University of Tartu

Faculty of Science and Technology

Institute of Ecology and Earth Sciences

Department of Geography

Master Thesis in Geoinformatics for Urbanized Society (30 ECTS)

Comparison of CDR and GPS data for estimating the individual activity space

JeongHwan Choi

Supervisors: Dr. Siiri Silm

Approved for defence:

Supervisor: /signature, date/

Head of Department: /signature, date/

Tartu 2020

CDR- ja GPS-andmete võrdlus inimeste tegevusruumi hindamiseks
Abstrakt

Selle uurimistöö eesmärk on hinnata CDR-andmete täpsust tegevusruumi hindamisel. CDR-andmete täpsust on hinnatud erinevate ajaühikute lõikes ning kõnetoimingute arvust ja inimeste sotsiaal-demograafilistest tunnustest lähtuvalt. Uurimistöös kasutatud andmed (CDR ja GPS) on kogutud ajavahemikus 5. september 2013 kuni 10. märts 2015. Uuringus on kasutatud 52 inimese andmeid kokku 8961 inimpäeva. Inimeste tegevusruumi on hinnatud kuue näitaja alusel. CDR-andmete põhiste tegevusruumi näitajate täpsuse hindamiseks on kasutatud kirjeldavat statistikat, korrelatsioonanalüüsi (Spearmani astakkorrelatsiooni) ning kombineeritud lineaarseid mudeleid.

CDR- ja GPS-andmete põhised näitajad on positiivses korrelatsioonis. Ringi raadiuse ala (*radius of gyration*) ja entroopia (*entropy*) näitajate puhul on CDR-andmete põhised näitajad sarnasemad GPS-andmete põhiste tegevusruumi näitajatega. Kerneli tiheduse (*kernel density*) puhul on CDR-andmete põhiste tegevusruumi näitajate täpsus absoluutarvuliste näitajate järgi kõige madalam. Uurimistöö tulemused osutavad sellele, et CDR-andmete põhiste tegevusruumi näitajate täpsust mõjutavad ajalistest tunnustest oluliselt ainult nädalapäevad. Lisaks sellele selgus, et kõnetoimingute arv ei mõjuta oluliselt CDR-andmete põhiseid tegevusruumi näitajaid, kui kõnetoimingute arv on üle nelja (st tegevusruumi näitajaid saab arvutada). Ükski analüüsitud sotsiaal-demograafiline tunnus CDR-andmete põhist tegevusruumi täpsust ei mõjuta.

Võtmesõnad: CDR, GPS, tegevusruum, ajaline kontekst, sotsiaal-demograafilised tunnused
CERCS-i kood: S230 sotsiaalgeograafia

Comparison of CDR and GPS data for estimating the individual activity space
Abstract

The aim of the research was to provide deeper understanding of the accuracy of CDR data for estimating individual activity space. The datasets (CDR and GPS) for the research had been collected from September 5, 2013 to March 10, 2015 and covered 52 people (8961 person-days). The individual activity spaces were analyzed by six major indicators: minimum convex polygon, ellipse, radius of gyration, kernel density, distance, and entropy. The absolute difference method was used to evaluate the accuracy of CDR-based measurements in comparison with GPS-based measurements. For statistical analysis, Spearman's rank correlation and linear mixed models were applied.

CDR and GPS-based measurements were positively correlated. Gyration and entropy were more closely related to GPS-based measurements whereas kernel density had the lowest accuracy based on the absolute difference between CDR and GPS-based measurements. The results from the study indicate that only days of the week factor significantly affects the accuracy of CDR-based measurements. Moreover, the number of CDRs per day was proven not to have a statistically significant effect on the accuracy of CDR-based measurements if the number of CDRs are four or more (i.e. it is possible to calculate the activity space indicators). Overall, none of the socio-demographic factors was proven to be significant to influence the accuracy of CDR-based activity spaces.

Keywords: CDR, GPS, Activity space, Temporal contexts, Socio-demographic factors
CERCS Code: S230 Social geography

Table of contents

Introduction

Human mobility and activity spaces are important to be studied because they provide a basis of application in urban planning and epidemic modeling. Human mobility information enables the government to design sustainable urban transport systems and take effective preventive measures in case of emergency like crowd evacuation and contagious diseases by considering the configuration of new buildings and public spaces (Barbosa et al., 2018; Wang et al., 2019). Recently, the outbreak of novel coronavirus (COVID-19) pandemic has a huge impact on the mobility patterns of individuals. The introduction of non-pharmaceutical interventions resulted in a reduction in human mobility patterns particularly in areas with a dense population (Askitas et al., 2020; Bryant & Elofsson, 2020). The reduction in mobility subsequently led to a decline in the spread of COVID-19 (Badr et al., 2020; Courtemanche et al., 2020).

The study of human mobility in public health crisis, traffic prediction, and migration flows greatly depends on contemporary mobile technologies. The introduction of these new technologies for communication has added another dimension to the mix as new data sources for modelling human activities are being developed daily (Yuan & Raubal, 2016). According to International Telecommunication Union (ITU), there were more than 7 billion mobile cellular subscriptions worldwide by the end of 2015 and more than 50 percent of the global population used the Internet at the end of 2018 (ITU, 2015, 2018). It is an undeniable fact that a mobile phone has become one essential tool for communication by people in many parts of the world and in performing various tasks in their daily lives. Moreover, there has been a steady penetration of mobile phone users in Estonia over the years which shows that in 2017 there were about 1.9 million subscribers and the number is projected to increase in the coming years (ITU, 2018).

Understanding the mobility patterns and people's use of space has been the focus of geographical research for many years (Xu et al., 2016). The term activity space is widely used in human mobility research to describe main places of interest of people where they carry out their daily routines such as residential dwellings, workplaces, and shopping centers (Gong et al., 2020).

In contemporary society, due to the advancement in the development of location–aware technologies, research on human mobility has gained popularity through access to a large

volume of individual tracking datasets which contributes to comprehension of an individual's activity space over time (Dobra et al., 2015; Williams et al., 2015; Xu et al., 2016). Call Detail Record (CDR) as a type of passive mobile positioning data is used in research to ascertain human mobility habits. This type of data provides low-level location information of origin and destination of call activities with attributes such as start and end time of calls, location of caller and receiver, duration of the call as well as, initiator and receiver ID (Xu et al., 2016; Lind et al., 2017; Vanhoof, Reis, et al., 2018). CDR data provide for largescale analysis of location and movement patterns, but they are scanty given that they have limited information pertaining to the time of a call (Vanhoof, Reis, et al., 2018). CDR data have been used for planning, policy and infrastructural developments by conducting analysis to determine anchor points in an individual's life (Amini et al., 2014). The accuracy of CDR based on mobile antennas is comparatively less accurate than Global Positioning System (GPS). Some researchers raise critiques about the accuracy of CDR data being varied in cities and rural areas because a higher density of population is directly related to a higher density of antennas and vice versa (Bengtsson et al., 2011). However, CDR data would provide information about mobility and activity space for more people over a longer period without additional activities for participants.

Alternatively, the precision and accuracy of GPS data are higher and reliable for the comprehension of human behavior compared to traditional collection methods such as survey, self-reporting, observation, etc. (Richardson et al., 2013). There have been prior studies on human spatial and temporal mobility patterns by relying on high-resolution smartphone-based GPS location datasets (Kwan 2012; Matthews & Yang 2013; Perchoux et al., 2013). Some criticisms of reliance on GPS datasets in human mobility research are the expensive data collection mode, battery drain and is not representative of an entire population but only considers a sample with GPS devices (Paz-Soldan et al., 2014; Xu et al., 2015). Besides, GPS is considered useless in indoor conditions due to the block of radio waves by physical objects (Cabric, 2017).

There appears to be no theoretical or empirical research on the comparison of CDR and GPS data for estimating human activity spaces. Moreover, there is a limitation of published materials in relation to accuracy of CDR-based activity space for various human mobility indicators. Therefore, the aim of the study is to provide deeper understanding of the accuracy of CDR data for estimating individual activity space.

The aim can be achieved through the following research questions:

Q1: How are the CDR and GPS-based activity space measurements related?

Q2: How different temporal scales affect the accuracy of CDR-based activity space?

Q3: How the number of call activities and socio-demographic factors of the people influence the accuracy of CDR-based activity space?

Q4: What is the accuracy of CDR-based activity space considering different time periods?

1. Theoretical overview

1.1 Activity Space

In the study of the spatial distribution of people's behavior and aggregated activity patterns of urban systems, activity space is important (Yuan & Raubal, 2016). According to Mazey (1981), activity space was defined as the local areas within which people travel during their daily activities. Other studies in the area focused on measuring the size, geometry, and inherent structure of human activity space (e.g., the randomness of activity patterns), as well as the reasons why activity space forms (Golledge & Stimson, 1997). However, in practice, investigating the quantitative properties of human activities often involves model fitting and an appropriate mathematical model provides insights for many application areas. These application areas range from building a smart system in urban planning and geography to a deeper understanding of the basic laws of human activity in physics (González et al., 2008; Song et al., 2010). On the other hand, the modelling of the distribution of activity space is still an ongoing process (Yuan & Raubal, 2016).

There are several related concepts to activity space. These are awareness space (Brown & Moore, 1970), action space (Horton & Reynolds, 1971), perceptual space (Relph, 1976) and mental maps (Lynch, 1960). But in general, an individual's activity space is usually conceptualized as the locations that have been visited as well as the travels among these locations (Schönfelder & Axhausen, 2003; Gong et al., 2020). People perform their daily routines mainly at a few activity locations such as home, school, workplace, supermarkets, favorite restaurants and so forth. These locations are often considered as anchor points of individual activity spaces (Golledge & Stimson, 1997; Ahas et al., 2010; Xu et al., 2016). Mobile phone location data give information of individual footprints recorded for people's major activity locations in space and time (Xu, 2015).

1.2 CDR and GPS data for measuring activity space

In measuring activity space, the CDR captures the phone activity of subscribers on the operator's network. CDR data are defined as non-continuous information because the data are only stored when text messages and calls are made or received (Vanhoof, Reis, et al., 2018). It is collected for billing and network maintenance purposes. Like the GPS, there are some studies conducted using the CDR data in recent times to analyze individual movement patterns

(regarding locations e.g., home and workplace) (Ahas et al., 2010; Xu et al., 2016). There have been studies in Estonia about measuring individual activity spaces based on CDR data to investigate the ethnic/racial segregation between different age groups as well as the monthly variability in the spatial travel behavior of people (Silm & Ahas, 2014a; Järv et al., 2014, 2015; Silm et al., 2018). Bogomolov et al., (2014) used mobile data to create a novel method to improve the crime prediction accuracy, and Schmitz and Cooper (2007) examined the activity spaces of offenders to assist investigating officers to solve criminal cases.

The GPS on the other hand has become widely adopted in understanding various aspects of urban dynamics such as individual commuting patterns (Shen et al., 2013), route choice behavior (Papinski et al., 2009) and spread of disease (Vazquez-Prokopec et al., 2009). The GPS has the capability to capture human movements with high spatiotemporal accuracy (Richardson et al., 2013), so GPS data have been accepted as a valuable source that can help enhance our understanding of human mobility and activity patterns in urban settings (Bazzani et al., 2010; Shoval et al., 2011). GPS has become a popular means of collecting tracking data for studying human travel and activity patterns two decades ago (Hirsch et al., 2014; Xu, 2015). Since then, various approaches have been applied to derive trips and important locations from individual GPS trajectories. Then, Schüssler and Axhausen (2009) also developed methods to derive individual trips and activities from GPS data. The cumulative effect of these results demonstrated the feasibility of using GPS for an understanding of individual activity patterns.

Overall, these studies indicate that using CDR and GPS data can be leveraged to understand the spatial distribution and movement patterns of individuals. Some comparative advantages and disadvantages of two data types (CDR and GPS) are summarized in Table 1.

Table 1. Comparisons of CDR and GPS data

Data type	Advantages	Disadvantages
CDR	Relatively cheap data collection mode Larger coverage regarding users, timespan, and spatial extent.	Positioning errors
GPS	High positioning accuracy; Fine-grained	Relatively expensive data collection mode High battery usage

1.3 Indicators of Activity Spaces

There are several methods of measuring activity spaces in research based on CDR and GPS data. Some of these methods are appropriate and applicable for both CDR and GPS datasets. The characterization of activity space gives a better understanding of the method used in the measurement. According to Chen and Dobra (2018), an individual's activity space includes: (i) estimating the spatial configuration and the frequency of the anchor locations; (ii) identifying other places of interest of an individual including anchor locations and assessing these places vis-à-vis the least visited; (iii) mapping the spatial configuration of these frequently visited places; and (iv) quantifying the spatial structure of the individual's activity space. Table 2 below gives major references that applied some indicators in CDR and GPS data analyses.

Table 2. Major indicators in measuring activity spaces

Metric/Indicator	Description	Reference
Minimum Convex Polygon	To describe a subscriber's activity space encompassing the spatial distribution of all activity places within a subscriber's movement pattern	Sherman et al., 2005 (GPS) Palmer et al., 2013 (CDR) Hirsch et al., 2014 (GPS) Dong et al., 2015 (CDR) Lee et al., 2016 (GPS)
Ellipse	To describe the activity location distributions	Sherman et al., 2005 (GPS) Chaix et al., 2012 (GPS) Hirsch et al., 2014 (GPS) Yuan and Raubal, 2016 (CDR) Puura, Silm, and Ahas, 2018 (CDR)
Radius of Gyration	To explore the individual's movement span	Yuan et al., 2012 (CDR) Barbosa et al., 2018 (GPS) Chen et al., 2018 (CDR) Pappalardo and Simini, 2018 (CDR, GPS)
Kernel Density	To measures a certain probability of visit to activity spaces which embraces all areas	Schönfelder and Axhausen, 2003 (GPS) Yuan et al., 2012 (CDR) Yuan and Raubal, 2016 (CDR) Chen and Dobra, 2018 (GPS)
Travel Distance	To explore the general trajectories of human movements	Zhao et al., 2016 (CDR) Burkhard et al., 2017 (CDR, GPS) Barbosa et al., 2018 (CDR) Pappalardo and Simini, 2018 (CDR, GPS)
Entropy	To describe individuals' visitation patterns	Yuan et al., 2012 (CDR) Comito et al., 2016 (GPS) Zhao et al., 2016 (CDR) Pappalardo and Simini, 2018 (CDR, GPS)

1.3.1 Minimum Convex Polygon

Minimum Convex Polygon (MCP) is occasionally regarded as a home range and describes an individual's activity space which shows the smallest convex polygon comprising the spatial distribution of all activity places within a traveler's movement pattern (Hirsch et al., 2014; Patterson & Farber 2015; Chen & Dobra 2018; Sharmeen & Houston 2019). Sharmeen and Houston (2019) further argue that the size of the polygon largely relies on the location sample size and sampling standardization. Figure 1 shows the weekly size of MCP of car-owning and non-working individuals.

This methodology has been widely used to study human activity spaces in order to determine how urban morphology affects the activity spaces of individuals (Buliung & Kanaroglou 2006; Fan & Khattak 2008; Lee et al., 2016). Some geographical features like river, valley, and hill affect the shape of activity space to be irregular since these features are usually inaccessible or undesirable for people (Lee et al., 2016; Chen & Dobra, 2018). Other limitations involve extreme vulnerability to outliers, trip chains are ignored, and only proximity to an area is considered (Li & Tong, 2016; Chen & Dobra, 2018).

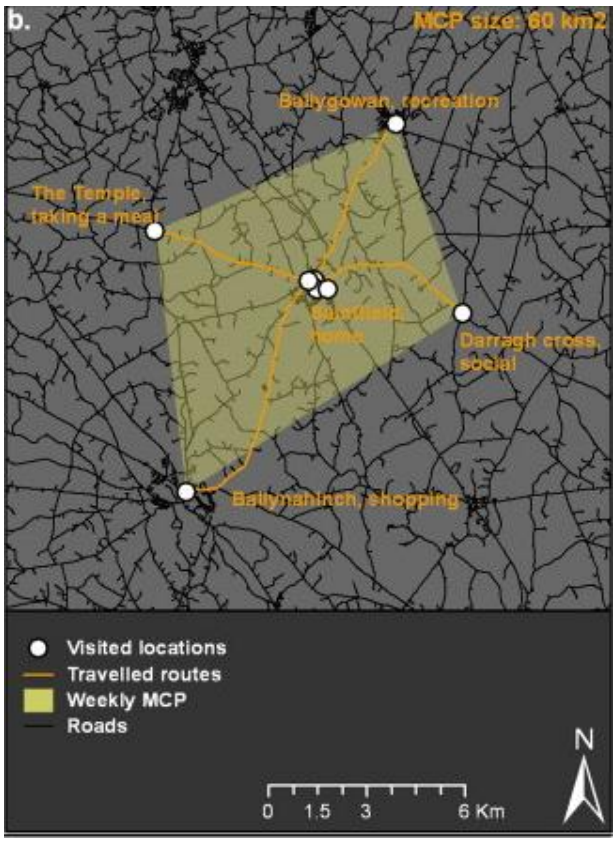

Figure 1. Example of Minimum Convex Polygon (Source: Kamruzzaman & Hine, 2012*)*

1.3.2 Ellipse

Ellipses focus on a set of visited locations for a specific area based on knowledge of the most relevant anchor locations in another space such as residence and workplace (Chen & Dobra, 2018). The confidence ellipse and home-work ellipse are two main types of ellipse (Chaix et al., 2012; Li & Tong, 2016). The confidence ellipse assumes that visited locations follow a bivariate normal distribution (Sherman et al., 2005; Chen & Dobra, 2018). The home-work ellipses relate to two anchor locations which become the two foci of the ellipse (Newsome et al., 1998).

Spatially, a confidence ellipse is used to describe the activity location distributions. The size of the area of an ellipse indicates the dispersion of visited locations and may be used in comparison to the dispersion between the mobility pattern of one or more travelers within different temporal space (Schönfelder & Axhausen, 2003). Figure 2 depicts the weekly size of ellipse of car-owning and non-working individuals. Some limitations of representing activity spaces through ellipses are when locations visited are only a few and in a straight line (Wong & Shaw, 2011); and thereby resulting in their relatively inflexible geometry (Chen & Dobra, 2018).

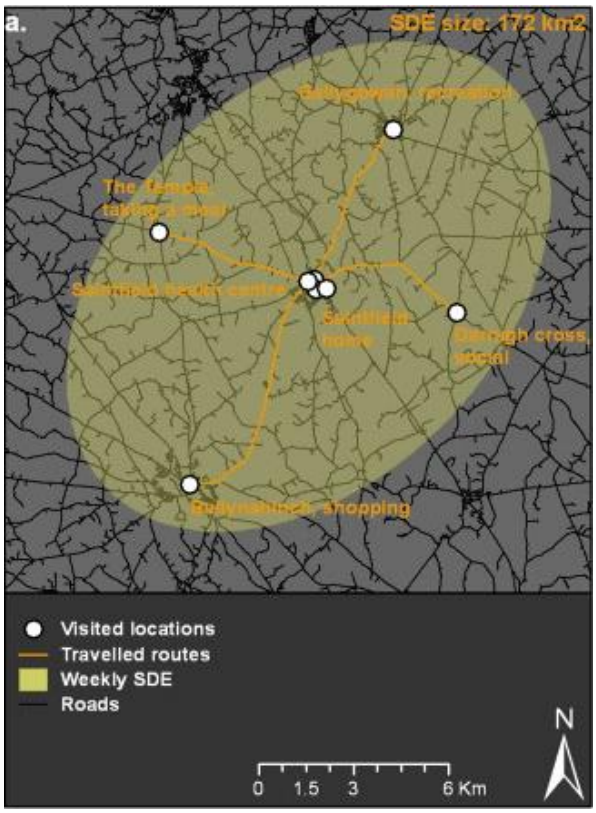

Figure 2. Example of Confidence Ellipse (Source: Kamruzzaman & Hine, 2012*)*

1.3.3 Radius of Gyration

One of the most predominantly used methods in measuring activity space is the radius of gyration. This is a measure that determines an individual's travel distance on the basis of the distance between the specific visited locations and the time spent in each location (Barbosa et al., 2018). The radius of gyration focuses mostly on the center of the home and work location for commuters and it is reflective of the range of activity space (Golledge & Stimson, 1997; Zhao et al., 2016). The application of the radius of gyration to CDR data is limited because commuters who travel long distances rarely use mobile phones which leads to an underestimation of this group (Zhao et al., 2016). A person's radius of gyration may be small even if the person has a longer travel distance regardless of repeated movements across all over different locations generally yielding a bigger radius of gyration (Chen et al., 2018).

1.3.4 Kernel Density

Kernel density analyzes the spatial density of a whole area by the distribution of point objects in the target region (Kang et al., 2018). This method can estimate activity spaces of any kind regardless of the shape and corresponding anchor locations (Chen & Dobra, 2018). Figure 3 reveals the kernel density estimation analysis of geo-tagged data from GPS devices.

It measures a certain probability or density of visit to activity spaces which includes all areas. Kernel densities involve a transformation of points represented continuously based on density in a wider area; the method generalizes the points to the area located and usually based on interpolation or smoothing technique (Schönfelder & Axhausen, 2003). The interpolation results in a value for any points in the entire area which defines the density. The shape of activity space could be refined to avoid unaccustomed daily activities like industrial areas, etc. (Schönfelder & Axhausen, 2004).

Kernel densities are most appropriate in the application to large cross-sectional datasets (Kwan, 2000; Buliung, 2001). However, kernel density does not always produce the most reliable results when applied to GPS data because it does not capture much of the underlying structure of the data (Chen & Dobra, 2018).

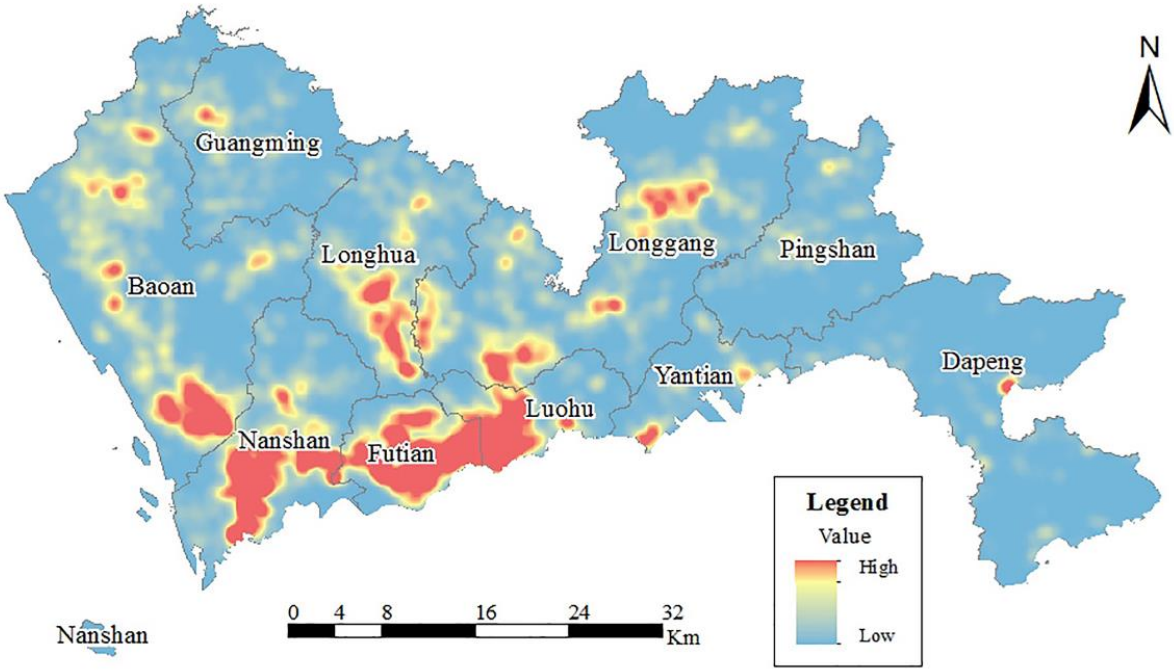

Figure 3. Example of Kernel Density (Source: Wu et al., 2016*)*

1.3.5 Travel Distance

Distance is usually measured in a one-dimensional way which needs knowledge of the specific points to measure (Sherman et al., 2005). The rationale of measuring activity space based on the travel distance covered by each subscriber is to mark the footprint of the subscriber and the distance traveled altogether considering Euclidean distance (Zhao et al., 2016). This provides a result of the consecutive footprints of the subscriber (Mooney et al., 2016). The longer the distance traveled daily by an individual could be directly related to a wider extent distance covered based on CDRs. One other reason is that, the longer the travel distance then it is likely that the number of subscribers is small (Zhao et al., 2016). The limitation in distance as a measure of activity space is that it generally ignores the preferences of people or other factors that instigate the direction of people toward a specific location (Sherman et al., 2005). There can also be directional bias because an individual may prefer a particular place over other places having equal distance due to the perception of the quality of the preferred place over others (Golledge & Stimson, 1997).

1.3.6 Entropy

Entropy describes the heterogeneity of human movement patterns which are recorded by data from mobile devices (Yuan et al., 2012; Vanhoof, Schoors, et al., 2018). Song et al. (2010) introduced three varied methods for computing entropy. These are random entropy, temporally uncorrelated entropy, and real entropy. All three methods depend on visited cell towers, but probabilities of being in a particular cell-tower are estimated in three different ways: (i) whether or not a person has made a prior visit to a given cell tower; (ii) the frequency of visits to a cell tower; (iii) the amount of time spent in the range of a cell tower.

It can be deduced that the more active a subscriber is at different places, the higher the movement entropy in comparison with a subscriber who visits fewer places many times (Zhao et al., 2016).

1.4 Evaluation of the accuracy of CDR data in activity space measures

The use of CDR data can be good based on the level of accuracy or there may be some uncertainty issues presenting some level of biases in its use. CDR data has the benefit of being significantly available for a large number and cover a significant proportion of the population where cell phone penetration is high (Burkhard et al., 2017). The individuals whose CDR are being collected for use in research are unaware of this and this makes it cheap and quick to obtain information from a large proportion of the population in a specified area for analysis (Steenbruggen et al., 2015). The accuracy of CDR is greater in metropolitan areas and those with denser networks of roads, whereas accuracy is lower in rural areas with a low density of population (Ahas et al., 2008; Chen et al., 2018). Furthermore, spatial errors in human mobility patterns computed from CDR and GPS data are higher for sparse CDRs (Hoteit et al., 2016).

The uncertainties surrounding the use of CDR data are numerous. It is considered that the spatial resolution of CDR is often limited to the specific location of a cell tower (Zhao et al., 2016; Lind et al., 2017). Thus, the precision of CDR data is given only by the availability of information about calls or messages routed from a cell tower location. This limits access to information about the exact location of individual mobile phone users and the precise time of call or message (Burkhard et al., 2017). One other issue with CDR data is the spatial uncertainty surrounding its use due to signal jump. This signal jump occurs when a mobile device switches

between neighboring cell towers due to the similar intensity of signal strength or closeness of cell towers (Iovan et al., 2013).

The effectiveness of CDR data in individual mobility study depends on the research question and the mobility measure selected to address those questions. The use of the CDR data in a study of human mobility is good enough in most cases when the radius of gyration measure is applied (Zhao et al., 2016). This radius of Gyration is good in analyzing human mobility based on CDR data depending on subscribers who make at least some phone communication throughout the day and travel frequently. The accuracy of Gyration could be enhanced for frequent CDR users than for rare CDR users (Chen et al., 2018). However, researchers must be cautious in the use of CDR data when it comes to problems relating to travel distance and heterogeneity of human mobility. This is because the validity of analysis largely depends on the activeness of subscribers engaged in phone communication. CDR data tend to significantly underestimate the total travel distance (Zhao et al., 2016). This is because the longer the travel distance, the wider the CDR data and one possible reason is that as the total travel distance increases, the number of subscribers decrease rapidly. CDR data can estimate the movement Entropy accurately for subscribers based on certain locations but may underestimate the movement Entropy for other subscribers in another location (Zhao et al., 2016).

1.5 Temporal Contexts of Activity Space

In activity space measurement, there are high movement patterns during weekend in terms of spatial coverage while working days have more regular and direct patterns of human spatial mobility (Kamruzzaman & Hine, 2012). The activity space of individuals on holidays differ from their daily routines (Wallendorf & Arnould, 1991; Gram, 2005). It is shown that people tend to travel longer distances over wider geographic extents on holidays than weekdays due to more time availability (Cools et al., 2009). Additionally, monthly variability in human movement patterns is based on seasons. Schönfelder and Axhausen (2016) found a clear distinction in seasonality of individual trips from their home where travel patterns are more spatially dispersed further away from home in spring and summer (April – July) in comparison to fall and winter in USA. Estonia has similar seasonal patterns regarding movements in winter and summer seasons since there is a higher mobility variation in summer (June – August) than in winter (Järv et al., 2014).

Time is one of the significant factors of activity space measurement research. Individual's movement preferences are influenced by time available to them and purposes for their mobility. Time-space concepts in activity space research are categorized into four; daily, weekly, holiday, and monthly. It can be linked to some studies that discovered a strong connection between individual's different movement patterns and their daily, weekly, monthly, and seasonal travel behaviors (Järv et al., 2014; Silm & Ahas, 2014b).

Humans' daily activities involve some activities such as work, school, home, etc. Primerano et al. (2008) defined as "a scheduling of activities in time and space" when it comes to daily human mobility from home to work and back. The space-time context of mobility of people as high and of particular interest to researchers in many fields such as urban planning, transportation, and business (Zeng et al., 2017).

However, comprehension of human spatial mobility and temporal context over a long period of time is also important to understand the impact of time context on people's spatial behaviors (Järv et al., 2014).

1.6 Socio-Demographic Factors and Activity Space

Socio-demographic factors such as gender, age, income, household composition and occupation, etc. have a critical impact on individual mobility patterns. Most previous empirical studies have proven that women tend to have shorter travel distances and smaller size of activity spaces compared to men (Fan & Khattak, 2008; Vich et al., 2017). However, Bajracharya and Shrestha (2017) discovered that women have similar travel behavior to men.

In terms of household membership mobility patterns, the larger activity space relates to the big households since they tend to travel more due to the ownership of private cars and visiting their family over a long distance. (Rubin et al., 2014; Kim & Ulfarsson, 2015). However, Dargay and Clark (2012) found that single member household has a longer total travel distance in comparison to those in larger household.

When it comes to age group variable as a socio-demographic aspect, young-employed and middle-aged groups have shown broader travel patterns; larger activity spaces and trip frequencies whereas older and younger groups exhibit the opposite (Yuan et al., 2012). Also, Fan and Khattak (2008) found that mature adults tend to have larger activity spaces than elderly and young adults but there is one research that indicates a similar trip pattern among people of

all age groups (Bajracharya & Shrestha, 2017). However, some studies found that younger groups have a large spatial mobility compared to elderly groups and the size of activity space declines with age (Silm et al., 2018; Masso et al., 2019).

Occupation and income are somewhat intertwined in many societies. The higher the occupational level, the higher your income and vice versa. This affects travel behaviors significantly through mode of transport. Some researchers discovered that the higher income and well-educated people are more likely to own private cars and can afford to travel over larger geographical extent (Davidov, 2007; Fan & Khattak, 2008; Mercado et al., 2012; Jones & Pebley, 2014; Klinger & Lanzendorf 2016; Tana, Kwan, & Chai 2016).; unlike unemployed or low-income groups who travel less (Vich et al., 2017). However, factors such as income, and education do not have significant effects on the size of activity space compared to other groups (Schönfelder & Axhausen, 2003; Zenk et al., 2011).

2. Data and Methodology

2.1 Data

The two datasets (CDR and GPS) were provided by the Mobility Lab of the University of Tartu. The data covered the whole of Estonia. As the name implies, CDR data is based on calls whereas GPS data is collected through an android mobile application known as "MobilityLog". Subscribers whose data were used in this study have the same date recorded for both CDR and GPS data for comparative analysis.

2.1.1 CDR data

In relation to the study, the timespan covering individuals in the dataset varies. The data was collected from September 5, 2013 to March 10, 2015. CDR data involve having a structure of mobile positioning identity number, unique individual identity number, the data record time, location, and date covering 52 people (8961 person-days) as shown in Table 3.

Four columns in the dataset were particularly extracted for further analysis such as unique individual identity number, the data record time, and location. The main rationale for being selective is because daily individual activity spaces need to be computed. These columns in the dataset give a more specific and easier estimation.

Table 3. Sample table for the CDR dataset.

pos_id	mps_usr_id	pos_time	lon	lat	x	y	date
1	10	2013-09-24 15:21:00	26.717558	58.372	658109	6473710	2013-09-24
2	10	2013-09-26 17:30:00	26.7150155	58.371	658851	6473252	2013-09-26
3	10	2013-09-26 17:54:00	26.7501205	58.373	660878	6473606	2013-09-26

2.1.2 GPS data

The data collected for mobile positioning spans from September 5, 2013 to March 10, 2015. It also varies in terms of the data coverage period for individuals. GPS data has a structure of mobile positioning identity number, unique individual identity number, time, point (location in

Well-Known Binary (WKB) format), accuracy, altitude, bearing, speed, and date for 52 people (8961 person-days) as depicted in Tables 4.

Table 4. Sample table for the GPS dataset.

mps_usr_id	id	counter	time_system	time_gps	time_system_ts	time_gps_ts
10	1	768062	1.37954E+12	1.37954E+12	24/09/2013 15:21:00 +03:00	24/09/2013 15:21:00 +03:00
10	2	768104	1.37954E+12	1.37954E+12	26/09/2013 17:30:00 +03:00	26/09/2013 17:30:00 +03:00
10	3	765096	1.37954E+12	1.37954E+12	26/09/2013 17:54:00 +03:00	26/09/2013 17:54:00 +03:00

mps_usr_id	point	accuracy	altitude	bearing	speed	date
10	0101000020E6100000184CD3D2A 2C13A40081AD8CF07314D40	42	174.4	72	0.75	24/09/2013
10	0101000020E6100000C795C4A39 EC13A40F0498A3C07314D40	30	157.6	76.4	0.5	26/09/2013
10	0101000020E610000007EB6247A 1C13A404959AFB707314D40	36	128.7	40	0.5	26/09/2013

Three attributes such as a unique individual identity number, the data record time, and location in the dataset were used to investigate the mobility patterns of the people. The point column is converted to x and y coordinates from WKB format for easy handling and analysis of the data (Table 5).

Table 5. Sample table for the conversion WKB format to x and y coordinates.

mps_usr_id	point	x	y
10	0101000020E6100000184CD3D2 A2C13A40081AD8CF07314D40	26.756390740000001	58.383050900000001
10	0101000020E6100000C795C4A39 EC13A40F0498A3C07314D40	26.756326900000001	58.383033339999997
10	0101000020E610000007EB6247A 1C13A404959AFB707314D40	26.756367170000001	58.383048019999997

The geographic coordinate system for both original datasets was WGS84. So, it was appropriate to convert to the Estonian national projected system (EPSG: 3301) in order to apply linear units in terms of meters, kilometers, and miles in measuring the activity spaces.

2.1.3 Socio-demographic data

Based on the purpose of this study, five demographic variables are selected. Gender has two main labels – male and female. Age group to which a person belongs, occupation, marital status, and the number of household members (Table 6). Table 6 has categories such as the number of people, person-days, and percentages based on the number of people.

Table 6. The description of socio-demographic factors

Socio-demographic factor		The number of people	The number of Person-days	Percentages per people
Gender	Male	16	2221	30.77%
	Female	36	6740	69.23%
Age Group	Young Adults (17 - 30)	23	4221	44.23%
	Middle-aged Adults (31 – 45)	11	2695	21.15%
	Old-aged Adults (Above 45)	18	2045	34.62%
Marital Status	Married	22	3847	42.31%
	Cohabitation	18	3342	34.62%
	Without partner	10	1728	19.23%
	Partnership without living together	2	44	3.85%
Occupation	Staff	18	3018	34.62%
	Students	11	2376	21.15%
	Unknown	23	3567	44.23%
The number of household members	One person	5	1075	9.62%
	Two people	15	2586	28.85%
	Three people	6	1358	11.54%
	Four people	2	350	3.85%
	Unknown	24	3592	46.15%

2.2 Methodology

2.2.1 Defining Individual Activity Space

This part gives all attention to assessing the representativeness of CDRs in an analysis of persons' daily patterns of movements. The primary target was to compute various measures of subscriber's mobility patterns to answer the research questions. The following six most common methods were selected: (i) minimum convex polygon; (ii) ellipse; (iii) gyration; (iv) kernel density; (v) distance; and (vi) entropy. These were chosen because they are commonly used methods in activity space measurements applied for both CDR and GPS data by other researchers.

In the assessment process, some steps were conducted to handle the set of CDR footprint and the set of GPS footprint in order to gain a general overview. Individual activity spaces were computed using RStudio software for both CDR and GPS data for each person on the same date using mentioned methods to analyze their daily movement pattern. A daily activity space was classified based on temporal patterns of days of the week, months, and holidays. This process aggregates the call activities of all 52 people. For instance, days are days of the week (Monday to Sunday) and months (January to December). Consequently, holiday data were extracted based on 9 national holidays in Estonia (Estonian Government Office, 2018) (Table 7).

Table 7. National holidays in Estonia

Public holidays and days off			
New Year's Day (January 1)	Easter Sunday (March 31, 2013) (April 20, 2014)	Victory Day (June 23)	Christmas Eve (December 24)
Independence Day (February 24)	Labor Day (May 1)	Midsummer Day (June 24)	Christmas Day (December 25)
Good Friday (March 29, 2013) (April 18, 2014)	Pentecost (May 19, 2013) (June 8, 2014)	Independence Restoration Day (August 20)	Boxing Day (December 26)

The activity space indicators were also constructed for a longer period so as to find out the period giving the best results. A different length of a period was categorized into seven such as 1-day, 5-days, 7-days, 10-days, 1-month, and 2-month in this study. All the points during each period were considered in the calculation when estimating the activity space.

2.2.2 Calculating activity space indicators

Minimum Convex Polygon

In the study, a 95 percent confidence level was used in order to alleviate the impact of large data points like outliers. Thus, MCP was more appropriate to estimate in individual activity spaces because it could better depict the shape of polygon based on the irregular range of data. The built-in function in R package called "adehabitatHR" was employed for the calculation (Calenge, 2006).

Ellipse

The confidence ellipse is determined by the following formula:

$$s_{xy} = \frac{1}{n-2}\sum_{i=1}^{n}(x_i - x)(y_i - y) \quad \text{(Eq.1)}$$

$$\mathrm{S} = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{bmatrix} \quad \text{(Eq.2)}$$

$$Ellipse\ size\ (Area) = 6\pi|S|^{1/2} \quad \text{(Eq.3)}$$

where *x* and *y* are referred to the arithmetic mean of all unique coordinates and n is the total number of activity locations (Schönfelder & Axhausen, 2003). In the study, a 95 percent confidence Ellipse was applied to describe the distribution of activity locations in space for both CDR and GPS data. The "car" package in R was adopted to estimate the area of an ellipse for individuals (Fox & Weisberg, 2019).

Radius of Gyration

In order to explore the individual's movement span, the radius of gyration was computed to determine how mobile phone subscribers moved widely along their travel trajectories:

$$r_g(t) = \sqrt{\frac{1}{K(t)}\sum_{x=1}^{K}(r_x - r_{cm})^2} \quad \text{(Eq.4)}$$

where K stands for the total number of detected sites, r_x indicates the x = 1, 2, …, K(t) location of an individual user, and r_{cm} states the centre of all observed locations during the experimental period (González et al., 2008; Chen et al., 2018). The author of the thesis developed the R script based on equation 4 to calculate the radius of gyration in measuring activity space.

Kernel Density

In this study, epanechnikov kernel function was used for comparison of subscribers' activity space size because its performance is considered the most efficient kernel function (Silverman, 1986; Wand & Jones, 1995).

An epanechnikov function is given by:

$$K(x) = \frac{3}{4}(1 - x^2) * 1 \; if \; x < 1 \qquad \text{(Eq.5)}$$

which leads to the following kernel density:

$$\hat{f}(x) = \frac{1}{ns^2}\sum_{j=1}^{n} K\left\{\frac{1}{s}(x - X_j)\right\} \qquad \text{(Eq.6)}$$

where *n* is the number of point observation, X_j is the location of j^{th} observation, and *s* is the smoothing parameter respectively (Vokoun, 2003; Vadrevu et al., 2018). Since the overlapping values are summed that produces the density, the smoothing parameter s is significant in the model where it manipulates the width of the kernel functions placed over each point (Schönfelder & Axhausen, 2003). In this research, a spatial bandwidth was calculated as follows:

$$\mathrm{s} = 1.77 \times \sigma \times n^{-\frac{1}{6}} \; where \; \sigma = 0.5 \times (\sigma_x + \sigma_y) \qquad \text{(Eq.7)}$$

where σ_x and σ_y are the standard deviations of the *x* and *y* coordinates of the locations, respectively (Silverman, 1986; Brunsdon & Singleton, 2015). The estimation of kernel density of individuals was done by using the R package called "adehabitatHR" (Calenge, 2006).

Total Travel Distance

The Euclidean distance method was employed to determine the general movement trajectories of subjects. It was calculated as follow:

$$Distance = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2} \qquad \text{(Eq.8)}$$

where A and B represent each pair of consecutive points of X and Y; $X = (A_1, \; A_2)$ and $Y = (B_1, \; B_2)$ (Kim et al., 2018). In this research, the summation of the subsequent recorded positions on the same trajectory was done based on the Euclidean distances between them. The

author of the thesis created the R script on the basis of equation 8 to calculate the total travel distance of each individual.

Entropy

Entropy was selected to analyze subscriber's visitation patterns. The formula of entropy was followed:

$$\mathrm{E} = -\sum_{x=1}^{n} p_x log_q p_x \qquad \text{(Eq.9)}$$

where p_x stands for the probability of visiting the location x, q indicates the number of unique locations, and n represents the total number of specific locations visited by users in the given movement pattern (Song et al., 2010; Zhao et al., 2016; Pi et al., 2018; Vanhoof, Schoors, et al., 2018). For CDR, the total number of visited locations was determined by estimating the number of times a person used mobile phones whereas the number of unique locations was calculated by quantifying the number of times an individual used mobile phones from a different cell tower than the previous cell tower. For GPS, the total number of visited locations was the overall GPS points for a user in a defined mobility range. The number of unique locations was estimated by a defined area of coverage by the user to determine the number of times the user performs activities within this defined scope. The defined area used for this calculation was the radius of 50m because it could be more appropriate to measure entropy. The author of the thesis came up with the script according to equation 9 for the calculation.

2.2.3 Data Wrangling

Prior to statistical analysis, pre-processing of data was conducted to filter data for both CDR and GPS. The filtering was done to remove data which have no values for activity space measurement methods and those who had a mobility coverage greater than the total area of Estonia. Individuals' CDRs recorded by a single cell tower are not considered for estimation but rather calculations made are based on CDRs collected for more than three different places. Therefore, based on this premise, most daily data were not sufficient to use kernel density and MCP indicators because they had a few number of CDRs (less than four). Since it is necessary to compare the accuracy of CDR-based measurements for all activity space indicators, data was subsequently reduced from 8961 to 477 (18.79%).

On the other hand, a definite age number of individuals was computed from the subtraction of data period from birth year (Table 6). Subjects were classified into three categories: “Young Adults”, “Middle-aged Adults”, and “Old-aged Adults”. For easier comparison, occupation was grouped into staff and students where staff includes head of the company, middle manager, top professionals, middle professionals, skilled worker, and office worker (Table 6). This was done because there are few data in specific occupation classes, so it is worth comparing between staff and students. Furthermore, the number of CDR data was categorized into four different groups to have a better understanding of the CDR data: (CDR <= 10), (11 <= CDR <= 20), (21 <= CDR <= 30), and (31 <= CDR).

2.2.4 Statistical Analysis

Descriptive statistics

Mean was deduced to identify the center of estimated activity space measurements. The formula of mean is as follows:

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} x \qquad \text{(Eq.10)}$$

where n indicates the sample size of each individual and x is the estimated activity space measurement (Holcomb, 2016).

Standard deviation was computed to explore the spread of the activity space measurements. If all estimated activity space measurements were closed to its mean, then the standard deviation would be smaller and vice-versa (Holcomb, 2016). This was calculated using the formula:

$$\text{Standard deviation} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2} \qquad \text{(Eq.11)}$$

where n is the sample size, x stands for the estimated activity space measurement, and $\bar{x}$ indicates the mean of all the estimated activity space measurements, respectively.

A 95% confidence interval is a range of values in which there is 95% certainty that the true value falls within. The confidence interval was employed to describe the reliability of the measurements. It was calculated as follow:

$$Confidence\ interval = \bar{x} - 1.96 * \frac{s}{\sqrt{n}} \qquad \text{(Eq.12)}$$

where n is the sample size, x stands for the standard deviation, and $\bar{x}$ indicates the mean of all the estimated activity space measurements, respectively (Siegel, 2012).

Absolute Difference (AD) was computed to determine how the accuracy of CDR-based activity space compares to that of GPS-based measurement. In the research, AD was deduced by the difference between CDR and GPS-based activity spaces for all indicators. The AD was calculated using the formula:

$$\text{AD} = |V_{CDR} - V_{GPS}| \quad \text{(Eq.13)}$$

where $V_{CDR}\ and\ V_{GPS}$ indicate the CDR and GPS-based measurements (Oracle, 2011; Weisstein, 2007). If AD is closer to 0, it could be concluded that the accuracy is higher; otherwise, the accuracy is lower.

Correlation Analysis

Spearman's rank correlation is regarded as more robust compared to Pearson's correlation coefficient because it is less sensitive to skewed data and outliers (Lehman et al., 2005; Dodge, 2008). Therefore, in this study, Spearman's rank correlation was employed to assess whether two different sorts of mobile positioning data (CDR and GPS) are correlated with each other and also measure the strength of association between the number of CDRs and the accuracy of CDR-based activity spaces.

Regression Analysis

To investigate factors affecting the accuracy of CDR-based human mobility patterns, linear mixed models (LMMs) were employed. LMMs are widely used to determine a causal relationship between variables when there is non-independence in the data (Van Dongen et al., 2004; Harrison et al., 2018). The linear mixed model incorporates both fixed and random factors in which fixed factors that are of interest in the research and can be controlled whereas random factors cannot be controlled experimentally (West et al., 2007; Winter, 2013).

In this study, measurements were obtained repeatedly from the same subjects and variables were not independent of each other. Accordingly, interesting factors on the accuracy of CDR-based measurements such as temporal variability, the number of CDR, and socio-demographic characteristics were referenced as fixed factors and the subjects were used as random factors, respectively. The absolute difference variables of each activity space indicator were used as

dependent variables and log-transformed because they did not contain negative values and showed right-skewed distributions. However, entropy was not used in this regression analysis because it does not show the extent of the activity space but the internal structure of data. So, it would not be appropriate to identify the effect of varied factors on the accuracy of CDR-based measurements.

Two linear mixed models were constructed regarding the types of interesting factors. All temporal factors (days of the week, months, and holidays) were put in LMM together for the analysis of the temporal effect on the accuracy of CDR-based measurements (Eq.14) whereas the number of daily CDR variables is added in the model of socio-demographic factors (gender, age group, marital status, occupation, and the number of household members) for the analysis of personal characteristics effect on the accuracy of CDR-based computations (Eq.15).

The following are the detailed model specifications:

$$\log(Y_{AD}) = \beta_0 + \beta_{Days\ of\ the\ week} * X_{Days\ of\ the\ week} + \beta_{Months} * X_{Months} + \beta_{Holidays} * X_{Holidays} + Z_{Subject} * U_{Subject} + \epsilon \quad \text{(Eq.14)}$$

$$log(Y_{AD}) = \beta_0 + \beta_{CDR} * X_{CDR} + \beta_{Gender} * X_{Gender} + \beta_{Age\ group} * X_{Age\ group} + \beta_{Marital\ status} * X_{Marital\ status} + \beta_{Occuppation} * X_{Occupation} + \beta_{Number\ of\ household\ members} * X_{Number\ of\ household\ members} + Z_{Subject} * U_{Subject} + \epsilon \quad \text{(Eq.15)}$$

where Y_{AD} is a vector of the continuous absolute difference of measurements for the subjects, β_0 represents the intercept, β_k is the regression coefficient for a specific fixed variable, $X_{Days\ of\ the\ week}$ is a variable based on days of the week, X_{Months} is a month variable, $X_{Holidays}$ is a holiday variable, X_{CDR} is the number of CDRs, X_{Gender} is a gender variable, $X_{Age\ group}$ is an age variable, $X_{Marital\ status}$ is a marital status variable, $X_{Occupation}$ is an occupation variable, and $X_{Number\ of\ household\ members}$ represents the family size correspondingly. Furthermore, $Z_{Subject}$ is a random intercept, $U_{Subject}$ shows a random effect for each subject, and ϵ represents a general error term, respectively.

In linear mixed-effects models, all fixed variables were transformed into dummy variables. To avoid the dummy variables trap, one feature from each of those dummy variables was used as a reference group. For example, CDR <= 10, female, young adults, single status, staff, and single household size features were not used for the regression analysis.

The perfect multicollinearity was identified while assessing an equation using IBM SPSS Statistics software which gives a correlation coefficient value of 1 or -1 if the model suffers from the perfect multicollinearity. However, the Variance Inflation Factor (VIF) was used to detect the imperfect multicollinearity which was computed as:

$$\frac{1}{(1-R^2)} \quad \text{(Eq.16)}$$

The model having the largest value of VIF greater than 10 or the mean of VIFs significantly larger than 1 is considered as evidence of the problem of the imperfect multicollinearity (Chatterjee & Hadi, 2012). Thus, if VIF exceeds 10 or its mean is greater than 1, the independent variables should be examined individually or removed from the model.

In this research. there was no sign that linear mixed models suffered from an imperfect multicollinearity problem since their mean VIFs were not significantly larger than 1 (2.1 for temporal factors and 3.4 for personal characteristics). Therefore, the linear mixed models were performed without regressing the independent factors on the dependent variable for each indicator separately. Furthermore, a five percent significance level was used to confirm or reject our tests. The significant level explains the probability of rejecting the null hypothesis when it is true (Stock & Watson, 2014). Correspondingly, an independent variable is significantly different from zero if its p-value is lower than 0.05.

3. Results

3.1 Activity spaces based on CDR and GPS data

Overall, all indicators show skewed distributions and the density of asymmetric graphs (Figure 4). All distributions of activity space indicators are positively skewed, in which more values fall toward the lower side of the scale and there are very few higher values. In general, CDR data have a longer peak and skinny tail which means that CDR based measurements would be smaller than GPS data. However, the distribution of Entropy is negatively skewed and both CDR and GPS-based measurements are similar to each other.

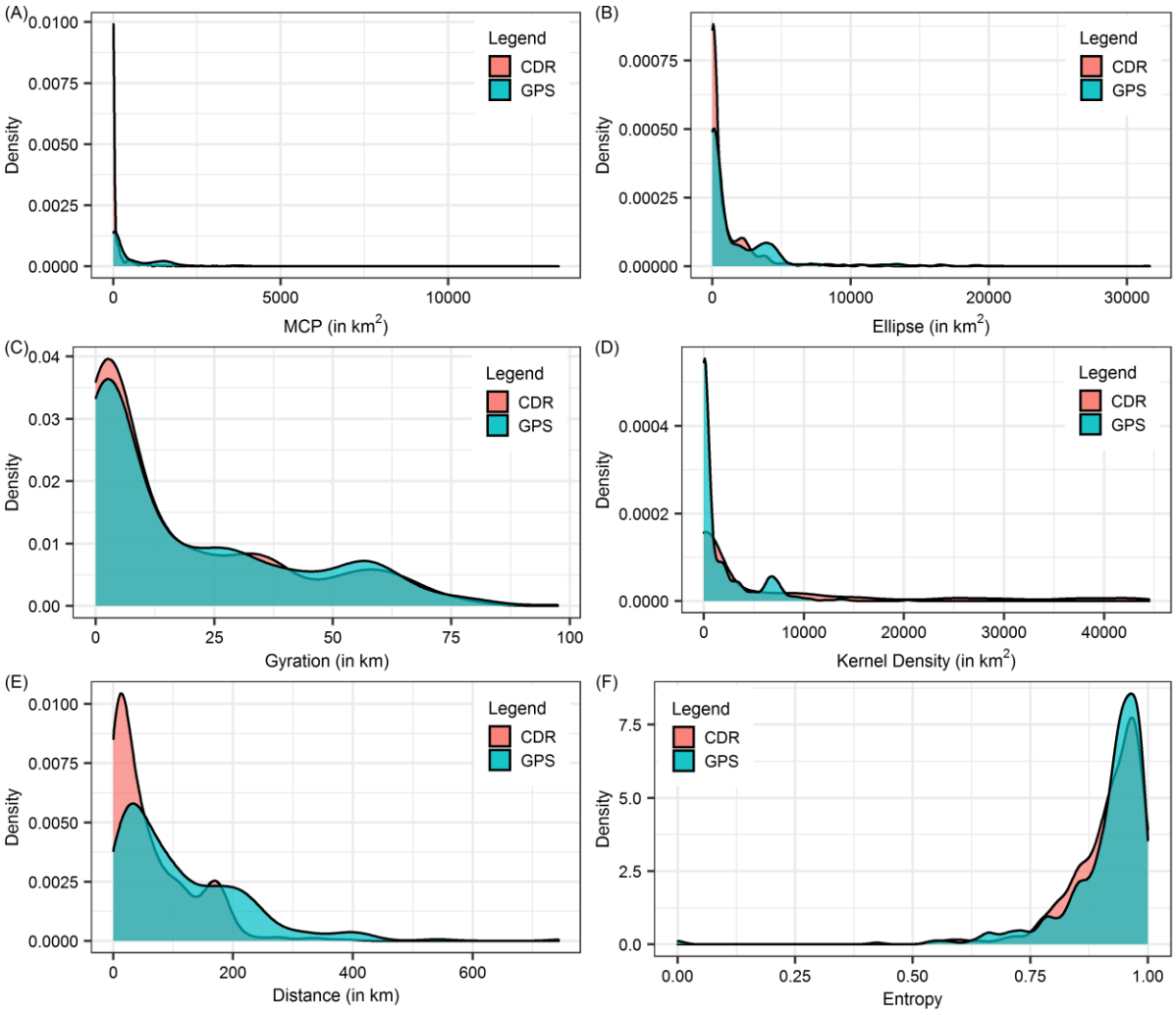

Figure 4. The density graph on activity space indicators regarding both CDR and GPS data

The positive relation appears for all indicators in which both CDR and GPS move in the same direction (Table 8). The apparent correlations are observed for the most indicators except entropy. The kernel density indicator had the highest correlation value of 0.734, followed by

gyration, ellipse, distance, MCP, and entropy. Besides, four indicators like kernel density, gyration, ellipse, and distance had similar correlation values to each other. The relation existing between CDR and GPS-based measurements regarding entropy was proven to be a weak positive association between variables as 0.170. However, the correlation coefficients were proven to be significant for all activity space indicators since the p-values were less than 0.05.

Table 8. Correlation between activity spaces based on CDR and GPS data

	MCP	Ellipse	Gyration	Kernel Density	Distance	Entropy
Correlation coefficient	0.587*	0.722*	0.730*	0.734*	0.702*	0.170*

** indicates 5% significance level*

To have a clear understanding of the relationship between CDR and GPS-based measurements, the scatter plot and summary of a given data set were constructed (Figure 5 and Table 9). As more points deviated above the diagonal line, CDR-based measurements were underestimated when it comes to MCP, ellipse, and distance compared to using GPS data. There was a similar pattern of estimation for gyration and entropy in both CDR and GPS. On the other hand, the activity space of kernel density was overestimated by CDR than GPS. Moreover, kernel density had such a large deviation in CDR data.

In general, entropy and gyration provide the lowest average absolute difference of the measurement as 0.1 and 8.3, respectively with low deviations in measurements followed by distance (60.1), MCP (527.6), ellipse (1191.7), and kernel density (5448.7). Kernel density had the lowest accuracy of CDR-based activity space and there was a large deviation (9588.5) in measurements as well.

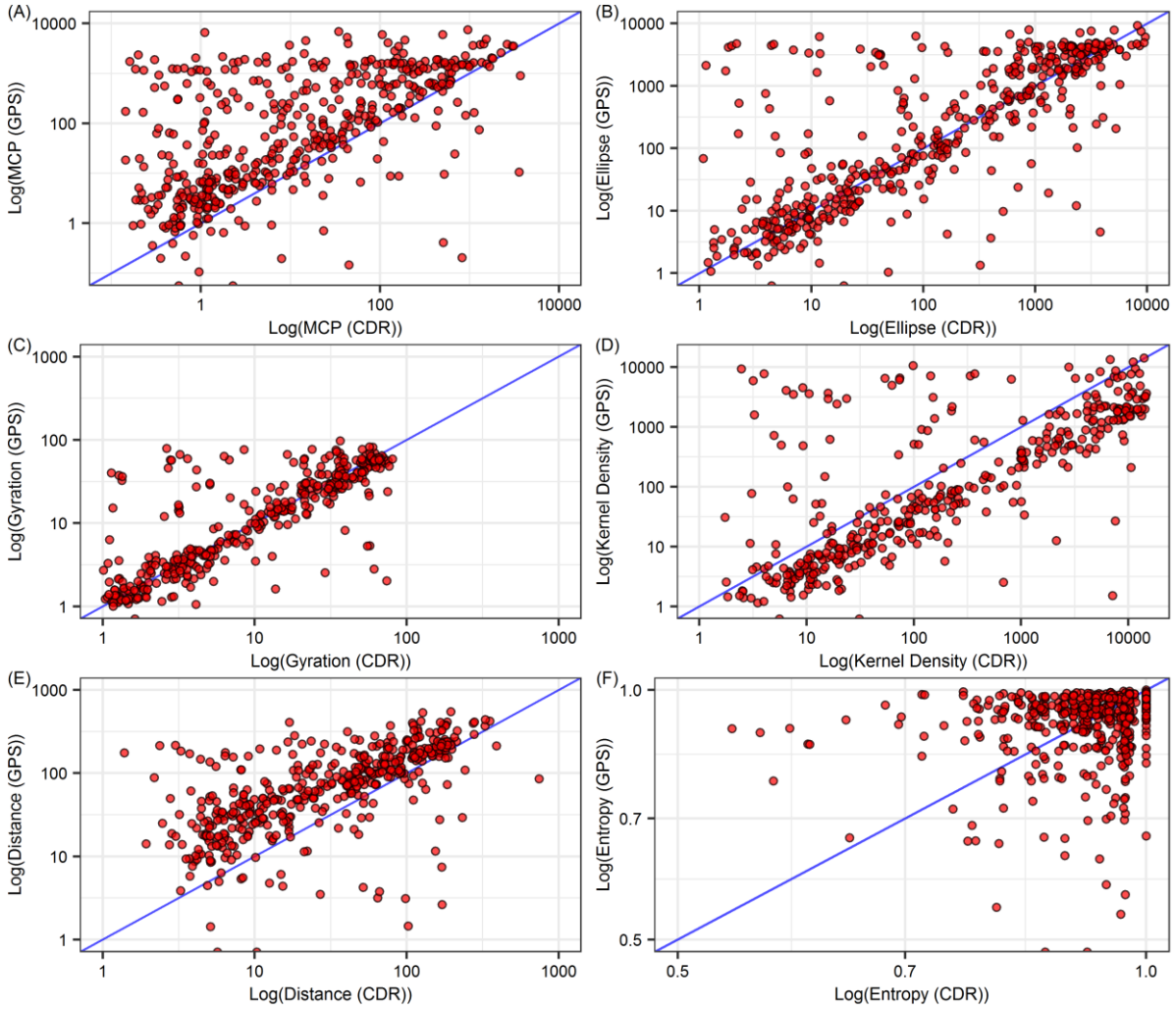

Figure 5. The scatter between CDR and GPS-based measures

Table 9. Descriptive statistics of CDR and GPS-based indicators

		MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)	Entropy
CDR	Mean	163.9	1191.3	17.2	6542.0	63.9	0.9
	S.D.	432.8	2551.2	20.8	11399.9	73.8	0.1
GPS	Mean	628.8	1735.7	18.9	1753.0	110.9	0.9
	S.D.	432.8	3324.6	22.0	2882.1	100.3	0.1
Absolute Difference	Mean	527.6	1191.7	8.3	5448.7	60.1	0.1
	S.D.	1123.6	2550.1	14.6	9588.5	69.9	0.1

3.2 Effects of Temporal variability

3.2.1 The days of the week

It can be observed that Mondays, Fridays, Saturdays, and Sundays have comparatively higher values for all indicators rather than Tuesdays, Wednesdays, and Thursdays (Figure 6). A similar pattern is depicted in Table 10 when comparing weekends to weekdays; weekends' values are noticeably higher. As depicted in Table 10, kernel density has the highest absolute difference among other indicators, in which the weekends have a value of 6585.3 while the weekdays are represented by the value of 5180.7. Moreover, there was enough statistical evidence that days of the week factors like Tuesday, Wednesday, and Thursday had lower absolute differences of the measurement for all activity space indicators compared to the reference factor of Sunday (Table 12).

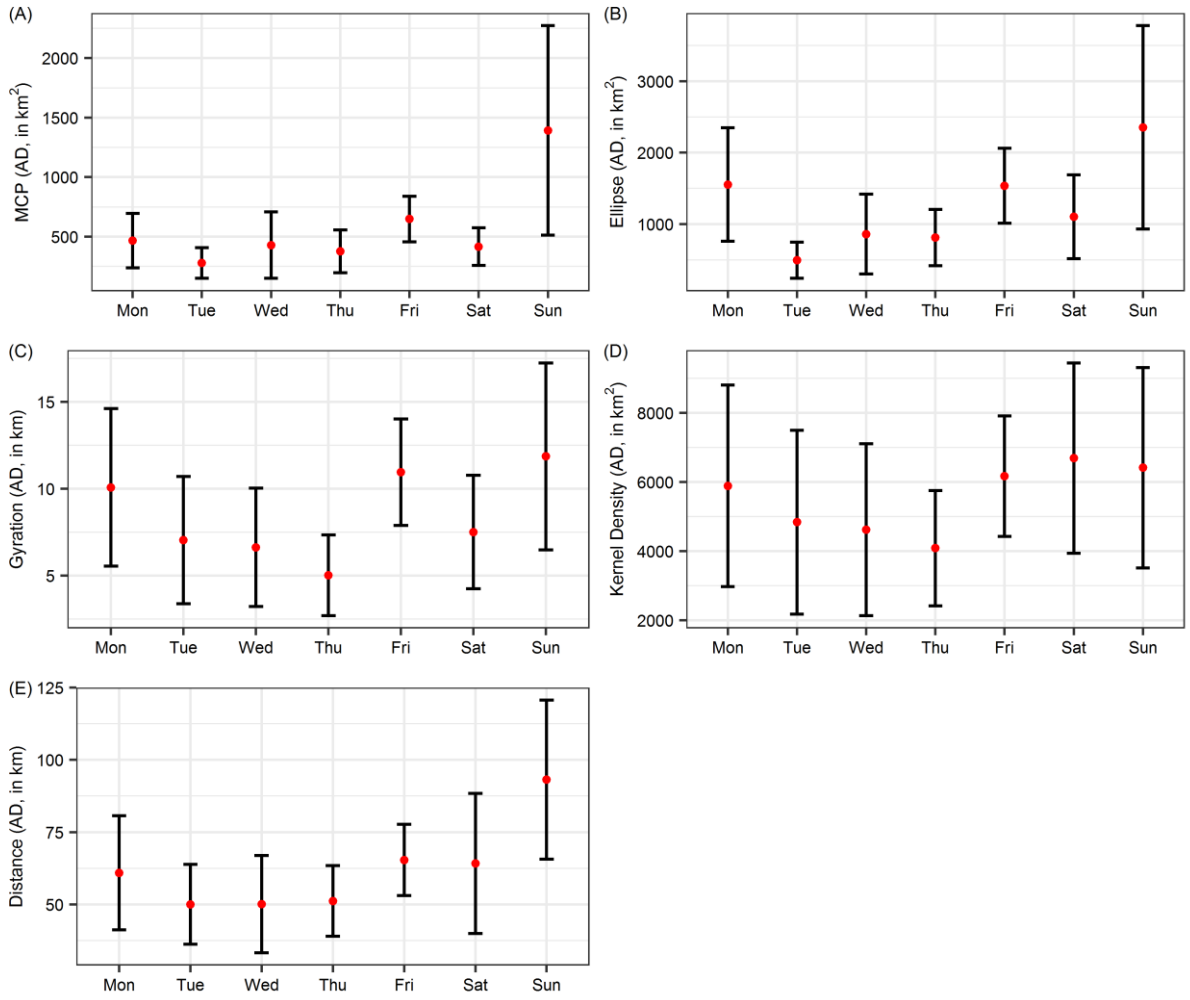

Figure 6. The mean value and confidence interval (CI, 95%) of the absolute difference on days of the week; red points - mean value and whiskers - confidence interval

Table 10. Descriptive statistics of the accuracy of CDR-based measurements on weekdays and weekends

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Weekdays	Mean	468.0	1102.6	8.1	5180.7	56.6
	S.D.	944.7	2404.1	14.8	9575.6	64.7
Weekends	Mean	780.4	1569.7	9.1	6585.3	75.0
	S.D.	1667.1	3080.8	13.6	9612.7	87.5

3.2.2 Months and Season

MCP has similar values for all months with the exception of January having the highest absolute difference value (Figure 7). Gyration and ellipse illustrate a similar monthly irregular pattern but show higher values for January. Kernel density seems to have closely related values across all months with September, October, November, and December having relatively higher values. Distance had higher values for January, July, and May respectively, but September and October had the lowest absolute difference values. However, the months of the year factors had a significant effect on the absolute difference of the measurement only for the distance indicator, while it was proven not to be significantly different from 0 at the 5% level for other indicators (Table 12). The absolute difference in distance was higher for January and August than for the reference factor of December.

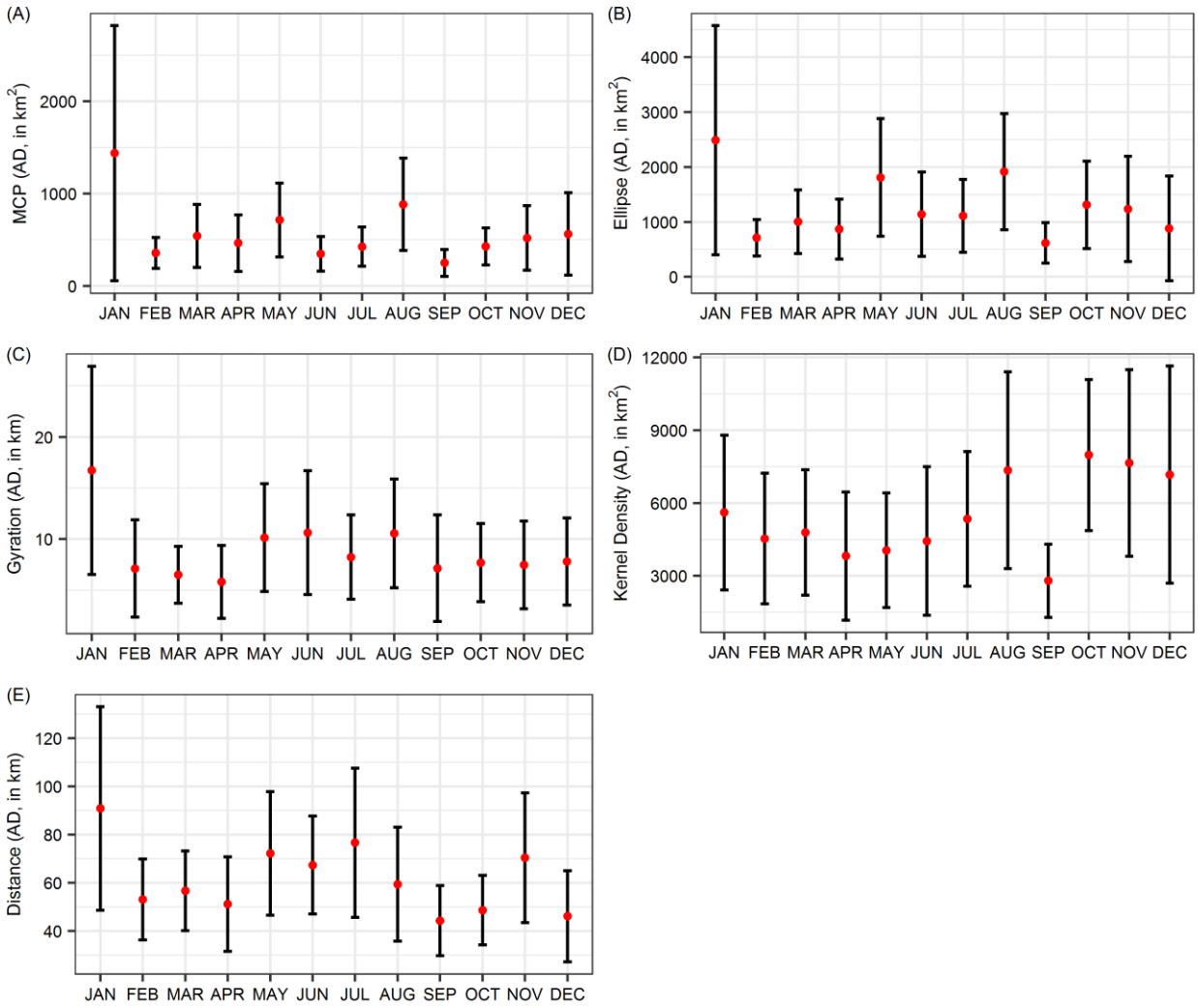

Figure 7. The mean value and confidence interval (CI, 95%) of the absolute difference in months; red points - mean value, whiskers - confidence interval.

In terms of the absolute difference for seasons, MCP revealed the highest value for winter (December – February), spring (March – May), summer (June – August), and autumn (September – November) individually (Figure 8). Ellipse depicted a high absolute difference values for summer, spring, winter, and autumn. Gyration showed the highest value for summer followed by winter, autumn, and spring, respectively. Kernel density had the highest values in autumn, winter, summer, and spring. Distance indicated the highest value for summer followed by spring, winter, and autumn, respectively.

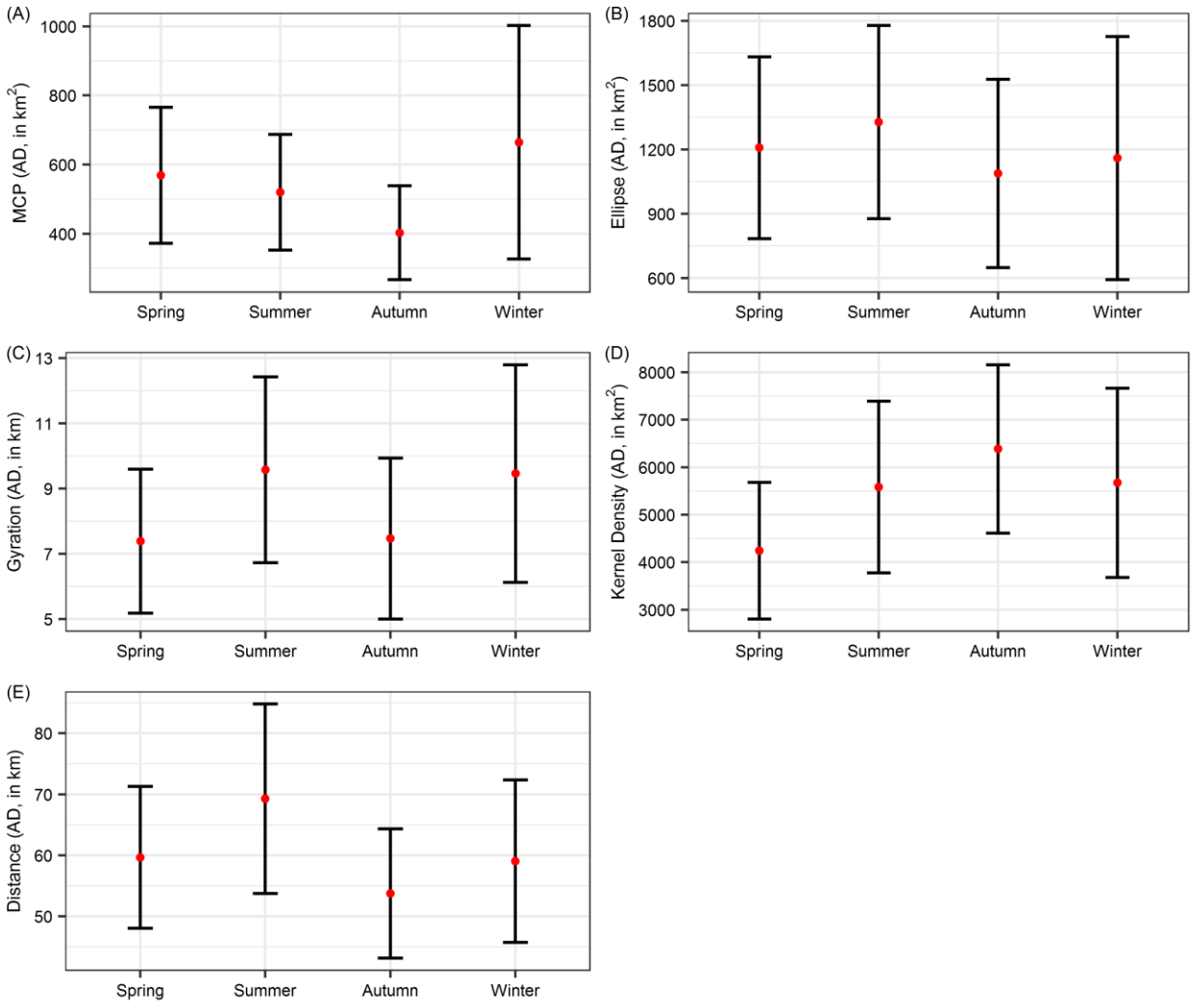

Figure 8. The mean value and confidence interval (CI, 95%) of the absolute difference for seasons; red points - mean value, whiskers - confidence interval

3.2.3 Holidays

There are similar trends in ellipse, kernel density, and MCP which showed the largest absolute difference happening on Good Friday and Easter Sunday, but lowest values were on New Year's Day, Independence Day, and Boxing Day (Figure 9). In terms of the gyration indicator, it shows a similar pattern with the Distance indicator, but the highest value occurred on Good Friday and other days such as New Year's Day, Independence Day, Christmas Eve, and Boxing Day were similar to each other. In terms of the distance indicator, Labor Day had the largest absolute difference followed by Good Friday and Easter Sunday, while the lowest value was on New Year's Day. However, the holiday factor of Good Friday had a significant effect on the absolute difference in the measurement at the 5% level regarding MCP, gyration, and kernel

density (Table 12). Good Friday had a higher absolute difference on average for these three indicators compared to the reference factor of Boxing Day.

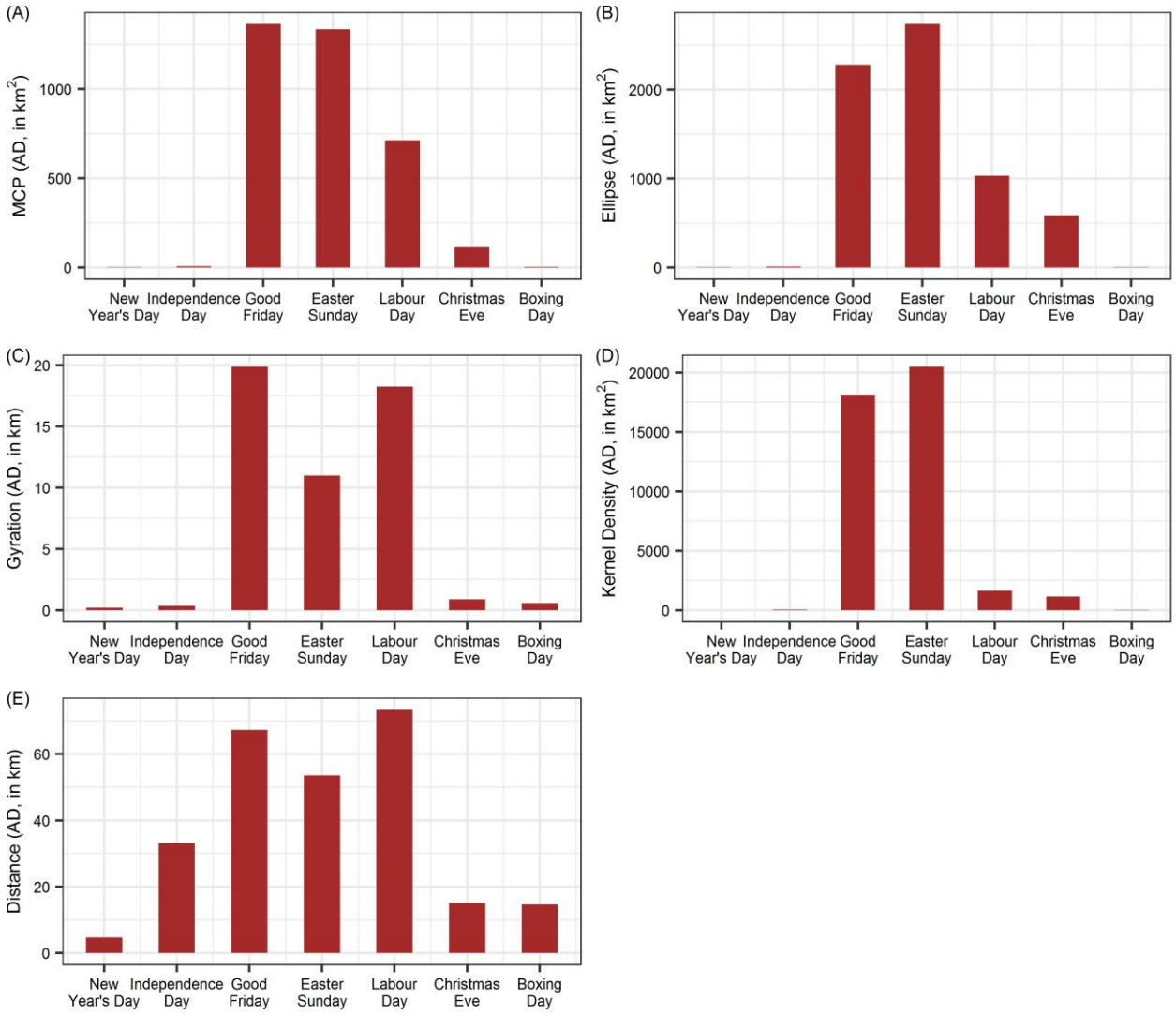

Figure 9. The mean absolute difference of the measurement on holidays

It can be seen that holidays tend to have low accuracy for all activity space indicators except distance (Table 11). Regarding gyration, it showed lower absolute difference with a value of 9.9 for holidays and 8.3 for regular days whereas kernel density indicated higher absolute difference with a value of 6813.0 for holidays and 5422.6 for regular days. Also, there was a large deviation in measurements for kernel density.

Table 11. Descriptive statistics of the accuracy of CDR-based measurements on holidays and regular days

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Holidays	Mean	623.9	1106.7	9.9	6813.0	44.7
	S.D.	712.7	1493.7	12.9	12425.9	43.5
Regular days	Mean	525.7	1193.3	8.3	5422.6	60.4
	S.D.	1130.4	2567.1	14.7	9541.0	70.3

Table 12. The summary of the linear mixed model: Absolute Difference of the measurement

	Log (MCP) β	Log (Ellipse) β	Log (Gyration) β	Log (Kernel Density) β	Log (Distance) β
Days of the week (ref.: Sunday)					
Monday	-1.531*	-1.022	-0.576	-0.902	-0.487
Tuesday	-2.093*	-2.147*	-1.172*	-1.672*	-0.431
Wednesday	-2.379*	-1.955*	-1.049*	-1.634*	-0.736*
Thursday	-2.143*	-2.041*	-1.353*	-1.666*	-0.585*
Friday	-1.120*	-0.948	-0.402	-0.751	-0.313
Saturday	-1.232*	-1.072	-0.587	-0.239	-0.251
Months of the year (ref.: December)					
January	1.167	1.789*	1.036	1.209	0.666
February	0.080	0.314	-0.774	-0.077	0.062
March	0.504	1.010	0.180	-0.031	0.274
April	-0.560	0.183	-0.423	-0.358	0.132
May	-0.040	0.883	-0.011	0.069	0.067
June	0.341	0.630	0.063	0.315	0.488
July	0.244	0.956	0.236	0.558	0.175
August	1.017	1.766*	0.616	1.224	0.216
September	-0.363	0.367	-0.378	-0.126	0.037
October	0.143	0.703	-0.073	0.619	0.050
November	0.446	0.921	-0.279	0.519	0.364
Holidays (ref.: Boxing Day)					
New Year's Day	-3.156	-3.641	-2.318	-4.391	-1.978
Independence Day	-0.411	0.047	0.313	0.306	0.108
Labor Day	0.828	1.129	1.908	0.770	0.541
Good Friday	4.080*	3.484	3.102*	4.261*	0.591
Easter Sunday	2.389	2.375	1.865	3.747	-0.152
Christmas Eve	0.653	2.553	-0.594	1.245	-0.930
-2LL	2226.224	2292.698	2000.200	2334.017	1526.243
Observations	477				

** indicates 5% significance level*

3.3 Effects of personal characteristics

3.3.1 The number of CDRs

Closer inspection of Figure 10 shows there is a positive association between the number of CDRs and accuracy of CDR-based measurements. As the number of CDRs increase, absolute differences of each activity space indicator decrease. However, it is not such a strong relationship. Nevertheless, the general tendency that the accuracy and the number of CDR increase together is indisputably present.

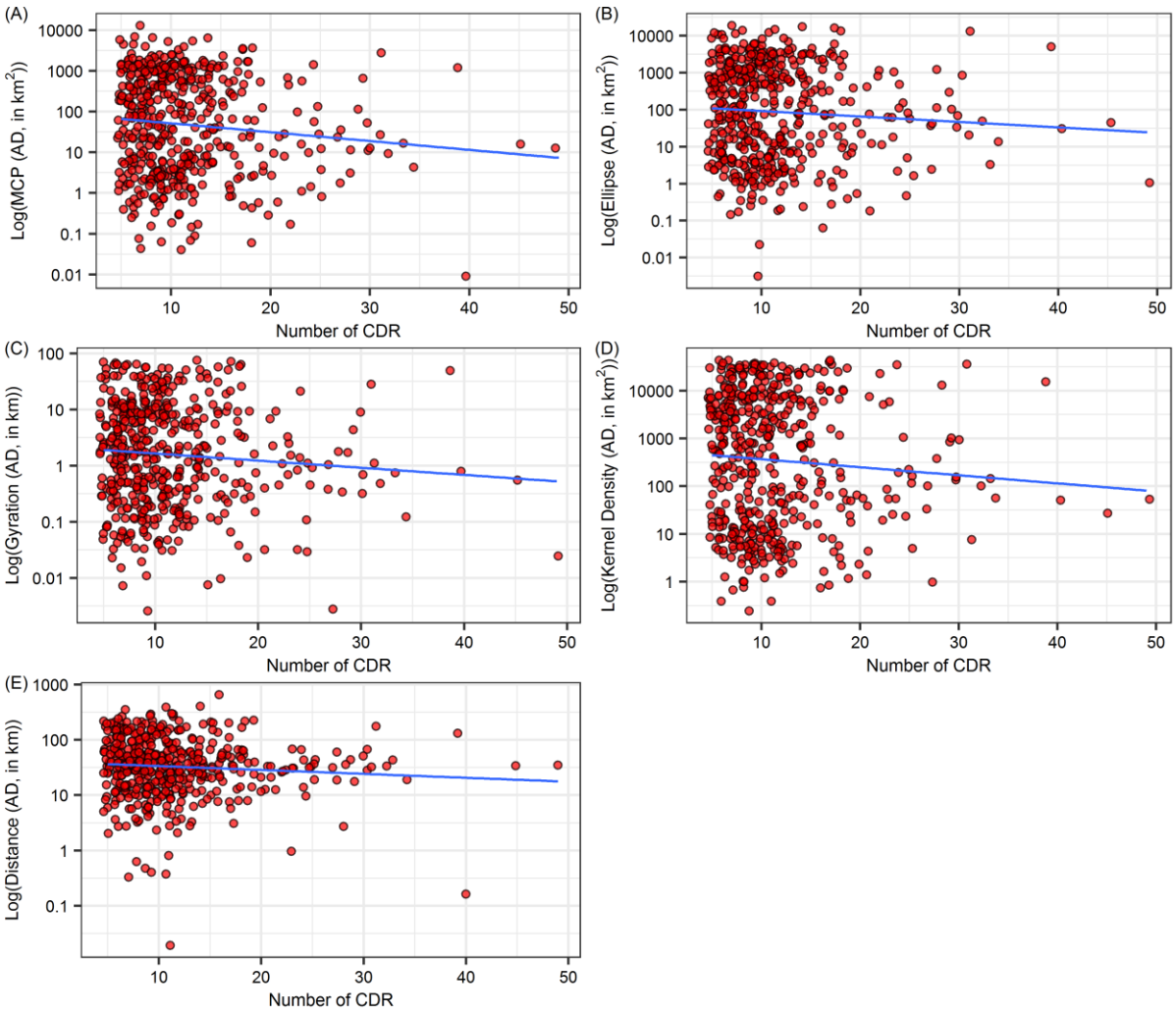

Figure 10. Relationship between the number of CDR and Absolute difference based on the number of CDRs

The downward trend appears for all indicators as the number of CDR gets bigger, but the correlation coefficients are proven to be very weak (Table 13). Additionally, only the correlation coefficient of distance was proven to be significantly different from zero. However,

the correlation does not imply a cause-and-effect relationship between variables. Thus, the regression analysis was conducted to investigate whether the number of CDR leads to an increase in the accuracy of CDR-based measurements or not.

Table 13. Correlation between the number of CDRs and accuracy of CDR-based measurements

	MCP	Ellipse	Gyration	Kernel Density	Distance
Correlation coefficient	-0.08	-0.064	-0.056	-0.066	-0.116*

** indicates 5% significance level*

In the subsequent table, the values related to the accuracy of activity space based on five indicators are going to be compared within these four above-mentioned groups (Table 14). Group 3 has the best estimation of activity spaces because they produce the smallest in the absolute difference of all indicators. However, the number of CDRs were proven not to be significantly different from 0 at the 5% level for all activity space indicators. The number of CDRs had no causal effect on the absolute difference of the measurement in general (Table 20).

Table 14. Descriptive statistics of the accuracy of CDR-based measurements based on the number of CDRs

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Group 1 (CDR <= 10)	Mean	589.9	1220.2	8.4	5831.8	64.0
	S.D.	1267.3	2425.7	14.4	9826.6	63.7
Group 2 (11 <= CDR <= 20)	Mean	501.5	1266.7	9.1	5240.3	59.0
	S.D.	968.1	2788.4	15.8	9370.3	83.2
Group 3 (21 <= CDR <= 30)	Mean	149.9	246.6	2.5	3176.3	32.3
	S.D.	315.6	379.4	4.3	7743.5	17.8
Group 4 (CDR >= 31)	Mean	452.5	2063.0	9.1	5748.6	56.5
	S.D.	957.7	4549.7	17.8	12388.2	58.2

3.3.2 Socio-demographic Factors

The result of the absolute difference value for different gender indicated that the accuracy of the mean value of females by each indicator is lower than males (Table 15). Regarding kernel density, the absolute difference has the biggest value of 5761.8 for females and 4816.6 for males, while gyration has the smallest absolute difference value of 8.9 for females and 7.1 for males. There was not enough statistical evidence that the regression coefficient for gender was

significantly different from 0 at the 5% level (Table 20). The mean absolute differences of all indicators for males and females were not statistically different from each other in general.

Table 15. Descriptive statistics of the accuracy of CDR-based measurements based on gender

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Female	Mean	591.7	1343.0	8.9	5761.8	63.9
	S.D.	1176.6	2739.7	14.9	9434.0	75.1
Male	Mean	398.2	886.3	7.1	4816.6	52.4
	S.D.	999.0	2091.1	13.9	9893.3	57.4

The absolute difference value of Old-aged adults is smaller in all the indicators and on the sharp contrast, the Young Adults group has higher values except in kernel density and distance (Table 16). The Middle-aged Adults group has the highest absolute difference values of 7384.8 and 63.4 in terms of kernel density and distance, respectively. From the regression analysis, middle-aged adults had a lower absolute difference on average compared to young adults at the 5% level regarding only gyration and distance, while it was proven not to be statistically significant for other indicators (Table 20).

Table 16. Descriptive statistics of the accuracy of CDR-based measurements based on the age group

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Young Adults	Mean	577.2	1333.2	9.2	5475.0	61.3
	S.D.	1237.6	2792.0	16.1	9794.6	74.7
Middle-aged Adults	Mean	520.3	1076.8	6.8	7384.8	63.4
	S.D.	895.3	2496.6	12.2	12643.3	69.3
Old-aged Adults	Mean	335.2	702.1	5.7	4114.2	53.1
	S.D.	675.5	1145.6	7.8	5571.1	47.1

Overall, the absolute difference values have a similar pattern for all activity space indicators. All groups except partners not living together have closely related values for all indicators (Table 17). Cohabitation group has the absolute difference values in terms of gyration, ellipse, and MCP as 9.4, 1337.9, and 594.1, respectively, whereas the single group has the highest values with respect to distance and kernel density as 68.0 and 5978.0 individually. However, the group of Partners (not living together) shows the lowest absolute difference values for all indicators except ellipse which is the lowest for Married people. On the other hand, none of the marital status factors were proven to be significantly different from 0 at the 5% level for all activity space indicators (Table 20).

Table 17. Descriptive statistics of the accuracy of CDR-based measurements based on marital status

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Married	Mean	422.3	886.8	6.1	5151.7	59.5
	S.D.	782.8	1824.3	9.7	8269.7	56.6
Cohabitation	Mean	594.1	1337.9	9.4	5491.3	58.5
	S.D.	1342.3	2869.6	16.2	10002.5	69.4
Without partner	Mean	501.6	1270.4	9.1	5978.0	68.0
	S.D.	755.2	2556.1	16.7	10700.9	93.9
Partners (Not living together)	Mean	210.0	1049.0	5.2	2820.1	28.4
	S.D.	175.2	1284.9	5.0	2436.8	18.3

The staff has the highest absolute difference values for all indicators compared to students. The difference in measurements is largest for kernel density as 11266.4 with a higher deviation (14362.7) in measurement whereas the gyration indicator represents the lowest difference for these two groups (12.4 for staff and 7.5 for students) (Table 18). However, students had a lower absolute difference on average compared to staff only for the gyration indicator, while it was proven not to be statistically significant at the 5% level for other indicators (Table 20).

Table 18. Descriptive statistics of the accuracy of CDR-based measurements based on occupation

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
Staff	Mean	800.4	1854.9	12.4	11266.4	80.5
	S.D.	1336.9	3353.3	17.5	14362.7	101.2
Student	Mean	442.1	1107.0	7.5	4445.4	55.3
	S.D.	817.1	2338.3	14.6	8485.1	64.4

Looking through the detail of table 19, it has been apparent that the absolute difference values based on each indicator follow a similar pattern. The absolute difference value for the single-family is significantly lower than families with more than two members. On the other hand, the family with three members tends to show the lowest accuracy of CDR-based measurements for all indicators except for distance which is the highest for the family with four members. However, the regression analysis indicated that the family with three members had a higher absolute difference of the measurement than for the single-family regarding only gyration and kernel density (Table 20).

Table 19. Descriptive statistics of the accuracy of CDR-based measurements based on the family size

	Absolute Difference	MCP (km^2)	Ellipse (km^2)	Gyration (km)	Kernel Density (km^2)	Distance (km)
One Person	Mean	202.3	308.0	4.2	2494.2	41.8
	S.D.	456.5	828.9	9.5	7431.4	51.4
Two People	Mean	451.2	1125.2	7.5	4520.9	56.2
	S.D.	844.2	2423.4	14.4	8555.7	65.8
Three People	Mean	917.2	1983.8	13.3	12783	75.6
	S.D.	1527.0	3232.7	16.8	14459.6	74.6
Four People	Mean	659.0	1898.5	11.9	8213.8	81.7
	S.D.	869.5	3225.8	20.1	12581.5	117.7

Table 20. The linear mixed model analysis of accuracy the measurement; Absolute Difference

	Log (MCP) β	Log (Ellipse) β	Log (Gyration) β	Log (Kernel Density) β	Log (Distance) β
Number of CDR (ref.: CDR <= 10)					
11 <= CDR <= 20	-0.155	-0.115	0.230	-0.215	-0.026
21 <= CDR <= 30	-0.752	-0.633	-0.639	-0.683	-0.246
31 <= CDR	-0.578	-0.221	-0.163	-0.217	-0.245
Gender (ref.: Female)					
Male	-0.910	-0.387	-0.446	-0.826	-0.401
Age Group (ref.: Young Adults)					
Middle-aged Adults	-1.589	-2.130	-2.225*	-2.357	-1.371*
Old-aged Adults	2.607	-0.516	-0.538	0.205	1.630
Marital status (ref.: Without partner)					
Married	-0.647	-0.961	-0.869	-1.128	-0.029
Cohabiting	-1.917	-1.495	-1.293	-2.193	-0.969
Partnership (Not living together)	-3.411	-0.042	-0.990	-0.348	-3.218
Occupation (ref.: Staff)					
Student	-1.456	-1.803	-1.786*	-1.295	-0.955
The size of family (ref.: 1 person)					
2 people	3.619	3.059	2.110	3.337	1.277
3 people	4.300	4.308	3.098*	5.690*	1.431
4 people	2.368	2.359	0.899	1.958	0.578
-2LL	1586.153	1626.353	1416.143	1652.559	1068.190
Observations	477				

** indicates 5% significance level*

3.4 Activity space indicators for a longer period

The higher the length of the period, the higher absolute differences of CDR-based measurements except for kernel density which shows the opposite (Figure 11). In kernel density, the absolute difference is the highest on a 5-days period and it starts declining as the length of the period becomes longer.

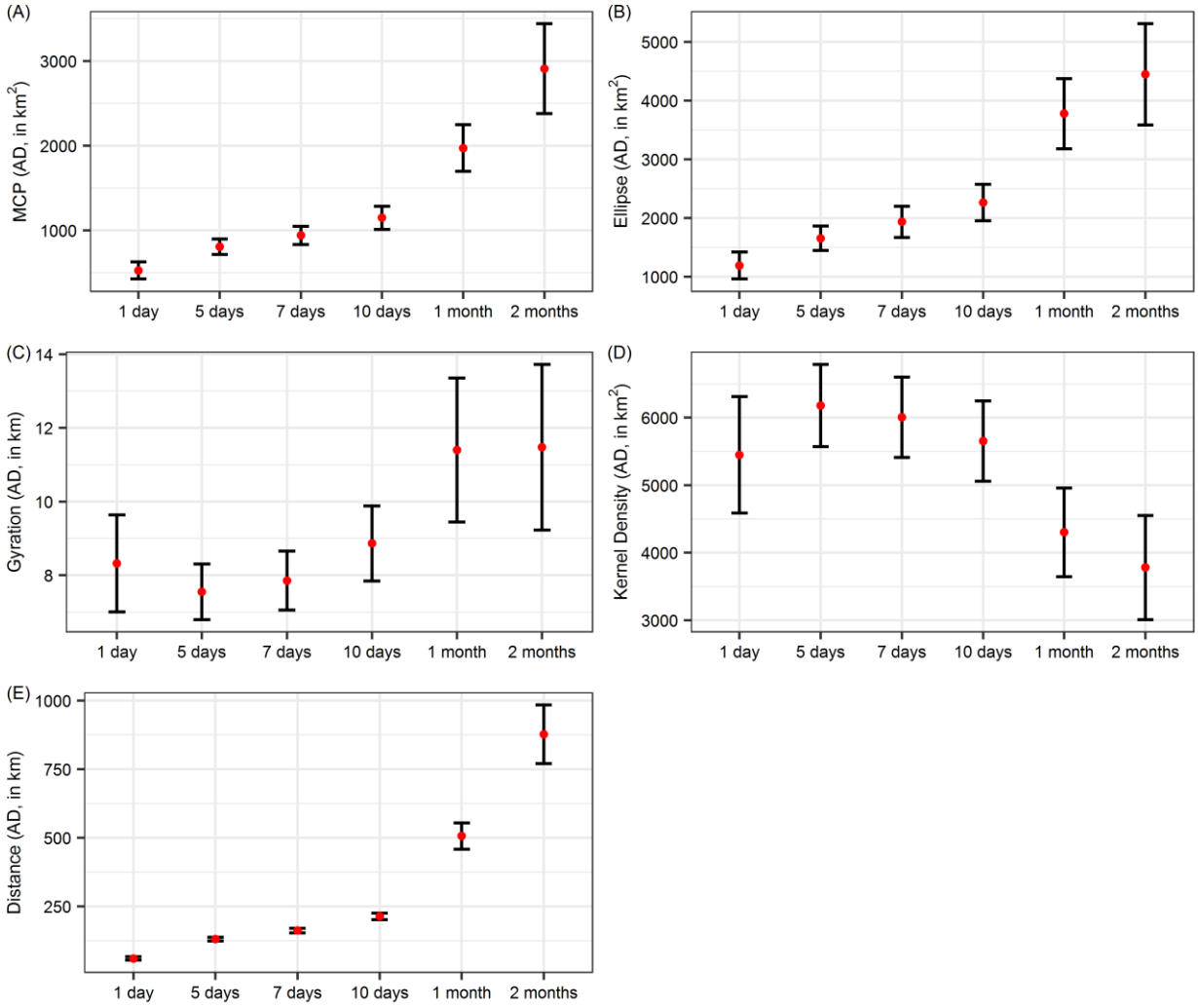

Figure 11. Graph of the mean value and confidence interval (CI, 95%) of the absolute difference; red points - mean value, whiskers - confidence interval.

4. Discussion

This study focused on CDR-based individual activity spaces in comparison with GPS-based measurements. The general relationship between CDR and GPS-based activity spaces was explored and analyzed the accuracy of CDR-based measurements based on selected activity space indicators. Moreover, how temporal factors and personal characteristics impact on the accuracy of CDR-based measurements were investigated. Previous research generally focused on the usage of CDR data to estimate the individual movement patterns and paid less attention to the empirical consensus on the comparison of CDR and GPS data for estimating individual activity spaces (Ahas et al., 2010; Järv et al., 2015; Xu et al., 2016; Vanhoof, Reis, et al., 2018). The results from this research can contribute to the existing studies related to CDR data and enhance the understanding of the individual spatial movement patterns as well.

This study has underlined four key discoveries. First, the general analysis shows both CDR and GPS-based measurements have similar skewed distributions for all activity space indicators and positive associations as well. In comparison to other indicators of CDR-based measurements, gyration and entropy were more closely related to GPS-based measurements. Similarly, Zhao et al. (2016) argued that CDR data are good enough to explore the human mobility pattern concerning gyration and entropy. Kernel density has the lowest accuracy and large deviations based on the absolute difference between CDR and GPS-based measurements. This is because the number of CDR is not frequent enough to capture the underlying movement patterns of subsets when kernel density is applied.

Second, the impact of different temporal scales on the accuracy of CDR-based activity spaces was considered. The results show that the accuracy of the measurements is generally observed higher during the weekdays compared to the weekends regarding all activity space indicators. This result coincides with previous findings (Kamruzzaman & Hine, 2012) that individuals tend to have higher movement patterns during the weekend compared to the weekday. This could make the accuracy of the measurement lower for the weekend as people tend to move around more. One other reason would be that people make fewer call activities on weekends, which means that they have less CDR location points (Abeele et al., 2016). On the other hand, the monthly factor was not proven to have a statistically significant effect on the accuracy of CDR-based activity spaces in general. This is in contrast with previous findings whereby they discovered apparent seasonality in the activity space (Järv et al., 2014; Schönfelder &

Axhausen, 2016). This is because people have a predictable lifestyle whereby, they visit places of interest regularly regardless of the month. Moreover, in terms of the overall outlook on the temporal analysis, it does not show any significant differences in the factor of holidays regarding the accuracy of CDR-based measurements. The results could be related to the limited number of holidays data.

Third, the role of the number of CDRs over defined categorized CDR groups in the accuracy of CDR-based measurements was investigated. There is a positive association between the number of CDRs and the accuracy of CDR-based activity space for all indicators. This means that the accuracy of the measurement increases as individuals use their phones more frequently, but its correlations were not statistically significant. This could be attributed to data wrangling regarding eliminating CDR location points having less than four per day. The correlation may have been higher assuming all CDRs were included. In the regression analysis, the number of CDRs was proven not to have a statistically significant effect on the accuracy of the measurements. This result is contrary to previous studies (Hoteit et al., 2016; Zhao et al., 2016; Chen et al., 2018). They argue that spatial errors in individual movement patterns are lower for people having more frequent CDRs.

Fourth, the role of socio-demographic factors in the accuracy of CDR-based measurements was explored. Overall, none of the factors was proven to be significant to influence the accuracy of CDR-based activity spaces. This could be attributed to a reduced amount of data and a small number of people in different categories. Also, there could be factors that can impact the accuracy of the CDR-based measurements but were unavailable to be used for this study. On the other hand, the activity space for a longer period tends to have a higher absolute difference probably due to absolute numbers being large for a longer period.

Lastly, we conclude the accuracy of CDR-based measurements depends on the type of activity space indicators applied as some indicators give higher accuracy of the measurements in comparison with others. Additionally, the accuracy of CDR-based measurements across different temporal scales is significantly influenced by only days of the week. On the other hand, personal characteristics are not considered important factors to explain the variability in the accuracy of CDR-based measurements.

Due to the encountered limitation in the process of research completed in this study, there is still room for improvement in further studies. Data was drastically reduced in order to make it

applicable to the scope of this research in terms of the activity space indicators. This could be resolved when we consider calculating individual activity spaces over a longer period such as weekly activity spaces or monthly activity spaces. Besides, the minimum value of the area of the mobile antenna could be taken to calculate the activity space if the person has been stationary so that there are more days with activity space values in the analysis.

Additionally, more socio-demographic factors like ethnicity, location of residential housing, etc. can be employed as a basis for assessing the accuracy of CDR-based measurements in the future. It would be possible to show a clearer causal connection between personal characteristics and the accuracy of CDR-based activity space.

Furthermore, it should be careful to generalize the findings of this study due to the limited scope of this research regarding the sample size and study area. Further research could focus on different countries, especially, the ones that are in Asia or America, to see if those countries have the same effect as in Estonia. It would give more reliable results when many countries are involved.

Conclusions

Many researchers in human mobility focus on measuring activity spaces using mobile positioning data due to its advantages compared to GPS data (Ahas et al., 2010; Järv et al., 2015; Xu et al., 2016; Vanhoof, Reis, et al., 2018). However, there are inadequate studies that investigated how CDR and GPS-based activity spaces are related and the accuracy of CDR-based measurements in comparison with GPS-based measurements. This is why the research explored the relationship between CDR and GPS-based activity spaces and investigated factors affecting the accuracy of CDR-based measurements regarding temporal factors and personal characteristics.

The researcher employed the six major activity space indicators such as MCP, ellipse, gyration, kernel density, distance, and entropy to determine the individual activity space on a daily basis and also the absolute difference method was used to evaluate the accuracy of the CDR-based activity spaces in comparison with GPS-based activity spaces. Spearman's rank correlation and linear mixed models were adopted for the statistical analysis. These two techniques were applied to find solutions for the research questions in this study.

It can be seen that both CDR and GPS-based measurements are positively correlated and have similar skewed distributions for all activity space indicators. In comparison to other indicators of CDR-based measurements, gyration and entropy were more closely related to GPS-based measurements. Kernel density has the lowest accuracy and large deviations based on the absolute difference between CDR and GPS-based measurements. Thus, the accuracy of CDR-based measurements depends on the type of activity space indicators applied as some indicators give higher accuracy of the measurements in comparison with others.

The temporal factor of days of the week in relation to the accuracy of CDR-based activity spaces proved that the accuracy of the measurements is generally higher during the weekdays compared to the weekends regarding all activity space indicators. However, other temporal factors such as months of the year and holidays have no statistically significant effect on the accuracy of the measurements. Thus, it can be concluded that only days of the week factor is considered a significant factor to explain the variability in the accuracy of CDR-based measurements in terms of the temporal scales.

The number of CDRs is negatively associated with the accuracy of CDR-based measurements, but its correlations are not statistically significant enough. The statistical analysis shows that the effect of the number of CDRs on the accuracy of CDR-based measurements is not statistically significant as well. It is therefore conclusive to say that the number of CDRs has no significant impact on the accuracy of CDR-based activity spaces.

Socio-demographic factors like gender, age, marital status, occupation, the family size can be concluded that they have no significant impact on the accuracy of CDR-based measurements. Therefore, the accuracy of CDR-based activity spaces does not vary among different socio-demographic factors in this study.

Kokkuvõte

CDR- ja GPS-andmete võrdlus inimeste tegevusruumi hindamiseks

Inimeste tegevusruumi uurimine võib aidata kaasa inimeste käitumise mõistmisele ning on olnud inimgeograafia fookuses juba palju aastaid (Xu et al., 2016). Tegevusruum on geograafilise terminina laialdaselt kasutusel, kirjeldamaks peamisi kohti, kus inimesed igapäevaselt oma asju ajavad (Gong et al., 2020).

Inimeste liikumise uurimisele on kasuks tulnud info- ja kommunikatsioonitehnoloogia areng, mis võimaldab koguda andmeid suure hulga inimeste liikumiste kohta ning täpse asukohaga. Seetõttu on CDR- ja GPS-andmeid inimeste tegevusruumi uurimiseks laialdaselt kasutatud (Richardson et al., 2013; Amini et al., 2014; Dobra et al., 2015; Williams et al., 2015; Xu et al., 2016; Vanhoof, Reis, et al., 2018). Siiski on ebapiisavalt uurimistöid, mis käsitleksid kuidas on CDR- ja GPS-andmete põhised näitajad seotud ning millised tegurid mõjutavad CDR-andmete põhisete tegevusruumi näitajate täpsust.

Selle uurimistöö eesmärk on hinnata CDR-andmete täpsust tegevusruumi hindamisel. CDR-andmete täpsust on hinnatud erinevate ajaühikute lõikes ning kõnetoimingute arvust ja inimeste sotsiaal-demograafilistest tunnustest lähtuvalt.

Mõlemad andmestikud (CDR ja GPS) on saadud Tartu Ülikooli Mobiilsusuuringute laborist. Uurimistöös kasutatud andmed on kogutud ajavahemikus 5. september 2013 kuni 10. märts 2015. Uuringus on kasutatud 52 inimese andmeid kokku 8961 inimpäeva. Analüüsis on kasutatud 477 inimpäeva andmeid, mis võimaldasid võrrelda CDR-andmete täpsust kõigi valitud tegevusruumi näitajate puhul.

Inimeste igapäevast ruumilist käitumist uuriti järgmise kuue tegevusruumi näitaja alusel: (1) minimaalne kumer polügoon (minimum convex polygon (MCP)), (2) ellips (*ellipse*), (3) ringi raadiuse ala (*radius of gyration*), (4) kerneli tihedus (*kernel density*), (5) vahemaa (*distance*) ja (6) entroopia (*entropy*). CDR-andmete põhiste tegevusruumi näitajate täpsuse hindamiseks on kasutatud kirjeldavat statistikat, korrelatsioonanalüüsi (Spearmani astakkorrelatsiooni), et hinnata CDR- ja GPS-andmete põhiste tegevusruumi näitajate vahelise seose tugevust ning kombineeritud lineaarseid mudeleid, et leida, kuidas mõjutavad CDR-andmete põhiste näitajate täpsust ajalised ja inimeste tunnused.

Analüüsi tulemused näitavad, et CDR- ja GPS-andmete põhised näitajad on positiivses korrelatsioonis ning neil on sarnane asümmeetriline jaotus kõigi tegevusruumi näitajate lõikes. Ringi raadiuse ala (*radius of gyration*) ja entroopia (*entropy*) näitajate puhul on CDR-andmete põhised näitajad sarnasemad GPS-andmete põhiste tegevusruumi näitajatega. Kerneli tiheduse (*kernel density*) puhul on CDR-andmete põhiste tegevusruumi näitajate täpsus absoluutarvuliste näitajate järgi kõige madalam. Statistiline analüüs näitas, et CDR-andmete põhiste tegevusruumi näitajate täpsust mõjutavad üksnes nädalapäevad. Lisaks sellele selgus, et kõnetoimingute arv ei mõjuta oluliselt CDR-andmete põhiseid tegevusruumi näitajaid, kui kõnetoimingute arv on üle nelja (st tegevusruumi näitajaid saab arvutada). Ükski analüüsitud sotsiaal-demograafiline tunnus CDR-andmete põhist tegevusruumi täpsust ei mõjuta.

Acknowledgements

I firstly would like to thank the Mobility Lab of the University of Tartu for providing the research data to me. The research would not be conducted without the relevant data. Additionally, I would like to thank my supervisor, Dr. Siiri Silm, for her brilliant guidance and insights through this research. Her feedback is very valuable and helpful while working on my thesis. Finally, I am grateful to my family, especially my parents. They have supported me all these years, through good times and bad times.

References

Abeele, M. Vanden, Beullens, K., & Roe, K. (2016). Measuring mobile phone use: Gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use . *Mobile Media & Communication*, *1*(2), 213–236. https://doi.org/10.1177/2050157913477095

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*, *29*, 469–486. https://doi.org/10.1016/j.tourman.2007.05.014

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), 3–27. https://doi.org/10.1080/10630731003597306

Amini, A., Kung, K., Kang, C., Sobolevsky, S., & Ratti, C. (2014). The impact of social segregation on human mobility in developing and industrialized regions. *EPJ Data Science*, *3*(1), 6. https://doi.org/10.1140/epjds31

Askitas, N., Tatsiramos, K., & Verheyden, B. (2020). *Lockdown Strategies, Mobility Patterns and COVID-19*.

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(20)30553-3

Bajracharya, A. R., & Shrestha, S. (2017). Analyzing Influence of Socio-Demographic Factors on Travel Behavior of Employees, A Case Study of Kathmandu Metropolitan City, Nepal. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, *6*(7), 111–119. www.ijstr.org

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human Mobility: Models and Applications. *Physics Reports*, *734*, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R., & Giovannini, L. (2010). Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *Journal of Statistical Mechanics: Theory and Experiment*, *2010*(05), P05001. https://doi.org/10.1088/1742-5468/2010/05/P05001

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), e1001083. https://doi.org/10.1371/journal.pmed.1001083

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data. *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, 427–434. https://doi.org/10.1145/2663204.2663254

Brown, L. A., & Moore, E. G. (1970). The Intra-Urban Migration Process: A Perspective. *Geografiska Annaler. Series B, Human Geography*, *52*(1), 1–13. https://doi.org/10.2307/490436

Brunsdon, C., & Singleton, A. D. (2015). *Geocomputation : a practical primer* (1st ed.). SAGE Publications Ltd.

Bryant, P., & Elofsson, A. (2020). Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. *MedRxiv*, 2020.04.13.20063644. https://doi.org/10.1101/2020.04.13.20063644

Buliung, Ron N., & Kanaroglou, P. S. (2006). Urban form and household activity-travel behavior. *Growth and Change*, *37*(2), 172–199. https://doi.org/10.1111/j.1468-2257.2006.00314.x

Buliung, Ronald N. (2001). Spatiotemporal patterns of employment and non-work activities in Portland, Oregon. *2001 ESRI International User Conference*. http://proceedings.esri.com/library/userconf/proc01/professional/papers/pap1078/p1078.htm

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*, *11*(2), 78–97. https://doi.org/10.1080/17489725.2017.1333638

Cabric, M. (2017). Security Inventions. In *From Corporate Security to Commercial Force* (1st ed., pp. 187–199). Elsevier. https://doi.org/10.1016/B978-0-12-805149-8.00023-6

Calenge, C. (2006). The package adehabitat for the R software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, *197*, 1035. https://cran.r-project.org/web/packages/adehabitatHR/index.html

Chaix, B., Kestens, Y., Perchoux, C., Karusisi, N., Merlo, J., & Labadi, K. (2012). An Interactive Mapping Tool to Assess Individual Mobility Patterns in Neighborhood Studies. *American Journal of Preventive Medicine*, *43*(4), 440–450. https://doi.org/10.1016/j.amepre.2012.06.026

Chatterjee, S., & Hadi, A. S. (2012). *Regression Analysis by Example*. John Wiley & Sons. https://www.wiley.com/en-us/Regression+Analysis+by+Example%2C+5th+Edition-p-9780470905845

Chen, G., Hoteit, S., Carneiro Viana, A., Fiore, M., & Sarraute, C. (2018). Enriching sparse mobility information in Call Detail Records. *Computer Communications*, *122*, 44–58. https://doi.org/10.1016/j.comcom.2018.03.012ï

Chen, Y.-C., & Dobra, A. (2018). *Measuring Human Activity Spaces from GPS Data with Density Ranking and Summary Curves*. https://arxiv.org/pdf/1708.05017.pdf

Comito, C., Falcone, D., & Talia, D. (2016). Mining human mobility patterns from social geo-tagged data. *Pervasive and Mobile Computing*, *33*, 91–107. https://doi.org/10.1016/j.pmcj.2016.06.005

Cools, M., Moons, E., & Wets, G. (2009). Investigating the Variability in Daily Traffic Counts through use of ARIMAX and SARIMAX Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2136*(1), 57–66. https://doi.org/10.3141/2136-07

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Affairs*, *39*(7), 1237–1246. https://doi.org/10.1377/hlthaff.2020.00608

Dargay, J. M., & Clark, S. (2012). The determinants of long distance travel in Great Britain. *Transportation Research Part A: Policy and Practice*, *46*(3), 576–587. https://doi.org/10.1016/j.tra.2011.11.016

Davidov, E. (2007). Explaining Habits in a New Context the Case of Travel-Mode Choice. *Rationality and Society*, *19*(3), 315–334. https://doi.org/10.1177/1043463107077392

Dobra, A., Williams, N. E., & Eagle, N. (2015). Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data. *PLOS ONE*, *10*(3), e0120449. https://doi.org/10.1371/journal.pone.0120449

Dodge, Y. (2008). *The concise encyclopedia of statistics*. Springer.

Dong, Y., Pinelli, F., Gkoufas, Y., Nabi, Z., Calabrese, F., & Chawla, N. V. (2015). Inferring Unusual Crowd Events from Mobile Phone Call Detail Records. *In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015*, *9285*, 474–492. https://doi.org/10.1007/978-3-319-23525-7_29

Estonian Government Office. (2018). *National, public and school holidays | Eesti.ee*. Ministry of Education and Research. https://www.eesti.ee/en/republic-of-estonia/republic-of-estonia/national-public-and-school-holidays/

Fan, Y., & Khattak, A. J. (2008). Urban Form, Individual Spatial Footprints, and Travel. *Transportation Research Record: Journal of the Transportation Research Board*, *2082*(1), 98–106. https://doi.org/10.3141/2082-12

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior : a geographic perspective*. Guilford Press. https://www.guilford.com/books/Spatial-Behavior/Golledge-Stimson/9781572300507/reviews

Gong, L., Jin, M., Liu, Q., Gong, Y., & Liu, Y. (2020). Identifying Urban Residents' Activity Space at Multiple Geographic Scales Using Mobile Phone Data. *ISPRS International Journal of Geo-Information*, *9*(4), 241. https://doi.org/10.3390/ijgi9040241

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

Gram, M. (2005). Family Holidays. A Qualitative Analysis of Family Holiday Experiences. *Scandinavian Journal of Hospitality and Tourism*, *5*(1), 2–22. https://doi.org/10.1080/15022250510014255

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *2018*(5). https://doi.org/10.7717/peerj.4794

Hirsch, J. A., Winters, M., Clarke, P., & McKay, H. (2014). Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis. *International Journal of Health Geographics*, *13*(1), 51. https://doi.org/10.1186/1476-072X-13-51

Holcomb, Z. C. (2016). *Fundamentals of descriptive statistics* (1st ed.). Routledge.

Horton, F. E., & Reynolds, D. R. (1971). Effects of Urban Spatial Structure on Individual Behavior. *Economic Geography*, *47*(1), 36–48. https://doi.org/10.2307/143224

Hoteit, S., Chen, G., Carneiro Viana, A., Fiore, M., & Viana, A. (2016). Filling the Gaps: On the Completion of Sparse Call Detail Records for Mobility Analysis. In *Proceedings of the Eleventh ACM Workshop on Challenged Networks*. https://hal.inria.fr/hal-01448821

Iovan, C., Olteanu-Raimond, A.-M., Couronné, T., & Smoreda, Z. (2013). Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies. In D. Vandenbroucke, B. Bucher, & J. Crompvoets (Eds.), *Geographic Information Science at the Heart of Europe* (pp. 247–265). Springer. https://doi.org/10.1007/978-3-319-00615-4_14

ITU. (2015). *ICT Facts and Figures 2015*. International Telecommunication Union (ITU). https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf

ITU. (2018). *Statistics*. International Telecommunication Union (ITU). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

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*, *38*, 122–135. https://doi.org/10.1016/j.trc.2013.11.003

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), 2680–2698. https://doi.org/10.1177/0042098014550459

Jones, M., & Pebley, A. R. (2014). Redefining Neighborhoods Using Common Destinations: Social Characteristics of Activity Spaces and Home Census Tracts Compared. *Demography*, *51*(3), 727–752. https://doi.org/10.1007/s13524-014-0283-z

Kamruzzaman, M., & Hine, J. (2012). Analysis of rural activity spaces and transport disadvantage using a multi-method approach. *Transport Policy*, *19*(1), 105–120. https://doi.org/10.1016/j.tranpol.2011.09.007

Kang, Y., Cho, N., & Son, S. (2018). Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. *PLOS ONE*, *13*(5), e0196845. https://doi.org/10.1371/journal.pone.0196845

Kim, J., Park, J., & Lee, W. (2018). Why do people move? Enhancing human mobility prediction using local functions based on public records and SNS data. *PLoS ONE*, *13*(2), e0192698. https://doi.org/10.1371/journal.pone.0192698

Kim, S., & Ulfarsson, G. F. (2015). Activity Space of Older and Working-Age Adults in the Puget Sound Region, Washington. *Transportation Research Record: Journal of the Transportation Research Board*, *2494*(1), 37–44. https://doi.org/10.3141/2494-05

Klinger, T., & Lanzendorf, M. (2016). Moving between mobility cultures: what affects the travel behavior of new residents? *Transportation*, *43*(2), 243–271. https://doi.org/10.1007/s11116-014-9574-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*, *8*(1–6), 185–203. https://doi.org/10.1016/S0968-090X(00)00017-6

Kwan, M.-P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*, *102*(5), 958–968. https://doi.org/10.1080/00045608.2012.687349

Lee, J. H., Davis, A., Yoon, S. Y., & Goulias, K. G. (2016). *Activity Space Estimation with Longitudinal Observations of Social Media Data*. https://pdfs.semanticscholar.org/21b0/78aceac91c081d9cba1c5225105027e9a0f2.pdf

Lehman, A., O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2005). *JMP for Basic Univariate and Multivariate Statistics: A Step-by-step Guide*. SAS Institute.

Li, R., & Tong, D. (2016). Constructing human activity spaces: A new approach incorporating complex urban activity-travel. *Journal of Transport Geography*, *56*, 23–35. https://doi.org/10.1016/j.jtrangeo.2016.08.013

Lind, A., Hadachi, A., & Batrashev, O. (2017). A new approach for mobile positioning using the CDR data of cellular networks. *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 315–320. https://doi.org/10.1109/MTITS.2017.8005687

Lynch, K. (1960). *The image of the city*. MIT Press.

Masso, A., Silm, S., & Ahas, R. (2019). Generational differences in spatial mobility: A study with mobile phone data. *Population, Space and Place*, *25*(2), e2210. https://doi.org/10.1002/psp.2210

Matthews, S. A., & Yang, T.-C. (2013). Spatial Polygamy and Contextual Exposures (SPACEs). *American Behavioral Scientist*, *57*(8), 1057–1081. https://doi.org/10.1177/0002764213487345

Mazey, M. . (1981). The effect of a physio-political barrier upon urban activity space. *Ohio Journal of Science*, *81*(5–6), 212–217.

Mercado, R. G., Paez, A., Farber, S., Roorda, M. J., & Morency, C. (2012). Explaining transport mode use of low-income persons for journey to work in urban areas: a case study of Ontario and Quebec. *Transportmetrica*, *8*(3), 157–179. https://doi.org/10.1080/18128602.2010.539413

Mooney, S. J., Sheehan, D. M., Zulaika, G., Rundle, A. G., McGill, K., Behrooz, M. R., & Lovasi, G. S. (2016). Quantifying Distance Overestimation From Global Positioning System in Urban Spaces. *American Journal of Public Health*, *106*(4), 651–653. https://doi.org/10.2105/AJPH.2015.303036

Newsome, T. H., Walcott, W. A., & Smith, P. D. (1998). Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation*, *25*(4), 357–377. https://doi.org/10.1023/A:1005082827030

Oracle. (2011). *Absolute Difference*. Oracle. https://www.oracle.com/webfolder/technetwork/data-quality/edqhelp/Content/processor_library/matching/comparisons/absolute_difference.htm

Palmer, J. R. B., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E., & Li, K. (2013). New Approaches to Human Mobility: Using Mobile Phones for Demographic Research. *Demography*, *50*(3), 1105–1128. https://doi.org/10.1007/s13524-012-0175-z

Papinski, D., Scott, D. M., & Doherty, S. T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(4), 347–358. https://doi.org/10.1016/j.trf.2009.04.001

Pappalardo, L., & Simini, F. (2018). Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, *32*(3), 787–829. https://doi.org/10.1007/s10618-017-0548-4

Patterson, Z., & Farber, S. (2015). Potential Path Areas and Activity Spaces in Application: A Review. *Transport Reviews*, *35*(6), 679–700. https://doi.org/10.1080/01441647.2015.1042944

Paz-Soldan, V. A., Reiner, R. C., Morrison, A. C., Stoddard, S. T., Kitron, U., Scott, T. W., Elder, J. P., Halsey, E. S., Kochel, T. J., Astete, H., & Vazquez-Prokopec, G. M. (2014). Strengths and Weaknesses of Global Positioning System (GPS) Data-Loggers and Semi-structured Interviews for Capturing Fine-scale Human Mobility: Findings from Iquitos, Peru. *PLoS Neglected Tropical Diseases*, *8*(6), e2888. https://doi.org/10.1371/journal.pntd.0002888

Perchoux, C., Chaix, B., Cummins, S., & Kestens, Y. (2013). Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. *Health & Place*, *21*, 86–93. https://doi.org/10.1016/j.healthplace.2013.01.005

Pi, M., Jeong, S., Yeon, H., & Jang, Y. (2018). Visual Analysis of Taxi Trajectory Data using Information Entropy. *KIISE Trans. on Computing Practices*, *24*(9), 0476–0481.

Primerano, F., Taylor, M. A. P., Pitaksringkarn, L., & Tisato, P. (2008). Defining and understanding trip chaining behaviour. In *Transportation* (Vol. 35, Issue 1, pp. 55–72). https://doi.org/10.1007/s11116-007-9134-8

Puura, A., Silm, S., & Ahas, R. (2018). The Relationship between Social Networks and Spatial Mobility: A Mobile-Phone-Based Study in Estonia. *Journal of Urban Technology*, *25*(2), 7–25. https://doi.org/10.1080/10630732.2017.1406253

Relph, E. C. (1976). *Place and placelessness*. Pion. https://openlibrary.org/books/OL4954978M/Place_and_placelessness

Richardson, D. B., Volkow, N. D., Kwan, M.-P., Kaplan, R. M., Goodchild, M. F., & Croyle, R. T. (2013). Spatial turn in health research. *Science*, *339*(6126), 1390–1392. https://doi.org/10.1126/SCIENCE.1232257

Rubin, O., Mulder, C. H., & Bertolini, L. (2014). The determinants of mode choice for family visits - evidence from Dutch panel data. *Journal of Transport Geography*, *38*, 137–147. https://doi.org/10.1016/j.jtrangeo.2014.06.004

Schmitz, P., & Cooper, A. (2007, September). Using Mobile Phone Data Records to Determine Criminal Activity Space. *IQPC International GIS Crime Mapping Conference*. www.csir.co.za

Schönfelder, S., & Axhausen, K. W. (2003). Activity spaces: measures of social exclusion? *Transport Policy*, *10*(4), 273–286. https://doi.org/10.1016/j.tranpol.2003.07.002

Schönfelder, S., & Axhausen, K. W. (2004). On the Variability of Human Activity Spaces. In M. Koll-Schretzenmayr, M. Keiner, & G. Nussbaumer (Eds.), *The Real and Virtual Worlds of Spatial Planning* (pp. 237–262). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-10398-2_17

Schönfelder, S., & Axhausen, K. W. (2016). Urban Rhythms and Travel Behaviour. In *Routledge* (1st Editio). Routledge. https://doi.org/10.4324/9781315548715

Schüssler, N., & Axhausen, K. W. (2009). Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT). *Arbeitsberichte Verkehrs- Und Raumplanung*, *568*, 1–22. https://www.research-collection.ethz.ch/handle/20.500.11850/19956

Sharmeen, N., & Houston, D. (2019). Spatial Characteristics and Activity Space Pattern Analysis of Dhaka City, Bangladesh. *Urban Science*, *3*(1), 36. https://doi.org/10.3390/urbansci3010036

Shen, Y., Kwan, M.-P., & Chai, Y. (2013). Investigating commuting flexibility with GPS data and 3D geovisualization: a case study of Beijing, China. *Journal of Transport Geography*, *32*, 1–11. https://doi.org/10.1016/j.jtrangeo.2013.07.007

Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M., & Arcury, T. A. (2005). A suite of methods for representing activity space in a healthcare accessibility study. *International Journal of Health Geographics*, *4*(1), 24. https://doi.org/10.1186/1476-072X-4-24

Shoval, N., Wahl, H.-W., Auslander, G., Isaacson, M., Oswald, F., Edry, T., Landau, R., & Heinik, J. (2011). Use of the global positioning system to measure the out-of-home mobility of older adults with differing cognitive functioning. *Ageing and Society*, *31*(5), 849–869. https://doi.org/10.1017/S0144686X10001455

Siegel, A. F. (2012). Confidence Intervals. In *Practical Business Statistics* (6th ed., pp. 219–247). Elsevier. https://doi.org/10.1016/B978-0-12-385208-3.00009-2

Silm, S., & Ahas, R. (2014a). Ethnic Differences in Activity Spaces: A Study of Out-of-Home Nonemployment Activities with Mobile Phone Data. *Annals of the Association of American Geographers*, *104*(3), 542–559. https://doi.org/10.1080/00045608.2014.892362

Silm, S., & Ahas, R. (2014b). The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset. *Social Science Research*, *47*, 30–43. https://doi.org/10.1016/j.ssresearch.2014.03.011

Silm, S., Ahas, R., & Mooses, V. (2018). Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *Journal of Ethnic and Migration Studies*, *44*(11), 1797–1817. https://doi.org/10.1080/1369183X.2017.1400425

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.

Song, C., Qu, Z., Blumm, N., & Barabasi, A.-L. (2010). Limits of Predictability in Human Mobility. *Science*, *327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170

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

Stock, J. H., & Watson, M. W. (2014). *Introduction to Econometrics*. Pearson.

Tana, Kwan, M.-P., & Chai, Y. (2016). Urban form, car ownership and activity space in inner suburbs: A comparison between Beijing (China) and Chicago (United States). *Urban Studies*, *53*(9), 1784–1802. https://doi.org/10.1177/0042098015581123

Vadrevu, K. P., Ohara, T., & Justice, C. O. (Christopher O. . (2018). *Land-atmospheric research applications in South and Southeast Asia*. Springer. https://books.google.co.kr/books?id=d1FTDwAAQBAJ&pg=PA179&lpg=PA179&dq=search+radius+%3D+0.9&source=bl&ots=t82LlESOTj&sig=ACfU3U2Mj1Y3SkOqiAni9WXa91g4rXbxvA&hl=ko&sa=X&ved=2ahUKEwjmo6a8r7fjAhWCGaYKHaXhAo8Q6AEwDXoECAcQAQ#v=onepage&q=search radius %3D 0.9&f

Van Dongen, H. P. A., Olofsen, E., Dinges, D. F., & Maislin, G. (2004). Mixed-Model Regression Analysis and Dealing with Interindividual Differences. *Methods in Enzymology*, *384*, 139–171. https://doi.org/10.1016/S0076-6879(04)84010-2

Vanhoof, M., Reis, F., Smoreda, Z., & Ploetz, T. (2018). *Detecting home locations from CDR data: introducing spatial uncertainty to the state-of-the-art*. http://arxiv.org/abs/1808.06398

Vanhoof, M., Schoors, W., Van Rompaey, A., Ploetz, T., & Smoreda, Z. (2018). Comparing Regional Patterns of Individual Movement Using Corrected Mobility Entropy. *Journal of Urban Technology*, *25*(2), 27–61. https://doi.org/10.1080/10630732.2018.1450593

Vazquez-Prokopec, G. M., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Kochel, T. J., Scott, T. W., & Kitron, U. (2009). Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to dengue virus. *International Journal of Health Geographics*, *8*(1), 68. https://doi.org/10.1186/1476-072X-8-68

Vich, G., Marquet, O., & Miralles-Guasch, C. (2017). Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour. *The Geographical Journal*, *183*(4), 426–439. https://doi.org/10.1111/geoj.12220

Vokoun, J. C. (2003). Kernel Density Estimates of Linear Home Ranges for Stream Fishes: Advantages and Data Requirements. *North American Journal of Fisheries Management*, *23*(3), 1020–1029. https://doi.org/10.1577/M02-141

Wallendorf, M., & Arnould, E. J. (1991). "We Gather Together": Consumption Rituals of Thanksgiving Day. *Journal of Consumer Research*, *18*(1), 13. https://doi.org/10.1086/209237

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. Chapman & Hall.

Wang, J., Kong, X., Xia, F., & Sun, L. (2019). Urban Human Mobility: Data-Driven Modeling and Prediction. *ACM SIGKDD Explorations Newsletter*, *21*(1), 1–19. https://doi.org/10.1145/3331651.3331653

Weisstein, E. W. (2007). *Absolute Difference*. Wolfram MathWorld.

https://mathworld.wolfram.com/AbsoluteDifference.html

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models : a practical guide using statistical software*. Chapman & Hall/CRC.

Williams, N. E., Thomas, T. A., Dunbar, M., Eagle, N., & Dobra, A. (2015). Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data. *PLOS ONE*, *10*(7), e0133630. https://doi.org/10.1371/journal.pone.0133630

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. http://arxiv.org/abs/1308.5499

Wong, D. W. S., & Shaw, S. L. (2011). Measuring segregation: An activity space approach. *Journal of Geographical Systems*, *13*(2), 127–145. https://doi.org/10.1007/s10109-010-0112-x

Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and Social Media Data Analytics of Housing Prices in Shenzhen, China. *PLOS ONE*, *11*(10), e0164553. https://doi.org/10.1371/journal.pone.0164553

Xu, Y. (2015). Mobility and activity space: understanding human dynamics from mobile phone location data. *Doctoral Dissertations*. https://trace.tennessee.edu/utk_graddiss/3619

Xu, Y., Shaw, S. L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, *42*(4), 625–646. https://doi.org/10.1007/s11116-015-9597-y

Xu, Y., Shaw, S. L., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z., & Li, Q. (2016). Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, *106*(2), 489–502. https://doi.org/10.1080/00045608.2015.1120147

Yuan, Y., & Raubal, M. (2016). Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *International Journal of Geographical Information Science*, *30*(8), 1594–1621. https://doi.org/10.1080/13658816.2016.1143555

Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior - A case study of Harbin, China. *Computers, Environment and Urban Systems*, *36*, 118–130. https://doi.org/10.1016/j.compenvurbsys.2011.07.003

Zeng, W., Fu, C. W., Müller Arisona, S., Schubiger, S., Burkhard, R., & Ma, K. L. (2017). A visual analytics design for studying rhythm patterns from human daily movement data. *Visual Informatics*, *1*(2), 81–91. https://doi.org/10.1016/j.visinf.2017.07.001

Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J. E., Wegrzyn, L., Gibbs, K., Braunschweig, C., & Stokes, C. (2011). Activity space environment and dietary and physical activity behaviors: A pilot study. *Health and Place*, *17*(5), 1150–1161. https://doi.org/10.1016/j.healthplace.2011.05.001

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*, *30*(9), 1738–1762. https://doi.org/10.1080/13658816.2015.1137298

Non-exclusive licence to reproduce thesis and make thesis public

I, JeongHwan Choi,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

"Comparison of CDR and GPS data for estimating the individual activity space"

supervised by Dr. Siiri Silm

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

JeongHwan Choi

12.08.2020