AN INTEGRATED APPLICATION OF ZONING FOR MOBILITY ANALYSIS AND PLANNING: THE CASE OF PARIS REGION

L. Miguel Martínez, CESUR, Department of Civil Engineering, Instituto Superior Técnico, Technical University of Lisbon (martinez@civil.ist.utl.pt)

Ariane Dupont-Kieffer, Department of Economics and Sociology of Transports (DEST), Institut National de Recherche sur les Transports et leur Sécurité (INRETS) (adupont@inrets.fr)

José Manuel Viegas, CESUR, Department of Civil Engineering, Instituto Superior Técnico, Technical University of Lisbon (viegas@civil.ist.utl.pt)

ABSTRACT

In most transport planning studies one of the first steps is the definition of a zoning scheme with which the study area is divided and the corresponding space is disaggregated. There are no clear rules on how to carry out this operation in an optimal way, and the dominating practice is to do it based on experience, trying to mix a certain degree of within-zone homogeneity and the convenience of using administrative borders as zone limits.

Firstly, this paper starts by presenting a set of quality criteria for a general zoning scheme and an algorithm that constructs an initial zoning based on a sample of geo-referenced trip extreme points and improves it in successive steps according to those criteria. This kind of zoning fits perfectly well to traffic assignment purposes.

But this paper will investigate an improvement of this approach in order to give a better understanding of the mobility determinants and its externalities on the environment. In doing so, the new zone is determined not only by the trips generation and distribution but also constrained by other indicators. In our case, we have selected a combination of the following: 1/ air pollution emissions, 2/ population density, 3/ work and study density and 4/ public transport accessibility. The integration of these 4 indicators allows us not only to picture the mobility within the region and to identify at the very precise level the main zones of activities and traffic exchanges. This integration relates the picture to the land use and the clustering of the economic activities location at a very discrete level. Furthermore, it relates the density of the mobility in dense, large and economically dynamic urban area to its externalities in terms of air pollution.

In order to be effective for mobility analysis and policy purposes, this kind of approach cannot only rely on the cell grid unit but a hierarchical aggregation should be set up. This

aggregation allows analyses within the administrative and political boundaries but with a more disaggregated perspective.
A case study based on the Mobility Survey for the Paris Metropolitan Area in 2001 is developed to illustrate this new approach to zoning. The magnitude of those resulting improvements is very significant especially when compared with the already existing zoning as the department zoning or the IAURIF zoning. It then shows that more attention should definitely be given to this initial process in the transport planning studies.

Keywords: zoning, transport demand modelling, policy design and assessment.

INTRODUCTION

TAZs [Traffic Analysis Zone, Transportation Analysis Zone, Traffic Assignment Zone] have been pointed out as one of the keystones on Transportation Planning studies. Nevertheless, defining a good set of TAZ is still one of the transportation unsolved problems.

Transportation Planning studies have used transport demand models over the past four decades to forecast travel demand for short and long term planning. Transport Demand models typically follow a four-step process of trip generation, trip distribution, modal split, and network or trip assignment (Ortúzar and Willusen 2001). The steps are chained in a sequence, and the outputs of each step become inputs of the following step.

The key elements of a transport demand model are a study area divided into zones (called TAZ), and a transportation network in which each zone is represented by a centroid. TAZs represent areas from which and to which trips are allocated in a transport demand model. A centroid represents the ‘centre of activity’ of a zone and the origin and destination for all trips to and from the zone (Chang, Khatib et al. 2002). Trips generated between zones are assigned to the transportation network through connectors joining centroids and the physical network. Because zones and centroids are defined and used at the beginning of the modelling process, they affect the subsequent outputs, particularly the trip assignments on the network (Chang, Khatib et al. 2002).

In most Transportation Planning studies, a lot of effort is put in data collection, estimation of parameters and sophistication of models, but the issue of zoning rarely merits similar attention, normally being done on top of administrative units or ‘by common-sense’.

Dividing the territory in zones (for the purpose of Transport Modelling, and namely for the definition of O/D matrices), is a process of discretisation of space:

- With great advantages for the simplicity of the models and for the interpretation and communication of results.
- But with insufficiently recognized (and managed) problems of loss of information in the process.

Geographical information is lost in the process of substituting (concentrating) the real origins and destinations of trips (which occur over a continuous space) by an artificial point in each zone, the centroid. In doing this, two types of loss occur:

- Intra-zonal trips cannot be processed (as they now start and end in the same point).
- The travelling paths close to the real origins and to destinations of the trips are subject to large errors, implying that traffic load estimates in the lower hierarchy links of the transportation network are highly unreliable.

If loss of information was only geographical, the solution would be simplified in terms of possible solutions: Adopting a high number of small zones. But this leads to a head-on collision with the loss of statistical precision, due to the fact that O/D matrices are always estimated through sampling processes, most frequently through direct survey of travellers and /or by traffic counts on network links (Viegas, Martinez et al. 2009). Adopting many small zones would lead to two problems:

- Width of Confidence intervals of the flow estimates.
- Percentage of matrix cells with zero flow.

At the same time, the TAZ concept has not been clearly defined, and their multiple uses sometimes are confused. Different zoning systems can be defined for different uses. Clear examples of that are the zoning system developed for sampling and computation of the expansion coefficient in a Mobility Survey, the zoning system developed for Transportation and Urban Planning analysis, and the zoning system defined for transportation gas emissions in Environmental assessments, which do not necessarily have to be exactly the same, but are often considered in the scientific literature as only one.

Lack of rigor about the TAZ concept led to a non uniform terminology in the scientific literature. Most authors use the Traffic Analysis Zones term, but the domain of this terminology is not clearly defined. Some authors use the Transportation Analysis Zones term, usually associated with the Transportation and Land Use Planning. The term of Traffic Assignment Zones is also often used in scientific literature and technical studies linked to traffic forecasting models (e.g. transportation 4 step classical model). Other terms as Traffic Zones, Travel Analysis Zones or just zones are also used by some authors, normally associated with the O/D matrixes establishment and traffic assignment in the transportation 4 steps model (Martínez, Viegas et al. 2007; Martínez, Viegas et al. 2009).

Through four decades of research, the scientific literature has established some guidelines and constraints to the definition of TAZ. Table I presents a complete set of the constraints discussed in the literature. Some contradictions and difficulties in their implementation are discussed in detail.

Table I – Constraints to the definition of TAZ

Constraints	References
Trip generation/attraction homogeneity	(Baass 1981; O'Neill 1991; Ding 1994; 1998; Ortúzar and Willusen 2001; Chang, Khatib et al. 2002)
Contiguity and convexity of zones	(O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; 1998)
Compactness of TAZ shapes	(Baass 1981; O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; Chang, Khatib et al. 2002)
Exclusiveness (no doughnuts or islands) of zones	(O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; 1998)
Equity in terms of trip generation (small standard deviation across zones)	(O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; 1998)
Adjustment of TAZ boundaries to political, administrative, or statistical boundaries	(Baass 1981; O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; 1998; Ortúzar and Willusen 2001; Chang, Khatib et al. 2002)
Respect of physical separators	(O'Neill 1991; Ding, Choi et al. 1993; Ding 1994; 1998)
Decision makers' preferences are considered in determining the number of TAZ	(Ding, Choi et al. 1993; Ding 1994; 1998; Ortúzar and Willusen 2001)
Avoid main roads as zone boundaries	(Ortúzar and Willusen 2001)
Zone size is selected such that the aggregation error caused by the assumption that all activities are concentrated at the centroid is not too large (geographical precision)	(Ortúzar and Willusen 2001)
Minimization of intra-zonal trips	(Baass 1981; Crevo 1991)
Maximization of the statistical precision of the estimation of the OD matrix cells	(Openshaw 1977)

It is difficult to consider and implement all of the above criteria in a single process of TAZ design because some rules contradict others (O'Neill 1991; Ding, Choi et al. 1993). Some of these rules are a consequence of the use of a fixed zoning scheme over time, and of different study scales and purposes. In the scientific literature, it has been stipulated that transport demand modelling requires constant spatial data aggregation in TAZ along the entire process, from data collection to the trip assignment step (Chang, Khatib et al. 2002).

The number of zones problem can be solved through the development of a hierarchical zoning system, as in London Transportation Studies (Ortúzar and Willusen 2001), where sub-zones are aggregated into zones, which in turn are combined into districts, traffic boroughs, and finally a sector. This facilitates an analysis of different types of decisions at the appropriate level of detail (Ortúzar and Willusen 2001).

Unfortunately, predetermined zoning systems do not take into account the ongoing changes of land use (spatially and temporarily), which can deeply affect TAZ homogeneity and compactness, producing significant misestimates of trip generation and OD matrices (Edwards 1992).

These ongoing spatial and temporal land use changes demonstrate that current traffic management models, which are sometimes directly based on the results of the data collection process, and suppress the trip generation and distribution steps of the classical travel demand model (Ding 1998), should use a different zoning scheme from the transportation demand forecasting models (Openshaw and Rao 1995).

A better solution to the data collection zoning system constraint is the development of survey processes, in which the trip ends are geocoded. This data collection process requires the establishment of an initial zoning system for the sampling process, as with all other data collection processes, for the determination of expansion coefficients of each trip (Ortúzar and Willusen 2001); however, after the conclusion of this process, travel data is not attached to this or any other zoning system, but only to their geospatial coordinates (Chapleau 1997). Each study that uses this database can then develop a new zoning scheme that better fits the study scale and goals, resulting in more flexibility for the transportation analyst and a higher utility of the available database (Chapleau 1997; Trepanier and Chapleau 2001).

The complexity involved in TAZ definition underlies the need of further research on this topic. This research, building on previous research by the authors, tries to develop a new TAZ delineation methodology and a GIS based application, which does not only suits to one purpose (e.g. traffic assignment or 4 steps model definition), but that can also integrate under the same tool, different zoning schemes for different types of studies, allowing a more comprehensive analysis of mobility within Transport Planning Studies.

The new TAZ delineation approach is determined not only by the trips generation and distribution patterns, but is also constrained by other indicators that characterise other mobility and accessibility components. The integration of these indicators allows a holistic picture of mobility within a region and to identify at the very precise level the main zones of activities and traffic exchanges.

In order to be effective for mobility analysis and policy purposes, this kind of approach cannot only rely on the original cell grid unit but a hierarchical aggregation should be set up to ensure integration between the several resulting zoning schemes for different types of analysis that can be performed to the mobility and land use patterns of a region..

This new methodology has been tested on the Mobility Survey for the Paris Metropolitan Area in 2001 (EGT 2001), allowing an illustration of the capabilities of this new modelling and planning tool.

PARIS REGION AND PRESENTATION OF THE DATA BASE

Ile-de-France is one of the twenty-six administrative regions of France, composed mostly of the Paris metropolitan area with the size of 12,211 Km^2 (0,86% in Paris Municipality). Created as the "District of the Paris Region" in 1961, it was renamed after the historic province of "Isle de France" in 1976, when its administrative status was aligned with the other French administrative regions created in 1972.

The Paris Region corresponds also to a NUTS2 EU statistical division of the territory, and it is formed by 8 NUT3, which do also correspond the department administrative division. The Ile-de-France region is formed by 1,281 communes (20 of them inside Paris Municipality – Arrondissements), which is the French smallest administrative division.

Figure I – Map of the Ile-de-France communes

Paris region appears as the major labour pool of France. The Parisian region is defined by 8 departments: Paris (75), and the narrow belt (Petite Couronne) including 3 departments (Val de Marne (94), Hauts de Seine (92), and Seine Saint Denis (93)) close to Paris and directly surrounding the main city of France and the large belt (Grande Couronne) including 4 departments Seine et Marne (77), Essonne (91), Yvelines (78), Val d'Oise (95) up to 100 km from Paris, but the activities of which depend on Paris, the more important administrative and business centre in the region.

Brief presentation of the Paris Region Patterns

In 2006, according to OECD Regional Statistical Database, Paris region with a population of 11,490,969 inhabitants**,** accounts for one fifth of the French population and with a GDP of 523,138 mil.USD is the wealthiest part of France (GDP/ capita: 46,797 USD). Even if other parts of France, especially in the South of the Loire Valley, are economically dynamic and attractive, Paris region remains a dynamic pole with a GDP yearly growth (1995-2005) of 2.5% and the major labour pool, with an employment rate of 59.5% and an unemployment rate of 9.5% (2006). The active population is highly qualified with 26.2% of the population with at most lower secondary education (0-2 years after graduate), and 33.5% of the population with an upper secondary education (3-4 years after graduate) and at last 40.3% of the population with a tertiary education (5-6 years after graduate) and the job market is mainly oriented towards tertiary activities.

The analysis of the number of inhabitants according to the residential location shows that in all the departments of the Parisian region, the residential population has increased from 1976 to 2001, except in Paris (75) where the population decreased by 5%. The most important increase occurred in the Seine-et-Marne department (+ 58% between 1976 and 2001), in the East of Paris region and in the North West area, in the Val d'Oise department (+ 41% between 1976 and 2001). Even so, Paris remains the most populated place with 1.97 million inhabitants followed by the Hauts-de-Seine department, close to the La Défense business area, with 1.31 million inhabitants.

These latter departments are those with the less important non working population. The share of non working individuals has significantly and continuously increased in the departments of Seine-Saint Denis and the Val d'Oise department, in the North of Paris.

The jobs are mainly located in the centre and the west of the region, as to say in Paris and the Hauts-de-Seine, and housing places in the east. In the recent years, the firms have relocated their headquarters from inside Paris to the suburbs due the high level of rents in the French capital.

Due to the high density of the Parisian Region and a cluster location of business and administrative activities and housing, mobility is a key issue.

Brief presentation of the Paris Mobility Survey

A large survey is regularly conducted on transportation behaviour of the households living in Paris region. It contains various survey instruments in order to describe the means of transport owned by the household, commuting lifestyle, daily mobility, long distance trips, etc. Data collected in each of these instruments can be analyzed separately. The purpose of this survey is to describe the trips made by households and individuals who are living in Paris and the nearest departments, as well as their use of public and private transport

means. The survey is conducted for a week day. We focus on internal trips as to say the trips which origin and destination are within the Parisian region (Ile de France).

The analysis of the trips, the travel time budget and the travel distance budget, of the most wealthy and densely populated region of France, is based on the data collected by 4 Household mobility surveys EGT carried on the Paris region in 1976/77, 1983/84, 1991/92 et 2001/02.

For the 4 surveys, the sample size is around 18,000 households representative of Ile de France population and the answer rate is around 57% with an average of 10,000 respondents (10,478 household answers in 2001). The questionnaire contains three parts:

1. Household composition (HSD) (e.g. address, income, motorization, number of persons);
2. Surveyed individuals within the household (e.g. work, age, place of work);
3. Trip characterisation (e.g. Origin, Destination, Mode, Time, purpose).

It shall be noticed that the origin and destination of each trip are geo-coded allowing the analysis of the Parisian mobility within different zonings.

Due the fact that only the 2001 survey was geocoded, our research was carried only with this survey. This mobility survey contains 23,656 respondents from different households, sampling a total of 81,386 trips (0.23% of the total trips). The trip ends, and the residential location of the respondent are geocoded using a 300 square meters grid cell.

Figure II presents the spatial distribution of the origins of trips of the EGT 2001 mobility survey. It is easy to notice a higher concentration of trip origins within the central departments of the study area and mainly within the Paris Municipality.

It is also possible to identify from the figure the existence of four main axes of mobility to reach Paris centre, corresponding to locations with better accessibility to the public transport network and the road network.

There also a significant number of communes that do present a low number of trips generated, reflecting locations with lower population density and concentration of activities within the Ile-de-France region.

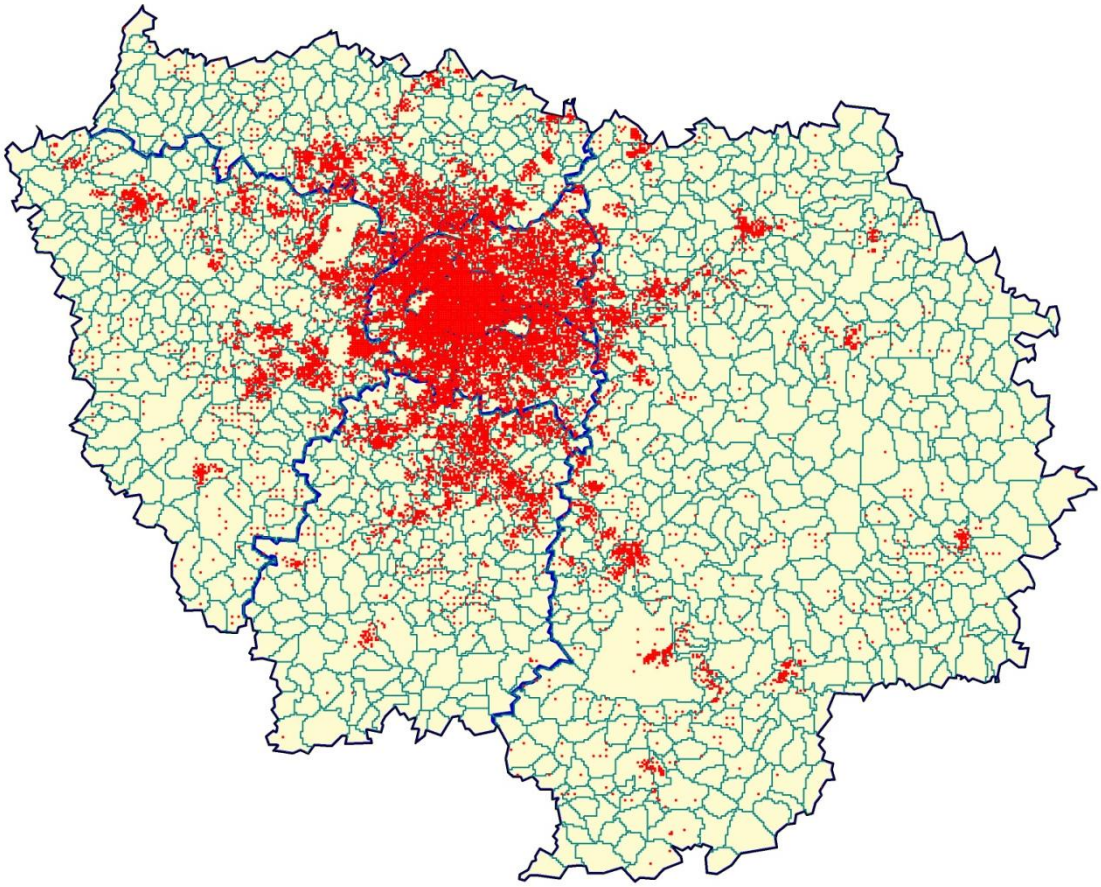


Figure II – Trip origins of the EGT 2001 mobility survey

TAZ DELINEATION ALGORITHM

This section of the paper presents a summary of the original TAZ Delineation Algorithm that has been developed in previous research of the authors (Martínez 2006; Martínez, Viegas et al. 2007; Martínez and Viegas 2009; Martínez, Viegas et al. 2009) and the application of the methodology to the Paris Region mobility survey (EGT 2001).

The methodology for zones delineation is defined to reduce the noise level of the data for traffic modelling, and at the same time, to minimise the geographical error of the trip end location.

The methodology defines zones such that:

- Zone boundaries correspond to places with very low generation of trips (reducing the probability of misallocating trips to zones near zones boundaries);
- intra-zonal trips are minimised;
- the definition of zones with a very low number of trips or very large area (high geographical error) is avoided;
- the density of trip production inside a zone should be as homogeneous as possible.

TAZ methodology

The methodology for the determination of zones starts by the aggregation of the geocoded trip ends (origin and destination) into a (relatively fine) cell grid. The cell grid can be variable and depends on the size of the study area and the precision intended for the study (Viegas, Martinez et al. 2009). As a reference, for the Lisbon Metropolitan Area, a square cell grid a 200 m side length was used.

A thin plate spline[1] was used to smooth the resulting surface and interpolate for cells without observations. The result of this analysis is presented in Figure III, where it is possible to identify the high concentration of trip ends in the Lisbon city centre and near some locations at the Lisbon municipality boundary.

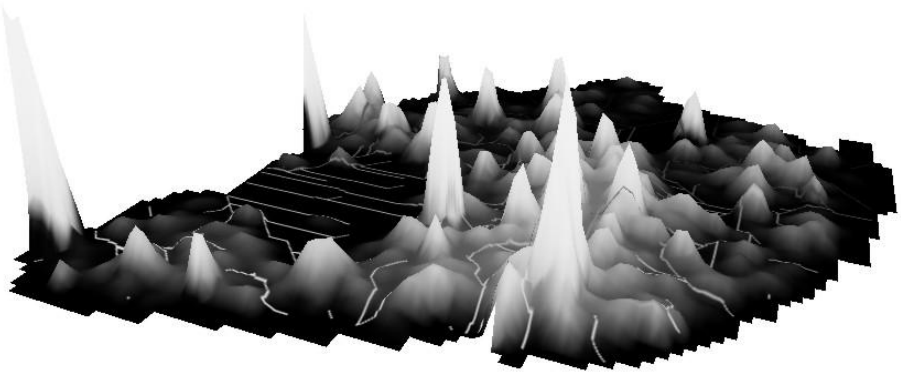


Figure III - Total origins and destinations in Lisbon Municipality 3D view

The TAZ delineation algorithm is defined by five different constraints and an objective function with two variables. These constraints can be divided in two groups: Those derived from the algorithm (4 constraints) and the geographic constraint for the TAZ border delineation (avoiding overlapping between zones). The constraints derived from the algorithm are:

1. The total origins (O_i) or destinations (D_i) of trips of each TAZ should be greater than 70%[2] of the average origins or destinations of trips by zone (total origins or destinations of trips divided by the number of zones). This is a requirement for quasi-homogeneity of trip quantities across zones, which indirectly controls the relative statistical error of the resulting zones.
2. Each TAZ area should be at least 70%[3] of the size of the influence area[4] of local predefined “highest peaks,”[5] which avoids the formation of zones with very low geographic precision.

[1] The thin plate spline is the two-dimensional analog of the cubic spline in one dimension. It is the fundamental solution to the biharmonic equation. The name "thin plate spline" refers to a physical analogy involving the bending of a thin sheet of metal.
[2] This value was obtained for the case study after calibration using different size grid cells.
[3] This value was obtained for the case study after calibration using different size grid cells.
[4] The surrounding area size is a parameter of the algorithm that is defined by the user.
[5] The concept of influence area of local “highest peaks” is defined as the minimum size that an analyst establishes for the specific modeling problem.

3. The average statistical (relative) error in the estimation of OD flow matrix cells should be lower than 50%, which directly controls the statistical precision of each TAZ (see (1) and (2)).

$$\left(\frac{SWCI}{p}\right)_{ij} = \sqrt{\frac{Z_\alpha^2}{n}\frac{(1-p_{ij})}{p_{ij}}} \tag{1}$$

$$\bar{x}_{STAT_i} = \frac{\sum_{i=1}^{N}\left(\frac{SWCI}{p}\right)_{ij}}{N} \tag{2}$$

where $\left(\frac{SWCI}{p}\right)$ is the relative error of OD matrix cell ij, p the probability of flow of the cell ij, n the total number of surveyed persons, Z_α is the z-score for having α% of the data in the tails, i.e., P(|Z| > z) = α (in the study α=0.05 and Z_α =1.96), $\bar{x}_{STAT_i}$ the average relative statistical error and N the total number of zones.

4. The number of zones should fall within the range previously defined by the analyst, which forces the algorithm to follow the analyst's preferences. This constraint is also used as one of the stopping criteria of the algorithm when it cannot be satisfied because the given range of is too high for the available number of cells and data.

The objective function of the TAZ algorithm contains two different components: The density of trips and the percentage of intra-zonal trips of each zone. This objective function tries, simultaneously, to optimise these variables by minimizing the standard deviation of the density of trips (across cells inside each TAZ), thus leading to more homogeneous zones and minimizing the sum of the percentages of intra-zonal trips across all zones.

These components can have minimum values at different points, with trade-offs solved through the use of a ranking function, which minimises the sum of the rankings of the two variables (see Figure IV).

If the ranking function retrieves the same result for different cases, the objective function considers the result with the lower trip density standard deviation to be the "most suitable". The decision tree used for the objective function is presented in Figure IV.

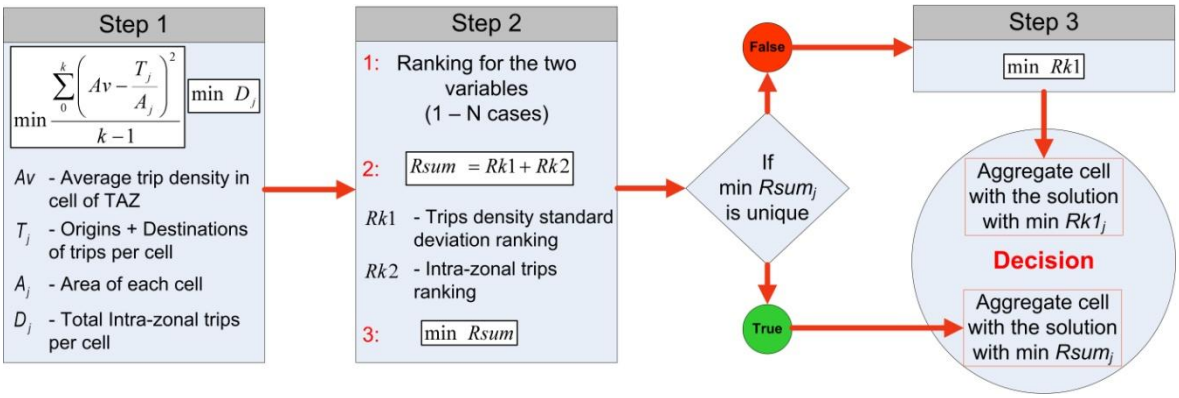


Figure IV – Decision tree for determining the optimum of the TAZ algorithm objective function

The algorithm requires setting some parameters, which are important for the establishment of constraints, as well as for the optimization of some additional features of the algorithm (e.g., the optimal number of zones for a given range, which depends on some macroscopic indicators that will be presented below).

These parameters are:

- The definition of the range of number of zones (searching for the most suitable solution) – compulsory input;
- minimum size of the influence area of local "highest peaks" for TAZ delineation;
- definition of a core problem area[6] – compulsory input;
- percentage of zones belonging to the core problem area;
- maximum proportion of areas between zones in the core problem area and in the "rest of the world;"
- some parameters of the indicators (macroscopic indicators) used to define the optimal number of zones of the given range. This optimal number is found with the help of a multi-criteria additive function, which is defined below.

If the user omits some of these parameters, and they are not compulsory, the algorithm uses the default values of those parameters[7] (e.g. percentage of zones belonging to the core problem area equal to 50%).

[6] The core problem area parameter allows the analyst to establish different levels of relevance inside the modelling area, which does not need to be used, and considers this parameter to be equal to the entire modelling area.
[7] For an overview of the TAZ delineation algorithm and a detailed presentation of the mathematical and data flow of the algorithm, presentation allowing a deeper understanding of its mechanics, see Martinez, L., Viegas, J., TAZ Delineation Application for Transport Planning Studies: A New approach Applied to the Ile-de-France-Technical report for DEST/INRETS, September 2009 (not published but available on request).

From the administrative zoning to the TAZ zoning: a better understanding of geography and mobility

After the initial presentation of the TAZ Delineation Algorithm, this point presents an assessment of the geography and mobility patterns of Ile-de-France region with two main goals:

- Assess the relevance of some indicators for the mobility analysis of the region in order to include them in the new TAZ Delineation Algorithm;
- Compare the resulting spatial structures with the existing planning zoning scheme for the Ile-de-France region (IAURIF).

With this purpose, a cluster analysis of these indicators was developed using the zoning scheme resulting from the TAZ Delineation Algorithm, having in mind the future application of this zoning scheme for the hierarchical procedure.

The zoning scheme used is presented in Figure V. This TAZ configuration presents 1500 zones and was developed for traffic assignment purposes. Table II presents some indicators of this zoning scheme. The zoning scheme presents overall good quality indicators, with the exception of the statistical precision indicators, derived from the large number of zones used for traffic assignment purposes.

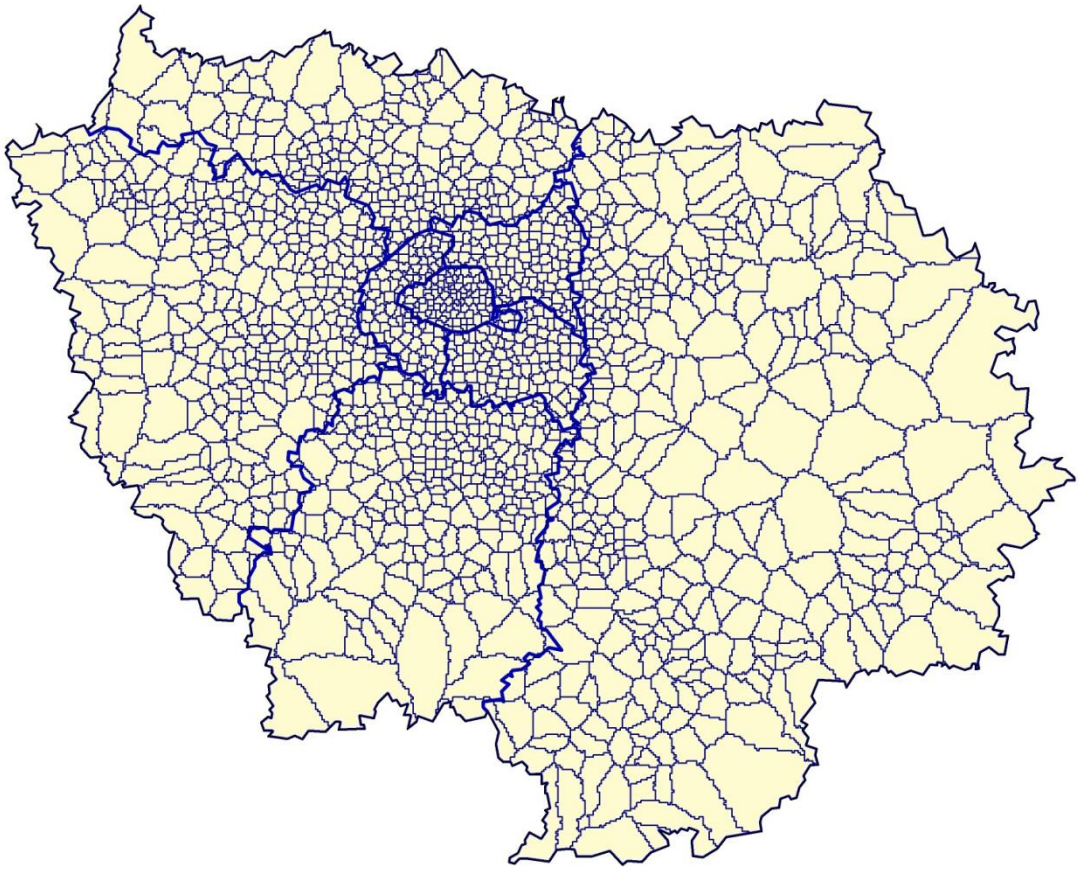


Figure V – 1500 zones scheme of the Ile-de-France region (TAZ Delineation Algorithm)

Table II – Summary of the 1500 zones scheme indicators (TAZ Delineation Algorithm)

Indicators	**Values**
Percentage of intra-zonal trips	28.701%
Average origin or destination per zone (%)	0.067%
Maximum origin or destination per zone (%)	0.513%
Maximum flow	61,325
Maximum flow percentage	0.178%
Average flow	15.347
Percentage of cells with null flows	97.166%
Percentage of trips in non significant O/D matrix cells	75.336%
Average of TAZ average statistical Relative error	41.077
Global Average Trip Density (trips/ha.)	279.581
Average Trips Density (Cv)	2.151
Average zone equivalent radius (m)	958.6

After a comprehensive analysis of a large set of indicators available from EGT data, the indicators selected to perform the clusters analysis resulted from the authors' observation of the main determinants to mobility within the available data and that might have transport and land use planning implications. These indicators are:

- Emissions factor (Kg CO2/Passkm.ha.)[8]
- Population Density (Inhabitant/ha.)
- Work and Study Trips Density (Trips/ha.)
- Average PT Accessibility (meters)

The population density is one of the key factors for mobility assessment due to its high correlation with trip generation/attraction. The spatial distribution of the population density is presented in Figure VI where a higher density within the narrow belt and some agglomerations along new development axis. The overlap between the population density and the public transport maps shows the relevance of this land use indicator and its relevance for the design of the public transport system (see Figure VI).

[8] In order to perform an environmental assessment of the EGT 2001 data, the data and coefficients defined on the EC TREMOVE model were used. TREMOVE is a policy assessment model. The TREMOVE 2.52 model has been developed by Transport & Mobility Leuven in a service contract for the European Commission, DG Environment. The most recent TREMOVE 2.7 version includes further developments made in the 6th Framework Programme GRACE project.

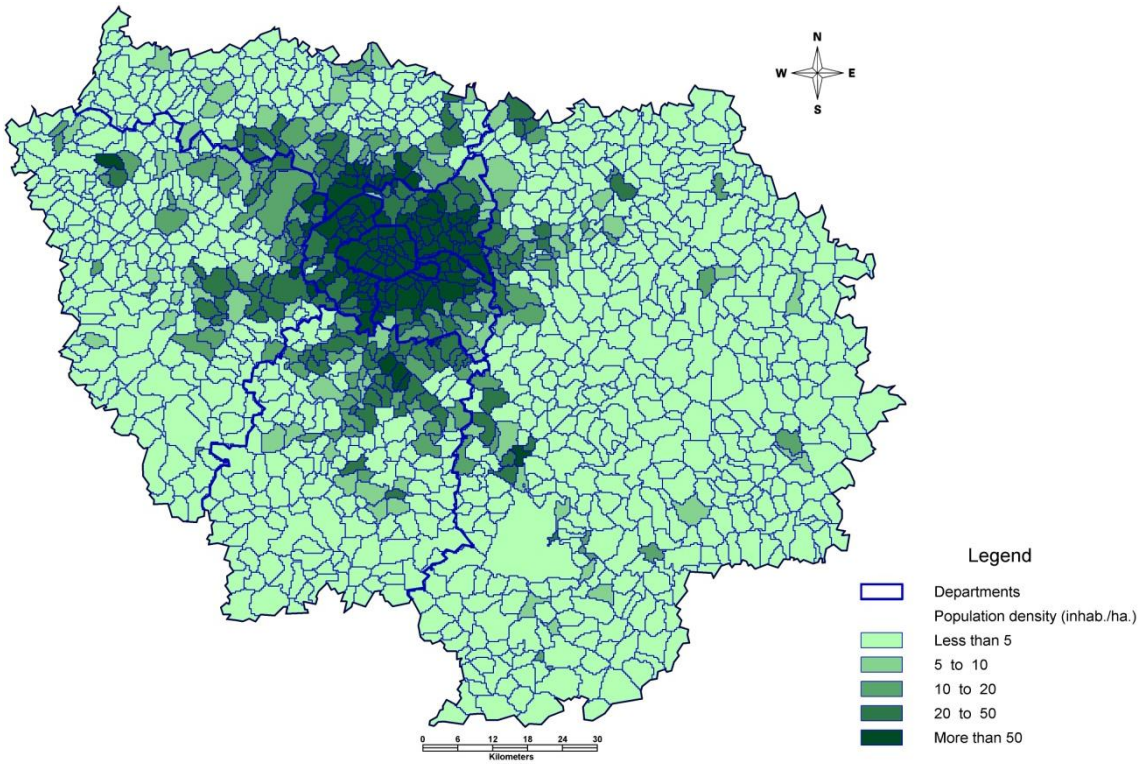


Figure VI – Population density spatial distribution of the Ile-de-France region

Public transport accessibility is another key variable for the assessment of mobility. This variables influences directly mobility from the supply side, but at the same time, influences largely households residential location and firms location options (Land Use-Transport feedback loop), presenting then also a key role in land use planning. The spatial distribution of this variable is presented in Figure VII, where we can observe a higher accessibility to public transport in Paris city and the first ring and along some main axes that are set by the RER service.

The work and study density is also a key determinant to mobility generation/attraction due to the importance of commuting trips in the total number of trips performed. Figure VIII presents the spatial distribution of this indicator, which overlaps significantly the population density map, showing a planning concern of developing mixed-use agglomerations in order to reduce the needs of mobility within the region. The greatest concentrations of work and studies are located inside Paris, mainly in its central Arrondissements.

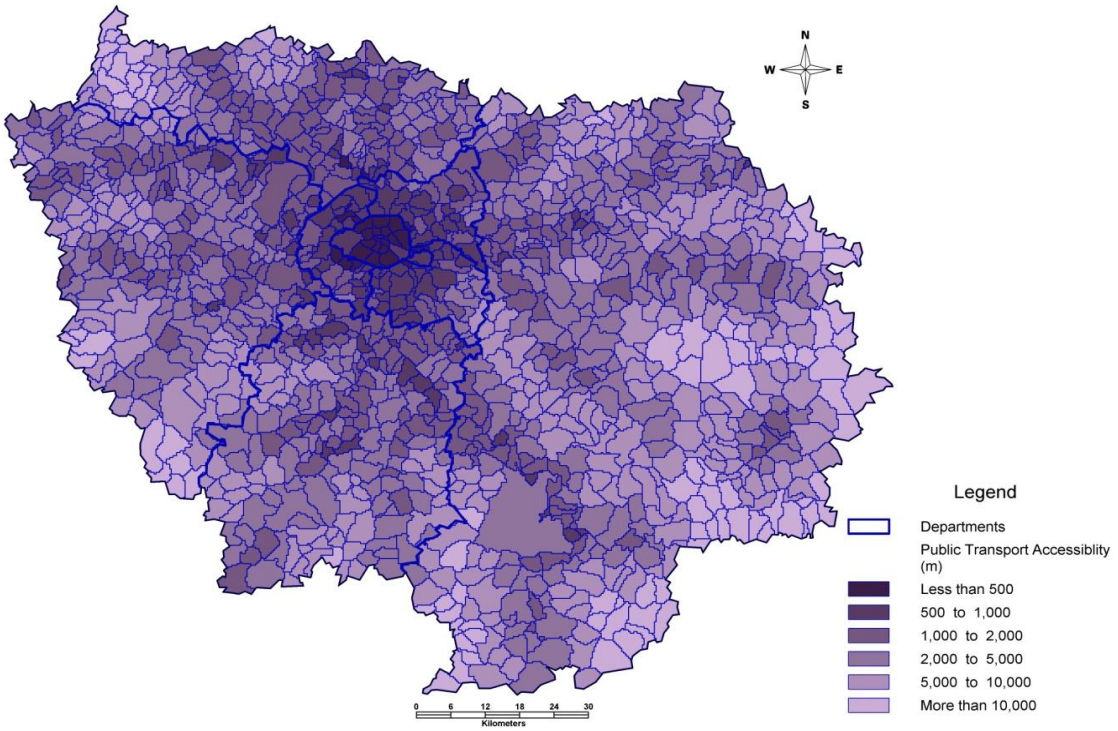


Figure VII – Public transport accessibility spatial distribution on the Ile-de-France region

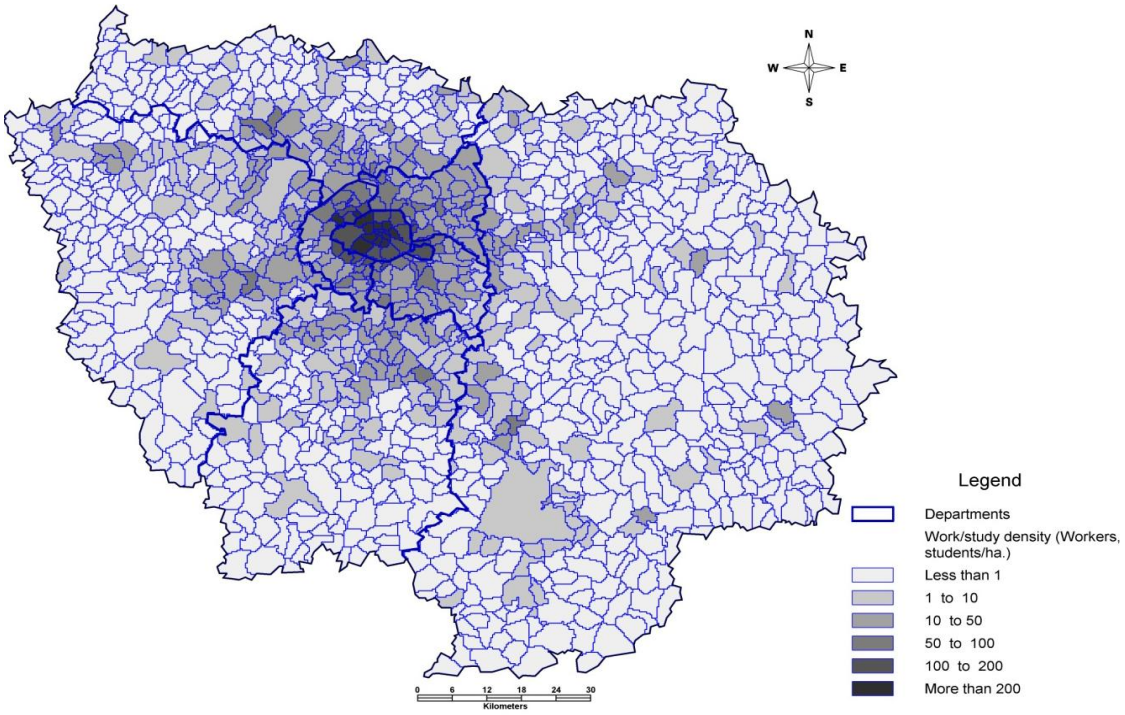


Figure VIII – Work and study density spatial distribution of the Ile-de-France

In order to introduce a more aggregate assessment of the data a percentile assessment of the different indicators was performed and three levels defined for the different indicators: high, medium and low. These values resulted from the following percentiles:

- High >= Percentile 75% of the indicator
- Percentile 25% of the indicator < Medium < Percentile 75% of the indicator
- Low <= Percentile 25% of the indicator

After an initial classification focusing on one of the 4 indicators, a factorial design was performed resulting in 81 different profiles (3^4 cases). These profiles result of the combination of the different levels of each zone for the different indicators.
Allocation of numbers to the profiles was ordered from high to low for each indicator, resulting that profile 1 includes high levels for the 4 indicators and profile 81 low levels for all the indicators.

The spatialisation of this analysis is presented in Figure IX. The figure confirms once again Paris and the agglomerations around it as the main centralities, followed by some secondary agglomerations developed along some new development axis. These results are aligned with the planning and functional zoning developed for the Ile-de-France region (IUARIF). This zoning scheme is presented in Figure X, where 8 types of zones are identified, showing the same developing trends identified in Figure IX. The main differences between the two zoning scheme results from the more disaggregate description of some mobility characteristics within the macro-zones identified in the IUARIF zoning.

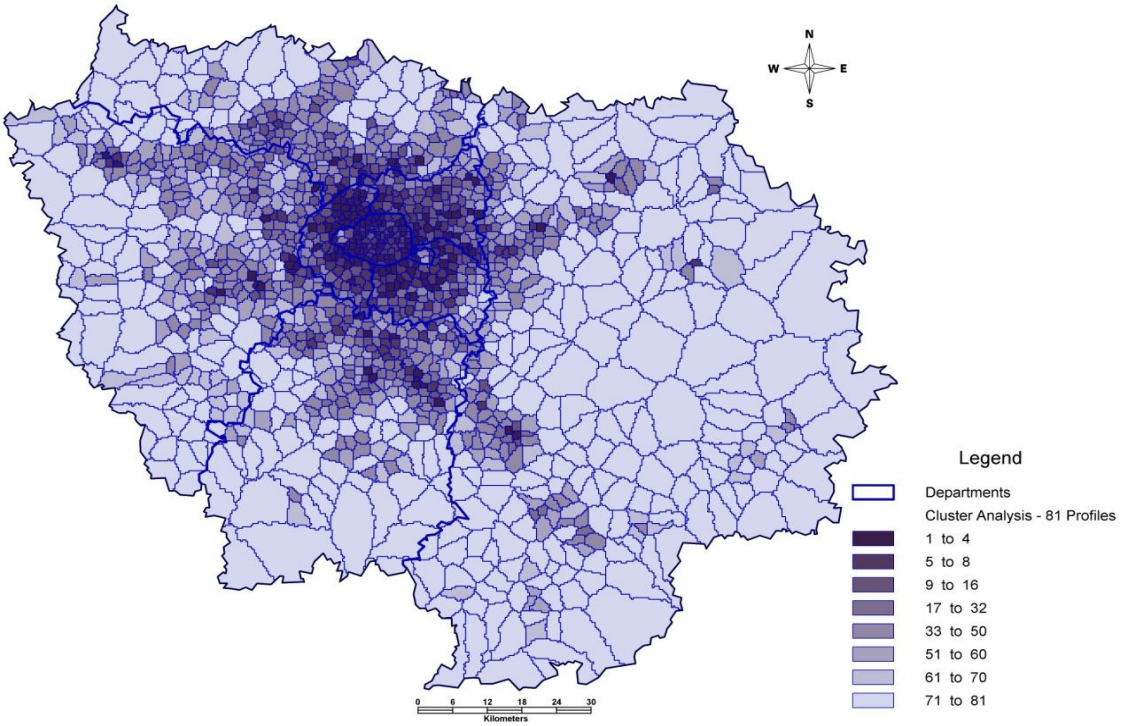


Figure IX – Cluster analysis with the 81 profiles for the Ile-de-France region

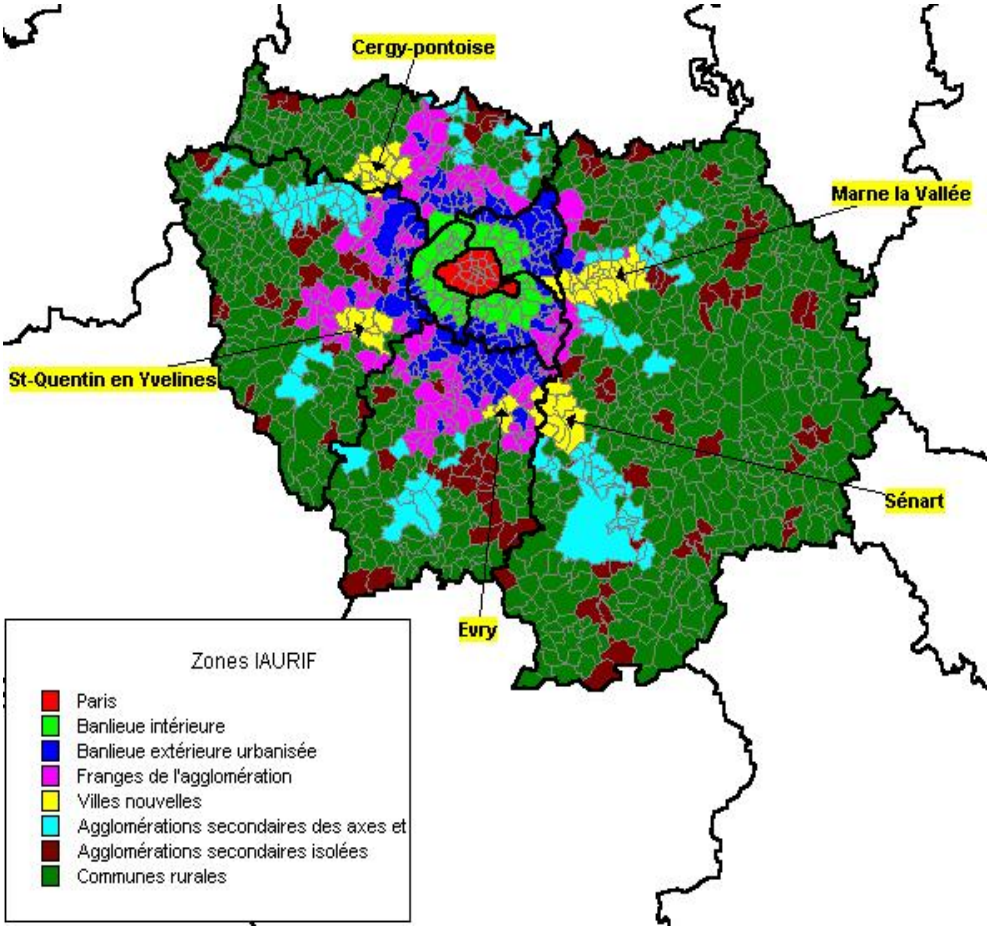


Figure X – IUARIF zoning of the Ile-de-France region

INTRODUCING A HIERARCHICAL ZONING APPROACH

The new approach introduces the concept of hierarchical zoning, due to the fact the zoning schemes result from previously aggregated zones and not from square grid cells. This new methodology allows integrating results from previous stages to make models computing and assessment compatible.

The concept of hierarchical zoning has been already use in literature and in practical application as in the London TAZ system (Ortúzar and Willusen 2001); however a systematic and algorithmic approach has not been already developed.

The new methodological approach for zones delineation is defined in order to keep the goals of the original algorithm (reduce the noise level of the data for traffic modelling, and at the same time, to minimise the geographical error of the trip end location), but also include some new indicators relevant for transport policy assessment in order to measure the homogeneity within the same zone.

The new methodology defines zones such that:

- Zone boundaries correspond to places with very low generation of trips (reducing the probability of misallocating trips to zones near zones boundaries);
- intra-zonal trips are minimised;
- the definition of zones with a very low number of trips or very large area (high geographical error) is avoided;
- the homogeneity of the selected indicators within the same zone is maximized.

The selected indicators to use to measure the zone homogeneity were defined in the previous section. These indicators are:

- Emissions factor (Kg CO2/Passkm.ha.)
- Population Density (Inhabitant/ha.)
- Work or Study Trips Density (Trips/ha.)
- Average Public Transport Accessibility (meters)

At the same time, the new aggregation unit are no longer grid cells, which will result in a different set of rules for the vicinity assessment.

New formulation of the TAZ

The main changes introduced by the new algorithm in order to use the new hierarchical approach are:

- The definition of new vicinity rules based on irregular boundaries overlap;
- The definition of a new objective function and constraints set for the algorithm.

The introduction of new vicinity rules in the TAZ hierarchical procedure is due to the possibility of use of irregular borders zones as seed for the aggregation process. The algorithm starts, as the original algorithm, by identifying the local “highest peaks” and their surrounding area, sorts them by decreasing magnitude, and then uses a local search algorithm for the design of the zones. This search is performed for each peak considering a defined set of rules. These search rules were developed to avoid the delineation of zones with complex spatial structures, which could undermine the applicability of the model in complex urban structures, as well as the assessment of its results, even for an experienced transportation analyst. This local search algorithm presents a new set of rules, which use as main indicator the percentage of TAZ boundary perimeter overlap between the already formed zones and the candidate zone, considering as a minimum overlap percentage 5% of the perimeter of the already formed TAZ.

The new rule for vicinity assessment, presents a simple mathematical definition, however the computational burden needed for the dynamic computing of the TAZ boundary perimeter overlap is considerably higher than with the original regular grid search rules.

The objective function of the Hierarchical TAZ algorithm contains two different components as original algorithm: the percentage of intra-zonal trips of each zone and an aggregated Euclidean distance measure of the new set of indicators within each zone.

The Euclidean distance function is computed after a hipergeometric standardisation of the indicators (N) as presented in (3). The distance function is then computed by the difference between the standardized indicators (NS) of the already aggregated zones and the zone under analysis. The distance of each indicator is then weighted by multiplying the standardized distanced by a weight factor of each indicator (wn) (considered equal among all indicators by default). This distance function (DNj) is presented in (4).

$$NS_i = \frac{max(N_i) - N_i}{max(N_i) - min(N_i)} \tag{3}$$

$$DN_j = \sum_i NS_i \cdot wn_i \tag{4}$$

This objective function tries, simultaneously, to optimise these variables by minimizing the distance measure of the new indicators (across cells inside each TAZ), thus leading to more homogeneous zones and minimizing the sum of the percentages of intra-zonal trips across all zones. These components can have minimum values at different points, with trade-offs solved through the use of a ranking function, which minimises the sum of the rankings of the two variables.

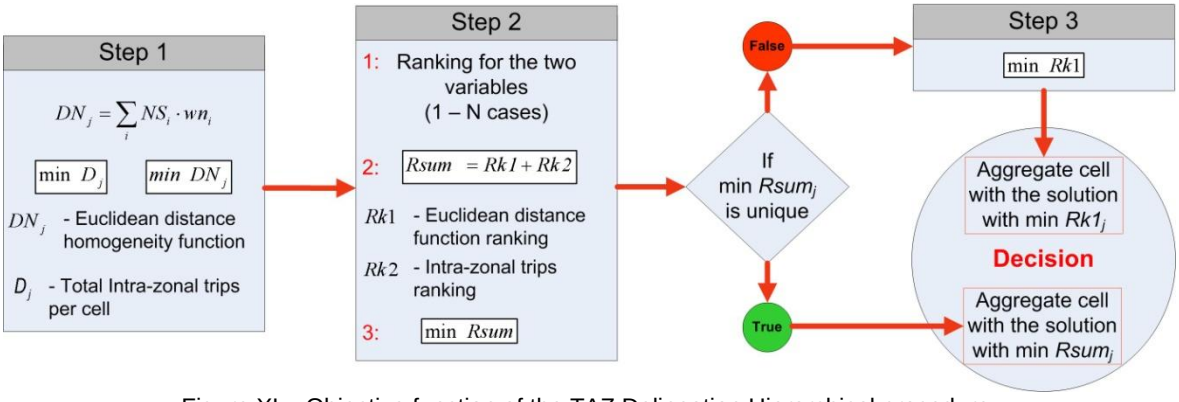


Figure XI – Objective function of the TAZ Delineation Hierarchical procedure

After the establishment of the new TAZ Delineation hierarchical procedure, an application of the new algorithm is presented below for the Ile-de-France region based on the EGT 2001 survey.

The hierarchical procedure will use as base zoning scheme the TAZ system with 1500 zones. This TAZ configuration was developed using the original TAZ Delineation Algorithm and had as main purpose the development of a zoning scheme appropriate to traffic assignment models.

In order to test the new hierarchical algorithm, two different tests were performed:

- A 50 zones scheme that could be applied for transport planning and energy consumption and emissions assessment in the transport sector;
- A 75 zones scheme for the same purposes.

A new mapping

The obtained results are presented in Figure XII and Figure XIII. The presented results map the spatial distribution of the emission factor, which leads to a configuration where the agglomeration ring around Paris and the some secondary agglomerations present the highest values of this coefficient. These results are coherent with the assessment presented in the previous section, which is due to the greater private car use in areas outside Paris.

It is relevant to notice that the borders of the resulting zones frequently cross the administrative limits of the departments, which indicates urban development and activity interactions that not always follow the administrative divisions.

The larger zones are located in more rural areas with more disperse land use patterns and lower trip generation/attraction rates. This fact leads to a trade-off between statistical precision and homogeneity of zones and the geographical precision, which for planning purposes is not as relevant as for traffic assignment models.

Comparing the two zoning schemes developed, a significant overlap between zones can be observed. The main differences result from the subdivision of some zones located within the narrow belt, which can present a higher geographical precision for a greater number of zones. The global indicators as the average equivalent radius change from 4646 m to 3842 m, and the average statistical error from 0.51 to 0.70. The new TAZ shows that the centroids of the zones are not the administrative centres but the economic centres which are close to transport networks nodes.

Both resulting zoning systems seem adequate for transport planning assessment.

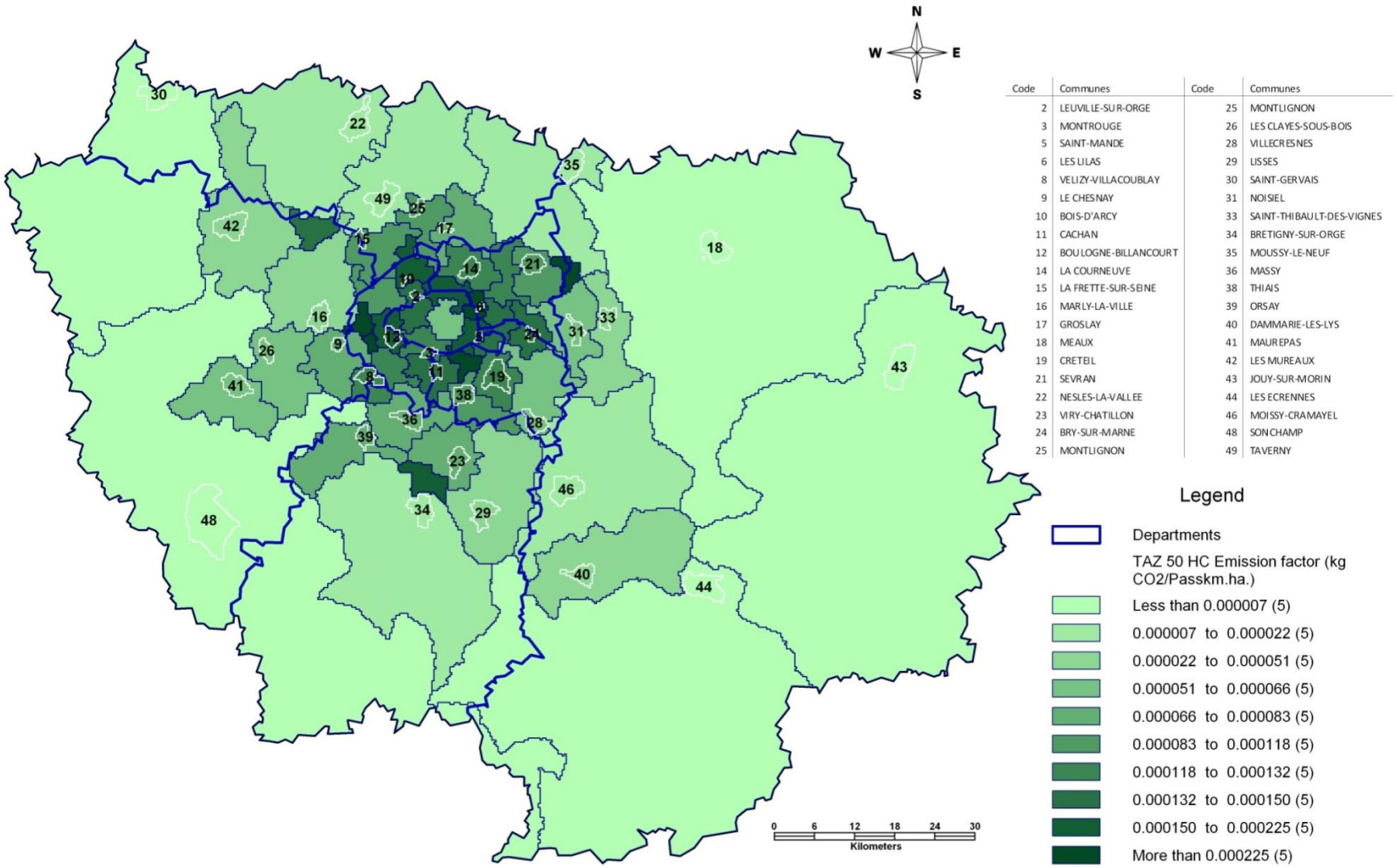
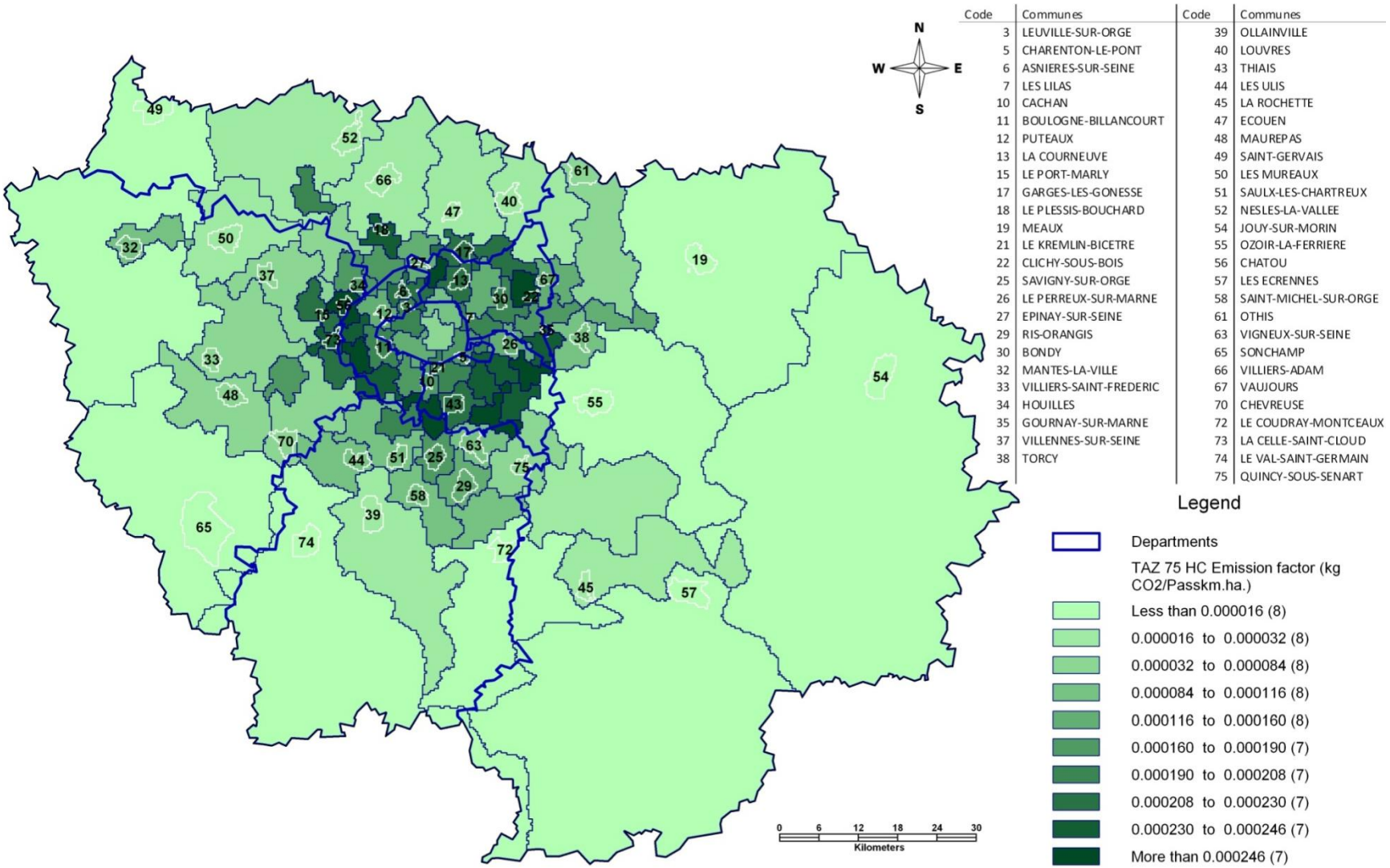


Figure XII – TAZ 50 zones (Application of the Hierarchical TAZ Delineation Algorithm)



Code	Communes	Code	Communes
3	LEUVILLE-SUR-ORGE	39	OLLAINVILLE
5	CHARENTON-LE-PONT	40	LOUVRES
6	ASNIERES-SUR-SEINE	43	THIAIS
7	LES LILAS	44	LES ULIS
10	CACHAN	45	LA ROCHETTE
11	BOULOGNE-BILLANCOURT	47	ECOUEN
12	PUTEAUX	48	MAUREPAS
13	LA COURNEUVE	49	SAINT-GERVAIS
15	LE PORT-MARLY	50	LES MUREAUX
17	GARGES-LES-GONESSE	51	SAULX-LES-CHARTREUX
18	LE PLESSIS-BOUCHARD	52	NESLES-LA-VALLEE
19	MEAUX	54	JOUY-SUR-MORIN
21	LE KREMLIN-BICETRE	55	OZOIR-LA-FERRIERE
22	CLICHY-SOUS-BOIS	56	CHATOU
25	SAVIGNY-SUR-ORGE	57	LES ECRENNES
26	LE PERREUX-SUR-MARNE	58	SAINT-MICHEL-SUR-ORGE
27	EPINAY-SUR-SEINE	61	OTHIS
29	RIS-ORANGIS	63	VIGNEUX-SUR-SEINE
30	BONDY	65	SONCHAMP
32	MANTES-LA-VILLE	66	VILLIERS-ADAM
33	VILLIERS-SAINT-FREDERIC	67	VAUJOURS
34	HOUILLES	70	CHEVREUSE
35	GOURNAY-SUR-MARNE	72	LE COUDRAY-MONTCEAUX
37	VILLENNES-SUR-SEINE	73	LA CELLE-SAINT-CLOUD
38	TORCY	74	LE VAL-SAINT-GERMAIN
		75	QUINCY-SOUS-SENART

Figure XIII – TAZ 75 zones (Application of the Hierarchical TAZ Delineation Algorithm)

CONCLUSIONS AND FUTURE DEVELOPMENTS

In most Transport Planning studies, a lot of effort is put in data collection, estimation of parameters and sophistication of models, but the issue of zoning rarely merits similar attention, normally being done on top of administrative units or "by common-sense". The results obtained show that zoning is not a trivial matter giving the significant consequences it may have for the generation of statistical and geographical errors. This study introduces a new methodology of zones delineation that deals with the existent trade-offs in the process.

The introduction of the new Hierarchical Delineation Algorithm enhances the capabilities of the original algorithm, leading to more flexible TAZ delineation process that can be used for multiple purposes.

The application developed for this study focuses on new indicators that are considered determinants of mobility. These new indicators are used to measure the homogeneity within a zone, leading to zoning schemes that are more suited for transport planning.

Applying the new hierarchical zoning procedure over the results of the original TAZ Delineation Algorithm, ensures the conservation of the "good traffic assignment properties" of the original algorithm and simultaneously gathers areas with similar values in relevant land use, accessibility and energy consumption attributes.

The results show a great potential in the use of this integrated approach that can easily combine previous studies to other layers of analysis and more comprehensive and holistic assessments. This step forward allows simultaneously achieving good zoning schemes for different purposes and models and using their outputs as inputs for different models due to the hierarchical design used.

The use of this algorithm for different case studies (Lisbon Metropolitan Area and the Ile-de-France region) means a step forward on the validation of this procedure. Testing in a different database from the original application allowed the tuning of the model and the assessment of the sensitivity of the results to different travel patterns and land use distributions.

Four main directions are envisaged for the next steps of this research:

- Look at the consequence of these different zoning strategies on the traffic load estimates on roads of intermediate and lower hierarchies, especially after the implementation of all the local refinements to the base algorithm.
- Investigate the consequences of these findings for the process of matrix estimation from traffic counts, being aware that the assignment of a trip to a zone directly affects the TAZ definition process. For this reason, studies based on traffic counts might have a simultaneous definition of the zoning system and the OD matrix, and not as a exogenous and pre-defined process. This iterative process until the convergence of

the zoning systems and OD matrix might have a similar structure to the k-clustering algorithms, where the centroids of the clusters change position with every iteration.

- Economically quantify the improvements in statistical precision and the reduction of percentage of intra-zonal trips obtained by the use of the algorithm. This significant decrease of information loss can lead to the use of smaller samples in data collection or to a significant increase of the quality of the data used in typical transportation planning studies, obtaining more robust results for the same input data and costs.
- Evaluate the possible application of the algorithm with other indicators resulting from zonal data rather then travel vector data (e.g., population vs. employment). Both resulting zoning systems seem adequate for transport planning assessment.

REFERENCES

Baass, K. G. (1981). Design of Zonal Systems for Aggregate Transportation Planning Models. *Transportation Research Record - Travel Demand Forecasting and Data Considerations* Vol.807, pp. 1-6.

Chang, K. T., Z. Khatib and Y. M. Ou (2002). Effects of zoning structure and network detail on traffic demand modeling. *Environment and Planning B-Planning & Design* Vol.29, No.1, pp. 37-52.

Chapleau, R. (1997). Conducting Telephone Origin–Destination Household Surveys with an Integrated Informational Approach. *International Conference on Transport Survey Quality and Innovation*, Grainau, Transportation Research Board.

Crevo, C. C. (1991). Impacts of zonal reconfigurations on travel demand forecasts. *Transportation Research Record* Vol.1477, pp. 72-80.

Ding, C. (1994). Impact analysis of spatial data aggregation on transportation forecasted demand. *Urban and Regional Information System Association (URISA) Conference*, Washington D.C, URISA.

Ding, C. (1998). The GIS-based human-interactive TAZ design algorithm: examining the impacts of data aggregation on transportation-planning analysis. *Environment and Planning B-Planning & Design* Vol.25, No.4, pp. 601-616.

Ding, C., K. Choi and T. J. Kim (1993). GIS-Based Traffic Analysis Zone Design. *3rd International Conference on Computers in Urban Planning and Urban Management*, Atlanta.

Edwards, J. D. (Eds.). (1992). *Transportation Planning Handbook*. Washington DC.: Institute of Transportation Engineers.

Martínez, L. and J. Viegas (2009). Effects of Transportation Accessibility on Residential Property Values. *Transportation Research Record: Journal of the Transportation Research Board*, No.2115, pp. 127-137.

Martínez, L., J. Viegas and E. Silva (2007). Zoning Decisions in Transport Planning and Their Impact on the Precision of Results. *Transportation Research Record* Vol.1994, No.1, pp. 58-65.

Martínez, L., J. Viegas and E. Silva (2009). A traffic analysis zone definition: a new methodology and algorithm. *Transportation* Vol.36, No.5, pp. 581-599.

Martínez, L. M. (2006). *TAZ Delineation and Information Loss in Transportation Planning Studies*. Master, Civil Engineering Department, IST- Technical Universty of Lisbon.

O'Neill, W. A. (1991). Developing Optimal Transportation Analysis Zones Using GIS. *ITE Journal-Institute of Transportation Engineers* Vol.61, No.12, pp. 33-36.

Openshaw, S. (1977). Optimal Zoning Systems for Spatial Interaction Models. *Environment and Planning A* Vol.9, No.2, pp. 169-184.

Openshaw, S. and L. Rao (1995). Algorithms for Reengineering 1991 Census Geography. *Environment and Planning A* Vol.27, No.3, pp. 425-446.

Ortúzar, J. and L. Willusen (Eds.). (2001). *Modeling Transport*. New York: John Wiley and Sons Inc.

Trepanier, M. and R. Chapleau (2001). Object-oriented and totally disaggregate analysis of origin-destination household survey data. *Canadian Journal of Civil Engineering* Vol.28, No.1, pp. 48-58.

Viegas, J. M., L. M. Martinez and E. A. Silva (2009). Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environment and Planning B-Planning & Design* Vol.36, No.4, pp. 625-643.