Check for updates

Available Online at www.e-iph.co.uk

E-B
Environment - Behaviour
Proceedings Journal

AicE-Bs2016Edinburgh

7th Asia-Pacific International Conference on Environment-Behaviour Studies,
St Leonard Hall, Edinburgh University, United Kingdom, 27-30 July 2016

e-IPH
e-International
Publishing House Ltd.,
United Kingdom

Are Euclidean Distance and Network Distance Related ?

K. Mert Cubukcu [1]*, Hatcha Taha [2]

[1] Department of City and Regional Planning, Faculty of Architecture, Dokuz Eylul University, Izmir, 35160, Turkey
[2] The Graduate School of Natural and Apllied Sciences, Dokuz Eylul University, Izmir, 35160, Turkey

Abstract

Although spatial distance is a very important concept for a wide variety of disciplines including social, natural, and information sciences, the methods used to measure spatial distance are not directly expressed and fully explained. In this study, we calculate and compare Euclidean distances and network distances for 10 randomly selected European cities. On the contrary to the findings reported in past research, we find that there is not a global straight forward relation between the Euclidian distance and network distance.

Keywords: Euclidean distance; network distance; network analysis

1. Introduction

It has been repeatedly referred that gravity models have been adapted to planning and geography studies more than any other mathematical model (Lee, 1973). Gravity-based urban models, also gravity-based accessibility measures, were originally derived from Physics, and have then been successfully applied to social sciences to analyze the interaction between various urban functions and human activities. According to the model, the interaction between the two areas increases with an increase in the size of the areas in question. Together with the size of the masses, distance is the predominant factor in determining the magnitude of the interaction. Thus, measuring distance is a major issue in these studies.

* Corresponding author. Tel.: +90-232-3018462
E-mail address: mert.cubukcu@deu.edu.tr

DOI: http://dx.doi.org/10.21834/e-bpj.v1i4.137

Recently, it has been argued that the Euclidean distance on a plane fails to generate realistic distances at the urban scale. However, the relation between the Euclidean distance and network distance has not been subject to an extensive empirical study.

In this study, we calculate and compare Euclidean distance and network distance for 10 randomly selected European cities: Berlin, Milan, Rennes, Madrid, Nuremberg, Glasgow, London, Brussels, and Paris. The shortest distances for a total of randomly generated 100 origin and 100 destination points are calculated using network distance and Euclidean distance using network analysis techniques and a GIS-database specifically constructed for this purpose. We use a geographic information system (ArcGIS) together with SANET, Spatial Analysis developed by the SANET Team, for data processing including maps and geographic representations. SPSS is used for the statistical analysis and graphic representation of data.

The rest of this paper is structured as follows. Section 2 presents a theoretical background on distance measurements and its relation with human behavior. Section 3 explains data collection and processing, together with a brief description on each city network. Section 4 presents analysis and discussion of the results. Finally, Section 5 concludes the paper.

2. Theoretical Background

Although spatial distance is a very important concept for a wide variety of disciplines including social, natural, and information sciences, the methods used to measure spatial distance are not directly expressed and fully explained. There are still many unanswered questions regarding the measurement of spatial distance and on the role of spatial distance on human activities (Sander, Ghosh, van Riper, & Manson, 2010).

In almost all quantitative urban models, the including gravity-based models, the urban area is divided into geographical units. The size of these units may differ with respect to the context of the study. In most cases, administrative boundaries such as districts and neighborhoods are utilized. A prerequisite in these models is the distance matrix, showing the distances between the geographical units in the study. The size of the matrix is n x n, where n is the number of geographical units. For a long period of time, these distance matrices are calculated using Euclidean distances between the centers of the pertaining geographical units. Euclidean distance is simply the distance between the points. If the distance Euclidean distance (*ED*) between the two points with the coordinates (x_1, y_1) and (x_2, y_2) is calculated as:

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

This formula is based on the Cartesian coordinate system and it can be used id the study area is small such as a district or a city. However, geodetic distance has to be calculated if the study area is large such as a state or a country (Wang, 2006). Geodetic distance takes into consideration the globe shape of the Earth. In this study, the Euclidean distances based on the Cartesian coordinate system are calculated, as the geographic units considered are at the city level.

In the recent years, measuring distance using spatial networks has gained popularity with the help of rapid spreading of geographic information systems and declining costs of digital spatial data. Levinson & El-Geneidy (2009) argue that simple distance metrics including the Euclidean distance should not be used in planning models, as people take network topology into consideration when making spatial choice decisions. The basic idea behind using network distances is that the Euclidean distance on a plane fails to generate realistic distances at the urban scale. It is assumed that a network embedded on a plane is a better approximation to the real world, where events or geographical units are represented by points on the network. In this approach, the distance between the two points is measured by the shortest-path distance using network analysis (Okabe, Okunuki, & Shiode, 2006).

The detour index (*DI*) is a measure of the amount of detour of the shortest route connecting two points. It is calculated by dividing the shortest network distance (*ND*) between the two points to the shortest distance between them and multiplying by 100. The shortest distance here is the Euclidean distance (*ED*) (Hammond & McCullagh, 1974):

$$DI = \frac{ND}{ED} \times 100$$

This index is called the detour index, as detours have to be made between the two points, assuming that no two points are directly connected (Hammond & McCullagh, 1974). The detour index summarizes the relation between the network distance and the Euclidean distance in a single index. A higher detour index value shows that the network distance deviates from the Euclidean distance at a greater extent, and vice verse. The detour index is used in this study to study the relation between the Euclidean distance and the network distance, as it is scaled by 100 and easier to comprehend and interpret. An alternative to the detour index is the circuity index, which has almost the same formula (Levinson & El-Geneidy, 2009):

$$Circuity = \frac{ND}{ED}.$$

Studies on the relation between Euclidean distance and network distance are limited in number. In an earlier study, Newell (1980) reported that the network distances in an urban area are about 1.2 times the Euclidean distances for a randomly selected set of points. O'Sullivan and Morrall (1996) focus on various transit station catchment areas and argues that the detour index varies between 121 and 123. Levinson & El-Geneidy (2009) report that the circuity, or the detour index, decreases among the randomly selected origin and destination pairs with the increase of both Euclidean and Network distances. It is speculatively argued that the difference is more than 20% when the Euclidean distances are below 400 meters (SANET Team, 2015). Apparicio, Abdelmajid Riva & Shearmur (2008) show that there is strong correlation between the Euclidean distance and network distance across the metropolitan area. This relation is reported to be weaker in suburban areas. These findings indicate that there is a measurable relation between the Euclidean distances and network distances.

3. Data Collection and Processing

In this study, we calculate and compare Euclidean distances and network distances for 10 randomly selected European cities. The cities in the sample include Berlin, Milan, Rennes, Madrid, Nuremberg, Glasgow, London, Brussels, and Paris. First, the digital street network data for each city was obtained in a 15-kilometer radius circular area. The center of this circular area for each city resides at the very city center, namely a famous public square or a public building.

Obviously, each city has specific characteristics, which affects the structure of its network, for that reason we calculated the network characteristics (Table 1), focusing on the edges or street segments (the basic element of the network structure) that would help us understand the physical characteristics of the cities in the sample. As seen in Table 1, the longest average edge length is observed in Berlin, followed by Glasgow and Copenhagen. Whereas, the shortest average edge length is observed in Milan, followed by London and Madrid. Madrid has the longest total line segment length, Rennes the shortest (Table 1). These edges within the study areas for the randomly selected cities are shown in Fig. 1.

Table 1. Descriptive Statistics for the edges (street segments) (in meters).

City	Count	Minimum	Maximum	Mean	Std. Dev.	Sum
Berlin	33,576	1.1367	9,496	251.6	360.3	8,446,566
Brussels	79,638	0.0075	3,573	97.2	124.7	7,743,976
Copenhagen	71,813	0.0100	2,944	117.5	137.0	8,441,147
Glasgow	53,916	0.0063	4,138	118.7	140.6	6,399,733
London	129,257	0.0038	2,219	81.7	87.9	10,557,271
Madrid	149,583	0.0033	3,440	82.0	115.5	12,266,775
Milan	111,031	0.0045	3,162	81.0	113.7	8,991,074
Nuremberg	69,245	0.0030	3,372	104.0	145.6	7,203,802
Paris	134,566	0.0030	3,286	87.3	101.5	11,751,452
Rennes	62,062	0.0029	3,494	94.0	150.1	5,836,883

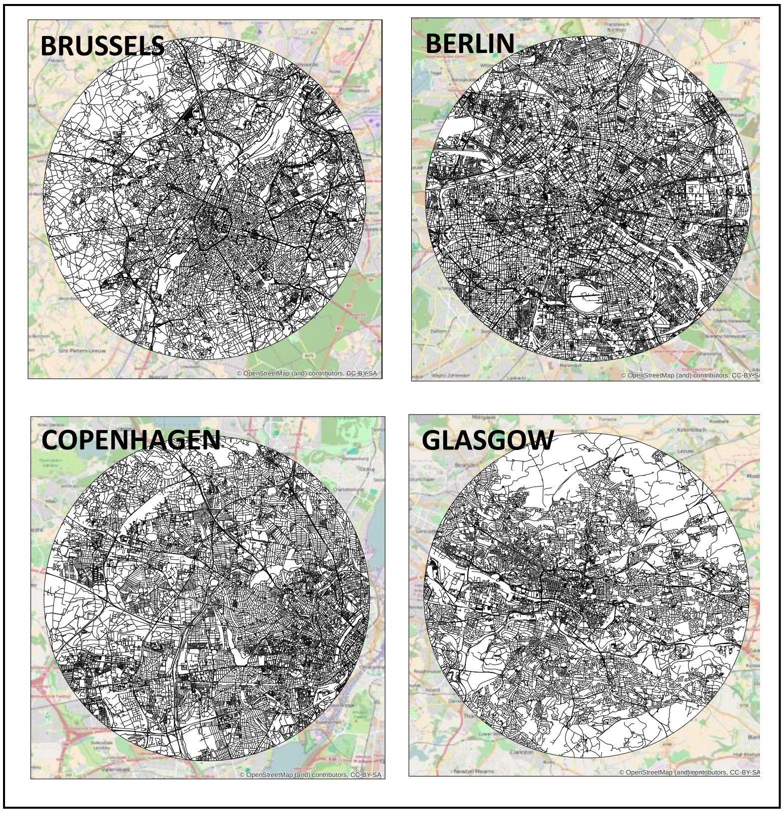


Fig. 1. The edges within the study areas for the cities in the sample.

Table 2. Descriptive Statistics for the Euclidean distances and network distances.

		Euclidean Distance				Network Distance			
City	Count	Minimum	Maximum	Mean	Std. Dev.	Minimum	Maximum	Mean	Std. Dev.
Berlin	90	871.4	17,239.1	9,737.3	3,649.8	1,087.4	25,999.1	14,066.3	5,581.3
Brussels	90	3,039.2	24,340.3	14236.1	5,825.5	3,236.5	27,266.0	16,602.7	6,556.2
Copenhagen	90	2,614.2	16,476.3	8,587.3	3,396.3	3,284.4	24,465.2	13,163.3	5,434.4
Glasgow	90	2,409.1	17,508.7	8873.1	3,735.8	3,329.9	24,149.2	12,066.0	4,947.7
London	90	4,091.3	16,425.3	9,780.4	3,573.7	5,470.9	24,122.0	13,833.1	5,423.9
Madrid	90	4,107.1	19,155.2	10,931.4	3,985.0	4,404.9	22399.1	12,534.2	4,803.4
Milan	90	2,803.7	17,197.6	9,873.3	3,714.1	3,516.5	22,632.5	12,211.0	4,713.4
Nuremberg	90	4,164.7	18,128.0	13,634.9	3,840.7	4,536.7	23,554.0	10,545.3	5,235.3
Paris	90	3,134.4	15,905.1	9,495.8	3,388.9	4,175.5	22,935.5	12,987.1	4,822.7
Rennes	90	3,361.8	17,804.3	10,316.1	3,645.7	4,098.4	24,253.2	12,973.2	4,977.7

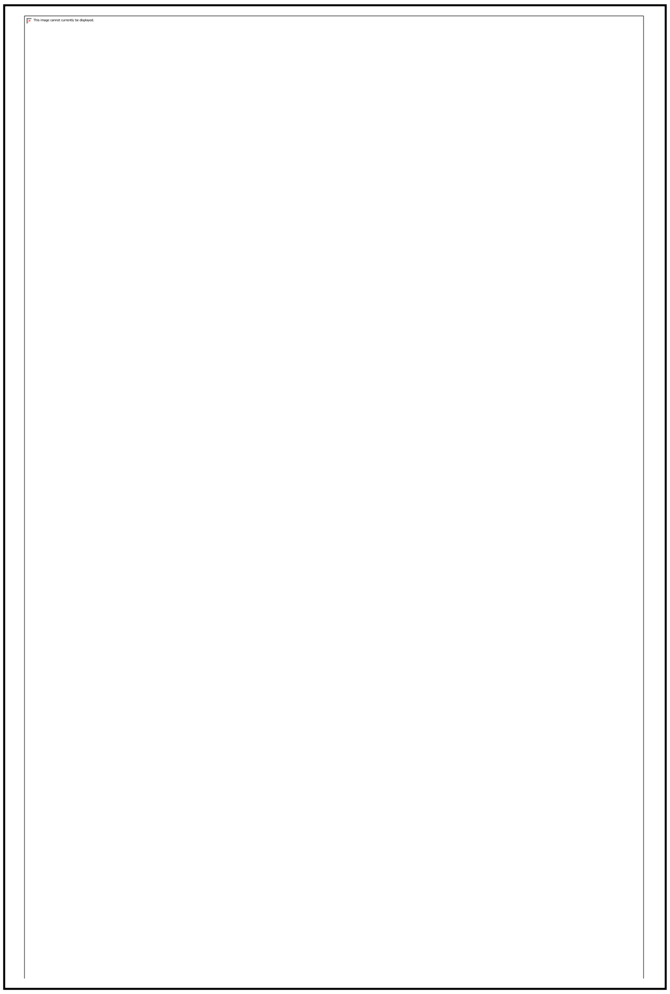


Fig. 1. The edges within the study areas for the cities in the sample.

Using the network database, 10 origin points and 10 destination points are randomly generated in each city, which makes a total of 90 (n x n-1) distance measures for each city, at varying distances from the city center. The shortest distances for a total of 900 origin and destination pairs are calculated using Euclidean distance and network distance. The coordinate system in the study is set as 'WGS 1984 Webercator Auxiliary Sphere, Geographic Coordinate System: GCS-WGS" for all cities, thought the base map by default was from ArcGIS® software by Esri. In Table 2, the descriptive statistics for the Euclidean distances and network distances are given for the 10 cities in the sample.

3. Analysis and Results

As expected, calculated average network distances are longer than the average Euclidean distances for all the cities in the sample. Remark that a network distance can never be shorter than an Euclidean distance. In Table 3, the descriptive statistics for the detour index values are presented. The means for the detour index values pertaining to the 10 cities is presented in Fig 2. As seen in Fig 2, Madrid has the lowest average detour index, as this city has the longest sum of edges and also the highest number of edges, providing a complex network.

The mean values for the detour index pertaining to the 10 cities in the sample are compared using one-way analysis of variance (ANOVA). The results show that the index values for the 10 cities do not belong to the same population. The null hypothesis that these 10 cities have equal means is then rejected. The results are statistically significant at the 0.001 level. The F-statistic for is equal to 159.7.

Table 3. Descriptive Statistics for the detour index values.

City	Count	Minimum	Maximum	Mean	Std. Dev.
Berlin	90	124.8	151.5	143.0	6.8
Brussels	90	100.0	135.0	117.8	8.9
Copenhagen	90	125.6	165.6	151.7	8.7
Glasgow	90	102.0	163.0	137.5	14.2
London	90	113.3	154.3	140.0	7.3
Madrid	90	102.0	125.0	114.0	6.2
Milan	90	101.9	136.4	123.6	8.2
Nuremberg	90	107.3	140.1	128.4	8.2
Paris	90	121.3	149.2	136.2	5.4
Rennes	90	107.0	139.0	124.8	8.4

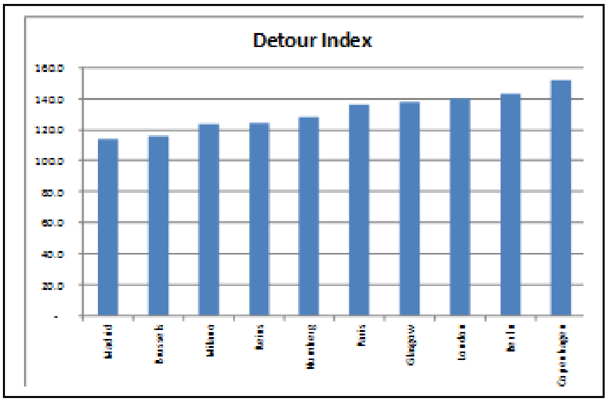


Fig. 2. The average detour index values for the cities in the sample.

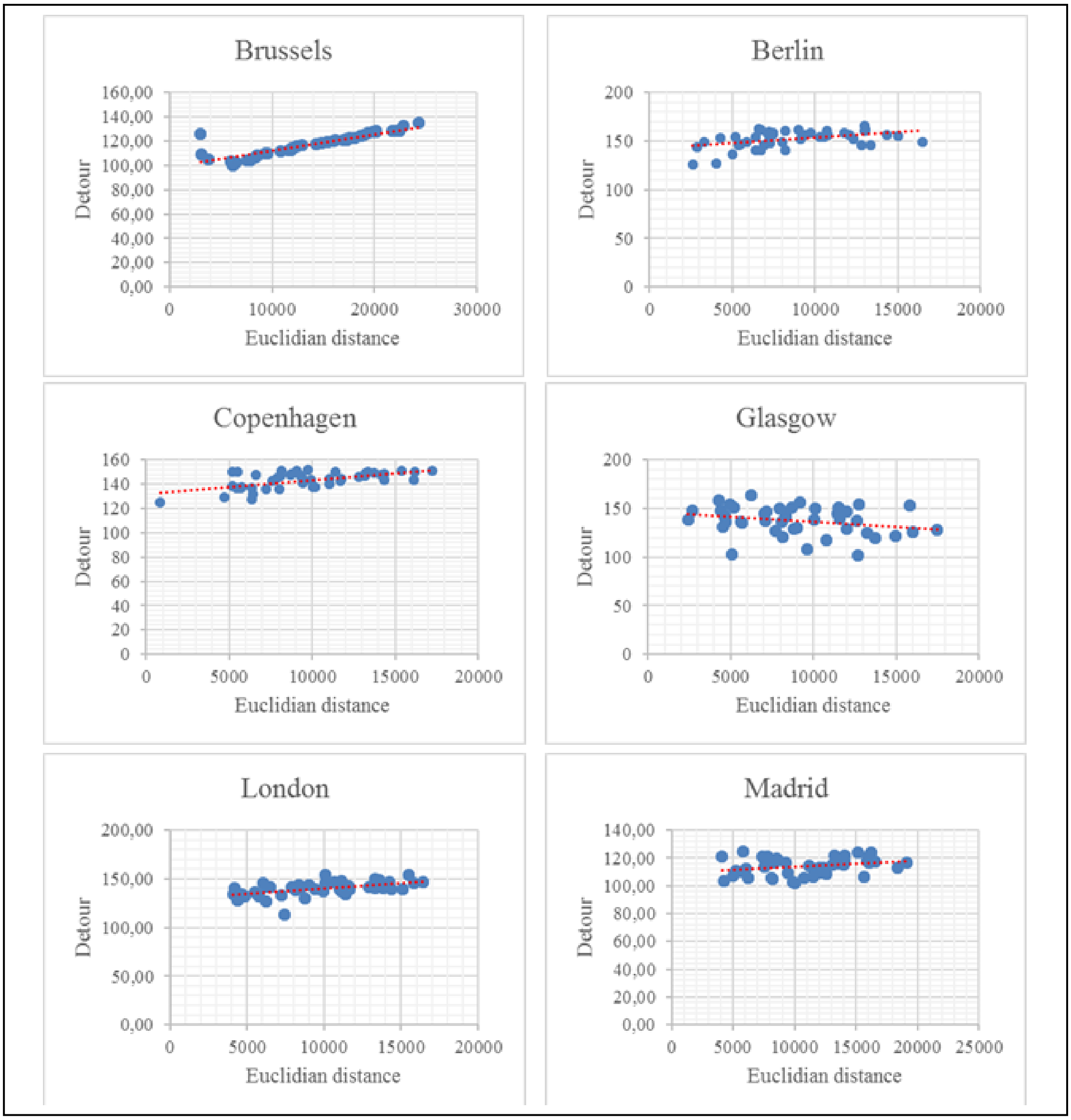


Fig. 3. Detour index versus Euclidian distance (in meters) for the cities in the sample.

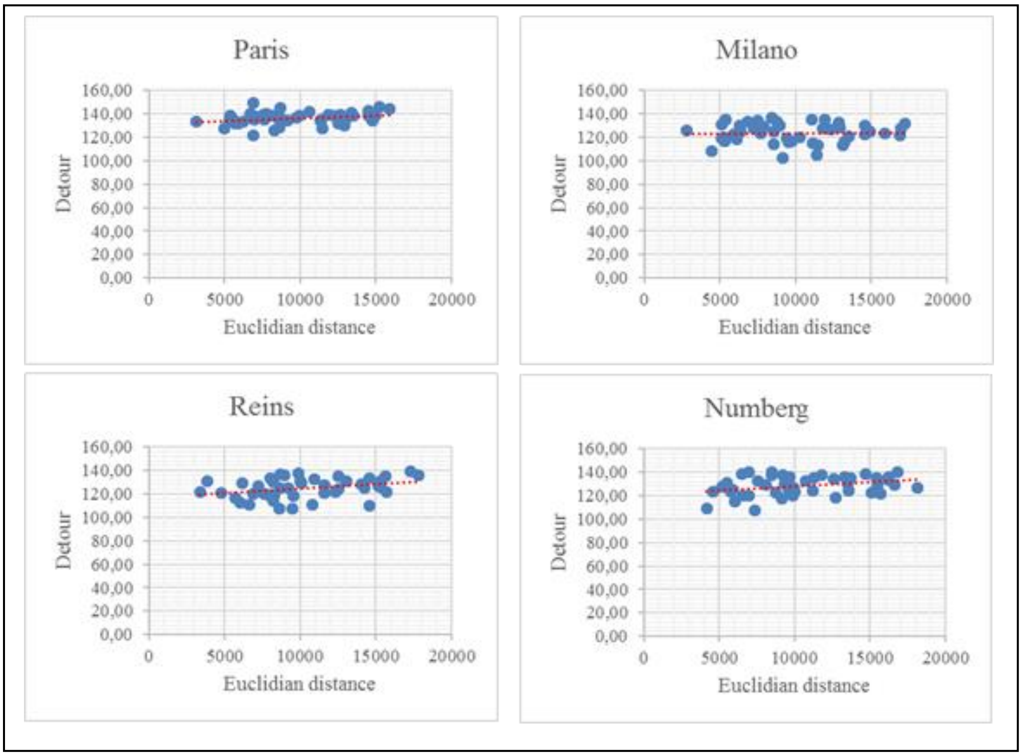


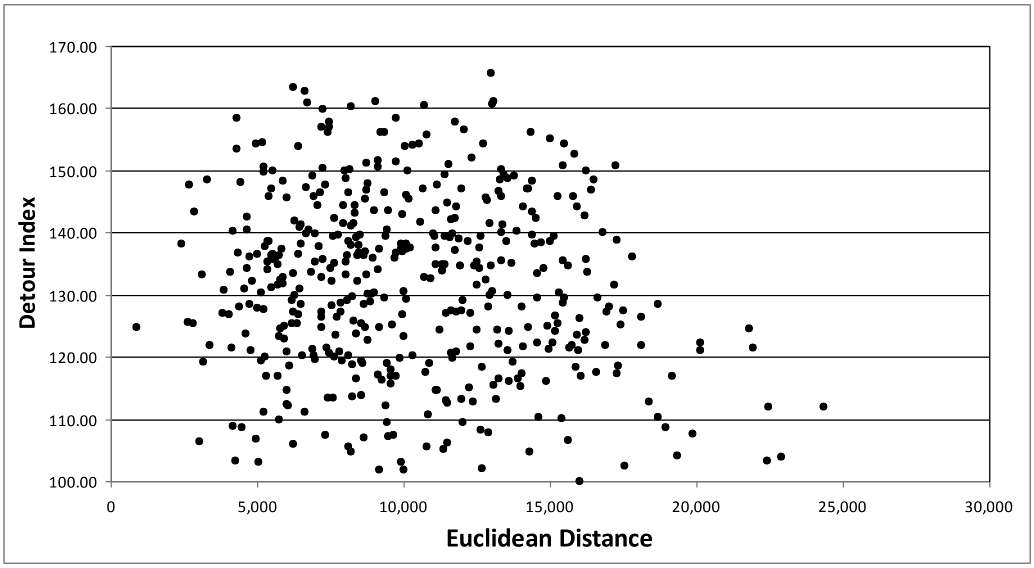
Fig. 3. Detour index versus Euclidian distance (in meters) for the cities in the sample.

In Fig. 3, the scatter plots for the detour index values versus the Euclidean distances are shown with their trend lines for the 10 cities in the sample. In all cities but in Milan and Glasgow, the detour index increases with distances. That is to say, the network distance deviates more from the Euclidean distance for further distances. This finding contradicts with the index values reported in Levinson & El-Geneidy (2009). In Milan, the detour index is indifferent from the distance, indicating a similar network pattern throughout the study area. Whereas in Glasgow, the difference between the Euclidean distance and the network distance narrows as the distance increases.

4. Conclusion

It has long been argued that the Euclidean distance on a plane fails to generate realistic distances at the urban scale. However, the relation between the Euclidean distance and network distance has not been subject to an extensive empirical study. We conclude that there is a relation between the Euclidean distance and network distance. However, this relation is rather local and can be observed at the neighborhood or city level. The detour index values calculated in 10 randomly selected European cities show that the detour index increase when the distance measured increases, implying that the network distance deviates more from the Euclidean distance for longer distances. However, this relation is rather local and specific for each city in the sample. Fig. 4 is a scatter plot showing the Euclidean distances versus the detour index values for all the origin and destination points derived from all the cities in the sample. Clearly, opposite to the patterns observed in city specific graphs in Fig 3, no pattern is observed and the distribution of points has a random distribution. On the contrary to the findings reported in past research, findings of this study reveals that there is not a quantifiable relation or a constant value to represent the relation between the Euclidian distances and network distances. Thus, as suggested in Levinson & El-Geneidy (2009), simple distance metrics including the Euclidean distance should not be used in planning models for better approximations to the real urban activities, and the network distances should be derived through spatial databases and applied in quantitative studies.

Fig. 4. Detour index versus Euclidian distance (in meters) for the distances in the sample.



References

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (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(1),* 1-14.

Hammond, R., & McCullagh, P. (1974). *Quantitative Techniques in Geography: An Introduction*. Oxford: Clarendon Press.

Lee, C. (1973). *Models in Planning: An Introduction to the Use of Quantitative Models in Planning*. Oxford: Pergamon Press.

Levinson, D., & El-Geneidy, A. (2009). The minimum circuity frontier and the journey to work. *Regional Science and Urban Economics, 39(6)*, 732-738.

Novaes, A. G. (2001). *Logística E Gerenciamento Da Cadeia De Distribuição Logistics And Supply Chain Management: Strategy, Operation And Assessment*. 3 ed. Rio de Janeiro: Campus.

Okabe, A., Okunuki, K. I., & Shiode, S. (2006). The SANET toolbox: new methods for network spatial analysis. *Transactions in GIS, 10(4)*, 535-550.

O'Sullivan, S. and Morrall, J., 1996. Walking distances to and from light-rail transit stations. *Transportation Research Record: Journal of the Transportation Research Board, (1538)*, pp.19-26.
Sander, H. A., Ghosh, D., van Riper, D., & Manson, S. M. (2010). How do you measure distance in spatial models? An example using open-space valuation. *Environment and Planning B: Planning and Design, 37(5)*, 874-894.

Team, SANET. (2015). SANET: Spatial Analysis along Networks, User Guide/Manual for SANET Standalone Beta. Tokyo.

Wang, F. (2006). *Quantitative Methods and Applications in GIS*. London: Taylor & Francis.