1 INTRODUCTION

Mobility need manifests itself by a trip between two points of territory anywhere located. A disaggregated analysis of these trips would produce insurmountable difficulties in modelling the transportation system (T).

Infact, a systemic approach to the study of mobility, provides for the modelling of territory that unfold in two subsequent steps: delimitation of study area and its following discreetization in traffic zones (Cascetta,1998) or TAZs (Traffic Analysis Zones) (McNally,2000, Dueker and Ton,2000).

In the hypothesis of discreetization of territory in zones and aggregation of origins and destinations of elementary trips by zone, it’s possible to realize a modelization coherent with the analysis level and that can be analytically dealt with, trough present computation resources.

In a network model, each zone is represented with a single point (zone centroid). In it all trips’ origins and destinations are located, which produce traffic flows among the zones (Meyer D.,Miller E.J.,2001), and physical, demographic and socioeconomic variables useful to define the system of activities (A) (McNally,2000).

Aggregation of origins and destinations of elementary trips by zone is not without consequences in analysis about system T, because in network models intrazonal trips are ignored. This implies an underestimate of trips’ number of the study area with a consequent underestimate of traffic flows on the network and effects on services’ planning.

A recent distribution model (Cascetta, 1998) considers the presence of a multiplicity of elementary destinations within a traffic zone through introduction of a “size” variable (ln M_d).

In this model one zone (d) is considered as a compound alternative of choice. From the assumed ipothesys it derives a formulation of systematic utility of zone (V_d), which will be influenced by the number of the possible destinations being in the zone:

\[ V_d = \Sigma_j \beta_j X_{jd} + \ln M_d \]

in which

M_d is the number of elementary destinations of zone;

X_{jd} is the j^{th} zone’s attribute.
It’s evident, then, as the number, the shape and not last, the homogeneity of socioeconomic characteristics (SE) (expression of capacity to generate and to attract trips), produce effects on the outcomes of analyses of the system.

In this note, it has been carried out a quantitative evaluation about effects of different levels of aggregation of socioeconomic and demographic data into TAZs on outcomes resulting from the simulation of a transportation system, by proposing a methodology for the design of zone number and for the aggregation of basic spatial units into homogeneous zones.

In Section 2 is described the state of the art about both zoning methodologies and the influence of zoning utilized on model’s results. In Section 3 is described the methodology adopted to solve TAZ design problem; it is also reported an application of the methodology applied to a real case with the analysis of the obtained numerical result; in section 4 are reported the conclusive considerations useful for following analysis and prospects of research on this subject.

2 THE STATE OF THE ART ABOUT TERRITORY MODELLING

The Zoning of a study area consists of two subsequent phases:

- Definition of spatial aggregation level of socioeconomic characteristics to achieve (determination of number of zones);

- determining shape and dimension of the zones respecting the spatial aggregation level determinated.

To determine the number of zones, Oppenheim (1995) asserts that no formal methods for defining the zones’ system exist. Characteristics of these zones must be decided on empirical bases in every specific situation.

In general the number of the zones inside the study area is closely connected to the analysis’ level that we want to achieve.

In short-term operations management programs, the number of zones to identify will be high. This will require the definition of a transportation network more detailed. In analysis of strategic problems a smaller number of zones, coherently with the supply model, is adopted (E. Cascetta, 1998; J. De D. Ortuzar, L.G. Willumsen, 1994).

Gehlke e Biehl already in 1934 had noted the tendency for correlation coefficients to increase with the level of aggregation of census tracts. Openshaw e Taylor (1979,1981), Openshaw (1984), Fotheringham e Wong (1991) have studied errors that can affect analysis based on aggregations of spatial data.

Openshaw (1977), in particular, presented a hierarchical heuristic procedure for Automatic Zoning Problem (AZP) by optimizing an objective function that was used to measure partition performance in terms of a predefined target value.

C. Ding (1998) analyzed impacts of socioeconomic data aggregation into traffic zones, at a national scale (South Korea), with gravitational demand models. He highlighted as the number of unsimulated intra-zonal trips is deeply affected from the number of zones especially when it is small. Ding deduced that exists a number of zones that isn’t useful to
exceed. In particular, with reference to network congestion index, Ding has remarked that exists a value of the number of zones under which the results calculated are quiet different from those obtained increasing zones’ number. Moreover, Ding noticed that over another value there is a little change of this index.

To physically delimit the zones, the criterions generally adopted (E. Cascetta, 1998; J. De D. Ortuzar, L.G. Willumsen, 1994) can be summarized in:

1. Omogeneity with respect to the socioeconomic characteristics;
2. compactness of TAZs’ shapes;
3. respect of administrative limits as census sections, municipal borders, etc.;
4. respect of physical geographic separators placed on territory as railways, rivers etc.
5. exclusiveness (no doughnut).

The elementary units in which are spatially organized socioeconomic data (basic spatial units - BSUs) (You J., Nedović-Budić Z., Kim T.J., 1996) in Italy coincide with ISTAT (National Institute of Statistics) census sections.

Procedures to design traffic zones through the aggregation of BSUs, on the base of definite criterions, are numerous and generally follow cluster analysis.

To aggregate two BSUs, their contiguity is necessary but not sufficient condition. In determining traffic zones is difficult that homogeneity and spatial contiguity exigencies can coincide. Choi and Kim (1995) individuated three fundamental methods with which you can face the BSUs’ aggregation:

a) the hierarchical heuristic approach;
b) the statistical approach;
c) the combined geographic information system (GIS) and statistical approach.

Cluster analysis techniques, to obtain the maximum homogeneity inside each zone, have instead been classified by M. Lorr (1983). Among these, the most efficacious are the agglomerative one and the iterative partitioning one. Jinsoo You, Zorica Nedović-Budić e Tschangho John Kim (1996) have carried out a traffic zones design method, given the spatial aggregation level to achieve, which is an hybrid of the two. Both methods rely on measuring the similarity between BSUs. Euclidean distance and correlation coefficient are the most commonly used similarity measures.

To obtain different zonings, C. Ding adopts an agglomerative methodology. Euclidean distance is the similarity measure used:

\[ d_{ij} = \left[ \sum w_r (X'_i - X'_j)^2 \right]^{0.5} \]

in which with X' are indicated the socioeconomic attributes of each BSUs while with w are indicated the weights associated to each attribute.

3 ADOPTED METHODOLOGY
The adopted methodology is articulated in subsequent steps:

1. construction of a geographic information system which will be of support to operations of BSUs aggregations;
2. individuation of a number $n$ of zonings each one characterized by a certain number $m$ of zones
3. supply model construction
4. calculation of $n$ OD matrix, one for each zoning defined
5. assignment to the transportation network of the travel demand determinated for each territory model adopted
6. comparison and evaluation of numeric results obtained with the proposition of a methodology for planning the number of zones to use in similar cases

In the first and second step, it is described the methodology adopted to aggregate socioeconomic data into homogenous traffic zones.

The study area considered for the application coincides with the whole municipal territory of the city of Bari. The area in this way delimited has an extent of 116 Km$^2$ and a resident population of about 342.300 accounting to the 1991 census. The road transportation network has been modeled with 677 real nodes and 1475 links.

### 3.1 Notations

In the continuation, will be denoted with:

- $i,j$ two generic BSUs localized in study area
- $X_{ik}^k$ the $K$th attribute of $i$th BSU
- $d_{ij}$ the euclidean distance from $i$ and $j$
- $\beta$ the generic weight given to the attributes used for aggregating BSUs into zones
- $G = G(N,L)$ the graph associated to transportation network
- $N$ the set of nodes of the network
- $L$ the set of links of the network $= \{(a,b)\}$ with $a,b \in N$
- $c_{il}$ the transportation link cost: $c_{il} = \alpha_1 t_{run} + \alpha_2 t_{wl}$
- $\alpha_y$ the generic homogeneity coefficient
- $f_i$ the link flow
- $\mathbf{C}$ the path cost vector
- $\mathbf{A}$ the link-path incidence matrix
- $\mathbf{C}$ the link cost vector
- $\mathbf{C}^{NA}$ the non-additive path cost vector
- $\mathbf{F}$ the path flow vector
- $\mathbf{f}$ link flow vector
- $G_s$ the grade of saturation
- $d_{od}$ the demand flow from the origin $o$ to the destination $d$
- $m$ the transport mode
- $s$ the purpose of the trip
- $h$ the time period of the trip
- $k$ the path of the trip
- $m_c(s)$ the mean number of trips for purpose $s$
3.2 The Geographic Information System

A general characteristic of Geographic Information Systems is their structure formed by layers overlapped. To each of these layers only a level of information is associated. The first layer, on which all the others are based, holds socioeconomic informations of the study area subdivided into census sections (1429 for the city of Bari accounting to the 1991 census). Their elaboration by aggregation procedures, allows to determine TAZs’ socioeconomic characteristics, useful to specify systematic utilities associated to zones in demand models.

The attributes immediately available from ISTAT census data, are: employees in industrial sector, employees in the public services sector, employees in the private services sectors, resident population older than 14 years, resident families, workers in industrial sector, workers in the public services sector, workers in the private services sectors, workers in the trade sector, resident population. Other attributes can be associated: places in upper secondary schools, places in lower secondary schools, places in primary schools etc. Moreover, for each zoning the routines of GIS software allow to determine and to associate to elements some physical characteristics of census sections, as area or the geographic co-ordinates of zone’s centroid.

The second layer holds the cartography of study area, georeferenced, that allows to locate the physical separators (railways, rivers, etc) and the road network.

On the third and the forth layer are represented respectively the graph of the network and the discretization of study area into traffic zones.

3.3 Zoning of study area

In the continuation, are explained the criterions adopted for the construction of a generic zoning and the procedure of clustering used to construct n zoning systems eachone characterized by a number m of zones.

At the beginning, a zoning characterized by an exiguous number of zones is considered through a procedure of aggregation of the BSUs based on the only criterion of physical separators. In subsequent phases, each zone is further subdivided into smaller zones. The criterion to adopt in aggregation of census sections, in these subsequent phases, must guarantee the construction of zones that show characteristics of homogeneity with respect to one or more attributes. The criterion of compactness is verified visually and with semplicity using GIS technology.

With this object in view, a procedure of K-means cluster analysis is used. This technique, to evaluate dishomogeneity among zones with respect to attributes, uses euclidean distance:

\[ d_{ij} = \left( \sum_k (X^k_i - X^k_j)^2 \right)^{0.5} \]
The attributes $X_i^k$ considered in this analysis are: number of employees, number of resident families, resident population, area of BSUs, geographic co-ordinates of the barycentre of the BSUs. The number of families is directly linked to the number of cars possessed while the geographic co-ordinates are indispensable to evaluate the contiguity among the census sections.

To each attribute $X_i^k$ a weight $\beta_i$ is associated, on the base of his relative importance, and a procedure of standardization is adopted to make sure that the results is not affected by the size of each attribute.

Consequently the expression of distance will become:

$$d_{ij} = [\beta_1(Z_{1j} - Z_{1i})^2 + \beta_2(Z_{2j} - Z_{2i})^2 + \beta_3(Z_{3j} - Z_{3i})^2 + \beta_4(Z_{4j} - Z_{4i})^2 + \beta_\nu [(X_j - X_i)^2 + (Y_j - Y_i)^2])^{0.5}$$

- $Z_1$ standardized variable of resident population ($\beta_1 = 0.15$)
- $Z_2$ standardized variable of employees ($\beta_2 = 0.15$)
- $Z_3$ standardized variable of number of families ($\beta_3 = 0.1$)
- $Z_4$ standardized variable of the area of BSU ($\beta_4 = 0.05$)
- $X_j, Y_j$ standardized geographic co-ordinates of the barycentre of the area of the BSU ($\beta_\nu = 0.55$).

The weights $\beta$ showed represent an hypothesis in which the criterion of contiguity is privileged. The results obtained in the application with the adoption of this values are coherent with the territorial reality used in application; a sensitivity analysis will can be developed to specify the value to attribute case by case.

The simulations, for the analysis of different alternatives of zoning, have been ten, characterized by a number of 10, 15, 20, 30, 40, 50, 60, 65, 70, 75 zones.

Two cases are shown in the following figures.

![Figura 1. 10 Traffic Analysis Zones](image1)

![Figura 2. 50 Traffic Analysis Zones](image2)

### 3.4 Modelling of transportation supply system

To analyze and to compare the ten hypothesis of zoning, it is necessary the construction of a supply model, which is considered constant by changing the level of discretization of territory and a demand vector $d$ variable with the number of zones.
The supply system, relative to the road network, is modeled with a synchronic network (Cascetta, 1998). The graph associated to the road network is constructed through a software GIS with the acquisition of the cartography normally available. This tool allows, with extreme facility, to select the transportation infrastructures which have an important role in connecting the traffic zones in the study area and external zones. GIS technology allows, besides, to associate to each element of the graph all the informations useful to describe the network: identification codes, cost functions, prominent flows, etc. With the hypothesis of congested network, the supply model is expressed formally by a relation that bind the link cost with path flows and the link flows with the path flows:

\[ C = A^T c(AF) + C^{NA} \]

\[ f = A F \]

\[ c_l = c_l (f_l) \]

In the application, non-additive costs \( C^{NA} \) have not been considered and as cost function has been adopted a separable one \( c_l = c_l (f_l) \).

The graph associated to the road network of the city of Bari is represented with 677 nodes and 1475 network links. The internal centroids and the relative connectors are variable with the number of zones considered in each simulation. The centroids outside the study area, in number of 11 constant for each simulation of zoning, are located about road sections of access to the study area and linked to the network with connectors.

3.5 The estimate of transportation demand and assignment to the network

The vector of demand \( d \), variable for each simulation, is estimated through a partial share model resulted from the application of four sub-models: emission model, distribution model, modal split model and path choice model:

\[ d_{od}(s,m,k) = d_o(s) \cdot \frac{SE_{T}}{T} \cdot p(d/osh) \cdot SE_{T} \cdot p(m/odsh) \cdot SE_{T} \cdot p(k/modsh) \cdot SE_{T} \]

For the model of emission is utilized a descriptive model:

\[ d_o(s) = \Sigma C_{m} c(s) \cdot N_o(C) \]

For distribution and modal, feet and cars, split models are utilized logit models:

\[ p(d/osh) = \frac{exp(\alpha V_d)}{\Sigma d' exp(\alpha V_{d'})} \]

\[ p(m/odsh) = \frac{exp(\alpha V_m)}{\Sigma m' exp(\alpha V_{m'})} \]

In both them the systematic utility \( V_d \) is valuable as a linear combination of attributes peculiar of the destination zone \( d \) or of the mode \( m \), on the base of parameters \( \gamma_k \):

\[ V = \Sigma_k \gamma_k X_k \]

For the model of choice of the path, on the contrary, the probit model is utilized.

The result of the application of the four step model are so much vectors of demand as TAZ alternatives.
Constrained the supply model and the demand model, the phase of assignment, in hypothesis of rigid demand, is carried out. The approach followed for the assignment is the User Equilibrium one, for congested network, with a SUE model
\[ f^* = f_{SNL}(c(f^*);d) \quad f^* \in S_f \]
and a resolutive algorithm MSA-FA.

### 3.6 Comparison and evaluation of numeric results.

For each TAZ alternatives produced the traffic flows on each link of the network have been estimated and network performance indicators as the total cost (TC = \( C^T F \)), the average cost (AC = \( C^T F / 1^T F \)), the congestion index PCP (\( I_c = \sum_i [(G_{s_i} f_i) / \sum f_i] \)), the number of saturated links (\( G_s > 80\% \)).

The results obtained for these indicators are represented in figures 3, 4, and 5 and are indicated in the subsequent table.

<table>
<thead>
<tr>
<th>n. zones</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total cost (hour)</td>
<td>75127</td>
<td>86750</td>
<td>18197</td>
<td>17847</td>
<td>15196</td>
<td>16907</td>
<td>16278</td>
<td>13898</td>
<td>14183</td>
<td>13348</td>
</tr>
<tr>
<td>Average cost (sec)</td>
<td>3734</td>
<td>4749</td>
<td>936</td>
<td>928</td>
<td>805</td>
<td>901</td>
<td>838</td>
<td>715</td>
<td>715</td>
<td>696</td>
</tr>
<tr>
<td>Average cost with C=0</td>
<td>940</td>
<td>772</td>
<td>616</td>
<td>569</td>
<td>588</td>
<td>614</td>
<td>622</td>
<td>501</td>
<td>492</td>
<td>492</td>
</tr>
<tr>
<td>Number of saturated links</td>
<td>341</td>
<td>200</td>
<td>140</td>
<td>158</td>
<td>89</td>
<td>99</td>
<td>86</td>
<td>86</td>
<td>99</td>
<td>101</td>
</tr>
<tr>
<td>Indicator of congestion PCP (%)</td>
<td>88.74</td>
<td>83.23</td>
<td>53.91</td>
<td>56.73</td>
<td>48.25</td>
<td>50.81</td>
<td>48.39</td>
<td>47.88</td>
<td>48.45</td>
<td>47.79</td>
</tr>
</tbody>
</table>

Table 1. Network performance indicators

The analysis shows a threshold of number of zones beyond which the values of network performance indicators present a stability of results included in a range of ±20% around the mean between the maximum and the minimum of results obtained in the simulations. This variability of results is coherent with the stochasticity of the model of assignment used (parameter of the variance of the costs of the Probit model equal to \( \theta = 10 \) s, n. SNL iterations= 50) and of the stop test of the algorithm MSA (M.G. Binetti, 1996).

For the city of Bari the threshold is individualized in 20 zones.

The results, moreover, show coherence with other analysis effected in the case of extraurban networks with estimation of demand effected with a gravitational model (Ding, 1998).
4 CONCLUSIONS
The discreetization of territory and the aggregation of origins and destinations of elementary trips in zone’s centroid produce effects on simulation of the transportation system. In this note these effects have been analyzed through different levels of discreetization (from 10 to 75 zones) of the study area (city of Bari) into traffic zones. The construction of the scenarios of simulation has been made possible thanks to use of a software GIS which has allowed the management, visualizations, aggregations and disaggregations of elementary informations about both the socioeconomic characteristic of territory and the elements of the transportation system. It has allowed, moreover, a first simple phase of initialization of zoning procedure thanks to the capacity to represent territorial data. The subsequent phases of construction of different zones, for each level examined, has been effected on the base of a cluster analysis through the use of a K-means technique. With the hypothesis of the same supply model, the esteemed transportation demand, in all analyzed cases, has been assigned to the network in the hypothesis of a SUE-PROBIT model and algorithm of the type MSA-FA. A variation of the supply model, for hypothesis constant
with the variation of the simulated scenarios, could be aim of extension of the simulations to acquire larger knowledge also about network effects.
The traffic flows esteemed (veic/h) on each link of the network have allowed to calculate network indicators as total cost, average cost on the network, the congestion index PCP the number of saturated links (Gs > 80%).
The analysis shows a threshold of number of zones beyond which the values of network’s indicator present a stability of results included in a range of ±20% around the mean between the maximum and the minimum of results obtained in the simulations.
The individualization of the threshold would allow a priori for the analyst a better definition of the modelling of the system and considerable economies of time, in elaborations, and resources, in the tools for calculation with a consequent reduction of the analysis’ costs.
For this end, an analysis extended to other cities of little, medium and big dimension could supply useful informations. Moreover, for the individualization of this threshold, from this kind of analysis relations between the level of discretization of territory and socioeconomic quantities, as resident population, demographic density, extent of study area, could be determined, if they exist.

REFERENCES

Baass K. G. Tecnique Dagregation de Zones en Planification des Trasports. Ecole Polytechnique, Campus de l’Université de Montreal

Bennion Wayne, M. And O’Neill, Wende A. Building Transportation Analysis Zones Using Geographic Information Systems. Transportation Research Record # 1429, pp. 49-56

Berglund S. Experience from GIS Based Travel Demand Modelling Royal Institute of Technology Stockholm, Sweden


Goodchild M.F. *GIS and Transportation: Status and Challenges* Department of Geography, University of California, Santa Barbara


