

## **Quantification of similarity between Land Cover categories**

Francesc Valls Dalmau<sup>1</sup>, Josep Roca Cladera<sup>2</sup>

### **ABSTRACT**

The transformation from forest and rural to developed land is growing dramatically worldwide. A measure of the underlying structure of this shift is needed to better understand the nature of this change and to be able to establish control mechanisms at a regional scale.

Landscape metrics evaluate the morphology of the spatial pattern at the object level (patch) through the categorization of Land Cover, generally without considering the similarities or dissimilarities between categories in the classification.

This work uses the Mean Edge Contrast Index (MECI) to quantify the structure of the counties around Columbus, OH. This index doesn't make a hard distinction between categories, since it uses a matrix that assigns a value to the contrast between each Land Cover class pairwise, and measures the degree of contrast between each patch and its immediate neighborhood.

The values of the matrix are obtained through the comparison of MECI and a studied variable, obtaining a measure of the similarity or dissimilarity of each class that better explains each variable. These relationships are visualized with different techniques to better understand their similarities.

### **KEYWORDS**

Spatial Pattern Analysis, Edge Contrast, Machine Learning, Data Visualization, Sprawl

---

<sup>1</sup> Centre of Land Policy and Valuations, Technical University of Catalonia (BarcelonaTech), e-mail: francescvalls@gmail.com

<sup>2</sup> Centre of Land Policy and Valuations, Technical University of Catalonia (BarcelonaTech), e-mail: josep.roca@upc.edu

## **1 Introduction**

The capacity of quantify underlying spatial structure of a territory is crucial to understand the complexity our environment and its changes over time.

The classification of remote sensed data provides us with categories of Land Cover and in some cases Land Use (Di Gregorio & Jansen, 2005), but to better understand how a region is structured, this classified data can be used to define multiple morphological indicators that allow us to measure, for example, its fragmentation or its diversity (Jaeger, 2000).

Beyond pixel level analysis (basic local, zonal, focal or global statistics), we can find many Synthetic Landscape Indicators, which usually base their calculation on three levels: “patch”, “class” and “landscape” (McGarigal & Marks, 1995).

Many of these indicators allow us to study growth patterns and therefore are a valuable tool to diagnose urban sprawl, and can be used to feed an Agent-Based Model to perform prospective analysis.

## **2 Objectives**

Categories are often organized hierarchically (Anderson, 1976) following a structure that is somewhat similar to the Carl Linnaeus classification of species, but without quantifying the magnitude of their relationship.

Some landscape indicators are based on the similarity or dissimilarity between Land Cover classes. The objective of this investigation is to perform this calculation in reverse, to get the contrast matrix that better explain any variable, instead of using a predefined contrast matrix to calculate the index.

The final result is a table with the dissimilarity between classes that yields the maximum correlation between the studied variables: in our case population and workplaces density. The results of this machine learning can then be fed to an Agent-Based Model.

## **3 Edge Contrast based Landscape Indicators**

There are several indices based on Edge Contrast between neighboring patches, which can be aggregated (with operations such as sum, differently weighted averages, statistical measurements, etc.) at the patch, class or landscape level.

These indicators are based on a contrast table, which define the amount of contrast (or dissimilarity) between each pair of Land Cover classes. If we assume that the contrast A-B is the same as the contrast between B-A, and that the contrast of each land use with itself is 0, the contrast table takes the form of a square symmetrical matrix with zeros on its diagonal.

### 3.1 Edge Contrast Index (ECON)

Edge Contrast Index is founded on the notion that all edges are not created equal (McGarigal, Cushman, Neel, & Ene, 2002). To account for this, the notion of edge “contrast” was created. This index is a relative measure of the amount of contrast along the patch perimeter.

ECON equals the sum of the patch perimeter segment lengths multiplied by their corresponding contrast weights, divided by total patch perimeter.

$$ECON_{ij} = \frac{\sum_{k=1}^m (p_{ijk} * d_{ik})}{p_{ij}}$$

Considering the following:

- $p_{ijk}$**  Length of edge of patch **ij** adjacent to patch type (class) **k**
- $d_{ik}$**  Dissimilarity (edge contrast weight) between patch types **i** and **k** (0% - 100%)
- $p_{ij}$**  Length of perimeter of patch **ij**

### 3.2 Mean Edge Contrast Index (MECI)

The Mean Edge Contrast Index (MECI or ECON\_MN), quantifies the average edge contrast for patches of a particular Land Cover in a landscape (class level) or for all patches in the landscape (landscape level).

The MECI inside a boundary we will call zone is:

$$MECI_k = ( \sum_{i \in I_k} \sum_{j \in J_{ki}} ( \sum_{r \in R_{ki}} (d_{kijr} * w_{ir}) / p_{kij} ) ) * 100 / N_k$$

Using the following sub-indexes:

- k** Zones
- i, r** Land use classes in zone **k**
- j** Land use patches with land use **i** inside zone **k**

Considering the following:

- $d_{kijr}$**  Edge length of patch **j** with land use **i** in contact with polygons with land use **r** inside zone **k**
- $w_{kir}$**  Edge contrast between land use **i** and land use **r** (for all zones)
- $p_{kij}$**  Perimeter of patch **j** with land use **i** inside zone **k**
- $N_k$**  Total number of patches in zone **k**
- $R_{ki}$**  Set of polygons with land use **r** in zone **k**
- $J_{ki}$**  Set of polygons with land use **i** in zone **k**
- $I_k$**  Set of land uses in zone **k**

### 3.3 Contrast

The following tables show the proposed contrast values (Table 1) and similarity values (Table 2) for six categories of land use (McGarigal K. , 2012):

	open	residential	water	forest	wetland	urban
open	0	0.75	0.5	0.75	0.5	0.75
resident	0.75	0	1	0.5	0.75	0.5
water	0.5	1	0	1	0.5	1
forest	0.75	0.5	1	0	0.75	0.5
wetland	0.5	0.75	0.5	0.75	0	0.75
urban	0.75	0.5	1	0.5	0.75	0

**Table 1: Contrast table (source: McGarigal)**

	open	residential	water	forest	wetland	urban
open	1	0	0	0	0.5	0
resident	0	1	0	0	0	0
water	0	0	1	0	0	0
forest	0	0.2	0	1	0.2	0
wetland	0.2	0	0.8	0	1	0
urban	0	0	0	0	0	0

**Table 2: Similarity table (source: McGarigal)**

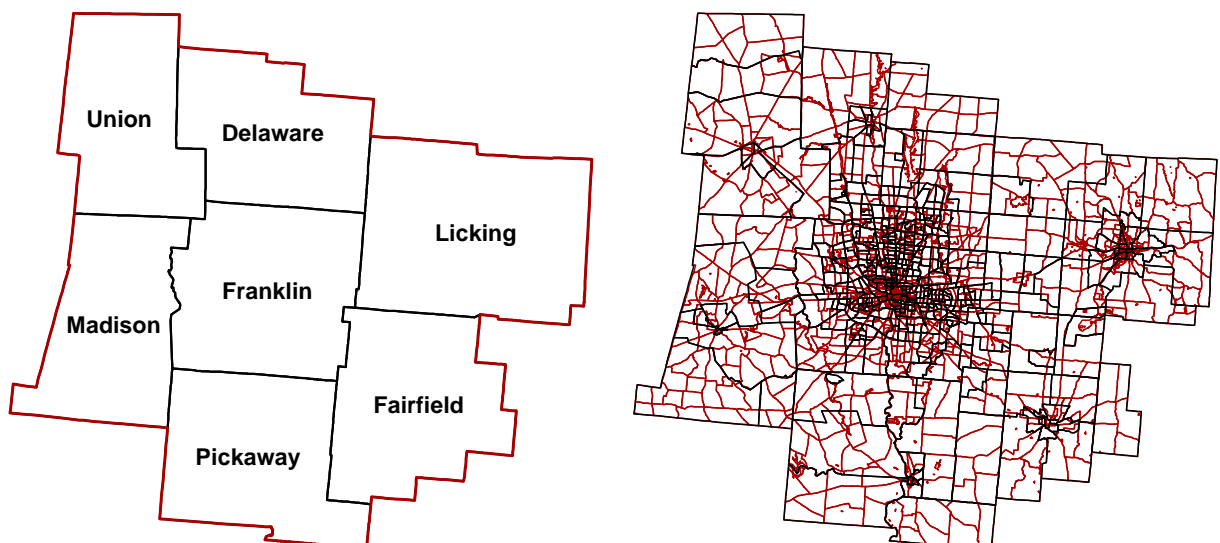
## 4 Source data

### 4.1 Traffic Analysis Zones (TAZ) and Census Tracts Cartography

The cartography used for demographic data (workplaces and population) was the following:

- 1990 Traffic Analysis Zones (TAZ)
- 2000 Census Tracts (U.S. Census Bureau, 2012)

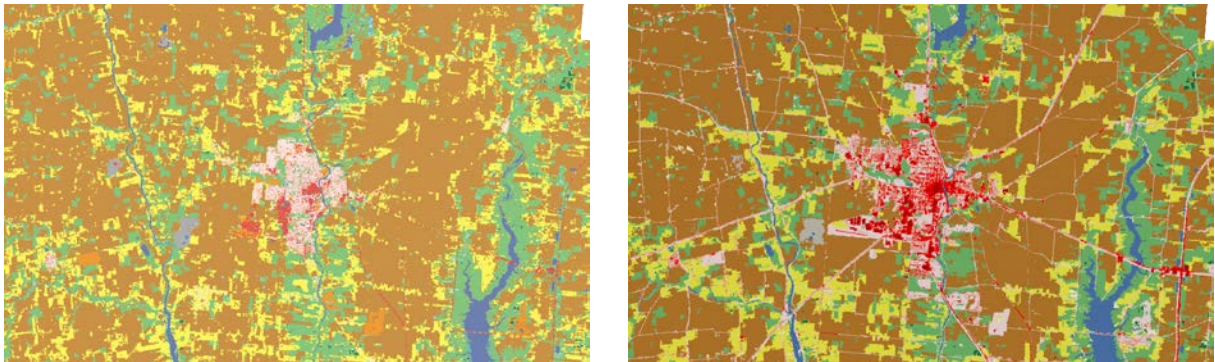
The following map show the the area study of the counties around Columbus (OH), and its Tracts drawn in black lines and TAZ drawn in red lines (Figure 1):



**Figure 1: Area of Study (left) and its Tracts in black and TAZs in red (right)**

## 4.2 Land Cover Raster

The Land Cover data used was from the 1992 (Multi-Resolution Land Characteristics Consortium, 2011) and 2001 (Multi-Resolution Land Characteristics Consortium, 2012). This two 30x30 resolution rasters are classified with different methodologies (United States Environmental Protection Agency, 2007) so it is not possible to perform a direct comparison (United States Environmental Protection Agency, 2007) of the results obtained. The following images (Figure 2) show the same zone classified with both methodologies, whose differences cannot be explained by the development between the elapsed years alone.



**Figure 2: 1992 (left) and 2001 (right) Land Use Classification**

They were classified with the following legends, for 1992 and 2001 data (Table 3), with the percentage of each Land Cover in the area of study shown in the rightmost column:

SYM.	CODE	DESCRIPTION	%	SYM.	CODE	DESCRIPTION	%
	11	Open Water	0.42		11	Open Water	1.13
	12	Perennial Ice / Snow	-		12	Perennial Ice / Snow	-
	21	Low Intensity Residential	3.98		21	Developed, Open Space	7.92
	22	High Intensity Residential	0.98		22	Developed, Low Intensity	5.81
	23	Commercial / Industry / Transp.	2.14		23	Developed, Medium Intensity	2.73
	31	Bare Rock / Sand / Clay	-		24	Developed, High Intensity	1.17
	32	Quarries / Strip Mines / Gravel Pits	0.13		31	Barren Land (Rock / Sand / Clay)	0.12
	33	Transitional Barren	0.03		41	Deciduous Forest	16.65
	41	Deciduous Forest	17.15		42	Evergreen Forest	0.31
	42	Evergreen Forest	0.48		43	Mixed Forest	0.10
	43	Mixed Forest	0.13		51	Dwarf Scrub (Alaska Only)	-
	51	Scrubland	-		52	Shrub / Scrub	0.14
	61	Orchards / Vineyards / Other	-		71	Grassland / Herbaceous	0.74
	71	Grassland / Herbaceous	-		72	Sedge / Herbaceous (Alaska Only)	-
	81	Pasture / Hay	20.22		73	Lichens (Alaska Only)	-
	82	Row Crops	52.28		74	Moss (Alaska Only)	-
	83	Small Grains	-		81	Pasture / Hay	10.63
	84	Fallow	-		82	Cultivated Crops	52.34
	85	Urban / Recreational Grasses	1.62		90	Woody Wetlands	0.16
	91	Woody Wetlands	0.33		95	Emergent Herbaceous Wetlands	0.05
	92	Emergent Herbaceous Wetlands	0.11				

**Table 3: Classification of 1992 (left) and 2001 (right) Land Cover and their percentage in the area of study**

## 5 Methodology

The methodology used involved the use of several software products to build the tool chain to obtain the desired results:

- ESRI ArcGIS for the preparation of the cartography (through Model Builder automation) and the visualization of results
- Microsoft Access for its capacity to aggregate large volumes of information
- Microsoft Excel to model the problem (through VBA automation) to feed the solver
- Frontline Systems Solver (bundled with Microsoft Excel) to perform the optimization

### 5.1 GIS cartography preparation

Since the calculations had to be done at the landscape level for every administrative unit, several steps were necessary to get the tables to be fed to the database engine.

#### 5.1.1 Cartography transformation

The cartography had to be transformed to perform the operations needed:

- The Land Cover raster was converted to vector data
- All cartography had to be projected to be able to take area measurements
- The raster data was clipped to the boundary of the area of study with an offset of 1km to minimize data usage. This was possible because the formulation does not take into account the area of the patches but only the edge of contact between them
- A Model Builder process was developed to split the cartography into counties, that were rejoined later in Access

#### 5.1.2 Overlay

Land Cover and Administrative Divisions were overlaid (intersected) to get a new layer made of homogeneous fragments with the attributes of the two original layers (Figure 3) with the following information:



Figure 3: Overlay of Land Cover and Administrative Divisions

- Unique ID of the fragment
- Land Cover codes and Administrative unit codes from the overlay layer
- Polygon perimeter and Polygon area of the fragment

### 5.1.3 Conversion to lines

The resulting overlay layer was decomposed to lines (ESRI, 2012). The following illustration shows the possible cases and the result table (Figure 4):

- Polygons in the boundary
- Disconnected polygons
- Polygons completely surrounding another
- Polygons completely surrounded by others
- Holes (no entity)

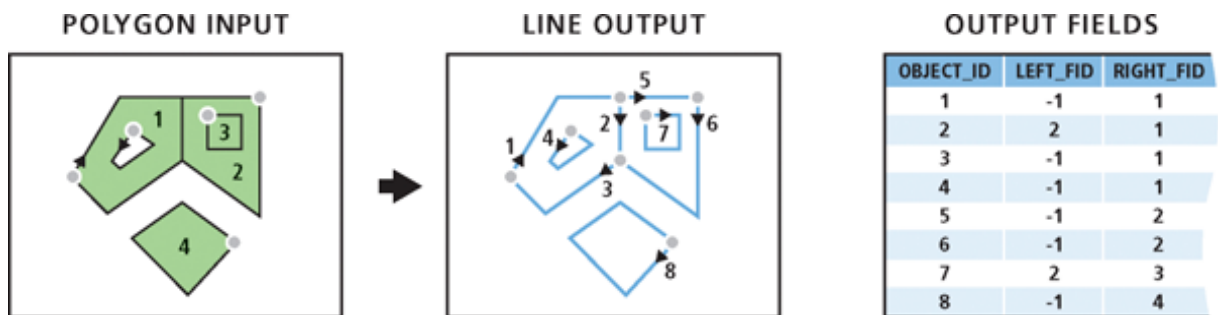


Figure 4: Polygon to line conversion process (source: ESRI)

## 5.2 Database preprocessing

The following diagram (Figure 5) shows the stacked SQL queries that were built to get the final tables to be processed by Excel.

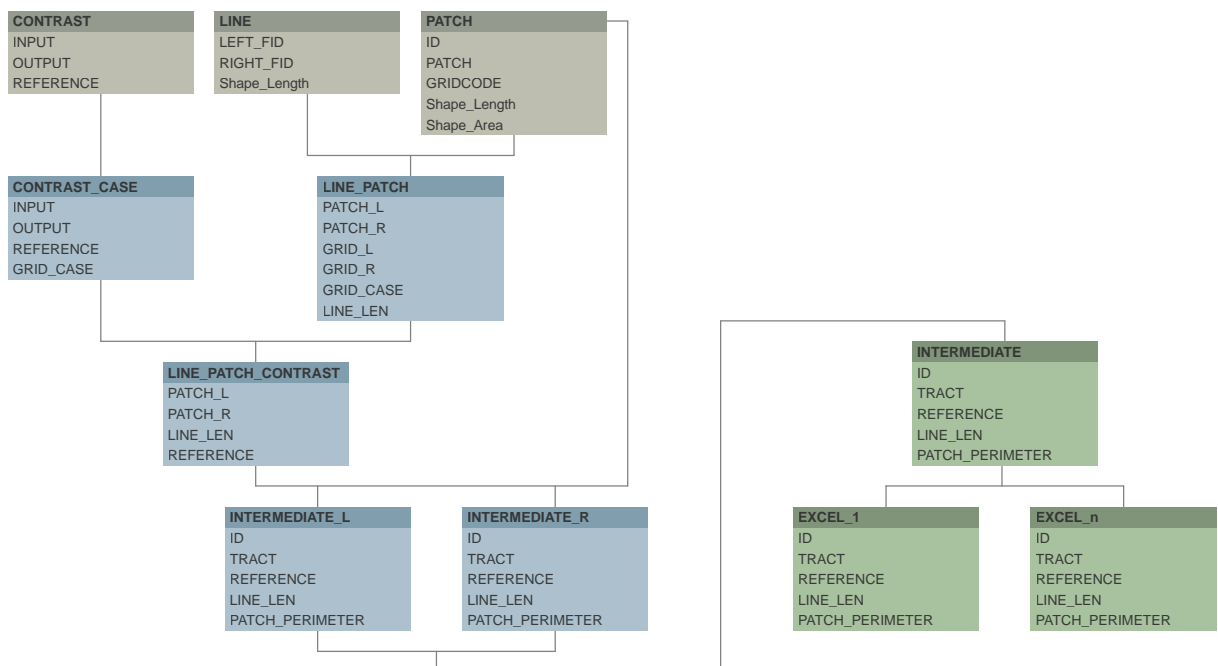


Figure 5: Database SQL queries diagram

The database created used the following tables from the GIS data to build the tables needed to feed the spreadsheet used for modeling:

- The polygon table resulting from the overlay operation
- The edge table resulting from the decomposition of the above layer into lines
- The contrast matrix, converted to table form (origin, destination, value) with a custom VBA routine (a reverse pivot table), whose values were the coordinates (pointers) of the cells that would hold the actual contrast values (variables)

The results were exported to Excel using a routine written in the Access macro language

### **5.3 Spreadsheet problem modeling**

Microsoft Excel was used to model the optimization problem instead of a custom optimization language. Several workarounds had to be made to overcome the limitation of 1,048,576 ( $2^{20}$ ) rows in Excel 2010, splitting the data in several sheets linked to a “master” one.

The implementation (Valls & Roca, 2011) of the formula working on the imported data was developed in Visual Basic for Applications (VBA), because Excel lacks aggregation functions common in databases:

- It calculates the product of the contrast by the segment length. The routine can choose the appropriate contrast for every edge because it knows beforehand the Land Cover on either sides of the edge thanks to the prior preprocessing in the database
- Performs an aggregated sum of the product calculated before for every patch in every Tract and divides it by the perimeter of the patch
- Calculates an aggregated average of the values calculated before for every Tract
- Performs a search in a table to retrieve the value of the indicator (population density, workplaces) to study and joins it to the calculated value for each Tract

### **5.4 Solver optimization**

#### **5.4.1 Limitations**

The current Solver (at the day of writing) that ships with Excel 2010 has the following restrictions (Frontline Systems, 2012):

- Maximum number of Constraints: 100
- Maximum number of Decision Variables: 200

To overcome these limits, the following assumptions were made:

- The contrast with water (class 11) has been considered to be 100%
- The contrast in the diagonal (with oneself) has been considered 0%
- The matrix is considered symmetrical, so contrast A-B is the same than contrast B-A



These assumptions reduced the number of variables to fit the limitations:

- For 1992 data (14x14 matrix) from  $14^2$  to  $(13^2-13)/2$ , thus from 225 to 91 variables
- For 2001 data (15x15 matrix) from  $15^2$  to  $(14^2-14)/2$ , thus from 196 to 78 variables

#### **5.4.2 Parameters**

The optimization was set up to maximize the Pearson Correlation coefficient between MECI and the variable of study, and the variables were assigned to the contrast values, limiting its scope between 0 and 1 (minimum and maximum contrast).

The problem was diagnosed as Non Linear, and the following algorithms were tested (GRG Non-Linear, Simplex LP and Evolutionary). The best results were obtained with the Generalized Reduced Gradient Non-Linear as expected, using automatic scaling.

To perform the calculation, the initial values were seeded with 100% values (maximum contrast) and 0% values (no contrast) yielding almost identical results. The contrast values that were more sensitive to the seed values were the ones that had a small presence in the area of study (Table 3) because the optimization routine had less information to guess its values and its modifications didn't alter the goal significantly.

#### **5.5 Visualization of the results**

The resulting contrast table was ordered top-down and left-right in ascending order using the Land Cover class number. This order can conceal patterns that may exist between the elements of the table. If we group the elements that are closer together (have less contrast between them) this patterns can be revealed.

To reveal these patterns, the "distance" between each class was computed to get a new table with the distances between each pair. A cluster analysis was performed on this data using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm.

To visualize the results, a dendrogram was built to show the phylogenetic distance between different Land Cover classes, for each variable studied:

- Classes that are more similar originate in the same branch, while dissimilar ones are shown in visually separate branches
- The order is modified so classes that are closer together appear listed together

The order extracted from this tree hierarchy was used to reorder the matrix to better reveal relationships between classes (Bertin & Bonin, 1975). In this case the shortest distances are closer to the diagonal, revealing the clustering in a stair-like effect.

Finally, the edges were colored according to the contrasts obtained after the optimization.

## 6 Results

### 6.1 Land Cover 2001 and Population 2000 (Tracts)

The Pearson correlation of the following variables was maximized through the optimization of the contrast matrix values, yielding a value of 0.89:

- MECI of each Tract (form 2001 Land Cover data and 2000 Tracts)
- Logarithm of the Population Density of each Tract (from 2000 population data)

The following images show the scatterplot between the two variables (Chart 1), and the resulting maps (Figure 6).

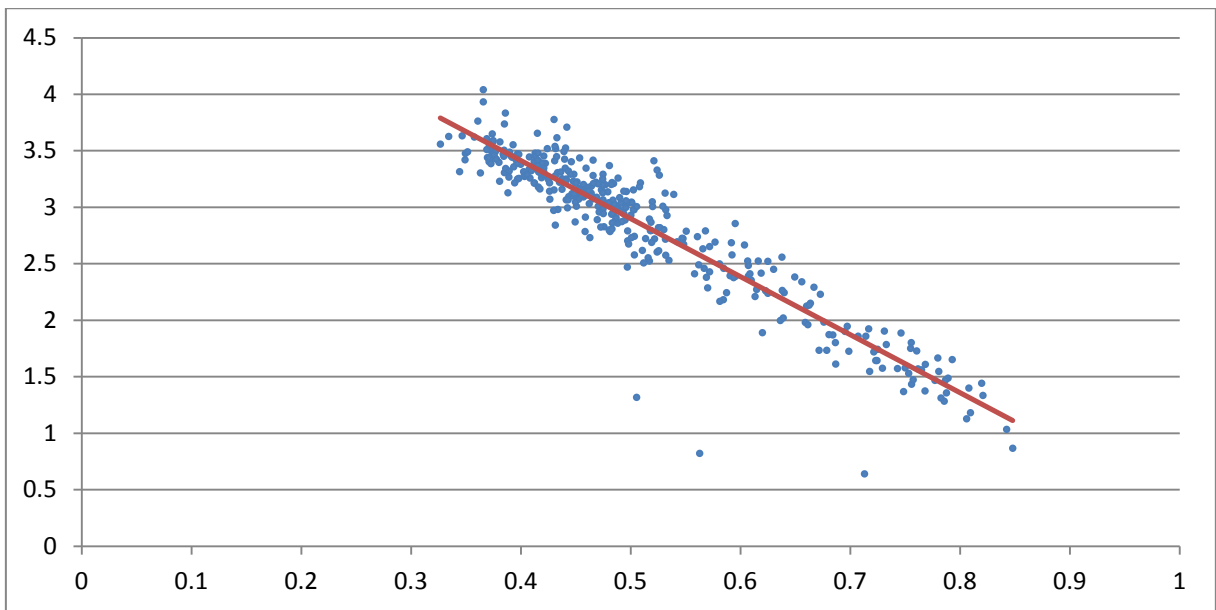


Chart 1: MECI (horizontal) and Population Density Log (vertical) for 2001 Land Cover and 2000 Tracts

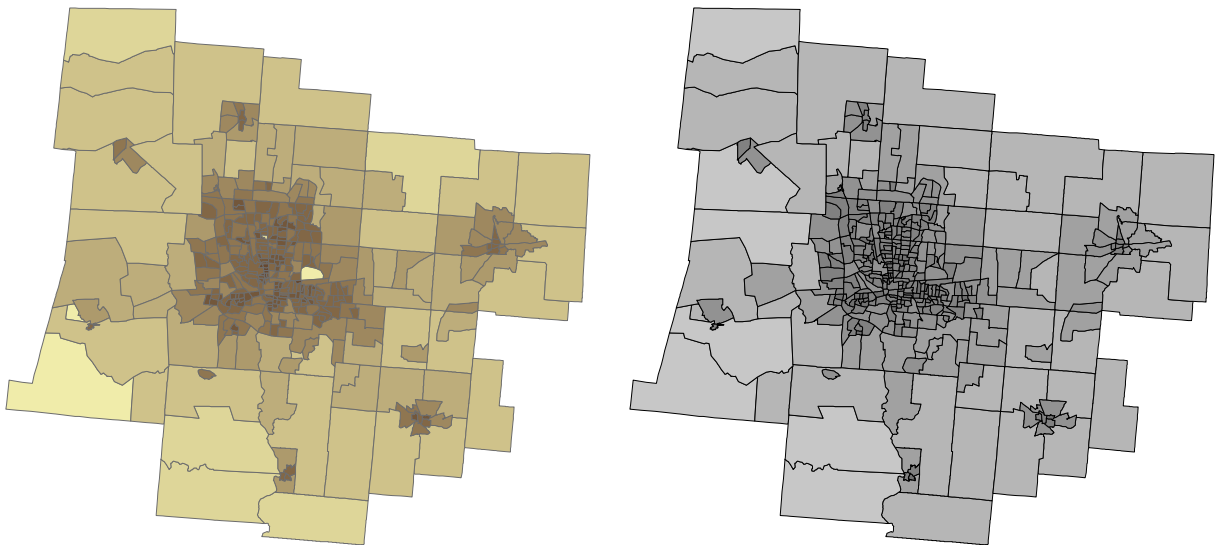


Figure 6: Maps of Population Density Log (left) and MECI (right) for 2001 Land Cover and 2000 Tracts

The dendrogram constructed from the contrast matrix was the following (Figure 7), and had a Cophenetic Correlation Coefficient of 0.63:

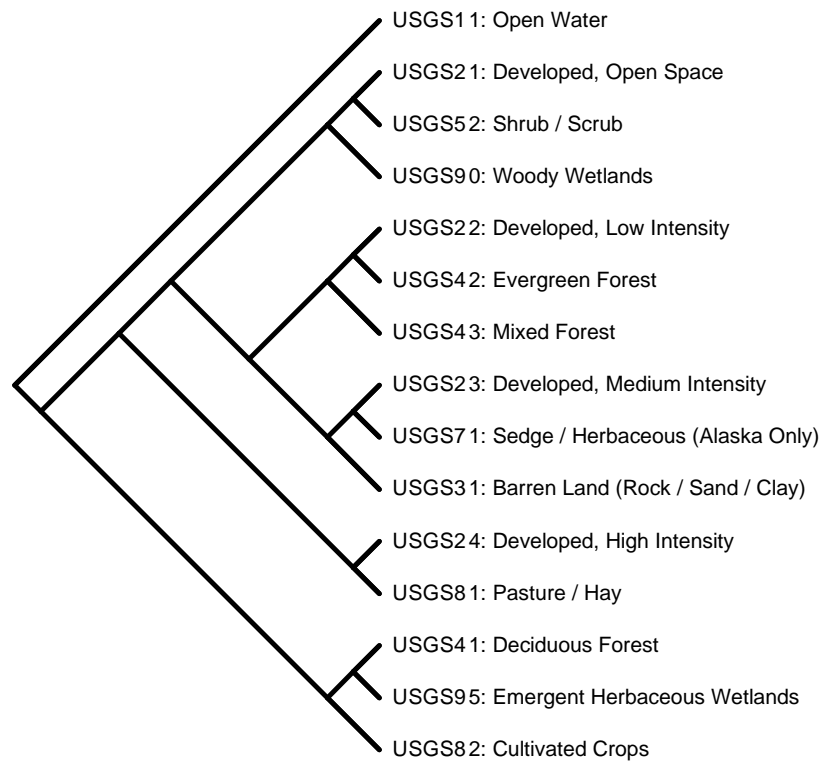


Figure 7: Dendrogram of the contrast for 2001 Land Cover and 2000 Tracts

The following tables show the optimal contrast matrix result of the optimization (Table 4), before and after reordering:

	11	21	22	23	24	31	41	42	43	52	71	81	82	90	95
11	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
21	1.0	0.0	0.5	0.9	1.0	1.0	0.6	0.3	1.0	0.0	0.5	0.8	1.0	0.0	1.0
22	1.0	0.5	0.0	0.4	1.0	0.8	0.1	0.0	0.0	0.0	0.0	1.0	0.5	1.0	1.0
23	1.0	0.9	0.4	0.0	0.4	0.3	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
24	1.0	1.0	1.0	0.4	0.0	0.4	1.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
31	1.0	1.0	0.8	0.3	0.4	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0
41	1.0	0.6	0.1	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.7	0.8	0.4	0.0	0.0
42	1.0	0.3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.8	1.0	0.0	1.0
43	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.2	1.0	0.0
52	1.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.9	0.0	1.0
71	1.0	0.5	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7	1.0	0.0	1.0
81	1.0	0.8	1.0	0.0	0.0	1.0	0.7	0.8	1.0	1.0	0.7	0.0	1.0	0.0	0.0
82	1.0	1.0	0.5	1.0	1.0	1.0	0.8	1.0	0.2	0.9	1.0	1.0	0.0	1.0	0.0
90	1.0	0.0	1.0	0.0	1.0	1.0	0.4	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
95	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0

	11	21	52	90	22	42	43	23	71	31	24	81	41	95	82
11	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21	1.00	0.00	0.00	0.00	0.51	0.33	1.00	0.90	0.53	1.00	1.00	0.81	0.55	1.00	1.00
52	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.93
90	1.00	0.00	0.00	0.00	0.96	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.38	0.00	1.00
22	1.00	0.51	0.00	0.96	0.00	0.00	0.00	0.37	0.00	0.77	1.00	1.00	0.07	1.00	0.52
42	1.00	0.33	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.77	1.00	1.00	1.00
43	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.20
23	1.00	0.90	0.00	0.00	0.37	0.00	0.00	0.00	0.00	0.31	0.44	0.00	1.00	0.00	1.00
71	1.00	0.53	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.65	1.00	1.00	1.00
31	1.00	1.00	0.00	1.00	0.77	0.00	0.00	0.31	0.00	0.00	0.39	1.00	1.00	1.00	1.00
24	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.44	1.00	0.39	0.00	0.00	1.00	1.00	1.00
81	1.00	0.81	1.00	0.00	1.00	0.77	1.00	0.00	0.65	1.00	0.00	0.00	0.75	0.00	1.00
41	1.00	0.55	1.00	0.38	0.07	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.00	0.00	0.77
95	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
82	1.00	1.00	0.93	1.00	0.52	1.00	0.20	1.00	1.00	1.00	1.00	1.00	0.77	0.00	0.00

Table 4: Contrast matrix before (left) and after (right) reordering for 2001 Land Cover and 2000 Tracts

## 6.2 Land Cover 1992 and Population 1990 (TAZs)

For comparison, the Pearson correlation of the following variables was maximized through the optimization of the contrast matrix values, yielding a value of 0.85:

- MECI of each Tract (form 1992 Land Cover data and 1990 TAZ)
- Logarithm of the Population Density of each TAZ (from 1990 population data)

The following images show the scatterplot between the two variables (Chart 1), and the resulting maps (Figure 8).

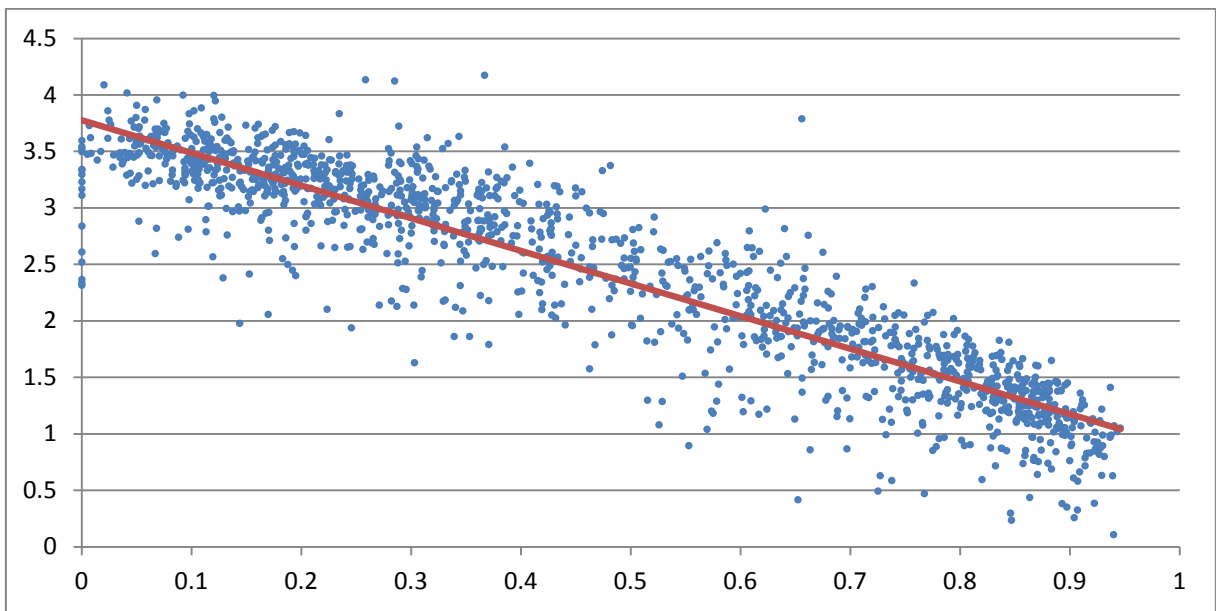


Chart 2: MECI (horizontal) and Population Density Log (vertical) for 1992 Land Cover and 1990 TAZ

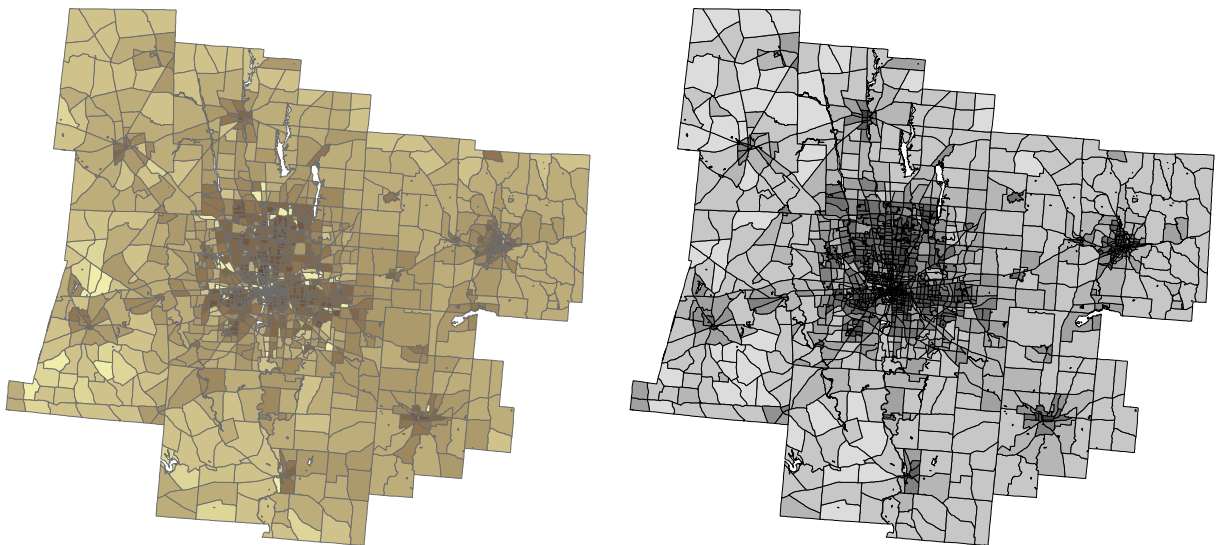


Figure 8: Maps of Population Density Log (left) and MECI (right) for 1992 Land Cover and 1990 TAZ

The dendrogram constructed from the contrast matrix was the following (Figure 9) and had a Cophenetic Correlation Coefficient of 0.82:

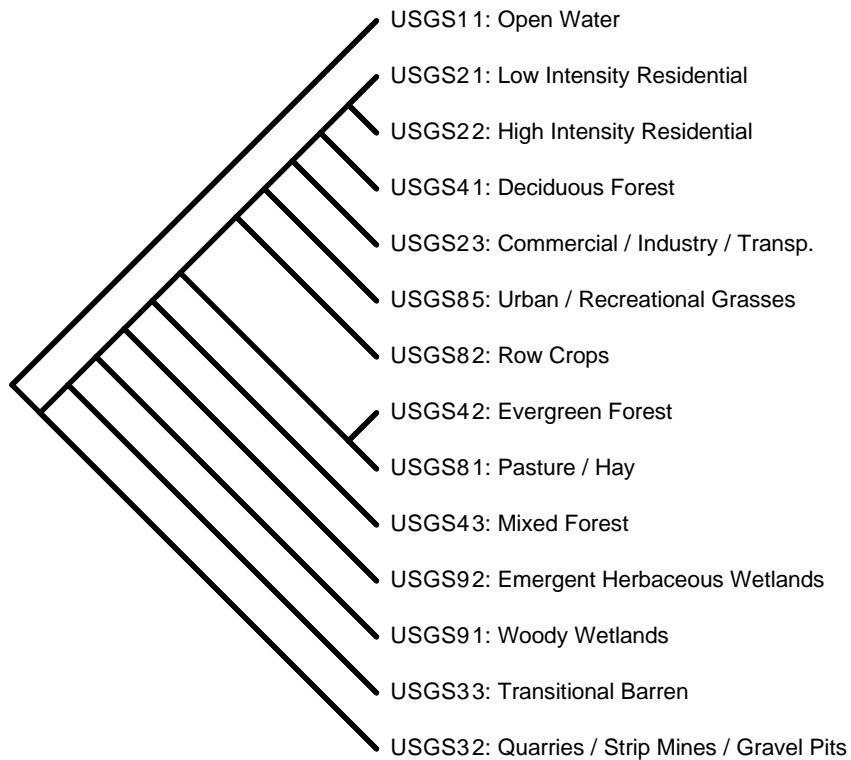


Figure 9: Dendrogram of the contrast for 1992 Land Cover and 1990 TAZ

The following tables show the optimal contrast matrix result of the optimization (Table 5), before and after reordering:

	11	21	22	23	32	33	41	42	43	81	82	85	91	92
11	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
21	1.0	0.0	0.0	0.2	1.0	1.0	0.0	0.7	0.9	0.0	0.3	0.0	1.0	1.0
22	1.0	0.0	0.0	0.6	1.0	1.0	0.0	0.9	0.9	0.6	0.2	0.5	1.0	1.0
23	1.0	0.2	0.6	0.0	1.0	1.0	0.3	0.9	1.0	1.0	0.9	0.8	1.0	0.8
32	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
33	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
41	1.0	0.0	0.0	0.3	1.0	1.0	0.0	0.7	0.9	0.7	1.0	0.7	1.0	1.0
42	1.0	0.7	0.9	0.9	1.0	1.0	0.7	0.0	1.0	0.9	1.0	1.0	1.0	1.0
43	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.0	1.0	1.0	1.0	1.0	1.0
81	1.0	0.0	0.6	1.0	1.0	1.0	0.7	0.9	1.0	0.0	1.0	1.0	1.0	1.0
82	1.0	0.3	0.2	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.8	1.0	1.0
85	1.0	0.0	0.5	0.8	1.0	1.0	0.7	1.0	1.0	1.0	0.8	0.0	1.0	1.0
91	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
92	1.0	1.0	1.0	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0

	11	21	22	41	23	85	82	42	81	43	92	91	33	32
11	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
21	1.00	0.00	0.00	0.00	0.16	0.00	0.27	0.69	0.00	0.91	0.98	0.98	1.00	1.00
22	1.00	0.00	0.00	0.00	0.55	0.53	0.22	0.93	0.64	0.94	0.97	0.98	0.99	1.00
41	1.00	0.00	0.00	0.00	0.29	0.69	1.00	0.72	0.70	0.86	0.96	1.00	1.00	1.00
23	1.00	0.16	0.55	0.29	0.00	0.79	0.94	0.93	1.00	0.99	0.81	1.00	1.00	1.00
85	1.00	0.00	0.53	0.69	0.79	0.00	0.76	1.00	0.98	1.00	1.00	1.00	1.00	1.00
82	1.00	0.27	0.22	1.00	0.94	0.76	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
42	1.00	0.69	0.93	0.72	0.93	1.00	1.00	0.00	0.91	0.95	1.00	1.00	1.00	1.00
81	1.00	0.00	0.64	0.70	1.00	0.98	1.00	0.91	0.00	1.00	1.00	1.00	1.00	1.00
43	1.00	0.91	0.94	0.86	0.99	1.00	1.00	0.95	1.00	0.00	1.00	1.00	1.00	1.00
92	1.00	0.98	0.97	0.96	0.81	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
91	1.00	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00
33	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00
32	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Table 5: Contrast matrix before (left) and after (right) reordering for 1992 Land Cover and 1990 TAZ

In this case the clusters appear more clearly than on the previous example.

### 6.3 Land Cover 2001 and Workplaces 2000 (Tracts)

Finally, the Pearson correlation of the following variables was maximized through the optimization of the contrast matrix values, yielding a value of 0.83:

- MECI of each Tract (from 2001 Land Cover data and 2000 Tracts)
- Logarithm of the Workplaces Density of each Tract (from 2000 data)

The following images show the scatterplot between the two variables (Chart 3), and the resulting maps (Figure 10).

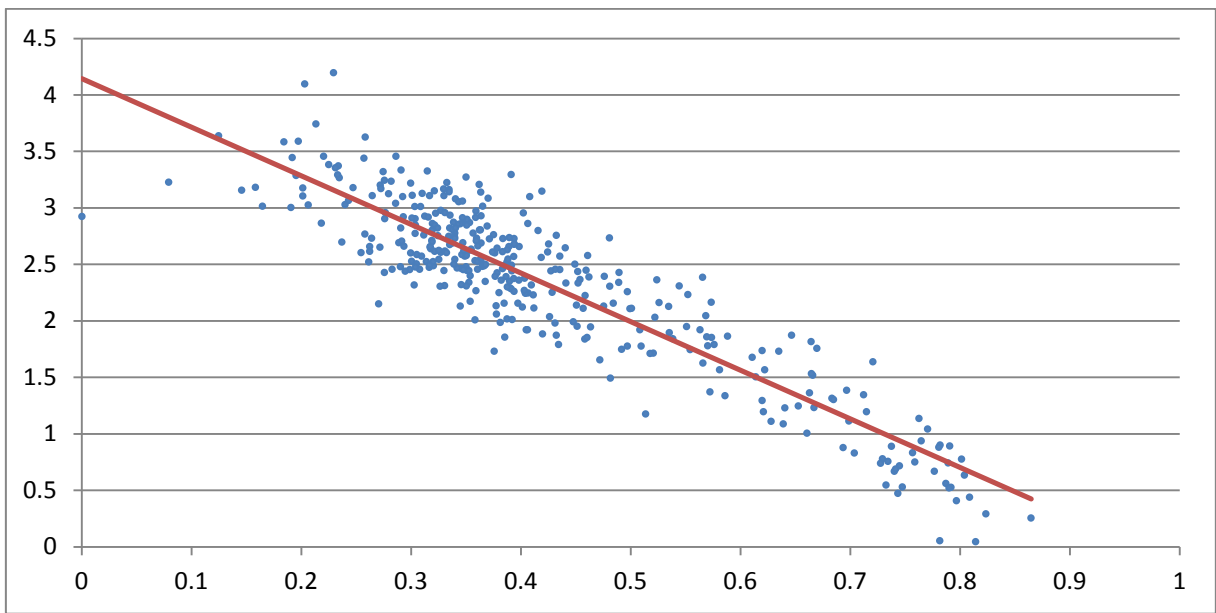


Chart 3: MECI (horizontal) and Workplaces Density Log (vertical) for 2001 Land Cover and 2000 Tracts

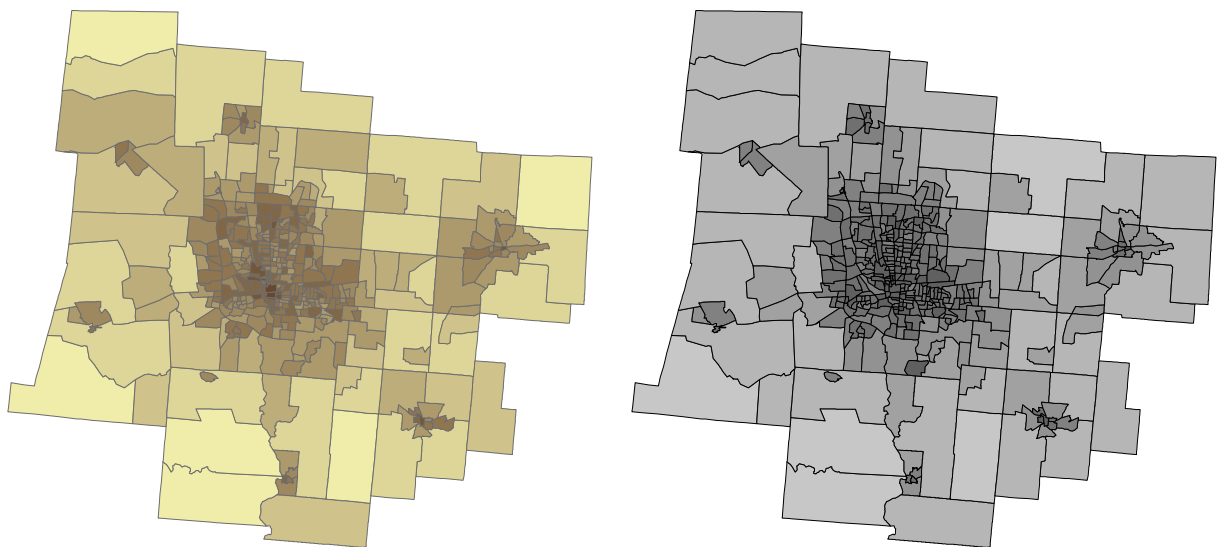


Figure 10: Maps of Workplaces Density Log (left) and MECI (right) for 2001 Land Cover and 2000 Tracts

The dendrogram constructed from the contrast matrix was the following (Figure 11) and had a Cophenetic Correlation Coefficient of 0.63:

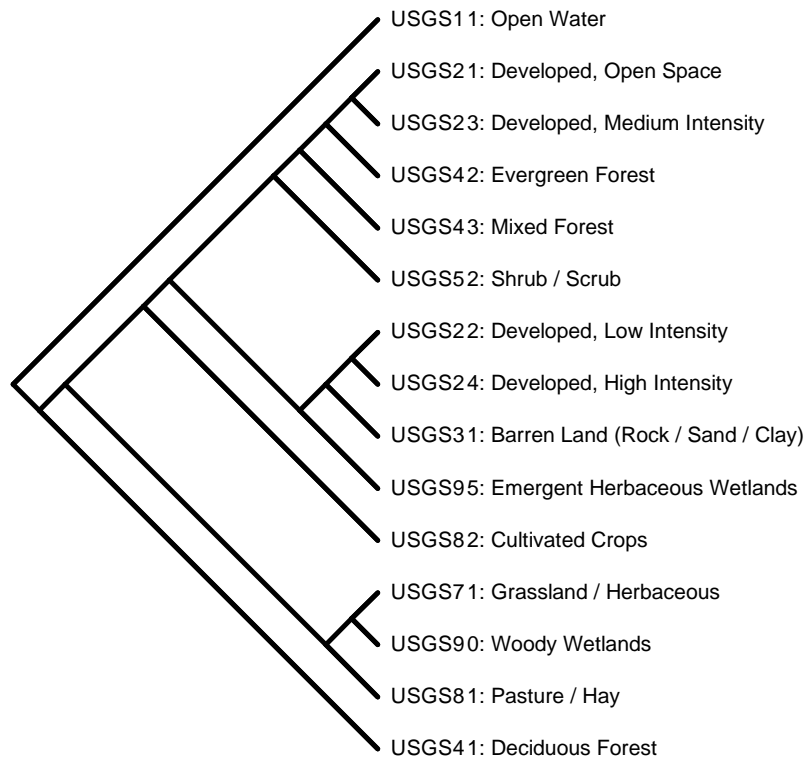


Figure 11: Dendrogram of the contrast matrix of MECI and Tract Worplaces Density Log

The following tables show the optimal contrast matrix result of the optimization (Table 6), before and after reordering:

	11	21	22	23	24	31	41	42	43	52	71	81	82	90	95
11	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
21	1.0	0.0	0.4	0.0	0.3	0.4	0.7	0.0	0.0	0.0	0.3	1.0	1.0	0.6	0.0
22	1.0	0.4	0.0	0.5	0.0	0.0	0.0	0.0	0.3	0.0	1.0	0.7	0.2	1.0	0.0
23	1.0	0.0	0.5	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
24	1.0	0.3	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
31	1.0	0.4	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0
41	1.0	0.7	0.0	0.2	0.0	1.0	0.0	1.0	1.0	1.0	0.4	0.7	1.0	0.5	0.0
42	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
43	1.0	0.0	0.3	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0
52	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.4	1.0	1.0	0.0
71	1.0	0.3	1.0	0.0	1.0	0.0	0.4	1.0	1.0	1.0	0.0	0.8	1.0	0.0	1.0
81	1.0	1.0	0.7	0.1	0.0	0.0	0.7	1.0	0.0	0.4	0.8	0.0	0.7	0.0	1.0
82	1.0	1.0	0.2	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.7	0.0	0.0	0.0
90	1.0	0.6	1.0	0.0	0.0	1.0	0.5	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
95	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0

	11	21	41	22	42	43	95	23	24	31	81	90	71	82	52	
11	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
21	1.00	0.00	0.00	0.00	0.00	0.00	0.45	0.30	0.42	0.00	1.00	0.31	0.56	1.00	0.70	
22	1.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.17
23	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	
24	1.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	
31	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.43	1.00
41	1.00	0.45	0.48	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.17	1.00	1.00	0.65	0.00	
42	1.00	0.30	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	
43	1.00	0.42	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00
52	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	
71	1.00	1.00	0.00	0.00	1.00	1.00	0.17	0.00	1.00	0.00	0.00	1.00	0.00	0.66	1.00	
81	1.00	0.31	0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.81	0.44	
82	1.00	0.56	0.00	0.00	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.47	
90	1.00	1.00	0.11	1.00	0.00	0.43	0.65	0.00	0.00	1.00	0.66	0.81	0.00	0.00	0.73	
95	1.00	0.70	0.17	1.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.44	0.47	0.73	0.00	

Table 6: Optimal contrast matrix before (left) and after (right) reordering

In this case the clusters appear even more clearly than on the two previous examples.

## 7 Sensitivity analysis

The previous results brought up some questions to about the influence of two factors:

- The geometry of the analysis unit (Tract and TAZ)
- The contiguity criteria in raster analysis

To answer these, a battery of tests was configured to better understand the influence of the change of those two variables using the 1992 Land Cover data and the 1990 TAZ for easy comparison to the previous results.

### 7.1 Hexagonal and Square grids

The TAZ data was converted into two regular grids, with equal surface cells:

- A square grid with a grid separation of  $1\text{km}^2$ , with each square cell measuring 100ha
- A hexagonal grid with hexagons measuring 100ha

Assuming that the population was spread evenly in every TAZ, a Model Builder module was built to break the administrative units and assign the population to every cell proportionally.

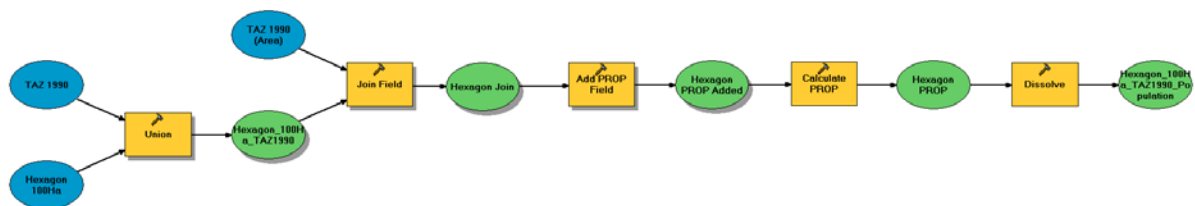


Figure 12: Diagram of the module that performs the tessellation

#### 7.1.1 Square grid results

Reusing the matrix obtained for TAZ, the resulting Pearson Coefficient was lowered to 0.39, with the corresponding maps shown below (Figure 13).

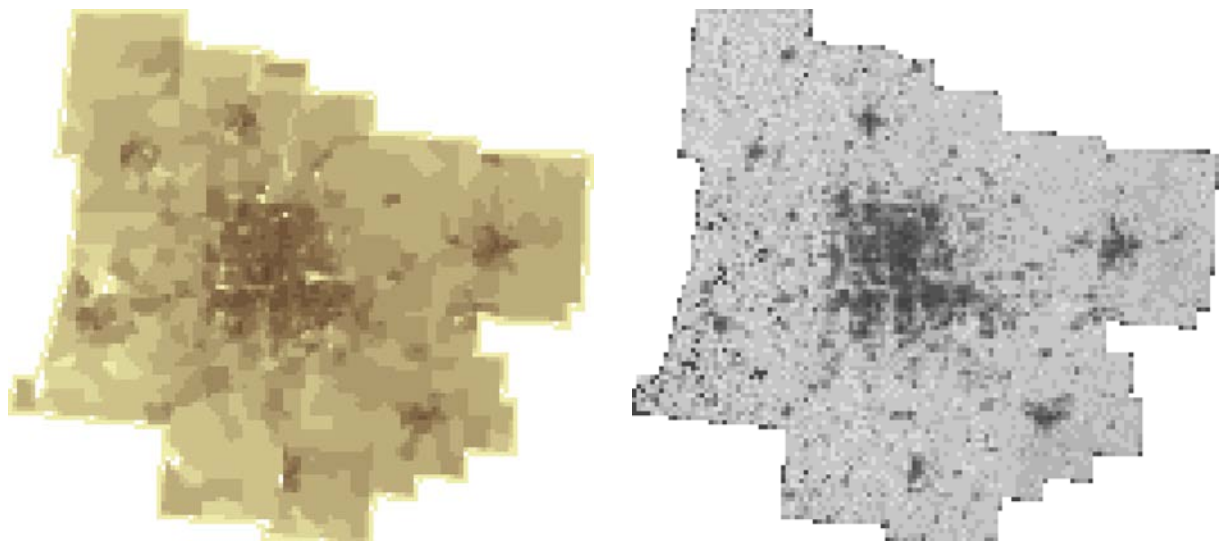


Figure 13: Maps of Population Density Log (left) and MECI (right) tessellated in a 100Ha square grid



### 7.1.2 Hexagonal grid results

Using the same matrix, the resulting Pearson Coefficient was the same 0.39. The corresponding maps below (Figure 14) show the similarities with the square grid.

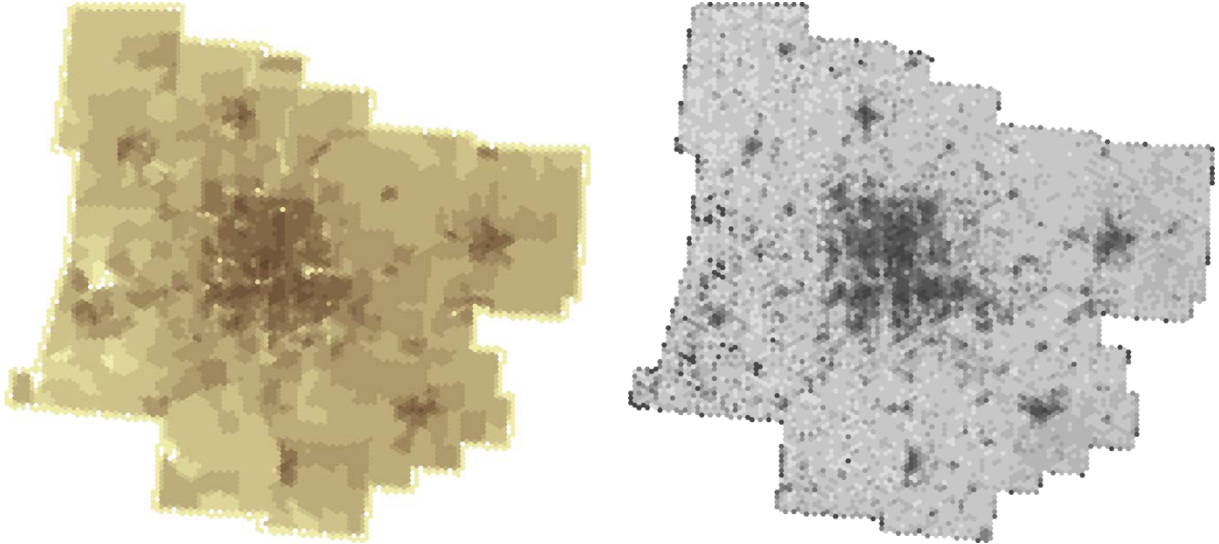


Figure 14: Maps of Population Density Log (left) and MECI (right) tessellated in a 100Ha hexagonal grid

## 7.2 Contiguity criteria

A final test was done to know the sensitivity to the raster contiguity criteria (Figure 15):

- Von Neumann neighborhood, where only the 4 orthogonal pixels are used to check the contiguity of patches
- Moore neighborhood, where in addition to the 4 orthogonal pixels, the 4 diagonal pixels are considered to check the continuity of patches

Reusing the matrix obtained for TAZ, the resulting Pearson Coefficient was almost the same for the 8 neighbor rule (0.84) that the one obtained with a 4 neighbor rule (0.85).



Figure 15: Individual patches, Von Neumann neighborhood (left) and Moore neighborhood (right)

### 7.3 Edge contrast visualization

Finally, to help visualize the relationships between Land Cover classes, a map was made to display the resulting contrasts of the edges involved in the calculations (Figure 16).

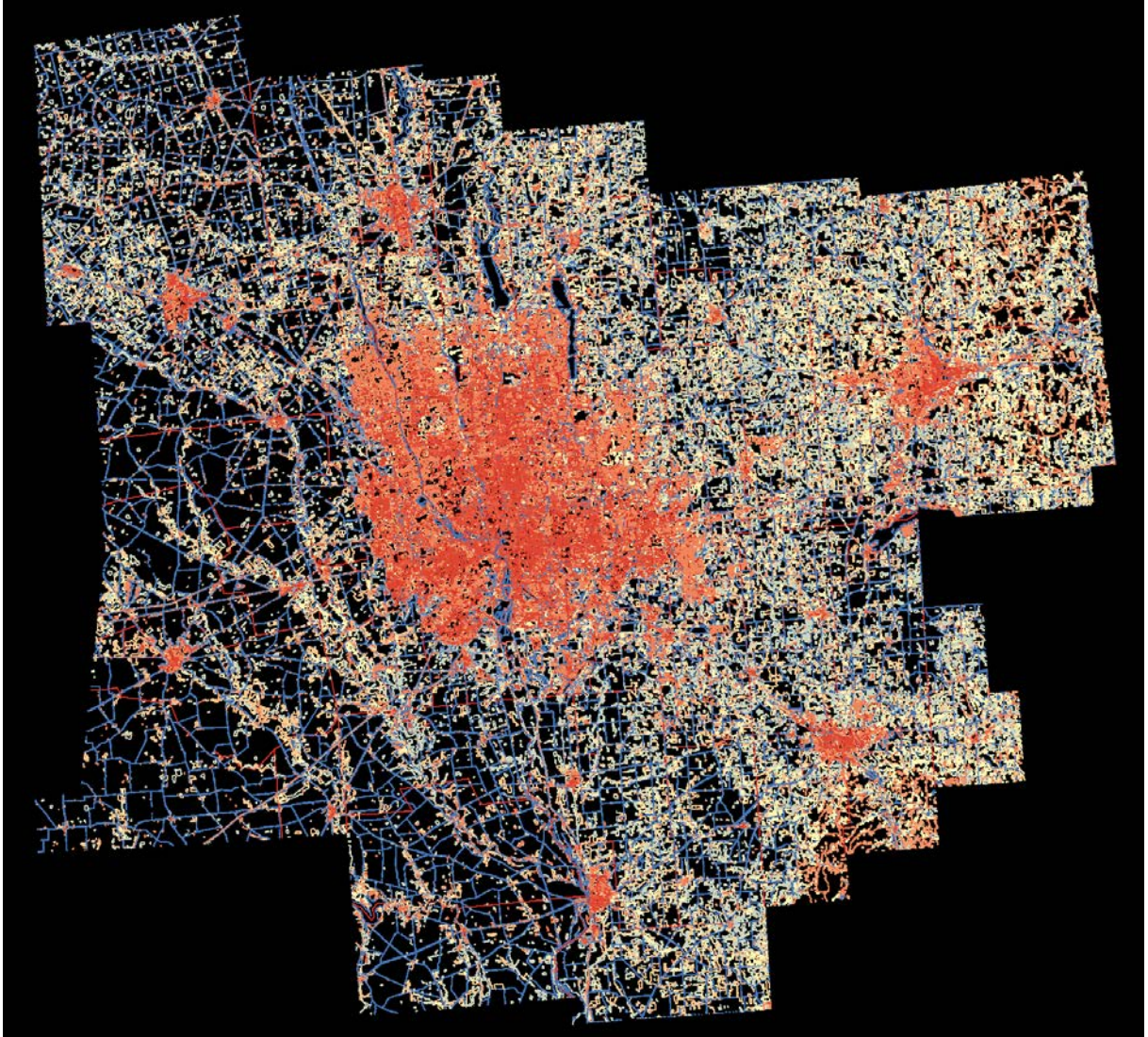


Figure 16: Edge contrast (Red = 0% / Yellow = 50% / Blue = 100%) for 2001 Land Cover and 2000 Tracts

## 8 Conclusions

The methodology used had some advantages compared to the use of off-the-shelf Spatial Pattern Analysis software. Most of these advantages arise from the fact that the methodology used was vector based instead of raster based.

- Projection can be considered to compare data in different latitudes or vast areas, instead of having a homogeneous rectangular grid
- It's possible to overlap grids of arbitrary separation and geometry to perform a multiscale analysis



In addition, since it was possible to repurpose the calculation engine, it was used in reverse to obtain the contrast values that better explain any given spatially distributed variable. Over the course of the investigation, the results obtained with the Large Scale SQP (Sequential Quadratic Programming) solver algorithm to perform the optimization were very promising.

The representation of the relationships with a Phylogenic Tree, and the matrix reordering technique allowed a better understanding of the optimal contrast data than when it was in a randomly ordered matrix form, and its applicability will be tested in future investigations.

We believe that the visualization of the edges with their contrasts in a map might be a helpful tool to develop a new, more robust indicator that is based on a contrast matrix, in order to incorporate it into the rule set of an Agent-Based model for the analysis of sprawl.

## 9 Bibliography

Anderson, J. H. (1976). A Land Use And Land Cover Classification System For Use With Remote Sensor Data. *U.S. Geological Survey Professional Paper(964)*. Washington, DC: USGS.

Bertin, J., & Bonin, S. (1975). *Graphique et le traitement graphique de l'information*. Paris: Flammarion.

Di Gregorio, A., & Jansen, L. J. (2005). *Land Cover Classification System, Classification concepts and user manual Software version 2*. Rome: Food and Agriculture Organization of the United Nations.

ESRI. (2012, July 13). *Polygon To Line (Data Management)*. Retrieved July 14, 2012, from <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//00170000003t000000>

Frontline Systems. (2012). *Standard Excel Solver - Dealing with Problem Size Limits*. Retrieved July 14, 2012, from <http://www.solver.com/suppstdsizelim.htm>

Jaeger, J. A. (2000). Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. *Landscape Ecology, 15*, 115-130.

Lu, G., & Guldman, J.-M. (2002, March 30). Landscape ecology, land-use structure, and population density: Case study of the Columbus Metropolitan Area. *Landscape and Urban Planning, 105*(1-2), 74-85.

McGarigal, K. (2012). FRAGSTATS 4.0 Tutorial.

McGarigal, K., & Marks, B. J. (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*. Gen. Tech. Report PNW-GTR-351. Portland (OR): USDA Forest Service, Pacific Northwest Research Station.

- McGarigal, K., Cushman, S., Neel, M., & Ene, E. (2002). FRAGSTATS v3: Spatial Pattern Analysis Program for Categorical Maps.
- Multi-Resolution Land Characteristics Consortium. (2011, October 27). *National Land Cover Dataset 1992 (NLCD1992)*. Retrieved July 14, 2012, from <http://www.mrlc.gov/nlcd1992.php>
- Multi-Resolution Land Characteristics Consortium. (2012, June 28). *National Land Cover Database 2001 (NLCD2001)*. Retrieved July 14, 2012, from <http://www.mrlc.gov/nlcd2001.php>
- U.S. Census Bureau. (2012, May 09). *Census Tracts Cartographic Boundary Files Descriptions and Metadata*. Retrieved July 14, 2012, from <http://www.census.gov/geo/www/cob/tr1990.html>
- United States Environmental Protection Agency. (2007, August 30). *NCLD 19920 and 2001 Land Cover Class Definitions*. Retrieved July 14, 2012, from <http://www.epa.gov/mrlc/definitions.html>
- United States Environmental Protection Agency. (2007, December 10). *NLCD Change (NLCD 1992 versus NLCD 2001)*. Retrieved July 14, 2012, from <http://www.epa.gov/mrlc/change.html>
- Valls, F., & Roca, J. (2011). *Optimizaci3n de la matriz de contraste de Mean Edge Contrast Index (MECI) con Excel Solver*. Barcelona: Technical University of Catalonia (BarcelonaTech).