Interacting Microsoft Visual Basic Procedures (Macros) and GIS tools in order to access optimal location and maximum use of railways and railway infrastructures

Vieira, J.; Cristovão, H.; Viegas, J.; Silva, E.

CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa

Abstract

Some parts of the Portuguese railway infrastructure have been neglected through time: Rural lines have been abandoned, investment in new infrastructure is sometimes delayed, and marketing strategies to keep or attract more users have not been pursued.

Simultaneously, problems with urban congestion, pollution and mobility for the young, the elderly, the poor, and the handicapped are putting forward the discussion about new or more sustainable modes of transportation. Common sense of public officials, other lobbying groups, and the locals demand new, trendy train lines. And while some axes may have the potential to justify rail lines, others seem to lack demand or funding to be enabled.

One major problem in order to evaluate the worthiness of these rail projects has been the fact that very often the studies of travel demand and physical implantation are done separately. Travel demand analysis is done based on the four-step model (trip generation, distribution, modal split, and network assignment) using survey data and the network system, using a relatively wide zoning. The importance of interacting with other, finer, information (i.e. slope, density of population, environmental sensitivity, or other socio-economic and land use information) with the development of the travel analysis demand will enhance the analysis/results and increase the chance of proposing lines that are both optimal in location and will have the maximum use by the citizens.

Off the shelf software is still unable to perform this kind of operations. Some perform the analysis using existing networks, and no information on the land is available besides the zoning system, other software propose lines accordingly to land slopes, but no trip information is included.

GIS packages have the capacity to include the land information and some have transportation analysis, but are lacking computation capabilities and algorithms to perform analysis similar to off-the-shelf transportation software.

In order to develop this kind of integrated analysis it is important to have a good knowledge of the algorithms and analysis required by transportation and of the tools/opportunities offered by the GIS packages.

This paper presents a methodology that integrates the transportation algorithms with the GIS functionalities, using excel macro-language. The result is an interaction of both travel demand analysis and site selection. The characteristics of the place constrain the travel demand analysis, but on its own the travel demand analysis define not only the buffer of the train line, but systematically enhance the shape of the line and the location of the stops each time the results of a phase of the travel demand analysis is outputted.

This paper offers guidelines for those developing travel demand analysis including some site selection criteria, and it can be a starting point for those of whom intend to develop further application of in the GIS fields.

Keywords: GIS, Transportation Models,

Conference Paper: Regional Science Association (RSA) and Associação Portuguesa de Desenvolvimento Regional (APDR); "Regions and Fiscal Federalism "; 44th European Congress of the European Regional Science Association, Porto, Portugal 25-29 August 2004

1. Introduction

Decisions concerning the transportation system imply major investments that constraint very strongly the mobility options of the populations. Therefore these decisions should be taken once studies are performed, a special care should be put on the "demand" side of the equation (Viegas, 2003 a, b). Because the evaluation of the impacts of developing/implementing new transportation policies and projects can't be done through experimentation or analogical processes, mathematical modelling play a central role.

The need to understand the mobility behaviour promotes detailed studies about consumer behaviour and consumer choices. Two different approaches can be observed: 1. observation of real behaviour (revealed choices) and 2. observation and study of possible choices and behaviour under hypothetical conditions (declared choices). The surveys play a central role in order to build this database on revealed and/or stated preferences, afterwards, the processing of the information is done using discrete choice models. The Logit model has been very successful processing the information, allowing through a simple mathematical equation to extract the probabilities of citizen's choice accordingly to the different options of transportation.

This paper presents a methodology and subsequent analysis in order to evaluate the preferences of local residents to the possible development of a new transportation infrastructure.

While, state and revealed preferences play a pivotal role to evaluate the probabilities of the successful use of a transportation infrastructure, other elements constraint the effective use of a new transport mode. This is to say that to the present-day characteristics and distribution of the population, its socio-economic characteristics, and the characteristics of the land (topography) all of them contribute to the final success of a wise, fair, and economically sustained decision.

The use of Geographic Information Systems has a common platform that enables to interact the previous information with the development of discrete choice models enhances the results of the model and expands its capacities. This is done through a step by step interactive process where results from the survey start feeding the GIS allowing to narrow down the land information required and the zone to be studied, on the other hand, once this is done, stated preferences became to be shaped accordingly to the information resulting from crossing survey data with attribute data imputed to the GIS, and this will lead to a new interaction process and to narrow down the best places to draw the infrastructure line.

While, this seems to be a relatively easy process, once GIS already allows to overlay and query different datasets, some major drawbacks were found when trying to model directly on the GIS the discreet choice model, as well as when trying to interact land information with the discreet choice model. Nevertheless, we think that in a near future it will be possible to have this as an operational process. The best option for the purpose of this study (time-wise and cost related options) was to develop some of the analysis through the use of Microsoft VB macros running on excel, whose tables would afterward feed the GIS, and vice versa.

As a consequence, this paper goes beyond the usual methodological approaches to evaluate the preferences of local residents to the possible development of a new transportation infrastructure, once it tried to constrain model accordingly to other characteristics together with local stated or revealed preferences.

A region in the north of Portugal will be used to implement the methodology. This region comprises four different regions, and ten different municipalities (Figure1). It is characterized by a steep slope landscape, and a dispersed urban pattern of population and activities, that has been changing its mobility patters through the reinforcement of private transportation.

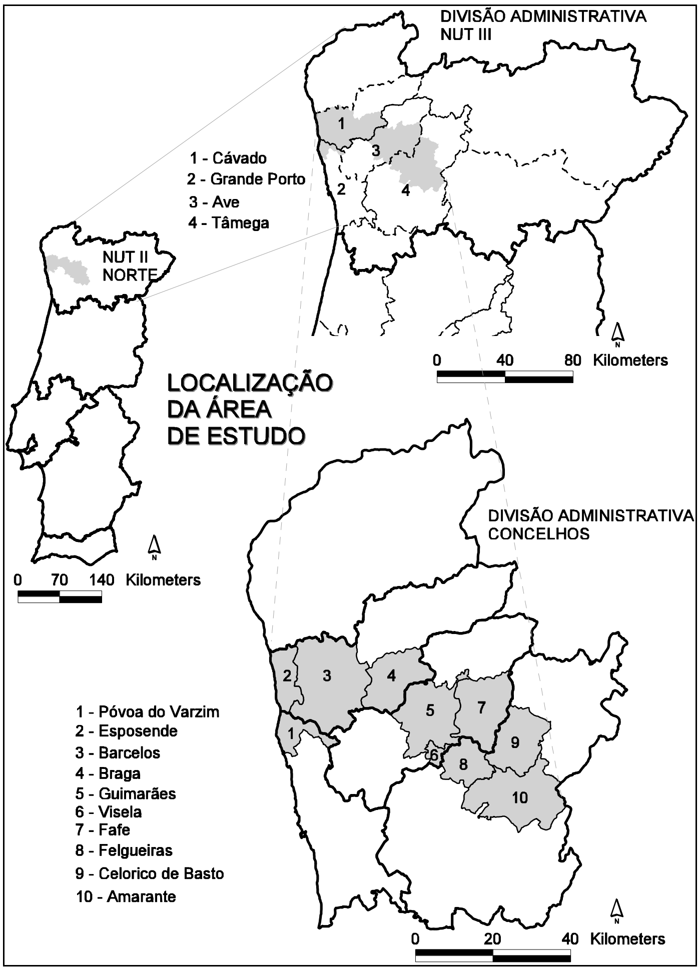

Figure 1 - Location of the study area

The next points will describe the different phases, first the interaction of GIS-Excel macros is detailed, afterwards the results of interacting this survey data with the land information set at the GIS are presented. Concluding remarks highlight the advantages of these procedures and some problems found and will point out future research directions.

2. Data requirements and data handling (TI e TC)

This point details the steps required to produce and group the information resulting from the survey done at the study area in order to calculate the potential of "demand" of the railway. Table 1 presents the succession of steps in order to reach the required results.

Zoning	Multiplicative Coefficients	Declared Preference	Disponibility	Maximum potential of demand	Estimation of final demand
1	2	3	4	5	6

Table 1 - Methodological steps to data handling

Step 1 – Zoning

Field data collection is preceded by an assessment of the best zoning size for the study purposes, usually done in transportation studies. This will also grant that we will have data collected and referenced to a spatial zoning that can be imported to the GIS (the spatial unity “Freguesia”-Parish has a size in the north of Portugal that allows detailed studies of transportation). The fact that the local residents can very easily identify the “Freguesia”-Parish of origin and destination, was an important element to grant the success of the surveys (no survey had to be deleted due to lack of information). Some exceptions had to be taken in what concerns zoning and the city centre of major cities (Braga and Guimarães). At these cities it was important to have a narrow zoning that would reflect different O/D at of the city.

Step 2 – Calculation of the multiplying coefficients

The main goal to calculate the multiplicative coefficients is that it will grant a quantitative weights to the statistic inference of the surveyed data (i.e. representativity of the resulting surveys accordingly to the surveyed population) , the following expression allows to calculate these coefficients:

Multiplying Coef. at a collection point and for a specific hour = Results from the traffic counts / number of surveys [at the specific collection point and hour]

As at the surveyed sample tax, for each hour and for each collection point it was calculated the equivalent multiplicative coeff.. Therefore, when considering the surveys to the TI (Individual Transportation) the results were 28 multiplicative coeff. To two collecting posts and 14 hours. In what concerns the TC (Public Transportation), there were 56 coefficients resulting from 14 hours and four collection posts.

Step 3 – calibration of the model of stated preferences

This point started with the grouping of surveyed that presented a fixed pattern of answers of stated preferences (kept the same answer throughout thee different questions of the survey). After quantifying this level of preferences, the surveys were deleted once they added no new knowledge to the factors that constrained the modal preferences of the travellers.

Step 4 – Evaluation of the availability of the service for each O/D pair of each trip accordingly to the GIS buffer zones

In order to evaluate, among pairs of trips of Origin and Destination (O/D) which of them would be served by the train line, the following process was enabled:

Classification of the zones of origin and destination of the trips, accordingly to a code 0 ,1, or 2. Zero if they are not served by the proposed lines, one if they are served but aren’t under the propose line, and three if they are under the proposed line.

Afterwards, it was performed a classification of each trip as 0 (if they are not served by an origin or destination), 11 if both are served by a service of type 1 (both O/D are at the buffer zone); 12 or 21 if the Origin is located at the buffer and the Destin is crossed by the train line (12), or in case the Origin is crossed by the train line and the Destin is at the buffer (21), the number 22 was given if both Origin and Destin are crossed by the train line.

Step 5 – Maximum potential of demand

The fifth step corresponds to the calculation of the maximum of demand potential, considering that the total number of trips and the zones of Origin and Destin were served both by the train line and the survey of the declared preference. In order to calculate this first all the trips that are not served by the train line (step 4), and among the ones that present a fix patter of answers favourable to the maintenance of TI or TC (step 3). This way, the maximum potential of demand corresponds to all the trips that are served by the train line, excluding also the ones that the surveyed residents sated that never would use the proposed train.

Step 6 – Final Demand Estimation

This step corresponds to the application of the model of discrete choice (parameters where calibrated accordingly to step 3) and it is only applied to each of the trips that the inquiry had stated as having a chance of changing to another transportation mode, and if the zones of origin and destination where close enough to the train line. Accordingly to the timelines and costs estimated to the transportation mode "Light train", and having used the parameters of the model, it was possible to estimate the probabilities of each of the inquired persons using this new transportation mode. Multiplying this new probability with the multiplicative coefficient of the collecting post/schedule corresponding to that specific survey-inquiry, it is possible to extract the contribution of that specific trip to the total demand of the system.

Application of the Model of Discrete Choice

Once the declared preferences are set, both to TI and TC, it is possible to calibrate the model of discrete choice.

If in the calibration phase preferences are set (and intensity of preference) among two alternatives of hypothetical transportation that are given (cards of stated preferences presented to the resident), at the phase of exploration of the model we first start by simulate the amplitude of the class-values at the model to the new transportation mode (train line) to the same trip it was surveyed.

The applied model is the Logit, it is based on the theory of the Stochastic utility. It is assumed that each consumer (traveller) is a rational decision-maker, but decision is made in a context of some randomness. Having in account that the attributes vary from trip to trip, and can't be known a priori. Once the model calibrated and applied to a vast number of travellers, there is always some degree of randomness corresponding to the variation of the preferences among the consumers.

The intensity of the preference to the alternative A against alternative B is reproduced through a coefficient among the probabilities of choosing one or another alternative. The Logit Model sets as main basis that the logarithm of the probability coefficient is a linear function of the utility (generalized costs) among two alternatives, expressed through the polinomium of the magnitudes that characterize the costs.

At the following expression the parameters are the common element of the scale [and the coefficients (replacement tax) that affect each of the variables corresponding to the varying dimensions of the generalized costs (times, monetary costs, etc.)]. To increase the easiness of interpretation of the meaning of the resulting values an option was made, at the calibration phase, to use the 'minute' has the time unity.

$$\ln\left(P_A / P_B \right) = \Theta.\left[\Delta X_0 + \alpha.\Delta X_1 + \beta.\Delta X_2 + \chi.\Delta X_3\right]$$

The variable X0 corresponds to the specific modal attributes (non measurable, but with value resulting from the calibration), while the variables X1, X2, and X3 are respectively the time of the rout-path , the time of the access/egress to the boarding on the public transportation, and the monetary cost by trip. This way, the value of the specific attributes is stated in minutes and the coefficient **α** equals to 1.0. The coefficient **β**, can take values different from 1.0, even if the corresponding variable can be expressed in minutes (even if the travellers tend to feel the access/eggress time more intensely than the time during the trip, and this might lead to an overestimation of time by the traveller during those transition periods). The coefficient $\mathbf{X}$ makes the conversion of the equivalent costs in minutes (of travelling time).

Calibration was performed through the minimum square in both cases (TI and TC)

3. Interacting the survey data sets of the excel with the land information in the GIS

In order to evaluate the viability and opportunity of a rail connection between the municipalities of Braga and Guimarães, with probable extensions to the west (Póvoa do Varzim) and to the east (Amarante), besides the estimation of the demand potential, it is important to perform an evaluation of the most probable channels that will allow to develop the line with minimum impact from topographic constraints. Consequently it is very important to characterize the region's topography, once the definition of the served urban nuclei, besides being determined by the demand potential it is also by the construction costs.

A first analysis of both characteristics (topography and density of population) seem to rank the following municipalities as being in the best position to be connected: Amarante, Felgueiras, Fafe (or Fafe and Celorico de Basto) , Guimarães, Braga, Barcelos e Póvoa de Varzim. The study of connecting these cities is justified by the topographic characteristics of the region, by the strengthening of the urban system (Figure 2), by the urban development of the Metropolitan Area of Porto (AMP) and by the increasing growth of trips through time.

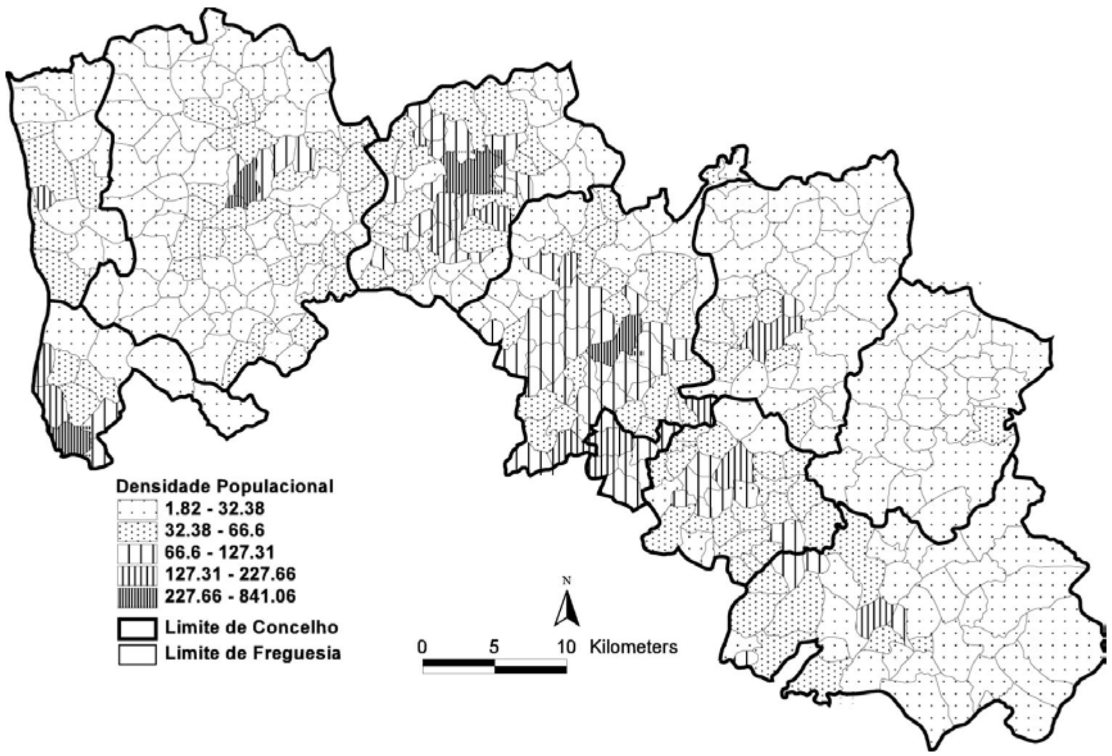

Figure 2 - Density of population by parish in the studied municipalities

The definition of the rail channel had two main constraints: the slope factor among the initial list of cities to be served, and also the opportunities of adjusting the line accordingly to the traffic demand studies.

The fist attempt in order to reach a final line location, was performed with the goal of searching for a liable channel that would link the cities of Póvoa do Varzim, Barcelos, Braga, Guimarães e Amarante. Topography was the main constraint, and the goal was to reduce the steep slopes and the required land movement. At a second phase other criteria were set ion order to optimise the location of the line accordingly to other constraints.

Therefore the first goal was to build a digital terrain model (DTM), using contour lines with a 100m equidistance.

Figure 3, presents the interaction of the proposed line with the slope characteristics, and with the results of the survey (origin and destination answers). At an initial phase several lines were proposed (the steep slop characteristics still allowed to make the same connection through several lines) , this is the case of the connection to Amarante and/or Fafe, that enable the existence of two extensions to the main line.

After the first attempt of linking the arc Povoa-Amarante, we juxtaposed the information of the demand studies. In order to perform that all the survey information was imported to the GIS: parish code numbers; information on the origins and destinations resulting from the survey (these ones already with previous cleaning and analysis – deletion of surveys considered erroneous, this analysis was done in order to impose multiplicative coefficients in order to estimate the real and the maximum demand potential).

Having this information, the resulting sum of the origins and destinations were overlapped to the DTM in order to refine the first line. The main goal was to search for alignments accordingly to the highest values of demand and the least steep slope locations (Figure 4).

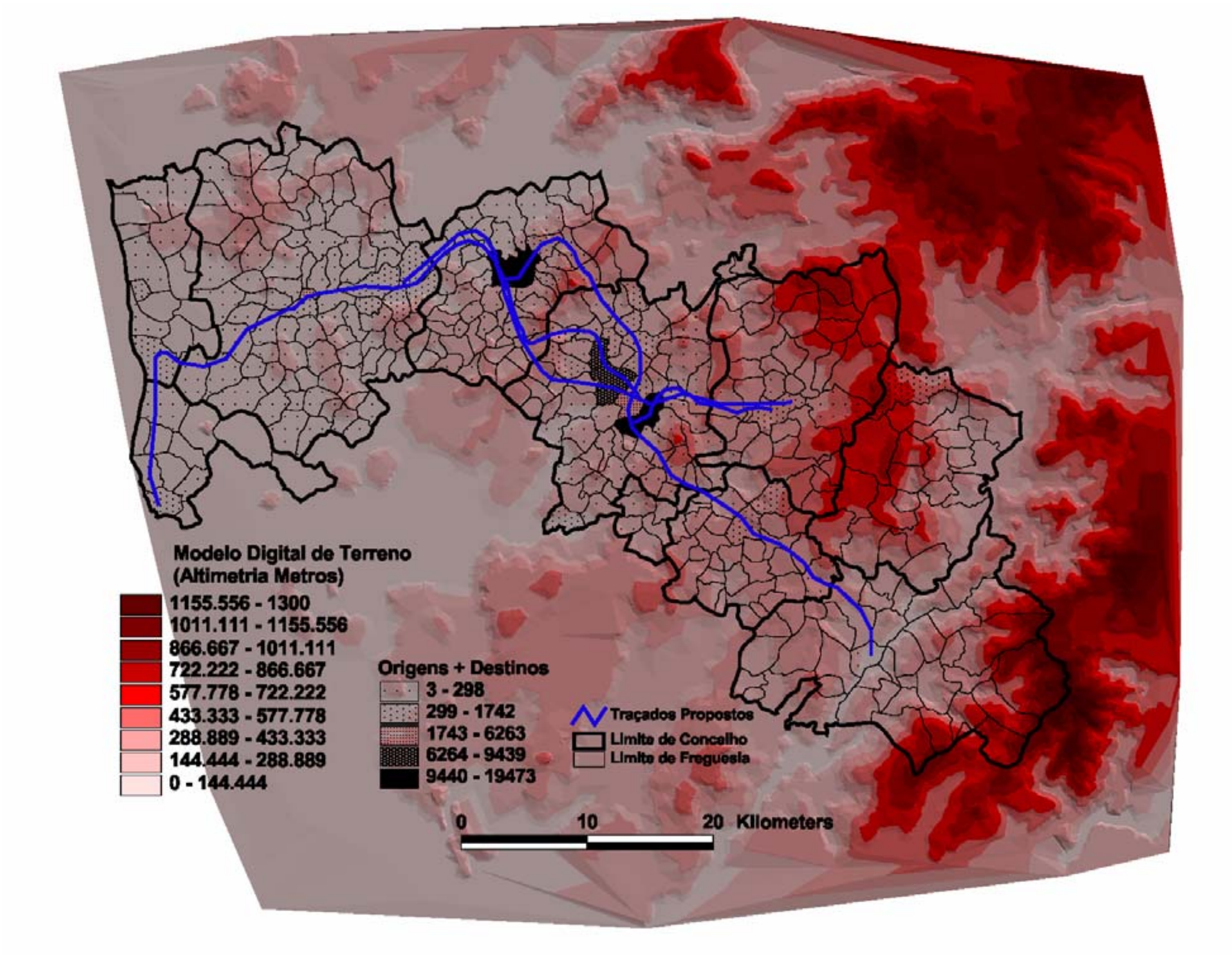

Figure 3 - Juxtaposing the different options of line location accordingly to slope and values of Origins and Destinations.

Figure 4 start to unveil the best location: demand at this methodology is as important as the topography being a constraint factor to the location the most probable line.

The reduced values of demand (the majority of the parishes has values from class 3 to 298 trips of origins+ destinations), putting aside the majority of the municipal parishes. While, the values between Braga and Guimarães are high, the extensions to the municipalities of Povoa and Amarante are strongly constrained. Also the locations of intermediate rail stations between the main municipalities is also constrained, due to the reduced potential of demand.

Figure 4 allows verifying that the municipalities of Amarante, Celorico de Basto, Felgueiras, Fafe, barelly have origins and destinations resulting from the counts. The municipalities of Barcelos and Póvoa do Varzim also have reduced values. Consequently, the options was to draw a main line that would link Barcelos-Braga-Guimarães, and a secondary line that would link Barcelos-Póvoa, and two optional lines that would link Guimarães-Felgueiras.

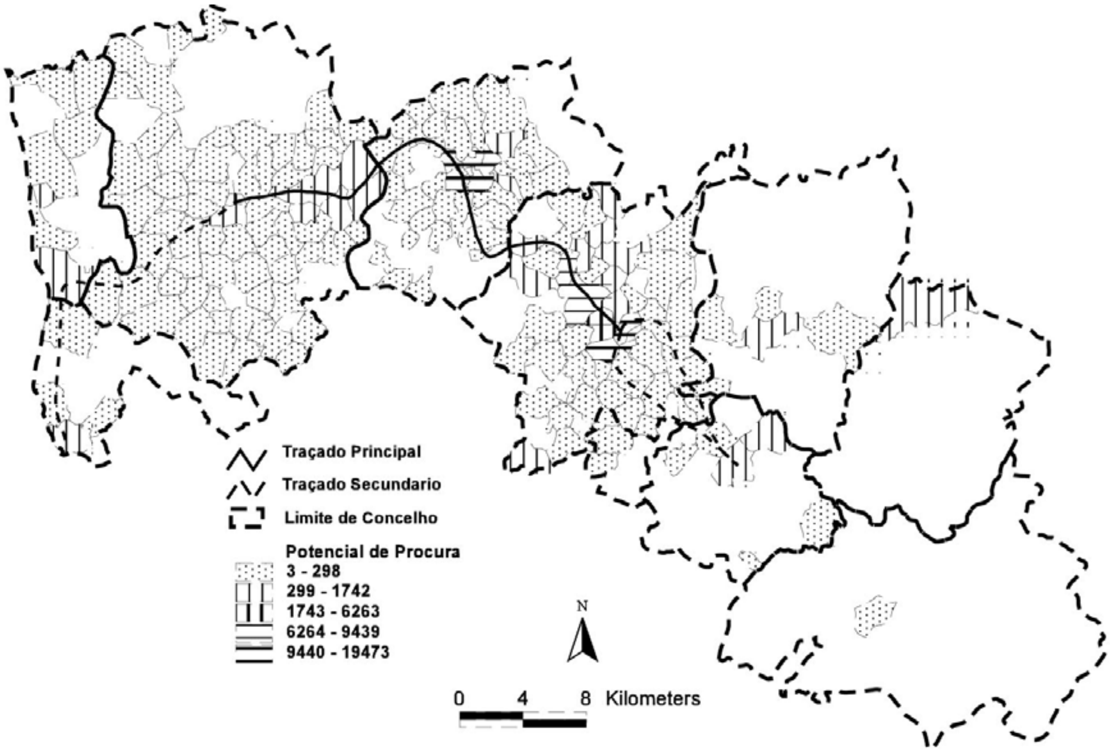

Figure 4 - Classification of the potential of mobility and suggestion of main rail line

The next phase of this analysis, included sorting out among the zones/parishes the ones that presented a demand potential and that would be interesting to the estimation of the demand potential of the rail services. The constraints include the distance to the proposed rail line, but also the weights given to each zone/parish accordingly to the proposed new rail line.

In order to filter the zones with interest, it we had to identify at this first line were the which were the zones served by the proposed line (1), to the zones not served a (0) value was given. It was required to define a proximity buffer zone to the proposed rail line; the option was to consider a buffer with 1km around the proposed line. The distance of 1km had to do with the fact that the considered walking distances bigger than 10minutes won't be attractive to the public transportation users. The Public transportation users usually have bus services with higher proximity and that time period of 10minutes allowed to walk around 1km around the proposed line (slope will constraint significantly this value).

This phase lead to another filtering of the served parishes/zones, reducing therefore the zones that potentially could had interest to the estimation of the potential of demand (Figure 5)

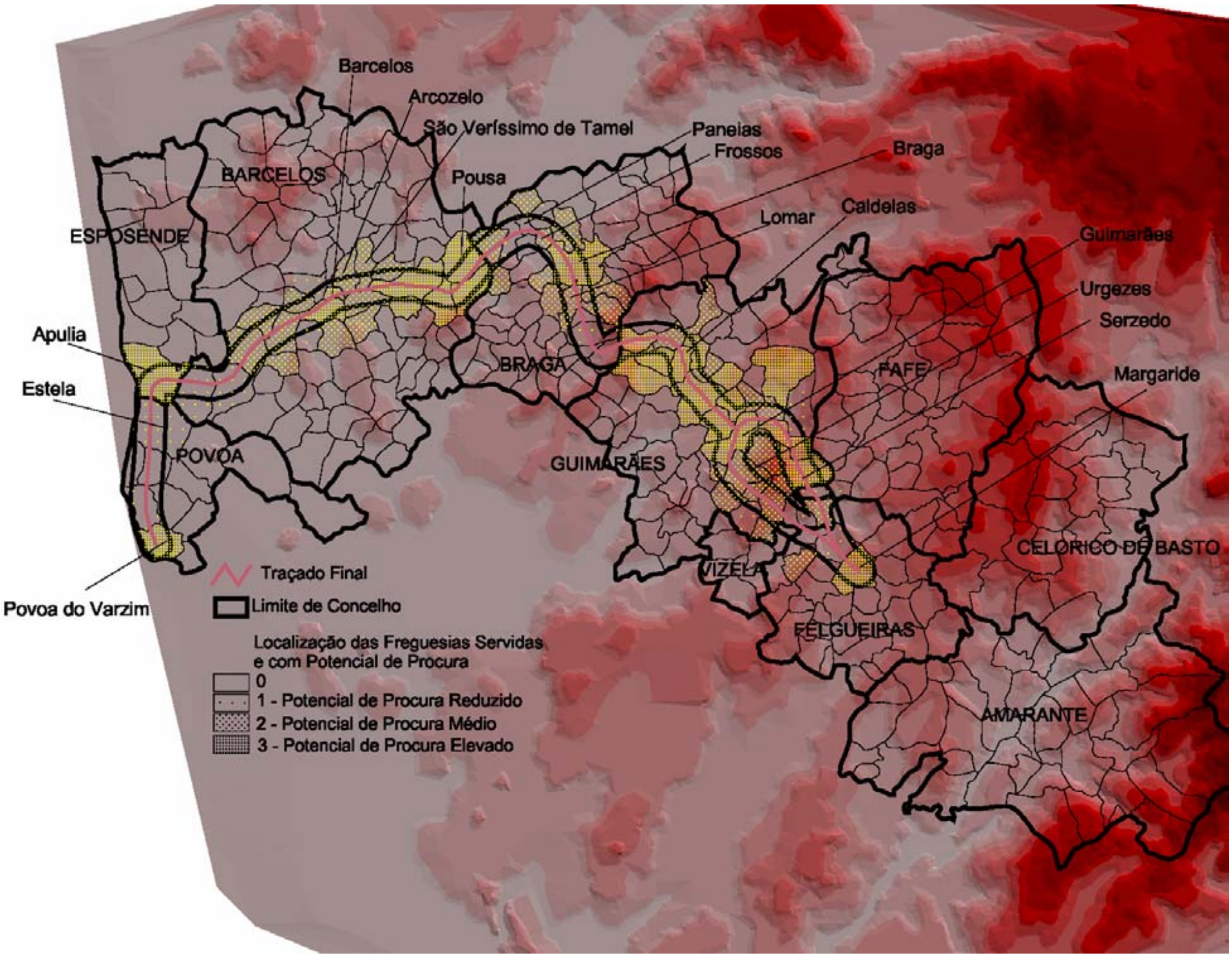

Figure 5 - Filtering of the parishes/zones served and with potential of demand (buffer zone around the line of 1km)

The last phase of the process of filtering of the zones includes a set of procedures to reach a better understanding of the zone's characteristics, both in terms of its demand potential and accordingly to its population's density .The used methodology in order included the identification of the percentile 25 and the percentile 75, both to the demand and to the population density. Once the percentiles were identified, it was possible to cross the information of both to establish different classes of potential of demand (high, average, reduced).

The association of the values of demand with the percentiles was done in a way that the matching of the lower values of the demand correspond to the lower values of the population's density (and *vice versa*). The extracted value of the percentile would give the limit of each class. Table 2 presents how the sorting of the percentile and the equivalent class was done; on the right side it is presented the key that allowed to match the classes and reclassify the potential of demand accordingly to the population's density.

	Demand	Density of Population
	A	A
Percentile 25	[26,75]	[20,07]
	B	B
Percentile 75	[266,5]	[57,79]
	C	C

AA	AB	BA	A [1]
AC	BB	CA	B [2]
BC	CC	CB	C [3]

Table 2 - Classification of the Percentiles of Demand and Density of Population

Afterwards, having the remaining zones that were contained by the proposed buffer, another filtering process was performed: identification of the zones that were directly under the proposed line and the zones being at the buffer of 1km but not under the proposed rail line.

Lastly, to the zones under the proposed rail line, a final calculation was performed: assessment of the average point of the proposed rail line at each zone, as well as accessibility to the centre to all the zones that were under the buffer. This information was used afterwards to estimate the tracks with light rail and the walking tracks in order to reach the light rail stops.

Concluding remarks

The used methodology in order to estimate the potential of demand of a new rail line, was based on a phased process of interaction of the results of the characterization of the region's mobility through the use of GIS. These processes included importing the mobility survey results into the GIS, and the chance of visually evaluate the origins and destinations of TC and TI at the regional scale, what by its turn enabled to propose a set of possible lines accordingly to the mobility results, characteristics of population and the topography of the region.

This process allows to set at an early phase, when considering all the options that might lead to the development of an infrastructure, the elements that are considered important in its costs structure (i.e. construction costs associated to the slope characteristics) and also in what concerns its potential earnings (once it allows to get the proposed line closer to the mobility needs of the populations).

The use of GIS allowed the identification of the multiple options of setting and expanding the proposed service. Using data from slope, mobility and demography it was possible to identify the main axis that allowed for some potential of demand. Therefore this tool can be used in the definition of priorities in what concerns future studies or in the deepening of existent studies.

The identification of the trips that might be significant to determine the potential of demand was easily done and enhanced through the use of GIS. As it is usual with this kind studies, among a set of trips of the region, only one part is covered by the proposed service, and the GIS allowed to construct the proposed buffer, identifying the zones served by this light rail service. Once the zones that effectively would contribute were filtered, service was classified accordingly to proximity to the line, what consequently would mean the chance to include at the demand studies, the capacity of including the factor "distance to the stop".

Once the region's mobility is characterized and a proposal of the line is set, the process of estimating a demand potential included a model of discrete choice in order to determine the behaviour of the population to the new proposed service. Once there isn't such a service at the region, it was required to perform a survey of declared demand as a data source in order to run the model of discrete choice. While the use of these surveys presents some drawbacks, the collected information allowed calibrating the model of discrete choice and giving some directions on the options made by the travellers in case they have more than one mode of transportation.

On the other hand, the application of the discrete choice model allows to identify the factors of the utility function that condition the most the option by a specific mode of transportation. As a fact, the calibration of the "replacement taxes" among the different attributes of the function utility can reveal a strong indication on the value given to the different attributes of the transportation alternatives (and this should be important information at the stage of conception of a transportation service).

During the development of the work some problems were identified: one important aspect when developing calculation on the demand potential has to do with the lack of information on mobility subjects. In order to perform this study data from the national institute of survey (INE) has been considered, nevertheless, the available information is barely existent and it isn't available at a disaggregated level (such as the parish – zoning unity of the study). Another issue is related with the difficulty of including geographical information in order to feed the GIS. A third problem, results from the inclusion of both census data with geographic data, some of the parishes defined at INE were not included at the geographic information.

In what concerns the data resulting from fieldwork, some of the surveyed people could not identify the origin and destination parishes/zones. This was the case at the city of Braga and Guimarães, whose centre is divided in several parishes unknown to most of its residents. Therefore the option was to group the different zones in one major zone. Some surveys also needed to be discarded due to inconsistent answers.

It seems fair to state that the developed methodology can be improved in some areas: the incorporation of more land use information might optimise the GIS use, improving the results. Information on the degree of consolidation of the urban centres could allow pinpointing interesting locations to the proposed rail line. In what concerns the model of discrete choice and to the study of the consumers behaviour once a new transportation mode is developed, it was not explored the possibility of calibrating the LOGIT model to different market segments. This analysis would improve the calibration of the model and estimate with increased detail not only the demand of the service but also the typology of the demand.

Acknowledgments

The authors wish to express their gratitude to REFER.EP, especially to Eng. Carlos Reis. Work was conducted while the primary authors were master students in Transportation (Master of Transportation, Department of Civil Engineering and Architecture, at the Instituto Superior Técnico (Lisbon Technical University) and were awarded a TRANSPOR scholarships.

References

Bhat, C., 1995. A Heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part 29: 471-483.

Brownstone, D., David, B., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B 34: 315-338.

Recker, W., 1995. Discrete choice with an oddball alternative. Transportation Research Part B 29B: 201-211.

Viegas, J., 2003a. Slides da Aulas de Modelação da Mobilidade Urbana, Sessão 1b. Instituto Superior Técnico, Mestrado em Transportes. http://cesur.civil.ist.utl.pt/~sgvct/mt/2_semestre/MMU/mmu_2002.htm

Viegas, J., 2003b. Slides da Aulas de Modelação da Mobilidade Urbana, Sessão 7b. Instituto Superior Técnico, Mestrado em Transportes. http://cesur.civil.ist.utl.pt/~sgvct/mt/2_semestre/MMU/mmu_2002.htm

Wen, C., Koppelman, F., 2001. The generalized nested logit model. Transportation Research Part B 35: 627-641.