MODIFIABLE AREAL UNIT PROBLEM (MAUP) EFFECTS ON TRAFFIC ANALYSIS ZONES (TAZ) DELINEATION

L. Miguel Martínez
CESUR
Department of Civil Engineering
Instituto Superior Técnico
Lisbon Technical University
Av. Rovisco Pais
1049 - 001 Lisboa, Portugal
E-mail: martinez@civil.ist.utl.pt

José Manuel Viegas,
CESUR
Department of Civil Engineering
Instituto Superior Técnico
Lisbon Technical University
Av. Rovisco Pais
1049 - 001 Lisboa, Portugal
E-mail: viegas@civil.ist.utl.pt

Elisabete A. Silva,
School of Engineering
Catholic University, Portugal
Estrada Octávio Pato
2635-631 Rio de Mouro, Sintra,
Portugal
E-mail: elisabetesilva@fe.ucp.pt

KEYWORDS: TAZ (Traffic Analysis Zones); MAUP (Modifiable Areal Unit Problem), Traffic Modeling; GIS

ABSTRACT

Transportation analysis is typically thought as one kind of spatial analysis. A major point of departure in understanding problems is the recognition that spatial analysis has some limitations and impediments. Among them, modifiable areal units and boundary problems are indirectly or directly related to transportation planning analysis through the design of traffic analysis zones (TAZ). The modifiable boundary and the scale issues all should be given specific attention during the predetermination of TAZ because of the effects these factors exert on statistical and mathematical of spatial pattern (MAUP). The results obtained from the study of spatial data are not independent of the scale and the aggregation affects implicit in the choice of zonal boundaries. The delineation of zonal boundaries of TAZs has a direct impact on the reality and accuracy of the results obtained from transportation forecasting models.

In this article, the MAUP effects over the TAZ delineation and the transportation assignment models are measured and analyzed using different grids (in size, and origin location). This analysis was developed building an application integrated in commercial GIS software and using a case study (Lisbon Metropolitan Area, AML) to test its implementability and performance. The results reveal the conflict between statistical and geographic precision, and its relationship with the loss of information in the traffic assignment step of the transportation forecasting models.

INTRODUCTION

The problem emerges from a common practice in transportation analysis to collect data by geographic zones, such as census blocks or census tracts. These datasets may also be further aggregated to a spatial zoning system that is chosen for a particular application. These spatial units are often arbitrary and reflect the needs of the analyst and not an empirical geographic process.

A transportation application where spatial aggregation is common practice is travel demand modeling. While many transportation textbooks provide good guidelines for designing these zoning systems (e.g., Ortúzar and Willusen 1994), the transportation analyst usually defines these systems based on intuition and other difficult to codify rules. This can affect the results from the travel demand analysis since each boundary change affects the proportion and pattern of inter-zonal trips versus intra-zonal trips, with the latter trips not captured in the modeling process (Ding C., 1994).

The modifiable area unit problem (MAUP) occurs when the spatial zoning system used to collect and/or analyze geographic data is “modifiable” or arbitrary. Since the spatial units are arbitrary, the results from the analysis based on these units may be arbitrary, that is, an artifact of the spatial units rather than reflecting the true underlying geographic process.

MAUP effects can be divided into two components. Scale effects result from spatial aggregation of the data. Zoning effects relate to the changes in the spatial partitioning at a given level of spatial aggregation (Openshaw and Taylor, 1979; Wong and Amrhein, 1996).

Openshaw and Taylor (1979 and 1981) explained that is possible for data users to exert some influence over or minimize the impact of the MAUP. To achieve this, it is recommended that analysts:

- Start from the smallest division available, or the smallest they can process;
- Aggregate these in a fashion relevant to their investigation; and
- Assess the repetitiveness of their results for several aggregations.

Although this does not completely solve the MAUP it does allow analysts to exert corrective influence over the problem rather than ignoring it. In the past, most analysts ignored MAUP effects since the data and tools to deal with these
effects were not available. However, the digital cartographic data and GIS tools are now available for assessing the effects of zoning systems and developing optimal zoning systems for particular applications as areal interpolation and boundaries delineation (Eagleson S. et al., 2002; Xie Y., 1995; Duckham M. et al., 2001; Hall O., Arnberg W., 2002), or more specific studies like road accidents (Thomas L., 1996) and residential location (Guo J., Bhat C. 2003). These methods can be used for traffic analysis zones (TAZ) design in transportation planning and analysis.

Some developments have been made on this field, for example Openshaw (Openshaw, 1977) devised the automated zone design program (AZP) for investigating the MAUP. During the mid 1990s using new technology, digital data and improved algorithms, AZP was further refined and extended forming the Zones Design System (ZDES) (Openshaw and Rao, 1995; Openshaw and Alvandies, 1999). These zone design systems have been developed to allow the data analyst the freedom to start with data at one scale and then reaggregate it to create a new set of regions designed to be suitable for a specific purpose independent of the collection boundaries used (Openshaw and Rao, 1995).

Another approach is to assess MAUP effects, in order to develop this method is to perform a sensitivity analysis by changing the zoning system and rerunning the analysis. For example, Chou (1991) recomputes a spatial autocorrelation method at different levels of map resolution to assess the aggregation effects. Moellering and Tobler (1972) develop a scale-dependent analysis of variance technique. Batty (1976) develops a general information statistic to measure aggregation affects in spatial analysis. Batty and Sikdar (1982a, 1982b, 1982c, 1982d) use these statistics to assess aggregation effects in spatial interaction models.

In this article the MAUP effects are analyzed on traffic zoning methods using elemental grids of different dimensions. A methodology of analysis of the MAUP scale effects is developed and its consequences for the traffic demand modeling, specially underlying the geographic and statistical conflict when dealing with this kind of data. For this analysis it was used as case study the AML (with approximately 320000 ha. of area), using as dataset a 1994 Mobility Survey of this metropolitan area (TIS, 1994).

The Mobility Survey used for this article has a sample of 30,681 individual surveys that describe daily individual trips along one week, resulting in a total of 58,818 trips, having all the origin and destination of these trips georeferenced. Each survey has a coefficient that reflects its representatively on the total daily week trips made on the AML (resulting in a total 11,124,776 trips on a week day).

ANALYSIS METHODOLOGY

The analysis of the MAUP effects described at this article, result of the need of minimization and categorize these effects on the TAZs delineation. For this purpose, it was developed a GIS based application in Geomedia Professional that makes its analysis over a grid of variable dimension and form.

The analysis of this transportation and at the same time spatial problem had as background information some traffic modeling basic concepts important for TAZs delineation:

- Geographic precision in the localization of origin and destinations of trips and statistical precision are inversely proportional;
- The intra-zonal trips are not assigned to the network, therefore these trips are lost information for the traffic demand modeling that should be minimized on the O/D matrixes;
- The cells of O/D matrix of low values have high statistical errors;

The first step of this methodology is the insertion of origin and destination of trips of the Mobility Survey on GIS software (in this case Geomedia Professional), the geographic shape and data available for the AML and the analysis grids.

At the establishment of the different dimension and forms of the analysis grid was a requirement. The initial cell-form chosen was the square, because the universal use of this form and its easiness of manipulation. This paper reports the analysis performed using square cells, nevertheless, as already tested, it can also be made for other geometric shapes as the hexagon. Relatively to the grid dimensions, it was defined an upper and a lower limit of the square side and a step size of variation. The lower bound of the side of the square used was 200 meters, that could represent the average distance of urban highway intersections, and also for computational reasons because a smaller size of grid would make the GIS based application developed for this study very slow (a square of 200 meters of side, represents a total of 250,000 elements to analyze for all th AML region). The upper bound chosen for the side of the grid square was 2000 meters, which could represent the average distance of inter-urban highway intersection, and is a multiple of the lower bound, resulting in a 200 meters variation step and 10 different grid of analysis.

With all the inputs available, the GIS application sets two kind of effect analysis: the aggregation and scale effects of the grid over the data (in this case trips) using different indicators to measure it (the application outputs), and the effect of the location of the grid origin on the indicator results.

The selected indicators that were used as base to measure the variation on the data aggregation were:

- Maximum cell value at the O/D matrix;
- Average cell value at the O/D matrix;
- Percentage of cell without flow in the O/D matrix;
• Percentage of intra-zonal trips of the O/D matrix;

• Percentage of cells with values lower than L1 (value established in 200 trips, because is the average trip coefficient, and that represents the cell with high statistical errors or no statistical mean);

• Percentage of cell with values higher than L2 (value estimated in 30% of the maximum cell value, and that represents the cells with low statistical errors or higher statistical mean);

• Ranking chart of the cells of the O/D matrix;

• Number of grid cells;

• Maximum origin or destination per grid cell;

• Average origin or destination per grid cell;

• Percentage of grid cells with no origin or destination of trips;

• Percentage of grid cells with origin or destination of trips lower than L1 (value established in 200 trips, because is the average trip coefficient, and that represents the lines and columns of the O/D matrix with higher error or no statistical mean);

• Percentage of grid cells with origin or destination of trips higher than L2 (value estimated in 30% of the maximum cell value, represents the lines and columns of the O/D matrix with lower error or no statistical mean);

• Ranking chart for total origin or destination per grid cell;

• The percentage of trips in non significant statistical O/D matrix cells.

Some of these indicators have an important correlation and later in the analysis were selected those who had an easier way to measure and compare values from different grid cell dimensions. All these indicators were also calculated for different grid origin locations to test the sensitivity of this variation. The effect of the location of the origin of the grid was tested considering variations in both axes of 25% of the side of grid cell for each dimension of analysis initially defined, resulting in 16 different values of all indicators for the same grid dimension.

With these indicators and the two different analysis made, additional analysis can be developed in order to determine the best grid dimension for the available data considering the basic principles of traffic demand modeling, and at the same time, the best location and/or orientation of the origin of the grid, to obtain the most favorable results of the indicators for traffic modeling. The definition of the optimal grid dimension and origin location could be obtained by defining a goal function using the indicator values developed, but once it wasn’t the main goal of this article whose main objective was to estimate and measure the MAUP effects prior to the TAZ delineation.

SOFTWARE INSTRUMENTS INTEGRATION

The GIS based application developed for this study, incorporates different kinds software. The integration of such a different suit of softwares was the key to obtain the results of the GIS application. Four different softwares were used: CAD software, worksheet software, text software, and GIS software (Geomedia Professional 5.1).

The CAD software was used to create the different grids of analysis, producing closed squares that later are imported to the GIS software. The grids were produced using macros written in Visual Basic as is showed in Figure 1.

Figure 1: Example of the grid obtained on CAD software and the macros code.

The GIS software produces the spatial analysis and aggregation of the data of the Mobility Survey with the case study boundaries and the imported CAD grids on their different origin location. The GIS application used for all the analysis needs as input: the start of trip points, the end of trip points, the grid of analysis, the boundaries of study area, and the definition of an iteration step, that was used as 25% of
side of each square grid. The user interface and all the inputs asked for the application are showed in the Figure 2. This spatial analysis is stored in a warehouse (Access file .mdb) and the other outputs are produced in text files and in Excel workbooks.

Figure 2: Example of GIS based application interface with the user and all the needed inputs.

The outputs from the text files are the O/D matrix and a resume of the indicators used in this analysis. Some of the indicators of this text file include: the percentage of intra-zonal trips, the maximum cell value at the O/D matrix, etc. Some of these indicators and the O/D matrix can be seen in Figure 3.

Figure 3: Example of the text flies outputs obtained by the GIS based application.

The Excel Workbook contains some analysis of the O/D matrix and the grid cells, and also the ranking charts produced and some trends of these charts. The matrix output was not produced in an Excel worksheet because the number of zones is always higher than the 256 columns (the maximum allowed by this kind of worksheets). The trends used to adjust to the ranking charts are an exponential curve that adjusts well at the lower values of the Ranking chart, and a geometric progression function that deals to better results at higher levels of Ranking chart, looking to be the more appropriated trend to use (see Figure 4).

Figure 4: Example of the Excel workbook outputs and one of ranking charts.

The good integration of these different softwares is possible due to OLE Automation technology used by the Visual Basic language code, which turns very easy the integration and the complex relationships between different softwares. Using different software instruments for the same analysis, can overcome some limitations of the unique use of each software.

MAUP SCALE EFFECTS ANALYSIS

After the run of the GIS based application for all grid dimensions and having all the outputs, an analysis of the MAUP scale effects was undertake, or in other words, the analysis of the grid size variation on the data aggregation results.

This analysis started by looking at the impact of the variation of the grid cell dimension in the spatial trips distribution. The spatial trips distribution consists in adding for each cell all the origins and destinations within its boundaries. Figure 5 presents the spatial trips distribution of the study area for the 2000 meters of side grid cell. As predicted the greater trips concentration are within the Lisbon city centre borders and in their environs. The Cascais and Sintra corridors (close by municipalities of Lisbon and included at the Lisbon Metropolitan Area), and Almada (a municipality of the south boundary of the Tagus River), also have a high quantity of trips and the in mediations of highway A1 (connects Lisbon with O Porto) until the municipality of Vila Franca de Xira that works as regional interchange.

If this analysis is made for other grid-cell dimensions, the number of zones with no origin or destination of trips has a very significant increase. The cells that maintain a higher value of origin or trip destinations are within the Lisbon city boundaries as it can be seen in the Figure 6, Figure 7 and Figure 8 (spatial trips distribution for the 1200, 800 and 200 meters of cell side grid trips respectively).
Figure 5: Spatial trips distribution of the Lisbon Metropolitan Area for 2000 meters of cell side grid.

Figure 6: Spatial trips distribution of the Lisbon Metropolitan Area for 1200 meters of cell side grid.

Figure 7: Spatial trips distribution of the Lisbon Metropolitan Area for the 800 meters of cell side grid.

Figure 8: Spatial trips distribution of the Lisbon Metropolitan Area for the 200 meters of cell side grid.
The normal conflict when establishing traffic zones in traffic demand modeling rests upon the base concept of the “inversely” proportional relationship between the statistical precision and the geographic precision. To obtain a good geographic precision, and locate with precision the origin and destination of trips, it is important to build small size zones, which induces a significant reduction in the statistical precision of the zone. This conflict is derived from the increase of the number of O/D matrix cells with a non significant quantity of trips with the number of zones (decrease of statistical precision). Although, with traffic zones of small size the percentage of traffic zones with no trips increase, permitting a better geographic precision.

A result from this conflicting goal when defining TAZ is the loss of information for the traffic assignment of the trip made at the intra-zonal. This information lost increases with the zones size (non assigned trips in the transportation forecasting models).

To evaluate this conflict for the different grid-cell dimensions used in this article, the indicators previously described were used in order to best-fit optimum cell-size. The most significant indicators were the percentage of intra-zonal trips of the O/D matrix, the percentage of trips in non significant statistical O/D matrix cells, and the percentage of zones with no origin or destination.

To measure the variation of the first indicator with the cell grid dimension, and estimation of the percentages of the average of the values obtained for 16 different locations of the grid origin for every grid dimension had to be performed. The values obtained are presented in the Table 1.

Table 1: Variation of the intra-zonal trips percentage with the grid dimension

<table>
<thead>
<tr>
<th>Grid dimension (m)</th>
<th>Difference to minimum value (%)</th>
<th>Average intra-zonal trips (%)</th>
<th>Difference to maximum value (%)</th>
<th>Elasticity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.09</td>
<td>17.87</td>
<td>0.29</td>
<td>0.00</td>
</tr>
<tr>
<td>400</td>
<td>0.21</td>
<td>18.70</td>
<td>0.34</td>
<td>0.0042</td>
</tr>
<tr>
<td>600</td>
<td>0.33</td>
<td>19.50</td>
<td>0.26</td>
<td>0.0040</td>
</tr>
<tr>
<td>800</td>
<td>0.30</td>
<td>20.35</td>
<td>0.28</td>
<td>0.0043</td>
</tr>
<tr>
<td>1000</td>
<td>0.46</td>
<td>21.34</td>
<td>0.29</td>
<td>0.0049</td>
</tr>
<tr>
<td>1200</td>
<td>0.45</td>
<td>22.38</td>
<td>0.47</td>
<td>0.0052</td>
</tr>
<tr>
<td>1400</td>
<td>0.35</td>
<td>23.25</td>
<td>0.44</td>
<td>0.0043</td>
</tr>
<tr>
<td>1600</td>
<td>0.60</td>
<td>24.17</td>
<td>0.52</td>
<td>0.0046</td>
</tr>
<tr>
<td>1800</td>
<td>0.34</td>
<td>25.28</td>
<td>0.37</td>
<td>0.0056</td>
</tr>
<tr>
<td>2000</td>
<td>0.83</td>
<td>26.39</td>
<td>0.78</td>
<td>0.0056</td>
</tr>
</tbody>
</table>

The variation of this indicator is significant (approximately 9% between the maximum and minimum value) and it is increasing in the same direction as the grid dimension. This fact reflects some scale effect over the data because the elasticity of this value related to the grid dimension tends to increase with it.

To measure the variation of the second indicator with the grid-cell dimension, the previous methodology was applied. The values obtained are presented in the Table 2.

Table 2: Variation of the percentage of trips in non significant statistical O/D matrix cells.

<table>
<thead>
<tr>
<th>Grid dimension (m)</th>
<th>Difference to minimum value (%)</th>
<th>Average percentage trips in non significant O/D matrix cells (%)</th>
<th>Difference to maximum value (%)</th>
<th>Elasticity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.32</td>
<td>73.05</td>
<td>0.11</td>
<td>0.0000</td>
</tr>
<tr>
<td>400</td>
<td>0.32</td>
<td>72.34</td>
<td>0.24</td>
<td>-0.0035</td>
</tr>
<tr>
<td>600</td>
<td>0.58</td>
<td>70.78</td>
<td>0.30</td>
<td>-0.0078</td>
</tr>
<tr>
<td>800</td>
<td>0.59</td>
<td>67.29</td>
<td>0.61</td>
<td>-0.0175</td>
</tr>
<tr>
<td>1000</td>
<td>0.65</td>
<td>62.34</td>
<td>0.83</td>
<td>-0.0247</td>
</tr>
<tr>
<td>1200</td>
<td>0.97</td>
<td>57.08</td>
<td>1.14</td>
<td>-0.0263</td>
</tr>
<tr>
<td>1400</td>
<td>0.81</td>
<td>52.03</td>
<td>1.02</td>
<td>-0.0253</td>
</tr>
<tr>
<td>1600</td>
<td>1.39</td>
<td>46.88</td>
<td>0.88</td>
<td>-0.0258</td>
</tr>
<tr>
<td>1800</td>
<td>1.23</td>
<td>41.75</td>
<td>1.27</td>
<td>-0.0257</td>
</tr>
<tr>
<td>2000</td>
<td>1.11</td>
<td>37.77</td>
<td>1.49</td>
<td>-0.0199</td>
</tr>
</tbody>
</table>

Another important indicator is the percentage of cells with no origin or destination of trips. As for the intra-zonal trips indicator, it was important to calculate for each grid dimension the average value of the different values observed for all the origins of the grid. In opposition to the previous indicator, the values of this indicator decrease with the grid dimension (as expected). The values obtained are presented in the Table 3.

Table 3: Variation of the percentage of zones with no origin or destination with the grid dimension.

<table>
<thead>
<tr>
<th>Grid dimension (m)</th>
<th>Difference to minimum value (%)</th>
<th>Average percentage of zones with no origin or destination of trips (%)</th>
<th>Difference to maximum value (%)</th>
<th>Elasticity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.38</td>
<td>96.79</td>
<td>0.74</td>
<td>0.0000</td>
</tr>
<tr>
<td>400</td>
<td>1.10</td>
<td>92.49</td>
<td>2.06</td>
<td>-0.0215</td>
</tr>
<tr>
<td>600</td>
<td>1.85</td>
<td>86.55</td>
<td>4.00</td>
<td>-0.0297</td>
</tr>
<tr>
<td>800</td>
<td>3.09</td>
<td>79.85</td>
<td>5.94</td>
<td>-0.0335</td>
</tr>
<tr>
<td>1000</td>
<td>3.87</td>
<td>72.89</td>
<td>7.60</td>
<td>-0.0348</td>
</tr>
<tr>
<td>1200</td>
<td>4.39</td>
<td>66.83</td>
<td>8.73</td>
<td>-0.0303</td>
</tr>
<tr>
<td>1400</td>
<td>3.80</td>
<td>60.46</td>
<td>9.66</td>
<td>-0.0318</td>
</tr>
<tr>
<td>1600</td>
<td>3.72</td>
<td>55.76</td>
<td>9.56</td>
<td>-0.0235</td>
</tr>
<tr>
<td>1800</td>
<td>4.63</td>
<td>51.66</td>
<td>9.84</td>
<td>-0.0205</td>
</tr>
<tr>
<td>2000</td>
<td>4.32</td>
<td>48.71</td>
<td>10.45</td>
<td>-0.0148</td>
</tr>
</tbody>
</table>

In order to obtain a good traffic zoning it is important to have simultaneously a good statistical precision and good geographic precision. These two requisites can be estimated through the intra-zonal trips percentage and the percentage of cells with no origin or destination of trips. The first indicator, if minimized, produces a better geographic precision and a
lower loss of information. The second indicator, if maximized, produces a better origin or destination of trips geographic location. But at the same time to obtain a good statistical precision the maximization of the grid cell dimension is required. It is not possible to reach the optimum for these three requests, but the existence of local or global optimum in this relationship can be analyzed.

The analysis of the relationship of the three variables was made using the concept of crossed elasticity between the variation of the percentage of cells with no origin or destination of trips and the variation of the intra-zonal trips percentage with the grid dimension. The results obtained from the available data for this study are presented at Figure 9.

By analyzing this indicator a conclusion can be pointed out: the minimization (absolute value) conducts to a grid dimension with a better geographic precision. Then, is pretended a local minimization of this elasticity but with the higher dimension of cell as possible. At Figure 9 it can be observed that the crossed elasticity has a local minimum at the cell side dimension of 1200 meters that could indicate that this size for the side of the square, conducts to a better equilibrium between geographic precision and O/D matrix loss of information. This conclusion is only valid for this case study and the data available, in order to extrapolate more case-studies and indicator testing should be performed.

Figure 9: Crossed elasticity between the variation of the percentage of cells with no origin or destination of trips and the variation of the intra-zonal trips percentage with the grid dimension.

In order to measure the relationship between the statistical precision and loss of information generated by the intra-zonal trips of the O/D matrix, a chart that presents the evolution of this value had to be built. The results are presented in Figure 10, where it is incorporated the variation of the grid dimension and grid origin positioning.

Observing this figure, it can be seen that is difficult to establish an optimal grid dimension due to the existing trade-off between both indicators with the grid cell size variation. The grid size choice should be based on the definition of maximum and minimum values allowed for each indicator, depending on the case study, study scale and goal, and the mobility survey or data available. It should always be therefore a sensible choice aware of the existing trade-off and its impacts for traffic assignment and transportation planning.

It is also important to notice that the grid origin location has a greater impact in the O/D matrix statistical error than in the percentage of information loss in the O/D matrix (approximately the double), although in the chart it looks the other way around because the different scale of both axes.

Other important indicators are the maximum origin or destination per grid cell, and the average origin or destination per grid cell. These values are estimated as the average of the 16 iterations made for each grid cell dimension. The values obtained for these indicators are presented in the Table 4 and in the Figure 11. In the Table 4 and in the Figure 11 it is possible to observe that an increase in the grid dimension doesn’t conduces always to an increase in the maximum origin or destination per grid cell value. The explanation to this fact resides in the high concentration of trips in some points that can all be contained in smaller cells, and for this study not all the origin grid positions have been tested (only 16 different positions).

Table 4: Maximum and the average origin or destination per grid cell variation with grid dimension.

<table>
<thead>
<tr>
<th>Grid dimension (m)</th>
<th>Maximum origin or destination per zone</th>
<th>Average origin or destination per zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>462270,78</td>
<td>265,89</td>
</tr>
<tr>
<td>400</td>
<td>625433,34</td>
<td>1039,51</td>
</tr>
<tr>
<td>600</td>
<td>409963,88</td>
<td>2289,55</td>
</tr>
<tr>
<td>800</td>
<td>534759,28</td>
<td>3990,37</td>
</tr>
<tr>
<td>1000</td>
<td>618473,19</td>
<td>6122,74</td>
</tr>
<tr>
<td>1200</td>
<td>596973,66</td>
<td>8638,75</td>
</tr>
<tr>
<td>1400</td>
<td>612476,44</td>
<td>11571,47</td>
</tr>
<tr>
<td>1600</td>
<td>661582,64</td>
<td>14829,69</td>
</tr>
<tr>
<td>1800</td>
<td>980089,69</td>
<td>18500,10</td>
</tr>
<tr>
<td>2000</td>
<td>970595,33</td>
<td>22530,52</td>
</tr>
</tbody>
</table>

The average origin or destination per grid cell, as expected, decreases with the increase of the grid dimension, but as it is presented in the Figure 11, with a quadratic trend. This information is important because the decrease in this value could mean less statistical precision for the traffic zoning.
For this reason, it should be seen as a constraint in the reduction of the grid dimension.

Figure 11 - Maximum and the average origin or destination per grid cell variation with grid dimension.

To characterize the flow degradation in the O/D matrix and the total origin or destinations per cell were used ranking charts. As for all other indicators, different charts for all the positions of the origin grid and for all grid sizes were built. As an example of these kind of charts, in the Figure 12 is showed a Ranking chart for total origin or destination per grid cell, for an origin grid position of a 2000 meters side cell grid. In this figure, it is possible to evidence that for a grid, even for a grid of greater dimensions, the values decrease very fast. This fact indicates that are few cells with high values of origin or destinations of trips.

For the resulting curve of this chart, two different trend lines were estimated: an exponential curve and a geometric progression function ($A \cdot R^{n-1}$). Both adjust well to the estimated curve in its lower slope part, but the geometric progression function adjusts better at the higher values, as can be seen at the Figure 12. Nevertheless the exponential curve has higher values of the $R^2$ parameter ($R^2=0.98$ versus $R^2 = 0.90$ of the geometric progression function) derived of its almost perfect adjustment at the lower values, but can not “explain” the variation at higher values. For this reason, the geometric progression function was selected as representative of the Ranking chart behavior.

Figure 12: Ranking chart for total origin or destination per grid cell, for a 2000 meters side cell grid.

To explain in an aggregated way the total origin or destinations per cell for each grid dimension, it was made an average of the parameters of adjustment of the geometric progression function ($A$ and $R$), for each grid cell dimension. With this simplification it was easier to analyze the trip degradation function for each grid cell dimension.

In the Figure 13, it can be seen how the function, as the grid dimension grows, reveals a bigger slope and higher values of the parameter $A$. The great concentration of origin or destination of trips in some few cells when the grid dimension is bigger is no surprise and a reason why the statistical precision is higher.

Figure 13: Geometric progression function for the different grid dimensions.

Another indicator that reinforces the information obtained from the ranking chart of the total origin or destinations per cell is the definition of intervals of this indicator. As it was presented in the methodology, two values for this indicator were defined in order to distinguish cells with poor statistical precision (200 trips per cell) and cells with good statistical precision (30% of the maximum value per cell). These two points define three intervals of values that are presented in the Figure 14.

With the percentages presented at the figure it is possible to conclude that the high values of this indicator are a little percentage of the total cells and the increase with the cell grid dimension is not very significant. Otherwise, the low values of the indicator, that are an important constrain to the statistical precision, have a great variation with the grid-cell dimension (-0,03% slope for the adjustment with a linear trend line). This information reinforces the conclusion of the importance of the grid cell dimension to the statistical precision.

Figure 14: Total origin or destination of trips per cell intervals variation with the cell grid dimension.
GRID ORIGIN SENSITIVITY ANALYSIS

To measure the influence of the variation of the grid origin location in the indicators results, a sensitivity analysis had to be made. The results point out to the fact that the indicators that presented a higher variation in their value were the intra-zonal trips percentage and the percentage of cells with no origin or destination of trips.

The sensitivity analysis for the intra-zonal trips percentage indicator was made making use, once more, of the ranking charts. In the Figure 15 and Figure 16 is showed for some grid dimensions (2000, 1200 and 200 meters side of the cell grid respectively), the resulting ranking charts. In these curves can be seen an approximated linear behavior and a negative slope that decrease with the grid dimension (reducing the indicator variance). The origin intersections of adjustment trend lines also decrease with the grid dimension as expected.

It also important to underline that these indicators do not have their maximum value at the same position, which makes impossible a joint sensitivity analysis.

For the indicator of percentage of cells with no origin or destination of trips it was used the same procedure as for the intra-zonal trips percentage indicator. In the Figure 17 and Figure 18 are showed some example of the resulting Ranking charts (2000, 1200 and 200 meters side of the cell grid respectively) where it can be seen that the curves have an approximated linear behavior and a negative slope that decrease with the grid dimension (reducing the indicator variance). The origin intersections of adjustment trend lines also decrease with the grid dimension as expected.

A final step of the developed methodology was to determine the distance and angles relatively to the starting grid location that could present the most favorable values of the intra-zonal trips indicators (considered the most important for this analysis).

The methodology used for this analysis was the determination, for each grid cell dimension, the best neighbor location from all the 16 analyzed cases. The goal of this methodology was the determination of the best direction when moving the grid from his actual position. The resulting analysis allows evaluating the average angle relatively to the starting position location, and the exact point location in polar coordinates.

The results of this analysis are presented in the Figure 19 where it can be seen that the distance to the best neighbor (in percentage of the grid cell size) is not biased with the grid cell size (approximately constant and around the 59% of the grid cell size), and also that the angle relatively to the starting point doesn’t have a regular behavior.
With these results, it can’t be established any behavior pattern for the best angle to minimize the intra-zonal trips percentage for the different grid cell dimensions (as a result of very low or inexistent correlation for these variables). The only information that can be obtained from this analysis is the average angle for the minimization of the indicator but independent of the grid dimension, nevertheless, it does not seem to look as its real behavior although the angle oscillates around this value (see Figure 19). Therefore it is impossible to determine a best location for the grid origin, because there is a need of one of the two polar coordinates and we only can estimate the distance to the starting point but not the angle.

![Figure 19: Angle and distance sensitivity analysis to minimize the intra-zonal trips percentage.](image)

**CONCLUSIONS**

This study focused on the research of the MAUP scale effects in some important traffic demand modeling variables in order to develop a future TAZ delineation methodology with less MAUP interferences. A new analysis methodology and a GIS based application was developed to this study to measure these effects over some indicators considered relevant for the future TAZs delineation, as the intra-zonal trips percentage and percentage of trips in non significant O/D matrix cells. The model was applied over the AML and a Mobility Survey of this area made in 1994.

This paper reports these analyses, and it was possible to extract some important conclusions about some MAUP scale effects over the aggregated travel data that can be an important knowledge for the TAZs delineation:

- The increase in the grid dimension, as expected, conducts to better indicator values from a statistical precision point of view, but produces a significant loss of trips that are not assigned to the transportation network.

- The grid cell size has a very significant impact in the percentage of intra-zonal trips (approximately 9%), what suggests that TAZ delineation and the number of zones selection should not be arbitrary as is usually made.

- It exists an important residual intra-zonal trips percentage even for very small grid cells (approximately 18%), which indicates a great trip concentration and the existence of an important percentage of trips made by walking or cycling, and that will not be assigned to the transportation network.

- The grid cell origin position has a very significant impact in the percentage of intra-zonal trips (approximately 9%).

- The grid cell size has a very significant impact in the percentage of grids cells with no origin or destination of trips (approximately 50%), what is a decisive factor for geographic precision.

- The grid cell origin position has a very significant impact in the percentage of cells with high values of origin or destination of trips (approximately 50%).

- The percentage of cells with with values of origin or destination of trips is approximately constant, which indicates a high concentration of trips in little zones of LMA.

- The majority of the developed indicators are very sensible to the grid origin location, but was not possible to establish a spatial bias or pattern to obtain the more favorable values of the indicators.

**AKNOWLEDGEMENTS**

This research was supported by Fundação para a Ciência e a Tecnologia (FCT) since 2004. The private company TIS has also provided support by making available the AML Mobility Survey (TIS 1994), and to the software company INTERGRAPH for the Geomedia Professional 5.1 licence.

**REFERENCES**


TIS. 1994. “Lisbon Metropolitan Area Mobility Survey”.

**BIOGRAPHY**

**L. MIGUEL MARTÍNEZ** was born in Lisbon, Portugal and went to the Instituto Superior Técnico (Lisbon Technical University), at their Lisbon site, where he studied Civil Engineering and obtained his degree in 2004. After finishing his degree, he started his work as researcher in the CESUR (Department of Civil Engineering - Instituto Superior Técnico) where he has been working since.