Master Thesis in Geographical Information Science nr 87

Effect of distance measures and feature representations on distance-based accessibility measures

Jeremy Azzopardi

2018
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Jeremy Azzopardi (2018). Effect of distance measures and feature representations on distance-based accessibility measures

Master degree thesis, 30 credits in Geographical Information Sciences

Department of Physical Geography and Ecosystems Sciences, Lund University

Effect of distance measures and feature representations on distance-based accessibility measures

---

Jeremy Azzopardi
Master thesis, 30 credits, in Geographical Information Sciences

Supervisor:
Harry Lankreijer
Lund University

Acknowledgements

My thanks go to my wife, Amanda, who has encouraged me to keep going throughout all these years.

Thanks also to Professor Maria Attard who facilitated the use of recreational area data and for her useful advice.

Abstract

Distance-based accessibility measures are often built using vector representations of origin and destination features, and Euclidean or network-based distances. There are few comparisons of how the choice of feature representations and distance types affects results. Existing comparisons often use Spearman's rank correlation coefficients. This study seeks to understand the effect of using different types of distance and feature representation on accessibility measures by comparing accessibility measures using the Bland-Altman plot, which measures the agreement between two variables.

Accessibility measures for recreational areas in Malta's Grand Harbour Area (GHA) were calculated. Two distance-based measures were compared: distance to nearest recreational area (DNRA) and Nearest Recreational Area ID (NRAID), measured from all residential blocks within the study area. Each of these two measures was calculated using two different vector representations for destinations (the recreational areas), access points and internal geometric centroids, and three different ways of measuring distance, Euclidean, network and full-network distance (network distance plus distance from feature representation to network). The combinations were compared for each measure. DNRA results were compared using the Bland-Altman plot and NRAID results were compared using percentage overlap.

Analysis showed that NRAID and DNRA results are especially affected by the selection of Euclidean distances versus network and full-network distances, and to a lesser extent the selection between centroids and access points, especially when using network or full-network distances.

An important assumption in this study was that full-network distance and access points are the most realistic alternatives for their respective groups. They are also the most demanding in terms of data collection, pre-processing, computation and analysis.

From these results one can conclude that it is important that the type of data and distance measure used and their possible effects on distance-based results are taken into account.

Keywords: Geography, GIS, Bland-Altman, Euclidean distance, geographic accessibility, network distance, recreational area

Contents

List of abbreviations

AAEF – average of DNRA using Euclidean distance and full-network distance with RA access points

AAEN – average of DNRA using Euclidean and network distance with RA access points

AANF – average of DNRA using network distance and full-network distance with RA access points

ACEN – average of DNRA using Euclidean and network distance with RA centroids.

ACEF – average of DNRA using Euclidean and full-network distance with RA centroids.

ACNF – average of DNRA using network and full-network distance with RA centroids.

DAEF – difference between DNRA using Euclidean distance and full-network distance with RA access points

DAEN – difference between DNRA using Euclidean and network distance with RA access points

DANF – difference between DNRA using network distance and full-network distance with RA access points

DCEF – difference between DNRA using Euclidean and full-network distance with RA

DCNF – difference between DNRA using network and full-network distance with RA centroid

DECA – DNRA difference between Euclidean distance using centroids and access points

DECN – difference between DNRA using Euclidean and network distance with RA centroids.

DFCA – DNRA difference between full-network distance using centroids and access point

DNCA – DNRA difference between network distance using centroids and access points

DNRA – Distance to Nearest Recreational Area

EA – Enumeration Area

GHA – Grand Harbour Area

GIS – Geographic Information System

NRAID – Nearest Recreational Area ID

OD – Origin Destination

OSM – OpenStreetMap

PI – Proximity Indicator

RA – Recreational Area

RU – Residential Unit

SI – Simple distance Indicator

SL – Sealing Layer

UTM – Universal Transverse Mercator

1 Introduction

Spatial analysis is essential in research containing a geographic component, regardless of the topic. Studies on public health, accessibility, and physical planning often contain a geographic component requiring the use of spatial analysis to test the study's hypotheses. A common focus of such studies is to investigate mobility: how people move, where they go, and how far they can and are willing to take themselves in order to avail of specific (usually public) facilities. Examples of such facilities are health care facilities, recreational areas, and schools. The concepts of fairness, segregation and policy are often present in such studies (Comber et al., 2008; Herzele and Wiedemann, 2003; Yang et al., 2015).

The key measure in such studies is accessibility, 'the quality of being able to be reached or entered' (Oxford Dictionaries, 2017). Accessibility to facilities can be composed of many elements, including distance, quality of facilities, density of facilities, size, type of facility, number of persons served by each facility and so on.

One of the most investigated elements of accessibility is distance – "the smallest separation between two features" (ESRI, 2017d). Distance is a fundamental component of many measures of accessibility, and it is therefore important to consider how one calculates distance and how this affects one's results. In many studies producing some kind of accessibility measures, the methodology used to calculate distance is often given little attention. Rather, studies often follow previous studies' methodologies, without investigating the assumptions and impacts concomitant to those methodologies. In such cases it is difficult to ascertain the quality of the resulting measures. A critical analysis of assumptions regarding methods is addressed in this study and discussed in chapter 2.

A few studies do take a closer look at their selected methodology for calculating distance in accessibility measures. This is usually done by comparing results from one method of calculating distance using different elements – for instance by using different types of distance or by using different ways of representing the features used in analysis. Such comparisons are often made with the use of a correlation measure, such as Spearman's rank correlation coefficient (e.g. Apparicio et al., 2008; Higgs et al., 2012; La Rosa, 2014). However, making

use of correlation measures is often insufficient to give an informed comparison of the results at hand, leading to potentially erroneous or at least incomplete conclusions.

1.1 Aim

This study aims to compare the effect of the use of six different combinations of accessibility measures based on distance. Specifically, the differences yielded when calculating six combinations of two simple and common measures of access, where the combinations are the result of variations in how distance is measured. The two measures of access in question can be formulated as the following questions:

1. Which facility is nearest to a specific residential unit?
2. What is the shortest distance from a specific residential unit to the nearest facility?

These two measures of access, although similar, produce different results. The first one produces a discrete value (the identity of the nearest facility) which may or may not be affected by distance changes in how distance is measured. The second measure produces a continuous numeric value, distance, which underlies the first measure, and which is more easily affected by changes in accessibility measures. These two measures of accessibility are two different ways of looking at distance. Both are included to give a more complete picture of the effects the choice how to measure distance has on results.

In this case, the facilities investigated are recreational areas (RA), while the residential units (RU) are residential blocks. Urban parks and other similar recreational, green and open areas are today seen as a necessity in a sustainable urban environment (Hiu Ming, 2014). The study is applied to the main urban region of the island republic of Malta, and the mode of transportation being considered is walking, since local recreational areas are often accessed this way in Malta (see Dai, 2011). The study could be repeated for different modes of transportation. Figure 1 gives two typical examples of what RU and RA look like in Malta - from street-level (fig. 1a and 1b), in aerial view (fig. 1c and 1d), and as feature abstractions (fig. 1e and 1f).

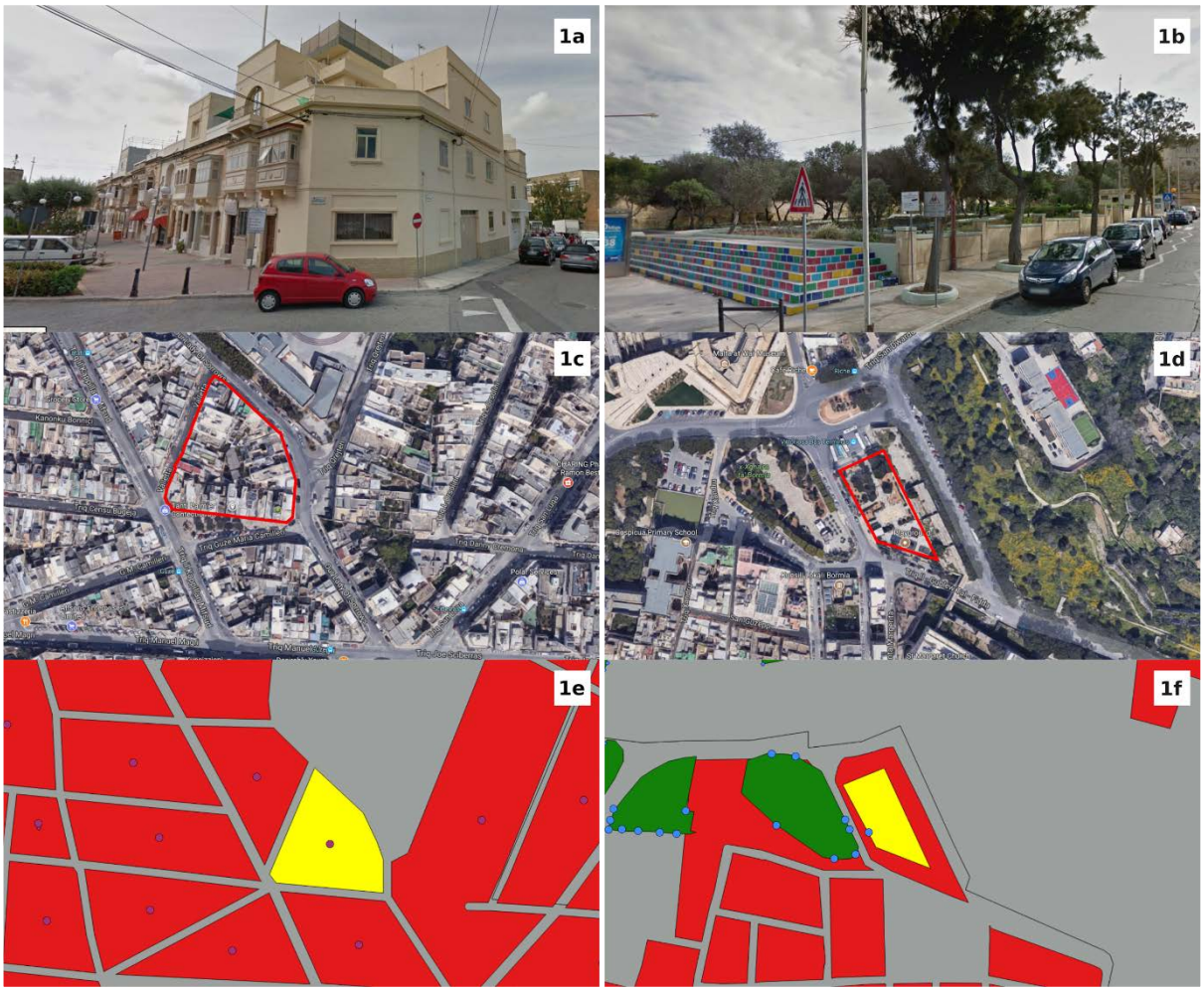


Figure 1 – Comparison of features photographed at street level, viewed from aerial imagery, and viewed as digitised polygons.

Key:

1a – Residential block as seen from street level.

1b – Recreational area as seen from street level.

1c – Residential block from aerial view.

1d – Recreational area from aerial view.

1e – Residential block as a digitized polygon (yellow) among other residential blocks (red) with geometric centroids (purple).

1f – Recreational area as a digitized polygon (yellow) among other residential blocks (red) with geometric centroids (purple).

The two measures of accessibility outlined above can be reformulated as:

1. Nearest recreational area
2. Distance to nearest recreational area

These reformulated measures can be used as building blocks to generate other measures, such as the mean distance to all available facilities and the number of facilities within a given distance (Apparicio et al., 2008).

Some of the elements used in these two measures will be changed and the results compared in order to explore how common choices of elements used in the production of the above-mentioned measures (both based on distance) can affect results. The elements tested are three types of distance:

- Euclidean distance – the straight-line distance between two points
- Network distance – the distance over a road network between two points
- Full-network distance - the distance over a road between two points added to the distance of the points from the network itself.

, and two types of feature representations:

- Geometric centroid – The geometric centre of a feature object
- Access point – A point representing an entry/exit point of a feature such as a building or public space

The choice of elements is discussed in the methodology section.

The main question of this study is therefore: How does the choice of distance measure and destination feature representation influence results when calculating two basic accessibility measures – the nearest RA and distance to the nearest RA?

In practice, each RA has its own identification number (or RA ID), and finding the nearest RA involved finding the nearest RA ID (NRAID). The distance to the NRAID for an RU is found in conjunction with the distance to the nearest RA (DNRA).

The question can be subdivided into several parts:

1) How does the use of different feature representations affect NRAID results?

2) How does the use of Euclidean distances, network-only and full-network distances affect NRAID results?

3) How does the use of different feature representations affect DNRA?

4) How does the use of Euclidean distances, network-only and full-network distances affect DNRA?

2 Background

This chapter gives a review of the literature used in the development of this study and will examine key concepts used to inform work. Using available literature, it will proceed to compare used definitions and methods and to determine the definitions for this particular study.

2.1 Studies featuring accessibility

Studies featuring accessibility which were reviewed in this study fall within the fields of urban studies (see Cetin, 2015; Comber et al., 2008; Curar, 2014; Curl et al., 2015; Herzele and Wiedemann, 2003; Higgs et al., 2015; La Rosa, 2014; Zhou an Kim, 2013), geography (see Boone et al., 2009), and public health (see Apparicio et al., 2008; Barr et al., 2016; Bauer et al., 2015; Dai, 2011). Accessibility also features in local plans (Malta Planning and Environment Authority, 2006; Carter, 2010). In most cases articles tend to have different approaches to accessibility, due to coming from different fields, having different purposes, and addressing different urban contexts. This means accessibility was defined in different ways, given different levels of importance, and explored using different methods. The most common definition of accessibility is as accessibility based on physical distance (Morar et al., 2016). Several studies only use accessibility based on physical distance, calling it 'spatial accessibility' (Barr et al., 2016; Ikram et al., 2015). Other approaches to defining accessibility also take into account non-spatial factors such as attractiveness, destination type, and neighbourhood population when calculating accessibility (Meng and Malczewski, 2015; Yang et al., 2015).

Since this study focuses on spatial accessibility, the studies consulted had spatial accessibility as their focus, or at least as a component.

2.2 Definitions of accessibility

A main element of many reviewed studies is an exploration of accessibility, specifically spatial accessibility. Although different studies make use of different definitions, the definition used often has a strong relationship to the methodology within a study. Some studies do not define accessibility at all (Baycan-Levent et al., 2009; Cetin, 2015) - this is due to either lack of focus on methodology (Cetin, 2015) or a high-level, multi-criteria evaluation which uses a large

number of factors (Baycan-Levent et al., 2009). In some cases, studies only had working definitions of accessibility, usually taken from third-party categorizations or tools (Oh and Jeong, 2007; Comber et al., 2008). Most studies have, however, a more detailed definition of accessibility (Giles-Corti et al., 2005; Apparicio et al., 2008; La Rosa, 2014; Morar et al., 2014; Meng and Malczewski, 2015; Yang et al., 2015).

In all studies reviewed, whether an explicit definition of accessibility was available or not, the term 'access' or 'accessibility' was used.

Definitions of accessibility range from basic working definitions, such as "Service area is the service range of a public facility which is equivalent to the accessibility to a public facility such as a park or school that supplies service via traffic networks" (Oh and Jeong, 2007: 26), to broader, multifaceted descriptions incorporating a multidimensional aspect to the concept of accessibility, describing accessibility in "terms of affordability, acceptability, availability and spatial accessibility…" (Apparicio et al., 2008) Apparicio et al. are not the only ones taking a multidimensional approach to accessibility - Morar et al. list monetary cost as being a deciding factor in accessibility (Morar et al., 2014), while Yang et al. and Meng and Malczewski list park-type and quality as deciding factors in accessibility. That said, distance – the spatial aspect of accessibility, is a deciding factor (La Rosa, 2014) since it imposes a physical limit on what can and cannot be reached (Morar et al., 2014).

Accessibility is measured in different ways, based on the aims of the study and the definitions used (themselves often partly relying on the aims of the study). Thus, one finds that studies not defining accessibility have a simplistic approach in their methodology – Baycan-Levent et al. (2009) only use simple statistical indicators (area per population, for instance) while Cetin uses only Euclidean distance (buffer zones) from RU's to green spaces as a measure of accessibility. As the aims start giving more prominence to the spatial aspect of accessibility, the method of analysis becomes more complex: studies start using a mix of statistical measures (average green space area per person, percentage of park area in relation to total area and service area analyses (Giles-Corti et al., 2005; Baycan-Levent et al., 2009; M'Ikiugu et al., 2012). Most service area analyses measure network distance from RU's to facilities, e.g. green spaces (often, Euclidean distance is used for comparison purposes - e.g. in Apparicio et al., 2008). Simple statistical measures and visual presentations are used for an initial exploration of the material (Oh and Jeong, 2007; Apparicio, et al., 2008; Morar et al., 2014).

Some studies use a more detailed methodology, notably La Rosa, who proposes two kinds of accessibility indicators, namely the Simple distance Indicators (SI) and the Proximity Indicators (PI) (La Rosa, 2014). SIs are a simple measure of all the people who have access to a green space, while PIs take the distance involved into account. Apparicio et al., 2008 take a slightly different approach, comparing various types of distance, aggregation areas, and five different accessibility measures, namely: "1) the distance to the closest service; 2) the number of services within n metres or minutes; 3) the mean distance to all services; 4) the mean distance to *n* closest services; and 5) the gravity model" (Apparicio et al., 2008). That said, these different measures of accessibility are not themselves explored in detail.

Ultimately, these more complex measures still have distance as a fundamental component. Therefore, the methodology used for calculating accessibility measures based on distance, and therefore the calculation of distance via Geographic Information Systems (GIS) is the next step. It is interesting to note that the different measures mentioned require different methodological approaches – the application of GIS varies very strongly, from the calculation of areas to find the total available area of facilities within a city to the creation of weighted maps, from the use a simple buffer zone to a more complex network analysis to measure distance.

This jungle of definitions and methodologies with varying degrees of depth means that there is no standard definition of accessibility and no standard methodology to choose. Rather, the needs of this study as well as the nature of the available data and the local context need to be used to define an own set of definitions and methodologies, keeping previous studies as a reference.

2.3 Defining Recreational Areas

Literature covers many different types of facilities used for recreational purposes, and there is a lot of overlap between definitions, as will be seen below. Defining these different types of spaces depends on the study's aim, availability of data and the local context. It is also useful to make explicit that all studies on such spaces focus on urban areas.

2.3.1 Availability of base data

Available data which is useful for spatial analysis: data on recreational spaces, urban open spaces, green spaces, and similar facilities being investigated, as well as road networks (if

applicable) and RU's varies greatly between and even within different studies, coming from open data such as OpenStreetMaps and Urban Atlas (European Environmental Agency, 2010; Geofabrik, 2016; La Rosa, 2014; Morar et al., 2014) to data provided by local authorities or companies (Comber et al., 2008; La Rosa, 2014; Morar et al., 2014; Meng and Malczewski, 2015; Yang et al., 2015) to the datasets created or updated by the authors themselves (Baycan-Levent et al., 2009; La Rosa, 2014; Cetin, 2015). In some cases, some of the data sources were vague or were not provided (Oh and Jeong, 2007; Apparicio et al., 2008; Cetin, 2015).

The form and quality of available data is an important issue (this is also influenced by resource limitations – time and money). In one case the authors collected some of the necessary data themselves, with other data being obtained from pre-existing sources (Cetin, 2015). This meant that the data collected was well-suited to the aims of the study. Most authors obtained data from relevant authorities. This of course helped shape their definition of green spaces (Dai, 2011; Morar et al., 2014), although not always or completely. Dai for instance, obtained the green space data from the Atlanta Regional Commission, but then followed the definition of green spaces used in previous studies (Dai, 2011).

Although limitations posed by data are at times discussed, many studies seem to imply that there were no other options available – usable and affordable spatial data for street networks, population units, and green spaces does not seem to be uniformly available. This is a shortcoming which cannot be avoided unless a dataset is created from scratch or from pre-existing datasets – often not an affordable option.

2.3.2 Terminology and definitions used to describe recreational areas

The term Urban Open Spaces is often used within urban studies and public health to describe zones within an urban area whose functions are community service, recreational, and sporting (Carter, 2010). The term is used to refer to such spaces as parks, playgrounds and vegetated areas, as opposed to "streets, plazas and squares" (Hiu Ming, 2014).

Another commonly used term is ‘green spaces’. Green spaces are often described as being "open, vegetated and permeable" (Taylor, 2013). It is most common for only public spaces to be taken into consideration under this term, although some have included private open spaces in their definition (see The Urge Team, 2004 in Taylor, 2013).

Other terms are used, such as Urban Green spaces whose function include improving liveability, adding to social justice, boosting the economy by attracting more investment and increasing property prices, and also improving the environment (Baycan-Levent et al., 2009). Higgs et al. (2012) use the term parks and recreational spaces in their study, while La Rosa and Privitera (2012) mention non-urbanised areas.

Local authorities often mention green spaces in planning documents, using different terms and definitions. In Taylor's (2013) report we find that the city of Berlin's definition of green spaces includes the following subcategories:

- Green and recreational areas,
- Public green spaces (including parks, allotments, cemeteries, sports fields, etc.)
- Forests
- Agricultural Land
- Water

Dublin's City Development Plan 2011-2017 (in Taylor, 2013) includes in its definition of green space such features as:

- Parks
- Gardens
- Institutional Grounds
- Allotments
- Communal gardens
- Green corridors
- Natural and semi-natural areas

The Marseille municipal area distinguishes between green spaces, white (rocky) and blue (sea/water) spaces (Taylor, 2013: 8). In this case the definition follows land cover rather than land use as a primary descriptor. This classification also does not address the differences between land use and land cover in the urban core and in the outskirts of the municipality in a clear manner (Taylor, 2013: 8-9).

Here it is interesting to note how studies deal with developed and undeveloped green spaces. While most studies do not explicitly differentiate between the two, some do make a distinction between developed and undeveloped (Comber et al., 2008; Morar et al., 2014; Cetin, 2015), although methodologies to deal with such differences are limited. On the other end, La Rosa (2014) lumps all green spaces into one category (La Rosa, 2014). Another aspect is that most studies assess only public spaces, although some have included private open space in their definition (see The Urge Team 2004 in Taylor, 2013: 2).

One function common to most definitions encountered above is recreation. Indeed, several authors mention recreation when describing open, green and urban spaces (among many: Giles-Corti et al., 2005; Comber et al., 2008; Baycan-Levent et al., 2009; La Rosa and Privitera, 2013; Zhou and Kim, 2013).

It is clear from the many definitions mentioned above that a unified definition is lacking. This is because the definition of such facilities and what it includes depend a lot on the definer's needs and the local context. Therefore, it is necessary to provide a term and definition specific to the Maltese context. The selected term is 'recreational areas'.

2.3.3 A local definition

As seen above, the definition of a study's working term for what we will call recreational areas is influenced by the data available, local legislation and policy, the aim of the study, and the methodology used, among many factors. In many cases, the definition used within a study depends on available data and/or local legislation. In some cases, available data is modified to better fit the aim of the article (e.g. Oh and Jeong, 2007), although the majority seems to use what is available and then adjust the methodology and assumptions accordingly. The circular relationship between available data, aim of article and method used means that each study reviewed uses a different approach and yields different results even if very similar methodology is used.

Taking all the above into consideration, the most generalised definition of recreational spaces is defined for this study by the following:

"Recreational spaces can be defined as any urban open space whose main functions includes public recreation. These spaces need to be open, physically accessible, and well-maintained.

Their use as recreational spaces can be both their intended use or an unintended use which is still permitted in practice."

The term is motivated because the function of recreation is a common theme in all the facilities used in this study, which are taken from the Valletta 2018 Cultural Mapping project[1] dataset's categories 'playing fields' and 'managed public parks and recreational spaces'.

2.4 Methodologies and aims for measuring accessibility

Methodology used for measuring accessibility to public facilities such as RA's is strongly influenced by the aim of the study, as well as the data available. The studies reviewed can be placed in several general groups:

1) Studies on distribution of recreational areas in relation to local population:

    a. Within a single study area (Oh and Jeong, 2007; Apparicio et al., 2008; Comber et al., 2008; La Rosa, 2014; Morar et al., 2014; Cetin, 2015; Meng and Malczewski, 2015; Yang et al., 2015) – these studies usually use a range of techniques, ranging from statistical measures such as proportion of green spaces per 1000 inhabitants, to distance-weighted accessibility models, with a critique of the use of Euclidean distance and a preference for the use of network distance. One study (Comber et al., 2008) take different ethnicities and religious beliefs into account, others focus on social fairness and equity (Oh and Jeong, 2007; Yang et al., 2015).

    b. Comparative analyses of several areas (Baycan-Levent et al., 2009) – these studies tend to compare different cities by comparing statistical measures and specific features via multi-criteria analysis. Studies like Yang et al. and Morar et al. also include a comparative analysis of several cities using one or more factors as part of the discussion and as a tool to compare their results to other cities.

[1] The Valletta 2018 Cultural Mapping project aimed 'to catalogue all spaces across the Maltese islands which are used for cultural purposes, ranging from established cultural venues (such as museums, theatres and heritage sites) to public and open spaces.' (Culture Map Malta, 2017)

2) Studies on accessibility to RA's as a factor in public health – although most studies qualify their research questions by stating that RA's have a positive health impact, most studies do not include a health aspect in their methodology (Giles-Corti et al., 2005).

Subgroup 1a can be further divided into three further groups based on how accessibility is approached and analysed:

1) Analysis of spatial accessibility based on distance – this is where all studies mentioned in 1a can be found, except for those in subgroup ii below. Most studies do not take obstacles into account when network distance analysis is made, although some do, such as Yang et al. (2015) who, since they use network analysis with time as impedance add 30 seconds for each road-network node along routes. Mode of transport considered in the reviewed studies is either walking (e.g. Yang et al., 2015) or driving (e.g. Dai, 2011). See section 3.2.4 for more information on how distance is used in the analysis of spatial accessibility.
2) Analysis of accessibility using measures based on distance as well as park size (e.g. Meng and Malczewski, 2015)
3) Studies which explore distance-based accessibility but also focus on the indicators/measures used – a comparative analysis of different types of indicators or measures of accessibility, as well as variation in the measure/indicator elements such as distance and population aggregation methods (Apparicio et al., 2008; La Rosa, 2014). In the case of Apparicio et al. (2008) the study focuses on accessibility to health services rather than Green spaces. However, the study is deemed to be very useful to this literature review due to its methodological explorations.

Although methodology is, at a basic level, dictated by aim, there are several other factors which influence it. These include availability of resources and data, data quality, and technical limitations.

In most cases with moments of spatial analysis, three basic elements are required to measure spatial accessibility based on distance – points of origin, destinations and a way to measure the distance between the two (Apparicio et al., 2008; La Rosa, 2014). These basic elements are

used regardless of which methodology is used. Their use in this study is discussed more in the methodology section.

3 Methodology

The methodology used for this study is divided into four main steps:

1. Selection of study area
2. Data collection and pre-processing
3. Data processing
4. Data analysis

3.1 Definition of study area

The study area selected for this study is the largest conurbation in the island nation of Malta. Malta has suffered from over-development, which has led to a lack of access to recreational spaces within its urban fabric (See Appendix I). It also has a very dense road network, with concomitant traffic congestion problems.

Malta is a small island republic situated in the central Mediterranean, about 100 kilometres south of Sicily. The country comprises three inhabited islands: Malta, Gozo, and Comino, along with several minor uninhabited islands. The total land area of the archipelago is 316 square kilometres. (Central Intelligence Agency, 2017)

The Maltese islands have a high population density as well as a large percentage of artificial surfaces (see Table 1). This is especially true for the Grand Harbour Area (GHA), which is heavily urbanised (see Figure 2). Here, 72.0% of the total land area is artificial (including artificially made green and recreational areas) and 24.5% agricultural. There is very little land left for natural areas, and therefore the availability of green and recreational areas within that 96.5% of land area is very important and must be planned as part of the urban fabric.

The high population density within such a restricted area creates a high demand on available green and recreational areas. The term 'Recreational Areas' covers all open areas which are used for recreational purposes: Green spaces, non-green recreational areas, and playing fields. These definitions are very fuzzy and would warrant a research project by themselves. The term 'Recreational Areas' therefore covers all open areas (open to the public) which are used for recreational purposes, as seen in the previous chapter.

		Maltese Islands		Grand Harbour Area	
Code	**Corine Category 1**	**km^2**	**% of total area**	**km^2**	**% of total area**
1xx	Artificial Surfaces	90.877	28.776	35.737	71.981
2xx	Agricultural Surfaces	164.554	52.105	12.166	24.504
3xx	Forest and Semi-natural Areas	60.131	19.040	1.745	3.515
4xx	Wetlands	0.251	0.080	0	0
	Total	315.813	100	49.647	100

Table 1 – Percentage of different types of surfaces in the Maltese islands.

Summarised from CORINE Land Cover data 2012

The GHA's characteristics as Malta's largest urban area with a high percentage of artificial surfaces and very little natural or agricultural areas, where the main access to RA's is to public gardens, playing fields, and similar urban recreational areas, led to its selection as a study area (Figure 3). An overview of the presence and distribution of recreational areas within the selected study area as well as a comparison of the GHA's situation to other cities around the world is available in Appendix I (overview in Figure 4).

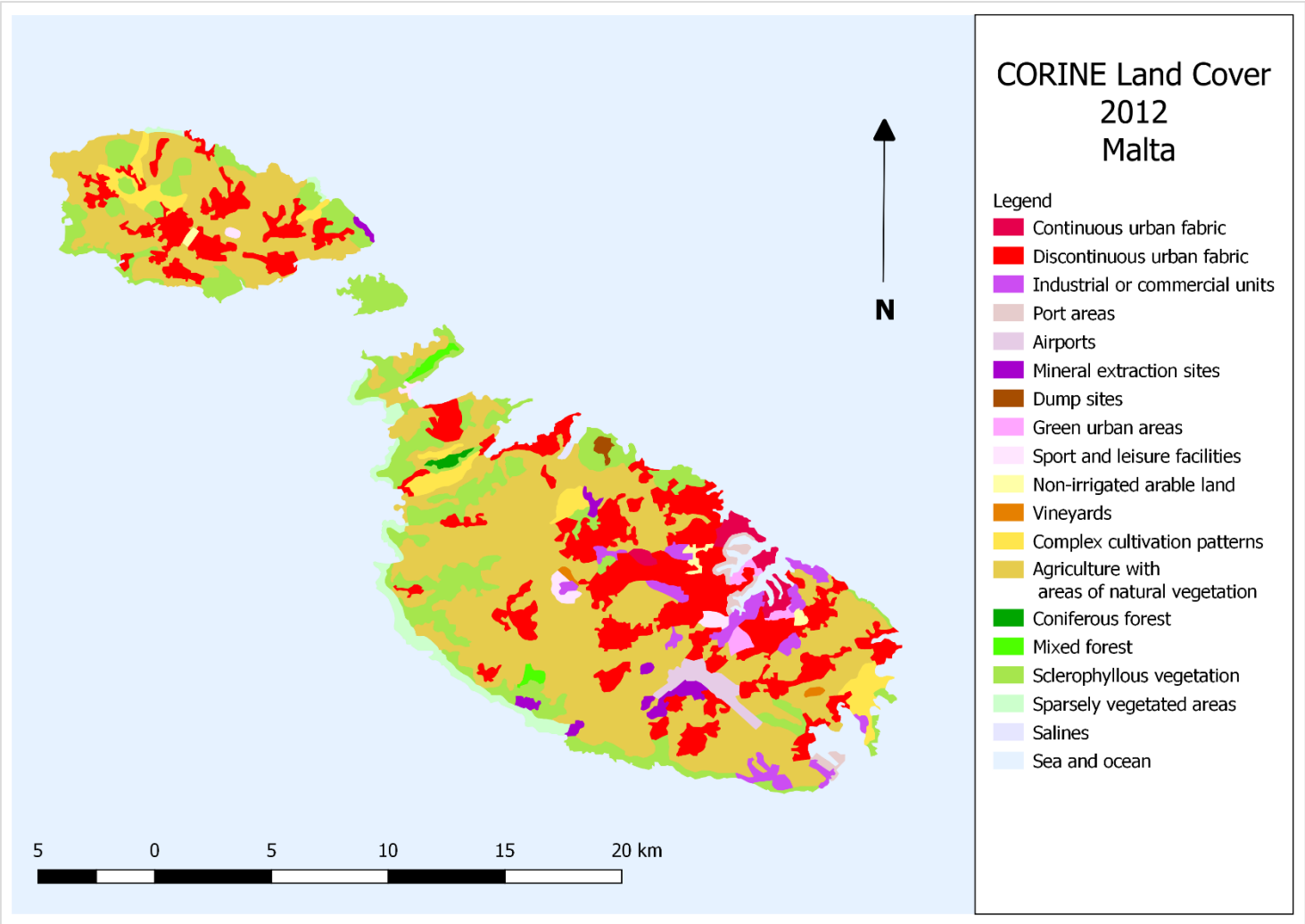


Figure 2 – CORINE Land Cover 2012 for the Maltese islands (MEPA, 2014)

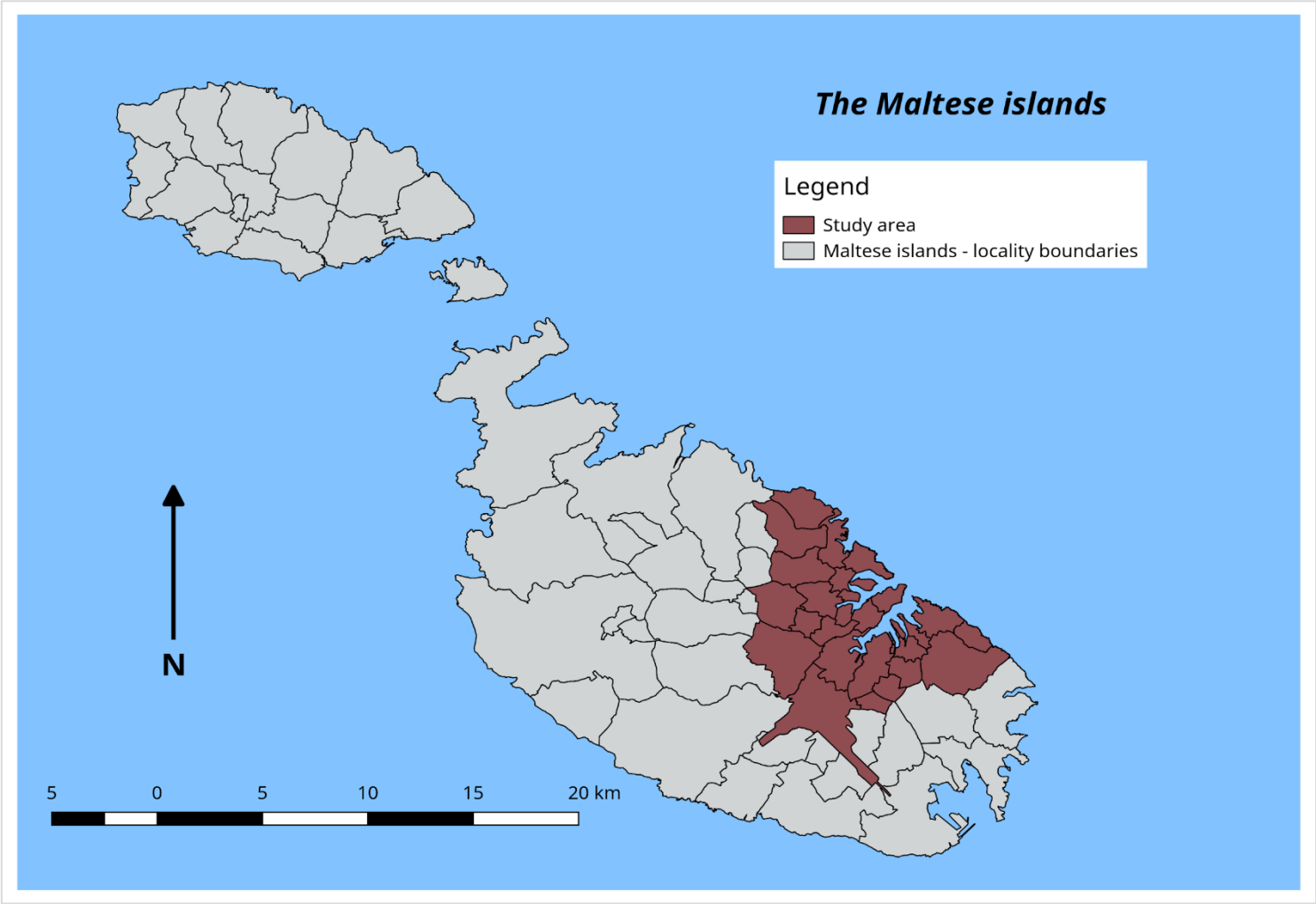


Figure 3 – The Maltese islands and the selected study area showing locality boundaries

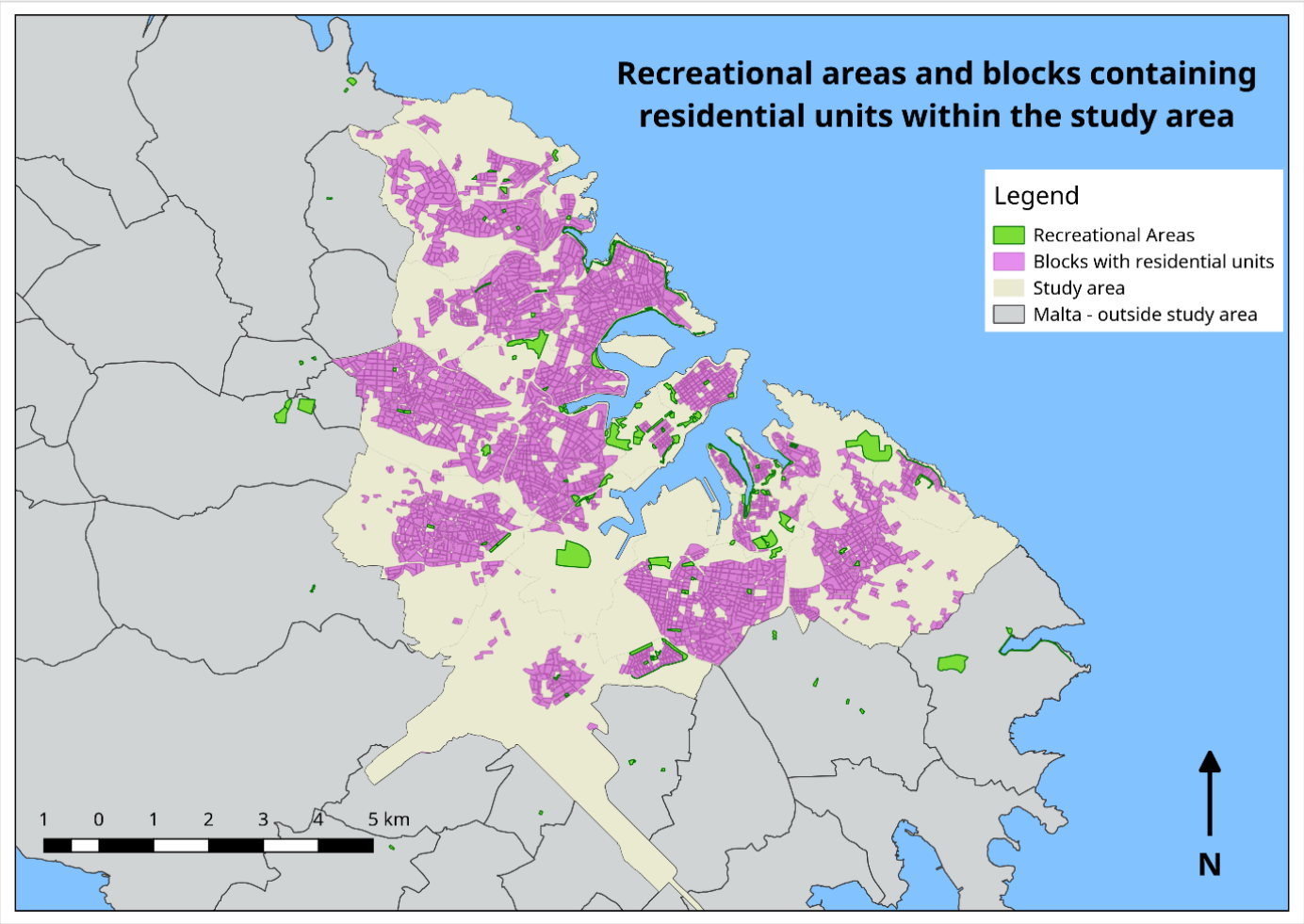


Figure 4 – Overview of the study area with recreational areas and blocks with residential units.

3.2 Calculating measures of accessibility

Before base data was collected and processed, it was important to define which measures of accessibility should be investigated and what data would be required to calculate them.

Two measures of accessibility were selected for their simplicity: ID of nearest recreational area to each residential block and the distance from each RU to the nearest recreational area. Both these measures have a single determining component: distance. In the case of the first measure, distance is relative – the nearest RA to each block is only so in relation to other RA's. It can still be very far away. In the second measure, distance is absolute and is measured directly.

Four elements are necessary for the calculation of measures of accessibility: i) a way of defining an RU as a unit; ii) a way of representing the distribution of a population within each RU; (iii) a "measure of accessibility", and; iv) a way of calculating distance between RU's and service features. (Apparicio et al., 2008)

A measure which is not mentioned by Apparicio et al. (2008) but which is as important as element iii is how destinations, in this case RA's, are represented. This will be included as the third element of a total of five.

3.2.1 First Element: Residential units

The European Environmental Agency's Urban Atlas dataset maps different urban areas within the European Union (EU), including data on "land use and land cover data for Large Urban Zones with more than 100.000 inhabitants […]" (European Environmental Agency, 2010)

The selection and extraction of relevant fields and features is mentioned in the Data pre-processing section.

3.2.2 Second Element: A way of representing population distribution within each residential unit

The Urban Atlas dataset divides data on buildings with RU's into blocks of buildings, but does not have information about distribution within each block. A typical solution is to take a centroid of each block as representing the population of the whole block. This assumes equal distribution of residences within each block. The problem with this solution is that blocks differ in size and therefore distance to the nearest road is influenced by the size of the building block.

Since residential blocks have access to the road, the use of regular points to represent a residential block boundary (base on the Urban Atlas dataset) was initially considered, with each polygon being represented by boundary points spaced 25 metres from each other. However, this solution creates data points from the false assumption that doorways (i.e. access points) in a residential block are evenly spaced – it does not take into account whether the residential block in question is an eight-storey apartment block with a single main door or a single-family villa. Although this method potentially reduces artificially large distances (created where the feature centroid is far away from the actual door or doors) between origin and destination points, it can instead create artificially short distances where there are more boundary points The advantage of this solution as a halfway point between using centroids as feature representation and using the location of each doorway and the number of RU's it leads to, is that it potentially reduces the distance from the representation of a residential block and the road network. A disadvantage is that artificially short distances can be created in the same way. Another disadvantage was that this approach required an unfeasible amount of computing resources. Therefore, it was decided that the only practical approach was to use internal centroids for residential blocks.

3.2.3 Third element: A way of representing the destination

When measuring distances, the destination is as necessary as the point of origin. How a destination is represented can influence the distance measured between it and other features. When using Euclidean measurements, measuring distance to a RA centroid versus distance to its entrance can give very different results, especially if the RA is large or has a non-uniform shape. When using a network to measure distance, the location of the destination point is also important, as it can influence routing.

For this study, two feature representations will be used and tested. These are internal geometric centroids and access points. Internal geometric centroids - as they are easily created and used to measure distance (see Apparicio et al., 2008; Higgs, Fry, and Langford, 2012; Ikram et al., 2015; La Rosa, 2014; Morar et al., 2014). Access points, that is entrances to recreational areas, will be the second destination representation used, as they are often the most realistic reflection of how people access RA's within the GHA – by walking into them through an entrance, be it a path, a gateway, or other. A third option was planned: boundary points, where a polygon is represented by its edge or by a large number of points approximating a continuous edge.

However, this model is unrealistic and unnecessary, since the vast majority of RA's in the GHA have boundaries, and the points where these boundaries can be traversed are already represented by access points. It is also extremely computationally demanding for network distances, although this particular hurdle can be cleared

3.2.4 Fourth Element: A Measure of Accessibility

There are many different measures of accessibility used by researchers in different studies. The vast majority of these are based on measuring distance. Some use simple measures based directly on distance, such as distance to nearest feature and identity of nearest feature (e.g. Apparicio et al., 2008; Higgs et al., 2012; Morar et al., 2014) or on time (walking speed * time, limiting results to a maximum walking distance (e.g. Higgs et al., 2015; Yang et al., 2015). Others use slightly more elaborate measures, taking into account more factors than distance, but which also include distance as a basic impedance measure. An example is as La Rosa (2014), who uses a measure of accessibility for green spaces where the population served by a green space is weighted by the distance to said population (La Rosa, 2014).

A basic building block is, therefore, distance. A simple and yet useful measure is distance to the nearest recreational area (DNRA) along with the concomitant ID of the nearest RA (NRAID). NRAID is therefore an element of the measure DNRA, even though for practical reasons both DNRA and NRAID are called 'measures' in this study.

DNRA is the shortest distance from a specific RU to the recreational area nearest to it. Which unit and type of distance are used and how it is measured is up to the user. The unit of measurement can be any unit which represents impedance: not only physical distance but also time. Thus, DNRA can be measured in metres, miles, seconds and so on. The type of distance used can vary from network distance, to Euclidean distance, to any other type of distance. The methodology of their measurement can again vary. The important thing is that in one study all above elements are kept constant.

NRAID is the element of DNRA which identifies the nearest RA 'hidden' in the DNRA. All RAs have a unique identifier, or ID. When the DNRA from a single RU is calculated, the identity of the nearest RA becomes the NRAID result for that specific RU. NRAID results must therefore use the same elements and methodology as those used for calculating DNRA if they are to make sense.

DNRA and NRAID are two measures which take into account several things. First of all DNRA is a simple measure which yields a meaningful result, unlike for instance mean distance to all or a set number of features (e.g. in Apparicio et al., 2008), where a larger number of results are bundled up together into a single figure. This measure can also be useful as an indicator of a minimum level of accessibility. Furthermore, the same method can yield distance to the second nearest RA and so on, and build more complex measures.

On the other hand, from an analytical point of view, knowing the nearest RA ID allows us to understand where the 'breaking point' is when comparing different methodologies. That is, when comparing two methods, in how many occasions does the ID to the nearest RA change, based on the methodology used? From a planning perspective, knowing the nearest RA allows for the identification of the best placed and most demanded RA's, and can be useful for measuring and comparing the potential load placed on different RA's.

However, the most important point is that these two simple measures allow for direct analysis of the effect on measured distance by different methodologies, since there are no other factors or elements involved in the calculations of the measures. There are no factors of desirability, walkability, public transport access, population weight, etc. to muddy the effects of changing methodologies for measuring distance and representing RA's.

3.2.5 Fifth Element: Distance

Apparicio et al. (2008), La Rosa (2014), and Higgs et al. (2012) collectively mention the following distance types as being often used in such analyses:

* Euclidean distance (e.g. via buffering);
* Manhattan distance (mentioned by La Rosa (2014) and Apparicio et al. (2008))
* Network distance (Shortest length);
* Network distance (Least time) (mentioned by La Rosa and Apparicio et al.)

Based on Apparicio et al., Bilkova et al., Higgs et al., and Sander et al., the most accurate (or realistic) type of distances is considered to be network distances. While Euclidean distance can approximate network distances, they can be less reliable in certain cases (e.g. suburban areas). (Apparicio et al., 2008)

Both Euclidean distance and network distance rely heavily on the methodology used. In theory, Euclidean distance is relatively simple – it is the shortest distance between two points, in this case between RU and RA representation. In practice this can be done by creating a buffer around points of origin and then selecting all destination points located within each point of origin's buffer zone, or vice versa. A buffer can be described as a zone or area surrounding a feature, with a size defined by a set distance *x* from the feature's points, so that it covers all points within distance *x*. Buffer zones can event be dissolved into each other to create buffer areas for multiple features. See Figure 5.

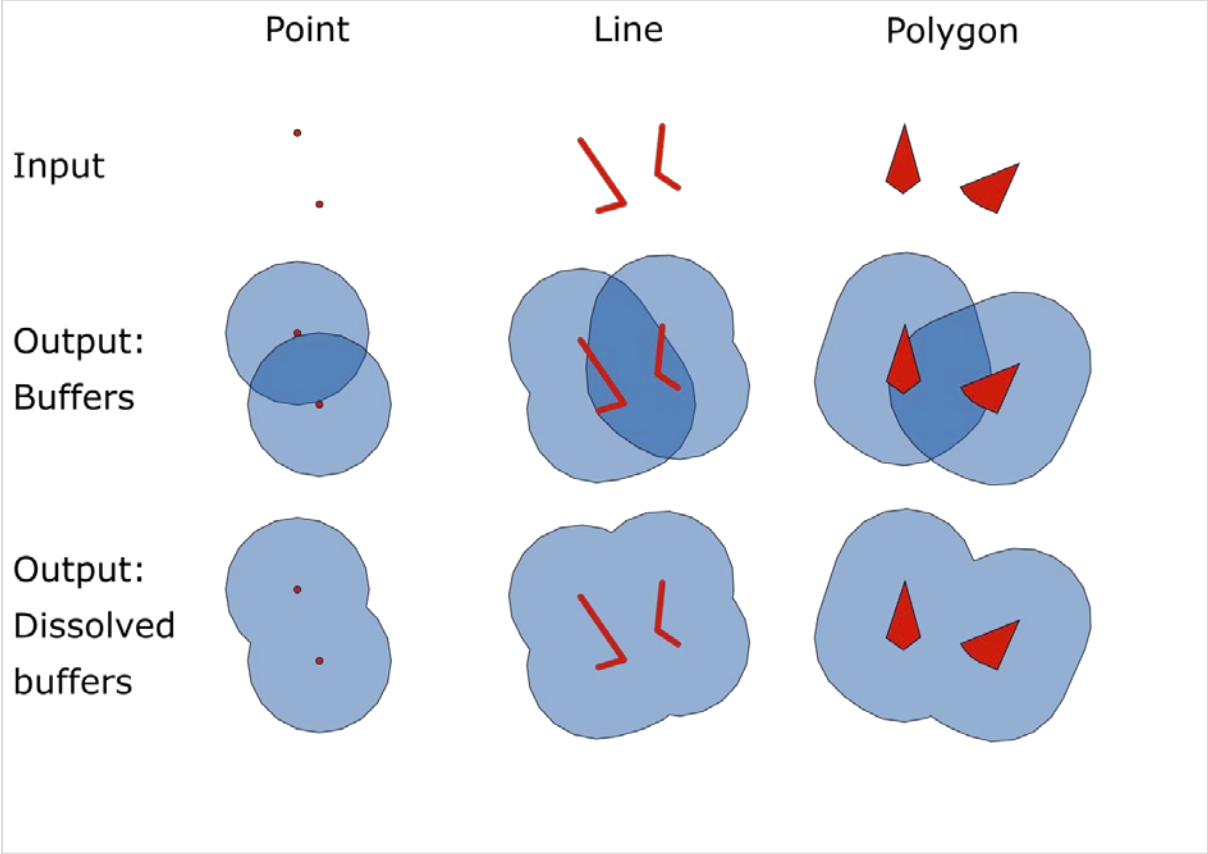


Figure 5 – Buffers and dissolved buffers for points, lines, and polygons.

Another way is to create a Euclidean distance table between all points of origin and all destinations. This is the method chosen for this study since it is the fastest way to produce Euclidean distances. The specific steps taken to produce the table are listed in the 'Analysis' section of this chapter. In all combinations using Euclidean distance (see Figure 6), distance X varies depending on the RA representation selected.

3.2.6 Measurement of Euclidean distances using Near tool

This study uses Euclidean distance tables produced by the ArcMap (ESRI, 2011) Near tool, found within the Proximity toolbox. According to ESRI's website, "The Near tool calculates the distance from each point in one feature class to the nearest point or line feature in another feature class." (ESRI 2017c)

In the case of this study, the Euclidean distances measured using this tool are between origin points representing RU's to destination points representing RA's (centroids and access points). Vector-based distance tools in ArcGIS follow the basic premise that distance is measured by drawing a straight line between two points. In mathematical terms, the x and y coordinates of two points are used to calculate the distance between them on a flat surface. The simplest way of calculating this is by using Pythagoras' theorem:

Equation 1

$$a^2 = b^2 + c^2$$

Looking at two points, one can adapt the concept of the right-angled triangle the points and use Pythagoras' theorem to find the shortest distance between the two points (Figure 6)

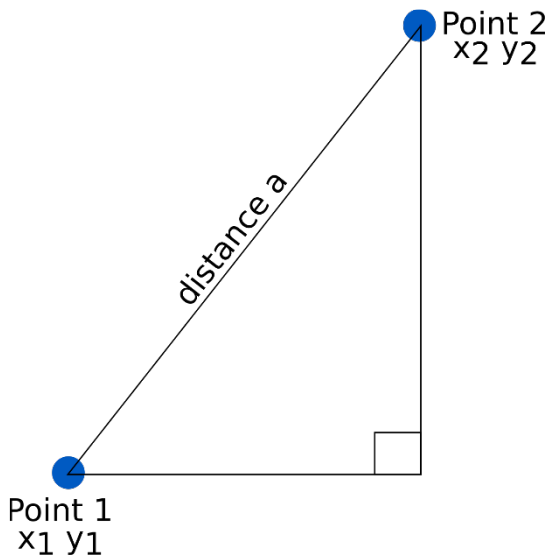


Figure 6 – Application of Pythagoras' theorem to calculate distance between two points on a flat plane.

Pythagoras' theorem can be adapted for use with two points to find the shortest distance between them:

Equation 2

$$a = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Where:

a = shortest distance between two points (point 1 and point 2)

x1 = x coordinate of point 1 x2 = x coordinate of point 2

y1 = y coordinate of point 1 y2 = y coordinate of point 2

In this case, geodesic distances were not taken into account since planar distances were measured.

Network distance is somewhat more complicated. When measuring network distance it is a path which is measured, in many cases, such as in this study, the shortest path (Shawi et al., 2014). Several algorithms are available for use, depending on the nature of the network being used, and selection of one algorithm over the other is often a question of calculation speed rather than consistency, accuracy or precision. (Cherkassky et al., 1996)

ArcMap (ESRI, 2011) makes use of Dijkstra's algorithm (Dijkstra, 1959) to find the shortest path between two points along a network (ESRI, 2017e). In the case of finding the shortest distance between a point of origin and a destination point, Dijkstra's algorithm's 'Problem 2' is especially applicable (Dijkstra, 1959). In this case, Dijkstra's algorithm works in the following way where the shortest distance between two nodes, P and Q, is calculated (Adapted from Dijkstra, 1959):

1. An assumption is made that if there is a shortest path between points P and Q, then the shortest paths from P to other nodes R on the shortest path to Q can be constructed successively until Q is reached.
2. All nodes apart from P and Q are divided into three sets:

   i. Nodes for which the shortest path from P has been found. Nodes are added to this set in order of minimum path length starting from P, since distance between P and P is 0, while distance from P to unvisited nodes is infinity.
   ii. Nodes from which the next node will be selected to be added to set 2(i).
   iii. All other nodes.
3. Branches, that is paths from P, are divided into three sets:
   i. Shortest paths from point P to nodes in set 2(i)
   ii. Branches from which the next branch will be placed into set 2(ii). Only one branch can possibly lead to a single node in 2(ii).
   iii. All other branches.
4. At the start of operations all nodes and branches belong to their respective set iii.
5. Two operations follow:
   i. The distance to all nodes connected to the last node placed in set 2(i) is considered, via their respective branches. If the new total distance from P to any node R is shorter than the one already known (at the start an unknown distance is set to infinity), then this new branch (with the shorter distance) is moved from set 3(iii) to set 3(ii) while node R is moved to set 2(ii) if it belongs to set 2(iii).
   ii. Taking the shortest branch from the branches from sets 3(i) and 3(ii), the node closest to P is moved from 2(ii) to 2(i), while its branch is moved from 3(ii) to 3(i).
6. Steps 1 and 2 are repeated until the destination point, Q, ends up in 2(i). The associated branch is then the shortest branch.

When creating Origin Destination cost matrices (OD cost matrices), maximum distance cutoffs as well as the maximum number of nearest destinations are applied (ESRI, 2017e). An OD cost matrix is essentially a table (created using ArcMap) containing network distances between all combinations of origin and destination points, and is calculated using a provided network dataset (ESRI, 2017b). ArcMap only improves the basic algorithm by using well-suited data structures such as d-heaps, and by snapping origin and destination points not only to junctions but to part of an edge, i.e. any part of the network (ESRI, 2017e).

As mentioned above, network tools in ArcGIS (ESRI, 2011) measure distances along a network. That means that distances between points outside the network and the corresponding location along the network are not included in calculations of distance. Origin and destination points are snapped onto a network, that is, placed on the nearest point along an edge feature, such as a line (ESRI, 2017f)

Figure 7 shows how network distances would look like if the distance between snapped point and nearest network point were taken into consideration, where:

X1 = distance between residential block representation and nearest point on network

X2 = distance between two points on the network

X3 = distance between destination point and nearest point on the network

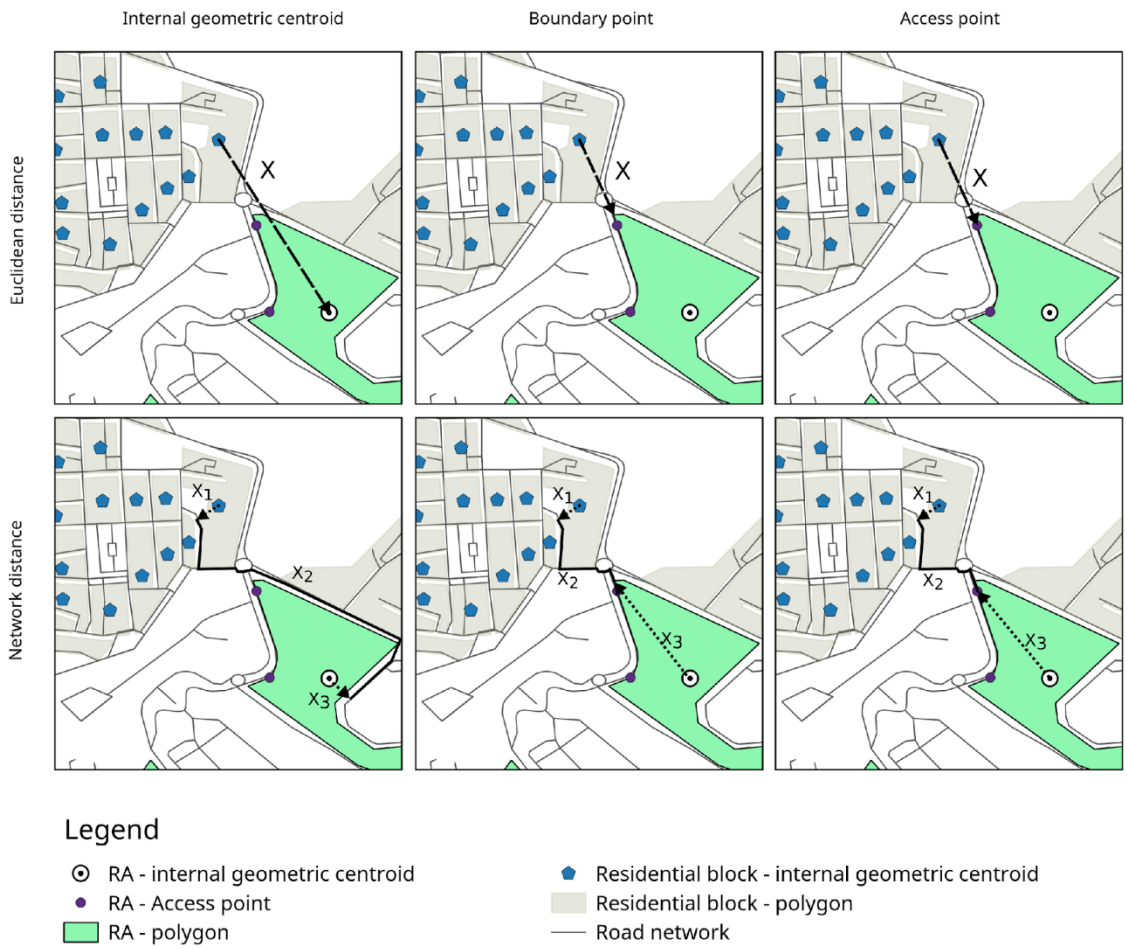


Figure 7 – Comparison of distance measures and feature representation combinations including discarded boundary points option (replicated from Higgs et al., 2012)

This concept of taking full-network distance into consideration is taken from Higgs et al. (2012), and is a good way of looking at how using simple network distance can influence results compared to when full-network distance is used, since it takes into account the whole chain of distances calculated during GIS analysis (X1, X2 and X3 are all calculated, but X1 and X2 are then discarded). Which one between network and full-network distance is most useful is a big question - comparing the two illustrated the effect of selecting one over the other on results.

Distance calculations will therefore use network distances (X2), full-network distances (X1+X2+X3) and Euclidean distances.

3.2.7 Acceptable walking distance

Since this study will be using walking distance to measure of accessibility, maximum walking distance is useful to consider since it is necessary to set an arbitrary cut-off point when preparing data – e.g. the maximum distance to nearest recreational area.

The fixed distances used by La Rosa in his calculations to produce distance-based accessibility measures (Simple distance Indicators and Proximity Indicators) were 300m, 600m and a non-fixed distance (La Rosa, 2014). Distance-based accessibility measures take into account all destination features within a specific distance (the fixed distance) from a point of origin. Non-fixed distance in this case means that all destinations within a dataset are taken into consideration, irrespective of how far away they are from a point of origin. The use of a non-fixed distance-based accessibility measure would assume the use of a motorized vehicle or some sort of public transport, since all destinations would be included, even those which are several kilometres away from a destination, and too far to walk to within a reasonable time. Higgs et al., on the other hand, mention a cut-off point of 300 to 400m for walking distances as significant limit (Higgs et al., 2012). Apparicio et al. use distances of 500m, 1000m, and 2000m (Apparicio et al., 2008). However, their calculations are made for hospital services, where the use of a car is more probable. These fixed distances are not appropriate as walking distance. The cutoff distance of 1000m, as a maximum limit walking distance, will be applied to this study where necessary.

3.3 Base data

Spatial datasets needed for the analyses in this study were selected based on the elements mentioned in the previous section. The datasets were obtained from different sources:

- Road network data was obtained from Open Street Map data, downloaded via Geofabrik (Geofabrik, 2016).
- The 2012 Corine landcover dataset (Copernicus, 2016). This dataset was used to produce Tables 1 and 2.
- Data on recreational areas in Malta was used with permission by the University of Malta[2] and the Valletta 2018 Foundation[3]. This data was compiled and digitized by the author between 01-12-2013 and 30-11-2014, using a methodology of field work and digitization. See Appendix II for considerations taken when reviewing which features to include and which to exclude when reviewing the available dataset.

The types of facilities included in the definition mentioned in the background chapter are visible in Table 2.

[2] The University of Malta is the largest post-secondary and tertiary education institution in Malta. (University of Malta, 2018)

[3] The Valletta 2018 Foundation was set up to prepare and run "the European Capital of Culture programme in Malta" by "driving cultural, social and economic regeneration in Valletta and the Maltese Islands through collaboration, exchange and innovative practice." (Valletta 2018 Project, 2017)

Recreational Areas		
Green Spaces	Playing Fields	Other Spaces
Gardens	Playing Fields	Artificial Open Spaces
Parks		Squares[4]
Urban Forests		'Gardens' in name but not in fact[5]
Green Areas[6]		
Green squares[7]		

Table 2 – Subcategories of recreational areas

The above table builds from the available V18 dataset and its original categories 'playing fields' and 'parks and recreational areas', using experience and local knowledge gained during fieldwork. Recreational areas under (re)development were also included, where possible.

- Maltese local council boundary data was used to define the study area. This data was digitized from maps showing the latest boundary designations during 2014. This data was used to delineate the study area used in the analyses.
- Urban Atlas Data (downloaded from https://www.eea.europa.eu/data-and-maps/data/urban-atlas 15th July 2016). This data was used to extract RU data for use in the project's analysis.

  Originally, data for RU's was thought to be available in the Enumeration Areas (EAs) dataset used by Malta's National Statistics Office (NSO). However, upon obtaining the

[4] In Malta, squares are usually non-green public open spaces with a recreational element.

[5] Although gardens are green by definition, in Malta there are often recreational areas called gardens but which in fact have very little vegetation due to badly planned renovation projects.

[6] 'Green areas' is a catch-all term for any green spaces which do not fit into another category

[7] Although squares can often be considered grey rather than green, having very little vegetation, some squares are an exception in that they can be considered green spaces.

data, it was found that the EAs where simply too irregular and stretched in shape to be of any use for spatial analysis. The Urban Atlas dataset was used instead.

3.3.1 Selection of projection

Since several different layers had to be superimposed and analysed together, it was important to determine a common geographic coordinate system and a projection for all data layers to be re-projected to. All spatial data was transformed to datum WGS84 and re-projected to WGS84 / UTM Zone 33N. This ensured that all data had a common geographic coordinate system while also having a locally suitable projection (see Malta Planning and Environment Authority, 2016 and EPSG Geodetic Parameter Registry, 2017).

3.4 Data pre-processing

The data pre-processing stage contained several steps to prepare the data for analysis. The software used was QGIS[8] due to its higher speed. (QGIS Development Team, 2009)

Steps taken during pre-processing included:

- Re-projecting all datasets to UTM 33N
- Local council boundary data:
    1. Exporting GHA localities to new layer
    2. Merging all GHA locality polygons into a single polygon for use in extracting recreational areas
- Recreational areas data:
    1. Calculating RA area and exporting all RA with size => 1500m².
    2. Creating a buffer of 1000m from the merged re-projected GHA boundaries shapefile and using the resulting buffer to clip the resulting layer from (1). All RA within the GHA and a 1000m buffer around it are selected. This is because 1000m

[8] QGIS versions 2.12 and 2.14 were used.

is the maximum walking distance being considered in this study, and the presence of RAs beyond the study boundary will help avoid boundary effects during analysis.

3. Creating and exporting the geometric centroids (internal) of the RAs from step 2.
4. Digitizing the access points for the RAs in step to and exporting the access points as a separate layer.

- Road network data:

1. Road network data is clipped to include only the island of Malta. Network data is not clipped to the study area to avoid boundary effects during analysis.
2. Since road network data obtained from OSM contains data about different types of roads and can be uneven in quality, road network data is visually checked for quality. The road network is also edited by removing or adjusting all network types unsuitable for walking, namely:

   i. trunk – Main roads (the largest type of road) without suitable sidewalks for walking.

   ii. trunk_link – Roads connecting main roads to other roads. This type of road does not have pedestrian access.

   Adjustments to the above categories were made by either removing the roads completely or else by editing them so that they better reflect walkability (for example by adding pedestrian crossings where missing, by removing roundabouts, and by joining double carriageways where tunnels and pedestrian bridges exist. Similar adjustments were made to 'primary' and 'primary_link' road categories where necessary.

   Test network analyses were also run using recreational area centroids and access points with RU centroids to detect errors. The majority of the resulting errors were due to the nearest road being disconnected from the main network. Such roads were joined to the main network where this reflected reality, while others were simply removed to allow access to the next nearest road feature – the assumption here being that all RU's have access to the main road network.

When digitization occurred aerial imagery (Bing and Google map layers via QGIS's OpenLayers plugin), and the Maltese planning authority's map server (https://www.mepa.org.mt/mepa-mapserver, accessed regularly from January 2015 to December 2016) were used as references for digitization.

3. Creation of network dataset. The network dataset is created from the road network dataset. This was done using ArcCatalog 10.1 (ESRI, 2011) using the following settings:

   i. Do not model turns in network – Since no turn data is available.

   ii. Connect to any vertex – This allows for roads to be connected to other roads along any point.

   iii. No elevation model was used to create z values

   iv. Network dataset attributes:

      1. Cost: Length in metres – This means that distances are measures in metre, not in, for example, time or kilometres.

      2. Oneways turned off – Oneway properties of roads are not considered when building the road network since pedestrians can walk in either direction along a pavement.

   v. Travel mode:

      1. Type: Walk – Walking mode was selected since it avoids hierarchies (e.g. where straight motorways are preferred to smaller, winding roads sine travelling on foot does not have the same requirements as driving)

      2. Impedance: Length (metres) – Metres were selected since distance is being measured in network analysis.

      3. U-Turns at Junctions: Allowed – This reflects the fact that pedestrians can turn around at any point.

   These settings allowed for maximum flexibility in order to reflect the flexibility of walking.

4. Obtaining RU data. This data was obtained from the European Environmental Agency's Urban Atlas dataset (European Environmental Agency, 2010), which maps land use of urban areas within the European Union. This dataset was the highest-resolution data available to the author.

   Several categories of land use are present within the downloaded dataset. Of these, all polygons within the study area which fell into the categories of Urban Fabric were extracted. Several subcategories were present in the dataset, each covering a range of percentage ratio of artificial surface cover, labelled Sealing Layer (S.L.) (Mequignon, 2011)

   i. Continuous Urban Fabric (S.L. > 80%)

   ii. Discontinuous Dense Urban Fabric (S.L. : 50% - 80%)

   iii. Discontinuous Low Density Urban Fabric (S.L. : 10% - 30%)

   iv. Discontinuous Medium Density Urban Fabric (S.L. : 30% - 50%)

   v. Discontinuous Very Low Density Urban Fabric (S.L. < 10%)

   Urban fabric is described as "built-up areas with their associated land [where] residential structures and patterns are predominant..." (Meirich, 2008: 18). The percentage figures represent the percentage of soil sealing, i.e. the amount of natural land cover sealed off by artificial surfaces (Meirich, 2008: 18).

   Another category which may have contained lone villas or farmhouses was called 'Isolated structures'. This category was not included because it was impossible to tell which structures were actually residential in nature without further investigation.

3.5 Data issues and limitations

3.5.1 Alignment issues

It was noted that the Urban Atlas data and street data do not fully align with other data, especially street data. The alignment is off by about 1 to 3 metres. A certain misalignment is also present (and expected) between RA centroids/access points and the road network. This

misalignment is to be expected and, being small and present in practically all measurements, has a rather small and even effect on analysis.

3.5.2 Road network accuracy

Road network data, having been obtained from Open Street Map (OSM), is not guaranteed to be one-hundred percent accurate or complete. Although basic checking and editing were made to adapt OSM's road network layer to this study's purposes, possible errors and issues may have gone unnoticed by the author.

3.5.3 Recreational areas

Recreational areas were collected by the author via fieldwork and desktop research. Several limitations exist, however:

- RA's were digitized with the help of Google maps' satellite imagery. This means that some level of inaccuracy and imprecision (e.g. due to lack of visibility or high walls creating an angled image.)
- RA access points were digitized with the help of MEPA's map server and of Google maps (Google Maps, 2015). Their accuracy is higher, but some entrances may exist on plan but be closed off in reality.

3.5.4 Residential units

This data was obtained from Urban Atlas layers and is somewhat generalized. RU's are divided into blocks, which means that it is impossible to check for quality without collecting field reference data. Although this would have limited an analytical study, this study focuses on methodologies rather than pure analysis and therefore its main requirement is internal consistency within the dataset.

Another limitation is that weighting centroid representations by population was impossible, as was creating access points for residential blocks. These limitations meant that variation in representations of polygons was only possible for RA's. Although using only RU centroids lessens the variety of analyses possible in this study as well as producing potentially less

accurate results when estimating distances, the use of RU access points is not crucial, since this study focuses on differences in methodology rather than accurate estimation of distances.

3.5.5 Use of the Universal Transverse Mercator Projection

The use of an equidistant projection is recommended for use with planar distance calculations (ESRI, 2017d). UTM was used in this study, which is a conformal rather than equidistant projection. That said, equidistant projections only retain true distances in specific directions (Shinker, 2017), while conformal projections suffer the most distance distortion in non-local scales, that between points/areas on different sides of a map (Furuti, 2016). The study area is relatively small compared to the projection's designated area, so that relatively little relative distortion of results is expected.

3.6 Analysis

Production of distance measures

Based on the elements mentioned at the start of the methodology chapter, the following analytical combinations exist (Table 3).

	Euclidean distance	Network distance (X2)	Full-network distance (X1 + X2 + X3)
Internal geometric centroid	Euclidean distance to nearest centroid	Network distance to nearest centroid	Full-network distance to nearest centroid
Access point	Euclidean distance to nearest access point	Network distance to nearest access point	Full-network distance to nearest access point

Table 3 – Combining distance measures and feature representations

These combinations can be applied to the two accessibility measures being used:

- Distance to nearest RA;
- Id of nearest RA.

The result is six combinations for each of the two measures.

Euclidean distances

Euclidean distances were measured in the following steps:

1. The Euclidean DNRA and NRAID for each RU was calculated using the ArcGIS Near tool found in the Proximity toolbox. Input features (i.e. points of origin) were the residential block centroids, while the Near features (i.e. destinations) were the RA representations. The option Location was checked so that nearest feature ID could be obtained. The option 'Method' was set to 'Planar', which meant that the curvature of the Earth (based on projection) was not taken into account, and that distances would therefore be calculated based on a flat surface (ESRI, 2017). The reason for using planar is that the area being studied is so small (see fig. 3) that the earth's curvature plays practically no role in comparison to such factors as height over sea level, which was not used in this study.
2. The results included the residential centroid IDs, the distance to and ID of the nearest RA representation for each RU centroid. The results were then exported to a CSV file for analysis.

The above steps were followed for each RA representation type, so that results for Euclidean distance to RA centroid and Euclidean distance to RA access point were calculated.

Network distances

Network distances were measured using the ArcGIS (ESRI, 2011) Network Analyst extension.

Using Network Analyst, OD cost matrices were created using the network dataset created in the pre-processing phase. Creating an OD cost matrix "finds and measures the least-cost paths along the network from multiple origins to multiple destinations." (ESRI, 2017b)

The OD cost matrices was created using residential block internal centroids as the origin layer and RA representations as destination layers. The following OD cost matrices were therefore produced:

1. Network distances from all RU centroids to RA centroids.
2. Network distances from all RU's centroids to RA access points.

It was ensured that all ID fields matched so that all origin and destination points would have the correct names in the results layer. Snapping of origin and destination points to the network for the aim of calculating network distances was allowed, with a search tolerance of 1000 metres to limit snapping to the maximum walking distance established for this study.

One-way restrictions were ignored while U-turns at junctions were allowed since the mode of transportation being considered is walking. The impedance measure was set to length in metres.

The resulting cost matrices were exported as CSV files for further processing, namely using a Python script to yield DNRA and NRAID results from the OD cost matrix containing network distances from all RU centroids to RA centroids and network distances from all RU's centroids to RA access points (script in appendix IIIb)

OD cost matrices contain one row for every combination of origin and destination points in the input layers. Each row contains the following basic fields:

1. Origin point ID – using an integer as an ID
2. Distance from origin point to destination point – using a decimal number to represent distance
3. Destination point ID – using and integer as an ID

The algorithm used to extract the DNRA and NRAID results for each residential block is the following:

1. Create two empty lists, list A (for temporary results) and list B (for final results).
2. Open the OD cost matrix.
3. For each origin ID number, loop through all rows in document once using a generator (to reduce memory load – a generator can load one row at a time into memory rather than the whole document).

i. For each row with a matching origin ID, save the row in a list A.

4. When the rows have been looped through once find the smallest distance value from all the rows in list A.

5. Add the result to list B.

Full-network distances

As discussed, full-network distance can be defined as:

- distance from population representation (X1) to nearest point on network, plus
- network distance (X2), plus
- nearest distance from network to destination point (X3).

Full-network distance is therefore X1 + X2 + X3.

Cost matrices for RA centroids and access points are reused from the network distances step. The values in full OD cost matrices are added to the matching distances X1 and X2 distances based on matching IDs, using the Python script included in Appendix IIIb

This process includes the shortest distance values between origin/destination points and the network which are excluded from the OD Cost matrix. The new cost matrices are then processed using the Nearest feature script included in Appendix IIIa.

Results contain both full-network distance to nearest RA and the nearest RA's ID.

3.7 Resource demand

As mentioned earlier on, different types of accessibility measures demand different amounts of resources (computational time, human hours and computational resources). This depends on the complexity of the calculations as well as on the amount, based on number of points of origin, destination points, and in case of network analyses, network complexity.

The approximate levels of demand that were observed are given in Table 4.

	Euclidean distance	Network-only distance	Full-network distance
Centroids	Low	Medium	Medium-High
Access points	Low	Medium-high	High

Table 4: Approximate and relative resource demand for production of accessibility measures based on combination of distance type and feature representation

Such a difference in resource demand means that a selection of distance and feature representation must take the resource-intensiveness of such a choice into account.

3.8 Comparing results

An important part of this study is to compare the resulting accessibility measures in a meaningful way.

Comparing accessibility measures can be a difficult task. Each type of measure demands a different kind of comparison to make sense of the effects the selection of a particular distance and RA representation has.

In the case of NRAID, the results are unique values, and results from different methods can be compared in absolute terms: either they match or they do not. Therefore, a simple percentage overlap comparison suffices to quantify the difference between methods. This follows Higgs, Fry and Langford's (2012) methodology for comparing results for NRAID using different elements. This operation was done using a Python script (found in Appendix III). This script records the number of cases where the resulting nearest RA ID of a residential block is the same for both methods being compared, and divides this by the total number of cases, returning a percentage value – a picture of the proportion of cases where the nearest RA affected by a change in method used. This operation can be condensed into an equation:

Equation 3

$$Overlap = \frac{\sum cases_{matching}}{\sum cases_{total}}$$

Comparing non-unique continuous values, such as distance to nearest RA is not as straightforward. Existing studies (Apparicio et al., 2008; Higgs et al., 2012; La Rosa, 2014) use Spearman's correlation rank. This method, is useful for showing how two measurements vary in relation to each other, thus helping to spot unaccounted for factors influencing the results. However, it is not useful for capturing the magnitude of the difference between two otherwise highly correlated variables, especially when all the factors influencing results are already accounted for (Bland and Altman, 1986, 2010). In other words, "the correlation coefficient measures the strength of the relationship, and it is incorrect to interpret it as a measure of agreement" (Magari 2002: 32). An example: the correlation of resulting Euclidean distance to nearest RA access point using Euclidean distance and full-network distance gives a Pearson correlation coefficient of 0.923, significant at the 0.01-level (2-tailed). The two variables have a strong linear correlation and vary at very similar rates. However, this says nothing about the absolute difference between the two variables. The difference between the two variables is plotted as a histogram in Figure 8.

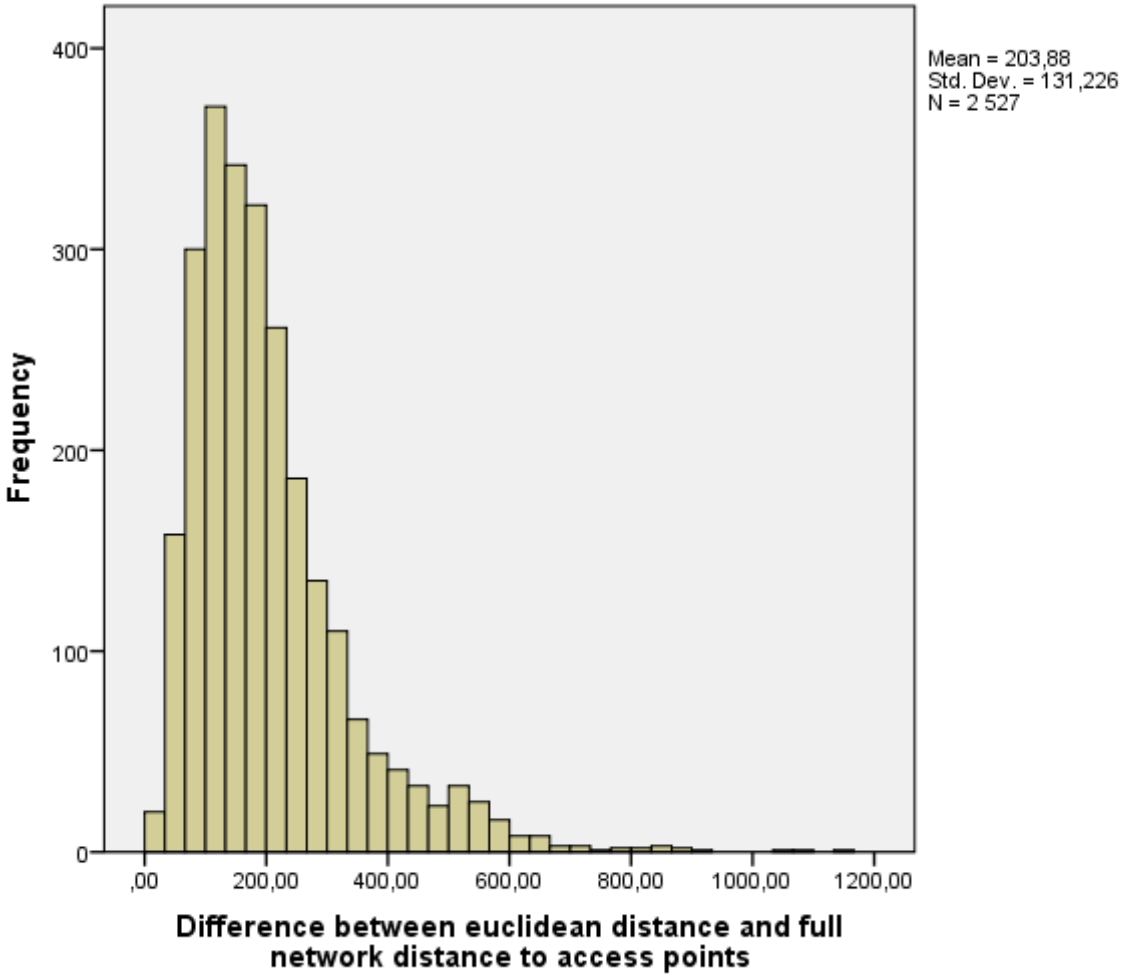


Figure 8 – Example histogram illustrating spread of differences between distance between residential block centroids and nearest RA access point using Euclidean and full-network distances.

The difference between Euclidean and full-network results is visible here, with an arithmetic mean of 204 metres. This difference was invisible in the correlation score, and is large enough to seriously influence results for both distance to nearest RA and the nearest RA's ID – in this case 19.23% of cases have a different nearest RA (see Table 5).

A way of viewing the differences (no matter how highly correlated) between two variables is to plot the differences between two methods against the average of two methods. According to Altman and Bland (1983), "it is much easier to assess the magnitude of disagreement (both error and bias), spot outliers, and see whether there is any trend" (Altman and Bland, 1983:

313). Plotting difference vs average of the results of the two measures mentioned above, we obtain the following scatter plot (Figure 9).

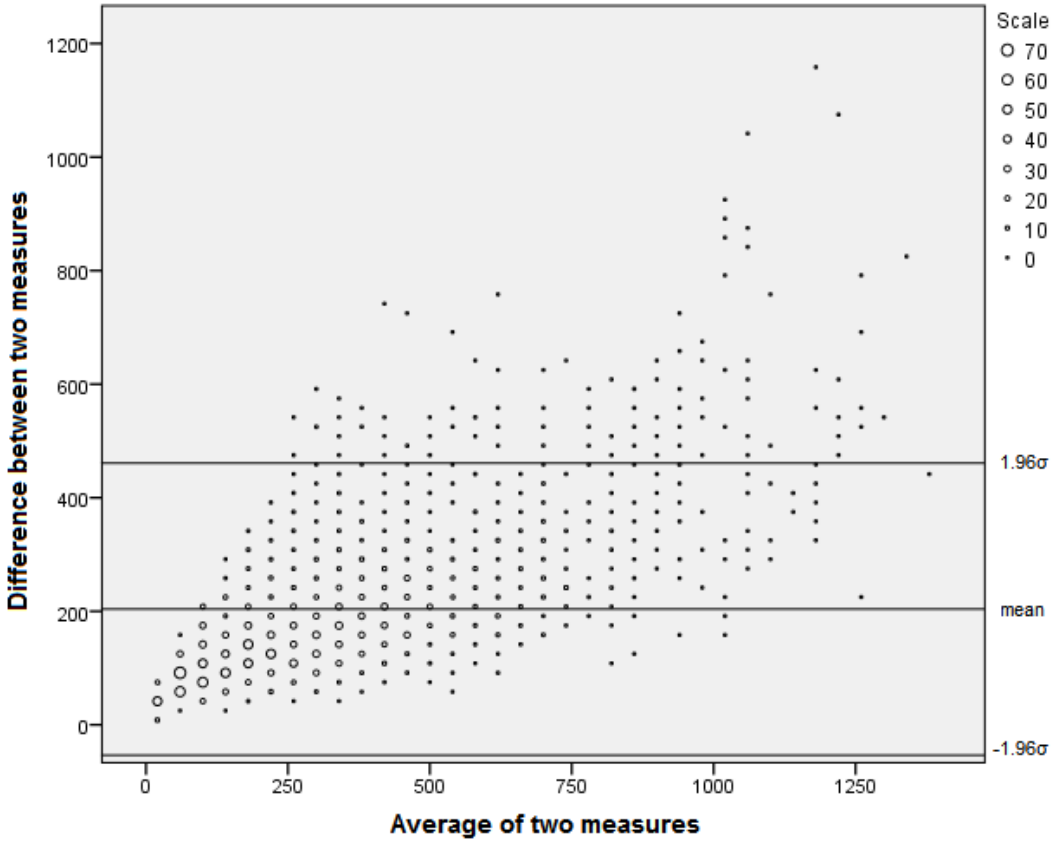


Figure 9 – Scatter plot of differences between and averages of two measures. Scale indicates that the size of a circle represents x points collected together to make reading the graph easier. Mean indicates the arithmetic mean[9].

Using such a graph makes it easy to see that most cases are within one standard deviation. The small number of cases with very large differences could possibly be related to larger RAs (since the measures compared both use Euclidean distance). It can be speculated that the differences are attributable to the distance between centroid and edge, although this is simply speculation.

[9] The arithmetic mean is defined as "the mean of *n* numbers is […] their sum divided by *n.*" (Freund, 1970: 29-30) This can be represented by the following formula, where is the mean the sum of numbers *x* divided by their number *n*.

The use of a scatter plot of average vs difference is proposed by Altman & Bland (1983) and Bland & Altman (1986 and 2010) and is named Bland-Altman Plot. The original purpose of such a plot is to compare clinical measurement methods, the plot can also be used for spatial applications, e.g. as used in Blaasaas & Tynes (2002) for the comparison of three ways of measuring distances. The Bland-Altman plot is suited for comparing methods of measurement, especially of continuous data. This is exactly the scenario seen in this study, where different methods are used to calculate, or 'measure', a number of distances (continuous numbers).

The Bland-Altman plot is meant as a visual tool showing level of agreement between two methods, especially when the relationship between difference and mean is complex (Bland and Altman, 1986). In more straightforward cases, hard limits of agreement can be set to 1.96 s.d. from its mean/mean bias. If 95% of the plotted cases fall within the hard limits of agreement, then the two methods can be considered to agree. However, this method was suggested by Bland and Altman (1986) for use when comparing instrument measurements, and depends on the differences in results being normally distributed. In this study's case it cannot be expected that differences between the results of two element combinations will be normal, since other factors than error are involved (e.g. difference between Euclidean distance and network distance can only increase as distance increases). Figure 8 for instance, is not normal, since it is positively skewed and not symmetrical around its mean (Freund, 1970: 136) which suggests that further processes other are also involved.

The Bland-Altman plot will therefore be used to compare results for distance to nearest RA.

Calculations of difference of measurements will follow following order for consistency:

Distances	*RA representations*
$d_{fullnetwork} - d_{network}$	$d_{accesspoint} - d_{centroid}$
$d_{fullnetwork} - d_{euclidean}$	
$d_{network} - d_{euclidean}$	

Not all possible comparisons are made – in some cases, comparing different measures does not yield a useful comparison. Since there are two elements involved in calculating accessibility measures: RA representation and distance measure, one of these elements is always kept constant when making comparisons.

3.9 Interpreting the Bland-Altman plot

Reading Altman and Bland (1983) and Bland and Altman (2010), one comes across several possible characteristics of a Bland-Altman plot which must be kept in mind during interpretation.

The first question to be asked is whether plotted points cluster around the mean line. Plotted points cluster around the mean when the only factor affecting results is the difference between the two methods being compared. Increasing or decreasing trends in the plot can indicate accumulating errors, resulting in cases where differences get larger as average values get larger. A positive trend between average measured distance and increasing differences in the method is visible in Figure 9.

Another key factor of Bland-Altman plots is the distribution of points. The wider the points of the plot are distributed, the lower the agreement between the methods being compared. Since distributions of differences in this thesis are not normal it can be interesting instead to compare distributions between different plot.

Distribution can vary within the plot. An example found in this study was funnelling (clearly visible in Figures 9 and 13). Funnelling indicates decreasing levels of agreement as the funnel grows wider. This means that the two methods may possibly be used interchangeably only up to a certain value, beyond which the two methods do not have acceptable levels of agreement.

Another important consideration is outliers. Clusters of outliers can indicate a subgroup of cases influenced by unaccounted-for external factors. In Figures 12a, 12b, and 12c, the majority of points lie along the mean. A small number of outliers, however, is present, indicating the presence of weak funnelling.

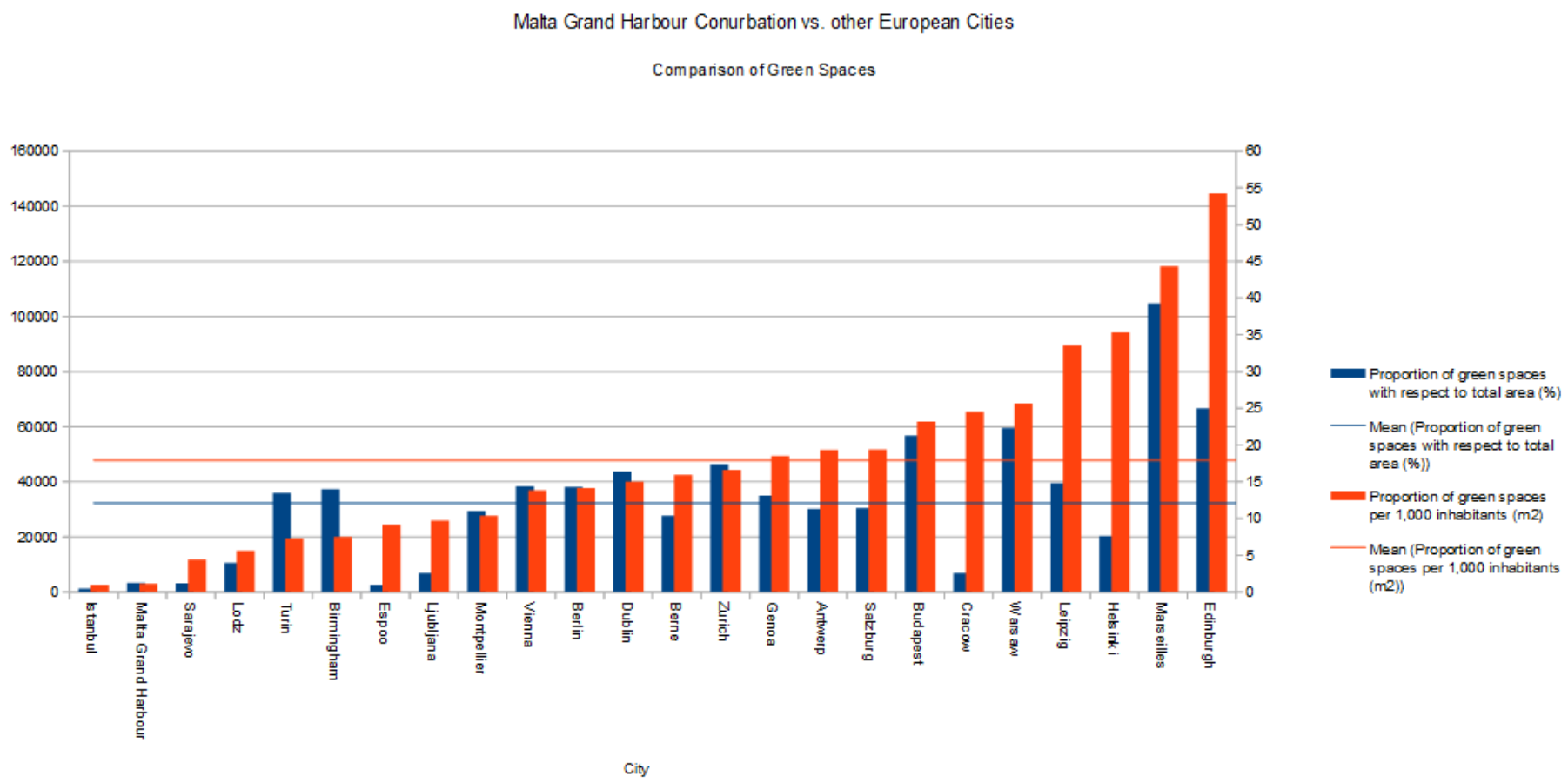
Other patterns may also be present. In Figures 12a, 12b, and 12c, several points cluster into horizontal sequences, possibly an artefact of network calculations.

4. Results

This chapter will present the results and describe them in a way that the reader can understand them and their interpretation. A comparison of accessibility measures resulting from different methods of measuring distance will be presented to determine how large an impact the selection of distance measures has on the calculation of an accessibility measure, in this case NRAID and DNRA.

An overview of how Malta's Grand Harbour Area compares to major European cities when it comes to availability of RA's can be found in Appendix A, and is summarized in Figure 10 below. Area-wise, Malta's GHA is next-to-last both in the proportion of RAs in relation to the total land area of an urban area, and in square metres of RA's per 1000 inhabitants of an urban area. The term green areas is used in the chart since it is based on data collected by Baycan-Levent et al. (2009) where the term 'green spaces' is used. Recreational areas is the equivalent term being used in the Maltese context, and can be considered to be equivalent, although, as can be seen in Appendix I, comparison between different cities can be somewhat shaky, since definitions of RAs and green spaces varies between cities, communes, regions and countries.

Figure 10 – Comparison of the Malta Grand Harbour area with other cities



4.1 Nearest Recreational Area ID: Comparing results

Table 5 presents percentage overlap between NRAID results using RA centroids and RA access points. For each distance type (represented by a single row in Table 5) the results were produced by finding the nearest recreational area to each residential unit using the relevant distance type. The nearest RA was calculated twice, once using distance from each RU to RA centroids and once using distance from each RU to RA access points. The result was two lists of NRAID for each RU within the study area (NRAID using shortest distance from RU to RA centroid, and NRAID using shortest distance from RU to RA access point) for each distance type (six in total for all distance types). For each distance type, the two relevant lists were compared and the proportion of NRAID was calculated as a percentage. Therefore, three comparisons for three different distance types were made, as is seen in Table 5.

Distance type	**Percentage overlap between NRAID results**
Euclidean distance	86.93%
Network distance	76.63%
Full-network distance	77.88%

Table 5 – Percentage overlap of nearest Recreational Area ID (NRAID) – comparison of NRAID results using RA centroids versus RA access points for each distance type.

When using Euclidean distance to calculate the nearest RA ID for each residential block, nearest RA ID overlaps in 87% of the cases. This is a large overlap, indicating that the selection of one representation over another has a rather small effect on NRAID.

When using network distance, the overlap between nearest RA ID is lower, at 77% overlap of results.

The selection between centroids and access points when using full-network distance has about the same impact, at 78% overlap.

	Centroids	**Access points**
Euclidean - network	71.79%	80.45%
Euclidean – full-network	71.04%	80.77%
Network - full-network	97.54%	90.83%

Table 6 – Percentage overlap of nearest Recreational Area ID (NRAID) – comparison of NRAID results using all combinations of distance types for each geometric representation type.

When RA representations are kept constant, a comparison of results obtained from different distance measures is possible.

In both cases, the biggest effect on results comes from using Euclidean distance versus one of the two network distances. The effect is noticeable – when using RA centroids, the overlap between NRAID calculated from Euclidean distance and NRAID calculated from network or full-network distance is of about 71-72%, while the overlap between network and full-network distances is of 97.5%

When using RA access points, selection between Euclidean and network/full-network distance give an overlap of about 80.5%, while choosing between network and full-network gives an overlap of almost 91%.

The factors influencing results are the following:

- Selecting RA centroid vs RA access points when calculating NRAID gives a difference in results of between 14-24%, with more uneven results when using network or full-network distances.
- When using RA centroids, choosing between Euclidean and network/full-network gives 28-29% difference in results.

- When using RA access points, selecting between Euclidean distance and network / full-network distance yields a difference in results of about 19%.
- Whether RA centroids or access points are used, the difference in results between using network or full-network distances is minimal, from about 2.5% to 9.3%.

4.2 Distance to nearest Recreational Area – Comparing results

Although looking at NRAID gives a good idea of how the selection of specific elements when measuring accessibility can affect results, NRAID results are influenced by the position of RA's relative to the point of origin and to each other. For any residential block, a difference in RA ID does not matter much if the closest two RA ID's to a RU are both 20 kilometres away and just 20 metres from one another. Comparing distances removes this layer of obfuscation – only absolute distances are measured, not relative positions. A comparison of DNRA results follows below, as well as a visual presentation of DNRA results (Figure 11).

Differences and averages of different combinations of distances and feature representations are presented in the results below, as already described in the methodology chapter. To make presentation easier, differences and averages of these combinations are described using the following abbreviations:

DCEN: difference between DNRA using Euclidean and network distance with RA centroids.

ACEN: average of DNRA using Euclidean and network distance with RA centroids.

DCEF: difference between DNRA using Euclidean and full-network distance with RA centroids.

ACEF: average of DNRA using Euclidean and full-network distance with RA centroids.

DCNF: difference between DNRA using network and full-network distance with RA centroids

ACNF: average of DNRA using network and full-network distance with RA centroids

DAEN: difference between DNRA using Euclidean and network distance with RA access points

AAEN: average of DNRA using Euclidean and network distance with RA access points

DAEF: difference between DNRA using Euclidean distance and full-network distance with RA access points

AAEF: average of DNRA using Euclidean distance and full-network distance with RA access points

DANF: difference between DNRA using network distance and full-network distance with RA access points

AANF: average of DNRA using network distance and full-network distance with RA access points

4.2.1 Comparing Recreational area representations

The use of either centroids or access points when calculating DNRA leads to some differences in results. A first look at mean differences between DNRA results using centroids or access points for each type of distance in Table 7 gives a similar picture to Figure 8, where there is a noticeable difference between results for centroids and access points when using network and full-network distances, but with lower difference when using Euclidean distances.

Difference type	Arithmetic mean (bias) (in metres)
DECA	-45.2350
DNCA	-100.5442
DFCA	-112.3696

Table 7 – Calculated bias for DECA, DNCA, and DFCA.

Using Bland-Altman representations (Figures 12a, 12b, 12c) one can see that for all distance measures, the difference between internal geometric centroids and access point results tends to be distributed along the mean. This indicated there are no unaccounted-for factors influencing these results. However, Euclidean distances vary much less than network and full-network distances. Almost all DECA values are within +- 200m.

This indicates a relatively high level of agreement, especially since the majority of cases lies along and rather close to the mean, with a smaller number of larger deviations driving the mean difference up. Network and full-network distance give much lower agreement, with most points having +200 to -500m difference.

Although not so clear, a funnel effect is visible, following a negative linear relationship. This means that as average distance increases, the difference between distances using different RA representations and using the same distance increases. Since differences are calculated using access point minus centroid, this means that while distance type is constant, using an access point tends to give shorter distances than using centroids. This is perhaps not so strange for Euclidean and full-network distances, but it is harder to account for in network distance, where distance between RA representation and the network is unaccounted for.

Since, except for outliers due to the rather weak funnelling, differences are distributed around the mean, this bias can be calculated and addressed by adjusting all distances with the bias value to obtain a mean of zero.

In the case of Euclidean distance adding a value of 45 would yield values which hover around a mean difference of zero.

This means that when using Euclidean distance, one could accept the compromise of choosing centroids over access points (or vice versa) for ease of computing, or due to limitations in source data, provided a slight correction of values (here set at 45.2m) is made. If using access points is somehow impractical, changing to centroids can be accepted if the limitations it brings with it are accounted for.

The values obtained from using network values and full-network values, although easily adjusted due to their even distribution around the mean, have a high level of disagreement that cannot be accounted for by a simple compensation for bias. Since access points can be assumed to be closer to reality than centroids, this means that using centroids with network or full-network distance will have a strong effect on results, which must be considered when interpreting results. The reason for this disagreement probably lies in the fact that road networks create a large amount of variation in results due to their complex nature.

This leads to the next question: can Euclidean, network and full-network distance be used interchangeably?

Figure 11 – Visual comparison of DNRA results using different element combinations.

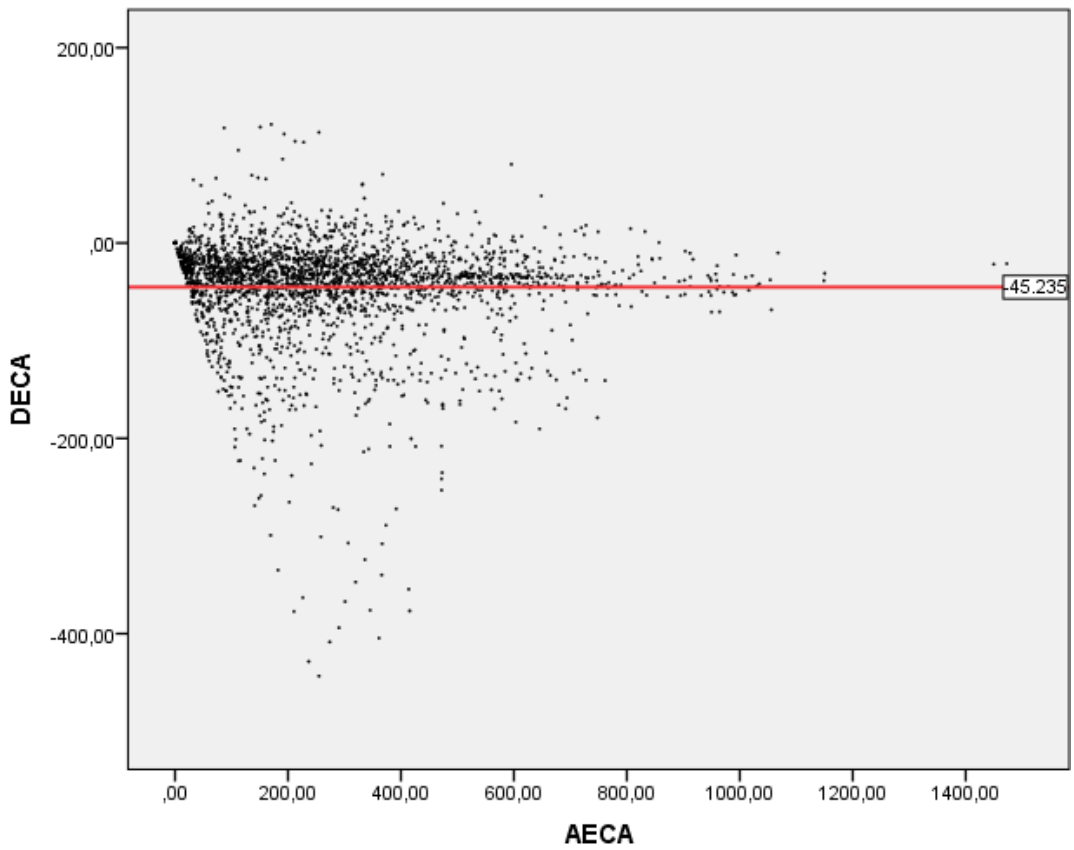


Figure 12a – Scatter plot of difference vs average of DNRA per block using RA centroids and access points with Euclidean distance.

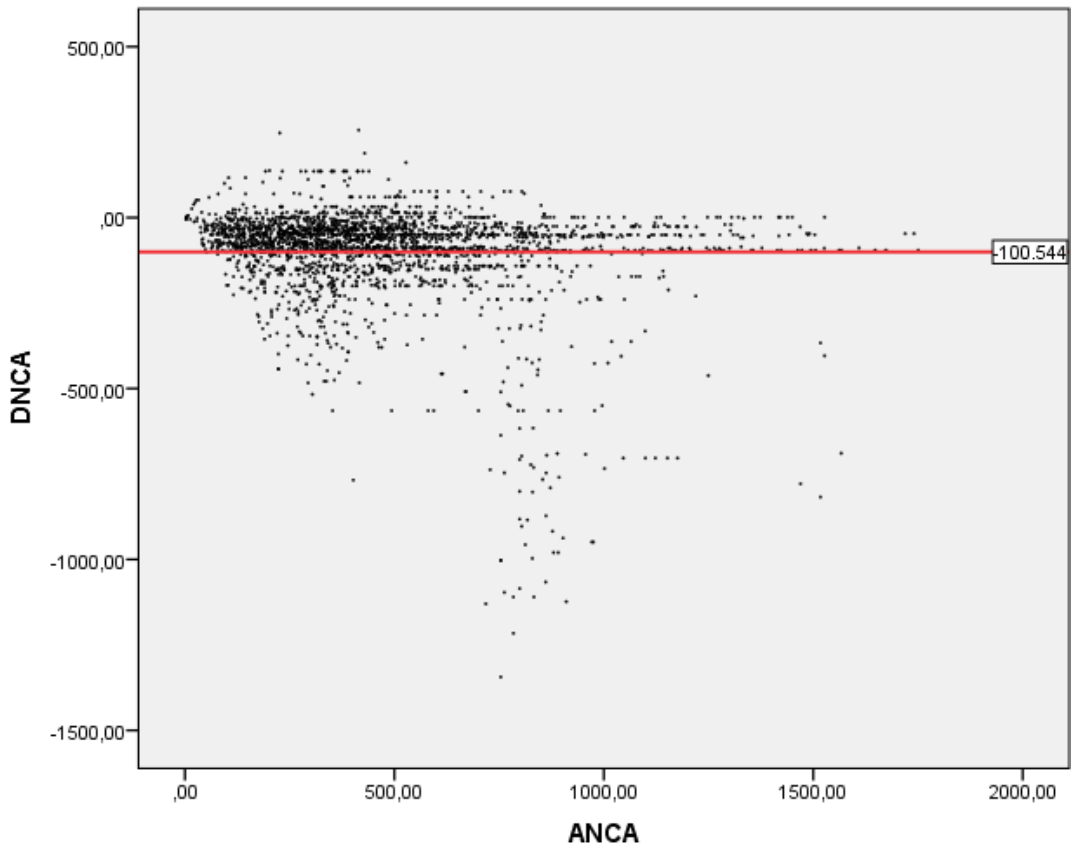


Figure 12b – Scatter plot of difference vs average of DNRA per block using RA centroids and access points with network distance.

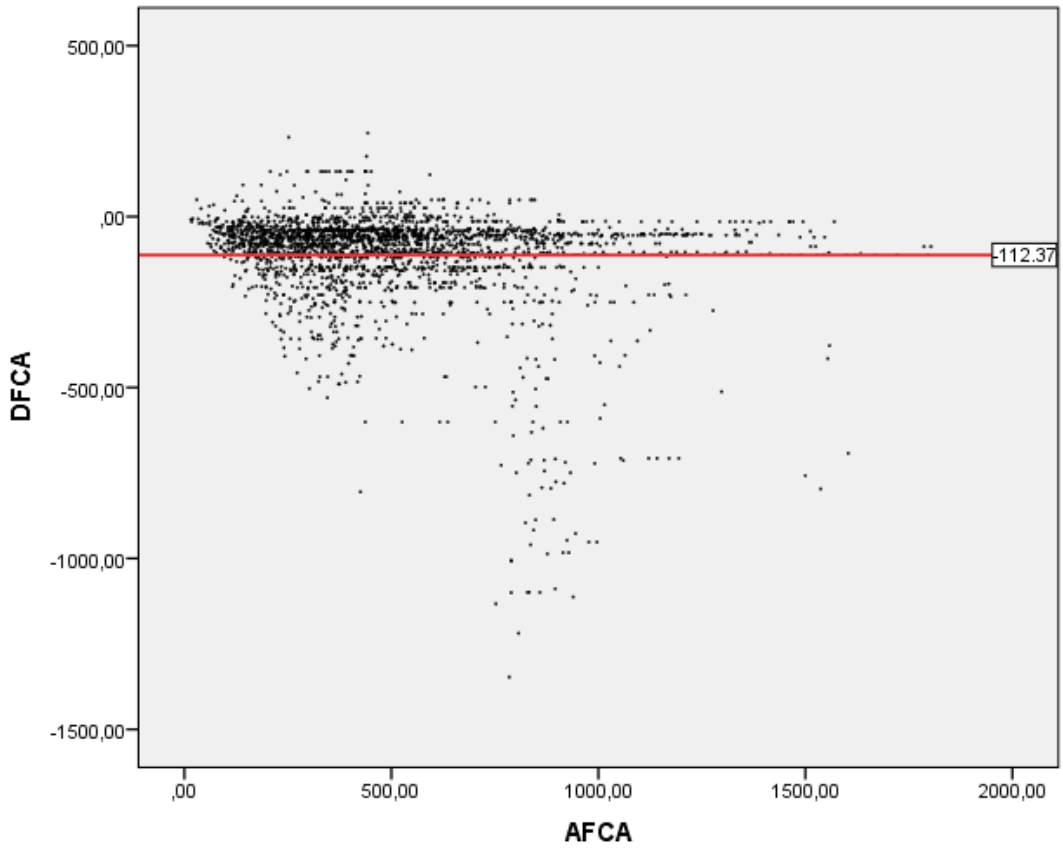


Figure 12c – Scatter plot of difference vs average of DNRA per block using RA centroids and access points with full network distance.

4.2.2 Comparing distance measures

Figure 13 below contains Bland-Altman plots for differences between distance measures when using centroids (top row) and access points (bottom row).

Agreement between Euclidean distance results and network / full-network results, both when using centroids and access points is very low, with differences strongly increasing as distances grow larger (a positive relationship). This is probably because differences between network distance and Euclidean distance (which is the shortest possible distance between two points) accumulate as objects get farther from each other. There is also funnel effect present, which points to a larger variation between results as distances grow larger. This is likely an effect of the semi-random nature of the local road network – different paths between different objects can have different lengths even though Euclidean distance is similar.

Choosing between Euclidean and network or full-network can therefore yield very large differences, especially as distances grow. Correcting for bias is not enough, since agreement diminishes so strongly as distance increases. If one assumes that network distances are more realistic than Euclidean distances, then the use of Euclidean distances to measure accessibility over a road network yields increasingly more inaccurate and imprecise results.

Difference type - difference between DNRA using:	**Arithmetic Mean (in metres)**
Euclidean and network distance with RA centroids (DCEN)	231.766
Euclidean and full-network distance with RA centroids (DCEF)	270.891
Network and full-network distance with RA centroids (DCNF)	39.125
Euclidean and network distance with RA access points (DAEN)	176.564
Euclidean distance and full-network distance with RA access points (DAEF)	203.884
Network distance and full-network distance with RA access points (DANF)	27.319

Table 8 – Number of cases and mean distance for DCEN, DCEF, DCNF, DAEN, DAEF, and DANF. See Figure 13 or glossary for explanation of terms.

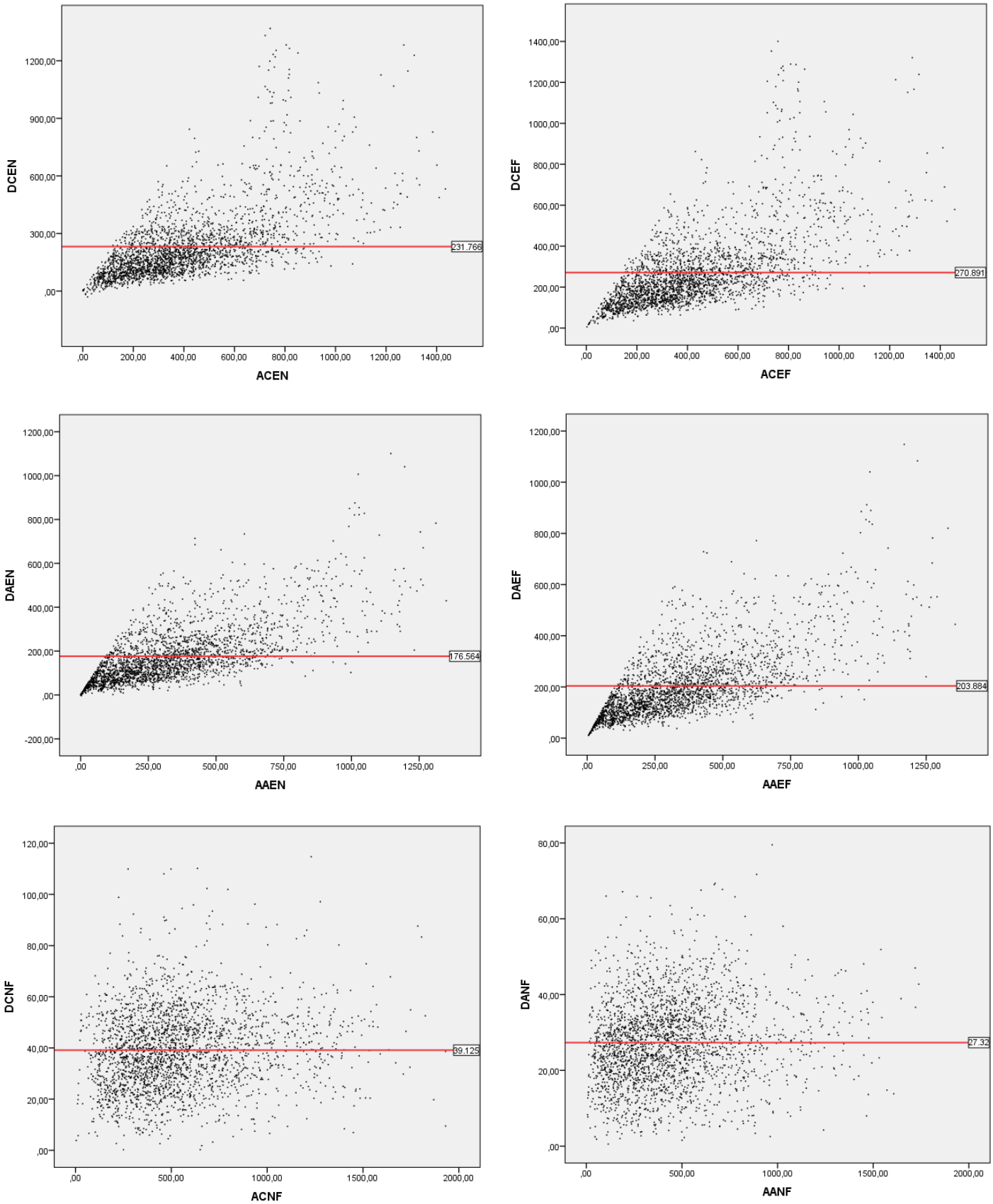


Figure 13 – Scatterplot of difference vs average of DNRA per block with RA representation kept constant.

Agreement between network and full-network results for both centroids and access points is much higher and more consistent. There is no hint of funnelling here. Since the values are distributed around a mean value, adjusting values for network distances around an average difference of 0 with full-network distances is easy: -39.1m for centroids and -27.3m for access points (see Table 8).

5. Discussion

Distance to facilities is a measure often used in varying forms by studies seeking to investigate accessibility to a specific kind of facility. Distance to facilities is too generic as a measure, and is therefore often adapted to the need of the authors, and given a concrete form such as nearest facility and distance to nearest facility, number of facilities within a specified radius, average distance to all facilities, and so on. These different measures all have a basic component as their base: distance, and by extension the identity of the facilities selected by using the resulting distance.

The measures explored in this study are distance to nearest recreational area (DNRA), equivalent to distance to nearest facility, and nearest recreational area ID (NRAID). Necessary elements for the calculation of the above-mentioned measures (both based on distance) are a point of origin (in this case represented by a residential block centroid), a way to measure distance (in this case either Euclidean distance, network distance, or full-network distance), and a destination (represented either by an RA centroid or RA access points). The element combinations tested here were the following:

- Euclidean distance and destination centroids
- Euclidean distance and destination access points
- Network distance and destination centroids
- Network distance and destination access points
- Full-network distance and destination centroids
- Full-network distance and destination access points

Results have shown that when selecting between Euclidean, network or full-network distance when calculating distance-based accessibility measures, the choice of distance can yield noticeable differences, especially since disagreement accumulates as distances grow, with the difference between network / full-network distances and Euclidean distances incrementally growing larger as distance between residential block and RA increases. The differences in NRAID result comparisons show this very clearly in percentages.

The choice between network and full-network has very little impact on the tested accessibility measures, and differences can be compensated for. Here the effect of using centroids versus access points becomes more obvious since the effect of choosing Euclidean versus network/full-network distance is removed.

The difference in results between using Euclidean and network/full-network distance is also visible in Higgs et al. (2012), who also use DNRA and NRAID. Unfortunately, DNRA result comparisons from their study cannot be compared to this study's results since Higgs, Fry, and Langford use Spearman's rank correlation to compare results, a method whose merits were discussed previously. It is possible, however, to compare comparisons of NRAID results. Significant differences in NRAID overlap are visible between the two studies, with Higgs, Fry, and Langford having an overall lower overlap than this study (Table 9).

NRAID overlap: Euclidean vs network distances using:	Study results	Higgs, Fry, and Langford (2012)
Centroids	71.79%	50.9%*
Access points	80.45%	70%*

Table 9 – Comparison of NDRAID overlap between network and Euclidean distance results, Higgs, Fry and Langford (2012) compared to this study.

* Results taken from Higgs et al. (2012).

Several different factors could be contributing to these differences. NRAID overlap differences are much larger for centroids, and this is possibly due to the fact that Higgs, Fry, and Langford use Cardiff as a study area, where recreational areas (called green spaces) are visibly larger than the ones present in this study's study area (Higgs, et al., 2012). Differences in the road networks of the two studies certainly also play a part, possibly being the main influencer of the

ca. 10% difference between NRAID overlap for access point results. Higgs et al. do not look at overlapping NRAID results in this case, and cannot be compared (Higgs et al., 2012).

Considering the effects of choosing Euclidean distance versus network/full-network distance, if one assumes that network distance is more realistic when measuring distance traveled along a network, and that full-network distance is even more realistic, then an acceptable middle way is to use network distance, since a considerable advantage is gained by using network distance, and little would be gained by using full-network distance over network distance, especially considering the increase of resource demands for using full-network distance. Using network distance over Euclidean distance is especially worth the effort in view of today's widespread availability of free road network datasets such as Open Street Map and ease of use of network tools. Bilkova et al. (2017); Higgs et al. (2012); as well as Sander et al. (2010) came to similar conclusions: choosing Euclidean distance over network distance is very likely to give different results which, if one considers that network distance is more realistic than Euclidean distance, can give rise to considerable error.

However, for Sander et al. (2010) the differences in time taken to achieve these results meant that the conclusion was less definitive. In a minority of cases they also found that Euclidean distances could be more accurate than network distances, where the real distance between origin and destination was zero or very close to zero.

Selecting between RA centroids and access points can also yield noticeable differences. When using Euclidean distances these differences are smaller than for network and full-network results – by 10% for NRAID and 100% for DNRA results. The findings in Higgs et al. (2012) do not match these results: in their study the differences in NRAID overlap between centroid and access points while using Euclidean distances were 64.8%, meaning that the NRAID overlap is much lower than what is seen in this study. Overlap between NRAID results using network distances is at 50.4%, again much lower than what is seen in this study. NRAID overlap results for full-network distances are not given.

Although access points are more realistic, these are not always available. If centroids are used as feature representation, it is important to take the potential difference in results into consideration. Some correction for bias is possible, but this does not attenuate the larger range of disagreement found in network and full-network distances. However, the fact that

disagreement is distributed around the mean with only weak funnelling makes taking differences into account easier. The difference in results when using RA centroids or access points can be partly explained by the shape of the recreational areas being investigated. In many cases, the shapes of RA's are uneven, often elongated or having shapes which cause the centroid to be position very far from access points.

A possible method for reducing the effect of using centroids over access points is to divide larger areas into smaller ones, so that the distribution of RA areas suddenly becomes smaller, since the distance from centroid to a potential (though unknown) access point would be smaller. That said, this could give rise to a situation where distances from larger RA's where access points are not evenly distributed are much shorter than in reality.

The above discussion about RA representations could also be extended to representations of origin features – the selection of one type of representation over another can be assumed to have similar effects on results as the selection of one destination representation over the other.

The question of whether using simple Euclidean distances and centroids can replace the more complex and demanding analyses using network / full-network distances and access points is worth investigating. Similar considerations are taken by Higgs et al. and Sander et al. Other factors such as available data and the amount of resources available for such analyses. This means that, for example, when using Euclidean distance, one could accept the compromise of choosing centroids over access points (or vice versa) for ease of computing, or due to limitations in source data, provided a slight correction of values is made. If using access points is somehow impractical, changing to centroids can be accepted if the limitations it brings with it are accounted for.

Available data for the study plays a huge role in which elements are chosen to measure accessibility, as is made clear in the literature review, where it is seen that many studies use second-hand data which they cannot really influence. In such cases there is no choice but to use what one has. However, it is important to state the quality of data available and the possible influence this can have on methodology and results, something not always present in the studies reviewed in this study, especially where distance-based accessibility measures were only a small part of the study.

Adjustments to the available road network data by removing trunk and trunk_link road sections and joining roads where pedestrian crossings were visible (see section 3.4 - Data pre-processing) has two identified potential effects:

- Shortest distances (calculated using network distances) between features may be affected in some cases.
- Shortest distance calculations using network distance are more accurate than when non-navigable zones are included in network distance calculations.

The effect of removing sections of the road network deemed unwalkable are unquantified in this study. Further adjustments to the available network data could have increased accuracy, such as field-testing and using multiple sources of road network data to ensure completeness of adjustments.

The amount of resources required to compute distance-based accessibility measures versus the available resources also plays a large role. Although the possibility to use network distance and accessibility points may exist, researchers may choose Euclidean distance and/or centroids to reduce computing time, especially if the production of accessibility measures is only a small part of a whole study.

When it comes to methodologies used for comparing distance measures and representations, most studies found use Spearman's correlation rank (e.g. Apparicio et al., 2008; Higgs et al., 2012), while others use Wilcoxon signed-rank tests (Sander et al., 2010) or Bland-Altman comparisons (Blaasaas and Tynes, 2002). The use of Spearman's correlation rank is not fully justified in the cases mentioned above, while Blaasaas and Tynes focus on different distance types and topic than this study's.

This has led to the author having to rely on resources from outside the field of study. The field of medical statistics seems to have tackled this issue a long time ago (Bland and Altman, 1983), and the application of the Bland-Altman plot method seems to be a common occurrence in this and medical science in general. The application of this method for the comparison of accessibility measures is therefore somewhat experimental, since nothing was found to compare it to in the study's field. Blaasaas and Tynes (2002) is an exception in that it takes up spatial questions within medical studies.

6. Conclusions

The questions to be investigated were set out in the Aims section of this study as the following:

How does the choice of distance measure and destination feature representation influence results when calculating two basic accessibility measures – nearest RA ID and distance to nearest RA?

The question was then split into the following sub-questions:

1. How does the use of different destination representations affect the identity of the nearest RA?
2. How does the use of Euclidean distances, network-only and full-network distances affect the identity of the nearest RA?
3. How does the use of different destination representations affect distance to the nearest RA?
4. How does the use of Euclidean distances, network-only and full-network distances affect distances to nearest RA?

The selection of destination feature representation (centroid or access point) has a noticeable impact on DNRA and NRAID results when using Euclidean distances, although this effect can be partly compensated for. However, when using network or full-network distances, differences in results from using centroids versus using access points are larger and more noticeable, and cannot be compensated for in a satisfactory manner. It is therefore important to take the effect of feature representation selection on results into account in all cases. By extension, one can safely assume that similar effects influence results based on what origin feature representation is chosen for a study.

The selection of one distance measure over the other also has a noticeable effect on DNRA and NRAID results when it comes to choosing between Euclidean distance and network or full-network distance. The largest differences are between Euclidean and network or full-network

distances for both NRAID and DNRA results. Agreement here is very low and differences in distances increase as distances grow larger. This is presumably due to the accumulating differences in distance between Euclidean distances and the road network. Correcting for such differences between Euclidean and network or full-network results is not possible, and differences are quite large and must be accounted for.

The same cannot be said for differences between results for network and full-network distances. Here, differences are relatively small (less than 10% for both centroids and access points) and can be partly corrected since agreement is high and there is no accumulative effect. It can be concluded that the selection of Euclidean distance over the more realistic network or full-network distances has a considerable impact on distance-based results, but the difference between network and full-network results is almost negligible. The small differences between network and full-network distance can be partly compensated for by adjusting network distances for bias, making the use of full-network distances unnecessary in most cases.

In this study DNRA and NRAID results often follow the same pattern – when DNRA results are affected, NRAID results are also affected. This is because NRAID results are dependent on distance to nearest feature, i.e. DNRA. In the same way one could expect other distance-based accessibility measured to be affected by the choice of feature representations and distance measures, and one could assume that the effects would be at least partially similar.

From these results one can conclude that the selection of both distance measures and feature representations has varying degrees of impact on distance-based accessibility measures, as well as any other distance-based result. It is therefore important that the type of data and distance measure used and their possible effects on distance-based results are taken into account. This could be done by simply mentioning the feature representations and distance measures used, by ensuring that the best available datasets are used (by best one means most realistic and most detailed), or even by conducting a small case-study in within the study area to quantify effects of the choice of feature representation and distance measure.

The comparison of distance-based accessibility measures conducted in this study is very limited in scope, since it compares only a small number of elements based on a single case study. A large problem with conducting comparisons with more factors or elements, accessibility measures and urban areas is that studies used different accessibility measures

based on available data for a single or at most a handful of urban areas. Even the methodology of the comparison itself could be expanded by exploring other possible methods of comparison. This leaves plenty of scope for future research.

The results presented in this thesis are very general. The application of the Bland-Altman method was by the non-normal distribution of the differences in the comparisons made. This meant that agreement could not be investigated fully. No definite results as to which method is most suitable for which case were provided. This makes comparison of results with those from other cities difficult. No existing studies were found using the same methods of comparison either, making it impossible to compare results with those studies.

Further limitations specific to base data used were mentioned in the methodology section. The issues mentioned were those of alignment of different datasets, limits in accuracy of road network data as well as the limited scope and detail of recreational area and RU datasets.

The applications of the methods used to explore the different results one obtains by using common elements for measuring accessibility can be applied to other features than recreational areas. The choice of types of origins and destinations is somewhat immaterial to the methodology, though they may influence interpretation.

There are also many different city configurations to explore, where the typical size of origin and destination feature varies, as does the network's configuration. There are also many other kinds of travel which can be explored, with a large number of possible elements to vary. For instance, public transport, car transport, bicycling, as well as the effect of roads and road conditions could be interesting to explore and compare to walking.

More work on finding local patterns of accumulation of difference between network distances and Euclidean distances as length of path grows could also be an avenue of exploration. This could give rise to ways of adjusting Euclidean distances to match network distances or full-network more closely, possibly by using a smaller number of network/full-network measurements to model changes in differences between Euclidean and network/full-network distances.

This also leads to another point: there are many different distance measures available apart from Euclidean, network and full-network distance. Manhattan distance and network distance with time-based increments are two oft-mentioned distance measures.

Comparing results of different methodologies is also possible. No mention was made here of impedance maps using raster data, for instance. Nor is the variability of points of origin addressed.

Finally, there is a multitude of accessibility measures which go beyond simple distance – where distance is only one element of many or where it does not feature at all. The comparison of such measures becomes more complex as more elements become involved.

7. References

Altman, D. G. and J. M. Bland. 1983. Measurement in Medicine: the Analysis of Method Comparison Studies. *Statistician* 32: 307–317. doi: 10.2307/2987937.

Apparicio, P., M. Abdelmajid, M. Riva, and R. Shearmur. 2008. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International journal of health geographics* 7: 7. doi: 10.1186/1476-072X-7-7.

Barr, A., R. Bentley, J.A. Simpson, J. Scheurer, N. Owen, D. Dunstan, L. Thornton, L. Krnjacki, et al. 2016. Associations of public transport accessibility with walking, obesity, metabolic syndrome and diabetes. *Journal of Transport and Health* 3: 141-153. doi: 10.1016/j.jth.2016.01.006.

Bauer, S. R., M.C. Monuteaux, and E.W. Fleegler. 2015. Geographic Disparities in Access to Agencies Providing Income-Related Social Services, *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 92(5): 853-863. doi: 10.1007/s11524-015-9971-2.

Baycan-Levent, T., R. Vreeker, and P. Nijkamp. 2009. A multi-criteria evaluation of green spaces in European cities. *European Urban and Regional Studies*, 16(2): 193–213. doi: 10.1177/0969776408101683

Bilkova K., F. Krizan, M. Hornak, P. Barlik, and P. Kita. 2017. Comparing two istance measures in the spatial mapping of food deserts: The case of Petržalka, Slovakia. Moravian Geographical reports 25(2): 95-103. doi: 10.1515/mgr-2017-0009

Bland, J.M. and D.G. Altman. 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet* 327: 307–310. doi: 10.1016/S0140-6736(86)90837-8.

Bland, J. M. and Altman, D.G. Altman. 2010. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies 47:* 931–936. doi: 10.1016/j.ijnurstu.2009.10.001.

Boone, C.G., G.L. Buckley, J.M. Grove, C. Sister.2009. Parks and People: An Environmental Justice Inquiry in Baltimore, Maryland. *Annals of the Association of American Geographers:* 99(4): 767–787. doi: 10.1080/00045600903102949

Carter, M. 2010. Public open space planning in Western Australia: new residential developments. Position paper. Parks and Leisure Australia (WA region). Perth, Western Australia, Australia.

Cetin, M. 2015. Using GIS analysis to assess urban green space in terms of accessibility: case study in Kutahya. *International Journal of Sustainable Development & World Ecology* 22(5): 420-424. doi: 10.1080/13504509.2015.1061066.

Central Intelligence Agency. 2017. *CIA World Factbook*. Retrieved 22 March, 2017, from https://www.cia.gov/library/publications/the-world-factbook/geos/mt.html/

Cherkassky, B. V, A.V Goldberg, and T. Radzik. 1996. Shortest paths algorithms: Theory and experimental evaluation. *Mathematical Programming* 73: 129-174).

Comber, A., C. Brunsdon, and E. Green. 2008. Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious group., *Landscape and Urban Planning* 86(1): 103–114. doi: 10.1016/j.landurbplan.2008.01.002.

Copernicus. Corine Land Cover 2012. Retrieved 6 January, 2017, from http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012/

Curl, A., J.D. Nelson, and J. Anable. 2015. Same question, different answer: A comparison of GIS-based journey time accessibility with self-reported measures from the National Travel Survey in England. *Computers, Environment and Urban Systems 4*9: 86–97. doi: 10.1016/j.compenvurbsys.2013.10.006.

Dai, D. 2011. Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? *Landscape and Urban Planning* 102(4): 234–244. doi: 10.1016/j.landurbplan.2011.05.002.

Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. *Numerische Matematik* 1: 269-271.

EPSG Geodetic Parameter Registry. Projected CRS [WGS 84 / UTM zone 33N]. Retrieved 15 September, 2017, from http://epsg-registry.org/?display=entity&urn=urn:ogc:def:crs:EPSG::32633/

ESRI. 2011. ArcGIS Desktop. Versions 10.1, 10.2, and 10.3. Redlands, CA, United States: Environmental Systems Research Institute.

ESRI. 2017a. Near tool. Retrieved 9 September, 2017, from http://pro.arcgis.com/en/pro-app/tool-reference/analysis/near.htm/

ESRI. 2017b. OD Cost Matrix. Retrieved 21 September, 2017, from http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/od-cost-matrix.htm/

ESRI. 2017c. Proximity analysis. Retrieved 16 October, 2017, from http://desktop.arcgis.com/en/arcmap/10.3/analyze/commonly-used-tools/proximity-analysis.htm/

ESRI. 2017d. How proximity tools calculate distance. Retrieved 16 October, 2017, from http://desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/how-near-analysis-works.htm/

ESRI. 2017e. Algorithms used by the ArcGIS Network Analyst extension. Retrieved 22 October, 2017, from http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/algorithms-used-by-network-analyst.htm/

ESRI. 2017f. GIS dictionary: Snapping. Retrieved 22 October, 2017, from http://support.esri.com/en/other-resources/gis-dictionary/term/snapping/

European Environmental Agency. 2010. Urban Atlas. Retrieved 21 November, 2014, from https://www.eea.europa.eu/data-and-maps/data/urban-atlas/

Geofabrik. 2016. Malta. Retrieved 6 January, 2016, from http://download.geofabrik.de/europe/malta.html/

Giles-Corti, B., M. Broomhall., M. Knuiman, C. Collins, K. Douglas, K. Ng, A. Lange, and R. Donovan. 2005. Increasing walking: How important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine* 28: 169-176. doi: 10.1016/j.amepre.2004.10.018.

Google Maps. 2014. *Malta (several locations)*. Retrieved January to November, 2014, from https://www.google.se/maps/place/Malta/@35.9423682,14.2379207,41910m/data=!3m1!1e3!4m5!3m4!1s0x130e45281d8647c5:0xf582d86136be4239!8m2!3d35.937496!4d14.375416

Freund, J. E. 1970. *Statistics: A first course*. Englewood Cliffs, New Jersey: Prentice-Hall.

Furuti, C. A. 2016. Conformal Projections. Retrieved 16 October, 2017, from http://www.progonos.com/furuti/MapProj/Dither/ProjConf/projConf.html/

Herzele, A. Van, and T. Wiedemann. 2003. A monitoring tool for the provision of accessible and attractive urban green spaces. *Landscape and Urban Planning* 63: 109–126. doi: 10.1016/S0169-2046(02)00192-5.

Higgs, G., R. Fry, and M. Langford. 2012. Investigating the implications of using alternative GIS-based techniques to measure accessibility to green space. *Environment and Planning B: Planning and Design* 39(2): 326–343. doi: 10.1068/b37130.

Higgs, G., M. Langford, and P. Norman, P. 2015. Accessibility to sport facilities in Wales: A GIS-based analysis of socio- economic variations in provision. *Geoforum* 62: 105–120. doi: 10.1016/j.geoforum.2015.04.010.

Hiu Ming, L. 2014. *Investigating the Small Public Urban Open Spaces at High-density Cities: A Case Study of Hong Kong*. Master Thesis. Uppsala, Sweden: Uppsala University.

M'Ikiugu, M. M., I. Kinoshita, and Y. Tashiro. 2012. Urban Green Space Analysis and Identification of its Potential Expansion Areas. *Procedia - Social and Behavioral Sciences*, 35: 449–458. doi: 10.1016/j.sbspro.2012.02.110.

Malta Planning and Environment Authority (MEPA). 2006. *Malta South local plan*. Malta Planning and Environment Authority, Marsa, Malta.

Malta Planning and Environment Authority (MEPA). 2014. Retrieved 22 April, 2017, from https://www.mepa.org.mt/infores-gi

Malta Planning and Environment Authority (MEPA). 2017. Help - MapServer Spatial Referencing System. Retrieved 15 September, 2017, from http://mapserver.mepa.org.mt/sites/malta_internet/help/neapoljs_help_en.htm.

Meirich, S. 2008. Mapping Guide for a European Urban Atlas. GSE Land Information Services, Report ITD-0421-GSELand-TN-01, Infoterra GmbH, Berlin, Germany.

Meng, Y., and J. Malczewski, J. 2015. A GIS-based multicriteria decision making approach for evaluating accessibility to public parks in Calgary, Alberta. *Human Geographies – Journal of Studies and Research in Human Geography* 9(1): 29–41. doi: 10.5719/hgeo.2015.91.3

Mequignon, L. 2011. Urban Atlas – Delivery of land use/cover maps of major European urban agglomerations. SIRS sas, report, Villeneuve d'Asq, France.

Morar, T., R. Radoslav, L. C. Spiridon, L. Păcurar. 2014. Assessing pedestrian accessibility to green space using GIS. *Transylvanian Review of Administrative Sciences* 42: 116–139.

Oh, K., and S. Jeong. 2007. Assessing the spatial distribution of urban parks using GIS. *Landscape and Urban Planning* 82: 25–32. doi: 10.1016/j.landurbplan.2007.01.014.

Oxford Dictionaries. Accessibility. Retrieved 21 April, 2017, from https://en.oxforddictionaries.com/definition/accessibility/

La Rosa, D. 2014. Accessibility to greenspaces: GIS based indicators for sustainable planning in a dense urban context. *Ecological Indicators* 42: 122–134. doi: 10.1016/j.ecolind.2013.11.011

La Rosa, D. and R. Privitera. 2013. Characterization of non-urbanized areas for land-use planning of agricultural and green infrastructure in urban contexts. *Landscape and Urban Planning* 109(1): 94–106. doi: 10.1016/j.landurbplan.2012.05.012.

QGIS Development Team. 2009. QGIS Geographic Information System. Versions 2.12 and 2.14. Beaverton, OR, United States: Open Source Geospatial Foundation.

Shawi, R., J. Gudmundsson, and C. Levcopoulos. 2014. Quickest path queries on transportation network. *Computational Geometry* 47(7): 695-709. doi: 10.1016/j.comgeo.2014.01.004

Shinker, J. J. Map projection properties. Retrieved 16 October, 2017, from http://www.geog.uoregon.edu/shinker/geog311/Labs/lab02/properties.htm/

Taylor, A. 2013. Green space issues in a comparative perspective: Berlin, Dublin, Marseille and Turin. University of Helsinki, Finland.

University of Malta. The University. Retrieved 13 February, 2018, from http://www.um.edu.mt/about/um/

Valletta 2018 Project. Culture Map Malta. Retrieved 4 September, 2017, from https://www.culturemapmalta.com/

Yang, J., C. Li, Y. Li, J. Xi, Q. Ge, X. Li. 2015. Urban Green Space, Uneven Development and Accessibility: A Case of Dalian′s Xigang District. *Chinese Geographical Science 2*5(5): 644–656. doi: 10.1007/s11769-015-0781-y.

Zhou, X., and J. Kim. 2013. Social disparities in tree canopy and park accessibility: A case study of six cities in Illinois using GIS and remote sensing. *Urban Forestry and Urban Greening* 12(1): 88–97. doi: 10.1016/j.ufug.2012.11.004.

8. Appendices

Appendix I - Prevalence and distribution of existing recreational areas

It is useful to compare the Maltese GHA scenario to other urban areas around the world to obtain some context to Malta's situation.

Baycan-Levent et al. compare several cities based on five different criteria:

"Quantity and availability of urban Green Spaces"; "Changes in Green Spaces"; "Planning of Green Spaces"; "Financing of Urban Green Spaces" and; "Level of Performance". (Baycan-Levent, Vreeker and Nijkamp, 2009)

The first criterion is interesting, since it helps place Malta's GHA within an international context. This criterion has three sub-criteria:

- "Proportion of green spaces with respect to total area" - This includes "gardens, urban parks, neighbourhood parks, historical gardens, green squares and plazas, green playgrounds and other city-specific green areas."
- "Proportion of green spaces per 1,000 inhabitants"
- "Existence of a regional green space system."

  (Baycan-Levent, Vreeker and Nijkamp, 2009)

Thus, the GHA can be compared to other European cities based on the above subcriteria.

Data for the above categories are provided for a number of cities in Baycan-Levent et al. (2009), and is shown in Table I below:

City	Proportion of green spaces with respect to total area (%)	Proportion of green spaces per 1,000 inhabitants (m2/1000 inhabitants)	Existence of a regional green space system
Antwerp	11.3	51,509	N
Berlin	14.3	37,786	Y
Berne	10.4	42,519	Y
Birmingham	14.0	20,000	Y
Budapest	21.3	61,800	Y
Cracow	2.6	65,455	Y
Dublin	16.4	40,000	Y
Edinburgh	25.0	144,592	Y
Espoo	1.0	24,465	Y
Genoa	13.1	49,394	N
Helsinki	7.6	94,154	Y
Istanbul	0.5	2,675	Y
Leipzig	14.8	89,617	Y
Ljubljana	2.6	25,966	Y
Lodz	4.0	14,947	Y
Malaga	59.3	4,614,815	Y
Malta	1.25[10]	3,106	N
Marseilles	39.3	118,225	Y
Montpellier	11.0	27,729	Y
Salzburg	11.4	51,755	Y
Sarajevo	1.2	11,818	Y
Turin	13.5	19,444	Y
Vienna	14.4	36,863	Y
Warsaw	22.3	68,499	Y
Zurich	17.4	44,253	Y

[10]Calculated from 2012 CORINE Landcover data

Table I – Green space sub-criteria (Baycan-Levent, Vreeker and Nijkamp, 2009: 208-8 except for Malta)

As mentioned earlier on, green areas are rather difficult to define in Malta, and a broader term, ‘Recreational Areas’, was selected which would function better with available data and local use.

	Proportion of RA with respect to total area (%)	Proportion of RA area per 1,000 inhabitants (m2)
Grand Harbour Area	2.12	5,275
North Harbour Area	1.69	3,321
South Harbour Area	2.52	8,274

Table II– Sub-criteria values for all recreational areas, green spaces and playing fields.

Table I and Figure 10 in the main text show that compared to other European cities, Malta's Grand Harbour conurbation has a low proportion of green spaces both in proportion to the population and as a percentage of the total land area. The same goes for the broader definition of RA (Table II), which is somewhat higher but varies considerably between the North and South divisions of the GHA. The North GHA has been subject to more recent, rapid and sprawling development, while the South GHA has been highly populated for longer, and had less space for rapid development.

That said, the above number may also be affected by the way green areas are categorised - Malta does not have the luxury of categorising large stretches of land on the edge of its conurbation as parks or green spaces, since most land is taken up by agriculture, industry, and smaller villages and towns. This makes comparison to cities within larger countries difficult.

It is also useful to note that no information on the minimum size of recreational areas taken into consideration in each city is given by Baycan-Levent, Vreeker and Nijkamp. In the case of Malta, any feature smaller than 0.15ha was not taken into consideration (see Appendix II).

Appendix II

Minimum size of recreational area considered

Another interesting aspect is the minimum size of facilities whose accessibility is being analysed, for example public spaces. The minimum size authors use varies. Morar et al. list three different studies where minimum sizes of green spaces analysed vary between 0.15 to 2 hectares (Morar et al., 2014). Ming, analysing small public spaces, considers all sizes, although the average area is seen to be between 1000-1500 square metres, with some recreational areas being as small as 60 square metres (Hiu Ming, 2014).

Field experience in the GHA has shown that many small recreational areas do not really provide a 'break' or 'refuge' for recreation, but are rather a sidewalk extension, a 'hole-in-the-wall' recreational area. Therefore, the minimum size of 1500 square metres is selected for this study to ensure that the recreational areas being considered are of sufficient size to provide enough space for recreation.

Accessibility and maintenance

Some spaces may be of the type being used for analysis (e.g. green spaces), but are not fully accessible due to lack of maintenance, danger of use, being closed off to the public, and so on. This means that green patches, fields, wasteland, inaccessible fortification ditches (a local Maltese phenomenon) and other inaccessible areas are not considered in this study, since they are not accessible to the general public.

Purpose of use versus usage in practice

Some spaces may be technically considered recreational spaces, but in truth are not used as such due to their use in practice being other than the original purpose. One such example is large chunks of the Marsa sports complex (see www.sportmalta.org.mt/facilities/marsa-sports-complex), a large area within the GHA which is often categorised as a green area, but which is in fact a series of specialised sport facilities with limited accessibility to the public.

Appendix IIIa – Nearest distance.py

```
#!/usr/bin/env python3
'''
Nearest distance.py
This script finds the shortest distance for each residential block (origin
feature) ID.
'''
import csv
import os
from collections import defaultdict
# final dataset to be exported
min_data = defaultdict(list)
# dataset created anew for each origin ID
single_res_dict = defaultdict(list)
def openCSV(filename):
    csv_file = open(filename, 'r')
    dialect = csv.Sniffer().sniff(csv_file.read(1024), delimiters=";,")
    csv_file.seek(0)
    reader = list(csv.reader(csv_file, dialect))
    return reader
def minimumLength(length_list):
    try:
        min_value = min(length_list)
        return min_value
    except ValueError:
        return None
def main():
    reader = openCSV('residences to rec centroids x1x2.csv')
    for c in range(17, 6000):  # origin ID - changed according to contents of file
        orig_subset = []
        length_subset = []
        for i, row in enumerate(reader):
            try:
```

```
                if int(row[0]) == c:  # value to be matched to c number, or the
origin id
                    orig_subset.append(reader[i])  # this builds a set of rows
with the same origin ID
                    length_subset.append(float(row[2]))  # row[2] for non-direct
dbf exports, 5 otherwise
            except ValueError:  # in case of empty row or row with text, i.e.
header
                continue
        shortest_l = minimumLength(length_subset)  # takes the shortest length
from the whole length subset
        try:
            shortest_index = length_subset.index(shortest_l)  # finds the list
index of the shortest length value
        except ValueError:
            continue
        try:
            print(orig_subset[shortest_index])  # print whole row of shortest
length for a particular origin ID
        except:
            continue
if __name__ == '__main__':
    main()
    print('Finished')
```

Appendix IIIb – Calculating full-network distance

```
#!/usr/bin/env python3
'''
x1x2x3.py
This file matches distances by ID from files containing x1, x2 and x3 distances.
It then adds these distances together and prints results.
'''
import csv
import os
from collections import defaultdict
# Opens CSV file
def openCSV(filename):
    csv_file = open(filename, 'r')
    dialect = csv.Sniffer().sniff(csv_file.read(1024), delimiters=";,")
    csv_file.seek(0)
    reader = list(csv.reader(csv_file, dialect))
    return reader
# Transforms selected CSV values to dictionary for easier use.
def CSVtoDict(csvpath):
    infile = openCSV(csvpath)
    reader = csv.reader(infile)
    for rows in enumerate(reader):
        print(rows[0])
    mydict = {rows[0]: rows[2] for rows in reader}
    reader.close()
    return mydict
def main():
    resd = openCSV('Distances_between_residential_centroids_and_road_network.csv')
    cosm = openCSV('residences to rec centroids.csv')
    recd =
openCSV('Distances_between_recreational_centroids_and_road_network.csv')
    resd_dict = {rows[0]: rows[2] for rows in resd}
```

```
    recd_dict = {rows[0]: rows[2] for rows in recd}
    out_file = open("x1x2x3 res centroids to rec centroids.csv", "w")
    writer = csv.writer(out_file)
    header_row = ['residential_centroid', 'rec area centroid',
                  'x1', 'x2_from_CM', 'x3', 'x1x2x3']
    writer.writerow(header_row)
    for i, row in enumerate(cosm):
        for v in resd_dict.keys():
            try:
                if int(v) == int(row[0]):
                    x1 = float(resd_dict.get(v))
                    break
                else:
                    continue
            except ValueError:
                continue
        for w in recd_dict.keys():
            if w == row[1]:
                x3 = float(recd_dict.get(w))
                break
            else:
                continue
        try:
            x1x2x3 = x1 + float(row[2]) + x3
        except UnboundLocalError:
            continue
        outrow = [row[0], row[1], x1, row[2], x3, x1x2x3]
        writer.writerow(outrow)
    out_file.close()
if __name__ == '__main__':
    main()
    print('Finished')
```

Appendix IIIc - Calculating overlap percentage of nearest RA IDs

```
#!/usr/bin/env python3
'''
percentoverlap.py

This script was created to calculate percent overlap (i.e. matching values)
between two sets of variables.

'''

import csv

#Opens csv file

def openCSV(filename):

    csv_file = open(filename, 'r')

    dialect = csv.Sniffer().sniff(csv_file.read(1024), delimiters=";,")

    csv_file.seek(0)

    reader = list(csv.reader(csv_file, dialect))

    return reader

if __name__ == "__main__":

    reader = openCSV('Network distance results-just nearestID.csv') #name of csv
file being analysed

    rowcount = 0 # number of rows iterated

    matchcount = 0 # number of matching rows

    for row in enumerate(reader):

        try:

            if row[1][2] == row[1][0]: # row[1][0] is the full-network access
points. The first index because a tuple is

                # returned by enumerate()

                matchcount += 1

            rowcount += 1

        except ValueError: #in case where text or empty cells are present

            continue

print(matchcount)

print(rowcount)

print(float(matchcount)/float(rowcount)*100)
```

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011).
12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011).
15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).
16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).
18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).
19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).
20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).
21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013).
22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).
23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013).
24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).
25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).
26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).
27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014).
28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).
29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).
30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).
31. *Sarah Tressel:* Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).
32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).
33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).
34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014).
35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).
36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).
38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).
39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).
40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).
41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).
42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015).
43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).
44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).
45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).
46. *Alejandra Narváez Vallejo:* Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).
47. *Anna Lundgren:* Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).
48. *Oluwatomi Esther Adejoro:* Does location also matter? A spatial analysis of social achievements of young South Australians (2016).
49. *Hristo Dobrev Tomov:* Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).
50. *Vincent Muller:* Impact of Security Context on Mobile Clinic Activities A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).
51. *Gezahagn Negash Seboka:* Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).
52. *Holly Buhler:* Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).
53. *Lars Ole Grottenberg*: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).
54. *Sean Grant:* The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).
55. *Irshad Jamal:* Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).
56. *Fulgencio Sanmartín:* Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. *Nezha Acil:* Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).
58. *Julia Hjalmarsson:* A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).
59. *Mathewos Tamiru Amato:* Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).
60. *Karim Alaa El Din Mohamed Soliman El Attar:* Bicycling Suitability in Downtown, Cairo, Egypt (2016).
61. *Gilbert Akol Echelai:* Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).
62. *Terje Slinning:* Analytic comparison of multibeam echo soundings (2016).
63. *Gréta Hlín Sveinsdóttir:* GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).
64. *Jonas Sjögren:* Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).
65. *Nadine Raska:* 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).
66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).
67. *Michaela Bobeck:* A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).
68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).
69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).
70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).
71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).
72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).
73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).
74. *Daniel Gardevärn:* PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).
75. *Kim Friberg:* Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).
76. *Viktor Svanerud:* Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).

77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).
78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).
79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).
80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).
81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).
82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).
83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).
84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).
85. *Luka Jovičić*: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).
86. *Christina Kourdounouli*: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).
87. *Jeremy Azzopardi*: Effect of distance measures and feature representations on distance-based accessibility measures (2018).