HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master's thesis in Geography
Geoinformatics

Modelling risk factors in urban residential fires in Helsinki

Sara Todorović

2020

Supervisors:

Petteri Muukkonen
Hanna Rekola
Venla Bernelius

Master's Programme in Geography
Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Faculty of Science	Osasto – Institution – Department Department of Geosciences and Geography

Tekijä – Författare – Author

Sara Katariina Todorović

Tutkielman otsikko – Avhandlings titel – Title of thesis

Modelling risk factors in urban residential fires in Helsinki

Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track

Master's programme in Geography, Geoinformatics

Tutkielman taso – Avhandlings nivå – Level of the thesis Master's thesis	Aika – Datum – Date May 2020	Sivumäärä – Sidoantal – Number of pages 74 pp. + 1

Tiivistelmä – Referat – Abstract

Fires in residential buildings can lead to significant personal injury and property damage, especially in cities. Fire incidence has been found to have a strong connection with the characteristics of neighbourhoods and their inhabitants, such as with socioeconomic status and the features of households and buildings. However, the influencing factors are complex and often interconnected, which has made it difficult to make accurate predictions. Risk modelling and spatial data analysis provide effective and practical means of studying the phenomenon, especially from the point of view of accident prevention and preparedness. To date, knowledge of the spatial risk factors affecting residential fire incidence is yet limited in Helsinki. Thus, this study has sought to bring new empirical evidence on the matter.

This study analysed residential fires in Helsinki from 2014 to 2018 at a 250 x 250 m grid level. The spatial dependence of fires was investigated by observing statistically significant clusters of fires. In this study, a risk model was created that sought to identify the underlying structural, socioeconomic, and household characteristics of neighbourhoods that affect the likelihood of residential fire incidence. The methods used were linear regression and the Geographically Weighted Regression (GWR), which takes spatial heterogeneity into account.

The results showed that residential fires are spatially clustered in Helsinki. A significant large concentration of fires was found in the inner-city area and smaller concentrations in eastern Helsinki. The results indicate that the structural features of the neighbourhoods, socioeconomic status, and household circumstances have an impact on the likelihood of residential fire incidence by both increasing and decreasing the risk of fire. At the neighbourhood level, statistically significant explanatory variables that increased fire risk were population density, low education, unemployment, occupancy rate of dwellings, and home ownership. A negative relationship with fire risk was found with residential building density, age of the buildings, high education, as well as home ownership. Overall, in the study area, these eight variables explained about half of the variance of residential fire incidence. In a comparison between the models, the explanatory power of the GWR was better than linear regression, and it was also able to identify significant local variations in the effects of explanatory variables on fire risk.

A comprehensive understanding of the factors influencing residential fire risk at local levels is important for rescue services, especially in terms of planning response readiness and efficient allocation of resources. In the future, more precise models should be developed in order to achieve a more comprehensive understanding of fire risk and the factors affecting it. Particular attention should be paid to the use of more precise and diverse data and methods in modelling, as well as to the temporal dimension and the consequences of fires.

Avainsanat – Nyckelord – Keywords

Fire risk, residential fire, Geographically Weighted Regression, GIS

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Luonnontieteellinen tiedekunta		Osasto – Institution – Department Geotieteiden ja maantieteen osasto
Tekijä – Författare – Author Sara Katariina Todorović		
Tutkielman otsikko – Avhandlings titel – Title of thesis Urbaanien asuinrakennuspalojen riskitekijöiden mallintaminen Helsingissä		
Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track Maantieteen koulutusohjelma, Geoinformatiikka		
Tutkielman taso – Avhandlings nivå – Level of the thesis Pro gradu	Aika – Datum – Date Toukokuu 2020	Sivumäärä – Sidoantal – Number of pages 74 s. + 1

Tiivistelmä – Referat – Abstract

Asuinrakennuksissa syttyvät tulipalot aiheuttavat merkittäviä henkilö- ja omaisuusvahinkoja erityisesti kaupungeissa. Palojen esiintyvyydellä on todettu olevan voimakas yhteys alueiden ja alueiden asukkaiden piirteisiin, kuten sosioekonomiseen asemaan sekä kotitalouksien ja rakennusten ominaisuuksiin. Vaikuttavat tekijät ovat kuitenkin monimutkaisia ja usein toisiinsa kytkeytyneitä, mikä on vaikeuttanut tarkkojen ennusteiden tekemistä. Riskimallinnus ja paikkatietoanalyysit tarjoavat entistä tehokkaampia ja käytännöllisiä keinoja ilmiön tutkimiseen, erityisesti onnettomuuksien ennaltaehkäisyn ja varautumisen näkökulmasta. Tähän mennessä asuinrakennuspalojen alueelliseen esiintyvyyteen vaikuttavien riskitekijöiden tuntemus Helsingissä on ollut rajallista, mihin tällä tutkielmalla on pyritty tuomaan uutta empiiristä tietoa.

Tässä tutkielmassa analysoitiin Helsingissä syttyneitä asuinrakennuspaloja vuosina 2014–2018 250 x 250 metrin ruututasolla. Tulipalojen alueellista riippuvuutta tutkittiin havainnoimalla tilastollisesti merkittäviä palojen keskittymiä. Lisäksi luotiin riskimalli, jolla pyrittiin tunnistamaan tulipalojen alueelliseen esiintyvyyteen vaikuttavia naapurustojen rakenteellisia, sosioekonomisia ja väestöllisiä piirteitä. Menetelminä käytettiin lineaarista regressiota ja spatiaalisen heterogeenisyyden huomioivaa Geographically Weighted Regression (GWR) -menetelmää.

Tulokset osoittivat, että asuinrakennuspalot ovat alueellisesti klusteroituneita Helsingissä. Merkittävä suuri keskittymä löytyi kantakaupungin alueelta ja pienempiä keskittymiä Itä-Helsingistä. Tulosten perusteella naapuruston rakenteellisilla piirteillä, sosioekonomisella asemalla ja kotitalouksien ominaisuuksilla on vaikutusta asuinrakennuspalojen esiintyvyyden todennäköisyyteen sekä paloriskiä lisäävinä että vähentävinä tekijöinä. Naapurustotasolla tilastollisesti merkittäviä paloriskiä lisääviä selittäviä muuttujia olivat väestöntiheys, alhainen koulutustaso, työttömyys, asumisväljyys sekä omistusasuminen. Negatiivisesti paloriskiin vaikuttavia tekijöitä olivat asuinrakennusten tiheys, alueen rakennuskannan ikä, korkea koulutustaso sekä myös omistusasuminen. Yleisesti tutkimusalueella tämä kahdeksan muuttujaa selittivät noin puolet asuinrakennuspalojen vaihtelusta. Mallien välisessä vertailussa GWR:n selitysaste oli lineaarista regressiota parempi, ja se myös pystyi tunnistamaan merkittäviä paikallisia eroja selittävien muuttujien vaikutuksissa paloriskiin.

Asuinrakennuspalojen riskiin vaikuttavien tekijöiden kokonaisvaltainen ymmärtäminen aluetasolla on tärkeää pelastustoimelle erityisesti valmiuden mitoittamisen ja resurssien tehokkaamman kohdentamisen kannalta. Jatkossa tulisikin kehittää tarkempia malleja, jotta saavutettaisiin entistä kattavampi kokonaiskuva paloriskistä ja siihen vaikuttavista tekijöistä. Erityisesti huomiota tulee kiinnittää tarkemman ja monipuolisemman aineiston ja menetelmien hyödyntämiseen, sekä myös tulipalojen ajallisen ulottuvuuden ja palojen seurauksien sisällyttämiseen mallinnuksessa.

Avainsanat – Nyckelord – Keywords

Paloriski, asuinrakennuspalo, Geographically Weighted Regression, GIS

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

Table of contents

Abbreviations

AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
GIS	Geographic Information System
GWR	Geographically Weighted Regression
KDE	Kernel Density Estimation
LISA	Local Indicators of Spatial Association
OLS	Ordinary Least Squares
OSF	Official Statistics of Finland
PRONTO	Accident and resource statistics of the Finnish rescue services
UTM	Universal Transverse Mercator
WHO	World Health Organization

1 Introduction

Fires are both natural and social phenomena that cause extensive harm to societies in terms of human lives, economic losses, and operational costs (Corcoran et al. 2011b; Corcoran & Higgs 2013; Jennings 2013; Špatenková & Virrantaus 2013). Fires also affect communities, their livelihoods and productivity, and can create serious damage to urban infrastructure (Jennings 2013; KC & Corcoran 2017). Of all types of fire, residential fires pose the greatest risk to human lives and the surrounding environment because of their higher likelihood to lead to fatal consequences (Ceyhan et al. 2013).

Fires are also strongly spatial phenomena, and studies have shown that residential fires exhibit spatial patterns especially in the most deprived neighbourhoods (Wuschke et al. 2013; Guldåker et al. 2018). Apart from showing spatial patterns in the city, residential fires are driven by multiple underlying spatial factors, including complicated direct and indirect relationships between individuals' and neighbourhood characteristics (Corcoran et al. 2011b). Low socioeconomic status, unstable family circumstances, and poor housing conditions have been found to increase not only the risk of fire occurrence, but also the risk of dying in a fire (Gunther 1981; FEMA 1997; Duncanson et al. 2002; Jennings 2013).

The complexity of people's behaviour at an individual and collective level in cities has made fire risk extremely complicated to model and theorize (Corcoran et al. 2011b; Jennings 2013; Špatenková & Virrantaus 2013). While the number of studies has been increasing in recent years, the current knowledge about the spatial aspects of fire risk is still limited to a few studies mostly from developed countries, such as the UK, Australia, Canada, Sweden and Finland (e.g. Corcoran et al. 2007; Chhetri et al. 2010; Asgary et al. 2010; Corcoran et al. 2011b; Špatenková & Virrantaus 2013; Wuschke et al. 2013; Guldåker et al. 2018; Ardianto & Chhetri 2019).

The methods in fire risk studies have varied greatly from exploratory visual methods to multiple regression models, and more recently, spatial statistics, but the follow-up has been very limited. Thus, there is a demand for new and more profound case studies to help better understand the residential fire risk and its spatial drivers (Corcoran et al. 2007; Jennings 2013; Clark et al. 2015).

In Finland, rescue departments are obliged by law to prevent and respond to fires (Rescue Act 379/2011). Nation-wide strategies have set goals for the rescue services to move towards more comprehensive and continuous risk analysis and mitigation practices (Ministry of Interior 2016; 2019). In order to achieve these goals, the underlying causes of residential fires should be better researched and comprehended. As fires are a phenomenon with life-threatening consequences, finding these local disparities is important for maintaining the health and sustainability of the society. By identifying areas where the impact of risk factors is significant, resources of the rescue services can be allocated and targeted more efficiently. The resulting risk maps are highly practical and effective visual products that can be directly used in planning and decision making (Krisp et al. 2005).

While risk factors associated with higher risk of dying in a residential fire have previously been analysed in Finland (e.g. Kokki & Jäntti 2009; Kokki 2014; Östman 2015), there is only limited understanding of the types of factors affecting the spatial patterns and variability of residential fire incidence regardless of the outcomes, especially in the city of Helsinki. Currently the knowledge of residential fires in Finland is based on a few empirical studies, along with GIS analyses and descriptive statistics conducted by the local rescue departments. There is a need for updated information, as the latest empirical studies from Helsinki were carried out with data from 2005–2008 (Tillander et al. 2010; Špatenková & Stein 2010; Špatenková & Virrantaus 2013). It is also unclear, whether the current internationally relatively low level of residential segregation, high standard of welfare services, and the generally high quality of housing relates to the internationally observed connections between socio-spatial deprivation and fire risk.

This study continues the recent trend of analysing residential fires from a spatial perspective by applying spatial analytical methods to detect potential risk factors and local variability of residential fire risk in a city environment (Špatenková & Virrantaus 2013; Ardianto 2018). The first aim of this study is to get an updated picture of the spatial patterns of residential fires in Helsinki by using measures of spatial autocorrelation. The second aim of this study is to create a spatial risk model by combining various aggregated census variables, and to identify the main urban factors influencing residential fires in Helsinki at neighbourhood level by using Ordinary Least Squares regression and Geographically Weighted Regression.

Thus, the thesis aims to answer the following research questions:

1. Are there spatial patterns in the distribution of residential fires?
2. What are the main factors predicting residential fire occurrence?
3. Can local variability be found in the explanatory factors and what kind of local patterns are discovered?

2 Background

2.1 Concepts and definitions

2.1.1 Fire risk in urban areas

Fire risk can be defined as the probability of fire occurrence and the negative consequences to be expected with it, such as injuries, deaths, and economic losses (Watts & Hall 2002, 2818-2820). A risk usually consists of exposure and consequences. The potential or likelihood of a hazard, such as fire, is measured as exposure to a source of risk, and the consequences become real if this exposure occurs. Therefore, fire risk can be divided into two parts: risk of fire occurrence (*exposure*) and risk of fire loss (*consequences*) (Watts & Hall 2002; Jennings 2013). Risk of fire occurrence includes all those factors that increase the possibility of being exposed to a fire regardless of the consequences. Risk of fire loss, on the other hand, considers factors that can increase the possibility of experiencing greater consequences in a fire, in terms of injuries, deaths, or economic losses. In this study, fire risk is defined as the likelihood or expected frequency of fire occurrence regardless of the consequences.

Studying fires in urban areas is relevant for several reasons. First, risks are closely associated with population density, because in places where there are more people, there is also a greater likelihood of events with unwanted consequences (November 2004). Residential buildings are also helpful to indicate people's location in cities. That is, by assuming that residential fires are usually caused by people, and because of the assumption that people live in buildings, fire datasets can give a good indication of people's locations in cities (Rohde et al. 2010). Consequently, fires are also more frequent and systematic in cities where the population density is high, which is why urban fires have been an increasingly studied topic (Wallace & Wallace 1984; Jennings 2013; KC & Corcoran 2017; Ardianto & Chhetri 2019).

Second, the proportion of residential fires of all fires is substantial. According to statistics, 70% of all building fires in Finland happen in residential buildings (Emergency Services Academy 2020). As the major urban area Finland, Helsinki's share of residential fires is considerable. In fact, around 10.5% of all residential fires between 2014–2018 happened in Helsinki (Emergency Services Academy 2020).

In addition, most of the negative consequences from fires happen in urban residential buildings. In Finland, 70% of all fire deaths and injuries in fires happen in residential buildings, and the estimated economic loss of destroyed property is yearly almost 48 million euros (Emergency Services Academy 2020). Moreover, at the current population level, more people die in building fires in Finland compared to other European countries, with an average of 75 fire deaths, and over 600 injures in recent years (Ministry of Interior 2019).

Figure 1 shows the trend of residential fires in total and per capita between 2009 and 2018 in Helsinki and Finland. Although the total number of residential fires and fires per capita are decreasing both in Helsinki and in the whole country, the threat of injuries and economic losses has remained at the same levels (Emergency Services Academy 2020).

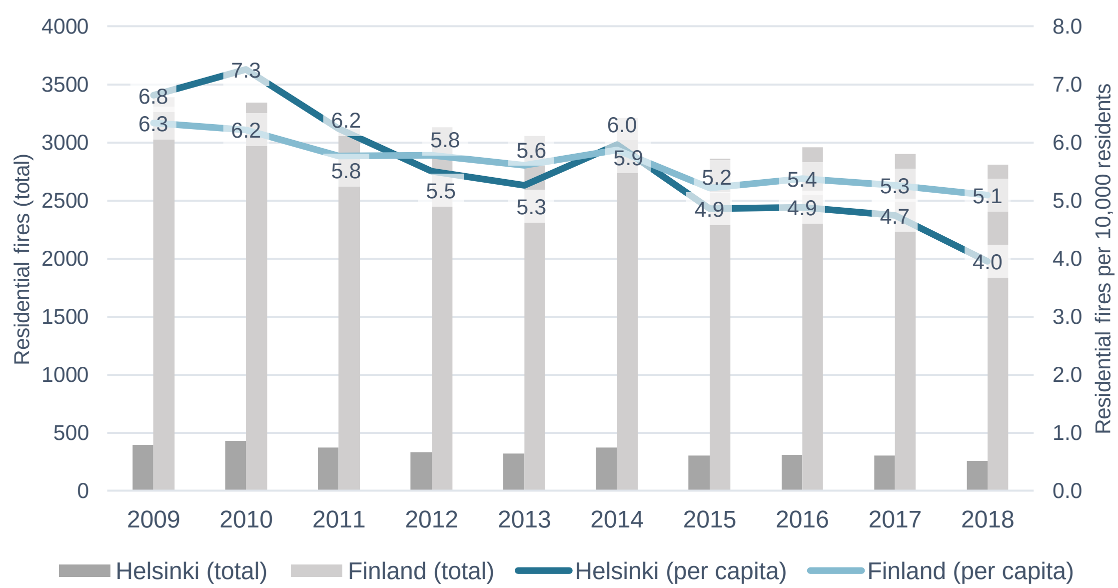

Figure 1. Residential fires in Helsinki and Finland in 2009–2018 with total number of fires and fires per capita. Sources: Emergency Services Academy (2020) & OSFb (2020).

2.1.2 Fire risk analysis

Fires are unexpected events, which is why analysing and managing them is important for the overall safety of the society. Fires can ignite due to multiple reasons that are hard to control.

That is why it can be difficult to avoid fires completely, but with effective fire management and risk modelling, their harmful consequences can be reduced (Ceyhan et al. 2013; Špatenková & Virrantaus 2013).

Fire management is part of the larger concept of risk analysis. Risk analysis is defined as the systematic process of identifying and understanding the sources of risk with the best available data and knowledge (Society for Risk Analysis 2018). Thus, fire risk analysis is not only the process of identifying fire related risk factors by using modern data and methods, but it also includes fire management, i.e. the tools to manage fire risk. According to Ceyhan et al. (2013, 226), fire management covers the "*systematic analysis, planning, decision making, assignment and coordination of available resources to manage fire related risks and includes interrelated sub-phases such as prevention, preparedness, response and recovery*".

These four sub-phases form the principles of emergency management (Godschalk 1991; Cova 1999). Each of these sub-phases affects one another, and they are linked to each other in a cycle. Together they form a general framework to guide the management of different emergencies, such as fires, and each phase plays an equally important role (*Figure 2*).

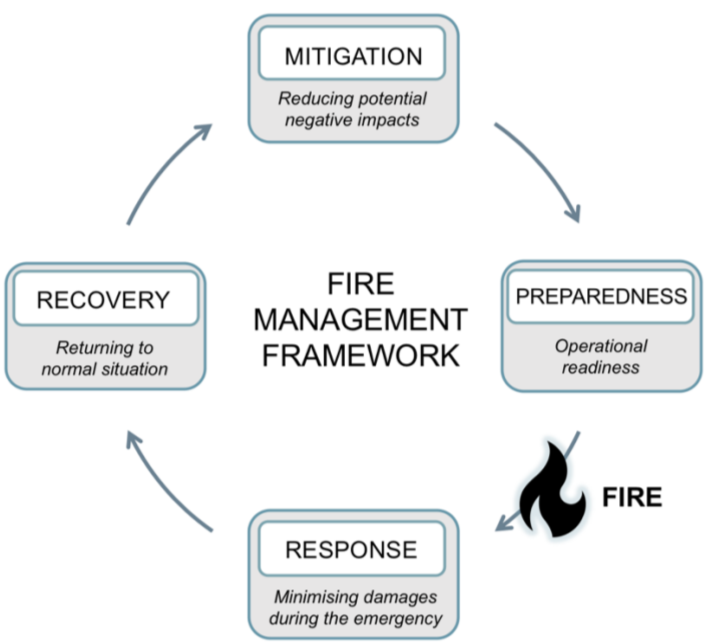

Figure 2. Emergency management framework with fire risk. Modified from Godschalk (1991) and Cova (1999).

In this context, prevention has been identified as the most efficient way to manage the fire problem (Jennings 2013). Prevention, or mitigation, is the process that takes place well before the emergency happens and consists of actions that aim to reduce the potential negative impacts of risks with effective risk assessment. Risk assessment, as part of risk analysis, uses scientific frameworks such as probability theory, modelling, and statistical analyses to identify and understand risk sources (Society for Risk Analysis 2018). In other words, risk assessment answers to questions such as “what could go wrong”, “what is the likelihood of it”, and “what are the consequences”.

To summarize, studying factors affecting fire risk in urban areas is not only important for the safety and public health in cities, but it also helps to mitigate the risk long term with effective fire management and prevention strategies. As the risk management process is cyclic, effective mitigation efforts also help rescue services to be better prepared for accidents. Better understanding of the spatial distribution of fires in the study area as well as investigating the main driving factors behind the variation can give valuable information for fire management purposes (Taylor et al. 2012; Ceyhan et al. 2013; Oliveira et al. 2014). In particular, statistical modelling and GIS analyses, such as risk mapping, are regarded as extremely practical and significant methods in risk mitigation, and they are also widely used in the rescue services worldwide (Cova 1999; Rohde et al. 2010; Tillander et al. 2010; Špatenková & Virrantaus 2013; Ardianto & Chhetri 2019).

2.1.3 Risk analysis at the Finnish rescue services

In Finland, the Rescue Act (379/2011) obliges the rescue services to ensure the safety and control of the rescue operations in their service area, prevent and prepare for fires and other accidents, take proper action in case of accidents, and limit the consequences of accidents.

The national Emergency Service Strategy (Ministry of the Interior 2016) and the new national agenda for a safe and accident free everyday life (Ministry of the Interior 2019) set certain goals to be met by 2025. The key goals according to these reports are to increase all kinds of cooperation between authorities and other stakeholders, with an emphasis on increasing research activity especially regarding accident prevention (Ministry of the Interior 2019). What is especially needed, according to both reports, is a more holistic understanding of risks. This understanding should be based on a continuous analysis, exploiting new technologies and methods to create more detailed analyses, and improving data quality and management (Ministry of Interior 2016; 2019).

Locally, municipal rescue departments are obliged to conduct a risk analysis covering their service area. In addition to the previous definition of risk analysis being a systematic process, it is also a document that is used to validate the service standard for the rescue departments that is updated every four years (see e.g. Helsinki City Rescue Department 2017). The aim of the risk analysis is to recognize risks and their underlying factors, improve prevention, preparedness, and response with targeted resource allocation, and support quick recovery, i.e. improve the total emergency management efficiency (Ministry of the Interior 2016).

Each rescue department is responsible for their own risk analysis. The rescue departments in Uusimaa region base their service standard on a co-operative risk analysis and common criteria. The risk analysis document involves a description of the operative environment and predicted development, assessment of different types of risks, their causes and consequences, and effectiveness of risk prevention activities. The risk assessment is done using accident statistics and GIS methods as well as literature.

For determining response readiness in different areas, the Finnish rescue services use a regression model that predicts the probability of a building fire. The model is calculated to a spatial grid covering the whole country in 1000 m resolution. This nation-wide building fire risk model is developed by Tillander et al. (2010), and it utilizes population density, floor area, and the combination of those explanatory variables in a simple linear regression. Tillander et al. (2010) also tested different spatial approaches such as spatial regression, but as linear regression performed better in 1000 m resolution on the country scale, the model was selected and implemented by the rescue services. Since there are only three variables in the simple regression, the risk level is simple to calculate for the whole country. In Helsinki, during the time of the study, population density, floor area and their combination explained around 68% of building fire variation (Tillander et al. 2010). However, this model predicts fire occurrence in all buildings, and lacks a more detailed analysis of factors driving specifically residential fire occurrence on smaller spatial scale. Tillander et al. (2010) suggest that due to the constant development of the methods, it is highly encouraged to develop spatial models for the risk analysis work in the rescue services. This need also motivates this study.

2.2 Residential fire risk factors

2.2.1 Risk factors detected in earlier studies

Earlier studies have found that risk of a fire at home is not the same for everyone. In fact, socioeconomic status, household circumstances, and neighbourhood characteristics have been found to have significant effect on the likelihood of urban residential fire incidence (Jennings 1999; Jennings 2013).

Early studies found several drivers associated with an increased risk of urban fire occurrence, including dense population, socioeconomic status, education level, household crowdedness, family stability, family lifecycle, residential structure, and household ownership status (Wallace & Wallace 1984; Jennings 1999). Recently, Jennings (2013) wrote a literature review specifically about urban residential fires and the associated socioeconomic risk factors. According to his review, findings in studies after the 1970’s showed little consistency in different risk factors until the 2000’s. Since then, studies have found an explicit connection between residents’ socioeconomic characteristics and higher risk of a residential fire, as well as a more indirect link between housing conditions, neighbourhood properties, and fire incidence (Jennings 2013).

On a general level, poverty and poor housing quality have been consistent factors affecting high fire rates since early research. Gunther (1981), with one of the first models to identify drivers behind higher fire incident rates in the US, found that with an increase in income level, the fire incident rates became smaller. The study also found an association between low education and higher risk of fire (Gunther 1981). Similarly, the Federal Emergency Management Agency in the US found in the late 1990’s that low income, unemployment, and other socioeconomic factors influence residential fire occurrence (FEMA 1997). More recently, numerous studies have achieved similar results, empirically confirming a relationship between social deprivation and a higher fire risk both at the individual level and neighbourhood level (e.g. Duncanson et al. 2002; Holborn et al. 2003; Chhetri et al. 2010; Corcoran et al. 2011b; Jonsson & Jaldell 2019).

Socioeconomic deprivation has been associated especially with a higher risk of an injury in a fire. For example, Duncanson et al. (2002) studied fatal residential fires in New Zealand between 1993–1998 at a population level and found an increased risk of fire among the most socially deprived population, pensioners, disabled people, and children. Broadly speaking, low socioeconomic status was seen as a combination of risky behaviour associated with disadvantage, such as high alcohol consumption and smoking (Duncanson et al. 2002). For instance, Runyan et al. (1992) studied risk factors associated with fire fatalities in the US and found in a case-control study that fatal fires are more likely in homes with alcoholics and without smoke detectors. Similarly, Holborn et al. (2003) studied fatal fires in London in the late 1990's and found that cigarette is a major cause for fire deaths. In Finland, while the number of fire deaths has been steadily decreasing, a typical person dying in a residential fire is a middle-aged man living alone, who usually smokes and consumes alcohol (Kokki & Jäntti 2009; Kokki 2014; Östman 2015).

However, an increased risk for fire fatality is not necessarily due to a more common occurrence of fires (Nilson et al. 2015). In other words, factors affecting fire occurrence can be different from those leading to a fire death. The perceived risk can therefore be different at individual, household, and neighbourhood levels. For example, Nilson et al. (2015) found in a cross-sectional study on fires reported by households, that for individual residents, having high education significantly increased the risk of having a fire in their homes. On the other hand, Duncanson et al. (2002) found the proportion of high education to decrease the risk of fire risk at neighbourhood level. Nilson et al. (2015) and Xiong et al. (2015) suggest that fires are more likely to occur in more advantaged families, but these families are more likely to survive than if a more disadvantaged family encounters a fire. Therefore, the effect of socioeconomic status, including levels of income, education, and employment situation, can be very different at individual and collective levels. However, it was understood early on, that socioeconomic factors predict fire rates best at neighbourhood level, and the actions and behaviour of humans inside the buildings are the main drivers for fire incidence (FEMA 1997).

In addition to socioeconomic status, family and household structure have been important indicators for fire occurrence (Corcoran et al. 2007; Chhetri et al. 2010). At a household level, it has been found that one-parent families increase the risk of fire, probably because the household income is not only smaller, but there is also a higher risk to leave a child unattended (Gunther 1981; Chhetri et al. 2010). Children can also be more likely to play with fire and cause accidental fires. Consequently, some studies found fires to be more frequent in families with children (Chhetri et al. 2010; Nilson et al. 2015; Turner et al. 2017). However, in another study it was found that households with children experienced less fires (Špatenková & Stein 2010). In a Finnish study, adult households were found to have a strong relationship with fire occurrence (Špatenková & Stein 2010).

Furthermore, according to Nilson et al. (2015), elderly people seem to have a decreased risk to experience a fire but at the same time a higher risk to die in a fire. A possible reason is that elder people live in a protected environment and are less likely to engage with risky behaviour, such as smoking or cooking. However, once a fire starts, they might have more difficulties to exit the building within three minutes, which is the standard time limit before a fire becomes dangerous (Muckett & Furness 2007, 190).

Household crowdedness has also been associated with a higher risk of fire in several studies. Studies have found that families with many members, especially with children, have a higher risk to experience a fire (Gunther 1981; FEMA 1997; Duncanson et al. 2002; Nilson et al. 2015). Crowdedness is usually associated with limited living space per household member. According to the World Health Organization (WHO 2018), household crowdedness is often used as an indicator for poverty and social deprivation.

Regarding household tenure and fire risk, home ownership has been found to decrease fire risk (Duncanson et al. 2002; Corcoran et al. 2007). In these studies, living in an owner-occupied dwelling has been a protective factor against fire risk. On the other hand, contradicting results have been found regarding rental housing and fire risk. Greene (2012) studied fires reported both by fire departments and households in the US and found a relationship between living in rented accommodation and a higher risk of fire incidence. Conversely, Nilson et al. (2015) found rental accommodation to decrease fire risk. A recent Swedish case-control study by Jaldell and Jonsson (2019), adds that rented housing also decreases the risk of a resident dying in a fire. Results suggest that living in a rented apartment can act both as a protective and a risk factor. Apartments can be accommodated by many different socioeconomic groups, such as students, middle-income families etc.; thus, a clear relationship is hard to establish. Different results between the US and Sweden might also be tied to pricing of the housing market as well as cultural differences.

Apart from factors related to households and residents, some structural variables, such as building materials, building age, and building type have also been associated with a higher risk of fire in some studies (Goodsman et al. 1987; Runyan et al. 1992; Vasiliauskas & Beconytė 2015; Xiong et al. 2015; Turner et al. 2017). For example, it was found that fires are more common in old buildings in poor condition (Shai 2006). Also, at household level, a link has been found between low-quality housing, low socioeconomic status, and higher fire rates (Shai 2006). The result suggests that more disadvantaged people might be financially less capable to invest in fire protection equipment, such as smoke alarms and fire-resistant furniture, which can further increase their vulnerability to fire consequences (Goodsman et al. 1987; Jennings 2013).

Spatial analysis and GIS methods became more popular in urban fire risk research in the 2000's, and it was understood that fires tend to cluster in certain neighbourhoods in many urban areas (Jennings 2013; Wuschke et al. 2013; Guldåker et al. 2018). Already in the early literature a link between population density and urban fires was found (Wallace & Wallace 1984), and the results of later studies confirm the association (Tillander et al. 2010; Špatenková & Virrantaus 2013). The results are not surprising given the strong connection between population density and fire risk in cities. Furthermore, the density of residential buildings in an area was also found to increase fire risk, as fires tend to cluster in areas with high proportion of residential buildings (Ceyhan et al. 2013; Ardianto 2018).

Overall, Chhetri et al. (2010) point out that fires are socially constructed, and they are closely connected to the way the society is unevenly divided. In fact, Chhetri et al. (2010) and Corcoran et al. (2011b) emphasize that residential fires are not just random and accidental events, but rather an outcome of the existing socioeconomic circumstances. The relationship is not surprising, given that urban areas are often spatially fragmented due to disparities in socioeconomic features that are shaping them (Chhetri et al. 2006). This was demonstrated in a recent study in Sweden by Guldåker et al. (2018), where it was found that the proportion of accidental residential fires, such as kitchen fires, were quite equally distributed in the city, while at the same time the proportion of intentional fires such as arson fires increased significantly with poorer living conditions (Guldåker et al. 2018). According to Chhetri et al. (2010), unstable family situation, socioeconomic disadvantage, and dwelling quality can provide a good starting point to investigate areal differentiation of urban fires. In other words, wellbeing of households is often linked to the wellbeing of neighbourhoods.

Nevertheless, the connection between socioeconomic deprivation and higher fire risk might not be the same in Helsinki as elsewhere, as Helsinki is characterized in particular by relatively low level of residential segregation compared to international levels, high standard of welfare services, and a generally high housing quality (Saikkonen et al. 2018). To compare, Stockholm, as the capital of the neighbouring country Sweden, has been identified as one of the most socioeconomically segregated capital cities in Europe (Musterd et al. 2017).

Socioeconomic segregation has not caused similar problems in Finland as in many other countries, as large Finnish cities including Helsinki have proactively tried to mitigate and manage socioeconomic inequalities in cities, for example with social mixing policies (Saikkonen et al. 2018). However, despite the political efforts, there is evidence of intensified and expanded accumulation of regional disadvantage (Kortteinen & Vaattovaara 2015; City of Helsinki 2019b), as well as of increased regional income disparities in the Helsinki Metropolitan area during the last few decades (Kortteinen & Vaattovaara 2015; Saikkonen et al. 2018). Thus, in this context, it is relevant to study how the results of earlier international studies relate to Helsinki in terms of residential fire risk.

2.2.2 Conceptualizing residential fire risk

Although there is a link between socioeconomic disadvantage and high risk of fire, deprivation does not itself cause fires. Different socioeconomic factors affect the exposure to fire risk in different ways, and there are many indirect factors leading to fire ignition, which is why a theory is needed to understand the complexities of fire risk.

However, despite the growing interest in modelling urban fire problem in recent years, there is still a lack of a universal and well-defined theory of fire risk. Recently many studies have pointed out this shortcoming (Corcoran et al. 2011b; Corcoran & Higgs 2013; Jennings 2013; Wuschke et al. 2013; Ardianto & Chhetri 2019). A comprehensive theory of fire risk would not only be a highly effective tool in fire risk research to understand causal relationships with varying factors, but it would also help rescue services to plan and target prevention measures based on certain standards.

Some researchers have argued that the current absence of theory has even been slowing down the expansion of knowledge of fires (Jennings 2013). In recent years there have been some proposed theoretical viewpoints; one even suggesting an ecological criminology theory approach (Wuschke et al. 2013).

Corcoran et al. (2011b) proposed a theoretical framework for fire risk, which connects different socioeconomic, behavioural, structural, and environmental factors that lead to the ignition of a fire. Jennings (2013) emphasized that this framework is suitable to guide empirical research, and it has been widely referenced in later urban fire research (e.g. Song et al. 2017; Zhang et al. 2018; Ardianto & Chhetri 2019; Nilson & Bonander 2020). Since the conceptual model of Corcoran et al. (2011b) comprises fire risk as a whole, it has been partly modified in this study in order to conceptualize residential fire risk and to guide variable selection for the fire risk model (*Figure 3*).

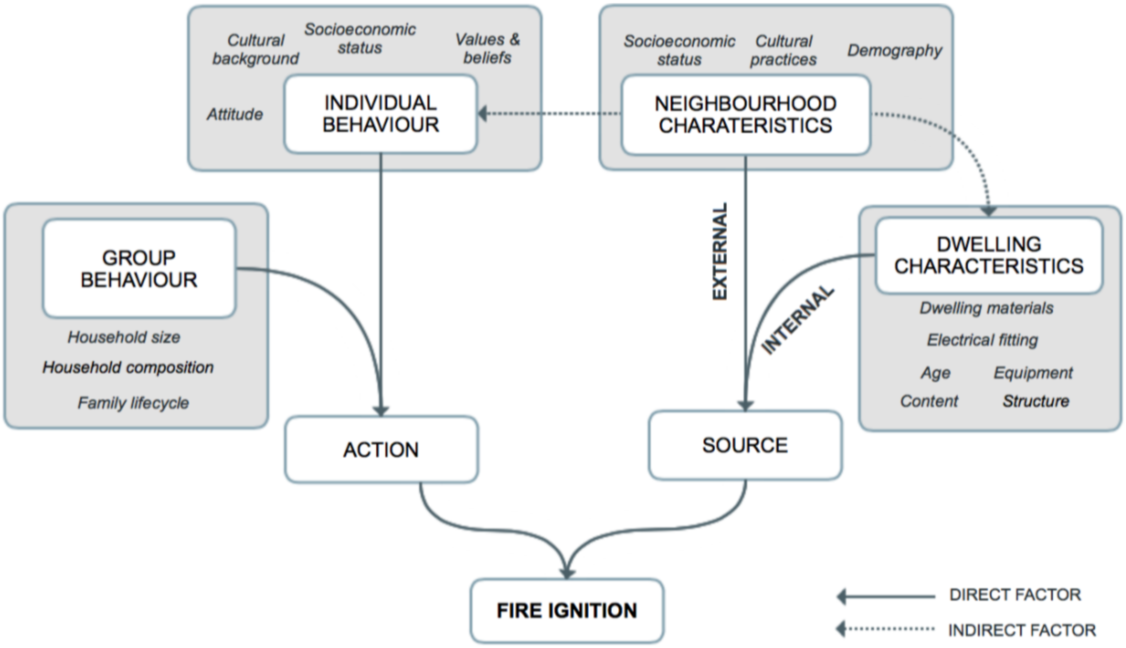

Figure 3. Conceptual model of residential fire risk. Partly adopted from Corcoran et al. (2011b).

According to the model of Corcoran et al. (2011b), fire ignition requires a source and an action. Sources and materials can be either internal or external. Internal sources can be the properties that either cause or aggravate the fire, such as building age, dwelling materials, or furniture. Neighbourhood characteristics, such as the socioeconomic and demographic status of the neighbourhood, account for external sources.

Apart from a source, fire ignition often requires an action by a human – unless the fire is caused by some technical fault. The human action is usually a result of different individual and group behaviours, shaped by values, socioeconomic status, cultural background, household composition and size, and family lifecycle (Corcoran et al. 2011b).

These different characteristics can then be connected to each other either directly or indirectly. For example, the behaviour of individuals is indirectly affected by the surrounding neighbourhood characteristics. Accordingly, it would be misleading to assume that if there is a correlation between poverty and an increased risk of fire, that poverty itself directly increases the risk of fire (Clark et al. 2015, 1123). Rather it can be an indirect factor associated, for example, with an increased exposure to sources of fire, poor housing condition or riskier behaviour, such as high alcohol consumption, all leading to a higher risk of fire incidence (Corcoran et al. 2011b; Ardianto & Chhetri 2019).

All these different factors lead to different contextual situations in a study area, which are not independent from each other. As Corcoran et al. (2011b, 197) state, "*for a fire to occur it is dependent upon certain conditions to be met and in their absence the likelihood of a fire is minimal"*. Therefore, the situational context plays an important role in understanding fire risk (Ardianto 2018).

2.3 Assumptions and conditions in fire risk modelling

2.3.1 Models in earlier studies

Many of the influential studies previously discussed have successfully identified factors associated with a higher risk of residential fire incidence with varying methods. An essential part of modelling fire risk is the definition and quantification of fire risk. Fire risk has been quantified in different ways in different studies, for example as the count of fires per spatial unit (Duncanson 2002; Corcoran et al. 2011a; Wuschke et al. 2013), fires per capita (Corcoran et al. 2007; Chhetri et al. 2010; Chhetri et al. 2018), fire density (Špatenková & Virrantaus 2013; Song et al. 2017), and as the likelihood of fire as a function of past fire history (Ardianto & Chhetri 2019). Having many different definitions of fire risk, however, has made it difficult to compare and validate results with each other.

Early literature used mainly descriptive methods, and in recent years different quantitative statistical methods to model urban fire risk have become increasingly popular (Jennings 2013). In particular, the recent technological advancement has enabled new methods to be used also in urban fire risk studies (Jennings 2013). The main advantage of using statistical methods to study quantitative fire risk is that the data is based on real fire incidents and the results are often simple to apply in practice (Ramachandran & Charters 2011). Typically used methods for studying urban fires include different visual clustering analysis methods, probabilistic models, regression models, and, in recent years also different spatial analytical models.

In some studies that have employed visual methods, Kernel Density Estimation (KDE) has been a popular method for detecting areas with a higher fire risk. KDE was used for example by Corcoran et al. (2007), Asgary et al. (2010), and Guldåker and Hallin (2014). In this approach, a weighted bandwidth, or a kernel, is created around each sample point, which then estimates the probability density function of the random variable (e.g. fires). For example, KDE allows detecting increased likelihoods of events, such as fires, inside defined clusters. Another visual method that has been applied in fire risk studies is a *comap*; a GIS-based method to explore spatial and temporal patterns (Chhetri et al. 2009; Asgary et al. 2010). This method, developed by Brunsdon (2001), allows the investigation of the locations of fire incidents and their relationship with conditioning variables, such as time of the incident and socioeconomic variables, which can then be mapped.

While some studies have detected clusters and densities of past fires, others have created probabilistic models to predict or forecast fire occurrence. Some of the probabilistic methods used in residential fire research are Bayesian network (Rohde et al. 2010), point pattern analysis (Špatenková & Virrantaus 2013), and modified kernel density functions such as Diggle's D- and Ripley's K-function (Ceyhan et al. 2013). In addition, recently Ardianto and Chhetri (2019) created a fire prediction model using a Bayesian approach, more specifically Markov chain Monte Carlo method, which was employed to predict fire occurrence in a city in Australia based on the location and time of past fires.

As a third group of popular methods, regression models have been widely used to model the association of fire with different explanatory variables (Corcoran et al. 2007; Chhetri et al. 2010; Tillander et al. 2010; Corcoran et al. 2011a; Taylor et al. 2012; Hu et al. 2019). Regression has been a popular approach because it estimates the intensity and significance of the relationships between variables. The method is simple to apply, and it is applicable to many different research questions.

In traditional regression, the dependent variable Y is modelled as a function of a matrix of explanatory variables X, the corresponding parameters, plus the error term. In fire studies, the dependent variable has been the quantified fire risk, e.g. the count of fires or fires per capita, regressed against multiple chosen explanatory variables, such as census variables.

For example, Chhetri et al. (2010) used multiple regression to explore the relationships between residential fires and socioeconomic characteristics in Australia. Similarly, in Finland, the current building fire risk model used in risk analysis is based on a traditional multiple regression (Tillander et al. 2010).

However, traditional global regression models often do not explain local processes over space, leaving out a lot of important and interesting information about the phenomenon. The Geographically Weighted Regression (GWR) is an extension to the traditional regression, which has been developed to capture the effects of non-stationarity and differences in the importance of the variables (Fotheringham et al. 2002). The GWR, as a spatial modelling method, has been more widely used in research on forest fires and fires in cities in general (Yamashita 2008; Koutsias et al. 2010; Oliveira et al. 2014; Song et al. 2017). To date, only a few empirical studies have used the GWR to study residential fires (Špatenková & Virrantaus 2013; Ardianto 2018).

Špatenková and Virrantaus (2013) applied the GWR in their research in Helsinki to study where and how much residential fires are influenced by explanatory variables such as building type, family structure, population density, and income. The authors not only found that the GWR model performed better than the traditional regression, but also the explanatory variables showed great variability across the study area. For example, a higher proportion of households with children decreased fire risk in the city centre of Helsinki, while in the same location a higher proportion of adult households increased the risk of fire incidence (Špatenková & Virrantaus 2013).

More recently, Ardianto (2018) used the GWR in his doctoral dissertation to study local spatial drivers of residential fires in Melbourne, Australia. Ardianto (2018) found, for example, that owning an apartment increased fire risk more in the eastern Melbourne and decreased it in the central business district. Results indicate that the explanatory variables show great variability not only in their predictive outcome, but also in their intensity and direction across the study area. Overall, these studies prove that residential fires are non-stationary, and the GWR is able to address this problem and to find spatial variations in the study area (Špatenková & Virrantaus 2013; Ardianto 2018).

Since the relationship between fire risk and its influencing factors may vary over space, traditional regression models – often used in fire studies – may not be adequate for examining the spatially varying relationships between multiple predictors and fire occurrence (Yamashita 2008; Špatenková & Virrantaus 2013; Song et al. 2017; Ardianto 2018). Therefore, it is suitable to use a local analysis method, such as the GWR, that can capture local variations, rather than a global model that assumes spatial stationarity. To further justify this selection, differences between global and local models are discussed in the next section in more detail.

2.3.2 Comparison of global and spatial models

Traditional non-spatial analysis methods, such as traditional regression models, are not always suitable for analysing spatial data (e.g. fires) (Brunsdon et al. 1996; Fotheringham et al. 2002). Spatial analysis methods are more suitable for that purpose, as they are used to identify the nature of relationships between variables, and to create new information out of spatial data (Anselin 1995; Brunsdon et al. 1996).

Traditional regression methods have several assumptions that cause problems with spatial data. First, they assume that data is homogeneous, meaning that both the observations and the model residuals are independent from one another, and the modelled relationships are the same everywhere in the study area. Essentially, the methods presume that the strength and magnitude of the coefficients are universal throughout the study area. In a traditional regression model, the dependent variable is assumed to be spatially stationary, which means that the results of the regression are global estimates, i.e. average values from the whole study area. This is why the method produces global statistics.

In reality, spatial independence is rarely the case with spatial data. As Tobler (1970) observed, "*everything is related to everything else, but near things are more related than distant things*". This notion, also called Tobler's first law of geography, suggests that spatial data is not independent from its neighbours. Nearby observations have similar values, and the values in one location are affected by the values of neighbouring locations (Tobler 1970).

Spatial dependence can occur either in the model's variables or the model's residuals (Fotheringham et al. 2002). Fotheringham et al. (2002) argue that if this kind of spatial structure exists and it is not properly addressed, it can lead to an increase in the model's standard error. That can in turn cause the model's estimated parameters to be either too high or too low compared to the unknown real values, causing bias in the prediction results.

While spatial dependence is one aspect of spatial data, another important characteristic is spatial heterogeneity. Spatial heterogeneity means that the phenomenon not only varies geographically, but also the processes that generate the variation (i.e. underlying spatial characteristics) vary across space, and thus are non-stationary (Anselin 1988; Charlton & Fotheringham 2009). Spatial non-stationarity means that the data does not have long term mean values, i.e. it behaves in an unpredictable way. In regression it means that the values of some coefficients can vary in different locations, depending on the mechanisms of certain non-stationary variables (Fotheringham et al. 2002). Fotheringham et al. (2002) further state that social processes are usually non-stationary, and that the measurements of relationships depend on the location in which they were done. For example, the effect of income on the probability of fire can be different in different parts of the study area, which cannot be addressed with a global regression model.

Consequently, heterogeneous spatial data is not appropriate with global models, which assume that data is homogeneous (Fotheringham et al. 2002). This notion violates the aforementioned assumptions of the traditional regression models, and inevitably leads to biased conclusions. Addressing spatial heterogeneity in a model can therefore reveal differences in the relationships in different parts of the study area, which is why calculating local statistics is preferred (Anselin 1995; Fotheringham et al. 2002).

Summary of the main differences between global (traditional) statistics and spatial (local) statistics by Fotheringham et al. (2002) is presented in *Table 1*. Global models produce global statistics, and local models produce local statistics. Global statistics create a single valued number, for example one mean value and one standard error for the whole study area. Local statistics, on the other hand, create multi-valued statistics for each different location in the study area, meaning that each different location has its own statistics: when the location changes, the statistics are also different. This also makes local statistics easier to map and use in a GIS, since it is possible to visualize the parameter estimates at different locations at the same time.

Similarly, global models emphasize similarities and search for averages across space, whereas local statistics try to find differences in space and local "hot-spots" or instabilities within the study area. In other words, global statistics are locally more or less meaningless (Fotheringham et al. 2002). A traditional regression is an example of a global model, and the Geographically Weighted Regression (GWR) is an example of a local model.

Table 1. Comparison of global statistics and local (spatial) statistics (by Fotheringham et al. 2002).

Global	**Local**
Summarize data for whole region	Local disaggregations of local statistic
Single value statistic	Multi-value statistic
Non-mappable	Mappable
GIS-unfriendly	GIS-friendly
Non-spatial or spatially limited	Spatial
Emphasize similarities across space	Emphasize differences across space
Search for regularities or "laws"	Search for exceptions or local "hot-spots"
Example: Traditional regression	*Example*: GWR

3 Materials

3.1 Study area

Study area is the city of Helsinki. Helsinki is of particular interest for the rescue services as it is a major urban agglomeration in Finland and an important political, economic, and cultural centre. Currently there are eight fire stations in Helsinki, seven of which operate throughout the year.

As the capital city, Helsinki is the most densely inhabited city in Finland. At the end of 2018, the population of Helsinki was 648,042, which accounted for 11.7% of the total population of Finland (City of Helsinki 2019a). Together with its neighbouring municipalities, Espoo, Vantaa, and Kauniainen, the region forms the Helsinki Metropolitan area with almost 1.2 million residents (OSF 2020b). Helsinki draws people from other parts of Finland with its multiple possibilities for work and education, and the population is growing yearly. The city of Helsinki estimates that the population of Helsinki will grow from 2018 with over 100 000 residents by 2034 (Vuori & Kaasila 2019).

3.2 Data

3.2.1 Fire incident data

Fire incident data was requested from the Emergency Service Academy Finland, which maintains the national accident database PRONTO (Emergency Services Academy 2020). The database contains data about incidents attended by the rescue services. Each incident is geotagged as a coordinate tuple. Incidents also contain information about the time of the accident, primary accident type, response times of emergency vehicles, and other details. In addition, each building fire contains information about the building type, which is used for selecting fires that happened in residential buildings.

Since 2009, building fires have been divided into two categories. PRONTO defines a *non-developed building fire* as a fire that has not spread from the source of ignition. A usual example of a non-developed building fire is a stovetop or a cooking fire. *Building fires*, on the other hand, are classified as fires that have spread from their ignition point to other rooms (Emergency Services Academy 2020). In this study, building fires and non-developed building fires are combined and considered together as residential fires.

In total, 1546 incidents were collected from the database with the following criteria: non-developed building fires and building fires that have happened in the Helsinki Rescue Service area between 1st of January 2014 and 31st of December 2018, including only fires in buildings classified as residential. Residential buildings include separate small houses (one and two apartment houses and other small houses), row houses, and apartment buildings.

A five-year interval was selected to get enough sample points and also to decrease the possible identification of individual incidents from the analysis. The timeframe is also close to the timeframe of the census datasets, published in 2017 and 2018, and thus justifies the selection of the time interval for incident data. The distribution of residential fires in Helsinki as a kernel density surface with a 500-meter bandwidth is shown in *Figure 4*. Density surface minimizes the risk of recognizing individual incidents for privacy protection. The figure also shows the locations and names of the eight rescue stations in Helsinki.

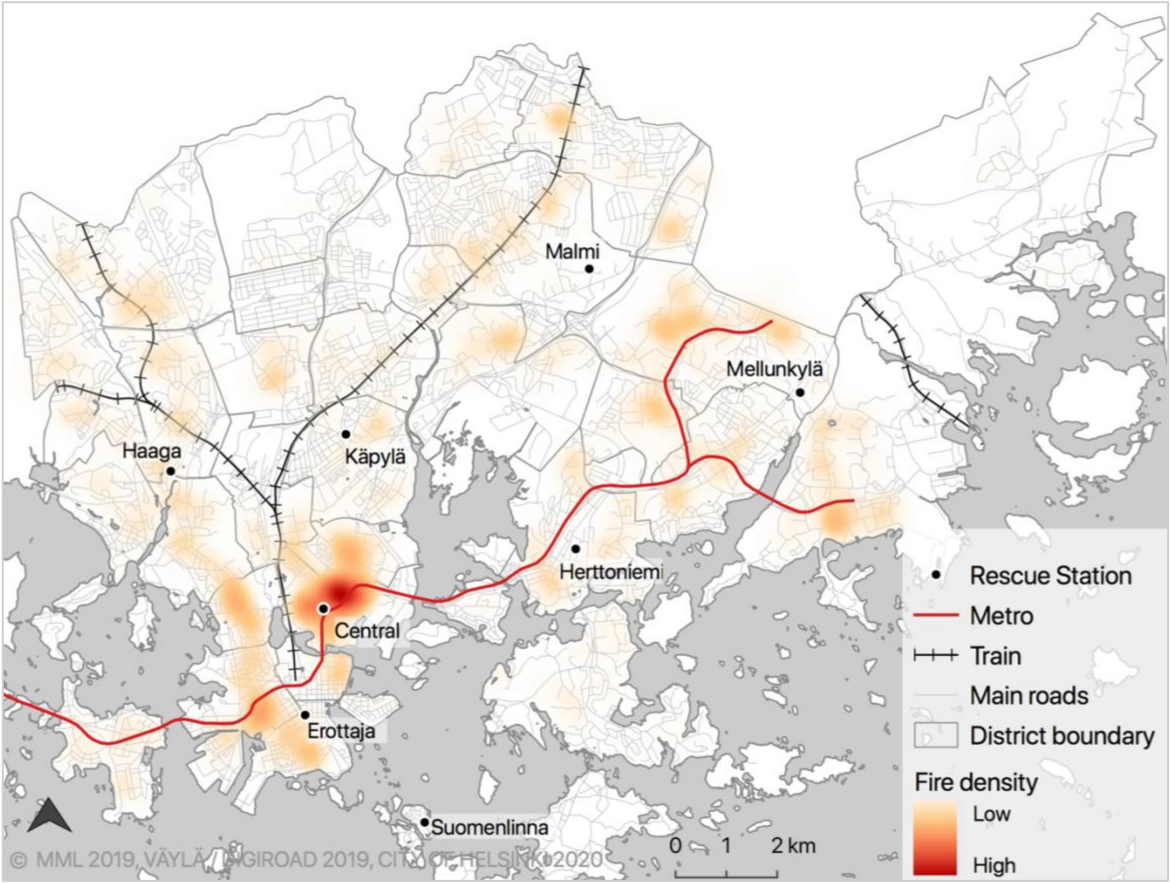

Figure 4. Distribution of residential fires in Helsinki in 2014–2018 as a kernel density surface (500 m bandwidth). Density surface minimizes the risk of recognizing individual incidents for privacy protection. Suomenlinna Rescue Station operates from May until September, other stations operate throughout the year. The Central Rescue Station is located in Kallio district.

3.2.2 Datasets

Background information was collected from two data sources, the Statistics Finland grid database (Statistics Finland 2019) and the YKR database (YKR 2019).

The Statistics Finland grid database provides census data from 2018, with detailed information about the socioeconomic characteristics of the population (Statistics Finland 2019). Data is aggregated into 250 x 250 m grid cells. Data obtained from the database includes information about the population, income and education level, unemployment rate, household tenure, and stage of life of the residents. Data protection applies to some of the variables, such as population density. This means that grid cells with less than three people are marked as protected, and hence were removed from the analysis.

The YKR database is a system for monitoring community structure, maintained by the Finnish Environment Institute SYKE (YKR 2019). Data contains rich information about different structural characteristics, such as building and dwelling properties. Data is also aggregated into grid cells with a cell size of 250 x 250 m, making it consistent with the Statistics Finland grid. Data obtained from the YKR database include the construction year of buildings and the number of rooms in dwellings from 2017.

3.2.3 Data preparation

Total of 1823 grid cells were used in the analysis. Fire incident data from 2014–2018 was aggregated into the census grid by counting the number of fires falling into each grid cell. Grid cells with no fire history during the five years were given a zero value. In 90% of the grid cells in the study area there were no residential fires between 2014–2018. In 8% of the cells there has been at least one fire. The maximum number of fires in one cell was 25 and the mean value was 0.83.

The number of fires was used as the dependent variable to represent fire risk, namely the risk of fire in a neighbourhood. In this study, quantifying fire risk as count of fires is justified because the size and shape of one aerial unit, i.e. one grid cell, is uniform (250 x 250 m). It also meets the definition of fire risk used in this study, which is the likelihood of fire occurrence in a neighbourhood. Neighbourhood is conveniently defined as the area of one grid cell. This allows also representing the count of fires as fire density, such as fires per square kilometre. For easier interpretation of results, count of fires in one grid cell was used to describe the risk in the results of the analysis.

Based on earlier literature, 14 variables were created from the Statistics Finland and YKR databases as candidates for modelling (*Table 2*). The conceptual model of fire risk was used to guide the variable selection and division into three categories.

The first group includes neighbourhood and structural characteristics accounting for the internal and external sources of fire risk. Variables include population density, residential building density, and the average construction year of buildings in a neighbourhood.

Population density and residential building density represent the total number of residents and the total number of residential buildings per square kilometre. An average of the construction years of the buildings in each grid cell was calculated and classified into 11 classes for each decade, starting from 1920 until 2010, as described in *Table 2*. Other variables about the internal dwelling characteristics were not possible to obtain in an aggregated form. The second group consists of socioeconomic variables including the proportion of low and high education, median income, and unemployment rate. Third group includes variables related to household, such as measures of crowdedness of dwellings, stage of life of households, and household tenure. According to Statistics Finland (2019), people living permanently in the same dwelling form a household, or by statistical definition, a household-dwelling unit.

While other variables were relatively straightforward to construct, household crowdedness needed careful consideration. Statistics Finland uses different indicators for housing standard. For example, household density is measured with several different norms. According to the classification of Statistics Finland, a household is crowded when there is less than one room per resident in a household, when the kitchen is excluded (OSF 2020a). Housing standard is also measured with average floor area per a household or a resident, and with the number of residents per room and a housing unit. Therefore, two variables for crowdedness were constructed, one for the average floor area per resident and the other for the average room size per resident.

Table 3 shows the descriptive summary of the different variables, and *Figures 5-7* present the observed distributions of the explanatory variables with a Natural Breaks classification scheme.

Table 2. Summary of candidate explanatory variables.

Variable	Description of measures	Year	Data source	Reference
Population density	Population density per km²	2018	Statistics Finland (2019)	Tillander et al. (2010) Špatenková & Virrantaus (2013)
Residential building density	Residential buildings per km²	2018	Statistics Finland (2019)	Ardianto (2018)
Building year	Construction years classified into classes: <u>1920</u> = up to 1920, <u>1929</u> = 1921-1929, <u>1939</u> = 1930-1939, <u>1949</u> = 1940-1949, <u>1959</u> = 1950-1959, <u>1969</u> = 1960-1969, <u>1979</u> = 1970- 1979, <u>1989</u> = 1980-1989, <u>1999</u> = 1990-1999, <u>2009</u> = 2000-2009, <u>2010</u> = 2010 onwards	2018	YKR (2019)	Shai (2006) Špatenková & Virrantaus (2013)
Education Low education High education	 Proportion of residents above 18 with only basic level education Proportion of residents above 18 with an academic or higher-level university degree	2018	Statistics Finland (2019)	Duncanson et al. (2002) Nilson et al. (2015)
Income	Median annual income of residents (€)	2017	Statistics Finland (2019)	Gunther (1981) Holborn et al. (2003)
Unemployment	Proportion of unemployed residents aged 15-64 of residents aged 18 or above	2018	Statistics Finland (2019)	FEMA (1997), Chhetri et al. (2010) Hastie & Searle (2016)
Household crowdedness Occupancy rate Room rate	Average floor area per resident in m² (total floor area of dwellings/total number of residents) Average number of rooms per resident (total number of rooms/total number of residents)	2017-2018	Statistics Finland (2019) & YKR (2019)	Gunther (1981) FEMA (1997) Duncanson et al. (2002) Nilson et al. (2015)
Family structure Households with children Adults Pensioners	 Proportion of households with children (aged 0-17) Proportion of households with only adults Proportion of households with at least one retired resident	2018	Statistics Finland (2019)	Chhetri et al. (2010) Corcoran et al. (2011a) Špatenková & Stein (2010) Turner et al. (2017)
Household tenure Owner-occupied dwellings Rented dwellings	 Proportion of households living in owner-occupied dwellings Proportion of households living in rented and household right of occupancy dwellings	2018	Statistics Finland (2019)	Corcoran et al. (2007) Corcoran et al. (2011a) Greene (2012) Wuschke et al. (2013)

Table 3. Descriptive statistics of explanatory variables (N = 1823[1]).

Variable	Min	Max	Median	Mean	Std.dev
Population density	320,00	44880,00	3680,00	5538,92	5338,35
Residential building density	0,00	1232,00	240,00	311,84	245,70
Building year	1920	2010	1989	1983	14,98
Low education	0,00	0,62	0,18	0,20	0,11
High education	0,00	0,67	0,21	0,23	0,13
Median income	10678,00	60292,00	25579,00	26075,46	5966,82
Unemployment	0,00	0,33	0,05	0,06	0,04
Occupancy rate	20,70	81,10	34,70	36,26	7,46
Room rate	0,15	3,83	1,43	1,45	0,28
Households with children	0,00	0,82	0,23	0,25	0,13
Adults	0,03	0,99	0,49	0,49	0,15
Pensioners	0,00	0,95	0,26	0,27	0,12
Rented dwellings	0,00	1,00	0,40	0,41	0,31
Owner-occupied dwellings	0,00	1,00	0,57	0,56	0,31

[1]Spatial scale is one grid cell (250 x 250 m).

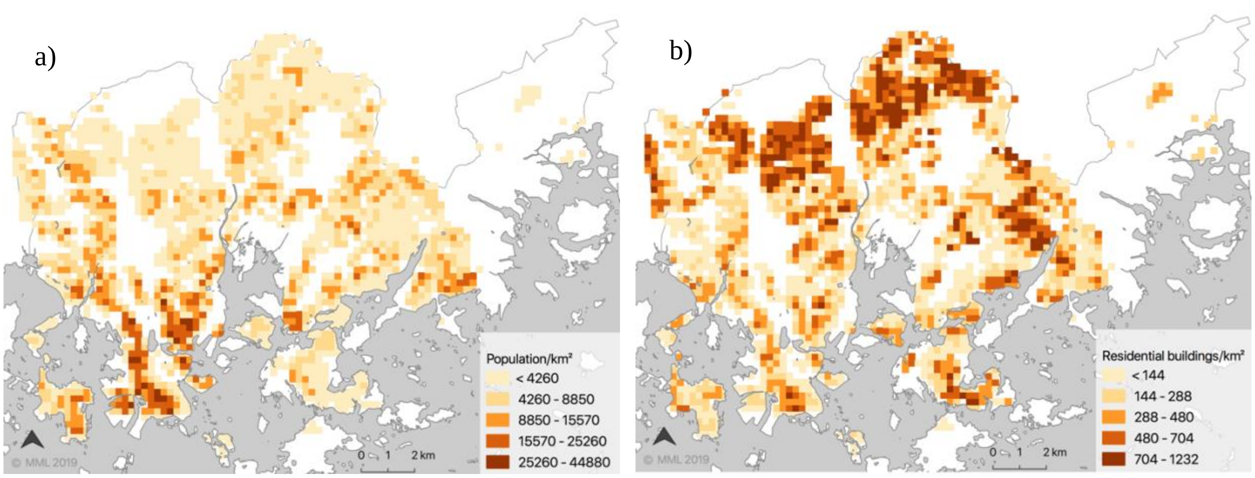

Figure 5. Observed distribution of a) population density per km², and b) residential building density per km². Natural Breaks is used as the classification scheme. Source: Statistics Finland (2019).

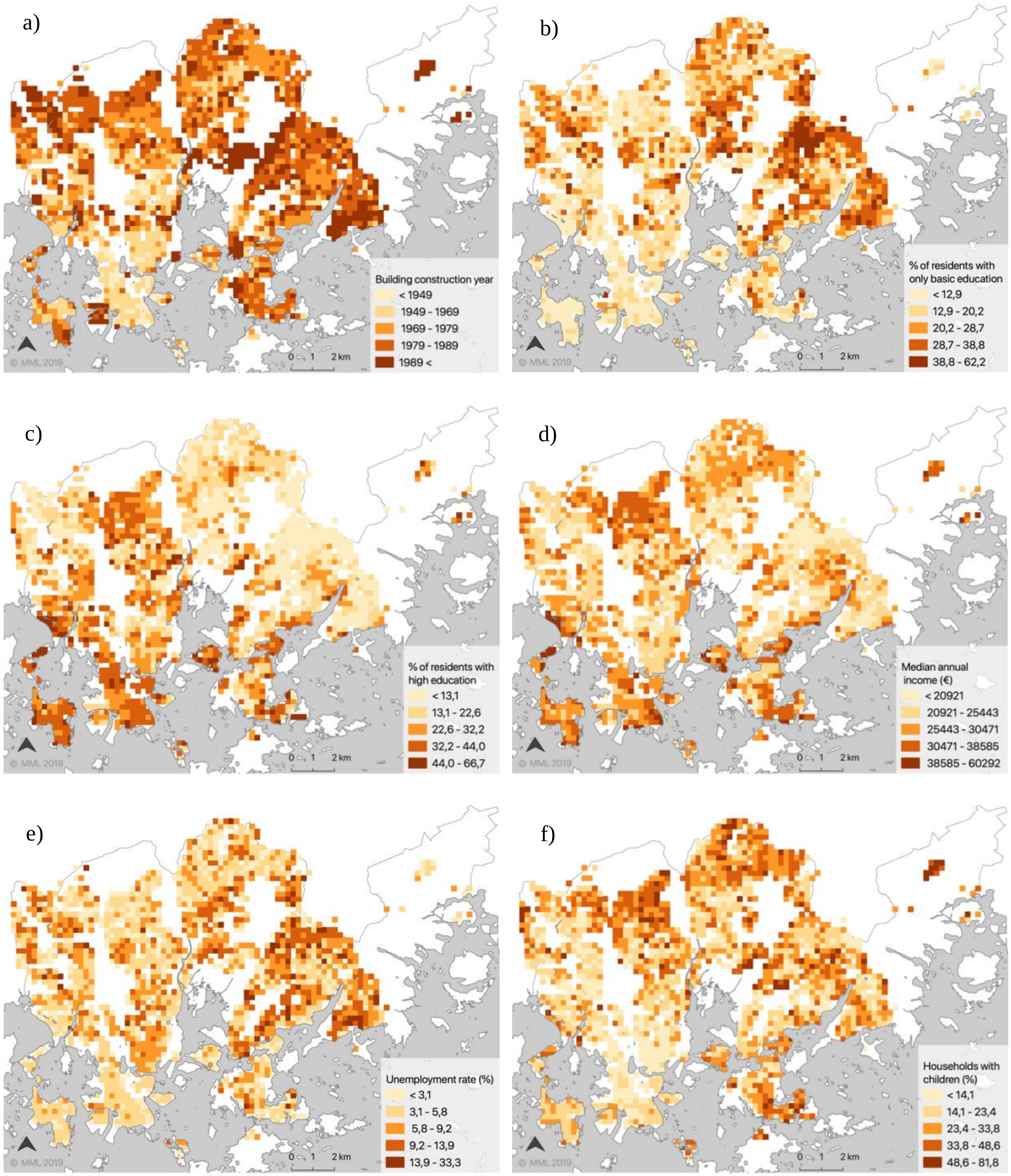

Figure 6. Observed distribution of a) building year; b) residents with low education; c) residents with high education; d) median annual income; e) unemployment rate; and f) households with children. Natural Breaks is used as the classification scheme. Source: Statistics Finland (2019) & YKR (2019).

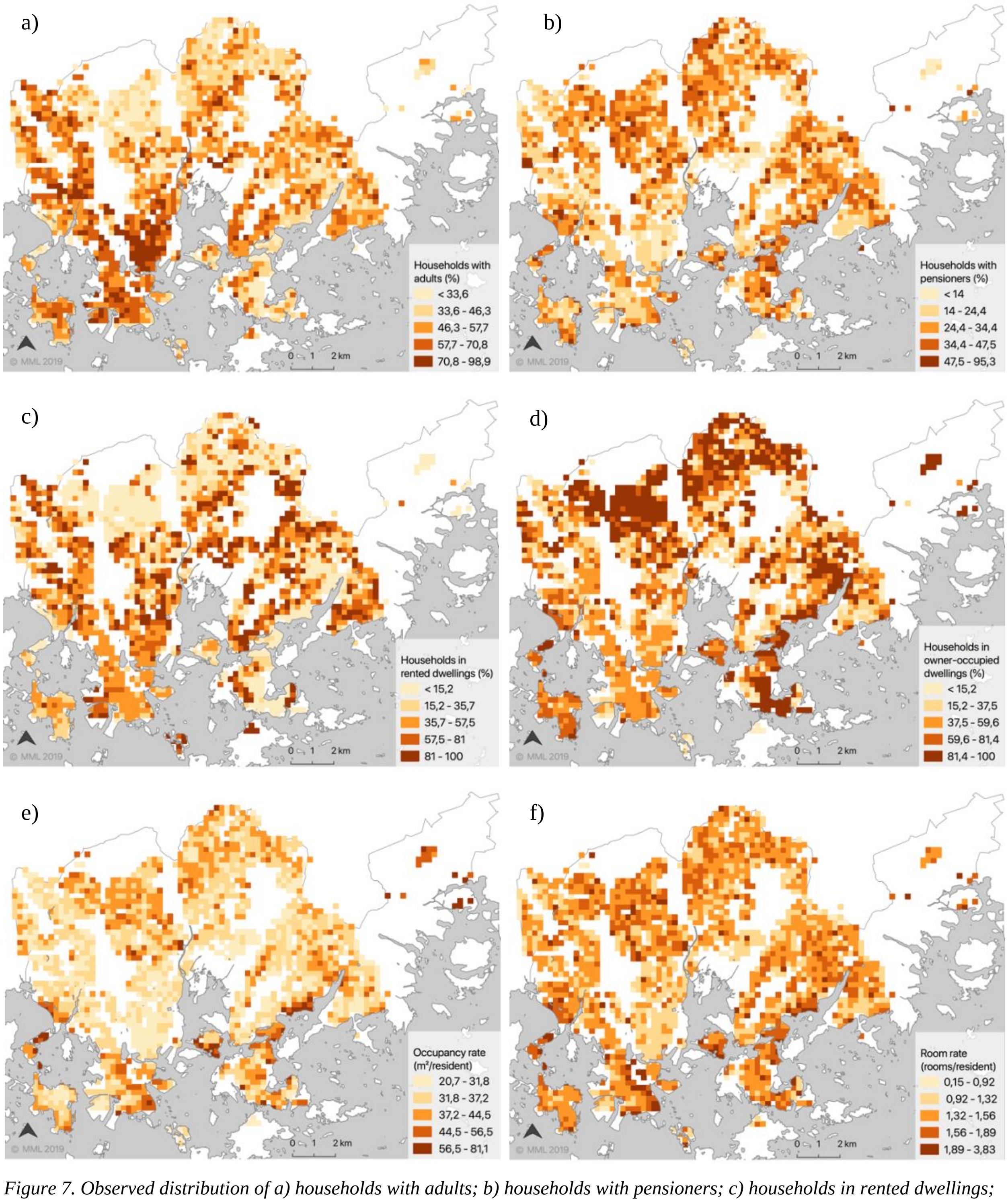

Figure 7. Observed distribution of a) households with adults; b) households with pensioners; c) households in rented dwellings; d) households in owner-occupied dwellings; e) occupancy rate; and f) room rate. Natural Breaks is used as the classification scheme. Source: Statistics Finland (2019) & YKR (2019).

4 Methods

4.1 Study design

The workflow of this study is presented in *Figure 8*.

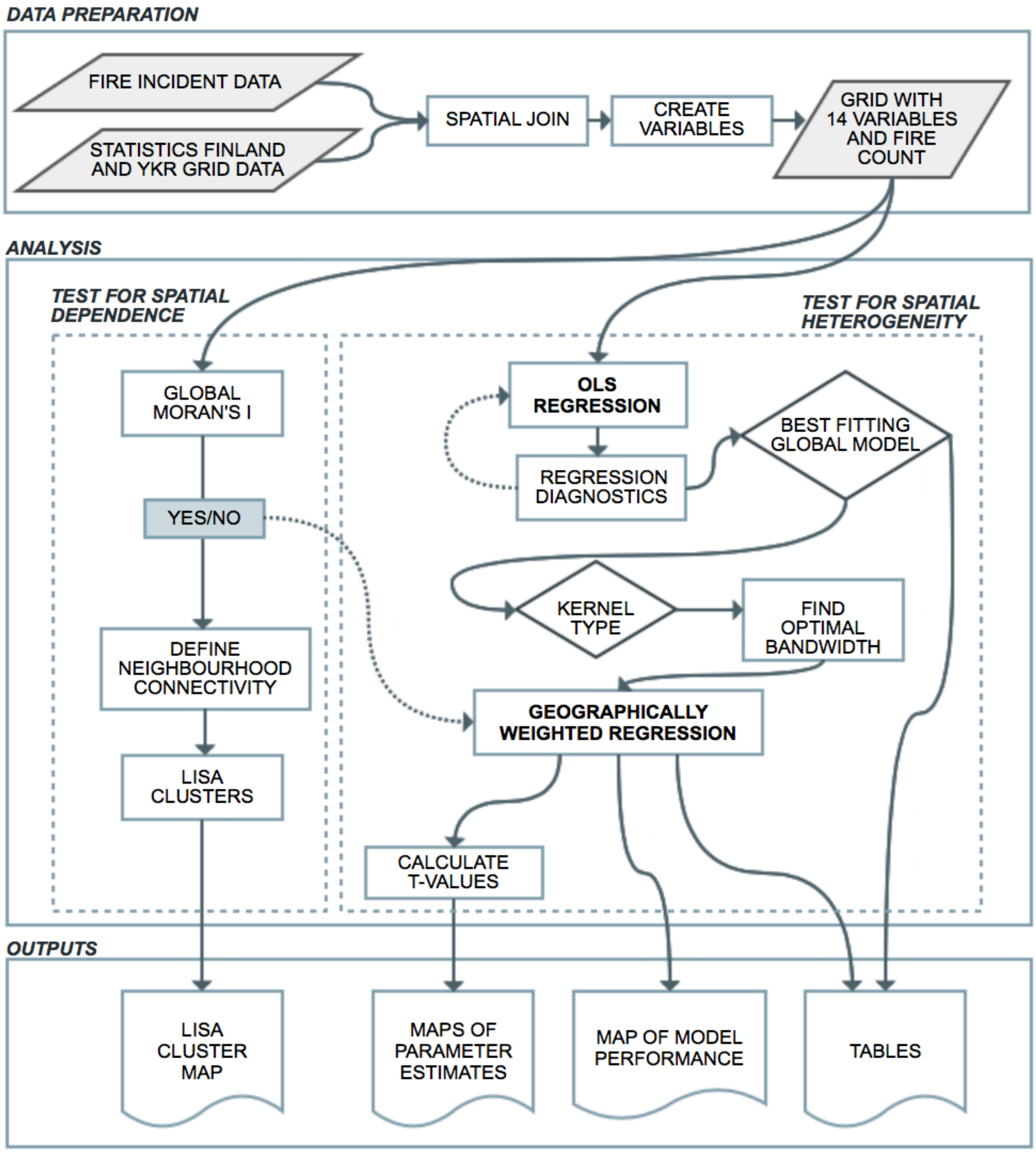

Figure 8. Workflow of the study.

Overall, the workflow consisted of data preparation, analysis, and creation of outputs. Data preparation included joining the datasets and creating a grid for analysis. The analysis phase was divided into two sub-phases: 1) testing for spatial dependence with a spatial autocorrelation analysis to answer the first research problem and 2) testing for spatial heterogeneity first with a global regression (i.e. Ordinary Least Squares, OLS) and then by a local regression (i.e. Geographically Weighted Regression, GWR) to answer the second and third research questions. The outputs are presented as tables and maps.

Construction of the grid and data pre-processing were done with custom Python scripts under an open source JupyterLab environment. Tests for spatial dependence were done in GeoDa (version 1.1.4) and the regression analyses in R Studio (RStudio Team 2015). An R package "spgwr" was used for calibrating the GWR. Detailed information about packages with references used in the analysis is presented in the appendix. Finally, visualizations and maps were done in QGIS 3.4. Section 4.2 introduces methods to test for spatial dependence used in this study, and 4.3 explains the relevant methodology of the GWR.

4.2 Spatial dependence

As Tobler's first law suggests, spatial phenomena are rarely independent, but are rather influenced by neighbouring values that cause the dependent variable to show spatial effects. These effects can be divided into spatial dependence and spatial heterogeneity, where the first is addressed by testing for spatial autocorrelation, i.e. whether observations are clustered in the study area.

To test the assumption of spatial independence of residential fires in the study area, Moran's I index for testing spatial autocorrelation was calculated. Moran's I statistic is a commonly used quantitative method to find whether there is statistically significant clustering in the data, or in other words, whether the values at one location are surrounded by similar values (Moran 1950). A Global Moran's I statistic calculates the degree of spatial autocorrelation. The value of Global Moran's I ranges from -1 to +1, where -1 indicates perfect dispersion of values, a zero value that values are randomly distributed, and +1 perfect positive autocorrelation, meaning that the values are clustered (Moran 1950).

While Global Moran's I statistic tells the strength of the spatial autocorrelation, it does not tell where the clustering occurs. Local clusters can be detected with Local Moran's I statistic or Local Indicators of Spatial Association (LISA), which points out the areas with significant local clusters with either similar or dissimilar values around the observations (Anselin 1995).

Prior to testing the null hypothesis of no spatial association with Moran's I statistics, neighbourhood connectivity was defined for each observation. There are many ways to define neighbourhood connectivity, for example based on contiguity (i.e. spatial units share a common edge or vertices), k-nearest neighbours (i.e. user defined number of neighbours), or distance (Anselin & Rey 2014). Distance-based weighting was selected to form a neighbourhood connectivity matrix. In distance-based weighting, polygons within a defined distance from the observed location are considered its neighbours. In this study, a distance of 300 meters was selected as it minimized the number of grid cells without neighbours and was small enough to imitate a neighbourhood in reality.

Global Moran's I and LISA statistics were calculated in GeoDa (version 1.1.4), an open source software for spatial analysis. The statistical significance of results was assessed with a pseudo p-value, which indicates how likely it is that the autocorrelation is different from randomness.

4.3 Geographically Weighted Regression

4.3.1 Definition and spatial weights matrix

Geographically Weighted Regression (GWR) is an exploratory spatial data analysis tool which addresses the issue of spatial heterogeneity (Brunsdon et al. 1996; Fotheringham et al. 2002). The GWR is an extension of the traditional regression, such as Ordinary Least Squares (OLS), with the difference that where the OLS takes the whole study area as one regression point, the GWR calculates regression coefficients for each individual data point separately. It is used to identify local variations in the relationships between the dependent variable (e.g. fire risk) and explanatory variables (e.g. underlying spatial characteristics). In other words, the method is used to find out aspects from the data that might otherwise be missed (Brunsdon et al. 1996).

As an extension of OLS, the definition of GWR follows the logic of a traditional regression (Brunsdon et al. 1996; Fotheringham et al. 2002; Charlton & Fotheringham 2009). In the formal definition of GWR, for each location u = 1, …, n, the GWR model is:

$$y(u) = \boldsymbol{\beta}(u)\boldsymbol{X}(u) + \boldsymbol{\varepsilon}(u),$$

where $y(u)$ is the dependent variable at location u, $\beta(u)$ is the regression multiplier at u marked as a column vector, $X(u)$ is a row vector of explanatory variables at location u, and $\varepsilon(u)$ is the random error at location u. The term u represents the location as a pair of coordinates. Assuming that nearby observations have more of an effect near the estimated location, the estimator for calibrating the parameter coefficients at each location u is done with a weighted least squares regression:

$$\hat{\beta}(u) = (\boldsymbol{X}^T\boldsymbol{W}(u)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u)y,$$

where $W(u)$ is a diagonal $n \times n$ geographical weights matrix calculated for each location u.

The weights matrix is defined before the analysis and is used to describe the spatial connectivity and the possible interdependence between all n observations at each location u. The weights matrix uses a kernel function that puts more weight into observations close to the observation at location u than to more distant ones. Defining the spatial weights is the essence of the GWR and allows revealing local differences by weighting the observations. Typically, this weighting function is a Gaussian curve which is fitted over each regression point, and the weight decreases with the distance from this point.

There are generally two different kernel functions, a fixed kernel and an adaptive kernel (Fotheringham et al. 2002). In a fixed kernel, the distance is the same around each regression point but the number of nearest neighbours under each kernel varies. In an adaptive kernel, the distance can vary, but the number of neighbours stays constant. The difference between the kernels is illustrated in *Figure 9*. Often an adaptive kernel is preferred if sample points are unevenly distributed in the study area, as it adopts to the differences in the density of the data points (Fotheringham et al. 2002; Charlton & Fotheringham 2009). However, as the study area consists of an almost uniform grid, a fixed Gaussian kernel was chosen as the spatial kernel in this study.

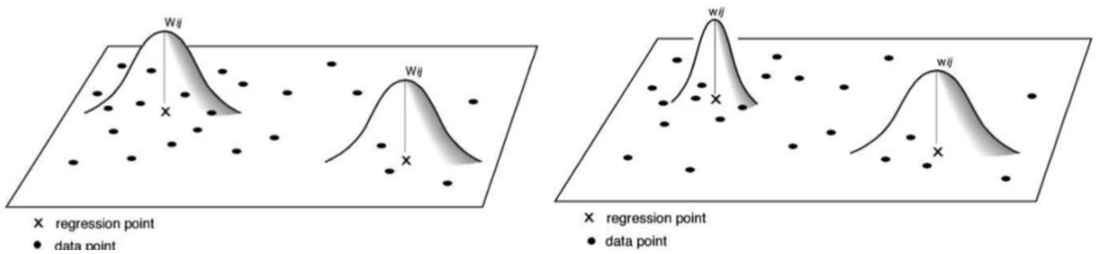

Figure 9. Fixed (left) and adaptive bandwidth kernel (right). Illustration by Fotheringham et al. (2002).

4.3.2 Selection of bandwidth

Often the problem is to define the scale of the analysis, i.e. the bandwidth or the size of the kernel function. If the bandwidth becomes too large, it would eventually cover the whole study area, turning the local GWR model into a global OLS model (Charlton & Fotheringham 2009). Thus, multiple algorithms exist for finding the optimal bandwidth of the spatial kernel, or it can be defined manually.

In this study, the optimal bandwidth was found with a function from the “spgwr” R-package which searches for the optimal value by cross-validating different bandwidths. The Akaike Information Criterion value (AIC) (Akaike 1974) was used as the optimization criterion to find the best fitting bandwidth. The bandwidth of the kernel function is expressed in the same units as the coordinates of the dataset. As the data is in UTM coordinate system, the bandwidth is in metres.

4.3.3 Outputs of the GWR

If the dependent variable is present for all locations in the analysis, the output of the GWR is equal to that in traditional regression, but for each location separately (Fotheringham et al.

2002). The usual set of regression diagnostics and statistics include parameter estimates, standard errors and t-values, as well as measures for model fitness.

For global multiple regression models, the goodness of fit is evaluated with the R-squared value (R^2), which is the proportion of how much of the variation the model can explain with the existing variables in the model. The adjusted R^2 adjusts itself to the number of variables in the model, and the value is often preferred in assessing model fitness of a multiple regression model or if multiple models are compared (Charlton & Fotheringham 2009).

In the GWR, goodness of fit is often assessed with the Corrected Akaike Information Criterion (AICc) (Hurvich et al. 1988). The lower the AICc value, the better fit the model is, and the better it can estimate the reality. The AICc value is relative, thus the value matters only in comparison between the models. In fact, its major benefit is that it can be used to compare the GWR model with OLS (Charlton & Fotheringham 2009). The GWR output also generates a global pseudo-R^2 value. Since each regression point has its own R^2 value, the values can be mapped in a GIS to indicate how well the model fits at each location separately.

Another way to estimate model performance is with t-values, which denote the relationships between the explanatory variables and dependent variable. Absolute t-values greater than 1.65, 1.96, and 2.58 are used to represent the confidence levels of 90%, 95%, and 99%, respectively. In a global model output, the significance of t-values is computed as a p-value, representing the statistical significance of the variable in the model. However, in the GWR output, local parameter estimates can show great variation in different areas, making it hard to evaluate their significance at a general level (Charlton & Fotheringham 2009). In fact, the same variable can be more significant in one regression point, but insignificant in another.

In this study, the approximate significance of the parameter estimates was evaluated by calculating t-statistics for each parameter estimate at each location, and by finding a cut-off value for 90% confidence level based on the degrees of freedom in the GWR model, which also makes it only an exploratory way to assess the statistical significance. A visualization approach suggested by Mennis (2006) was used, where the insignificant values were masked out from the output map leaving only the significant values visible.

5 Results

5.1 Spatial dependence

Global Moran's I statistic with 999 random permutations was calculated for fires to test for spatial autocorrelation. The global value for Moran's I is 0.34, indicating positive spatial autocorrelation. The result means that high fire incident values are surrounded with similar values and the spatial distribution of fires is more clustered than would be if the underlying spatial processes were random. The pseudo p-value is $p < 0.001$ indicating that significant clustering exists.

Supported by the Global Moran's results, LISA statistics with 999 random permutations were calculated for fires, and the clusters are mapped in *Figure 10*. The map shows the significant ($p < 0.05$) clusters of areas with high fire incident values surrounded by high values (high-high), areas with high values surrounded by low values (high-low), areas with low values surrounded by high values (low-high), and low values surrounded by low values (low-low).

The distribution of residential fires in Helsinki shows evident local differences. A large high-high cluster can be found in the inner-city area around the Central Rescue Station (see *Figure 4* for reference). Other areas with concentrated high-high values can be found in the south parts of the study area near the Erottaja station and smaller ones in eastern Helsinki around the Mellunkylä station. A cluster of high-high values means that also the neighbouring areas have significantly more fires. High-low and low-high values are more scattered in the study area and do not form any larger patterns, and no clusters of low-low values exist in the study area. Some high-high areas are surrounded with low-high values like in the inner-city area.

The results indicate that residential fires are indeed spatially non-stationary, and exhibit clustering in Helsinki with significant high-high clusters around the inner city with smaller hotspots in east and north east. Thus, the null hypothesis of no spatial association can be rejected. As the calibration of the GWR is reasonable only if spatial structure can be found in the data, the result justifies the use of the GWR to examine spatial heterogeneity.

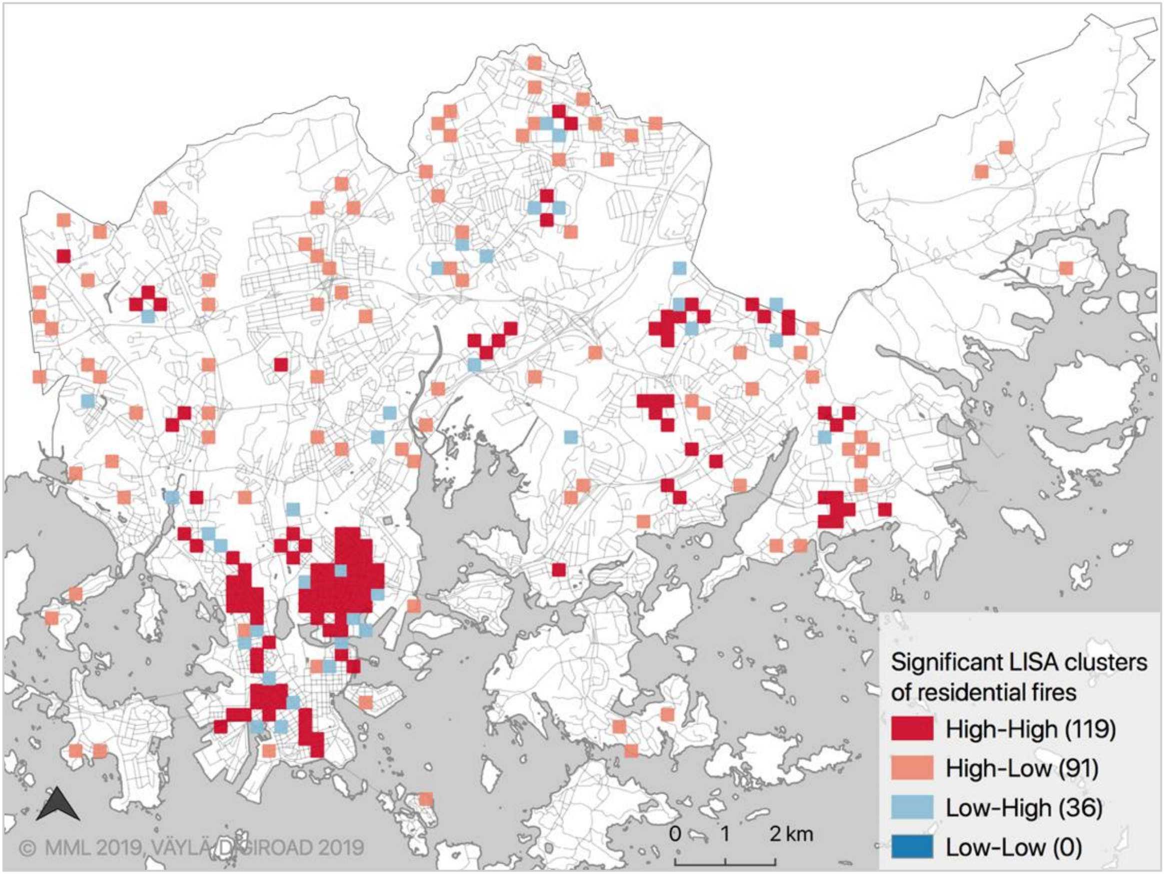

Figure 10. Significant LISA clusters of residential fires (p <0.05). The pseudo p-value was calculated with 999 random permutations of the data.

5.2 Global regression

5.2.1 Full model

First all the 14 explanatory variables were regressed against the dependent variable in a global OLS model. *Table 4* shows the association between residential fire risk and the selected variables. The global R^2 for the full model is 0.47 and adjusted R^2 is 0.47, meaning that the model does not explain 53% of the variation in the study area.

Not all variables are significant at global level. Only population density, proportion of low education, unemployment rate, residential building density, and building year are statistically significant in 95% confidence level ($p < 0.05$).

Table 4. Summary statistics of the full Ordinary Least Squares model.

Variable	Estimate	Std. Error	t-value	p-value[1]	VIF[2]
Intercept	12.6937	4.6012	2.759	0.006 **	---
Population density	0.0002	0.0000	30.464	0.000 ***	1.5
Residential building density	-0.0003	0.0001	-2.121	0.034 *	1.5
Building year	-0.0086	0.0020	-4.289	0.000 ***	1.3
Education					
Low education	1.4467	0.4725	3.062	0.002 **	3.6
High education	-0.5906	0.4120	-1.433	0.151	4.0
Income	-0.0000	0.0000	-0.310	0.756	4.7
Unemployment	4.0249	0.9903	4.064	0.000 ***	2.0
Household crowdedness					
Occupancy rate	0.0089	0.0059	1.516	0.12964	2.8
Room rate	0.1615	0.1336	1.209	0.227	2.1
Family structure					
Households with children	2.1925	2.1907	1.001	0.317	113.2
Adults	2.6914	2.2378	1.203	0.229	173.4
Pensioners	2.1702	2.2249	0.975	0.329	107.2
Household tenure					
Owner-occupied dwellings	1.0601	1.1087	0.956	0.339	167.7
Rented dwellings	0.7244	1.0921	0.663	0.507	170.8

[1]p <0.05 (*), p<0.01 (**), p<0.001 (***)

[2]VIF values above 5 indicate severe multicollinearity.

Variance Inflation Factor values (VIF) of the variables indicate that severe multicollinearity exists between some of the model's variables. Multicollinearity occurs when the independent (explanatory) variables in a regression model are correlated. Multicollinearity can be a problem, because independent variables should be independent, and not correlated, otherwise the assumptions of a traditional regression model are violated. VIF values above five indicate severe multicollinearity where the coefficients are poorly estimated, and p-values cannot be trusted (James et al. 2014). As is visible from *Table 4*, all variables in the family structure and household tenure groups have very large VIF values that need to be addressed before further interpretation of the OLS results.

5.2.2 Addressing multicollinearity

A correlation matrix was constructed to further assess highly correlated variables (*Figure 11*). The significance of correlation is presented with colours, where the insignificant correlations in 95% confidence level are coloured blank, significant positive correlation in red, and negative in blue, respectively. As some of the variables are categorical or by logical terms linked to each other, such as education and income levels, strong correlations between some of the variables were expected.

The strongest positive correlations are between median income and high education, median income and households in owner-occupied dwellings, and between the average floor area and room rate. Results are expected given that high education usually leads to better paying jobs, and owning an apartment is tied to the level of income. Similarly, with an increase in the average number of rooms per resident, obviously also the floor area per resident increases.

The strongest negative correlation is found between owner-occupied and rented dwellings with perfect negative correlation, meaning that their relationship is always negative. Also, high education and low education, and median income and low education show strong negative correlation. Building year shows insignificant correlation with almost all variables except a weak one with education level.

Variables that have either very strong positive or negative correlation show that there is multicollinearity in the independent variables. In this case median income and high education, owner-occupied and rented dwellings, as well as high and low education, should not exist in the model at the same time, because their correlation value exceeds the threshold of 0.75 used in previous fire risk studies (Oliveira et al. 2012; Song et al. 2017).

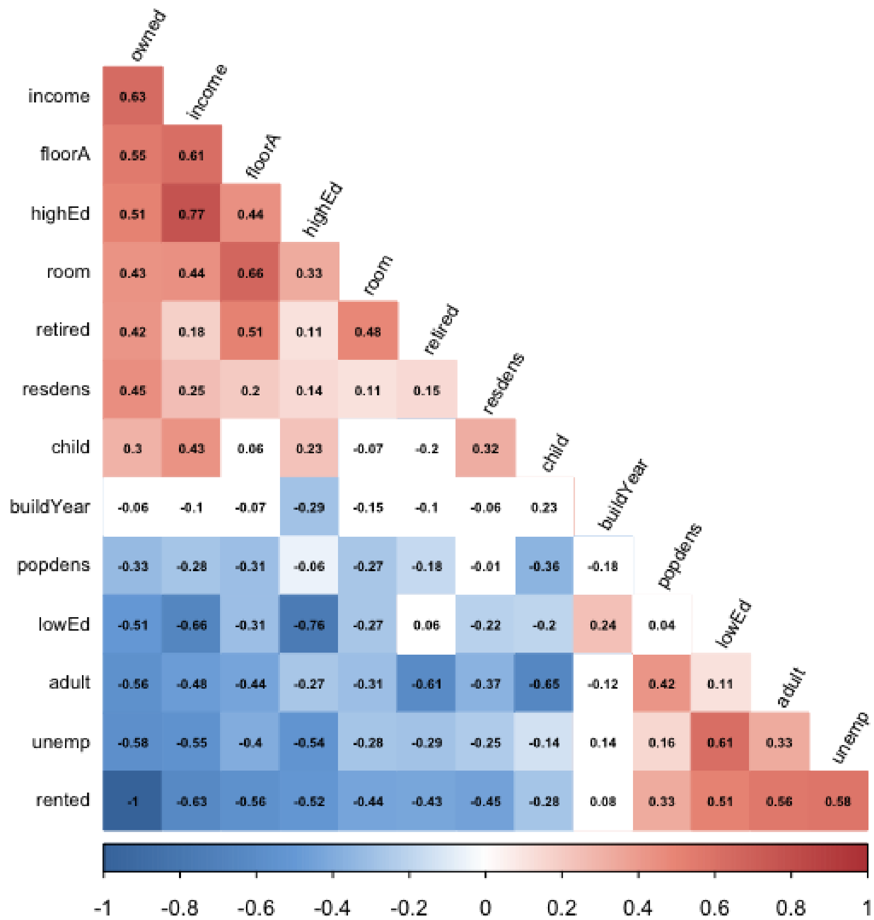

Figure 11. Correlogram of intercorrelations of independent variables. Positive correlations are coloured in red and negative in blue. Insignificant correlations in 95% confidence level are coloured blank.

5.2.3 Simplified final model

The complexity of the model was simplified step by step by removing the variable with the smallest t-value (i.e. largest p-value), in order to increase the model's reliability. Finally, the best fitting model was found by finding the model with the lowest AICc value and without affecting the model's predictive power (i.e. the adjusted R^2). As a result, proportion of high education and households living in owner-occupied dwellings became significant variables. However, although the correlation matrix in *Figure 11* indicated that high and low education are highly correlated, the model's performance was weaker if either of the variables were removed. This might suggest that the model managed to separate the two variables.

VIF values of all the remaining independent variables were also decreased to below five. The coefficient of determination and adjusted R^2 remained the same in the simplified OLS, both being 0.47. The simplified model did not improve the variation accounted for by the models, but it is simpler and a better fit with a lower AICc value. Summary of the simplified OLS is presented in *Table 5.*

Table 5. Summary statistics of the simplified Ordinary Least Squares model.

Variable	**Estimate**	**Std. Error**	**t-value**	**p-value[1]**	**VIF**
Intercept	17.5588	3.7243	4.715	0.000 ***	---
Population density	0.0002	0.0000	33.860	0.000 ***	1.3
Residential building density	-0.0004	0.0001	-2.881	0.004 **	1.3
Building year	-0.0093	0.0019	-4.990	0.000 ***	1.1
Low education	1.1234	0.4205	2.672	0.007 **	2.9
High education	-0.8628	0.3517	-2.453	0.014 *	2.9
Unemployment	4.3926	0.9532	4.608	0.000 ***	1.9
Occupancy rate	0.0107	0.0045	2.397	0.017 *	1.6
Owner-occupied dwellings	0.2662	0.1351	1.971	0.049 *	2.5

[1]p < 0.05 (*), p < 0.01 (**), p < 0.001 (***)

All variables in the simplified OLS are statistically significant in 95% confidence level ($p<0.05$). Both positive and negative global estimates exist in the OLS. Positive estimates are population density, low education, unemployment, occupancy rate, and owner-occupied dwellings meaning that with a unit increase in those variables, fire risk increases. On a global scale, residential building density, building year, and high education have a decreasing effect on fire risk.

Population density is a positive and significant factor with the highest t-value and $p < 0.001$. Residential building density, however, has a negative effect on fire risk, indicating that if the density of residential buildings increases, there would be less fires in the area. Building year has also a negative impact on fire risk, also being highly statistically significant ($p < 0.001$).

On a global scale, education level of the residents plays a significant role in fire occurrence. With one percentage point increase in the proportion of residents with only basic education, the estimated number of fires would increase by 1.2 fires. Similarly, an increase in highly educated population would decrease the risk of fire by 0.9 fires. Unemployment rate has also a significant positive effect on fire risk with $p < 0.001$. As for the variables related to household status, both occupancy rate and owner-occupied housing have a positive effect on fire risk with $p < 0.05$.

5.3 Geographically Weighted Regression

5.3.1 Model selection and validation

Variables in the simplified OLS were used to fit the GWR. Best fitting GWR model was found by selecting the bandwidth with the lowest AIC value, and the software suggested 1752 metres. By comparing the AICc values of the full OLS model, the simplified OLS model, and the GWR model, the GWR model provided the best fit with the smallest AICc value (*Table 6*). Compared to the full OLS, AICc value decreased by 5.4 for the simplified OLS model and by 108.6 for the GWR model, respectively.

Table 6. Comparison of the full OLS, simplified OLS and GWR with measures for goodness of fit. The best fitting model is bolded.

Model	**Global R^2**	**AICc**
OLS full	0.47	5592.726
OLS simplified	0.47	5587.371
GWR with bandwidth 1752 m	**0.56**	**5484.121**

As the GWR produces a set of regression diagnostics for each location, the explanatory power can also be assessed at each location separately. Local R^2 values are mapped to show how much of the variation the model explains at each location (*Figure 12*). Values in yellow-orange-red indicate values that are higher than the global R^2.

From the map can be seen that the explanatory power of the GWR model varies from 0.30 to 0.68, showing significant local improvement in some areas compared to the global model. Local R^2 values are high in the centre of the study area around the Central Rescue Station and Käpylä and Erottaja stations, spreading to north-east and decreasing gradually towards the outer edges. Yet in some areas, the local R^2 is much lower than the global average. Patches where the local R^2 is significantly lower than the global average are in north-west and north, as well as in some parts in east, south-east, and south-west. It means that the local regression model performs badly in these areas.

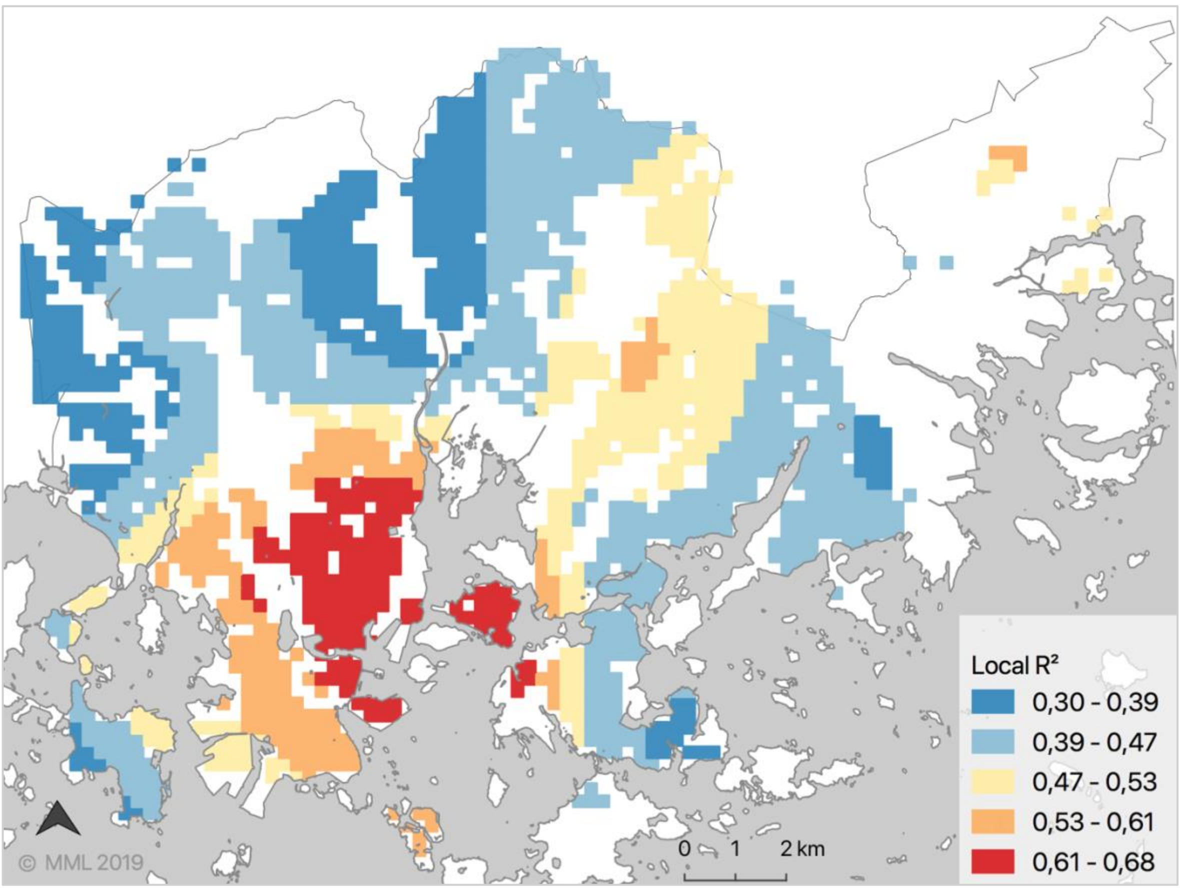

Figure 12. Local R^2 values of the Geographically Weighted Regression. Yellow-orange-red colours indicate locations where the GWR model explains more variance compared to the global model.

5.3.2 Local variation

Variations in the parameter estimates are shown in *Table 7*. For each parameter in the model, the GWR gives the local minimum, median, and maximum value in the study area, as well as the first and third quartiles. For reference, the global estimate is shown in the last column. The table shows that there is large variation in the ranges of the estimates across the study area depending on the location. Only population density has a constant positive effect on fire risk, other variables range from negative to positive.

However, not all of the estimated values are necessarily statistically significant, which is why t-values were calculated for each parameter at each location by dividing the estimate with its standard error. Using a two-tailed t-test, a cut-off t-value 1.646 was found to correspond with at least 90% confidence level ($p <= 0.1$) with 1709.282 degrees of freedom in the GWR model. The insignificant values ranging from -1.646 to +1.646 were masked out, leaving only the areas with the approximated significant estimates visible on the map.

Table 7. Parameter estimates of the GWR model.

Variable	**Min**	**1st Quartile**	**Median**	**3rd Quartile**	**Max**	**Global**
Intercept	-18.9788	6.5783	12.8200	21.8843	48.6382	17.5588
Population density	0.0001	0.0001	0.0002	0.0002	0.0003	0.0002
Residential building density	-0.0022	-0.0007	-0.0004	-0.0002	0.0008	-0.0004
Building year	-0.0255	-0.0113	-0.0070	-0.0038	0.0097	-0.0093
Low education	-1.3611	0.9285	1.3181	1.6773	2.2116	1.1234
High education	-7.1624	-0.9051	-0.3731	0.2897	1.4796	-0.8628
Unemployment	-5.7226	2.6149	3.7479	4.9407	27.8873	4.3926
Occupancy rate	-0.0151	0.0048	0.0149	0.0209	0.0378	0.0107
Owner-occupied dwellings	-1.0454	-0.0814	0.1831	0.5599	1.9451	0.2662

Local parameter estimates are mapped in *Figures 13-20*. All the values were classified into five classes with the Natural Breaks classification scheme. Positive values are coloured in yellow-orange-red and negative values in two shades of blue. Positive values mean that as the value of the explanatory variable increases, so does residential fire risk. Similarly, negative values mean that with a unit increase in the explanatory variable, residential fire risk decreases. Only the significant values ($p <= 0.1$) are visible in the figures.

Figure 13 shows the influence of population density on fire occurrence at local level. According to the figure, the estimates are positive and significant almost everywhere in the study area, except in a few cells in the north-east. This means that an increase in population density increases the residential fire risk almost everywhere in Helsinki, and the effect is the strongest in the inner-city area around the Central Rescue Station, as well as in the north and north-east around the Malmi and Mellunkylä stations.

Local estimates for residential building density are mapped in *Figure 14.* Only negative values are significant, indicating that an increased proportion of residential buildings in a neighbourhood seems to be a protective factor against fire risk. The strongest effect is in the centre of the study area around the Central and Erottaja Rescue Stations, with a clear direction towards north-east while decreasing in intensity.

Also, for building year, only negative values are significant (*Figure 15*). Building year appears to be a protective factor, indicating that the newer the buildings in the area, the less fires are to be expected. The impact of buildings' construction year is significant only in the south and centre of the study area, with the strongest negative impact in a large area around the Central Rescue Station and expanding outwards while decreasing in intensity.

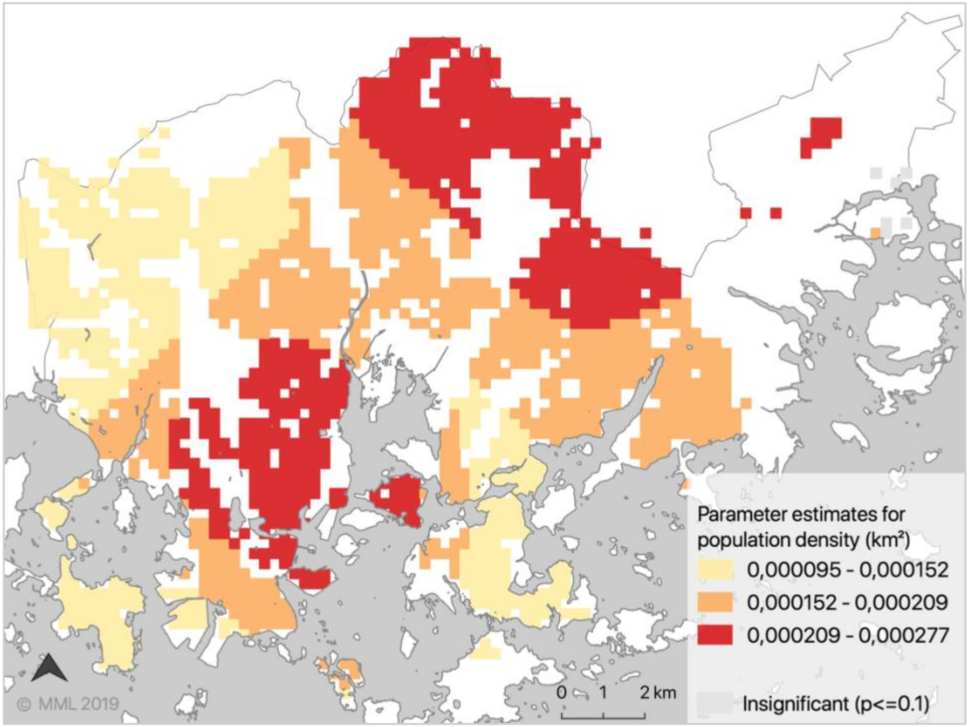

Figure 13. Geographically Weighted Regression local parameter estimates for population density per km².

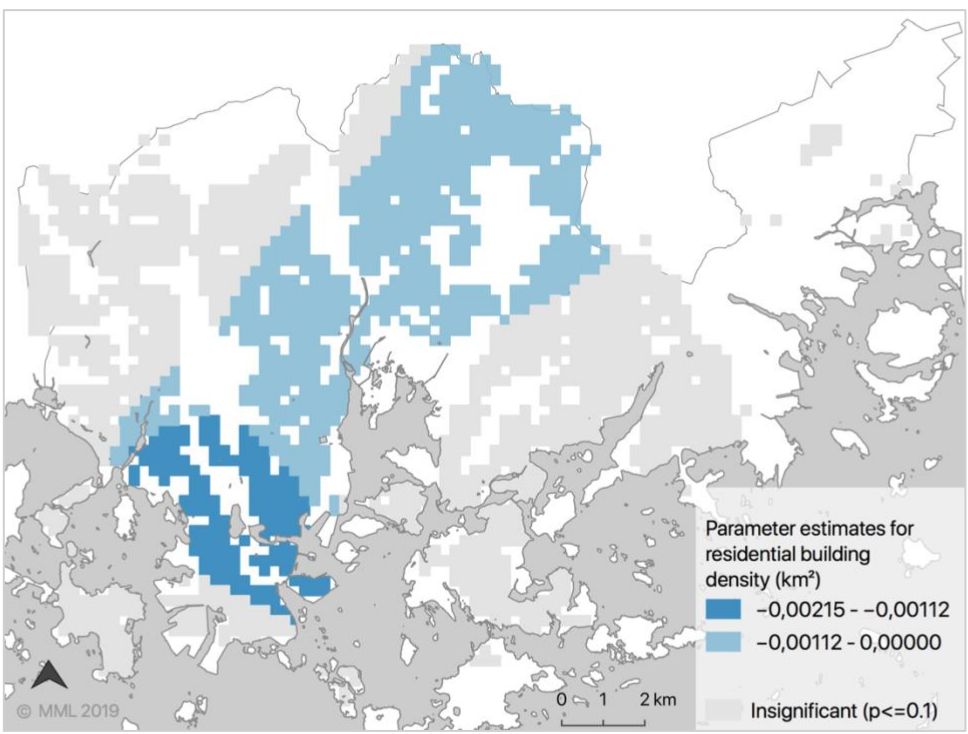

Figure 14. Geographically Weighted Regression local parameter estimates for residential building density per km².

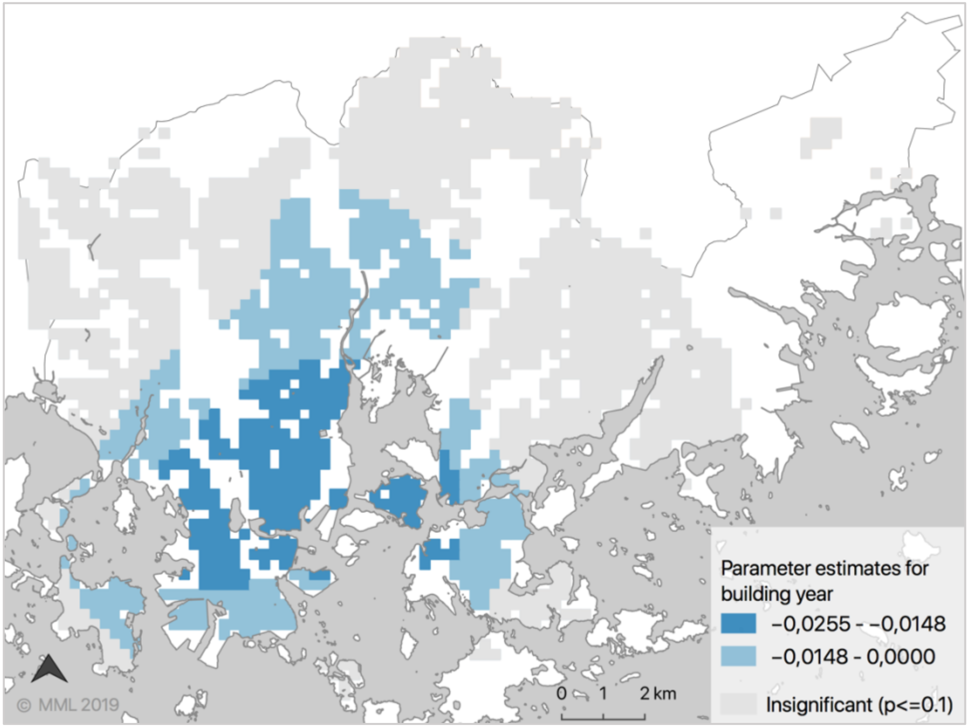

Figure 15. Geographically Weighted Regression local parameter estimates for average building construction year.

Low education and high education show interesting variations in their impact on residential fire occurrence. In *Figure 16*, the local effects of the proportion of residents with only basic level education are mapped. Although the parameter estimates vary from negative to positive, only the values in the two highest positive categories are significant. Two clear separate areas are revealed where the proportion of low education has a strong positive influence on residential fire occurrence; one in the western side of the study area north from the Haaga Rescue Station, and the other between the Herttoniemi and Mellunkylä stations in the east.

Figure 17 shows the effect of the proportion of highly educated residents on residential fire occurrence. For high education, only negative values are significant. Showing contrasting patterns with low education, high education seems to be a protective factor especially in the central parts of the study area around the Central and Erottaja Rescue Stations. This indicates that in these areas, if the proportion of highly educated residents increases, the fire risk decreases dramatically.

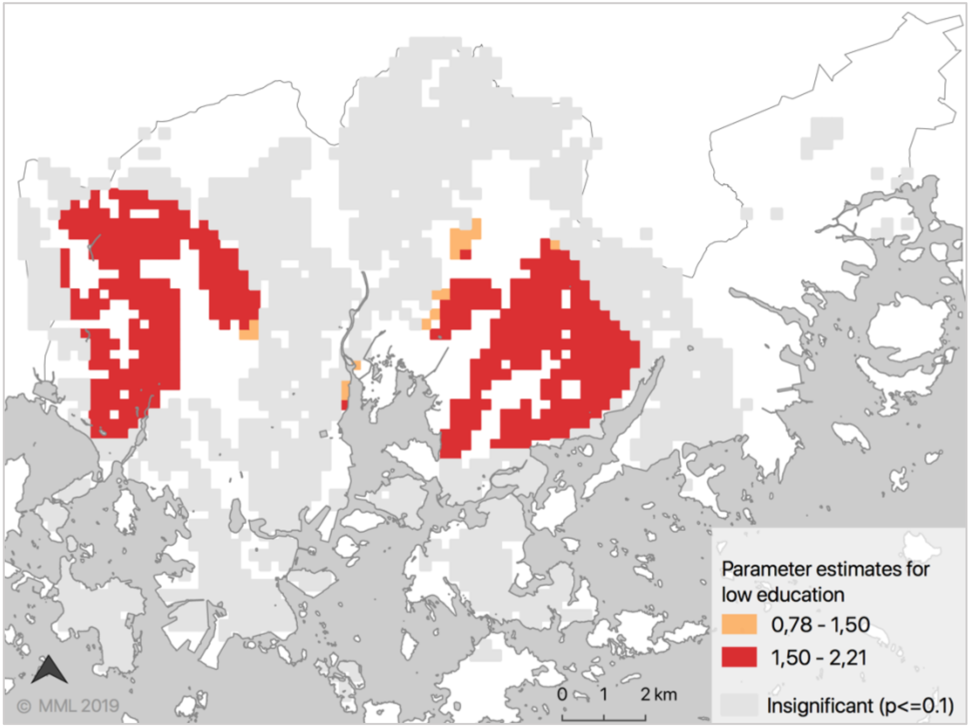

Figure 16. Geographically Weighted Regression local parameter estimates for proportion of residents with only basic level education.

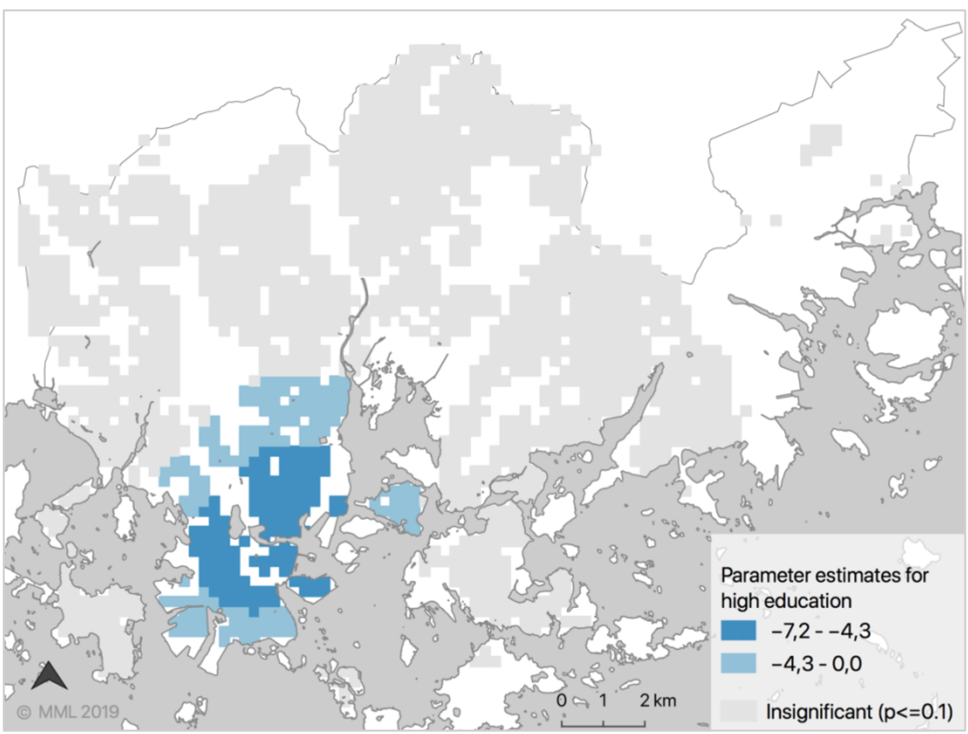

Figure 17. Geographically Weighted Regression local parameter estimates for proportion of residents with high education.

Figure 18 shows that the positive influence of unemployment rate on fire risk is significant on a large area, with the strongest effect in the south parts of Helsinki around the Central and Erottaja Rescue Stations. The estimated values decrease and move towards north and north-west. Only one grid cell in the north-east has a significant negative effect on fire occurrence.

Occupancy rate, presented in *Figure 19*, statistically has only a positive effect on fire occurrence in the study area. The effect of the average floor area is very directional with the strongest effects in the eastern part of the city centre around the Central and Erottaja Rescue Stations, and in a large area in the north-east surrounding the Mellunkylä station. In those areas, an increase in the average floor area per resident means a higher risk of fire.

Lastly, *Figure 20* shows that living in owner-occupied dwellings appears to have both significantly positive and negative effects on fire risk. The strongest positive effects are around the Central Rescue Station and west from the Erottaja station, with a clear visible direction towards north. In contrast, higher proportion of households in owner-occupied dwellings seems to mitigate the fire risk in the west in an area north from the Haaga station.

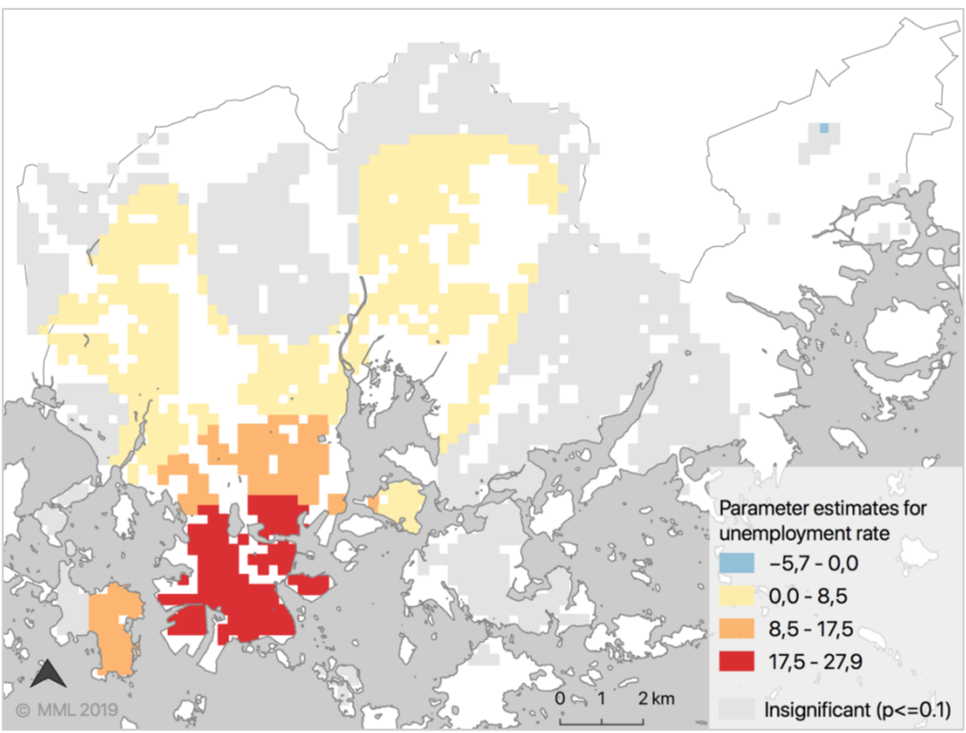

Figure 18. Geographically Weighted Regression local parameter estimates for unemployment rate.

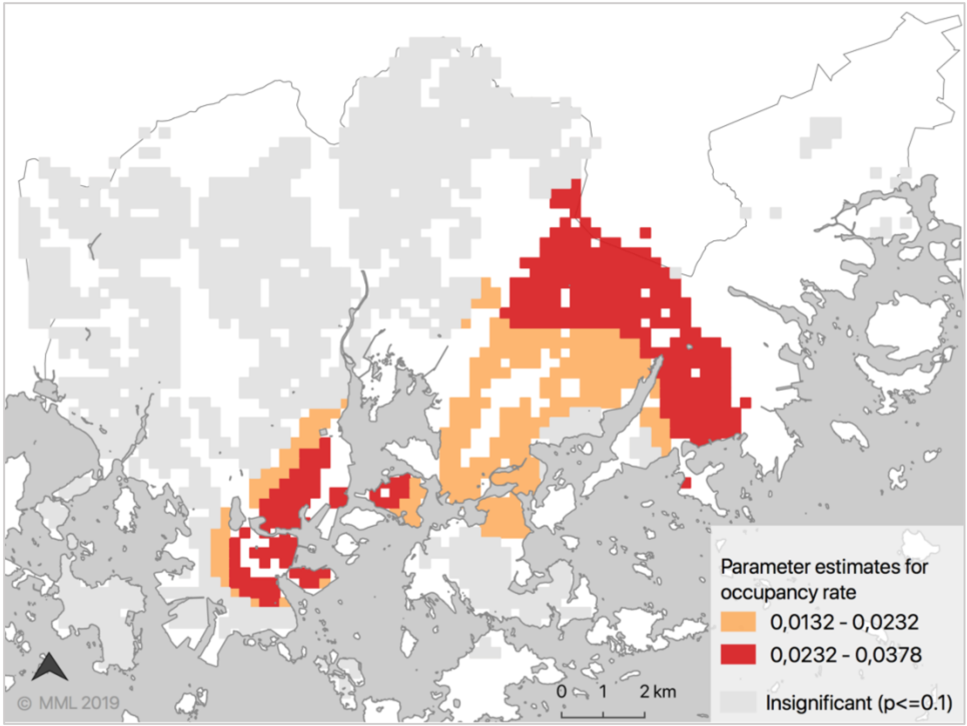

Figure 19. Geographically Weighted Regression local parameter estimates for occupancy rate (average floor area in m2/resident).

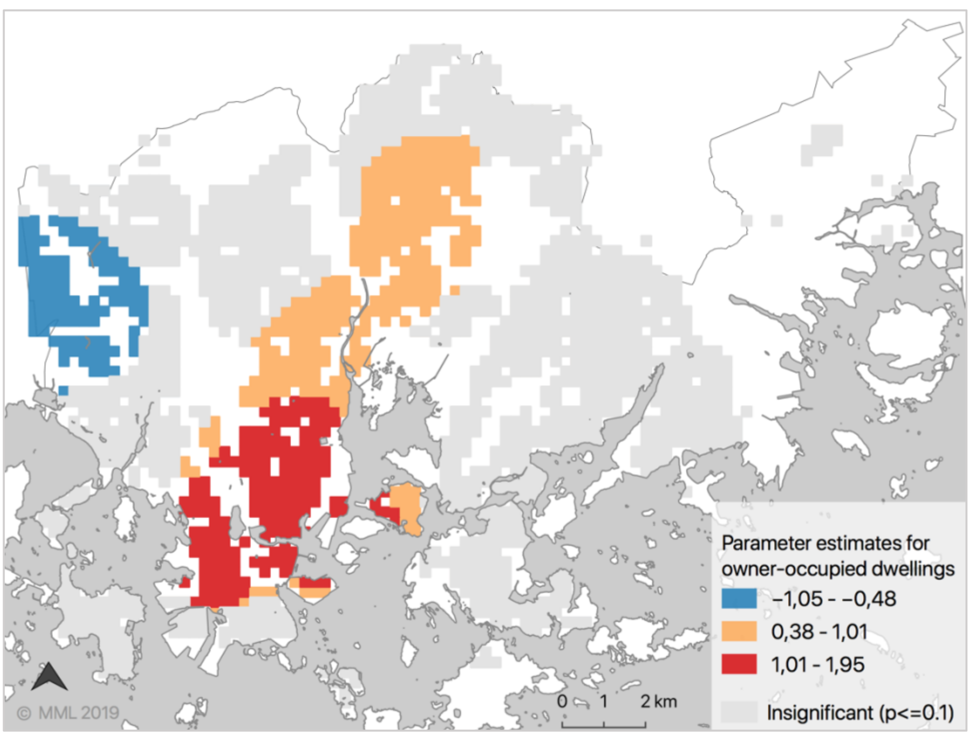

Figure 20. Geographically Weighted Regression local parameter estimates for households in owner-occupied dwellings.

6 Discussion

6.1 Factors associated with residential fire risk

6.1.1 Key explanatory variables

The results of this study show that residential fires exhibit spatial effects in Helsinki, and that neighbourhood and structural characteristics, socioeconomic status, and household circumstances are strong determinants of residential building fire incidence on a neighbourhood scale. The first key finding is that residential fires are spatially clustered, and some neighbourhoods seem to have more fires than others. In addition, several risk factors increasing residential fire risk were found, but also factors associated with a decreased risk of fire were identified (*Table 8*). To be exact, variables that increase residential fire risk are population density, proportion of low education, unemployment rate, occupancy rate, and proportion of households in owner-occupied dwellings, which, in some areas, also decreases fire risk. In addition, a decreasing effect on fire risk was found with residential building density, building year, and proportion of high education. Furthermore, the Geographically Weighted Regression (GWR) model was able to identify significant local variation in the effects of the explanatory variables on fire risk.

Table 8. Factors associated with an increased and decreased residential fire risk in Helsinki at neighbourhood (250 x 250 m) level.

Increased residential fire risk	**Decreased residential fire risk**
High population density (per km²)	High residential building density (per km²)
Low education (proportion of residents aged above 18 with only basic level studies)	Newer buildings (average building construction year)
Unemployment (proportion of unemployed residents aged 15-64 of residents aged 18 or above)	High education (proportion of residents aged above 18 with an academic/higher-level university degree)
More living space per occupant (occupancy rate/average floor area in m² per resident)	Home ownership (proportion of households living in owner-occupied dwellings)
Home ownership (proportion of households living in owner-occupied dwellings)	

At neighbourhood level, population density was found to have the most significant and positive effect on fire risk throughout the study area, which is in line with previous studies from Helsinki (Tillander et al. 2010; Špatenková & Virrantaus 2013). The result was expected since residential fires happen at people's homes, and a high concertation of people is closely related to higher fire rates (Wallace & Wallace 1984; Corcoran et al. 2007). Population has been considered the most important variable in risk analysis ar the Finnish rescue services and the variable is also a key factor in the recent building fire model in Finland (Krisp et al. 2005; Tillander et al. 2010). Thus, the result confirms the strong relationship between population and residential fires, which is an important factor to acknowledge as the population is estimated to grow in Helsinki from 2018 with over 100 000 residents by 2034 (Vuori & Kaasila 2019).

While the results for population density were expected, results for residential building density were contradicting with a few recent studies. In this study, residential building density had a moderate but negative effect on fire incidence especially in the inner-city area, around the Central and Erottaja Rescue Stations. This result contradicts with the results of Ardianto (2018) in Australia and Ceyhan et al. (2013) in Turkey, who found higher residential building density to increase the fire risk. In this study, the result is most likely connected to the historical structure of Helsinki, where the highest densities of population and thus residential buildings can be found in the most prosperous and old downtown areas. In other words, high population density and high residential building density can also serve as indicators of the area's well-being in terms of income, education level, employment status, etc. On the scale of this study, the density of residential buildings in a neighbourhood hardly correlates with fires as such, although, as an independent phenomenon it could be even positively related to fire risk. However, in Helsinki, the negative statistical relationship between residential building density and fire risk is most likely explained with the connection between more prosperous downtown areas and high densities of people and residential buildings.

Higher likelihood of urban fires has been associated also with building characteristics (Goodsman et al. 1987; Shai 2006; Špatenková & Virrantaus 2013; Vasiliauskas & Beconytė 2015). In particular, older buildings have found to be risk factors for higher risk for occupants to die in a fire (Runyan et al. 1992; Xiong et al. 2015). In this sense, findings in this study partly support earlier findings, since newer buildings seem to decrease fire risk in Helsinki. In fact, building year was found highly statistically significant ($p < 0.001$) especially in the inner-city area around the Central and Erottaja Rescue Stations, where not only the fire density is the highest, but where also many of the oldest buildings in Helsinki are located (see *Figure 6a* for reference). In older buildings, the fire safety and fire-resistant building materials can be insufficient, while in newer buildings these aspects are most likely better considered. However, to better investigate the effect of each different building ages on fire risk, the variable for building year in this study could have been categorical, and classified into old, middle-aged, and new buildings. Given this shortcoming of the variable, the result still reveals that newer buildings in a neighbourhood provide protection against fires, either through better dwelling materials or due to the fact that the residents living inside are less likely to be exposed to fire.

Socioeconomic status, and especially socioeconomic deprivation, have been steady factors associated with urban residential fires at neighbourhood level (FEMA 1997; Duncanson et al. 2002; Chhetri et al. 2010; Corcoran et al. 2011a; Corcoran et al. 2011b; Jennings 2013). Many earlier studies have used a composite “social index” to measure the levels of social deprivation, combining low education, income, unemployment, and other variables into one variable (e.g. Duncanson et al. 2002; Corcoran et al. 2007; Chhetri et al. 2010; Guldåker et al. 2018). This study, however, used separate socioeconomic variables to study their relationship with residential fires in Helsinki. Income was not found statistically significant, although it has been consistent in many earlier studies (Gunther 1981; Duncanson et al. 2002; Holborn et al. 2003; Špatenková & Virrantaus 2013). Apart from income, all other socioeconomic variables were globally significant and showed considerable variation at local levels, supporting findings in earlier fire studies.

The results regarding socioeconomic variables are interesting also because Špatenková and Virrantaus (2013), who previously studied residential fires in Helsinki, found fire incidence to be significantly affected by income, but not education level or unemployment rate. The authors used census data by Statistics Finland collected in 2006 (Špatenková & Virrantaus 2013). Differences in results might reflect how income is closely connected with other socioeconomic variables, such as unemployment and levels of education (FEMA 1997). This relationship is also visible in the correlation matrix in this study, and the high correlation can be a reason why the income variable was ultimately excluded from the full model.

On the other hand, since 2000 it has been noted that disparities between advantaged and disadvantaged neighbourhoods in Helsinki have been slightly widening (Kortteinen & Vaattovaara 2015; City of Helsinki 2019b). According to the City of Helsinki, this development has been due to faster welfare development in already advantaged areas, while at the same time, deprivation has become increasingly multi-layered with low income, low levels of education, and unemployment concentrating in certain districts (City of Helsinki 2019b, 26). While it is still unclear how this socioeconomic and structural development in Helsinki has affected the manifestation of the factors affecting fire incidence in the long term, it is certainly important to account for while planning the future response readiness and resource use at the rescue services.

Variations in education levels have been found to affect fire rates also in earlier studies (Duncanson et al. 2002; Corcoran et al. 2007). In this study, low education was found to have a globally significant positive effect on fire risk, but at the same time, high education acts as a protective factor against fire occurrence at neighbourhood level. Both variables also show significant local variations in separate areas. The result for high education is supported by the results of Duncanson et al. (2002), but it contradicts with the results of Nilson et al. (2015), who found high education level of residents increasing the risk of fire occurrence at individual level. This suggests that the behaviour and properties of individuals can differ from the neighbourhood average. A possible reason for a decreasing effect could be that with a higher education level, residents would ideally also have higher incomes, and thus better homes equipped with fire-resistant materials, and be more self-educated and aware of fire-related risks. Similarly, poor education level is tied to lower income – as the correlation matrix in this study revealed – which in turn can act as an indirect factor leading to risky behaviour and an increased exposure to fire.

Unemployment rate was found a highly significant positive parameter at global level and also in a very large area at local level. The result supports previous findings where unemployment rate was connected with higher fire incidence rates (Gunther 1981; FEMA 1997; Chhetri et al. 2010; Hastie & Searle 2016). As most of the residential fires have happened around areas near the Central and Erottaja Rescue Stations, the results suggest that by increasing the proportion of highly educated people and by decreasing the unemployment rate in these areas, a lot of fires could possibly be prevented. In one grid cell in the north-east, unemployment rate had a decreasing influence on fire risk. Although conclusions cannot be drawn from only one grid cell, the result implies that the effect of unemployment on fire risk is not unambiguous. In fact, there is some uncertainty related to the unemployment variable, as also students are counted into this variable. From this variable it is impossible to distinguish, for example, unemployment due to full-time studies from general unemployment, which in turn can be an indicator for poor life management acting as the basis for a higher fire risk.

The results for household crowdedness contradict with earlier studies, where crowdedness – or having less living space per resident – was found to increase the fire risk (Gunther 1981; FEMA 1997; Duncanson et al. 2002; Nilson et al. 2015). On the contrary, while the original assumption was that the effect of household crowdedness acts as a risk factor, the results revealed that having more living space actually increases the risk of fire incidence, especially in east from the inner-city towards the Mellunkylä station area. However, genuine household crowdedness is impossible to measure at aggregate level since the original definition by WHO (2018) considers sex, age and marital status of the residents in a household. Therefore, it is possible that the variable measures something else than household crowdedness. The positive result can be connected, for example, with risks associated with living in larger homes that have more possible sources of fire ignition (e.g. more electronic appliances etc.) or with other unknown reasons associated with individual and household characteristics.

The last significant variable found in this study was owner-occupied housing, and the results are interesting especially at local level as it has a both positive and negative effect on fire risk. Home ownership was found to decrease fire risk in earlier studies at neighbourhood level (Duncanson et al. 2002). The mitigating effect can be related to the positive correlation between owner-occupied housing and the income level and high education of the residents – also apparent in the correlation matrix in this study – which can indicate generally better living standards and higher awareness of risks. Results from a recent study also show that during the past two decades, the connection between income level and owner-occupied housing has strengthened in Helsinki (Saikkonen et al. 2018). Moreover, homeowners can be more aware of the risks in the neighbourhoods due to active interaction and long period of residence, which can act as a mitigating factor (Ardianto 2018).

However, the effect of home ownership as a risk factor has not been that often demonstrated in earlier fire risk studies. There are some possible explanations for the variation. For example, the strong relationship between income and owner-occupied dwellings can be associated with larger houses, garages, saunas, etc.; or, in other words, with more potential sources of ignition leading to a fire. On the other hand, the results also indicate that there is a possibility of ecological fallacy regarding the effect of home ownership as a factor increasing fire risk. That is, the share of owner-occupied housing is often higher in more affordable (i.e. less well-off) areas and therefore it can be an indicator of deprivation at neighbourhood level. Yet, the fire risk can be different for individual households, as the fires may not necessarily occur in the owner-occupied dwellings of those neighbourhoods. Rather they can occur in some other types of housing, which go unnoticed due to the aggregated form of data. As a result, in those parts of the city where owner-occupied housing increases fire risk, it may be linked to lower real estate prices and thus be an indicator of higher degree of socio-spatial segregation.

All in all, home ownership status both as a risk factor and as a risk mitigator implies that the manifestation of the variable across the study area is affected to a large extent by different characteristics and assumptions associated with the variable, and that certain bias is expected due to the nature of aggregated data. For example, in the conceptual model by Corcoran et al. (2011b), dwelling characteristics comprised dwelling materials, equipment, electrical fitting, etc.; all of which are impossible to measure at aggregate level. Thus, in some areas, the combination of these dwelling characteristics can increase the fire risk rather than decrease it, and vice versa.

6.1.2 Complexity and spatial variety of explanatory variables

Altogether, these eight variables account for 47% of the global variation of residential fire occurrence, meaning that over half is not explained. Locally, the model performance reaches up to 68% in the central study area, around the Central Rescue Station. The result can be considered a great improvement to the global model, also because most of the residential fires in Helsinki happened in those areas. Simultaneously, the model predicts fire occurrence poorly around the outer boundaries of Helsinki and especially in the northern and eastern Helsinki. These areas mostly correspond to grid cells with low fire values or no past fire history; in fact, the model performed the best in areas where there is more fire data available. Most importantly, the results highlight how the GWR can find areas where the model works better or weaker, and how the explanatory power of the same variables varies across the city.

The large proportion of unexplained variation indicates that there is possibly a lot of *omitted variable bias* in the model. It means that some important variables are not included in the model, and therefore, some existing variables might act as proxies for others by explaining some of the characteristics of other variables. For example, this could be the case with occupancy rate or household ownership status, as demonstrated above. Many aspects possibly affecting fire occurrence cannot be measured at aggregate level, although their effect on the likelihood of fire might be significant. Hence, we can only postulate what is happening in reality given that some key variables are not included in the model. However, considering the strong role of human activity and behaviour in residential fire occurrence, Merrall (2002) suggests that a lot of unpredictable and arbitrary variance is to be expected. In the context of this study, the overall explanatory power of the models can be viewed as notable.

As a result, while being aware of the complexities and uncertainties behind the explanatory variables, the question arises as to whether we can infer what variables can predict residential fire risk? By definition, causal variables correspond to the theoretical definition of the dependent variable (e.g. fire risk) and can thus be assumed to have a causal relationship with the dependent variable (Bollen & Bauldry 2011). Accordingly, causal factors directly influencing fire risk cannot be inferred because the theory of fire risk is not definite.

While it is easier to assume causal relations at individual level, such as the cause of fire death, causality at neighbourhood level is hard to establish. In fact, in an aggregate level study, drawing conclusions about the relationship between the likelihood of fire and the characteristics of individuals leads to the problem of ecological fallacy (Corcoran et al. 2007; Clark et al. 2015). For example, while in some areas higher unemployment rate increases fire risk significantly, it does not mean that being unemployed directly causes the fire. Similarly, as stated earlier, while in some neighbourhoods it appears that the proportion of owner-occupied dwellings increases fire risk, the fires do not necessarily occur in those dwellings. Thus, correlation at neighbourhood level does not necessarily indicate association at individual level.

To summarize, it is difficult to conceptualize fire risk at aggregate level, as the variables can have many meanings. However, several variables related to socioeconomic and household status as well as population and structural characteristics were found to influence residential fire incidence in Helsinki. As Jennings (1999) concludes in his early review, the precise combination of variables influencing fire risk is not yet known, but different influencing factors are possible to identify. Identifying these local predictors is important especially when planning the response readiness and resource allocation, as finding legalities at the neighbourhood level can help predicting the occurrence of accidents.

6.2 Validity and reliability of the study

Modelling residential fires and the reliability of results depend on the accuracy of fire incident data and other spatial datasets. Thus, interpretation of results should always be done with careful consideration. Currently the fire incidents are uploaded to the PRONTO database by an officer at the incident site. Information is added according to the permanent building code of the building where the incident happened, which also automatically updates the coordinates of the incident into the database. Although there are some known inconsistencies in statistics, building fires have been mostly filed correctly into the PRONTO database (Majuri & Kokki 2010). Further aggregation of the incident points into the grid cells also helped to mitigate the possible precision errors.

In addition, the data consisted only of fires attended by the Helsinki Rescue Services. However, it can be assumed with some certainty that in reality there are considerably more fires in buildings, as some smaller ignitions can be extinguished by the residents themselves. In fact, one study in the US estimated that over 95% of unwanted residential fires were not reported to the fire departments (Greene 2012). Therefore, fires in this study are only a subset of all domestic fires, but they are more likely to be those with a higher probability of personal injury or property damage.

Perhaps the largest issue was the quantification of fire risk. The sample size, spatial scale, and ultimately the research questions drive the operationalization of fire risk. It is well established that residential fires are more frequent in more populated areas, which is why fire count might give a misleading picture of the probability of fire, especially if the sizes of the spatial units are not uniform (Chhetri et al. 2010; Ardianto 2018). In other words, the ignition of a fire in a dwelling or the fact that a person "encounters" a fire is not mediated by this variable. Therefore, some studies have used fires per capita as the dependent variable (e.g. Chhetri et al. 2010). In this study, the difference between fire count and fires per capita was also tested in the regression model. However, the model fitness was bad for fires per capita, as it explained only around 9% of the variation compared to fire count, which explained 47%, respectively. From the result, it can be concluded that the model explained weakly fire risk to residents at neighbourhood level. Identifying factors that contribute to a higher risk of fire to a resident would be helpful especially from accident prevention point of view, but on the other hand, knowing where and how many fires happen is important for planning the response readiness.

However, analysing fires per capita might work on a larger spatial scale, since larger spatial units tend to control the influence of model's outliers, resulting in more explanatory power (Hastie & Searle 2016). Adding more samples from different years or enlarging the spatial units would allow more stable fire rates to be calculated, but at the same time it would change the scale of the analysis both spatially and temporally. For example, five more years of fire data in this study would not have been representing the reality as reliably. In fact, fires were examined as static phenomena, not as time series, thus this study does not account for temporal changes. Therefore, the results are just a snapshot of reality, covering a few years of data.

This study used local analysis methods to examine the spatial patterns and relationships between residential fire and multiple census variables. The assumption prior the analysis was that since the relationship between fire and its influencing factors may vary across space, traditional linear models such as the OLS are not adequate to capture the spatially varying effects of multiple factors and residential fire incidence. Spatial autocorrelation at neighbourhood level and local significant LISA clusters gave enough evidence that spatial structure exists in the distribution of residential fires, and thus justified the use of the GWR as the analysis method.

In this study, the local GWR model provided a better fit compared to the global OLS model, as it was able to address spatial heterogeneity by revealing differences in the effects of explanatory variables on residential fire occurrence in Helsinki. Previous studies on fire risk that have used the GWR all have all confirmed that GWR is a better fit than the global OLS model, as it allows the coefficients to vary across space (Yamashita 2008; Špatenková & Virrantaus 2013; Oliveira et al. 2014; Song et al. 2017; Ardianto 2018).

The main advantage of the GWR is that it allows visualizing local statistics as well as the model’s local explanatory power. On the other hand, a major limitation of the GWR is that it cannot precisely decide on the statistical significances of the parameter estimates. Thus, using the GWR focuses more on data analysis and interpretation rather than on prediction (Oliveira et al. 2014). The results of the GWR model also highly depend on the selection of bandwidth and the spatial weights function, all of which affect the model’s performance and results (Wheeler & Tiefelsdorf 2005; Bidanset & Lombard 2014). In earlier studies on fire risk, different kernel types have been used. The fixed kernel used in this study was selected as it is the same that Špatenková and Virrantaus (2013) used in their study in Helsinki.

Limitations of the GWR also include the problem of multicollinearity in the local parameter estimates. Although the multicollinearity problem was adequately addressed in the OLS, addressing it in the GWR parameters was beyond the scope of this study. There are methods to deal with local multicollinearity (Wheeler & Tiefelsdorf 2005), but in this study the problem has been sufficiently addressed with the visualization technique by Mennis (2006), where the insignificant values were masked out from the map.

Modelling with spatial data in linear regression models is tricky for numerous reasons. If some of the variables are spatially correlated, the way OLS solves the beta coefficient is not trustworthy. Similarly, if there is spatial autocorrelation in the model's residuals, it means that the significance of the variable estimates is not to be trusted. The residuals of the models can also be compared and mapped as a way to test the model assumptions and reliability of the model's estimates (Brunsdon et al. 1996; Charlton & Fotheringham 2009). Therefore, ideally the OLS model would also be run with maximum likelihood spatial autoregressive methods, which are used to address the spatial autocorrelation in the model's residuals (Anselin 1988).

7 Conclusions

Urban residential fires are an increasingly studied topic due to their devastating consequences in terms of human lives and economic losses, and because of the increasing availability of modern analysis methods (Ceyhan et al. 2013; Corcoran & Higgs 2013; Jennings 2013). In this study, the spatial patterns and the underlying spatial factors of 1546 residential fires in Helsinki between 2014–2018 were analysed. Prior to this study, empirical evidence about the spatial characteristics of residential fires in Helsinki has been limited to a few studies with data from 2005–2008 (Tillander et al. 2010; Špatenková & Virrantaus 2013). Thus, the objectives of this study were to bring new empirical evidence and an updated picture of the spatial patterns of residential fires, as well as to create a spatial risk model to identify the main factors influencing residential fires in Helsinki.

All in all, the data and methods used in this study were sufficient to address the study questions. First, this study has confirmed that residential fires are non-stationary and non-random spatial phenomena that cluster in space and exhibit spatial effects. The test for spatial dependence revealed that certain neighbourhoods have significant clusters of residential fires in Helsinki, namely the inner and southern parts of the city around the Central and Erottaja Rescue Stations, with a few smaller clusters in eastern Helsinki.

Second, the global OLS model found eight significant variables affecting residential fire occurrence in Helsinki. Furthermore, the GWR model revealed local variations in the distribution of explanatory variables which were overlooked by the global OLS model. Using local statistics in modelling residential fires improved the model performance and prediction outcomes significantly compared to the global OLS model as the local GWR model took spatial dependency into account. The significant parameter estimates were mapped and the patterns of each variable revealed variations in the intensity and direction of the effect on residential fire incidence.

The key contribution of this study is an updated picture of the spatial distribution and drivers of residential fires in Helsinki. The results of this study support, to a large extent, previous studies on the topic, as a connection between low socioeconomic status and residential fire incidence was found at neighbourhood level. Thus, the results have also value from a theoretical point of view. So far only a few studies, such as Špatenková and Virrantaus (2013) and Ardianto (2018), have employed the GWR to study residential fires at aggregate level. Therefore, another contribution of this study was to employ the GWR to understand the underlying behaviour of different factors associated with fire risk. This knowledge and the produced maps can be used for enhancing the strategic objectives of the rescue services, such as preparedness planning and allocation of resources.

In this study, openly available statistical and GIS software, namely R Studio, QGIS and GeoDa, were used. Using open source software is increasingly common also in the rescue services. Although the datasets are not directly openly available, the rescue services have, as an internal security official, access to many otherwise restricted data. Using newest possible available data also enables continuous risk analysis, which accounts for the national strategy for the rescue services (Ministry of Interior 2016; 2019). Furthermore, the methods used in this study provide great possibilities for further research, as the input data and scripts are easily modifiable to produce models for different types of fires or different accidents that rescue services respond to, for example.

While this study demonstrated the advantages of using spatial analytical methods to model residential fires, more case studies need to be done to identify causal variables, as well as to build and refine the fire risk theory. Further studies should focus on refining the model in this study in order to get a more comprehensive understanding of the fire risk in Helsinki, specifically in the areas where the model did not perform that well. This can be done by splitting the study area into smaller areas, and by adding more explanatory variables into the model. For example, seasonal and environmental effects have been proven to have an impact on fire risk (Corcoran et al. 2011b, Jennings 2013; KC & Corcoran 2017). The effect of environmental variables on indoor fires is not that well known, especially in Finland, therefore studying how outdoor variables relate with indoor fires is reasonable. From the climate change point of view, one could study the effect of heatwaves on the behaviour of fires by adding information about the temperature at the time of the event.

This study focused on modelling the spatial distribution of residential fires. Thus, other potential avenues for future research include accounting for the temporal aspect of fire risk, as urban fires have shown great hourly, daily, monthly, and seasonal variations (Asgary et al. 2010; Corcoran et al. 2011b; Špatenková & Virrantaus 2013; Rekola & Itkonen 2016; Song et al. 2017; Ardianto & Chhetri 2019). The results in this study are applicable only in Helsinki at the time of the analysis, so they cannot be generalized. Each city has its own unique characteristics, and therefore, further studies could focus on comparing different cities in Finland at different spatial levels.

Fire risk was defined in this study as the likelihood of fire occurrence. However, as fires have serious consequences, more attention should be also paid in developing models that can estimate fire consequences at local levels. As Jennings (1999, 28) observed, "*limited effort has been directed at micro-level studies of fire incidents to simultaneously reveal occupant characteristics, common patterns of behaviour, and causal factors underlying losses*". In addition, fires could be disaggregated by their cause and deliberateness, as has been done e.g. in Sweden (Guldåker et al. 2018). In Finland, while the PRONTO database provides vast amounts of exploitable information about the causes and deliberateness of fires, yet more detailed data about the estimated economic losses and human casualties is needed.

The quality and reliability of data and methods are directly reflected in the results. Consequently, by improving those aspects, more precise estimations and a comprehensive understanding of fire risk can be achieved. As fires are phenomena no one wishes to encounter in their lives, continuous analysis of risks helps to predict their occurrence, reduce the overall operational costs and economic losses, and potentially save lives.

Acknowledgements

I owe great thanks to my supervisor and research manager at the Helsinki City Rescue Department, Hanna Rekola, first for suggesting the interesting topic, and for giving helpful comments and endless support throughout the process. Thanks also to Johannes Ketola at the Emergency Services Academy for providing me data for this thesis.

I also want to thank my supervisors at the University of Helsinki, Petteri Muukkonen for valuable comments and for pushing me to finish in time, and Venla Bernelius for constructive comments and inspiring conversations. I owe special thanks to Athanasios Votsis for his massive help during the analysis phase.

Finally, I want to thank my family and friends, both here in Finland and overseas for their encouragement and emotional support during the journey.

Literature

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), pp. 716-723.

Anselin, L. (1995). Local Indicators of Spatial Association - LISA. Geographical Analysis, 27(2), pp. 93-115.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht, Springer Netherlands: Imprint: Springer.

Anselin, L. & Rey, S. J. (2014). Modern Spatial Econometrics in Practice, a Guide to Geoda, Geodaspace and Pysal. Chicago, IL: GeoDa Press.

Ardianto, R. (2018). Modelling spatial temporal patterns and drivers of urban residential fire risk. Doctoral dissertation, 167 p. School of Business IT and Logistics, College of Business, RMIT University.

Ardianto, R. & Chhetri, P. (2019). Modeling Spatial–Temporal dynamics of urban residential fire risk using a Markov Chain technique. International Journal of Disaster Risk Science, 10(1), pp. 57-73.

Asgary, A., Ghaffari, A. & Levy, J. (2010). Spatial and temporal analyses of structural fire incidents and their causes: A case of Toronto, Canada. Fire Safety Journal, 45(1), pp. 44-57.

Bidanset, P. E. & Lombard, J. R. (2014). The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal. Journal of Property Tax Assessment & Administration, 10(3), pp. 5-14.

Bivand, R., Keitt, T. & Rowlingson, B. (2019). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-8. URL: *<https://CRAN.R-project.org/package=rgdal>*.

Bivand, R. & Lewin-Koh, N. (2019). maptools: Tools for Handling Spatial Objects. R package version 0.9-9. URL: *<https://CRAN.R-project.org/package=maptools>*.

Bivand, R., Pebesma, E. & Gomez-Rubio, V. (2013). Applied spatial data analysis with R, Second edition. Springer, NY.

Bivand, R. & Rundel, C. (2019). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2. URL: *<https://CRAN.R-project.org/package=rgeos>*.

Bivand, R. & Yu, D. (2020). spgwr: Geographically Weighted Regression. R package version 0.6-33. URL: *<https://CRAN.R-project.org/package=spgwr>*.

Bollen, K.A. & Bauldry, S. (2011). Three Cs in Measurement Models: Causal Indicators, Composite Indicators and Covariates. Psychological Methods, 16(3), pp. 265-283.

Brunsdon, C. (2001). The comap: exploring spatial pattern via conditional distributions. Computers, Environment and Urban Systems, vol. 25, pp. 53-68.

Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial non-stationarity. Geographical Analysis, 28(4), pp. 281-298.

Ceyhan, E., Ertuğay, K. & Düzgün, Ş. (2013). Exploratory and inferential methods for spatio-temporal analysis of residential fire clustering in urban areas. Fire Safety Journal, 58, pp. 226-239.

Charlton, M. & Fotheringham, A. S. (2009). Geographically Weighted Regression. White Paper, 14 p. National Centre for Geocomputation, National University of Ireland Maynooth Maynooth, Co Kildare, Ireland.

Chhetri, P., Corcoran, J. & Stimson, R. (2009). Exploring the spatio-temporal dynamics of fire incidence and the influence of socio-economic status: A case study from South East Queensland, Australia. Journal of Spatial Science, 54(1), pp. 79-91.

Chhetri, P., Corcoran, J., Shafiq, A. & Kiran, K. C. (2018). Examining spatio-temporal patterns, drivers and trends of residential fires in South East Queensland, Australia. Disaster Prevention and Management, 27(5), pp. 586-603.

Chhetri, P., Corcoran, J., Stimson, R. & Inbakaran, R. (2010). Modelling potential socio-economic determinants of building fires in South East Queensland. Geographical Research, 48(1), pp. 75-85.

Chhetri, P., Stimson, R. & Western, J. (2006). Modelling the factors of neighbourhood attractiveness reflected in residential location decision choices. Studies in Regional Science, 36, pp. 393–418.

City of Helsinki (2019a). Helsinki facts and figures 2019. City of Helsinki, Executive Office, Urban Research and Statistics. Available at: <*https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/19_06_14_HKI-taskutilasto2019_eng_w.pdf*>. Accessed on: 30.3.2020.

City of Helsinki (2019b). Helsinki's present state and development 2019. City of Helsinki, Executive Office, Urban Research and Statistics. Available at: <*https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/19_08_27_Helsinkis_presentstate_and_development_2019.pdf*>. Accessed on: 13.5.2020.

Clark, A., Smith, J. & Conroy, C. (2015). Domestic fire risk: A narrative review of social science literature and implications for further research. Journal of Risk Research, 18(9), pp. 1113-1129.

Corcoran, J. & Higgs, G. (2013). Special issue on spatial analytical approaches in urban fire management. Fire Safety Journal, 62, pp. 1-2.

Corcoran, J., Higgs, G., Brunsdon, C., Ware, A. & Norman, P. (2007). The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study. Computers, Environment and Urban Systems, 31(6), pp. 623-647.

Corcoran, J., Higgs, G. & Higginson, A. (2011a). Fire incidence in metropolitan areas: A comparative study of Brisbane (Australia) and Cardiff (United Kingdom). Applied Geography, 31(1), pp. 65-75.

Corcoran, J., Higgs, G., Rohde D. & Chhetri, P. (2011b). Investigating the association between weather conditions, calendar events and socio-economic patterns with trends in fire incidence: an Australian case study. Journal of Geographical Systems, 13(2), pp. 193-226.

Cova, T. (1999). GIS in Emergency Management. In: Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds). Geographical Information Systems: Principles, Techniques, Applications, and Management. John Wiley & Sons, New York.

Duncanson, M., Woodward, A. & Reid, P. (2002). Socioeconomic deprivation and fatal unintentional domestic fire incidents in New Zealand 1993–1998. Fire Safety Journal, 37, pp. 165-179.

Emergency Services Academy (2020). Statistics system of Finnish rescue services (PRONTO). URL: <*http:/prontonet.fi*>. Retrieved: 2.1.2020.

FEMA = Federal Emergency Management Agency (1997). Socioeconomic factors and the incidence of fire. National Fire Data Center, United States Fire Administration, Report no. FA170.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, Chichester.

Fox, J. & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <*https://socialsciences.mcmaster.ca/jfox/Books/Companion/*>.

Godschalk, D. R. (1991). Disaster mitigation and hazard management. In Drabek T. E. & Hoetmer G. J. (eds). Emergency management: principles and practice for local government. Washington D. C., International City Management Association 131–60.

Goodsman, R. W., Mason, F. & Blythe, A. (1987). Housing factors and fires in two metropolitan boroughs. Fire Safety Journal, 12, pp. 37–50.

Greene, M. A. (2012). Comparison of the characteristics of fire and non-fire households in the 2004–2005 survey of fire department-attended and unattended fires. Injury Prevention, 18(3), pp. 170–175.

Guldåker, N. & Hallin, P-O. (2014). Spatio-temporal patterns of intentional fires, social stress and socio-economic determinants: A case study of Malmö, Sweden. Fire Safety Journal, 70, pp. 71-80.

Guldåker, N., Hallin, P-O., Nilsson, J. & Tykesson, M. (2018). Bostadsbränder i storstadsområden. (In Swedish). Malmö universitet och Lunds universitet: Myndigheten för samhällsskydd och beredskap.

Gunther, P. (1981). Fire-cause patterns for different socioeconomic neighborhoods in Toledo, Ohio. Fire Journal, 75(3), pp. 52-58.

Helsinki City Rescue Department (2017). Helsingin pelastustoimen alueen palvelutasopäätös 2017–2018. Helsingin kaupungin pelastuslaitos. (In Finnish). Available at: <*https://www.hel.fi/static/liitteet/pela/Helsingin_pelastustoimen_alueen_palvelutasopaatos_2017-2018.pdf*>. Accessed on: 24.4.2020.

Holborn, P. G., Nolan, P. F. & Golt, J. (2003). An analysis of fatal unintentional dwelling fires investigated by London fire brigade between 1996 and 2000. Fire Safety Journal, 38, pp. 1-42.

Hu, J., Shu, X., Xie, S., Tang, S., Wu, J. & Deng, B. (2019). Socioeconomic determinants of urban fire risk: A city-wide analysis of 283 Chinese cities from 2013 to 2016. Fire Safety Journal, 110, Article 102890.

Hurvich, C. M., Simonoff, J. S. & Tsai, C-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of Royal Statistical Society, Series B, 60, pp. 271-293.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2014). An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated.

Jennings, C. (2013). Social and economic characteristics as determinants of residential fire risk in urban neighbourhoods: a review of the literature and commentary. Fire Safety Journal, 62, pp. 13-19.

Jennings, C. (1999). Socioeconomic characteristics and their relationship to fire incidence: a review of the literature. Fire Technology, 35(1), pp. 7-34.

KC, K. & Corcoran J. (2017). Modelling residential fire incident response times: a spatial analytic approach. Applied Geography, 84, pp. 64-74.

Kokki, E. (2014). Palokuolemat vähentyneet - Suomen palokuolematilastot 2007–2013. (In Finnish). Julkaisu/Tutkimusraportti, 24 p. Emergency Services Academy Finland, Kuopio. Available at: <*https://www.pelastusopisto.fi/wp-content/uploads/2016/12/52749_B2_2014.pdf*>. Accessed on: 8.3.2020.

Kokki, E. & Jäntti, J. (2009). Vakavia henkilövahinkoja aiheuttaneet tulipalot 2007–2008. Research report, 119 p. (In Finnish). Emergency Services Academy Finland, Kuopio. Available at: <*http://info.smedu.fi/kirjasto/Sarja_B/B2_2009.pdf*>. Accessed on: 8.3.2020.

Kortteinen, M. & Vaattovaara, M. (2015). Segregaation aika. (In Finnish). Yhteiskuntapolitiikka, 80(8), pp. 562-574.

Koutsias, N., Martínez-Fernández, J. & Allgöwer, B. (2010). Do factors causing wildfires vary in space? evidence from geographically weighted regression. GIScience & Remote Sensing, 47(2), pp. 221-240.

Krisp, J. M., Virrantaus, K. & Jolma, A. (2005). Using explorative spatial analysis to improve fire and rescue services. In P. van Oosterom, S. Zlatanova & E. M. Fendel (eds.), Geo-information for disaster management, pp. 1283-1296. Berlin, Heidelberg: Springer Berlin Heidelberg.

Majuri, M. & Kokki, E. (2010). Pronton luotettavuus. Pelastusopiston julkaisu: B-sarja. Research report, 60 p. (In Finnish). Emergency Services Academy Finland, Kuopio. Available at: <*http://info.smedu.fi/kirjasto/Sarja_B/B4_2010.pdf*>. Accessed on: 20.4.2020.

Mazerolle, M. J. (2019) AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.2-2. URL: <*https://cran.r-project.org/package=AICcmodavg*>.

Mennis, J. (2006). Mapping the Results of Geographically Weighted Regression. The Cartographic Journal, 43(2), pp. 171–179.

Merrall, S. (2002). Anthropogenic Accidental Dwelling Fire: Incident Distribution, Theory and the Fire Service. Doctoral Dissertation, University of Liverpool.

Ministry of Interior (2016). A Safe and Resilient Finland - Rescue Services Strategy 2025. Internal Security - Ministry of the Interior Publication 20/2016. Ministry of Interior, Helsinki. Available at: <*http://urn.fi/URN:ISBN:978-952-324-101-5*>. Accessed on: 9.3.2020.

Ministry of the Interior (2019). Turvallinen ja onnettomuuksista vapaa arki 2025: Pelastustoimen toimintaohjelma onnettomuuksien ehkäisemiseksi. (In Finnish). Internal Security - Ministry of the Interior Publication 33/2019. Available at: <*http://urn.fi/URN:ISBN:978-952-324-304-0*>. Accessed on: 9.3.2020.

Moran, P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika, Vol. 37, No. 1/2, pp. 17-23.

Muckett, M. & Furness, A. (2007). Introduction to Fire Safety Management. 440 p. Routledge.

Musterd, S., Marcińczak, S., van Ham, M., & Tammaru, T. (2017). Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. Urban Geography, 38(7), pp. 1062-1083.

Nilson, F. & Bonander, C. (2020). Household Fire Protection Practices in Relation to Socio-demographic Characteristics: Evidence from a Swedish National Survey. Fire Technology, 56, pp. 1077-1098.

Nilson, F., Bonander, C. & Jonsson, A. (2015). Differences in Determinants Amongst Individuals Reporting Residential Fires in Sweden: Results from a Cross-Sectional Study. Fire Technology, 51(3), pp. 615–626.

November, V. (2004). Being close to risk. From proximity to connexity. International Journal of Sustainable Development, 7(3), pp. 273-286.

Official Statistics of Finland = OSF (2020a). Housing and housing standards. (In Finnish). ISSN=1798-6745. Helsinki: Statistics Finland. Available at: <*http://www.stat.fi/til/asas/kas.html*>. Accessed on: 8.3.2020.

Official Statistics of Finland = OSF (2020b). Municipal key figures 1987-2019. Statistics Finland's PxDatabases. Helsinki: Statistics Finland. Available at: <*https://pxnet2.stat.fi/PXWeb/pxweb/fi/Kuntien_avainluvut/*>. Accessed on: 30.3.2020.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A. & Pereira, J.M.C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. Forest Ecology and Management, 275, pp. 117–129.

Oliveira, S., Pereira, J. M., San-Miguel-Ayanz, J. & Lourenço, L. (2014). Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression. Applied Geography, 51(C), pp. 143-157.

Ramachandran, G., & Charters, D. A. (2011). Quantitative risk assessment in fire safety. London; New York: Spon Press.

Rekola, H. & Itkonen, P. (2016). Spatiotemporaalisia tarkasteluja pelastustoimen tehtävistä Helsingissä 2011-2015. (In Finnish). Helsinki: Helsingin kaupungin pelastuslaitos. Helsingin kaupungin pelastuslaitoksen julkaisuja, 2/2016, 67 p. Available at: <*https://www.hel.fi/static/liitteet/pela/Spatiotemporaalisia%20tarkasteluja%20pelastustoimen%20tehtavista%20Helsingissa%20web.pdf*>. Accessed on: 13.5.2020.

Rescue Act 379/2011. Helsinki, 1.7.2011. Available at: <*https://www.finlex.fi/fi/laki/ajantasa/2011/20110379*>. Accessed on: 5.4.2020.

Rohde, D., Corcoran, J., & Chhetri, P. (2010). Spatial forecasting of residential urban fires: A Bayesian approach. Computers, Environment and Urban Systems, 34(1), pp. 58-69.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL: <*http://www.rstudio.com/*>.

Runyan, C., Bangdiwala, S., Linzer, M., Sacks, J. & Butts, J. (1992). Risk factors for fatal residential fires. The New England Journal of Medicine, 327(12), pp. 859-863.

Saikkonen, P., Hannikainen, K., Kauppinen, T., Rasinkangas, J. & Vaalavuo, M. (2018). Sosiaalinen kestävyys: asuminen, segregaatio ja tuloerot kolmella kaupunkiseudulla. (In Finnish). Terveyden ja Hyvinvoinnin Laitos, Raportti 2/2018, 172 p. Available at: <*http://www.julkari.fi/bitstream/handle/10024/136125/URN_ISBN_978-952-343-084-6.pdf?sequence=1&isAllowed=y*>. Accessed on: 13.5.2020.

Shai, D. (2006). Income, housing and fire injuries: A census tract analysis. Public Health Report no. 121, pp. 149-154.

Špatenková O. & Stein, A. (2010). Identifying factors of influence in the spatial distribution of domestic fires. International Journal of Geographical Information Science, 24(6), pp. 841-858.

Špatenková, O. & Virrantaus, K. (2013). Discovering spatio-temporal relationships in the distribution of building fires. Fire Safety Journal, 62, pp. 49-63.

Society for Risk Analysis (2018). Society for Risk Analysis Glossary. Available at: <*https://www.sra.org/sites/default/files/pdf/SRA%20Glossary%20-%20FINAL.pdf*>. Accessed on: 19.2.2020.

Song, C., Kwan, M. & Zhu, J. (2017). Modeling fire occurrence at the city scale: A comparison between geographically weighted regression and global linear regression. International Journal of Environmental Research and Public Health, 14, 296.

Statistics Finland (2019). Grid Database 2019. Statistics Finland, Helsinki. Retrieved: 10.1.2020.

Taylor, M. J., Higgins, E., Lisboa, P. J. G. & Kwasnica, V. (2012). An exploration of causal factors in unintentional dwelling fires. Risk Management, 14(2), pp. 109-125.

Tillander, K., Matala, A., Hostikka, S., Tiittanen, P., Kokki, E. & Taskinen, O. (2010). Pelastustoimen riskianalyysimallien kehittäminen. (In Finnish). VTT tiedotteita - research notes 2530. 133 p. Available at: <*https://www.vtt.fi/inf/pdf/tiedotteet/2010/T2530.pdf*>. Accessed on: 9.3.2020.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, Vol. 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods, pp. 234-240.

Turner, S., Johnson, S., Weightman, A., Rodgers, S., Arthur, G., Bailey, R. & Lyons, R. (2017). Risk factors associated with unintentional house fire incidents, injuries and deaths in high-income countries: A systematic review. Injury Prevention, 23(2), 131.

Vasiliauskas, D. & Beconytė, G. (2015). Spatial analysis of fires in Vilnius city in 2010-2012. Geodesy and Cartography, 41(1), pp. 25-30.

Vuori, P. & Kaasila, M. (2019). Helsingin ja Helsingin Seudun Väestöennuste 2019-2050. Ennuste alueittain 2019-2034. (In Finnish). Helsingin kaupunki, kaupunginkanslia, kaupunkitutkimus ja -tilastot. Available at: <*https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/19_10_25_Tilastoja_14_Vuori_Kaasila.pdf*>. Accessed on: 30.3.2020.

Wallace, D. N. & Wallace, R. (1984). Structural fire as an urban parasite: population density dependence of structural fire in New York City, and its implications. Environment and Planning, 16(2), pp. 249–260.

Watts, J. M. & Hall, J. R. (2002). Introduction to Fire Risk Analysis. In: Hurley M.J. et al. (eds.) SFPE Handbook of Fire Protection Engineering. Springer, New York, NY.

Wei, T. & Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from: *<https://github.com/taiyun/corrplot>*.

Wheeler, D. & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems, 7, pp. 161–187.

WHO = World Health Organization (2018). WHO Housing and health guidelines. Geneva. Available at: *<https://apps.who.int/iris/bitstream/handle/10665/276001/9789241550376-eng.pdf >*. Accessed on: 10.3.2020.

Wuschke, K., Clare, J. & Garis, L. (2013). Temporal and geographic clustering of residential structure fires: A theoretical platform for targeted fire prevention. Fire Safety Journal, 62, pp. 3-12.

Xiong, L., Bruck, D. & Ball, M. (2015). Comparative investigation of 'survival' and fatality factors in accidental residential fires. Fire Safety Journal, 73, pp. 37–47.

Yamashita, K. (2008). Understanding urban fire: Modeling fire incidence using classical and geographically weighted regression. Thesis. University of Colorado at Boulder.

YKR = Yhdyskuntarakenteen seurantajärjestelmä (2019). Asuinhuoneistot ja rakennukset. Finnish Environment Institute (SYKE) and Statistics of Finland. Retrieved: 20.12.2019.

Zhang, X., Yao, J. & Sila-Nowicka, K. (2018). Exploring spatiotemporal dynamics of urban fires: A case of Nanjing, China. ISPRS International Journal of Geo-Information, 7(1), 7.

Östman, L. (2015). Varför omkommer det fler personer i bostadsbränder i Finland än i Sverige? (In Swedish). Thesis, 77 p. Department of Fire Safety Engineering, Lund University, Sweden.

Appendix: R packages

Package name	Usage	Reference
AICcmodavg	Computing AICc values	Mazerolle (2019)
car	Computing VIF values	Fox & Weisberg (2019)
corrplot	Correlogram of predictor intercorrelations	Wei & Simko (2017)
maptools	Tools for handling spatial objects	Bivand & Lewin-Koh (2019)
rgdal	Reading and writing shapefiles	Bivand et al. (2019)
rgeos	Geometry tools for spatial objects	Bivand & Rundel (2019)
spdep	Spatial dependence, weighting schemes, statistics	Bivand et al. (2013)
spgwr	Geographically Weighted Regression	Bivand & Yu (2020)