LOCATION MODELS FOR MAIN INDUSTRIES AT MACRO-GEOGRAPHICAL LEVEL USING OPEN GEOSPATIAL DATA AND SOFTWARE: CASE OF ROMANIAN COUNTIES

Cornel GHITA
Romanian Academy - Institute for World Economy, Romania
cornel.ghita8@yahoo.com
Gheorghe MILITARU
University "Politehnica" of Bucharest, Romania
ghmilitaru.militaru@gmail.com

Abstract
This paper reveals the mathematical models to explain the location of industries within macro-geographical areas (Romania's counties), by means of GIS (Geographical Information System) metrics. While the explained variable (output) was a composite indicator equally weighting employment and turnover quotients computed from statistical data, the explaining variables (inputs) were a set of GIS metrics, computed on geospatial open data, using open GIS software. The preferred method for modelling was the multiple linear regression; different nonlinear functions were tested to provide the best fit. The GIS metrics are an alternative to statistical data, having the advantage of being procured and updated easier by the automatic import of the GIS database. For all industries, the study delivered relevant models. This study is part of a spatial decision support system (SDSS) for the location of enterprises, including both a macro-geographical layer and a micro-geographical layer of factors. This paper's results can be used independently (the micro-geographical layer was addressed by a previous study). The macro-geographical and the micro-geographical layer could share the 100% weight according to the desired level of location. For example, a governmental organization could grant all weight to the macro layer, while an enterprise could weight more the micro layer.

Key words: *economic and social GIS metrics; geospatial factors; industry location models; location of industries in Romanian counties.*

JEL Classification: *R30; R50; E17.*

I. INTRODUCTION

The goal of this paper is to reveal the results of a secondary research study, based on open geospatial data, extracted from OpenStreetMap GIS (Geographical Information System) database, and on statistical data, extracted from the available territorial statistics. The results consisted of a set of functions, one for each industry, having a composite location indicator as the dependent variable (output), and a set of GIS computed indicators as the independent variables (inputs).

These functions can be used by public authorities, to assess the conditions that favor certain industries in certain counties, and target them through their development policy. Also, when considering a location decision, an entrepreneur in a certain industry can calculate county scores to find out which county is appropriate to accommodate his / her business.

The location factors can be grouped into two layers – a macro-geographical layer and a micro-geographical layer. That means there are also micro-geographical (local) factors that apply to the location decision of an enterprise. Micro-geographical factors were covered by a previous research (Ghiţă, 2014, pp. 101-120). Actually, the results of this research and the results of the previous one can be integrated in a complete Spatial Decision Support System (SDSS) aimed for finding proper business locations for enterprises. Each layer of location factors, both the macro-geographical layer and the micro-geographical layer can be weighted by the decision person according to the importance he / she grants to each layer, and converge to a single and complete location decision.

This paper refers only to the macro-geographical layer of location factors. This layer offers useful insights, because there are statistical data available to calculate the output variable; therefore, a mathematic relationship between the output and the input variables could be determined. Conversely, a relevant sample of micro-geographical data on enterprises, such as geographical coordinates (location), industry, number of employees, and turnover, could not be obtained at the time of the research. While the macro-geographic layer results were based on aggregated statistical data of the whole enterprise population, the micro-geographical results were based on a survey on the location behavior applied to a sample of enterprises.

The next section of the paper describes the research methodology, followed by the results section, where the location functions were revealed and statistically tested. The last section discusses the implications of the findings, limitations and the possibility of integration within a complete SDSS for location decision.

II. METHODOLOGY

The study regions were the counties of Romania, European Union NUTS-III (Nomenclature of Units for Territorial Statistics) administrative level. The reason for territorial segmentation of Romania at this level was the fact that there were 41 counties (Bucharest city and its satellite county - Ilfov were aggregated), which provided a reasonable number of observations, compared to the number of NUTS-II regions (only 8 development regions). Moreover, the development regions are heterogeneous; for example, in the South-Muntenia Region, there is a very big gap in economical welfare and structure between the northern counties, such as Prahova or Argeș, compared to southern counties, such as Giurgiu or Teleorman. The same heterogeneity between counties can be observed across all the development regions in Romania.

The industry panel consisted of the sections of European Industrial Activity Classification (NACE Rev.2), excluding public services and administration. In Romania, education and health are still dominantly owned by public institutions; the data regarding these sections counted only private entities.

Having in mind the goal of building deterministic location models for each main industry (NACE section), this study addressed quantitative factors that could be measured. This kind of location factors can be derived from the classical and neo-classical approaches to location theory, based on the principle of the “economic man”, who acts rationally in self economic benefit. All these (neo) classical location variables derived from literature, that cause uneven distribution of economic activity and income across space (Krugman, 1991, pp. 484–499; Venables, 2008) can be grouped into four main concepts (factors):

- *Access to markets* condensing principles such as transportation cost to the market (Thünen and Hall, 1966; Alonso, 1964), population size and its purchasing power generating a hierarchy of localities and the need for market areas to sustain certain industries (Christaller, 1966; Lösch, 1973);
- *Access to labor force* as an important production factor, inducing distortions on the models based on other production factors (Weber and Friedrich, 1929);
- *The quality of transportation infrastructure* affecting the cost of transport of raw materials and products (Weber and Friedrich, 1929);
- *Specific natural conditions / access to natural resources* inducing the coagulation of land-use areas (Thünen and Hall, 1966).

But these location factors couldn’t be measured as a whole. In order to measure the intensity of these factors, a set of indicators (independent variables) had to be built.

Regarding data sources, as mentioned in the introduction section, the study used two types of data sources:

- *Statistical data* used to compute the *output variable*;
- *Geospatial (GIS) data* used to compute *input variables*.

The output variable

The output was a composite indicator, calculated as the arithmetic average of two indicators:

$$Output = \frac{LQ + TQ}{2} \tag{1}$$

where LQ is the *location quotient* (Baer and Brown, 2006) of the county, and TQ is a *turnover per employee quotient*.

The location quotient (LQ) was calculated as follows:

$$LQ = \frac{e_i / e}{E_i / E} \tag{2}$$

where:

e_i = local (county) employment in industry i;

e = total local employment;

E_i = reference area (Romania) employment in industry i;

E = total reference area employment.

The turnover per employee quotient (TQ) was calculated as follows:

$$TQ = \frac{t_i}{T_i} \tag{3}$$

where:

t_i = local (county) turnover per employee in industry i

T_i = reference area (Romania) turnover per employee in industry i.

Thus, the output equally weighted the concentration of employees and the productivity per employee, for a certain industry in a certain county. The location quotient measures the employment share of the industry in the county, relative to nation-wide industry share, while the turnover quotient measures the turnover per employee of the industry in the county, relative to nation-wide industry turnover per employee.

Actually, the first phase of the study used only the location quotient as output. The results were straight-forward, but the simple fact that an industry had a larger employment share in a county doesn't necessarily mean that the region is better for doing that kind of business, than others. Having in mind the transitional phase of the Romanian economy with the consequence of dis-industrialization, for example, in case of agriculture, forestry and fishing industry, the bigger share of employees in the predominantly plain counties could show the lack of employment alternative and the practice of subsistence agriculture, rather than the attractiveness to entrepreneurs. Then, in the second phase, the location quotient was replaced by the turnover per employee quotient, considering the “economic man” principle, and the desire for economic benefit of the entrepreneurs. In case of turnover per employee, disparities between counties existed, but for many important industries they were not significant enough (not even when squared) to point out standing models. Therefore, in the end, a balanced solution was preferred, to keep both quotients in an equally weighted relation.

The data concerning employment and turnover for each county and industry were collected from territorial statistics (Romanian National Institute of Statistics, 2013).

The input variables

The input variables were exclusively GIS metrics (calculated by means of geospatial computing). In order to make these geospatial calculations possible, first, it was necessary to build the GIS infrastructure and to import the GIS database, then to prepare database queries to extract GIS information having the proper format and content.

The *GIS infrastructure* consisted of a computer, with the necessary software installed to perform geospatial computing. The *GIS database* was acquired from OpenStreetMap (OpenStreetMap contributors, 2013). This study used only open GIS data and software. The procedure of installing the required software and database import is described in Appendix A. The *GIS database queries* used to extract the required information for the calculation of GIS metrics, are depicted in Appendix B.

In correlation with the concepts (factors) mentioned in the previous section, the GIS metrics were designed to measure features regarding population (number, density, income, education), transportation network (road and railway density), and natural conditions (land-uses). A number of 20 GIS metrics were computed for each county:

1. Urban area quotient (UAQ);
2. Residential area quotient (RAQ);
3. Urban-residential area quotient (URAQ);
4. Residential area in the main city quotient (RAMCQ);
5. Commercial area quotient (CAQ);
6. Density of buildings quotient (DBQ);
7. Density of shops quotient (DSQ);
8. Density of amenities quotient (DAQ);
9. Density of universities quotient (DUQ);
10. Density of main roads quotient (DMRQ);
11. Density of roads quotient (DROQ);
12. Density of railways quotient (DRAQ);
13. Farmland area quotient (FAQ);

14. Forest area quotient (FOQ);
15. Meadow area quotient (MAQ);
16. Vineyard area quotient (VAQ);
17. Orchard area quotient (OAQ);
18. Grass area quotient (GAQ);
19. Industrial area quotient (IAQ);
20. Quarry area quotient (QAQ).

The definition and details of land-uses can be consulted on OpenStreetMap wiki website (OpenStreetMap Wiki contributors, 2013).

The GIS metrics were computed by the same principle of the location quotient, but replacing the employment shares with the measured indicator shares. For example, urban area quotient (*UAQ*) was computed according to the following formula:

$$UAQ = \frac{u_j / a_j}{U / A} \qquad \textbf{(4)}$$

where:

u_j = urban area in county *j*;

a_j = total area of county *j*;

U = total urban area in Romania;

A = total area of Romania.

These GIS metrics are an alternative to statistical data, with an advantage in case of a business location SDSS, namely the fact that the metrics could be updated automatically, with every database update. The database update is an all purposes operation, necessary for the relevance of the entire SDSS that could also be made automatically, so there would be no additional effort for updating GIS metrics alone. The only statistical data that should be updated manually are those used to compute the output variable, i.e. employment and turnover data.

Regression analysis

Multiple linear regression was the preferred method for drawing the functions out of existing data. Calculations were made with Microsoft Office Excel 2010 (Harmon, 2011) using the standard data analysis tools, such as Scatter Plots charts, correlation, regression, and Trendline. Before performing the regression, a correlation analysis was undertaken, to select the relevant input variables.

First, in order to avoid multi-collinearity of input variables (that would result in coefficients with opposite signs to the correlations signs) a correlation analysis of the input variable between themselves was performed. This showed that the GIS metrics from 1 to 12, and 19 were in very strong positive correlation (all correlation coefficients were greater than 0.7). This can be explained by the fact that population income, transportation infrastructure quality, urbanization, and population concentration are interdependent. The GIS metric number 19 (industrial area quotient) was rather surprising though, but when looking at the industrial areas in the cities, it showed up that they consisted mostly of the former industrial sites in major cities, but during last 15-20 years, due to dis-industrialization process, the industrial facilities were replaced by or transformed into commercial sites (malls, hypermarkets etc.). Moreover, grass area quotient was positively correlated with forest and orchard quotients, and negatively correlated with farmland, though in a weaker relation (correlation coefficients around 0.5). These findings limited the possible combinations of many explainable variables that would add extra relevance to the models. Anyway, computing so many correlated variables was not in vain, because they enabled the selection of the best correlated variable.

Second, a correlation analysis of each input variable with the output variable was performed in order to select the most correlated input variables, from the groups of collinear variables identified in the first phase.

When data showed a very strong correlation between the output and one of the selected input variables, but weak correlation with the rest of variables, other types of nonlinear functions were tested to find the best fit. Also, to obtain the best out of regression analysis, when statistical tests were not good enough, the most relevant factors were grouped in composite input variables, by multiplication, squares, average and other arithmetic procedures.

III. RESULTS

The results consisted of the mathematical functions for each industry (NACE sections, excluding public administration and services, NGO's, and households) explaining the location output. The multiple linear regression revealed the following results, as shown in **Error! Reference source not found.**.

Table 1. Mathematic models for the location of the main industries

Industry	Mathematic function of the output variable	R^2	Adjusted R^2	Significance of F	p-values
Agriculture, forestry and fishing	0.867 – 0.081*DBQ + 0.497*FAQ	0.498	0.472	0.000002	Y-intercept = 0; DBQ = 0.0013; FAQ = 0.00003
Mining and quarrying	0.358 + 0.237*DBQ + 0.237*GAQ + 0.315 *QAQ	0.677	0.651	0.00000003	Y-intercept = 0.057; DBQ = 0.0003; GAQ = 0.079; QAQ = 0
Manufacturing	0.715 + 0.218*DRAQ + 0.093*OAQ – 0.024*QAQ	0.556	0.520	0.000001	Y-intercept = 0; DRAQ = 0.000003; OAQ = 0.003; QAQ = 0.048
Electricity, gas, steam and air conditioning supply	0.714 + 0.151*DBQ – 0.195*MAQ + 0.081*QAQ	0.559	0.523	0.000001	Y-intercept = 0.000003; DBQ = 0.000005; QAQ = 0.0003; MAQ = 0.08
Water supply, sewerage, waste management and remediation activities	0.791 + 0.388*DRAQ – 0.183*MAQ	0.375	0.342	0.0001	Y-intercept = 0.00007; DRAQ = 0.0004; MAQ = 0.096
Construction	0.754 + 0.129*UAQ	0.882	0.879	0	Y-intercept = 0; UAQ = 0
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.742 + 0.103*RAMCQ	0.949	0.947	0	Y-intercept = 0; RAMCQ = 0
Transportation and storage	0.304 + 0.587*DMRQ	0.644	0.635	0	Y-intercept = 0.001; DMRQ = 0
Accommodation and food service activities	0.857 + 0.061*DAQ	0.533	0.521	0	Y-intercept = 0; DAQ = 0
Information and communication	0.481 + 0.079*CAQ	0.934	0.932	0	Y-intercept = 0; CAQ = 0
Financial and insurance activities	0.633 + 0.053*DUQ	0.927	0.925	0	Y-intercept = 0; DUQ = 0
Real estate activities	0.582 + 0.124*DAQ	0.905	0.902	0	Y-intercept = 0; DAQ = 0
Education [1)]	–0.013 + 0.029*URAQ	0.988	0.987	0	Y-intercept = 0; URAQ = 0
Human health and social work activities [1)]	–0.0018 + 0.0184*RAMCQ	0.985	0.985	0	Y-intercept = 0.24; RAMCQ = 0
Other service activities	0.871 + 0.082*RAMCQ	0.541	0.529	0	Y-intercept = 0; RAMCQ = 0

[1)] Includes only private organizations

Source of statistical data: Romanian National Institute of Statistics (Romanian National Institute of Statistics, 2013)
Source of geospatial data: OpenStreetMap (Geofabrik GmbH, OpenStreetMap Contributors, 2015)

In case of *agriculture, forestry and fishing* the explaining variables were the farmland area quotient (FAQ), understanding farmland mainly as arable land, with a correlation coefficient of 0.583, and the density of buildings quotient (DBQ), with a negative correlation coefficient of -0.448. Statistical tests' results were acceptable, to be considered a relevant model.

The function for *mining and quarrying* industry had good significance tests, the output being positively related to quarry area quotient (QAQ), grass area quotient (GAQ), and density of buildings quotient (DBQ). Though, QAQ is an effect of mining and quarrying, rather than an explaining variable. The GAQ suggest a correlation with hills and mountains, where the most of grass areas are. Anyway, mining and quarrying is an industry open in general to big companies, which are not the main target of the SDSS. A more relevant model could be obtained by overlaying a natural resources GIS layer, but these data were not open.

The results for *manufacturing* show positive correlation with density of railway quotient (DRAQ) with a coefficient of 0.625, a rather strange positive correlation with orchard area quotient (OAQ), coefficient 0.361, and a weaker negative correlation with the quarry area quotient (QAQ).

In case of *electricity, gas, steam and air conditioning supply*, even if the significance tests were not excellent, the input variables explaining this model have an economic logic, i.e. the density of buildings quotient (DBQ), and the quarry area quotient (QAQ) are connected to coal quarries (after all, coal is used to produce energy to heat and power buildings).

For the industry of *water supply, sewerage, waste management and remediation activities*, the significance tests recommend caution for the application of this model. The best correlated inputs were density of railway quotient (DRAQ), correlation coefficient 0.572, and meadow area quotient (MAQ) with a negative correlation coefficient -0.358. Due to the effort of public authorities during last years to bring utilities to all households, even to those in the countryside, this industry is rather dispersed across the whole country.

The results for *construction* industry provided only one relevant input variable, i.e. the urban area quotient (UAQ), a pure urbanization variable, but the correlation was very strong (coefficient 0.939). Other nonlinear functions were tested and an even better fit was found, the following quadratic function: $Output = -0.0043*UAQ^2 + 0.2101*UAQ + 0.6883$, with $R^2 = 0.8884$. The smooth convex parabola of this function has the logic that first, constructions record increasing output growth to constant growth in UAQ, and then, when regions become overcrowded, they tend to decrease growth and even decline.

In case of this *wholesale and retail trade, repair of motor vehicles and motorcycles*, as well, there is only one but very strong correlated input variable, namely the residential area in the main city quotient (RAMCQ), which stands in accordance with economic logic. RAMCQ measures the concentration of population in a major city. The other nonlinear functions were tested, though with no significant improvement in the R^2.

Again, in case of *transporting and storage* industry, the best correlated input variable, the density of main roads quotient (DMRQ) stands also in accordance with the economic logic. This quotient was based on the density (length per km^2) of roads belonging to the following OpenStreetMap categories: motorway, motorway link, trunk, trunk link, primary, and primary link.

The output for *accommodation and food service activities* recorded the best correlation with the density of amenities quotient (DAQ), having a correlation coefficient of 0.730. The amenities (points of interest) are logically connected with the location of this industry.

For the location of *information and communication* industry, the best correlated input variable was commercial area quotient (CAQ), with a correlation coefficient of 0.966. Logically, the most relevant input variable should have been the residential area in the main city quotient (RAMCQ), according to the hypothesis of the location preference of this industry for major cities. Moreover, RAMCQ was very close, recording a correlation coefficient of 0.961. But CAQ was preferred due to the fact that it was also belonging to the same collinear variables group addressing population (density, urbanization, income, and education), but unlike RAMCQ, it put some weight on population income as well.

Regarding *financial and insurance activities*, the best correlated of the input variables addressing population was the density of universities quotient (DAQ), with a correlation coefficient of 0.963. Again, the most logical connection, namely with the residential area in the main city quotient (RAMCQ) was close (correlation coefficient 0.957), but DAQ was preferred for a plus of accuracy.

For the *industry of real estate activities* the output that recorded the best correlation was with the density of amenities quotient (DAQ), with a correlation coefficient of 0.951.

In case of *education* the results showed an almost perfect correlation of the location output with the urban-residential area quotient (URAQ), with a correlation coefficient of 0.994. This quotient is more relevant than the simple urban area quotient, because often the administrative territories of towns include large areas with other land-uses than residential, such as industrial or farmland.

In respect of *human health and social work activities*, the residential area in the main city quotient (RAMCQ) was the best correlated input variable, with a correlation coefficient of 0.993. The high p-value of Y-intercept of 0.24 could be explained by the low value of Y-intercept (close to 0), thus, its variance would not imply a significant variance in the output.

The industry of *other services activities* also showed a preference for major cities, depending on residential area in the main city quotient (RAMCQ), but because of the heterogeneity of sub-industries included in it, the statistical tests of the results were not so good, but they were acceptable.

IV. CONCLUSION

The conclusion of the results is that the models are relevant for all industries. The determination coefficient R^2 was around or over 0.5, except for one industry, namely water supply, sewerage, waste management, and remediation activities. In this particular industry case, the model could be applied, but with caution (should be validated by variables data). Conversely, in most of the services industries, R^2 was over 0.9, showing a preference for counties with the population concentrated in major cities.

Moreover, the general economic logic of the correlations of the output with the input variables validate the input variables selection, and the accuracy of the SQL queries used to extract relevant data from the GIS database. The location models for the main industries (NACE sections) were computed for the counties of Romania. The GIS metrics used herein as input variables could be used with no alteration to other study areas (e.g. other countries, whole European Union), but the output functions should be re-computed, because the relationships between output and input variables might change. The GIS metrics are good alternatives to statistical data, having the advantage of being procured and updated easier by automatic import of the GIS database, having the SQL queries at hand, that and can be built in the form of SQL views (stored SQL queries).

The output variable combined two main aspects regarding the location decision: the employment figures on one hand, and the productivity per employee in terms of turnover, on the other hand. Employment is a rather macro-economic target, while turnover is more relevant to enterprises.

The location models can be used to compute scores for each county, in order to find the most suitable options for the location decision of a single enterprise, or for public institutions to find the best county candidates for industrial policies. This is an important issue; the re-industrialization of Romania is strongly debated now at institutional level.

The administrative territory of Romania includes 41 counties, plus the city of Bucharest (that is 42 administrative entities). The first attempt to draw the location models used all 42 observations, but it wasn't very successful. The noise was coming from the fact that in case of most industries the input in Bucharest generated the output in Ilfov. For example, the huge urbanization and population income figures in Bucharest boosted constructions and real estate in Ilfov. The key decision for obtaining relevant models was to aggregate Bucharest and its satellite county – Ilfov, into a single observation. If all other counties followed the geographical model of a central major city and its surrounding area, there was no point in separating Bucharest from its satellite surrounding county.

The results of this study are part of a complete solution spatial decision support system (SDSS) for the location of enterprises, but the macro-geographical models can be used independently as well. The SDSS include two layers of location factors, a macro-geographical layer to discern between large geographic objects (e.g. counties) on one hand, and a micro-geographical layer to assess the local business conditions at street and building level, on the other hand. The macro layer and the micro layer share the 100% weight, according to the desired level of location. For example, a governmental organization could grant all weight to the macro layer, while an enterprise could weight more the micro layer.

APPENDIX A: GIS INFRASTRUCTURE AND DATABASE IMPORT

These operations were performed in the following sequence (OpenStreetMap and contributors, 2013):

1. Installed Linux operation system – Ubuntu 14.04 LTS (Ubuntu Documentation Team, 2014)
2. Installed the database service – PostgreSQL version 9.3 (The PostgreSQL Global Development Group, 2014) and its GIS extension – PostGIS version 2.1 (OSGeo - Open Source Geospatial Fundation, 2014), from packages with all required dependencies:

```
sudo apt-get install postgresql postgresql-contrib postgis postgresql-9.3-postgis-2.1
```

3. Created a database named *gis* and a user:

```
sudo -u postgres –i
createuser username # answer yes for superuser (although this isn't strictly necessary)
createdb -E UTF8 -O username gis
exit
```

4. Created Unix user and assigned to the created database:

```
sudo useradd -m username
sudo passwd username
```

5. Set up PostgreSQL database with PostGIS:

```
sudo -u postgres psql
\c gis
CREATE EXTENSION postgis;
ALTER TABLE geometry_columns OWNER TO username;
ALTER TABLE spatial_ref_sys OWNER TO username;
\q
```

6. Downloaded *planet.osm.pbf* file for Romania; there are websites that maintain regional files, such as country extracts (OpenStreetMap and contributors, 2006), (Geofabrik GmbH and OpenStreetMap Contributors, 2015):

```
sudo mkdir /usr/local/share/maps/planet/Romania
cd /usr/local/share/maps/planet/Romania
sudo wget http://download.geofabrik.de/europe/romania-latest.osm.pbf
```

7. Installed *osm2pgsql* – a software tool for importing *planet.osm.pbf* files to a PostgreSQL / PostGIS database

```
sudo apt-get install osm2pgsql
```

8. Imported the *planet.osm.pbf* file to local database using *osm2pgsql*:

```
osm2pgsql --slim -d numebdg -C 3000 --number-processes 3 romania-latest.osm.pbf
```

APPENDIX B: GIS DATABASE SQL QUERIES

County view, including geometry and size (area) in Stereo 70 projection (this projection is used by cadaster services in Romania):

```
CREATE VIEW bl_judete AS SELECT name, way,
   ST_Area(st_transform(way,31700))/1000000 As kmp
   FROM planet_osm_polygon
 WHERE planet_osm_polygon.boundary = 'administrative' AND
 planet_osm_polygon.admin_level = '4';
```

Major land-uses totalizing over 100 km^2 nation-wide:

```
SELECT landuse,
    sum(st_area(st_transform(way, 31700)) / 1000000::double precision) AS kmp
  FROM planet_osm_polygon
  GROUP BY landuse
  ORDER BY kmp DESC;
```

This query returned 9 major land-uses, as shown in **Error! Reference source not found.**:

Table 2. Major land-uses in Romania

LAND-USE	KM^2	% in TOTAL nation-wide
farmland	102532	43.09
forest	70350	29.51
meadow	25863	10.85
residential	13555	5.69
vineyard	3703	1.55
orchard	3689	1.55
grass	3048	1.28
industrial	1344	0.56
quarry	251	0.11
TOTAL land-use	**224532**	**94.19**
TOTAL nation-wide	**238391**	

GIS metrics (input variables queries) by county:

1. Urban area by county:

```
CREATE VIEW bl_judete_urban AS SELECT b.name,
    sum(st_area(st_transform(p.way, 31700)) / 1000000::double precision) AS kmp,
    st_union(st_intersection(p.way, b.way)) AS way
  FROM planet_osm_polygon p JOIN bl_judete b ON st_intersects(p.way, b.way)
  WHERE p.admin_level = '6'::text AND (p.place = ANY (ARRAY['city'::text, 'town'::text])) GROUP
BY b.name ORDER BY b.name;
```

2. Residential area:

```
SELECT j.name,
    sum(st_area(st_transform(st_intersection(p.way, j.way), 31700)) / 1000000::double precision)
AS kmp,
    st_union(st_intersection(p.way, j.way)) AS way
   FROM planet_osm_polygon p
     JOIN bl_judete j ON st_intersects(p.way, j.way)
  WHERE p.landuse = 'residential'::text
  GROUP BY j.name ORDER BY j.name;
```

3. Urban-residential area:

```
SELECT a.name,
    sum(st_area(st_transform(st_intersection(a.way, b.way), 31700)) / 1000000::double precision)
AS kmp,
    st_union(st_intersection(a.way, b.way)) AS way
   FROM bl_judete_urban a
     JOIN bl_judete_residential b ON a.name = b.name
  GROUP BY a.name ORDER BY a.name;
```

4. Residential area in the main city:

```
SELECT a.name, sum(st_area(st_transform(st_intersection(a.way, b.way), 31700)) / 1000000::double
precision) AS kmp
   FROM planet_osm_polygon a
```

```
    JOIN planet_osm_polygon b ON st_intersects(a.way, b.way)
  WHERE a.admin_level = '6'::text AND a.place = 'city'::text AND b.landuse = 'residential'::text
  GROUP BY a.name ORDER BY a.name;
```

5. Commercial area:

```
SELECT j.name, sum(st_area(st_transform(st_intersection(p.way, j.way), 31700)) / 1000000::double
precision) AS kmp
   FROM planet_osm_polygon p
     JOIN bl_judete j ON st_intersects(p.way, j.way)
  WHERE p.landuse = ANY (ARRAY['comercial'::text, 'commercial'::text, 'retail'::text])
  GROUP BY j.name ORDER BY j.name;
```

6. Number of buildings:

```
SELECT b.name, count(a.osm_id) AS buildings
   FROM planet_osm_polygon a
     JOIN bl_judete b ON st_intersects(a.way, b.way)
  WHERE a.building IS NOT NULL
  GROUP BY b.name ORDER BY b.name;
```

7. Number of shops:

```
SELECT b.name, count(a.osm_id) AS shops
   FROM planet_osm_point a
     JOIN bl_judete b ON st_intersects(a.way, b.way)
  WHERE a.shop IS NOT NULL
  GROUP BY b.name ORDER BY b.name;
```

8. Number of amenities:

```
SELECT b.name, count(a.osm_id) AS amenity
   FROM planet_osm_point a
     JOIN bl_judete b ON st_intersects(a.way, b.way)
  WHERE a.amenity IS NOT NULL GROUP BY b.name ORDER BY b.name;
```

9. Number of universities:

```
SELECT b.name, count(a.osm_id) AS university
   FROM planet_osm_point a
     JOIN bl_judete b ON st_intersects(a.way, b.way)
  WHERE a.amenity = ANY (ARRAY['university'::text, 'college'::text])
  GROUP BY b.name ORDER BY b.name;
```

10. Length of main roads:

```
SELECT j.name,
    sum(st_length(st_transform(st_intersection(p.way, j.way), 31700)) / 1000::double precision)
AS km
   FROM planet_osm_line p
     JOIN bl_judete j ON st_intersects(p.way, j.way)
  WHERE p.highway = ANY (ARRAY['motorway'::text, 'motorway_link'::text, 'trunk'::text,
'trunk_link'::text, 'primary'::text, 'primary_link'::text])
  GROUP BY j.name ORDER BY j.name;
```

11. Length of roads = Length of main roads (10) + Length of secondary roads. Length of secondary roads was computed as follows, excluding residential streets:

```
SELECT j.name, sum(st_length(st_transform(st_intersection(p.way, j.way), 31700)) / 1000::double
precision) AS km
   FROM planet_osm_line p
     JOIN bl_judete j ON st_intersects(p.way, j.way)
  WHERE p.highway = ANY (ARRAY['secondary'::text, 'secondary_link'::text, 'tertiary'::text,
'tertiary_link'::text])
  GROUP BY j.name ORDER BY j.name;
```

12. Length of railways:

```
SELECT j.name,
    sum(st_length(st_transform(st_intersection(p.way, j.way), 31700)) / 1000::double precision)
AS km
   FROM planet_osm_line p
     JOIN bl_judete j ON st_intersects(p.way, j.way)
  WHERE p.railway = 'rail'::text GROUP BY j.name ORDER BY j.name;
```

13. Farmland area:

```
SELECT j.name, sum(st_area(st_transform(st_intersection(p.way, j.way), 31700)) / 1000000::double
precision) AS kmp
   FROM planet_osm_polygon p
     JOIN bl_judete j ON p.landuse = 'farmland'::text
  WHERE st_intersects(p.way, j.way) GROUP BY j.name ORDER BY j.name;
```

14-20. The rest of GIS metrics were computed similar to 13, but replacing *landuse = 'farmland'* with *'forest', 'meadow', 'vineyard', 'orchard', 'grass', 'industrial',* and *'quarry'*.

V. ACKNOWLEDGMENT

This paper has been financially supported within the project entitled "**Horizon 2020 - Doctoral and Postdoctoral Studies: Promoting the National Interest through Excellence, Competitiveness and Responsibility in the Field of Romanian Fundamental and Applied Economic Research**", contract number POSDRU/159/1.5/S/140106. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. **Investing in people!**"

VI. REFERENCES

1. Alonso, W. (1964) *Location and Land Use: Towards a General Theory of Land Rent.* Cambridge, Massachusetts: Harvard University Press.
2. Baer, C., Brown, T. (2006) *Location quotients: A tool for comparing regional industry compositions*. Advanced Economic and Market Analysis Group, Strategic Research and Development, Indiana Department of Workforce Development.
3. Christaller, W. (1966). *Central Places in Southern Germany* (translated by Baskin, Charlisle W. after Christaller, W. *Die Zentralen Orte in Süddeutschland*, 1966). Englewood Cliffs, NJ: Prentice Hall.
4. Geofabrik GmbH, OpenStreetMap Contributors. (2015) *Download.geofabrik.de (OpenStreetMap Data Extracts)*. http://download.geofabrik.de/europe/romania-latest.osm.pbf, accessed May 10, 2015.
5. Ghiță, C. (2014) *A Decision Support System for Business Location Based on Open GIS Technology and Data*. Managing Global Transitions, 12(2), pp. 101-120, http://www.fm-kp.si/zalozba/ISSN/1581-6311/12_101-120.pdf.
6. Harmon, M. (2011) *Advanced regression in excel.*
7. Krugman, P. (1991) *Increasing returns and economic geography*. Journal of Political Economy, 99, pp. 484–499.
8. Lösch, A. (1973) *The Economics of Location* (2nd edn., revised, translated by Woglom, W. H. and Stolper, W. F. after Lösch, A. *Die räumliche Ordnung der Wirtschaft*). New Haven: Yale University Press.
9. OpenStreetMap and contributors. (2006) *Planet.osm.* http://wiki.openstreet map.org/wiki/Planet.osm, accessed April 10, 2015.
10. OpenStreetMap and contributors. (2013) *Manually building a tile server (14.04)*. https://switch2osm.org/serving-tiles/manually-building-a-tile-server-14-04/, accessed May 7, 2015.
11. OpenStreetMap contributors. (2013) *OpenStreetMap*. http://www. openstreetmap.org, accessed February 2, 2015.
12. OpenStreetMap Wiki contributors. (2013) *Map Features.* http://wiki.openstreetmap.org/wiki/Map_Features, accessed June 6, 2015.
13. OSGeo - Open Source Geospatial Fundation. (2014) *PostGIS 2.1.3 Manual.*
14. Romanian National Institute of Statistics. (2013) Regional social and economic figures: Territorial Statistics (Institutul Național de Statistică. Repere sociale și economice regionale: Statistica Teritorială). Bucharest, http://www.insse.ro/cms/files/publicatii/Statistica%20teritoriala/Statistica%20teritoriala%202013.pdf
15. The PostgreSQL Global Development Group. (2014) *PostgreSQL 9.3.9 Documentation.*
16. Thünen, J. H., & Hall, P. (1966) *Von Thünen's Isolated State: an English edition of Der isolierte Staat* (translated by Wartenberg, C. M. after von Thünen, J.H. *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*). (P. Hall, ed.) London: Pergamon Press.
17. Ubuntu Documentation Team. (2014) *Ubuntu Server Guide.* http://help.ubuntu.com, accessed December 7, 2014.
18. Venables, A. J. (2008) *New Economic Geography*. (Durlauf, S. and Blume, L., eds.) The New Palgrave Dictionary of Economics. 2nd edn.
19. Weber, A., Friedrich, C. J. (1929) *Alfred Weber's theory of the location of industries* (translated by Friedrich, C.J. after Weber, A. *Über den Standort der Industrie*). (C. J. Friedrich, ed.) Chicago: University of Chicago Press.