European Journal of Geography Volume 6, Number 3: 69-78, September 2015

ASSESSING THE IMPACT OF THE GEOGRAPHICAL SCALEON THE MAXIMUM DISTANCE ERROR: A PRELIMINARY STEP FOR QUALITY OF LIFE STUDIES

Paolino DI FELICE
Department of Industrial and Information Engineering and Economics
University of L'Aquila, 67100 L'Aquila, Italy
paolino.difelice@univaq.it

Abstract

Improving citizens' quality of life is a relevant research topic from long time. Several papers emphasized the geographical dimension of such a multidimensional problem. In a sub-category of those studies, the degree of satisfaction of the citizens' needs is obtained by calculating the distance between the centroid of the polygon that expresses the boundary of the administrative district, within the city, where they live and the location of relevant public services in the area. An open issue is up to which geographical scale it is meaningful to push this approach. Our opinion is that the answer to such a question depends on the geography of the administrative units one refers to. So, the only way to know what to do consists in conducting a preliminary computation devoted to investigate the geometrical structure of the administrative units. In connection with this issue, our paper reports the findings of a case study regarding the three administrative levels of Italy.

Keywords: *quality of life, citizens' needs satisfaction, Italian administrative units, centroid, distance measurement error*

1. INTRODUCTION

Quality Of Life (QOL) has a wide range of interpretations as stressed for instance by Sirgy et al. (2006) and Das (2008). In the present empirical study, we refer to the QOL meant as "*citizens' needs satisfaction*", as it is called in (Mohit, 2013). Meet the needs of people contributes to increase their QOL, e.g., (Sirgy et al., 2008). Brereton et al. (2008) found that the impact of public services (namely, hospitals, schools, universities, banks, post offices, ...) on life satisfaction is a function of *distance*. In fact, the measurement of the degree of satisfaction of the citizens' needs is obtained by calculating the distance between the place where they live and the location on the territory of the public services.

(Tesfazghi et al., 2010) and (Brereton et al., 2008) report studies that have used QOL measures of distance of the type mentioned above, at the urban geographical scale. However, in both those studies, authors assume that citizens' dwelling coincides with the centroid of the polygon that expresses the boundary of the administrative district, within the city, where they live. Such an assumption implies that, within the administrative district, *all* the geographic locations where the citizen could actually reside are equivalent to each other. "*This introduces a maximum measurement error equal to the greatest distance between the centroid and the border of the administrative district*.", (Brereton et al., 2008).

Obviously, we can foresee that the value of the maximum measurement error grows as the geographical scale increases. Vice versa, what it is not possible to predict is the entity of the error, for a specific geographic area. Nevertheless, the knowledge of this data represents a mandatory precondition before being able to decide whether studies such as those reported in (Tesfazghi et al., 2010) and (Brereton et al., 2008) can be repeated at geographical scales bigger than the urban one or not. This paper reports about a case study aimed at quantify for Italy the maximum measurement error that arise in case the geographical scale of assessment of the level of citizens' needs satisfaction coincides with one of its three administrative levels, namely: municipal, provincial and regional. In the experiments, we ignore the geographic location of the dwelling of the citizens, as done in (Tesfazghi et al., 2010) and (Brereton et al., 2008).

The numerical results extracted from the case study refer to Italy and, therefore, they can not be exported to other countries, however, the paper has the merit of proposing a methodological-technological framework that can be replicated in studies referring to the territory of other countries all over the world.

2. NOTATIONS

AdmLevel denotes a whole administrative level (e.g., municipalities) of a given geographic area (e.g., Sardinia) subject of the study of QOL of a certain State (e.g., Italy), while *U* denotes the generic (administrative) *unit* in *AdmLevel*. |*AdmLevel*| represents the number of units composing *AdmLevel*, while *GeoU* is the geometry of the boundary of *U*. *GeoU* may be either a (single) polygon or a multipolygon. *NumGeoU* (>=1) denotes the number of polygons composing *GeoU*. C_i, i=1, 2, …, *NumGeoU*, stands for the centroid of the i-th polygon of *GeoU*, while C_{GeoU} denotes the centroid of *GeoU*. In the literature very often the centroid is adopted as an abstraction of a whole adminstrative unit, e.g., (Photis, 2012).

For a given *U*:

a) $d^{U}_{MAX,i}$, i=1, 2, …, *NumGeoU*, is the maximum distance between the centroid C_i and the boundary of the i-th polygons composing *GeoU*;

b) $d^{U}_{MAX,AVG} = \left(\sum_{i=1}^{NumGeoU} d^{U}_{MAX,i}\right) / NumGeoU$ is the value of the average of the distances $d^{U}_{MAX,i}$;

c) d^{U}_{MAX} is the maximum distance between the centroid C_{GeoU} and the boundary of *GeoU*.

For the whole *AdmLevel*:

d) $MAX(d^{AdmLevel}_{MAX,i})$ denotes the maximum distance among the values $d^{U}_{MAX,i}$, when *all* the components *U* in *AdmLevel* are taken into account;

e) $MAX(d_{MAX}^{AdmLevel})$ denotes the maximum distance among the values d_{MAX}^{U}, when *all* the components *U* in *AdmLevel* are taken into account;

e) $d_{MAX,i,AVG}^{AdmLevel} = \left(\sum_{j=1}^{|AdmLevel|} d_{MAX,AVG}^{U}\right) / |AdmLevel|$ denotes the value of the average of the distances defined at point "b)" extended to all the units *U* in *AdmLevel*;

f) $d_{MAX,AVG}^{AdmLevel} = \left(\sum_{j=1}^{|AdmLevel|} d_{MAX}^{U}\right) / |AdmLevel|$ denotes the value of the average of the distances defined at point "c)" extended to all the units *U* in *AdmLevel*.

3. THE ACTIVITIES PLAN

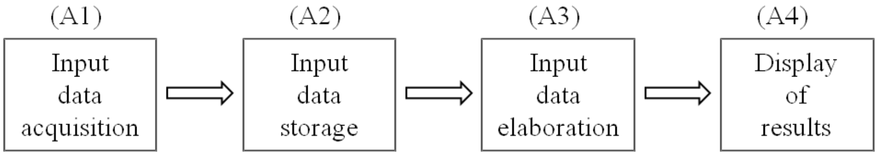

Figure 1. The activities plan.

Fig.1 outlines (at a high level of abstraction) an activities plan for the determination of the maximum distance between the centroid of the administrative units in *AdmLevel* and their border. A brief description of each of them follows.

Activity A1

It concerns the acquisition of data about the geometry of the administrative units in *AdmLevel*, together with their description (i.e., name of the units, their identification code, etc.). The ESRI's shapefile is the more frequently available format for those data.

Activity A2

It implies: a) the design of a Spatial DB (SDB) suitable to accommodate the input data (both the geographical and the descriptive component), b) its implementation (with the SQL `CREATE TABLE` statement); c) loading into the tables of the SDB of the data input (step facilitated by the `import` command offered by the SDBMSs).

Activity A3

This phase returns the values of the following parameters. For a given *U*: $d_{MAX,i}^{U}$, $d_{MAX,AVG}^{U}$, d_{MAX}^{U} and for the whole *AdmLevel*: $MAX(d_{MAX,i}^{AdmLevel})$, $MAX(d_{MAX}^{AdmLevel})$, $d_{MAX,i,AVG}^{AdmLevel}$, $d_{MAX,AVG}^{AdmLevel}$. To carry out these calculations, it is necessary to design ad hoc algorithms devoted to investigate the geometry of the boundary of the administrative units *U* in *AdmLevel*. These algorithms have to be coded and executed. Finally, the results have to be saved on a permanent support. The phases of design and coding of the algorithms, as well as the permanent storage of the results, can be significantly facilitated by making use of the technology of the SDBMSs, in fact such a technology allows us:

- to keep together in a single repository both the geographical and descriptive data about the administrative units in *AdmLevel*. Data often dispersed in several independent files;
- to store in the SDB the voluminous results returned by the processing phase (A3);

- to reduce considerably the manifold difficulties involved in the implementation of activity (A3) by calling the spatial built-in functions of SQL which are a relevant capability of SDBMSs;
- to code ad hoc views and User Defined Functions (UDFs) to be exposed as database objects, both joinable in SQL queries of extraordinary expressiveness.

Activity A4

To simplify the interpretation of the results, it is advisable to show them in terms of charts, tables and maps. Software tools suitable to assist in this phase are: software for manipulating spreadsheets (such as Microsoft Excel able to read **`.csv`** files), the SDBMS (it allows the tabular display of the results), and a geographical viewer (it allows displaying the results as maps).

4. THE CASE STUDY

4.1 Data and methods

Study area and input data sets

The study concerned the three levels of nesting of the Italian administrative units, namely (proceeding from top to bottom) the (20) regions, the (110) provinces, and the 8,094 municipalities. So, *AdmLevel*={*Mun*, *Pro*, *Reg*} where *Mun*, *Pro*, *Reg* denote, respectively, the set of Italian municipalities, provinces and regions.

We downloaded all the input data files from the ISTAT homepage (namely the Italian Institute of Statistics http://www3.istat.it/ambiente/cartografia/versione_non_generalizzata.html),specifically, we acquired the boundaries of the Italian municipalities, provinces, and regions as shapefiles.

The spatial database

The supporting SDB is structured in terms of three tables that fit with the hierarchy of the Italian administrative units:

```
municipality(id, name, provinceId, istatCode, regionId, geom)
province(id, name, regionId, geom)
region(id, name, geom).
```

The database has been implemented in PostgreSQL/PostGIS. As the first step, the data in the ISTAT shapefiles were imported into the corresponding column **`geom`** of the tables above. Fig.2 shows (from left to right) the workflow of the case study.

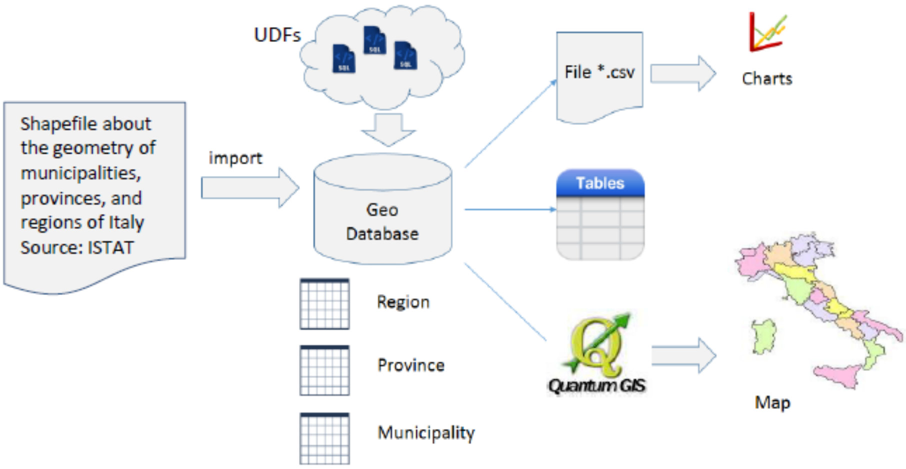

Figure 2. The workflow of the case study.

4.2 Results and discussion

For each of the three administrative levels in *AdmLevel*, we have calculated the values of the seven parameters being part of activity (A3). The data processings were performed by running spatial queries, many of them making use of (PL/pgSQL) UDFs and (SQL) views.

We made recourse to the **ST_Centroid()** function to compute the centroid of the administrative units. **ST_Centroid()** admits as input geometry any shape type (and hence also multipolygons), in spite of the theoretical notion of centroid, and returns the geometric center (computed as the center of the minimum bounding rectangle of the given geometry), as a point. **ST_Centroid()** is largely used in the literature, e.g., (Deakin et al., 2002). The maximum distance between two given geometries been calculated with the **ST_MaxDistance()** PostGIS function.

In the experiments carried out, one of the two geometries is always a point while the other is *GeoU*, therefore, this latter may be either a polygon or a multipolygon. Tab.1, taken from (Di Felice, 2014), tells us in *how many* cases the boundary of the Italian administrative units is a multipolygon instead of a polygon. Moving from municipalities to regions these numbers become more and more somehow amazing. In fact, 50% of the provinces and 75% of the regions fall into such a category. Likely, the same happens in most countries all over the world.

Table 1. Numerical findings about the structure of the boundary of the Italian administrative units.

Municipality	Province	Region
5.9%	50.0%	75.0%
476 (out of 8,094)	55 (out of 110)	15 (out of 20)

Tab.2 and Tab.3 summarize the results of the experiments.

Table 2. The values (in Km) of $d^{AdmLevel}_{MAX,i,AVG}$ and $d^{AdmLevel}_{MAX,AVG}$ for the administrative levels in *AdmLevel*.

	Municipality	Province	Region
$d^{AdmLevel}_{MAX,i,AVG}$	4.9	28.9	36.1
$d^{AdmLevel}_{MAX,AVG}$	5.3	50.9	124.7

From Tab.2 we have confirmation of what expected, in details we see that:

a) the value of parameters $d^{AdmLevel}_{MAX,i,AVG}$ and $d^{AdmLevel}_{MAX,AVG}$ is low for municipalities, while it grows rapidly for provinces and regions. Specifically, the error that is done assimilating the residence of the citizen with the centroid of the administrative unit where he lives is of a few kilometers at the level of municipalities, while it becomes remarkable at the other two levels;

b) the value of parameter $d^{AdmLevel}_{MAX,i,AVG}$ is always below that of $d^{AdmLevel}_{MAX,AVG}$, but the difference between them rises passing from municipalities to regions (0.4Km, 22Km, 88.6Km).

Table 3. The values (in km) of the maximum distance centroid-boundary.

AdmLevel	MAX($d^{U1}_{MAX,i}$)	Name of *U1*	$d^{U1}_{MAX,AVG}$	MAX(d^{U2}_{MAX})	Name of *U2*
Municipality	32.7	Rome	17.7	50.6	Lipari
Province	81.3	Bolzano	81.3	234.4	Agrigento
Region	198.8	Apulia	4.6	279.6	Sicily

Tab.3 shows the values of the maximum distance centroid-boundary, that is, it provides information about the "extreme case" without saying how many times it occurs. For Italy, the value of the maximum error in the extreme cases is huge (up to 198.8Km, column 2), when *all* the components of *all* the administrative units are taken into account. Such errors deteriorate (up to 279.6Km, column 5) if we consider only the "aggregate" geometries of the various administrative units. Few remarks about the results in Tab.3 are necessary to prove their correctness.

The first. The municipality of Rome (Fig.3) is composed of two subareas whose extension is very different (10.9 km^2 vs. 1,276.5 km^2). That is why the value of parameter $d^{U1}_{MAX,AVG}$ is about half that of parameter $d^{U1}_{MAX,i}$ (*U1* denotes Rome).

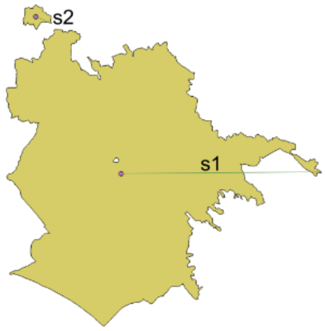

Figure 3. The two components of the municipality of Rome. Segments s1 and s2 visually show the values of the terms $d^{Rome}_{MAX,1}$ and $d^{Rome}_{MAX,2}$. As we can see, s1>>s2.

The second. The province of Bolzano is composed of a single entity, for this reason $MAX(d^{U1}_{MAX,i}) = d^{U1}_{MAX,AVG}$ (where, this time, *U1* denotes Bolzano).

The third. The Apulia region is composed of 42 polygons. The area of one of them is much larger than the remaining areas which, therefore, have a negligible weight compared to the total area of the region. This is the reason of the enormous difference between the values of $MAX(d^{U1}_{MAX,i})$ and $d^{U1}_{MAX,AVG}$ (198.8km vs. 4.6km).

Finally, one can observe that the value of $MAX(d_{MAX,i})$ (Tab.3, second column) is significantly lower than the value of $MAX(d_{MAX})$ (fifth column). This result is valid for any of the three levels in *AdmLevel*.

Fluctuations of the maximum error in function of NumGeoU

Fig.4, 5 and 6 show, in sequence, the fluctuations of the parameters d^{U}_{MAX} (blue), $MAX(d^{U}_{MAX,i})$ (red) and $d^{U}_{MAX,AVG}$ (green) for regions, provinces and municipalities, as a function of the value of *NumGeoU* (the horizontal axis).

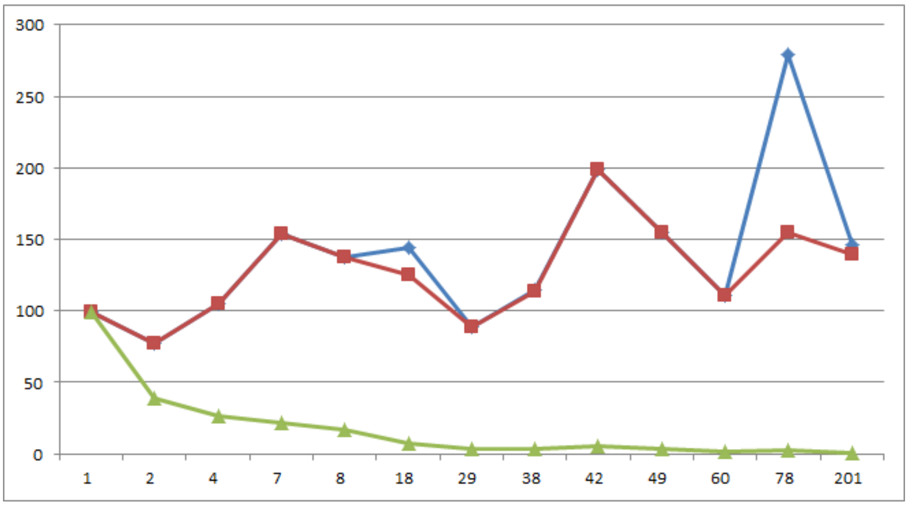

Figure 4. The plot of the maximum error in function of *NumGeoU* for the Italian regions.

Fig.4 has in abscissa thirteen values, as many as the regions that have distinct values of the number of components (Tab.4).

Table 4. The value of *NumGeoU* for the twenty Italian regions.

Region name	*NumGeoU*	Region name	*NumGeoU*
Sardinia	201	Basilicata	4
Sicily	78	Marche	4
Liguria	60	Lombardy	4
Tuscany	49	Umbria	2
Apulia	42	Abruzzo	2
Campania	38	Molise	1
Friuli-Venezia Giulia	29	Veneto	1
Lazio	18	Aosta Valley	1
Calabria	8	Piedmont	1
Emilia Romagna	7	Trentino Alto Adige	1

From Fig.4 we can draw two general conclusions:

a) the value of parameter $d^{U}_{MAX,AVG}$ decreases as the value of *NumGeoU* increases. This happens because most of the polygons that make up the boundary of the Italian regions have very small area and this reduces dramatically the maximum error that we make in approximating the dwelling of the citizen with the centroid of the polygon that contains it;

b) for *all* the Italian regions happens that the value of d^{U}_{MAX} is greater or equal to that of $MAX(d^{U}_{MAX,i})$. The maximum deviation between these two values concerns Sicily (279.6Km vs. 155km).

Fig.5 and Fig.6 referred to, respectively, provinces and municipalities, confirm the previous two conclusions. In addition, it can be observed that the starting value of the parameter $d^{U}_{MAX,AVG}$ (namely that for *NumGeoU* = 1) is smaller than that of the regions (45Km and 5Km, respectively). The reduction is due to the lower extension of the area of the polygons involved.

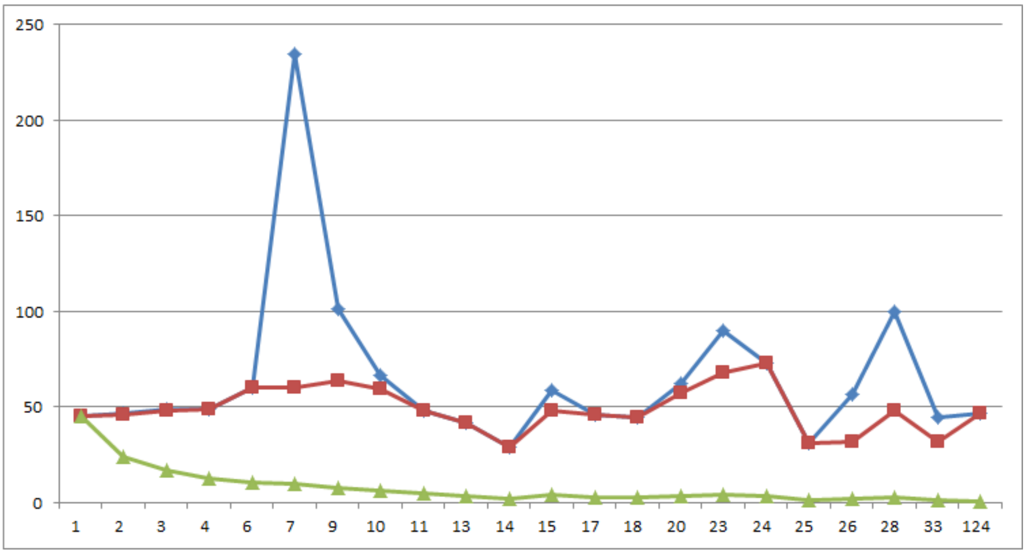

Figure 5. The plot of the maximum error in function of *NumGeoU* for the Italian provinces

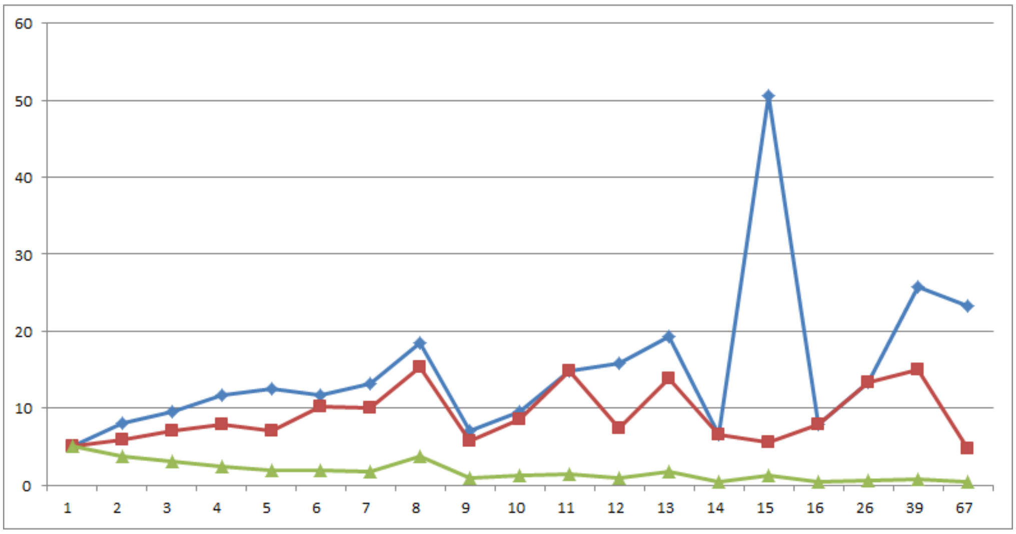

Figure 6. The plot of the maximum error in function of *NumGeoU* for the Italian municipalities.

5. CONCLUSIONS

The case study about Italy enabled us to know the exact extent of the *maximum measurement error* that arise in case the geographical scale of assessment of the level of citizens' needs satisfaction coincides with one of its three administrative levels, namely: municipal, provincial and regional.

Such an investigation represents a preliminary step to be accomplished before being able to decide whether studies such as those reported in (Tesfazghi et al., 2010) and (Brereton et al., 2008) can be started at geographical scales bigger than the urban one they referred to. In the case of Italy, thanks to the outcomes of the case study, we are now able to state that those studies can be repeated at the municipal scale without compromise the correctness of the interpretation of the final results, *but not* at the provincial and regional scale.

The numerical results extracted from the case study refer to Italy and, therefore, they can not be exported to other countries; however, it is worthwhile to point out that the technological framework we have implemented to carry out the experiments reported in the paper, as well as the adopted activities plan, can be replicated to specific geographical areas of other countries all over the world.

REFERENCES

Brereton, F., Clinch, J. P., and Ferreira, S., (2008). Happiness, geography and the environment. *Ecological Economics*, 65(2), 386–396.

Das, D., (2008). Urban quality of life: A case study of Guwahati. *Social Indicators Research*, 88, 297–310.

Deakin, R.E., Bird, S.C., and Grenfell, R.I., (2002). The centroid? Where would you like it to be? *Cartography*, 31 (2), 153–167.

Di Felice, P., (2014). Investigation of the geometry of the Italian administrative units: a preliminary step for quality of life studies. 4th *International Conference on Integrated*

Information, Madrid, Spain. In *AIP Conference Proceedings* Vol. 1644, pp. 78-82 (2015); http://dx.doi.org/10.1063/1.4907820. ISBN: 978-0-7354-1283-5.

Mohit, M.A., (2013). Quality of Life in Natural and Built Environment. An Introductory Analysis. AMER *International Conference on Quality of Life*, Langkawi, Malaysia. In *Procedia - Social and Behavioral Sciences* 101, 33 – 43.

Photis Y. N., (2012), Redefinition of the Greek Electoral Districts through the Application of a Region-Building Algorithm, *European Journal of Geography*, Vol. 3:2, 72-83; http://www.eurogeographyjournal.eu/articles/EJG01206_photis.pdf.

Sirgy, M. J., Michalos, A. C., Ferriss, A. L., Easterlin, R. A., Patrick, D., and Pavot, W., (2006). The quality-of-life (QOL) research movement: past, present and future. *Social Indicators Research*, 76, 343-466.

Sirgy, M. J., Gao, T., and Young, R. F., (2008). How Does Residents' Satisfaction with Community Services Influence Quality of Life (QOL) Outcomes? *Applied Research Quality Life*, 3:81–105. DOI 10.1007/s11482-008-9048-4

Tesfazghi, E. S., Martinez, J. A., and Verplanke, J. J., (2010). Variability of Quality of Life at Small Scales: Addis Ababa, Kirkos Sub-City. *Social Indicators Research*, 98:73–88. DOI 10.1007/s11205-009-9518-6.