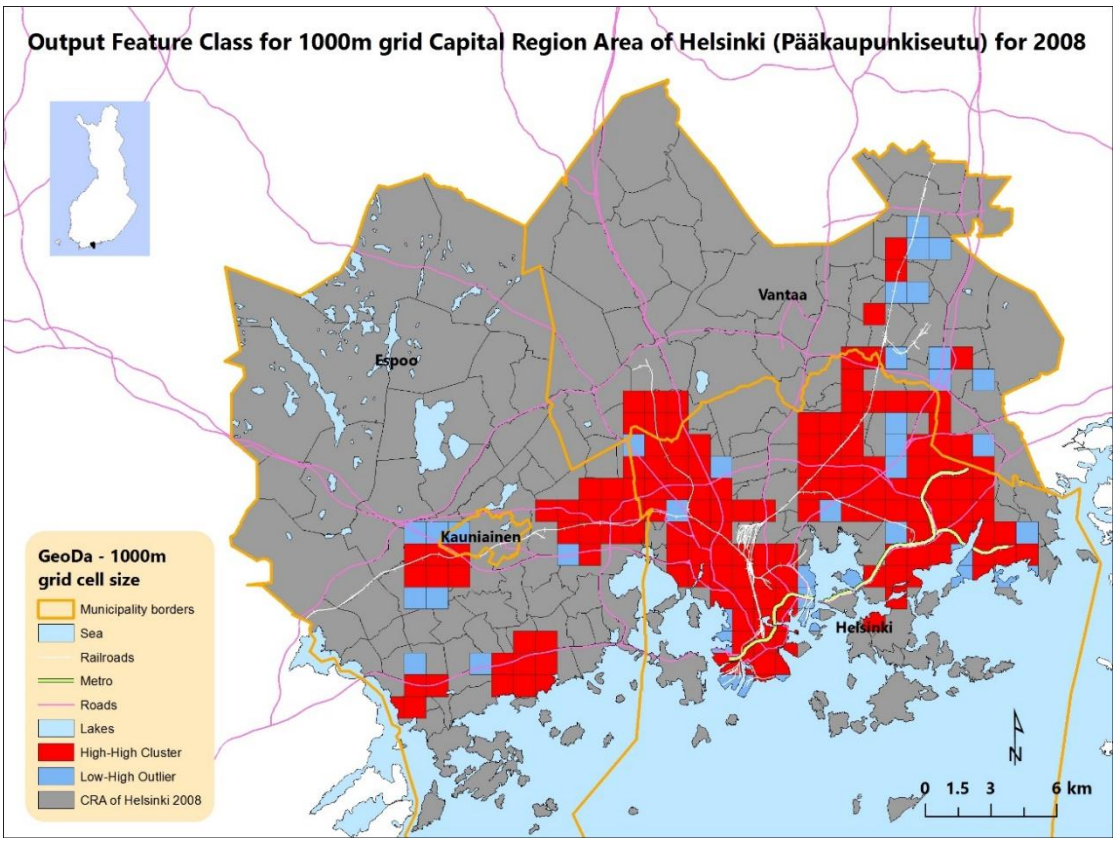


Figure 20. Clustering values and spatial behavior of immigrant population for lattice grid level size 1000×1000m GeoDa LMI map (Kekez, 2014).

6.2.5. Mapping immigrant population clusters with grid cell size of 500×500m using ArcGIS

Map of lattice level 500×500m grid cell size produced in ArcGIS (Figure 21) is in comparison with the map of lattice size of 1000×1000m also produced in ArcGIS, providing us with more specific representation of the results of clusters of High-High values. Representation of spatial concentration of immigration population spread out throughout HMA, is forming more specific spatial clusters. They are crossing over administrative borders of Helsinki, Vantaa and Espoo following concept of contiguity and spatial autocorrelation of similar values. Low-Low cluster values are not presented as well as outlier's value in ArcGIS. Specific clusters of High-High values are formed throughout the area of HMA forming more specific clusters of smaller sizes but more specific spatial locations. Clusters formed Espoo are located in the same spatial locations as in map of lattice size of 1000×1000m but they are taking into account more specific areas, defining spatial locations more precisely, which is also the case for clusters formed in Vantaa.

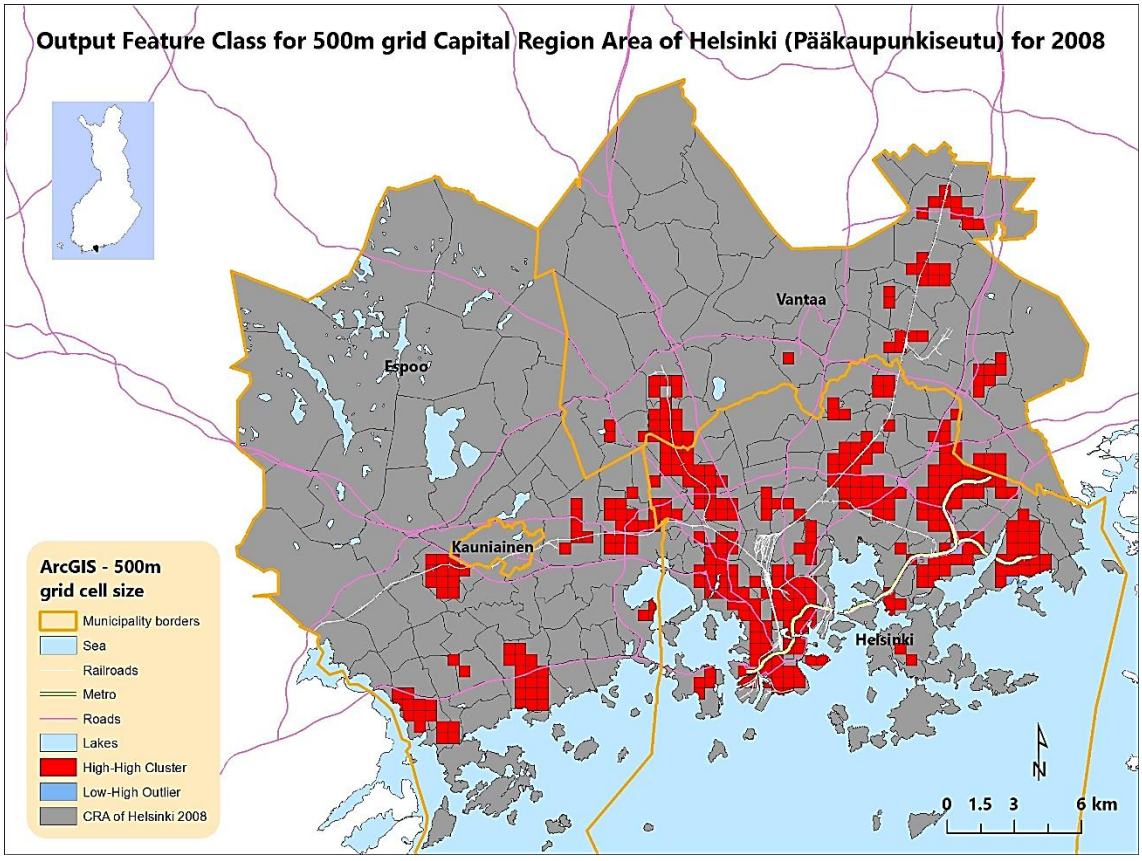


Figure 21. Clustering values and spatial behavior of immigrant population for lattice grid level size 500×500m ArcGIS LMI map.

6.2.6. Mapping immigrant population clusters with grid cell size of 500×500m using GeoDa

Lattice level map of 500×500m size grid cells produced in GeoDa (Figure 22) is represented by formation of clusters of High-High values throughout CHR area. Exhibited patterns are coinciding with map of lattice level of 1000×1000m size of the grid cells produced in GeoDa but they resemble more specific spatial clusters. Clusters are formed more compact, than in ArcGIS map of the same level including bigger number of cells which are forming clusters of High-High values. They resemble larger amount of territory. Clusters are surrounded with High-Low outlier cells with several outliers developed around big clusters, representing higher concentrations of native population in these certain areas.

There is only one cell that represents High-Low outlier located in North of Espoo municipality, not having any kind of statistical importance of affecting developed spatial processes going on in HMA area.

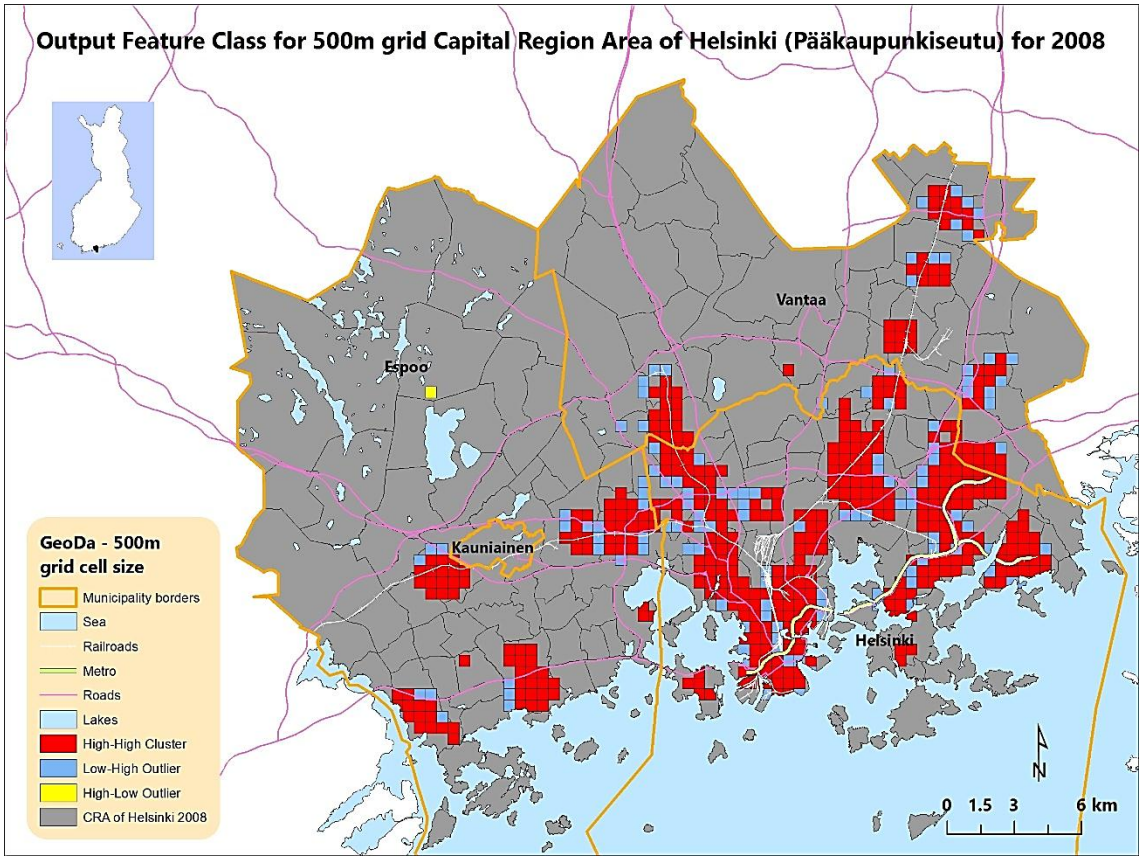


Figure 22. Clustering values and spatial behavior of immigrant population for lattice grid level size 500×500m GeoDa LMI map (Kekez, 2014).

6.2.7. Mapping immigrant population clusters with grid cell size of 250×250m using ArcGIS

Map of lattice level 250×250m grid size cells produced in ArcGIS (Figure 23) is exhibiting clusters of High-High values formed in Espoo, following the trend of spatial locations as in map of lattice size of 500×500m, but in Helsinki clusters are formed in more specific and more accurate manner. Previously located huge cluster in West part of Helsinki covering triangle border area of Helsinki, Vantaa and Espoo in 1000×1000m grid is now dispersed into three significant clusters located in West of Helsinki, cross-border cluster located on the West border of Helsinki and East border of Espoo and cross-border cluster located on the West border of Helsinki and South border of Vantaa. Cluster located in East of Helsinki, noticed in 1000×1000m grid is now dispersed into four different clusters. One is located in East part of Helsinki, other cross-border one located on the East border of Helsinki and South-East border of Vantaa and two located in South-East of Helsinki. There is only one outlier cell attached to the biggest cluster. High-Low outlier cell is located in administrative cross-border cluster located in East of Helsinki and South of Vantaa, with cell being located in South of Vantaa.

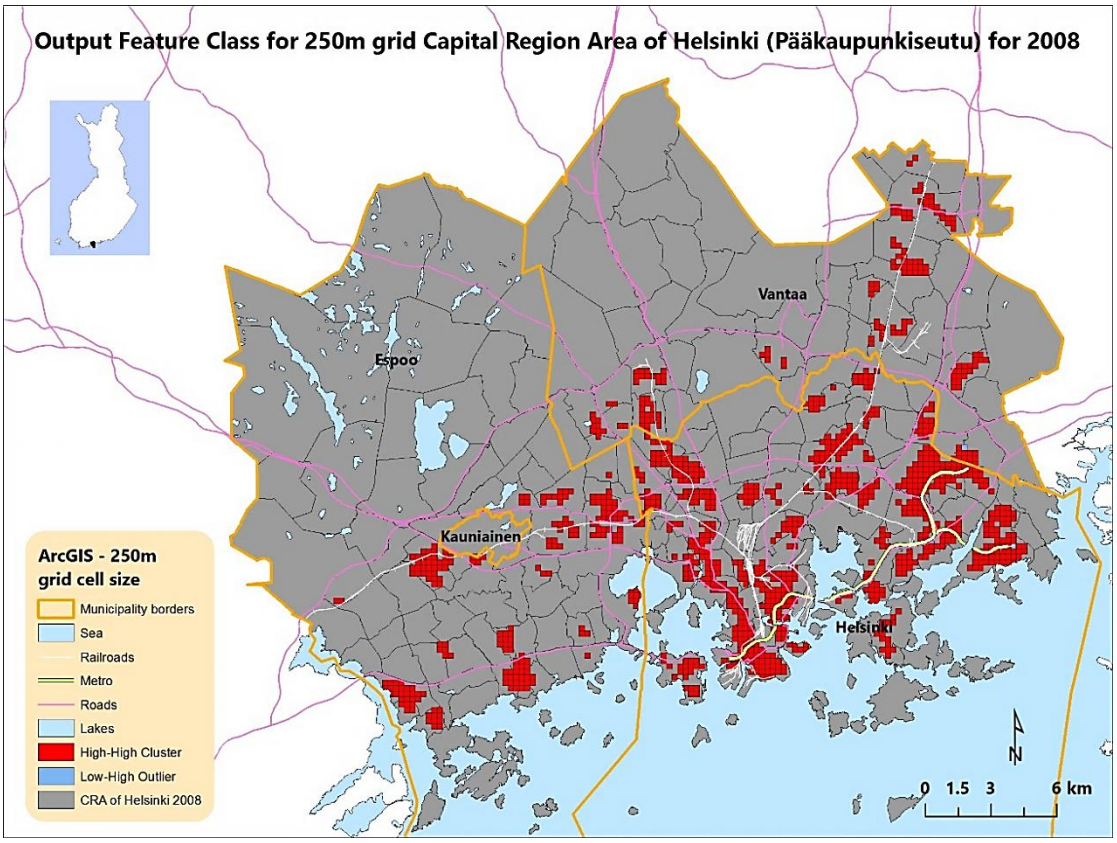


Figure 23. Clustering values and spatial behavior of immigrant population for lattice grid level size 250×250m ArcGIS LMI map (Kekez, 2014).

6.2.8. Mapping immigrant population clusters with grid cell size of 250×250m using GeoDa

Map of lattice level 250×250m grid size cells produced in GeoDa (Figure 24) is producing more precise clusters and outliers around HMA area. Following up territorial patterns formed in GeoDa maps of coarser lattice levels (1000m and 500m), formed clusters represented in Figure 24 are more accurate and precise spatial processes. High-High clusters are located on the same spatial location like in the ArcGIS map (250m) but taking into account more cells. Core of their spatial locations is the same like in the results produced in ArcGIS, spread out through HMA area. Formation of outliers as a significant spatial trend is not noticeable, but Low-High outlier cells are concentrated around massive High-High clusters. Random occurrences of High-Low clusters cannot be described as continuous spatial processes, but still they are processing meaningful information about certain cell size behaving of population. Interesting aspect of this phenomenon is that they are exhibited only in Vantaa and Espoo where occurrences of clustering of High-High values are spatially more scattered than how it is the case in Helsinki where processes is more spatially compact.

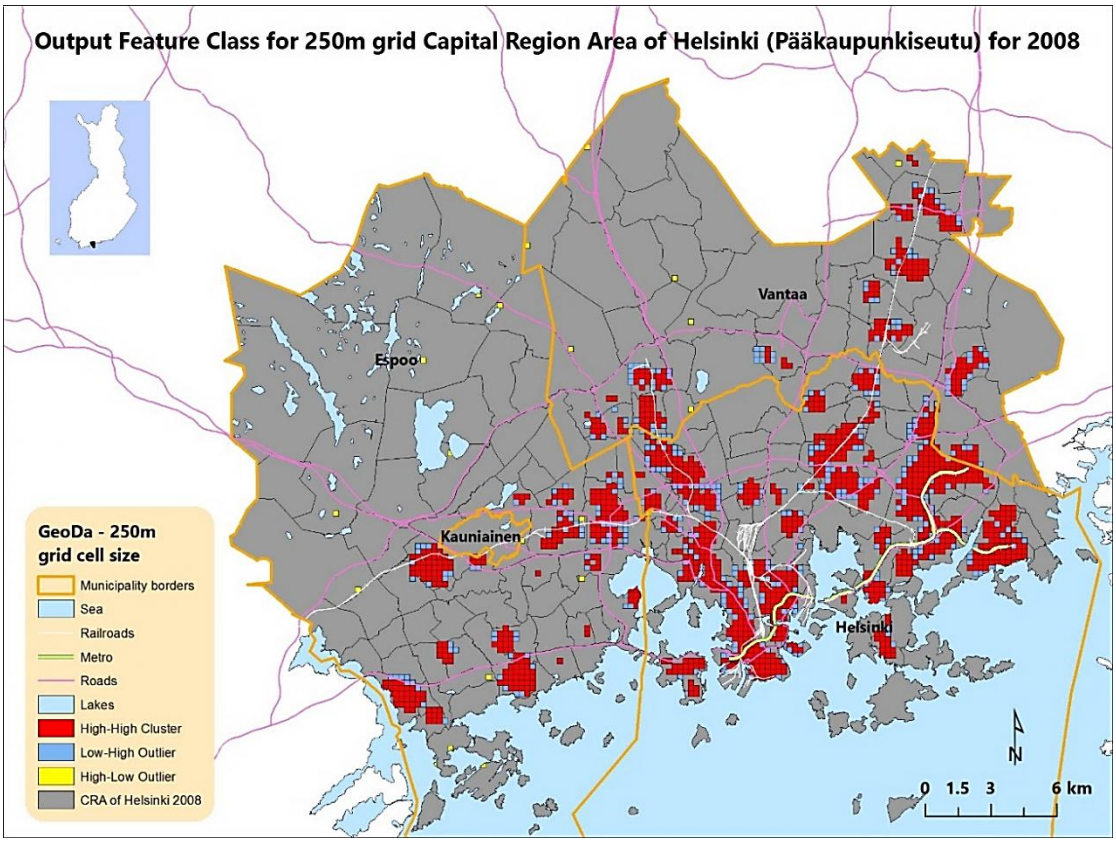


Figure 24. Clustering values and spatial behavior of immigrant population for lattice grid level size 250×250m GeoDa LMI map (Kekez, 2014).

6.2.9. Mapping immigrant population clusters with grid cell size of 50×50m using ArcGIS and GeoDa

New, previously unused lattice level size of 50×50m was also used in this study and it seems to reveal new interesting results for immigrant population cluster analysis in the study area. Maps of lattice level 50×50m grid size cells produced in ArcGIS (Figure 25) and GeoDa (Figure 26) are exhibiting completely changed view on the formation processes of clusters and outliers in HMA area.

Clusters of Low-Low areas are not presented. One cell is produced in GeoDa map representing statistically and spatially insignificant result. High-High clusters are scattered around cores of previously detected High-High cluster areas in maps of different levels (small areas, 1000m, 500m and 250m). They are more dispersed spatially forming small significant areas in central, East and West area of Helsinki; South-West, North-East and South-East of Vantaa; South, East and central part of Espoo.

Formation of outliers is not any different than how it is detected in the previous maps of higher lattice grid cell size. There is exhibited difference in computational capabilities of producing and exhibiting spatial outliers of Low-High values between ArcGIS and GeoDa, resulting in remarkable different visual representations of outlier values (Figures 25 and 26). Outlier cells are occurring around some clusters located around the most significant clusters of High-High values in HMA area in ArcGIS and on the results presented in GeoDa they are occurring all around High-High value clusters of different sizes being presented in much larger number than what is the case with ArcGIS.

Manifestation of High-Low values is happening also on random basis without any specific spatial processes defining and depicting it. There is higher number of cells presented in both maps (Figures 25 and 26) than what was the case with maps of previous lattice levels (small areas, 1000m, 500m and 250m). One of the reasons for this spatial behaving can be explained through greater amount of cells being analyzed in this certain set-up.

Results provided by Local Moran’s Index spatial cluster maps are providing us with meaningful locations of occurrence of spatial clusters and outliers which can help us to explain formations of clusters of immigrant population in Helsinki Metropolitan Area.

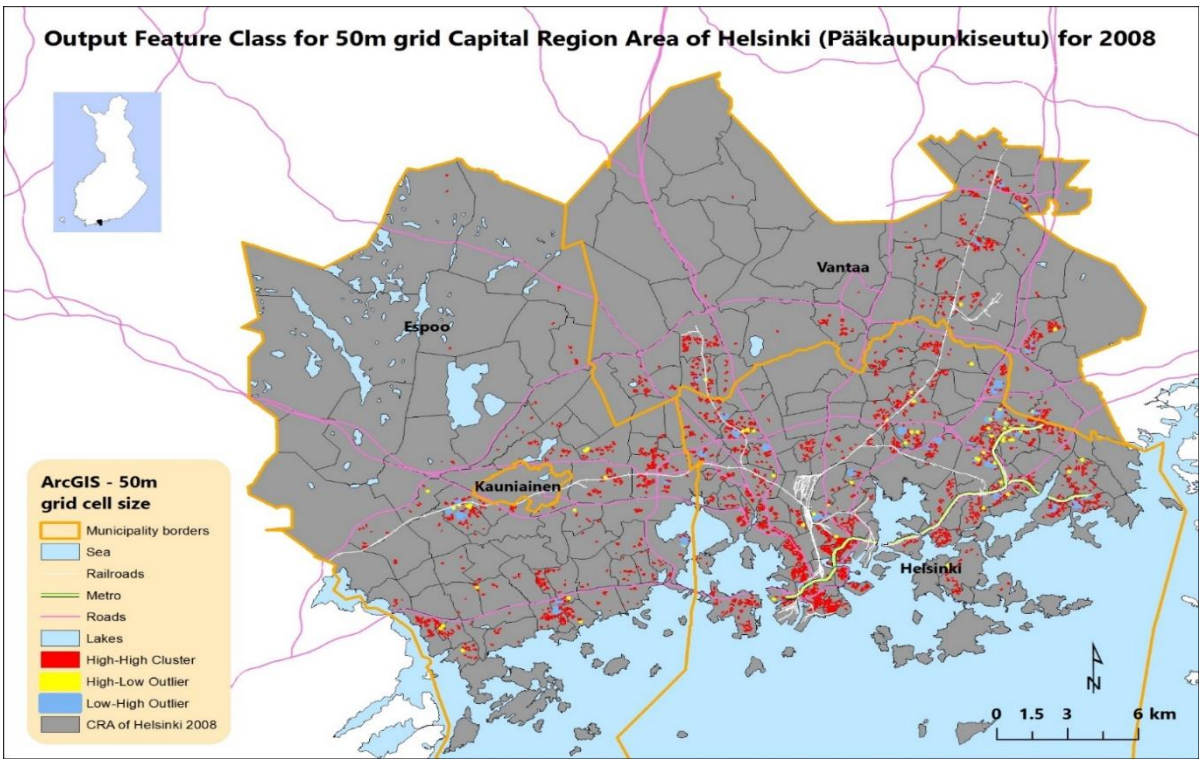


Figure 25. Clustering values and spatial behavior of immigrant population for lattice grid level size 50×50m ArcGIS LMI map (Kekez, 2014).

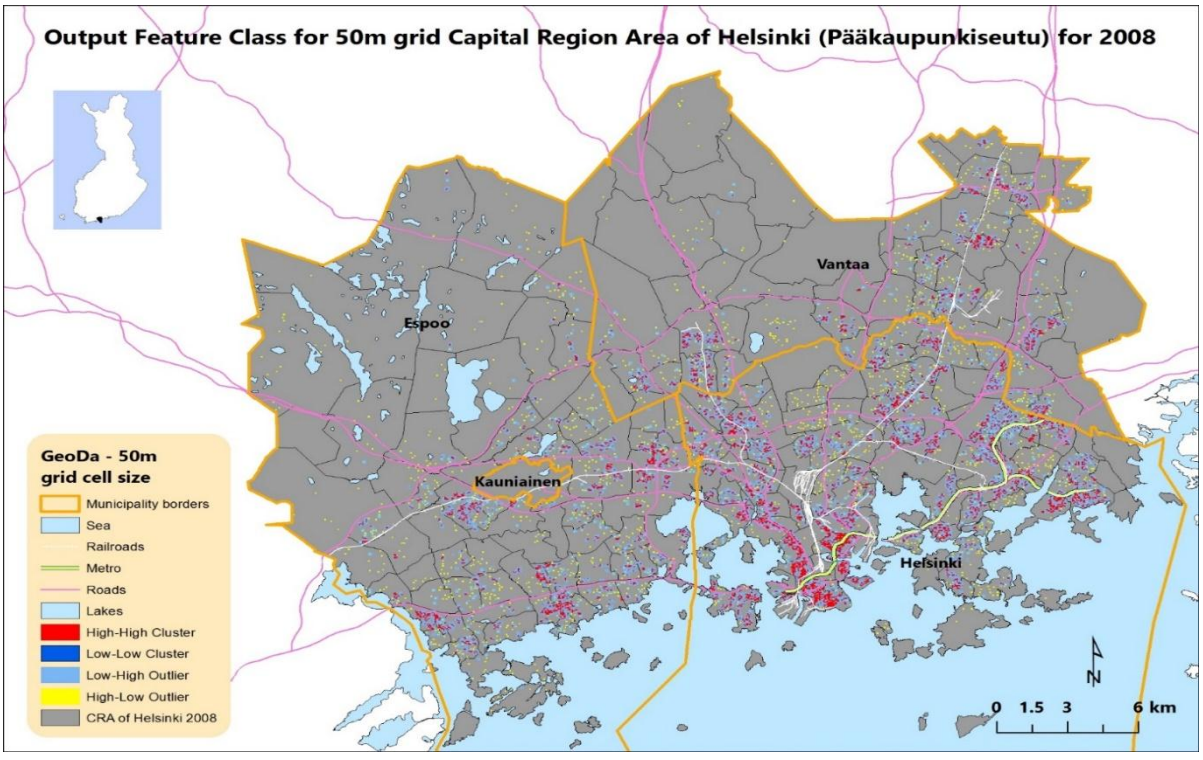


Figure 26. Clustering values and spatial behavior of immigrant population for lattice grid level size 50×50m GeoDa LMI map (Kekez, 2014).

7. DISCUSSION

7.1. Methodology of immigration studies in Finland and Helsinki Metropolitan Area

Immigration studies conducted in Finland, dealing with problems of immigration population are using different set of methods (geographical, economical and sociological) and most of the previous studies e.g. (Jasinskaja-Lahti, 2000; Phinney et al., 2001; Lehti & Aromaa, 2002 Heikkilä & Peltonen, 2002; Koivukangas, 2003; Heikkilä & Järvinen, 2003; Gulijeva, 2003; Forsander, 2003; Musterd et al. 2008; Söderling, 2010; Łobodzińska, 2011) are not using advanced, inferential statistical or GIS methods in the research and representation of the results. It must be stated that most of the researchers producing these previous studies are human geographers or sociologists, which are probably not trained to use state-of-the-art, advanced GIS based spatial statistical methods such as spatial autocorrelation calculation, and therefore limiting them to the use more traditional statistical methodologies in human geography or sociology.

Finnish population studies employing methods of spatial autocorrelation are conducted in recent years by Vasanen (2009) and Lehtonen & Tykkyläinen (2010). These two studies are employing global and local methods of spatial autocorrelation in research, indicating more advanced possibilities of analyzing spatial patterns of clustering and clusters in the practice of population studies.

They are not specifically focusing on the research of immigration population, specifically not in Helsinki Metropolitan Area. Need for employment of these methods is unquestionable. From the perspective of use of descriptive statistic in representing the results gained by different methods, hiring up inferential statistical methods for gaining results after right employment of proper methods should make study more GIS usable and efficient.

Main hypothesis of this thesis was that by use of advanced ESDA methods in discovering and mapping potential clusters of immigration population of HMA completely new visual, statistical and presentational capabilities of clustering of immigration population are changing and improving, giving us more precise information of the level of clustering and its physical distribution throughout certain specific areas and HMA area as a whole.

7.1.1. Descriptive statistical studies of immigration population

City of Helsinki, Urban facts represents probably the most important institution, when it comes to processing and exhibiting statistical and spatial statistical data of immigrants in Helsinki. Perspective from which City of Helsinki, Urban facts is explaining spatial concentrations of immigration population in Helsinki area is analysis of immigration population in comparison with native population. Statistical methods used in this study are belonging to descriptive statistics, dealing with a notion of space as a restriction of certain spatial process, describing it in one spatial territory without taking in consideration First Law of Geography (Tobler, 1970) and possible effects of surrounding neighbors. In Finland, especially in HMA where majority of immigration population lives, this demographical issue is almost always being analyzed in a comparison with a native population, depicting percentage number of immigrants living in a certain small area (pienalue) of Helsinki out of the number of all population living in the area (Figure 27).

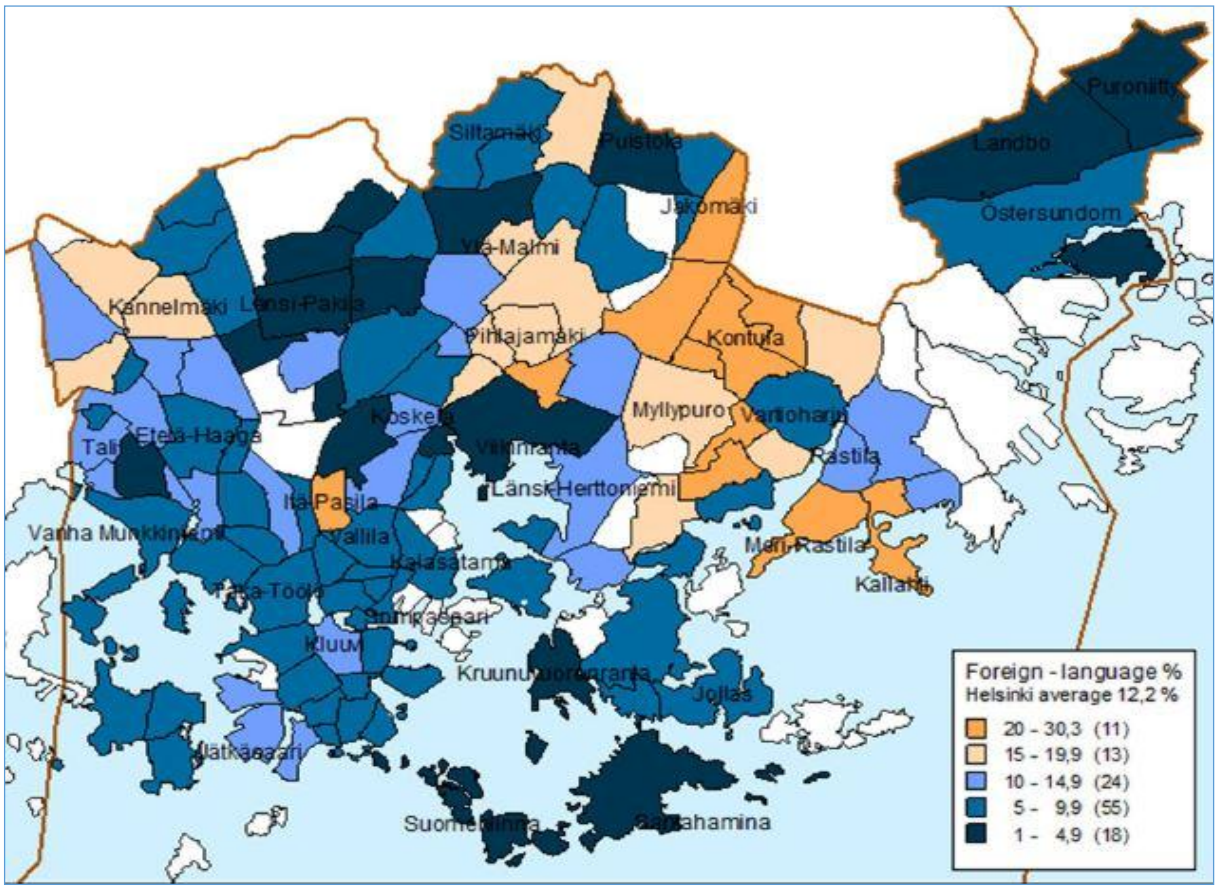


Figure 27. Proportion of foreign-language residents in the population of Helsinki sub districts on 1 January 2013 (City of Helsinki, Urban Facts, 2013).

Interpretation of the results gained from this kind of studies is done in following manner:

"Those residents with a foreign mother tongue most typically live in Helsinki's Eastern-Major District–28 percent of them do. The proportion of foreign-language residents has grown fast in the Eastern and the North-Eastern Major Districts. This proportion was smallest in the Northern Major District and in Östersundom Major District, both of which predominantly have detached and terraced houses." (City of Helsinki, Urban Facts, 2013)

Interpretation of these results resembles quite limited perspective on phenomenon of clustering and specific spatial locations of immigration population (citizens not speaking Finnish, Swedish or Sami as a native language). It provides information of huge concentrations of immigrants living in suburban areas. It leads to conclusion that immigrants are only concentrated in these certain areas.

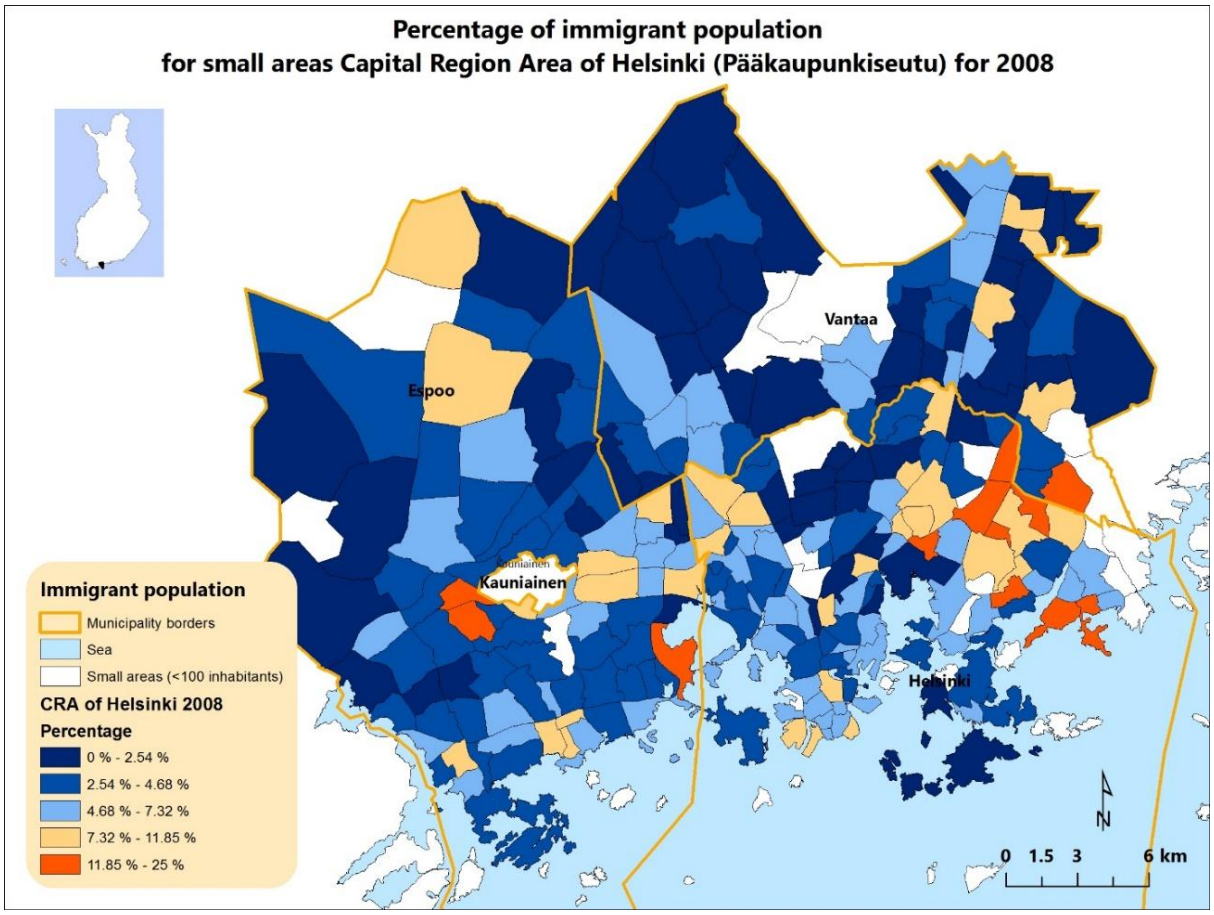


Figure 28. Proportion of foreign-language residents in the population of Helsinki Metropolitan Area in small areas (pienalue) in 2008 (Kekez, 2014).

In above map (Figure 28) same interpretation is performed for HMA as whole for the year 2008. It points out that the process of concentrations of immigrants described by descriptive

statistical methods has not been changed since 2008. What occurs as a change is that the highest percentage level of 25% which is slightly lower in comparison with previous map (Figure 27) indicates higher concentration in Helsinki if analyzed separately from HMA. What is not noticed is a high concentration of immigrant population in a neighboring small area (pienalue) in Vantaa. Administrative organization of municipalities is providing limitations in presentation of concentrations of immigrant population. Spatial process is not fully represented, giving a limited perspective about its size and volume.

7.1.2. Inferential statistical studies of immigration population

Use of statistical methods, precisely descriptive statistical methods in explanation processes and representations of measured and exhibited problems (social housing, ethnic segregation, integration, etc.) is presented in studies dealing specifically with immigration population of HMA area (Vaattovaara 1998, 2001, 2002; Vilkama 2007, 2011; Vilkama & Dhalmann 2009). Notification of importance of use of more advanced GIS methods in immigration studies is done in PhD dissertation of prof. Mari Vaattovaara (1998). She is making an excellent point that "*use of GIS in the examination of social spatial patterns is crucial"* Vaattovaara (2001), marking examination of social spatial patterns and segregation examined by various GIS methods conducted in ArcGIS at the end of that decade. Computing and analyzing power have changed a lot since that time and nowadays there is much more powerful tools and methods allowing more advanced processing and representing of data on immigration population. Most important work, which inspired creation of this thesis is done by PhD Katja Vilkama whose work was concentrated on explanation of social patterns of concentration of immigrant population through concepts of social housing and ethnic segregation of immigrant population (Vilkama 2007, 2011; Vilkama & Dhalmann, 2009). Vilkama used MapInfo and less advanced GIS methods in her work using for presentation of the gained results of the spatial concentrations of immigration population (Vilkama 2007, 2011; Vilkama & Dhalmann 2009).

Inspired by the previous works conducted by Vaattovaara (1998, 2001, 2002), Vilkama (2007, 2011) and Vilkama & Dhalmann (2009) this thesis is trying to analyze processes of spatial concentrations of immigration population on slightly different basis, using primarily concept of spatial autocorrelation as a method of explanation of formation of the clusters of immigration population. Conceptualization of spatial concentration of immigrants in this thesis is conducted on the basis of spatial autocorrelation (Goodchild, 1987; Haining, 2009;

Fotheringham, 2009) through the concept of contiguity based up on the ***"First law of geography"*** defined by Tobler (1970)*: "Everything is related to everything else, but near things are more related than distant things"*. Inferential statistical methods (ESDA methods) are going beyond obvious, analyzing and explaining processes going on in the area from different perspective. By implementing new methodological processes, new results are expecting to occur and new perspective about immigration population spatial concentrations should be gained. First time inferential statistical methods (Global and Local Moran's Index), are employed in the study of immigration population in HMA area analyzing spatial patterns (clusters and outliers) conceptualized on the basis of spatial autocorrelation.

7.2. Data

Data creation

Data used in this study is provided and created by HSY and its main purpose is to asses Finnish municipalities with information for conducting planning processes. Because, MapInfo is software mainly used by Finnish municipalities for purposes of analyzing, planning and mapmaking as the final outcome of the processes data is produced in different MapInfo formats. There is several reasons which are explaining massive use of MapInfo: one of the first GIS software introduced in municipalities, cheaper licensing than ArcGIS, long period of use with difficulties in shifting to the use of more powerful software and similar issues. They are depicting this software as dominant on a market used in processing, analyzing and mapping exactly the same or similar data, like used in this thesis. Nowadays computing and analytical, as well as visual capacities of Mapinfo are almost on the same level like free GIS software (Quantum GIS, GeoDa), or even worse in the case of ArcGIS, data is still produced almost exclusively in Mapinfo format.

Inferential statistical methods, especially spatial statistical methods are not included in MapInfo package, leaving majority of planners out of capacities and capabilities to analyze processes with more precise and powerful methods and tools. Mistakes in data produced for MapInfo could also be more avoidable if more powerful software would be used, like ArcGIS.

Data mistakes

Data within itself contains a lot of mistakes which was noticed in the process of digitizing. Editing tools where used for replacing incorrectly edited features representing borders of certain areas (pieni_aluet) in the case of municipalities of Helsinki and Kauniainen.

Population data is produced in a point pattern manner. All the points should match center of living unit which they are representing and conceptualized PKS_VAKI shape file containing population data is full of displacement of points, stepping out of the building areas. This could lead to the assumption of precision and quality of the produced results. Problem may also lay in a file representing buildings.

After performing of spatial joining simple statistics was conducted for provision of the gained results and performed quality of this relationships. Lattice and population files were both produced by HSY. Only the smallest lattice level (50×50m) was created by the author, for testing hypothesis of specific clustering unnoticed before. Table 4 is representing results of spatial fit of the number of immigrants varying in different lattice levels:

Table 4. Fit of immigrant population analyzed in certain lattice levels (Kekez, 2014).

Fit difference of immigration population in different lattice levels

Population	Immigrants	Difference (compared with pks_vaki 08)
pks_vaki 08	59875	
small areas	56099	3776
1000×1000m lattice	56280	3595
500×500m lattice	56340	3535
250×250m lattice	57812	2063
50×50m lattice	47143	12372

There was no information provided by the publisher of the data that some data may be lost during the process of spatial joining and it happened due to unknown reasons. Unfortunately, there was no indication in the previous works conducted by other researchers about this misfit. Table 4 is providing meaningful information on differences in population numbers for different lattice levels. Provided results are indicating that the best statistical fit of number of population as well as the number of immigrant population is provided by 250×250m lattice level size which is also providing us with the most significant statistical as well as visual formation of spatial clusters and outliers. Newly imposed lattice level 50×50m is providing the least satisfying fit. It provides the smallest value of immigration population as well as

native population out of all lattice levels used in the thesis. This could be effect of computation due to the enormous spatial joining operation creating 637340 observations, grid cells. But even if the percentages of population in lattice level of 50×50m would improve results would not change.

Creation of the lattice level of 50×50m (fishnet) helped in creation and calculation of new clusters representing previously not exhibited results on the lattice level of 250×250m. Process of creation was operated in Quantum GIS. Attempts of creation of the lattice in ArcGIS failed probably due to complications in operating capabilities of running such enormous operation from a personal computer connected over VPN Internet connection to ArcGIS.

7.3. Comparison of computational capabilities of ESDA methods

Comparing results gained by ESDA methods (Global Moran's Index and Local Moran's Index) produced in ArcGIS and GeoDa is going to show operating, analytical and computational capabilities of these two software. Comparison is made on the use of the same methods, which are using the same equation, written by the same authors (LISA, Luc Anselin). Weight matrix is conceptualized in the same manner, using the same methodology.

7.3.1. Computing capabilities of Global Moran's Index in ArcGIS and GeoDa

Results gained from Global Moran's Index reports are showing same statistical values and patterns being produced by both software. These reports are showing statistically positive results, confirming presence of clustering as a processes being present in the area in all different levels of conceptualizing units (small areas (pienalue), 1000×1000m, 500×500m, 250×250m and 50×50m).

Results produced by Global Moran's Index performed in ArcGIS are generalized within spatial autocorrelation report (seen in Figure 15). Report is providing statistical values on Moran's Index, Expected Index, Variance, z-score and p-values, see also Table 2. They are not offering that much information about spread and distribution of clustering in area being analyzed. Statistical results are scrutinized to a minimum information being provided in comparison with the results produced by GeoDa's Global Moran's Index report.

Report produced in GeoDa is much more detailed in statistical information provided on processes as well as visual interpretation of the distribution of the values produced in scatter plot graph. It produces scatter plot graph which is already at this stage of analysis offering

spatial distribution of the values (Figure 29) indicating visually possible locations of the clustering values in the area. It provides us with information about spatial clusters before even producing LMI map, providing a meaningful input about necessity to proceed with a creation of LMI map and possible spatial distribution of cluster and outlier values in it. Randomization is additional option which is providing more statistical information on the processes. Statistical values produced by GBI reports (Table 3) in GeoDa are: #obs, R^2, const a, std-err a, t-stat a, p-value a, slope b, std-err b, t-stat b, p-value b.

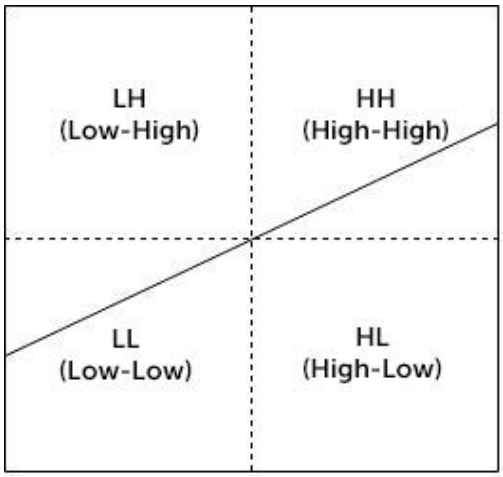


Figure 29.Scatter plot graph of spatial spread of the results in GMI and LMI in GeoDa (Kekez, 2014).

Compared ArcGIS and GeoDa reports, they are providing same information on the processes. Main difference is in the amount of information provided by reports and statistical quality. Values of Moran's Index, Expected Index, p-values (significance is set up on the level of 0.50 so the results could be comparable) and z-score (with a slight difference in the results of small areas level) are the same confirming usage of the same equation in production of the gained results (Table 2 and 3).

Report created in GeoDa is creating much more statistically and visually significant information on the processes. It employs higher amount of detailed information with a possibility of rechecking that the process is not occurring randomly (Randomization). It provides immediately visual representation of the possible clusters and outliers values occurring in analyzed area, making it more visually clear to user what is the distribution of the values of LMI maps before even creating it.

7.3.2. Computing capabilities of Local Moran's Index in ArcGIS and GeoDa

Final results created in ArcGIS and GeoDa are the visual representations of cluster and outlier values of immigration processes going on in Helsinki Metropolitan Area. It is created by the implementation of LMI method, mapping out cluster and outlier values of certain areas being analyzed in relationship towards each other. Results gained from previous GBI in both software indicated presence of clustering, which led to implementation of LMI and creation of LMI maps.

Small areas (pienalue)

Visual representation of the results of LMI for ArcGIS and GeoDa is shown in Figure 17 and 18 which reports similar, but also different visual representation of the produced results.

Results produced in GeoDa are showing presence of clusters of High-High and Low-Low values while in ArcGIS there is only presence of clusters of High-High values. Presence of neighborless area (municipality of Kaunianen for which there is no data), which was undetectable in the same level map in ArcGIS is providing us input on more precise visual representation of statistical and computational results produced by GeoDa.

High-High cluster results are almost the same with few differently included or excluded small areas (pienalue) within a certain cluster created by ArcGIS or GeoDa.

Low-Low cluster is presented only on map produced by GeoDa and it undetected in ArcGIS map. It spreads contiguously throughout the area crossing over administrative municipality border line, located in the outskirts of North of Espoo and Vantaa taking into account massive area of both municipalities. There is few small areas (pienalue) located South of Helsinki and Vantaa with Low-Low values representing locations with extremely small number or complete non-presence of immigrant population. Most of these detected areas are islands.

Outliers are not forming any spatially significant areas in both maps, but they are spread out throughout HMA area, mostly attached to the clusters of High-High and Low-Low values.

Low-High outlier in GeoDa map, located in Espoo is indicating an area in which low value of immigration population is surrounded by high values of native population. Outliers presented in municipality of Helsinki are spread throughout the municipality. In ArcGIS there is only one small area noticed in East of Helsinki, connected to the huge cluster of High-High values. In GeoDa two are attached to spatially significant cluster presented in East part of Helsinki,

purely representing small areas with a huge amount of concentration of native population. One area is located next to standalone High-High area located in south of the municipality. There is two other areas one located in downtown of municipality of Helsinki and one located in South-East, representing Low-High outliers. Two High-Low small areas (pienalue), representing huge value of immigrant population surrounded by low values of immigration population are located in both maps but in a different locations. One is located in the same location (Hakunila in Vantaa) representing an exceptional outlier where huge amount of immigration population is surrounded by huge amount of native population in surrounding small areas (pienalue).

1000×1000m lattice grid cell size

Usage of lattice starts to improve results of clustering of High-High value into more specifically spatially oriented locations of clusters and outliers where spatial behaving of the processes is changing its spatial form, defining some new areas of high clustering of immigrant population in Helsinki Metropolitan Area. It covers much more specific territorial units of all the municipalities including certain areas which have not been detected in previous analyses which were using small areas (pienalue) as a basic unit for analysis.

Majority of newly present clusters of High-High value (Figure are covering areas, which were detected as "Non-Significant", when the basic unit for analyses was small areas (pienalue). In Espoo, three small clusters of High-High values are occurring in both ArcGIS and GeoDa forming clusters from three to eight square kilometers. Sizes of formed clusters are taking into account wider areas than in small area (pienalue) unit level size, explaining better computational capabilities of GeoDa in comparison to ArcGIS.

Clusters of High-High value are represented in a different manner in ArcGIS and GeoDa. GeoDa is creating two huge clusters centrally located mainly in Helsinki but spreading over administrative border of municipalities to Espoo and Vantaa, while ArcGIS is producing one instead of two clusters. Clusters of High-High values are formed also separately in Espoo and Vantaa. In Espoo one is located in the central part, sharing the West border of the municipality of Kaunianen and other two are located in South spreading throughout different small areas (pienalue) of Espoo. In Vantaa clusters of High-High values are formed in ArcGIS which is not the case with results produced in GeoDa. Three small clusters are formed (each consisted of 3 square kilometers) two in the North-East and one in the East of Vantaa.

Outliers are different in spatial occurrence, production and spread. GeoDa is not producing specific clusters of High-High value but occurrence of the Low-High outlier values are indicating different results which was not the case in the map produced by ArcGIS. There is one relatively significant cluster of Low-High outlier values located in the North of the municipality of Vantaa (three square kilometers) and random occurrences of Low-High outlier grid cells representing continuation of huge cluster located in East of Helsinki municipality.

General difference between results produced in GeoDa and ArcGIS is that clusters are occupying bigger and more specific areas (with exception of clusters located in Vantaa) and all of them have attached cells of Low-Low values. Also, GeoDa is producing significant number of outlier cells (High-Low value) and even one cluster in Vantaa.

500×500m lattice grid cell size

This level of lattice grid cells starts to indicate occurrence of realistic spatial processes going on in Helsinki Metropolitan Area. Even if the size of analyzed cells is still too big to explain real spatial processes it represents an improvement in categorization of spatial locations of clusters of High-High value of immigration population. Processes is covering more refined territories of occurrences of clusters following up trends of spatial locations from previous levels defining and discovering some new locations.

As in the previous levels results produced in ArcGIS and GeoDa are different. In ArcGIS previously located huge cluster of High-High values in Helsinki of lattice size of 1000×1000m is now dispersed into six clusters located in different parts of Helsinki: downtown and central part of (the biggest one) Helsinki, cross-border cluster located on the West border of Helsinki and East of Espoo, East part of Helsinki, cross-border one located on the East border of Helsinki and South-East border of Vantaa and one in South-East of Helsinki. GeoDa is producing similar but different results. Instead of six it is forming five significant clusters of High-High values in Helsinki integrating downtown and central part of Helsinki and cross-border cluster located on the West border of Helsinki and East of Espoo into one. Other clusters have same locations but GeoDa is producing more cells within clusters. In Espoo and Vantaa cluster produced in the previous level map (1000×1000m) in ArcGIS are following same spatial trends but defining spatial locations more precisely. GeoDa is producing clusters in the same locations in Espoo but in Vantaa clusters are defined

differently than in previous level map (1000×1000m) following the same trends and locations as in the ArcGIS map of lattice level 500×500m.

Outliers are produced completely differently in ArcGIS and GeoDa. There is only two outlier cells located in South-East of Helsinki noticed in ArcGIS. In GeoDa all clusters are surrounded by outlier cells of Low-High value connecting some of them into one integral unit (clusters conceptualized in Helsinki) following the concept of contiguity. There is no significant outlier of Low-High values in both of maps. Only one cell representing High-Low outlier value is produced in GeoDa and located in North of Espoo.

In comparison with ArcGIS map of lattice level 500×500m, GeoDa is producing higher amount of High-High cluster cells in all areas were clusters are located, giving more specific and spatially larger information about location of clusters. Other important difference is production of Low-High outlier cells surrounding all clusters.

250×250m lattice grid cell size

On the maps of 250×250m lattice grid cell size spatial process of clustering is represented by best fit of virtually created data (lattice) and actual physical processes going on in HMA area. Shapes of clusters are represented most realistic in best level scale, measuring most appropriately processes going, with quantitative catchments of population informing about concentrations of immigrant population within clusters giving a precise input about clustering of immigration population.

Map of lattice level 250×250m grid size cells produced in ArcGIS is in comparison with the map of lattice size of 500×500m is providing better fit and visual representation of the results of the processes of clustering of High-High values spread out throughout HMA. Specific clusters of High-High values are formed throughout the area, representing more accurate specific spatial locations in comparison with 500×500m map. They are taking into account previously undetected areas in the High-High cluster value grids of previous lattice levels (1000m or 500m). These specific spatial clusters crossing over administrative borders of Helsinki, Vantaa and Espoo following concept of contiguity of spatial autocorrelation of similar values are representing most accurate representations of clusters formed in HMA area. Clusters of High-High values are formed in a finer manner, shaping out more understandable areas, to which reader of the map can more relate too. Results conceptualized on the map of lattice level 250×250m grid size cells produced in GeoDa is representing most accurate and spatially significant results produced by both software in any lattice level (1000m, 500m,

250m and 50m). As in previous maps GeoDa is taking into account more cells during the formation of the clusters of High-High value. Locations of clusters are the same in both ArcGIS and GeoDa confirming that this level of lattice is the best fit even computationally.

Low-High outlier cells around High-High resembles, already a typical characteristic for the results produced in GeoDa. Same kind of behaving is exhibited in previous maps of different lattice size (1000m and 500m) in GeoDa. Some of the clusters of High-High value are connecting among themselves with outlier cells of Low-High value. There is not a significant outlier formed in the HMA area. ArcGIS is producing only two cells with Low-High outlier cells. Production of High-Low outlier cells is done only in GeoDa and pattern seemed to be random like in previous maps which exhibited the production of High-Low value outliers.

50×50m lattice grid cell size

Map of 50×50m lattice grid cell size was specially created for purpose of this study. Population data for this study, was conceptualized as a point pattern data representing center of the living unit (house, building, etc.), representing the number and variety of residents (nationality, women, men, children, different age groups, etc.) living inside of them. Realizing that smallest analyzing unit, in which calculation of spatial autocorrelation and contiguity concept was created is 250×250m, new lattice level size was imposed.

Lattice level of 50×50m tends to represent data in more realistic manner. Size of the most of the living units in HMA is either on the level of this lattice size grid cells or it's even smaller. Following the concept of contiguity units of analyses tend to cluster more tightly.

Catchment areas are more clustered where physical units of living (houses, buildings, etc.) are built up attached to each other. Represented area within a catchments of clusters and outliers is taking into account built up areas more concentrated, than in the case of other levels i.e. small areas (pienalue), 1000×1000m, 500×500m, 250×250m and 50×50m levels, where majority of space represented in clusters or outliers is space surrounding units of living (houses, buildings, etc.).

Following explained need for creation of new lattice level of 50×50m, results represented in Figures 25 and 26 are exhibiting new input in representation of cluster and outlier values of immigrant population on territory of HMA. Patterns developed on the map are representing trends from previous lattice level maps (500m and 250m), with more dispersed locations of cluster and outlier values, now presenting tight, very specific locations.

Map of lattice level 50×50m grid size cells produced in ArcGIS is producing the biggest amount of clusters of High-High value dispersed all around HMA area. Clusters of High-High value are following locations from previous lattice level of 250m but, they are more specific, with catchment areas being less spacious, more focused and concentrated on locations of living units.

Results produced in GeoDa are quite similar, but the catchments areas of High-High value clusters are accounting bigger amount of cells. Computing power of GeoDa is accounting more cells within a clusters, than ArcGIS in all the unit levels analyzed and mapped. There is only one cluster cell of Low-Low value produced in GeoDa while ArcGIS is not producing any cells with Low-Low value.

Outliers are produced in the higher quantities out of all produced maps. Number of Low-High value outlier cells produced in this lattice level map in ArcGIS is the biggest in comparison with all the previously produced maps in the same software.

They are not creating any spatially significant territory in comparison with clusters of High-High value and at the same time they are not that frequent, but their occurrences is higher.

Random appearances of Low-High outlier cells is noticeable and concentrated mostly on the edges of clusters of High-High value. Results produced in GeoDa are following the same spatial pattern but occurrences of Low-High outlier cells is highly frequent concentrated all around HMA area.

Some of the outliers are forming significant spatial areas in comparison with clusters of High-High values surrounding them completely without exception.

High-Low outlier cells are formed randomly around HMA area appearing in ArcGIS map attached to the clusters of High-High value or in their immediate surroundings. In GeoDa results of High-Low outliers are different. Frequency of appearance is much higher than in ArcGIS.

Spatially they are scattered throughout the HMA area. Majority of them is surrounded with Non-Significant cells, but certain number of this type of cells is surrounded by High-Low outlier cells surrounding them.

7.3.3. Effect of lattice level (cell size) on spatial distribution of clusters and outliers

Both software, GeoDa and ArcGIS are producing statistically significant, computationally comparable but quantitatively and visually different results.

Computational results for small areas (pienalue) could be more qualitatively than quantitatively compared, because of the conceptualization and number of observed units.

Representation of computational capabilities of the results produced by usage of Local Moran's Index produced on the maps of lattice levels, can be compared visually and quantitatively. Visual representation of the results is shown in Figures 30 and 31.

Numbers of computed units for different level of lattice data is presented in Table 5.

Results presented in Table 5 and Figures 30 and 31 are clearly showing that more computationally and statistically significant results are produced in GeoDa than in ArcGIS. Differences in computational levels are significant.

Values of the results produced are differing from 63 square kilometers in 1000×1000m lattice, 45.75 square kilometers in 500×500m lattice, 34.31 square kilometers in 250×250m and 42.26 square kilometers in 50×50m lattice in favor of GeoDa.

Differences in computational results are the smallest in 250×250m lattice, due to best fit of the results and computational capabilities of software.

Comparison of the results produced in ArcGIS and GeoDa is proving that for more specific analyses of clustering processes, formation of clusters and closer examination of results forming clusters preferably results processed and computed by GeoDa are going to be used.

Lattice levels of 250×250m and 50×50m are going to be used for further examination and location of specific areas of HMA area where clusters are occurring.

Table 5. Computational differences of produced results of LMI maps in ArcGIS and GeoDa (km²) (Kekez, 2014).

Number of units (clusters and outliers) and given size (km²) in different lattice levels

	Computational capabilities					
ArcGIS	HH	LL	LH	HL	Σ (cell)	Σ (km²)
small areas (pienalue)	15		1	2	18	
1000×1000m	122		2		124	124
500×500m	367		3		370	92.50
250×250m	1074		3		1077	67.31
50×50m	6234		163	68	6465	16.16
GeoDa	HH	LL	LH	HL	Σ (cell)	Σ (km²)
small areas (pienalue)	15	30	6	2	53	
1000×1000m	147		40		187	187
500×500m	421		131	1	553	138.25
250×250m	1194		412	20	1626	101.62
50×50m	6472	1	17608	2287	23368	58.42
Software differences	HH	LL	LH	HL	Σ (cell)	Σ (km²)
small areas (pienalue)		30	5		35	
1000×1000m	25		38		63	63
500×500m	54		128	1	183	45.75
250×250m	120		409	20	549	34.31
50×50m	238	1	17445	2219	16 903	42.26

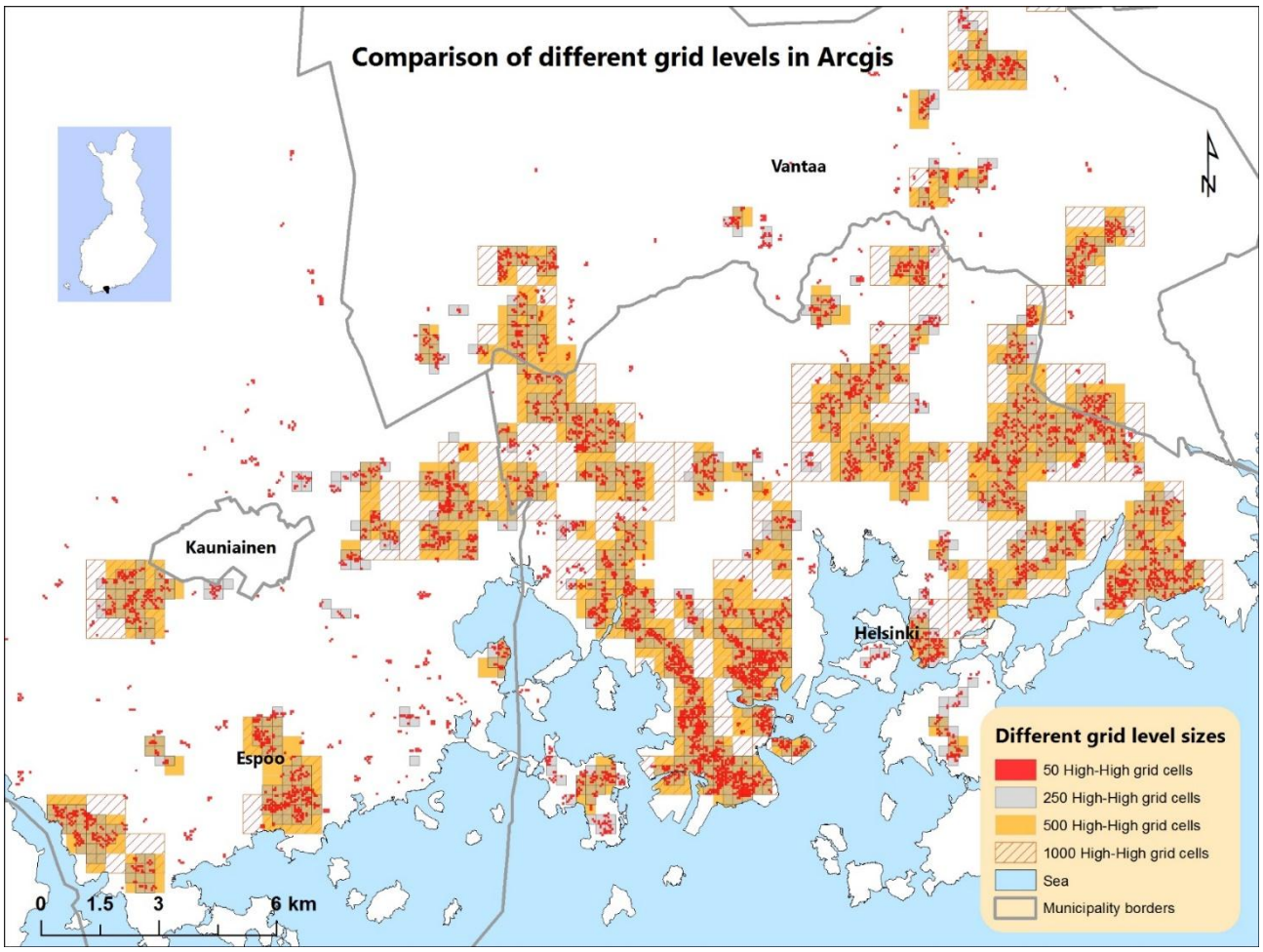


Figure 30. Results of different level lattice for High-High cluster values in ArcGIS (Kekez, 2014).

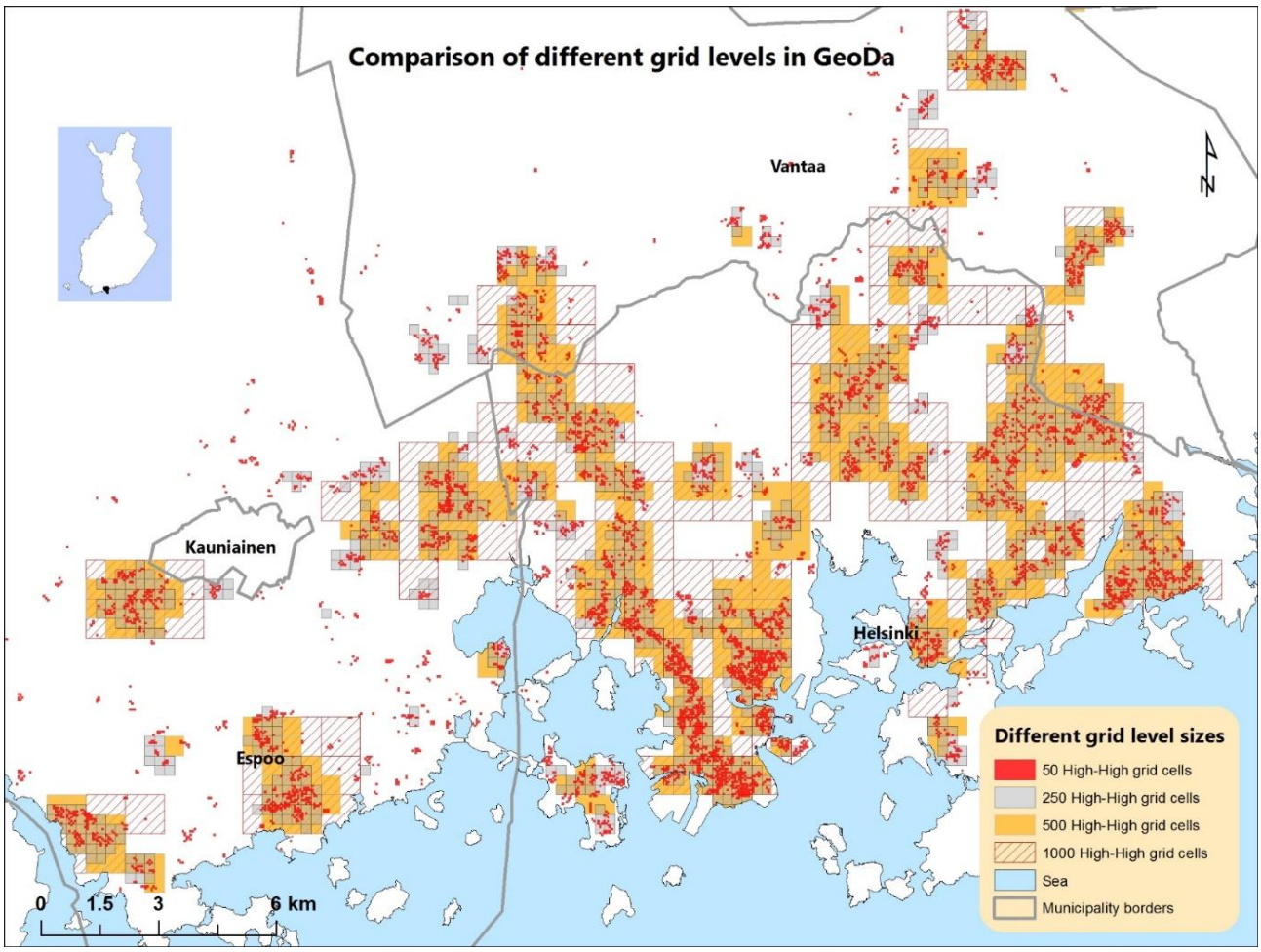


Figure 31. Results of different level lattice for High-High cluster values in GeoDa (Kekez, 2014).

7.4. Influence of scale and MAUP on formation of clusters

It is out of highest importance to measure spatial autocorrelation accurately (O'Kelly, 1994). Imposing certain scale will have direct influence on the magnitude of manifestation of spatial autocorrelation in certain physical space. Availability of tools for measurement of spatial autocorrelation doesn't increase analyzing capabilities. Analyst must use proper statistical indicators and at the same time conscious of the role of the units and scale of analysis effecting final results (Chou, 1991).

As Vaattovaara (2001) is pointing out that *"appearance of spatial development is possible only if the spatial unit of analysis is small enough; the use of GIS in this task is essential"*. Size of the lattice cells (250×250m) of spatial unit used in most of the immigration studies in HMA (Vaattovaara 1998, 2001, 2002; Vilkama 2007, 2011; Vilkama & Dhalmann 2009)

analyzing socio-economic patterns of spatial distributions of immigrants is satisfying for this type of analysis, because of the size of the area being analyzed (most of the time are even bigger than Helsinki Metropolitan Area) and magnitude of the phenomena being analyzed.

Gaining descriptive results (maps, graphs, etc.) from usage different methods of spatial autocorrelation in contemporary GIS environment is not a problem. But for creating and interpreting proper results, understanding of background calculating processes and theory on which it is conceptualized final user has to fully understand statistical operators for different measures of spatial autocorrelation (Haining, 1978). The use of spatial autocorrelation tools can be easily misleading and incorrectly interpreted. Conceptualization of the weight matrix, choice of the proper scale and interpretation of the gained results are steps on creation and interpretation of proper spatial autocorrelation analysis. Getis (1991) and Chou (1991) are pointing out the there is a little to be gained from creating spatial autocorrelation analysis and descriptive statistics collected (maps, graphs, etc.) if analytical and grasping capabilities of the final user are limited.

Question of spatial autocorrelation and occurrence of clustering and clusters can be analyzed with the same grid cell size. That is employing a certain scale of the area we are analyzing. Lattice of 250×250m of represents huge area for analyzing spatial autocorrelation pattern. This area is representing block level size. This work already proved significance and importance of that lattice level, but it is also importing new lattice level (50×50m), because of the specificity of analyzed problem. Problem being analyzed is conceptualized in the contiguous manner following of distribution of the same values (clusters and outliers) with huge effect of influence of neighboring values. New lattice level is discovering unseen patterns in analyzed area showing huge concentrations of clustering of immigrant populations in unexpected locations in HMA. Initiation of smaller grid cell size for lattice with value of 50×50m, which will more accurately represent data provided on building level and investigation and interpretation of the results gained by it is representing a challenge and a new perspective on possible occurrence of spatial clusters.

Modifiable Areal Unit Problem (MAUP)

Modifiable Areal Unit Problem (MAUP) *is "a problem arising from impositions of artificial units of spatial reporting on continuous geographical phenomena resulting in generation of artificial spatial patterns"* (Heywood, 1998). This problem represents observed scientific

investigation of errors created when data are grouped into the units used for analysis. Most forms of spatial analysis are sensitive to both variations in the zoning systems used to collect data and the scale at which data are reported. These effects have been known for some time. Robinson's (1950) demonstration of a positive relationship between the level of aggregation and the magnitude of the correlation between race and illiteracy is a classic. Prior to Robinson's study, Gehlke & Biehl (1934) had noted the tendency for correlation coefficients to increase with the level of aggregation of census tracts. Renewed attention on this basic problem of aggregate spatial analysis was provided by Openshaw & Taylor (1979, 1981) and Openshaw (1984), and more recently by Fotheringham & Wong (1991). But as Goodchild (2011) is pointing out: *"The power of GIS lies in its ability to transform, analyze, and manipulate geographic data, but all transformations, analyses, and manipulations must also be scale-specific".* Represented in Table 6 occurrence of specific mistakes is presented:

Table 6. Size of errors produced by different scale levels (Kekez, 2014).

Represented population in analyzed lattice level of 250×250m

Population	All	Immigrants	Σ	ALL %	Σ Immigrants %	Σ %
pks_vaki 08 (point data)	999679	59875	1059554	100	100	100
small areas	971775	56099	1027874	97.21	93.69	97.01
1000×1000m	974205	56280	1030485	97.45	94.00	97.26
500×500m	975260	56340	1031600	97.56	94.10	97.36
250×250m	998284	57812	1056096	99.86	96.55	99.67
50×50m	753380	47143	800523	75.36	78.74	75.55

7.5. Spatial locations of clusters of immigration population in HMA area

Locations of clusters of immigration population formed in the HMA area are created out of lattice level of 250×250m which can be seen in Figure 32. They are exhibiting the most realistic representation of clustering process, which forms spatial clusters of High-High values in specific locations. Clusters are formed throughout different small areas (pienalue) of HMA.

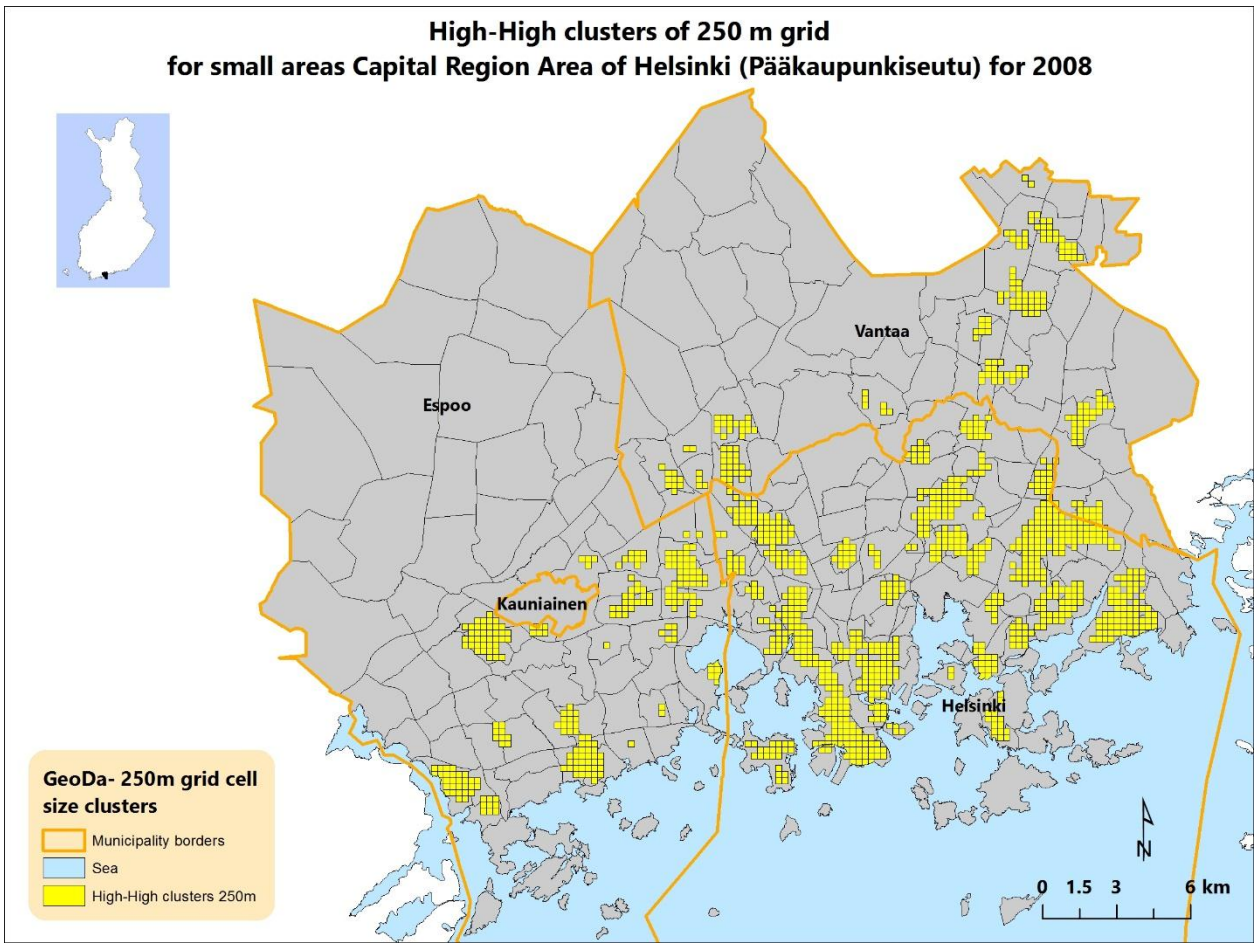


Figure 32. Clusters of High-High values for lattice of 250m for HMA area (Kekez, 2014).

Clusters are forming new spatial units integrating various small areas (pienalue) or their parts into newly formed cluster units. Visually, noticeable is a huge cluster of immigration population in a center and a downtown of Helsinki. Presumably, high concentration of immigration population is always presented and noticed in suburb areas of East Helsinki as well as peripheral units of Vantaa and Espoo (Figures 27 and 28). In comparison to administrative borders of municipalities in HMA area clusters are exhibiting different characteristics. There are two cross-border clusters (catchment areas are formed on the territory of two municipalities), overcoming administrative borders which would present limitation if concept of contiguity would not be applied. Other clusters are located within territories of Helsinki, Espoo and Vantaa.

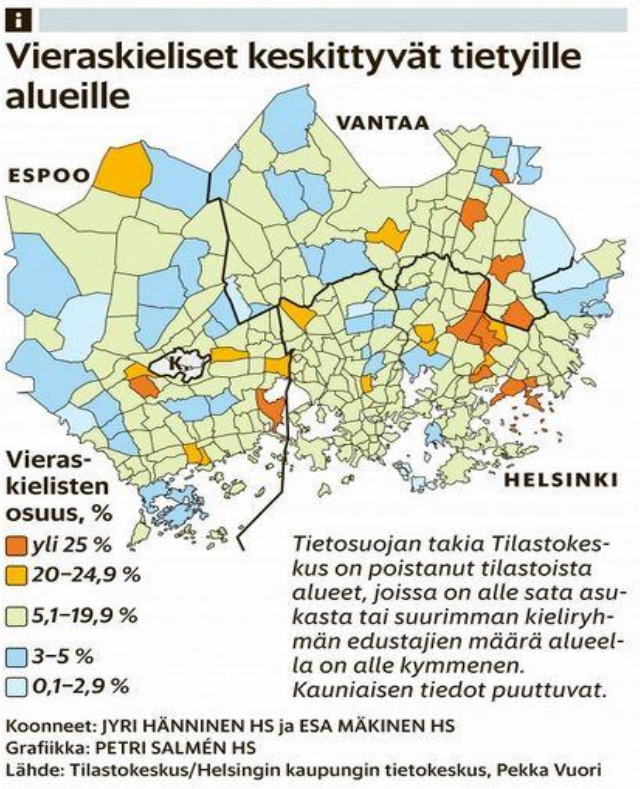


Figure 33. Concentration of population with a foreign language as a maternal language (Helsingin Sanomat, 2014.)

Cross-border clusters of 250×250m lattice

Cross-border clusters formed between small areas (pienalue) of Helsinki and Vantaa are:

Cluster 1 is formed from parts of Helsinki (Kurkimäki, Puotila, Marjaniemi, Jakomäki, Kivikko, Roihuvuori, Itäkeskus, Myllypuro, Kontula, Mellunmäki, Puotinharju and Vartioharju) and a parts of Vantaa (Länsimäki and Rajkylä). It represents the biggest cluster of immigrant population consisted of 8000 individuals, formed in the whole HMA area. Cluster 1 is representing almost one fifth of all of the immigration population living in the High-High value cluster areas. It resembles the most accurate representation of concentration and location of immigration population, with specific spatial catchments of the areas being processed and analyzed in comparison with results in Figures 28 and 33.

Cluster 2 is taking into account catchment areas parts of small areas (pienalue) in Helsinki (Lassila, Malminkartano, Kannelmäki, Pohjois-Haaga) and in Vantaa (Myyrmäki). Results presented in Table 7 are showing significance and magnitude of cross-border clusters.

Table 7. Table of crossborder clusters formed between Helsinki and Vantaa (Kekez, 2014).

Population of 250×250m High-High administrative cross-border clusters (Helsinki and Vantaa)

Cross-border clusters (Helsinki and Vantaa)		Helsinki Metropolitan Area			
			All	Immigrants	High-High 250×250m
	Population	Values	1056096	57812	42719
			(%)	(%)	(%)
Cluster 1	Immigrants	8000	0.76	13.84	18.73
	All	70424	6.67		
Cluster 2	Immigrants	2928	0.28	5.06	6.85
	All	39284	3.72		
Cross-border clusters (part of Helsinki)		Municipality of Helsinki			
			All	Immigrants	High-High 250×250m
	Population	Values	596074	34251	28099
			(%)	(%)	(%)
Cluster 1	Immigrants	6184	1.04	18.05	22.01
	All	62918	10.56		
Cluster 2	Immigrants	1575	0.26	4.60	5.61
	All	22822	3.83		
Cross-border clusters (part of Vantaa)		Municipality of Vantaa			
			All	Immigrants	High-High 250×250m
	Population	Values	251224	10069	6854
			(%)	(%)	(%)
Cluster 1	Immigrants	809	0.32	8.03	11.80
	All	7506	2.99		
Cluster 2	Immigrants	1353	0.54	13.44	19.74
	All	17815	7.09		

Clusters of 250×250m lattice in Helsinki

There is five significant clusters formed within a territory of Helsinki. They are enlisted in Table 8, and according to size of immigrant population they present:

1. ***Cluster 3*** (Eira, Talinranta, Ruoholahti, Kaivopuisto, Etelä-Haaga, Munkkivuori, Vanha Munkkiniemi, Jätkäsaari, Munkkisaari, Ullanlinna, Punavuori, Kamppi, Taka-Töölö, Meilahti, Laakso, Ruskeasuo, Etu-Töölö)
2. ***Cluster 4*** (Keski-Pasila, Itä-Pasila, Länsi-Pasila, Vallila, Alppila, Linjat, Hermanni, Sörnäinen, Siltasaari, Kluuvi, Kruununhaka, Katajanokka)
3. ***Cluster 5*** (Mustavuori, Keski-Vuosaari, Rastila, Meri-Rastila, Aurinkolahti, Kallahti)
4. ***Cluster 6*** (Viikinmäki, Pihlajisto, Pihlajamäki, Viikin Tiedepuisto, Latokartano)
5. ***Cluster 7*** (Pukinmäki, Tapaninvainio, Tapanila, Ylä-Malmi, Puistola, Malmin Lentokenttä, Ala-Malmi)

Table 8. Table of High-High value clusters of immigrant population formed within Helsinki (Kekez, 2014).

Values of 250×250m clusters High-High clusters in Municipality of Helsinki

Clusters in Helsinki			Helsinki Metropolitan Area			Municipality of Helsinki		
			All	Immigrants	High-High 250×250m	All	Immigrants	High-High 250×250m
Nr.	Population	Values	1056096 (%)	57812 (%)	42719 (%)	596074 (%)	34251 (%)	28099 (%)
3	Immigrants	5059	0.48	8.75	11.84	0.85	14.77	18.00
	All	103000	9.75			17.28		
4	Immigrants	3264	0.31	5.11	7.64	0.55	9.53	11.62
	All	63405	6.00			10.64		
5	Immigrants	2953	0.28	2.90	6.91	0.50	8.62	10.51
	All	34004	3.22			5.70		
6	Immigrants	1679	0.16	2.53	3.93	0.28	4.90	5.98
	All	18804	1.78			3.15		
7	Immigrants	1462	0.14	2.53	3.42	0.25	4.27	5.20

Clusters of 250×250m lattice in Espoo

In Espoo there is also five significant clusters formed within a territory of Helsinki. They are enlisted in Table 9, and according to size of immigrant population they present:

1. ***Cluster 8*** (Kaupunginkallio, Kirkkojärvi, Suvela, Tuomarila)
2. ***Cluster 9*** (Friisilä, Tiistilä, Kuitinmäki, Olarinmäki, Matinmetsä, Matinlahti)
3. ***Cluster10*** (Pohjois-Leppävaara, Lintukorpi, Etelä-Leppävaara, Perkkaa, Mäkkyllä, Lintulaakso)
4. ***Cluster 11*** (Kivenlahti, Laurinlahti, Espoonlahden Keskus)
5. ***Cluster 12*** (Nuijala, Kuninkainen)

Table 9. Table of High-High value clusters of immigrant population formed within Espoo (Kekez, 2014).

Values of 250×250m clusters High-High clusters in Municipality of Espoo

Clusters in Espoo			Helsinki Metropolitan Area			Municipality of Espoo		
			All	Immigrants	High-High 250×250m	All	Immigrants	High-High 250×250m
Nr.	Population	Values	1056096	57812	42719	251224	13492	7766
			(%)	(%)	(%)	(%)	(%)	(%)
8	Immigrants	1860	0.18	3.22	4.35	0.74	13.79	23.96
	All	15110	1.43			6.01		
9	Immigrants	1710	0.16	2.96	4.00	0.68	12.67	12.67
	All	23034	2.18			9.17		
10	Immigrants	1150	0.11	1.99	2.69	0.46	8.52	8.52
	All	17689	1.67			7.04		
11	Immigrants	961	0.09	1.66	2.25	0.38	7.12	7.12
	All	14312	1.36			5.70		
12	Immigrants	658	0.06	1.14	1.54	0.26	4.88	4.88
	All	6995	0.66			2.78		

Clusters of 250×250m lattice in Vantaa

In Vantaa there is also five significant clusters formed within a territory of Helsinki. They are enlisted in Table 10, and according to size of immigrant population they present:

1. ***Cluster 13*** (Hakunila, Itä-Hakilla)
2. ***Cluster 14*** (Asola, Rekola, Koivukylä, Havukoski)
3. ***Cluster15*** (Korso, Matari, Metsola,Mikkola)
4. ***Cluster 16*** (Maartinlaakso)
5. ***Cluster 17*** (Viertola, Tikkurila, Jokiniemi, Hiekkaharju)

Table 10. Table of High-High value clusters of immigrant population formed within Vantaa (Kekez, 2014).

Values of 250×250m clusters High-High clusters in Municipality of Vantaa

Clusters in Vantaa			Helsinki Metropolitan Area			Municipality of Vantaa		
			All	Immigrants	High-High 250×250m	All	Immigrants	High-High 250×250m
Nr.	Population	Values	1056096	57812	42719	25122 4	10069	6854
			(%)	(%)	(%)	(%)	(%)	(%)
13	Immigrants	1161	0.11	2.01	2.72	0.46	8.61	14.96
	All	10623	1.01			4.23		
14	Immigrants	1006	0.10	1.74	2.35	0.40	7.46	12.96
	All	16501	1.56			6.57		
15	Immigrants	914	0.09	1.58	2.14	0.36	6.77	11.78
	All	8264	0.78			3.29		
16	Immigrants	636	0.06	1.10	1.49	0.25	4.71	8.19
	All	9273	0.88			3.69		
17	Immigrants	370	0.04	0.64	0.87	0.15	2.74	4.77
	All	5923	0.56			2.36		

Comparison of the results

Results of the formed clusters of 250×250m throughout HMA area are significant representation of concentrations of immigrant population. Expectedly the biggest concentration of immigration population can be fined in East part of Helsinki (Cluster 1), already noticed and point out in studies using descriptive statistical methods (Figures 27 and 33). Inferential statistical methods, concepts of spatial correlation and contiguity with practical appliance of ESDA methods, specifically Global and Local Moran's Index have allowed us to prove that other most important clustering area is a center area of Helsinki.

Cluster 3 and 4 are showing significant concentrations of immigrant population occurring in central area of Helsinki. This results are proving incredible high and previously unnoticed concentrations of immigration population. These results are detecting and resembling processes unnoticed, slightly opposite leading towards formation of different conclusions then presented by official statistical institutions are doing.

Clusters of 50×50m lattice in Helsinki

Results gained from the analysis of 250×250m lattice, specifically in Helsinki led towards employment of final lattice level 50×50m, for further investigation of the formation of clusters.

Restrictions imposed by the use of this lattice level are presented in (Table 4). Amount of population being analyzed in this grid cell level is smaller significantly in comparison with 250×250m lattice level (Table 4). This type of lattice level tends to favor more compactly built unit blocks, which is most of the time case with buildings in downtown and not in suburban areas where they could be more scattered. At the same time this level of units is the closest to the level of actual living units (buildings, houses, etc.) which could potentially represent even more practical example how to interpret correlation among units, discovering spatially-correlational relationship on the basic of unit level.

From previously established patterns we discovered that for specific local measurements of spatial autocorrelation of values exhibiting same patterns 50×50m represents a valid instrument for explanation of concentration of immigration population in Helsinki.

Clusters quantitatively defined in Table 11 are defined by following small areas or parts of them:

1. ***Cluster 1*** (Eira, Munkkisaari, Ullanlinna, Punavuori, Kamppi)
2. ***Cluster 2*** (Torkkelinmäki, Linjat, Harju, Sörnäinen, Siltasaari)
3. ***Cluster 3*** (Meri-Rastila)
4. ***Cluster 4*** (Aurinkolahti, Kallalahti)
5. ***Cluster 5*** (Etu-Töölö, Kamppi, Taka-Töölö)
6. ***Cluster 6*** (Kannelmäki)
7. ***Cluster 7*** (Taka-Töölö, Laakso)
8. ***Cluster 8*** (Kontula)
9. ***Cluster 9*** (Hermanni, Vallila)
10. ***Cluster 10*** (Itäkeskus)

Table 11. Table of 50×50m cluster concentrations of immigrant population in Helsinki (Kekez, 2014)

Quantitative representation of clusters of 50×50m in Helsinki

Number of cluster	CL1	CL2	CL3	CL4	CL5	CL6	CL7	CL8	CL9	CL10
Population	1115	942	695	678	573	470	405	383	321	319

Opposite to previous results on lattice level of 250×250m, results produced in GeoDa on lattice of 50×50m are exhibiting higher concentration of immigrant population in the center of Helsinki. The biggest concentration of clustered immigrant population is living in the core center area of Helsinki (Eira, Munkkisaari, Ullanlinna, Punavuori, Kamppi). Out of ten, five biggest clusters are concentrated in a central area of Helsinki proving that immigration population is living more specifically spatially concentrated in the central area of the town than in suburban areas. Specific clusters and concentrations are also the consequence of physical shape of the neighborhood.

This level of analyses is offering possibilities to explore spatial processes from a different perspective. Even if computing possibilities were pushed to the limit in both software, especially in GeoDa result produced results still don't resemble completely accurate spatial concentration.

8. CONLUSIONS

The results produced in this thesis have intention to explain, prove and introduce capabilities, capacities and possibilities of inferential statistical methods, specifically Exploratory Spatial Data Analysis (ESDA) methods, which are best represented with Global and Local Moran's Index. Computational capabilities of the most well-known commercial GIS software ArcGIS in comparison with standalone, free software like GeoDa have proved to be insignificant. GeoDa is producing more precise, informative and meaningful statistical and visual representation of data. Results could be more operationally manipulated especially with significance levels which are offering deeper insight in analyzed patterns. Visual representation and manipulation of data performed in production of the maps in GeoDa is still on unsatisfying level in comparison with ArcGIS, but statistical production of the results, analytical capabilities and computational power are much better and stronger.

Thesis manage to prove that clustering of immigrant population is not exclusive trend of suburbs, infact it proved that cluster formations are the biggest in central and downtown part of Helsinki. It proved that constant representations of immigration population through the prism of complete neighborhood level percentage are not completely rightfully interpreted. Number of people living in suburbs is smaller than number of people living in more urban areas due to different reasons. Measuring amount of immigrants in areas where majority of people are belonging to native population through the percentage of sum of all immigrant population is not representing quite accurate measure of the processes of creation of clusters of high concentration of immigrant population. Administrative borders, borders of small areas (pienalue) have proved not to have any meaning in this study infect they were creating more problems in conceptualizing of clusters.

Future studies dealing with this problem could improve analytical possibilities by employing more advanced methods like regression analysis which could be performed in GeoDa and R free, programing language or geographically weighted regression analysis which can be performed in ArcGIS and GWR4, free, standalone software. One of the possibilities of future analysis could be more specific analysis of characteristics (education, income, country of origin, etc.) of immigration population living within catchments of clusters formed during the production of this thesis. More specific analysis of this criteria could provide meaningful information about immigration population which could deal with more detailed study of social aspects of that specific population.

ACKNOWLEDGMENTS

First of all I would like to thank my mentor PhD Mika Siljander for all the time, effort and patience he put in helping me producing this work. I could not imagine anybody else helping me this much, being there and saying right things at the right time. Thank You!

Second I would like to thank my family my wife Maarit, my son Vuk, my mother Božana, my father Maksa, my brother Marko and the rest of the Hohteri family for all the years of support, love and for being there when nobody else was. Without all of them I would not be what I am. My family is my world.

I would also like to express thankfulness to PhD student and special consultant Athanasios Votsis, as a mentor, teacher and friend who helped to understand deeper and more meaningful side of spatial statistical analysis.

I would also like to thank PhD Gareth Rice for unconditional support and friendship in times of despair. Conversations helped me to look at the writing problems from a different perspective.

I reserve special thanks to PhD Katja Vilkama for inspiring work on the topic of immigration and couple of important conversations which helped me to conceptualize my topic of analysis and PhD Rami Ratvio who was helping me to understand some concepts in Finnish language as well to be ready to discuss themes and topics considering this work with open mind.

I would also like to thank adjunct Prof. Tuuli Toivonen for all the support during my studies. Without Tuuli's understanding and extreme amount of help in creating and guiding my personal study plan and studies in general I wouldn't be where I am today.

Special thanks to to Prof. Petri Pellikka for creating GIMP programme.

I would also like to thank the last generation of GIMP studies my dear friends Oula Seitsonen and Heikki Vesanto. It was my privilege to be there together with you two in the same study programme. There is one more fellow student, which deserves special thanks, my dear friend Katri Tegel. Thank you for keeping us all as a team and a unit!

Also, special thanks to prof. Maari Vaattovaara for previous work on the topic of immigration in Helsinki Metropolitan Area.

At the end I would also like to thank and Prof. Tommi Inkinen for examining my thesis.

REFERENCES

Anselin, L. (1989). What is special about spatial data?: alternative perspectives on spatial data analysis. *Symposium on Spatial Statistics, Past, Present and Future* National Center for Geographic Information and Analysis, University of California, Santa Barbara, USA, 63–77.

Anselin, L. (1992). Spatial data analysis with GIS: an introduction to application in the social sciences. *Technical report 92–10* National Center for Geographic Information and Analysis, University of California, Santa Barbara, USA, 1–54.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, Volume* 27, *Issue* 2, Blackwell Publishing Ltd., 93–115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS*, edited by Fischer, M. et al., Taylor & Francis, London, 111–125.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P. Longley et al., Geoinformation Int., Cambridge, 253–266.

Anselin, L. (2008). Geoda: Training, E-Slides, Spatial Autocorrelation (Background), GeoDa Center for Geospatial Analysis and Computation, Arizona State University, USA. 2.28.2008.
<https://geodacenter.asu.edu/spatial-autocor-1>

Anselin, L. & S. Bao (1997). Exploratory Spatial Data Analysis: Linking SpaceStat and ArcView. In *Recent Developments in Spatial Analysis, Chapter* 3, edited by Fischer, M. & Getis, A., Springer-Verlag, Berlin and New York, 35–59.

Anselin, L. & A. Getis (1992). Spatial Statistical Analysis and Geographical Information Systems. *The Annals of Regional Sciences* 26 (1), Springer-Verlag, Berlin, 19–33.

Anselin, L., I. Syabri & Y. Kho (2006). GeoDa: An Introduction to Spatial Data Analysis.*Geographical Analysis* 38, *issue* 1, Blackwell Publishing Ltd., 5–22.

Anselin, L. & S. Rey (2010). PySAL: A Python Library of Spatial Analytical Methods, In *Handbook of Applied Spatial Analysis*, edited by Fischer, M. and Getis, A., Springer-Verlag,Berlin, 175–193.

Bivand, R. (2009). Applying Measures of Spatial Autocorrelation: Computation and Simulation. *Geographical Analysis*, *Volume* 41 , *Issue* 4,375–384.

Borrie, W. D. (1970). The Growth and Control of World Population. Weidenfeld and Nicolson, London.

Castles, S. (2011). Migration, Crisis, and the Global Labour Market. *Globalizations*, *Volume* 8, *Number* 3, Taylor and Francis Group, Routledge, 311–324.

Chapain, C., K. Stachoviak & M. Vaattovaara (2010). Beyond Cluster Policy? Birmingham, Poznan and Helsinki. In *Making Competitive Cities* edited by S. Musterd and A. Murie. Willey-Blackwell. 263–284

Chou, Y. H. (1991). Map resolution and spatial autocorrelation. *Geographical Analysis*, *Volume* 21, *Issue* 3, Blackwell Publishing Ltd., 228–246.

City of Helsinki, Urban facts 2013, Statistics 40, 2013, Foreigners in Helsinki (2013)

<http://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/13_12_18_Tilastoja_40_Selander.pdf >

Cliff A.D. & J.K. Ord (1969). The problem of spatial autocorrelation. In *Studies in Regional Science,* edited by A.J.Scott, Pion Press, London, 25–55.

Cliff A.D. & J.K. Ord (1970). Spatial Autocorrelation: A review of existing and new measures with applications. *Economic Geography*, *Volume* 46, International Geographical Union. Commission on Quantitative Methods, Clark University, 269–292.

Cliff A.D. & J.K. Ord (2009). What Where We Thinking?. *Geographical Analysis*, *Volume* 41, *Issue* 4, Blackwell Publishing Ltd., 351–363.

Dhalmann, H. & S. Yousfi, (2010). Immigration flows, policies and practices in Finland. In *Immigration, housing and segregation in Nordic welfare states*, *Chapter 3: Immigration flows, policies and practices in Finland*, University of Helsinki, Faculty of Science, Department of Geosciences and Geography, University Print, Helsinki, 222–232.

Diffen, (2014). Comparison of the terms, Emigrate versus Immigrate

<http://www.diffen.com/difference/Emigrate_vs_Immigrate>

ESRI (2011). ArcGIS Desktop Help: How cost distance tools work. Environmental Systems Research Institute, Redlands, California. 1.12.(2013. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z00000025000000.htm>.

ESRI (2013a). ArcGIS Support: GIS dictionary, Spatial Statistics. Environmental Systems Research Institute, Redlands, California.

<http://support.esri.com/en/knowledgebase/GISDictionary/term/spatial%20statistics>

ESRI, (2013b). ArcGIS Support: GIS dictionary, Spatial Autocorrelation. Environmental Systems Research Institute, Redlands, California.

<http://support.esri.com/en/knowledgebase/GISDictionary/search>

ESRI, (2013c). ArcGIS Desktop Help: Spatial Statistics toolbox, What is a z-score? What is a p-value? Environmental Systems Research Institute, Redlands, California. 4.18.2013.

<http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_a_z_score_What_is_a_p_value/005p00000006000000/>

ESRI, (2013d). ArcGIS Desktop Help: Spatial Statistics toolbox, Annalyzing Patterns toolset, How Spatial Autocorrelation (Global Moran's I) works. Environmental Systems Research Institute, Redlands, California. 4.18.2013.

<http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Spatial_Autocorrelation_Global_Moran_s_I_works/005p0000000t000000/>

ESRI, (2013e). ArcGIS Desktop Help: Spatial Statistics toolbox, Modeling Spatial Relationships toolset, Spatial weights. Environmental Systems Research Institute, Redlands, California. 4.18.2013.

<http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000035000000>

ESRI, (2013f). ArcGIS Desktop Help: Spatial Statistics toolbox, Mapping Cluster toolset, How Cluster and Outlier Analysis (Anselin Local Moran's I) works. Environmental Systems Research Institute, Redlands, California. 4.18.2013.

<http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Cluster_and_Outlier_Analysis_Anselin_Local_Moran_s_I_works/005p00000012000000/>

Fawcett, J.T. (1989). Networks, linkages, and migration systems. *International Migration Review, Volume* 23, *Number* 3, Special Silver Anniversary Issue: International migration, an assessment for the 90's, The Center for Migration Studies New York Inc., New York. 671–680

Forsander, A. (2001). Immigrants in the Finnish Labour Market – Is There Ethnic Segmentation?. In Muuttoliikkeet vuosituhannen vaihtuessa –halutaanko niitä ohjata?, Muuttoliikesymposium, edited by E. Heikkilä, 250–266.

Forsander, A. (2003). Insiders of Outsiders Within? Immigrants in the Finnish Labor Market. *Yearbook of Population Research in Finland* 39, 55–72.

Fotheringham, A.S. (1998). Trends in quantitative methods II: stressing the computational, *Progress in Human Geography, Volume* 22, Sagepub, 283–292.

Fotheringham, A.S. (2009). "The Problem of Spatial Autocorrelation" and Local Spatial Statistics. *Geographical Analysis, Volume* 41, *Issue* 4, Blackwell Publishing Ltd, 398–403.

Fotheringham, A.S. & C. Brunsdon (1999). Local forms of spatial analysis. *Geographical Analysis, Volume* 31, *Issue* 4, Blackwell Publishing Ltd, 340–358.

Fotheringham, A.S., C. Brunsdon & M. Charlton (2000). Quantitative Geography: perspectives on spatial data. Sage, 1–267

Fotheringham, A.S. & D.W.S. Wong (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, Volume* 23, *Issue* 7, Pion Publication, Great Britain, 1025–1044.

Frank, A. I. (2003). Using measures of spatial autocorrelation to describe socio-economic and racial residential patterns in US urban areas. In *Socio Economic Applications in Geographical Information Science, Chapter* 11, edited by Kidner D., Higgs G. and White S., Taylor & Francis, London, 146–161.

Geary, R. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician, Volume* 5, JSTOR, 115–45.

Gehlke, C.H. & K. Biehl (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association, Volume* 29, 169–170.

GeoDa: Glossary Key of Terms, Weight Matrix
<https://geodacenter.asu.edu/node/390#w>

Getis, A. (1991). Spatial interaction and spatial autocorrelation: a cross-product approach. *Environment and Planing A, Volume* 23, *Issue* 9, Pion Publication, Great Britain, 1269–1277.

Getis, A. & J.K. Ord (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, Volume* 34, *Issue* 3, Blackwell Publishing Ltd, 189–206.

Goodchild, M.F. (1987). A spatial analytic perspective on geographical information systems. *International Journal of Geographical Information Systems*, *Volume* 1, *Issue* 4, Taylor & Francis, 327–334.

Goodchild, M.F. (2011). Scale in GIS: An overview. *Geomorphology*, *Volume* 130, *Issues* 1–2, Elsevier, 5–9.

Goodchild, M.F., R.P. Haining, S. Wise and 12 others (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems, Volume* 6, *Issue* 5, Taylor & Francis, 407–423.

Graham, E. (1997). Philosophies Underlying Human Geography Research. In *Methods in Human Geography: A Guide for Students doing a Research Project*, edited by R. Flowerdew & D. Martin, Harlow: Longman, 6-30.

Griffith, D. A. (2009). Modeling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems, Volume* 11, *Issue* 2, Springer, 117–140.

Gulijeva, A. (2003). Ingrian immigration to Turku after 1990 – Case study in Turku. University of Turku, Department of Geography/ Baltic Sea Region Studies. Thesis, 1–72.

Haining, R. P. (1978). Specification and estimation problems in models of spatial dependence. Department of Geography, Northwestern University, Evanston, Illinois.

Haining, R. P. (2003). Spatial Data Analysis, Theory and Practice. Cambridge University Press, Cambridge, UK.

Haining, R. P. (2009). Spatial Autocorrelation and the Quantitative Revolution. *Geographical Analysis, Volume* 41, *Issue* 4, Blackwell Publishing Ltd, 364–374.

Hannikainen, L. (1996). The status of Minorities, Indigenous Peoples and Immigrant and Refugee Groups in Four Nordic States. in *Nordic Journal of International Law, Volume* 65, 1-71.

Heikkilä, E. & T. Järvinen (2003). Migration and Employment of Immigrants in the Finnish Local Labor Markets. *Yearbook of Population Research in Finland* 39, 103–118.

Heikkilä, E. & S. Peltonen (2002). Immigrants and integration in Finland. Institute of Migration, Turku, 1–10.

Heywood, D., S. Cornelius & S. Carver (1998). An introduction to geographical information systems, Addison Wesley Longman, Harlow, UK.

Higazi, S. F., D.H. Abdel-Hady & S.A. Al-Oulfi (2013). Application of Spatial Regression Models to Income Powerty Ratios in Middle Delta Contigous Counties in Egypt. *Pakistan Journal of Statistics and Operation Research, Volume* 9, *Number* 1, 93–110.

Helsingin Sanomat, (2014)

<http://www.hs.fi/kaupunki/a1414298338550>

Helsingin Seudun Ympäristöpalvelut , Regional and environmental information (2014)

<http://www.hsy.fi/en/regionalinfo/urban/gis/Pages/default.aspx>

Inkinen, T. & M. Vaattovaara (2007). Technology and knowledge-based development. Helsinki metropolitan area as a creative region. Pathways to creative and knowledge-based regions. ACRE report WP2.5., 1–77.

Jasinskaja-Lahti, I. (2000). Psychological Acculturation and Adaptation Among Russian-speaking Immigrant Adolescents in Finland. *Social psychological studies,* Department of Social Psychology, University of Helsinki, 1–72.

Kepsu K., M. Vaattovaara, V. Bernelius & E. Eskelä (2009). Helsinki: An attractive metropolitan region for creative knowledge workers? The view of transnational migrants. ACRE report WP7.5., 1–144.

Koivukangas O. (2003). European Immigration and Integration: Finland. Paper presented at *Conference: The Challenges of Immigration and Integration in the European Union and Australia*, University of Sydney, 18–20 February 2003.

Kokko, K. (2002). Maahanmuuttajien Suomen sisäinen muuttoliike, Tapaustutkimuksena Turku, Turun yliopisto, Maantieteen laitos, Pro gradu–tutkielma, 1–92.

Kuuma (2013). Helsinki Metropolitan Area and surrounding municipalities (map)

<http://makery.fi/en/solutions-business/development-projects/>

Lehti, M. & K. Aromaa (2002). Trafficking in Human Beings, Illegal Immigration and Finland. European Institute for Crime Prevention and Control, 38, 1–77

Lehtonen, O. & Tykkyläinen, M. (2010). Self-reinforcing spatial clusters of migration and socio-economic conditions in Finland in (1998-2006), *Journal of Rural Studies, Volume* 26, *Issue* 4, Elsevier, 361–373.

Leitner, M. & H. Brecht (2007). Software Review: Crime Analysis and Mapping with GeoDa 0.9.5-I. *Social Science Computer Review*, *Volume* 25, *Number* 2, Sage Publications, 265–271.

Łobodzińska, A. (2011). Immigrants and immigration policy in ageing Finland. *Journal of Nicolaus Copernicus, University Torun, Bulletin of Geography, Socio–economic Series*, *Volume* 15, *Issue* 15, De Gryuter, 43–55.

Messner S., Anselin L., Baller R., Hawkins D., Deane G and Tolnay S. (1999). The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis. *Journal of Quantitative Criminology*, Volume 15, *Number* 4, Springer, 423–450.

Milanovic, B. (2007). Global income inequality: what it is and why it matters?. Flat World, Big Gaps: Economic Liberalization, Globalization, Poverty & Inequality, edited by K.S Jomo & J. Baudot, ZED Books Ltd., New York, 1–23.

Milligan, G. & Cooper, M. (1987). Methodology Review: Clustering Methods, Applied Psychological Measurement. *Applied Psychological Measurement, Volume* 11, *Number* 4, Sage publications, 329–354.

Monasterio, L. M. (2006). Wages and Industrial Clusters in Rio Grande do Sul (Brazil). *The Review of Regional Studies*, *Volume* 36, *Number* 3, 304 – 323.

Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of Royal Statistical Society Series B* 10, 243–251.

Moran, P.A.P. (1950). A test for the serial independence of residuals. *Biometrika, Volume* 37, Number 1–2, 178–181.

Musterd, S., R. Andersson, G. Galster & T.M. Kauppinen (2008). Are immigrants' earnings influenced by the characteristics of their neighbours? *Environment and Planning A, Volume* 40, *Issue* 4, Pion Publication, Great Britain, 785–805.

Norden, (2014).

<http://www.norden.org/en/the-nordic-region/population>

OECD, (2007). International Migration Outlook: Annual Report 2007. Organization for Economic Cooperation and Development, Paris, 63–67.

Oliveau S. & Guilmoto, C.Z., (2005). Spatial autocorrelation and demography. Exploring India's demographic patterns. *communication au Congrès International de la Population*, Tours, juillet.

Openshaw, S. (1983). The Modifiable Areal Unit problem. *The concepts and Techniques in Modern Geography*, *Number* 38, Geo Books, 1–39.

Openshaw, S. & P. J. Taylor (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley, *ed. Statistical applications in the spatial sciences, Volume* 21, Pion Publication, Great Britain, 127–144.

Oxford Dictionaries, (2014).

<http://www.oxforddictionaries.com/definition/english/immigration>

O'Kelly, M.E. (1994). Spatial analysis and GIS, In *Spatial analysis and GIS*, edited by Fotheringham, S. & P. Rogerson, Taylor & Francis, London, 62–79.

Peters G. & R. Larkin (1999). Migration and Mobility Population, In *Geography: Problem, Concepts and Prospects*, *Chapter* 8, Kendall/Hunt, USA, 193–213.

Phinney, J.S., G. Horenczyk, K.Liebkind & P.Vedder (2001). Ethnic Identity, Immigration, and Well-Being: An Interactional Perspective. *Journal of Social Issues, Volume* 57, *Number* 3, Willey-Blackwell, 493–510.

Population Register Center of Finland, (2014). 31.1.2014.

<http://vrk.fi/default.aspx?docid=7809&site=3&id=0>

Register data: Source SeutuCD'08 (2008). Espoon, Helsingin, Kauniaisten, Vantaan, Kirkkonummen, Keravan, Nurmijärven, Järvenpään ja Tuusulan mittausosastot sekä HSY, Helsingin Seudun Ympäristöpalvelut – kuntayhtymä.

Rice, S. (2003). Sampling in Geography. In *Key methods in Geography*, *Chapter* 15, edited by Clifford, N.J. & G. Valentine, Sage publications, UK, 223–248.

Robinson, W.S. (1950). Ecological correlations and the Behavior of Individuals. *American Sociological Review, Volume* 15, *Number* 3, American Sociological Association, JSTOR, 351–357.

Salt, J., Clarke J. & Smith S. (2000). Patterns and Trends in International Migration in Western Europe. Eurostat, Luxembourg.

Samers, M. (2010). Migration. Routledge, UK.

Schabenberger, O. & C. A. Gotway, (2005). Statistical Methods for Spatial Data Analysis. Chapmen and Hall, USA.

Scott, L. & Getis, A. (2008). Spatial statistics. In *Encyclopedia of geographic information science*, edited by K. Kemp, Sage Publications, USA, 439.

Scott, L. & M. Janikas (2010). Spatial statistics in ArcGIS. In *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, *Part* A.1, edited by M.M. Fischer & A. Getis, Springer-Verlag, 27–42.

Sokal, R., G. Jacquez & M. Wooten (1988). Spatial Autocorrelation Analysis of Migration and Selection. *Genetics, Volume* 121, *Number* 4, Genetics Society of America, 845–855.

Stalker, P. (2003). Migration Trends and Migration Policy in Europe. *International Migration Volume* 40, *Issue* 5, Blackwell Publishers Ltd., 151–179.

Stalker, P. (1994). The Work of Strangers: A Survey of International Labour Migration. International Labour Organization, Geneva.

Statistics Finland, (2013). 22.3.2013.

<http://www.stat.fi/til/vaerak/2012/vaerak_2012_2013-03-22_tie_001_en.html>

Statistics Finland, (2014). 18.12.2014.

<http://www.stat.fi/til/vamuu/2014/11/vamuu_2014_11_2014-12-18_tie_001_en.html>

Söderling, I. (2010). Factors affecting population size in Finland – the role of immigration and population policies. Institute of Migration, Turku, 1–14.

Taylor, P. J. (1977). Quantitative methods in Geography – An introduction to Spatial analysis, Probability Theory and Geographical Research Inferences. Houghton Mifflin Company, Boston.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, *Volume* 46, *Issue* 2, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods, 234–240.

Tukey, J. W. (1977). Exploratory Data Analysis, Addison-Wesley, Reading.

Urban Audit report (2006). Eurostat.

<http://circa.europa.eu/Public/irc/dsis/urbstat/library?l=/urban_audit_reports/urban_audit_(2006/final_reportpdf_18/_EN_1.0_&a=d>

Vaattovaara, M. (1998). Pääkaupunkiseudun sosiaalinen erilaistuminen (Residential differentiation within the metropolitan area of Helsinki, Finland – environment and spatiality), City of Helsinki, Urban Facts, Research Series, 1–178.

Vaattovaara, M. (2001). Residential differentiation studies by GIS, Statistical Commission and Economic Commission for Europe, invited Paper for Conference of European Statisticians, joint UNECE/Eurostat Work session on methodological issues involving integration of Statistics and Geography, 1–10.

Vaattovaara, M. (2002). Future Developments of Residential Differentiation in the Helsinki Metropolitan Area: Are We Following the European Model? *Yearbook of Population Research in Finland* 38, 107-123.

Vasanen, A. (2009). Deconcentration versus spatial clustering: changing population distribution in the Turku urban region, 1980–2005. *Fennia, Volume* 187, *Number* 2, Helsinki, 115–127.

Vilkama, K. (2011), Yhteinen kaupunki, eriytyvät kaupunginosat? Kantaväestön ja maahanmuuttajataustaisten asukkaiden alueellinen eriytyminen ja muuttoliike pääkaupunkiseudulla, Akateeminen väitöskirja, Tutkimuksia; Helsingin kaupungin tietokeskus.

Vilkama, K. & Dhalmann, H. (2009). Housing policy and the ethnic mix in Helsinki, Finland: perceptions of city officials and Somali immigrants. *Journal of Housing and the Built Environment,Volume* 24, *Issue* 4, 423–439.

Whittle, P. (1954), On stationary processes in the plane, Biometrika, 41, (pp. 431–449)