HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master's thesis
Geography
Geoinformatics

UNCOVERING POPULATION DYNAMICS USING MOBILE PHONE DATA: THE CASE OF HELSINKI METROPOLITAN AREA

Claudia Bergroth

May 2019

Supervisors:

Tuuli Toivonen, Olle Järv, Henrikki Tenkanen

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
GEOGRAPHY

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty		Laitos/Institution– Department
Faculty of Science		Department of Geosciences and Geography
Tekijä/Författare – Author		
Claudia Bergroth		
Työn nimi / Arbetets titel – Title		
Uncovering population dynamics using mobile phone data: the case of Helsinki Metropolitan Area		
Oppiaine /Läroämne – Subject		
Geoinformatics		
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Master's Thesis	May 2019	138 + appendices

Tiivistelmä/Referat – Abstract

Understanding the whereabouts of people in time and space is necessary for unraveling how our societies function. Regardless, our understanding of human presence is predominantly based on static residential population data, which is often outdated and excludes certain population groups, such as commuters or tourists. In the light of development towards 24-hour societies and the needs for promoting sustainable and equitable urban planning, reliable data of population dynamics are needed. To this end, ubiquitous mobile phones provide an attractive source for estimating the spatiotemporal digital footprints of people.

In this study, I set out to investigate 1) the feasibility of three different aggregated network-based mobile phone data – the number of voice calls, data transmission and general network connection attempts – as a proxy for human presence, 2) how does the population distribution vary in Helsinki Metropolitan Area over the course of a regular weekday and 3) the role of temporally-sensitive population data when analysing dynamic accessibility to grocery stores and transport hubs. To my best knowledge, this is the first attempt when mobile phone data is used to reveal population dynamics for scientific purposes in Finland.

Mobile phone data collected by the mobile network operator Elisa in 2017–2018 and ancillary data about land cover, buildings and a time use survey were used to estimate the 24-hour population distribution of the Helsinki Metropolitan Area. The mobile phone data were allocated to statistical 250 m x 250 m grid cells using an advanced dasymetric interpolation method and validated against population register data from Statistics Finland. The resulting 24-hour population was used to map the pulse of the city and to introduce the first fully dynamic accessibility model in the study area.

The results show that data use is a good proxy for people and outperforms voice calls or overall network connection attempts. During daytime, the static population overestimates the population in residential areas and underestimates the population in work and service areas. In general, the 24-hour population reveals the pulse of a city, which is highlighted especially in the inner city of Helsinki, where the relative share of population of the study area increases by 50 % from the share at night-time to its peak at noon. The results of the case study suggest that integrating dynamic population data to location-based accessibility analysis provides more realistic results compared to static population data, but the significance of dynamic population data depends on the study context and research questions.

In summary, aggregated network-driven mobile phone data is a feasible alternative for dynamic population modelling, however, different mobile phone data types vary in representativeness, which should be taken into account when using mobile phone data in research. To this end, critical evaluation of data and transparent data description are essential. Overall, understanding 24-hour societies and supporting sustainable urban planning necessitates dynamic population data, but advancements in data policy and availability are needed to harvest these possibilities. The results of this study also provide new empirical insights of the population dynamics in the study area, which can be used to advance planning and decision making.

Avainsanat – Nyckelord – Keywords

Mobile phone data, dynamic population, population distribution, accessibility, GIS, space-time modelling, big data, Helsinki Metropolitan Area

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Matemaattis-luonnontieteellinen	Geotieteiden ja maantieteen osasto

Tekijä/Författare – Author
Claudia Bergroth

Työn nimi / Arbetets titel – Title
Matkapuhelinaineistot väestödynamiikan tarkastelussa: tutkimuskohteena pääkaupunkiseutu

Oppiaine /Läroämne – Subject
Geoinformatiikka

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Pro gradu -tutkielma	Toukokuu 2019	138 + liitteet

Tiivistelmä/Referat – Abstract

Ymmärrys väestön alueellisen jakautumisen ajallisesta vaihtelusta on keskeistä yhteiskuntamme toiminnan ymmärtämiseksi. Tästä huolimatta ymmärrys ihmisten läsnäolosta on vähäistä ja perustuu pääasiassa staattisiin asuinpaikkakohtaisiin väestötietoihin, jotka ovat usein vanhentuneita ja saattavat johtaa eräiden väestöryhmien, kuten työmatkalaisten tai turistien, sivuuttamiseen. Kehityksen kohti ympärivuorokautista yhteiskuntaa ja kestävän ja tasa-arvoisen kaupunkisuunnittelun edistämisen tarpeiden valossa tarvitaan luotettavia tietoja väestön dynamiikasta.

Tässä tutkimuksessa tarkastelin 1) kolmen eri verkkopohjaisen matkapuhelinaineiston – puheluiden, tiedonsiirtoyhteyksien ja verkkoyhteyksien muodostusyritysten lukumäärän – soveltuvuutta ihmisen läsnäolon kuvaajana, 2) miten väestöjakauma vaihtelee pääkaupunkiseudulla säännöllisen arkipäivän aikana ja 3) temporaalisten väestötietojen käytön roolia saavutettavuusmallinnuksessa tarkasteltaessa ruokakauppojen ja liikenteen solmukohtien saavutettavuutta joukkoliikenteellä. Parhaan tietämykseni mukaan tämä on ensimmäinen kerta, kun matkapuhelinaineistoja käytetään väestön dynamiikan tarkasteluun tieteellisiin tarkoituksiin Suomessa.

Matkapuhelinoperaattori Elisan keräämiä matkapuhelinaineistoja (2017–2018) sekä aineistoja maankäytöstä, rakennuksista ja ajankäyttötutkimuksen tuloksia käytettiin pääkaupunkiseudun 24 tunnin väestöjakauman arvioimiseen. Matkapuhelimen tiedot allokoitiin 250 m x 250 m tilastoruutuihin käyttäen edistynyttä dasymetristä interpolointimenetelmää ja validoitiin Tilastokeskuksen väestörekisteritietoja käyttäen. Tuloksena saatua 24 tunnin väestöaineistoa käytettiin kaupungin pulssin analysointiin ja ensimmäisen täysin dynaamisen saavutettavuusmallin toteuttamiseen tutkimusalueella.

Tutkimuksen tulokset osoittavat, että matkapuhelinten tiedonsiirto on hyvä kuvaaja ihmisten sijainnille ja parempi kuin puhelut tai verkkoyhteyksien muodostusyritykset. Päivän aikana staattinen väestöaineisto yliarvioi väestöä erityisesti asuinalueilla samalla aliarvioiden väestöä alueilla, joilla on työpaikka- tai palvelukeskittymiä. Yleisesti katsottuna 24 tunnin väestö paljastaa kaupungin pulssin, mikä korostuu erityisesti Helsingin keskustassa, jossa tutkimusalueen väestön suhteellinen osuus kasvaa 50 %:lla yöstä sen huippuun keskipäivällä. Tapaustutkimuksen tulokset havainnollistavat kuinka dynaamisen väestötietojen integroiminen sijaintipohjaiseen saavutettavuustarkasteluun tarjoaa realistisempia tuloksia verrattuna staattiseen väestöaineistoon, mutta dynaamisten väestötietojen integroimisen merkitys riippuu tutkimuksen kontekstista ja tutkimuskysymyksistä.

Yhteenvetona voidaan todeta, että aggregoitu verkkopohjainen matkapuhelinaineisto on hyvä vaihtoehto dynaamisen väestön mallintamiseen, mutta soveltuvuus vaihtelee aineistojen välillä, mikä on tärkeä huomioida käytettäessä matkapuhelinaineistoja tutkimuksessa. Tätä vasten aineiston kriittinen tarkastelu ja läpinäkyvä aineiston dokumentointi on olennaista. Kaiken kaikkiaan 24 tunnin yhteiskuntien ymmärtäminen ja kestävän kaupunkisuunnittelun tukeminen edellyttävät dynaamisia väestötietoja, mutta tietopolitiikan ja aineistojen saatavuuden edistäminen on välttämätöntä tämän toteutumiseksi. Tämä työ tarjoaa myös uutta empiiristä tietoa väestön dynamiikasta pääkaupunkiseudulla, jota voidaan käyttää suunnittelun ja päätöksenteon tukena.

Avainsanat – Nyckelord – Keywords
Matkapuhelinaineisto, dynaaminen väestö, väestöjakauma, saavutettavuus, paikkatieto, aika-paikkamallinnus, massadata, pääkaupunkiseutu

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AFT	Activity Function Type
AW	Areal Weighting (Interpolation)
BS	Base Station
CDR	Call Detail Record
EHP	Estimated Human Presence
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
HMA	Helsinki Metropolitan Area
HSPA	High-Speed Packet Access
MFD	Multi-temporal Function-based Dasymetric (Interpolation)
MNO	Mobile Network Operator
OSM	OpenStreetMap
PT	Public Transport
RAB	Radio Access Bearer
RRC	Radio Resource Control

1. INTRODUCTION

Human behaviour and existence is ultimately bound to space and time. In other words, all activities of an individual occur and can only occur at a specific time in a specific location (Hägerstrand, 1970; Pred, 1977; Golledge & Stimson, 1997). Consequently, understanding the whereabouts of people in time and space is necessary for unravelling how our societies function (Sheller & Urry, 2006; Kwan, 2013).

Indeed, accurate information of the dynamic population distribution – where people are and when – is of high importance in a number of application fields. These include land use and transportation planning (Schmitt, 1956; Ahas et al., 2015), estimating human pressure on the environment (Levin et al., 2015), disaster preparedness and response planning (Dobson et al., 2000; Ahola et al., 2007; Bagrow et al., 2011; Bengtsson et al., 2011; Bian & Wilmot, 2015) and studying the spreading of diseases, ideas and other forms of human interactions (e.g. Wesolowski et al., 2012).

Regardless, our understanding of human presence is scarce and predominantly based on static population data (McPherson & Brown, 2004; Tatem & Linard, 2011; Deville et al., 2014; Wardrop et al., 2018). In most cases, population data is collected by national population censuses, which are typically carried out once every 10 years. Due to this, official census-based population data is often outdated, and even newly collected data represents only a snapshot in time lacking information of temporal variation in people's whereabouts. Another relatively commonly used source of population data are national population registers (Wardrop et al., 2018), available for example in Finland and other Nordic countries (UNECE 2007). Although updated more frequently, registry-based data provides similarly only a static picture of the population distribution. Using population census or register data as a proxy for people may also lead to exclusion of certain population groups, such as tourists, commuters, migrant workers or the unsheltered, as these are mostly residence-based (Smith, 1989; Dobson et al., 2000; Ricciato et al., 2017). For instance in Luxembourg, inbound cross-border commuters are estimated to account for over 40 % of the total workforce (Eurostat, 2017a). Indeed, several studies suggest that census and register-based data are sooner a proxy for "night-time population" than a representation of the actual population distribution (Schmitt, 1956; Dobson et al., 2000; e.g. Ahola et al., 2007; Bhaduri et al., 2007; Greger, 2015; Martin et al., 2015; Järv, Tenkanen & Toivonen, 2017a; Moya-Gómez et al., 2017).

Simultaneously, the mobilities of people are becoming increasingly complex. Developments in information and communication technologies (ICT) followed by increased flexibility in work-time arrangements (e.g. telecommuting and liberation of night-shift working) and increased leisure time have led to diversified, fragmented and more flexible daily rhythms and mobilities of people (Schlich et al., 2004; Sheller & Urry, 2006; Banister, 2008; Glorieux et al., 2008; Hubers et al., 2009). These changing population dynamics are also shaping our societies at large and have paved the way to the ongoing public and scientific discussion of a shift towards 24-hour-driven societies (Glorieux et al., 2008; Hubers et al., 2009; Malmberg, 2017).

Increasing complexities of daily rhythms and flows are particularly evident in cities, which already account for more than half (55 % in 2018) of the world's population (United Nations, 2018). The importance of understanding how people behave in space and time is becoming increasingly vital in the light of the rapid growth of cities. By 2050, the share of people residing in urban environments is estimated to rise to almost 70 % (United Nations, 2014, 2018). High levels of urbanization is linked to urban sprawl and service centralization, which in turn have been associated with increased mobilities of people (Banister et al., 1997; Næss, 2005; Zhao, 2010). These developments pose challenges for decision making and planning. To better understand how our cities function and promote sustainable planning, reliable measures that account for these dynamic realities are needed.

One widely applied conceptual and methodological framework to gain understanding of urban processes is spatial accessibility (Geurs & van Wee, 2004; Bertolini et al., 2005). Traditionally, accessibility has been defined as "the potential of opportunities for interaction" (Hansen, 1959), although there does not exist a clear consensus in literature on how to define it (see Geurs & van Wee, 2004). Due to its ability to bridge together land use and transportation together with environmental and social factors, spatial accessibility is particularly suitable for solving urban issues (Bertolini et al., 2005). Accessibility has proven useful in planning healthy and more sustainable cities that promote equitable access to services (van Wee & Geurs, 2011). Consequently, accessibility has become a central tool in urban planning and has been adapted to various fields of research.

Spatial accessibility is an intrinsically dynamic concept (Järv et al., 2018). The main components of accessibility – people, activities and transport – all fluctuate in time and space. Regardless, time has seldom been incorporated in accessibility modelling (Kwan, 2013). Fuelled by developments in computing power and geographic information systems (GIS), integration of time has seen a rise in so-called person-based approaches stemming from time geography (Hägerstrand, 1970; Miller, 1991; Kwan, 1998; Neutens, Versichele, et al., 2010; Widener et al., 2015). These advancements have, however, not poured to location-based approaches. The vast majority of location-based approaches have largely been treated as fixed in time, albeit static approaches often lead to oversimplification and errors (Kwan, 2013; Järv et al., 2018). Indeed, incorporation of time-sensitive data into accessibility models in form of e.g. service opening hours and public transport schedules has proven to provide more realistic results (Tenkanen et al., 2016; Järv et al., 2018).

In recent years, dynamic approaches have started to emerge also in location-based accessibility research for instance in the domain of health (Farber et al., 2014; Tenkanen et al., 2016). In these studies, multi-temporality of the transport network and service opening hours have been considered. Yet, the dynamism of the people-component in location-based accessibility models is still missing as a rule (for exceptions, see Järv et al., 2018; Moya-Gómez et al., 2017), much due to the difficulties of acquiring time-sensitive population data. Consequently, using home locations as a proxy for people is the standard of activity in spatial accessibility research despite the widely acknowledged criticism of the approach (e.g. Næss, 2006; Schönfelder & Axhausen, 2010; Kwan, 2013).

The proliferation of novel data sources and the advent of 'the big data revolution' (Kitchin, 2014a) has opened new opportunities for understanding our environment and societies. Big data derived for instance from mobile phones (e.g. Reades et al., 2007; González et al., 2008; Ahas, Silm, et al., 2010; Deville et al., 2014), smart cards, such as payment cards, public transit cards or loyalty cards (e.g. Bagchi & White, 2005; Hasan et al., 2013; Ma et al., 2013; Sobolevsky et al., 2014) or geo-located social media posts and other user generated geographical information (e.g. Noulas et al., 2012; Hawelka et al., 2014; Heikinheimo et al., 2017; Toivonen et al., 2019) can provide digital footprints of people in space and time. This wealth of spatiotemporal data has proven its feasibility and potential in dynamic population mapping in many fields of research.

During the last decades, mobile phones have become one of the most widely adopted pieces of technology by people (Townsend, 2000; Oliver et al., 2015). The number of mobile phone subscriptions outnumbered the fixed-line subscriptions in 2002 and exceeded the human population in 2014 (ITU 2018a). Although the fastest phase of the mobile phone adoption has already been passed, the global mobile phone penetration rate of 103.6 % (mobile subscriptions = SIM cards per 100 inhabitants) in 2017 is still expected to rise in the forthcoming years (ITU 2018a). Due to this ubiquity, mobile phones can provide a large sample size of the population. Thus not surprisingly, mobile phones have been widely used as a proxy for people in scientific research across disciplines to gain understanding of population distribution and spatial mobility patterns on both individual and aggregate levels, different spatial scales and temporal granularities (Ratti et al., 2006; González et al., 2008; Deville et al., 2014; Järv et al., 2014; Louail et al., 2014; Ahas et al., 2015). Of the vast amount of research based on mobile phone data, most studies are based on Call Detail Records, although this type of data is generated only when the user actively uses the phone (Ricciato et al., 2017). Furthermore, only few studies have reported using records of mobile data use as a proxy for people or compared their feasibility to other mobile phone data, although the amount of data transmission on mobile phones is becoming more and more significant alongside the increasing penetration of smart phones (European Commission, 2018; ITU 2018b).

In Finland, at least to the author’s knowledge, scientific research based on mobile phone data to reveal population dynamics or accessibility patterns has not yet been conducted, despite the internationally high mobile penetration and use rates (FICORA 2018; ITU 2018a; Statistics Finland, 2018a). Additionally, the public discussion of 24-hour society has been active in Finland in the Helsinki Metropolitan Area (HMA) (Malmberg, 2017), and the country also has one of the highest shares of employment working at night in Europe (8.5 % of the total employment, European Union average 6.1 %) (Eurostat, 2017a), which further highlight the need for understanding the spatiotemporal population patterns in the HMA. The high rates of night workers also raise questions of social (in)equality regarding access to services. Although several studies focusing on location-based accessibility have been conducted in the HMA (Salonen et al., 2012, 2014; Salonen & Toivonen, 2013; Tenkanen et al., 2016; Kujala et al., 2018), so far all approaches have relied on static population data. These notions combined with the availability of suitable temporally sensitive spatial data and sophisticated tools for accessibility modeling (see Toivonen et al., 2014) make the HMA an attractive study area for analyzing dynamic population patterns and dynamic accessibility.

In the light of the above, this study aims to make methodological and empirical contributions. Methodologically, the objective of this study is to assess the suitability of network-driven data types to study population dynamics and propose an open access multi-temporal dasymetric model to the study area, HMA, by extending approach developed by Järv et al. (2017a). Finally, the extended method is used to create a 24-hour population dataset for the study area. Empirically, the aim of this study is to reveal new insights from HMA regarding population dynamics and dynamic accessibility.

These objectives are pursued through the following research questions:

1) How well do different network-driven mobile phone data – the number of voice calls, data transmission and overall network connection attempts – correspond to the spatiotemporal patterns of human presence?
2) How does the hourly population distribution vary in Helsinki Metropolitan Area over the course of a regular weekday?
3) To what extent does the multi-temporal population data influence the results compared to static population data when measuring spatial accessibility to grocery stores and transportation hubs using public transport?

In this study, these questions will be analysed in the geographical context of Helsinki Metropolitan Area using network-driven cellular mobile phone data aggregated on base station level.

This thesis is structured as follows: chapter 2 presents the theoretical background, defines the key concepts of the study and summarizes the current scientific discussion of related topics. The study area, data and methods used in this study are presented in chapters 3–5, followed by the results in chapter 6. Finally, the significance of the results, their limitations and suggestions for further research are discussed in chapter 7.

2. BACKGROUND

2.1. Uncovering spatiotemporal population dynamics

2.1.1. Time geography as a framework for analysing human presence

Human behaviour is inherently tied to time and space. This notion is at the core of time geography, a conceptual framework originally developed by the Swedish geographer Torsten Hägerstrand (1970) in his seminal paper "*What about people in regional science?*". Since its introduction, the framework of time geography has been further developed by scholars (e.g. Lenntorp, 1976; Burns, 1979; Miller, 1991, 2005a). Time geography has become widely acknowledged across disciplines as a theoretical and methodological tool for understanding the interplay of human behaviour and various constrains in a spatiotemporal context (for a review, see Neutens et al., 2011; Sui, 2012).

One of the key concepts of time geography is the *space-time path* (Figure 1). A space-time path is essentially a trajectory that consists of consecutive points and path segments in the space-time continuum, which can be used to visualize movements of an individual in three-dimensional space (Hägerstrand, 1970). In essence, the space-time path is an illustration of an individual's realized *spatial mobility*. Spatial mobility has traditionally been defined as geographic displacement, i.e. "movement along a trajectory that can be described in terms of space and time" (Kaufmann et al., 2004). Consequently, the concept of space-time paths have been widely adopted in person-based mobility research (Kwan, 1999, 2000; Kveladze et al., 2013).

Our possibilities to conduct activities, and therefore also our space-time paths are, however, steered and limited by different spatiotemporal constraints. According to Hägerstrand (1970), three kinds of constraints can be distinguished: 1) *capability constraints*, which refer to an individual's biological limitations and availability of tools to trade time for space (e.g. need for sleep or car ownership), 2) *coupling constraints*, which refer to the limitations caused by participating in joint activities (e.g. meetings) and 3) *authority constraints*, which refer to restrictions set by different authorities to access locations in time (e.g. opening hours or restricted zones). Due to these constraints, different activities are fixed to space and time on different levels of rigidity (Cullen & Godson, 1975; Vilhelmson, 1999; Næss, 2005). For instance grocery shopping is typically bounded both in space and time since stores tend to be situated in specific locations and can be accessed only during predefined opening hours, whereas taking a walk is generally considered flexible both in space and time (Vilhelmson, 1999). Consequently, the space-time fixity of activities determines the “pegs” or anchor points in time and space (Cullen & Godson, 1975), which delimit the *time budget* of an individual (Carlstein et al., 1978; Miller, 2005a). Hence, more fixed activities delineate when and where more flexible activities can occur.

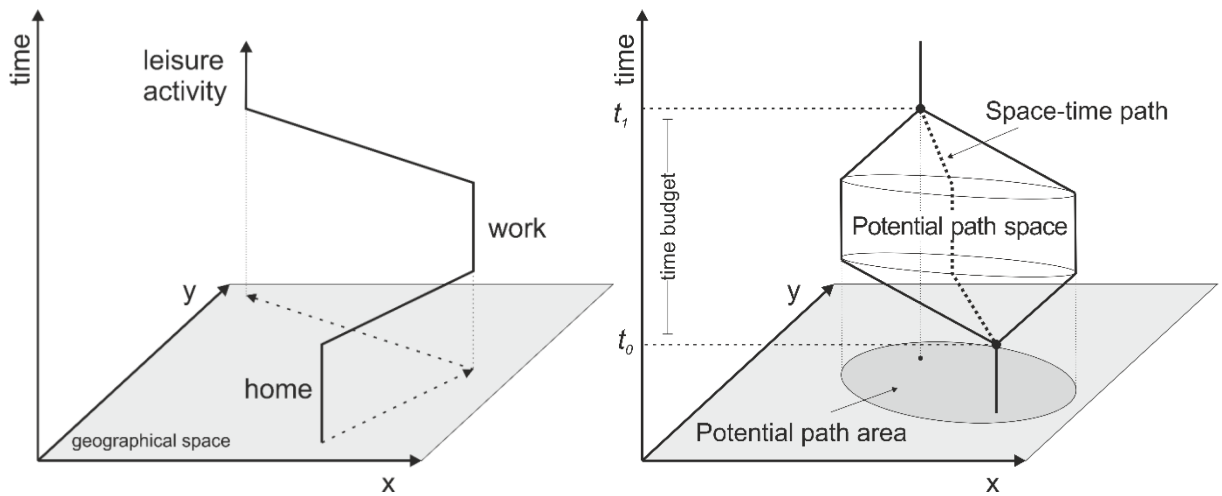


Figure 1. A space-time path (left) and a space-time prism (right). Adapted from Miller (2005a).

All locations that can be reached within a given time budget form the *potential path space* of an individual (Lenntorp, 1976). In other words, the potential path space is a sum of all potential space-time paths. A person’s potential path space can be visualized using a so-called *space-time prism*, which is another central concept of classical time geography (Figure 1). Projecting the potential path space on a geographical plane forms an individual’s *potential path area*, which can be used as a measure of the ability or freedom to move and access places in a space-time context (Lenntorp, 1976; Miller, 1991; Järv et al., 2014). Thus, the space-time prism and the potential path area are in fact graphical representations of an individual’s *potential mobility*, also considered as (person-based) *spatial accessibility* (Salomon & Mokhtarian, 1998; Kobayashi et al., 2011). Where the potential path area approaches the potential mobility from the perspective of a geographical plane, the space-time prism includes also the temporal dimension. The space-time path and space-time prism are in fact tightly interwoven concepts showing two sides of the same coin – the path a representation of mobility and the prism that of accessibility.

Sometimes the case of interest does not lie in understanding continuous movement from an origin to a destination, but simply in inferring the momentary state of a specific phenomenon. This is the situation for instance when mapping the population distribution at a specific moment in time in a given location. Where a trajectory of a series of consecutive units in time and space can be used to analyse mobility, a *snapshot* of a space-time trajectory at a given location and moment in time can be seen as a measure of (physical) *presence* (Figure 2). Vice versa, a continuous sequence of such snapshots in space-time forms the space-time path of an individual.

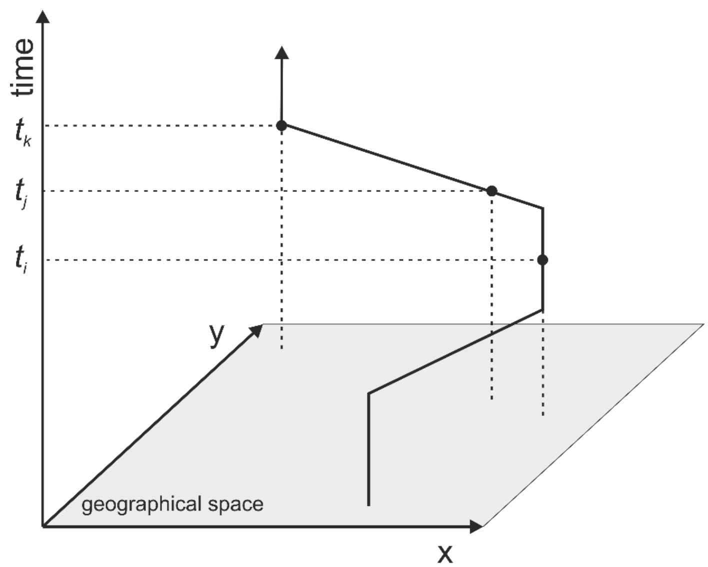


Figure 2. The concepts of presence and mobility. Mobility is represented by the trajectory (space-time path) and presence as snapshots of the trajectory at a specific moment in time (t_i, t_j, t_k) in a specific location.

Although the concept of physical human presence can be derived from classical time geography, only few attempts have been made to conceptualize it graphically. Efforts to picture the momentarily relationship of time and space have primarily concerned human interaction between multiple individuals. Parkes and Thrift (1980) and Golledge and Stimson (1997) proposed three spatiotemporal scenarios for presence relationships: 1) *co-location in time*, 2) *co-location in space* and 3) *colocation in time and space* or *co-existence* (Figure 3). Co-location in time refers to a setting where two or more individuals are in different locations at the same time. Co-location in space on the other hand occurs when people visit the same location at different times. If multiple individuals are in the same location at the same time, they have a relationship of co-existence. Also extensions to accommodate virtual interaction to classical time geography have been proposed including the states of *synchronous* and *asynchronous tele-presence* and *physical presence* (Janelle, 2004; Yu & Shaw, 2007).

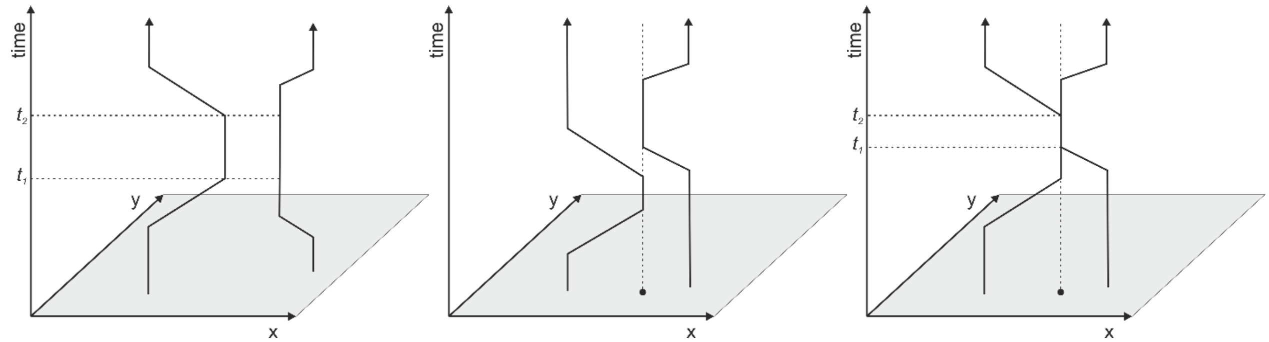


Figure 3. The relationship of time and space in presence of multiple individuals (adapted from Golledge & Stimson (1997)). In order from left, the images illustrate i) co-location in time, ii) co-location in space and iii) co-location in time and space (co-existence).

Even though time geography originally stems from an individual or person-based perspective, it can be used as a foundation to comprehend human activity also on an aggregated level. When considering presence at an aggregated level of multiple individuals, we arrive at the concept of spatial distribution or *population* (Zandvliet & Dijst, 2005). This aggregated form of presence can be visualized for instance as a time-series (Peuquet & Duan, 1995; Kang et al., 2012) and used to analyse change in time, assuming that the temporal sequence of the snapshots is known. Figure 4 shows the relationship between person-based and location-based approaches. One snapshot (S_i) represents the momentarily population, which stems from Golledge & Stimson's (1997) co-location in time. Mobility cannot be derived from the aggregated level presentation but is indirectly echoed through the change in cell values.

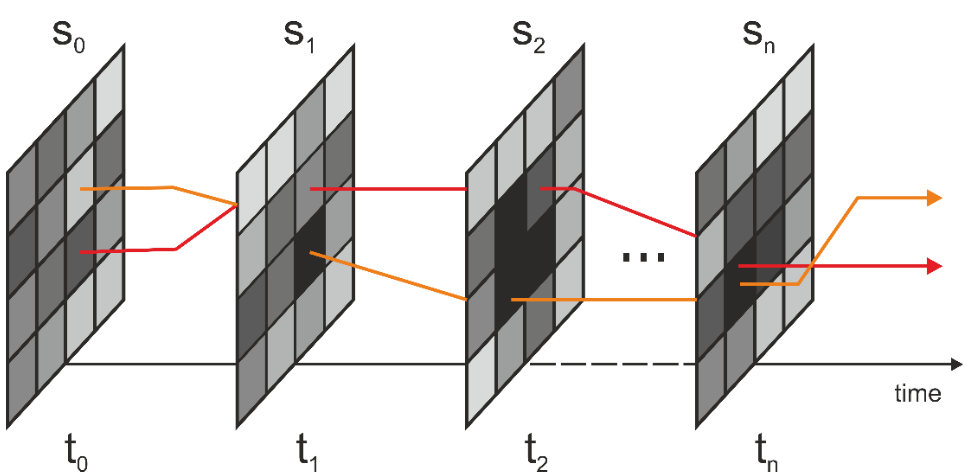


Figure 4. Graphical conceptualization of aggregated space-time paths as a time-series. The arrows represent the mobility of individuals (person-based perspective). Each snapshot (s_i) illustrates the state of the geographical phenomenon (e.g. population distribution) on aggregated level at a moment in time (t_i) (place-based perspective).

From a person-based viewpoint, if multiple individuals enter the same space at the same time, an occurrence also known as bundling (Hägerstrand, 1970), there is a higher likelihood for opportunities for interaction. This notion applies also to location-based approaches, however in such a case it is crucial to consider the impact of the spatial units on the likelihood of interaction.

2.1.2. Spatiotemporal rhythms of everyday life and the city

The space-time constraints and fixity of different activities shape the daily mobilities of individuals (see chapter 2.1.1). When these mobilities are examined from an aggregated perspective, the collective spatiotemporal patterns of individuals create a dynamic rhythm or "pulse" (Froehlich et al., 2008; Batty, 2010; Vaccari et al., 2010). In other words, the pulse of a spatial entity can be considered as an aggregated product of the underlying daily routines by individuals. Understanding this pulse is central especially in urban contexts for understanding how cities function (Batty, 2010). The term *pulse* builds on the continuum of biological metaphors that have been used to describe cities and society in general in literature, where cities have been for instance compared to organisms with road networks portraying the arteries (Girardet, 2004; Samaniego & Moses, 2008; Batty & Marshall, 2009; Batty, 2012). Another, more technocratic an approach for conceptualizing cities as dynamic systems is picturing them as machines (see e.g. Lynch, 1984).

Physical space is fundamentally interlinked with the social dimension (Lefebvre, 2004). As the pulse of a place is shaped by the mobilities or lack of mobilities of individuals, the spatiotemporal dynamics of presence can be considered as a measure of *social time* (Ahas et al., 2015). Social time is a concept, which encompasses human activities and interaction in the notion of time, unlike standard or solar time, which see time as merely quantitative (Sorokin & Merton, 1937). From the perspective of social time, for example the term "midnight" is rather seen as a functional moment defining the end of the active day, than a fixed hour based on the solar system. In this light, understanding the time use of individuals is key to unravelling the dynamics on city level, and vice versa, how people occupy space in time can give an understanding of the (social) functions areas represent. A place that is frequently visited can give an indication of its importance and characteristics. High levels of human presence have been used as a measure of the attractiveness (Moya-Gómez et al., 2017) and vitality (Sulis et al., 2018) of a place by scholars. Indeed, understanding which locations people spend time in and when is essential for urban planners.

However, attaining this type of information can be rather difficult and burdensome with traditional means. Approaches based on public participation geographic information systems (PPGIS) to map important locations have been applied by scholars and urban planners (Brown & Kyttä, 2014; Salonen et al., 2014), but these approaches are often limited in their temporal granularity. Moreover, important places collected by PPGIS do not necessarily reflect well the places people visit on a regular basis, and the results may be subject to sampling issues, for example by an underrepresentation of the elderly (Brown & Reed, 2009).

Understanding the spatiotemporal dynamics of cities has for long attracted scholars across disciplines. During the last decade, studies on the pulse and rhythmicity of cities have been seen a new surge largely thanks to novel data sources (Ratti et al., 2006; Reades et al., 2007; Froehlich et al., 2008, 2009; Vaccari et al., 2010; Ahas et al., 2015; Dobler et al., 2015; Graells-Garrido et al., 2017). Topical research questions and focus areas of city dynamics include for example event detection and analysis of human patterns during special events (Reades et al., 2007; Vaccari et al., 2010; Traag et al., 2011), identification of areas that people use for recreation (Levin et al., 2015), hourly hot-spot detection for land use planning (Louail et al., 2014), land use classification (Soto & Frías-Martínez, 2011) and analysis of urban morphology, such as polycentricity (Louail et al., 2014). Knowledge of people's whereabouts in time derived from novel data also enable the development of various smart city applications, such as on-demand transportation and services (Ahas et al., 2015).

The pulse of a city, and spatiotemporal rhythms of human behaviour in general, can and have been analyzed on various temporal scales. Most studies have focused on analysing the diurnal patterns in cities or differences between weekdays and weekends. For example Reades et al. (2007) analysed the weekly pulse of Rome using mobile phone data and applied K-means clustering to investigate the temporal profiles of different parts of the city. Studies also include variation caused by seasonality and special events. Vaccari et al. (2010), for example, used mobile phone data to analyse how the distribution of people fluctuated before, during and after the presidential inauguration of president Barack Obama in Washington D.C. in January 2009 based on the amount of hourly calls per base station. Deville et al. (2014), on the other hand, were able to extract seasonal patterns of population distribution from mobile phone data in France and Portugal and found that the population density decreased in most urban areas and increased rural areas and touristic destinations during the holiday season compared to work-driven autumn months.

Several studies suggest that spatiotemporal patterns of human behaviour are highly predictable (Song, Qu, et al., 2010) and have a high spatiotemporal regularity (González et al., 2008). In general, people tend to conduct more activities during daytime than during the night (Roenneberg et al., 2003). Relatively high rates of mobility can be found during morning hours and the late afternoon, whereas people tend to be more stationary during the night (Song, Koren, et al., 2010). In accordance, the likelihood to visit less regular locations in an individual's life is smallest during the night and highest during daytime, especially on weekend days (Song, Koren, et al., 2010). This regularity of human mobility and spatiotemporal patterns in conducting activities are visible also on city level in how people use space in time.

Previous research suggests that different activities have distinctive temporal rhythms, and these are often tied to spatial characteristics, such as land use. Several studies have shown for instance that areas with highest activity rates during the late evening, night and early morning and lower rates during late morning and afternoon on weekdays are likely to have residential functions (e.g. Froehlich et al., 2008; Soto & Frías-Martínez, 2011; Greger, 2015; Qi et al., 2015; Ma et al., 2017). Correspondingly, areas that exhibit high rates of activity during the typical work hours during the workweek are likely to have work or educational functions (e.g. Froehlich et al., 2008; Soto & Frías-Martínez, 2011; Greger, 2015; Qi et al., 2015; Ma et al., 2017). Places that offer leisure activities, on the other hand, appear to have highest concentrations of human activity during weekday evening hours and weekends, which is also when people typically conduct leisure activities (e.g. Froehlich et al., 2008; Soto & Frías-Martínez, 2011). Certainly, land use in cities is rarely monofunctional, and one place may contain more complex combinations of different types of functions, either simultaneously or at different hours of the day (see e.g. Ahas et al., 2015). Nonetheless, land use and time use patterns seem to be highly interconnected and places to have a pulse of their own, a phenomenon Goodchild & Janelle (1984) call 'temporal specialization' of areas, and has also been referred to in literature as 'chronotypes' (Bertolini & Dijst, 2003).

Overall, differences in the dynamics between the five weekdays appear to be relatively small (e.g. Dobler et al., 2015). On the other hand, several studies indicate that weekday patterns differ from weekend patterns (Reades et al., 2007; Candia et al., 2008). Weekend patterns tend to be for example more random and varying in time and space compared to weekdays (Dobler et al., 2015). This is also in line with results from time-use surveys (Novak & Sykora, 2013).

Although the diurnal variation of human presence in different land use areas seem to follow similar patterns in western cities on a broad scale, studies have shown that the patterns are not uniform across cities or within them on neighbourhood level (Reades et al., 2007; Ahas et al., 2015). Differences between cities can be influenced for example by urban form and cultural and political factors, which are inevitably reflected to daily time use routines (Ahas et al., 2015), a renowned example being siesta.

2.1.3. From static to dynamic population distribution mapping

Population mapping has traditionally relied on national population and housing censuses and sample surveys (Deville et al., 2014; Wardrop et al., 2018). A *census* is essentially a method for systematically enumerating all units in a statistical population (Eurostat, 2017b), although it is sometimes used interchangeably with the procedure of obtaining the count and other key statistics of the human population with conventional means. In other words, a census can also be used as an approach to collect other information than human demographics and while usually form- and interview-based (Valente, 2010), a census of the population can be recorded also using different techniques. A *sample survey*, on the other hand, is carried out only for a subset of the population, and then used to estimate the results of the whole target population (Eurostat, 2017b). Another, more sophisticated a method used to collect a census of the population is to extract data from existing *registers* – continuously updated databases of data flows (Eurostat, 2017b) – or use combinations of the above (see Valente, 2010).

Due to their nature, traditional population censuses are burdensome and expensive to carry out, which is why they are commonly taken only once every ten years (Valente, 2010). In addition to their low temporal resolution, questionnaire-based population census and survey data may also be subject to sampling issues related to non-uniform response rates (Vigdor, 2004). Moreover, official population census and register data are usually residence-based, which may lead to exclusion of certain population groups, such as commuters, migrant workers, tourists or the unsheltered (Dobson et al., 2000; Ricciato et al., 2017). Furthermore, these data lack information on the spatiotemporal variation of people's whereabouts occurring for example on a daily, weekly or seasonal basis. Regardless of these multiple shortcomings, our understanding of the population distribution is still largely based on these static data estimating the "sleeping population", despite the growing needs for fine-grained, or sometimes even real-time, spatiotemporal data of the whereabouts of people (Wardrop et al., 2018).

The emergence of remote sensing techniques opened new possibilities to overcome issues related to prior population estimation methods. Following the advancements brought by dwelling recognition from aerial photography, satellite-based remote sensing has become a widely applied method to collect data on population distribution (for a review, see Wu et al., 2005). The main advantages of satellite-based remote sensing over official population census counts are its higher temporal accuracy, lower data collection burden and the ability to map large areas (Sutton et al., 2001). Indeed, several initiatives to map population on a continental or global level arose in the 1990s (Balk et al., 2006; Deville et al., 2014). Many scholars use nightlights to estimate the distribution of people (e.g. Sutton et al., 2001), although previous research suggests that nightlight-based measures overestimate population in urban areas and underestimate smaller settlements (Tatem et al., 2005; Balk et al., 2006; Potere & Schneider, 2007). Furthermore, as nightlight data is specific to night-time, it is a fundamentally limited source for inferring the diurnal variation of population distribution.

The first known scholar to address the need to differentiate daytime and night-time population was sociologist Louis Wirth (1938) in his seminal paper "*Urbanism as a way of life*" (Ma et al., 2017). For long, daytime population was analysed using origin-destination (OD) matrices, travel surveys and census data on travel to work, often ignoring the spatial dimension (Foley, 1954; Goodchild & Janelle, 1984). To date, most studies that account for both the spatial and the temporal perspectives in population estimation are limited to a dichotomous approach that considers only night-time and daytime distribution of people and ignores the variation during the day, between days of the week, seasonally or during special events (Bian & Wilmot, 2015).

Another approach to ameliorate the issues related to using residential population as a proxy for people emerged at the turn of the 21st century, when Oak Ridge National Laboratories developed a method for analyzing the so-called *ambient population*. Coined by Dobson et al. (2000), the term ambient population is defined as the diurnal average population in a given area or location (Sutton et al., 2003; Martin et al., 2015). Accordingly, the concept of ambient population should be differentiated from the concept of *dynamic population, w*hich usually refers to multi-temporal population distribution (see e.g. Deville et al., 2014; Bian & Wilmot, 2015; Ma et al., 2017) and from *temporary population* or *visitor population* (see Smith, 1989; Zandvliet & Dijst, 2005) that typically refer to the non-permanent population (caused e.g. by flows of inbound daily or seasonal commuters) and do not necessarily contain a multi-temporal dimension. Perhaps the best known attempt to map ambient population is the first – the global LandScan population database (Dobson et al., 2000) with the spatial resolution of 30 by 30 arc seconds, which significatly improved the spatial accuracy of the earlier Gridded Population of the World (GPW) dataset (Tobler et al., 1997). Ambient population mapping has later become a commonly used concept for example in the field of crime analysis (e.g. Andresen, 2011; Malleson & Andresen, 2015, 2016). A significant shortcoming of ambient population data is that it does not represent the population distribution at any specific moment in time, and thus caters poorly to the needs of sudden or other time-specific purposes, such as service-allocation or evacuation (Aubrecht et al., 2013).

According to Greger (2015), the first time-specific population estimation method to move beyond the day-night dichotomy in the field of population geography was developed by Martin et al. (2009) as part of the Population 24/7 project. To achieve this, ancillary datasets, including statistics of the number of employers and employees, tourists, prisoners, hospital inpatients and traffic and passenger flow data, were used to enrich census data (Martin et al., 2009). Since then, new efforts to map the dynamic population using traditional data have surfaced on different spatial and temporal scales reaching both hourly level and extent of multiple continents (e.g. Ahola et al., 2007; Greger, 2015; Stevens et al., 2015).

These advancements have taken place largely thanks to advancements in development of geospatial analysis tools and interpolation methods, such as dasymetric interpolation, which enables combining ancillary data to refine the accuracy of existing census data (Deville et al., 2014; Tatem, 2017). The aim of efforts to downscale the existing population data from larger units, an approach also known as *spatial disaggregation* is twofold: to improve the spatial accuracy and usability of the data (e.g. Järv, Tenkanen & Toivonen, 2017a). Census data, for example, is typically available as aggregated data on tract-level with varying size and shape, which makes integration to other data challenging. Yet, the sophisticated spatial disaggregation methods still largely rely on static census or remotely sensed data as the basis, missing out on the potential of temporally and spatially sensitive novel data sources (Deville et al., 2014; Järv, Tenkanen & Toivonen, 2017a).

2.1.4. Novel data sources as a proxy for people

The advent of novel and "big" data sources have fundamentally changed how spatiotemporal dynamics of human presence can be studied on an individual and aggregated level. The emergence of unprecedentedly large volumes of diverse and fine-grained spatiotemporal data, often called *big data*, allow approaches and ways of science that were previously demanding or even impossible (Kitchin, 2014a). Simultaneously, these novel data sources also pose new challenges to the whole geographical analysis work flow from data collection, storing and processing to analysis, evaluation and conclusion.

One of the most commonly used definitions of big data is the one introduced by Laney (2001), where big data is described through three "Vs": volume (referring to the large size of big data), velocity (referring to the high, nearly real-time speed at which new data is created), and variety (referring to the diversity of data sources and structures). Scholars have later extended the definition by additional characteristics, such as fine-grained in resolution, scalable, exhaustive and relational in nature (Kitchin, 2013). Big data often contains a spatial reference, which is why the 'big data revolution' has opened new avenues for conducting geographic research (Goodchild, 2013; Graham & Shelton, 2013). Several developments in the ICT sector and their adoption in everyday life have paved the way towards the growth of data. One central factor that has fuelled both the growth of big data and its utilization in science is the increase in computational power and availability of affordable storage capacity (Kitchin, 2014b). Another related issue is the emergence of Web 2.0, which has enabled users to produce user-generated content on the web, such as through social media platforms, which are significant contributors to big data (Kitchin, 2014b). Moreover, the surge of various sensors embedded to everyday life, such as smart phones, motion detector -equipped street lights, surveillance cameras or Wi-Fi-enabled vehicles, provide a cornucopia of new sources of data (Kitchin, 2014b).

Consequently, various kinds of spatiotemporally-enabled novel big data sources have been used to extract digital footprints of users to infer population dynamics on different spatial and temporal scales (for a brief listing of novel spatiotemporal population data sources, see Järv et al., 2018).

Perhaps the most commonly used novel data source as proxy for people to date is **mobile phone data**. This is largely due to the ubiquity of mobile phones in people's daily lives (Townsend, 2000). Mobile phones are widely adopted across different population groups (Townsend, 2000), and customarily carried along and used throughout daily activities (Järv et al., 2014; Tranos & Nijkamp, 2015) making mobile phone data an attractive source to use as a proxy for people. Mobile phone data can also be used to overcome the traditional issue of data collection burden as it is often passively and automatically generated (Ahas, Silm, et al., 2010). Moreover, since the data is originally collected in digital format, the data is not subject to memory bias as traditional surveys and does not require manual digitalization, which is another potential source of human error (Ahas, Aasa, et al., 2010). In addition, mobile phone data can reach very high levels of spatial and temporal accuracy (see, section 2.2), the lack of which has been one of the largest drawbacks of traditional data sources used in dynamic population mapping. Furthermore, mobile phone data can provide the possibility to extend the spatial and temporal scope of traditional approaches of population mapping to large geographical areas and to longitudinal perspectives (Järv et al., 2014). Depending on the dataset, mobile phone data can also enable analysis of the population dynamics of certain groups, such as daily commuters, tourists or ethnic groups (e.g. Kuusik et al., 2009; Järv et al., 2012, 2015).

Consequently, initiatives to improve official population statistics by combining traditional census data with mobile phone data have surfaced in recent years (United Nations, 2017). One of the earliest examples in literature of using mobile phone data to analyze the dynamic population distribution was presented by Ratti et al. (2006), where population dynamics were analyzed in the city of Milan, and later studies have also extended the coverage to national level (e.g. Deville et al., 2014). From a temporal perspective, the majority of studies using mobile phone data analyse population distribution on an hourly level and mostly on weekdays (e.g. Oliver et al., 2015; Järv, Tenkanen & Toivonen, 2017a), but research covers also weekly variation including weekends and special events and phenomena (e.g. Pulselli et al., 2008; Vaccari et al., 2010; Graells-Garrido et al., 2017).

In addition to mobile phone data, also other novel data sources have been used or have the potential to provide information of human presence and mobility in time and space. **Geo-located social media data** from various platforms has been used as a proxy for people or human activities, such as Twitter (Frias-Martinez et al., 2012; Moya-Gómez et al., 2017; Patel et al., 2017), Foursquare (Cranshaw et al., 2012; Noulas et al., 2012) or Instagram (Heikinheimo et al., 2017; Tenkanen et al., 2017). Contrary to mobile phone data, which typically lacks or is limited in socio-demographic attributes of users (e.g. Järv et al., 2014), social media data can provide useful insight to the questions *who* (e.g. age, gender or nationality or other socio-demographic information of the user), *what* (activity that was conducted in the location) and *why* (reasons for visiting the location), the lack of which are typical drawbacks of cellular mobile phone data (Calabrese et al., 2014; Heikinheimo et al., 2017; Toivonen et al., 2019). In addition, social media data is often openly available through Application Programming Interfaces (APIs), unlike mobile phone data, which is typically difficult to access. However, geo-located social media data also has its drawbacks in dynamic population mapping. Firstly, social media users, for example of Twitter, tend to be biased towards certain population groups (Sloan et al., 2015) causing sampling issues when sampling the population dynamics of the whole population. Secondly, the post frequency is typically uneven and the spatial accuracy of the geotags raise issues of uncertainty (Toivonen et al., 2019). Furthermore, even if data can currently be accessed over open APIs, the data is still owned by private enterprises and access and use rights may be subject to change or APIs to be closed, as happened to the Facebook Pages API in April 2018 (Lomborg & Bechmann, 2014; Freelon, 2018).

Another novel data source that has been used to estimate the dynamic population distribution is **smart card data**, such as transportation, bank or loyalty card transactions. For example Ma et al. (2017) used subway smart card data to analyse the hourly distribution and variation of people in Beijing. Other potential data that have so far mostly been used for analysing human mobility instead of presence include positioning **data from Global Navigation Satellite System (GNSS) - enabled devices** and vehicles, such as taxi trajectories (Ashbrook & Starner, 2003; Romanillos et al., 2016) and **location-based sensors**, such as Bluetooth (Stange et al., 2011; Versichele et al., 2012), Wi-Fi and RFID (Kontokosta & Johnson, 2017). Novel data in terms of applicability to population mapping are non-satellite-based night-light data (Dobler et al., 2015), and points of interest (POIs) such as accommodation establishment data to analyse tourism intensity (Batista e Silva et al., 2018).

Different kinds of novel data sources have different strengths and weaknesses due to their characteristics. Subsequently, scholars have tried to overcome the limitations of certain novel data sources through *data fusion*, namely by combining them with traditional and/or other novel data sources. Recent studies, where traditional data have been used to refine the population distribution from novel data sources include a study by Järv et al. (2017a), who used an advanced dasymetric interpolation method to estimate the dynamic population in Tallinn, Estonia on hourly level using mobile phone data in combination with data on buildings, land use and time use statistics and a study by Patel et al. (2017), who in turn used geo-located Twitter posts together with census data and several ancillary datasets including land cover data, roads and remotely sensed night-lights to estimate the ambient population of Indonesia. Examples, where different types of novel data have been used together, include for example a study by Tu et al. (2017), who combined mobile phone data with social media data to analyse the spatiotemporal variation of urban functions and their relation to land use in Shenzhen, China and a study by Yu et al. (2018), where taxi trajectories based on GNSS tracking were used in conjunction with remotely sensed night-time lights and census data to infer the population distribution of Shanghai on 500 m grid cells.

Alongside new opportunities, large volumes of diverse and fine-grained data pose also challenges to researchers. Firstly, data does not equal knowledge. Hence, useful information needs to be mined from big data using appropriate methods. Secondly, existing methods and processes tailored for traditional data may not be suitable to big data requiring adaptation of current and development of new approaches (see Mennis & Guo, 2009 for an introduction to geographic knowledge discovery and spatial data mining). To extract knowledge and useful information from voluminous and complex data, raw data needs to 1) be preprocessed and cleaned, 2) transformed, 3) mined and the results to be 4) interpreted and evaluated in an iterative fashion (Fayyad et al., 1996; Mennis & Guo, 2009). Additionally, to communicate the results and share the extracted knowledge data also often needs to be visualized. Furthermore, understanding the limitations and quality of data is crucial throughout the process of spatial data mining (Goodchild, 2013; Graham & Shelton, 2013), which is linked to the challenge of finding suitable ground truth data to validate big data -derived information. In conclusion, sufficient technical and analytical skills are required to handle big data (Kitchin, 2013).

2.2. Mobile phone data

2.2.1. Mobile phone data in research

Mobile phones have rapidly become an established source of data in science across disciplines, just when mobile phones were introduced to society (Blondel et al., 2015). Research related to mobile phones has since seen a significant continuous growth retelling the pattern of the global rise of mobile penetration rates (Figure 5, Figure 6). Accordingly, mobile phone data has been used in a myriad of application fields, including human mobility and identification of meaningful places (Ahas, Silm, et al., 2010; Isaacman et al., 2011; Järv et al., 2014, 2015), social network analysis (Eagle & Pentland, 2006; Onnela et al., 2007; Eagle et al., 2009), land-use detection (Soto & Frías-Martínez, 2011; Toole et al., 2012; Pei et al., 2014), traffic flow estimation (Caceres et al., 2012) and epidemiology to study the spread of infectious diseases (Wesolowski et al., 2012; Bengtsson et al., 2015; Finger et al., 2016) among many other domains. Analysis based on mobile phone data have also been commonly used outside the scientific community for example in digital forensics (Mylonas et al., 2012). Most studies have used data from a time period of one day to a few weeks, but also longitudinal studies with data from six months (González et al., 2008) and the span of a year (Phithakkitnukoon et al., 2012; Wesolowski et al., 2012; Järv et al., 2014) up to five years (Kuusik et al., 2009) exist in literature.

Although mobile phone data is sometimes referred to as one data source, a variety of different types of mobile phone data exist and have been used in literature. The most common type of mobile phone data used in urban sensing and travel behaviour research is *cellular network data* (Froehlich et al., 2009; e.g. Becker et al., 2011; Isaacman et al., 2011), also known as *mobile network data* (e.g. Calabrese et al., 2014; Oliver et al., 2015; Rojas et al., 2016). Cellular network data is collected by mobile network operators and automatically generated when mobile phones interact with the cellular network infrastructure using long-range radio-frequency signals (Steenbruggen et al., 2015; Zhenzhen Wang et al., 2018). Cellular network data can be either *passively* or *actively* collected (Ahas, Silm, et al., 2010). Active mobile positioning refers to tracking a mobile phone using for example a radio wave query, where the location of the phone is typically determined through triangulation. Passively collected mobile phone data, on the other hand, refers to data automatically stored as logfiles by mobile network operators (MNO) (Ahas, Silm, et al., 2010). In passive mobile positioning, the location of the phone is usually stored on base station or antenna-level (see, section 2.2.2). Most studies are based on data captured using a passive approach, as is the case in this study.

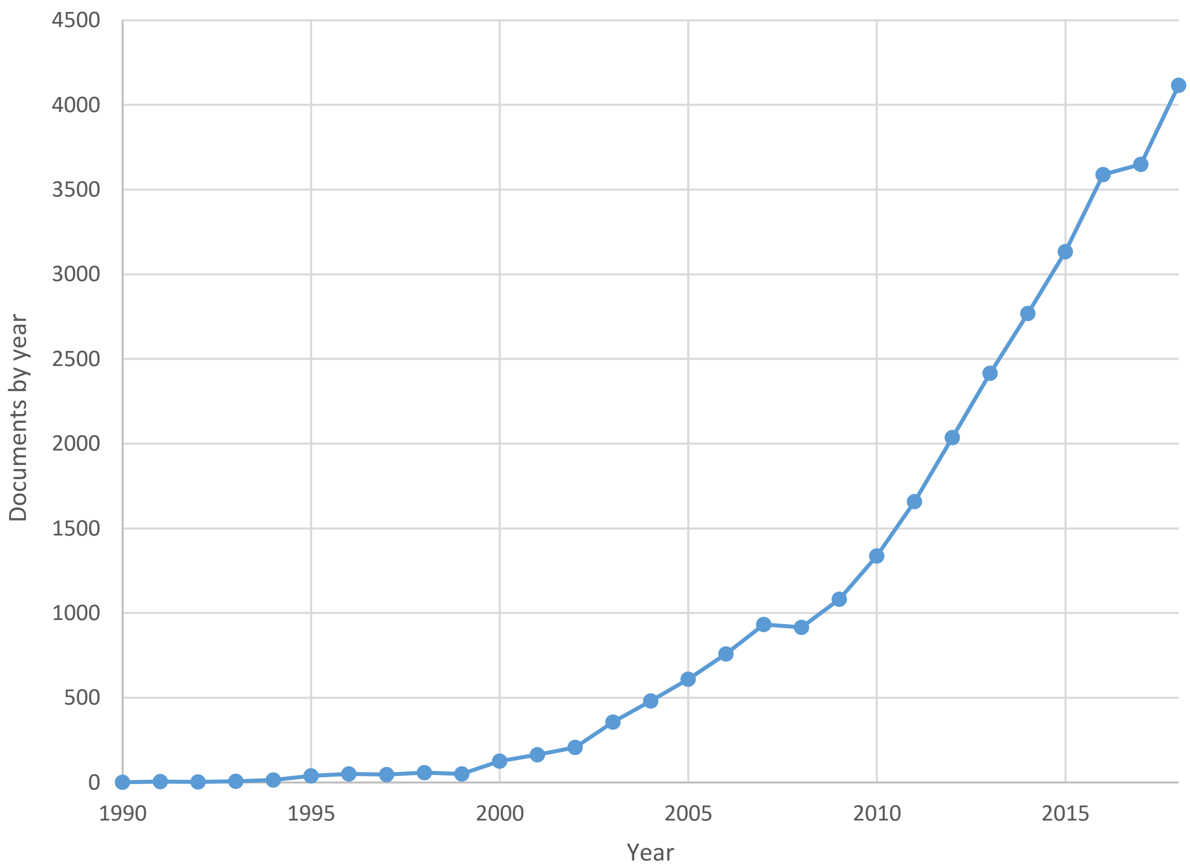


Figure 5. The general development of mobile phone research measured by the yearly number of published documents. The results are based on a literature search in Scopus abstract and citation database using the search query “TITLE-ABS-KEY (("mobile phone" OR "cellular phone" OR "cell phone" OR "smartphone" OR "mobile positioning") AND data))”. The search was conducted on 1.2.2019. Results before the year 1990 and from 2019 were excluded.

Another type of mobile phone data based on long-range radio frequency positioning is data based on the Global Navigation Satellite System (GNSS) or assisted GNSS (commonly referred to as A-GPS), which combines satellite-based and cellular mobile phone positioning (Ahas & Mark, 2005; Ratti et al., 2006). The benefit of GNSS-based mobile phone data is in particular its high spatiotemporal accuracy (Renso et al., 2008), but such data can usually provide only a limited sample of the population as the data collection requires active involvement of the participants, unlike passive data collection methods. Human mobility and presence has also been analysed using short-range radio frequency data, such as Bluetooth and Wi-Fi (Stange et al., 2011; Delafontaine, Versichele, et al., 2012; Versichele et al., 2012, 2014; Kontokosta & Johnson, 2017). Short-range radio frequency data can be used to passively derive human behaviour both on individual or aggregated level using a beacon-based tracking system, although they are also used in active positioning. Bluetooth and Wi-Fi data have been applied particularly in analysing spatiotemporal human behaviour during mass events and in indoor environments, where GNSS is not applicable (Versichele et al., 2012).

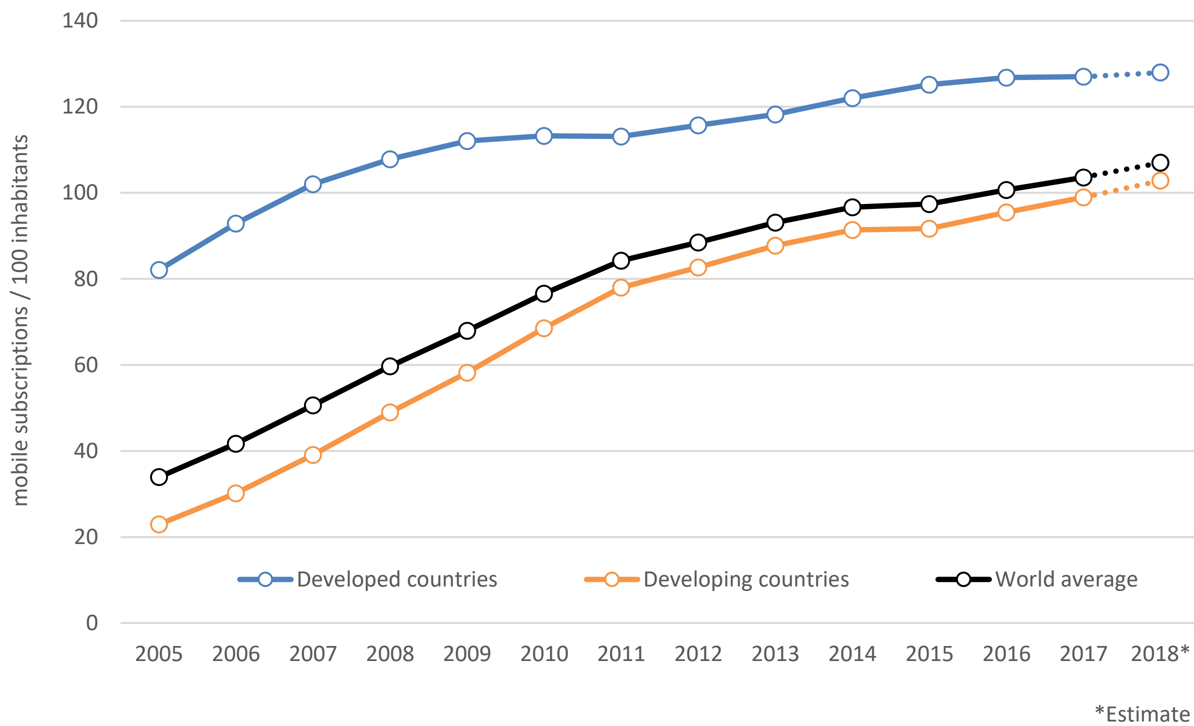


Figure 6. Global development of mobile penetration rates 2005–2018 (ITU 2018a).

Thanks to the proliferation of smartphones, also new kinds of mobile phone data are becoming available to researchers. These include RFID data, which is based on short-range radio frequency positioning similarly as Bluetooth and Wi-Fi (Ni et al., 2003), and data collected by active positioning methods using for example mobile application -based tracking with various positioning methods, such as A-GPS or Wi-Fi and non-radio frequency -based sensor data, generated for example by magnetometers, gyroscopes or accelerometers embedded in smartphones (Pei et al., 2013; Zhenzhen Wang et al., 2018). A comprehensive review on different mobile phone data sources used in travel behaviour research can be found from Zhenzhen Wang et al. (2018).

Only little research has been conducted to compare different mobile phone datasets and their suitability for analysing human presence (Pinelli et al., 2015). Passively collected cellular network data has nonetheless several advantages over other types of mobile phone data. Firstly, it is collected automatically by mobile network operators and can provide a large sample size without requiring to set up a separate infrastructure for tracking phones or collecting a sample of the population, as in active positioning (Ahas, Silm, et al., 2010). Secondly, passive cellular mobile phone data is collected regardless of if users have a specific mode activated, unlike GNSS, Wi-Fi or Bluetooth data. Thirdly, the generation of cellular mobile phone data is not dependent on mobile phone type, unlike data collected by certain censors embedded to smartphones, and is free of sampling issues caused by smartphone ownership. Furthermore, passively collected cellular mobile phone data is free from potential bias caused by users being aware of tracking that may affect data collected by active approaches.

2.2.2. The structure of a cellular network

A *cellular network* (or mobile network) is an infrastructure that enables wireless communication of mobile devices, such as mobile phones. The term cellular network refers to the modular structure of the network infrastructure, which consists of interconnected *base stations*. Base stations, also referred to as "base transceiver stations" in second (2G) or "node B" in third generation (3G) telecommunication networks, act as links between the users and the core network. The wireless communication between base stations and user equipment (e.g. mobile phones) is based on radio wave signals on different frequency bands. (Kaaranen et al., 2001; Ricciato et al., 2015).

Most base stations are equipped with two or more directed antennae that form sectors based on the direction they are pointing at (Oliver et al., 2015). This improves the traffic capacity of the base station. The optimal and most applied solution is three antennae per station with 120-degree sectors. A base station can alternatively have one omnidirectional antenna, which can send signals to all directions (Janecek et al., 2015). Each antenna in a cellular network has a known location and a unique identifier, which enables positioning of mobile phones and therefore spatial analysis of network activities (Ahas, Silm, et al., 2010). In most studies based on cellular network data, mobile phones are located on base station or antenna level, but also different triangulation approaches are used (for an overview of different positioning methods, see Adams et al., 2003; Renso et al., 2008;

Calabrese et al., 2014). Sometimes, the term base station is used interchangeably with the word antenna or cell (Oliver et al., 2015; Ricciato et al., 2015), although most scholars consider these are separate concepts (e.g. Ahas, Silm, et al., 2010). The base station itself can be both a tower-like structure (Figure 8) or individual antennae can be mounted on the sides of a building, which is common especially in densely built-up areas (Figure 7).

Figure 7. A base station antenna mounted on the side of a building.

The coverage areas of base stations are typically schematically visualized in a simplified manner as adjacent hexagons (see e.g. Caceres et al., 2007; Bayir et al., 2009; Calabrese et al., 2014). In reality, the coverage areas of base stations are not fixed and overlap each other (approximately to 20–30 %) to avoid gaps in the coverage (Ahas, Silm, et al., 2010; Caceres et al., 2012). Also the size of the coverage areas varies in across space according on the density driven by the demand of mobile network capacity and transmitting capabilities of base stations (Oliver et al., 2015). The base station network is typically densest in urban areas, where it can reach block level in city centres or similar hot-spots (Caceres et al., 2007; Järv, Tenkanen & Toivonen, 2017a; Ricciato et al., 2017). In rural areas, on the other hand, where the network traffic load is usually lower, the radius of the theoretical coverage areas can reach a few tens of kilometers (Ricciato et al., 2015).

Each base station belongs to an entity of one or more base stations called a location area, which is controlled by a radio network controller (RNC) in 3G telecommunication networks (Calabrese et al., 2014; Ricciato et al., 2015). One of the tasks of the RNC is to control for situations when a user moves within the network, for example from the coverage area of a base station or location area to another (Kaaranen et al., 2001; Calabrese et al., 2014; Zhao et al., 2016). The network also generates updates on regular intervals that provide information to which base station the user equipment is connected to at the given moment (Calabrese et al., 2014). The cellular network can also records situations when the user for example switches between the system type (e.g. transition from 3G to 4G network) or network operator (Kaaranen et al., 2001; Calabrese et al., 2014; Pinelli et al., 2015).

Figure 8. A base station with antennae attached to a tower.

Related to the temporal variation of base station and antenna coverage areas, the user equipment is not always served by the closest base station. If the nearest base station (or antenna) is congested, the user equipment may be automatically switched to a base station further away to optimize the signal quality and distribution of traffic load (Ahas, Silm, et al., 2010). The nearest base station may not necessarily either have the best visibility or widest signal (Ahas, Silm, et al., 2010). Due to the optimization algorithms in the network, the user equipment may shift between base stations even when the device is not actually moving. This phenomenon is known as oscillation, the "ping-pong effect" or "tossing" and is a potential source of error inferring human mobility or presence from mobile phone data (Bayir et al., 2009; Ahas, Silm, et al., 2010; Wang & Chen, 2018).

2.2.3. Passive cellular network data types

All cellular network -based mobile phone data originates from interactions between mobile phones and the cellular network infrastructure (for an overview of the structure of the cellular network and related central concepts and terminology, see section 2.2.2). Similarly as the term mobile phone data, the term cellular network data does not stand for merely a single data source, but a collection of different data products stored and collected for specific purposes by MNOs, such as billing or network monitoring, optimization and maintenance.

Passively collected cellular network data has commonly been divided into network-driven and event-driven mobile phone data based on how the data is generated (Calabrese et al., 2014; Oliver et al., 2015; Pinelli et al., 2015). Event-driven data is produced when a user creates an "event", i.e. makes a request to use network resources, such as makes phone call, sends an SMS or uses mobile data (Internet access) (e.g. Calabrese et al., 2014; Oliver et al., 2015). This type of data is usually used for billing purposes (Ahas, Aasa, et al., 2010; Zhao et al., 2016). Network-driven data, on the other hand is automatically generated by the cellular network even if a user does not actively use the phone for example when moving within the network (see section 2.2.2 and Calabrese et al., 2014) and used primarily for monitoring and optimizing the load of traffic of the telecom network.

Regarding event-driven data, one of the most commonly used data sources are Call Detail Records CDR (Janecek et al., 2015; Steenbruggen et al., 2015; Ricciato et al., 2017; Zhenzhen Wang et al., 2018). The content and the level of aggregation of CDR data is not uniform across studies and the data products provided for researchers can vary between different mobile network operators

(MNO) or research topic (Calabrese et al., 2014; Oliver et al., 2015; Ricciato et al., 2017). Typically, CDR data tends to contains at least the timestamp and duration of each event or transaction, the transaction type (e.g. phone call), an identifier and coordinates of the base station or antenna that the user or users were connected during the transaction and an anonymized user ID of the initiator and possible recipient of the transaction (e.g. Ahas, Silm, et al., 2010; Calabrese et al., 2014; Oliver et al., 2015).

There exist also differences in terminology between scholars, for example related to the inclusion of data use in CDR data. Some studies have separated CDR data from mobile data transmission and refer to the latter as Internet Protocol Detail Record (IPDR) data (Calabrese et al., 2014; Pinelli et al., 2015; United Nations, 2017), whereas others see Internet access logs as a part of CDR data (Järv et al., 2015; e.g. Graells-Garrido et al., 2017; Ricciato et al., 2017). In addition to IPDR, also other related terms including Data Detail Record (DDR) (Saluveer & Ahas, 2014, p. 225; Ahas et al., 2015; Srivatsa et al., 2017) and x-Detail Record (xDR) (de Montjoye et al., 2018) have been proposed in literature, although the definitions of these have been ambiguous. Regardless, studies that have reported using data transmission records as a proxy for people have so far been few (exceptions include Graells-Garrido et al., 2017; Yihong Wang et al., 2018) despite the increased rates of smartphone ownership and Internet use on mobile phones (ITU 2018b).

The concept of network-driven data appears in literature as even more diverse and ambiguous than that of event-driven data. Most studies that refer to network-driven data include some form of location updates (see Calabrese et al., 2014). These updates can be triggered by the movements of the user (e.g. regular update, mobility location update or handover) (e.g. Caceres et al., 2012; Sagl et al., 2012; Calabrese et al., 2014; Zhao et al., 2016) or generated on predefined intervals to query the locations of phones connected to the network (Calabrese et al., 2014; Janecek et al., 2015; Pinelli et al., 2015; Huang et al., 2018; Ni et al., 2018; Wang & Chen, 2018). The latter process is also called sighting (Chen et al., 2014; Wang & Chen, 2018). Another common network-driven data source is Erlang data (Reades et al., 2007; Sevtsuk & Ratti, 2010; Kang et al., 2012). Erlang is a standard unit for measuring the traffic load of the cellular network. One Erlang equals to the traffic load caused by one user calling for one hour (e.g. Ratti et al., 2006). By the broadest definitions, network-driven data encompasses event-driven updates, such as data and time of calls or data use (e.g. Oliver et al., 2015).

Although only few attempts have been made to compare different cellular network data (see Kang et al., 2012; Pinelli et al., 2015; Zhao et al., 2016), the current body of literature already suggests that the variety of different cellular network dataset are suitable for different research purposes. For example Zhao et al. (2016) found that different cellular network data including various network-driven datasets vary significantly in terms of temporal patterns. Kang et al. (2012) in turn found the aggregated hourly number of voice calls to be a better proxy for people than Erlang data in their study in Harbin.

Several scholars have also addressed the issue of sparse temporal accuracy of CDR data (e.g. Ahas et al., 2015; Burkhard et al., 2017; Ricciato et al., 2017). In contrast to event-driven data, which saves the location of the user only when an event is made (Phithakkitnukoon et al., 2010), network-driven data can save the locations also of phones in idle mode (e.g. Janecek et al., 2015). In addition, some studies suggest that phone calls, which are usually the main content of CDR data, are not evenly distributed in time (e.g. Candia et al., 2008; Zhao et al., 2016), although also contradictory results have been found by Kang et al. (2012). In the seminal paper by González et al. (2008), the average interval between call activities was reported to be as high as eight hours. In a study by Widhalm et al. (2015), the temporal frequency of data was found to be higher in the network-driven dataset compared to CDR data, especially during night time and mornings, which suggests that network-driven data may be more feasible for dynamic population mapping than event-driven data. However, the compared datasets in the study by Widhalm et al. (2015) were from different geographical areas and the used CDR dataset did not include data use records, which may have had an impact on the results. Nonetheless, including data use records to CDR data has the potential to mitigate the limitation of temporal sparsity in event-driven data for population mapping, as pointed out by Burkhard (2017). These findings highlight the importance of harmonized terminology and transparent description of data and methods as well as the need for reliable comparisons between datasets for arriving at an understanding of the limitations and feasibility of different data sources as a proxy for people or more broadly.

As presented, both event-driven and network-driven cellular network data have been used to analyse population dynamics. Often a more useful approach to categorise cellular mobile phone data from the perspective of analysing human mobility and presence is, however, to consider its level of spatial, temporal and user-level aggregation. For example, if data is not available on individual level, it is not suited for deriving trajectories for person-based mobility research but can be a useful source for analysing human presence. Furthermore, even though event-driven cellular network data has mostly been available on an individual-level enabling origin-destination (OD) analysis (see e.g. Järv et al., 2014), it can be aggregated. Likewise, network-driven data, which in many studies has been aggregated (e.g. Louail et al., 2014), may be available on individual level (e.g. Kang et al., 2012; Widhalm et al., 2015; Huang et al., 2018) and used to analyse for example mobility flows and get a deeper understanding of the dynamics behind population distribution. Hence, the feasibility of data should not be determined solely by how the data was initiated, which is further stressed by the existing ambiguity of event- and network-driven data but weighed from the viewpoint of the research questions.

2.2.4. Assumptions and limitations of using cellular mobile phone data as a proxy for people

Regardless of the level of aggregation or the means of generation of the data, several assumptions and limitations are related to using mobile phone data as a proxy for people. Firstly, it is generally assumed that one SIM card or mobile phone in the dataset represents one individual (Järv et al., 2015). Although studies have shown this approach to be generally feasible (e.g. Ahas, Silm, et al., 2010; Järv et al., 2014, 2015; Tranos & Nijkamp, 2015), this assumption can also be a potential cause of uncertainty and error. One person may have multiple SIM cards and devices connected to the cellular network, devices can be shared, some people do not own or carry mobile phones and not necessarily all devices that are connected to the network are devices associated to an individual but devices intended for machine-to-machine communication, such as wireless surveillance cameras, “smart” vehicles or weather sensors (Ricciato et al., 2015). Secondly, the mobile phone data records are typically assumed to represent the presence and mobility of users in time and space (Järv et al., 2015). Another common assumption is that frequently visited locations or locations, where more time is spent, are more likely to be of higher importance to users. This is at the core especially in individual-based analysis of for example human activity spaces (e.g. Järv et al., 2014) or when inferring home, work or other key locations of users (Ahas, Silm, et al., 2010; Isaacman et al., 2011), but also on aggregated level for example when analysing urban vitality (Sulis et al., 2018). However, individuals may systematically not carry or use mobile phones in certain locations or at certain times (Järv et al., 2014). Similarly, previous research suggests that mobile phone use may be systematically higher in certain situations or locations, such as at home, in transit or during special events (Nylander et al., 2009; Ahas, Silm, et al., 2010; Järv, Tenkanen & Toivonen, 2017a).

Cellular mobile phone data in general is also prone to several limitations. Firstly, it usually lacks socio-demographic attributes, which would allow getting a deeper understanding of who the users are and potentially shed light onto the motivations of users to be present in a certain location. Secondly, the spatial resolution of the data is uneven due to the varying density of the cellular network (e.g. Järv, Tenkanen & Toivonen, 2017a), which complicates fusion with other data sources. Different approaches in literature have been taken to tackle the issue, such as reallocation to statistical units using different interpolation methods (see Järv, Tenkanen & Toivonen, 2017a). Cellular mobile phone data may also be subject to sampling issues caused by mobile phone ownership, operator-selection and user characteristics influencing mobile phone data use (Castells et al., 2009). For example, different operators may attract different users for example due to pricing policies and mobile phone use or spatiotemporal behaviour can vary between these groups. Furthermore, finding suitable data for validation of cellular network data can be challenging (Calabrese et al., 2014). Moreover, technical issues related to data quality, such as oscillation, or lack of transparency of data processing can cause uncertainties. Finally, data access can be a barrier for mobile phone data research due to legislation, privacy concerns or other data policy issues, which may limit data availability and use.

2.3. Spatial accessibility

2.3.1. Spatial accessibility as a concept

Although the concept of *spatial accessibility* (hereinafter accessibility) has inspired scholars for decades, there does not seem to exist a consensus in literature of how it should be defined. Indeed, as Peter Gould (1969) states, accessibility is a "slippery notion" and the definition and approach to measure accessibility should be chosen by the case at hand (Geurs & van Wee, 2004). A few commonly referenced definitions, which are relevant from the perspective of this thesis include "*the potential of opportunities for interaction*" (Hansen, 1959, p. 73), "*the ease of an individual to pursue an activity of a desired type, at a desired location, by a desired mode, and at a desired time*" (Bhat et al., 2000, p. 1), "*the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)*" (Geurs & van Wee, 2004, p. 128) and "*the amount and the diversity of places of activity that can be reached within a given travel time and/or cost*" (Bertolini et al., 2005, p. 209).

Stemming from the various definitions in literature, Geurs & van Wee (2004) define four components of accessibility, which are interlinked: 1) the *land-use component*, which reflects the land-use that provides the spatial distribution of demand and supply of opportunities (i.e. the origin and destination locations), 2) the *transportation component*, which reflects the transport system that enables the transition from the origin to the destination location(s) using a given mode of transport 3) the *temporal component*, which reflects the variation of opportunities in time based on the temporal constraints and 4) the *individual component*, which reflects an individual's needs and capabilities. From a purely theoretical perspective, incorporation of all four components would be the preferred means to comprehend and analyse accessibility (Geurs & van Wee, 2004). However, such an extensive approach would be utterly data-hungry and methodologically demanding (Geurs & van Wee, 2004). Furthermore, the high level of complexity would hamper the interpretability and communicability of the results, which has been raised as a key issue for adoption in research and decision-making (Bertolini et al., 2005). Indeed, a central challenge in accessibility modelling is finding a feasible way of balancing with sophistication and understandability (Handy & Niemeier, 1997; Geurs & van Wee, 2004; Curtis & Scheurer, 2010).

As a result of the various existing definitions and facets of accessibility, several approaches have been taken to measure it (Salonen, 2014). Geurs and van Wee (2004) distinguish four types of accessibility measures, namely 1) *infrastructure-based*, 2) *location-based*, 3) *person-based* and 4) *utility-based* measures. Also alternative classifications of accessibility measures have been proposed by several scholars, including those by Handy & Niemeier (1997), Geurs & Ritsema van Eck (2001), Curtis & Scheurer (2010) and Páez et al. (2012). Perhaps the coarsest division among different conceptualizations in literature is the dichotomy of location-based and person-based accessibility measures (e.g. Kwan, 1998; Miller, 2005b; Neutens, Schwanen, et al., 2010; Li et al., 2011; Delafontaine, Neutens, et al., 2012; Horner & Downs, 2014; Boisjoly & El-Geneidy, 2016; Tenkanen, 2017; Järv et al., 2018). Location-based (also known as place-based, area-based or zone-based) accessibility measures describe accessibility to one or more predefined locations, whereas person-based (also known as people-based or individual-based) approaches scrutinize accessibility from the perspective of the abilities and constrains of individuals in space and time (e.g. Miller, 2005b). Although lacking the individual dimension, location-based approaches are well suited for analysis over larger geographical areas, which is a known challenge of person-based approaches (Geurs & van Wee, 2004; Boisjoly & El-Geneidy, 2016). Furthermore, location-based measures are generally considered easier to both produce and interpret than person-based measures, making location-based accessibility a more common option in decision making (Geurs & van Wee, 2004).

Location-based accessibility measures, which are the focus of this study, can be further classified to i) *distance (or connectivity)* and *contour (or isochronic) measures,* ii) *potential (or gravity-based) measures* and iii) *balancing factors* (Geurs & van Wee, 2004). Regarding the scope of this thesis, the most relevant location-based measures are contour measures. Contour measures can be used for example to analyze the cumulative amount of people that access a location in a given time. Metric-wise, studies employing contour measures, and location-based accessibility studies overall, typically use travel distance (e.g. Euclidean or network distance), travel time or the number of opportunities as an indicator of accessibility (Delafontaine, Neutens, et al., 2012). Nowadays travel time has become the most prevalent metric in accessibility literature (Tenkanen, 2017) overcoming the limitations of assuming human movement to follow Euclidean (straight-line) distance (Olsson, 1965) while still remaining intuitive and communicable (Geurs & van Wee, 2004).

2.3.2. Time in accessibility research

The concept of spatial accessibility is dynamic by nature (Hodge, 1997). Even though this fact has been recognized by a large number scholars, the majority of studies assume accessibility to be fixed in time ignoring the fluctuations of people, the transport network and activity locations (Li et al., 2011; Kwan, 2013; Moya-Gómez et al., 2017; Järv et al., 2018).

Although the need for incorporating time beside space in analysis of urban systems and human behaviour was addressed by scholars already in the late 1960s and early 1970s (Hägerstrand, 1970; Batty, 1971), empirical implementation of dynamic models – including those of accessibility – have emerged only fairly recently (Kim & Kwan, 2003). Person-based accessibility research, which is rooted in time geography (Hägerstrand, 1970, see also section 2.1.1), has traditionally included time in accessibility models on a conceptual level (see e.g. Neutens et al., 2012; Widener et al., 2015). Operationalization of these models was, however, postponed for decades due to technical limitations until advancements in geographic information systems (GIS) and computational power in the 1990s enabled bringing the theories into practice (Kwan, 1998). Location-based measures, however, are still predominantly treated as static (Moya-Gómez et al., 2017; Järv et al., 2018), which has led to a temporal divide between location-based and person-based accessibility approaches. The magnitude of this gap and the share of static approaches in the current body of literature is further highlighted, as the majority of accessibility studies are founded on location-based measures (Kwan, 2013; Järv et al., 2018).

To mitigate the issue of static location-based accessibility measures, Järv et al. (2018) proposed a generic conceptual framework for temporally-enabled location-based accessibility modelling, referred by scholars also as *dynamic accessibility* (Li et al., 2011; Moya-Gómez et al., 2017; e.g. Järv et al., 2018). The dynamic accessibility framework distinguishes three core components of accessibility: 1) the whereabouts of people, 2) activity locations (e.g. grocery stores) and the 3) transport network connecting these (Figure 9). Mathematically, dynamic accessibility (DA) can be formulated as:

$$DA = \int (PTA)_{st} \tag{1}$$

where P represents the distribution of people, T the transport network and A the activity locations as a function of space (s) and time (t) (Järv et al., 2018).

In recent years, location-based accessibility studies that consider the temporality of the transport network and opening hours of activity locations have surfaced (e.g. Li et al., 2011; Farber et al., 2014; Tenkanen et al., 2016). Yet, static data representing home and work locations are still as a rule considered as a proxy for origins of people, despite the widely acknowledged criticism of the approach (e.g. Næss, 2006; Miller, 2007). Although home is undoubtedly a central location in people's daily lives, several studies suggest that people seldom stay in only one place, such as home, over the course of a day (e.g. Schönfelder & Axhausen, 2010; Kwan, 2013). Exceptions to static population approaches in place-based accessibility modelling include those by Moya-Gómez et al. (2017), who used Twitter data as a proxy for people to analyse the dynamic accessibility in Madrid and Järv et al. (2018), who analysed the accessibility to grocery stores in Tallinn using mobile phone data (CDR) to estimate whereabouts of people on an hourly intervals. Both studies suggest that dynamic population data provides a more realistic view on accessibility. Similarly as in population mapping overall, the lack of suitable population data has partly hampered the emergence of dynamic approaches (Järv et al., 2018). As shown by the examples above, novel data sources have potential to overcome this issue, and bring more reliable tools for decision makers.

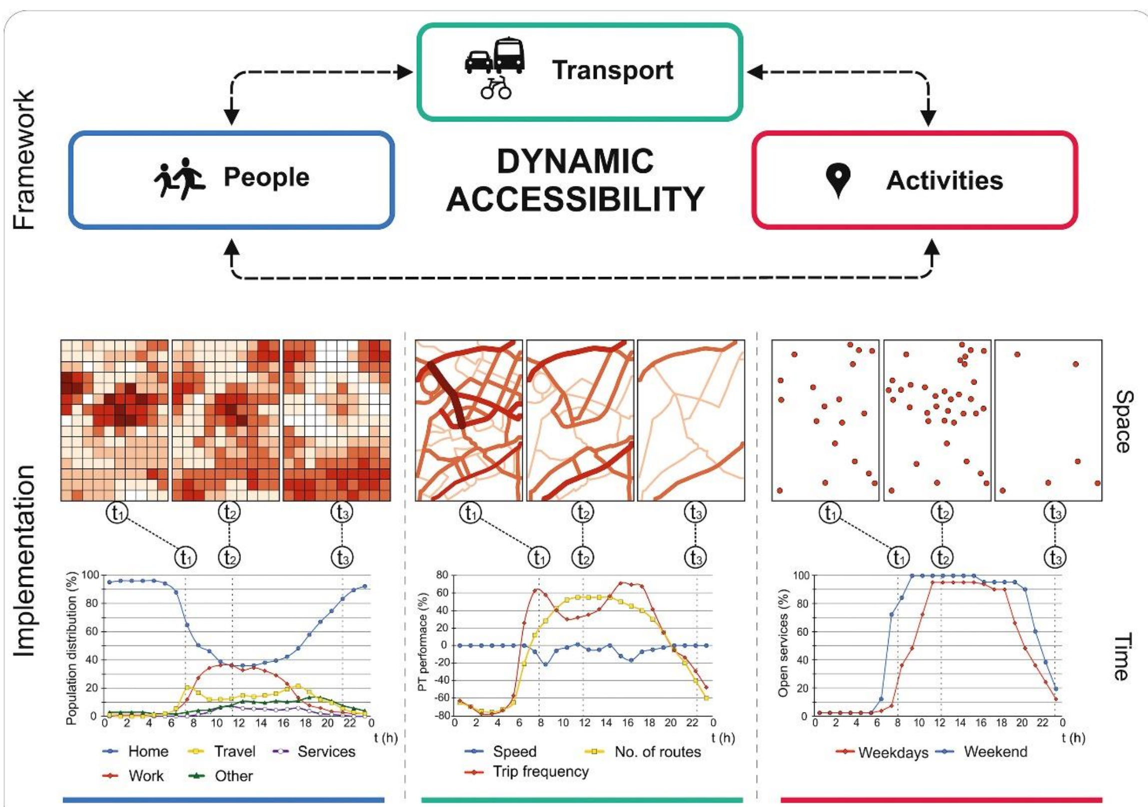


Figure 9. The conceptual framework of dynamic accessibility (adopted from Järv et al. (2018, p. 102)).

3. STUDY AREA

3.1. Helsinki Metropolitan Area as a study area

The study was conducted in the Finnish capital region, which consists of four municipalities: Helsinki, Vantaa, Espoo and Kauniainen (Figure 10). The study area also referred to as Helsinki Metropolitan Area (HMA) covers a total land area of approximately 771 km^2 (NLS 2018c) and has a population of over 1.1 million inhabitants (1 154 967 on 31.12.2017), which represents roughly a fifth (21 %) of the total Finnish population (Statistics Finland, 2018a).

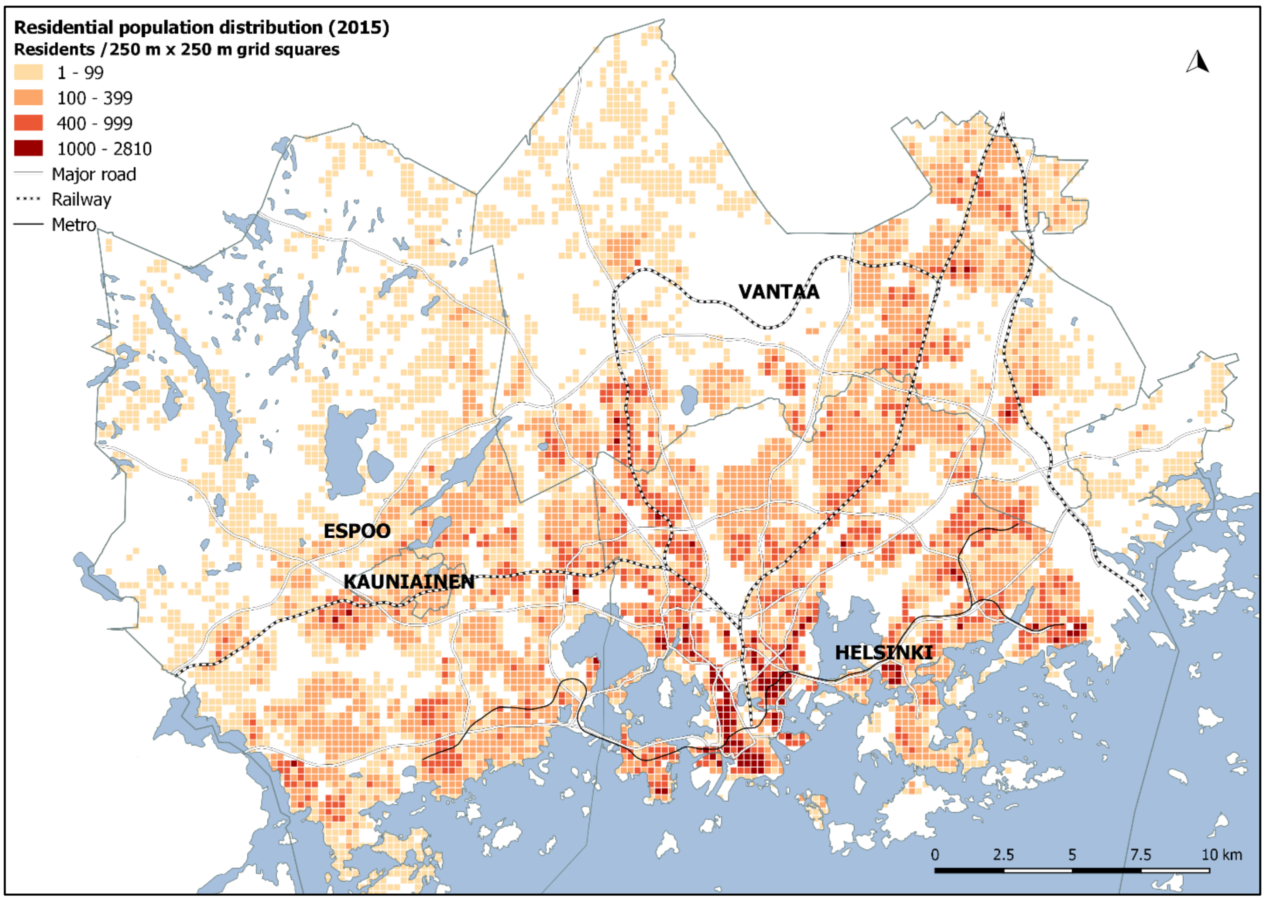


Figure 10. The study area and the number of residents per inhabited 250 m x 250 m statistical grid squares (n=8253) in 2015 (Statistics Finland, 2016; HSY 2018a).

The study area coincides with the existing accessibility tools and datasets developed by the Digital Geography Lab at the University of Helsinki, such as MetropAccess Reititin (Salonen & Toivonen, 2013; Tenkanen et al., 2016), MetropAccess Digiroad (Jaakkola, 2013) and Helsinki Region Travel Time Matrix (Toivonen et al., 2014; Tenkanen et al., 2018).

The average population density in the study area based on residential data is approximately 1,500 people/km^2 (NLS 2018c; Statistics Finland, 2018a), being highest in the inner city of Helsinki, which is located on the peninsula in the southern part of the study area (Figure 10). Relatively high residential concentrations can also be found along the railway and metro stations. The north-western parts of Espoo and the easternmost parts of Helsinki and Vantaa are comprised mainly of forest and are thus relatively sparsely populated. Other areas with a low residential use include the two airports in the region, namely Helsinki-Vantaa airport in the central part of Vantaa and Helsinki-Malmi airport in north-eastern Helsinki, and their surrounding industrial areas and other airport-driven functions, as well as the Helsinki Central Park that stretches from Töölö Bay in the inner city of Helsinki to river Vantaanjoki in the north at the border of Helsinki and Vantaa.

Together with the railway network, the major radial roads and ring roads (I-III) form the main arteries for transportation in the study area. The local public transport network comprising of bus, tram, train, metro and ferry connections is dense (Figure 11) and the system has also gained international acclaim (Curtis & Schreurer, 2015). Although the modal share of public transport (27 %) has started to grow during the last decade, private car is still the dominant mode of transport (37 %) in the region (HRT 2013).

In general, the accessibility of local services, such as libraries, can be considered relatively good in HMA (Salonen et al., 2012). Previous research shows, however, that the accessibility realities differ significantly in terms of transport mode (e.g. Salonen et al., 2012; Tenkanen et al., 2016). Indeed, the statistical grid cells with best (10 %) accessibility by private car are located in central part of the study area along the major roads, whereas the corresponding public transport cells are mostly located in the Helsinki inner city, where the public transport network based on the amount of departures is densest, and along the railway and metro stations (Toivonen et al., 2014; Figure 11).

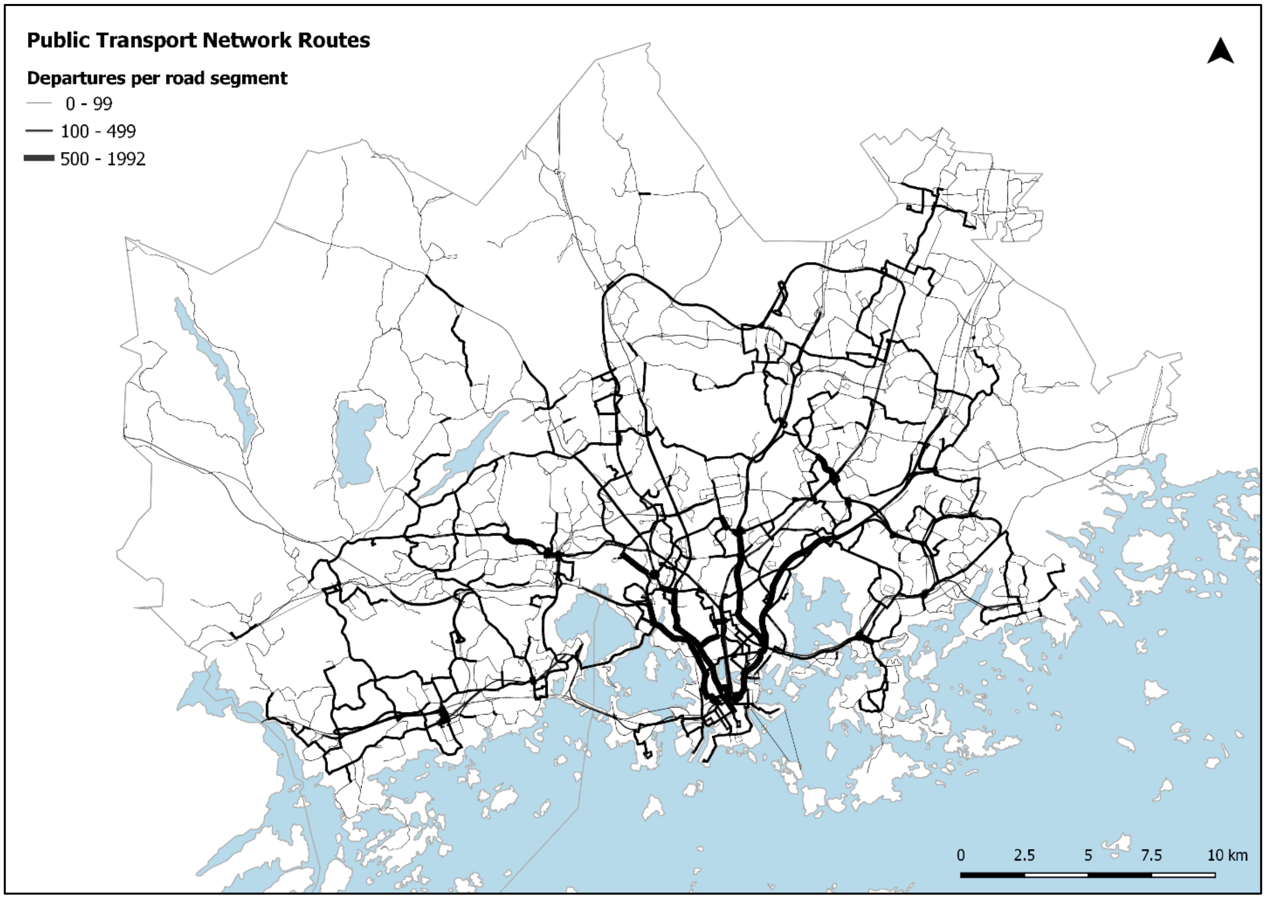


Figure 11. Public transport routes and their service densities as the number of daily departures per road segment (between intersections) in the Finnish capital region in 2016 (HRT 2018). Service density data of the metro line were not available.

Helsinki, and the capital region in general, is a nationally significant travel-to-work area, which has a direct impact on the daily population dynamics in the form of commuting flows. According to recent employment commuting statistics produced by Statistics Finland (2018b), approximately 19 % of the over 600 000 employed persons working in the HMA commute from outside the area (situation 31.12.2015) (HSY 2018b). The shares of commuters to the HMA are highest in the surrounding municipalities, out of which in Kirkkonummi (60.5 %), Nurmijärvi (53.1 %), Sipoo (52.7 %), Kerava (52.2 %), and Tuusula (50.1 %) the majority of the employed residents have their work place in the HMA (HSY 2018b).

Overall, the highest fluctuation in population caused by daily commuting can be found in Helsinki with over 87 000 net commuters (Table 1). The biggest share of commuters of the total employed workforce in the municipality is in Kauniainen (64.8 %), followed by Vantaa (57.5 %), where similarly over half of the workers are inbound commuters. The major workplace areas are located in the inner city of Helsinki and along the main transport ways for example in Pitäjänmäki, Leppävaara and Keilaniemi in South-Eastern parts of Espoo and Tikkurila and areas close to the airport in Vantaa (Figure 12).

Table 1. Residents, employed labour force and daily commuting in Helsinki Metropolitan Area, 2016 (Statistics Finland, 2018b).

Municipality	Residents (2017)	Persons working in the area	Persons working in their municipality of residence	Inbound commuters	Outbound commuters	Net commuting[a]	Commuters of persons working in the area [%]	Change to (residential) population by commuting [%]
Espoo	279 044	117 007	62 269	54 738	65 441	-10 703	46.8	-3.8
Helsinki	643 272	388 005	233 504	154 501	66 929	+87 572	39.8	+13.6
Kauniainen	9 624	2 336	833	1 513	3 165	-1 652	64.8	-17.2
Vantaa	223 027	110 784	47 084	63 700	55 837	+7 863	57.5	+3.5

[a] The difference between inbound and outbound commuters

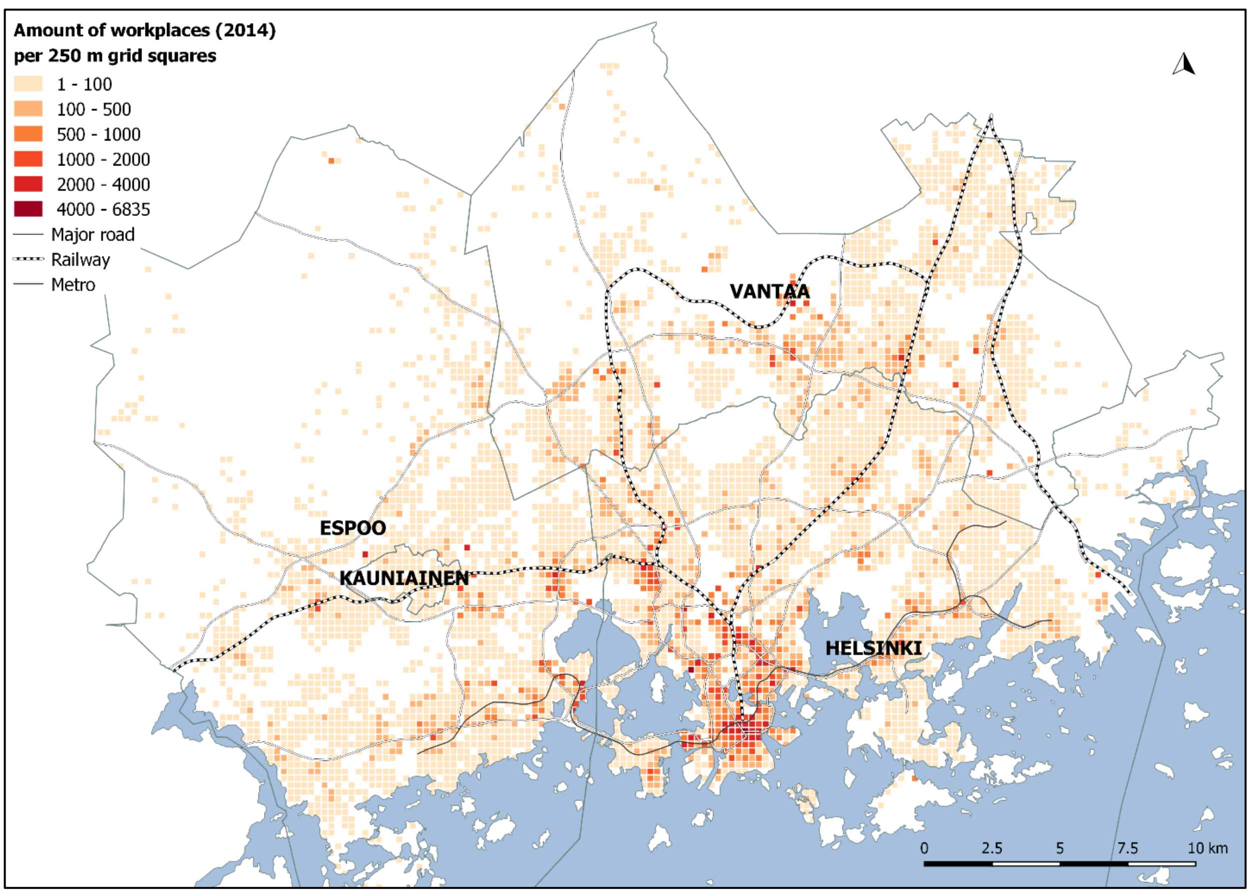


Figure 12. The distribution of workplaces is the study area on 250 m grid squares in 2014 (Statistics Finland, 2016; HSY 2018a).

3.2. Mobile phone use in Finland

At the end of 2017, the mobile phone penetration rate (mobile subscriptions = SIM cards/100 inhabitants) of Finnish households was 126 % with approximately 6 960 000 mobile subscriptions (FICORA 2018), which is above the global and the European average rates (103.6 %, 120.4 %) based on the global statistics by the International Telecommunication Union (ITU 2018a). All in all, households make up approximately 73 % of the total amount of Finnish mobile subscriptions (FICORA 2018). If subscriptions by private businesses are included, the mobile phone penetration rate is over 170 %. Also the adoption of smartphones has been rapid in recent years both in the study area and Finland more broadly. According to a survey carried out by Statistics Finland on the use of information and communications technology in 2018, it is estimated that 80 % of 16–89-year-olds own a smartphone in Finland (Statistics Finland, 2018c). The corresponding rate for the Finnish capital region is 89 % (Statistics Finland, 2018c). In the study, a smartphone was defined as a mobile phone equipped with 3G or 4G broadband and a touch screen.

In addition to a high degree of mobile subscriptions, mobile phones are also actively and widely used in Finland. According to the same survey on ICT use published by Statistics Finland in 2018, 98 % of the total Finnish population and 99 % of the population residing in the Finnish capital region had used a mobile phone at least once during the last three months (Statistics Finland, 2018c). The results of the survey suggest that there is no significant difference between women and men in terms of the given statistics.

Since 2007, mobile phone use has seen a significant change in Finland. The amount of transmitted data has seen an accelerated growth, while the number of text messages and mobile phone calls have simultaneously decreased at a steady pace (Figure 13–Figure 15). The mobile data transmission volume statistics echo the recent findings from the survey on ICT use, according to which mobile phones have become the most common way to access the internet (Statistics Finland, 2018c). It is probable that the proliferation of smartphones and mobile applications that offer voice call and instant messaging capabilities over the internet, fuelled by the development of telecommunication networks, have led to the decline in text messages and “traditional” phone calls over the cellular network. The amount phone calls made from mobile subscriptions by businesses have, however, seen only a gentle decrease.

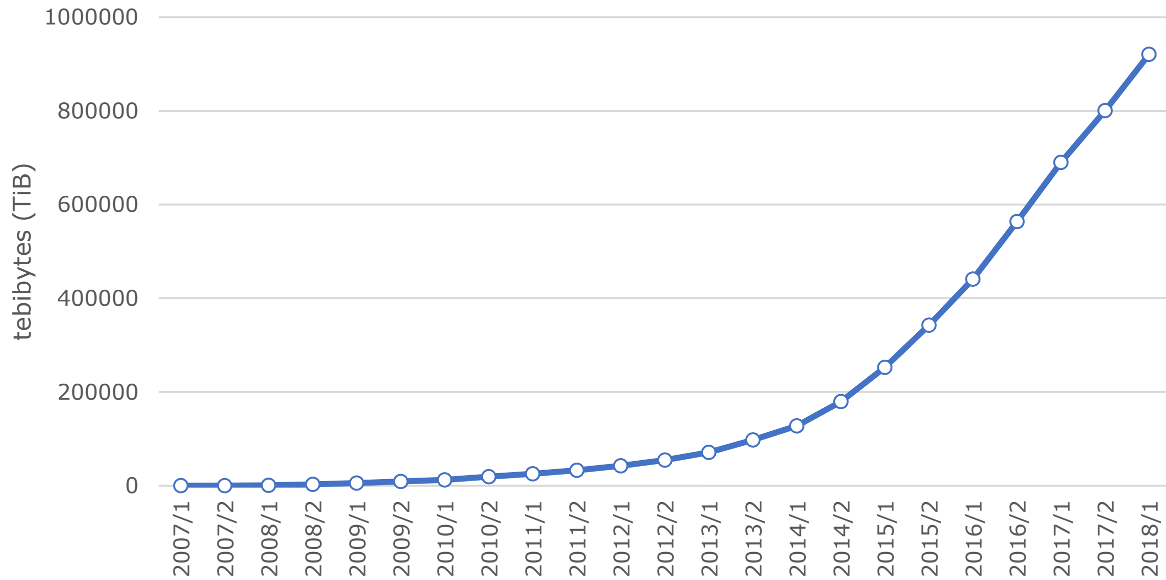


Figure 13. Mobile data transmission volume in Finnish telecommunication networks excluding roaming activity 2007–2018 (FICORA 2018). One tebibyte (TiB) is approximately equivalent to 1.1 terabyte (TB).

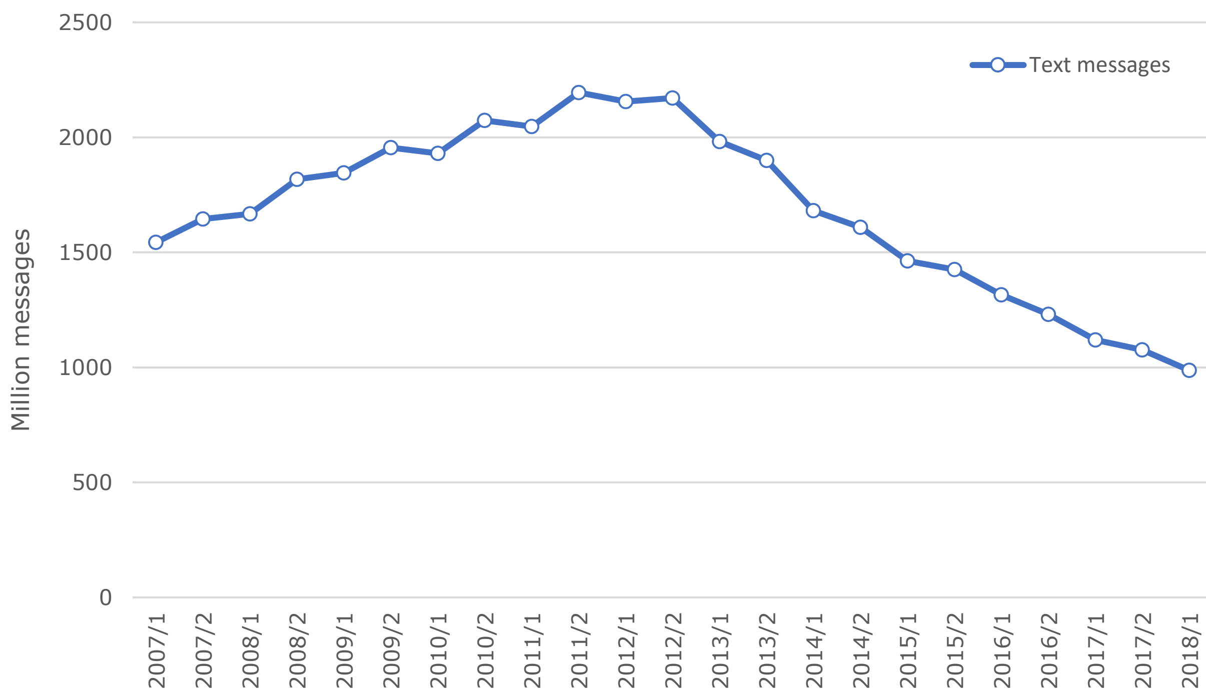


Figure 14. Text message volume in Finnish telecommunication networks 2007–2018 (FICORA 2018).

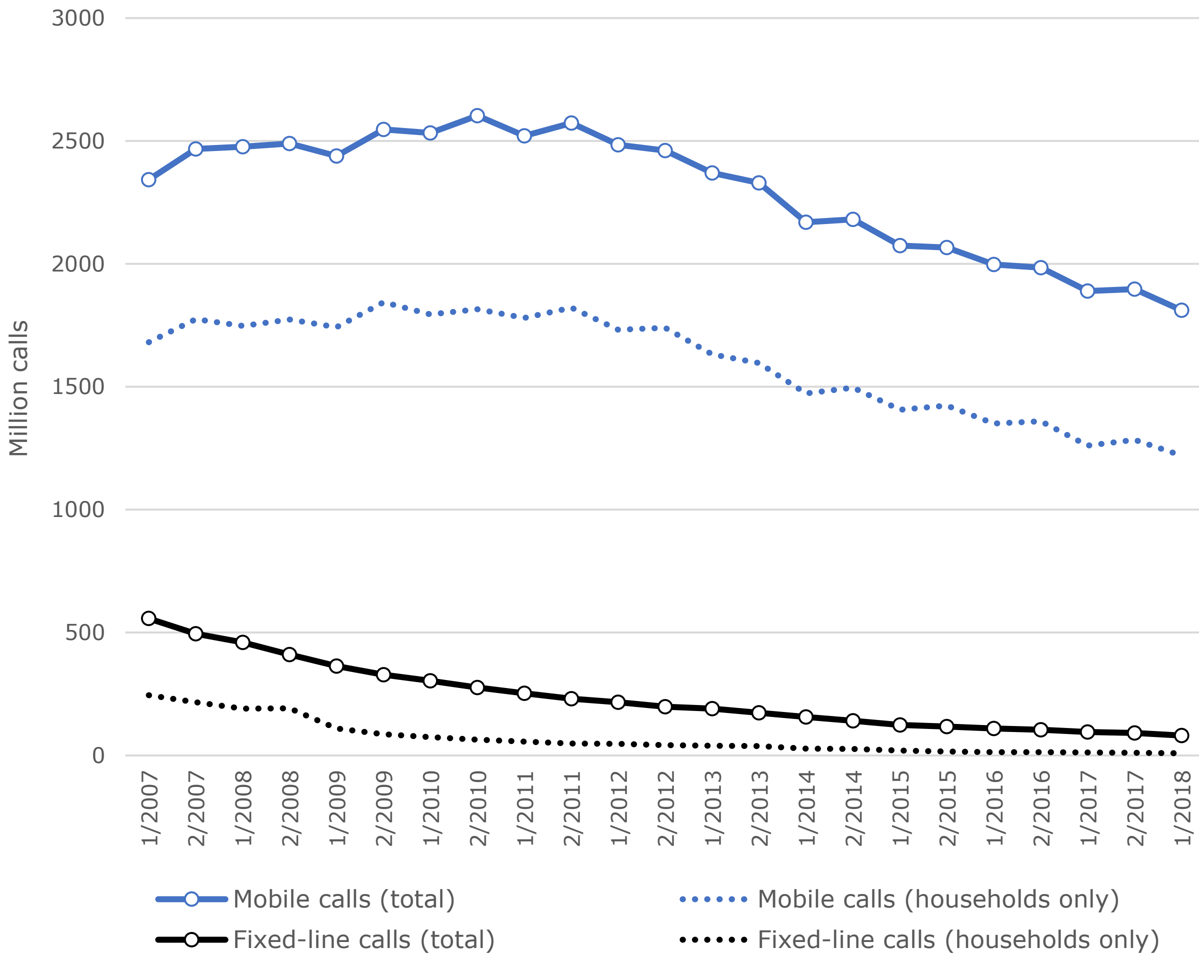


Figure 15. The development of mobile and fixed-line calls volumes in Finnish telecommunication networks 2007–2018 (FICORA 2018).

4. DATA

4.1. Overview of used data sources

Several datasets from various sources were used in this study. The used datasets are listed below in Table 2 and Table 3. The datasets are described in more detail in sections 4.2 Mobile phone data and 4.3 Ancillary data.

Table 2. Original data sources used in the study for dasymetric interpolation of mobile phone data.

Dataset name	Description	Source	Phase of analysis	Open data
Mobile phone data	Hourly network-driven mobile phone data records aggregated on BS level (2017/10–2018/01).	Elisa Oyj (2018)	Extracting spatio-temporal population patterns	No
Time use data	The average time use by activity and location on weekdays (Mon–Thu) on 10 min intervals in HMA. The data is based on a decennial time use survey carried out in 2009-2010.	Statistics Finland (2010)	Creating a probability matrix for human presence	No
Land cover data	Corine Land Cover 2012 raster (20 m x 20 m). Updated every six years.	Finnish Environment Institute (2012)	Allocating mobile phone data to land use areas based on time use data	Yes
Building data	Building polygon footprints and use type from the national topographic database. Updated annually.	National Land Survey of Finland (2018)	Allocating mobile phone data to buildings based on time use data	Yes
	Building polygon footprints, floor area and floor count of buildings in Helsinki. Updated daily.	City of Helsinki (2018)	Assigning floor area to buildings	Yes
	Building polygon footprints, floor area and floor count of buildings in Espoo. Updated weekly.	City of Espoo (2018)	Assigning floor area to buildings	Yes
	Building polygon footprints, floor area and floor count of buildings in Vantaa. Updated weekly.	City of Vantaa (2018)	Assigning floor area to buildings	Yes
	Buildings polygon footprints and use type of buildings.	OpenStreetMap (2018)	Refining functionality type of buildings	Yes
YKR grid squares	Empty 250 m x 250 m grid squares (n=13231)	Digital Geography Lab, Statistics Finland (2015)	Target zones for refining mobile phone data	Yes
Grid Database 2016	Registry-based residential population data (2015) and workplace data (2014) on 250 m x 250 m statistical grid cells	Statistics Finland (2016)	Evaluation of dynamic population data	No

Table 3. Data sources used in the study for dynamic accessibility modelling.

Dataset name	Description	Phase of analysis	Source	Open data
Kalkati XML	Public transport (PT) routes, stops and schedules (25.1.2018)	Travel time calculations for public transport with MetropAccess-Reititin	Helsinki Region Transport (2018)	Yes
OpenStreetMap (OSM)	Pedestrian road network	Travel time calculations for walking with MetropAccess-Reititin	OpenStreetMap (2018)	Yes
YKR grid squares	Empty 250 m x 250 m grid squares (n=13231)	Origins for travel time calculations.	Digital Geography Lab[a]	Yes
Grocery store data	Addresses and opening hours of grocery stores by the three largest chains (15.9.2018)	Destinations for travel time calculations.	Websites of grocery stores[b,c,d]	Yes
Transportation hub locations	Coordinates of the Helsinki Central railway station and Helsinki-Vantaa Airport (T2)	Destinations for travel time calculations.	OpenStreetMap (2018)	Yes
Dynamic population data	Dynamic 24-h population data on 250 m x 250 m statistical grid cells	Dynamic population for accessibility curves.	see Table 2	-
Grid Database 2016	Residential population data (2015) on 250 m x 250 m statistical grid cells	Static population for accessibility curves.	Statistics Finland (2016)	No

[a] http://www.helsinki.fi/science/accessibility/data/MetropAccess-matka-aikamatriisi/MetropAccess_YKR_grid.zip
[b] https://www.s-kanava.fi/web/s/myymalat-ja-palvelut
[c] https://www.kesko.fi/asiakas/kaupat/#paivittaistavarakauppa
[d] https://www.lidl.fi/fi/myymalahaku.htm

4.2. Mobile phone data

Network-driven cellular mobile phone data from a two-and-a-half-month study period (28.10.2017–9.1.2018) provided by mobile network operator (MNO) Elisa Oyj was used to analyse the distribution of people in space and time. Elisa Oyj has the largest market share of mobile subscriptions (38 %) in Finland followed by Telia Finland Oyj (34 %) and DNA Oyj (27 %) (FICORA, 2017). The original dataset contains approximately 3.8 million rows of data and covers all base stations by the given operator in Uusimaa region. The dataset was later cropped to the study area extent (Figure 10) and filtered to cover only average weekdays (Monday-Thursday).

The mobile phone data used in this study was (passively) collected, extracted and anonymized by the MNO prior to providing the data. The data was further aggregated to base station (BS) level with the temporal accuracy of one hour. In the case of base stations equipped with multiple directional antennae, the BS coordinates were approximated using the coordinates of the antenna with the maximum X-coordinate value. This only has an impact on the spatial accuracy of the BS coordinates when the antennae were not attached to a mast-like cell tower. In addition, due to confidentiality reasons, a randomized error of ±100 meters was set to BS coordinates in the inner city of Helsinki. Outside the inner city, the error was set to ±200 meters (Figure 16). In general, the spatial accuracy of the data is dependent on the density of the base station network (highest in the city centre and other densely populated areas, where use rates are highest) (Caceres et al., 2007; Järv, Tenkanen & Toivonen, 2017a; Ricciato et al., 2017). The median theoretical coverage area based on Voronoi polygons of the base stations cropped to the study area is 0.24 km^2.

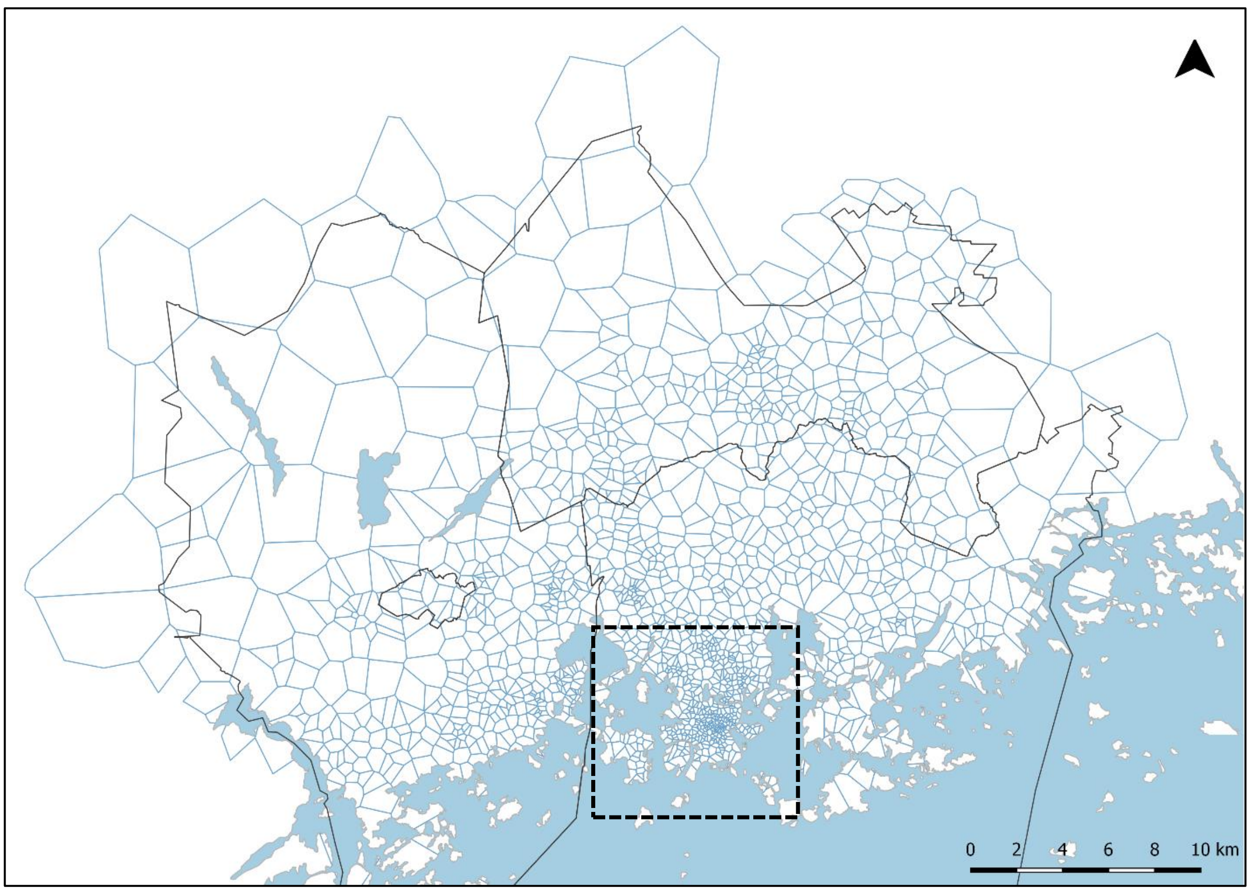


Figure 16. The theoretical coverage areas of all base stations that were used in the study calculated by Voronoi polygons. The black square indicates the approximate borders of the area, within which the coordinates of base stations were applied a random error of ±100 m by the MNO. The error outside the square is ±200 m.

4.2.1. Data content

The dataset contains three types of network-driven mobile phone data: the hourly number of 1) High-Speed Packet Access (HSPA) calls, 2) Radio Access Bearer (RAB) attempts and 3) Radio Resource Control (RRC) connections per base station. In addition, the dataset contains X and Y coordinates (ETRS-TM35FIN) and ID of the base station and a timestamp (YYYY-MM-DD hh:mm:ss) (Table 4). The dataset contains all given information in third generation (3G) telecommunication networks.

Table 4. Illustration of the examined dataset including eight variables (artificially generated pseudodata).

ID	Timestamp	Base Station ID	RRC Connections	RAB Attempts	HSPA Calls	X	Y
1	2018-11-15 08:00:00	000001	50	22	145	369876	6671234
2	2018-11-15 08:00:00	000002	432	178	982	368765	6672345
3	2018-11-15 08:00:00	000003	132	68	545	366543	6673456
4	2018-11-15 09:00:00	000001	89	5	320	369876	6671234

An RRC connection is a prerequisite for using any network resources, such as making a voice call, or browsing the Internet on a mobile phone. In 3G telecommunication networks, RRC connections are generated between the user equipment (UE), such as mobile phone, and the Radio Network Controller (RNC). Once an RRC connection has successfully been established, an RAB attempt can be generated between the UE and the core network. Where RRC is primarily used for controlling the network in form of signalling, radio access bearers are responsible for transmitting voice or data in a cellular network. As a rule, one RAB is created for each resource request, unless the connection is interrupted and a new one needs to be established. HSPA, on the other hand, is a collection of downlink (HSDPA) and uplink (HSUPA) protocols, which enables faster data transmission in a Universal Mobile Telecommunications System (UMTS) cellular network. If HSPA is supported by the network, data transfer can be replaced by HSPA bearers, which are prompted by HSPA call requests. Thus, the RAB attempts in the dataset consists primarily of voice calls, whereas HSPA calls encompass the majority of 3G mobile data transmissions. (Kaaranen et al., 2001; Holma & Toskala, 2006).

According to the operator, the data contains an exhaustive list of voice calls in networks managed by the MNO. The data transfer is however not complete, as content from 4G networks are not included. Hence, the completeness of RAB attempts is higher than that of HSPA calls.

4.2.2. Temporal distribution of data

All three different types of mobile phone data, HSPA calls, RRC connections and RAB attempts, show clear temporal patterns both on weekly and daily level.

On the scale of the whole study period from late October to early January, a recurring weekday-weekend rhythm can be distinguished (Figure 17). The amount of network activity is relatively similar between the weekdays from Monday to Friday. In the case of RRC connections and RAB attempts, Fridays typically exhibit slightly higher activity rates than other weekdays. The amount of generated data typically drops during the weekend days compared to the weekdays with lowest rates on Sunday. The weekly pattern is disrupted during the holiday season, which is displayed commonly as decreases compared to the day of the week average. Examples include Independence Day (Wednesday 6.12.), New Year's Day (Monday 1.1.) and the Christmas days, which all have lower rates than average. These temporal patterns can be identified from all three data types.

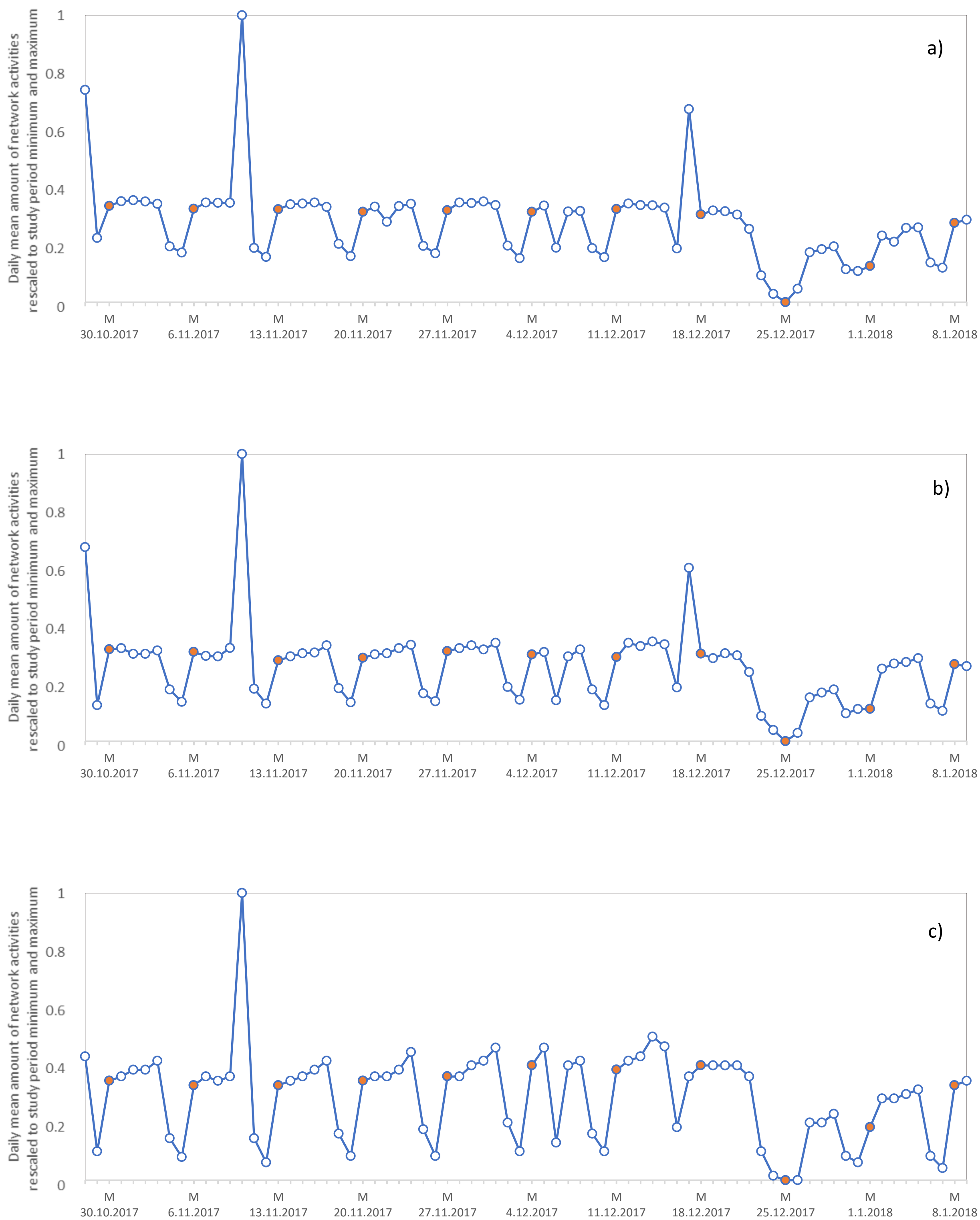


Figure 17. Daily mean amount of a) HSPA calls, b) RRC connections and c) RAB attempts during the study period. The values are rescaled using min-max normalization. Orange markers indicate Mondays (M).

There exists a distinct pattern in the temporal distribution of network activities also on a diurnal level. Within an average weekday (Monday–Thursday), the three different types of mobile phone data follow a similar pattern with lowest values during night-time between 2 AM and 4 AM and highest during the afternoon between 3 PM and 4 PM (up to 10 % of daily total network activity) (Figure 18). Voice calls (RAB attempts) have the highest daily fluctuation and they are strongly concentrated to the typical work hours, particularly to the early afternoon. The highest relative increase of voice calls takes place between the hours starting at 7 AM and 8 AM. Not surprisingly, the amount of voice calls during night-time (midnight–6 AM) is below 1 % of total daily voice call activity. The share of internet use (HSPA calls) and connection requests (RRC connections), on the other hand, are more evenly distributed over the course of the day with hourly shares ranging from approximately 2 % during the night to slightly above 6 % in the early afternoon. The hourly share of generated data of the daily total is higher in these two data types in comparison to voice calls in the early morning, late evening and during the night. The RRC and HSPA rates during the night can be partly due to passive mobile phone use, such as email synchronization of smartphones.

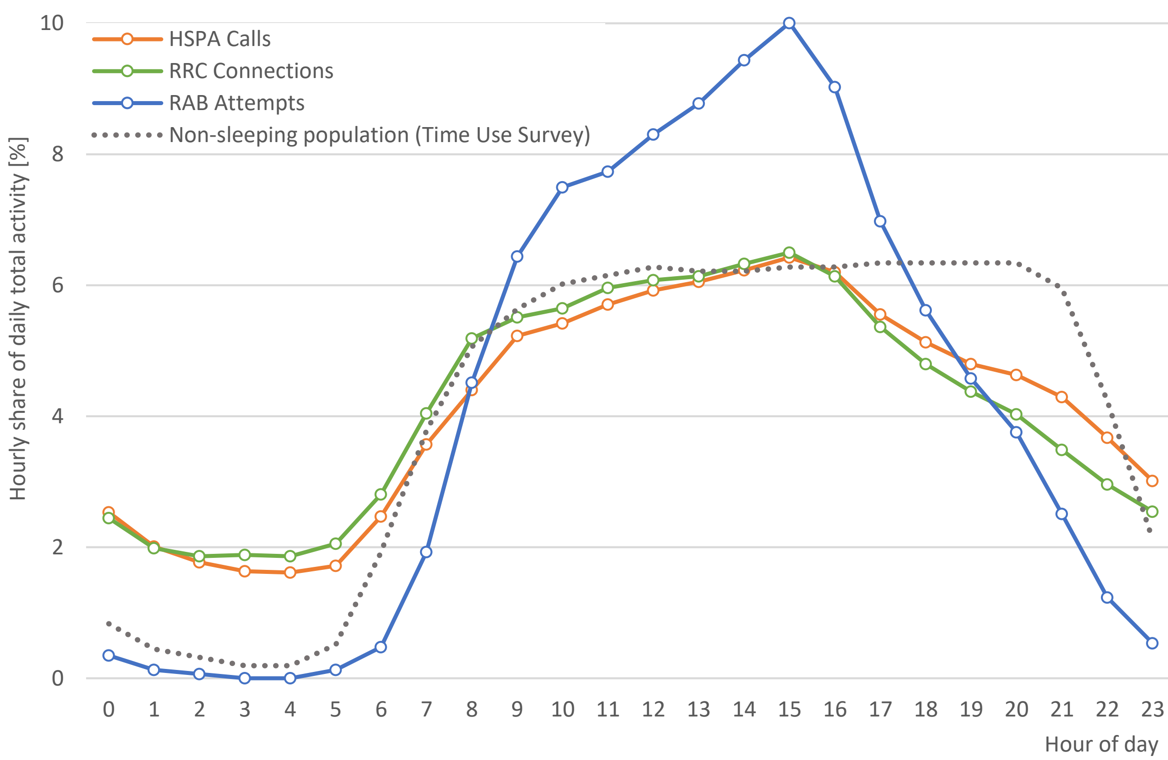


Figure 18. Daily temporal patterns of mobile phone data use and statistical non-sleeping population on an average weekday (Mon–Thu) in HMA.

4.3. Ancillary data

In addition to mobile phone data, several ancillary data sources were used in the study. The ancillary datasets used in the dasymetric interpolation of mobile phone data to 250 m x 250 m statistical grid squares can be categorized into three main categories: 1) land cover, 2) building and 3) time use data. The content of these datasets and their role in the analysis workflow are presented in more detail in chapter 5. Additionally, residential population data and workplace data produced by Statistics Finland (2016) were used for evaluation and validation of the mobile phone data - derived population data. The population dataset is based on the national population information system and contains the sum of permanently residing inhabitants at the end of the year (31.12.2015) on 250 m x 250 m statistical grid cells (Statistics Finland, 2015). The number of workplaces on the same spatial units is based on the number of employed persons in the given square in 2014.

To analyse dynamic accessibility to grocery stores, all grocery stores in the study area by the three largest grocery store chains by market share in 2017 were chosen for the analysis (PTY 2018). These included the grocery stores by the S Group (n=178; Alepa, S-market, Prisma, Stockmann herkku) with the market share of 46.7 %, K Group (n=205; K-Market, K-Supermarket, K-Citymarket, Neste K) with the share of 35.8 % and Lidl (n=35) with the share of 9.3 %. Altogether 418 stores were chosen for the analysis (Figure 19). The adresses and opening hours of the grocery stores were retrieved from their websites on 15.9.2018 (see Table 3). As shown in Figure 19, the grocery store density is highest in the inner city of Helsinki and along the railway and metro lines, such as in Itäkeskus, Tikkurila and Leppävaara. All 418 stores are open between 9 AM and 8 PM, whereas 32 stores (8 %) are open 24 hours a day (Figure 20).

Regarding the accessibility analyses to transport hubs, the destination coordinates of the main entrance to the central railway station and Helsinki-Vantaa Airport (Terminal 2) were derived from OpenStreetMap.

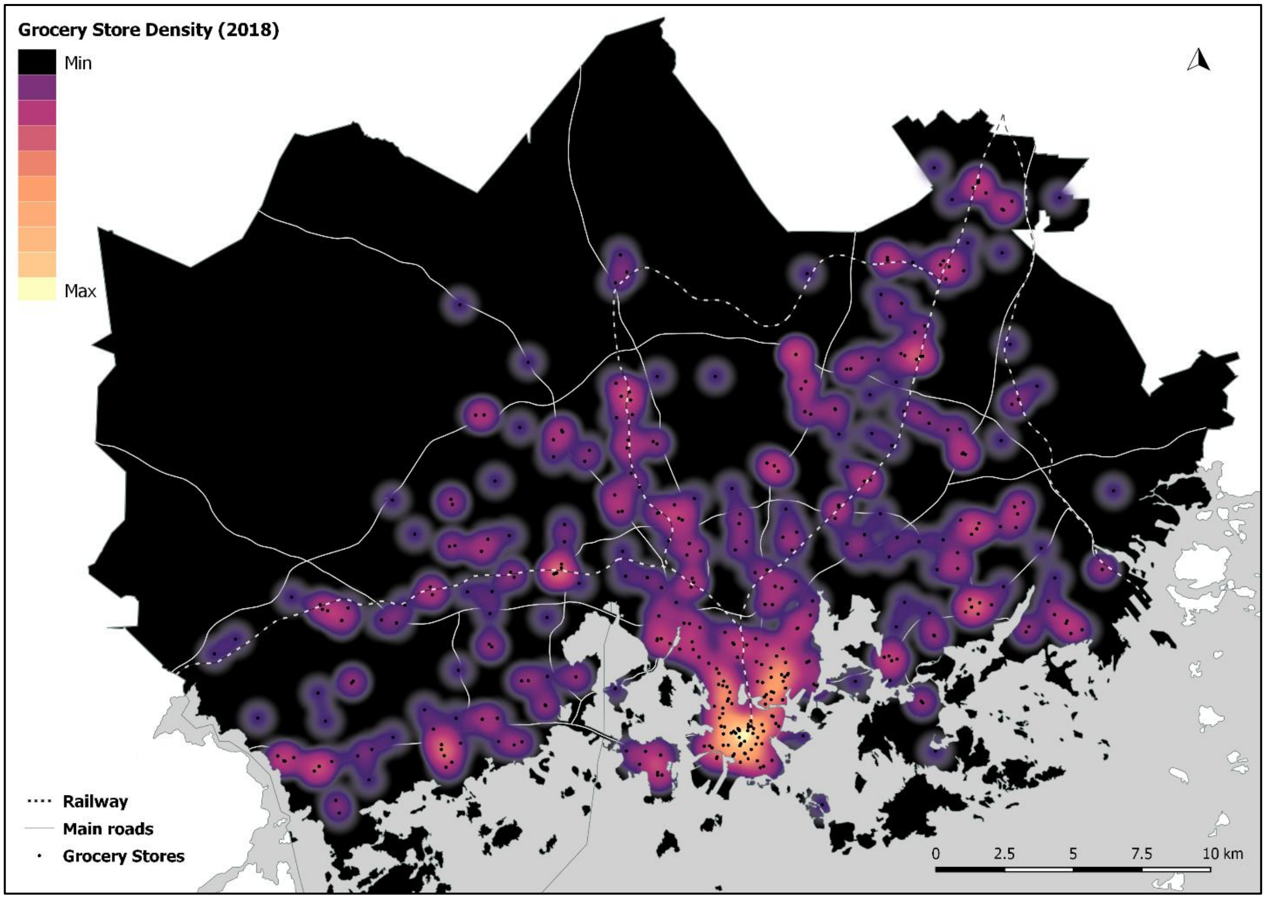


Figure 19. The locations and density of grocery stores (n=418) in the study area by the three largest chains in the study area (15.9.2018). The density surface is based on kernel density estimation.

Additionally, openly available supplementary spatial datasets from Seutukartta (HSY 2018a), such as local sea and lake areas, railway lines and the road network were used for visualization purposes. Municipal division data on the level of highest available spatial accuracy (1:10 000) by the National Land Survey of Finland (NLS 2018a) was used both to crop datasets to the study area and for visualization purposes.

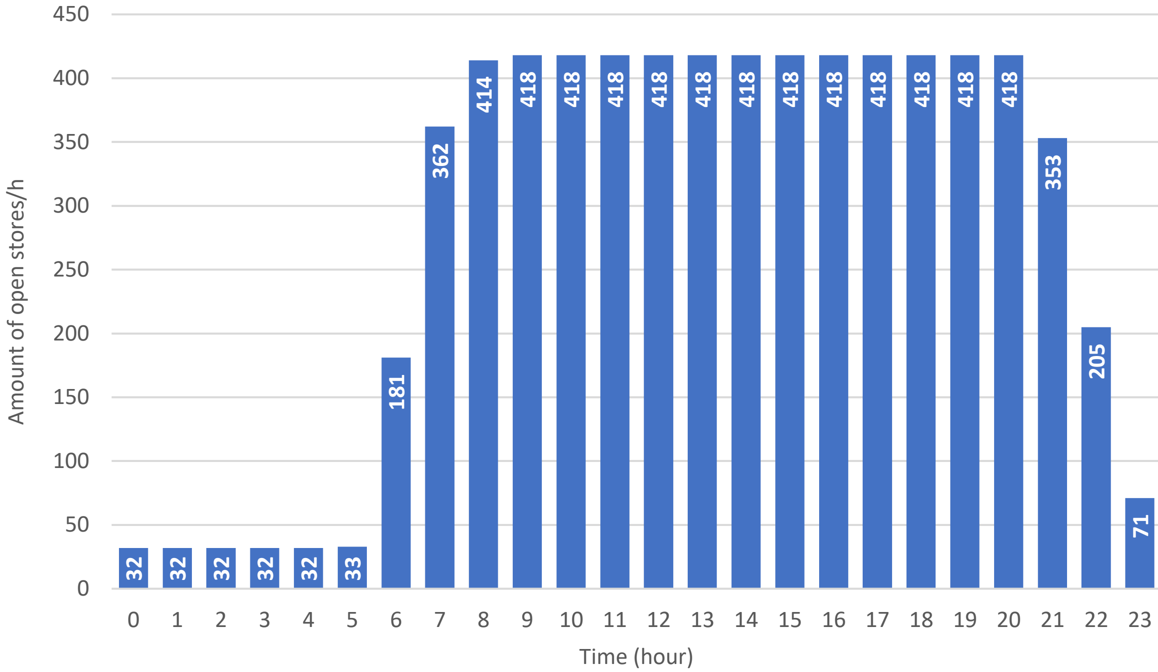


Figure 20. Opening hours of grocery stores (n=418) in the study area. The opening hours were retrieved from the websites of stores on 15.9.2018. Approximately 8 % (n=32) of all stores are open 24 hours a day.

5. METHODS

5.1. Study design

The empirical study was conducted in six steps. First, the data required for the analysis were acquired and pre-processed (1). Secondly, the pre-processed data was used as input to estimate hourly weekday population distribution in the study area using two interpolation methods a) simple areal weighting and b) an advanced dasymetric interpolation method (2). The hourly weekday population distribution was calculated using both methods for all three network-driven mobile phone datasets in the study: HSPA calls, RRC connections and RAB attempts. Thirdly, the resulting six dynamic population datasets were validated and evaluated to assess the best interpolation method and dataset combination as a proxy for people (3). The data was validated against official population register data representing the residential population and workplace data. Next, the most feasible population dataset was used to analyse the population dynamics of the study area on different spatial levels (4).

To assess the significance of dynamic population data in accessibility modelling, a case study was conducted to analyse the dynamic accessibility to transport hubs and grocery stores in the study area (5). The role of the population data was analysed by comparing cumulative reaching population to the selected activity locations using both dynamic and static population data (6). Lastly, the outputs of all steps were visualised as maps, charts and tables and used to answer the research questions and to draw conclusions. The steps in the empirical study were conducted primarily using Python and QGIS. The workflow of the study is illustrated in Figure 21.

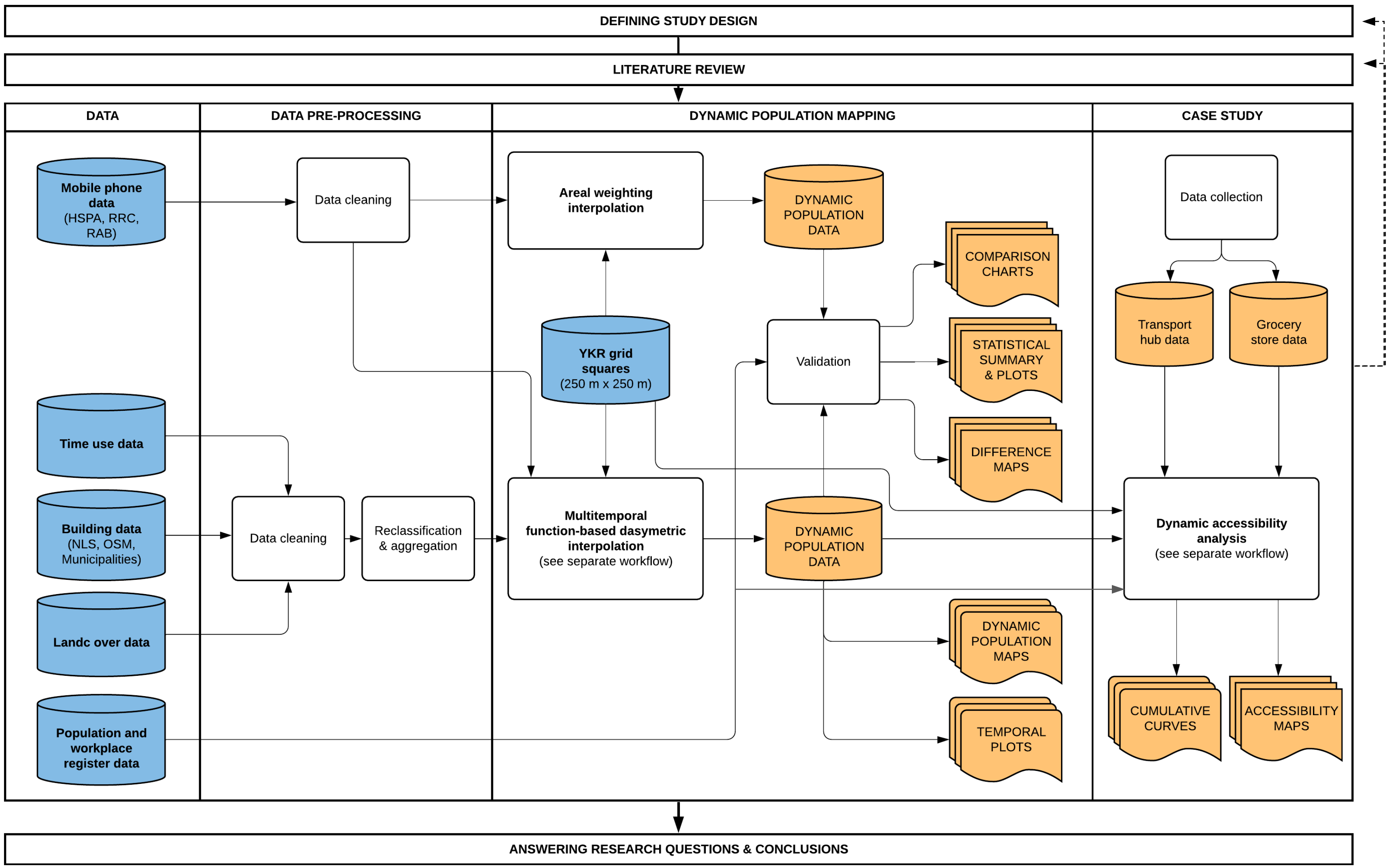


Figure 21. The general workflow of the study. The steps of the MFD method and dynamic accessibility analysis are presented in more detail in separate flow charts. Study outputs are shown in yellow, input data sources (if not produced in this study) are shown in blue.

5.2. Pre-processing of mobile phone data

The mobile phone data was prepared for constructing the dynamic population by filtering, cleaning, manipulating and aggregating the original data. The pre-processing of the mobile phone data was done using Python.

Firstly, the mobile phone data was cropped to the extent of the study area. To assess the spatiotemporal distribution of people more realistically, the area of interest was extended to cover all base stations, whose theoretical coverage area intersects the study area. This was done to enhance the accuracy of the result and eliminate the bias, which otherwise would occur in the outskirts of the study area.

Secondly, the data was cleaned to remove any values that might distort or skew the results. BS with no activity during the whole study period (n=8) were removed from the dataset (dropped rows = 14). Two cases were identified, where two BS in different locations had the same identifier. In both cases, the values of the two BS were merged, since the base stations did not have overlapping values. The BS with a higher original amount of network activity was kept and the other was removed. In addition, two cases were identified, where two BS with different identifier had the same coordinates. The values of the BS were summed, after which the other BS was removed. Also, a systematic error was found from the data. On all days except 29.10.2017, no RAB attempts or RRC connections were registered for any BS between midnight and 1 AM. New values were interpolated based on the values of the prior (11 PM – midnight) and following hour (1 AM – 2 AM).

After cleaning and editing the data, it was filtered to cover only weekdays between Monday and Thursday, since the focus of the study was to analyse the population distribution on average weekdays. After the filtering, data of 42 days was left for further analysis. Finally, the data was aggregated to get median number of HSPA calls, RAB attempts and RRC connections of the whole study period for every BS for every hour of the day.

5.3. Constructing the dynamic population from mobile phone data

5.3.1. Areal weighting interpolation

Two approaches were used to interpolate mobile phone data from base station locations to statistical 250 m x 250 m grid cells in this study, namely *areal weighting* (AW) and *multi-temporal function-based dasymetric* (MFD) *interpolation*. First, the simpler AW method was applied to allocate the mobile phone data. In short, the AW method disaggregates the data from the points of origin and redistributes it within the source zones to desired target zones based on their relative size (Figure 22).

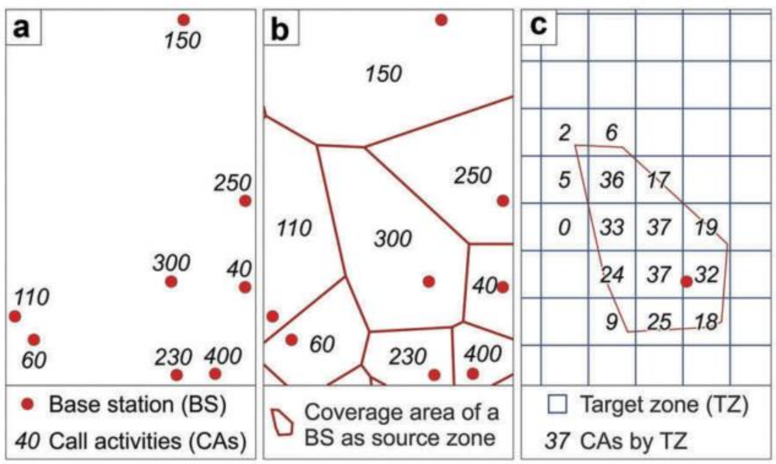


Figure 22. Workflow of the areal weighting interpolation (adapted from (Järv, Tenkanen & Toivonen, 2017a, p. 1635). The mobile phone data as points representing the base stations, b) Voronoi polygons as theoretical coverage areas of the base stations c) aggregation to the target zones based on their relative size.

In the first step of the AW method, coverage areas were created for the base stations to disaggregate the mobile phone data. Since the actual coverage areas of base stations were unknown, Voronoi polygons (also known as *Voronoi/Thiessen/Dirichlet diagrams or tessellation*) was selected as the method to estimate the theoretical coverage areas of base stations (see Figure 16). A Voronoi polygon of a point is formed of all points in space that are closer to the given point than to any other point, typically based on Euclidean distance on a plane (Figure 23). Voronoi polygons are dual to Delaunay triangulation (Okabe et al., 2000). Voronoi polygons is a generally accepted and the most common approach in studies that use cellular network data (e.g. Candia et al., 2008; González et al., 2008; Sevtsuk & Ratti, 2010; Soto & Frías-Martínez, 2011; Yuan & Raubal, 2012;

Deville et al., 2014). Also alternative approaches, such as improved Voronoi (Traag et al., 2011; Ricciato et al., 2015) and best serving cell or rasterization methods (e.g. Girardin et al., 2009; Steenbruggen et al., 2015) have been developed, but data required for these were not available in this study.

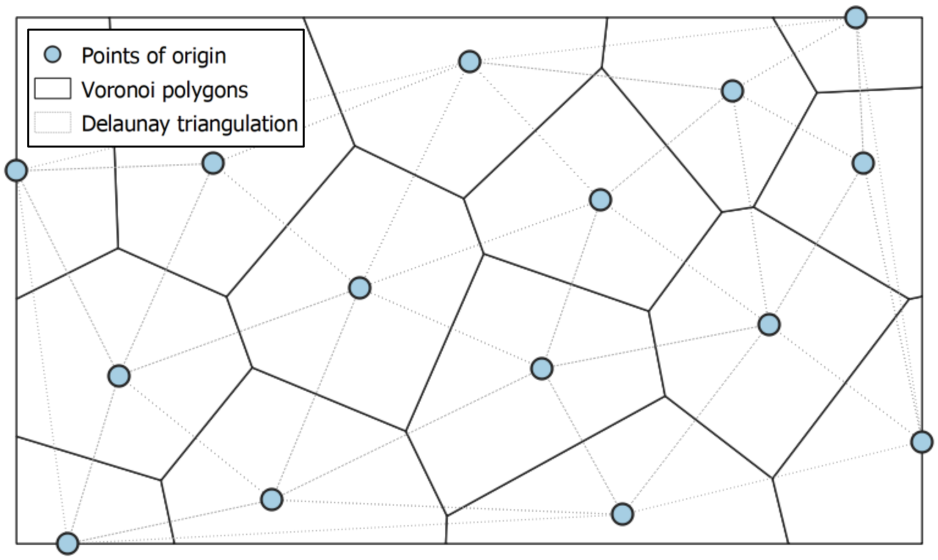


Figure 23. The principle of Voronoi polygons. The polygons with black borders indicate the Voronoi polygons and ones with dotted grey borders indicate the Delaunay triangulation of the points.

The twenty-four hourly values of each base station were scaled to the hourly sum of all base station in the study area, which were then assigned to the Voronoi polygons. Next the Voronoi polygon layer was overlaid with the target zones, statistical 250 m x 250 m grid squares. Finally, the area of each target zone subunit was multiplied by the share of the hourly activity in the related base station.

5.3.2. Multi-temporal function-based dasymetric interpolation

In addition to areal weighting, a more advanced interpolation method, namely *multi-temporal function-based dasymetric (MFD) interpolation* was used to distribute the mobile phone data from the base stations to the target zones.

The MFD-method is a *dasymetric interpolation* method, which belongs to same family of *areal interpolation* methods as areal weighting (Wu et al., 2005). Dasymetric interpolation differs from areal weighting in that it uses ancillary data to improve the interpolation (Mennis, 2003). The MFD method is thus equivalent to the AW method enriched with ancillary data. Accordingly, dasymetric interpolation has been regarded as one of the most feasible methods for refining the spatial resolution of population data in literature and has been widely applied in different application fields (Wu et al., 2005; Järv, Tenkanen & Toivonen, 2017a). An overview of the concept of dasymetric interpolation has been presented e.g. in Eicher and Brewer (2001) and Mennis (2003).

In summary, the MFD method disaggregates input (population) data within given source zones and aggregates it to chosen target zones using ancillary datasets (see Järv et al. (2017a) for a more detailed description). The MFD method consists of five main steps:

1) The creation of the physical surface layer
2) The spatial disaggregation by source and target zones
3) The integration of temporal human activity data
4) The integration of mobile phone data
5) The spatial aggregation to target zones

The three first steps involve the creation of a probability layer that can be used to assign the likelihood of population in certain areas at certain times or to restrict the allocation to certain areas. More specifically, the probability layer consists of time-dependent human activity data and data of land cover and buildings that are reclassified to match the human activity data based on their activity function type (Greger, 2015; Järv, Tenkanen & Toivonen, 2017a). In this study, the input data were reclassified to six activity function types, based on Järv et al. (2017a) (see Table 5). Each of the five steps in the MFD method are described in more detail in the following paragraphs. The general workflow of the MFD method in described graphically in Figure 24 and in more detail Figure 25.

Table 5. Reclassification table used in the MFD interpolation. The classification is adapted from Järv et al. (2017a) to suit local data.

Activity Function Type[a] (AFT)	Time Use Data	Land Cover Data	Building Data
Residential	at home or accommodation	residential area	residential and leisure buildings, hotels, prison
Work	at work or school	industrial and commercial area, construction sites	offices, industrial buildings, schools, public buildings not for in situ services (e.g. hospital, city hall)
Retail & Service	shopping or using services	-	shopping centres, services (e.g. hairdresser, gas station)
Transport	travelling	transport networks (e.g. roads, harbours, parking spaces)	Passenger terminals, station buildings
Restricted	(no activity)	water, wetlands, arable land, dump sites, restricted areas	-
Other	other activity, e.g. leisure activities	other areas, e.g. forest, parks, cemeteries and sport and leisure facilities	other buildings e.g. restaurants, entertainment, religious buildings, public buildings for in situ services (e.g. libraries, sport facilities)

[a] Target class, common key

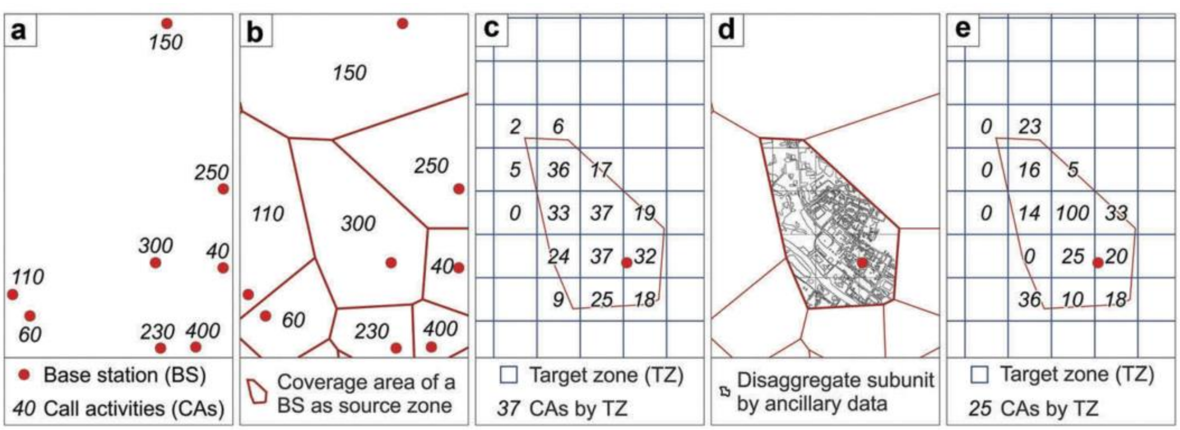


Figure 24. Workflow of the MFD interpolation (adopted from (Järv, Tenkanen & Toivonen, 2017a, p. 1635). The mobile phone data as points representing the base stations, b) Voronoi polygons as theoretical coverage areas of the base stations c) aggregation to the target zones based on their relative size, d) integration of ancillary data and e) final interpolated data based on the dasymetric interpolation.

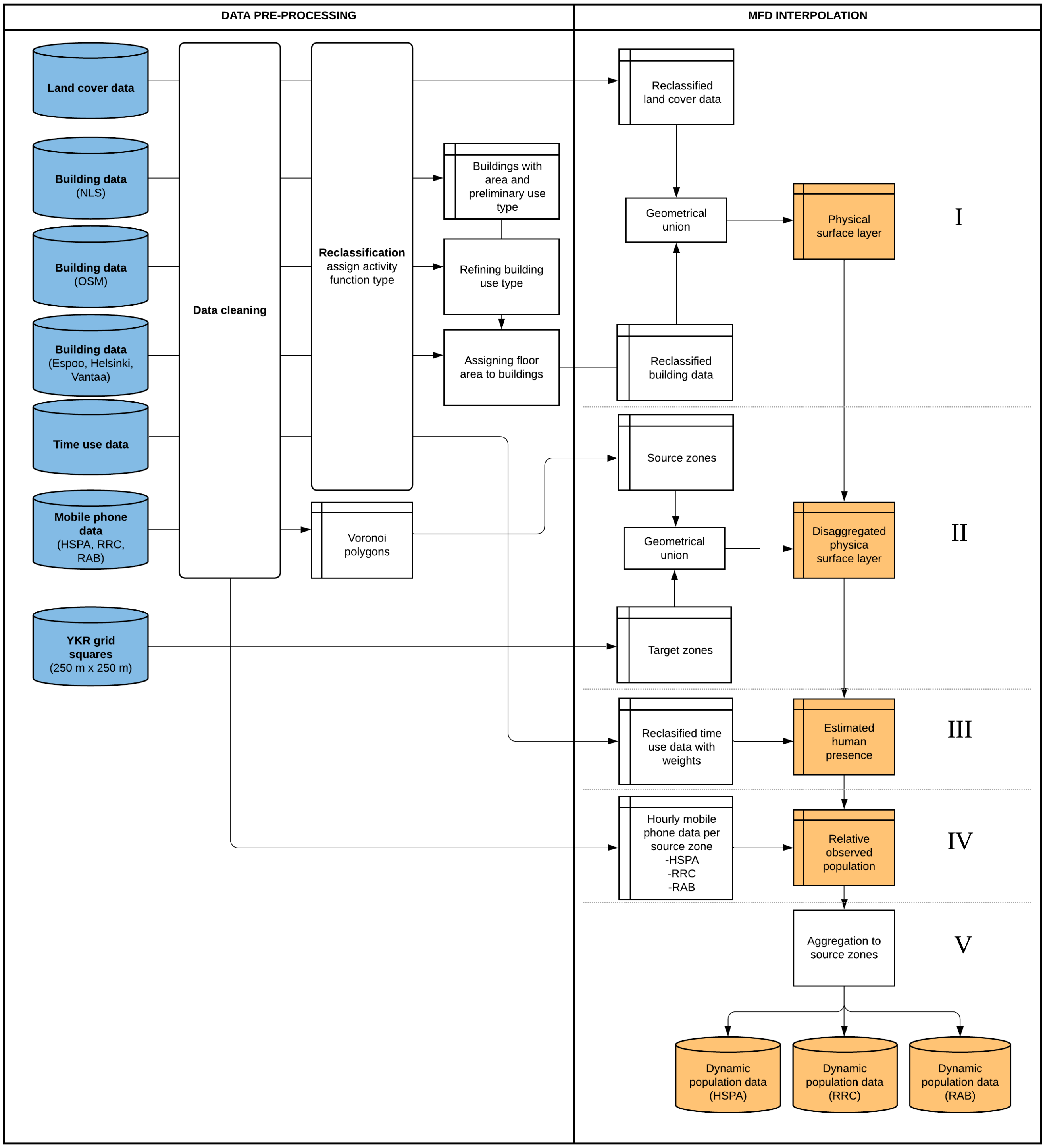


Figure 25. General workflow of the multitemporal function-based dasymetric (MFD) interpolation method. Steps I–V indicate the phases of the MFD interpolation method. Original input data sources are shown in blue and study outputs in yellow.

1) The creation of the physical surface layer

In the first step of the MFD method, land cover and building data were pre-processed and combined to create the physical surface layer. Each feature in the physical surface layer was assigned an activity function type, which enables the linking with time use data at a later stage of the MFD method (Greger, 2015; Järv, Tenkanen & Toivonen, 2017a).

Regarding land cover data, the most recent version of the CORINE Land Cover (2012) land cover raster dataset with a spatial accuracy of 20 m x 20 m was used to determine the land cover of the study area. The spatial accuracy of the more broadly available Pan-European CORINE Land Cover vector dataset was too coarse (25 ha) for the study purposes. Similarly, the more recent openly available land cover data provided by the National Land Survey of Finland and Helsinki Region Environmental Services Authority HSY were rejected due to insufficient content or too low spatial accuracy.

To prepare the land cover data for the MFD method, the dataset was transformed to vector format, reclassified and cropped to the extent of the study area. Following the approach in Järv et al. (2017a), the land cover data was reclassified from the original classes (n=48) to five classes based on their activity function types: 1) residential, 2) work, 3) transport, 4) restricted and 5) other (see Table 5; Appendix 1; Figure 26). To improve the classification, the Vuosaari harbour area was separated from the other passenger transport -driven harbour areas to the ‘work’ class due to its work-driven functions. Similarly, the Helsinki-Vantaa airport area and Malmi airport were separated to the ‘work’ (prior) and ‘transport’ (latter) classes to better correspond with their activity function types.

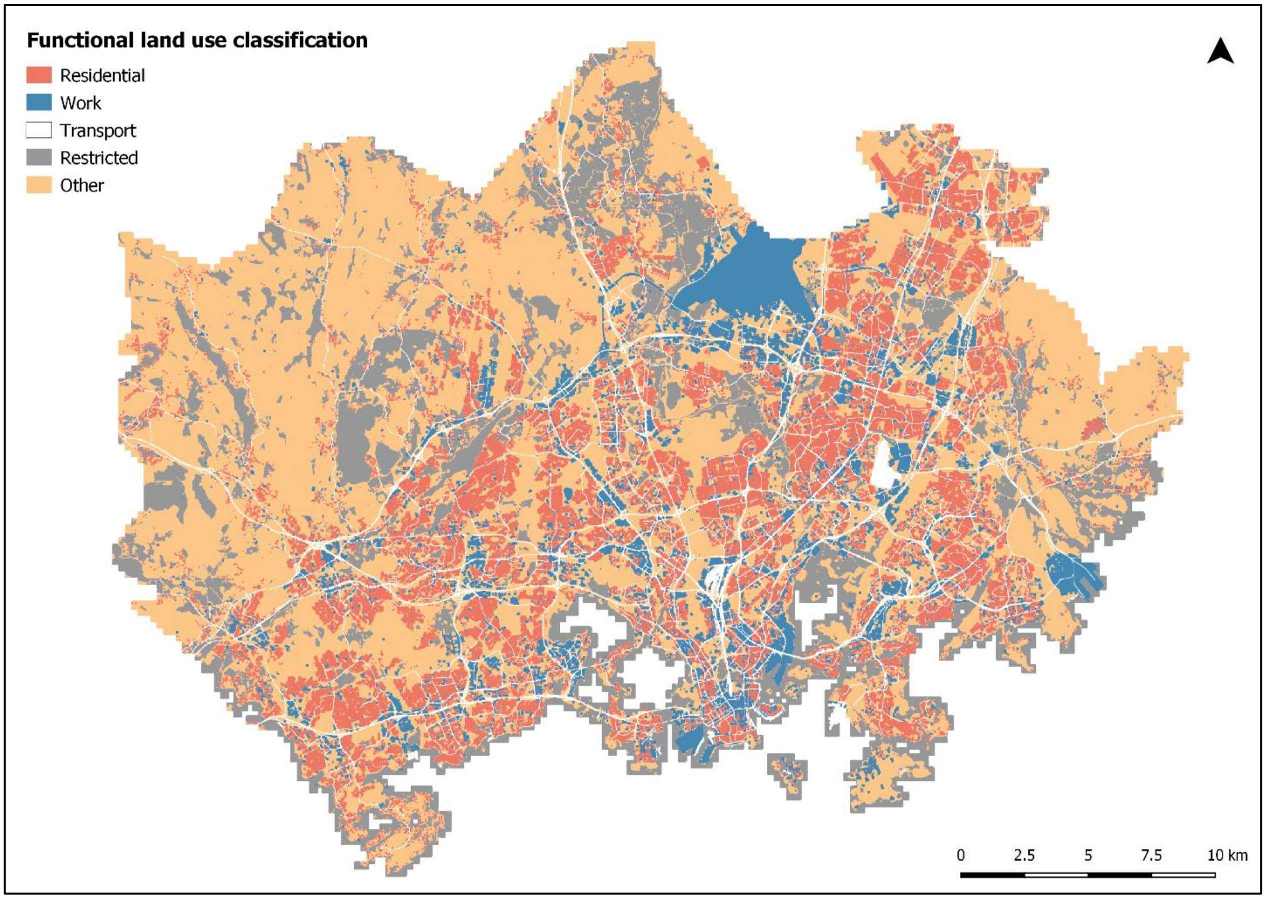


Figure 26. Reclassified land cover data based on CORINE Land Cover 2012 (Finnish Environment Institute, 2012).

In terms of the building data, building polygons were extracted from the National Topographic Database (NLS 2018b) and cropped to the extent of the study area. The cropped dataset contained 160 490 buildings in total. Next, the area based on the geometry of the building footprints was calculated. The data was further cleaned by filtering out basin polygons (n=273) and buildings with an area below 20 m^2 (n=6860) leaving 153 357 buildings left for further analysis. The buildings in the National Topographic Database were reclassified according to their primary activity function type (AFT) to residential, work and other buildings (see Table 5; Appendix 2; Figure 27). Similarly as in Järv et al. (2017a) non-classified building were assumed to be work buildings.

To refine the AFT classification of building polygons, additional building polygons were retrieved from OpenStreetMap (OSM) using OSMnx (Boeing, 2017), which is a tool for extracting vector data from OSM in Python. The OSM buildings (n = 72 574) were cropped to study area extent and reclassified based on their AFT (see Table 5; Appendix 3). The classification was assigned to the buildings dataset retrieved from the National Topographic Database using k-nn nearest neighbour search and spatial join.

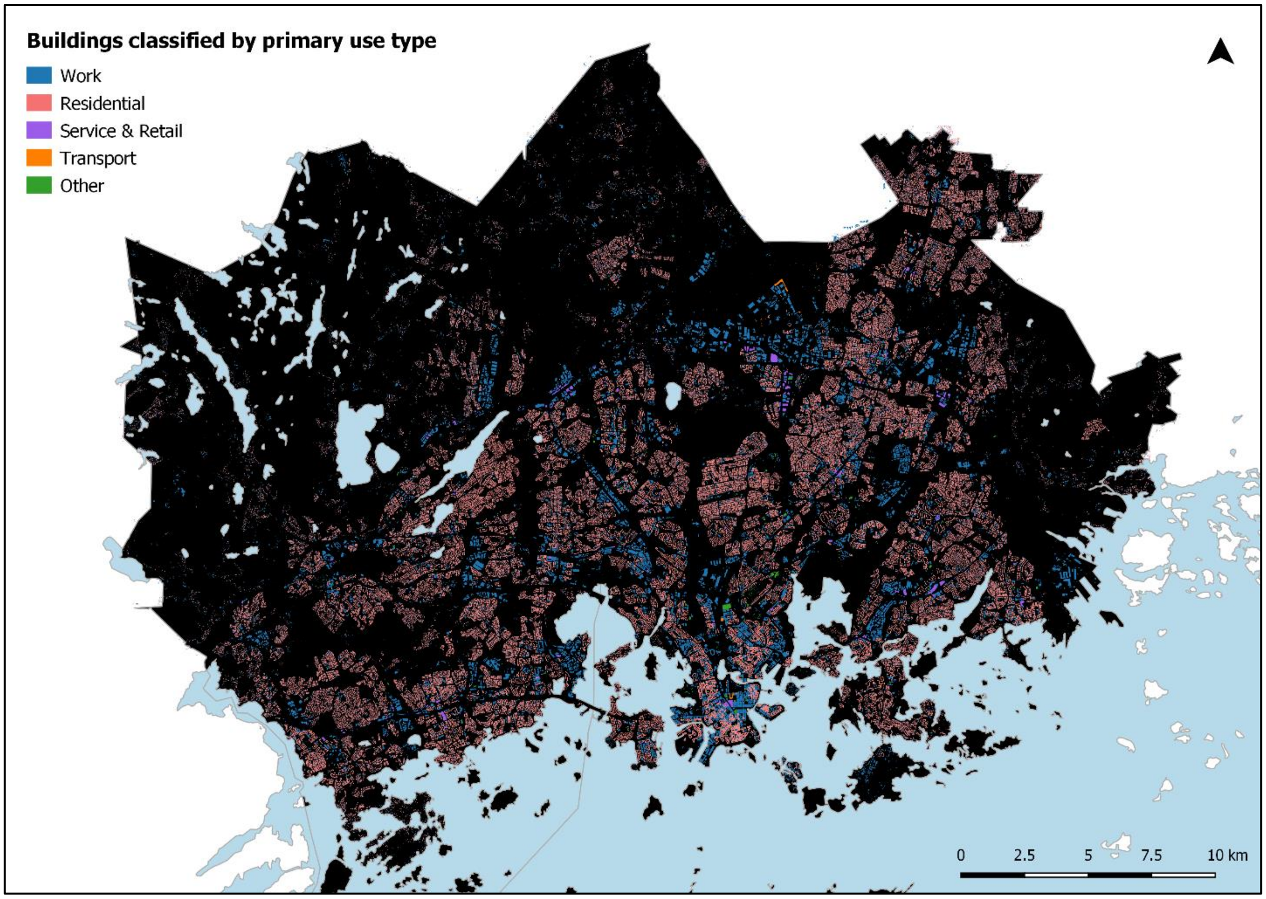


Figure 27. The reclassified buildings dataset (n = 153 357) based on primary activity function type.

Only one AFT was assigned for each building. The crudeness of the selected approach is recognized as buildings may have multiple use types either simultaneously or at different times. However, the current level of accuracy is expected to be feasible for the purpose of this study. The MFD method also performed well against validation data despite this limitation in Järv et al. (2017a). The final classification of buildings per AFT is presented in Figure 27 and Table 6.

The physical surface layer also takes account for the vertical dimension in the likelihood of human presence. To retrieve the vertical dimension, open data from Vantaa, Helsinki and Espoo containing building footprints, floor area (m^2) and floor counts based on the national building register were acquired (not available from Kauniainen). The municipal data was further cleaned, combined and joined to the original building dataset. In contrast to the Järv et al. (2017a), the floor area calculations were not performed until the second step of the MFD method due to differences in source data. Finally, a geometric union was performed to combine the reclassified building and land cover layers.

Table 6. Reclassified land cover and building data used in the MFD interpolation per activity function type.

Activity Function Type	Land cover Area (km^2)	Buildings (count)
Residential	161.79	104 988
Work	81.35	47 268
Service & Retail	0	302
Transport	43.54	55
Restricted	144.54	0
Other	406.02	744
Total	837.24	153 357

2) The spatial disaggregation by source and target zones

After creating the physical surface layer, a geometric union was performed between the physical surface layer, source zone and target zone layers. As in the AW method, Voronoi polygons were used to estimate the theoretical coverage areas of base stations and 250 m x 250 m statistical grid cells were used as the source zones. As a result, the study area was divided into 345 917 subunits, each with a designated AFT and spatial unit type (building or land) and floor area. The area of each subunit was recalculated after the overlay operation.

Next, the *relative floor area* of each subunit was calculated to include the vertical dimension to the interpolation. First, the absolute floor area was assigned to the subunits based on their spatial unit type and AFT. For subunits with the spatial unit type ‘land’, the geometric area of the subunit was set as the floor area. For subunits with the spatial unit type ‘building’, the floor area was calculated based on openly available building data from the municipalities of Espoo, Helsinki and Vantaa containing the building register -based floor areas and floor counts. Use of actual floor areas provides a more accurate estimate than the LiDAR-based approach applied in Järv et al. (2017a), where the floor area was estimated from the building height extracted from the digital surface model (DSM).

In the case of buildings, where population register data was not openly available (e.g. Kauniainen) the floor area was estimated based on the actual or mean floor count and a floor area coefficient. The floor area coefficient was calculated as the median ratio between the actual floor area and the product of the building footprint area and the floor count of those buildings, where the actual floor area and floor count was available. Both the mean floor count and the floor area coefficient were calculated separately for buildings of each AFT. The mean floor count was 2 for residential and service and retail buildings and 1 for others. The floor area coefficient was 0.95 for residential buildings, 0.91 for service and retail buildings and 0.98 for other buildings. Finally, the relative floor area (RFA) was calculated for each subunit within a source zone, based on the following formula:

$$RFA = \frac{FA_s}{\sum FA_s \ \in j} \tag{2}$$

where

RFA= relative floor area
FA = floor area
s = spatial subunit
j = source zone

As a result, the sum of the relative floor area of all subunits within one source zone (Voronoi polygon) equals to 1. The higher the relative floor area of the subunit, the higher the likelihood that activity of the BS is allocated to that subunit.

3) The integration of temporal human activity data

In the third phase of the MFD method, time use data was used to integrate the physical surface layer to create a probability matrix for allocating the mobile phone data to target zones within each source zone. As a result, each spatial subunit was assigned an hourly likelihood rate of human presence based on its activity function type.

The *estimated human presence* (EHP) in each subunit was calculated using human activity data based on the latest Finnish time-use survey carried out in 2009 by Statistics Finland (2010) (Figure 28). The survey was carried out according to the guidelines for Harmonised European Time Use Surveys (HETUS) issued by Eurostat (2009). The human activity data contained the time use based on the activity location on 10-minute intervals during weekdays (Monday–Thursday) of over 10-year-olds in the HMA.

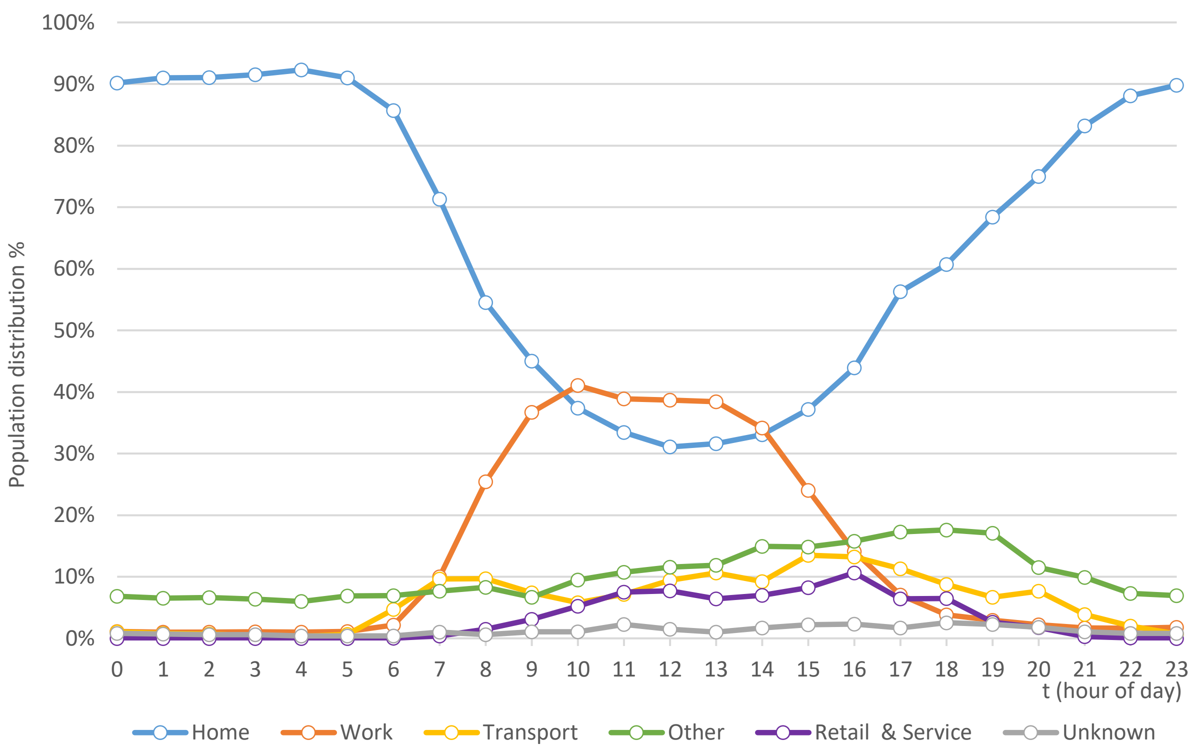


Figure 28. Time use by location and activity per hour on weekdays (Monday-Thursday) of over 10-year-olds in the Finnish Capital Region 2009–2010 (Statistics Finland, 2010).

To calculate the estimated human presence, the human activity data was first pre-processed by aggregating the data to hourly level. Secondly, the data and reclassified to the following classes based on the location, where the activity was undertaken to join it with the physical surface layer: 1) residential, 2) work (incl. education), 3) transport, 4) retail & service, 5) unknown and 6) other (e.g. recreational areas) (see Table 5).

An *hourly probability coefficient* (H) was assigned for every hour of a day based on the time use data. In addition, a *seasonal probability coefficient* (M) was assigned to account for the impact of the season on the distribution of people indoors and outdoors. According to a study conducted by Hussein et al. (2012), people were found to spend approximately 90 % of the day indoors in Helsinki in March. Similarly as in Järv et al. (2017a), the results are assumed to be suitable for the dasymetric interpolation, since the mobile phone data used for estimating the population distribution was also collected during the winter season. The seasonal factor was applied for three of the activity function types (residential, work & education, other). Thus, a subunit of for example the work activity function type would receive a coefficient of 0.9 if the spatial unit type was 'building' and a coefficient of 0.1. if the spatial unit type was 'land'. Subunits with the other activity function types were assigned the factor 1, except restricted areas, which were assigned factor 0. This way, the MFD method prevents population to be allocated to a subunit of restricted type. Overall, the estimated human presence per every spatial subunit at a given time unit (hour) was calculated using the following formula:

$$EHP_t = (H_{au}M_{au}) \times RFA_{sj} \tag{3}$$

where

EHP = estimated human presence
t = time unit
H = hourly factor
M = seasonal factor
RFA = relative floor area
a = activity function type
u = spatial unit type (building or land)
s = spatial subunit
j = source zone

4) The integration of mobile phone data

In the fourth phase of the dasymetric interpolation, the mobile phone data was integrated to the physical surface layer enriched with hourly and seasonal human activity data. The mobile phone data containing the hourly median number of the different network activities was linked to the physical surface layer based on the BS identifier. First, the mobile phone activity per spatial subunit was normalized by dividing it by the sum of the corresponding value of all spatial subunits in the study area. In other words, a sum of the relative share of mobile phone data of all subunits in the study area is 1. The *relative share of mobile phone data* per spatial subunit of study area total at given hour was calculated using the following formula:

$$RMP = \frac{MP_{tj}}{\sum MP_{tj} \in S} \qquad (4)$$

where

RMP = relative share of mobile phone data
MP = mobile phone data
t = time unit
j = source zone
S = study area

The formula was calculated separately for each of the three mobile phone datasets (HSPA calls, RRC connections, RAB attempts). Secondly, the hourly normalized mobile phone data was multiplied by the hourly estimated human presence to allocate the population to the subunits based on the physical surface layer and time use statistics. The *relative observed population* was calculated as follows:

$$ROP_t = EHP_t \times RMP_t \qquad (5)$$

where

ROP = relative observed population
EHP = estimated human presence
RMP = relative share of mobile phone data
t = time unit

5) **The spatial aggregation to target zones**

In the fifth and final phase of the MFD method, the spatial subunits were aggregated to the statistical 250 m x 250 m grid cells. The aggregation was performed by dissolving the subunits based on the target zone ID. As a result, each target zone was assigned the sum of the relative observed population of all spatial subunits within the given target zone. The aggregation to the target zones can be summarized as follows:

$$ZROP_t = \sum_{s \in z} ROP_t \tag{6}$$

where

$ZROP$ = spatially aggregated relative observed population per target zone
ROP = relative observed population
t = time unit
s = spatial subunit
z = target zone

As the final result of the MFD method, three normalized population data layers for each hour of the day based on each type of mobile phone data (HSPA calls, RRC connections, RAB attempts) were created. The script used to run the MFD method is based on (Järv et al. 2017b) and openly shared in GitHub[1].

5.4. Validation and evaluation of the dynamic population data

One of the most common ways to validate mobile phone data -derived population data is to compare it to residential population data (Calabrese et al., 2014; Deville et al., 2014; Järv, Tenkanen & Toivonen, 2017a). In accordance, the population distribution based on the interpolated mobile phone data was validated by comparing it against the population distribution based on official population register data (Grid Database 2016) during night-time. The night-time window (2 AM – 5 AM) was selected based on the hours when people are most likely to be at home according to the time use survey (Statistics Finland, 2010). Using evening and night-time for comparing mobile phone data to residential population data is a generally accepted method for

[1] https://github.com/cbergroth, see also https://github.com/AccessibilityRG/MFD-mobile

validation of mobile phone data as a proxy for people, although the used time interval varies between studies (see e.g. Deville et al., 2014; Järv, Tenkanen & Toivonen, 2017a; Ma et al., 2017).

The three different types of mobile phone data (HSPA calls, RRC connections, RAB attempts) were evaluated against the population register data using the 1) Pearson correlation coefficient and standard error of the estimate, 2) mean absolute error and 3) coefficient of variation. The chosen evaluation methods were used recently for example by Järv et al. (2017a) and Mennis (2016) in similar research contexts. Scatterplots were created to graphically represent the correlation between the interpolated mobile phone data and population register data. The evaluation methods were used to assess and compare the feasibility of the three different types mobile phone data as a proxy for people. The mean absolute error (MAE) was calculated using the following formula:

$$MAE = \frac{1}{n}\left(\sum_{i=1}^{n} |y_i - x_i|\right) \qquad (7)$$

where

y = normalized population per grid square (n=13231) based on population register data

x = normalized estimated population per grid square (n=13231) based on mobile phone data

The coefficient of variation (CV) based on the root mean square error (RMSE) was calculated using the following formula:

$$CV = \frac{\sigma}{\mu} \qquad (8)$$

where

σ= the standard deviation of the RMSE

μ = the mean RMSE

The statistical evaluations were performed separately for the results of both interpolation methods (MFD, AW) to assess their feasibility. The correlation was in addition calculated for all twenty-four hours between the interpolated mobile phone data and population register data as well as the number of workplaces per statistical 250 m x 250 m grid square. To further analyse the feasibility of the used interpolation methods, the distribution of population per spatial subunits grouped by their activity function type class (residential, work, other, transport, service & retail, restricted) were compared to the time use statistics during night-time (2 AM – 5 AM) and daytime (3 PM – 4 PM). Lastly, the absolute difference in the relative share of population per grid square between the mobile phone data -derived population distribution and the population register data were visualized as difference maps for day and night-time.

5.5. Extracting spatiotemporal patterns from the dynamic population data

The dynamic population distribution was analysed on the level of non-aggregated and aggregated statistical grid cells. Firstly, the relative population distribution within the study area on grid cell level was visualised as twenty-four hourly maps. The maps were also compiled to an animation to better visualize the changes between consequent hours and the overall patterns in the diurnal variation. To analyse the difference in the population distribution between daytime and night-time, the absolute and relative difference in share of population of HMA total between night-time (3–4 AM) and daytime (3– 4 PM) were calculated for every grid cell and visualised as difference maps. The diurnal variation was also analysed by the standard deviation of the hourly relative share of HMA total population during 24-hours per statistical grid cell. Furthermore, the pulse of individual grid cells was extracted from six different locations: Tapaninkylä, representing an area with residential functions (YKRID: 5902076), Pitäjänmäki industrial area (YKRID: 5934906) and Kumpula Campus (YKRID: 5949389), representing both an industrial and a service-driven area with work functions, the junction of Ring road III and E12 representing an area with transport functions (YKRID: 5876192), and Citycenter shopping centre (YKRID: 5975375) and Paloheinä recreational area representing areas with shopping and leisure functions (YKRID: 5903895).

The grid-level population dynamics were also analysed on three different levels of spatial aggregation: 1) on municipal level, 2) within and outside the inner city of Helsinki and 3) clusters based on K-means (k=5) clustering, which reveals grid cells with similar diurnal patterns in the relative share of hourly present population. The spatial extent of the inner city of Helsinki was delimited by the neighbourhoods that the city of Helsinki uses as the statistical boundaries of the inner city (City of Helsinki, 2015). The grid cells of the inner city of Helsinki constitute 5 % of all grid cells (n = 13231).

5.6. Case study: Modelling dynamic accessibility using dynamic and static population data to grocery stores and transport hubs in Helsinki Metropolitan Area

5.6.1. Construction of the dynamic accessibility model

The importance of the dynamic population data was assessed with a case study of dynamic accessibility in HMA. In the case study, the share of reached static and dynamic population were compared to two types of activity locations, 1) major transport hubs and 2) grocery stores. The selected activity locations represent functions that are important to access regardless of the time of the day, for example from the perspective of night-workers. The workflow of the accessibility analysis is presented in (Figure 29).

Accessibility to **grocery stores** was analysed both on study area level including all stores and on individual store level. Regarding the latter, two grocery stores were selected for separate analysis and they represent the perspective of a local small store in neighbourhoods, which experience a significant variation in the amount of accessible population during the day. These stores were Alepa Länsimäki in Vantaa on the fringe of the study area located in a residential-driven neighbourhood and K-Market Pasaati in a neighbourhood next to a major work place area in the inner city of Helsinki. Regarding the transportation hubs, chosen locations were **Helsinki-Vantaa airport** (YKRID: 5855721) and **Helsinki city centre** represented by the Helsinki Central Railway Station (YKRID: 5975375), which both have significant night-work functions.

A step-wise implementation of the dynamic accessibility model based on the conceptual framework presented by Järv et al. (2018) was applied to analyse the accessibility of the given activity locations. The temporality of the first dynamic accessibility component, **activity locations**, was incorporated by taking into account the opening hours of the selected activity location. The locations and opening hours of the grocery stores were collected from the websites of stores and geocoded using Python. Helsinki-Vantaa airport and Helsinki city centre can be considered open 24 hours of a day, so separate opening hours were not applied to these activity locations. The second dynamic accessibility component, **transport network**, was represented by Kalkati route and schedule data (public transport) and routes for walking extracted from OpenStreetMap. Finally, the temporal variation in the **locations of people** were derived from the 24-hour population distribution dataset driven from mobile phone data constructed in this study.

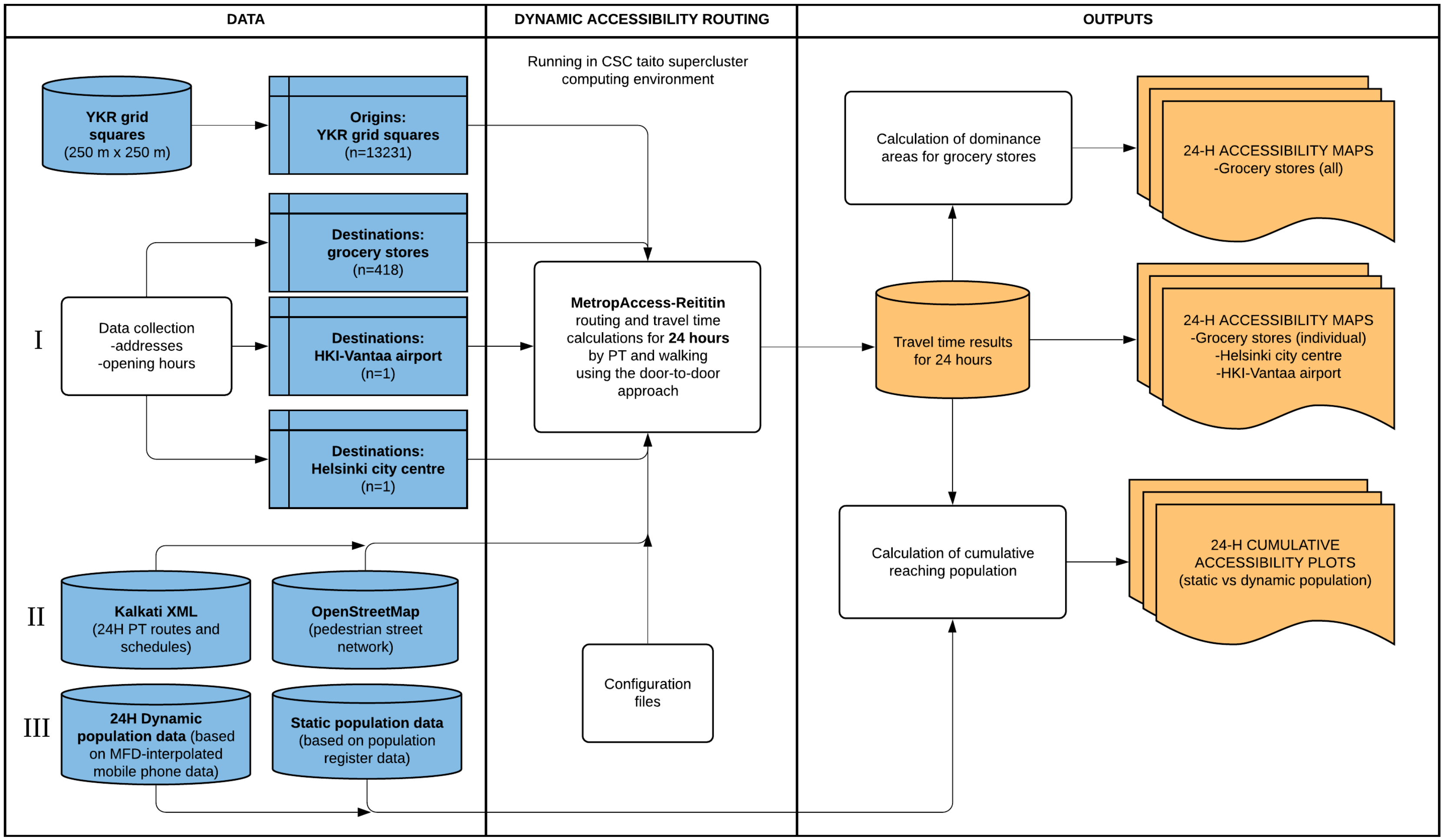


Figure 29. Workflow of the dynamic accessibility modelling. The Roman numbers I–III indicate the three components of the dynamic accessibility model: the temporally sensitive activity locations (I), temporally sensitive transport network data (II) and temporally sensitive population data (III). The opening hours were collected only for grocery stores as the airport and the city centre are open 24 hours of the day. Input data sources are shown in blue and study outputs in yellow.

5.6.2. Door-to-door travel time calculations

Accessibility to the selected activity locations was measured using MetropAccess-Reititin – an open tool developed by the Digital Geography Lab at the University of Helsinki for calculating the routes, travel times and distances for public transport and walking in the HMA (Toivonen et al., 2014). The tool is based on a so called door-to-door approach (Salonen & Toivonen, 2013), which takes into account the whole travel chain from the origin to the destination including 1) walking from the origin location to the transit stop, 2) waiting for the public transport vehicle to arrive and depart, 3) time spent for the traveling including potential transfers and waiting time and 4) walking time from the public transport stop to the destination (Figure 30). Accordingly, the results for public transport and walking are comparable to door-to-door travel times by car. If walking is a faster mode than public transport, such as on short distances, walking travel times are used for the whole trip. MetropAccess-Reititin uses a modified version of Dijkstra's (1959) algorithm for calculating the fastest route between given origin and destination points (Järvi et al., 2014).

The travel times were calculated for every hour of the day from 250 m x 250 m statistical grid squares (n = 13231) to the given activity locations. Travel time was selected as the measure of accessibility as it is commonly used in literature and is regarded as intuitive and easily interpretable (Olsson, 1965). The routing was based on the Kalkati public transport schedules of an average weekday (Thursday 25.1.2018). The walking speed in the accessibility analysis to all destinations was set to the default speed of 70 m/min (4.2 km/h). The same speed has been used previously for example by Tenkanen et al. (2016) for analysing the accessibility of grocery stores in the HMA. Ten different departure times based on Golomb's ruler (Bloom & Golomb, 1977) were used to find the optimal routes for every hour of the day (e.g. 00:00, 00:01, 00:06, 00:10, 00:23, 00:26, 00:34, 00:41, 00:53 and 00:55). The routing calculations were performed using Taito supercluster computing environment. Regarding the accessibility to all grocery stores, additional dominance areas were formed by calculating the travel time to the closest store for each grid cell.

The accessibility to the grocery stores and transportation hubs were visualized as hourly maps in QGIS. The hourly cumulative share of the total population in the HMA reaching the destination based on both dynamic and static population data was calculated using Python and visualised as cumulative curves and bar charts.

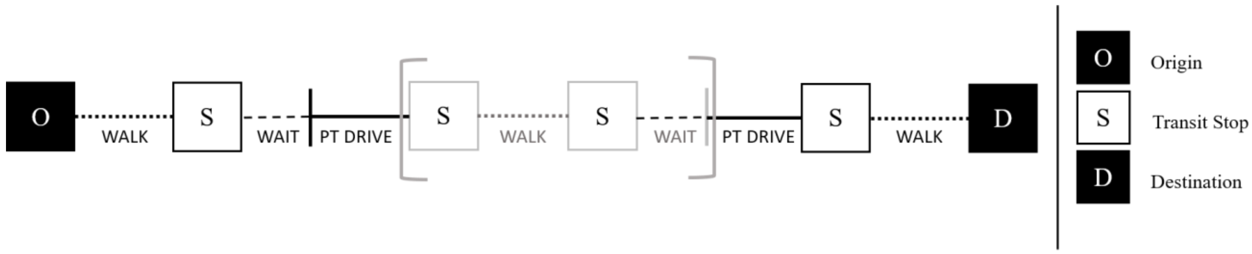


Figure 30. The principle of the door-to-door approach by public transport (PT) (adapted from Salonen & Toivonen (2013)).

6. RESULTS

6.1. Evaluation of mobile phone data as a proxy for people

6.1.1. Statistical evaluation of different mobile phone data sources

Out of the three types of network-driven mobile phone data, HSPA calls correspond best to the population register data during night-time (2 AM – 5 AM) based on all applied statistical evaluation methods (Table 7).

Regardless of the interpolation method, HSPA calls show the highest correlation (ρ = 0.683) with lowest standard error (0.000125) with the reference data, which shows a moderate to strong positive linear relationship with population register data (Figure 31–Figure 32). RRC connections (ρ = 0.489) and RAB attempts (ρ = 0.367) indicate at best a moderate to low positive correlation with the residential population. The Pearson correlation coefficients of all three different types of mobile phone data is statistically significant ($p < 0.001$). In addition, HSPA calls have the smallest mean absolute difference and coefficient of variation of the RMSE between the relative share of estimated present population and the residential population on 250 m x 250 m grid cell level (n = 13231). RRC connections exhibit the second-highest correspondence to the residential population at night of the three data sources, while RAB attempts differ most from the population register data.

Table 7. Statistical evaluation of the population data based on interpolated mobile phone data (HSPA calls, RRC connections, RAB attempts) against official population register data (2015) during night-time (2 AM - 5 AM) on 250 m x 250 m statistical grid squares (n=13231). The results are presented separately for Multi-temporal Function-based Dasymetric (MFD) and Areal Weighting (AW) interpolation.

Method of Evaluation		**MFD**			**AW**		
		HSPA	RRC	RAB	HSPA	RRC	RAB
Linear Regression	Correlation Coefficient (Pearson)	0.683***	0.489***	0.367***	0.533***	0.388***	0.285***
	Standard Error of the Estimate	0.000125	0.000149	0.000159	0.000145	0.000158	0.000164
Mean Absolute Error (MAE)		0.000052	0.000069	0.000074	0.000072	0.000084	0.000088
Coefficient of Variation (based on RMSE)		0.000145	0.000219	0.000325	0.000203	0.000248	0.000381

*** Correlation is significant at the 0.001 level (2-tailed)

Furthermore, based on solely the statistical evaluation, HSPA calls perform better than RRC connections or RAB attempts even if HSPA calls are interpolated using the AW method and the others with the MFD method. This highlights the superiority of HSPA calls as a proxy for the population during night-time. Consequently, HSPA calls are selected as the data source for the further steps of the analysis.

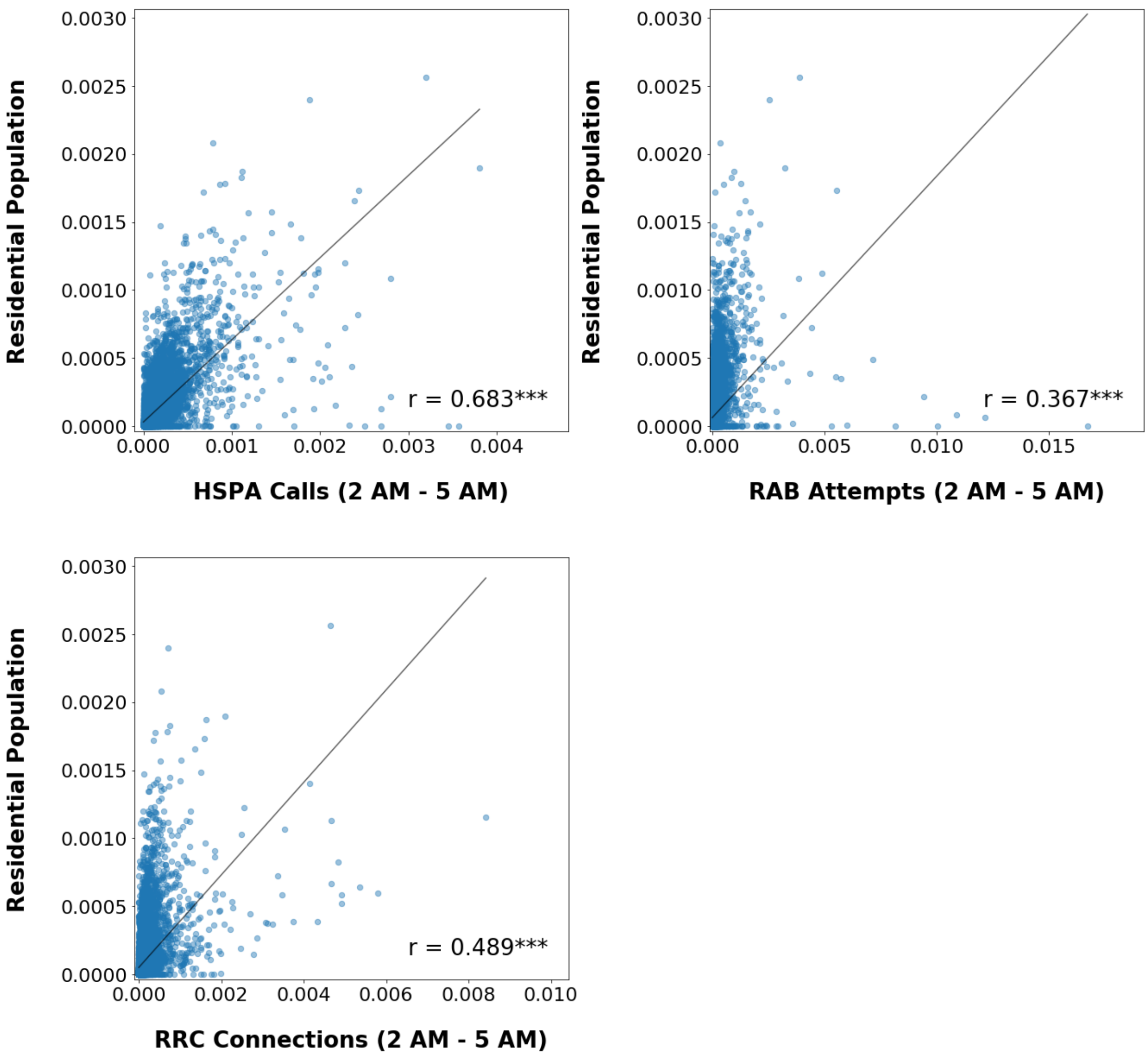


Figure 31. Scatterplots showing the Pearson correlation coefficients between the population distribution based on interpolated (MFD) mobile phone data during night-time (2 AM – 5 AM) and population register data. The axis values represent the relative share of population in a given statistical 250 m x 250 m grid square of all grid squares in the HMA (n = 13231).

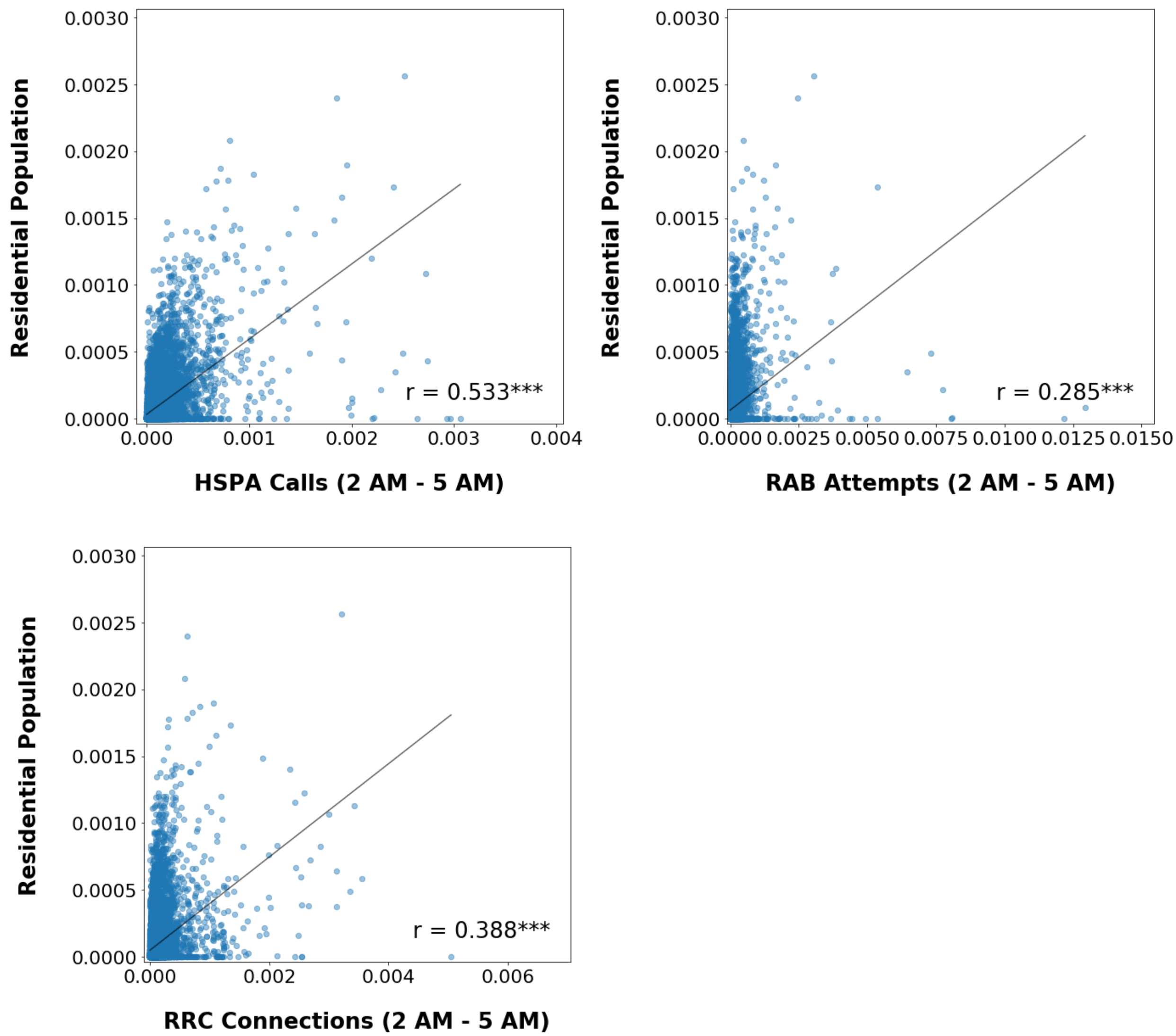


Figure 32. Scatterplots showing the Pearson correlation coefficients between the population distribution based on interpolated (AW) mobile phone data during night-time (2 AM – 5 AM) and population register data. The axis values represent the relative share of population in a given statistical 250 m x 250 m grid square of all grid squares in the HMA (n = 13231).

6.1.2. Statistical evaluation of the interpolation methods

Concerning the feasibility of the means of interpolation, the results of the statistical evaluation suggest that the more advanced MFD interpolation method is the preferred option for each mobile phone data type (Table 7). In the case of HSPA calls for example, the correlation coefficient decreases by 0.15 from 0.683 to 0.533 during night-time when using the AW interpolation method instead of MFD interpolation. Figure 33 shows that the correlation between the interpolated mobile phone data and residential population varies significantly during the day on an average weekday regardless of the chosen interpolation method (see also Appendix 4). HSPA calls interpolated using both the MFD and AW method follow the same overall pattern with highest correlation against reference data during the late evening and night-time, when people are most likely to be at home according to the time use survey (see Figure 28). The correlation coefficient of individual hours is highest between 10 PM – 11 PM for the MFD method ($\rho = 0.717$) and one hour later for the AW method ($\rho = 0.523$). The correlation with residential population data starts to decrease after the peak at night being steepest between 7 AM – 9 AM and reaches the lowest point during the afternoon between 1 PM – 2 PM (MFD: $\rho = 0.384$, AW: $\rho = 0.327$) in the case of both interpolation methods.

The difference between the two interpolation methods in terms of correlation with residential population is highest during night-time, when the MFD clearly outperforms the AW method. During daytime, the difference between the two interpolation methods is smaller (see also Appendix 5). However, as Figure 35 illustrates, the AW method allocates a significant share of the population to restricted areas unlike the MFD method, which accounts for ancillary data including land use weights in the modelling phase. The population distribution classified by activity function type sees only little change between night-time and daytime using the AW interpolation method, while the MFD method produces a clear difference between night-time and daytime. Overall, the results of the MFD interpolation method are more in line with the time use survey results. According to the time use survey, approximately 90 % of the population are at home at night (2 AM – 5 AM) during on an average weekday.

In comparison, the AW method allocates approximately only a third of the population to residential areas, whereas the corresponding share in the case of the MFD method is over 85 %. Furthermore, when analysing the relationship between interpolated mobile phone data and work place locations, the correlation is higher with the MFD method during typical work hours ($\rho = 0.648$) compared to the AW method ($\rho = 0.608$) (Figure 34; Appendix 4). The results suggest that the MFD method is a more feasible alternative for refining mobile phone data and is consequently used in this study for further analysis of mobile phone data.

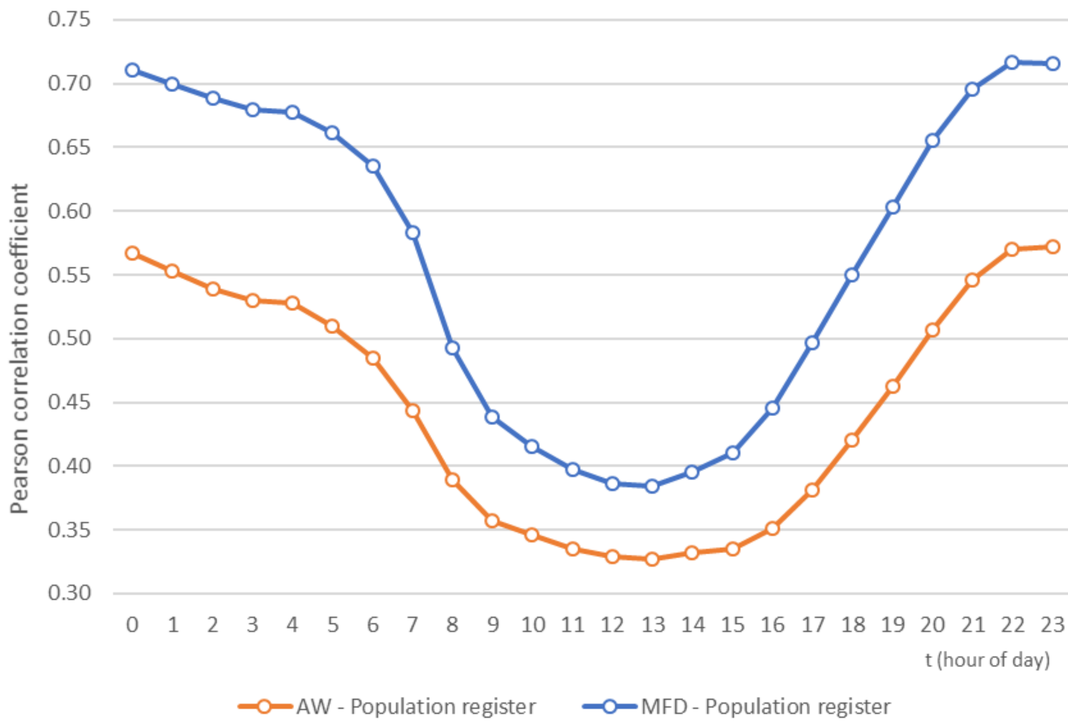


Figure 33. Correlation coefficient (Pearson) between population register and interpolated 24h population (HSPA calls) on statistical 250 m x 250 m grid squares (n = 13231) on an average weekday.

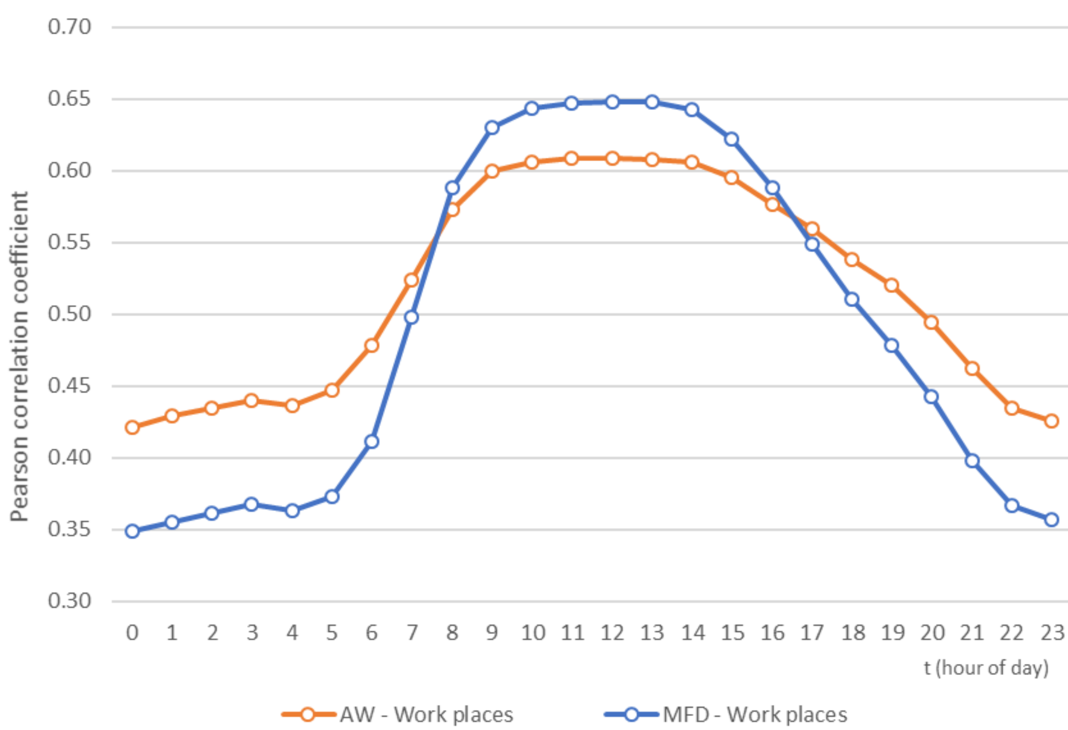


Figure 34. Correlation coefficient (Pearson) between number of work places and interpolated 24h population (HSPA calls) on statistical 250 m x 250 m grid squares (n = 13231) on an average weekday.

The presented statistical measures are based on global statistics. Hence, the performance of the two interpolation methods and the mobile phone data types may vary spatially within the study area. Consequently, the AW method or RRC connections or RAB attempts may perform better on an individual grid cell level or in certain areas compared to the MFD method and HSPA calls. However, on study area level, HSPA calls interpolated using the MFD method is clearly the most feasible approach.

Figure 35. Population distribution of the reallocated population by activity function type during night-time (upper) and daytime (lower) using the MFD and AW interpolation methods for all three types of mobile phone datasets (HSPA, RRC, RAB) compared to the original time use survey data.

6.1.3. *Difference in population distribution between mobile phone and reference data*

The comparison between the spatial distribution of the population based on interpolated (MFD) mobile phone data (HSPA calls) and the population register data uncovers distinct differences between the datasets. The differences in the relative share of population are present both during daytime and night-time.

During night-time, the population register data underestimates the population especially in the Helsinki city centre, the area around Helsinki-Vantaa airport and Pasila-Ilmala logistics area when compared to the mobile phone data (Figure 36). These areas have night-time work and service functions. The population register data overestimates the share of present population mostly in residential areas, such as Eira, Töölö, Kallio and Pakila. There is little difference between the two datasets in the sparsely populated areas of Northern Espoo and in the easternmost part of the HMA.

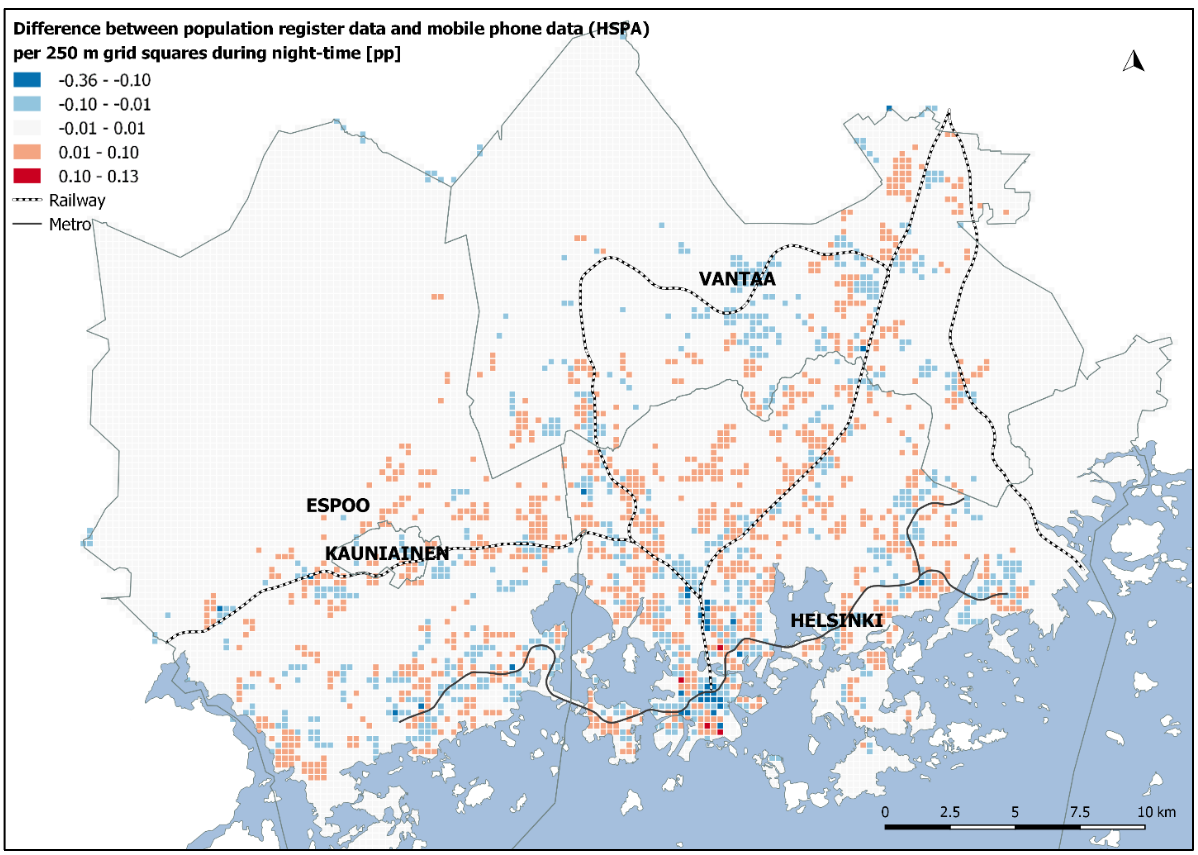


Figure 36. Absolute difference in population distribution (percentage points, pp) between population register and mobile phone data (HSPA) during night-time (2 AM – 5 AM). The values represent an average weekday. Blue grid squares (250 m x 250 m) indicate areas, where the population register underestimates the relative share of population compared to the mobile phone data. The mobile phone data was interpolated using the MFD method. See Appendices 6 and Appendix 7 for the corresponding maps for RRC and RAB.

During daytime, the population register underestimates primarily typical work place areas (Figure 37, see also Figure 12) compared to mobile phone data. These areas include for example Helsinki city centre, Pasila-Ilmala north of the city centre, Pitäjänmäki, Ruoholahti, Kalasatama and Itäkeskus in Helsinki. In Espoo, corresponding areas are focused along the railway and the metro line, including Leppävaara, Espoo city centre and Kera logistics area, Aalto University Campus, Tapiola and Matinkylä. In Vantaa, these areas include Tikkurila and the area around and south of Helsinki-Vantaa airport. On the contrary, the population register overestimates residential areas in many parts of the study area predominantly in Helsinki, such as in Töölö, Eira and Kallio in the inner city of Helsinki, large parts of Lauttasaari and Vuosaari and the detached-house oriented areas in northern parts of Helsinki, such as Pakila and Puistola.

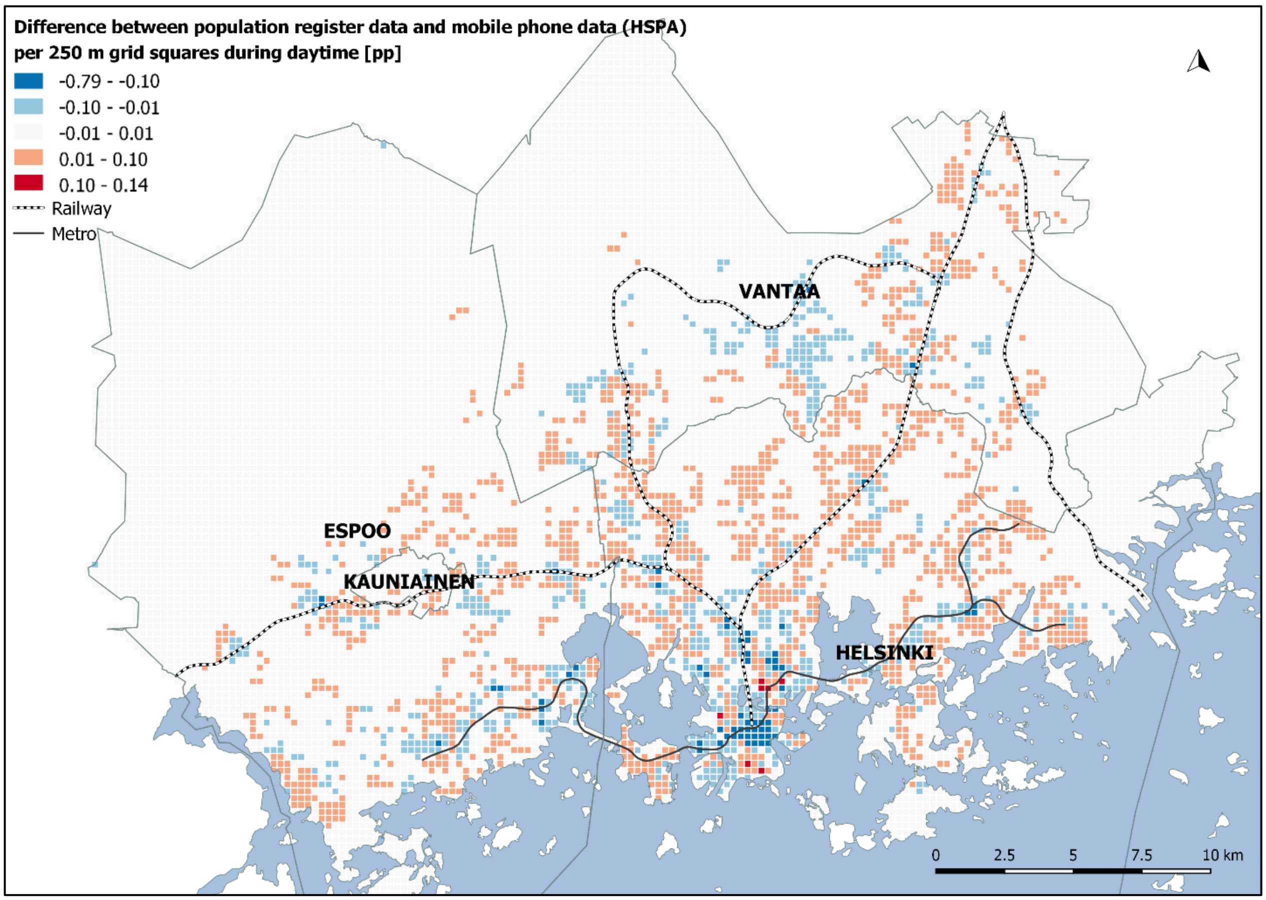


Figure 37. Absolute difference in population distribution (percentage points, pp) between population register and mobile phone data (HSPA) during daytime (3 PM – 4 PM). The values represent an average weekday. Blue grid squares (250 m x 250 m) indicate areas, where the population register underestimates the relative share of population compared to the mobile phone data. The mobile phone data was interpolated using the MFD method. See Appendices 6 and Appendix 7 for the corresponding maps for RRC and RAB.

Overall, the spatial differences between the population distribution based on interpolated mobile phone and static population register data appear larger during daytime compared to night-time. This supports the findings of the correlation results, where the relationship between the two datasets is weakest during the typical working hours. Based on a visual inspection, the differences between the mobile phone data and the reference data follow, however, a largely similar spatial pattern during both day and night-time. In both cases, large parts of the inhabited areas of the study area are either over- or underestimated by the static population data. The population register data overestimates mainly residential areas, while the underestimated areas typically exhibit work and service functions.

6.2. Diurnal spatiotemporal patterns of population dynamics

6.2.1. Non-aggregated grid-level patterns

Various spatiotemporal patterns of population distribution can be extracted from mobile phone data in the HMA on an average weekday on grid-level. Figures 38–39 illustrate the diurnal variation of the hourly population distribution in the study area on an average weekday (Monday–Thursday) based on interpolated (MFD) mobile phone data (HSPA calls)[2].

Overall, the results show that the population distribution varies over the course of the day in the HMA. Based on visual interpretation, the population is distributed more evenly in space during night-time than during daytime, when the population is more concentrated to specific locations. At night, the population distribution follows largely the locations of residential areas, as already indicated by the comparison to residential population data. However, a significant concentration of people can also be found in the Helsinki city centre, which has a relatively low share of residential functions.

Similarly as during night-time, the highest concentration of people during daytime is found in the Helsinki city centre, but also other hotspots can be distinguished. These include locations with major work and service functionalities, such as Pasila and Kallio, Itäkeskus and Malmi in Helsinki, Tikkurila and Helsinki-Vantaa airport in Vantaa and Leppävaara and the city centre in Espoo. Accordingly, the daytime hotspots and how their sizes vary in time reveal the polycentric urban structure of the study area. The Helsinki city centre, appears as the most significant centre in the study area throughout the whole diurnal cycle, while most other daytime hotspots fade out towards the night.

[2] The maps of all 24 hours have been compiled to an animation that shows more clearly the differences between the subsequent hours. The animation is available at: https://blogs.helsinki.fi/accessibility/2018/10/09/the-24-h-population-dynamics-of-the-finnish-capital-region-uncovered/

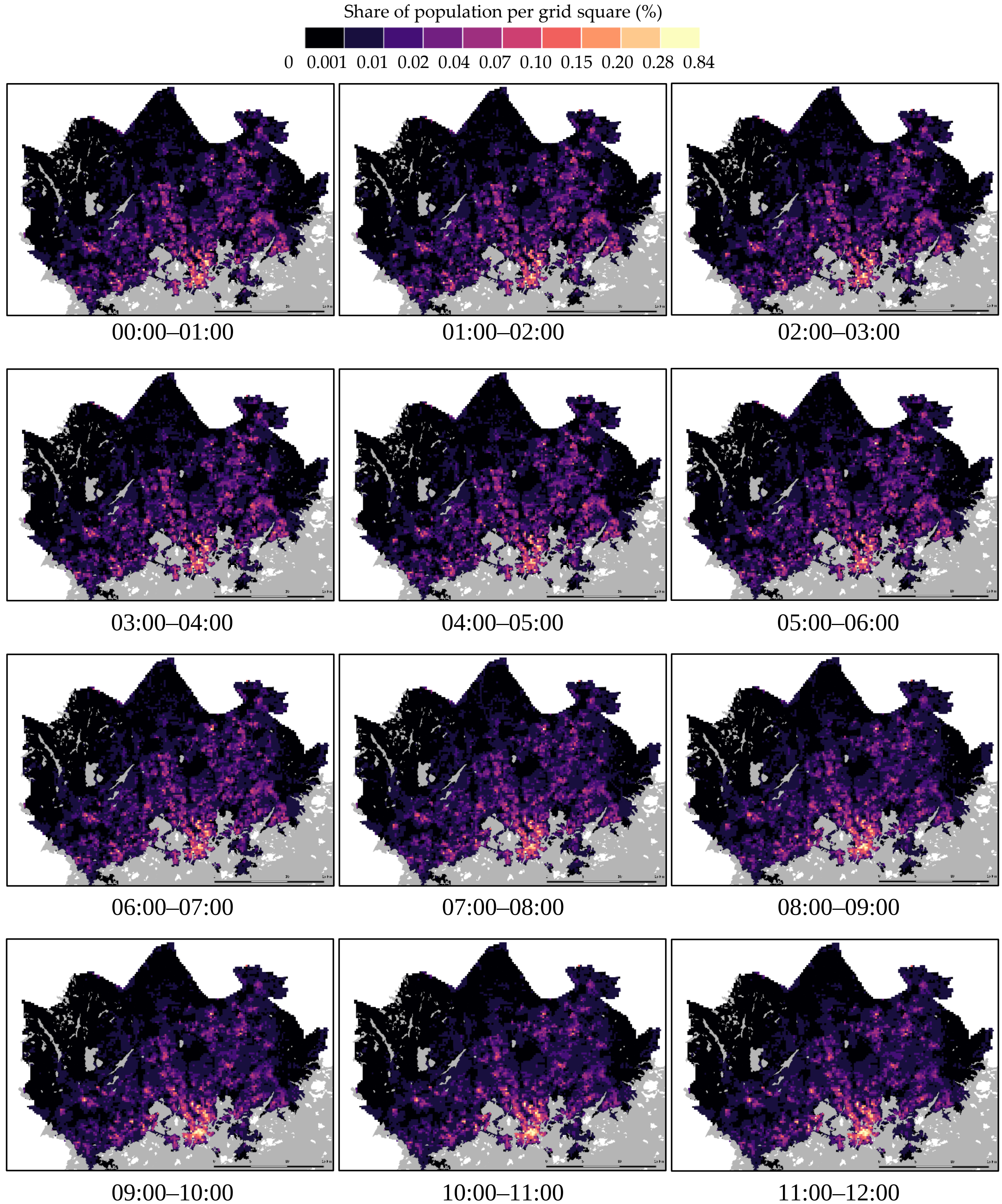


Figure 38. Hourly population distribution in the HMA based on mobile phone data (HSPA calls) between midnight and noon. The values represent the relative share of population of the HMA total (%) per given hour on statistical 250 m x 250 m gird cells. The mobile phone data is interpolated using multi-temporal function-based dasymetric interpolation (MFD) method. For a more detailed version of the hourly maps, see Appendix 8.

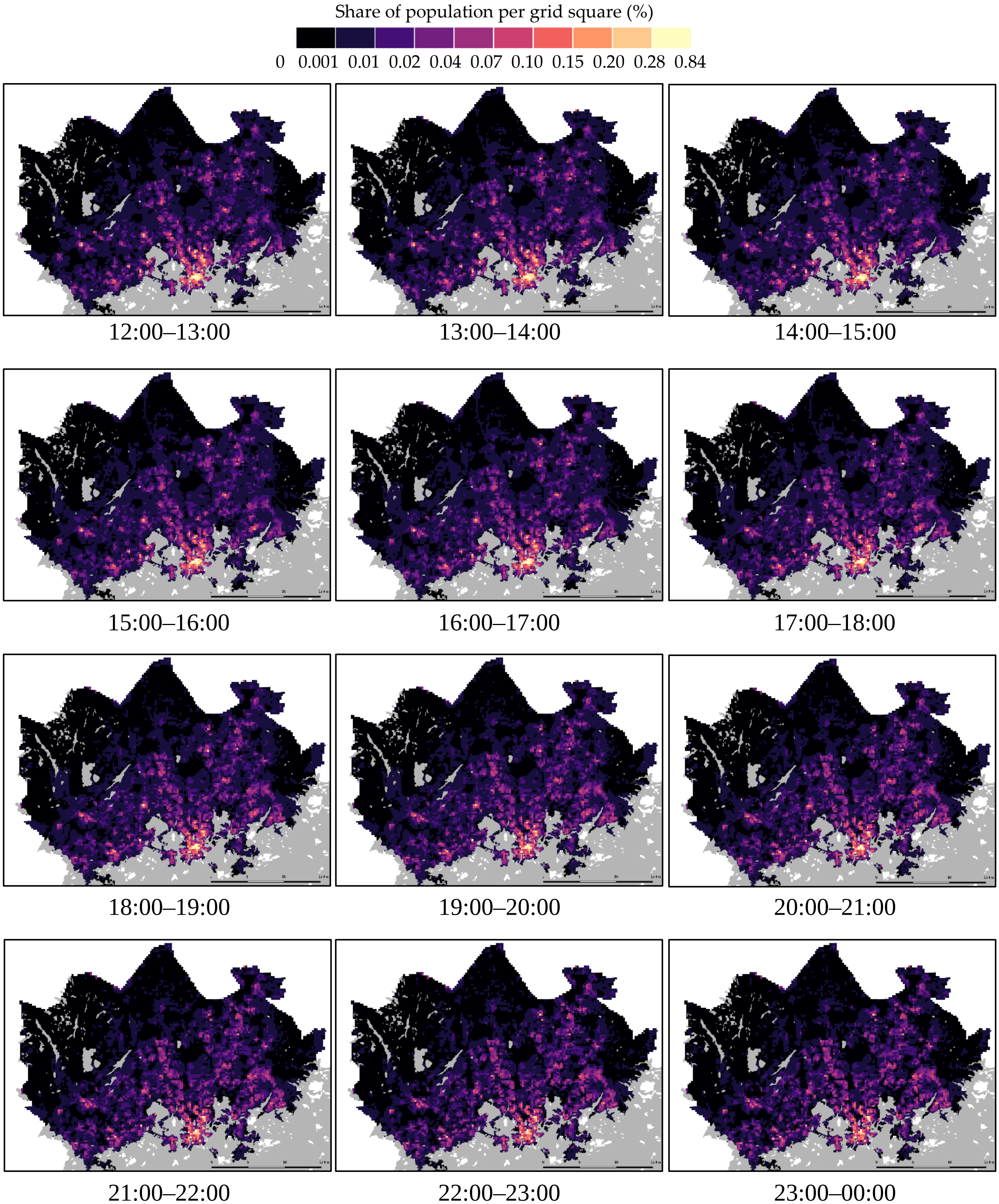


Figure 39. Hourly population distribution in the HMA based on mobile phone data (HSPA calls) between noon and midnight. The values represent the relative share of population of the HMA total (%) per given hour on statistical 250 m x 250 m gird cells. The mobile phone data is interpolated using multi-temporal function-based dasymetric interpolation (MFD) method. For a more detailed version of the hourly maps, see Appendix 8.

Figure 40 and Figure 41 show how the population distribution differs between night and day in the HMA on the level of individual statistical grid squares. Both figures illustrate that there is a clear difference in the population distribution between day and night in most parts of the study area. Areas with a higher share of population of the HMA total during night-time (3 AM – 4 AM) compared to daytime (3 PM – 4 PM) comprise mainly of areas with a relatively high share of residential functions, such as Vuosaari, Kontula and Mellunmäki in eastern Helsinki. In residential areas in general, the decrease during daytime is typically between 10–50 % compared to the night-time value (see Figure 41).

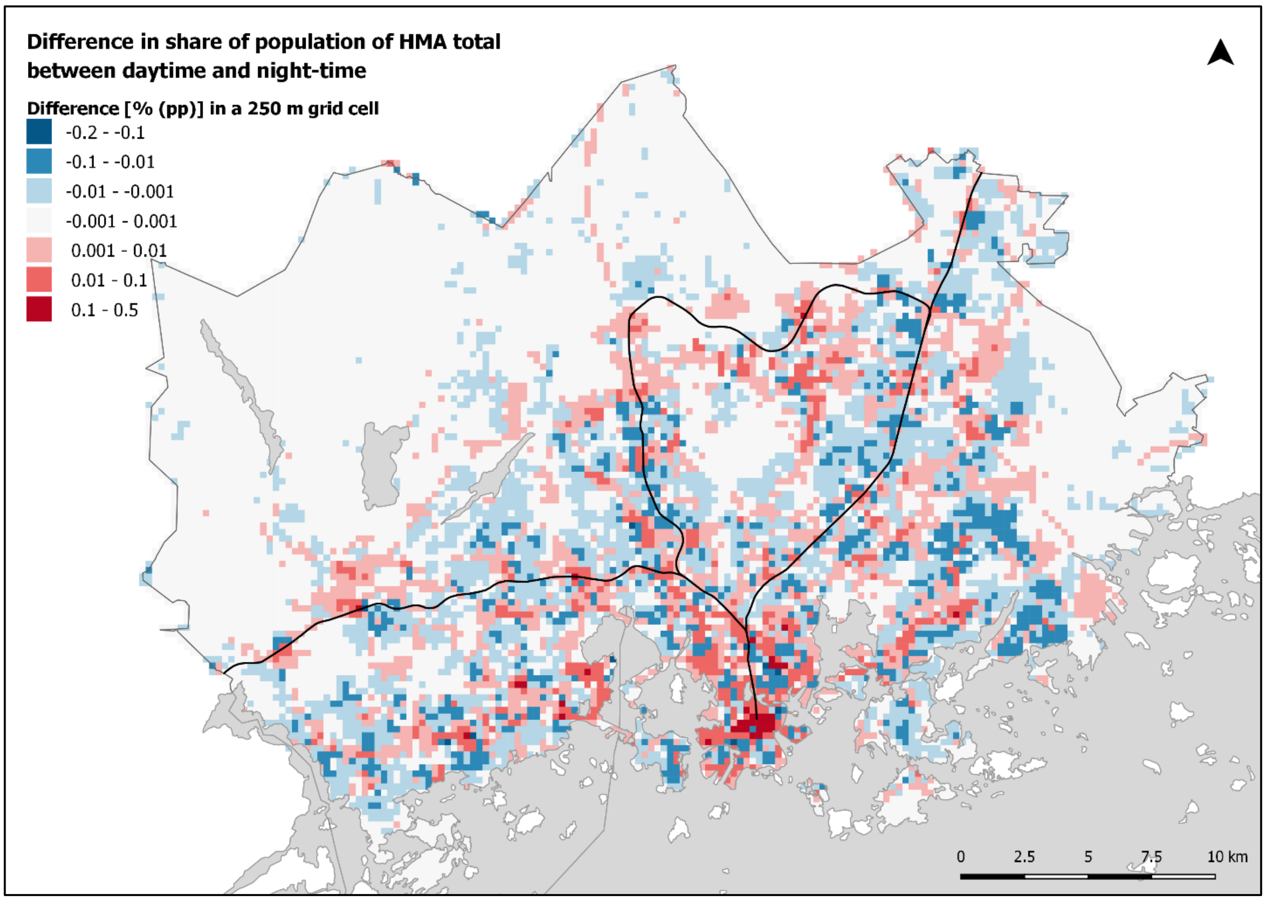


Figure 40. The absolute difference (percentage points, pp) in the share of present population of HMA total during given hour between daytime (3 PM–4 PM) and night-time (3 AM–4 AM) on statistical grid squares (250 m x 250 m). Daytime population is used as the baseline. Red squares indicate a higher presence of population during daytime compared to night-time and blue squares a higher value during night-time compared to daytime. The values are based on hourly median of HSPA calls during weekdays (Monday–Thursday).

Areas with a higher relative share of the study area population during the day compared to the night, in turn, include the city centre of Helsinki and other areas with major work, education or service functions, such as Pasila and Otaniemi. For example by the central railway station in Helsinki city centre (YKRID: 5975375), the relative share of the total population in HMA increases from night-time (approx. 0.1 %) to daytime (approx. 0.5 %) by over 250 % (267.7 %). If assumed that the population of the study area is constant during these hours and equal to the statistical residential population within the area (1 154 967 on 31.12.2017), the increase corresponds to approximately 4000 persons. The highest relative increase can be seen along the major roads, which indicates the start of the afternoon rush hour. In addition to major work areas, also most green areas (e.g. Helsinki Central Park) experience a significant increase in the share of population of the HMA total during daytime compared to night-time, although the absolute increase in these and other sparely populated areas is relatively low (Figure 40–Figure 41).

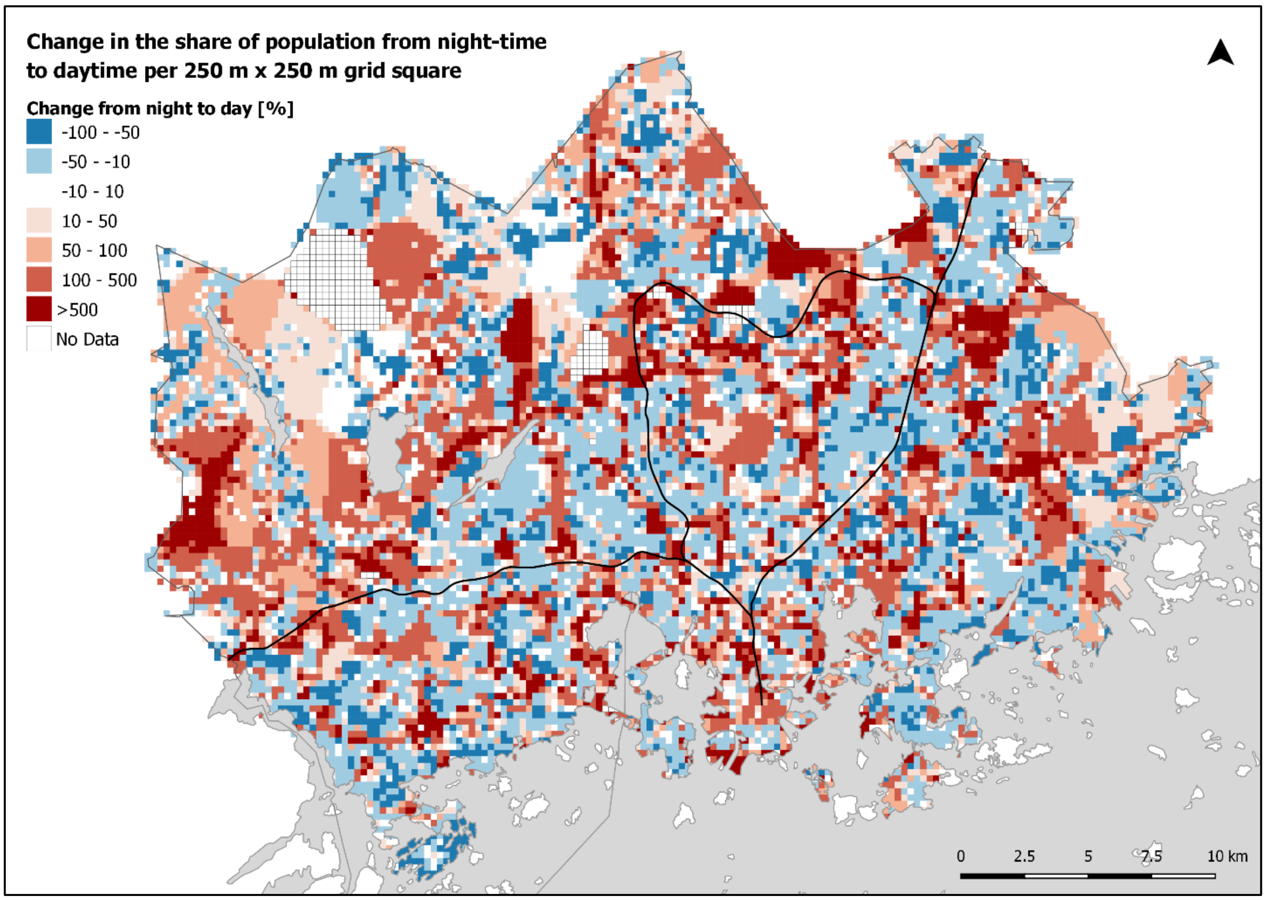


Figure 41. The change in the share of present population of HMA total (%) during to daytime (3 PM–4 PM) from night-time (3 AM–4 AM) on statistical grid squares (250 m x 250 m). Red squares indicate a higher presence of population during daytime compared to night-time and blue squares a higher value during night-time compared to daytime. The values are based on hourly median of HSPA calls during weekdays (Monday–Thursday).

The overall diurnal variation and its magnitude on individual grid cell level is presented in Figure 42. The map shows, that the variation in the share of population within the study area is not equally distributed in space and that the 24-hour variation follows the same overall pattern as the difference maps between night-time and daytime. The variation is highest in the inner city of Helsinki and especially in the city centre, Pasila and Teollisuuskatu area in Vallila, all of which have a relatively low rate of residential functions in comparison to work functions. The residential areas show up as areas with moderate variation and green areas and other sparsely populated areas have a relatively low diurnal variation.

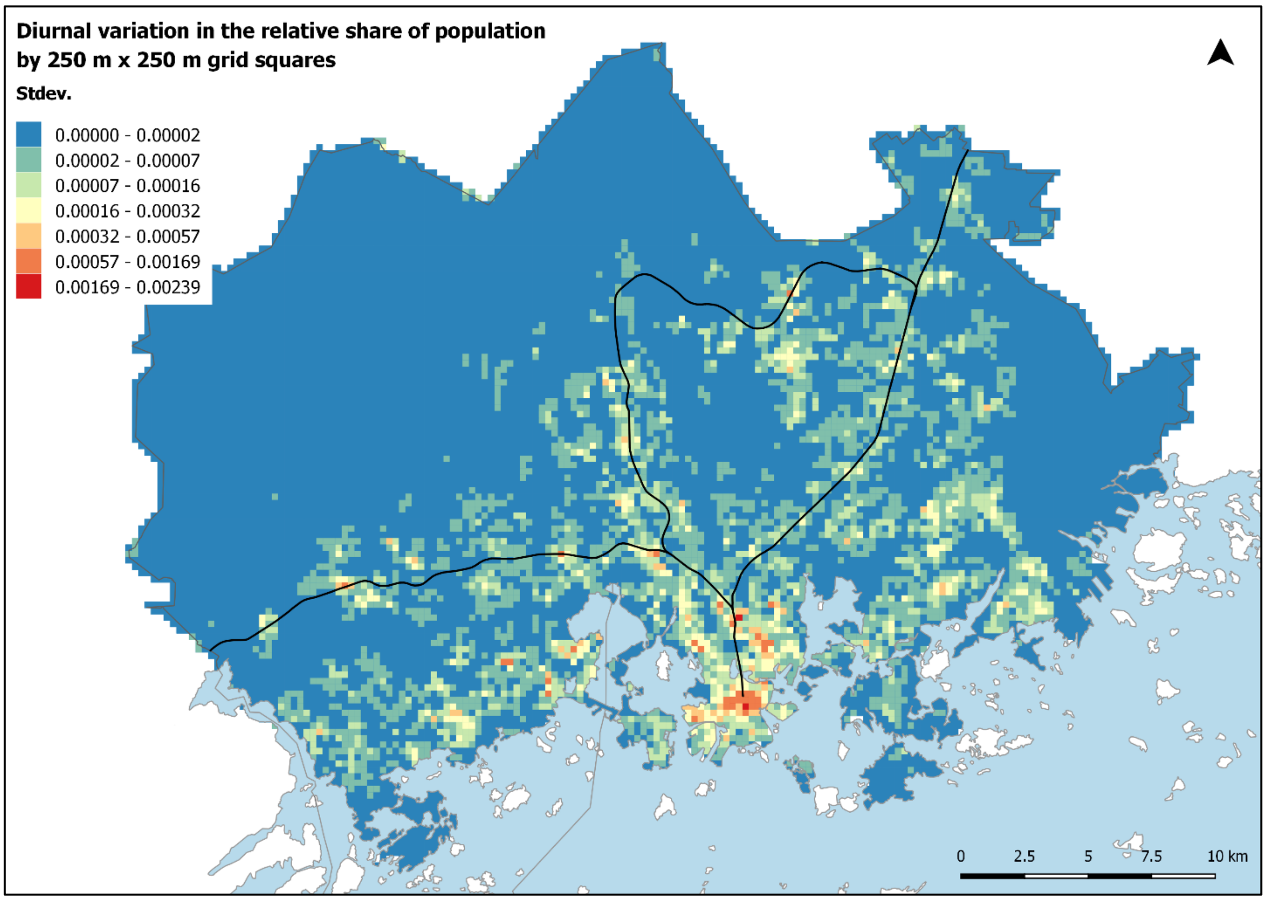


Figure 42. The diurnal variation in the relative share of population in HMA per 250 x m 250 m grid squares. The values represent the standard deviation of the daily mean share of population within a given grid cell. Blue squares indicate areas, where the relative share of population of the HMA total has a low variation, while red squares indicate areas with highest variation.

In addition to the differences between night and day or general daily variation, the hourly maps also show how the population distribution fluctuates between them throughout the diurnal cycle. During the morning hours, particularly between 7 AM – 10 AM, the share of total population in the HMA increases significantly in the Helsinki city centre and other areas with major work functions, while the relative importance of the residential areas decrease in comparison to the night (Figure 38). In the afternoon, the importance of the work place areas starts to gradually decrease, while the share of population is starting to increase in the residential areas, although the transition is more fragmented compared to its counterpart in the morning (Figure 39).

Although the data does not contain information of origin-destination flows of people, it is likely, that the simultaneous increase and decrease patterns in the share of population are related. For example the decrease of work place areas during the day is likely to be partly a product of the night-time population of residential areas, where the population is lower during the typical work hours. The typical home-work-home pulse can be seen for instance in the Aalto University campus area in Otaniemi. During daytime, the concentration of population is highest in the campus area, but as the evening draws closer, the concentration shifts to the residential area at the end of the cape. Similarly, population in recreational areas and locations with service-driven functions, such as shopping centres, starts to increase as the share of population decreases in work place areas in the late afternoon (Figure 38). This finding is in line with the time use statistics, which suggest that people undertake shopping or leisurely activities after work hours and these activities could be chained for example to the trip back home from work.

Certainly, not all population behind the patterns are from within the HMA, but for example daily visitors due to commuting. The incoming and outgoing population flows can also be distinguished from the maps along the major highways (e.g. Hämeenlinnanväylä) in the morning (7 AM – 9 AM) and afternoon (1 PM – 6 PM), especially during rush hour time. This finding highlights that the daily population within the study area is affected by in- and outbound population flows or by people passing through. Consequently, the absolute number of people in the study area may vary between hours. Nevertheless, the results based on the relative shares of present population in the HMA are equally relevant, as visiting persons are included, unlike in residential population register data.

Temporal patterns of the population distribution can also be analysed on the level of individual grid squares. Figure 43 shows how the share of the HMA total population varies within six individual grid squares in the study area each representing a main activity function type: a) transport, b) work c) recreation, d) education, e) residential use and f) service and retail. The six plots illustrate that areas with different functions have different types of temporal patterns matching the patterns found on study area level. Areas with work and education functions have an inverted u-shaped curve, with low activity outside the typical working hours. Residential areas follow the opposite shape of the work and education areas with the highest share of population is present at night and lowest during midday. The relative population concentrations in the shopping malls and recreational areas are highest in the evening, whereas the transportation ways follow the patterns of the rush hour times in the morning, at midday and in the late afternoon.

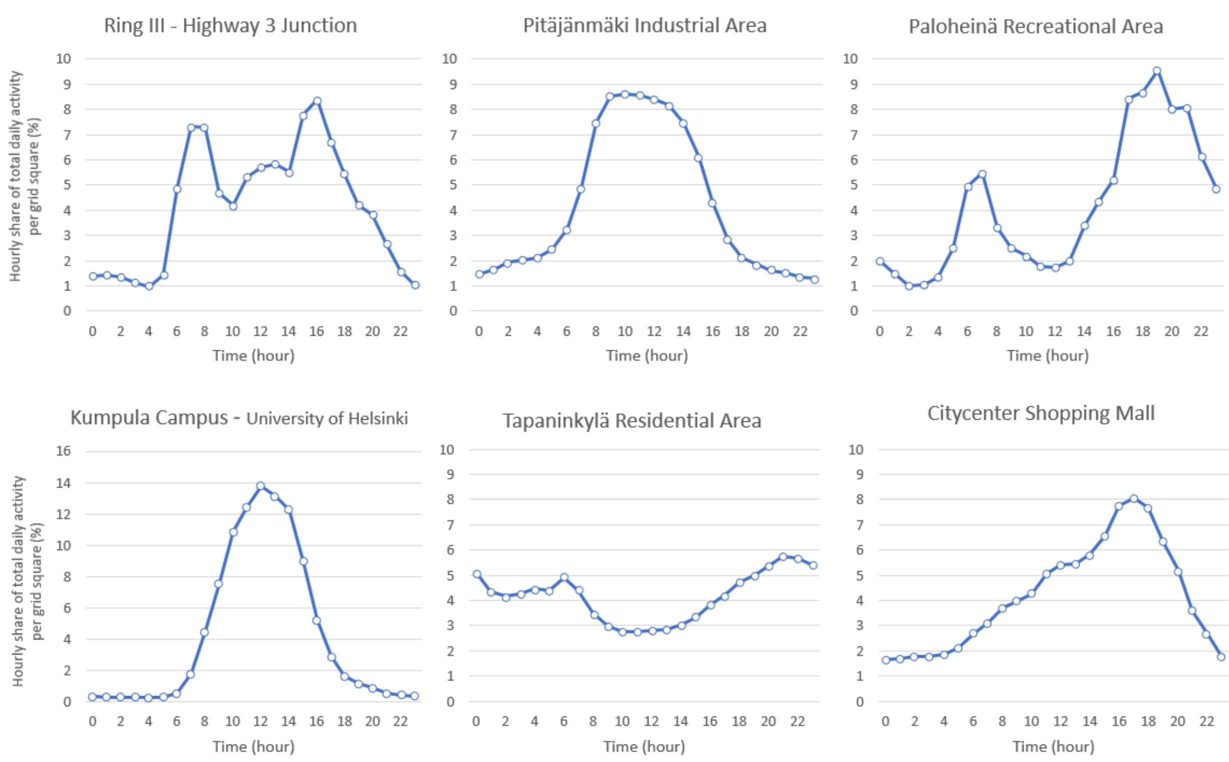


Figure 43. The hourly variation of the share of present population of HMA total in six individual 250 m x 250 m grid cells. The values represent the hourly share of the daily sum of activity in the given grid cell.

6.2.2. Aggregated grid-level patterns

The diurnal population distribution patterns can also be analysed on an aggregated level by combining individual grid cells. Here the population dynamics are analysed on 1) municipal level, 2) within municipality level and 3) based on clustering.

The municipal level analysis reveals that the municipalities in the HMA have distinct patterns. **Helsinki** contains the majority of the population within HMA during every hour of the day (Figure 44). The relative share of the study area total population is highest at noon (~59 %), after which it gradually decreases staying above 56 % until 6–7 PM. During the late evening and night, the range varies between 53–56 % being lowest between 5–7 AM. When assumed that the population in HMA total is equal to the statistical residential population of the whole study area (1 154 967 on 31.12.2017, see Table 1), the mean amount of present population in Helsinki during night-time (2–5 AM) of the HMA total (~55 %) is approximately 638 000 people. This corresponds well with the residential population of Helsinki based on the official statistics (643 272 on 31.12.2017). The correspondence is even higher at night-time when using the values of 10–11 PM in the evening (approx. 645,000).

In **Vantaa**, the early morning drop in Helsinki is mirrored as the highest peak of the diurnal cycle (Figure 44). The preceding and following drop suggest that population is passing through the city during the morning, as Vantaa is located along major transport ways including the airport and the timing coincides the morning rush hour. Since the values are shares of the HMA total, an increase in one area shows up as a relative decrease in another, unless the increase is higher in all other gird cells. Thus, the drop in Helsinki can be caused by a higher increase in Vantaa, Espoo or Kaunianen, regardless if the absolute increase of population would be positive in all municipalities. As opposed to Helsinki, the relative share of population in Vantaa is lowest during the day between 11 AM and noon (~18 %). In other words, the share of population in Vantaa of the HMA total is higher during night-time (~21 %) than during the day. When transforming the mean night-time share of population in Vantaa to a number of people based on the number of inhabitants (223 027 on 31.12.2017), the estimate is approximately 237 000. This overestimates the official statistics by 6 %. As in Helsinki, the correspondence with the number of inhabitants in Vantaa is higher between 10–11 PM (approx. 222 000) than at night.

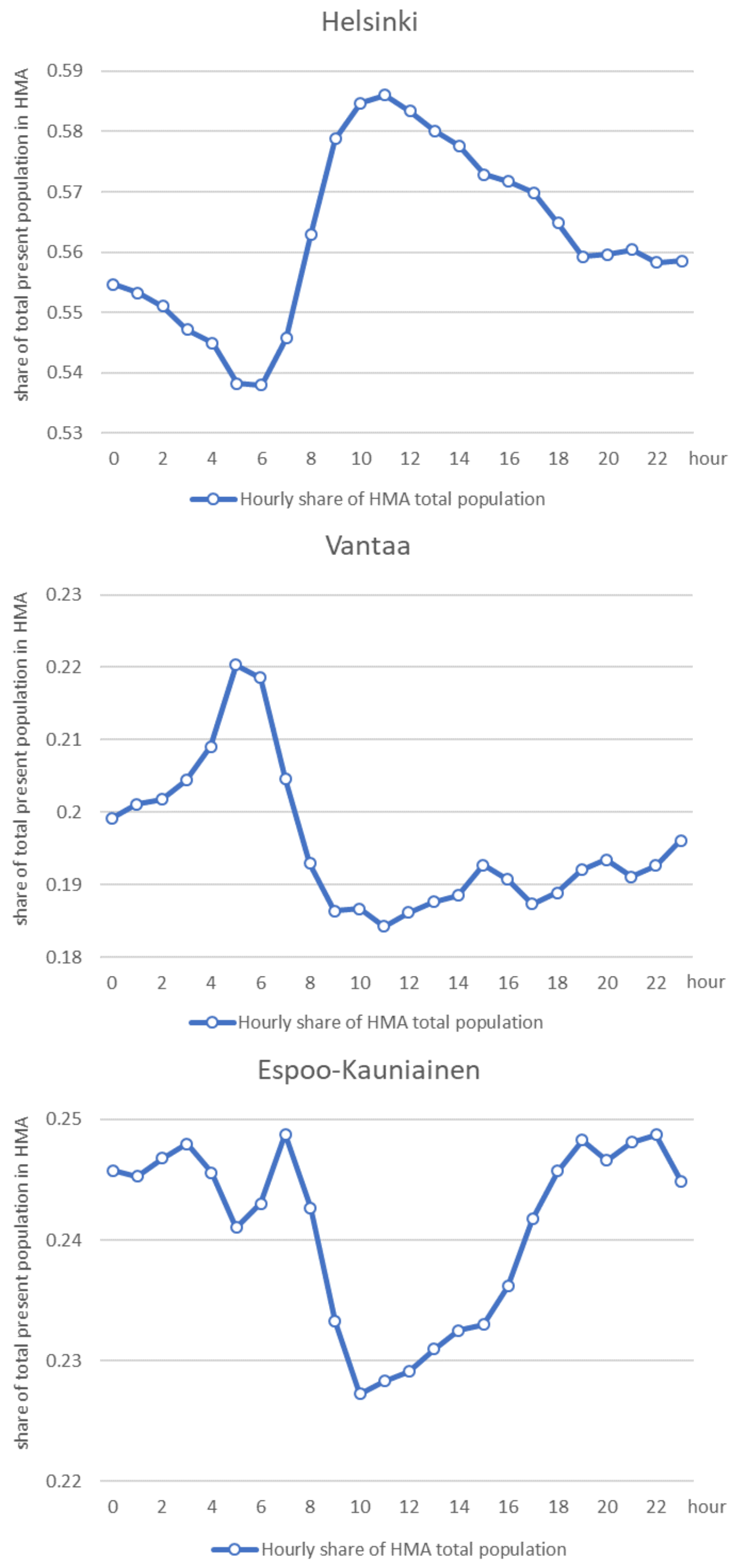


Figure 44. The estimated hourly share (1=100 %) of present population during an average weekday (Mon–Thu) in Helsinki, Espoo-Kauniainen and Vantaa compared to the hourly HMA total.

Similarly as in Vantaa, the share of the total population of the HMA is lowest during the daytime in **Espoo** (including the enclave of Kauniainen) (Figure 44). During the evening and night-time, the relative share of population varies between 24–25 % of the HMA total. The share of population is highest between 7AM and 8 AM, after which it decreases and reaches the lowest point between 10 AM to 11 AM. The share of the HMA total continues to increase until the evening, forming a u-shaped curve during the typical working-hours. If assumed that the population in HMA total is equal to the statistical residential population of the whole study area, the estimated share of population in Espoo and Kauniainen combined during night-time is approximately 285 000 people (~25 %). This corresponds well with the residential population of Helsinki based on the official statistics of the two municipalities combined, which is 288 668 (31.12.2017). Again, the estimated population using the values of 10–11 PM (approx. 287 000) have a higher correspondence with the residential population than night-time between 2–5 AM. Overall, the daily fluctuation in the share of population of the HMA total is smallest in Espoo and highest in Helsinki and.

The data allows also zooming into smaller aggregated units, such as the **inner city of Helsinki**, which provides a more fine-grained view to the population dynamics within Helsinki. The relative share of present population of the HMA total in the inner city of Helsinki increases significantly during the daytime compared to the night-time (Figure 45). The estimated hourly share of present population increases by approximately 52.4 % from 5–6 AM, when the relative share of the diurnal population within the area compared to the HMA total is lowest (21 %), to its peak between 11 AM–midday (32 %). After midday, the share of population decreases continuously until 5–6 AM. The increase is highest in the morning between 7–8 AM to 8–9 AM being approximately 14.6 %. The night-time share of present population is approximately 22 % of the HMA total. The inner city represents approximately 5 % of the whole study area based on the statistical grid cells.

The increase in the inner city of Helsinki may be a result of mobility within the HMA or incoming population flows from outside the study area, such as commuters and tourists. If, however, assumed that the population of the HMA is constant and equals to the residential population (1 154 967 on 31.12.2017), the population in the inner city of Helsinki increases by approximately 130 000 people from the lowest point at night-time to the peak at noon. This corresponds to approximately a fifth of the residential population in the city (643 272 on 31.12.2017).

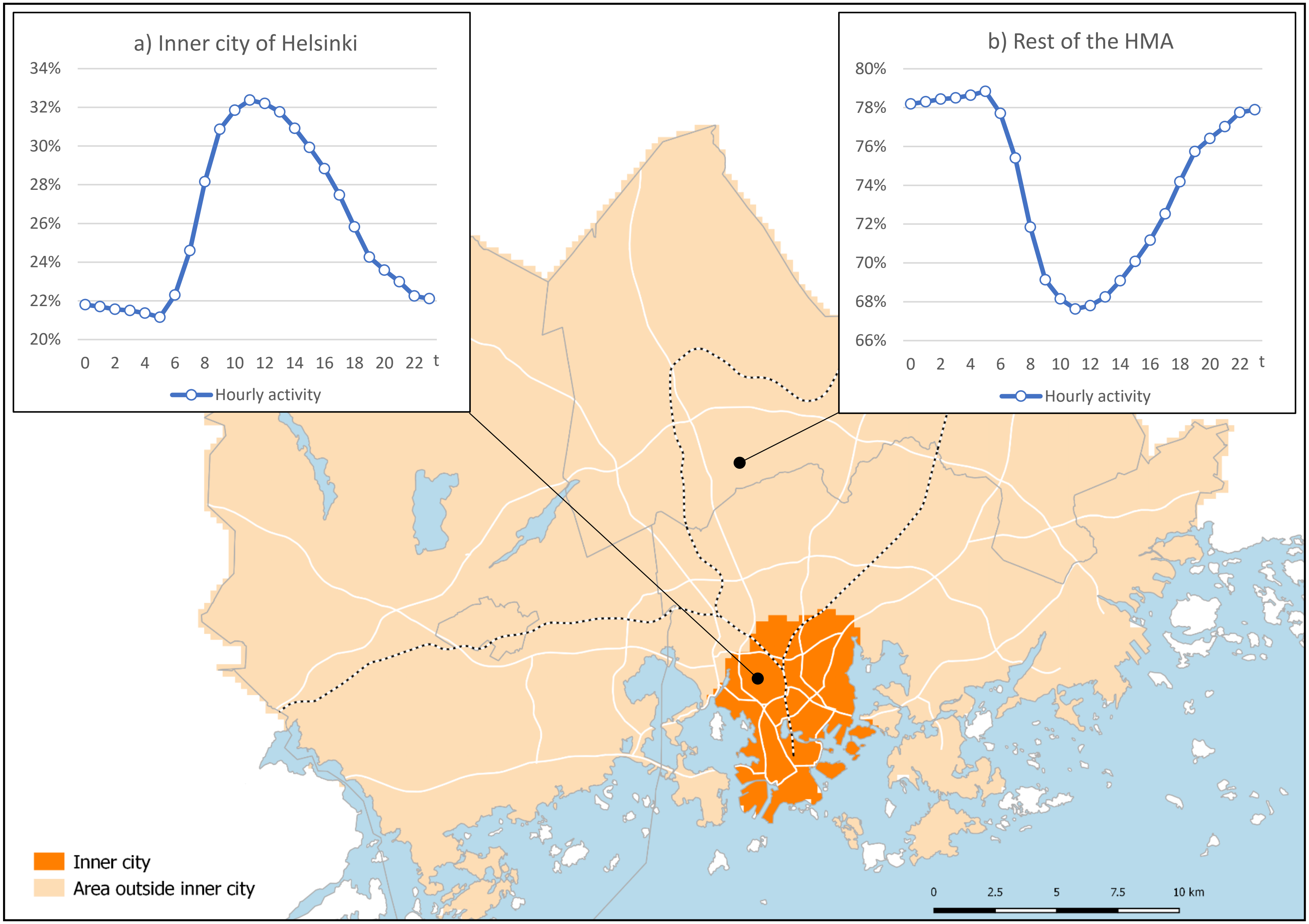


Figure 45. The estimated hourly share of present population during an average weekday (Mon–Thu) within (a) and outside (b) the inner city of Helsinki compared to the HMA total. The inner city grid cells (n=615) represent approximately 5 % of all grid cells in the study area (n=13231).

When analyzing the temporal population distribution on an aggregated level using k-means clustering, similar patterns emerge that support earlier results (Figure 46). **Cluster 1** (n = 4426) has the highest peak in the late afternoon and evening, referring to activities that take place after work. Accordingly, green areas such as Helsinki Central Park fall within this cluster. During the peak between 5 PM – 6 PM, over 28 % of the population is distributed to grid cells of this type. **Cluster 2** (n = 2246) shows a similar pattern with a stronger focus the earlier afternoon and a lower during the evening highlighting the Helsinki city centre and major transport ways. The highest hourly share of the HMA total population within this cluster is approximately 18 %, which falls between 4 PM – 5 PM. **Cluster 3** (n = 4799) has a distinct u-shape and is the only cluster where the relative share of the population decreases during daytime compared to night-time. Cluster 3 comprises largely of areas with residential functions repeating the patterns captured by the hourly maps and individual grid cells. During night-time, almost 70 % of the population is within grid squares that belong to cluster 3. The final **cluster 4** (n = 1581), in turn, shows an inverted u-shape that corresponds with the typical work hour pattern. Accordingly, many significant education or work place areas belong to cluster 4, including the cargo harbour in Vuosaari, Pasila, Pitäjänmäki and Otaniemi. The relative importance of the cluster is highest between 11 AM and midday, when approximately 21 % of the population of the HMA falls within the cluster.

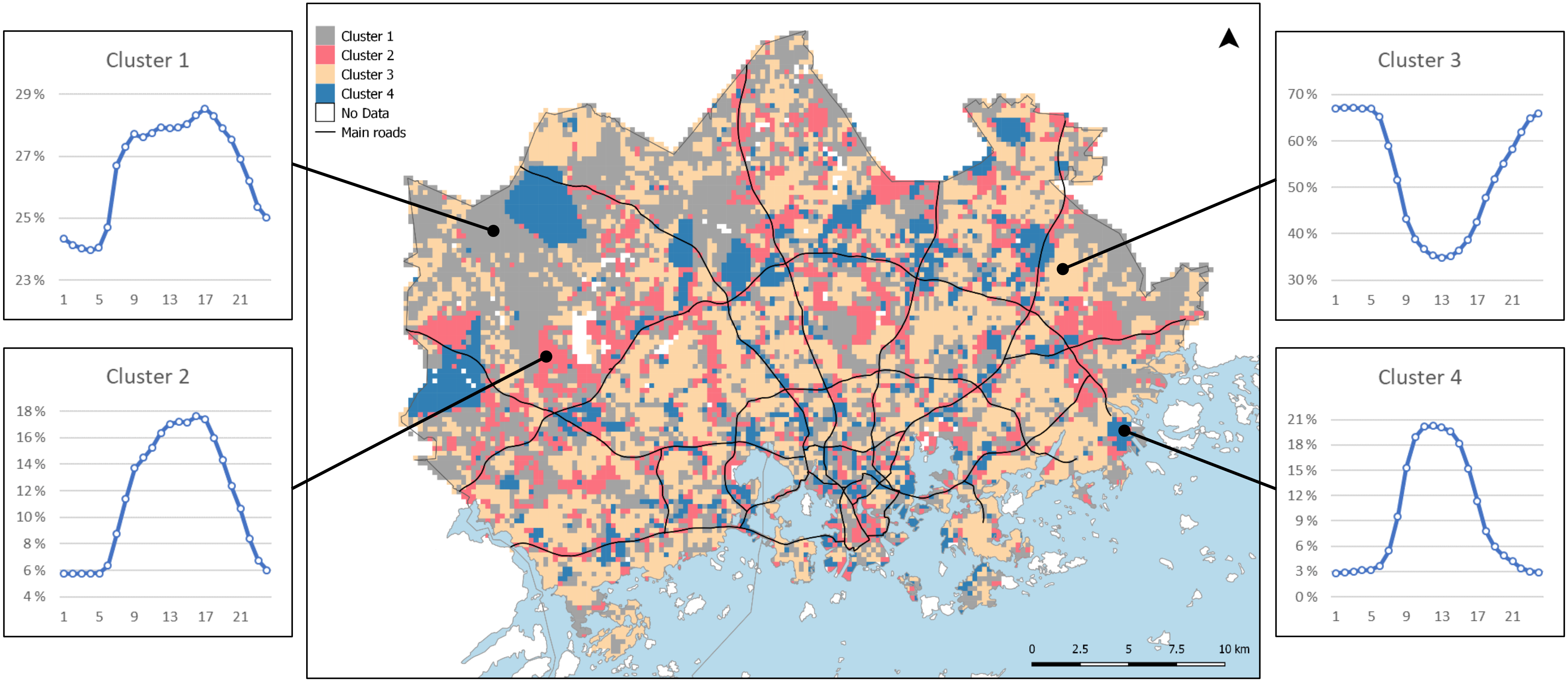


Figure 46. Statistical grid squares divided to clusters based on their diurnal temporal profiles. The classes are based on k-means clustering (k=5). One cluster with a size smaller than 15 was excluded. The temporal mean patterns of each cluster are presented in the adjacent figures. The values represent the hourly share of present population within cluster of the hourly HMA total.

6.3. Case study: Dynamic accessibility in Helsinki Metropolitan Area

6.3.1. Dynamic accessibility to major transport hubs

The results of the accessibility to major transport hubs using public transport (PT) show that the role of dynamic population varies depending on the time of day and study setting. In **Helsinki city centre** located in the junction of the central railway station and metro and tram lines, the accessibility can be considered generally good throughout the day. Between midday and 1 PM, almost 100 % of the population in the HMA can reach the Helsinki city centre within 60 minutes both in the case of dynamic and static population. Most parts of Helsinki and parts of Espoo and Vantaa along the metro and railway lines belong can reach the Helsinki city centre in less than 30 minutes during daytime. The accessibility is poorest during night-time between 3–4 AM, due to the lack of frequent PT connections (for 24-hour accessibility maps, see Appendix 9). Yet, still almost 60 % of the population in the HMA can reach the Helsinki city centre in one hour.

Static population data clearly underestimates the amount of population that can reach the Helsinki city centre within less than 40 minutes using PT (Figure 47). The difference in the share of reached population between static and dynamic population data is highest between the travel time marks of 20–30 minutes. For instance, only 25 % of the population in the HMA can reach the Helsinki city centre in 25 minutes when using static population data, while based on dynamic data, the amount of reaching population is 41 %. If assumed, that the population at noon equals to the number of residents (1 154 967 on 31.12.2017), the difference between the datasets is approximately 181 000 people. After 40 minutes, the difference between the dynamic and static population data is only small the static population slightly overestimating the accessibility.

The difference in the share of reached population between the population datasets is likely a result of the relative increase of population in the inner city of Helsinki during the day and the simultaneous decrease in the residential areas (see Figure 40; Figure 45). As a significant share of the population in HMA is already in the inner city of Helsinki during daytime, the cumulative travel time to the city centre becomes relatively shorter compared to night-time, when a larger share of people is in residential areas located further away. This is further emphasized by the low share of residential functions near Helsinki city centre. Accordingly, the difference between the datasets is lowest at night, when also the PT service interval and extent is lower.

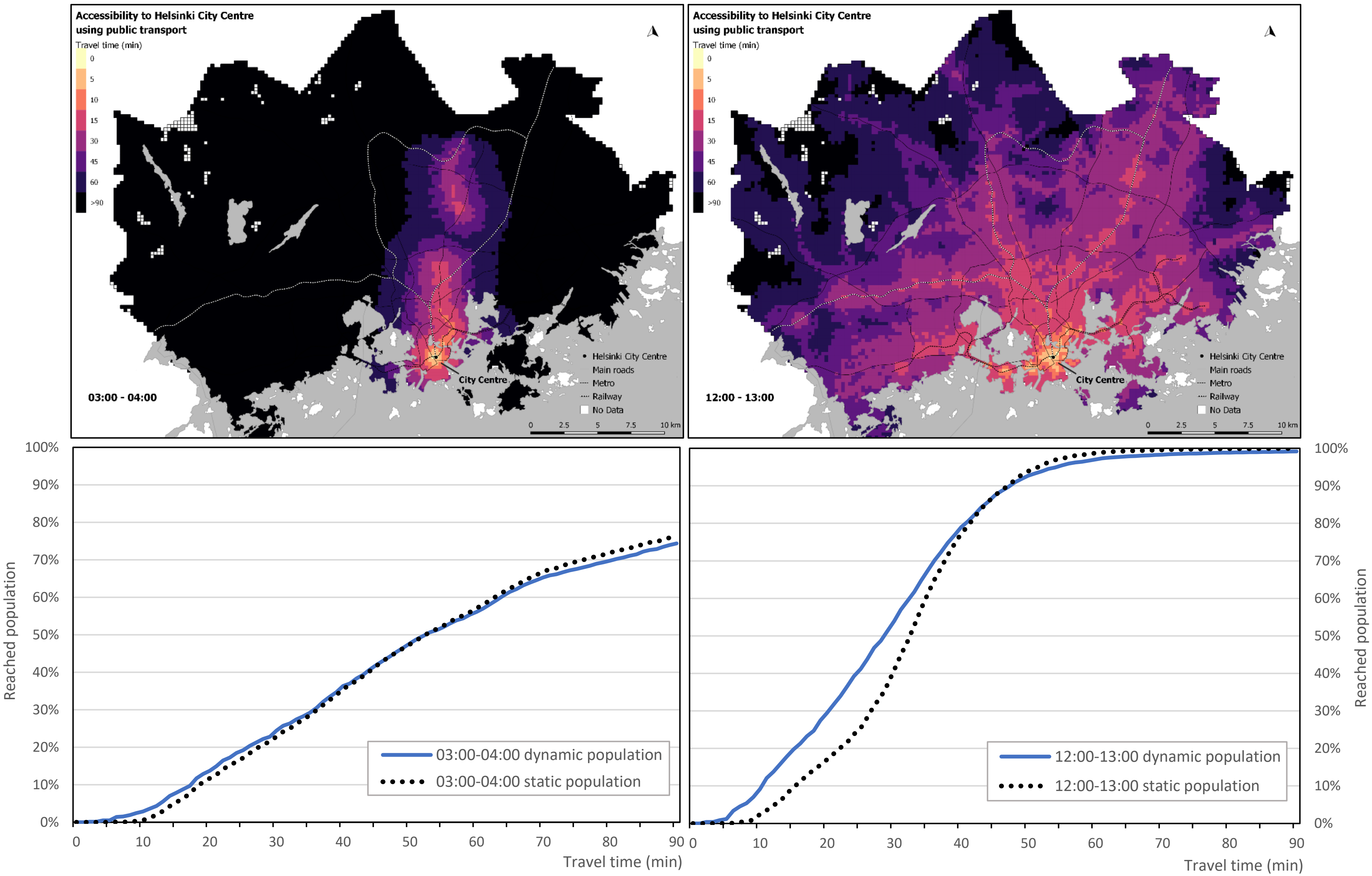


Figure 47. Accessibility to Helsinki city centre (central railway station) using public transport between a) 03:00-04:00 and b) 12:00-13:00. The plots show the cumulative sum of population in the HMA reaching Helsinki city centre by public transport between a) 03:00-04:00 and b) 12:00-13:00.

The accessibility to the **Helsinki-Vantaa Airport** using public transport varies greatly depending on the time of the day. Between midday and 1 PM, 66 % of the population in the HMA can reach the airport in 60 and 98 % in 90 minutes using according to the dynamic population data, while between 3 AM – 4 AM the corresponding rates are only 21 % and 37 % (Figure 48). The accessibility drops significantly after 1 AM and stays low in most parts of the study area until 4 AM – 5 AM while the railroad connection to the airport is out of operation (Appendix 9). The accessibility on the study area level is poorest between 3 AM – 4 AM, when only the strip stretching south of the airport to the Helsinki city centre is within the 60-minute accessibility zone. During midday, the airport can be reached within 45 minutes from most parts of Vantaa and within one hour in Helsinki, while the southern and central parts of Espoo and Vuosaari in Helsinki belong to the 90-minute zone.

Despite the differences in the overall accessibility between night and day, there is only a relatively small difference in the cumulative amount of people that can reach the airport between static and dynamic population data. This suggests that the variation in accessibility between night-time and midday is mainly caused by the low service frequency of public transport at night compared to daytime. Unlike in the case of Helsinki city centre, the difference between the static and dynamic population reaching the airport is also relatively even throughout the day. The static population generally underestimates the cumulative share of dynamic population reaching the airport, especially between the 50–80-minute marks, although the difference is at most only 5 percentage points. If assumed that the population in the study area is constant and equals to the number of inhabitants (1 154 967 on 31.12.2017), the difference corresponds to approximately 58 000 people.

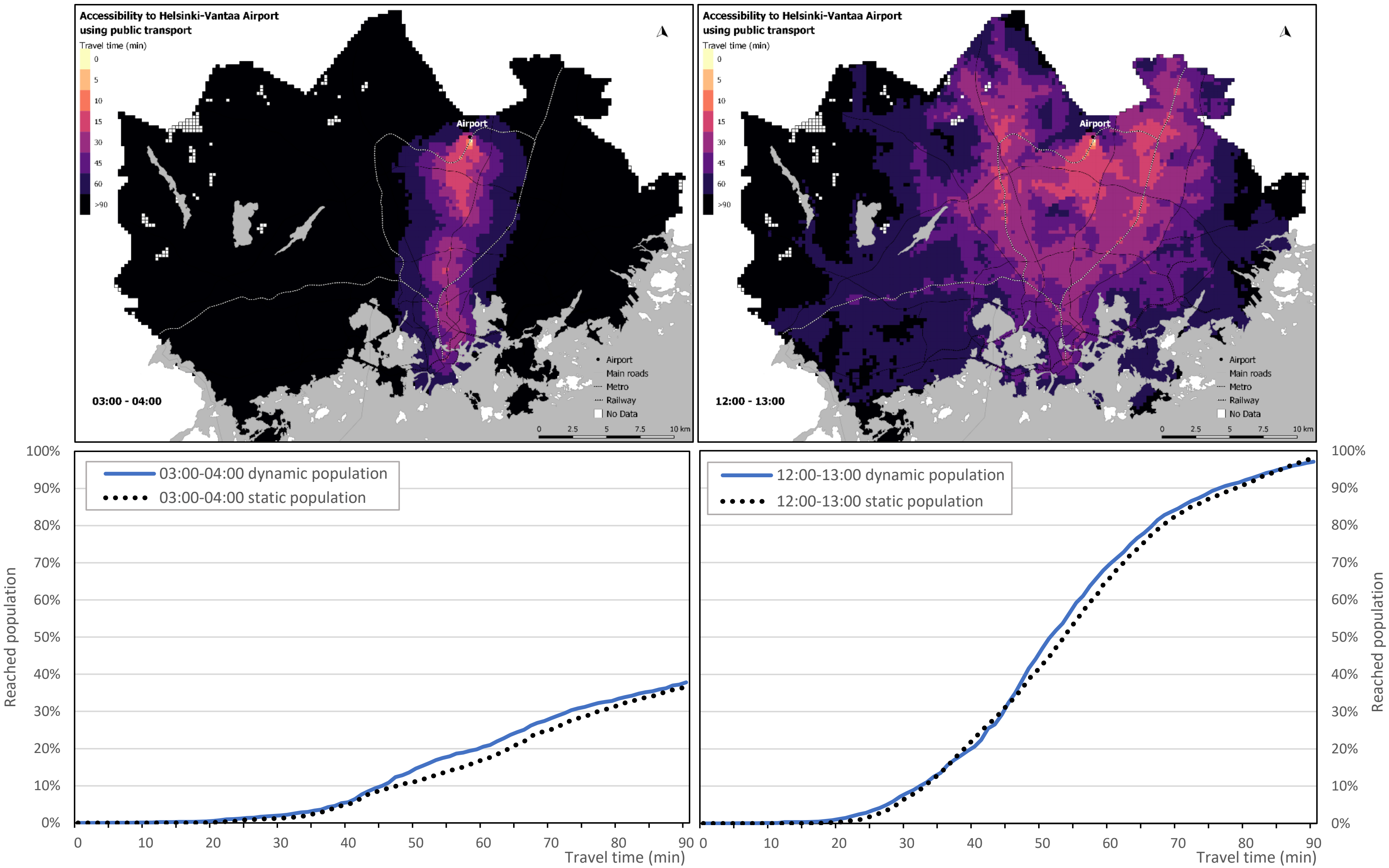


Figure 48. Accessibility to Helsinki-Vantaa airport using public transport between a) 03:00-04:00 and b) 12:00-13:00. The plots show the cumulative sum of population in the HMA reaching the airport by public transport between a) 03:00-04:00 and b) 12:00-13:00.

6.3.2. Dynamic accessibility to grocery stores

On the level of the **whole study area**, there is only little difference in the cumulative amount of people that can reach the closest grocery store between static and dynamic population data. Regardless of the time of day, the difference in the share of reached population is less than 10 % between the two datasets, although the static population data tends to overestimate the accessibility to the closest store on time distances above 10 minutes during daytime (Figure 49). The network of open grocery stores is dense and widely distributed across the study area both during the day and night with 32 grocery stores are open round the clock. Hence, the distance to the nearest grocery store is rather constant in the HMA, although the population distribution fluctuates in space. The difference in accessibility between hours is mainly caused by the variation in PT supply and opening hours of stores.

Overall, the accessibility to grocery stores is good during the day. From 7 AM to 10 PM, most parts of the study area belong to the 10-minute accessibility zone (Appendix 9). The accessibility starts to gradually decrease after 9 PM, when number of open stores and PT service level decreases. Between 5 PM – 6 PM, which according to the time use survey is a common hour for shopping activities, almost the whole population (97 %) can reach the nearest grocery store based on dynamic population data in 15 minutes. The corresponding share of reaching people between 10 PM – 11 PM is still very high (93 %), although over 50 % of the stores are closed. The accessibility to the closest grocery store is poorest between 3 AM – 4 AM, when the public transport network service level is at its lowest. Thus, the areas with best accessibility are generally within a walkable distance from open stores. Regardless, approximately 26 % of the population in HMA can access the closet grocery store within 15 minutes, based on the dynamic and 23 % based on static population data.

However, when analyzing accessibility from the perspective of an **individual grocery store**, the difference between the dynamic and static population is more significant. For example in the case of Alepa Länsimäki, the static population overestimates the amount of people that can reach the store in 15 minutes between 5 PM – 6 PM by PT by almost 10 000 people, if assumed that the population in the study area is constant and equals to the number of inhabitants (1 154 967 on 31.12.2017) (Figure 50). Approximately 37 000 people can reach the store in 15 minutes based on the static population data, whereas the corresponding amount based on the dynamic population data is only 28 000 people. The selected grocery store is located in an area with primarily residential functions, where the share of HMA total population decreases during daytime from that of night-time.

The opposite is the case of K-market Pasaati grocery store in eastern Pasila, where the static population data underestimates the number of people that can reach the store in 15 minutes between 5 PM – 6 PM. According to the dynamic population data, approximately 59 000 people can reach the grocery store in 15 minutes using PT (Figure 51). The static population data underestimates the accessibility of the store by approximately 9000 people compared to the dynamic population data. Opposite of Alepa Länsimäki, K-market Pasaati is located near areas, where the daytime population increases significantly compared to the night-time, but where the supply of grocery stores is limited. Similar examples are likely to exist also elsewhere in the study area – stores, where the static population underestimates the dynamic population during daytime are likely to be found in neighbourhoods, where the share of population during night-time is low compared to day-time and vice versa.

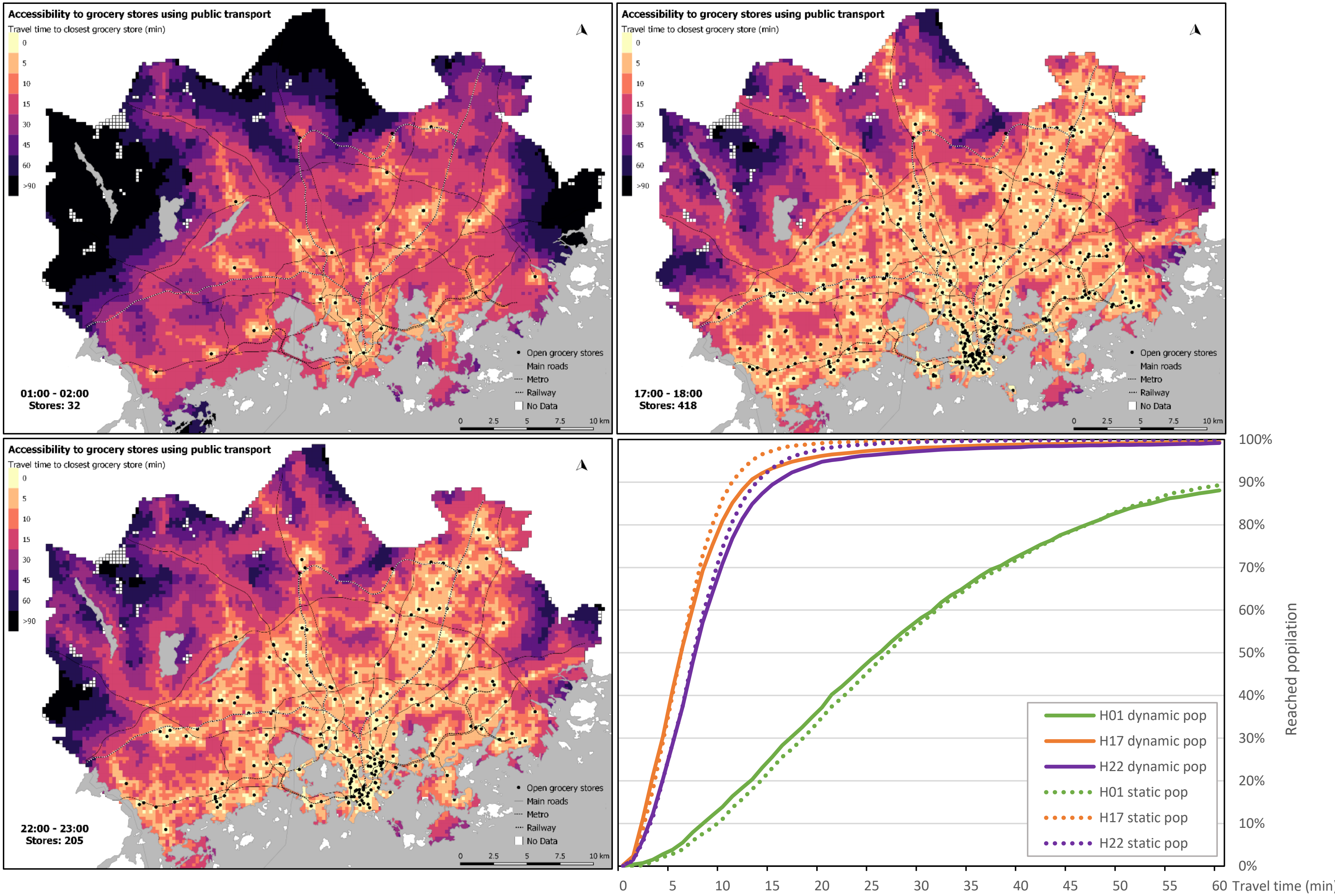
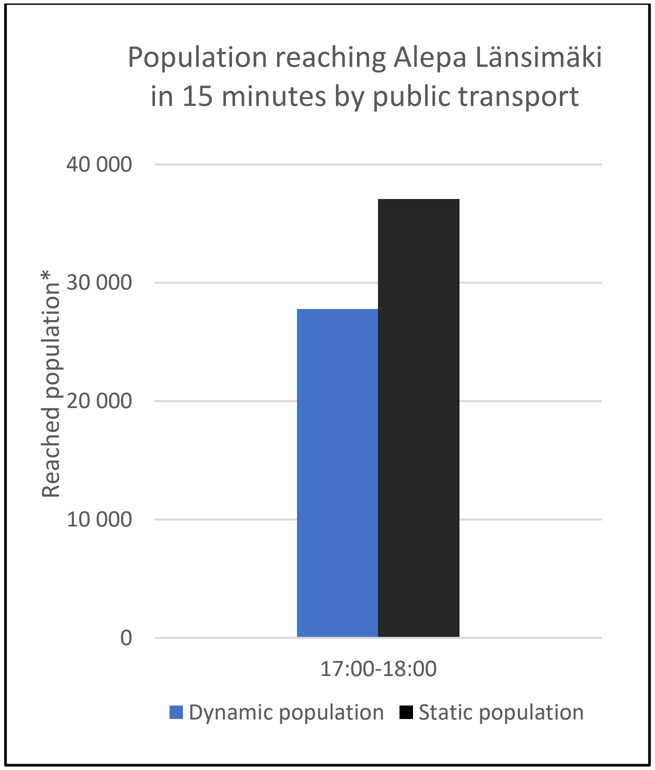


Figure 49. Accessibility to the closest grocery store (n=418) using public transport between a) 01:00–02:00, b) 17:00–18:00, c) 22:00–23:00. The final plot shows the cumulative sum of population reaching the closest grocery store by public transport.



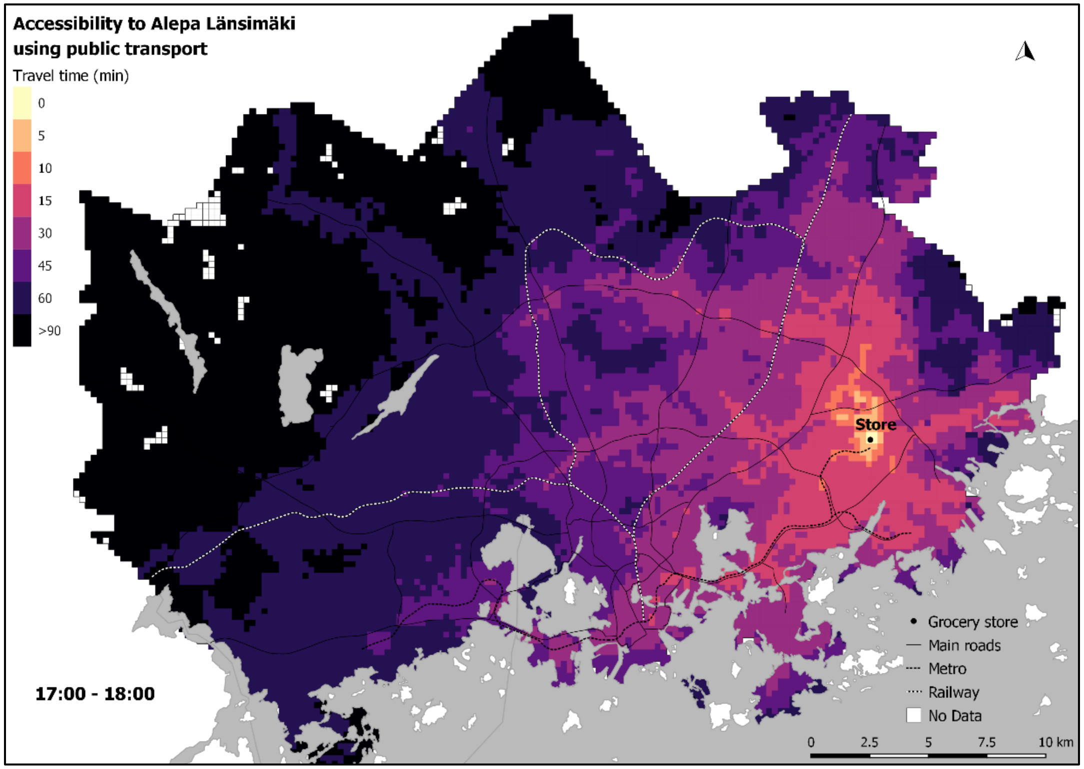


Figure 50. Population accessing Alepa Länsimäki in 15 minutes by public transport between 5 PM–6 PM. The number of reached population (*) is proportional to the number of inhabitants in the HMA (1 154 967 on 31.12.2017). The static population data overestimates the accessibility by approximately 9000 people.

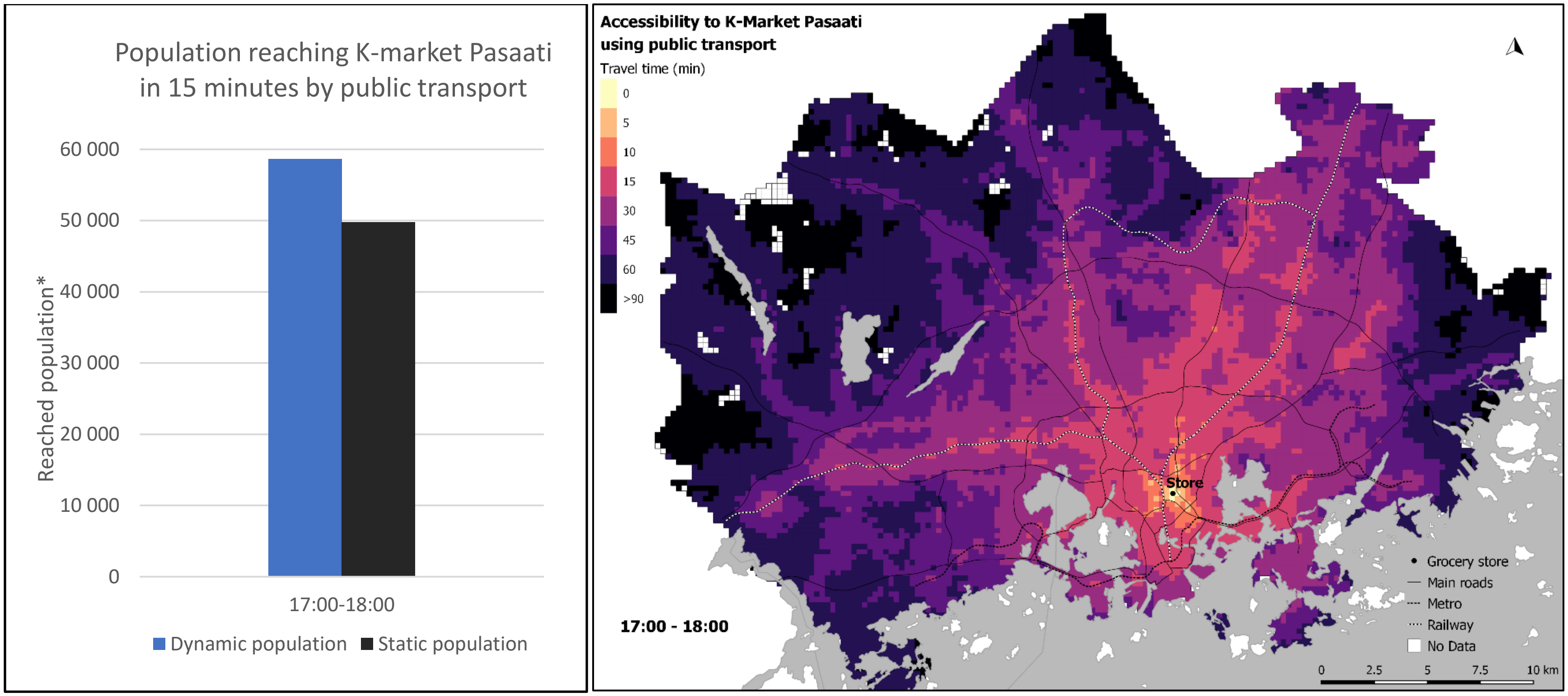


Figure 51. Population accessing K-Market Pasaati in 15 minutes by public transport between 5 PM–6 PM. The number of reached population (*) is proportional to the number of inhabitants in the HMA (1 154 967 on 31.12.2017). The static population data underestimates the accessibility by approximately 9000 people.

7. DISCUSSION AND CONCLUSIONS

7.1. Mobile phone data reveals the pulse of the city

This study shows that mobile phone data can be used to reveal diurnal spatiotemporal patterns of the population distribution in the HMA. The results demonstrate how the population distribution varies within the study area over the course of the day on different spatial levels and highlights the hourly hotspots of human presence. By this, the findings of this study expand previous knowledge of the dynamic population distribution in the study area. Prior understanding of the temporal fluctuation of human presence within the study area was limited to the study by Ahola et al. (2007), which covered only the inner city of Helsinki. Although the general pulse pattern of the inner city of Helsinki can be recognized from both studies, the time-sensitive population data and interpolation method used in this study are more sophisticated and extensive, while the spatial extent is also broader.

Overall, the empirical findings support and are in line with previous research on diurnal patterns of human presence in urban areas. For example, the share of present population in residential areas tends to be lower during daytime compared to night-time, while areas with work functions typically have the opposite pattern (e.g. Froehlich et al., 2008; Soto & Frías-Martínez, 2011; Ma et al., 2017). However, conclusions of the extent the land cover or built environment explain the temporal patterns should be made with caution, since the likelihood of human presence on different land cover areas or building types in time is inbuilt in the dasymetric interpolation method used in this study. Nevertheless, the interpolation only impacts the distribution of population within each Voronoi polygon, leading to rather accurate results in the inner city and elsewhere, where the density of base stations is high compared to the target zones.

This study focused on analysing the dynamic hourly population distribution only on weekdays. Previous research suggests that population distribution patterns vary between weekdays and weekend days, which are generally considered as less bounded (Vilhelmson, 1999; Soto & Frías-Martínez, 2011). This can also be detected on a temporal level from the local time use survey. Thus, further research is needed for understanding the population dynamics during weekends in the HMA and to analyse how the weekday and weekend patterns potentially differ from each other. In addition to weekends, also the patterns during special events and seasonal variation would provide valuable information for planning and decision making, although the latter would require longitudinal time-sensitive population data. The methods presented in this study provide a good foundation and directions for further research of dynamic population distribution also on these different temporal scopes.

A second significant topic for further research raised by this study is the magnitude and influence of the visiting population, such as commuters or tourists, on the diurnal population dynamics of the HMA. The aggregated data used in this study could not be used to infer the absolute number of population within the study area per hour or to analyse the origin-destination flows of mobile phone users, which would be valuable insights for transportation planning and for local tourism. Such information could shed light onto questions, such as *How large a proportion of the daily population is coming from outside the study area?*, *Where these people are coming from?*, *What is the regularity of visits?* and *How do the visitors use space in time?* According to previous research, suitable data sources to answer these questions are for example non-aggregated CDR data (Ahas et al., 2008) or travel surveys (Zandvliet & Dijst, 2005), however the latter is prone to smaller sample sizes and burdensome to collect. Furthermore, analysing the hourly number of population within the study area would also enrich the results of this study, since the relative shares per grid cell could be transformed to the hourly number of people within grid a cell.

Finally, cellular mobile phone data is only one of many time-sensitive population data sources. A key limitation of passive cellular mobile phone data is its lack of socio-demographic factors of users, which limits the research applications of this data. Other novel data of human whereabouts, such as geotagged social media content or creative fusion of multiple data sources could mitigate this gap and provide answers to *who* are present in a given location at a given time, *what* activities they are undertaking and *why* (Toivonen et al., 2019). This information could shed light onto the population distribution and use of space of different population groups, which could aid for example social segregation research. Furthermore, using other data sources to study the dynamic population in the study area provides the possibility to compare if different temporally-sensitive population data sources provide different insights into the spatiotemporal patterns of the population distribution and how these differ. This would not only bring interesting empirical findings from the perspective of the HMA but also input for the scientific discussion of novel population data sources.

7.2. Data transmission of mobile phones is a good proxy for people

This study assessed the feasibility of using aggregated network-driven mobile phone data as a proxy for human presence. The empirical results show that network-driven data aligns well with the population register data in HMA during night-time. During daytime, the mobile phone data highlights how the population register data underestimates population in areas with workplace functions and overestimates population in residential areas. Out of the three types of network-driven mobile phone data, the hourly number of data transmission attempts is the best proxy for people's whereabouts based on all statistical validation methods outperforming the overall network connection attempts and voice calls.

Interestingly, the correlation of the data transmission data against population register data was even higher during the late evening (10 PM – 11 PM) than during night-time (2 AM – 5 AM), further highlighting the fitness of this data for population mapping. A possible explanation may be that people use their phones more actively during the evening compared to the night, while still being at home. Also, the share of present population on municipal level corresponded best with the number of inhabitants during the late evening. This finding provides interesting input to the scientific discussion on the validation of novel population data sources, since night-time is commonly used as the time window for validation (e.g. Järv, Tenkanen & Toivonen, 2017a).

Somewhat surprisingly, the use of data transmission records of mobile phones to study population distribution has so far been scarce in literature, despite the increasing rates of data transmission and limitations of voice call and SMS frequencies (e.g. Ricciato et al., 2017). Studies limited to voice calls and SMS, for example fail to include widely adopted instant messaging applications, which rely on an Internet connection. Similarly, only few studies have compared the feasibility of different mobile phone data sources as a proxy for people (Pinelli et al., 2015). To this end, this study adds to prior research of mobile phone data for population mapping, since to the author's knowledge, no prior literature exists where the use of voice calls, data transmission or overall network connection attempts would have been compared for studying human presence. Comparing the feasibility of different mobile phone data types is however faced with challenges, as data between different studies may differ. In conclusion, to reliably assess the suitability and limitations of different mobile phone data sources to estimate human presence and mobility, it is essential to provide a transparent and adequate description of used data and aim towards harmonized terminology.

Despite the fitness of the mobile transmission data in this study, the results are not automatically scalable to other study settings. The suitability of different mobile data sources can vary between geographical regions and population groups due to socio-economic or cultural factors. For example data transmission costs, which are relatively low in Finland, may be high in some countries, which may lead to voice calls or SMS being the preferred more of communication, an issue also raised by Deville et al. (2014). Political factors are also relevant in this regard. Within Europe, the prices of data use have declined both on domestic level and among roaming use since the adoption of new EU legislation regarding roaming charges in 2017, which was further followed by an increase in data use (European Commission, 2018). Especially in the case of data use, the penetration of smartphones and how equally they are distributed in society are also likely to affect the reliability of the results. All in all, generalization of the results of this study require an understanding of the study context where they are applied or adapted to.

Since the data was provided in aggregated format, potential biases caused by individual users, such as very active individual users or telemarketing companies that are likely to produce a load higher than average, could not be detected or filtered out from the aggregated data. Another source of uncertainty of the mobile phone data and the resulting dynamic population data is the randomized error (±100/200 m) applied to the base station coordinates by the mobile network operator. This processing may influence how the mobile phone datasets perform against reference data. In other words, undistorted mobile phone data may provide even better results in the validation. A further limitation inherently related to mobile phone data is that the active use of mobile phones is likely to overestimate the presence of the non-sleeping population during night-time and in central locations, where communication rates are higher. This issue was also noticed by Järv et al. (2017a) in Tallinn when using interpolated CDR data as a proxy to estimate dynamic population. Furthermore, human presence may be underestimated in locations, where mobile phone data use is limited or restricted, such as schools or hospitals (Järv et al., 2014). In comparison to voice calls, data transmission data is, however, likely to be less sensitive to the potential active use and passive use bias, since mobile phones can transmit data also when the phone is not actively used, for instance due to email synchronization (Janecek et al., 2015).

The spatial accuracy of the results could be further improved with antenna-level data, which was not available in this study. In addition, inclusion of 4G data or data by multiple operators would improve the completeness of the mobile data transmission records and reduce the uncertainty related to sampling issues due to varying behaviour of users of different operators or subscription types. This study also highlights the challenge of validation and finding the ground truth in the context of time-sensitive population data, a notion that has also been raised in previous literature (e.g. Calabrese et al., 2014; Chen et al., 2014). Finding the suitable level of correlation between residential population data and mobile phone data during evening or night-time is for example not straightforward since it is difficult to assess which data is more reliable. Further research and scientific discussion are also needed to assess the impact of Wi-Fi use on the completeness of the data transmission records.

Overall, the findings suggest that mobile transmission data should be considered when selecting a data source for estimating population distribution, although simultaneously considering the special characteristics of this data and their impacts on the results.

7.3. Refining the spatial resolution provides more realistic population estimates, but is data-hungry

The results of this study show that refining the spatial resolution with ancillary data improves the estimation of the dynamic population distribution. According to the results, the more advanced MFD method outperforms the areal weighting method in the case of all three tested mobile phone datasets. During night-time, the MFD allocates most of the population to residential areas, whereas the AW method distributes the population relatively evenly to different activity function types, since it does not account for the hourly likelihood of human presence in different locations. The difference between the performance of the methods is smaller during daytime, but even then, the AW method allocates population to restricted areas unlike the MDF method and results of the MFD method coincide better with the time use survey. These findings support outcomes of previous research conducted in Tallinn by Järv et al. (2017a), where the MFD method also outperformed the simpler AW method. Furthermore, as noted by Batista e Silva et al. (2013), dasymetric interpolation, and thus the MDF method, is free of the possible distortions caused by the modifiable areal unit problem (MAUP), which is a source of statistical bias that can occur when aggregating data to (arbitrary) larger spatial units and lead to false interpretations due to selected scale or target zone geometry (Openshaw, 1984).

The share of population allocated to service and retail subunits during the afternoon was relatively low when compared to the time use survey even when using the MFD method. This is likely due to the small amount of buildings that were classified as service or retail, since reliable data was difficult to find for the model. Point of interests (POI) derived from OpenStreetMap (OSM) were initially used to aid this limitation similarly as in Järv et al. (2017a), but these were discarded from the final model, as the POIs turned out to skew the results towards certain activity function types and street level services. Several shopping centres or office buildings were for example misclassified as ‘other’ due to overrepresented eatery POIs. This limitation could be overcome and the MFD method could be further improved by determining a secondary activity function type for buildings, since buildings especially in cities may have multiple use types either simultaneously or at different times of the day. Multiple activity function types could still be derived for example from points of interest in OSM. However, to diminish the bias towards street level services, the activity function type could be assigned separately for the ground floor and other floors, if floor count data is available. Other means to improve the results and the MFD method include introduction of new data sources, such as traffic data used for example in Ahola et al. (2007), however the level of detail of the time use data largely defines the limits of the multi-temporal interpolation. As pointed out by Järv et al. (2017a), a way to improve the accuracy of the MFD method would be to use a more sophisticated approach for estimating the coverage areas of base stations. Current approaches, however, often require additional information of the network or data of actual footprints of base station antennae (e.g. Ricciato et al., 2015), which may not be available for researchers.

Indeed, a general challenge related to the MFD method is data availability. Although building footprints, use type and land cover can be at least partly derived from OSM, the scalability of the method relies strongly on the availability of the time use data, even if time-sensitive population data would be available. In this study, statistics based on a harmonized European-level time use survey were used, but similar data may not be available outside Europe. Also building volume data can be difficult to acquire, if such information is not registered or shared by authorities. In this study, building volume data was based on building register data, but also aerial imagery, LiDAR data or readily available normalized digital surface models (nDSM) could be used to derive building heights to estimate floor areas (see Alahmadi et al., 2013; Järv, Tenkanen & Toivonen, 2017a). Recently, 3D city models have emerged in several cities, which could also be used to derive building volumes. A drawback of using remotely sensed data is that the floor areas are always estimates unlike in register-based data. The temporal resolution of the data is also of importance when selecting the data source, since outdated data can lead to underestimation of population in newly built-up areas.

Due to the challenges related to data availability, the AW method may be a sufficient alternative if ancillary data is not available. This applies particularly in areas where the base station network is dense, as the role of the ancillary data sources in the redistribution of mobile phone data is smallest in these areas. However, understanding the uncertainties and limitations of the interpolation is important if a dynamic population data is to be constructed using the simplified approach.

7.4. Dynamic population provides more realistic results in accessibility modelling

Stemming from the recent literature on dynamic accessibility (Moya-Gómez et al., 2017; Järv et al., 2018), this study employed a multi-temporal accessibility model to analyse the role of incorporating temporally-sensitive population data to location-based accessibility modelling. The empirical results show, that using dynamic population data as the origins of people provides more realistic results compared to static data of home locations, which have predominantly been used in place-based accessibility research as the proxy for people's whereabouts. The significance of including temporally-sensitive population data in accessibility modelling does, however, depends on the study question and setting.

On a general level, the findings of the study show that static population may both over and underestimate the population in accessibility analysis, validating the critique of previous research (e.g. Kwan, 2013). However, the results are sensitive to the selected activity location, time of day and spatial level. The sensitivity of location-based accessibility analysis to selected travel-time thresholds and spatial scale was also noted by Pereira (2019). When analyzing the accessibility on study area level to the closest grocery store, the difference between the reached population using static and dynamic population were only small and the hourly differences were mostly caused by diurnal variation in the public transport service levels and the opening hours of services. Similar results were also found in Tallinn in the previous implementation of the dynamic accessibility model (Järv et al., 2018), which suggests that the dense store network is always at a close distance, regardless of differences in population distribution, which is further supported by the extensive public transportation network.

When analyzing accessibility on the level of individual stores or to the Helsinki city centre, however, the difference between static and dynamic population is significant. Both grocery stores selected for individual analysis in this study represent small local stores with a limited supply of products, whose primary customer base likely relies on the short travel time to the store. Consequently, an over or underestimation of accessibility by static population data can be of high significance from the perspective of a store, although the differences on study area level suggest that the differences between static and dynamic population are only small. In the case of Alepa Länsimäki, for example, the static population data overestimated the number of people that can reach the store in 15 minutes by 33 % corresponding to approximately 9000 people compared to the dynamic population data.

These findings also raise critical points for discussion and further research. Firstly, the number of reached population is only a rough estimate based on the number of inhabitants. In fact, the amount of present population within the HMA is likely to increase during the day due to commuters and other visiting population (Statistics Finland, 2018b), which suggests that the over- or underestimated number of reached population by static data may be even higher. Secondly, although an individual grocery store or transport hub is reachable in 15 minutes by public transport, it is not necessarily the closest option for the reached population. Research suggests that accessibility does not solely determine individuals' store choice, but for example individual factors and constraints, the selection of products or price range have an impact of the choice of store (Briesch et al., 2009). Thus, further research on how accessibility meets the actual mobility of people is needed for gaining a deeper understanding of the impacts of temporal population data. Nonetheless, a more realistic view of the potential is already valuable information in planning equitable services and selecting optimal store locations, where dynamic population clearly provides a more realistic results than static residential population.

On broader level, incorporating dynamic population data to location-based accessibility modelling has potential to bridge the gap between location-based and person-based accessibility. Although temporally-sound, person-based approaches are generally considered more complex to implement and interpret than location-based models and often limited to small sample sizes (e.g. Geurs & van Wee, 2004). Therefore, as also stated by Li et al. (2011), harnessing location-based accessibility approaches with temporally-sensitive data might serve as a trade-off between the strengths and limitations of the two approaches. Certainly, both location-based and person-based accessibility approaches are needed for solving different research problems. To account for the constraints of individuals for example, person-based approaches are still the preferred option, while large-scale planning and decision making may be better supported by location-based approaches.

All in all, the findings of this study suggest that temporally sensitive population, similarly as other components of accessibility, should be integrated to accessibility modelling whenever possible, and when not, the potential impacts of static population data should be considered if making conclusions based on such data.

7.5. Accessibility to grocery stores and transport hubs is generally good in Helsinki Metropolitan Area, but travel times grow during night-time

Accessibility to grocery stores and transport hubs is generally good in HMA, as the findings of this study show. During the day and evening, over 50 % of the population present in the HMA can reach the closest grocery store by public transport or walking within less than 10 minutes and 90 % in 15 minutes. Even between 1 AM – 2 AM, only the sparsely populated northwestern parts of Vantaa and Espoo and Östersundom in the eastern part of the study area appear as areas with poor accessibility (> 60 min with PT) to the closest grocery store at night. Furthermore, as grocery stores or public transport connections outside the study area were not considered, the accessibility in these and other areas on the fringe of the study areas may be underestimated. Overall, significant food deserts (see Farber et al., 2014; Widener et al., 2015; Tenkanen et al., 2016) are not identified in the HMA, largely thanks to the distributed network of grocery stores open 24 hours of the day and extensive public transport connections to most inhabited parts of the study area.

When comparing the results to previous studies on the grocery store accessibility in the HMA by Tenkanen et al. (2016) and Saarsalmi (2014), the findings suggest that accessibility to grocery stores using public transport has increased after the liberation of grocery store opening hours in 2016. The difference is largest during the late evening and night-time outside 7 AM – 9 PM, when only stores smaller than 400 m^2 could stay open during regular weekdays without special permits prior to the amendment to legislation. Currently, approximately 89 % of the population in the HMA can reach the closest grocery store in 15 minutes by public transport between 10 PM – 11 PM on an average weekday, while in 2015, the corresponding share was only 73 % (see Tenkanen et al., 2016). The share of population reaching the closest grocery store has also increased during the other two time slots used in Tenkanen et al. (2016), 1 AM (15 % to 26 %) and 5 PM (90 % to 93 %) whether using the dynamic or static population as a measure of reached population.

The findings of this study are in line with previous research claiming that temporality matters when analysing accessibility (Neutens et al., 2012; Saarsalmi, 2014; Widener et al., 2015; Tenkanen et al., 2016; Moya-Gómez et al., 2017; Järv et al., 2018). Although the accessibility to the selected activity locations is generally good during the day, the accessibility is significantly poorer in all cases at night, when public transport connections are scarce. The study by Geiger (2007) found that grocery shoppers during night-time are mainly night-workers. This highlights that the development towards 24-hour-driven societies should be considered when planning equitable services.

Far-reaching conclusions between the results of the grocery store accessibility in this analysis and by previous studies in the HMA should however be made with caution. This study only included the grocery stores of the three largest grocery store chains, while the previous studies also included some small stores. Many of these stores have closed since the liberation of opening hours when the advantage of smaller stores to stay open ceased or are located within the same statistical grid cell or in the direct vicinity of other stores that are included in this analysis. Thus, the differences in the datasets are likely to have only a small or local impact on the results. Yet, the relative influence of the public transport network service level and grocery store locations and opening hours between the results is difficult to differentiate.

The findings also raise topics for further research. Firstly, since accessibility results depend on selected mode of transport (e.g. Tenkanen et al., 2016), employing a multimodal approach that considers also the dynamic population is needed to uncover potential differences in accessibility realities between different modes of transport and to assess the competitiveness of these modes. Multimodality enables inclusion of different preferences and opportunities for mobility of individuals and can provide more robust insights for equitable and sustainable transport and service planning. Another topic that needs further research is analysis of the accessibility of different population groups, as the accessibility realities may not be equitable across all population groups. However, acquiring temporally sensitive data of the whereabouts of these population groups may be difficult.

7.6. Understanding how our societies function requires temporally-sensitive population data

Planning socially and environmentally sustainable and healthy cities call for reliable data and tools to comprehend the complex processes shaping our environment. This need is further stressed by the transition towards 24-hour societies, which is strongly interwoven with the spatiotemporal dimension of human behaviour. Consequently, accurate data of the population distribution is essential for unravelling how our societies function. So far, the whereabouts of people have been predominantly treated as static. Thus, decisions based on static population may lead to erroneous conclusions. This was also demonstrated in this study, where the static population was shown to over- or underestimate the population dynamics. Overall, the findings of this study underline that temporally-sensitive population data provide a more realistic picture of the population dynamics. The dynamic population datasets created in this study and the analyses based on the diurnal variation of people provide new information of the population dynamics in the Helsinki Metropolitan Area. As reliable data of population distribution is relevant across disciplines, the results of this thesis can be beneficial for multiple application fields. More specifically, the results can be used to aid decision making and planning of for example more equitable public services and on-demand transportation in the study area or to improve existing risk models for preparedness planning building upon the previous work of Ahola et al. (2007).

The need for reliable and cost-effective approaches to estimate population is not limited to dynamic population mapping. In fact, even the knowledge of the static population distribution in many areas of the world is still poor (Tatem & Linard, 2011; Wardrop et al., 2018). For instance, in developing countries, up-to-date population census data may not be available and collecting data by traditional means through surveys might be too expensive and laboursome (Deville et al., 2014).

Although the need for understanding the spatiotemporal variation of population distribution is evident, the acquisition of temporally sensitive population data remains a challenge for researchers. This leads to a critical question of who has access to such information in contemporary societies. Currently, mobile phone data and other fine-grained spatiotemporal data of human presence are primarily owned and used by private companies for business purposes. This sets a conspicuous contrast against the surge of big data, which is a notion that has also been raised in previous research (e.g. Boyd & Crawford, 2012; Kitchin, 2013). As stated by Boyd and Crawford (2012), "limited access to Big Data creates new digital divides" – those who have access to data and those who are data poor. Consequently, to respond to the needs cities all over the world are facing related to sustainable planning, it is crucial that researchers can access relevant data.

Current literature suggests that data availability is also directing research. For example Willberg (2019) found that most studies in the domain of bike-sharing are conducted in cities, where data is openly available. Thus, by providing data either openly or with reasonable expenses for researchers and public decision making, cities are likely to attract scholars and gain new information of its functioning. Obviously, the ethical considerations should always be considered when sharing that can potentially be connected to individuals or groups. Concerns of privacy are focused particularly on the sharing of individual-level data. As this study and a recent study by Deville et al. (2014) have shown, already aggregated mobile phone data can provide useful insights of population dynamics. Therefore, sharing aggregated data is one potential direction for improvement. Yet, as aggregated data can answer only a limited set of questions important for societal needs, also individual-level data should be brought available to research. At its best, increased sharing of temporally-sensitive population data can lead to new scientific advances and improved decision making, both within the domain of population dynamics and beyond.

8. ACKNOWLEDGEMENTS

I want to thank Elisa for providing access to the mobile phone data used in this study. I am also thankful for the Helsinki Metropolitan Region Urban Research Program (KATUMETRO) and University of Helsinki for funding my thesis work and the latter for supporting my work by assigning me the Geography Master’s programme travel grant.

I am deeply grateful for my supervisors Tuuli Toivonen, Olle Järv and Henrikki Tenkanen for inspiring me and always having time for me. I have learned so much from you! A big thank you goes also to my family, friends and everyone at Digital Geography Lab for your help and good times that kept me going even when things did not go according to plan. Finally, I want to thank you, Elias, for coping with me despite too many long evenings spent with this thesis and for your unconditional support throughout the whole process.

9. REFERENCES

Adams, P. M., Ashwell, G. W. B. & Baxter, R. (2003) Location-Based Services - An Overview of the Standards, *BT Technology Journal*, 21(1), pp. 34–43. doi: 10.1023/A:1022572210026.

Ahas, R., Aasa, A., Roose, A., Mark, Ü. & Silm, S. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study, *Tourism Management*. Pergamon, 29(3), pp. 469–486. doi: 10.1016/J.TOURMAN.2007.05.014.

Ahas, R., Aasa, A., Silm, S. & Tiru, M. (2010) Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data, *Transportation Research Part C: Emerging Technologies*. Pergamon, 18(1), pp. 45–54. doi: 10.1016/J.TRC.2009.04.011.

Ahas, R., Silm, S., Järv, O., Saluveer, E. & Tiru, M. (2010) Using mobile positioning data to model locations meaningful to users of mobile phones, *Journal of Urban Technology*, 17(1), pp. 3–27. doi: 10.1080/10630731003597306.

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M. & Zook, M. (2015) Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn, *International Journal of Geographical Information Science*. Taylor & Francis, 29(11), pp. 2017–2039. doi: 10.1080/13658816.2015.1063151.

Ahas, R. & Mark, Ü. (2005) Location based services - New challenges for planning and public administration?, *Futures*, 37(6), pp. 547–561. doi: 10.1016/j.futures.2004.10.012.

Ahola, T., Virrantaus, K., Krisp, J. M. & Hunter, G. J. (2007) A spatio-temporal population model to support risk assessment and damage analysis for decision-making, *International Journal of Geographical Information Science*, 21(8), pp. 935–953. doi: 10.1080/13658810701349078.

Alahmadi, M., Atkinson, P. & Martin, D. (2013) Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data, *Computers, Environment and Urban Systems*. Pergamon, 41, pp. 167–176. doi: 10.1016/J.COMPENVURBSYS.2013.06.002.

Andresen, M. A. (2011) The Ambient Population and Crime Analysis, *The Professional Geographer*, 63(2), pp. 193–212. doi: 10.1080/00330124.2010.547151.

Ashbrook, D. & Starner, T. (2003) Using GPS to learn significant locations and predict movement across multiple users, *Personal and Ubiquitous Computing*, 7(5), pp. 275–286. doi: 10.1007/s00779-003-0240-0.

Aubrecht, C., Özceylan, D., Steinnocher, K. & Freire, S. (2013) Multi-level geospatial modeling of human exposure patterns and vulnerability indicators, *Natural Hazards*, 68(1), pp. 147–163. doi: 10.1007/s11069-012-0389-9.

Bagchi, M. & White, P. R. (2005) The potential of public transport smart card data, *Transport Policy*. Pergamon, 12(5), pp. 464–474. doi: 10.1016/J.TRANPOL.2005.06.008.

Bagrow, J. P., Wang, D. & Barabási, A.-L. (2011) Collective Response of Human Populations to Large-Scale Emergencies, *PLOS ONE*. Public Library of Science, 6(3), p. e17680. Available at: https://doi.org/10.1371/journal.pone.0017680.

Balk, D. L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I. & Nelson, A. (2006) Determining global population distribution: methods, applications and data, *Advances in parasitology*, 62, pp. 119–156. doi: 10.1016/S0065-308X(05)62004-0.

Banister, D. (2008) The sustainable mobility paradigm, *Transport Policy*, 15(2), pp. 73–80. doi:

10.1016/j.tranpol.2007.10.005.

Banister, D., Watson, S. & Wood, C. (1997) Sustainable cities: Transport, energy, and urban form, *Environment and Planning B: Planning and Design*, 24(1), pp. 125–143. doi: 10.1068/b240125.

Batista e Silva, F., Marín Herrera, M. A., Rosina, K., Ribeiro Barranco, R., Freire, S. & Schiavina, M. (2018) Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources, *Tourism Management*. Pergamon, 68, pp. 101–115. doi: 10.1016/J.TOURMAN.2018.02.020.

Batista e Silva, F., Gallego, J. & Lavalle, C. (2013) A high-resolution population grid map for Europe, *Journal of Maps*, 9(1), pp. 16–28. doi: 10.1080/17445647.2013.764830.

Batty, M. (1971) Modelling Cities as Dynamic Systems, *Nature*. Nature Publishing Group, 231, p. 425. Available at: https://doi.org/10.1038/231425a0.

Batty, M. (2010) The pulse of the city, *Environment and Planning B: Planning and Design*, 37(4), pp. 575–577. doi: 10.1068/b3704ed.

Batty, M. (2012) Building a science of cities, *Cities*. Pergamon, 29, pp. S9–S16. doi: 10.1016/J.CITIES.2011.11.008.

Batty, M. & Marshall, S. (2009) Centenary paper: *The evolution of cities: Geddes, Abercrombie and the new physicalism*, *Town Planning Review*, 80(6), pp. 551–574. doi: 10.3828/tpr.2009.12.

Bayir, M. A., Demirbas, M. & Eagle, N. (2009) Discovering spatiotemporal mobility profiles of cellphone users, *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops 2009*, pp. 1–9.

Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A. & Volinsky, C. (2011) A tale of one city: Using cellular network data for urban planning, *IEEE Pervasive Computing*. IEEE, 10(4), pp. 18–26.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & von Schreeb, J. (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti, *PLoS Medicine*, 8(8), pp. 1–9. doi: 10.1371/journal.pmed.1001083.

Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K., Rebaudet, S. & Piarroux, R. (2015) Using Mobile Phone Data to Predict the Spatial Spread of Cholera, *Scientific Reports*. The Author(s), 5, p. 8923. Available at: http://dx.doi.org/10.1038/srep08923.

Bertolini, L., le Clercq, F. & Kapoen, L. (2005) Sustainable accessibility: A conceptual framework to integrate transport and land use plan-making. Two test-applications in the Netherlands and a reflection on the way forward, *Transport Policy*, 12(3), pp. 207–220. doi: 10.1016/j.tranpol.2005.01.006.

Bertolini, L. & Dijst, M. (2003) Mobility environments and network cities, *Journal of urban design*. Taylor & Francis, 8(1), pp. 27–43.

Bhaduri, B., Bright, E., Coleman, P. & Urban, M. L. (2007) LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics, *GeoJournal*, 69(1–2), pp. 103–117. doi: 10.1007/s10708-007-9105-9.

Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q. & Weston, L. (2000) *Development of an urban accessibility index: Literature review*. University of Texas at Austin. Center for Transportation Research.

Bian, R. & Wilmot, C. G. (2015) Spatiotemporal Population Distribution Method for Emergency Evacuation, *Transportation Research Record: Journal of the Transportation Research Board*, 2532, pp.

99–106. doi: 10.3141/2532-12.

Blondel, V. D., Decuyper, A. & Krings, G. (2015) A survey of results on mobile phone datasets analysis, *EPJ Data Science*, 4(1), pp. 1–55. doi: 10.1140/epjds/s13688-015-0046-0.

Bloom, G. S. & Golomb, S. W. (1977) Applications of numbered undirected graphs, *Proceedings of the IEEE*. IEEE, 65(4), pp. 562–570.

Boeing, G. (2017) OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, *Computers, Environment and Urban Systems*. Pergamon, 65, pp. 126–139. doi: 10.1016/J.COMPENVURBSYS.2017.05.004.

Boisjoly, G. & El-Geneidy, A. (2016) Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures, *Journal of Transport Geography*. Pergamon, 52, pp. 73–81. doi: 10.1016/J.JTRANGEO.2016.03.004.

Boyd, D. & Crawford, K. (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, *Information Communication and Society*, 15(5), pp. 662–679. doi: 10.1080/1369118X.2012.678878.

Briesch, R. A., Chintagunta, P. K. & Fox, E. J. (2009) How does assortment affect grocery store choice?, *Journal of Marketing research*. SAGE Publications Sage CA: Los Angeles, CA, 46(2), pp. 176–189.

Brown, G. G. & Reed, P. (2009) Public participation GIS: a new method for use in national forest planning, *Forest science*. Oxford University Press, 55(2), pp. 166–182.

Brown, G. & Kyttä, M. (2014) Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research, *Applied Geography*. Elsevier, 46, pp. 122–136.

Burkhard, O., Ahas, R., Saluveer, E. & Weibel, R. (2017) Extracting regular mobility patterns from sparse CDR data without a priori assumptions, *Journal of Location Based Services*. Taylor & Francis, 9725(November), pp. 1–20. doi: 10.1080/17489725.2017.1333638.

Burns, L. D. (1979) *Transportation, temporal, and spatial components of accessibility*. p. 152. Lexington, Mass. : Lexington books 1979.

Caceres, N., Romero, L. M., Benitez, F. G. & del Castillo, J. M. (2012) Traffic flow estimation models using cellular phone data, *IEEE Transactions on Intelligent Transportation Systems*. IEEE, 13(3), pp. 1430–1441.

Caceres, N., Wideberg, J. P. & Benitez, F. G. (2007) Deriving origin destination data from a mobile phone network, *IET Intelligent Transport Systems*, 1(1), pp. 15–26. doi: 10.1049/iet-its:20060020.

Calabrese, F., Ferrari, L. & Blondel, V. D. (2014) Urban Sensing Using Mobile Phone Network Data: A Survey of Research, *ACM Computing Surveys*, 47(2), pp. 1–20. doi: 10.1145/2655691.

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G. & Barabási, A.-L. (2008) Uncovering individual and collective human dynamics from mobile phone records, *Journal of Physics A: Mathematical and Theoretical*, 41(22), p. 224015. doi: 10.1088/1751-8113/41/22/224015.

Carlstein, T., Parkes, D. & Thrift, N. (1978) *Timing space and spacing time. Human activity and time geography*. Edited by T. Carlstein, D. Parkes, and N. Thrift. London: Arnold.

Castells, M., Fernandez-Ardevol, M., Qiu, J. L. & Sey, A. (2009) *Mobile communication and society: A global perspective*. Mit Press.

Chen, C., Bian, L. & Ma, J. (2014) From traces to trajectories : How well can we guess activity locations from mobile phone traces ?, *Transportation Research Part C*. Elsevier Ltd, 46, pp. 326–337.

City of Helsinki (2015) *Helsinki by District 2014. City of Helsinki Urban Facts. Helsingin kaupunki, Tietokeskus.* Available at: https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/15_02_23_Hki_alueittain2014_verkko.pdf.

Cranshaw, J., Schwartz, R., Hong, J. & Sadeh, N. (2012) The livehoods project: Utilizing social media to understand the dynamics of a city, in *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, pp. 58–65.

Cullen, I. & Godson, V. (1975) Urban networks: The structure of activity patterns, *Progress in Planning*, 4(PART 1), pp. 1–96. doi: 10.1016/0305-9006(75)90006-9.

Curtis, C. & Scheurer, J. (2010) Planning for sustainable accessibility: Developing tools to aid discussion and decision-making, *Progress in Planning*, 74(2), pp. 53–106. doi: 10.1016/j.progress.2010.05.001.

Curtis, C. & Schreurer, J. (2015) Accessibility Instruments in Planning Practice. Spatial Network Analysis for Multi-Modal Transport Systems (SNAMUTS): Helsinki. Available at: https://blogs.helsinki.fi/saavutettavuus/files/2015/08/AccessibilitySummerSeminar_SNAMUTS_Helsinki Presentation_2015_10_07.pdf.

Delafontaine, M., Versichele, M., Neutens, T. & Van de Weghe, N. (2012) Analysing spatiotemporal sequences in Bluetooth tracking data, *Applied Geography*. Pergamon, 34, pp. 659–668. doi: 10.1016/J.APGEOG.2012.04.003.

Delafontaine, M., Neutens, T. & Van de Weghe, N. (2012) A GIS toolkit for measuring and mapping space–time accessibility from a place-based perspective, *International Journal of Geographical Information Science*. Taylor & Francis, 26(6), pp. 1131–1154.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D. & Tatem, A. J. (2014) Dynamic population mapping using mobile phone data, *Proceedings of the National Academy of Sciences*, 111(45), pp. 15888–15893. doi: 10.1073/pnas.1408439111.

Dijkstra, E. W. (1959) A note on two problems in connexion with graphs, *Numerische mathematik*. Springer, 1(1), pp. 269–271.

Dobler, G., Ghandehari, M., Koonin, S. E., Nazari, R., Patrinos, A., Sharma, M. S., Tafvizi, A., Vo, H. T. & Wurtele, J. S. (2015) Dynamics of the urban lightscape, *Information Systems*. Pergamon, 54, pp. 115–126. doi: 10.1016/J.IS.2015.06.002.

Dobson, J. E., Bright, E. A., Durfee, R. G. & Worley, B. A. (2000) LandScan: A global population database for estimating population at risk, *Photogrammetric Engineering and Remote Sensing*, 66(7), pp. 849–857. doi: 10.1016/j.scitotenv.2008.02.010.

Eagle, N. & Pentland, A. (2006) Reality mining: Sensing complex social systems, *Personal and Ubiquitous Computing*, 10(4), pp. 255–268. doi: 10.1007/s00779-005-0046-3.

Eagle, N., Pentland, A. S. & Lazer, D. (2009) Inferring friendship network structure by using mobile phone data, *Proc. National Academy of Sciences of the United States of America (PNAS)*, 106(36), pp. 15274–15278. doi: 10.1073/pnas.0900282106.

Eicher, C. L. & Brewer, C. A. (2001) Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation, *Cartography and Geographic Information Science*, 28(2), pp. 125–138. doi: 10.1559/152304001782173727.

European Commission (2018). *Report from the Commission to the European Parliament and the Council on the implementation of Regulation (EU) 531/2012 of the European Parliament and of the Council of 13 June 2012 on roaming on public mobile communications networks within the Union as amended by*

Regulation (EU) 2015/2120 and Regulation (EU) 2017/920. Available at: http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2018/0822/COM_COM(2018)0822_EN.pdf.

Eurostat (2009) *Harmonised European time use surveys: 2008 guidelines*. European Communities, Luxembourg. Available at: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-08-014-EN.pdf.

Eurostat (2017a) *Labour mobility. Statistics on commuting patterns at regional level*, *EU Labour Force Survey 2016*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_commuting_patterns_at_regional_level.

Eurostat (2017b) *Statistical concepts: Survey, census and register. Eurostat statistics explained, Statistics 4 beginners.* Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Beginners:Statistical_concept_-_Survey,_census_and_register.

Farber, S., Morang, M. Z. & Widener, M. J. (2014) Temporal variability in transit-based accessibility to supermarkets, *Applied Geography*. Elsevier Ltd, 53, pp. 149–159. doi: 10.1016/j.apgeog.2014.06.012.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996) From data mining to knowledge discovery in databases, *AI magazine*, 17(3), p. 37.

FICORA = Finnish Communications Regulatory Authority (2017) *Market shares of mobile subscriptions 06/2017*. Available at: https://www.viestintavirasto.fi/en/statisticsandreports/statistics/2013/marketsharesofmobilesubscriptions.html (Accessed: 18 November 2017).

FICORA = Finnish Communications Regulatory Authority (2018) *Statistics and comparisons - Internet and telephone*. Available at: https://www.viestintavirasto.fi/attachments/toimialatieto/FicoraSTATs.xlsx (Accessed: 8 April 2018).

Finger, F., Genolet, T., Mari, L., Magny, G. C. de, Manga, N. M., Rinaldo, A. & Bertuzzo, E. (2016) Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 113(23), pp. 6421–6426. doi: 10.1073/PNAS.1522305113.

Finnish Environment Institute (2012). *CORINE Land Cover 2012*. Retrieved 18.1.2018.

Foley, D. L. (1954) Urban Daytime Population: A Field for Demographic-Ecological Analysis*, *Social Forces*, 32(4), pp. 323–330. doi: 10.2307/2574113.

Freelon, D. (2018) Computational Research in the Post-API Age, *Political Communication*. Routledge, pp. 1–4. doi: 10.1080/10584609.2018.1477506.

Frias-Martinez, V., Soto, V., Hohwald, H. & Frias-Martinez, E. (2012) Characterizing urban landscapes using geolocated tweets, *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pp. 239–248. doi: 10.1109/SocialCom-PASSAT.2012.19.

Froehlich, J., Neumann, J. & Oliver, N. (2008) Measuring the pulse of the city through shared bicycle programs, *Proceedings of the International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems, UrbanSense08*, pp. 1–5. Available at: http://sensorlab.cs.dartmouth.edu/urbansensing/papers/urbansense08_proceedings.pdf#page=22.

Froehlich, J., Neumann, J. & Oliver, N. (2009) Sensing and predicting the pulse of the city through shared bicycling, *IJCAI International Joint Conference on Artificial Intelligence*, (3), pp. 1420–1426. doi: 10.1.1.150.4370.

Geiger, S. (2007) Exploring night-time grocery shopping behaviour, *Journal of Retailing and Consumer Services*. Elsevier, 14(1), pp. 24–34.

Geurs, K. T. & Ritsema van Eck, J. R. (2001) *Accessibility measures: review and applications: Evaluation of accessibility impacts of land-use transport scenarios, and related social and economic impacts*. Rijksinstituut voor Volksgezondheid en Milieu.

Geurs, K. T. & van Wee, B. (2004) Accessibility evaluation of land-use and transport strategies: review and research directions, *Journal of Transport Geography*, 12(2), pp. 127–140. doi: 10.1016/j.jtrangeo.2003.10.005.

Girardet, H. (2004) The metabolism of cities, *The sustainable urban development reader*. Routledge: London, UK, pp. 125–132.

Girardin, F., Vaccari, A., Gerber, A., Biderman, A. & Ratti, C. (2009) Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate, in *Intl. Conference on Computers in Urban Planning and Urban Management*.

Glorieux, I., Mestdag, I. & Minnen, J. (2008) The Coming of the 24-hour Economy?: Changing work schedules in Belgium between 1966 and 1999, *Time & Society*, 17(1), pp. 63–83. doi: 10.1177/0961463X07086310.

Golledge, R. G. & Stimson, R. J. (1997) *Spatial behavior : a geographic perspective*. Edited by R. J. Stimson. New York: Guilford Press.

González, M. C., Hidalgo, C. A. & Barabási, A. L. (2008) Understanding individual human mobility patterns, *Nature*, 453(7196), pp. 779–782. doi: 10.1038/nature06958.

Goodchild, M. F. (2013) The quality of big (geo)data, *Dialogues in Human Geography*, 3(3), pp. 280–284. doi: 10.1177/2043820613513392.

Goodchild, M. F. & Janelle, D. G. (1984) The City around the Clock: Space—Time Patterns of Urban Ecological Structure, *Environment and Planning A*, 16(6), pp. 807–820. doi: 10.1068/a160807.

Gould, P. R. (1969) *Spatial Diffusion, Resource Paper No. 4*. Association of American Geographers, Washington, D.C.

Graells-Garrido, E., Ferres, L., Caro, D. & Bravo, L. (2017) The effect of Pokémon Go on the pulse of the city: a natural experiment, *EPJ Data Science*. The Author(s), 6(1). doi: 10.1140/epjds/s13688-017-0119-3.

Graham, M. & Shelton, T. (2013) Geography and the future of big data, big data and the future of geography, *Dialogues in Human Geography*, 3(3), pp. 255–261. doi: 10.1177/2043820613513121.

Greger, K. (2015) Spatio-temporal building population estimation for highly urbanized areas using GIS, *Transactions in GIS*, 19(1), pp. 129–150. doi: 10.1111/tgis.12086.

Hägerstrand, T. (1970) What About People in Regional Science?, *Papers in Regional Science*, 24(1), pp. 7–24. doi: 10.1111/j.1435-5597.1970.tb01464.x.

Handy, S. & Niemeier, D. (1997) *Measuring Accessibility: An Exploration of Issues and Alternatives*, *Environment and Planning A*. doi: 10.1068/a291175.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, *Journal of the American Planning Association*, 25(2), pp. 73–76. doi: 10.1080/01944365908978307.

Hasan, S., Schneider, C. M., Ukkusuri, S. V. & González, M. C. (2013) Spatiotemporal Patterns of Urban Human Mobility, *Journal of Statistical Physics*, 151(1–2), pp. 304–318. doi: 10.1007/s10955-012-0645-0.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P. & Ratti, C. (2014) Geo-located Twitter as proxy for global mobility patterns, *Cartography and Geographic Information Science*. Taylor & Francis, 41(3), pp. 260–271. doi: 10.1080/15230406.2014.890072.

Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J. & Toivonen, T. (2017) User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey, *ISPRS International Journal of Geo-Information*, 6(3), p. 85. doi: 10.3390/ijgi6030085.

Hodge, D. C. (1997) Accessibility-related issues, *Journal of Transport Geography*, 5(1), pp. 33–34. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0042595685&partnerID=40&md5=fc6e9be2a76e37a67c3d08367d27f4d2.

Holma, H. & Toskala, A. (2006) *HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications*. John Wiley & Sons, Ltd. doi: 10.1002/0470032634.

Horner, M. W. & Downs, J. (2014) Integrating people and place: A density-based measure for assessing accessibility to opportunities, *Journal of Transport and Land Use*, 7(2), p. 23. doi: 10.5198/jtlu.v7i2.417.

HRT = Helsinki Region Transport (2013) *HLJ 2015 Liikkumistottumukset Helsingin seudulla 2012*. Available at: https://www.hsl.fi/sites/default/files/uploads/liikkumistottumukset_helsingin_seudulla2012.pdf.

HRT = Helsinki Region Transport (2018) HSL:n linjat ja vuorovälit. Available at: https://www.hsl.fi/avoindata.

HSY = Helsinki Region Environmental Authority (2018a) Seutukartta 2017. Helsingin kaupunkimittauspalvelut, Helsingin seudun kunnat ja HSY. Available at: https://www.hsy.fi/fi/asiantuntijalle/avoindata/Sivut/AvoinData.aspx?dataID=37.

HSY = Helsinki Region Environmental Authority (2018b) Työmatkasukkulointi pääkaupunkiseudulla. Sukkulointiaineisto 2018. HSY Seutu- ja ympäristötieto - Tietoyhteistyöyksikkö. Available at: https://www.hsy.fi/fi/asiantuntijalle/seututieto/tyopaikat/_layouts/15/WopiFrame.aspx?sourcedoc=/fi/asiantuntijalle/seututieto/tyopaikat/Documents/Työmatkasukkulointi pääkaupunkiseudulla 2015_versio 26.4.2018.pdf&action=default.

Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T. & Wang, F.-Y. (2018) Modeling real-time human mobility based on mobile phone and transportation data fusion, *Transportation Research Part C: Emerging Technologies*. Pergamon, 96, pp. 251–269. doi: 10.1016/J.TRC.2018.09.016.

Hubers, C., Schwanen, T. & Djist, M. (2009) ICT and Temporal Fragmentation of Activities: An analytical framework and initial empirical findings, *Tijdschrift voor economische en sociale geografie*. Wiley/Blackwell (10.1111), 99(5), pp. 528–546. doi: 10.1111/j.1467-9663.2008.00490.x.

Hussein, T., Paasonen, P. & Kulmala, M. (2012) Activity pattern of a selected group of school occupants and their family members in Helsinki - Finland, *Science of the Total Environment*. Elsevier B.V., 425, pp. 289–292. doi: 10.1016/j.scitotenv.2012.03.002.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J. & Varshavsky, A. (2011) Identifying Important Places in People’s Lives from Cellular Network Data, in *9th International Conference, Pervasive 2011*, pp. 133–151. doi: 10.1007/978-3-642-21726-5_9.

ITU = International Telecommunication Union (2018a) *ICT Facts and Figures 2018. Key Indicators for 2005-2018*. Available at: https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2018/ITU_Key_2005-2018_ICT_data_with LDCs_rev27Nov2018.xls (Accessed: 1 December 2018).

ITU = International Telecommunication Union (2018b) Measuring the Information Society Report 2018 - Volume 1. Geneva, Switzerland., *ITU Publications*, 1, p. 189. Available at: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR-2018-Vol-1-E.pdf.

Jaakkola, T. (2013) *Paikkatietopohjainen menetelmä autoilun ajoaikojen ja kokonaismatka-aikojen mallintamiseen – esimerkkinä pääkaupunkiseutu*. p. 84. Master's Thesis. Department of Geosciences and Geography, University of Helsinki. Available at: http://www.helsinki.fi/science/accessibility/publications/TimoJaakkola_Paikkatietopohjainen_menetelma_autoilun_ajoaikojen_ja_kokonaismatka-aikojen_mallintamiseen.pdf.

Janecek, A., Valerio, D., Hummel, K. A., Ricciato, F. & Hlavacs, H. (2015) The Cellular Network as a Sensor: From Mobile Phone Data to Real-Time Road Traffic Monitoring, *IEEE Transactions on Intelligent Transportation Systems*, 16(5), pp. 2551–2572. doi: 10.1109/TITS.2015.2413215.

Janelle, D. (2004) Impact of Information Technologies, in The Geography of Urban Transportation, in Hanson, S. and Giuliano, G. (eds) *The Geography of Urban Transportation*. Guilford Publications (Geography, urban studies), pp. 86–112.

Järv, O., Ahas, R., Saluveer, E., Derudder, B. & Witlox, F. (2012) Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour Traffic Analysis Using Call Detail Records, *PLOS ONE*. Public Library of Science, 7(11), p. e49171. Available at: https://doi.org/10.1371/journal.pone.0049171.

Järv, O., Müürisepp, K., Ahas, R., Derudder, B. & Witlox, F. (2015) Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia, *Urban Studies*, 52(14), pp. 2680–2698. doi: 10.1177/0042098014550459.

Järv, O., Tenkanen, H., Salonen, M., Ahas, R. & Toivonen, T. (2018) Dynamic cities: Location-based accessibility modelling as a function of time, *Applied Geography*. Pergamon, 95, pp. 101–110. doi: 10.1016/J.APGEOG.2018.04.009.

Järv, O., Ahas, R. & Witlox, F. (2014) Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records, *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd, 38, pp. 122–135. doi: 10.1016/j.trc.2013.11.003.

Järv, O., Tenkanen, H. & Toivonen, T. (2017a) Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation, *International Journal of Geographical Information Science*. Taylor & Francis, 31(8), pp. 1630–1651. doi: 10.1080/13658816.2017.1287369.

Järv, O., Tenkanen, H. & Toivonen, T. (2017b) *Multi-temporal function-based dasymetric interpolation tool for mobile phone data*. doi: 10.5281/ZENODO.252612.

Järvi, J., Salonen, M., Saarsalmi, P., Tenkanen, H. & Toivonen, T. (2014) Reititin: an open source tool for analysing accessibility by public transport in Greater Helsinki, in *3rd open source geospatial research & education symposium OGRS 2014*.

Kaaranen, H., Ahtiainen, A., Laitinen, L., Naghian, S. & Niemi, V. (2001) *UMTS Networks Architecture, Mobility and Services*. p. 302. John Wiley & Sons. doi: 10.1002/047001105X.

Kang, C., Liu, Y., Ma, X. & Wu, L. (2012) Towards Estimating Urban Population Distributions from Mobile Call Data, *Journal of Urban Technology*, 19(4), pp. 3–21. doi: 10.1080/10630732.2012.715479.

Kaufmann, V., Bergman, M. M. & Joye, D. (2004) Motility: mobility as capital, *International Journal of Urban and Regional Research*, 28(4), pp. 745–756. doi: 10.1111/j.0309-1317.2004.00549.x.

Kim, H.-M. & Kwan, M.-P. (2003) Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration, *Journal of geographical Systems*.

Springer, 5(1), pp. 71–91.

Kitchin, R. (2013) Big data and human geography: Opportunities, challenges and risks, *Dialogues in Human Geography*, 3(3), pp. 262–267. doi: 10.1177/2043820613513388.

Kitchin, R. (2014a) Big Data, new epistemologies and paradigm shifts, *Big Data & Society*, 1(1), p. 205395171452848. doi: 10.1177/2053951714528481.

Kitchin, R. (2014b) The real-time city? Big data and smart urbanism, *GeoJournal*, 79(1), pp. 1–14. doi: 10.1007/s10708-013-9516-8.

Kobayashi, T., Miller, H. & Othman, W. (2011) Analytical methods for error propagation in planar space-time prisms, *Journal of Geographical Systems*, 13(4), pp. 327–354. doi: 10.1007/s10109-010-0139-z.

Kontokosta, C. E. & Johnson, N. (2017) Urban phenology: Toward a real-time census of the city using Wi-Fi data, *Computers, Environment and Urban Systems*. Elsevier Ltd, 64, pp. 144–153. doi: 10.1016/j.compenvurbsys.2017.01.011.

Kujala, R., Weckström, C., Mladenović, M. N. & Saramäki, J. (2018) Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys, *Computers, Environment and Urban Systems*, 67, pp. 41–54. doi: 10.1016/j.compenvurbsys.2017.08.012.

Kuusik, A., Ahas, R. & Tiru, M. (2009) Analysing Repeat Visitation on Country Level with Passive Mobile Positioning Method: an Estonian Case Study, *Estonian Discussions on Economic Policy*, 17, pp. 140–155. doi: 10.15157/tpep.v17i0.912.

Kveladze, I., Kraak, M.-J. & van Elzakker, C. P. J. M. (2013) A Methodological Framework for Researching the Usability of the Space-Time Cube, *The Cartographic Journal*, 50(3), pp. 201–210. doi: 10.1179/1743277413Y.0000000061.

Kwan, M.-P. (1998) Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework, *Geographical analysis*, 30(3), pp. 191–216. doi: 10.1111/j.1538-4632.1998.tb00396.x.

Kwan, M.-P. (1999) Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities, *Economic Geography*. Routledge, 75(4), pp. 370–394. doi: 10.1111/j.1944-8287.1999.tb00126.x.

Kwan, M.-P. (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set, *Transportation Research Part C: Emerging Technologies*. Pergamon, 8(1–6), pp. 185–203. doi: 10.1016/S0968-090X(00)00017-6.

Kwan, M.-P. (2013) Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility, *Annals of the Association of American Geographers*, 103(5), pp. 1078–1086. doi: 10.1007/978-94-017-9205-9_4.

Laney, D. (2001) 3D data management: controlling data volume, velocity, and variety. Meta Group. Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (Accessed: 30 October 2018).

Lefebvre, H. (2004) *Rhythmanalysis: Space, time and everyday life*. p. 112. London: Continuum cop. 2004.

Lenntorp, B. (1976) Paths in space-time environments: A time-geographic study of movement possibilities of individuals, *Lund Studies in Geography Series B Human Geography*. Royal University of Lund Department of Geography Lund Sweden 1976., (44).

Levin, N., Kark, S. & Crandall, D. (2015) Where have all the people gone? Enhancing global conservation using night lights and social media, *Ecological Applications*, 25(8), pp. 2153–2167. doi: 10.1890/15-0113.1.

Li, Q., Zhang, T., Wang, H. & Zeng, Z. (2011) Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach, *Journal of Transport Geography*. Elsevier Ltd, 19(3), pp. 379–393. doi: 10.1016/j.jtrangeo.2010.07.003.

Lomborg, S. & Bechmann, A. (2014) Using APIs for Data Collection on Social Media, *Information Society*, 30(4), pp. 256–265. doi: 10.1080/01972243.2014.915276.

Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J. & Barthelemy, M. (2014) From mobile phone data to the spatial structure of cities, *Scientific Reports*. The Author(s), 4, p. 5276. Available at: http://dx.doi.org/10.1038/srep05276.

Lynch, K. (1984) *Good city form*. MIT press.

Ma, X., Wu, Y. J., Wang, Y., Chen, F. & Liu, J. (2013) Mining smart card data for transit riders' travel patterns, *Transportation Research Part C: Emerging Technologies*, 36, pp. 1–12. doi: 10.1016/j.trc.2013.07.010.

Ma, Y., Xu, W., Zhao, X. & Li, Y. (2017) Modeling the Hourly Distribution of Population at a High Spatiotemporal Resolution Using Subway Smart Card Data: A Case Study in the Central Area of Beijing, *ISPRS International Journal of Geo-Information*, 6(5), p. 128. doi: 10.3390/ijgi6050128.

Malleson, N. & Andresen, M. A. (2015) Spatio-temporal crime hotspots and the ambient population, *Crime Science*, 4(10). doi: 10.1186/s40163-015-0023-8.

Malleson, N. & Andresen, M. A. (2016) Exploring the impact of ambient population measures on London crime hotspots, *Journal of Criminal Justice*. Pergamon, 46, pp. 52–63. doi: 10.1016/J.JCRIMJUS.2016.03.002.

Malmberg, L. (2017) Apulaispormestari Nasima Razmyar haluaa Helsinkiin ”yöpormestarin” – ”Täällä ajatellaan edelleen, että yöllä ihmiset nukkuvat”, *Helsingin Sanomat. 28.10.2017*. Available at: https://www.hs.fi/kaupunki/art-2000005426863.html.

Martin, D., Cockings, S. & Leung, S. (2009) Population 24/7: building time-specific population grid models, in *Proceedings of the European Forum for Geostatistics Conference*. The Hague, The Netherlands, pp. 1–11.

Martin, D., Cockings, S. & Leung, S. (2015) Developing a Flexible Framework for Spatiotemporal Population Modeling, *Annals of the Association of American Geographers*, 105(4), pp. 754–772. doi: 10.1080/00045608.2015.1022089.

McPherson, T. & Brown, M. (2004) *Estimating daytime and nighttime population distributions in U.S. Cities for emergency response activities*, *84th AMS Annual Meeting: Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone, At Seattle WA*. doi: 10.1215/9780822384625-001.

Mennis, J. (2003) Generating Surface Models of Population Using Dasymetric Mapping, *The Professional Geographer*. Routledge, 55(1), pp. 31–42. doi: 10.1111/0033-0124.10042.

Mennis, J. (2016) Dasymetric Spatiotemporal Interpolation, *The Professional Geographer*. Routledge, 68(1), pp. 92–102. doi: 10.1080/00330124.2015.1033669.

Mennis, J. & Guo, D. (2009) Spatial data mining and geographic knowledge discovery-An introduction, *Computers, Environment and Urban Systems*, 33(6), pp. 403–408. doi: 10.1016/j.compenvurbsys.2009.11.001.

Miller, H. (2005a) A measurement theory for time geography, *Geographical Analysis*, 37(1), pp. 17–45.

doi: 10.1111/j.1538-4632.2005.00575.x.

Miller, H. (2005b) Place-Based Versus People-Based Accessibility, in Levinson, D. and Krizek, K. J. (eds) *Access to Destinations*. London: Elsevier, pp. 63–89. doi: 10.1108/9780080460550-004.

Miller, H. (2007) Place-Based versus People-Based Geographic Information Science, *Geography Compass*, 1(3), pp. 503–535. doi: 10.1111/j.1749-8198.2007.00025.x.

Miller, H. J. (1991) Modelling accessibility using space-time prism concepts within geographical information systems, *International Journal of Geographical Information Systems*, 5(3), pp. 287–301. doi: 10.1080/02693799108927856.

de Montjoye, Y.-A. *et al*. (2018) On the privacy-conscientious use of mobile phone data, *Scientific Data*. The Author(s), 5, p. 180286. Available at: https://doi.org/10.1038/sdata.2018.286.

Moya-Gómez, B., Salas-Olmedo, M. H., García-Palomares, J. C. & Gutiérrez, J. (2017) Dynamic Accessibility using Big Data: The Role of the Changing Conditions of Network Congestion and Destination Attractiveness, *Networks and Spatial Economics*. Networks and Spatial Economics, pp. 1–18. doi: 10.1007/s11067-017-9348-z.

Mylonas, A., Meletiadis, V., Tsoumas, B., Mitrou, L. & Gritzalis, D. (2012) Smartphone Forensics: A Proactive Investigation Scheme for Evidence Acquisition, in Gritzalis, D., Furnell, S., and Theoharidou, M. (eds) *Information Security and Privacy Research*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 249–260.

Næss, P. (2005) Residential location influences travel - but how and why?, *Progress in planning*, 63(1), pp. 167–257.

Næss, P. (2006) Accessibility, activity participation and location of activities: Exploring the links between residential location and travel behaviour, *Urban Studies*, 43(3), pp. 627–652. doi: 10.1080/00420980500534677.

Neutens, T., Schwanen, T., Witlox, F. & de Maeyer, P. (2010) Equity of urban service delivery: A comparison of different accessibility measures, *Environment and Planning A*, 42(7), pp. 1613–1635. doi: 10.1068/a4230.

Neutens, T., Delafontaine, M., Schwanen, T. & Weghe, N. Van de (2012) The relationship between opening hours and accessibility of public service delivery, *Journal of Transport Geography*, 25, pp. 128–140. doi: 10.1016/j.jtrangeo.2011.03.004.

Neutens, T., Schwanen, T. & Witlox, F. (2011) The prism of everyday life: Towards a new research agenda for time geography, *Transport Reviews*, 31(1), pp. 25–47. doi: 10.1080/01441647.2010.484153.

Neutens, T., Versichele, M. & Schwanen, T. (2010) Arranging place and time: A GIS toolkit to assess person-based accessibility of urban opportunities, *Applied Geography*. Elsevier Ltd, 30(4), pp. 561–575. doi: 10.1016/j.apgeog.2010.05.006.

Ni, L. M., Liu, Y., Lau, Y. C. & Patil, A. P. (2003) LANDMARC: indoor location sensing using active RFID, in *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*. IEEE, pp. 407–415.

Ni, L., Wang, X. (Cara) & Chen, X. (Michael) (2018) A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data, *Transportation Research Part C: Emerging Technologies*. Elsevier, 86(April 2017), pp. 510–526. doi: 10.1016/j.trc.2017.12.002.

NLS = National Land Survey of Finland (2018a) Kuntajako 2018 1:10 000. Retrieved 18.1.2018.

NLS = National Land Survey of Finland (2018b) National Topographic Database. Retrieved 8.2.2018.

NLS = National Land Survey of Finland (2018c) Suomen pinta-ala kunnittain 1.1.2018, p. 8. Available at: https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2018/01/Suomen_pa_2018_kunta_maakunta_0.pdf.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. (2012) A tale of many cities: Universal patterns in human urban mobility, *PLoS ONE*, 7(5). doi: 10.1371/journal.pone.0037027.

Novak, J. & Sykora, L. (2013) A City in Motion : Time-Space Activity and Mobility Patterns of Suburban Inhabitants and the Structuration of the Spatial Organization of the Prague, *Geografiska Annaler: Series B, Human Geography* 89(2), pp. 147–168.

Nylander, S., Lundquist, T. & Brännström, A. (2009) At home and with computer access: why and where people use cell phones to access the internet, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 1639–1642.

Okabe, A., Boots, B., Sugihara, K. & Chiu, S. N. (2000) *Spatial tessellations: concepts and applications of Voronoi diagrams*. 2nd Editio. Chichester: Wiley.

Oliver, N., Matic, A. & Frias-Martinez, E. (2015) Mobile Network Data for Public Health: Opportunities and Challenges, *Frontiers in Public Health*, 3(August), pp. 1–12. doi: 10.3389/fpubh.2015.00189.

Olsson, G. (1965) *Distance and human interaction : a review and bibliography*. Philadelphia, PA: Regional Science Research Institute.

Onnela, J.-P., Saramaki, J., Hyvonen, J., Szabo, G., Lazer, D., Kaski, K., Kertesz, J. & Barabasi, A.-L. (2007) Structure and tie strengths in mobile communication networks, *PNAS*, 104(18), pp. 7332–7336. doi: 10.1073/pnas.0610245104.

Openshaw, S. (1984) The modifiable areal unit problem, *Concepts and techniques in modern geography*. GeoBooks.

Páez, A., Scott, D. M. & Morency, C. (2012) Measuring accessibility: positive and normative implementations of various accessibility indicators, *Journal of Transport Geography*. Pergamon, 25, pp. 141–153. doi: 10.1016/J.JTRANGEO.2012.03.016.

Parkes, D. & Thrift, N. J. (1980) *Times, Spaces, and Places: A Chronogeographic Perspective*. Chichester: J. Wiley (Chronogeographic Perspective).

Patel, N. N., Stevens, F. R., Huang, Z., Gaughan, A. E., Elyazar, I. & Tatem, A. J. (2017) Improving Large Area Population Mapping Using Geotweet Densities, *Transactions in GIS*, 21(2), pp. 317–331. doi: 10.1111/tgis.12214.

Pei, L., Guinness, R., Chen, R., Liu, J., Kuusniemi, H., Chen, Y., Chen, L. & Kaistinen, J. (2013) Human behavior cognition using smartphone sensors, *Sensors (Switzerland)*, 13(2), pp. 1402–1424. doi: 10.3390/s130201402.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S. L., Li, T. & Zhou, C. (2014) A new insight into land use classification based on aggregated mobile phone data, *International Journal of Geographical Information Science*. Taylor & Francis, 28(9), pp. 1988–2007. doi: 10.1080/13658816.2014.913794.

Pereira, R. H. M. (2019) Future accessibility impacts of transport policy scenarios: equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro, *Journal of Transport Geography. Forthcoming.* Elsevier, 74, pp. 321–332.

Peuquet, D. J. & Duan, N. (1995) An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data, *International Journal of Geographical Information Systems*, 9(1), pp. 7–24. doi: 10.1080/02693799508902022.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R. & Ratti, C. (2010) Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data, in Salah, A. A. et al. (eds) *Human Behavior Understanding*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 14–25.

Phithakkitnukoon, S., Smoreda, Z. & Olivier, P. (2012) Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data, *PLOS ONE*. Public Library of Science, 7(6), p. e39253. Available at: https://doi.org/10.1371/journal.pone.0039253.

Pinelli, F., Lorenzo, G. Di & Calabrese, F. (2015) Comparing Urban Sensing Applications Using Event and Network-Driven Mobile Phone Location Data, *Proceedings - IEEE International Conference on Mobile Data Management*, 1, pp. 219–226. doi: 10.1109/MDM.2015.33.

Potere, D. & Schneider, A. (2007) A critical look at representations of urban areas in global maps, *GeoJournal*, 69(1–2), pp. 55–80. doi: 10.1007/s10708-007-9102-z.

Pred, A. (1977) The Choreography of Existence: Comments on Hägerstrand’s Time-Geography and Its Usefulness, *Economic Geography*. Routledge, 53(2), pp. 207–221. doi: 10.2307/142726.

PTY = Päivittäistavarakauppa ry (2018) *Päivittäistavarakauppa 2018. Nielsen Päivittäistavaramyymälärekisteri.* Available at: https://www.pty.fi/fileadmin/user_upload/tiedostot/Julkaisut/Vuosijulkaisut/FI_2018_vuosijulkaisu.pdf.

Pulselli, R. M., Romano, P., Ratti, C. & Tiezzi, E. (2008) Computing urban mobile landscapes through monitoring population density based on cell-phone chatting, *International Journal of Design & Nature and Ecodynamics*, 3(2), pp. 121–134. doi: 10.2495/D&NE-V3-N2-121-134.

Qi, W., Liu, S., Gao, X. & Zhao, M. (2015) Modeling the spatial distribution of urban population during the daytime and at night based on land use: A case study in Beijing, China, *Journal of Geographical Sciences*, 25(6), pp. 756–768. doi: 10.1007/s11442-015-1200-0.

Ratti, C., Frenchman, D., Pulselli, R. M. & Williams, S. (2006) Mobile landscapes: Using location data from cell phones for urban analysis, *Environment and Planning B: Planning and Design*, 33(5), pp. 727–748. doi: 10.1068/b32047.

Reades, J., Calabrese, F., Sevtsuk, A. & Ratti, C. (2007) Cellular Census: Explorations in urban data collection, *Pervasive computing*, 6(3), pp. 30–38. doi: 10.1109/MPRV.2007.53.

Renso, C., Puntoni, S., Frentzos, E., Mazzoni, A., Moelans, B., Pelekis, N. & Pini, F. (2008) Wireless network data sources: tracking and synthesizing trajectories, in *Mobility, data mining and privacy*. Springer, pp. 73–100.

Ricciato, F., Widhalm, P., Massimo, C. & Pantisano, F. (2015) Estimating Population Density Distribution from Network-based Mobile Phone Data, *JRC Technical Report*, p. 65. Available at: http://dx.doi.org/10.2788/162414.

Ricciato, F., Widhalm, P., Pantisano, F. & Craglia, M. (2017) Beyond the ‘single-operator, CDR-only’ paradigm: An interoperable framework for mobile phone network data analyses and population density estimation, *Pervasive and Mobile Computing*. Elsevier, 35, pp. 65–82. doi: 10.1016/J.PMCJ.2016.04.009.

Roenneberg, T., Wirz-Justice, A. & Merrow, M. (2003) Life between Clocks: Daily Temporal Patterns of Human Chronotypes, *Journal of Biological Rhythms*. SAGE Publications Inc, 18(1), pp. 80–90. doi: 10.1177/0748730402239679.

Rojas, M. B., Sadeghvaziri, E. & Jin, X. (2016) Comprehensive Review of Travel Behavior and Mobility Pattern Studies That Used Mobile Phone Data, *Transportation Research Record: Journal of the Transportation Research Board*, 2563(January), pp. 71–79. doi: 10.3141/2563-11.

Romanillos, G., Zaltz Austwick, M., Ettema, D. & De Kruijf, J. (2016) Big Data and Cycling, *Transport Reviews*, 36(1), pp. 114–133. doi: 10.1080/01441647.2015.1084067.

Saarsalmi, P. (2014) *Päivittäistavarakaupan spatio-temporaalinen saavutettavuus pääkaupunkiseudulla*. p. 107. Master's Thesis. Department of Geosciences and Geography, University of Helsinki. Available at: http://urn.fi/URN:NBN:fi-fe2017112252288.

Sagl, G., Resch, B., Hawelka, B. & Beinat, E. (2012) From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments, in *Proceedings of the GI-Forum*. Herbert Wichmann Verlag Berlin, pp. 54–63.

Salomon, I. & Mokhtarian, P. L. (1998) What happens when mobility-inclined market segments face accessibility-enhancing policies?, *Transportation Research Part D: Transport and Environment*. Pergamon, 3(3), pp. 129–140. doi: 10.1016/S1361-9209(97)00038-2.

Salonen, M. (2014) *Analysing spatial accessibility patterns with travel time and distance measures: novel approaches for rural and urban contexts*. University of Helsinki. Available at: http://urn.fi/URN:ISBN:978-952-10-9466-8.

Salonen, M., Broberg, A., Kyttä, M. & Toivonen, T. (2014) Do suburban residents prefer the fastest or low-carbon travel modes? Combining public participation GIS and multimodal travel time analysis for daily mobility research, *Applied Geography*. Pergamon, 53, pp. 438–448. doi: 10.1016/J.APGEOG.2014.06.028.

Salonen, M. & Toivonen, T. (2013) Modelling travel time in urban networks: Comparable measures for private car and public transport, *Journal of Transport Geography*. Elsevier Ltd, 31, pp. 143–153. doi: 10.1016/j.jtrangeo.2013.06.011.

Salonen, M., Toivonen, T. & Vaattovaara, M. (2012) Arkiliikkumisen vaihtoehdoista monikeskuksistuvassa metropolissa: Kaksi näkökulmaa palvelujen saavutettavuuteen pääkaupunkiseudulla., *Yhdyskuntasuunnittelu*, 3, pp. 8–27.

Saluveer, E. & Ahas, R. (2014) *Using call detail records of mobile network operators for transportation studies*, *Mobile Technologies for Activity-Travel Data Collection and Analysis*. IGI Global. doi: 10.4018/978-1-4666-6170-7.ch014.

Samaniego, H. & Moses, M. E. (2008) Cities as Organisms: Allometric Scaling of Urban Road Networks, *Journal of Transport and Land Use*, 1(1), pp. 21–39. doi: 10.5198/jtlu.v1i1.29.

Schlich, R., Schönfelder, S., Hanson, S. & Axhausen, K. W. (2004) Structures of Leisure Travel: Temporal and Spatial Variability, *Transport Reviews*, 24(2), pp. 219–237. doi: 10.1080/0144164032000138742.

Schmitt, R. C. (1956) Estimating Daytime Populations, *Journal of the American Planning Association*, 22(2), pp. 83–85. doi: 10.1080/01944365608979227.

Schönfelder, S. & Axhausen, K. (2010) *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*. 1st edn. London: Routledge.

Sevtsuk, A. & Ratti, C. (2010) Does Urban mobility have a daily routine? Learning from the aggregate data of mobile networks, *Journal of Urban Technology*, 17(1), pp. 41–60. doi: 10.1080/10630731003597322.

Sheller, M. & Urry, J. (2006) The new mobilities paradigm, *Environment and Planning A*, 38(2), pp. 207–226. doi: 10.1068/a37268.

Sloan, L., Morgan, J., Burnap, P. & Williams, M. (2015) Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data, *PloS one*. Public Library of Science, 10(3), pp. e0115545–e0115545. doi: 10.1371/journal.pone.0115545.

Smith, S. K. (1989) Toward a methodology for estimating temporary residents, *Journal of the American Statistical Association*, 84(406), pp. 430–436. doi: 10.1080/01621459.1989.10478787.

Sobolevsky, S., Sitko, I., Des Combes, R. T., Hawelka, B., Arias, J. M. & Ratti, C. (2014) Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain, *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*. IEEE, pp. 136–143. doi: 10.1109/BigData.Congress.2014.28.

Song, C., Qu, Z., Blumm, N. & Barabasi, A.-L. (2010) Limits of Predictability in Human Mobility, *Science*, 327, pp. 1018–1021. doi: 10.1126/science.1177170.

Song, C., Koren, T., Wang, P. & Barabási, A.-L. (2010) Modelling the scaling properties of human mobility, *Nature Physics*. Nature Publishing Group, 6, p. 818. Available at: https://doi.org/10.1038/nphys1760.

Sorokin, P. A. & Merton, R. K. (1937) Social Time : A Methodological and Functional Analysis THE AMERICAN, *The American Journal of Sociology*, 42(5), pp. 615–629.

Soto, V. & Frías-Martínez, E. (2011) Automated land use identification using cell-phone records, *Proceedings of the 3rd ACM international workshop on MobiArch - HotPlanet '11*, p. 17. doi: 10.1145/2000172.2000179.

Srivatsa, M., Ganti, R. & Mohapatra, P. (2017) On the Limits of Subsampling of Location Traces, *Proceedings - International Conference on Distributed Computing Systems*. IEEE, pp. 1032–1041. doi: 10.1109/ICDCS.2017.82.

Stange, H., Liebig, T., Hecker, D., Andrienko, G. & Andrienko, N. (2011) Analytical workflow of monitoring human mobility in big event settings using Bluetooth, *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness - ISA '11*, (November), p. 51. doi: 10.1145/2077357.2077368.

Statistics Finland (2010) Time use survey 2009-2010. Helsinki, Tilastokeskus.

Statistics Finland (2015) YKR 250 m x 250 m statistical grid cells. Digital Geography Lab. Available at: https://blogs.helsinki.fi/accessibility/helsinki-region-travel-time-matrix-2018/.

Statistics Finland (2016) *Grid Database 2016.* Helsinki, Tilastokeskus.

Statistics Finland (2018a) *Suomen virallinen tilasto.* Väestörakenne. Väestöllisiä tunnuslukuja alueittain 1990 - 2017. Helsinki, Tilastokeskus. Available at: http://www.stat.fi/til/vaerak/index.html (Accessed: 5.4. 2018).

Statistics Finland (2018b) *Suomen virallinen tilasto.* Työssäkäynti. Helsinki, Tilastokeskus. Accessed 15.12.2018. Available at: http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__vrm__tyokay/statfin_tyokay_pxt_009.px/?rxid=c20851ad-3cbb-4fd6-8f1a-a8ba5205e684.

Statistics Finland (2018c) *Suomen virallinen tilasto.* Väestön tieto- ja viestintätekniikan käyttö 2018. Helsinki: Tilastokeskus. Accessed 15.12.2018. Available at: http://www.stat.fi/til/sutivi/2018/sutivi_2018_2018-12-04_tie_001_fi.html.

Steenbruggen, J., Tranos, E. & Nijkamp, P. (2015) Data from mobile phone operators: A tool for smarter cities?, *Telecommunications Policy*. Elsevier, 39(3–4), pp. 335–346. doi: 10.1016/j.telpol.2014.04.001.

Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. (2015) Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data, *PLoS ONE*, 10(2), pp. 1–22. doi: 10.1371/journal.pone.0107042.

Sui, D. (2012) Looking through Hägerstrand's dual vistas: towards a unifying framework for time geography, *Journal of Transport Geography*. Pergamon, 23, pp. 5–16. doi: 10.1016/J.JTRANGEO.2012.03.020.

Sulis, P., Manley, E., Zhong, C. & Batty, M. (2018) Using mobility data as proxy for measuring urban vitality, *Journal of Spatial Information Science*, 16(16), pp. 137–162. doi: 10.5311/JOSIS.2018.16.384.

Sutton, P., Roberts, D., Elvidge, C. & Baugh, K. (2001) Census from Heaven: An estimate of the global human population using night-time satellite imagery, *International Journal of Remote Sensing*, 22(16), pp. 3061–3076. doi: 10.1080/01431160010007015.

Sutton, P., Elvidge, C. & Obremski, T. (2003) Building and Evaluating Models to Estimate Ambient Population Density, *Photogrammetric Engineering & Remote Sensing*, 69(5), pp. 545–553. doi: 10.14358/PERS.69.5.545.

Tatem, A. J. (2017) WorldPop, open data for spatial demography, *Scientific Data*, 4, pp. 2–5. doi: 10.1038/sdata.2017.4.

Tatem, A. J., Noor, A. M. & Hay, S. I. (2005) Assessing the accuracy of satellite derived global and national urban maps in Kenya, *Remote sensing of environment*, 96(1), pp. 87–97. doi: 10.1016/j.rse.2005.02.001.

Tatem, A. & Linard, C. (2011) Population mapping of poor countries, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 474(7349), p. 36. doi: 10.1038/474036d.

Tenkanen, H., Saarsalmi, P., Järv, O., Salonen, M. & Toivonen, T. (2016) Health research needs more comprehensive accessibility measures: integrating time and transport modes from open data, *International Journal of Health Geographics*, 15(23), pp. 1–12. doi: 10.1186/s12942-016-0052-x.

Tenkanen, H. (2017) *Capturing time in space: Dynamic analysis of accessibility and mobility to support spatial planning with open data and tools*. Doctoral Thesis. University of Helsinki, Department of Geosciences and Geography 2017.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L. & Toivonen, T. (2017) Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas, *Scientific Reports*. Springer US, 7(1), pp. 1–11. doi: 10.1038/s41598-017-18007-4.

Tenkanen, H., Londoño Espinosa, J., Willberg, E., Heikinheimo, V., Tarnanen, A., Jaakkola, T., Järvi, J., Salonen, M. & Toivonen, T. (2018) *Dataset: Helsinki Region Travel Time Matrix 2018*. doi: 10.13140/RG.2.2.20858.39362.

Tobler, W., Deichmann, U., Gottsegen, J. & Maloy, K. (1997) World population in a grid of spherical quadrilaterals, *International Journal of Population Geography*, 3(3), pp. 203–225. doi: 10.1002/(SICI)1099-1220(199709)3:3<203::AID-IJPG68>3.0.CO;2-C.

Toivonen, T., Salonen, M., Tenkanen, H., Saarsalmi, P., Jaakkola, T. & Järvi, J. (2014) Joukkoliikenteellä, autolla ja kävellen: Avoin saavutettavuusaineisto pääkaupunkiseudulla, *Terra*, pp. 27–136. Available at: http://www.helsinki.fi/science/accessibility/publications/Toivonen_etal_2014_terra.pdf.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H. & Di Minin, E. (2019) Social media data for conservation science: A methodological overview, *Biological Conservation*.

Elsevier, 233, pp. 298–315. doi: 10.1016/J.BIOCON.2019.01.023.

Toole, J. L., Ulm, M., González, M. C. & Bauer, D. (2012) Inferring land use from mobile phone activity, in *Proceedings of the ACM SIGKDD international workshop on urban computing*. ACM, pp. 1–8.

Townsend, A. M. (2000) Life in the Real-Time City: Mobile Telephones and Urban Metabolism, *Journal of Urban Technology*, 7(2), pp. 85–104. doi: 10.1080/713684114.

Traag, V. A., Browet, A., Calabrese, F. & Morlot, F. (2011) Social event detection in massive mobile phone data using probabilistic location inference, *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pp. 625–628. doi: 10.1109/PASSAT/SocialCom.2011.133.

Tranos, E. & Nijkamp, P. (2015) Mobile phone usage in complex urban systems: a space–time, aggregated human activity study, *Journal of Geographical Systems*, 17(2), pp. 157–185. doi: 10.1007/s10109-015-0211-9.

Tu, W., Cao, J., Yue, Y., Shaw, S. L., Zhou, M., Wang, Z., Chang, X., Xu, Y. & Li, Q. (2017) Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns, *International Journal of Geographical Information Science*. Taylor & Francis, 31(12), pp. 2331–2358. doi: 10.1080/13658816.2017.1356464.

UNECE = United Nations Economic Commission for Europe (2007) *Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics*. United Nations Publications.

United Nations (2014) *World Urbanization Prospects: The 2014 Revision, Highlights.* Department of Economic and Social Affairs, Population Division, New York. Available at: https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.pdf.

United Nations (2017) *Handbook on the use of Mobile Phone data for Official Statistics.* UN Global Working Group on Big Data for Official Statistics. November 2017., p. 83. Available at: https://unstats.un.org/bigdata/taskteams/mobilephone/Handbook on Mobile Phone Data for official statistics - Draft Nov 2017.pdf.

United Nations (2018) *World Urbanization Prospects: The 2018 Revision, Key Facts. Department of Social Affairs, Population Division, New York*. Available at: https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf.

Vaccari, A., Martino, M., Rojas, F. & Ratti, C. (2010) Pulse of the city: Visualizing Urban Dynamics of Special Events, *Proceedings of the 20th International Conference on Computer Graphics and Vision, GraphiCon'2010*, pp. 1–9.

Valente, P. (2010) Census taking in Europe: how are populations counted in 2010?, *Population & Societies*. Institut National d'études Démographiques, 467(1), p. 4.

Versichele, M., Neutens, T., Delafontaine, M. & Van de Weghe, N. (2012) The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities, *Applied Geography*. Pergamon, 32(2), pp. 208–220. doi: 10.1016/J.APGEOG.2011.05.011.

Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I. & Van de Weghe, N. (2014) Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium, *Tourism Management*. Pergamon, 44, pp. 67–81. doi: 10.1016/J.TOURMAN.2014.02.009.

Vigdor, J. L. (2004) Community composition and collective action: Analyzing initial mail response to the

2000 census, *Review of Economics and Statistics*, 86(1), pp. 303–312. doi: 10.1162/003465304323023822.

Vilhelmson, B. (1999) Daily mobility and the use of time for different activities. The case of Sweden, *GeoJournal*, 48, pp. 177–185.

Wang, F. & Chen, C. (2018) On data processing required to derive mobility patterns from passively-generated mobile phone data, *Transportation Research Part C*. Elsevier, 87(November 2017), pp. 58–74. doi: 10.1016/j.trc.2017.12.003.

Wang, Y., Correia, G. H. de A., van Arem, B. & Timmermans, H. J. P. (Harry) (2018) Understanding travellers' preferences for different types of trip destination based on mobile internet usage data, *Transportation Research Part C: Emerging Technologies*. Pergamon, 90, pp. 247–259. doi: 10.1016/J.TRC.2018.03.009.

Wang, Z., He, S. Y. & Leung, Y. (2018) Applying mobile phone data to travel behaviour research: A literature review, *Travel Behaviour and Society*. Hong Kong Society for Transportation Studies, 11, pp. 141–155. doi: 10.1016/j.tbs.2017.02.005.

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V. & Tatem, A. J. (2018) Spatially disaggregated population estimates in the absence of national population and housing census data, *Proceedings of the National Academy of Sciences*, 115(14), p. 201715305. doi: 10.1073/pnas.1715305115.

van Wee, B. & Geurs, K. T. (2011) Discussing Equity and Social Exclusion in Accessibility Evaluations, *European Journal of Transport and Infrastructure Research*, 11(11), pp. 350–367.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W. & Buckee, C. O. (2012) Quantifying the impact of human mobility on malaria, *Science*, 338(6104), pp. 267–270. doi: 10.1126/science.1223467.Quantifying.

Widener, M. J., Farber, S., Neutens, T. & Horner, M. (2015) Spatiotemporal accessibility to supermarkets using public transit: An interaction potential approach in Cincinnati, Ohio, *Journal of Transport Geography*. Elsevier Ltd, 42, pp. 72–83. doi: 10.1016/j.jtrangeo.2014.11.004.

Widhalm, P., Yang, Y., Ulm, M., Athavale, S. & González, M. C. (2015) Discovering urban activity patterns in cell phone data, *Transportation*, 42(4), pp. 597–623. doi: 10.1007/s11116-015-9598-x.

Willberg, E. (2019) *Bike sharing as a part of urban mobility in Helsinki - A user perspective*. p. 94. Master's Thesis. Department of Geosciences and Geography, University of Helsinki.

Wirth, L. (1938) Urbanism as a Way of Life, *American Journal of Sociology*. The University of Chicago Press, 44(1), pp. 1–24. doi: 10.1086/217913.

Wu, S., Qiu, X. & Wang, L. (2005) Population Estimation Methods in GIS and Remote Sensing: A Review, *GIScience & Remote Sensing*, 42(1), pp. 80–96. doi: 10.2747/1548-1603.42.1.80.

Yu, B., Lian, T., Huang, Y., Yao, S., Ye, X. & Chen, Z. (2018) Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement, *Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement*. Taylor & Francis, 00(00), pp. 1–20. doi: 10.1080/13658816.2018.1555642.

Yu, H. & Shaw, S.-L. (2007) Revisiting Hägerstrand's Time-Geographic Framework for Individual Activities in the Age of Instant Access, in *Societies and Cities in the Age of Instant Access*, pp. 103–118. doi: 10.1007/1-4020-5427-0_7.

Yuan, Y. & Raubal, M. (2012) Extracting dynamic urban mobility patterns from mobile phone data, in

International Conference on Geographic Information Science. Springer, pp. 354–367.

Zandvliet, R. & Dijst, M. (2005) Research Note—The Ebb and Flow of Temporary Populations: The Dimensions of Spatial-Temporal Distributions of Daytime Visitors in the Netherlands, *Urban Geography*, 26(4), pp. 353–364. doi: 10.2747/0272-3638.26.4.353.

Zhao, P. (2010) Sustainable urban expansion and transportation in a growing megacity: Consequences of urban sprawl for mobility on the urban fringe of Beijing, *Habitat International*. Elsevier Ltd, 34(2), pp. 236–243. doi: 10.1016/j.habitatint.2009.09.008.

Zhao, Z., Shaw, S. L., Xu, Y., Lu, F., Chen, J. & Yin, L. (2016) Understanding the bias of call detail records in human mobility research, *International Journal of Geographical Information Science*. Taylor & Francis, 30(9), pp. 1738–1762. doi: 10.1080/13658816.2015.1137298.

APPENDIX

Appendix 1. Reclassification of the CORINE land cover data (2012) to activity function types (AFT) for the MFD method. The table contains the names of all classes in the original dataset on level 4.

Source class	Source class name (original)	Source class name (translated)	Target class (AFT)
1	Kerrostaloalueet	Continuous urban fabric (apartment house area)	Residential
2	Pientaloalueet	Discontinuous urban fabric (detached house area)	Residential
3	Palveluiden alueet	Commercial units	Work
4	Teollisuuden alueet	Industrial units	Work
5	Liikennealueet	Road and rail networks and associated land	Transport
6	Satama-alueet	Port areas	Transport*
7	Lentokenttäalueet	Airports	Transport*
8	Maa-ainesten ottoalueet	Soil extraction sites	Restricted
9	Kaivokset	Mines	Restricted
10	Kaatopaikat	Dump sites	Restricted
11	Rakennustyöalueet	Construction sites	Work
12	Vapaa-ajan asunnot	Leisure buildings	Residential
13	Muut urheilu- ja vapaa-ajan toiminta –alueet	Other areas for sport and leisure activities	Other
14	Golfkentät	Golf courses	Other
15	Raviradat	Race tracks	Other
16	Pellot	Non-irrigated arable land	Restricted
17	Hedelmäpuu- ja marjapensasviljelmät	Fruit trees and berry plantations	Restricted
18	Laidunmaat	Pastures	Restricted
19	Luonnon laidunmaat	Pastures	Restricted
20	Käytöstä poistunut maatalousmaa	Disused pastures	Other
21	Puustoiset pelto- ja laidunmaat	Agro-forestry areas	Restricted
22	Lehtimetsät kivennäismaalla	Broad-leaved forest on mineral soil	Other
23	Lehtimetsät turvemaalla	Broad-leaved forest on peatland	Other
24	Havumetsät kivennäismaalla	Coniferous forest on mineral soil	Other
25	Havumetsät turvemaalla	Coniferous forest on peatland	Other
26	Havumetsät kalliomaalla	Coniferous forest on bare rock	Other
27	Sekametsät kivennäismaalla	Mixed forest on mineral soil	Other
28	Sekametsät turvemaalla	Mixed forest on peatland	Other
29	Sekametsät kalliomaalla	Mixed forest on bare rock	Other
30	Luonnonniityt	Natural grassland	Other
31	Varvikot ja nummet	Moors and heathland	Other
32	Harvapuustoiset alueet , cc <10%	Transitional woodland/shrub, cc** <10%	Other
33	Harvapuustoiset alueet, cc 10-30%, kivennäismaalla	Transitional woodland/shrub, cc 10-30%, mineral land	Other
34	Harvapuustoiset alueet, cc 10-30%, turvemaalla	Transitional woodland/shrub, cc 10-30%, on peatland	Other
35	Harvapuustoiset alueet, cc 10-30%, kalliomaalla	Transitional woodland/shrub, cc 10-30%, on bare rock	Other
36	Harvapuustoiset alueet, sähkölinjan alla	Transitional woodland/shrub, under a power line	Other
37	Rantahietikot ja dyynialueet	Beaches, dunes, and sand plains	Other
38	Kalliomaat	Bare rock	Other
39	Niukkakasvustoiset kangasmaat	Sparsely vegetated areas	Other
40	Sisämaan kosteikot maalla	Inland marshes on land	Restricted
41	Sisämaan kosteikot vedessä	Inland marshes in water	Restricted
42	Avosuot	Peatbogs	Restricted
43	Turvetuotantoalueet	Peatbogs	Restricted
44	Merenrantakosteikot maalla	Salt marshes	Restricted
45	Merenrantakosteikot vedessä	Salt marshes	Restricted
46	Joet	Rivers	Restricted
47	Järvet	Lakes	Restricted
48	Meri	Sea and ocean	Restricted

Appendix 2. Reclassification of the building data from the National Topographic Database (2018) to activity function types (AFT) for the MFD method. The table contains the names of the main classes in the original dataset.

Source class ID	Source class name (original)	Source class name (translated)	Target class based on AFT
4221x	Asuinrakennus	Residential buildings	Residential
4222x	Liike- tai julkinen rakennus	Commercial and public buildings	Work
4223x	Lomarakennus	Buildings for leisure use	Residential
4224x	Teollinen rakennus	Industrial buildings	Work
42270	Kirkko	Churches	Other
4225x	Kirkollinen rakennus	Other religious buildings	Other
4226x	Muu rakennus	Other buildings	Work

Appendix 3. Reclassification of the building data from OpenStreetMap to activity function types (AFT) for the MFD method.

```
#reclassified buildings based on AFT
residential = ['residential', 'apartments', 'cabin', 'duplex', 'triplex', 'house', 'detached',
'semidetached_house', 'semi', 'prison', 'bungalow', 'hotel', 'hostel', 'dormitory', 'farm',
'garage', 'manor', 'mansion', 'villa', 'estate', 'terrace', 'terraced']

work = ['industrial', 'warehouse', 'school', 'office', 'public', 'civic','public_building',
'childcare', 'education', 'townhouse', 'university', 'construction', 'utility', 'roundhouse',
'commercial', 'Commercial', 'hospital', 'kindergarten', 'logistics', 'storage', 'PK-yritykset',
'hangar', 'gatehouse', 'greenhouse', 'glasshouse', 'farm_auxiliary', 'cowshed', 'barn',
'stable', 'silo', 'stables', 'guard_booth', 'guard', 'manufacture']

service = ['retail', 'shop', 'mall', 'service', 'carwash', 'store', 'supermarket', 'kiosk']

other = ['library', 'stadium', 'cafe', 'Cafe', 'sports_centre', 'swimming hall', 'sauna',
'museum', 'sport', 'horse arena', 'event_space', 'pavilion', 'hall', 'cathedral', 'church',
'chapel', 'social_facility', 'hut', 'shed', 'workshop', 'play_hut','playhut', 'manege',
'outhouse', 'bird_hide']

transport = ['parking', 'train_station', 'station', 'transportation', 'underground_entrance']
```

Appendix 4. Correlations coefficients between interpolated mobile phone data (HSPA calls) and population register data and workplace locations. The results are shown separately using MFD (1) and AW (2) interpolation methods.

		MFD_H0	MFD_H1	MFD_H2	MFD_H3	MFD_H4	MFD_H5	MFD_H6	MFD_H7	MFD_H8	MFD_H9	MFD_H10	MFD_H11
pop_norm	Pearson Correlation	,711**	,699**	,689**	,680**	,678**	,661**	,635**	,583**	,493**	,439**	,415**	,397**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231
workp_norm	Pearson Correlation	,349**	,356**	,362**	,368**	,364**	,373**	,411**	,498**	,588**	,630**	,644**	,648**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231

		MFD_H12	MFD_H13	MFD_H14	MFD_H15	MFD_H16	MFD_H17	MFD_H18	MFD_H19	MFD_H20	MFD_H21	MFD_H22	MFD_H23
pop_norm	Pearson Correlation	,386**	,385**	,395**	,410**	,445**	,497**	,550**	,604**	,656**	,696**	,717**	,716**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231
workp_norm	Pearson Correlation	,648**	,648**	,643**	,623**	,588**	,549**	,511**	,479**	,443**	,398**	,367**	,357**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231

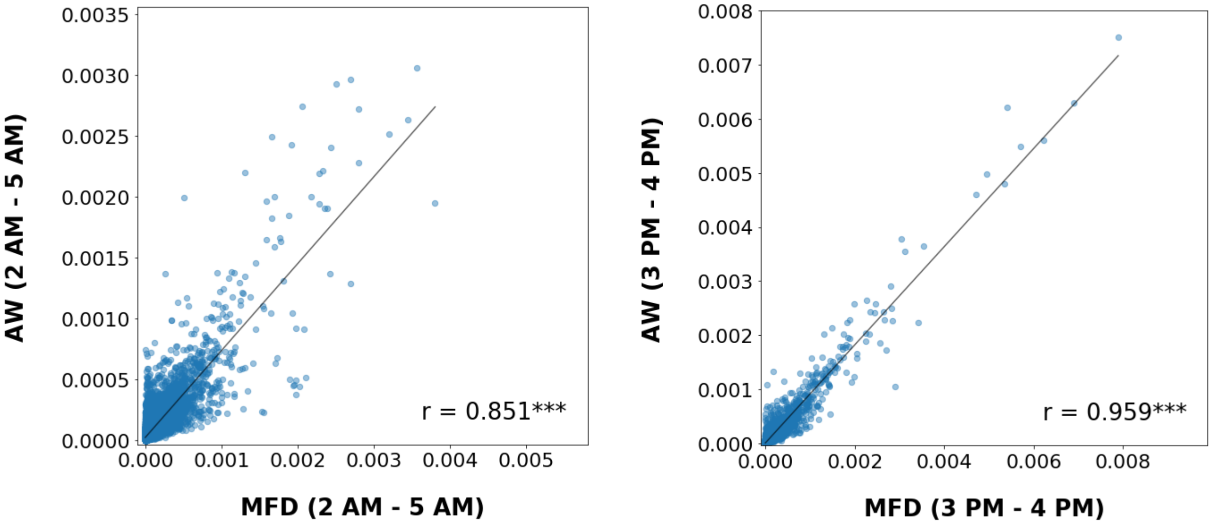
**. Correlation is significant at the 0.01 level (2-tailed).

		AW_H0	AW_H1	AW_H2	AW_H3	AW_H4	AW_H5	AW_H6	AW_H7	AW_H8	AW_H9	AW_H10	AW_H11
pop_norm	Pearson Correlation	,567**	,553**	,539**	,530**	,527**	,510**	,485**	,443**	,389**	,357**	,346**	,335**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231
workp_norm	Pearson Correlation	,421**	,430**	,435**	,440**	,437**	,447**	,479**	,524**	,573**	,600**	,606**	,608**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231

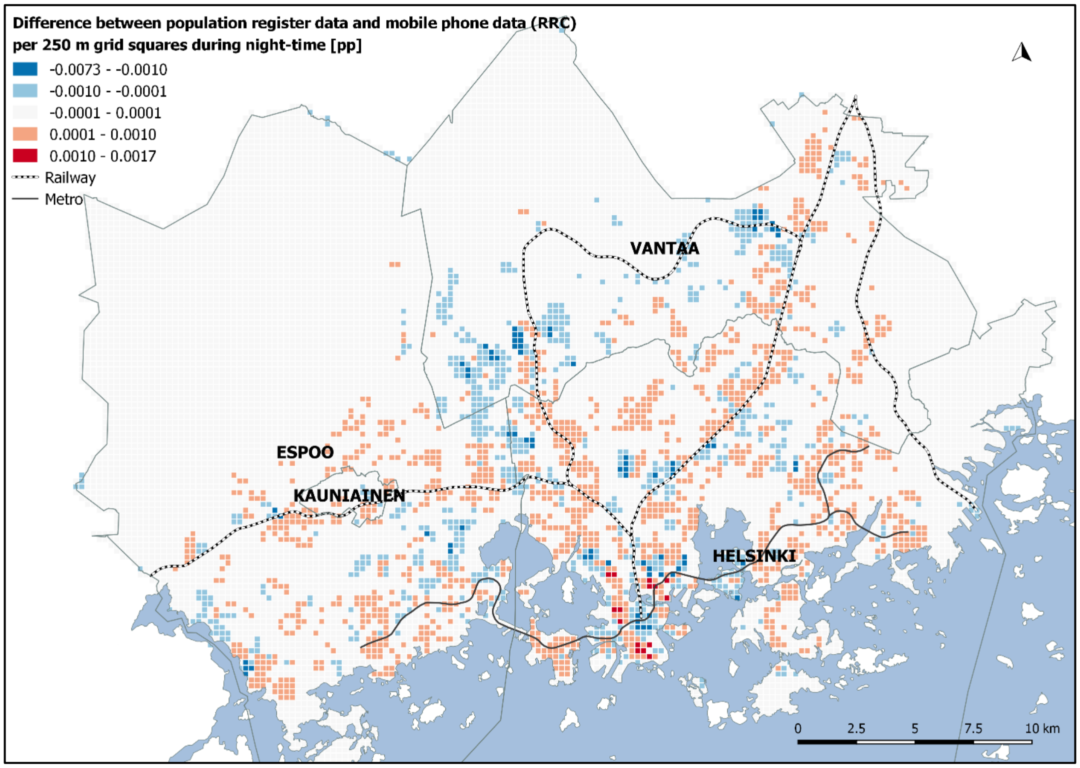
		AW_H12	AW_H13	AW_H14	AW_H15	AW_H16	AW_H17	AW_H18	AW_H19	AW_H20	AW_H21	AW_H22	AW_H23
pop_norm	Pearson Correlation	,329**	,327**	,332**	,335**	,351**	,381**	,421**	,463**	,507**	,546**	,570**	,573**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231
workp_norm	Pearson Correlation	,608**	,608**	,607**	,595**	,577**	,559**	,538**	,520**	,495**	,462**	,435**	,425**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231	13231

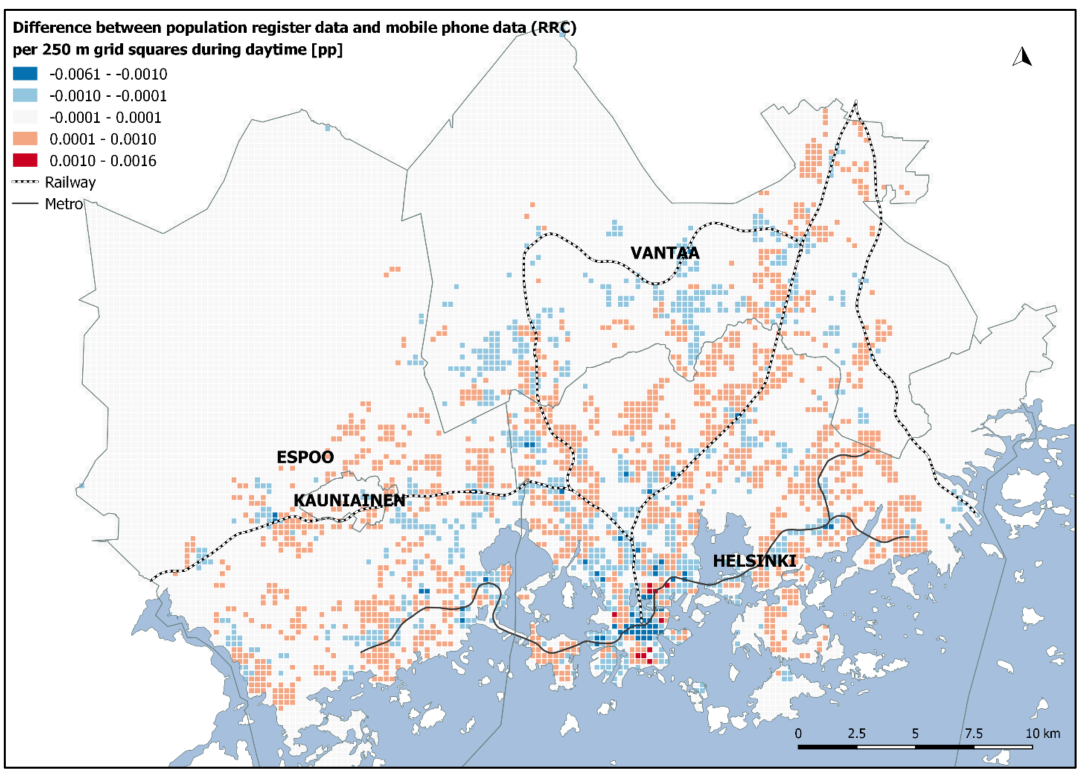
**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 5. The correlation (Pearson) between the population distribution based on the MFD and AW methods during night-time (left) and daytime (right).

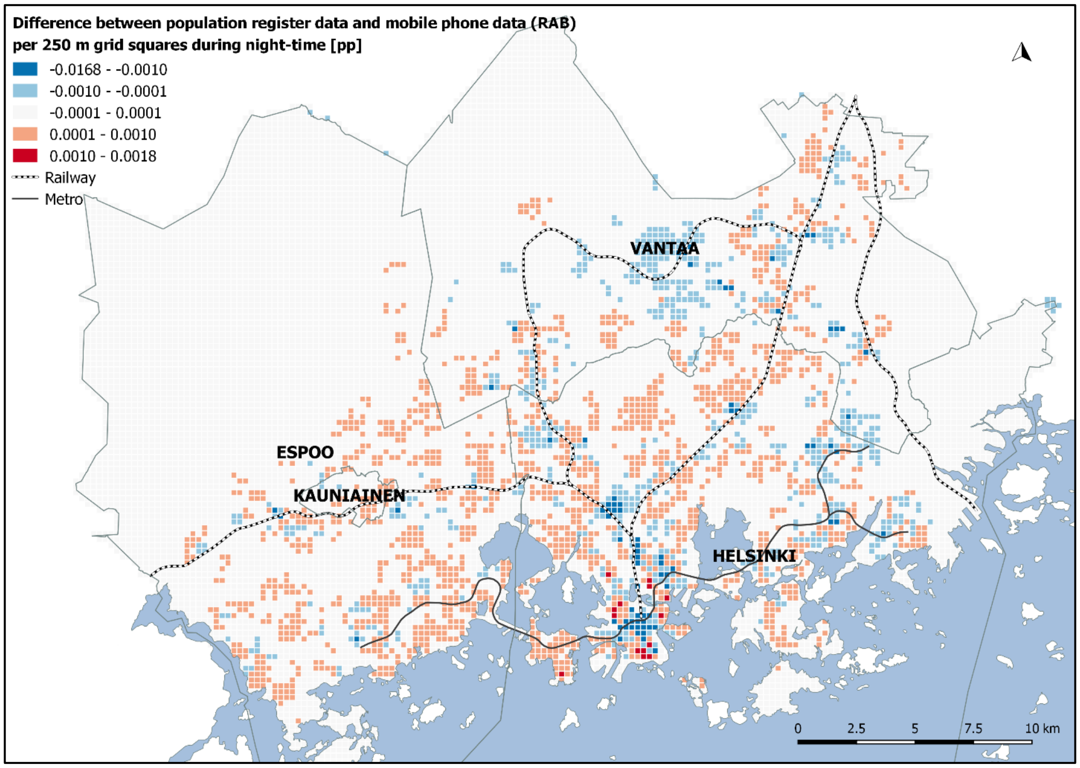


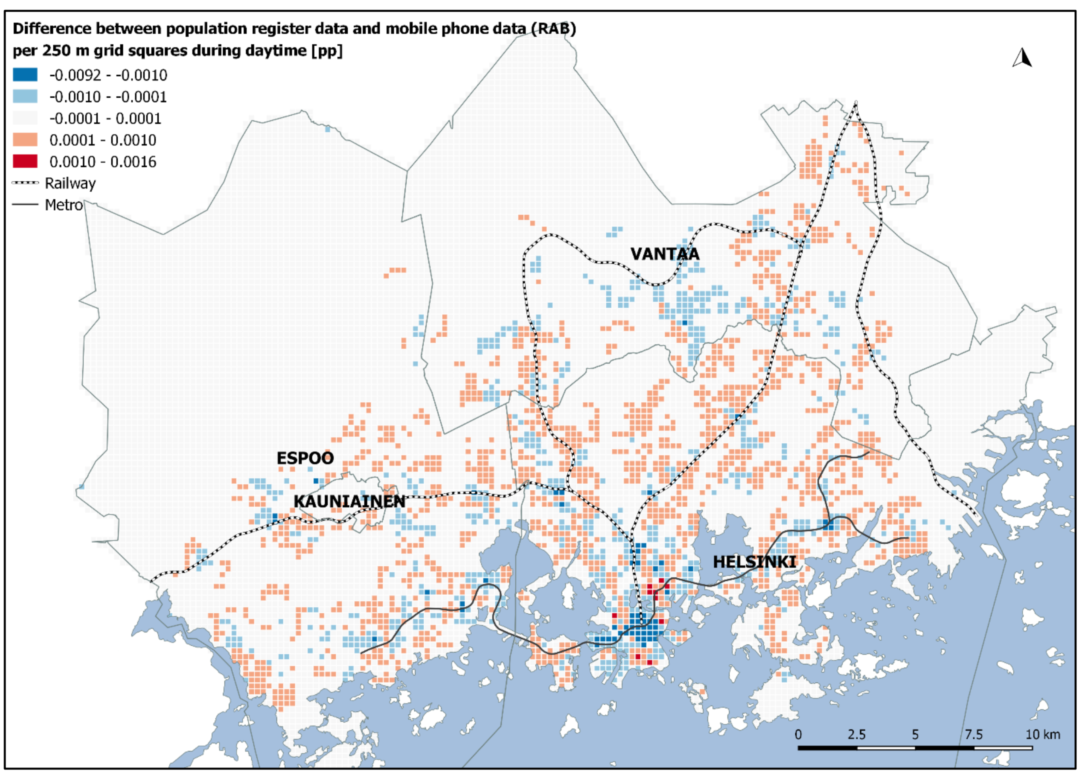
Appendix 6. Difference maps between population register data and interpolated mobile phone data (RRC connections) during night-time (2 AM –5 AM) (upper) and daytime (3 PM – 4 PM) (lower).



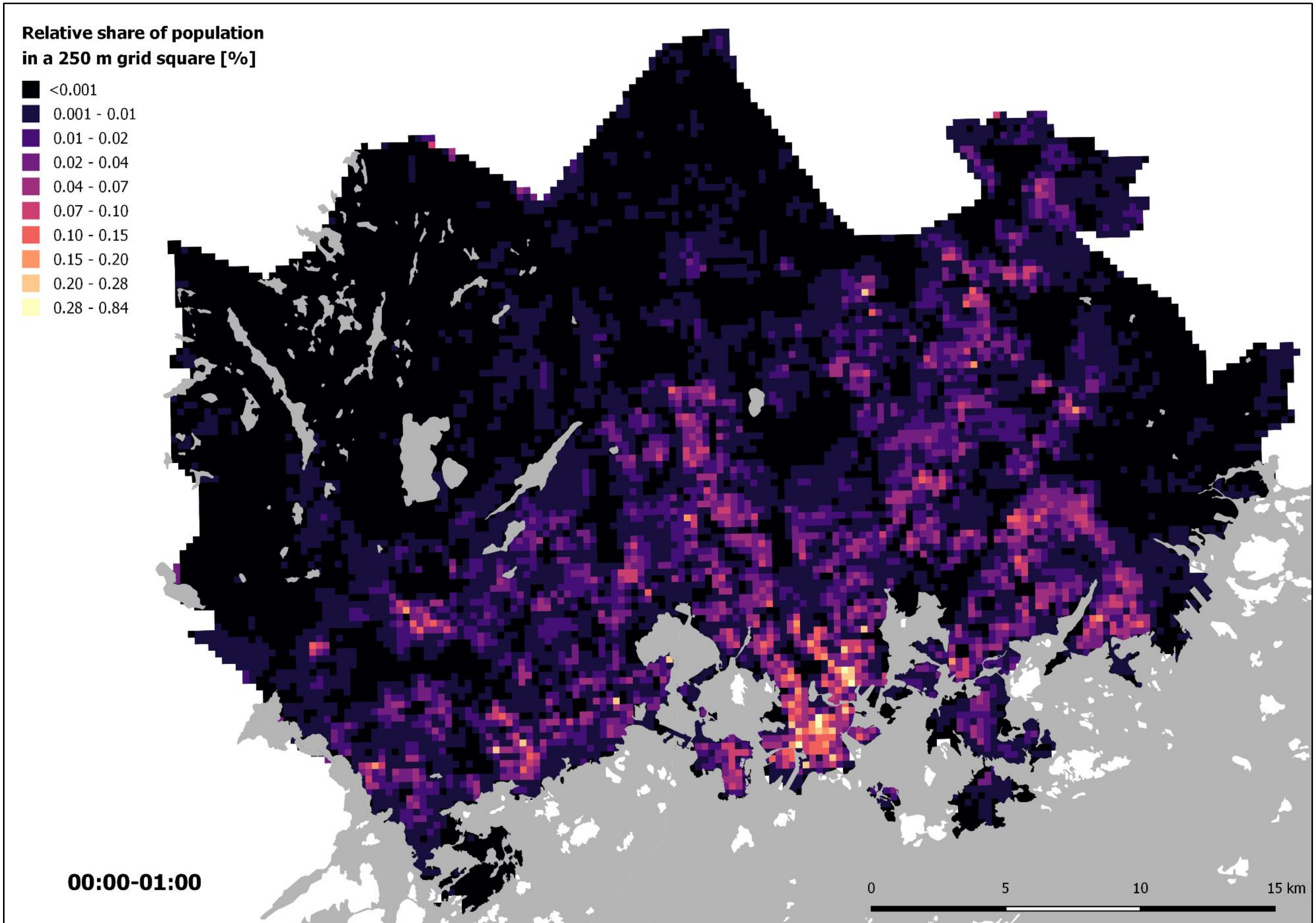


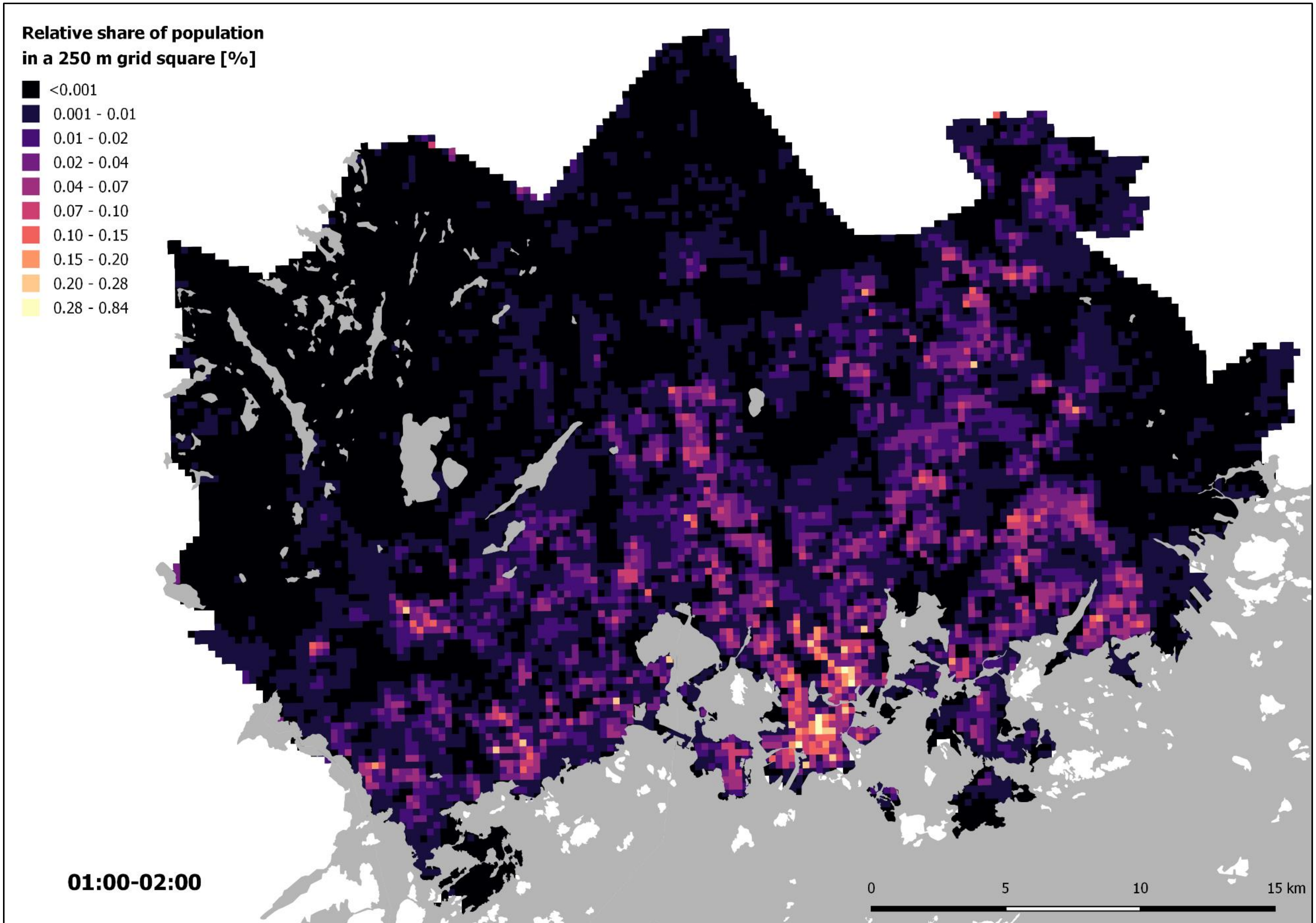
Appendix 7. Difference maps between population register data and interpolated mobile phone data (RAB attempts) during night-time (2 AM –5 AM) (upper) and daytime (3 PM – 4 PM) (lower).

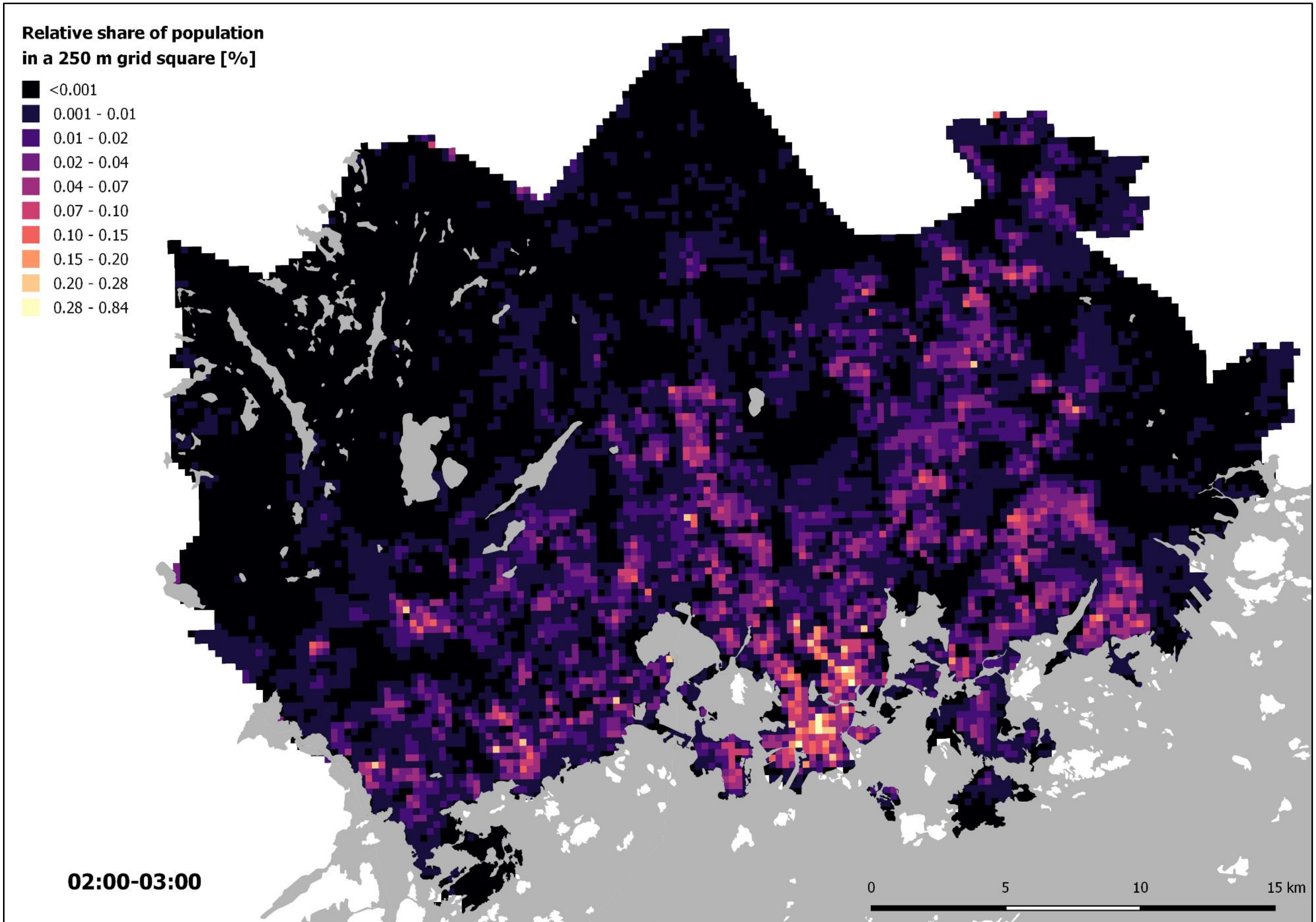


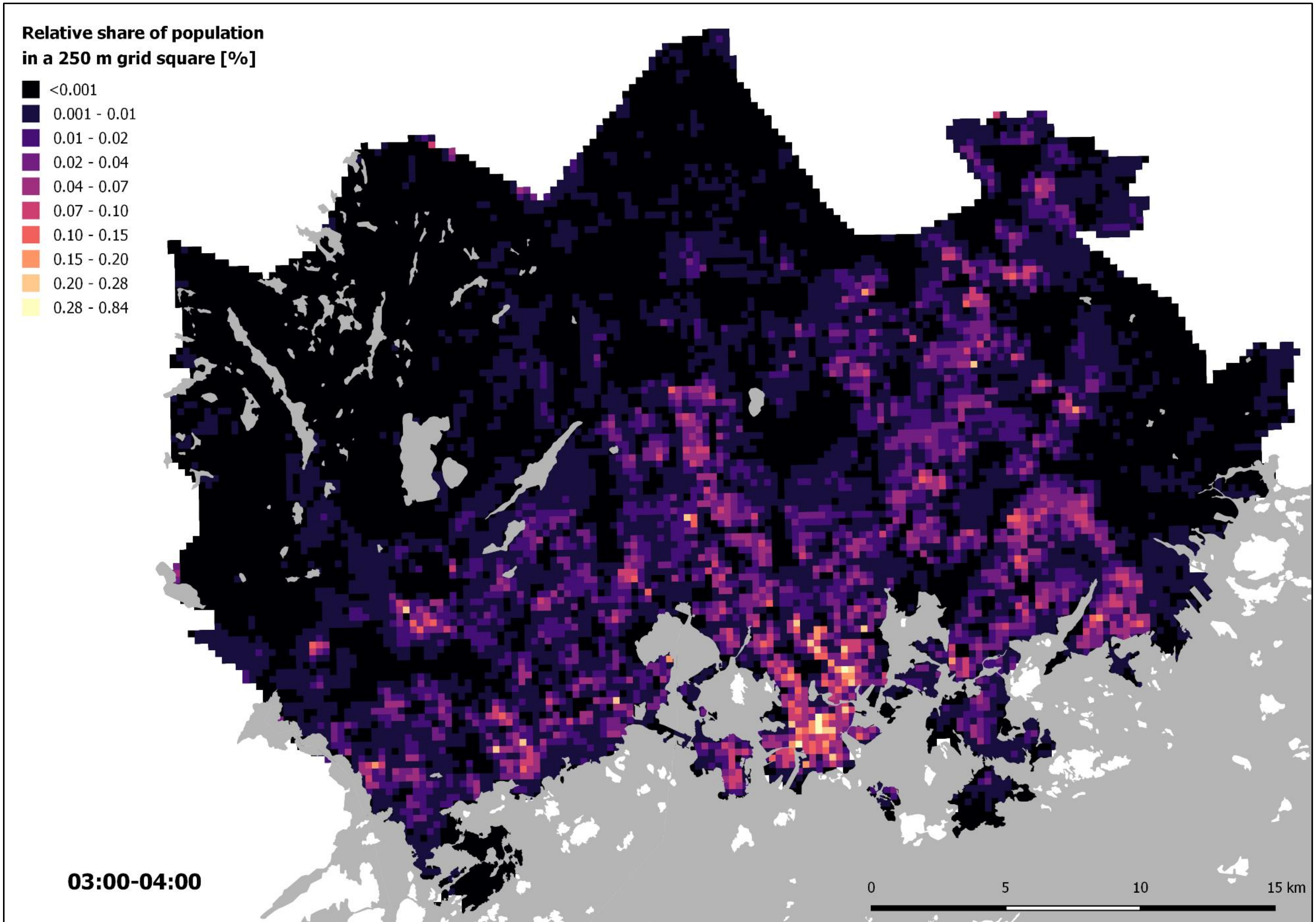


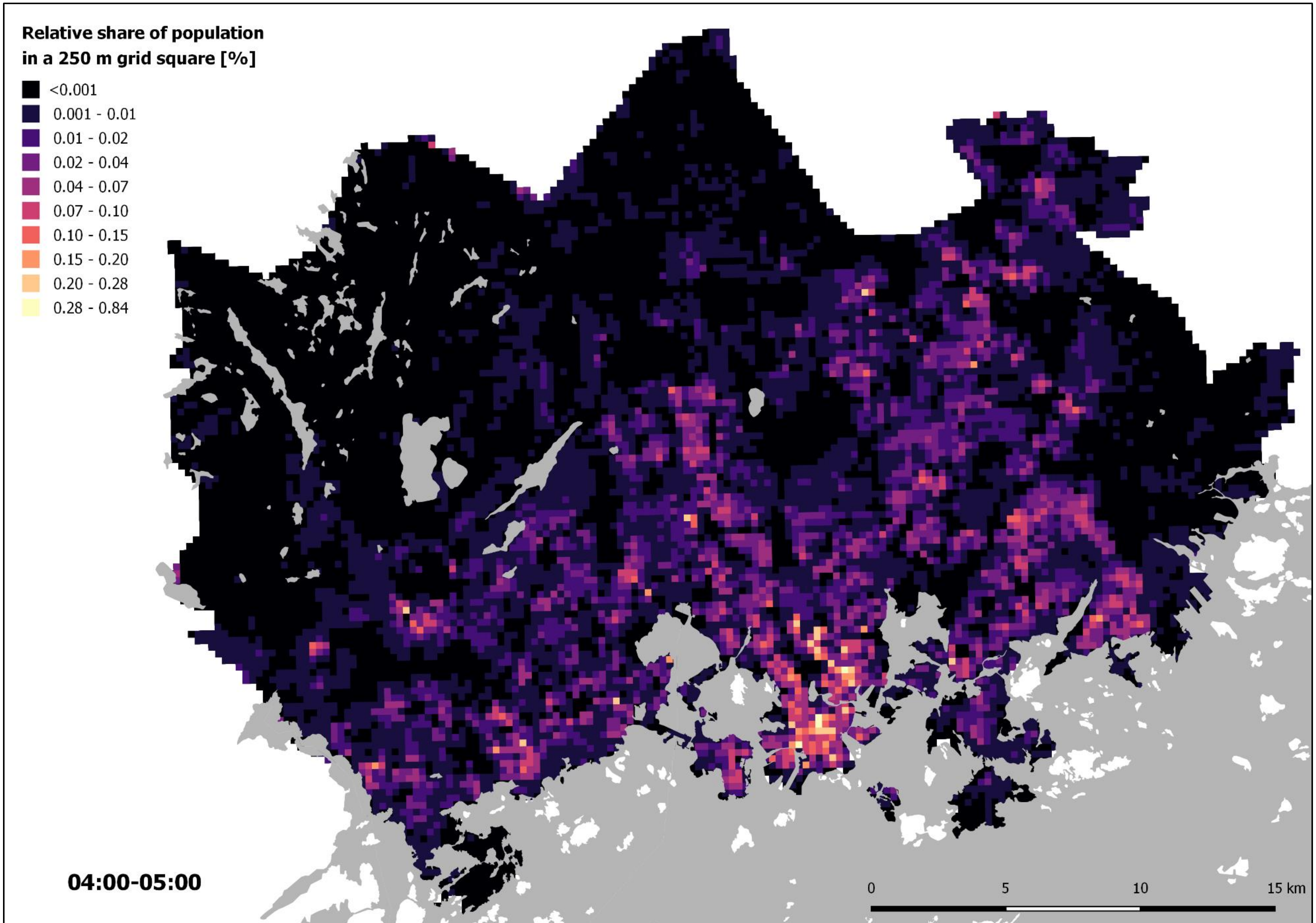
Appendix 8. The estimated hourly distribution of people on 250 m x 250 m statistical grid cells. The values are based on mobile phone data (HSPA calls) interpolated using the MFD method on an average weekday (Monday-Thursday).

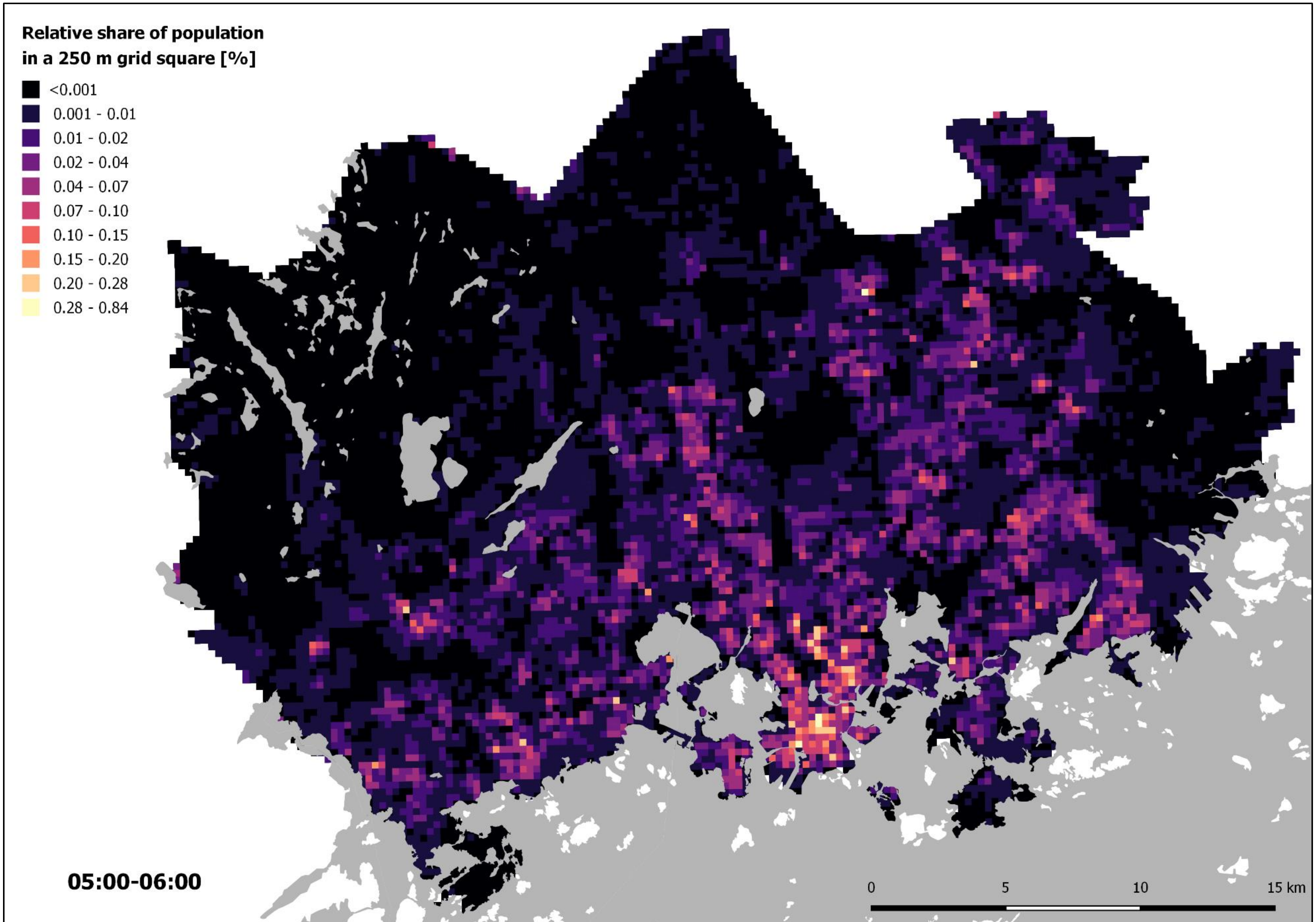


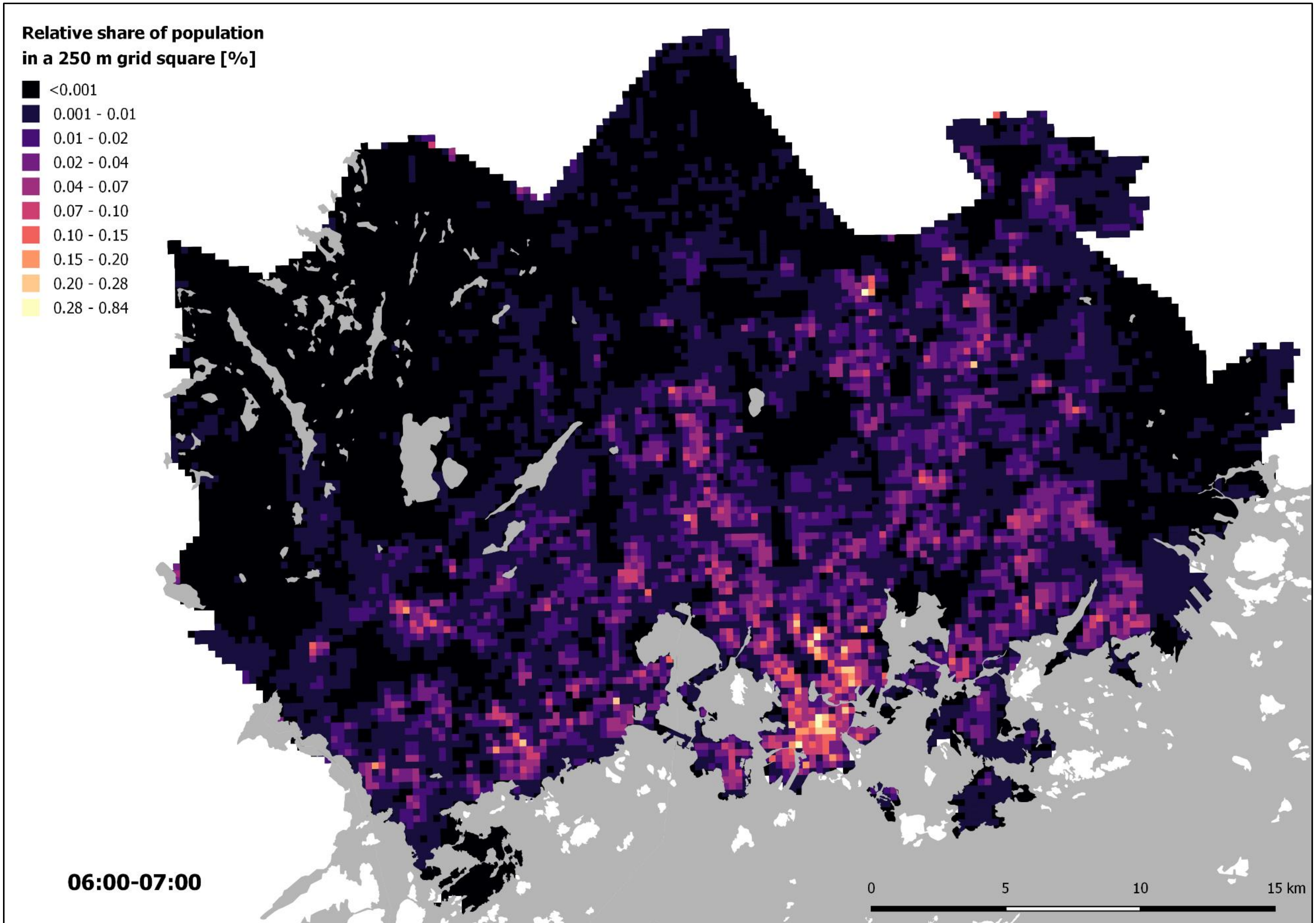

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
01:00-02:00
0
5
10
15 km

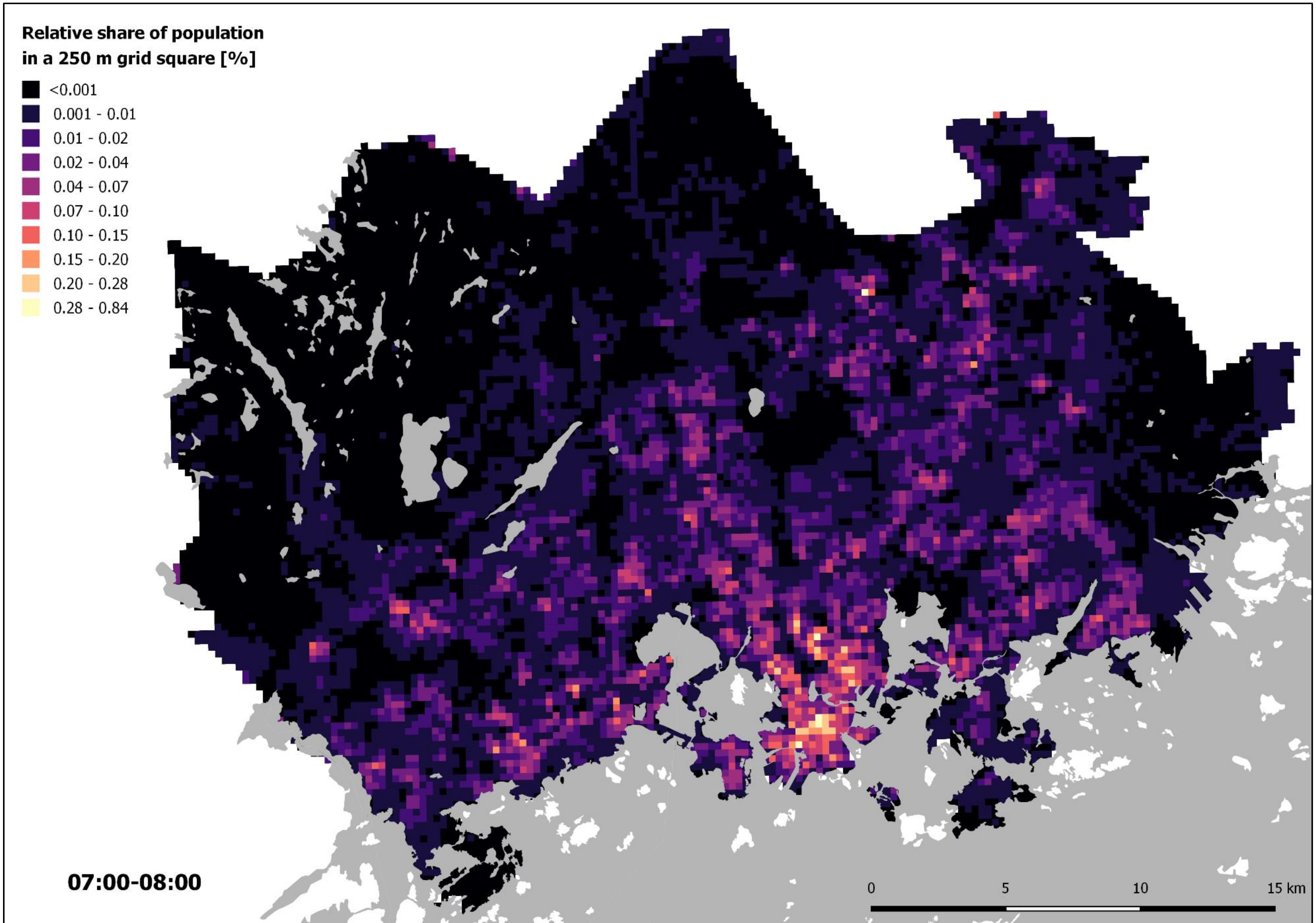

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
02:00-03:00
0
5
10
15 km

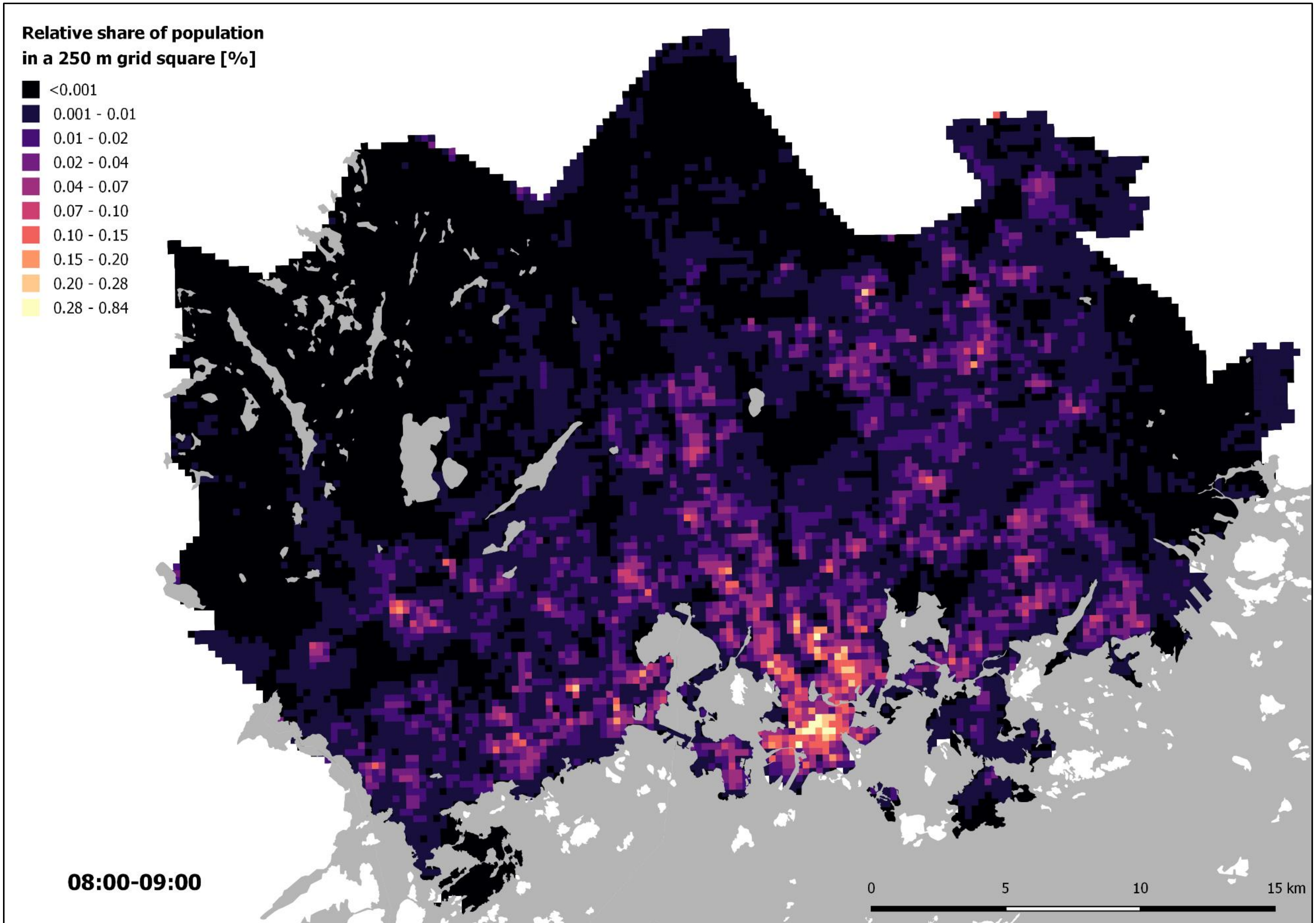

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
03:00-04:00
0
5
10
15 km

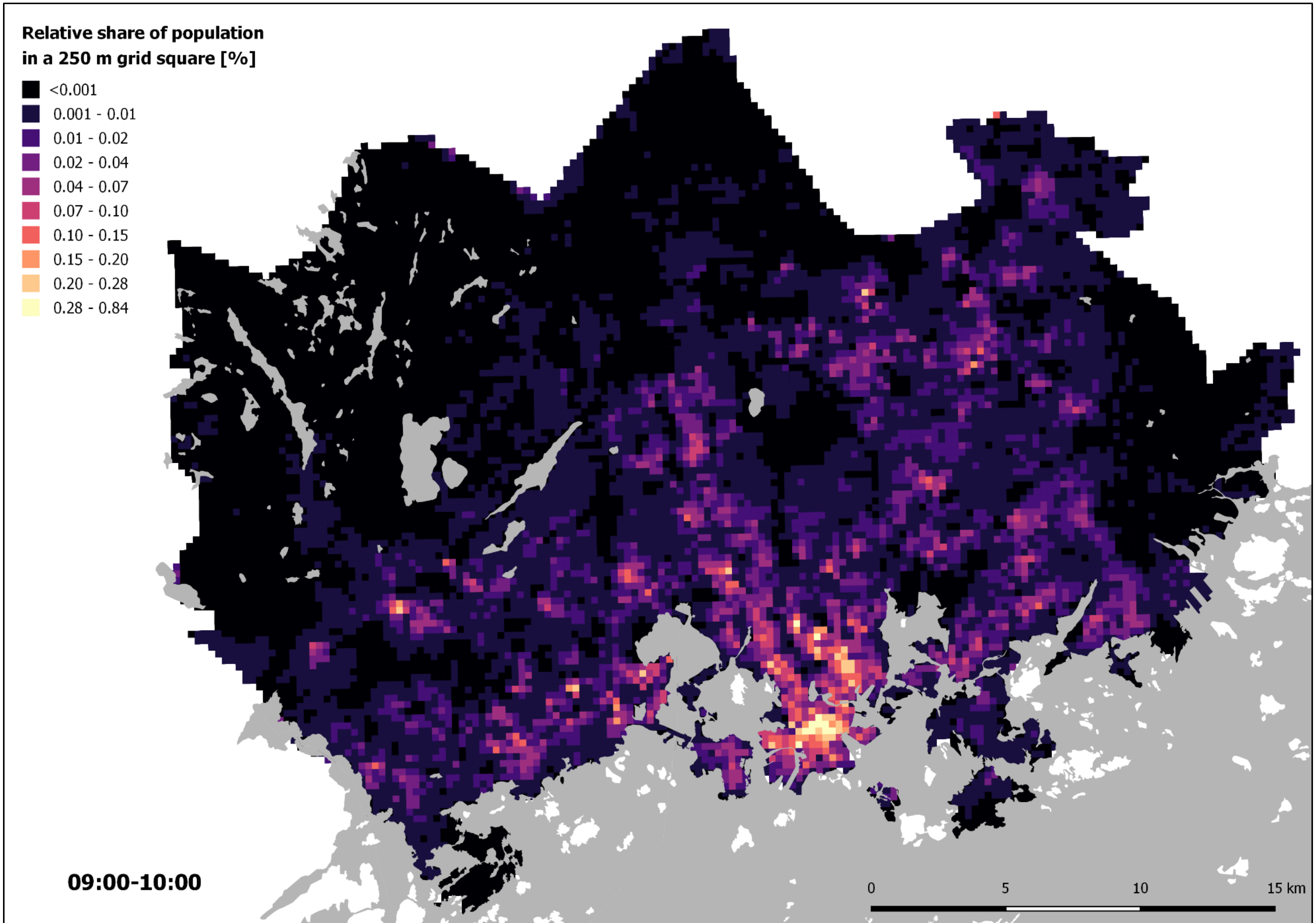

Relative share of population in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
04:00-05:00
0
5
10
15 km

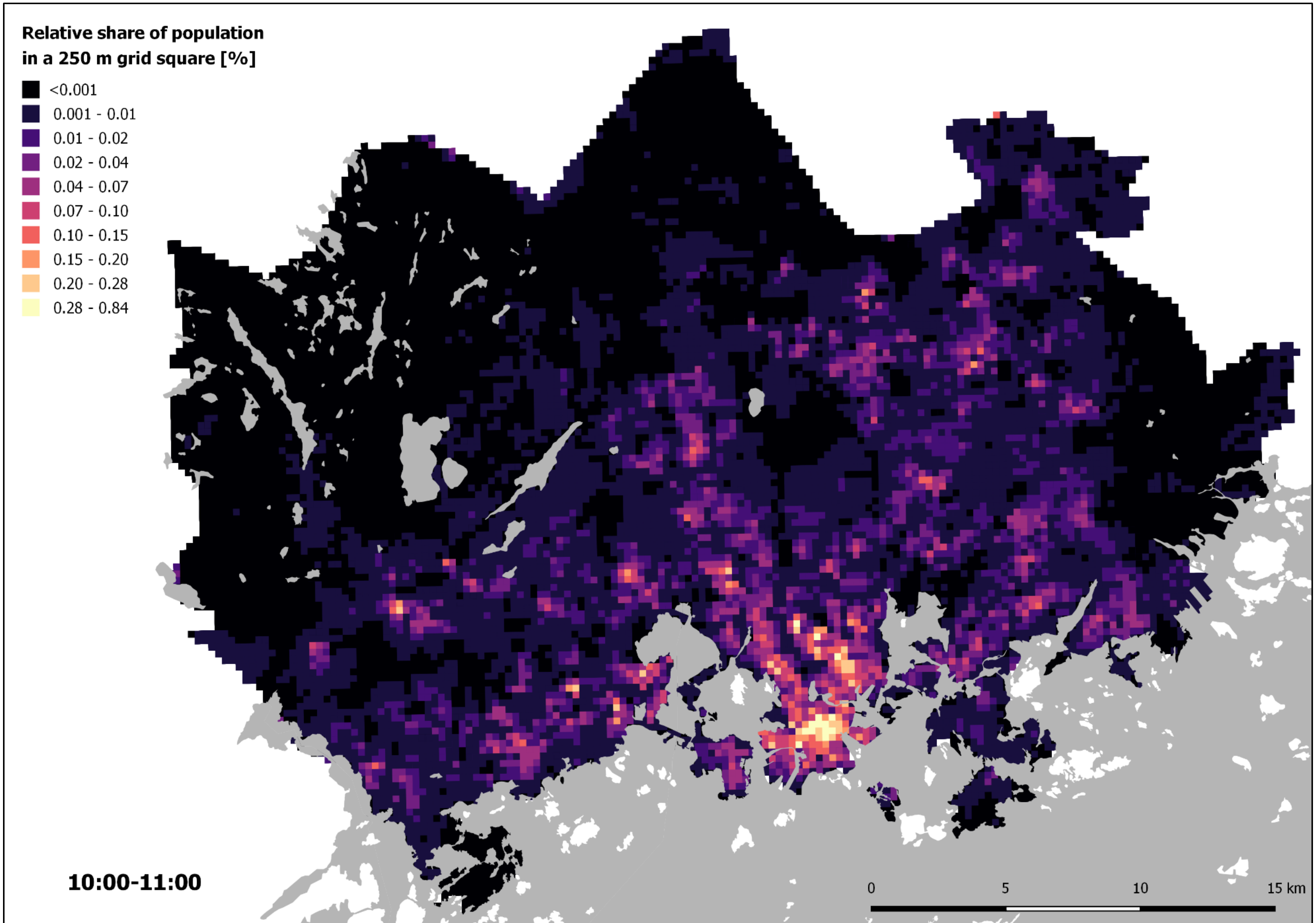

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
05:00-06:00
0
5
10
15 km


Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
06:00-07:00
0
5
10
15 km

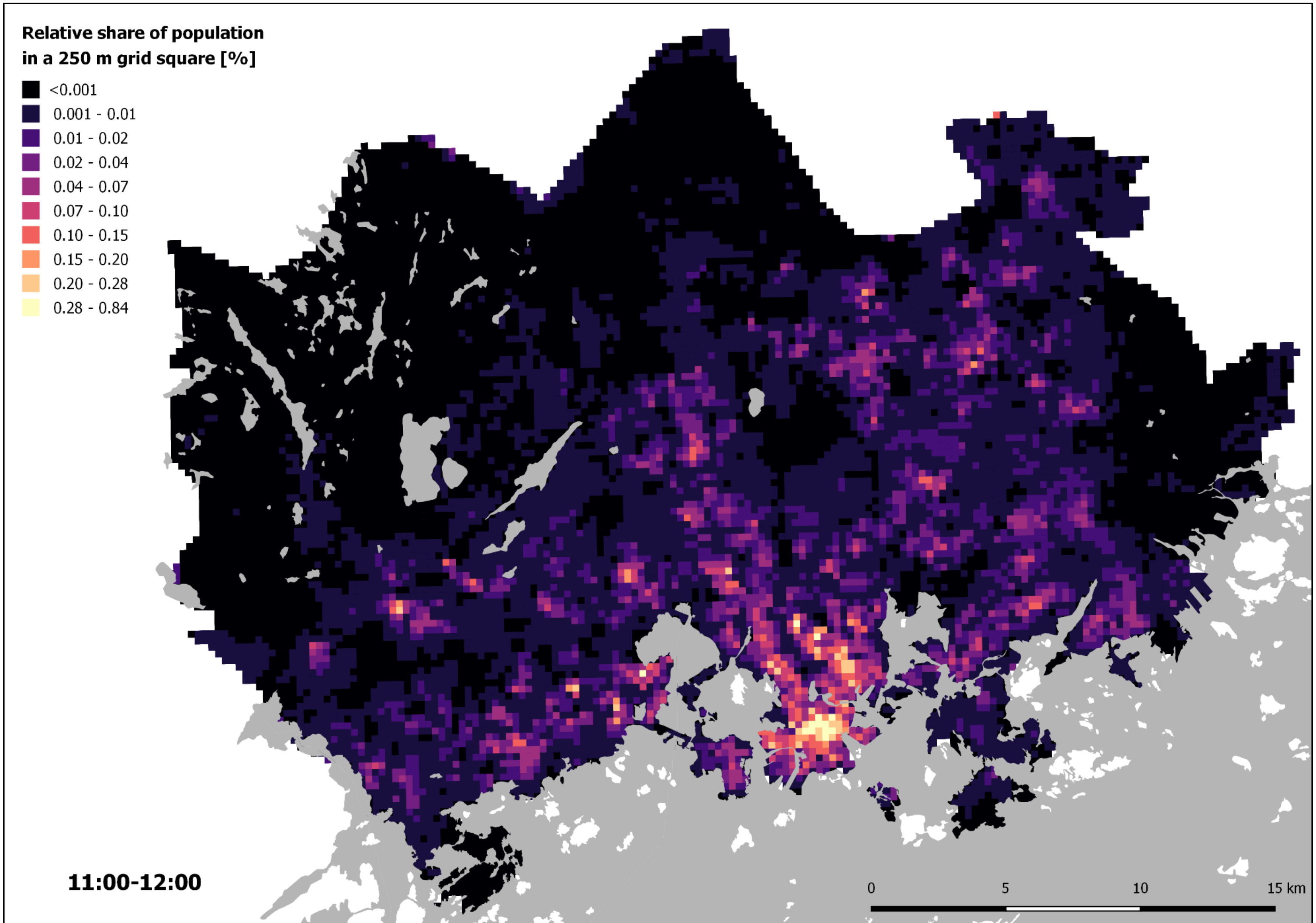

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
07:00-08:00
0
5
10
15 km

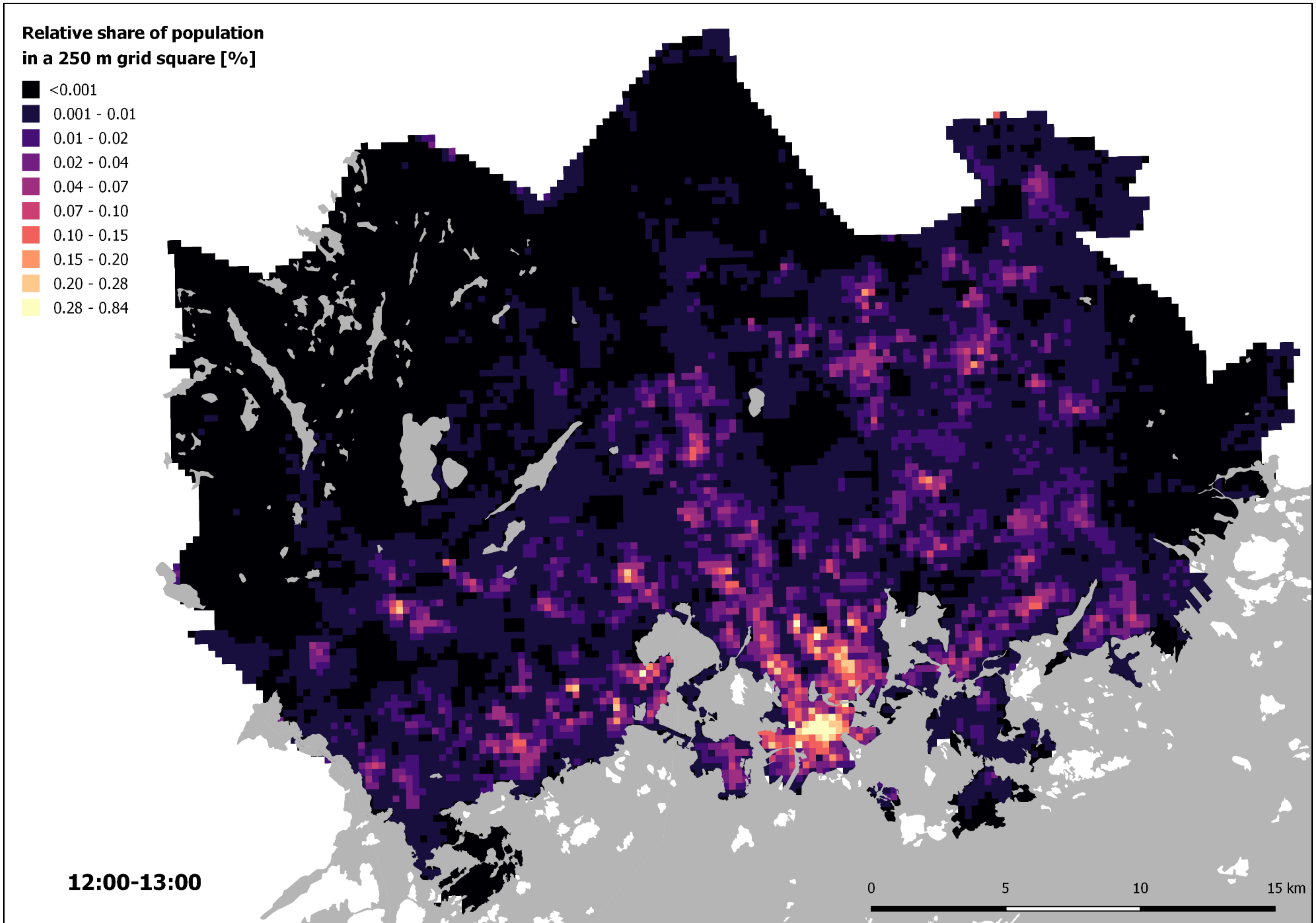

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
08:00-09:00
0
5
10
15 km

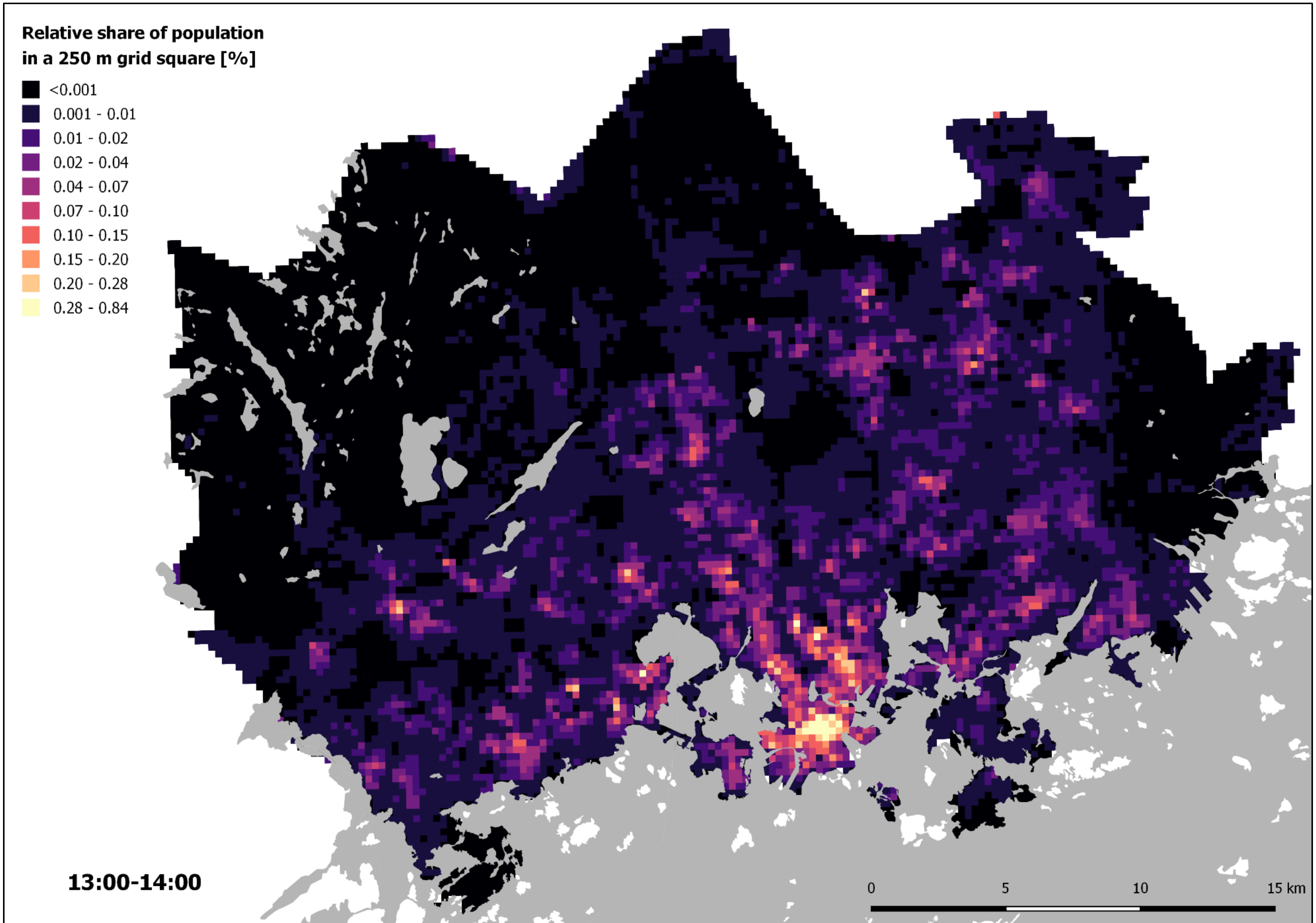

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
09:00-10:00
0
5
10
15 km

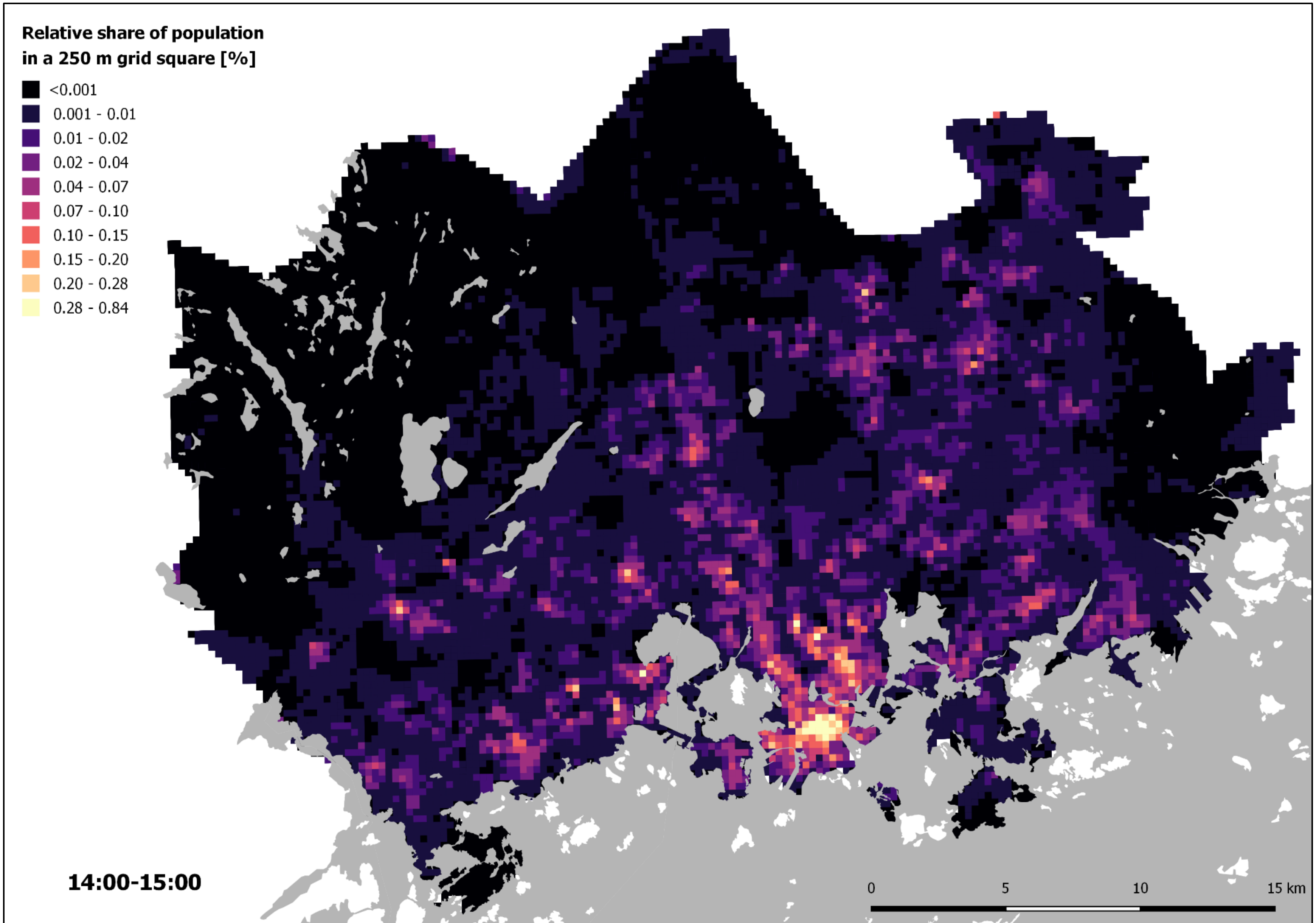


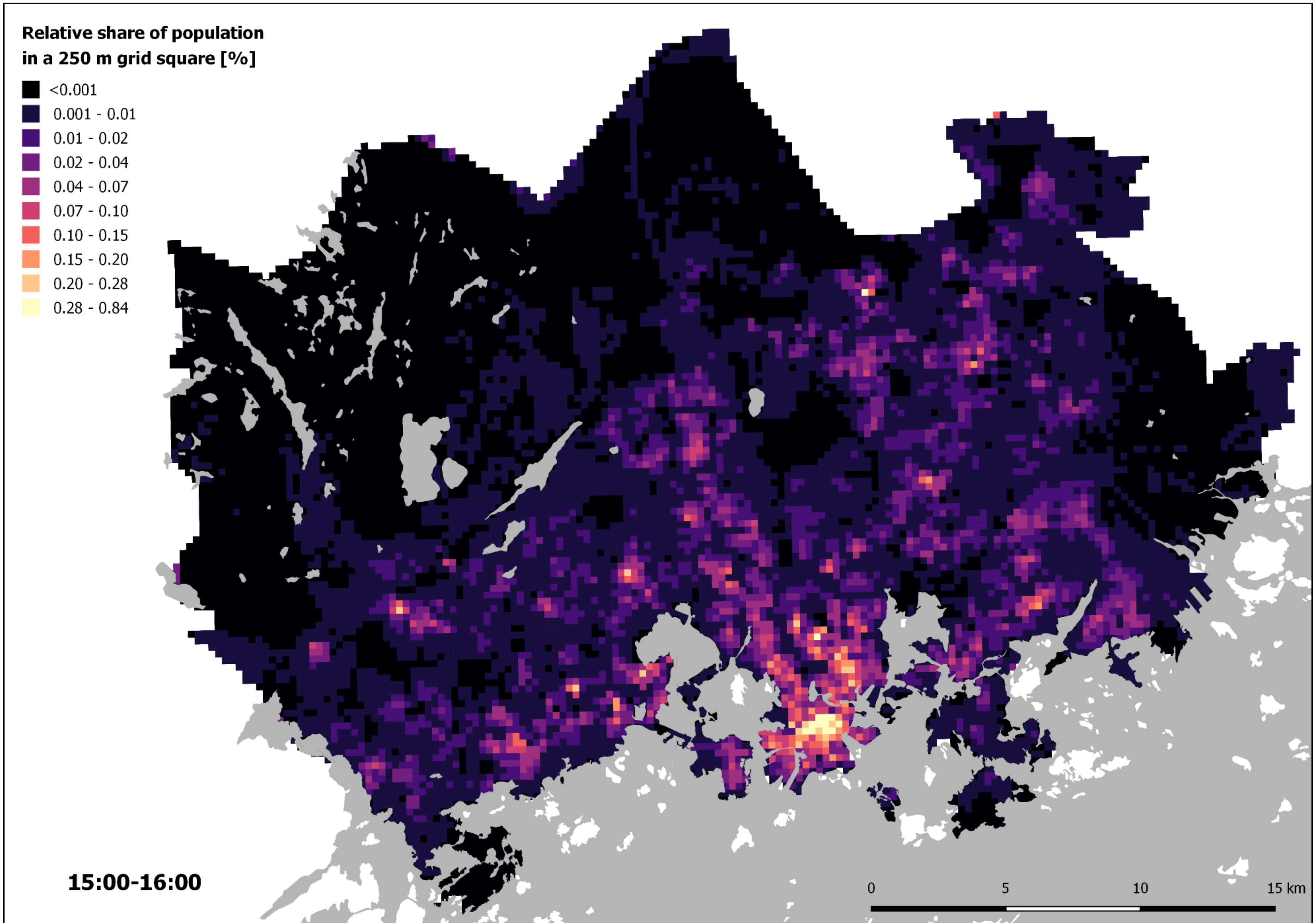
Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
10:00-11:00
0
5
10
15 km

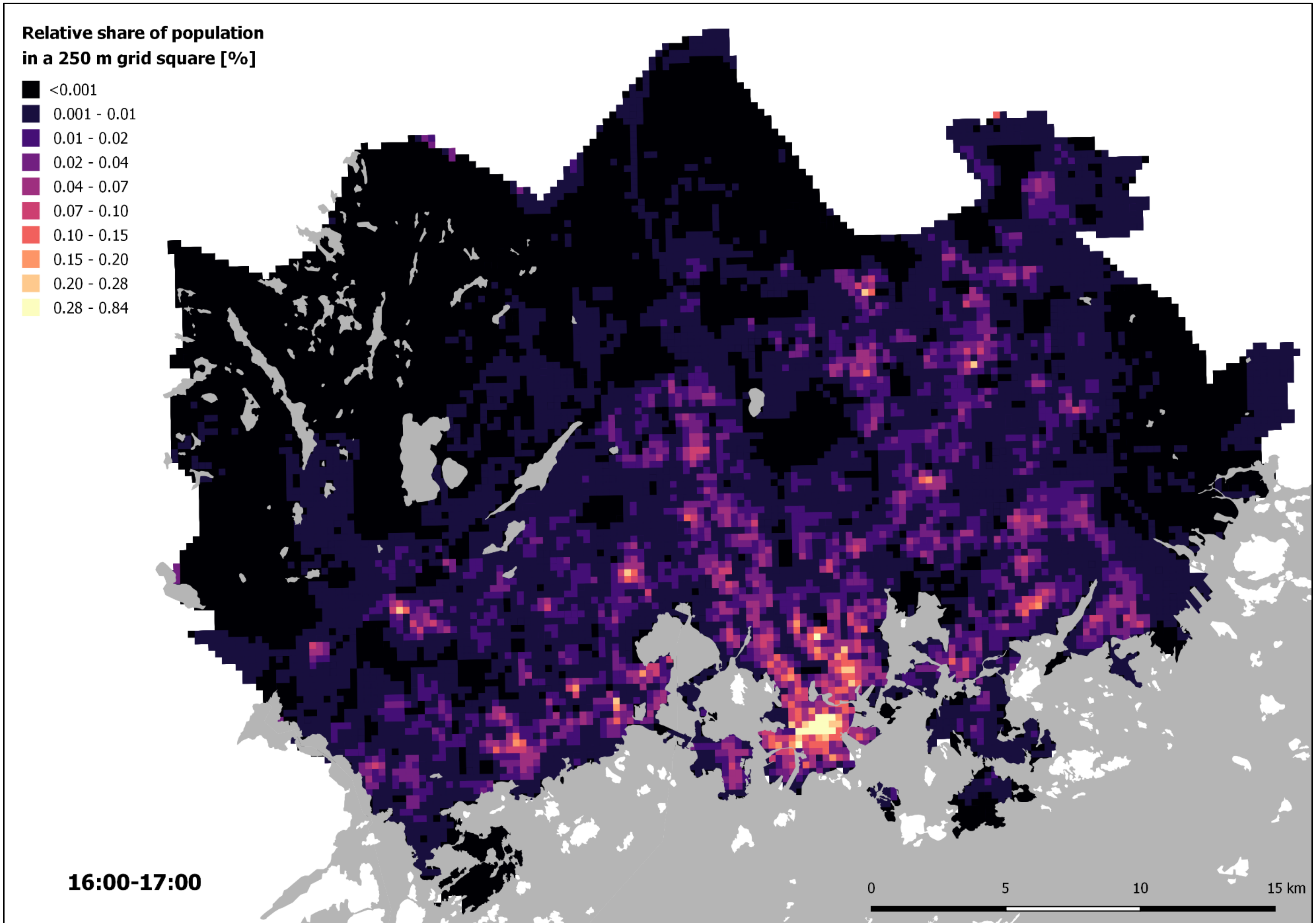

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
11:00-12:00
0
5
10
15 km

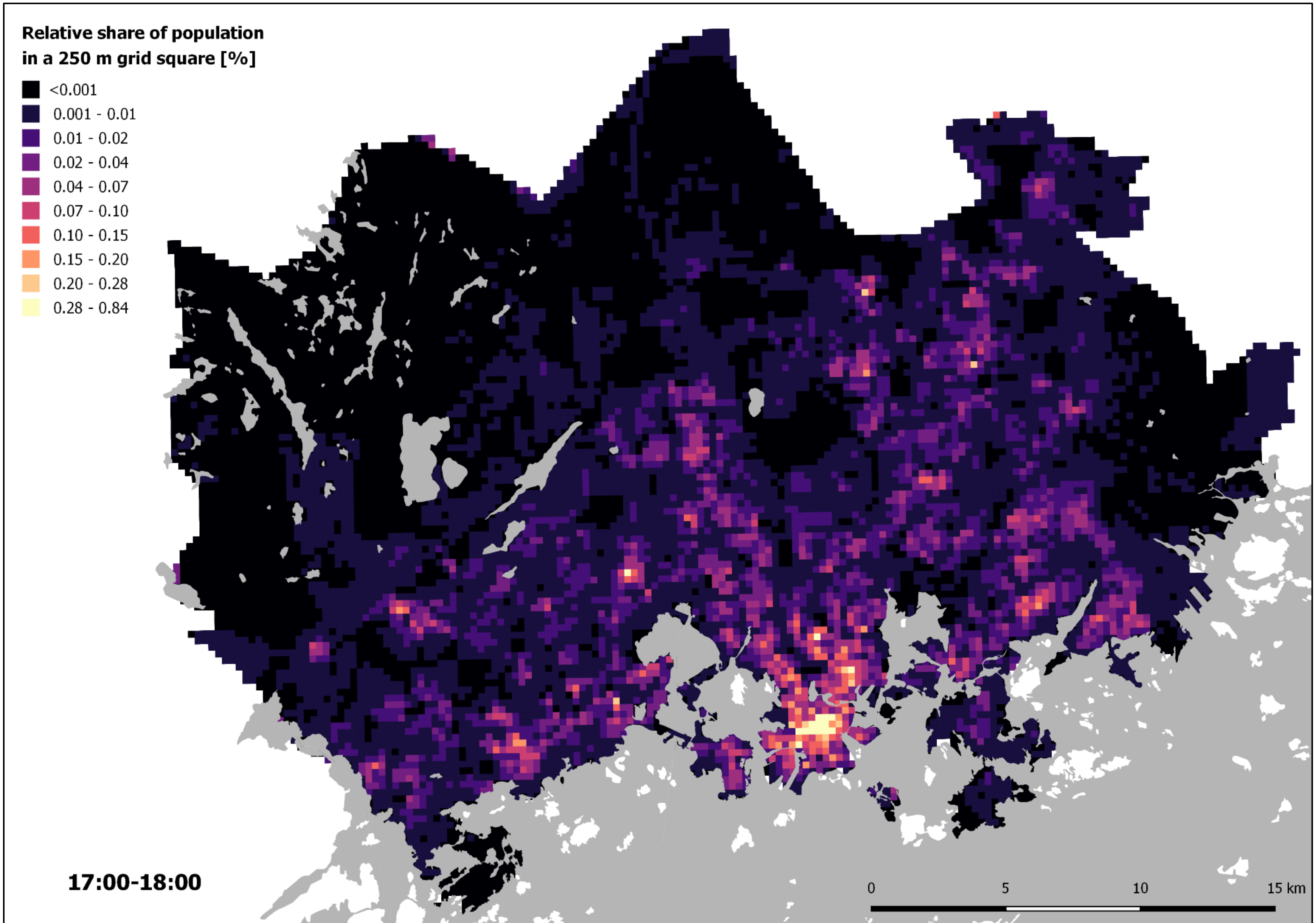

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
12:00-13:00
0
5
10
15 km

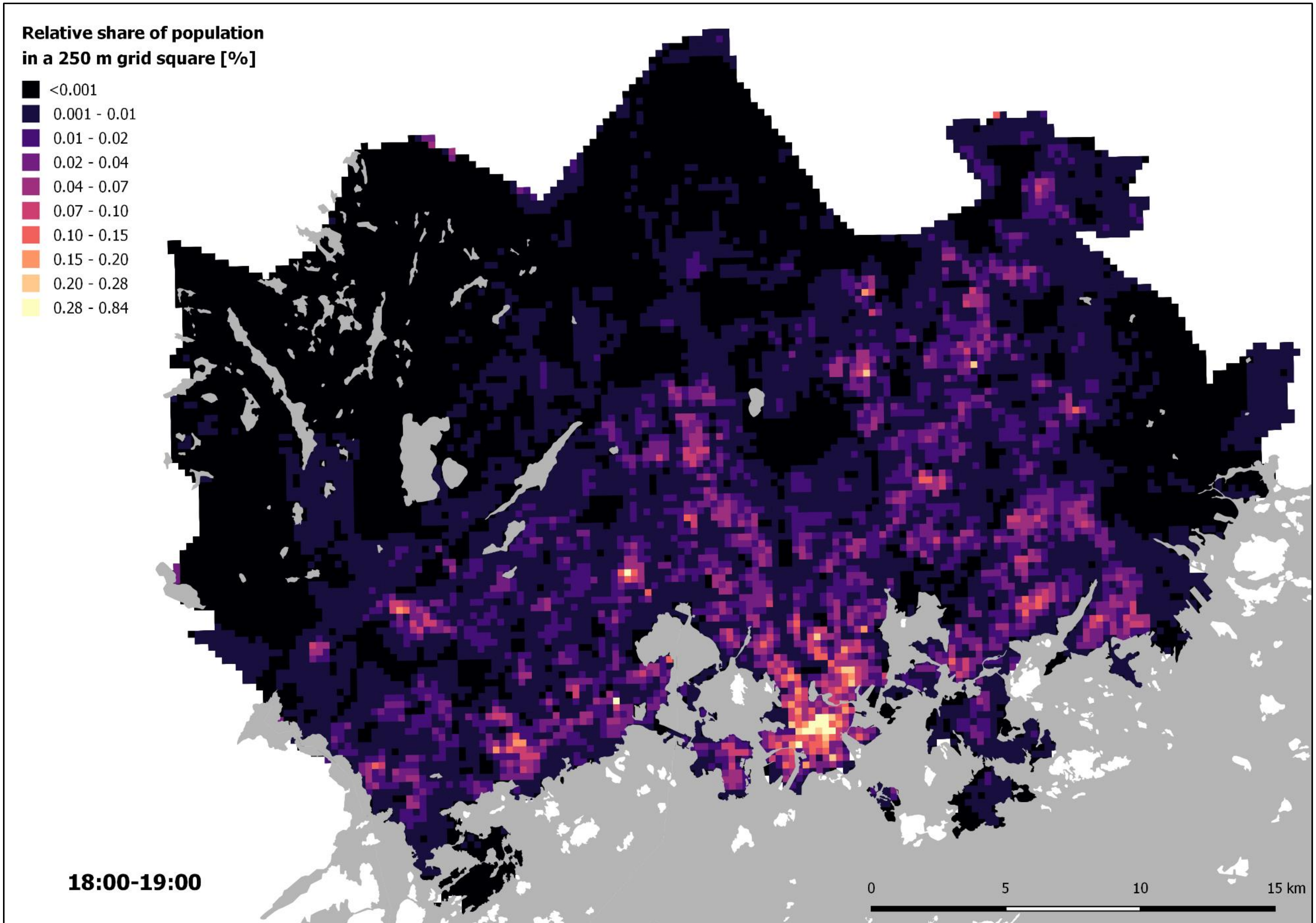

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
13:00-14:00
0
5
10
15 km


Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
14:00-15:00
0
5
10
15 km

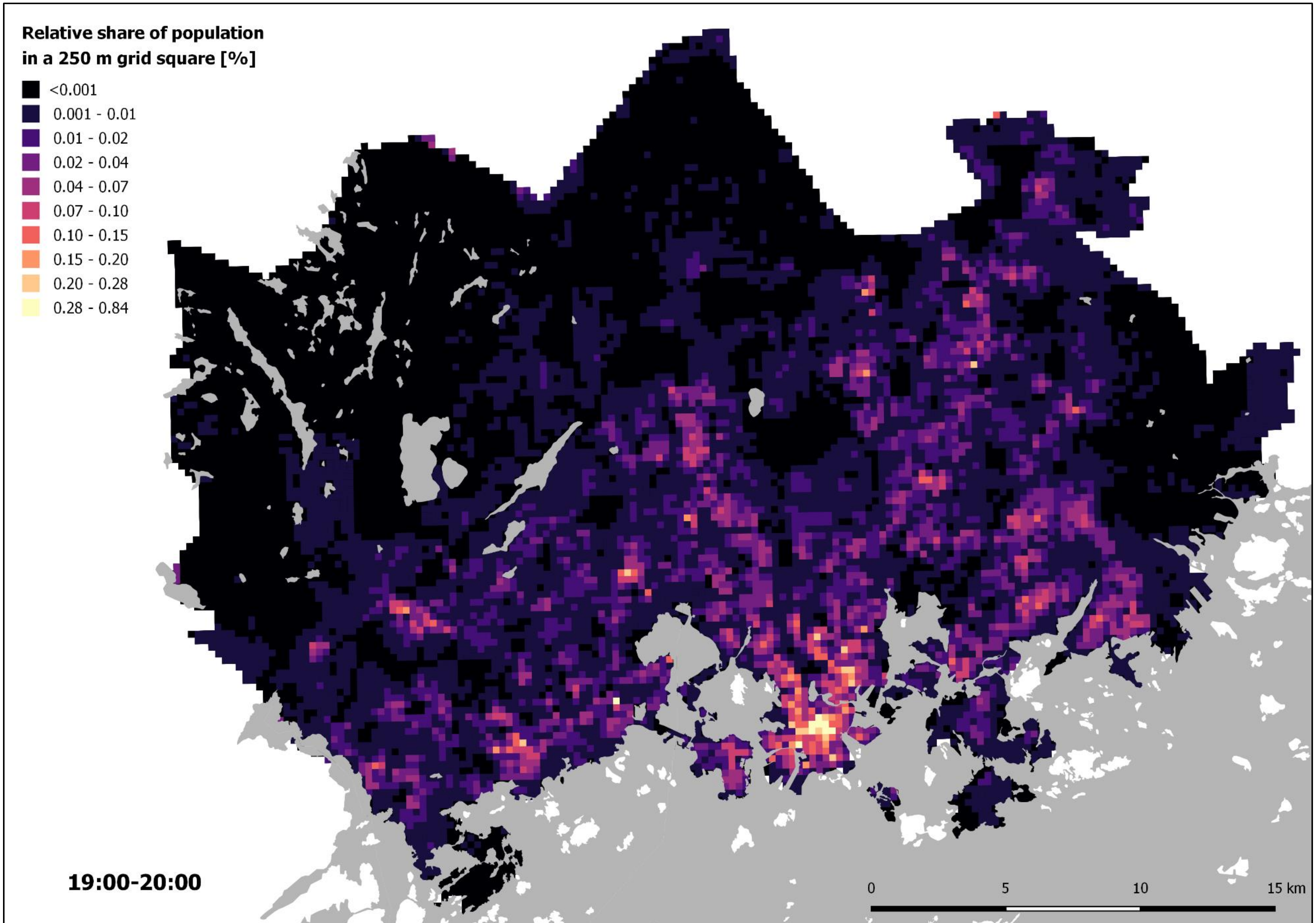

Relative share of population in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
15:00-16:00
0
5
10
15 km

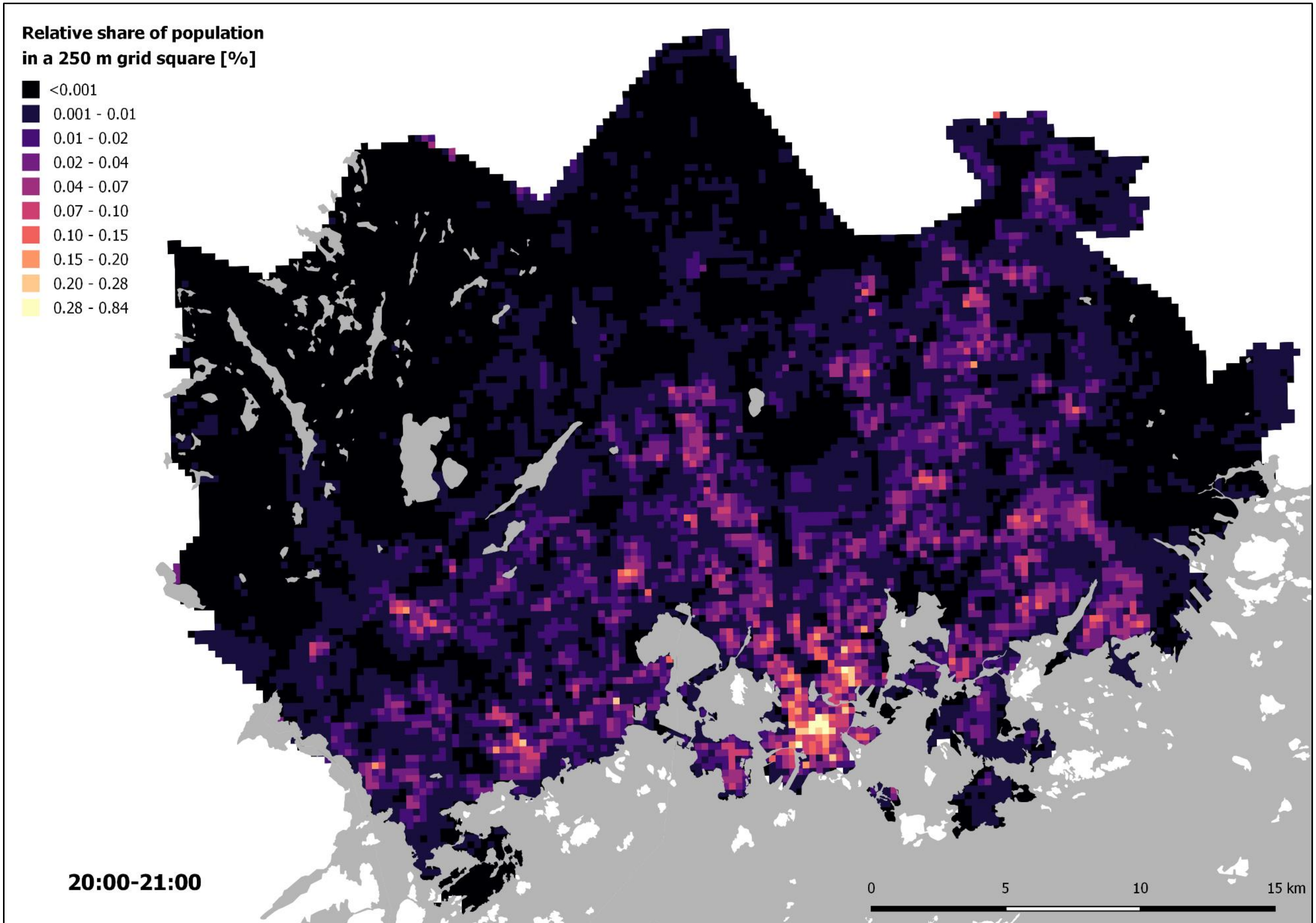

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
16:00-17:00
0
5
10
15 km

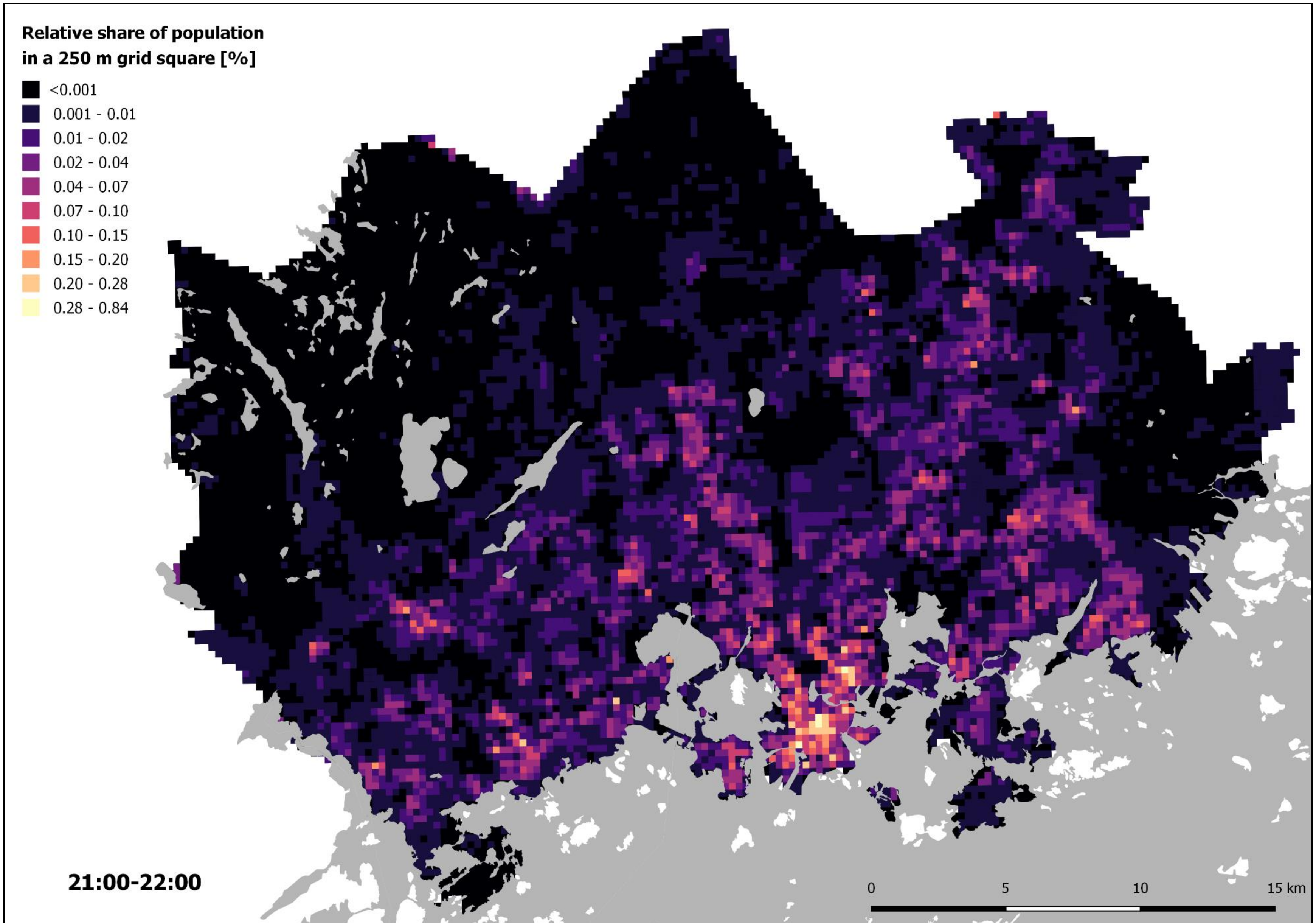

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
17:00-18:00
0
5
10
15 km



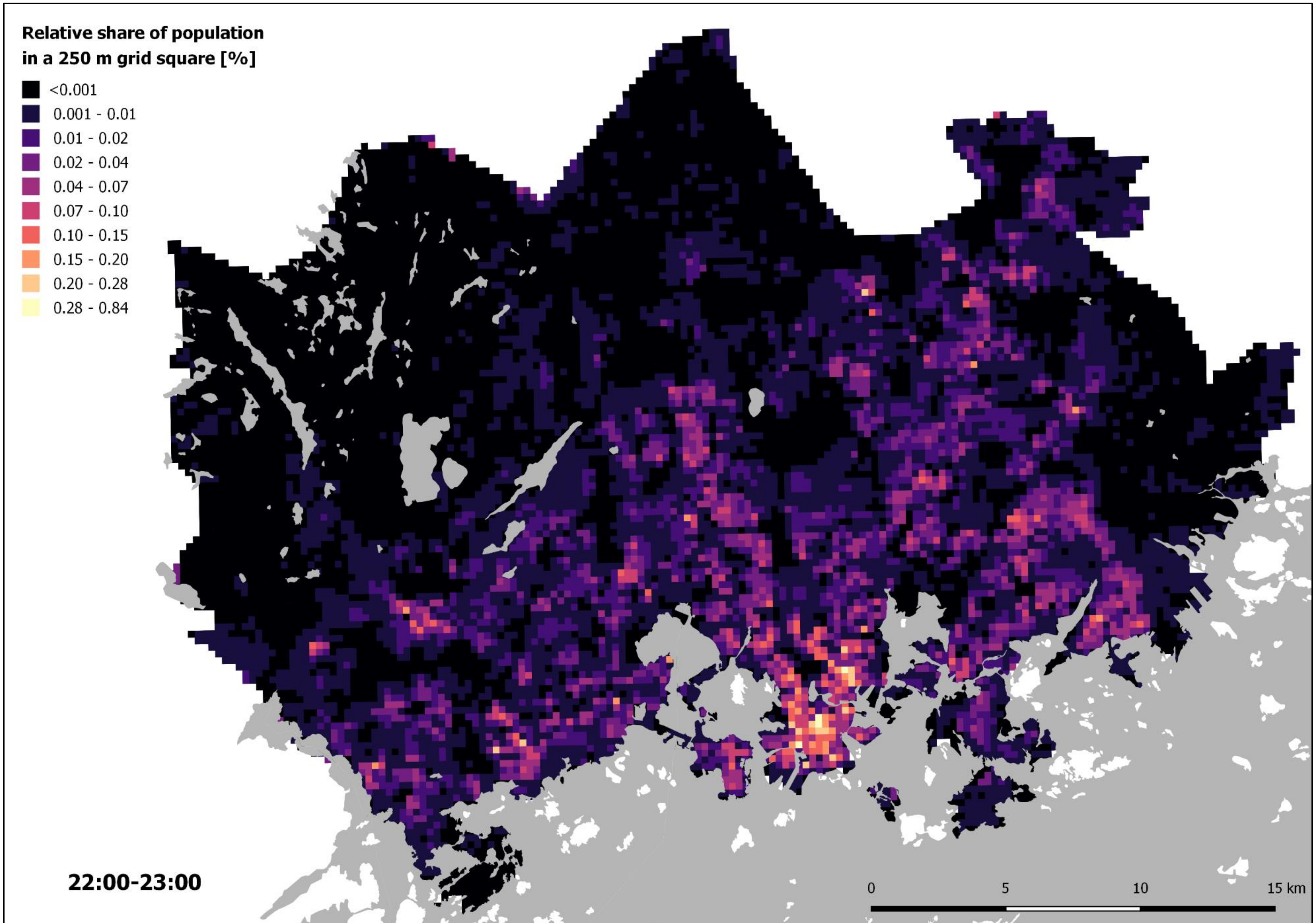
Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
18:00-19:00
0
5
10
15 km


Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
19:00-20:00
0
5
10
15 km

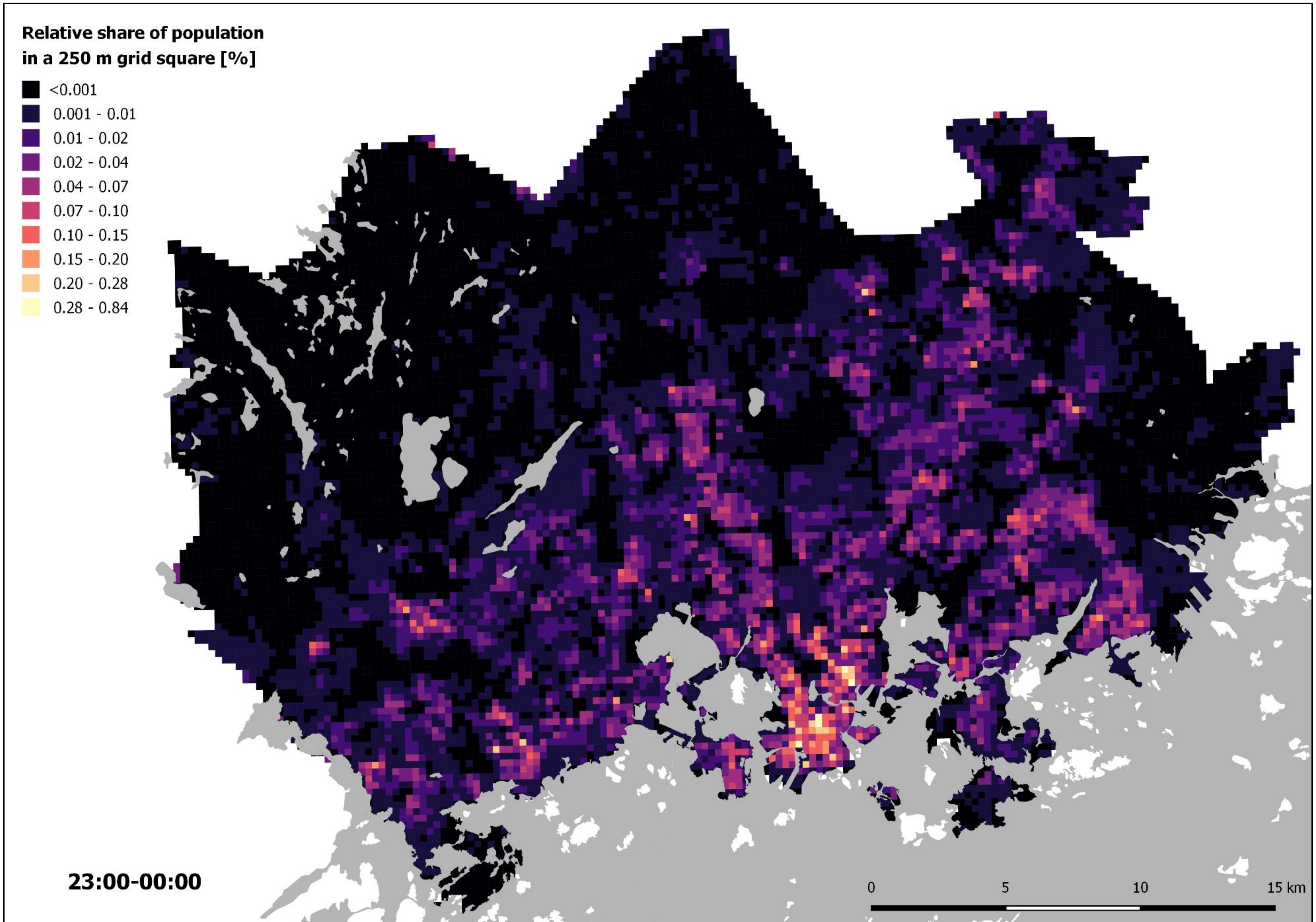

Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
20:00-21:00
0
5
10
15 km



Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
21:00-22:00
0
5
10
15 km

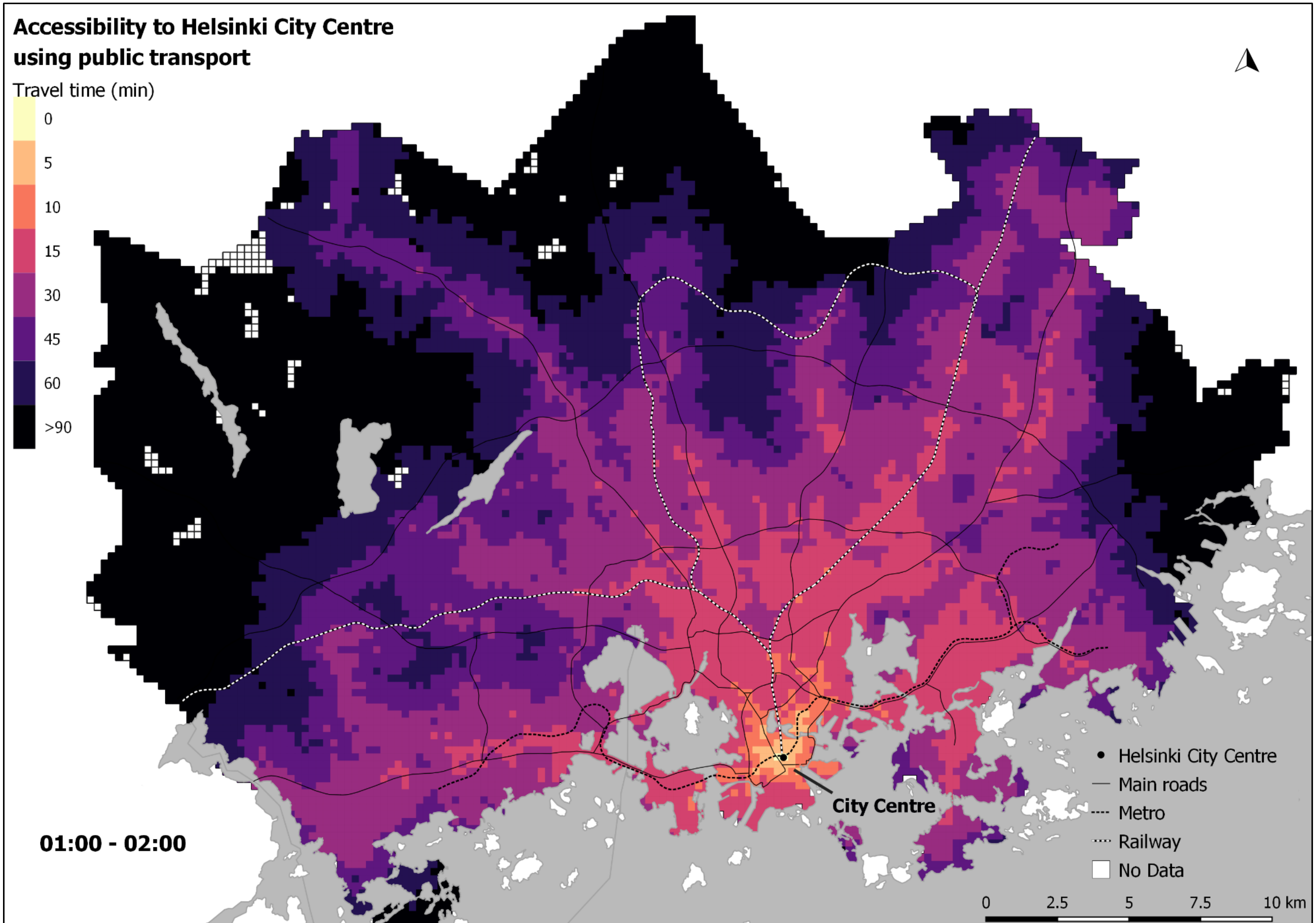


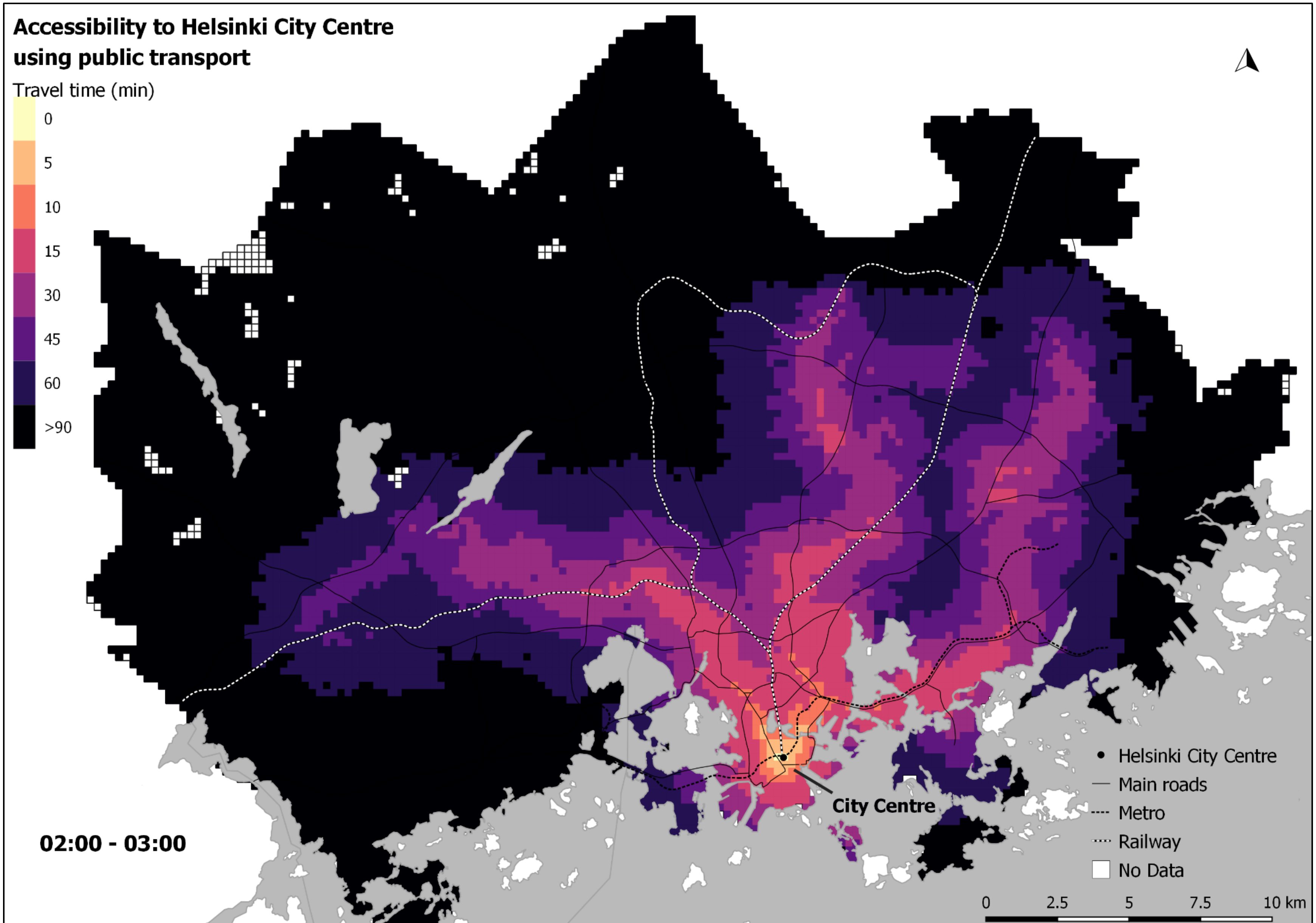
Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
22:00-23:00
0
5
10
15 km

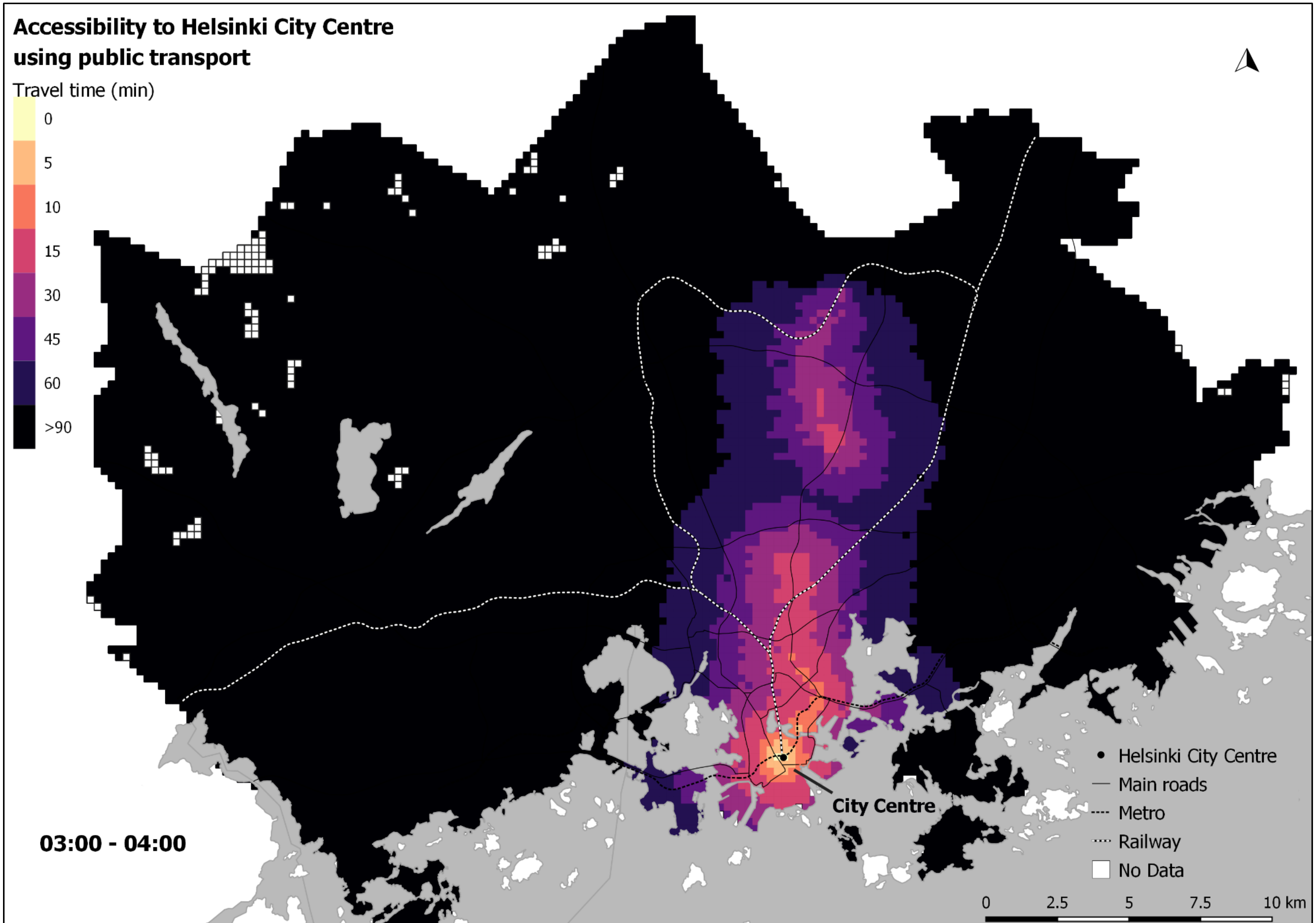


Relative share of population
in a 250 m grid square [%]
<0.001
0.001 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.28
0.28 - 0.84
23:00-00:00
0
5
10
15 km

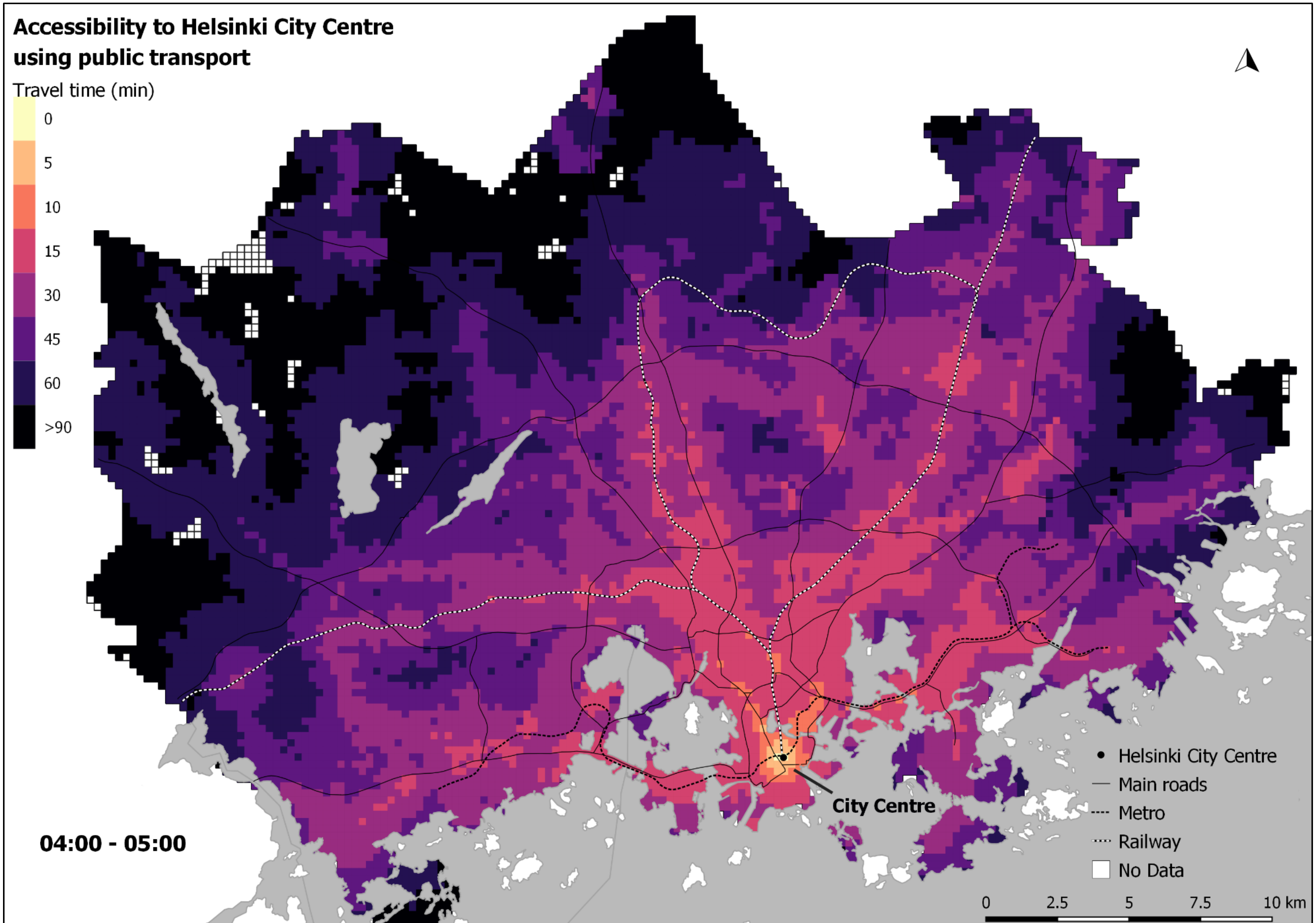
Appendix 9. The hourly accessibility maps to grocery stores and transport hubs using public transport. The results are based on door-to-door travel times from all 250 m x 250 m statistical grid cells to the target location.



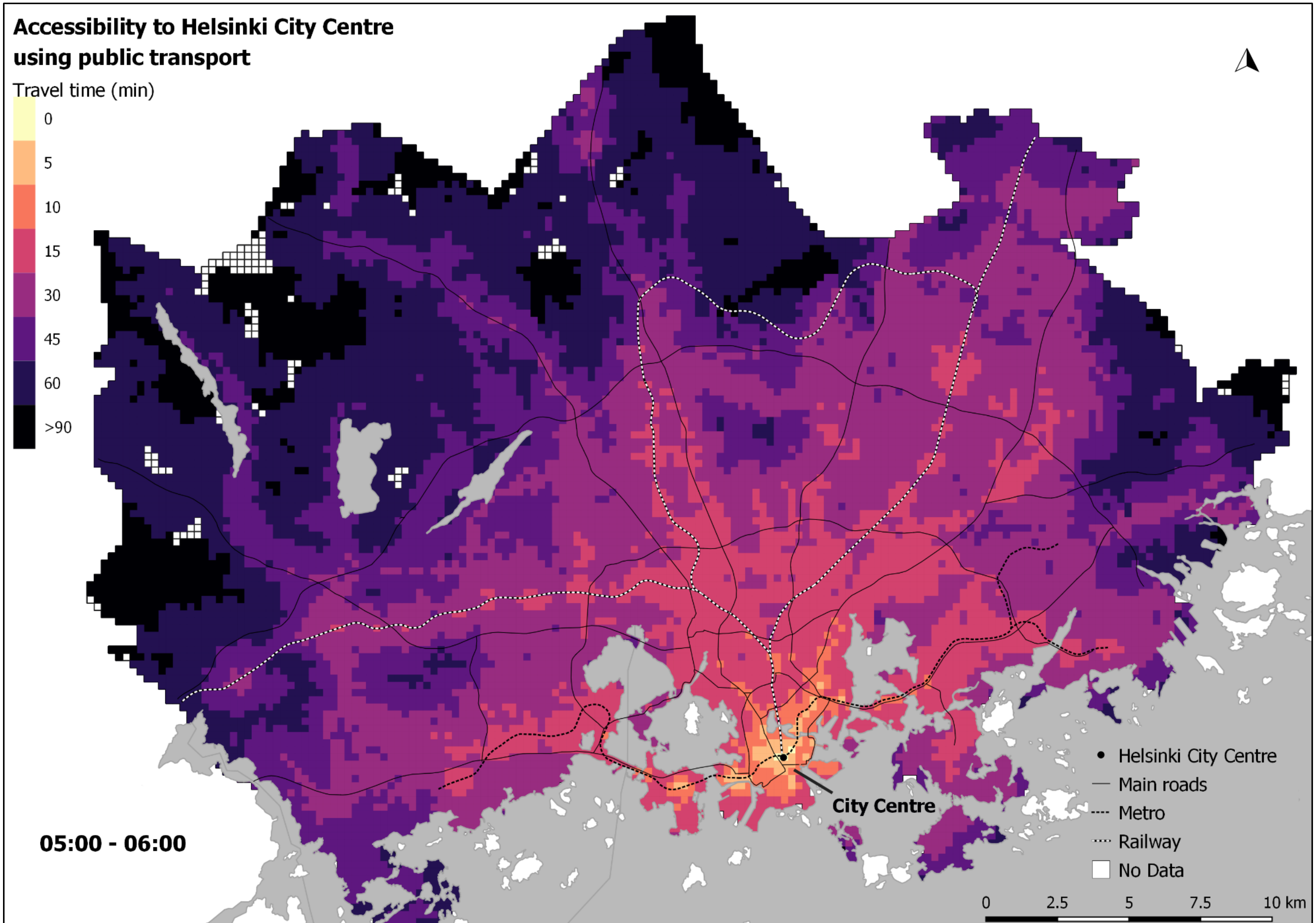

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
02:00 - 03:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

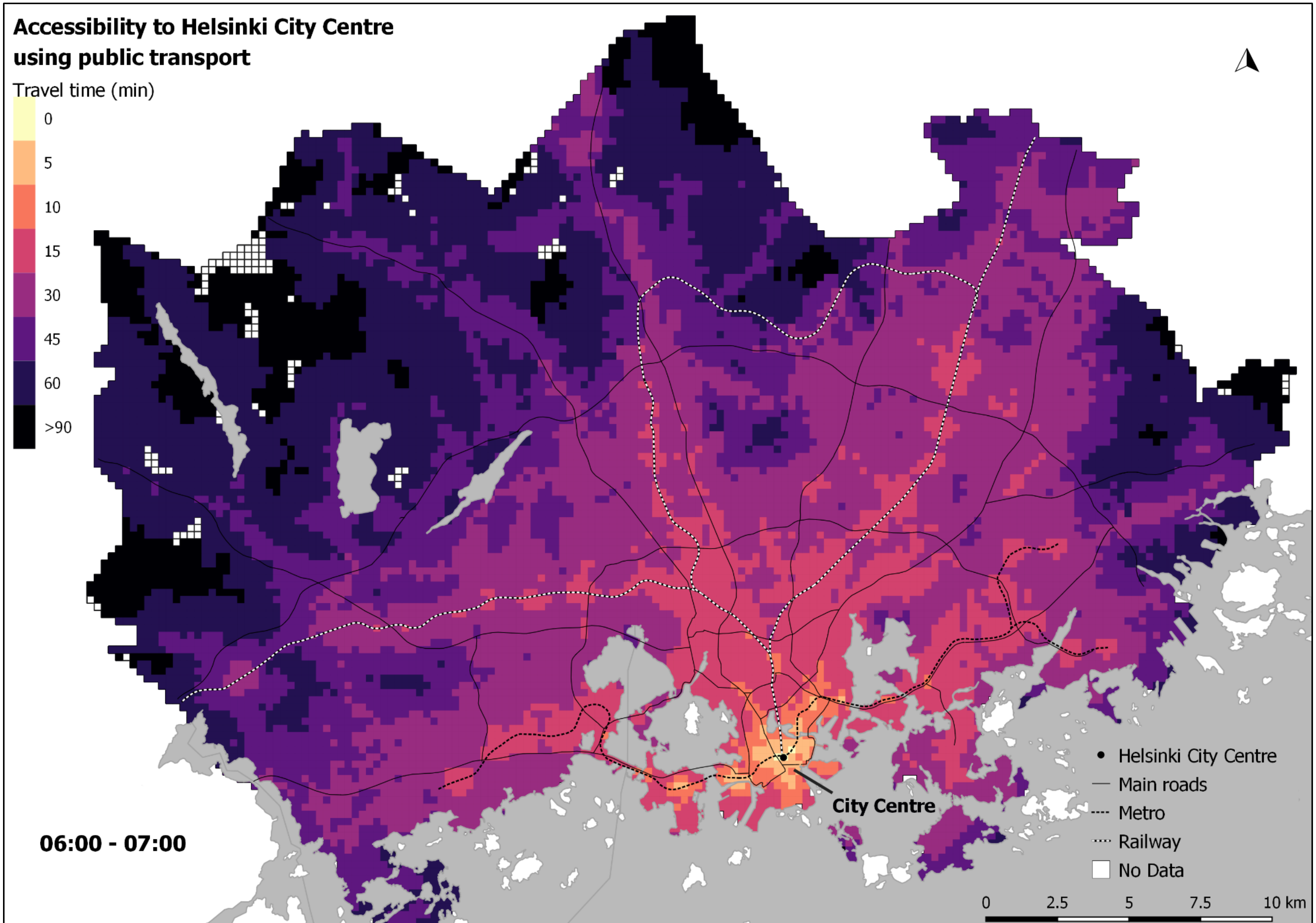


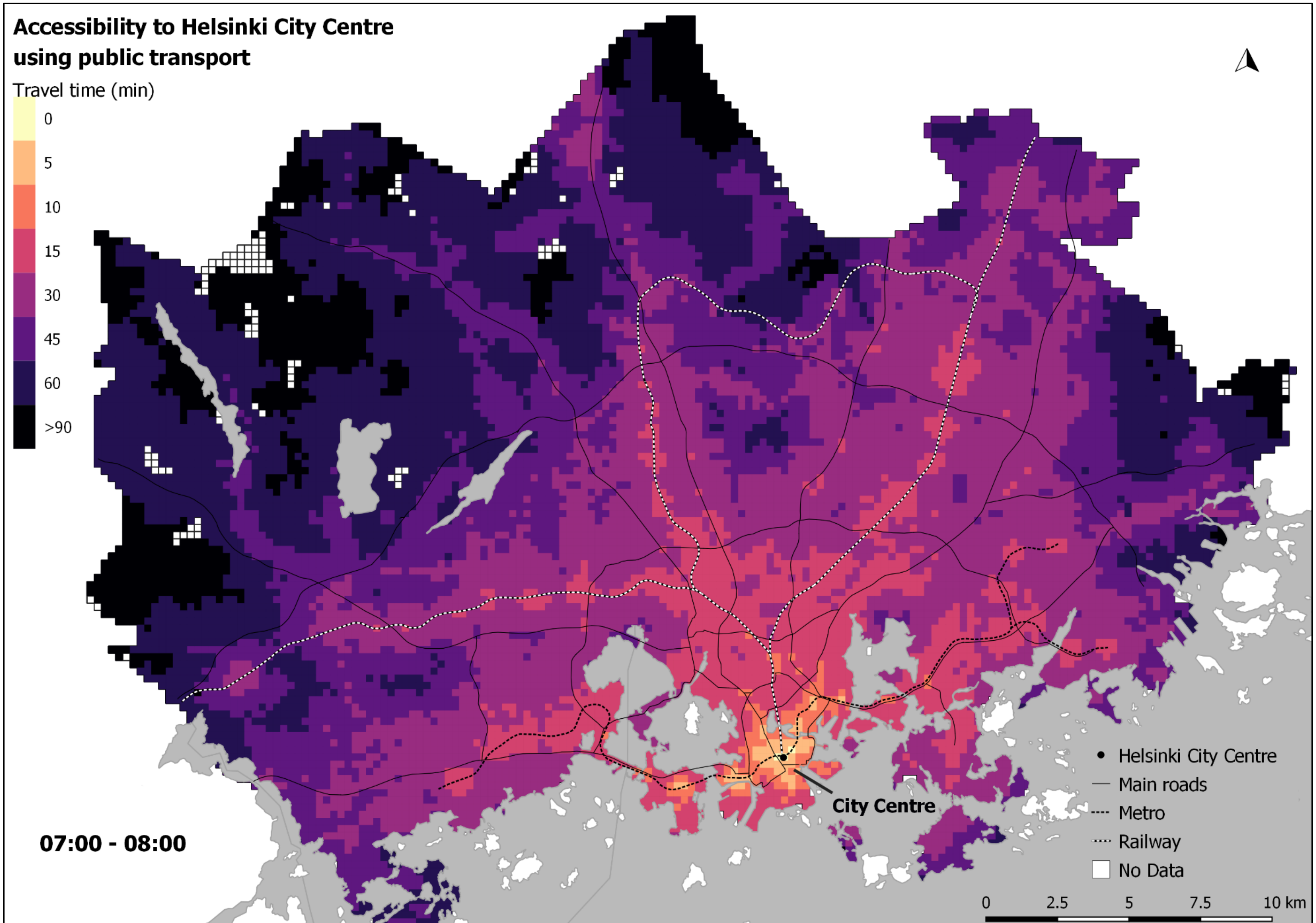
Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
03:00 - 04:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km



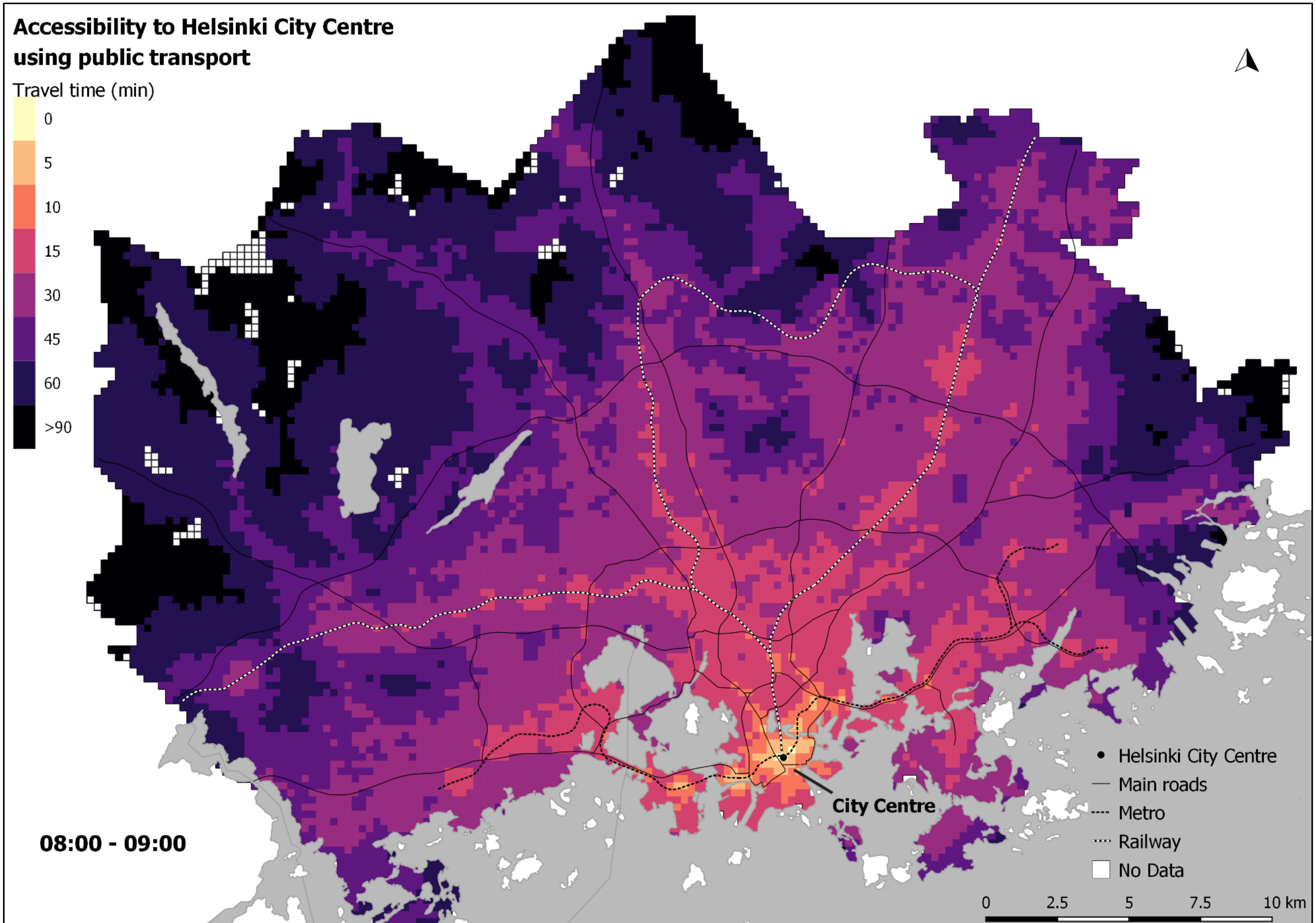
Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
04:00 - 05:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

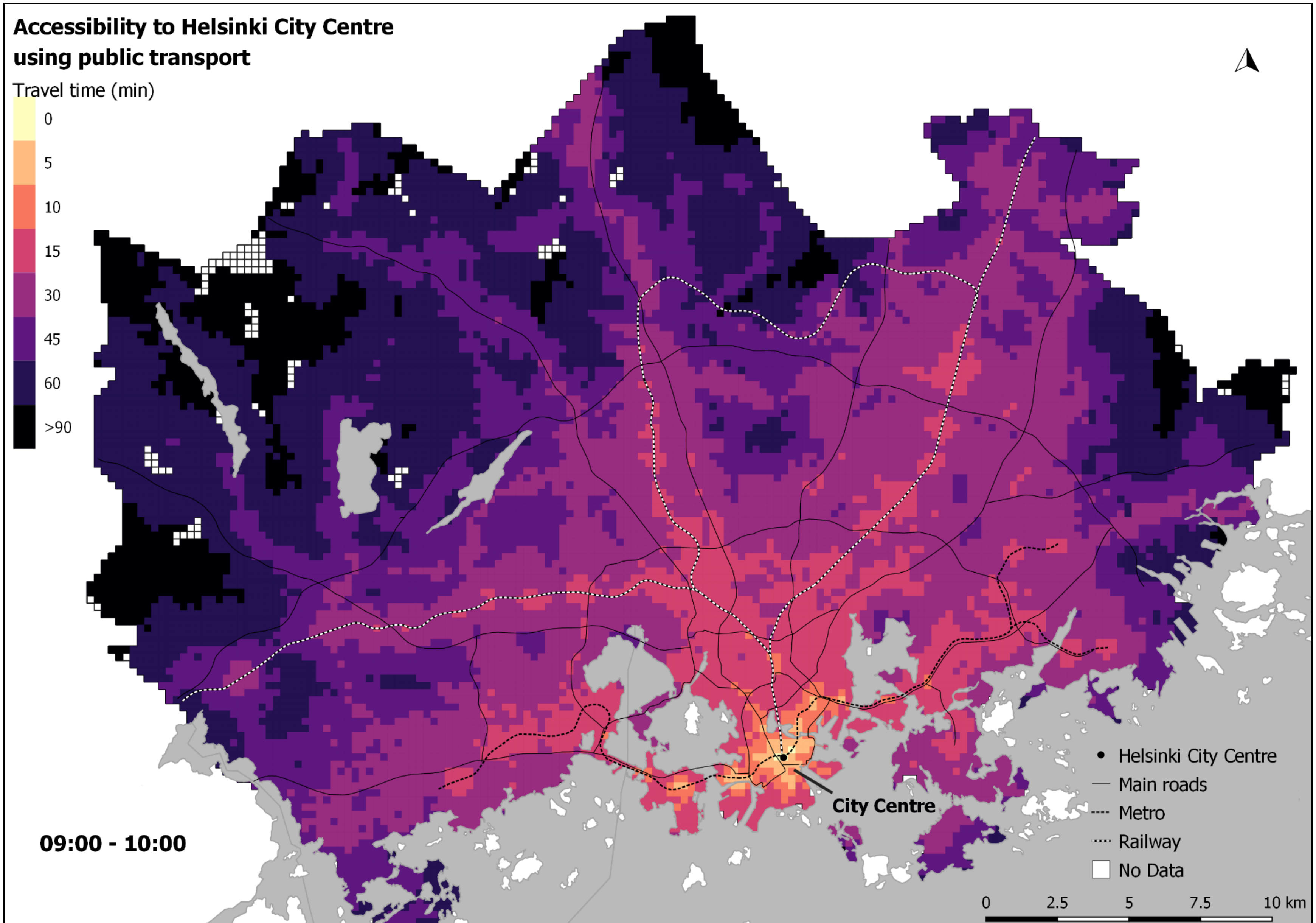

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
05:00 - 06:00
City Centre
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

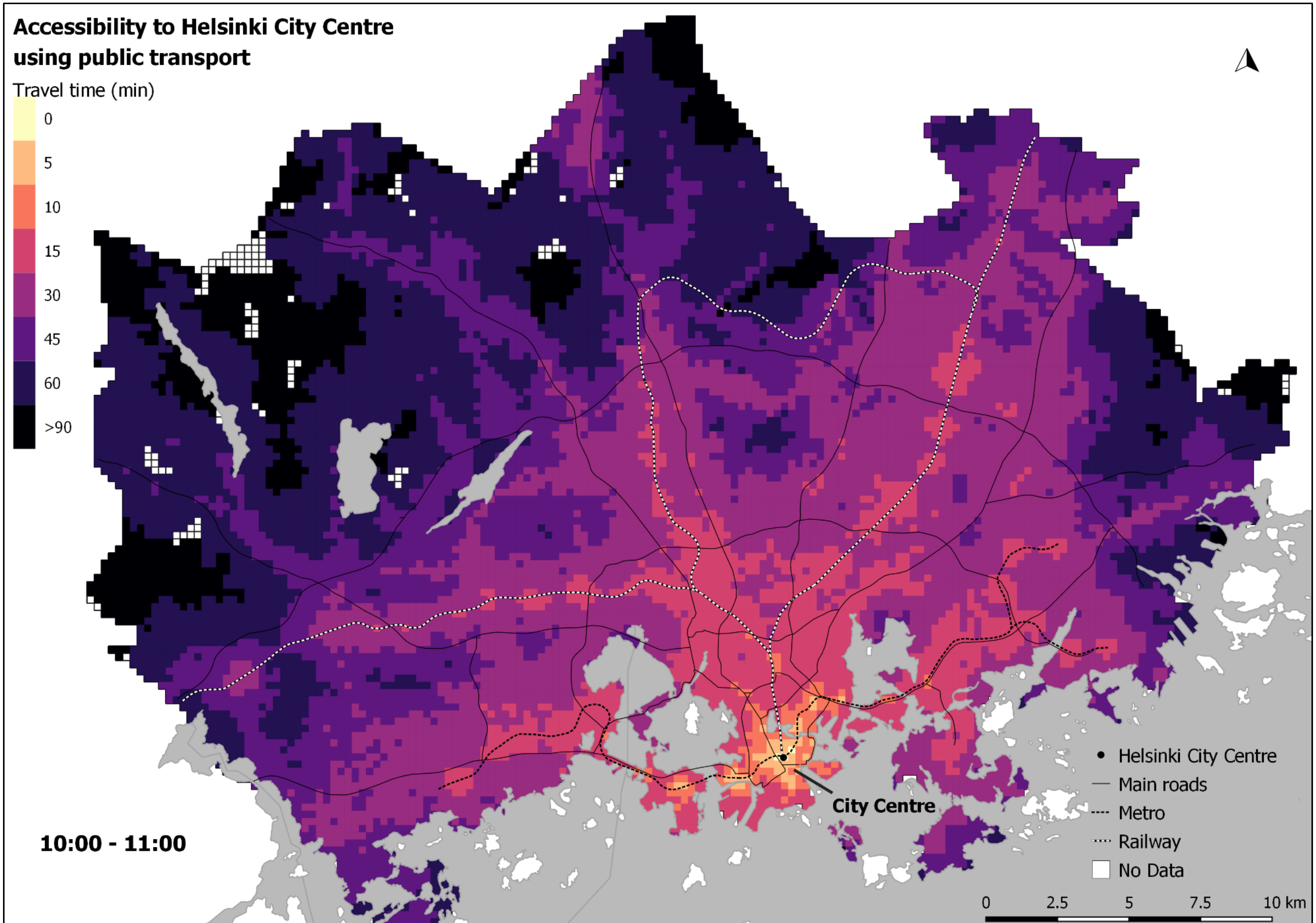

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
06:00 - 07:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

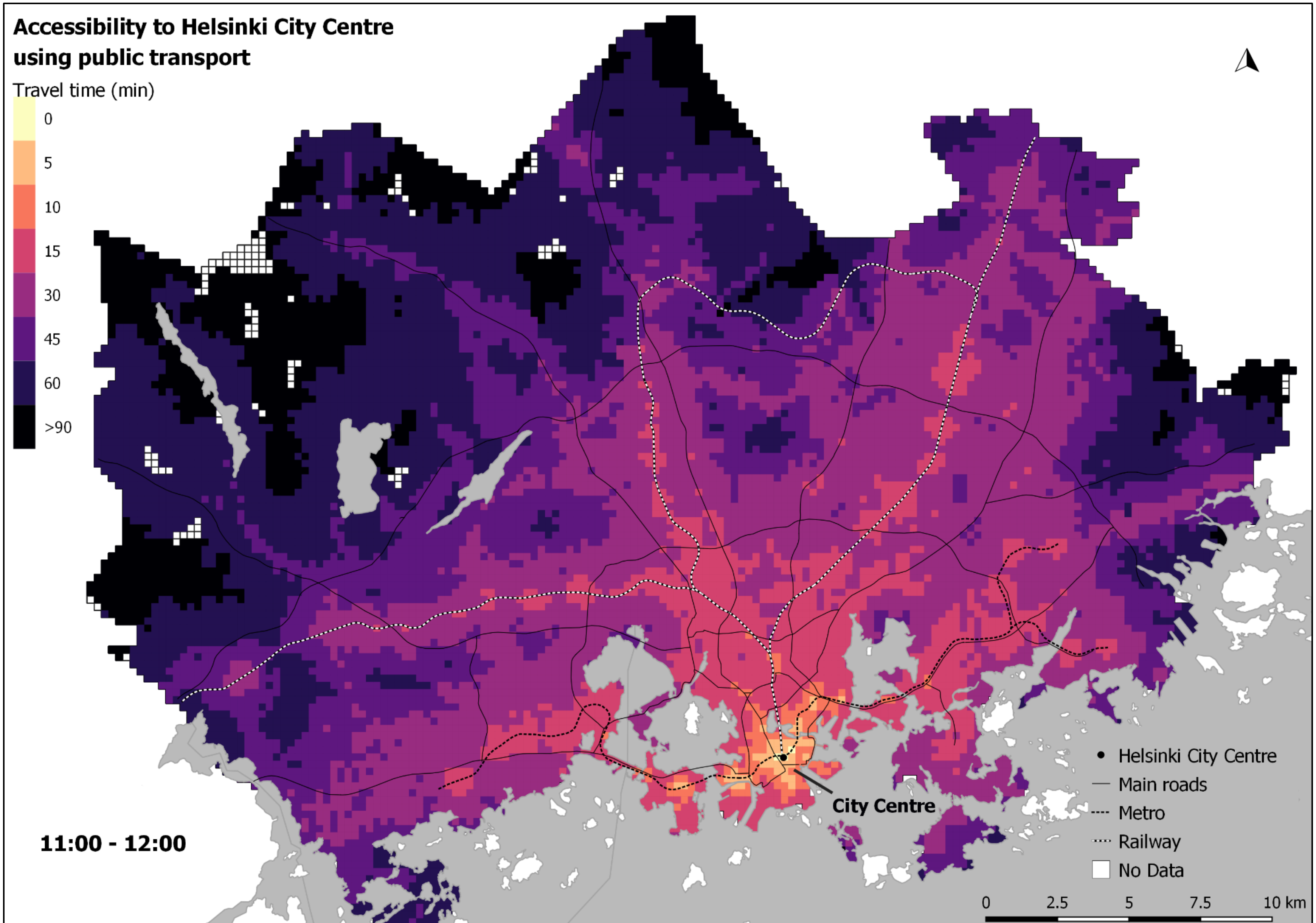


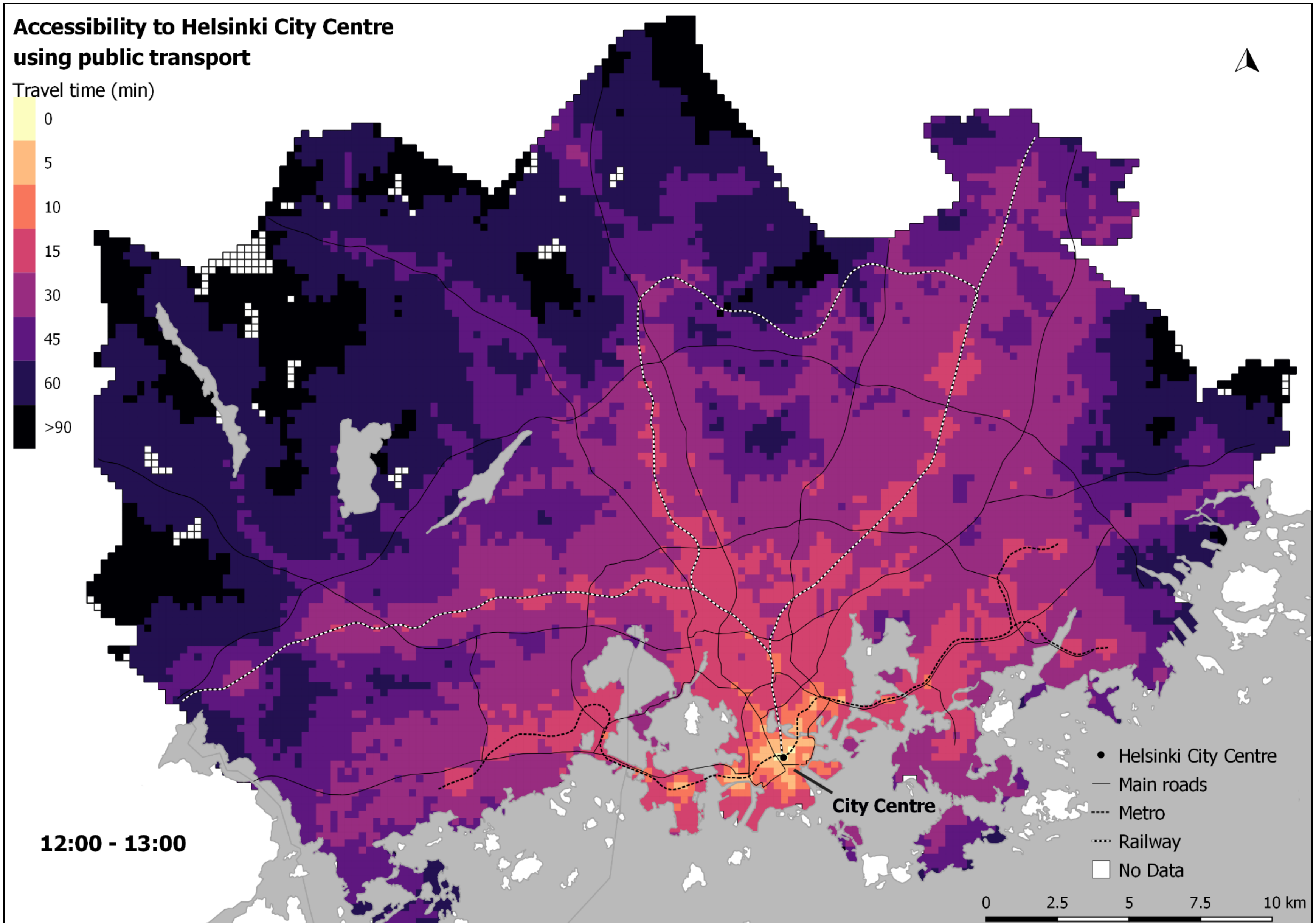
Accessibility to Helsinki City Centre using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
07:00 - 08:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

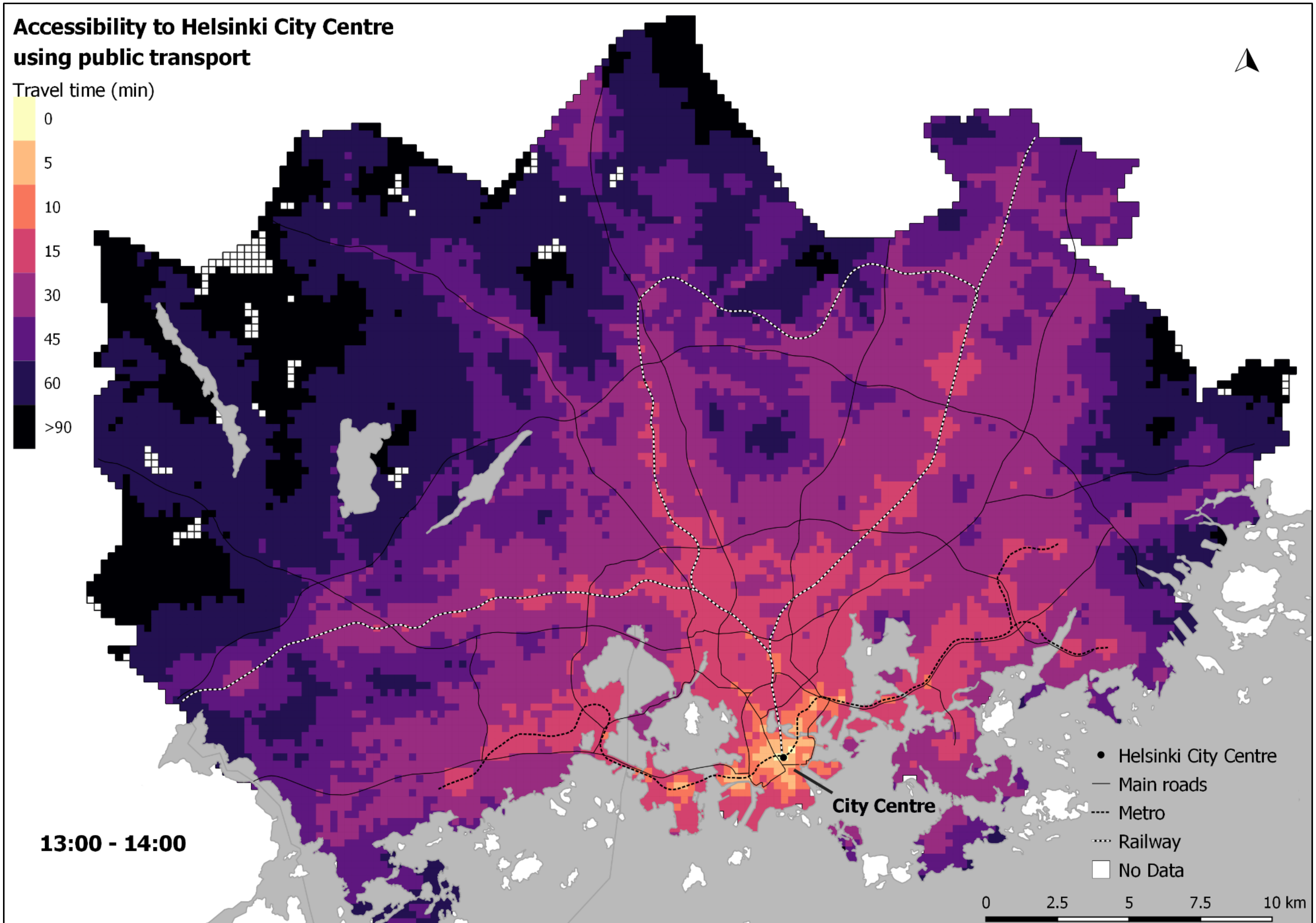

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
08:00 - 09:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

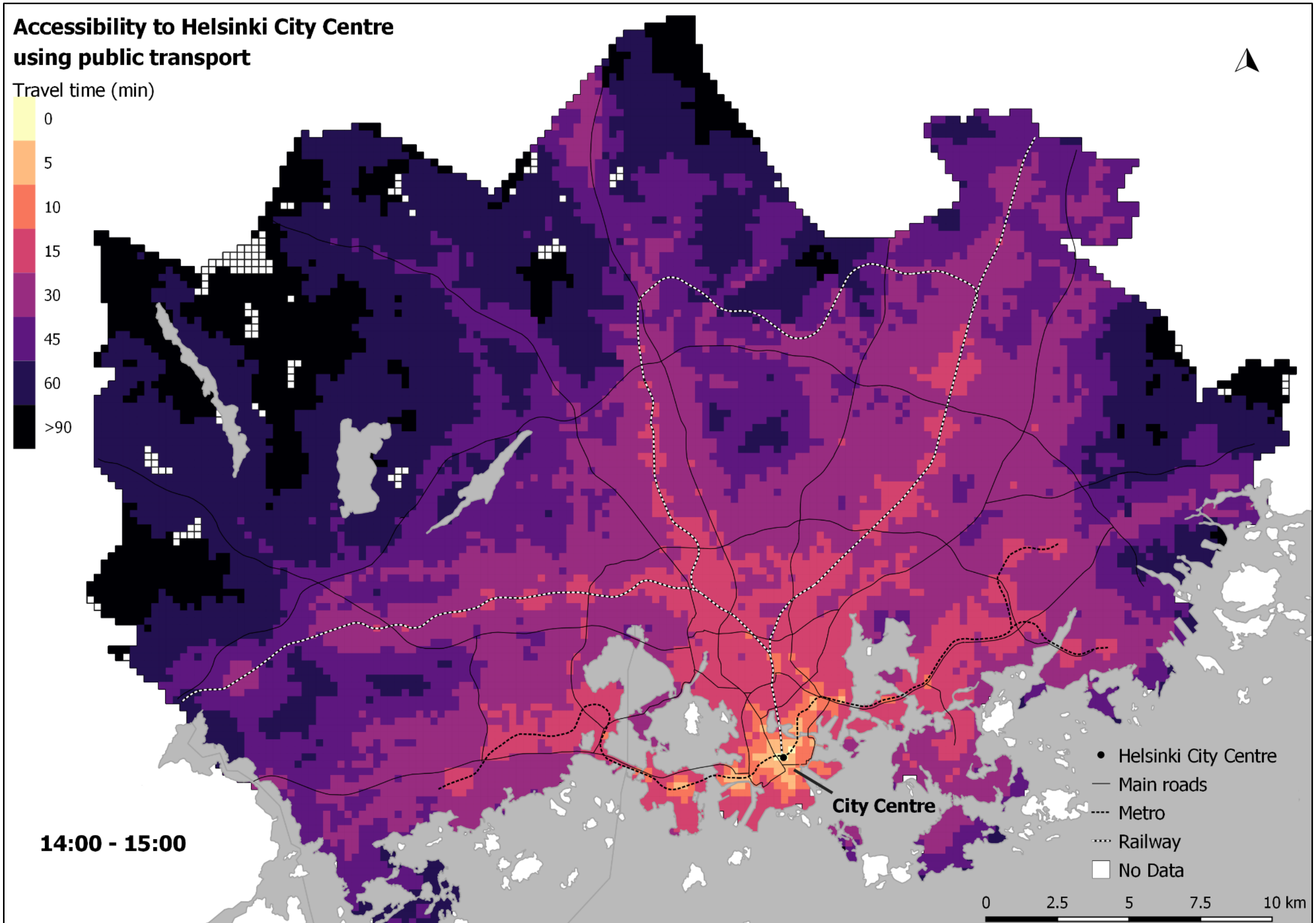

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
09:00 - 10:00

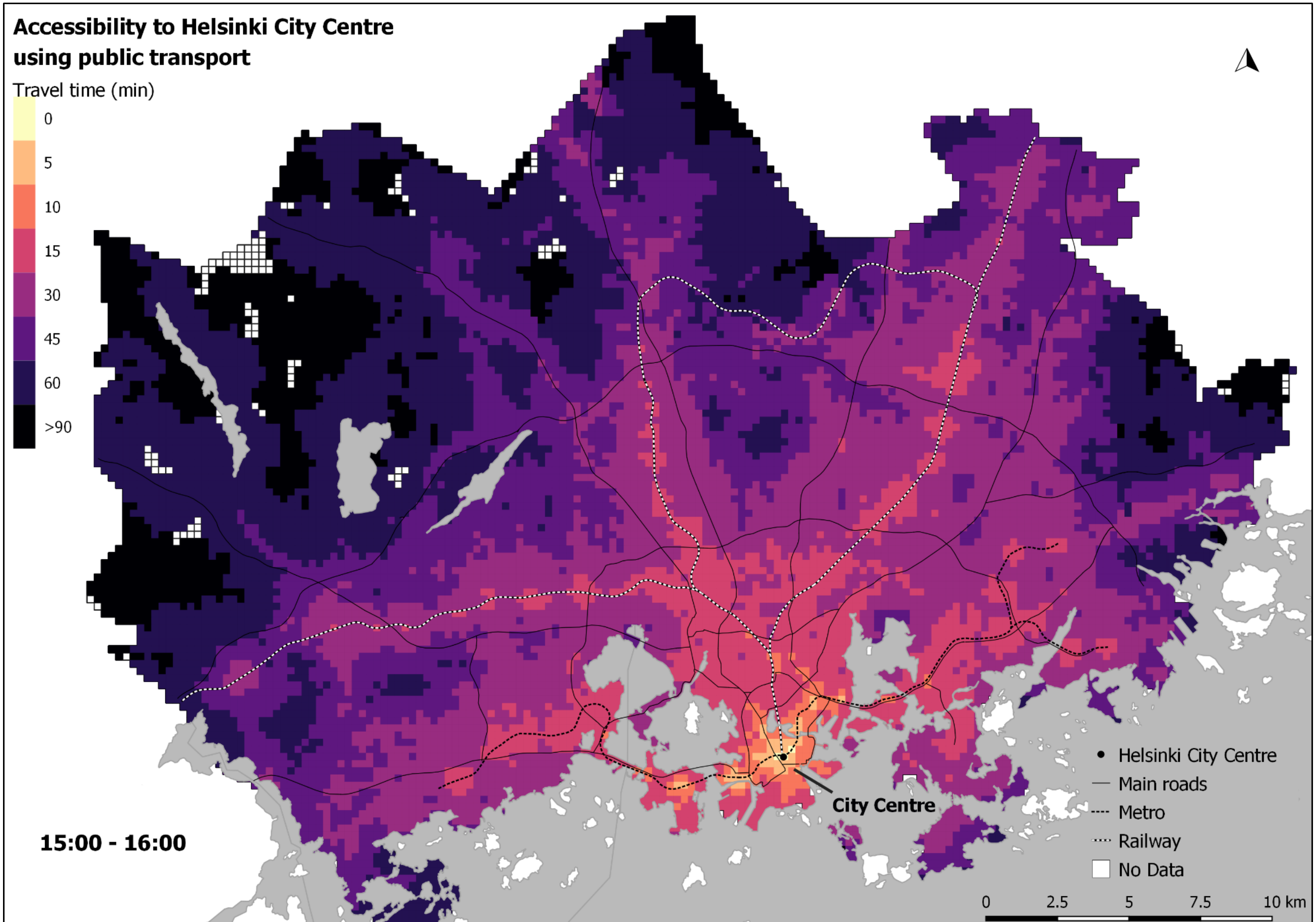

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
10:00 - 11:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

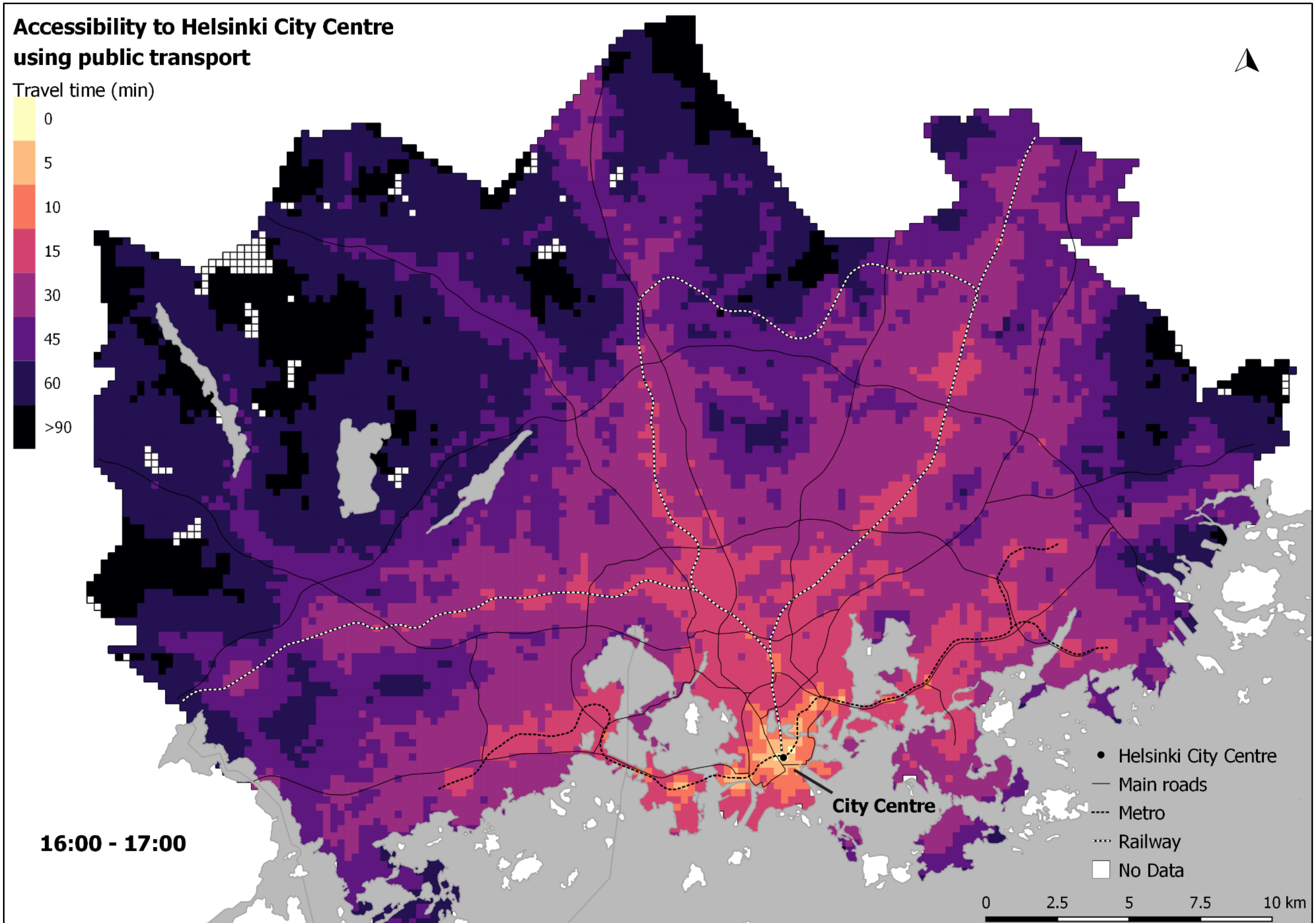

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
11:00 - 12:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

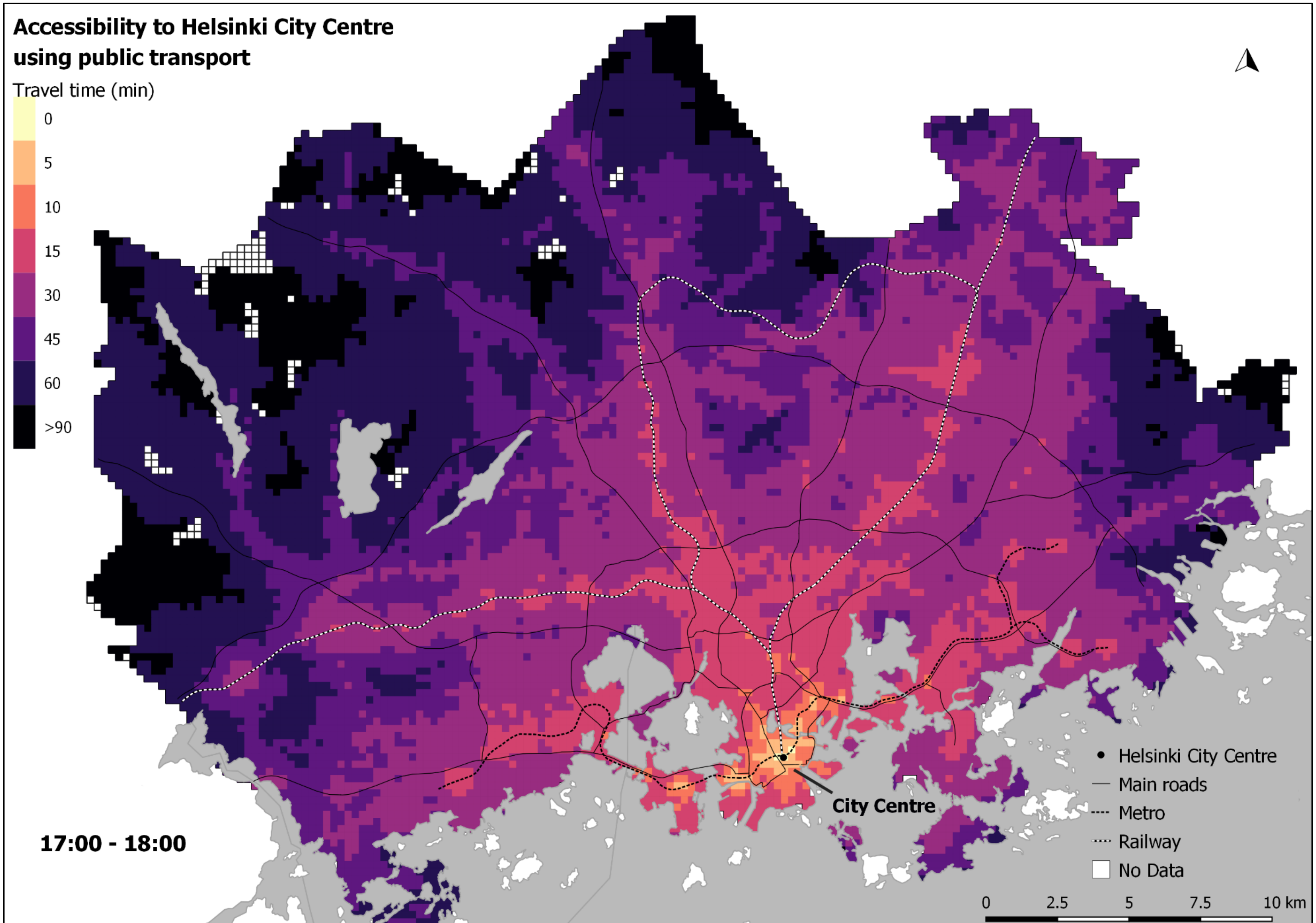

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
12:00 - 13:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

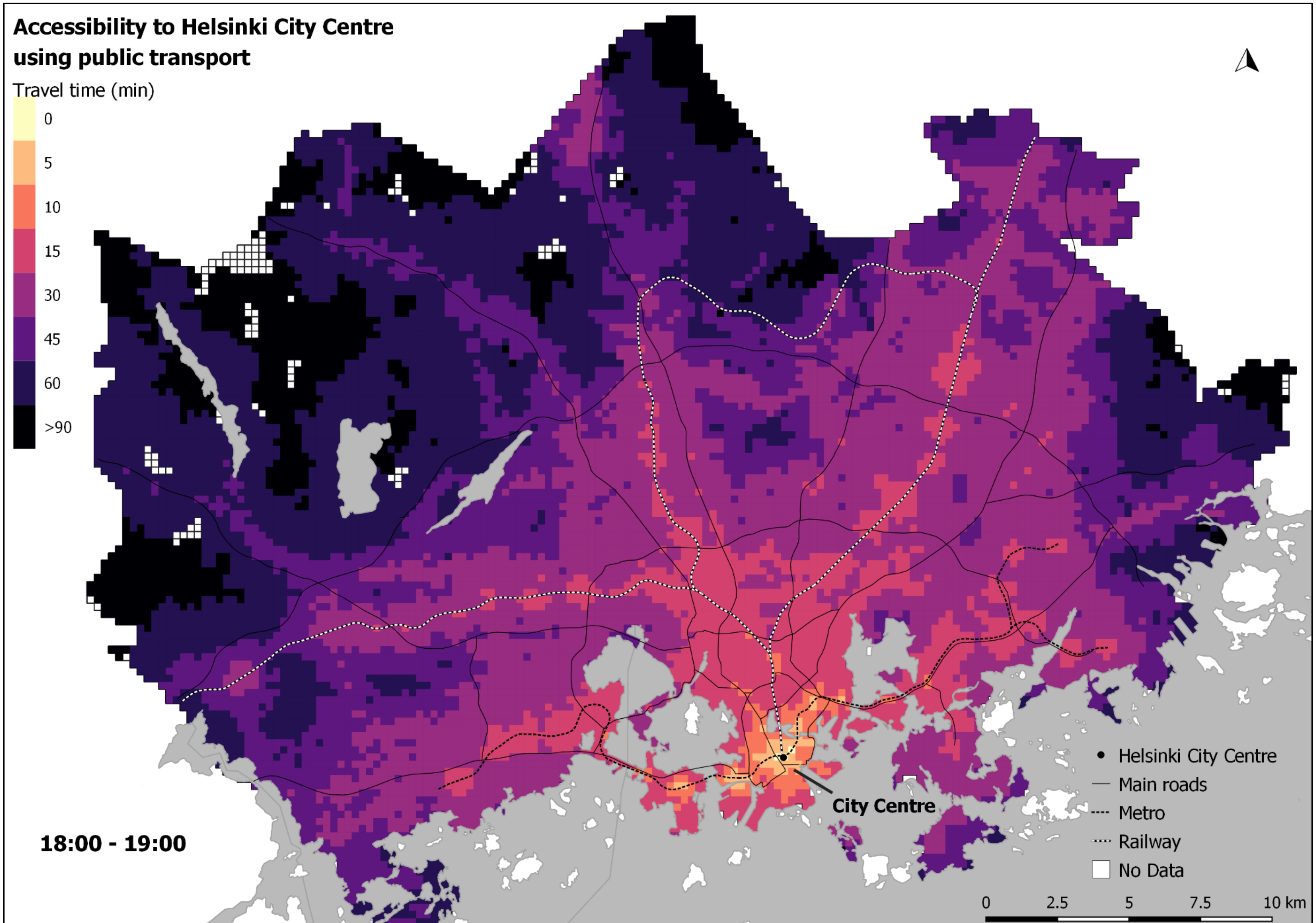

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
13:00 - 14:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

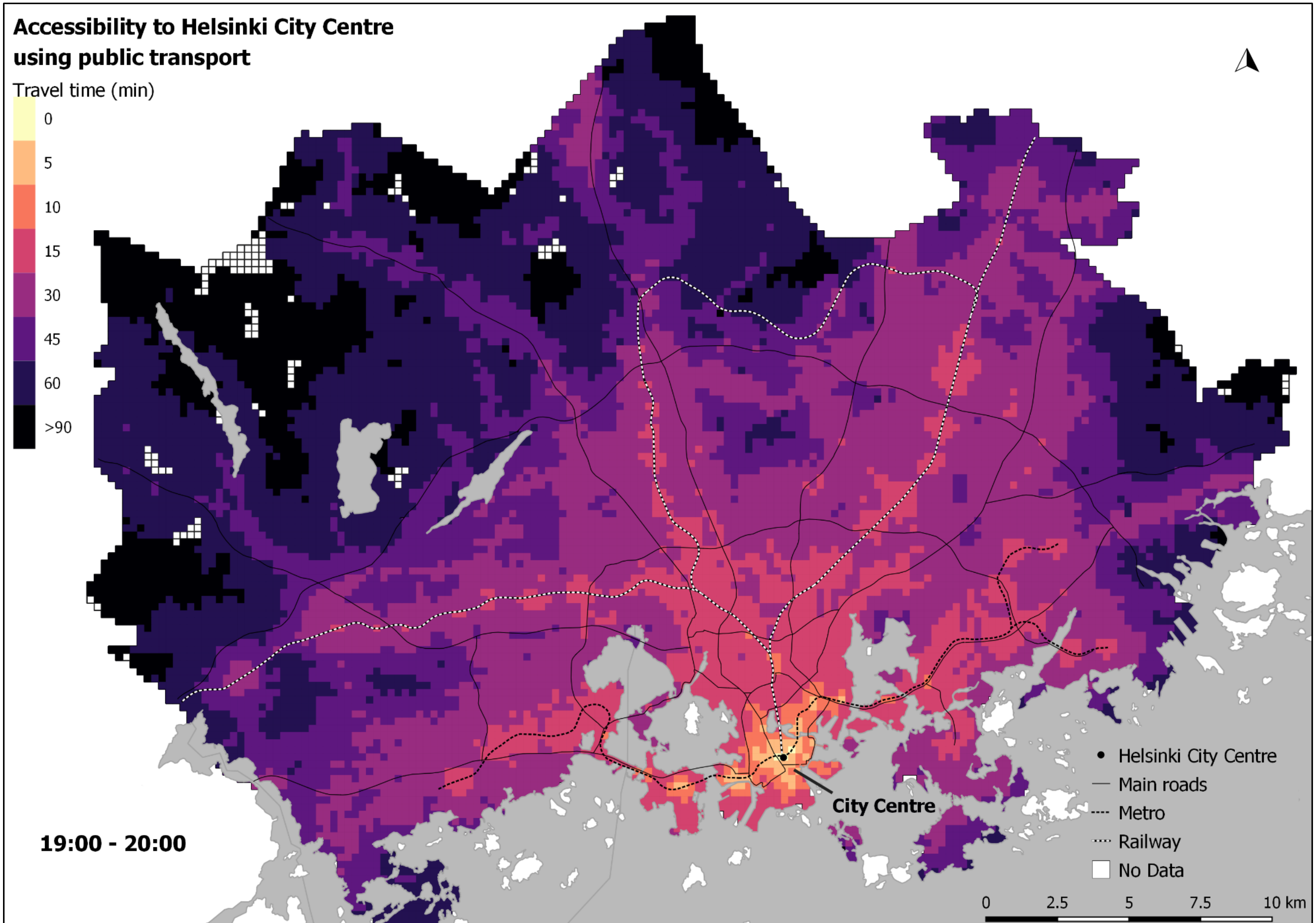

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
Helsinki City Centre
Main roads
Metro
Railway
No Data
14:00 - 15:00
0
2.5
5
7.5
10 km

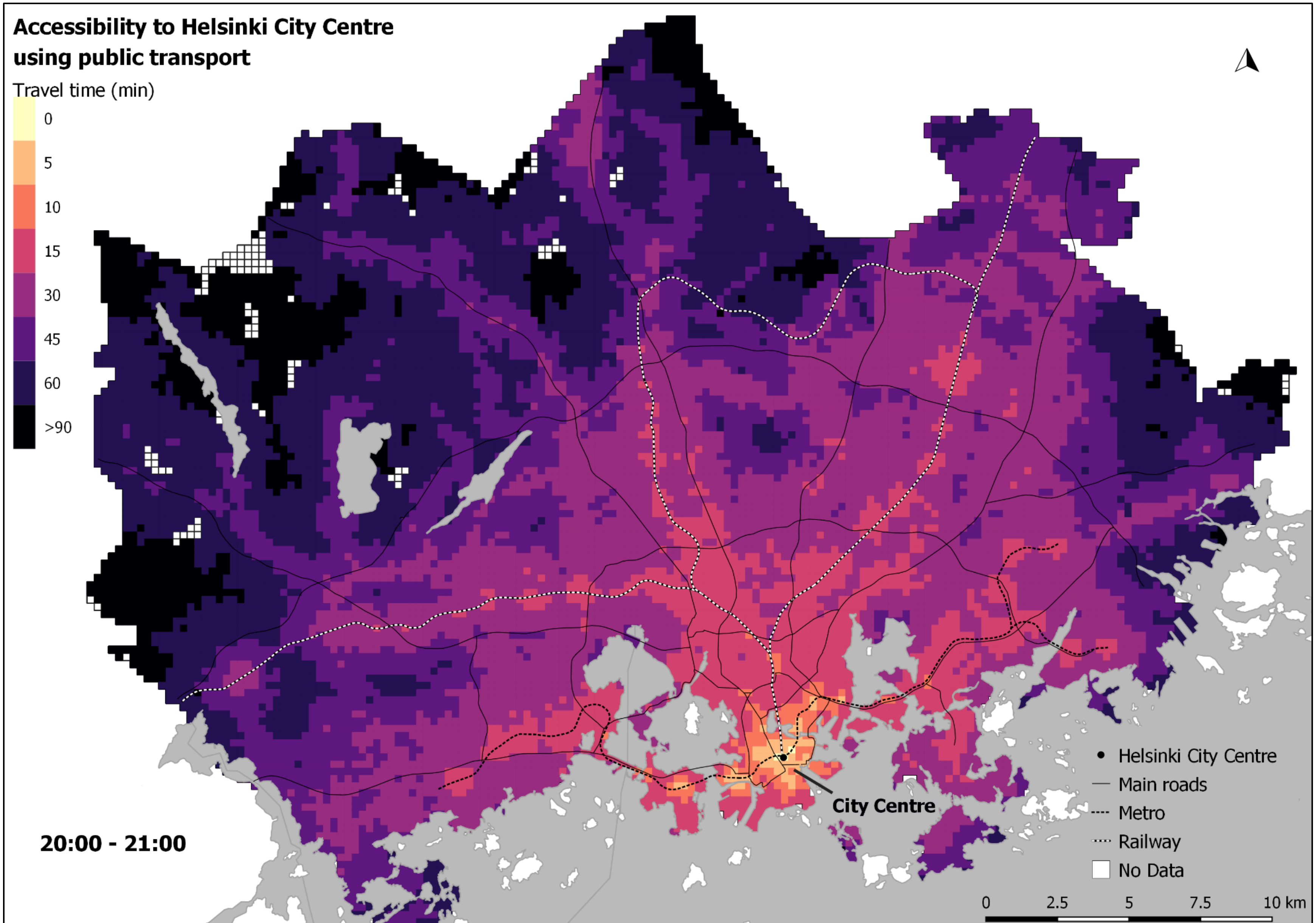

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
15:00 - 16:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

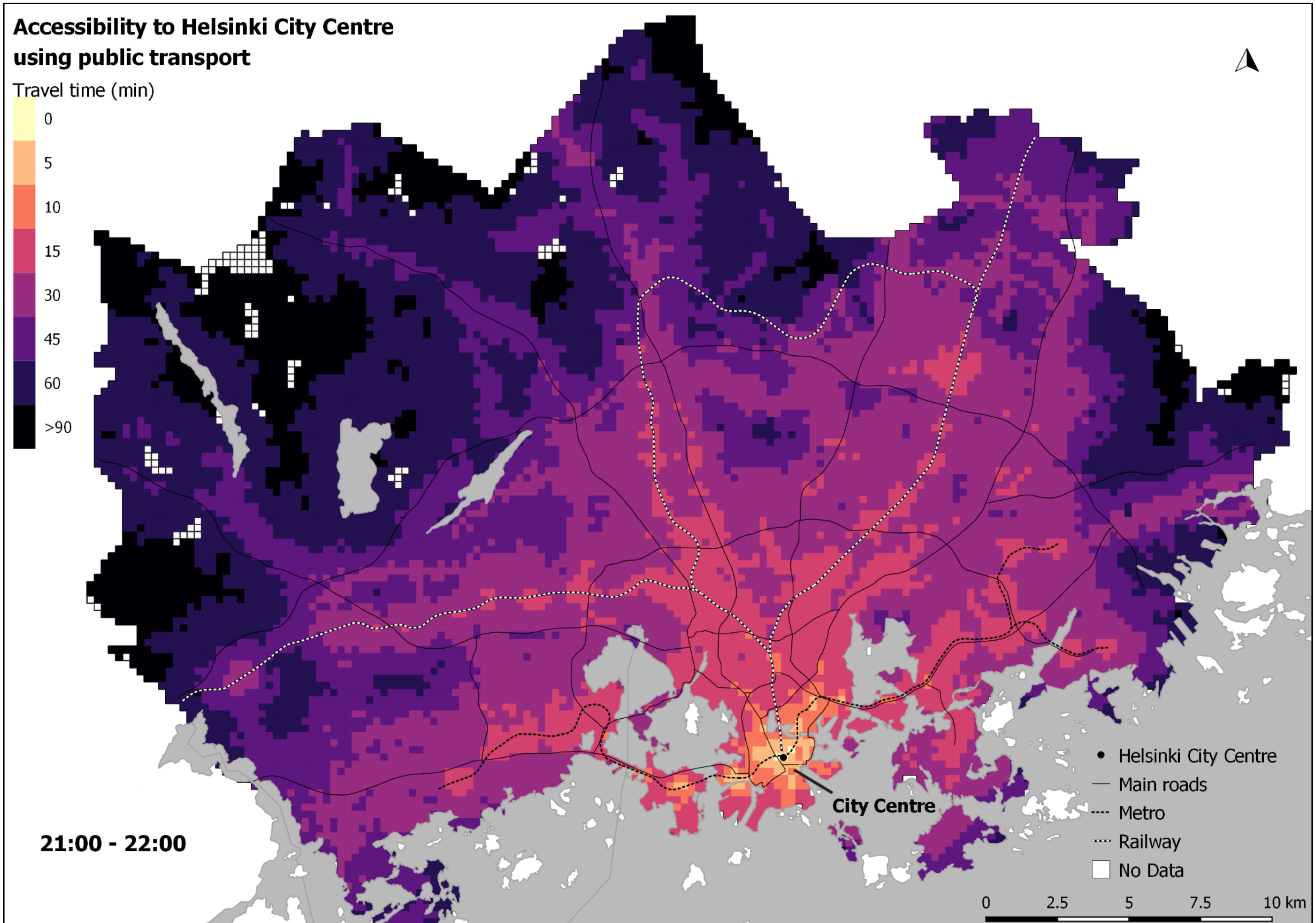

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
16:00 - 17:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
17:00 - 18:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

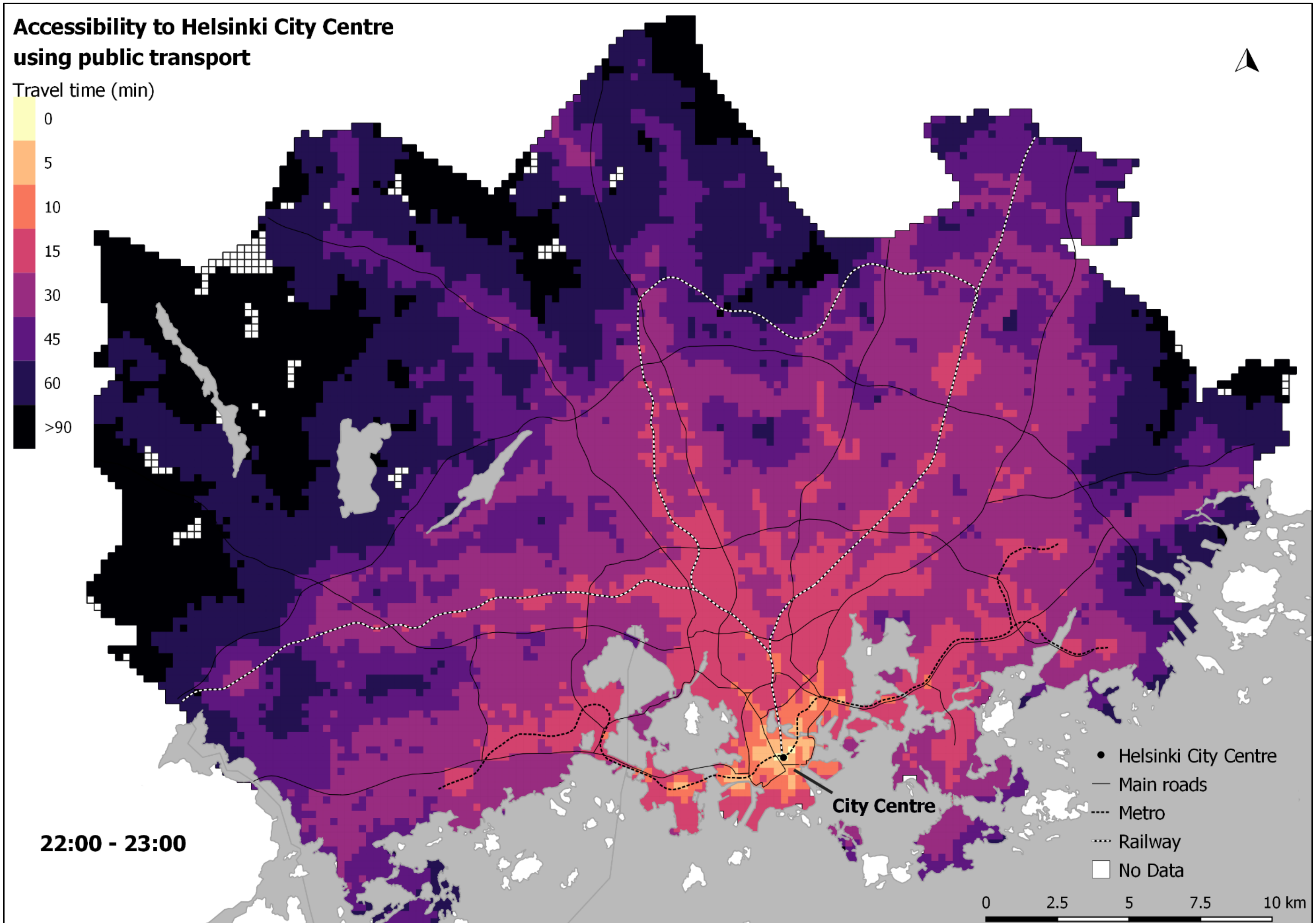

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
18:00 - 19:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

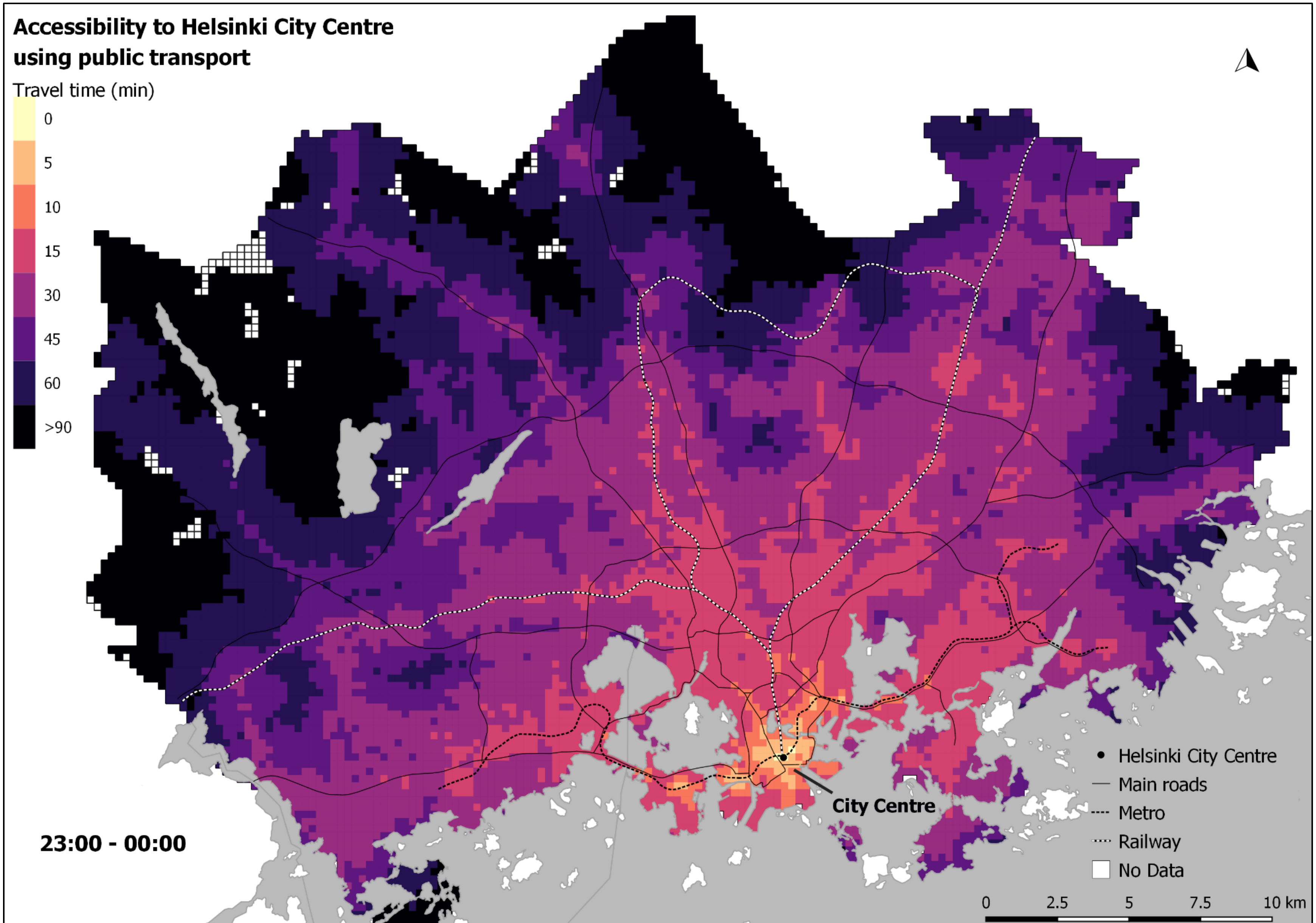

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Helsinki City Centre
Main roads
Metro
Railway
No Data
City Centre
19:00 - 20:00
0
2.5
5
7.5
10 km


Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
Helsinki City Centre
Main roads
Metro
Railway
No Data
20:00 - 21:00
0
2.5
5
7.5
10 km

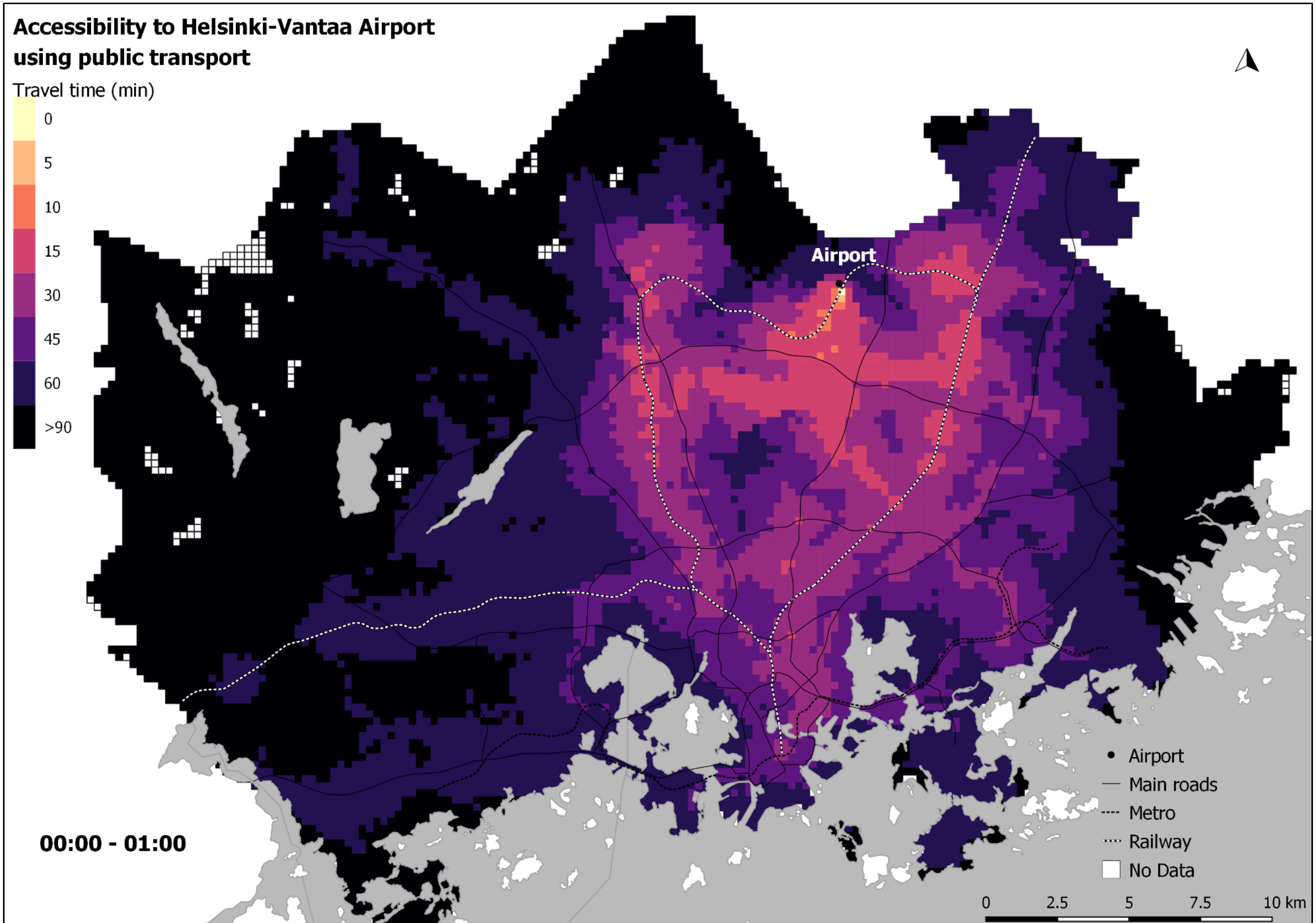


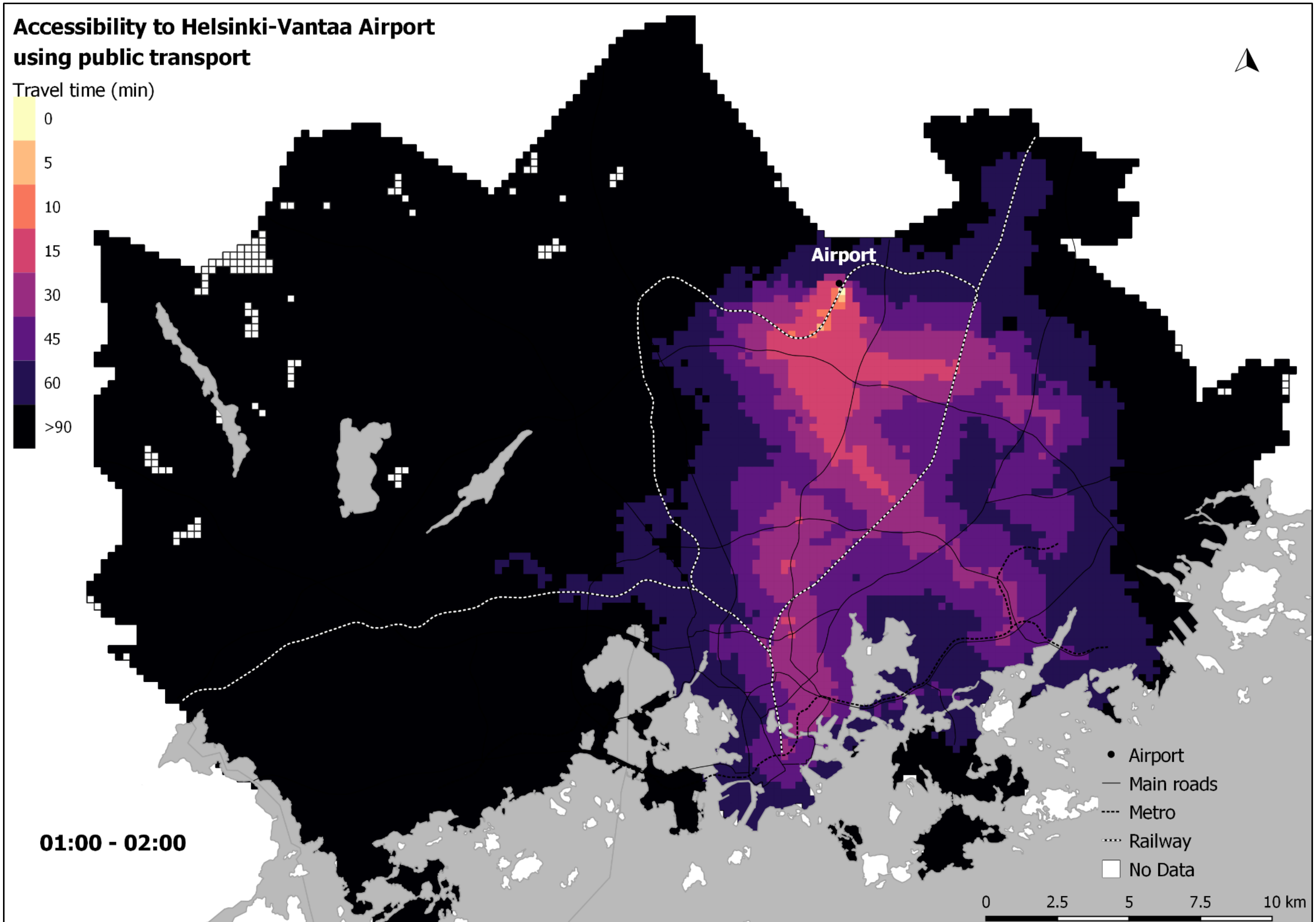
Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
21:00 - 22:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

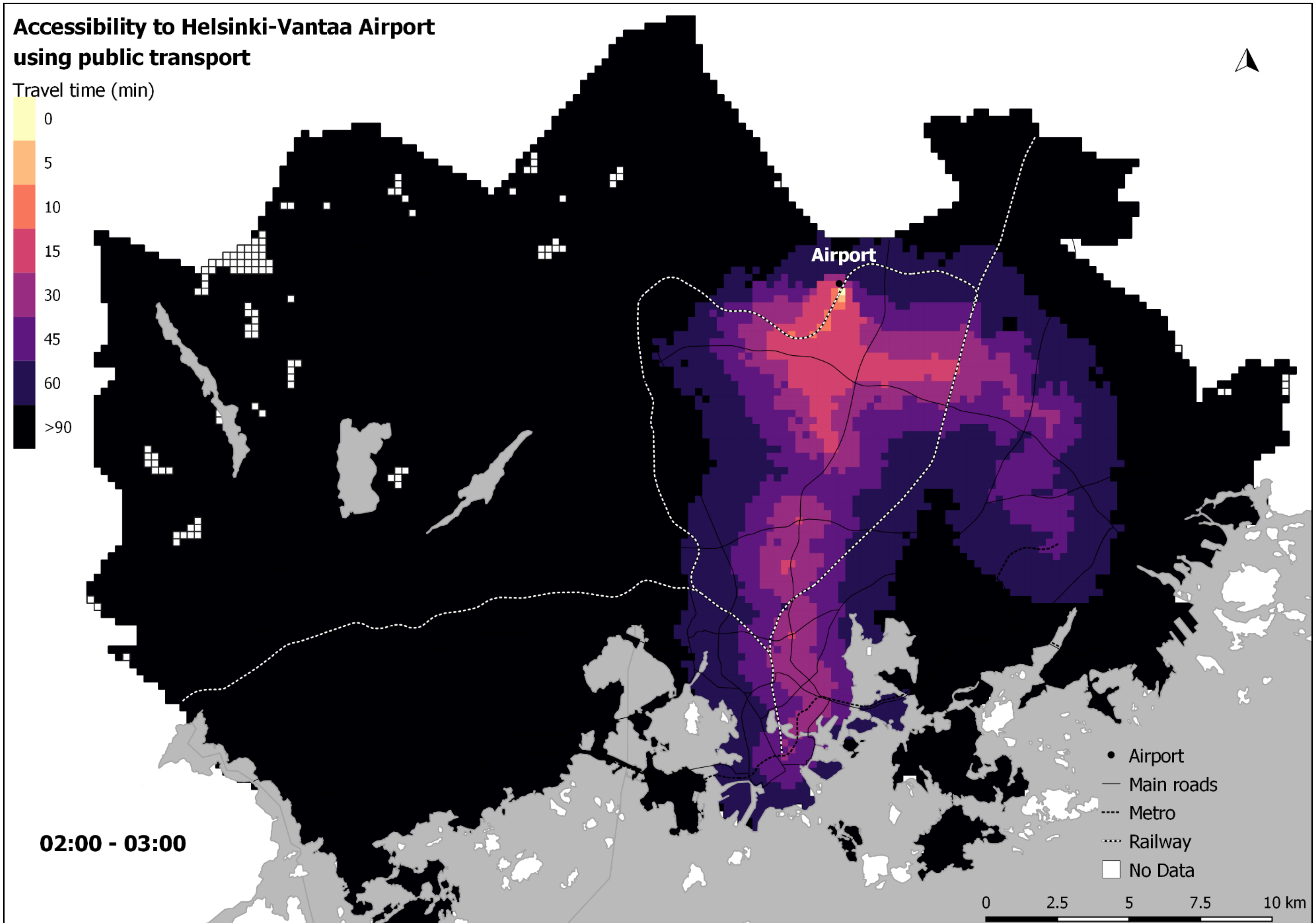

Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
22:00 - 23:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

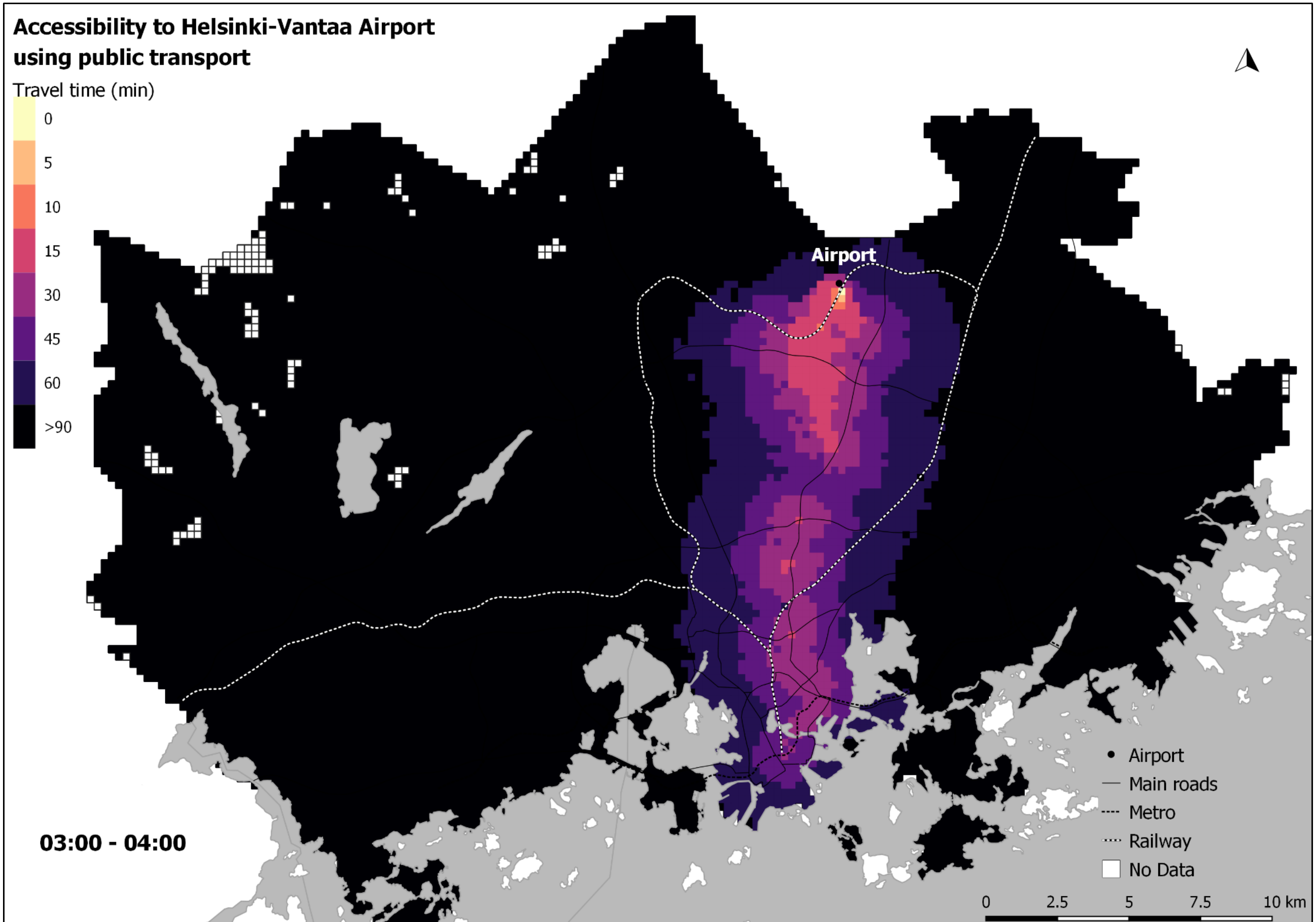


Accessibility to Helsinki City Centre
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
City Centre
23:00 - 00:00
Helsinki City Centre
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

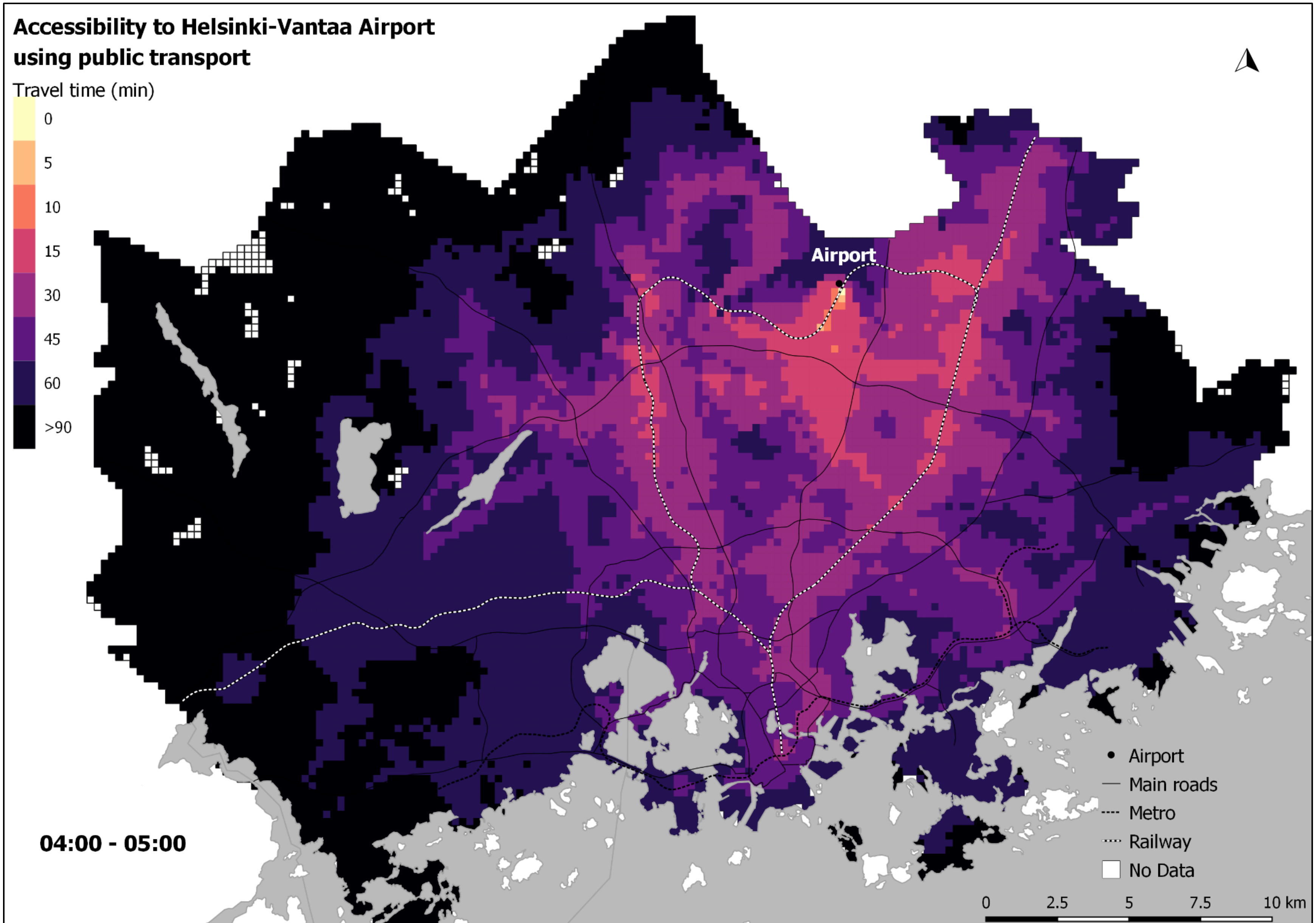

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
00:00 - 01:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

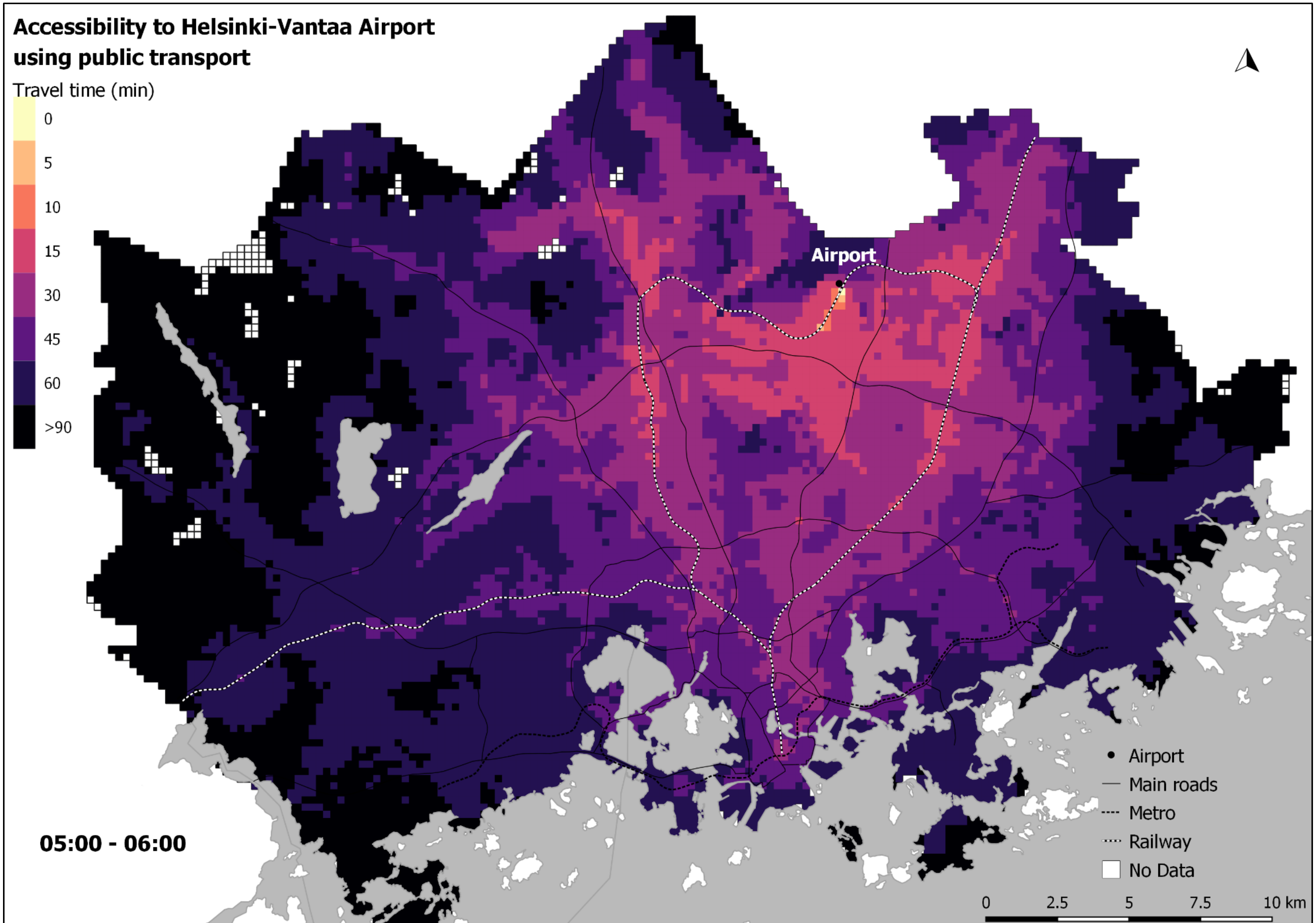

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
01:00 - 02:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

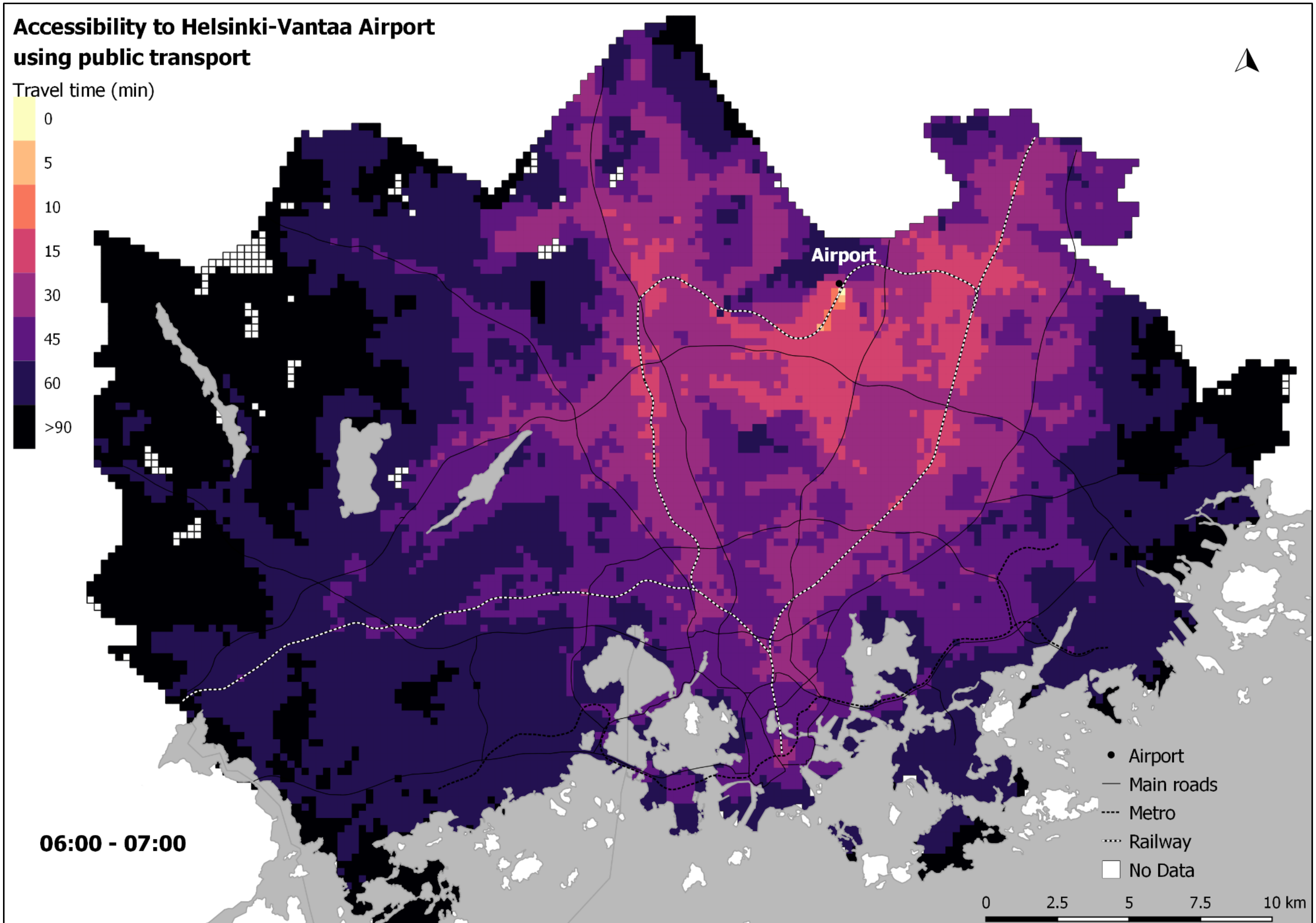

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
02:00 - 03:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

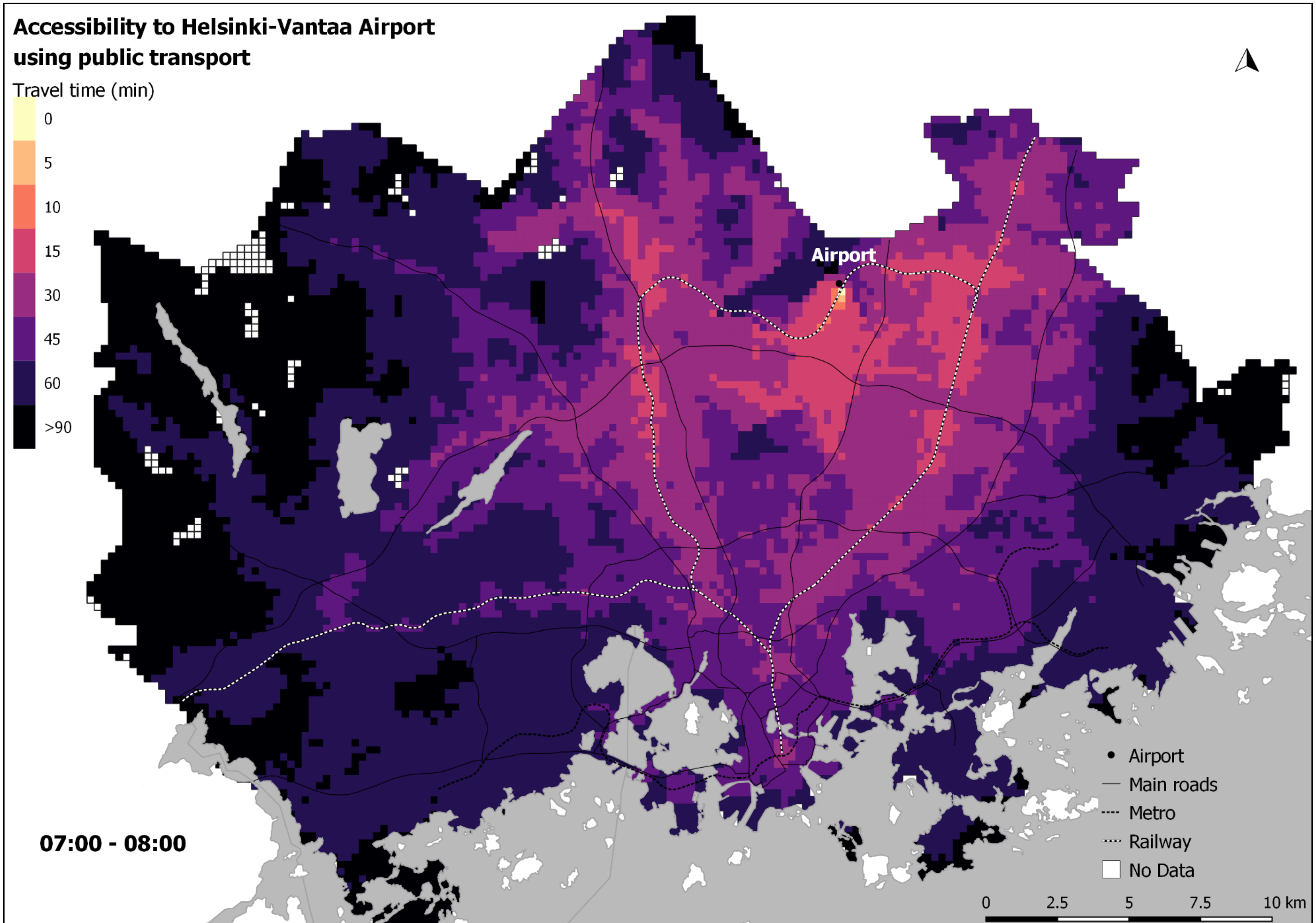


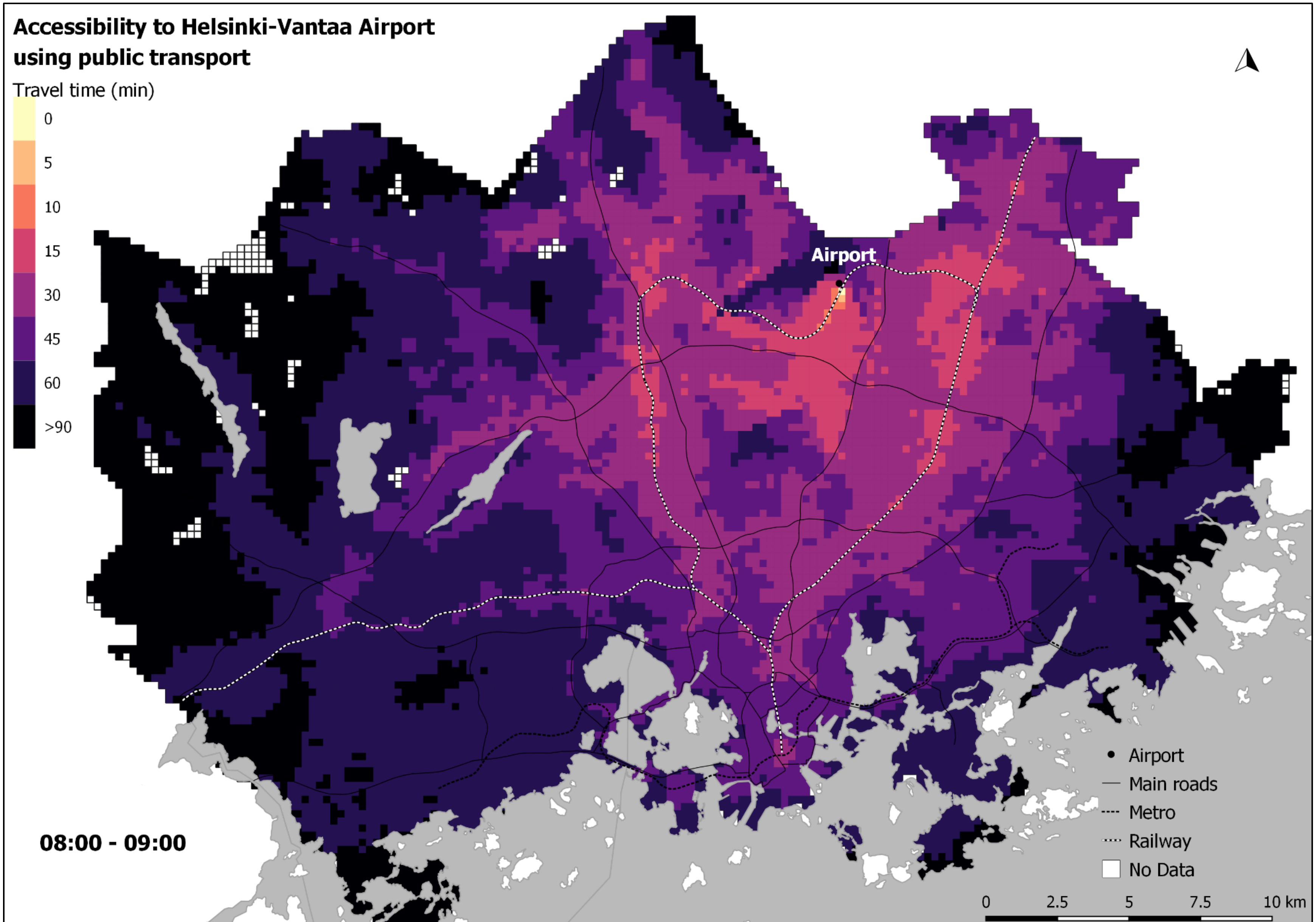
Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
03:00 - 04:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

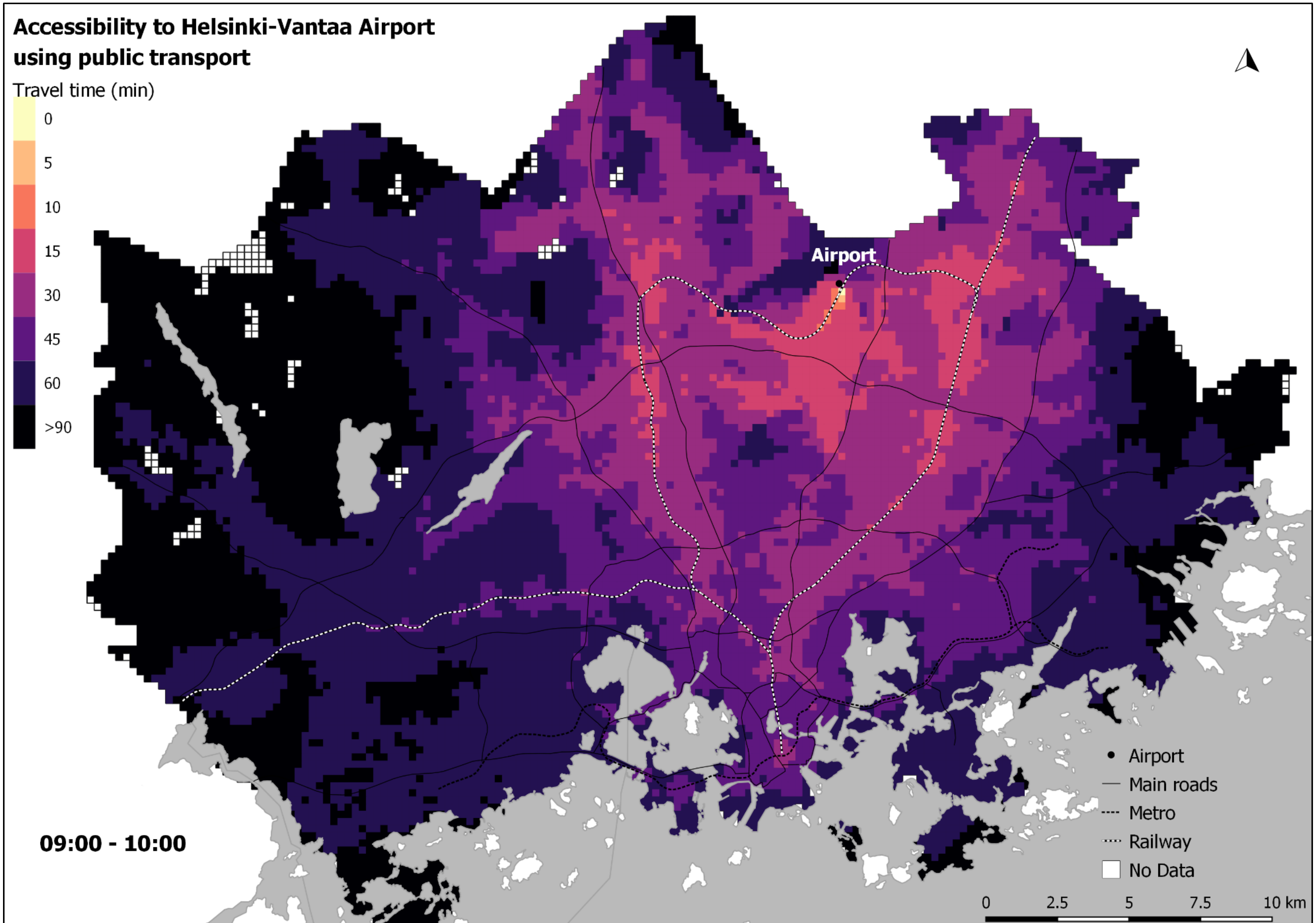

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
04:00 - 05:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

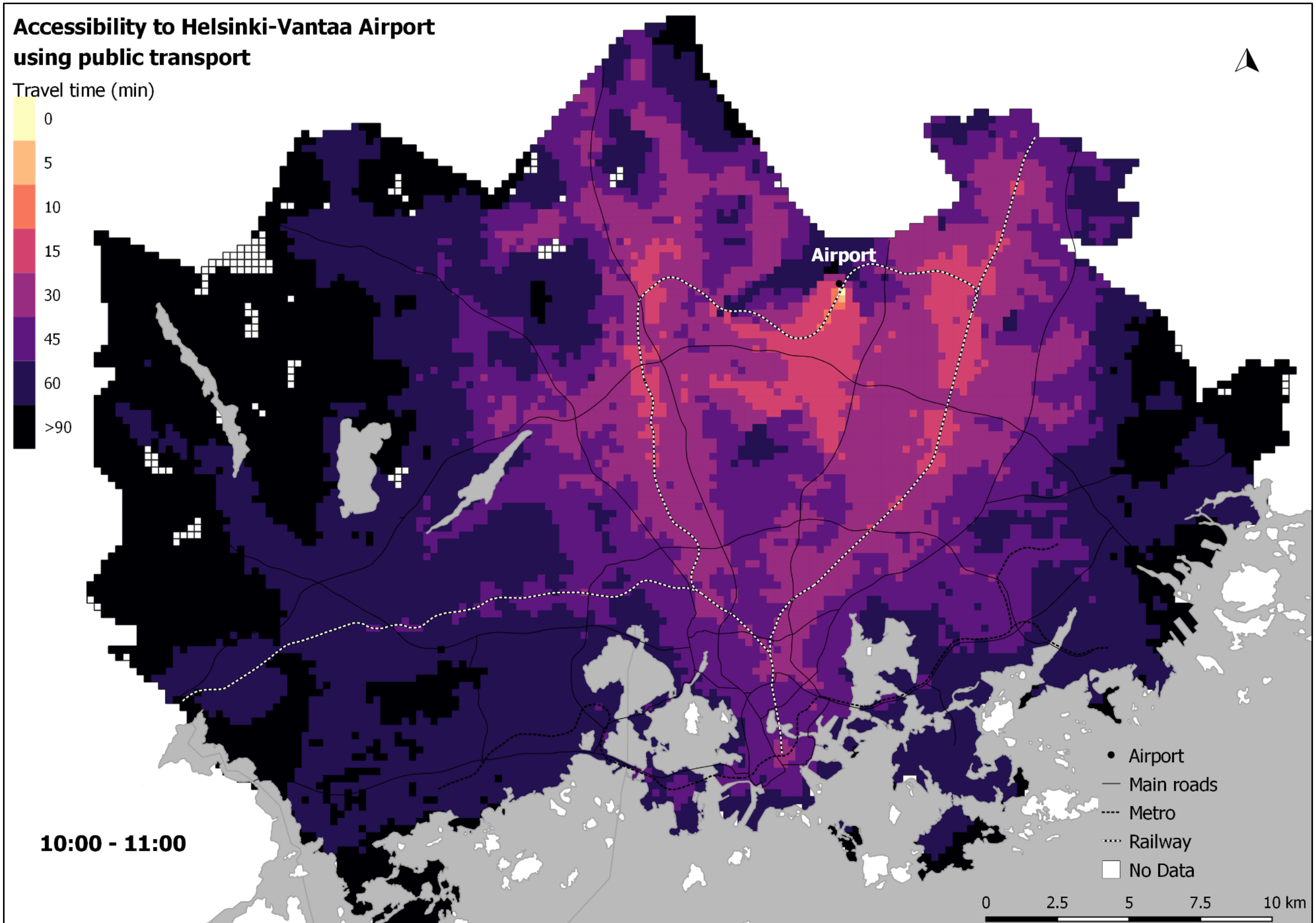

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
05:00 - 06:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

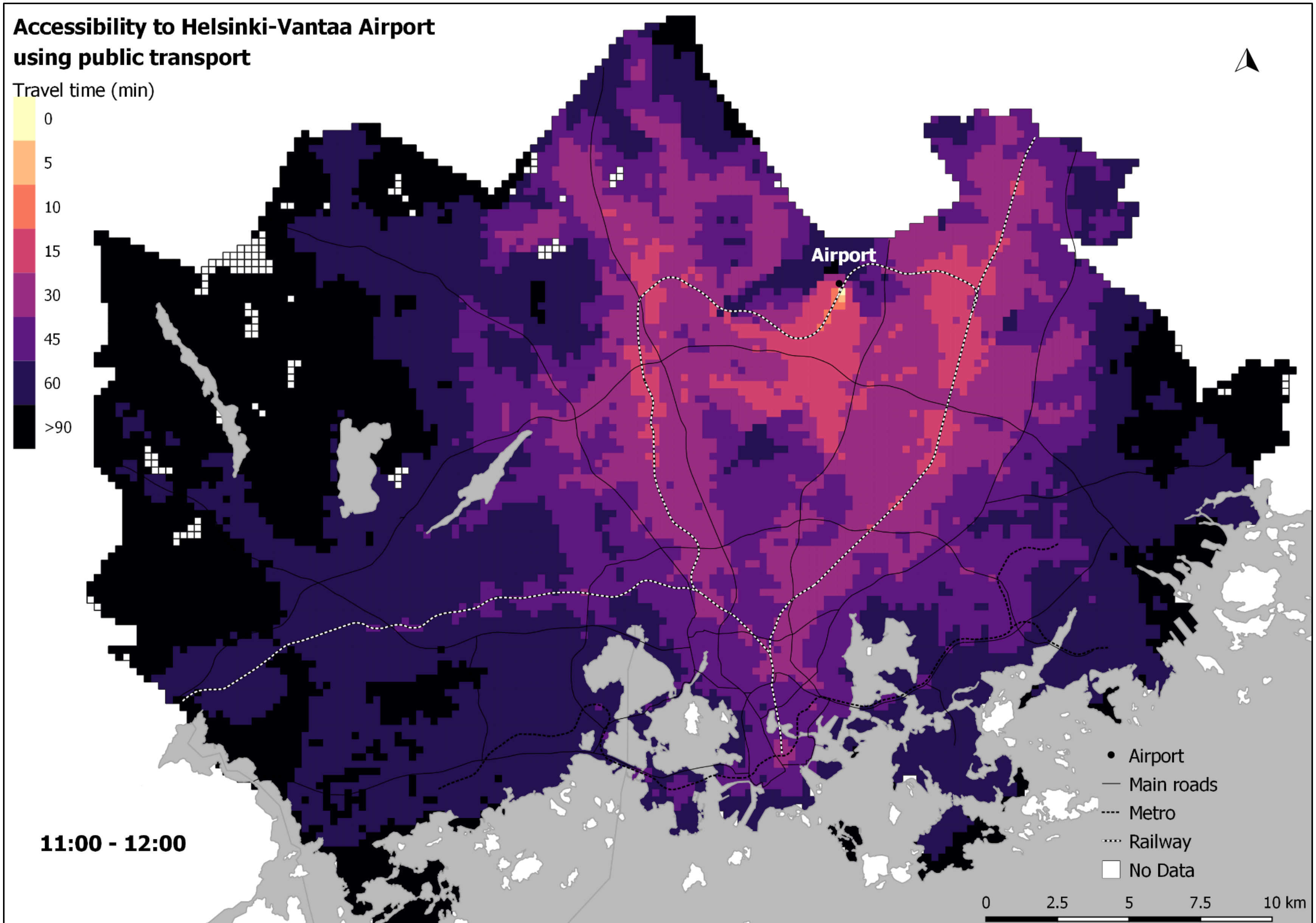

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
06:00 - 07:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
07:00 - 08:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

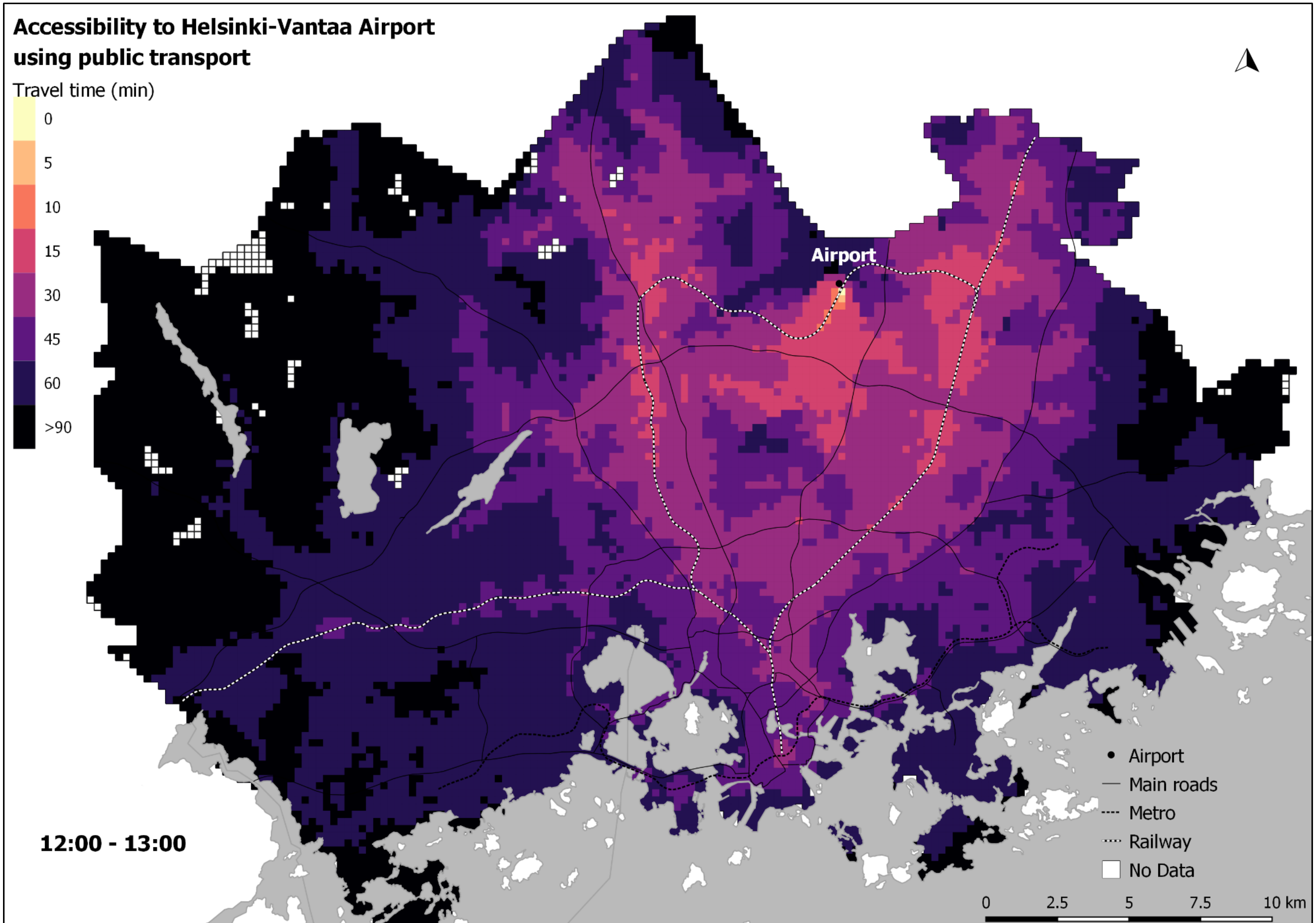

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
08:00 - 09:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
09:00 - 10:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

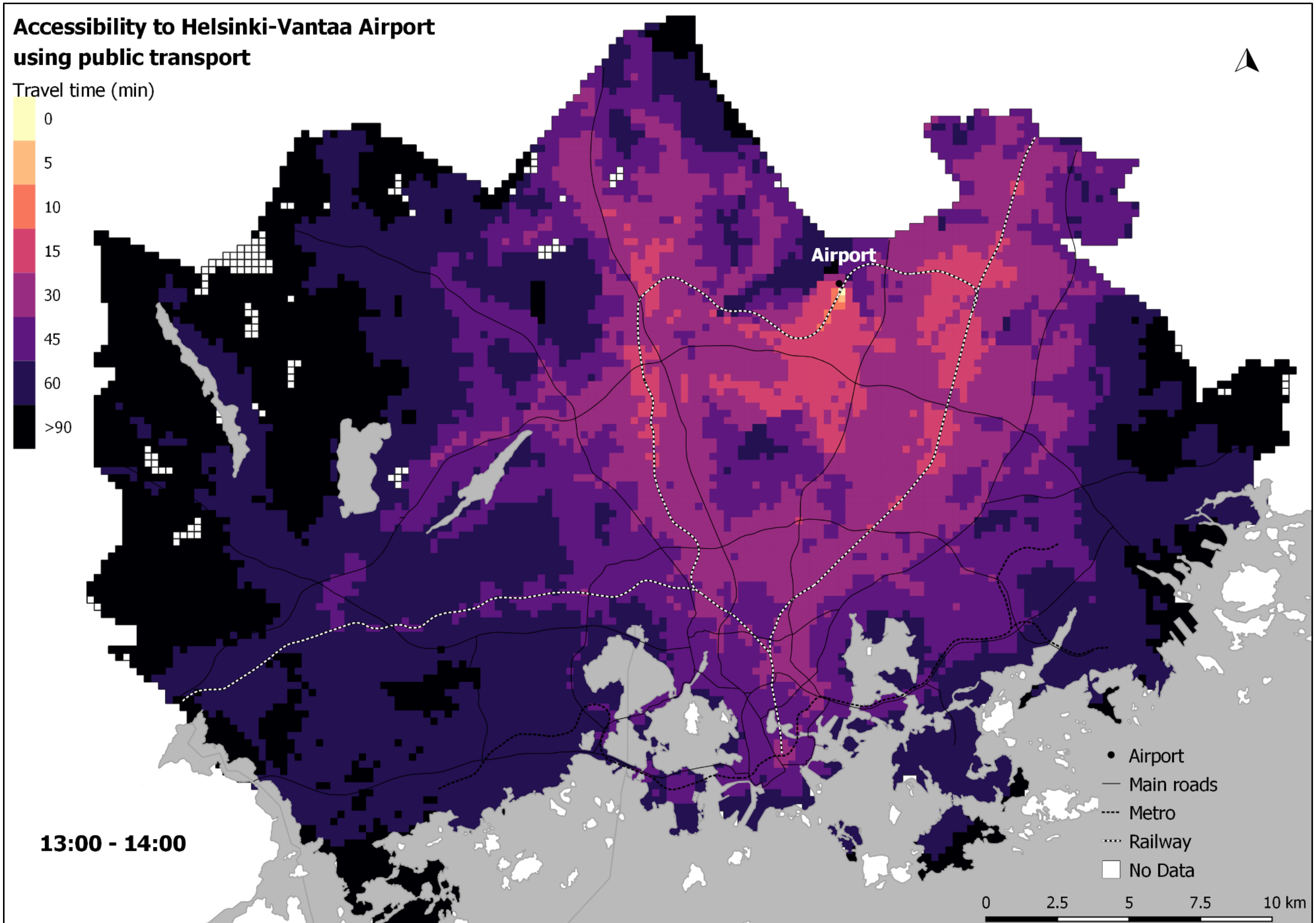

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
10:00 - 11:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

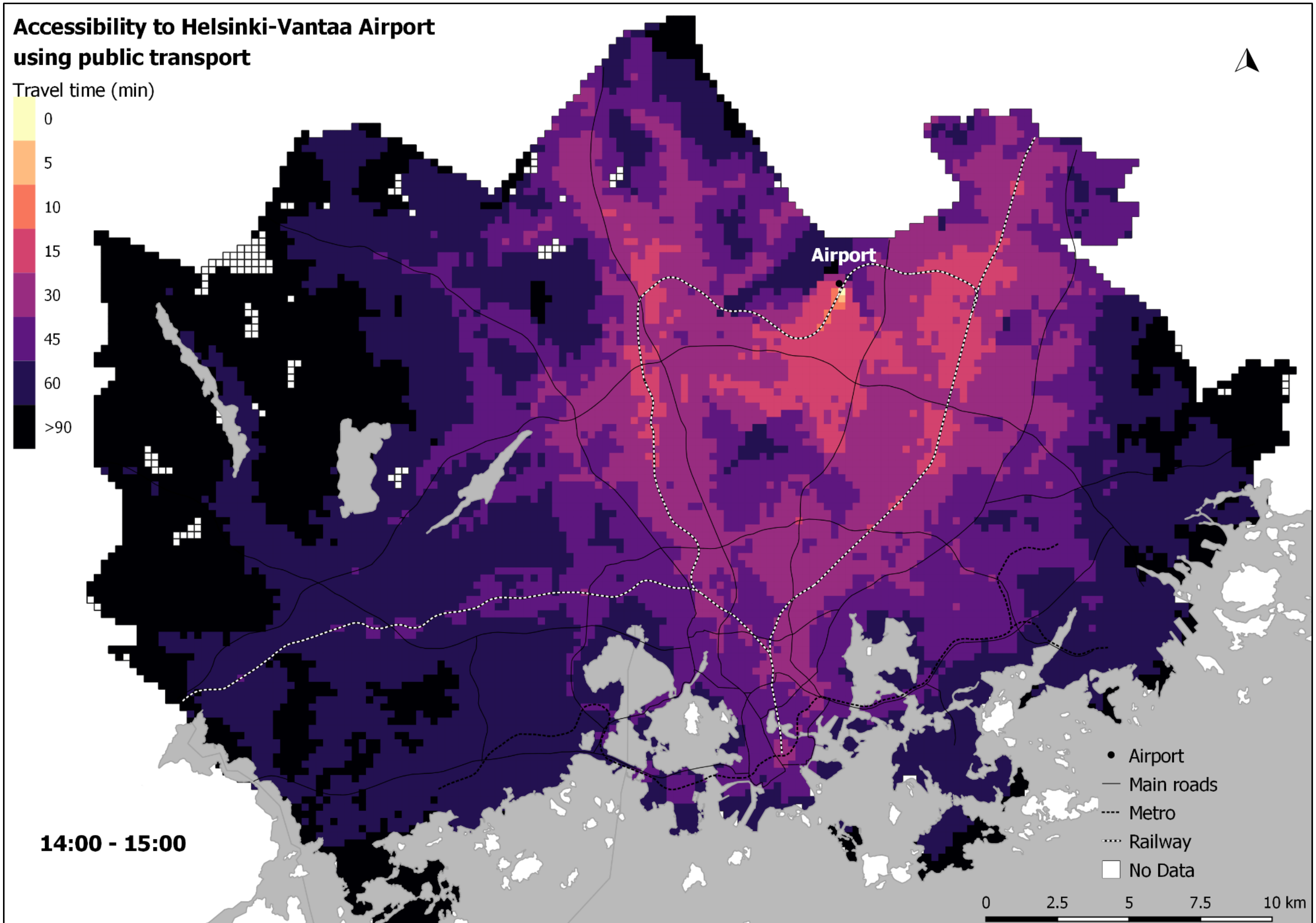


Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
11:00 - 12:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

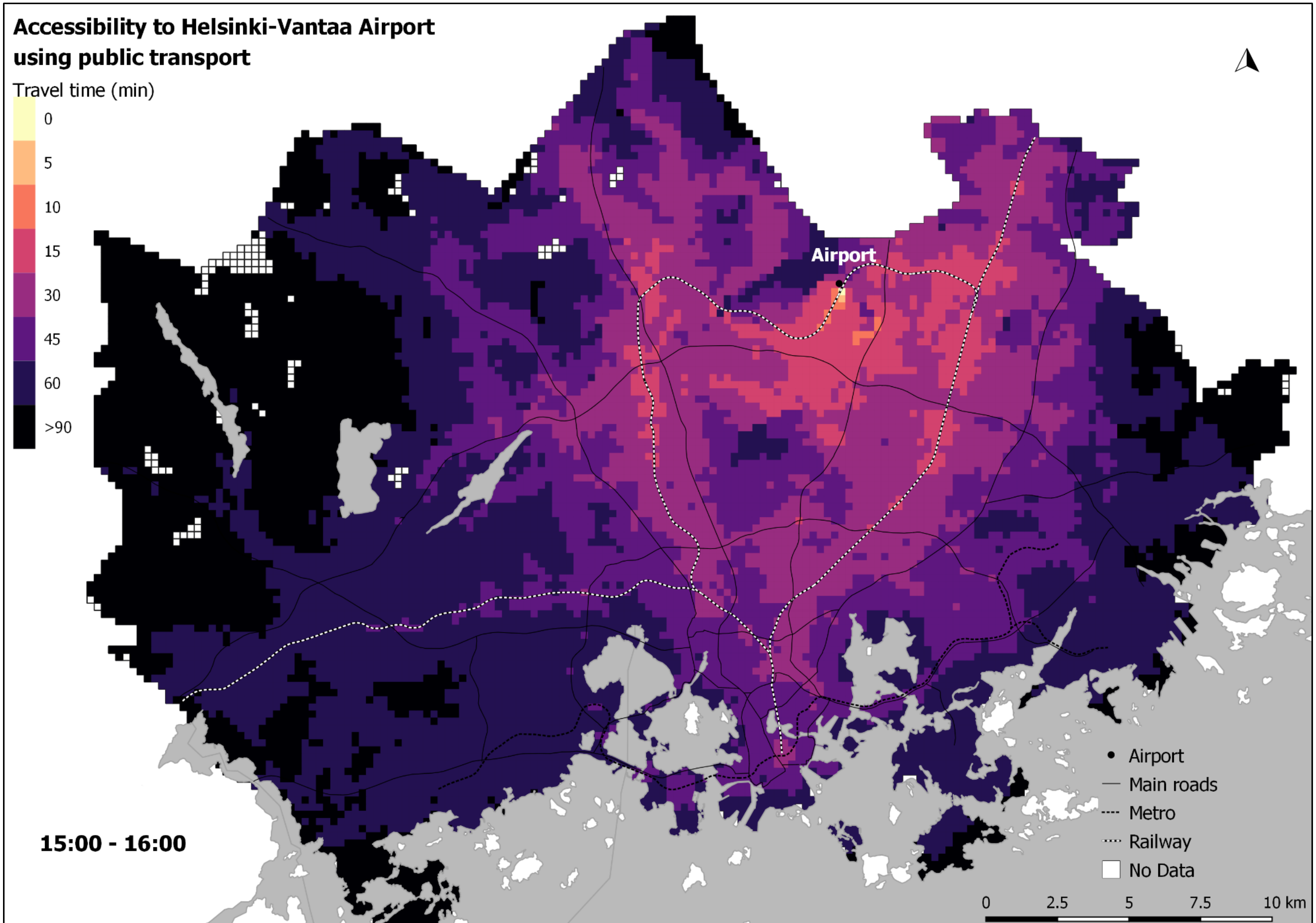


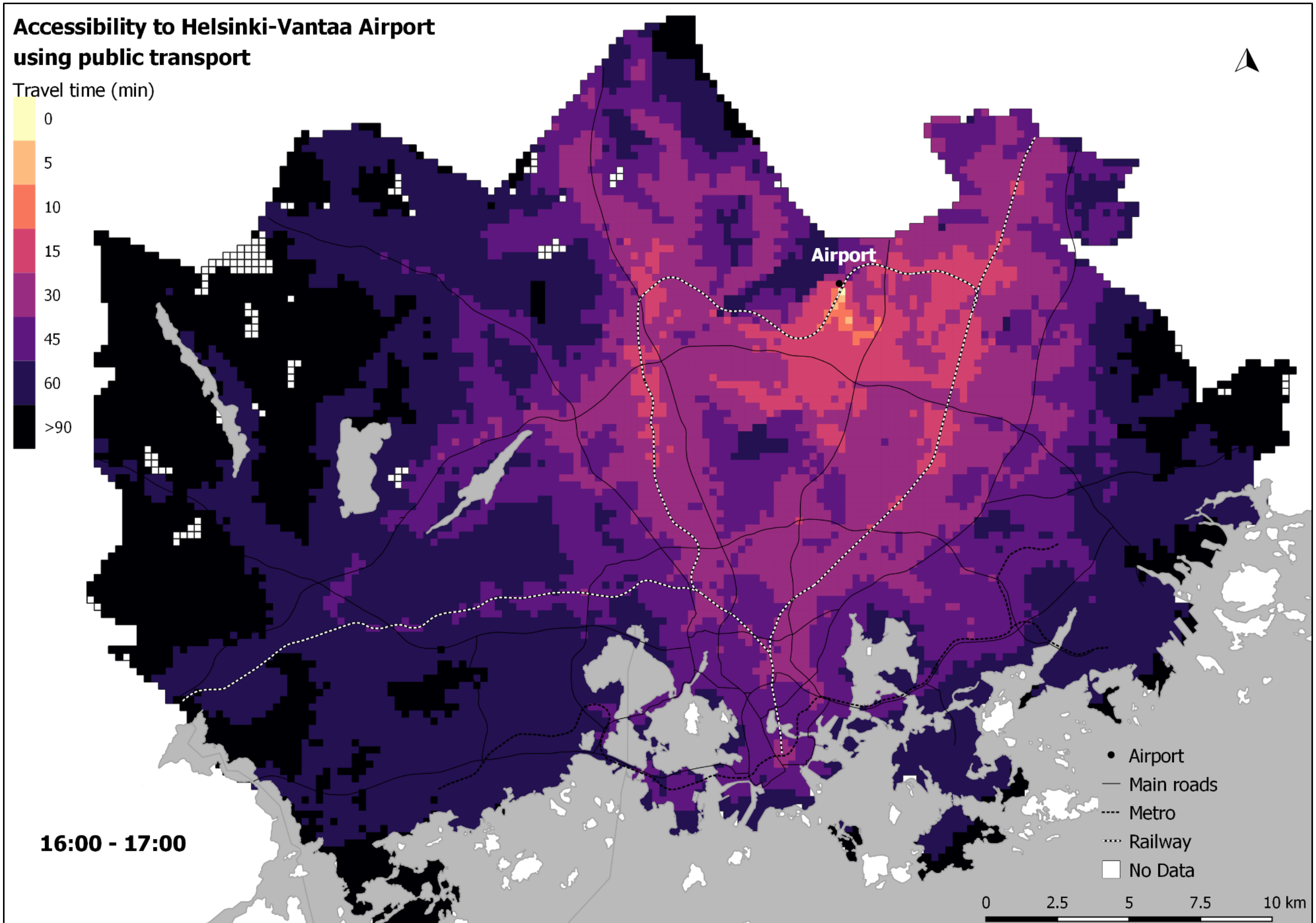
Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
12:00 - 13:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

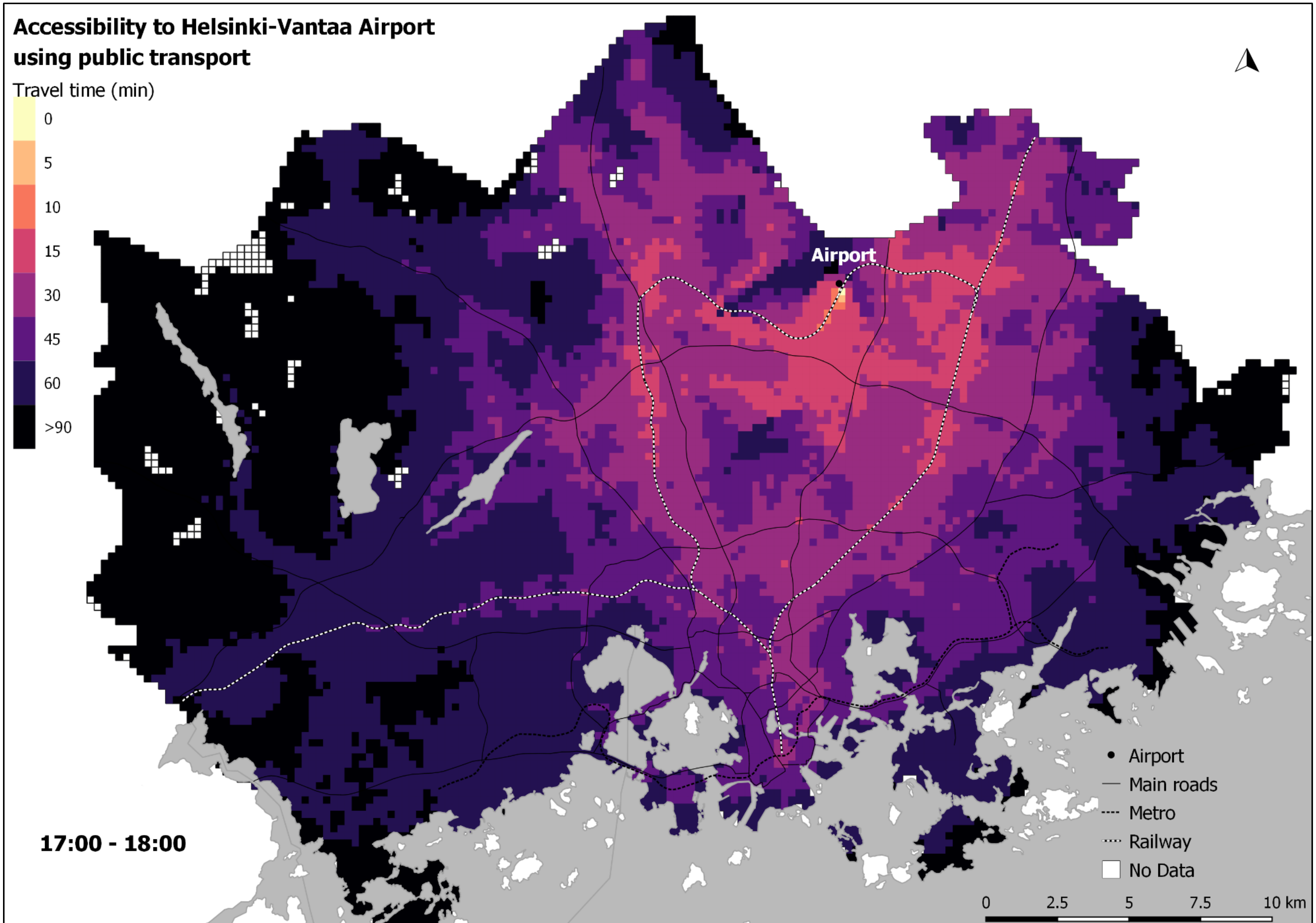

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
13:00 - 14:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

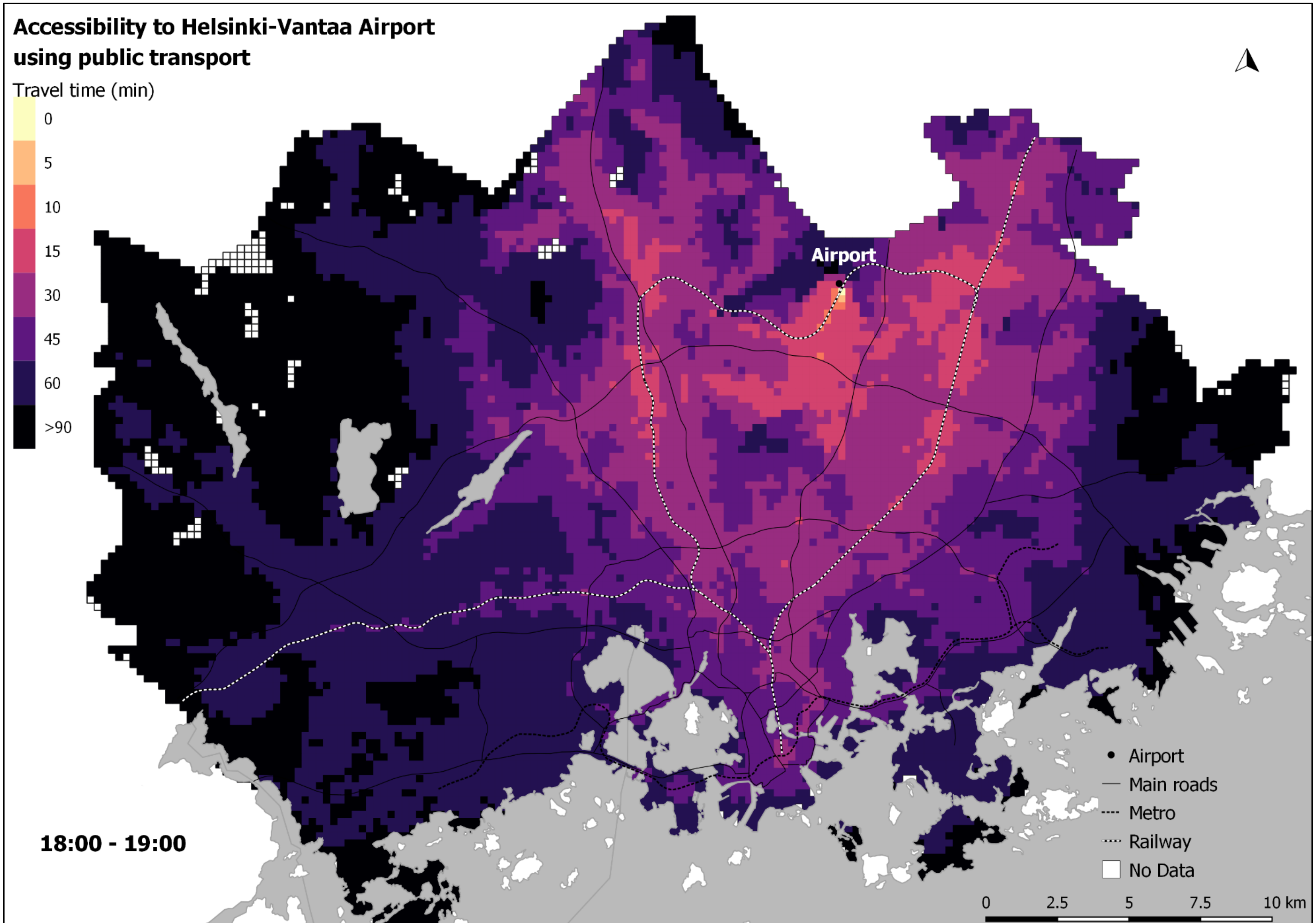


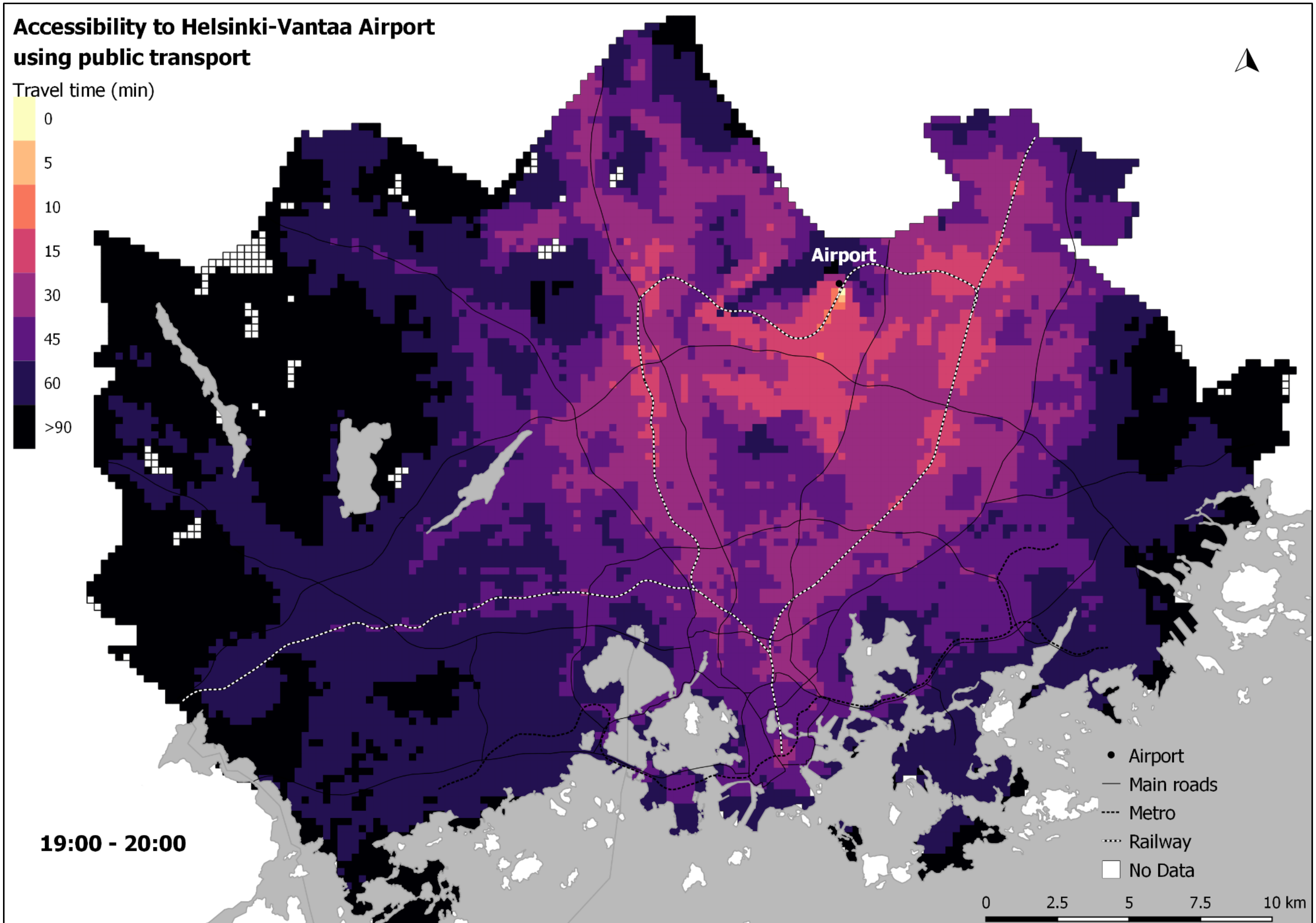
Accessibility to Helsinki-Vantaa Airport using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
14:00 - 15:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

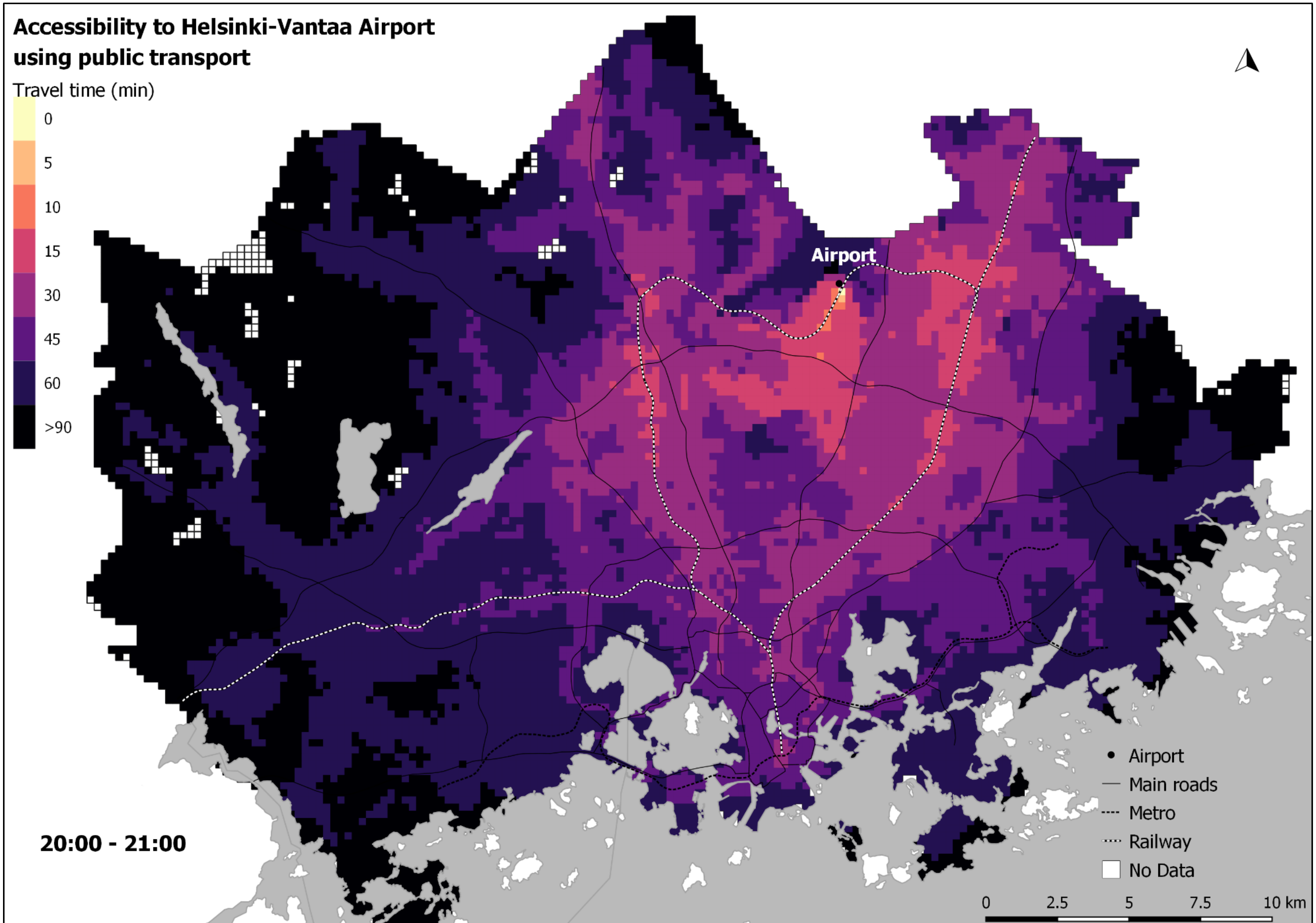

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
15:00 - 16:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

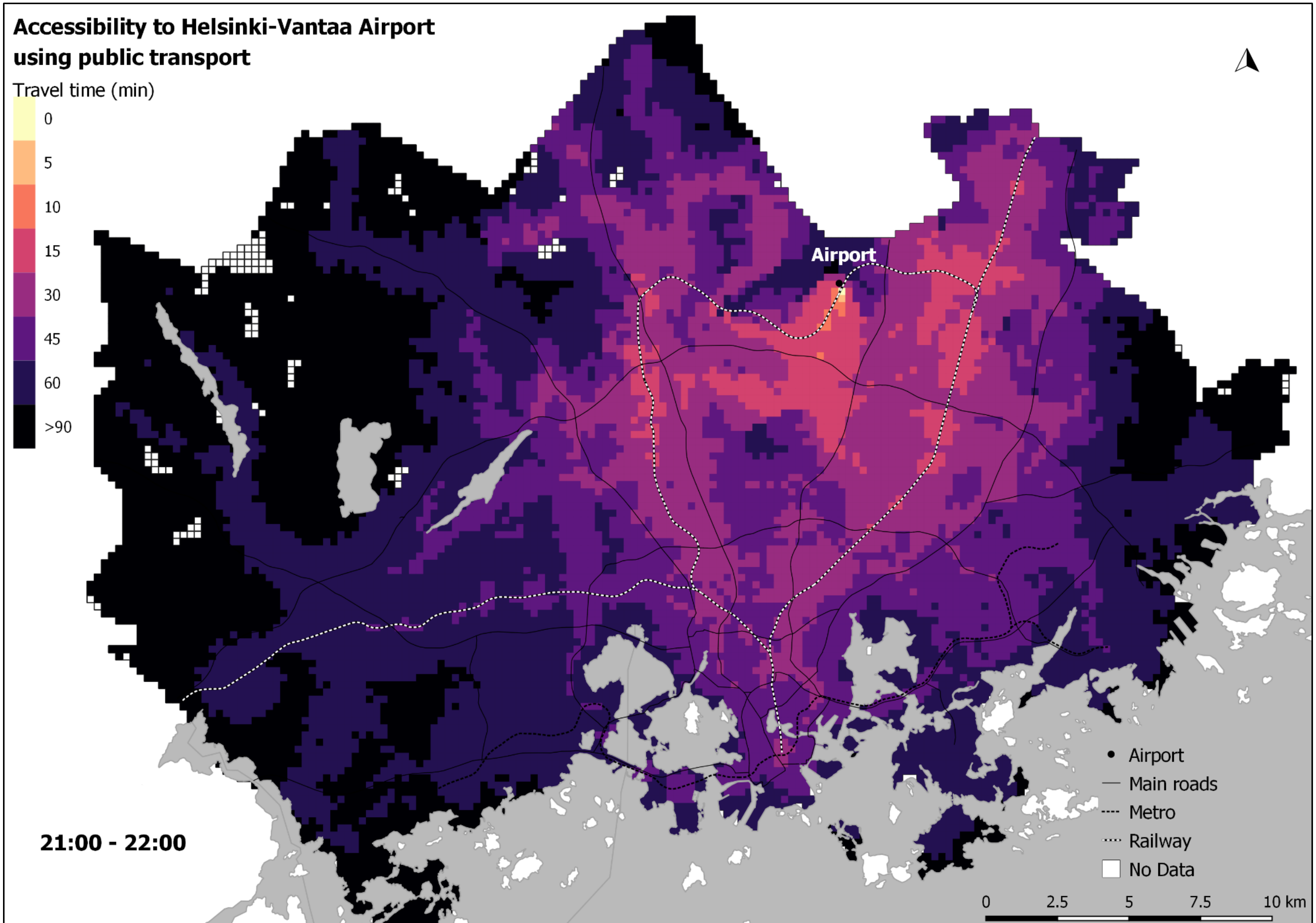

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
16:00 - 17:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

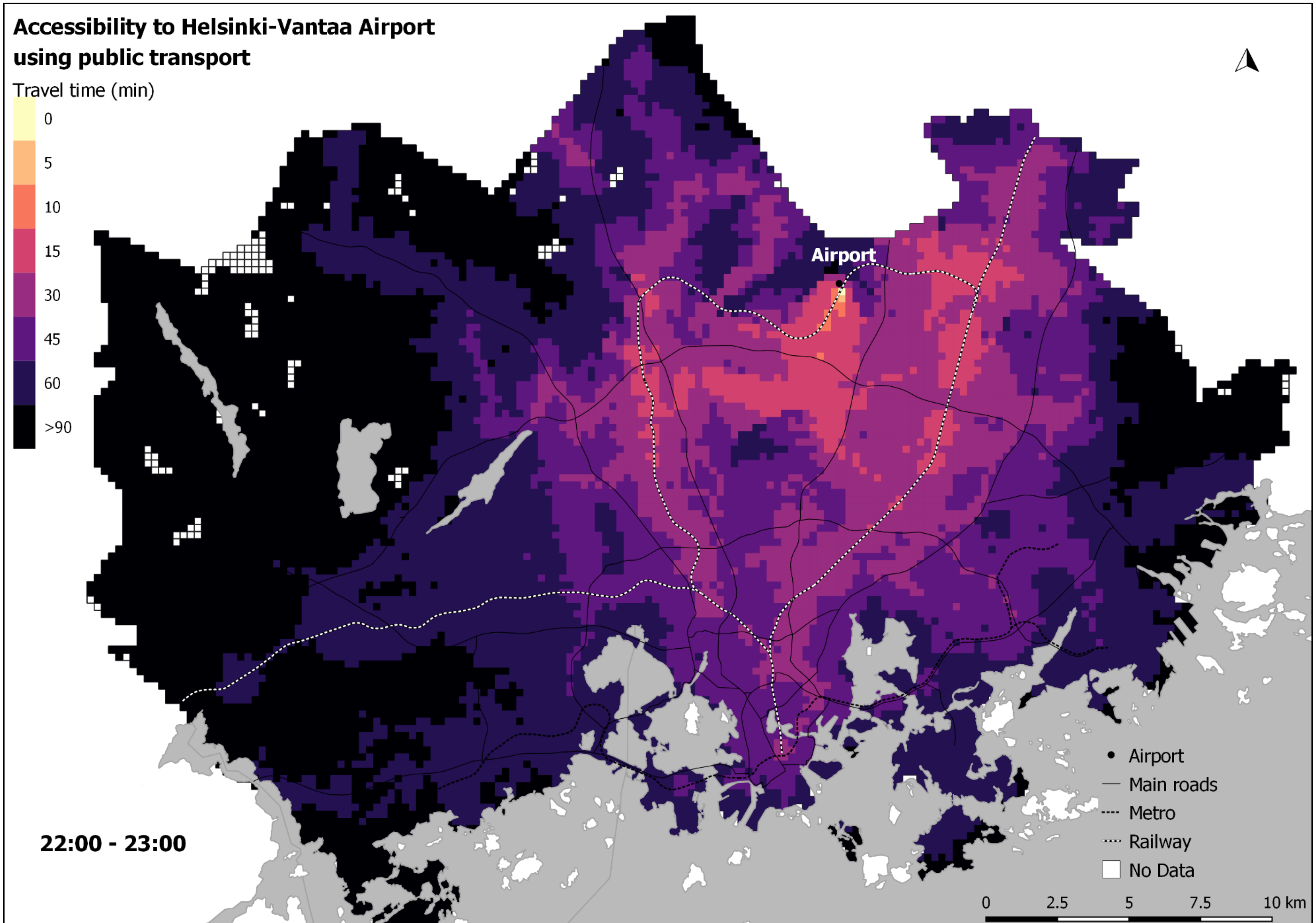

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
17:00 - 18:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

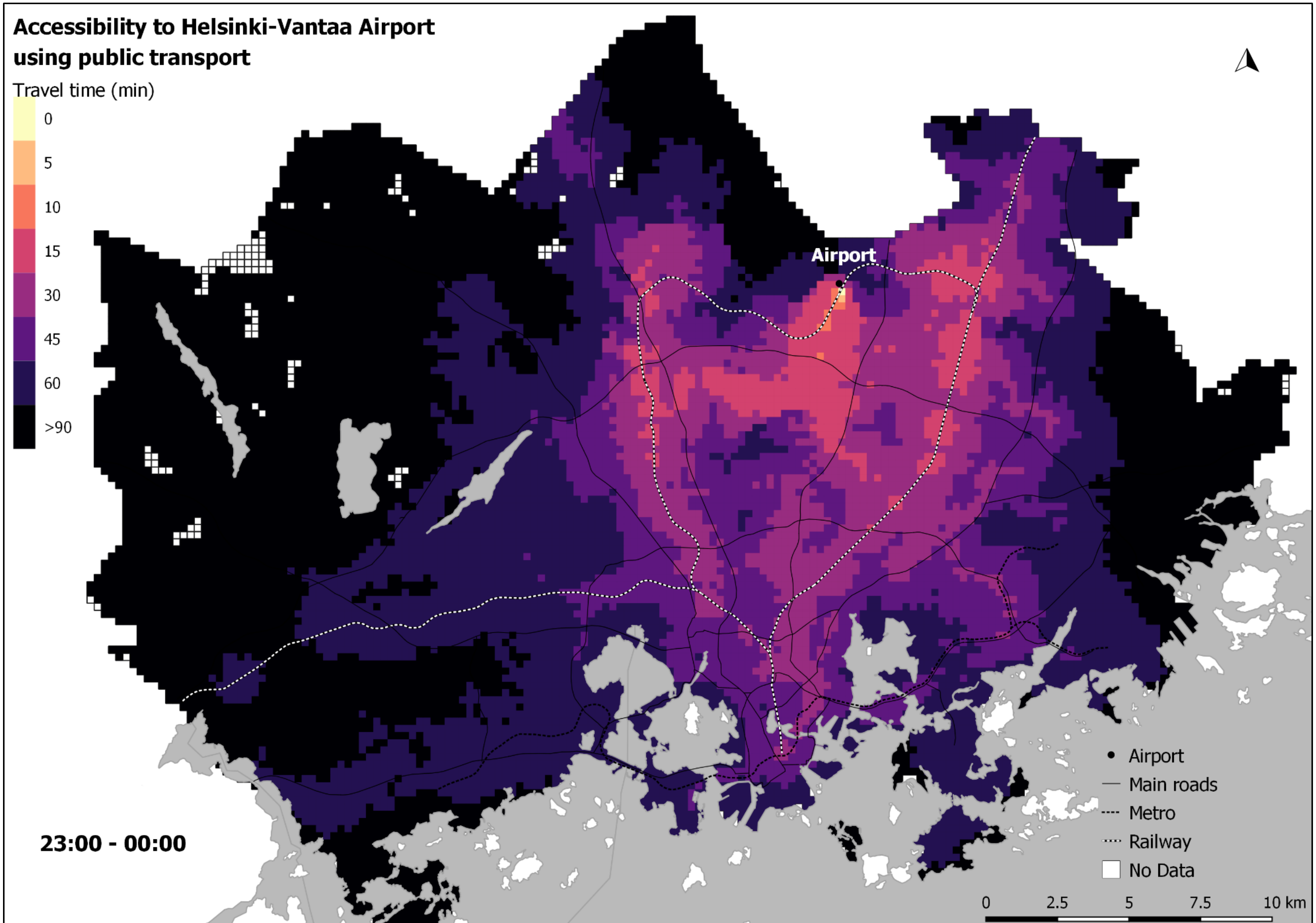

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
18:00 - 19:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
19:00 - 20:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

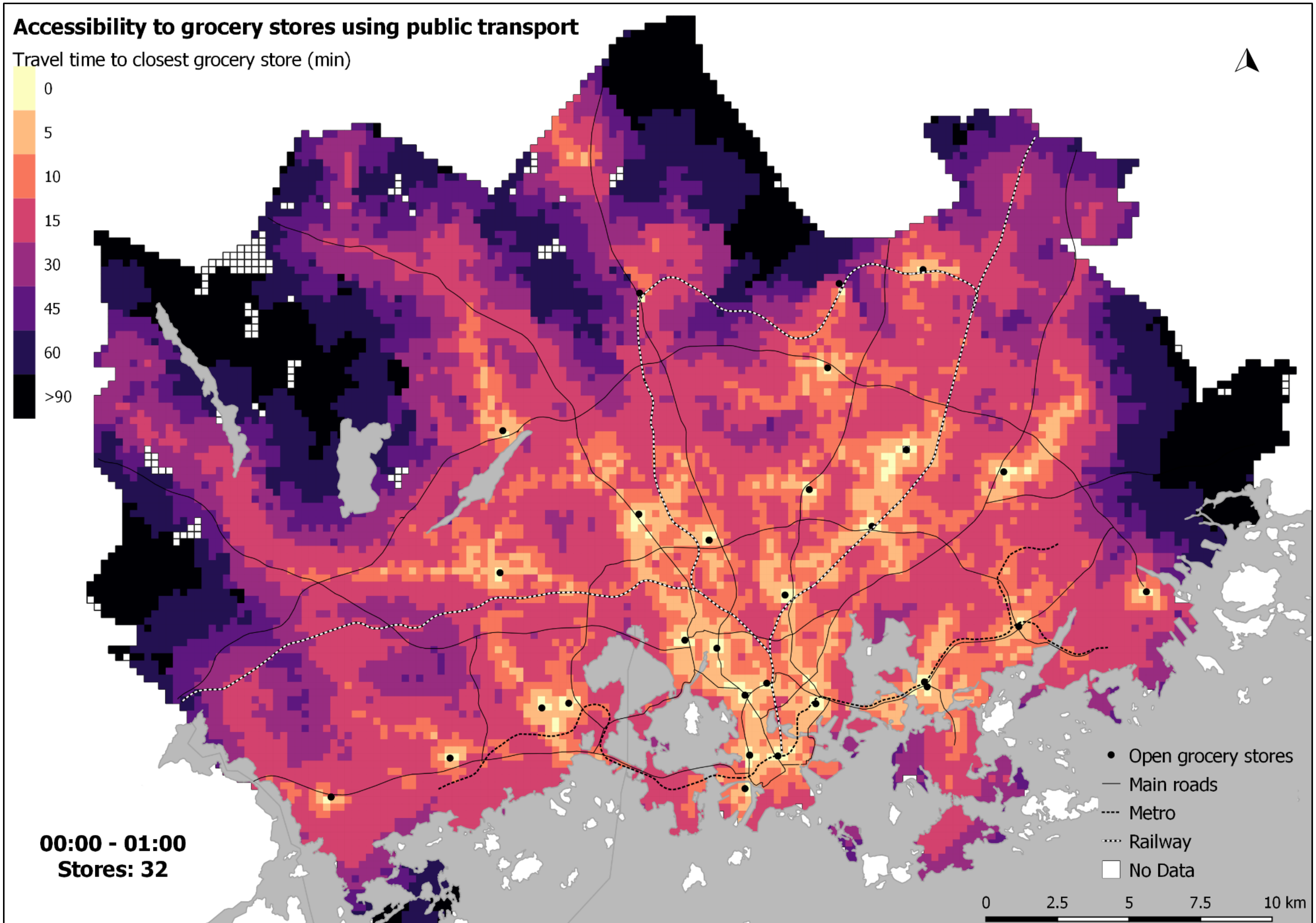

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
Airport
Main roads
Metro
Railway
No Data
20:00 - 21:00
0
2.5
5
7.5
10 km


Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
21:00 - 22:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

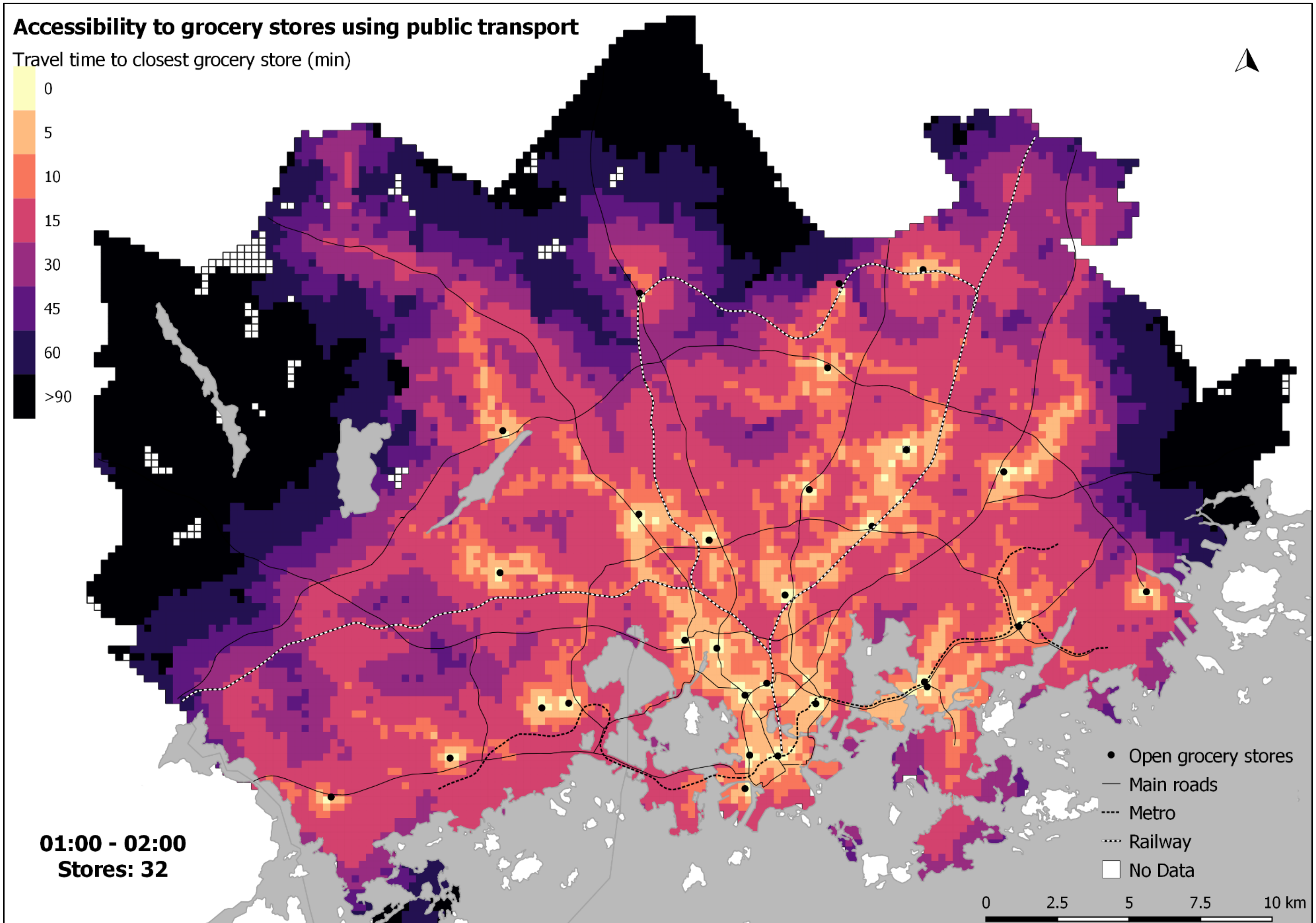

Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
22:00 - 23:00

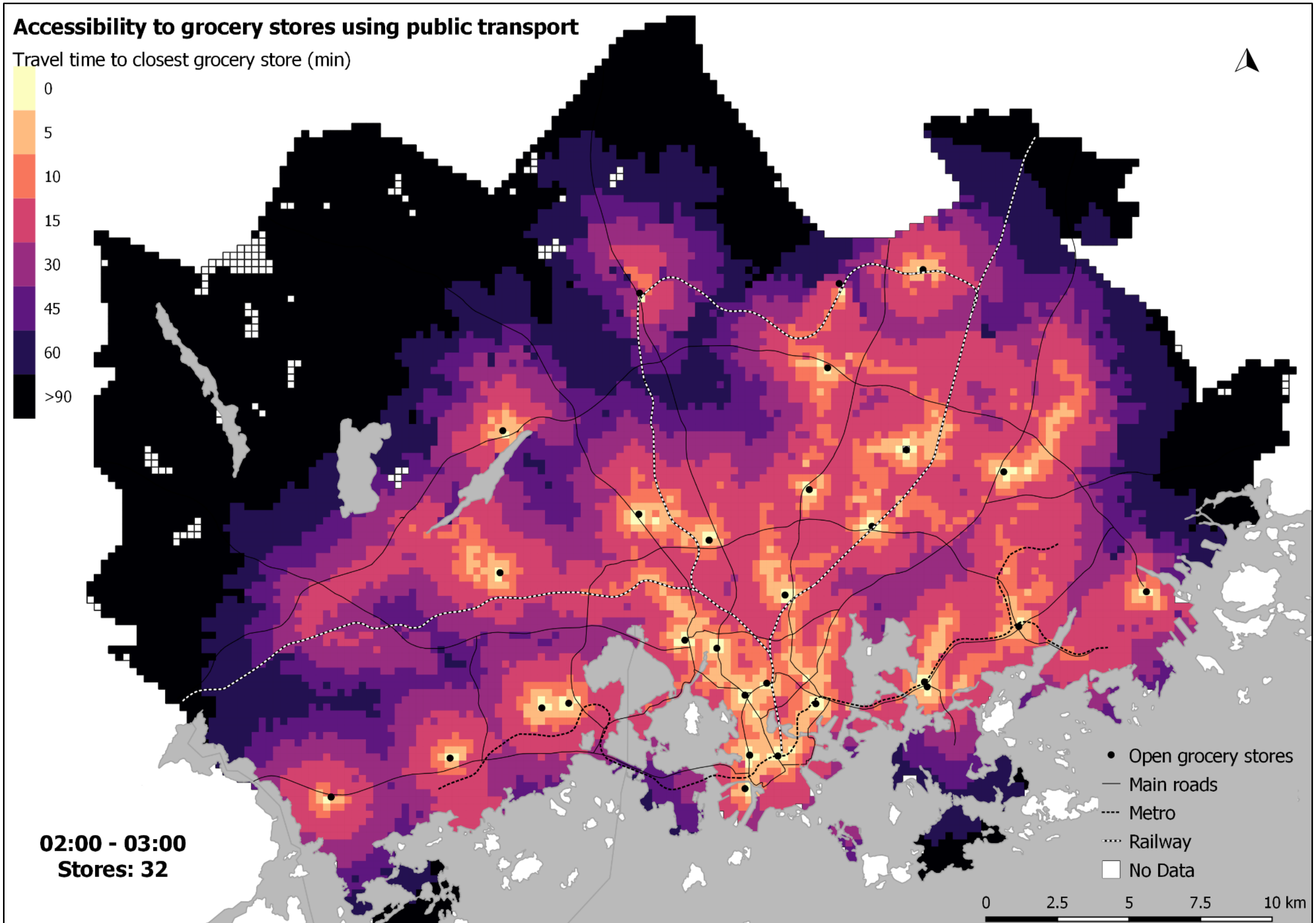


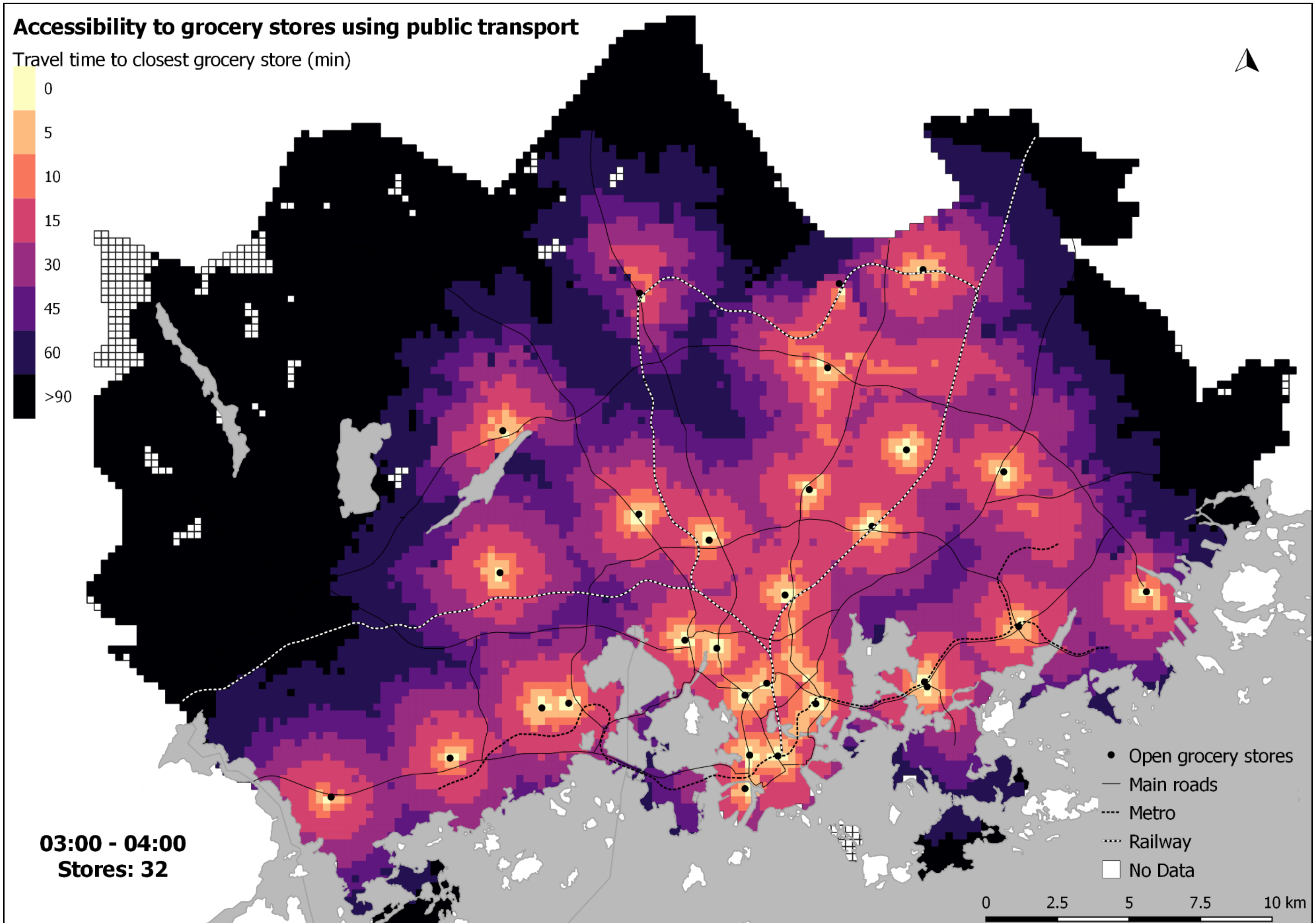
Accessibility to Helsinki-Vantaa Airport
using public transport
Travel time (min)
0
5
10
15
30
45
60
>90
Airport
23:00 - 00:00
Airport
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

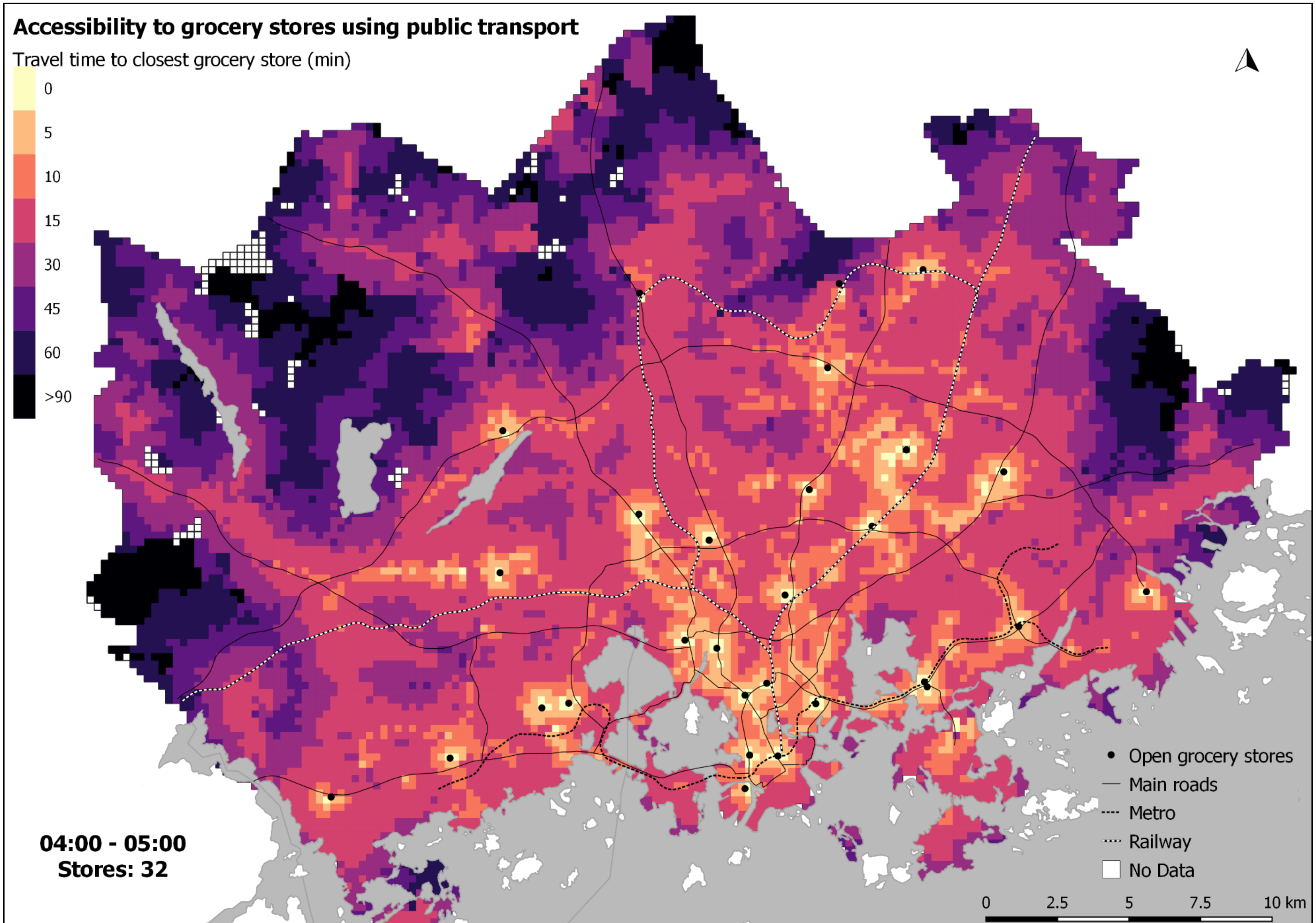


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
00:00 - 01:00
Stores: 32
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

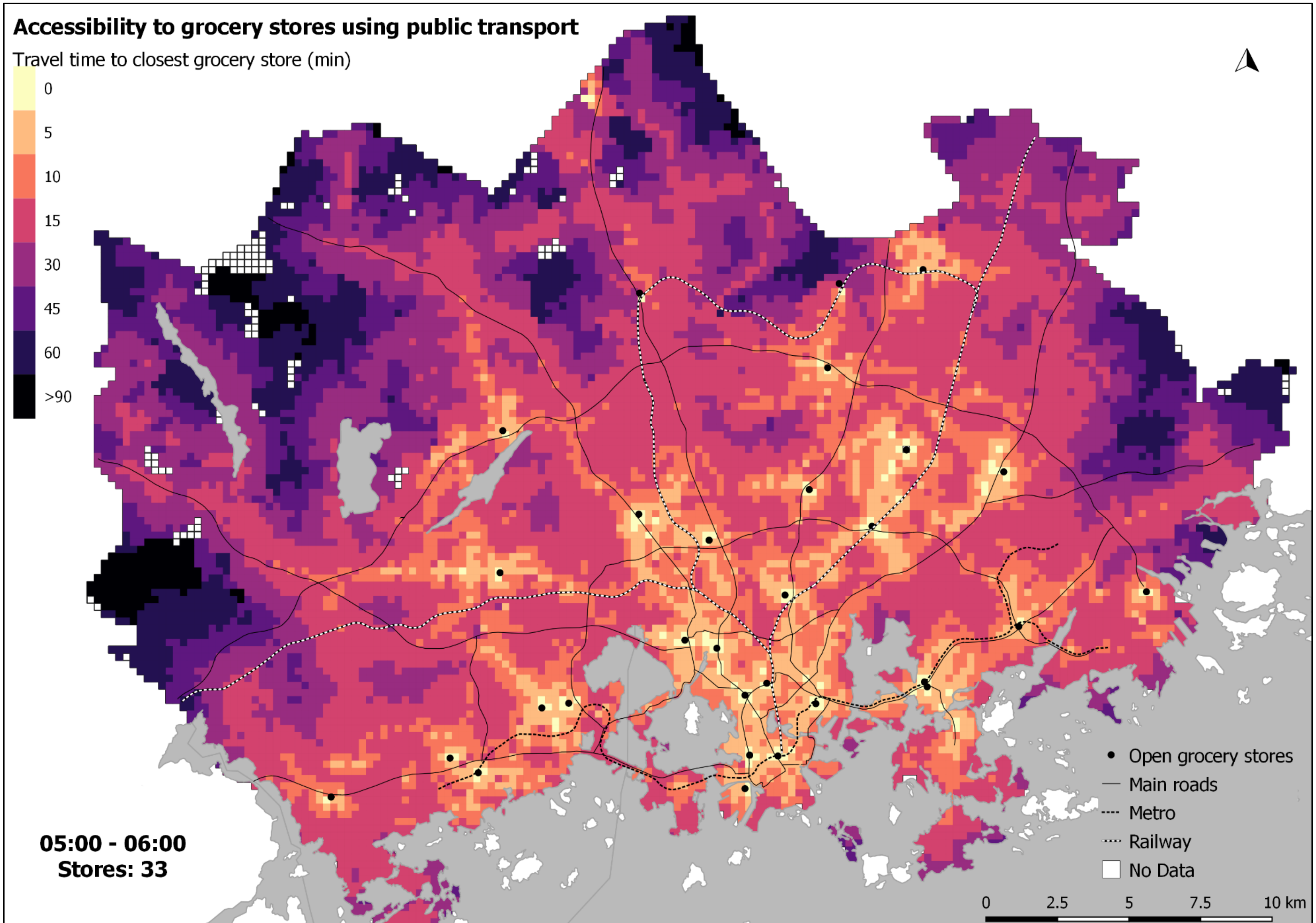

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
01:00 - 02:00
Stores: 32

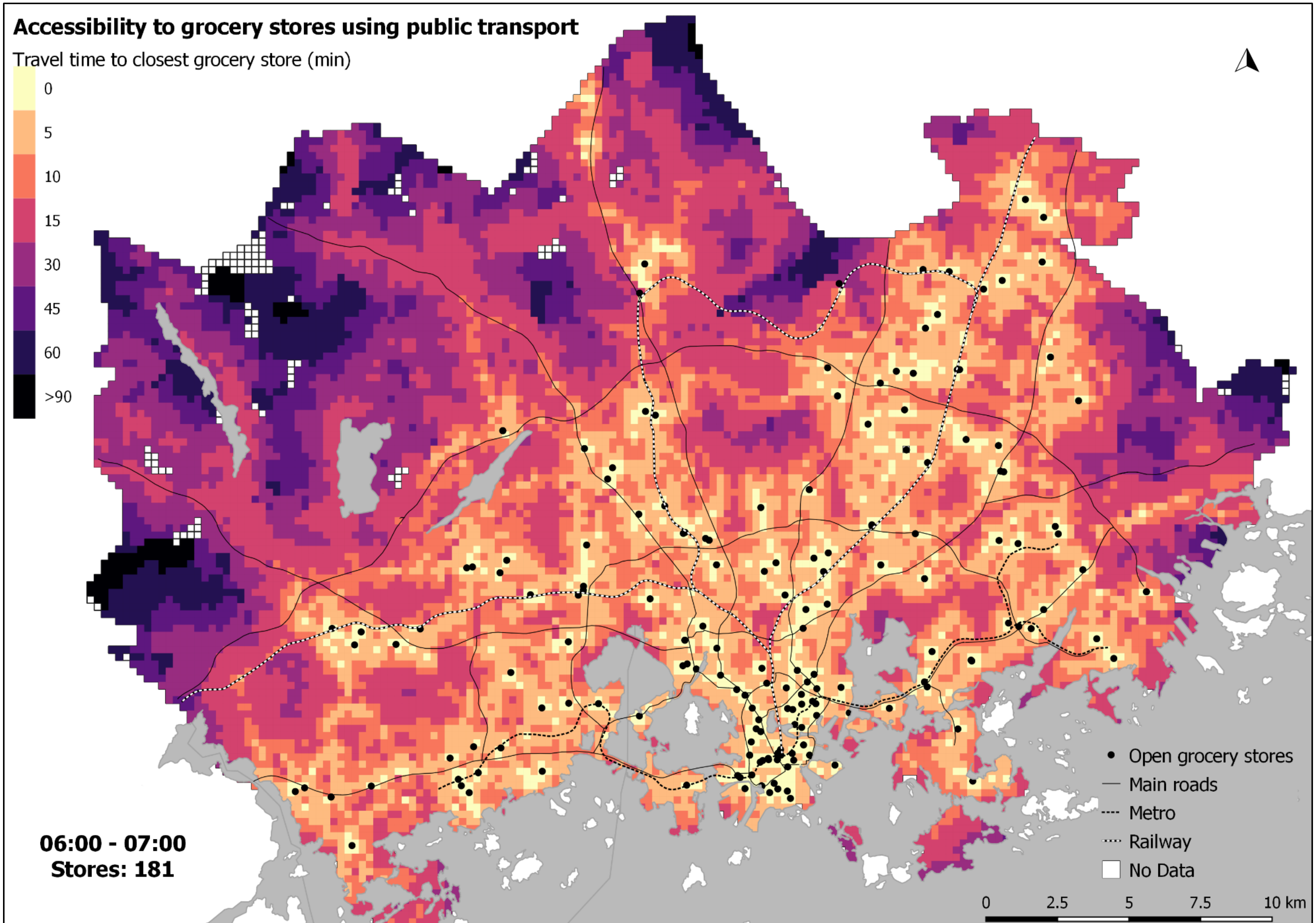

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
02:00 - 03:00
Stores: 32
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

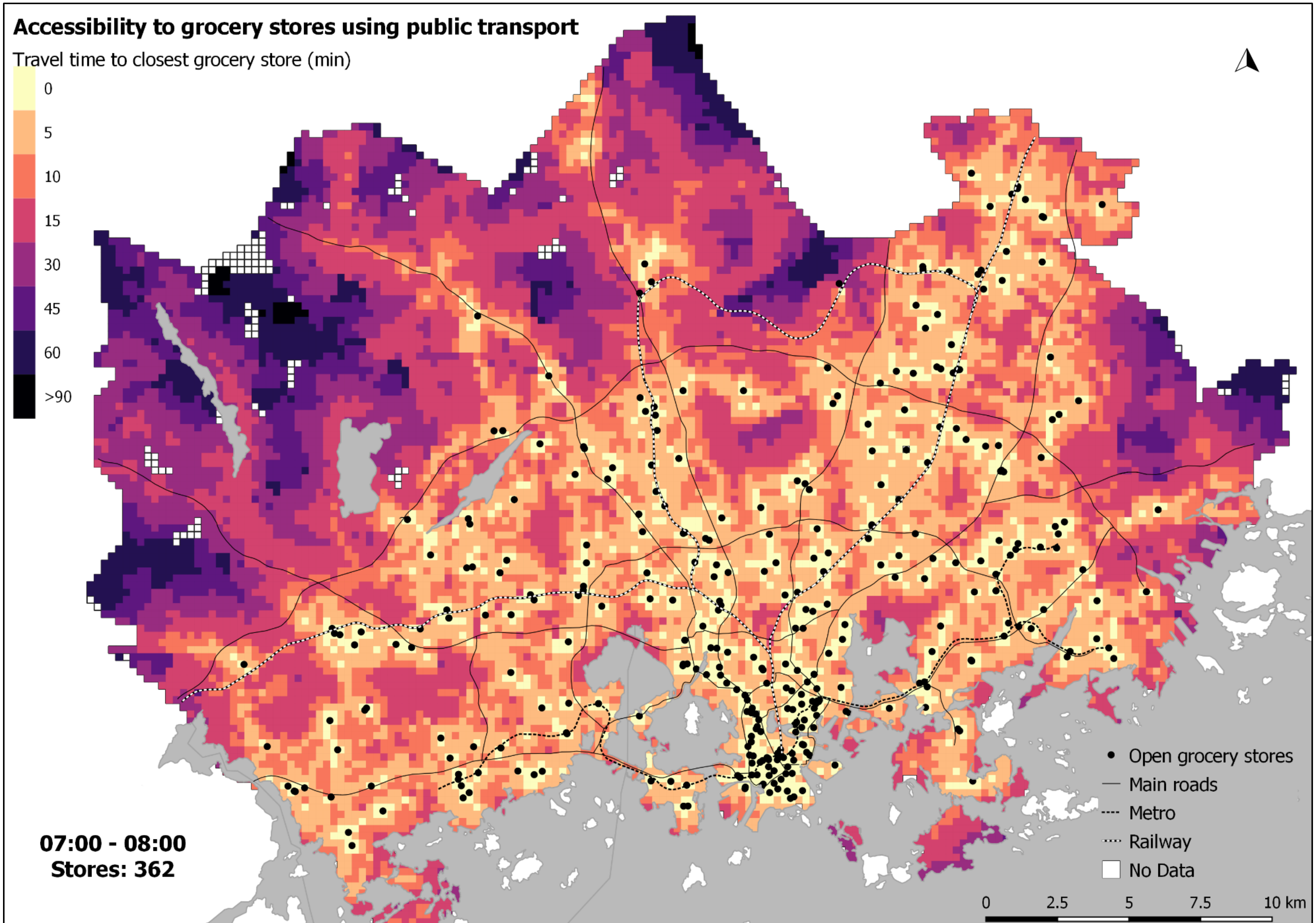

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
03:00 - 04:00
Stores: 32

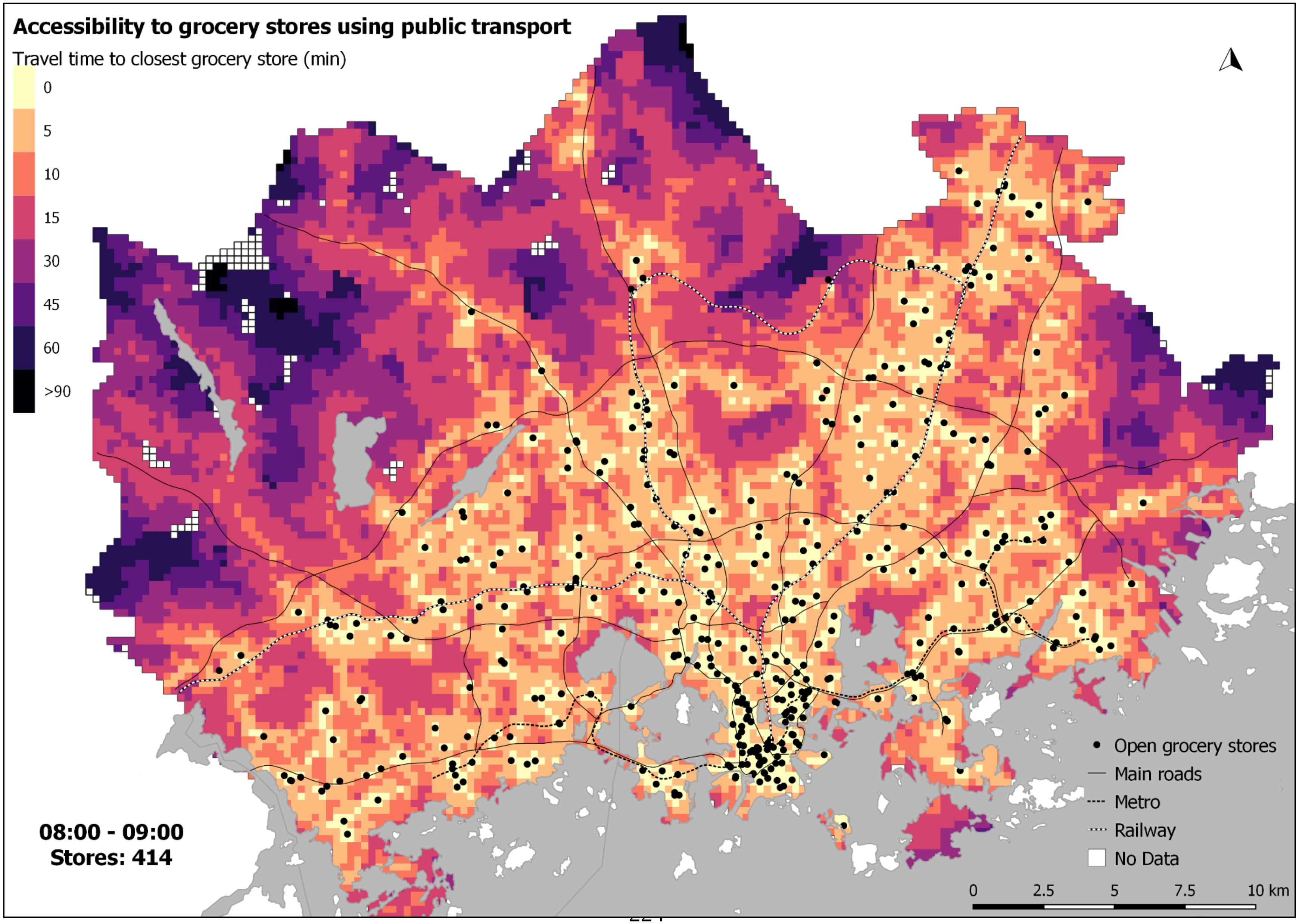


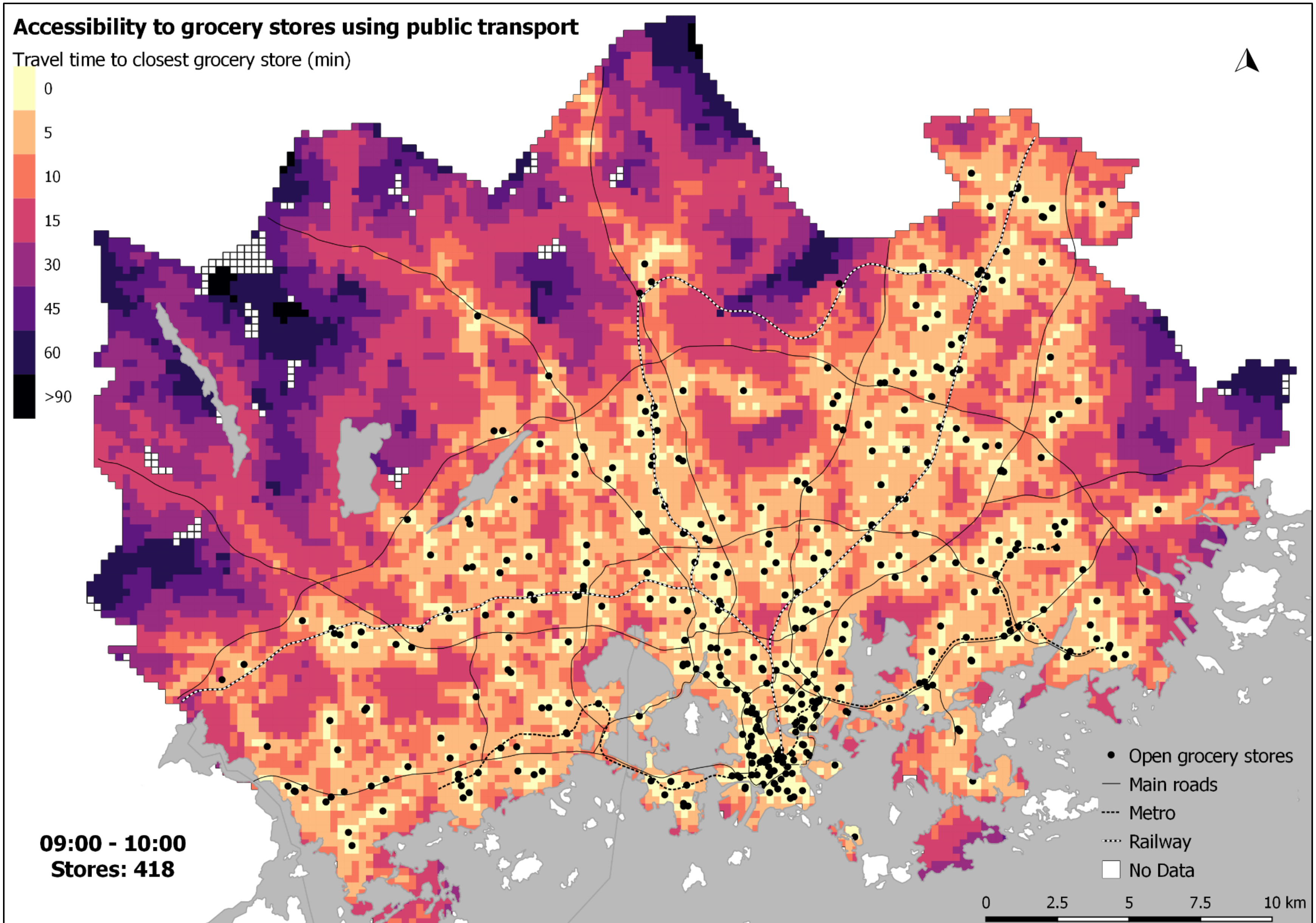
Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
04:00 - 05:00
Stores: 32
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

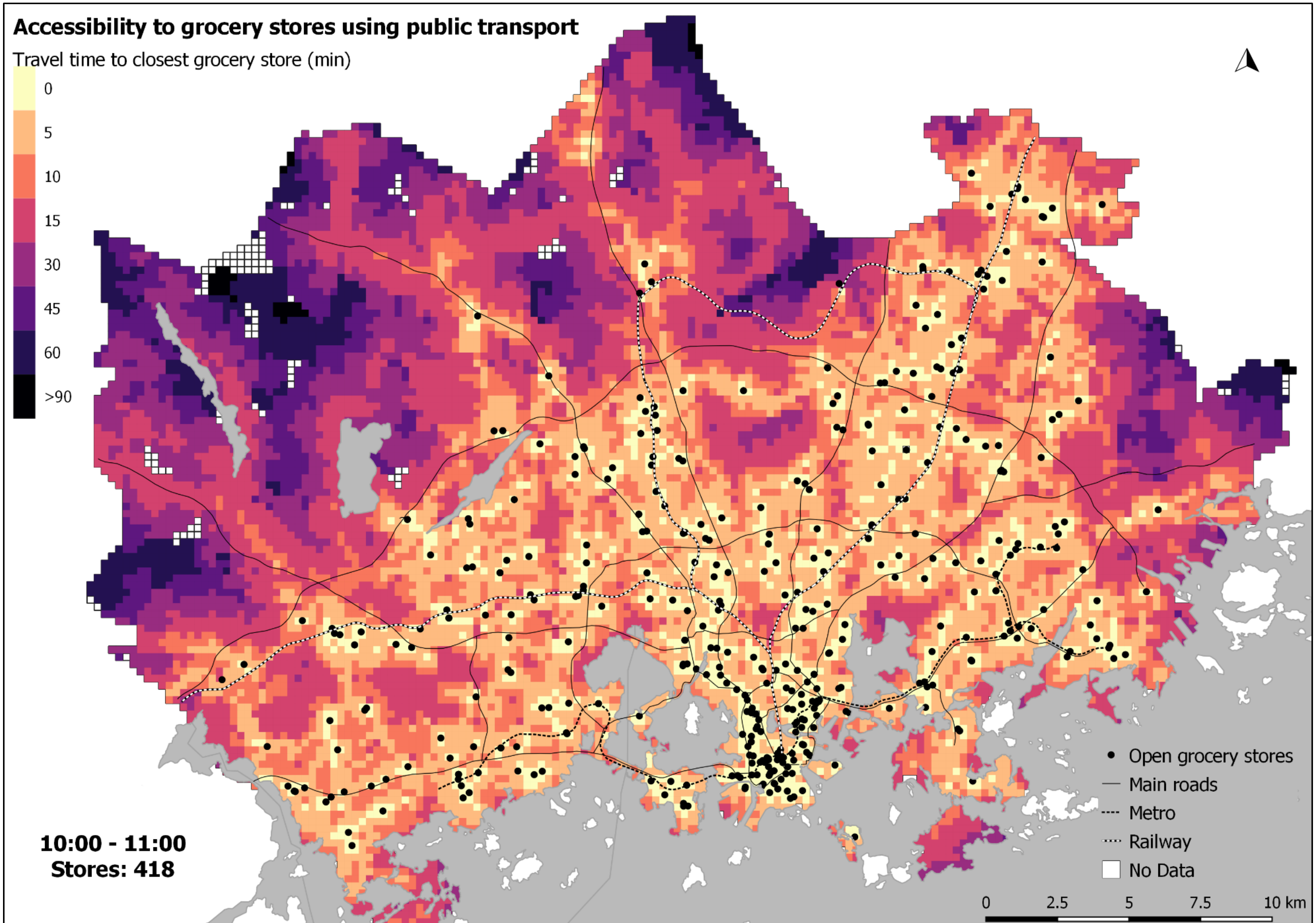

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
05:00 - 06:00
Stores: 33

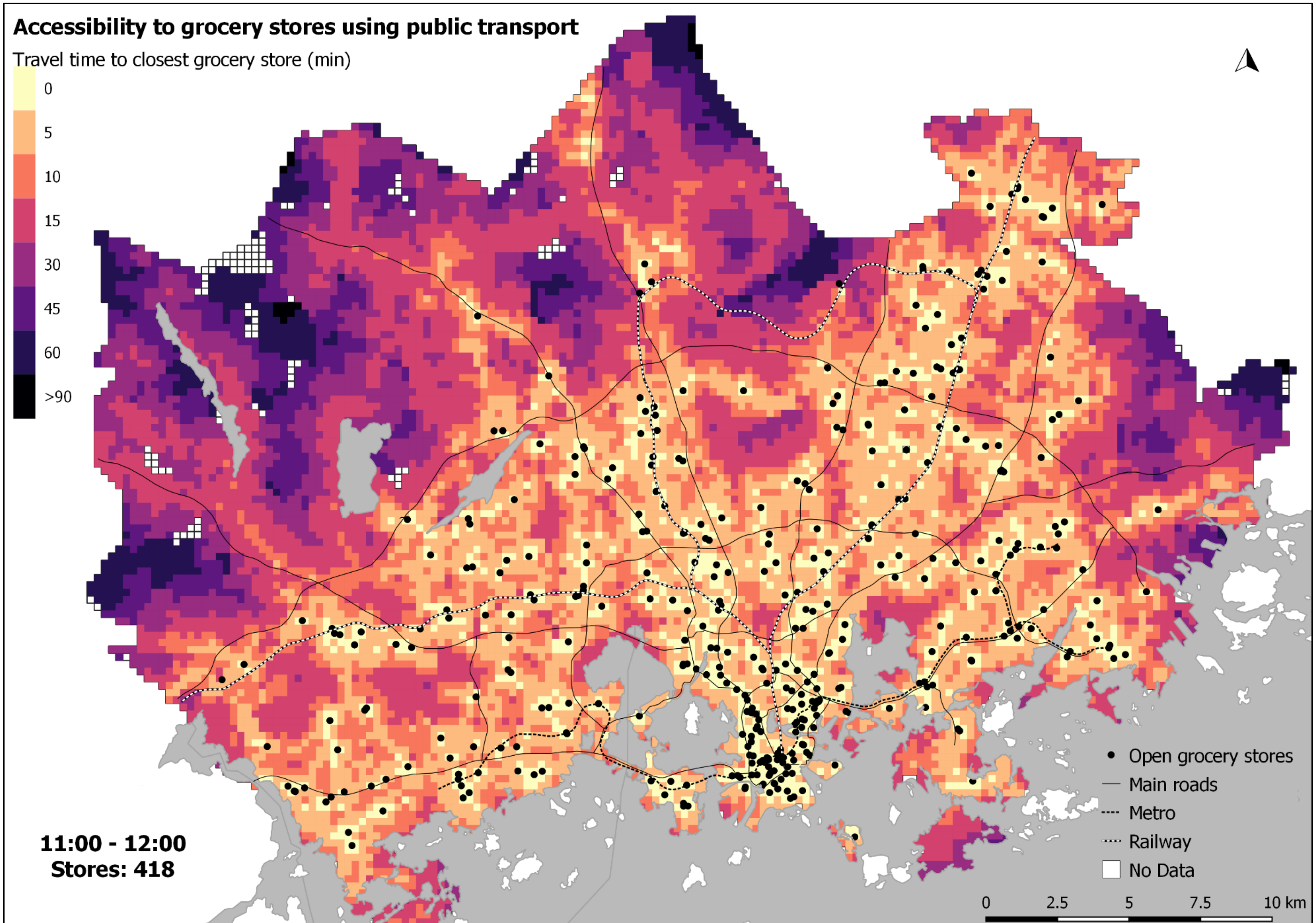

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
06:00 - 07:00
Stores: 181

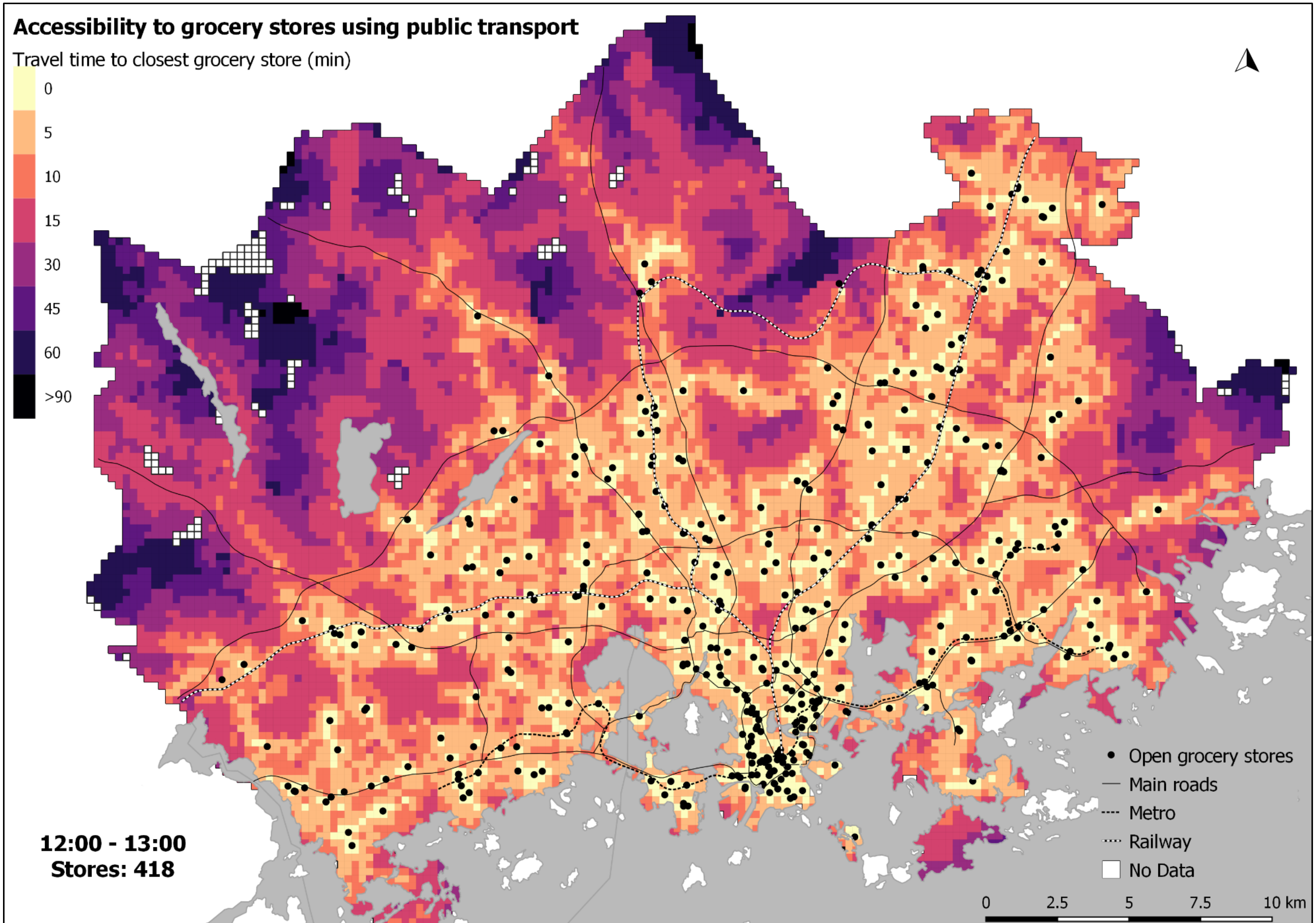

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
07:00 - 08:00
Stores: 362


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
08:00 - 09:00
Stores: 414

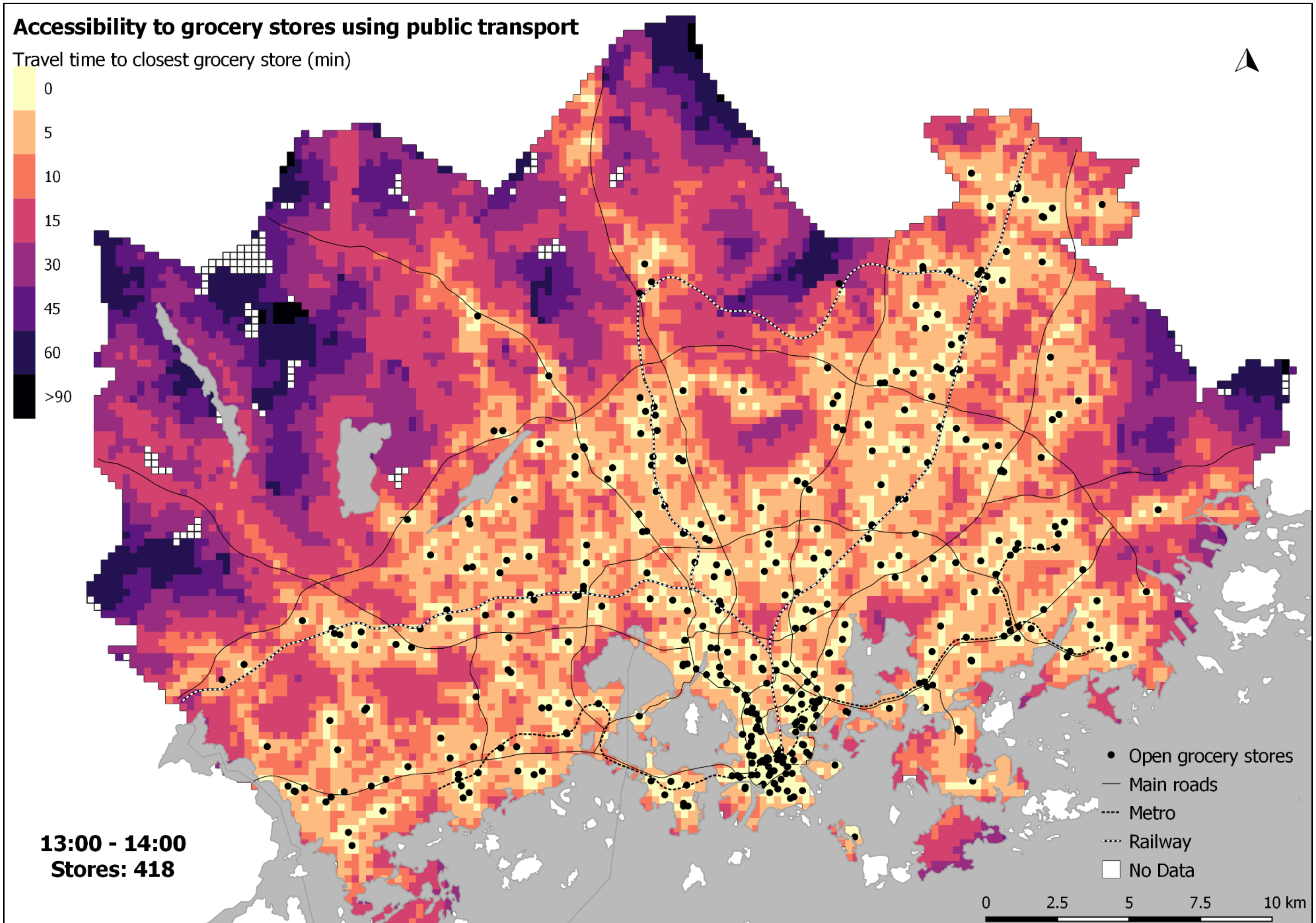

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
09:00 - 10:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

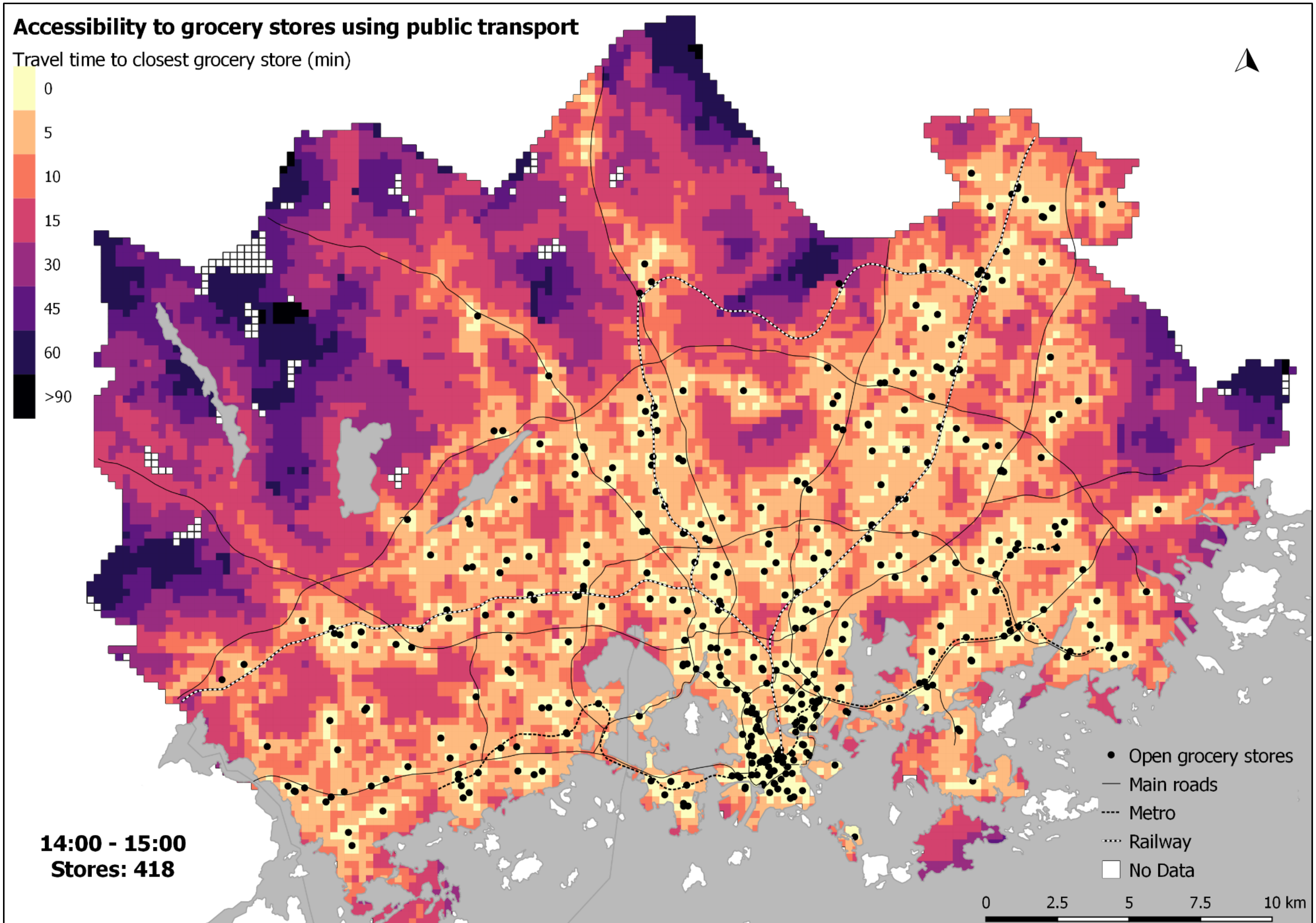

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
10:00 - 11:00
Stores: 418

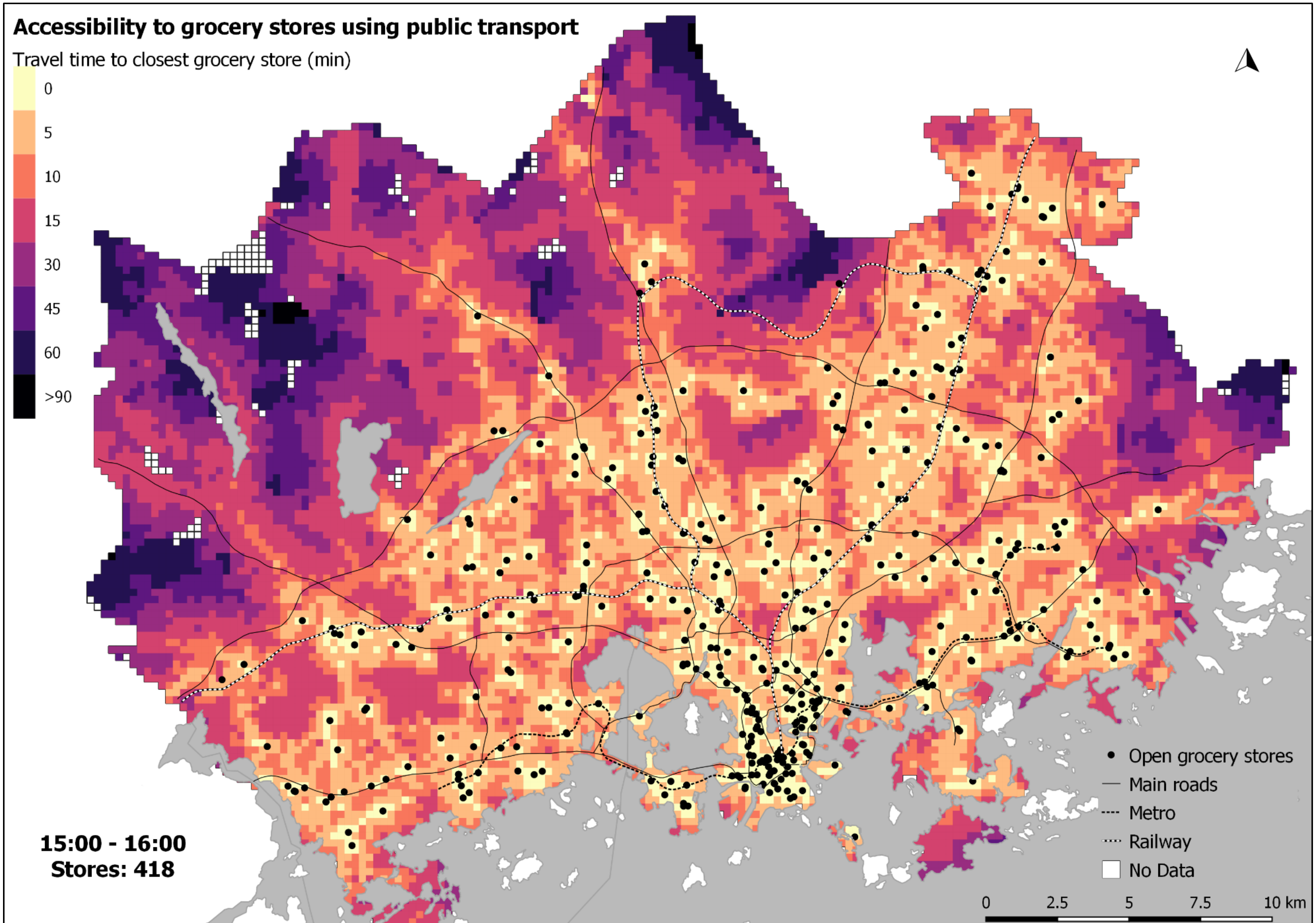

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
11:00 - 12:00
Stores: 418

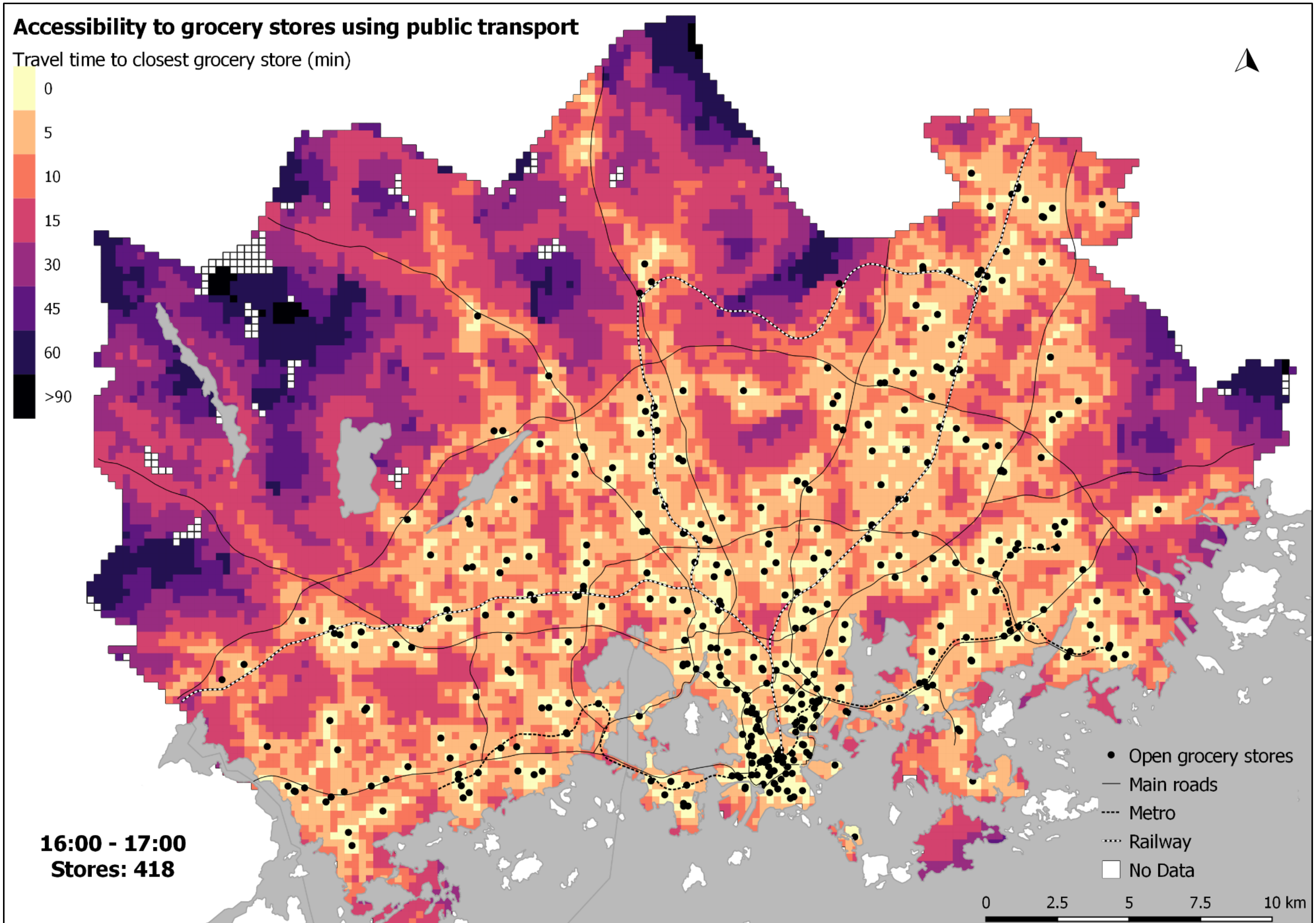


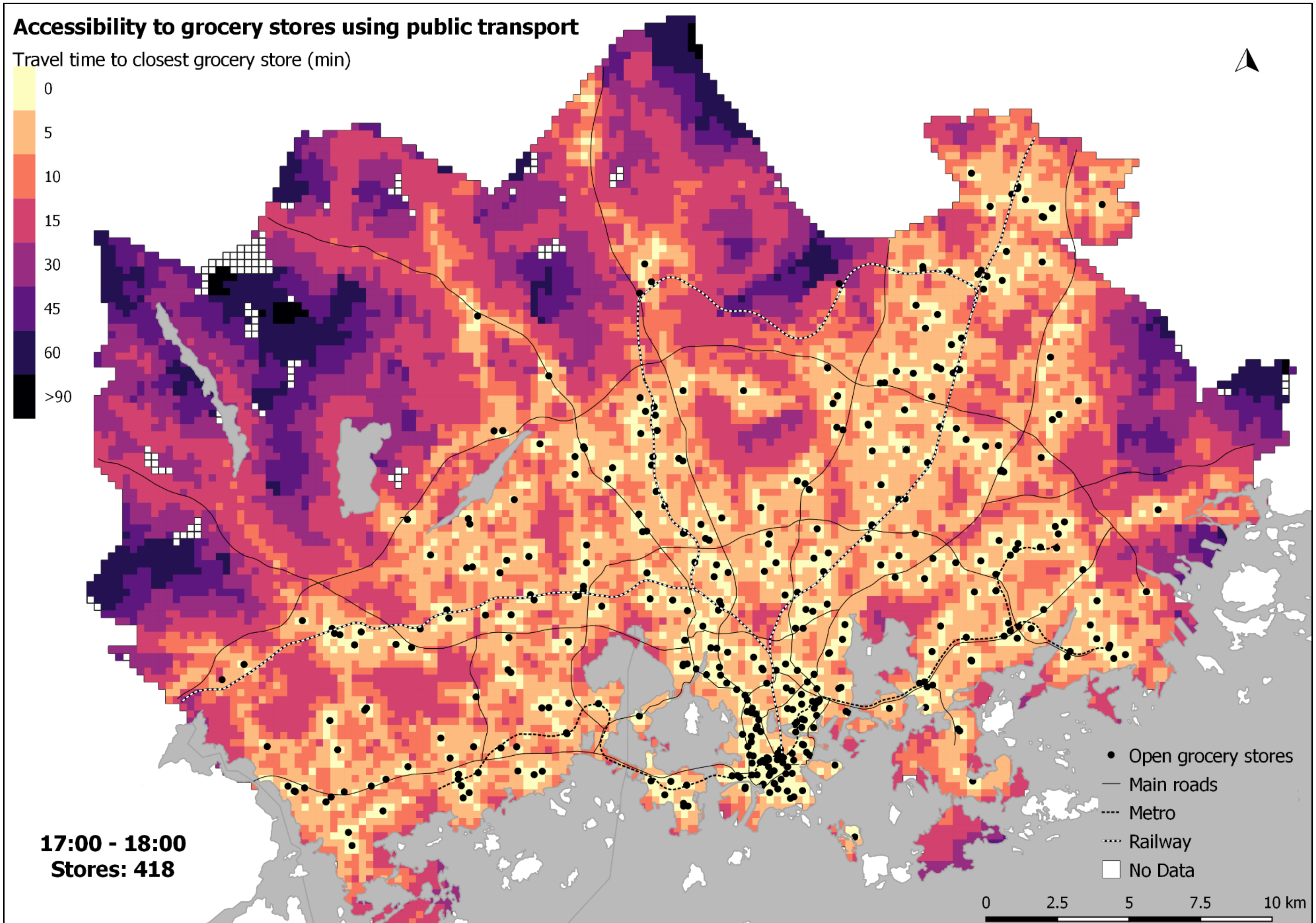
Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
12:00 - 13:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

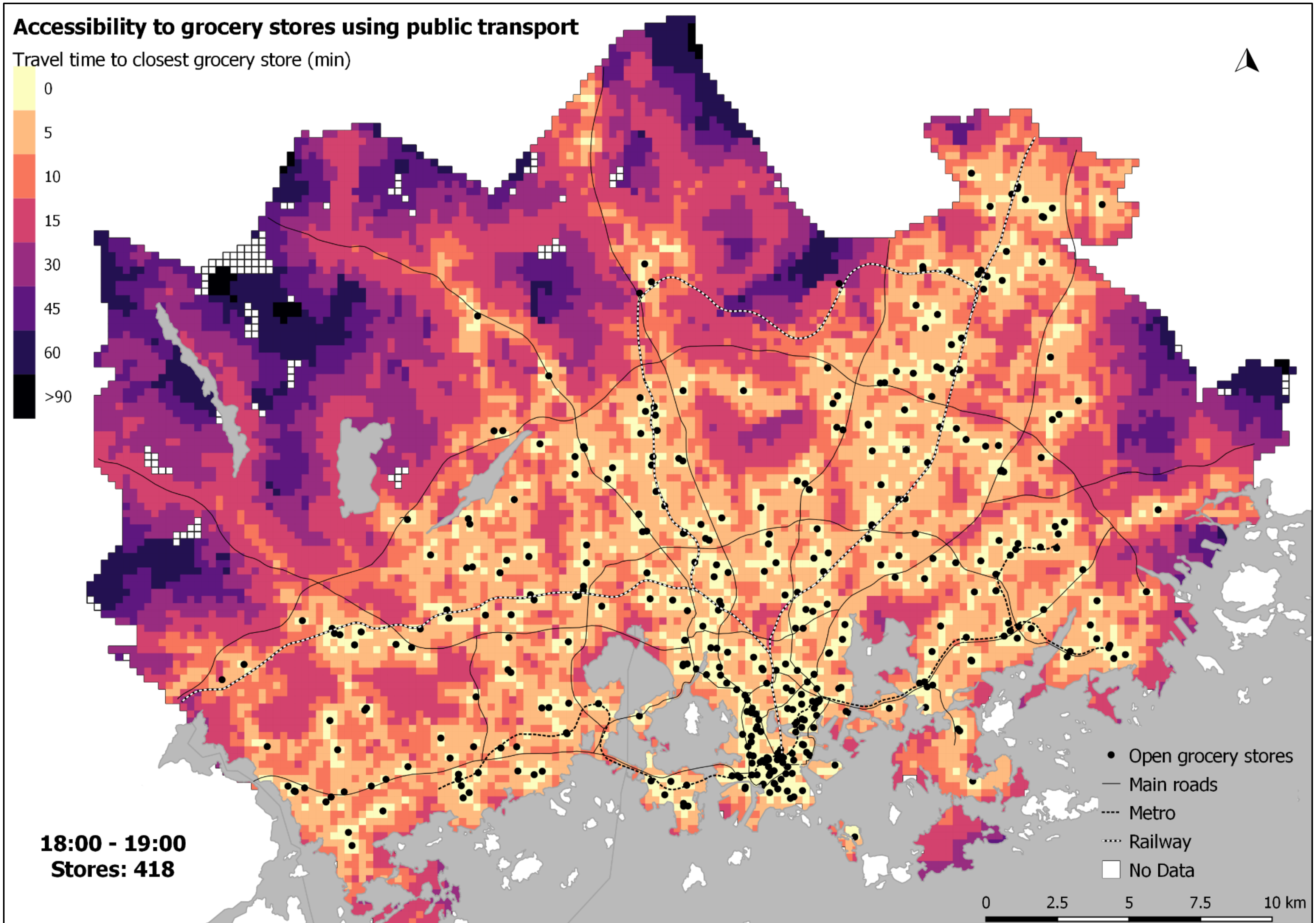

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
13:00 - 14:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
14:00 - 15:00
Stores: 418

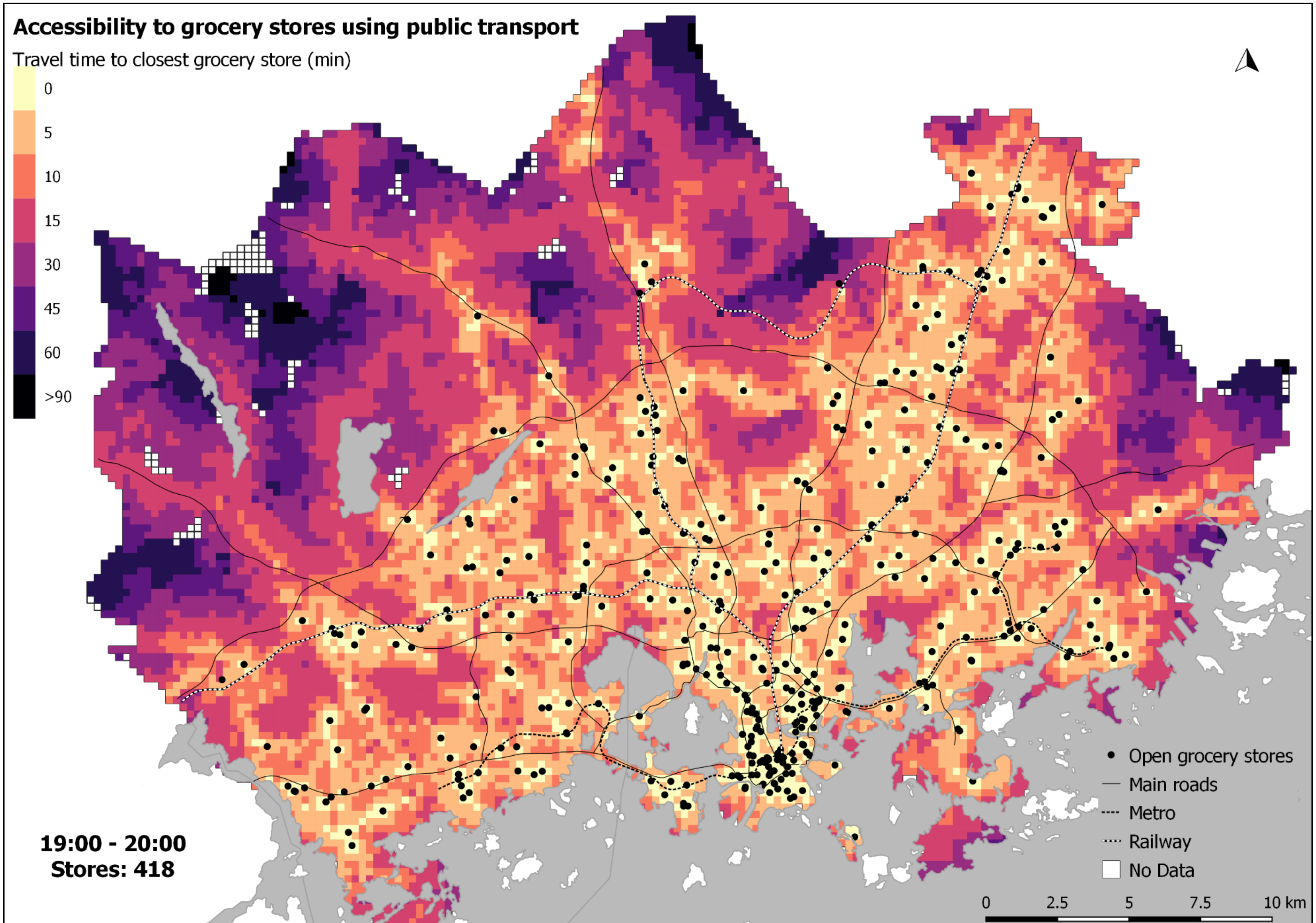

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
15:00 - 16:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

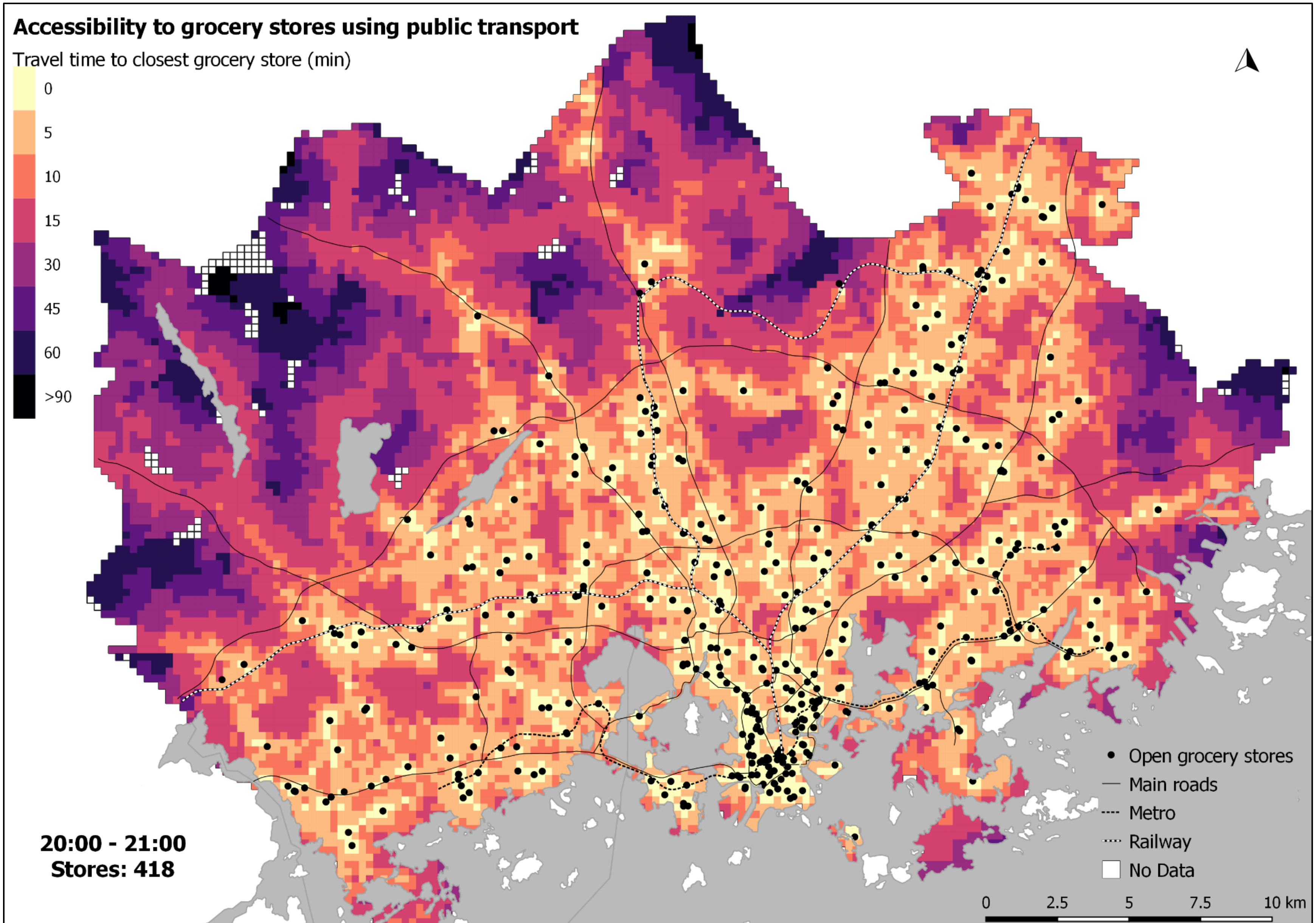

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
16:00 - 17:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

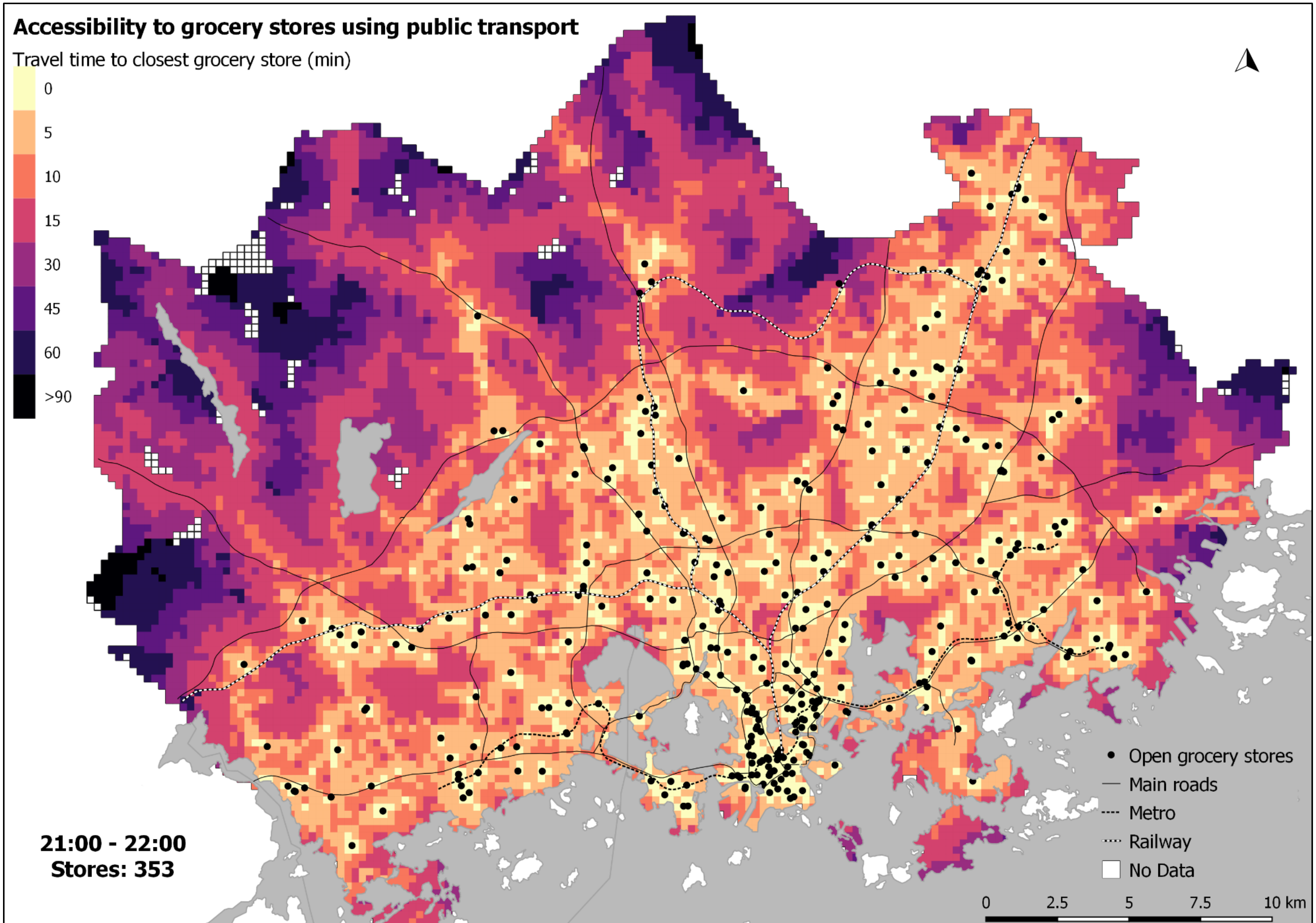

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
17:00 - 18:00
Stores: 418



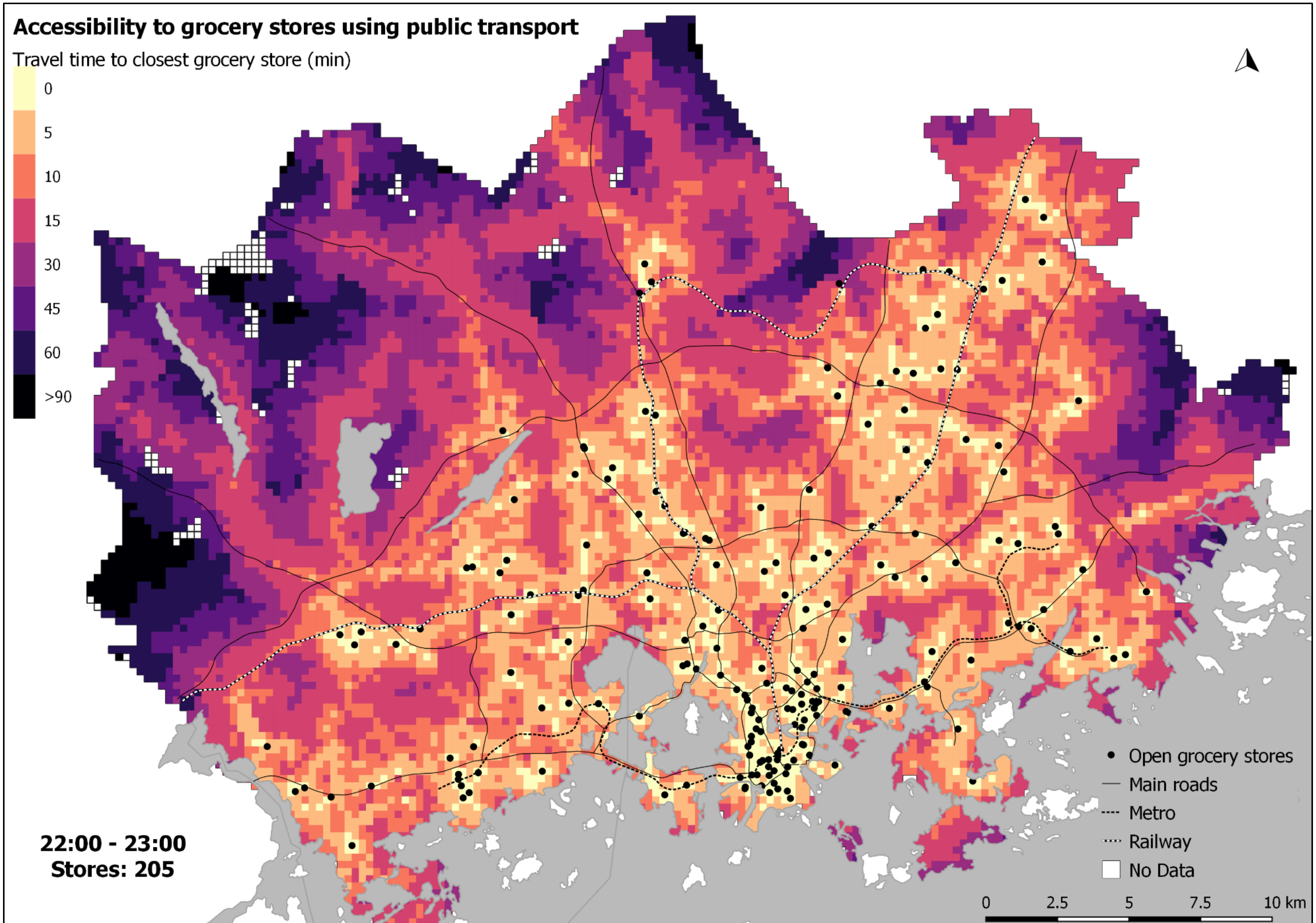
Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
18:00 - 19:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km

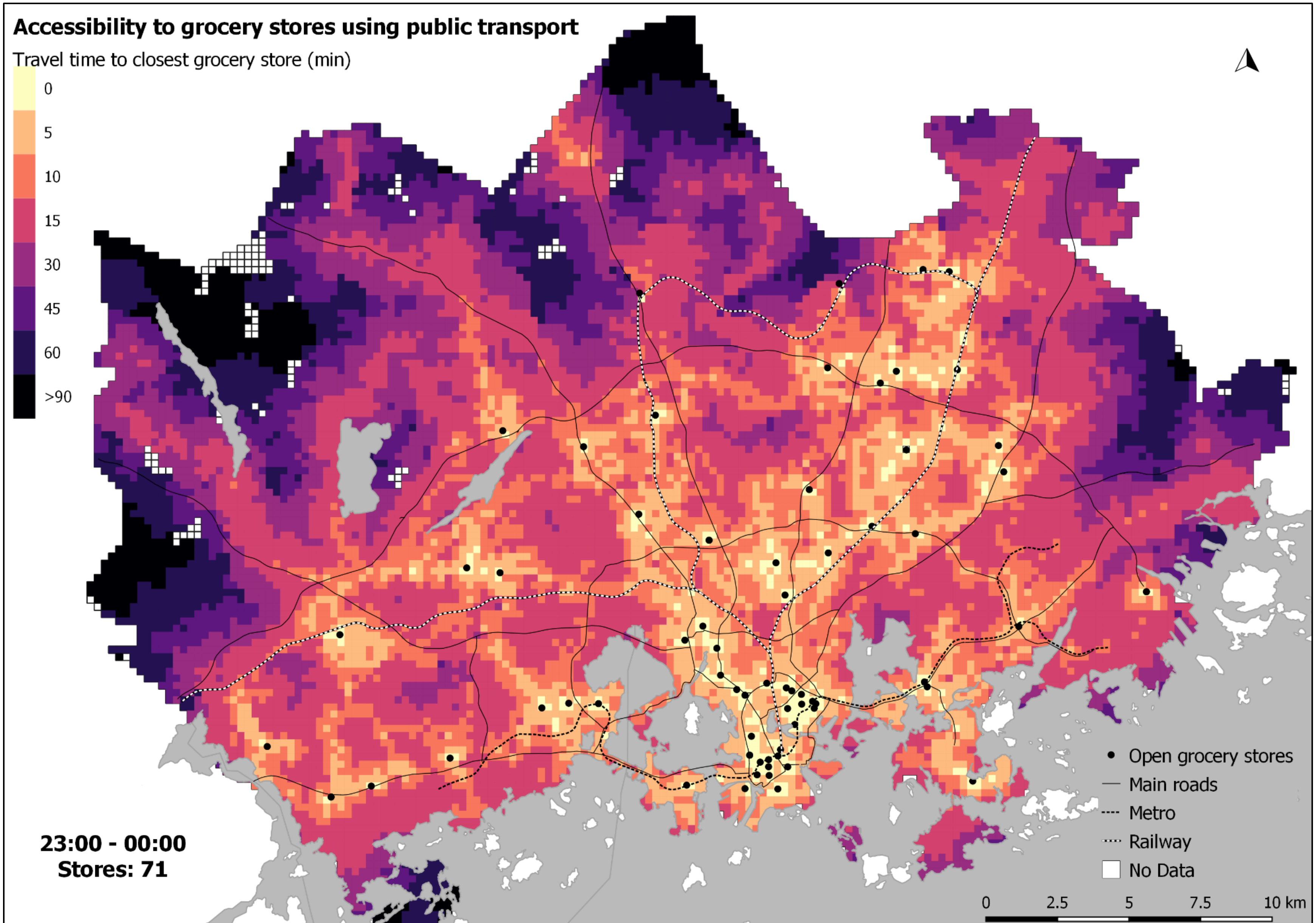

Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km
19:00 - 20:00
Stores: 418


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
20:00 - 21:00
Stores: 418
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km



Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
21:00 - 22:00
Stores: 353
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
22:00 - 23:00
Stores: 205
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km


Accessibility to grocery stores using public transport
Travel time to closest grocery store (min)
0
5
10
15
30
45
60
>90
23:00 - 00:00
Stores: 71
Open grocery stores
Main roads
Metro
Railway
No Data
0
2.5
5
7.5
10 km