SSD

SCIENCE FOR A SUSTAINABLE DEVELOPMENT

MOBILITIES AND LONG TERM LOCATION CHOICES IN BELGIUM

"MOBLOC"

P. TOINT, T. EGGERICKX, P. GERBER, S. CARPENTIER, E. CORNELIS,
L. DAL, S. KLEIN, X. PAULY, F. WALLE

ENERGY

TRANSPORT AND MOBILITY

AGRO-FOOD

HEALTH AND ENVIRONMENT

CLIMATE

BIODIVERSITY

ATMOSPHERE AND TERRESTRIAL AND MARINE ECOSYSTEMS

TRANSVERSAL ACTIONS

.be

SCIENCE FOR A SUSTAINABLE DEVELOPMENT
(SSD)

Transversal Actions

FINAL REPORT

MOBILITIES AND LONG TERM LOCATION CHOICES IN BELGIUM

"MOBLOC"

SD/TA/04

Promotors

Philippe TOINT
FUNDP, GRT

Thierry EGGERICKX
UCL, GéDAP

Philippe GERBER
CEPS/INSTEAD, GEODE

Authors

Samuel CARPENTIER
CEPS/INSTEAD, GEODE
Eric CORNELIS
FUNDP, GRT
Luc DAL
UCL, GéDAP
Sylvain KLEIN
CEPS/INSTEAD, GEODE
Xavier PAULY
FUNDP, GRT
Fabien WALLE
FUNDP, GRT

D/2011/1191/30
Published in 2011 by the Belgian Science Policy
Avenue Louise 231
Louizalaan 231
B-1050 Brussels
Belgium
Tel: +32 (0)2 238 34 11 – Fax: +32 (0)2 230 59 12
http://www.belspo.be

Contact person: Georges Jamart
+32 (0)2 238 36 90

P. Toint, T. Eggerickx, P. Gerber, S. Carpentier, E. Cornelis, L. Dal, S. Klein, X. Pauly, F. Walle - ***Mobilities and long term location choices in Belgium "MOBLOC" -*** Final Report. Brussels : Belgian Science Policy 2011 – 81 p. (Research Programme Science for a Sustainable Development)

Table of content

SUMMARY **5**
1. INTRODUCTION **9**
2. METHODOLOGY AND RESULTS **11**
A. Global methodology **11**
B. Accessibility models **14**
Methodology steps for the setting up of the Private Vehicle accessibility modelling 14
Observations data sets and accessibility models for comparisons 16
Road network modelling 17
Principles of traffic assignment and model comparisons 19
Off-peak hours accessibility modelling 24
Morning peak hours accessibility model 26
Communal accessibility indicators 29
C. Travel demand model **32**
Gravity model 32
Introduction of border effect 33
Model estimating 34
Model assessment 35
Residual analysis 36
Conclusion 38
D. Propensity to move model **39**
Residential migration: a concise overview 39
Pre-selection of variables for the propensity model 40
Methodological choices to build the model 43
Transformation of variables 45
Sample design 46
Model(s) 47
E. Localization model **53**
Preliminary analysis 53
Methodology 56
Models and results 63
3. POLICY SUPPORT **67**
4. DISSEMINATION AND VALORISATION **69**
A. MOEBIUS - Mobilities, Environment, Behaviours Integrated in Urban Simulation **69**
B. SimBelgium - NAXYS **70**
C. Population and households forecasting of Belgian municipalities **72**
5. PUBLICATIONS **73**
6. ACKNOWLEDGMENTS **75**
7. REFERENCES **77**

ANNEX 1 : Copy of the publications

ANNEX 2 : Minutes of the follow-up committee meetings

The annexes are available on our website
http://www.belspo.be/belspo/ssd/science/pr_transversal_en.stm

SUMMARY

A. Context

Mobility and transport evolve with time and the passing generations. Interactions are numerous between daily mobility and household migration (here defined as house moving implying a municipality change). The evolution of the transport system has deeply modified the barrier of distance and has largely opened the choices in terms of residence place. The continuing urban sprawl phenomenon resulting from these modifications has itself resulted in a strengthening of the property and housing market in certain territories, pushing people (young couples in particular) towards a residential localization which is further and further away from the traditional, urban activity centres. The tensions between daily and residential mobility have therefore increased, notwithstanding the recent rise in energy costs. This in turn generates unsustainable effects on society and environment.

But these new residential choices have in parallel induced new mobility behaviours, based on an extensive (and probably excessive) use of the private car in daily trips (home-work/school, shopping, leisure ...). Social life itself (visits to friend and family) has become more spatially dispersed. One already knows that the propensity to change residence is determined by a number of individual or household characteristics such as age, citizenship or income, but the effects of long-term trends as population ageing, the evolution of the household/family structure on both residential choices and mobility behaviours remain so far largely unanticipated.

This research project aimed at analyzing interactions between demographics and the evolution of mobilities at different time-scales. In particular, localization choice for household, daily accessibility and internal migrations appear to have strong relations.

B. Objectives

The project objective was so to investigate the link between long-term society evolution, residential choice, transportation demand and the resulting accessibility evolution. On the societal trends side, particular attention had to be paid to the population ageing effects, evolution of family cell structure and inter-generational relationships, but other variables such as land-use and standards of living trends have also been considered for potential inclusion in the developed models.

On the transportation side, emphasis had to be put on the evolution of transportation demand (gravity model/mobility demand model more lately described) and resulting traffic conditions (accessibility models).

The project objectives also included simulations of scenarios for the future in Belgium but as researchers faced a lack of time, such simulations could not be achieved. Nevertheless, the main models are available for these simulations. Furthermore, the links between those models are ready to use.

C. Conclusions

Major conclusions concern the main models, wihch are, a propensity to move model, a localization model, a transportation demand model and accessibility models.

(i) The transportation demand model allows the computation of the origins and destinations of home-work travels flows in order to provide an estimation for the accessibility model. Indeed, from the margins of an origin-destination matrix, this model, calibrated on exhaustive data from the national census of 2001, can compute the flow between each pair of Belgian municipalities. The accuracy could be improved to better model the long distance trips but without consequences on the shorter ones being currently well estimated. This transportation demand model can feed the accessibility models developed in order to get municipal accessibility indicators. These indicators take account, on one hand, of the accessibility to jobs during morning peak hours and, on the other hand, of the accessibility to other daily activity places during off-peak hours. These indicators rely on travel times computed thanks to an accessibility model calibrated on a highways and national roads network. This network was characterised with a typology based on the crossing of urbanized areas from the CORINE land cover GIS layer. Furthermore, a check of the speed, the length of the travel and the travel time difference between declared speed, length and travel time of the MOBEL survey in off-peak hours was achieved.

(ii) The propensity to move model shows us that this propensity depends on the life cycle of individuals, and more particularly their family trajectories. One can observe this link through the well known correlation with the age and the transition of the household structure. In fact, the transitions leading to move are break-up situations, new family-units compositions and leaving parents' home. It brings out that the more stable situation concerns people who are in married couple (with or without children); this situation which is often associated with an owner status, which is another factor of residential stability. In other words, the likely evolutions of family situations marked by the rise of less stable households (cohabitation situations, one-parent family…) should still generate higher propensity to move rates in the coming years.

(iii) The localization model mainly shows us that people tend to settle down in municipalities where "similar" people live (regarding e.g. the household structure, the age). The relocation often takes place on short distance. People rarely leave their residential municipalities to find a new dwelling in a far away place. Accessibility indicators are also significantly explanatory for the residential choice, although, as it turns out, not dominant.

D. Contribution of the project in a context of scientific support to a sustainable development policy

The project's objective was to investigate the links between long-term residential choice, accessibility and daily mobility, with the ambition to provide better understanding of the behaviour of Belgian households regarding these issues. In particular, the respective importance of several categories of factors is crucial for the establishment of land-use and accessibility policies. The mixing of long-term decisions such as housing and short-term ones such as daily mobility turns out to be a challenging issue.

The accessibility model outputs provide a few relevant accessibility indicators; they allow the characterization and comparison of municipalities and we feel that this could have direct consequences on municipality management.

The model describing the propensity of the Belgian individuals to move their residence (which incorporates a number of explanatory factors at the individual and household levels) shows us that the choice of changing residence is mainly caused by alterations in the household structure, accessions to a different position in the household (from child to head e.g.) and age classes. This reinforced the idea that societal trends (as opposed to material infrastructure evolution) are crucial to explain internal migration within a country. In particular, population aging and the increase of "narrow" households may present specific challenges in urban planning and land-use in general during the forthcoming years.

Finally, the residential localization model is central to the design of suitable land-use regulations at the regional level. Remarkably, analysis indicates that the dominant factors are, by decreasing level of importance, the distance between the previous residence and the new one, the perceived quality of life in the new municipality of residence, the household structure, and, in fourth position, the accessibility of the new municipality of residence. Accessibility is therefore less important than expected at the start of the project.

A merely interpretation of the results from the MOBLOC project is that migration within the country is less determined by infrastructural factors (within which accessibility is an important example) than by factors related to societal life in a more general sense: household structure and its evolution, closeness to one's relations, age class and quality of life score indeed higher in our results than purely transport related factors.

We clearly believe that these conclusions are important for any forecast on the development of land-use and transportation systems. They will be notably discussed in the new regional prospective study group (SRP) established under the leadership of the Institut Destrée and the Institut Wallon d'Evaluation, de Prospective et de Statistiques (IWEPS). We also plan to disseminate those conclusions more widely, via scientific publications but also aiming the municipality managers and the general public.

E. Keywords

Accessibility, daily mobility demand, propensity to move, residential localization, household structure, municipalities.

1. INTRODUCTION

Interactions between daily mobility and household migration are complex. The aim of the MOBLOC project is precisely to study, from one side, the impacts that daily mobility can have on residential localization (places, accessibility and the way it influences residential choices), and, from the other side, the repercussions that household migration can have on daily trips (transport demand and congestion) and the linking feedbacks. Indeed current residential choices (periurban localization) induce new mobility behaviours, often based on an extensive (and probably excessive) use of the private car for daily trips (home-work/school, shopping, leisure ...) which generate unsuitable effects. Social life itself (visits to friend and family) has also become more spatially dispersed.

One already knows that the propensity to change residence depends on a number of individual or household characteristics such as age, nationality or income, but the effects on both residential choices and mobility behaviours of long-term trends such as population ageing or the evolution of the household/family structure remain so far largely unanticipated. In this context, modelling can surely help the understanding of the behaviours and allows simulation. This is the way followed by the MOBLOC project; the models which have been developed participate to this objective and are important constitutive elements of this even more ambitious project.

The present report will successively address the global methodology of the project, the results of different modelling exercices achieved to reach the objectives without missing out the methodological aspects of these different steps and their respective conclusions. A first model concerns the accessibility at municipalities' level and calculates travel times between each pair of municipalities (during off-peak hours and morning peak hours). These travel times allow the computation of accessibility potential to employment and services (later used as explanatory variables in the residential localization model). Prior to this accessibility model, a transport demand matrix has have to be calibrated thanks to a gravitational model taking employment repartition and residential localization of active people into account. Then a two-fold model addresses the residential changes. First, a propensity to move model models the fact that an individual decides to settle down in a new municipality between two successive 1st of January or not. A residential localization model follows and estimates several parameters modelling the choice of a new residential municipality. This latest model notably uses the output of the accessibility model. The report will successively describes all these developed models.

A policy support section follows; this section underlines the main outcomes and their interests in a frame of sustainable development. As the MOBLOC project can be the start point for new researches, some future projects will also be briefly described: these will rely on the new expertise available thanks to the MOBLOC project. Useful references finally conclude the report.

2. METHODOLOGY AND RESULTS

A. Global methodology

Let us so first describe the global methodology of the project. This description will rely on a presentation of the data flows through the inputs and outputs of each model and the interactions between them. For that purpose, the FIGURE 1 (Pattern of the MOBLOC Project) below could be very helpful.

The main objective of MOBLOC is to draw up a link between residential migration and daily mobility. Thus the two main bricks of the project are the residential migration model and the accessibility model. As the figure shows, other models are necessary to reach this goal. Indeed, we need to determine some inputs for these two main models and to establish the links relating these inputs

The inputs of the residential migration model (1a) are individual data. They include age, gender, level of education, household evolution, previous migration and the current residential municipality (at time Y). The final output of this model is the new residential municipality (one year after, Y+1). This model has been split up into two sub-models: a propensity to move model and a localization model.
The first one uses some of the individual information to forecast if a migration occurs (i.e. if an individual moves) from year Y to year Y+1 (from a municipality to another – not inside a given municipality). If the answer is yes, then the localization model will simulate the localization choice between the 588 remaining Belgian municipalitiessince moving individuals may eventually not choose their original municipality.
The technique used for the propensity model is a binary logistic regression while the localization model will resort to a more sophisticated discrete choice technique. These methodological issues will be further detailed in following sections.

At the first step of the transportation models, the travel demand model (2) is based on two inputs: one from the migration model (1a): the number of working people and one from the evolution model (1b):.the number of jobs per municipality.
These figures constitute the margins of the O/D matrix to be calculated by the gravity model. Let us remark that for now, this matrix (demand) only takes into account the home-work and home-school trips because of the used inputs. Even if this assumption is very restrictive, it is a first good approximation considering the available data.

The next step of the transportation models is the modal split model (3). Its purpose is to compute from the "global" (all modi) O/D matrix another one concerning only the trips made by car. This step is necessary to provide a feasible input for the private vehicle accessibility model which is an essential brick of the project.
The private vehicle (PV) accessibility model (4) uses Wardrop equilibrium paradigm to assign the traffic (car trips) on the roads network. It provides a matrix of the travel times between each municipality's pair (a first matrix during off-peak hours and a second one during the morning peak hours on a working day) by car.

At earlier stages of the project a public transport (PT) accessibility model was also forecast. It would have computed the travel times by public transport. between each municipality's pair This model has not been developed due to the unavailability of a complete dataset for public transport timetables all around Belgium. Whatever, let us also point out that these times would not have been directly related to the travel demand because they would have been based on the supply. The PT accessibility model could not have been updated at each step of the simulation as we do not know how the public transport level of service will evolve in the next years. So as we did not manage even to receive complete information on the current supply, this model was not built, even if it would have been interesting to test its contributions for the localization model.

At the last step of the transportation model, the PV travel times between municipalities are used to compute different accessibility indicators (to employment and to services) per municipality. They are included as covariates of the localization model (the second step of the residential migration model) to measure the attractiveness of the municipality.

This chain of models could then be used in a prospective way. Therefore it will be necessary to update the input values for the residential migration model. Let us remind that this model works at an individual level which implies to find adapted techniques. This would be the role of the evolution models (1b). They could include different kinds of techniques at different levels: aggregate and disaggregate, e.g. demography trends, synthetic population… Although these developments are clearly interesting, they are postponed to further research. The research group is aware that the original program did include steps in these directions, but the determination of the accessibility model, propensity to move and residential localization turned out to be enough challenge for the contract duration. This research agenda is however considered in another research project relying on MOBLOC (see section 4, Dissemination and valorisation).

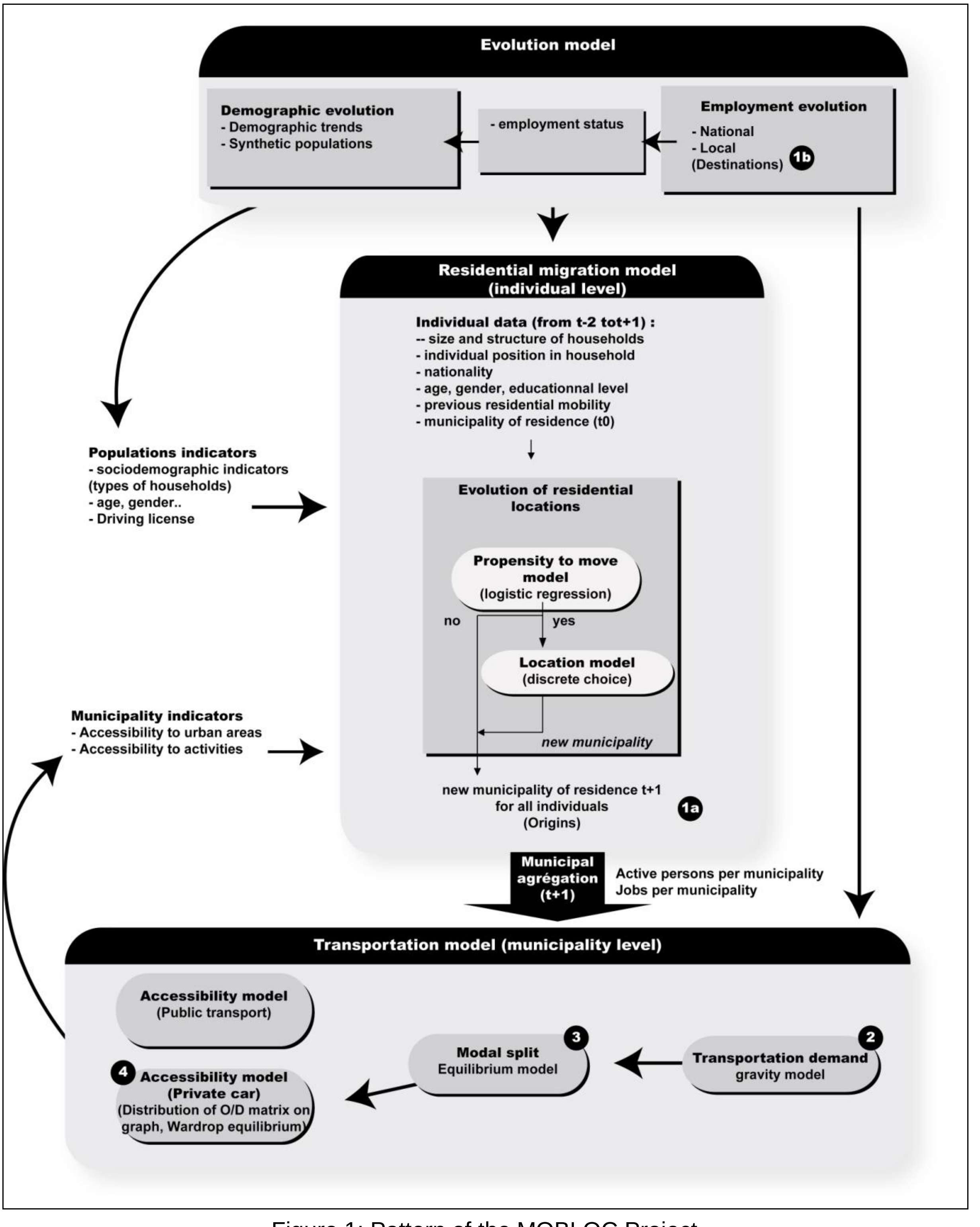

Figure 1: Pattern of the MOBLOC Project

B. Accessibility models

This part deals with the setting up of a road accessibility model (step 4 of the previous figure), at the scale of Belgium, and its validation. For the whole country, all the required parameters are not available to calibrate and validate the off-peak and peak hours accessibility models. Considering this lack of data, several goodness-of-fit statistics were applied on the basis of different comparisons between our model values and, on one hand, those of observations (MOBEL, 1999), and, on the other hand, those of other models (Vandenbulcke et al.,2009).

Methodology steps for the setting up of the Private Vehicle accessibility modelling

As shown on Figure 2, several steps have been carried out. At first, a road network and municipal centroids representation was set up.

As the modelling is to be done throughout the Belgian territory, the first data set up step consists in choosing a representative point for each of the 589 municipalities. If the choice of single municipal centroids constitutes an approximation, this appeared to be necessary for both a methodological consistency reason -the model coupling of MOBLOC is done at the municipalities level - and a practical reason, this being the aggregation level at which the origin/destination matrix was available (ESE2001 for the first run, gravity model for future simulation). This work was accomplished in the first phase of the MOBLOC project and fully described in its final report (Cornélis et al. ,2009) and therefore not further described here.

Afterwards, the mathematical representation of the road network was set up as follows:

G is a directed graph where G = (V, E),

with V: the set of vertices representing the road intersections

E, the set of edges representing the road sections with homogeneous properties.

It notably consists in describing these edges' length, free-flow speed and capacity. Since this information is not available for the whole country, we used a network typology drawn up by crossing with a land-use GIS layer as described further in subsection "Road network modelling".

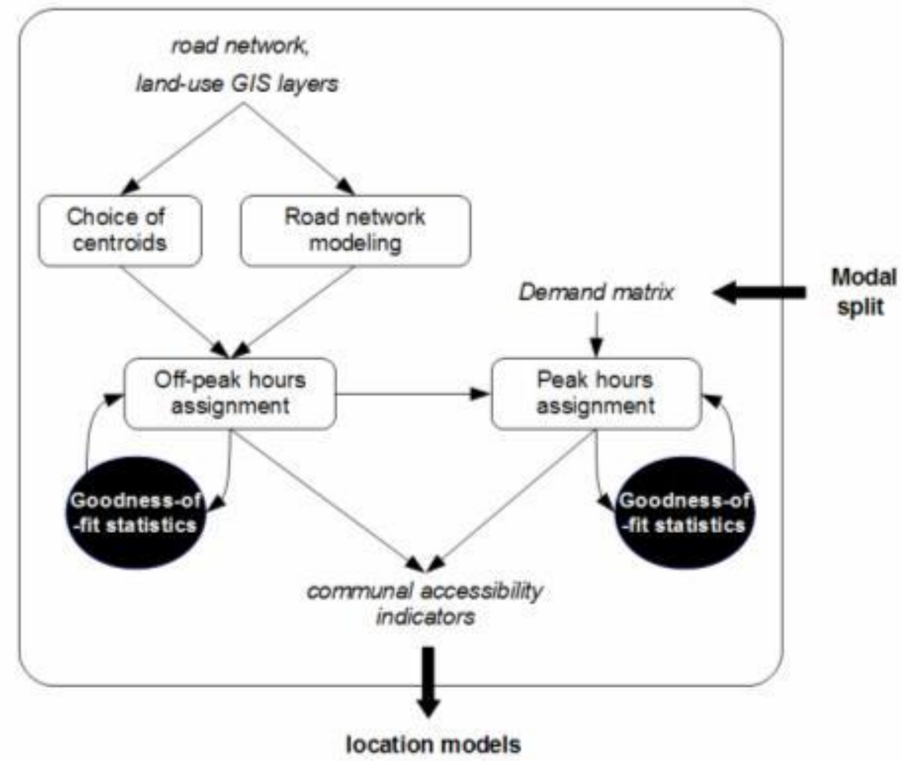

Figure 2: Methodology steps for the setting up of the private car accessibility modelling

In a second step of the modelling procedure, two road accessibility models were set-up, the former for off-peak hours, the later for morning peak hours (7 to 9 A.M.). The off-peak hours model is based on a free-flow calculation of the shortest path on the network. Morning peak hour's model involves a travel demand matrix, that is to say, an origin-destination exchange matrix of workers and students using a car (as main driver) for their home to work or home to school trips. This matrix allows us to include later the new localization outputs from the MOBLOC localization model and is extracted from the National Socio-economic Survey of the INS (2001) for the first run. The results of the two models are then compared to observed data from the MOBEL survey and to results from other available models for Belgium. Thanks to goodness-of-fit statistics measuring the accuracy of the model according to the observed values, it is then possible to calibrate its parameters and its accuracy compared to other models.

Finally, once both models (on-peak and off-peak) are validated, it is possible to calculate the municipal indicators. Then these indicators can be introduced in the localization model, therefore modelling the effect of daily mobility on the residential choices.

We assume here, that for their residential choices, workers primarily consider their travelling to work and related travel-time to ensure the feasibility of their activities schedule. In other words, people have to maintain their time-budget within some limits, as assumed by Zahavi's stable travel-time budget (1979)[1]. The main constraint influencing this choice is therefore the travel-time during the peak hours of a working day.

Observations data sets and accessibility models for comparisons

In this part we describe the data sets and accessibility models that we used for the model building and validation.

Demand estimation in peak hours: ESE 2001

Source: The demand matrix is drawn from ESE 2001 National Socio-Economic Survey.
Use: demand matrix for peak-hours assignment
Data processing: We take the drivers into account for the two following purposes: home to work and home to school travels. They represent the vehicles flows. The share of the trips within morning peak hours flows is estimated from MOBEL survey to 33.7 % (of total daily traffic) between 7 and 9 am.

MOBEL survey (1999)

Source: The available observation data set used for the validation process is taken from the national survey about mobility in Belgium conducted in 1999 in the framework of the MOBEL project (Hubert and Toint, 2001). We thus have, at the scale of the country, observations about travel times during a working day (excluding school holidays). The number of observations meeting these criteria is 10,036.
Use: This data set will be used as a reference for the comparisons in the morning peak hours as well as in off-peak hours. It is also used for the network modeling checking step and to calculate the morning peak-hours ratio of the home to work and home-to-school trips.
Data processing: From MOBEL observations were removed those out of the field of our study, such as the observations with location out of the study area (origin or destination in a foreign country for instance) or with inappropriate values (null values or with distances which where very lower than the euclidean distance between two communal centroids, as well as the abnormal values detected through integrity checks). In off-peak hours, that is to say, out of the two in-peak period between 7 and 9 am and between 3 and 6 pm (Hubert and Toint, 2001) we then have a total of 1,125 origin/destination couples and 598 couples for the morning peak hours (7 to 9 am).

[1] Let us note that this rule about a constant time-budget is subject to discussions (Joly I., 2003).

An internet road model: Google Maps (2010)

Source: http://maps.google.com

Use: for the comparison of off-peak hours models and for the network modeling checking step

Data processing: It is possible to query Google Maps servers to calculate road itineraries and get travel time values for the selected origin-destination couples. At first, it was necessary to geocode origins and destinations and check their matching with the municipality centroids of our road model. The road network used by Google maps comes from the Ministerie van de Vlaamse Gemeenschap (Flemisch Ministry) and from the Ministère de l'Équipement et des Transports[2] (Walloon Ministry). This model is used during the stage of network checking and for the off-peak hours comparisons. It will be thereafter labelled GMAP.

Accessibility model from the UCL (2007)

Source: In the project 'Accessibility indicators to places and transports' an accessibility model was developed by Vandenbulcke et al. (2009) at the scale of Belgium. Their approach is noticeably different as it involves the calculation of impedance based on the communal population and jobs densities for each of the municipalities' road sections. Their road network contains 627,856 edges. Thanks to the authors, the travel time calculation for all the municipalities in off-peak hours was made available for us. On the other hand, the travel times in peak hours were available only for the 53 top level cities according to an updated urban Belgian hierarchy (Van Hecke, 1998). This model is thereafter labeled UCL.

Use: for the comparison of off-peak hours and morning peak-hours models

Data processing: selection of the OD couples available in MOBEL as well

After the description of the datasets and models used for the building and comparison of our model, let us continue with a quick insight about the road network setting-up.

Road network modelling

This section is about the road network setting-up represented on Figure 1. This step involves the use of several GIS layers to correct and complement the road network when necessary.

Our original road network comes from the Service public fédéral Mobilité et Transports. It has a range of roads from highways to the third level of Belgian national roads and is generally accurate enough for an inter-communal modeling (FIGURE 3). There although remained a bit of work to be fully ready for the modeling. The representative municipal centroids were linked to the network by connectors

[2] It is thus a priori the same network as ours with less detail l. http://www.google.com/intl/fr_fr/help/legalnotices_maps.html

when they were not farther than 200 m from the road network. If farther than 200 m, we supposed that the network was not enough detailed and therefore additional roads have been digitized. This was done when necessary on the basis of Google Maps and thanks to urbanized and commercial zone CORINE land cover 2001 layer.

The following step consists in setting up the free flow speed which is one of the traffic assignment model parameter. These speeds are based on a road sections typology taking into account the number of road lanes, the presence of segregated lanes, or the type of urbanization crossed by the road.

We present here three road networks discussed thereafter under the names Mob_a, Mob_b and Mob_c.

- The first model Mob_a is built on a road typology taking into account the crossing of urbanized municipalities according to the communal classification of Van der Haegen (1996).
- The second model Mob_b is built on a road typology based on the crossing of urbanized or commercial areas from the CORINE land cover GIS layer at a subcommunal level.
- The last model Mob_c is built on the basis of Mob_b after checking the speed, the length of the travel and the travel time difference with declared speed, length and travel time of the MOBEL survey in off-peak hours.

The Google Maps values were used as well, in order to compute shortest path, and faster travel time between two municipalities. We mainly investigated the records for which the declared distances between two municipalities were not too small. Indeed, the origins and destinations of MOBEL can be located anywhere in the municipality. We thus calculated that, in MOBEL trips, the average distance not going outside a municipality is four kilometres. This gives an idea about the uncertainty range related to our aggregated level of analysis compared to the intermunicipal distances. The different corrections varied from the additions of links in the road network, where the road hierarchy of our data set did not allow a realistic representation of the municipalities' access, up to changes in the values of road links parameters (i.e number of lanes, with the help of aerial photograph) or typology.

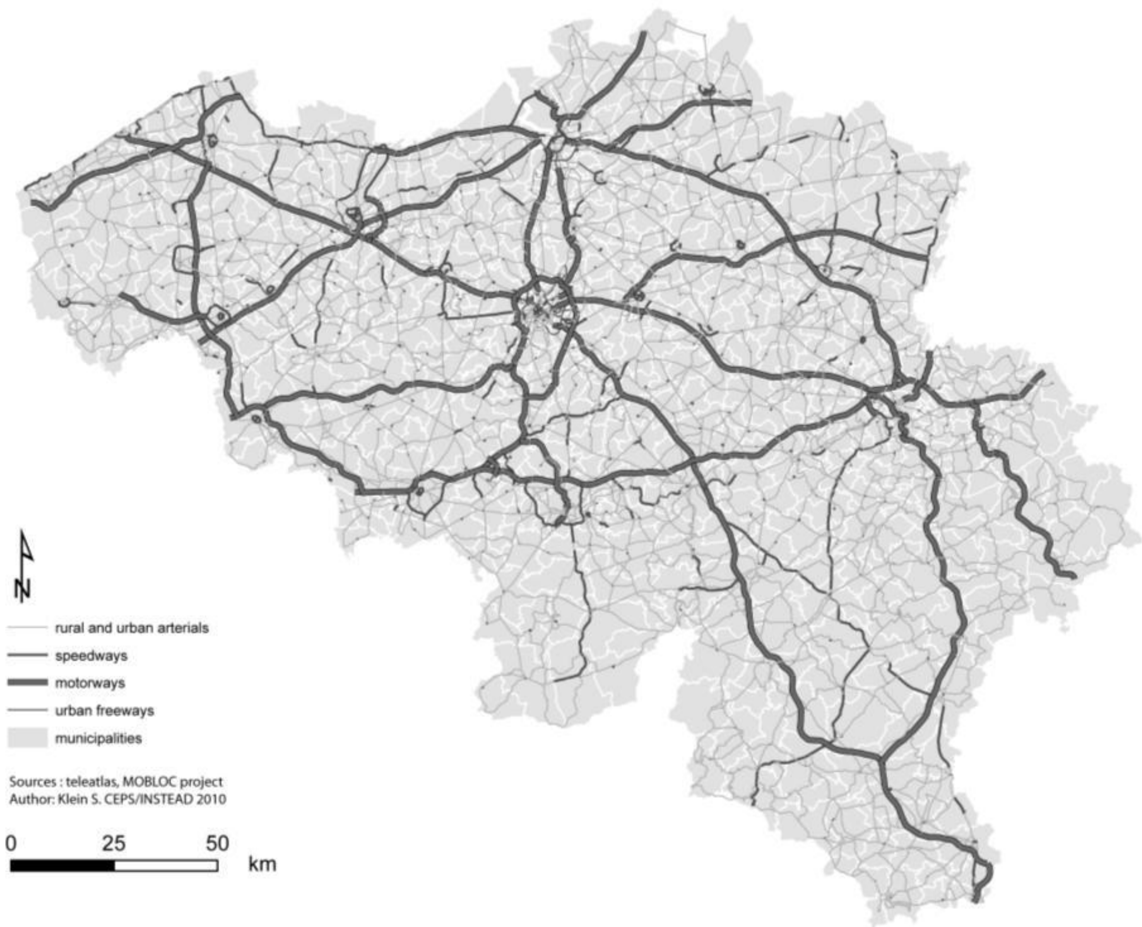

Figure 3: Belgian road network used in the accessibility models

The last important parameter concerns road capacity. According to the Ministry of Transports, the capacity of a highway is 2000 UVP per hour and traffic lane. This can also vary according to the road geometry and other factors. As far as roads in urban environment are concerned, the conventional capacity is 1200 UVP per hour and per traffic lane. For countryside roads this value ranges between 1400 and 2000 UVP per hour and per traffic lane.

On the map (FIGURE 3) one can notice the privileged position of Brussels at the center of the highway network connecting the major cities of the country: Antwerp up North, Ghent and Bruges up West. and the major cities of the Walloon Region with from East to West: Liege, Namur, Charleroi and Mons. The south of the country appears less equipped in infrastructures, reflecting thus its lower population density.

Principles of traffic assignment and model comparisons

After the centroids and network data are gathered and checked, the traffic modeling principles have to be settled. They will be presented in the following section. First, the principles of traffic assignment will be presented. Then, we will introduce the statistics used for the matrix comparison involved in the validation step, as well as the unipolar mapping of the estimated travel times.

Principle of traffic assignment

The main objective of the traffic modelling is to compute estimated travel time values between the Belgian municipalities for trips with private vehicle in order to calculate accessibility indicators. To do so, our methodology relies on a classical assumption of traffic engineering. This assumption specifies the decrease of the speed when the traffic approaches the maximal capacity of the road. Several functions describe this relationship and also define the parameters that have to be set. The assignment itself consist then on the use of a particular algorithm which will distribute the demand matrix (flows) on the network iteratively, taking into account the congestion effects on this distribution.

The modeling is done in two stages, the first for off-peak hours and the second for morning peak hours. During off-peak period, we assume that travels can be done at free-flow speed. This allows us to apply an algorithm to solve the minimum cost, as Dijkstra algorithm (1959) does, and thus to determine the shortest cost path. The cost function used here, is a simple minimization of travel times.

During morning peak hours, we use a model allowing assigning the demand matrix on the shortest cost paths, taking account of the relationship between the travel time on a road and the vehicles flow. These shortest-paths being beforehand determined in free-flow conditions, it is then possible to use the off-peak calibrated model for the morning peak model's setup. For each road network's edge a speed-flow function should be defined, specifying maximal speed decay while the number of vehicles approaches the maximal capacity of the road. A lot of functions were developed and discussed (Branston, 1976); so are BPR-like functions developed by the American Bureau of Public Roads:

$$T = T_0 (1 + \alpha (V/Q)^{\beta})$$

With T: travel time,

V: flow,

Q: road edge capacity,

T_0: free-flow travel time and

α and β : parameters of the function according to the road.

This function has the following curve (Figure 4):

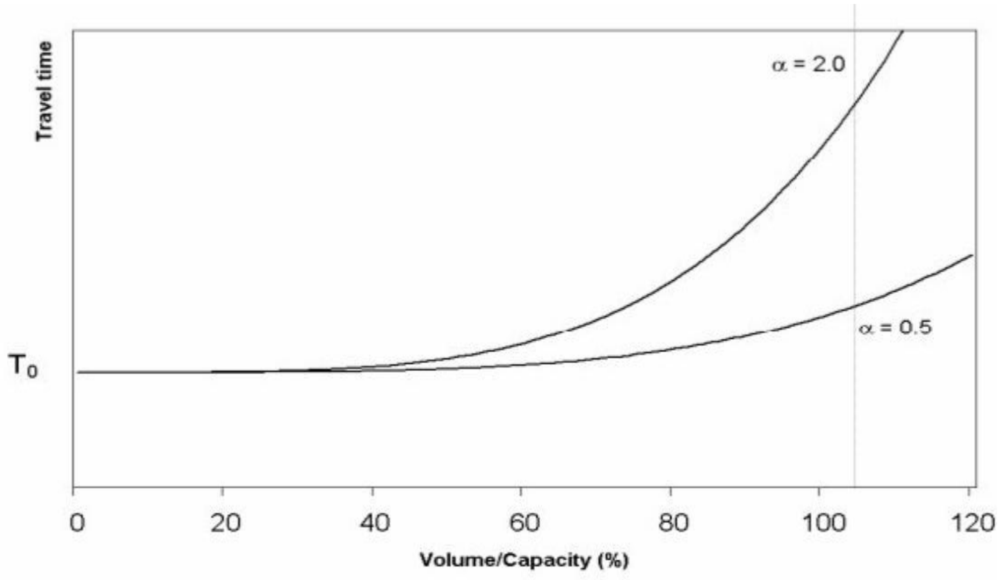

Figure 4: BPR curve (Bureau of Public Roads (1964))

The chosen model for the peak hours assignment is based on Wardrop's first hypothesis (1952) and specifies the conditions allowing reaching a user equilibrium. Such equilibrium stipulates that no any road-user can unilaterally improve his travel time by modifying his itinerary. An approximate of this equilibrium is reached by the method of successive averages.
At the end of this balancing process, a new speed is calculated for every road section. The travel time matrix is then calculated.

Validation principle by model comparison

After the setup of the accessibility model raised the question of its validation. We choose to compare our estimated values with declared travel times from the MOBEL survey and with results from other models based on distinct methodologies, in order to validate our model. This proceeding allows us the validation of our model, throughout goodness-of-fit statistics computed at the different steps of its development.
In order to test the accuracy of the model and its estimation, many statistics comparing estimated matrix with observed matrix are available (Knudsen and Fotheringham, 1986). The following statistics were selected considering their linear sensitivity to the error level for a complete matrix.

- SRMSE, the Standardized Root Mean Square Error:

$$SRMSE=\{\sum_{i=0}^{n}(p_i-q_i)^2/n\}^{\frac{1}{2}}/(\sum_{i=0}^{n}p_i/n)$$

with pi and qi, the observed and estimated travel times.

The lower limit of this statistic is zero indicating perfectly accurate prediction and its upper limit is generally one, though values higher than one arise whenever the average error is greater than the mean.

- phi statistic:

$$\phi=\sum_i p_i|\ln(p_i/q_i)|$$

with pi and qi, the probabilities of the observed and estimated flows. This statistic has limits of zero and plus infinity.

- absolute value of the psi statistic:

$$\bar{\Psi}=\sum_i p_i|\ln(p_i/s_i)|+\sum_i q_i|\ln(q_i/s_i)|$$

with pi and qi the probabilities of the observed and estimated flows, and si = (pi + qi) / 2. This statistic has no known theoretical distribution.

We tested two other statistics:

- AED: Absolute Entropy Difference, defined as the absolute value of the difference of the entropies of the observed and predicted probability values.

$$AED=|H_p-H_q|$$

with Hp and Hq Shannon's entropy (Shannon (1948)) so that, for example, $H_p=-\sum_i p_i \ln(p_i)$

- the chi-square statistic: χ2

$$\chi^2=\sum_i \frac{(p_i-q_i)^2}{q_i}$$

We tested the sensitivity of these statistics on a MOBEL sample for off-peak hours, introducing errors:

$$q_i = p_i + \delta \times (p_i \cdot rnd \cdot fact)$$

with qi, the estimated travel time values, pi, the observed travel time values, δ, random numbers {-1,1}, rnd, a random number between zero and one and fact, the percentage of error divided by 100. For each level of error the values of the following statistics: AED, χ2, SRMSE, φ and ψ, were calculated.

This process was repeated a hundred times and the average of the statistics were calculated. The sensitivity of each statistic to error level is drawn on Figure 5, where the horizontal axis represents error levels and the vertical axis represents the standardized value of the average statistics.
The SRMSE, φ and ψ statistics present a linear relationship with error level, on the contrary with AED and χ^2 that underestimate the error rate for low random error levels.

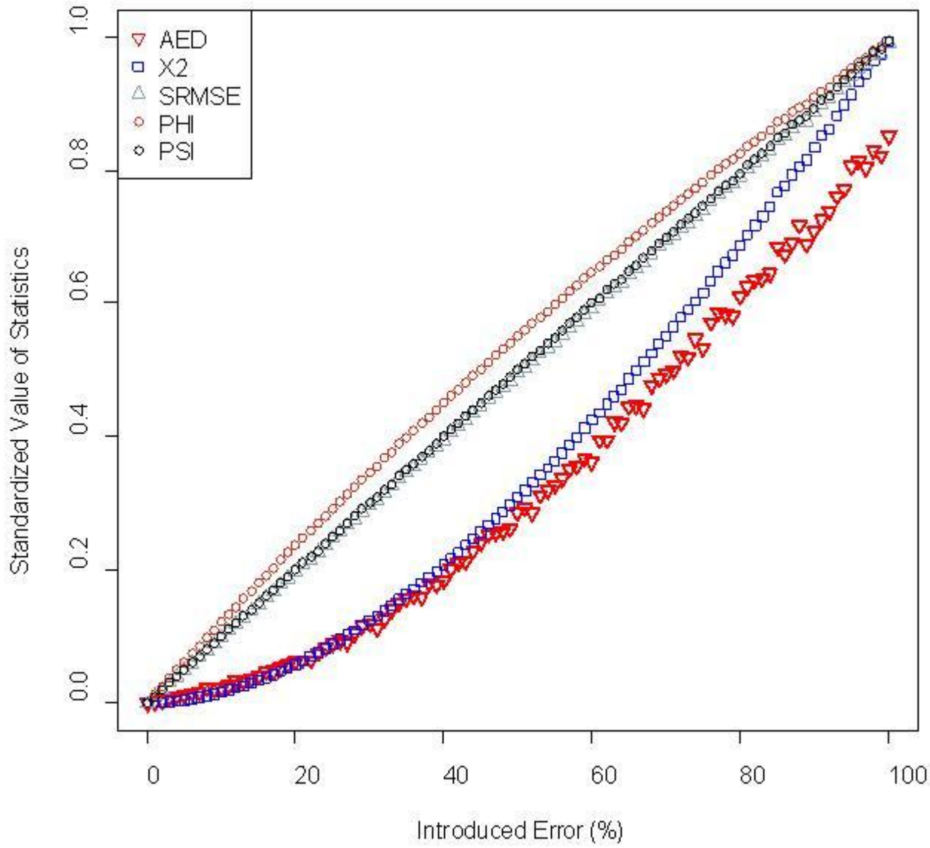

Figure 5: Error sensitivity of five Goodness-of-fit statistics

After this insight in the theoretical aspects of our method, lets continue with the results of the modelling, first in off-peak (free-flow) conditions, then during the morning peak hours.

Off-peak hours accessibility modelling

Off-peak hours assignment

The road model was imported in OmniTRANS software[3]. This software allows us to perform the traffic assignment for peak hours and off-peak hours.

The three travel time matrices (Mob_a, Mob_b, Mob_c) were calculated for off-peak hours with an all-or-nothing assignment based on the three road networks typology presented earlier (see subsection "Road network modelling"). The checking consisted in cartography of the estimated travel times compared with the observed travel times, as well as the goodness-of-fit statistics calculation of our model with the observed values from MOBEL and the statistics from other models.

The cartography of travel times from each municipality to Brussels destination is presented in Figure 6. We notice that the travel times spread amongst the country in different ways depending on the models. The Mob_a and UCL models have a large spread of the classes from Brussels to the periphery. This illustrates quite high speeds. Models Mob_b and Mob_c show a similar distribution to the one from Google Maps as well as a more tighten spread corresponding to lower speeds. Besides, we can see through the spread of the classes along highway axis an higher sensitivity to the road hierarchy compared to Mob_a and UCL.

[3] Version 5.0.28. This software is developed in the Netherlands by OmniTRANS International. http://www.omnitrans-international.com

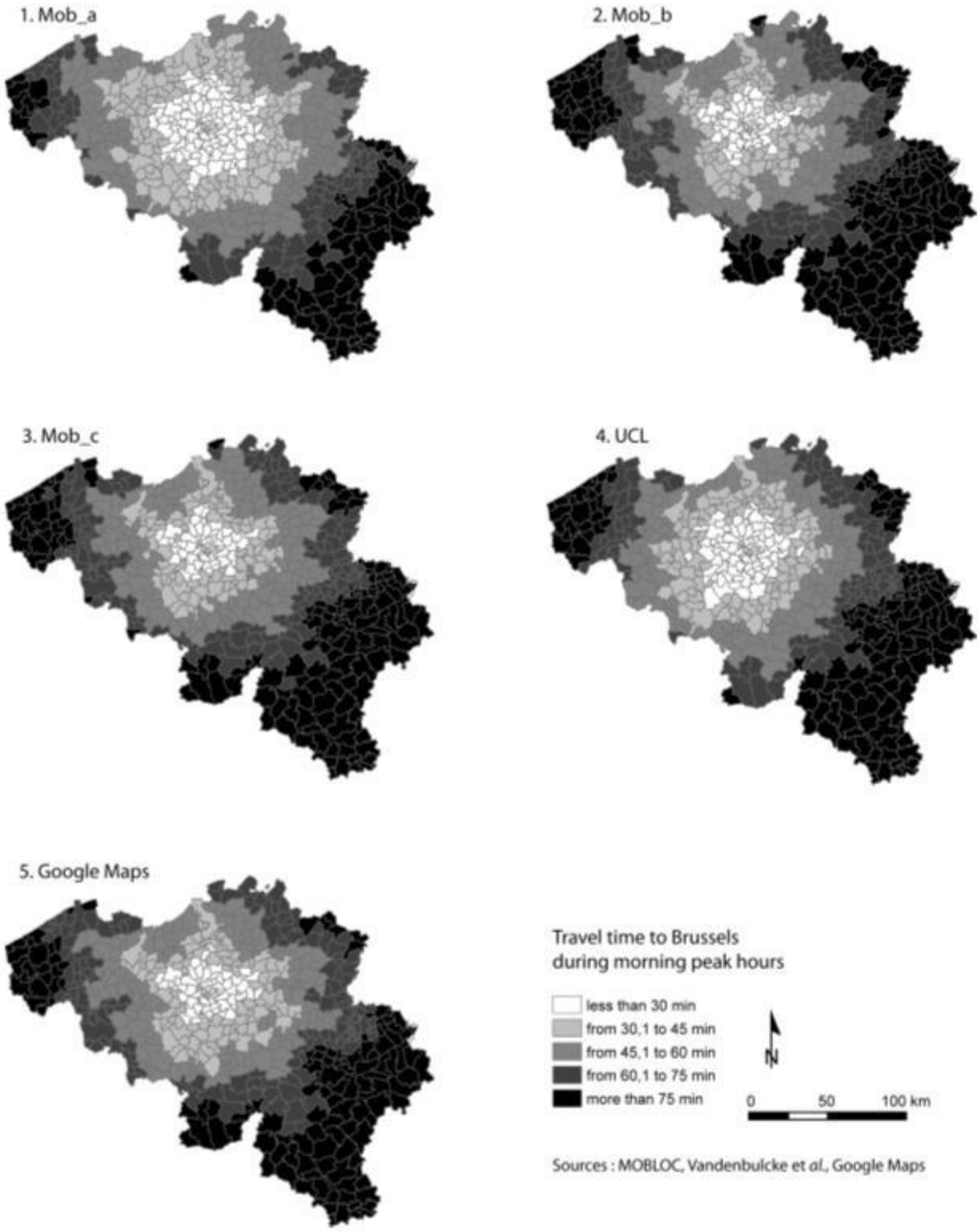

Figure 6: Representation of travel time in off-peak hours from different models

Goodness-of-fit in off-peak hours

The goodness-of-fit statistics of the models for off-peak hours are presented in Table I.

Table I: Goodness-of-fit values in off-peak hours

Model	AED	X^2	SRMSE	PSI	PHI
Mob_a	0,026	12670	0,84	0,30	0,33
Mob_b	0,037	5779	0,68	0,27	0,30
Mob_c	0,036	5789	0,67	0,27	0,29
UCL	0,048	11400	0,80	0,30	0,33
GMAP	0,024	6281	0,76	0,28	0,29

As far as AED statistic is concerned, Mob_a and GMAP have better values than Mob_c, UCL and Mob_c. On the contrary,Mob_b, Mob_c and GMAP models have quite the same magnitude regarding X^2, SRMSE, PSI and PHI and we notice better global accuracy than Mob_a and UCL. The last two statistics show quite little differences between the models.

From the map and the goodness-of-fit statistics from Table 1, we conclude that taking into account the crossed urbanization type of the road section in models Mob_b and Mob_c brings a better accuracy level. With respect to these results we decided to build the peak hours model on the basis of Mob_c.

Morning peak hours accessibility model

Morning peak hours assignment

Two models were tested for the morning peak hours. They were built on Mob_c model. The parameters of the speed-flow function had to be defined for each road section. It has been done according to their typology. The demand matrix based on the ESE 2001 census was brought back to a one hour period and assigned on the edges to reach user equilibrium[4].

Two sets of parameters were tested from Mob_c off-peak model. They are named Mob_c1 and Mob_c2. These parameters (Table II) were at first set according to the literature (Barton-Aschman Associates, Inc.; Cambridge Systematics, Inc., 1997) and then adjusted according to their results.

[4] the O/D demand matrix is assigned without taking account of its diagonal

Table II: BPR function parameters values of Mob_c1 and Mob_c2

Category of roads	Speed in km/h	Mob_c 1 alpha	Mob_c 1 bêta	Mob_c2 alpha	Mob_c2 bêta
connector	15	0.50	2.0	0.50	2.0
separated directions	95	0.83	2.7	2.50	2.5
separated directions in urban zones	40	0.83	2.7	2.50	2.5
not seperated directions	70	0.50	2.0	2.50	5.0
not seperated directions in urban zones	30	0.50	2.0	2.50	5.0
motorways in urban zones	100	0.88	9.8	1.50	3.0
motorways	110	0.88	9.8	1.50	3.0

The travel times (from every municipality to Brussels) computed with these parameters are represented on maps (Figure 7).

To facilitate the comparison between off-peak and morning peak models, the travel-time classification scheme is kept across the different maps. On the map of Mob_c2 model, the isochrones are quite narrow compared to the situation of Mob_c1. The travel time in the case of the UCL model in peak-hours have a concentric distribution around Brussels, on the contrary with the models developed in MOBLOC that more favour the Antwerp-Brussels axis. The explanation could rely in the presence of the two highways, A12 and A1, connecting the two cities and therefore better able to distribute the flows towards Brussels during the peak hours than in the cases of other municipalities that are mostly connected by only one highway.

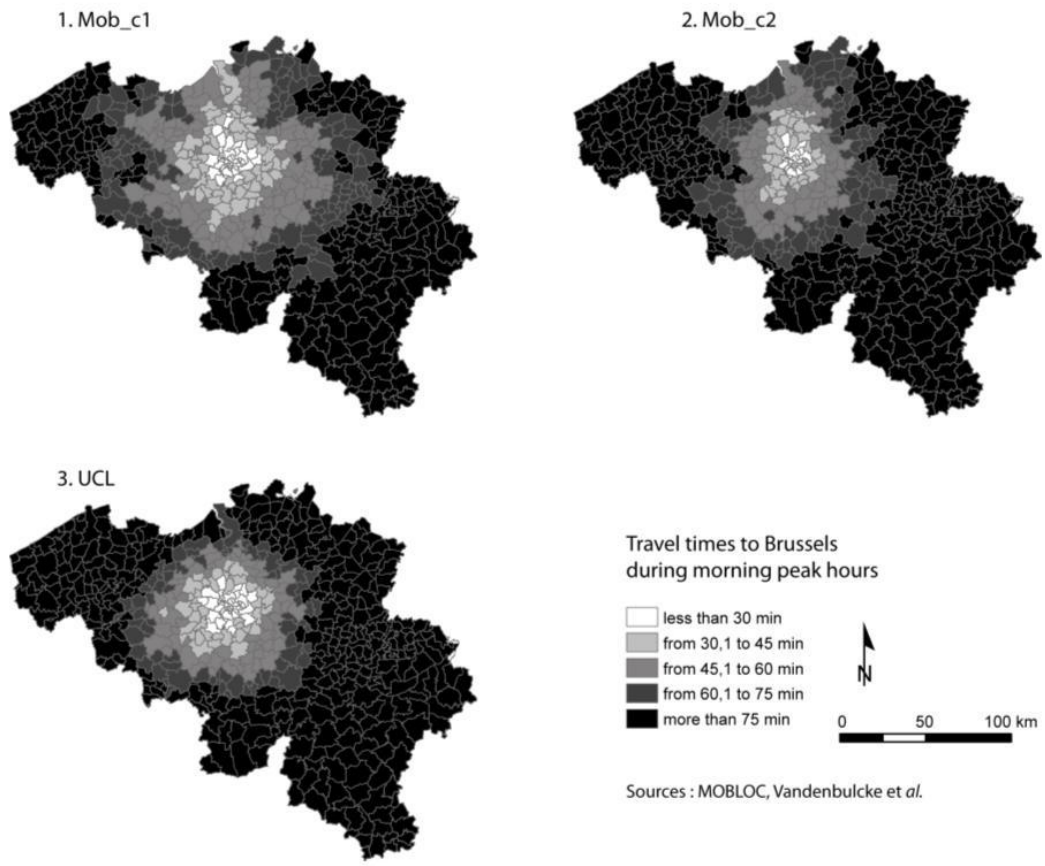

Figure 7: Travel time to Brussels in rush hours for the different models

Goodness-of-fit in peak hours

The results of the statistics, comparing the MOBEL observations in morning peak hours with travel times matrices of the different models, are presented in Table III. This comparison was made on the base of 230 origin/destination couples[5].

Table III: Goodness-of-fit values in morning rush hours

Model	AED	X^2	SRMSE	Abs_PSI	PHI
Mob_c1	0,019	801	0,312	0,234	0,241
Mob_c2	0,024	883	0,376	0,276	0,279
UCL	0,068	872	0,420	0,262	0,272

The three models in Table III show a close magnitude for the different statistics. The best results are exhibited by Mob_c1 model. Therefore this model will be used for the calculation of the accessibility indicators allowing the interaction with the migration model of MOBLOC.

[5] This size is lesser than the one of the off-peak analysis for two reasons. On one hand, the considered period is shorter. On the other hand, the origin/destination couples available in the UCL model don't concern every destination but only those of the municipalities of the three upper classes of Van Hecke urban hierarchy.

Communal accessibility indicators

Now that the travel times matrices of the traffic models have been validated, it is possible to calculate the accessibility indicators needed for the localization model.

Jobs potentially accessible by car during morning peak-hours

We decided to use a gravity-based indicator to take the employment into account. The advantage of this indicator compared to simpler indicators (number of jobs accessible in 30 min for instance) is to account for the negative influence of travel-time on the attraction of the destination as well as for the relative importance of the number of jobs.

We then calculated potential job accessibility for each municipality using the following formula.

$$Ai = \sum Dj.\exp(-\beta.dij)$$

with Ai, the potential accessibility to jobs of the municipality i,
Dj the jobs in the municipality of destination j in year 2004,
β a parameter calibrated according to MOBEL database, and
dij the travel time between municipalities i and j [6].

The value of beta is calibrated according the distributions of travel time values from the MOBEL data set during morning peak hours. Its calculated values is 0.0715 and its Pearson correlation coefficient with cumulative probability of moving is 0.996.

On Figure 8, the potential accessibility to jobs during morning peak hours is represented. The highest potentials are met in Brussels and Antwerp and have important consequences for the potential of municipalities linked by motorways to those major employment centres. A group of municipalities with high potential values surrounds a North-South axis between Antwerp and Brussels. Here the impact of the motorways network is clearly visible. On the contrary, the southern part of the country appears to have a poorer job potential accessibility due its important distance to a major employment centre[7].

[6] For computer time consideration dij is less or equal 60 min.

[7] We should stress out that this methodology doesn't take into account jobs available in bordering countries which could have impact the results especially for the Walloony/Luxembourg border.

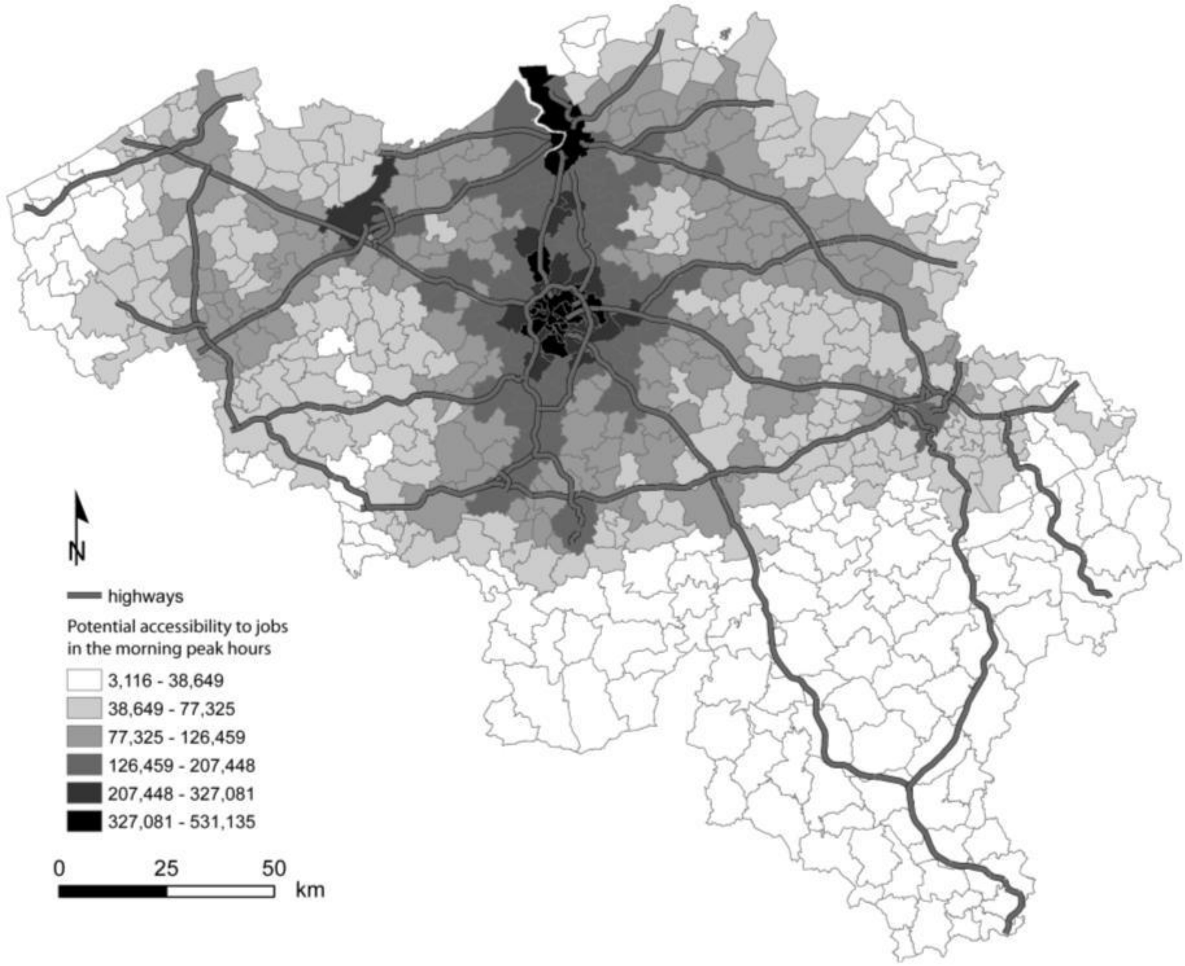

Figure 8: Potential accessibility to jobs during morning peak hours

Accessibility potential to services during off-peak hours

We present here a second accessibility indicator taking into account the services of Belgian cities. These services are supposed to have to be reached occasionally and during off-peak hours. The level of equipment is measured by the equipment score calculated by van Hecke in his work on urban typology. The equipment score encompasses eight functions (health, leisure and sports, communication, public service offices, public service administrations, culture, education and retail) by the mean of several quantitative and qualitative indicators (Van Hecke (1998)).

A gravity measure will put in perspective the weight of nearby equipped cities according to their equipment score, their distance to the municipality of origin. To do so, we decided to build a continuous variable on the basis of this equipment score.

$$A_i = \sum D_j.\exp(-\beta.d_{ij})$$

with A_i potential accessibility to jobs of the municipality i,

D_j the equipment score in the municipality of destination j,

β a parameter calibrated according to MOBEL database, and

d_{ij} the travel time between municipalities i and j with d_{ij} less or equal 60 min.

The value of β is calibrated according the distribution of travel time values from the MOBEL data set in off-peak hours. Its calculated values is 0.0735 and its Pearson correlation coefficient with cumulative probability of moving is 0.998.

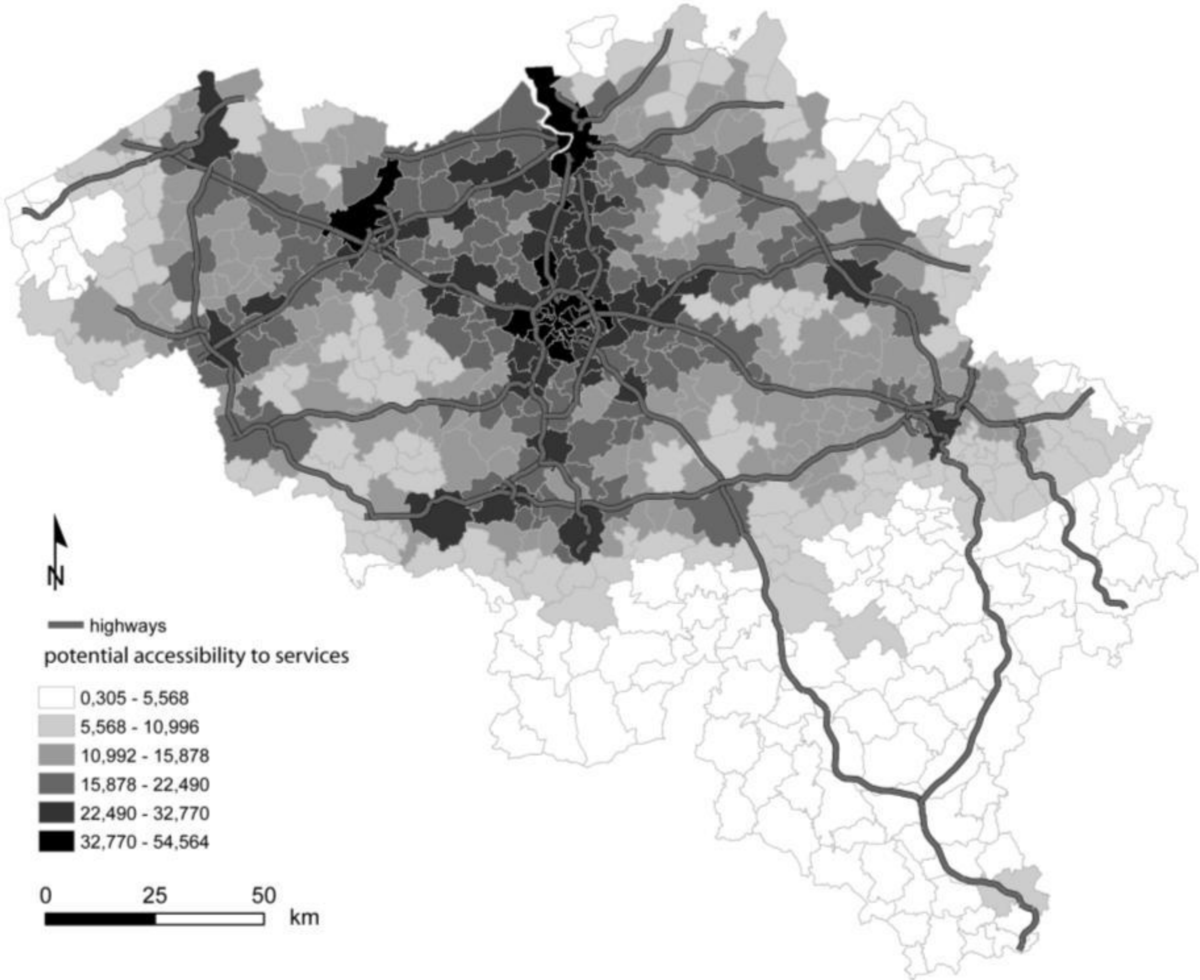

Figure 9: Potential accessibility to services in off-peak hours

This potential is represented on a map (Figure 9). The highest potential values clearly spots on the well-equipped cities such as Brussels, Antwerp, Gent, Liege as well as the regional cities of the hierarchy (Van Hecke (1998)) which also show values in the upper-class. The effect of the motorway network is stronger than for the accessibility potential to jobs and the attractiveness of the cities is visible far along the motorways axis. The southern part of Belgium appears once again as having a weak potential accessibility to services.

The two gravity based accessibility indicators (i.e. peak hours accessibility to jobs and off-peak hours accessibility to services) reveal similar spatial structures. These two indicators should thus influence the localization model in a quite similar way. Finally, the accuracy of estimated travel times seems to be correct for the model coupling and the related simulation.

C. Travel demand model

The modelling of peak hours accessibility needs simulating journeys to work distribution between all origins and destinations (i.e. 589*589 Belgian municipalities). Considering the available data and the accuracy needed for forecasting, the chosen model is a doubly constrained gravity model. Indeed, in our project, such a model as to be coherent with other MOBLOC models providing the margins of the Origin/Destination matrix which justifies the choice of a doubly constrained model. Moreover, the model has to take into account the diagonal and also local effects affecting the distance-decay function and reflecting on one hand the major linguistic dichotomy (Flanders vs Wallonia) and, on the other hand, the strong polarization around Brussels. To achieve these goals a doubly constrained gravity model with regional belonging effect and using mixed distance-decay function (combining exponential and Pareto function) has been developed with MATLAB.

Gravity model

From the knowledge of places of residence of the employed people and the number of jobs per municipality, forming the margins of the flow matrix (Figure 10), it is possible to model the origin/ destination flows for journeys to work. To do so, we implement a so-called gravity model, by analogy with the theory of universal gravitation. This model assumes that the intensity of flows between i and j depends on the respective masses of the spatial units (jobs/employed people) and is inversely proportional to the distance between them (thus regarded as an obstacle to the interaction).

	1	j	n	O_i
1				
i		F_{ij}		
n				
D_j				T

Fij: flow between i and j
Oi: number of employed residents in municipality i
Dj: number of persons working in j
T: total workers/jobs

Figure 10: Data: flow matrix

In order to ensure consistency with other MOBLOC models (including the localization model), it is necessary to respect the margins of the O/D matrix and thus to constrain the marginal totals. The doubly constrained gravity model is then conventionally written as:

$$\widetilde{F_{ij}} = A_i B_j O_i D_j f(d_{i,j}, \beta)$$

where *f* is an exponential or Pareto function or a combination of them; and Ai and Bj are the margin constraints.

It could be formalized as:

$$\begin{cases} \forall i : A_i = \dfrac{1}{\sum_{j \leq n} B_j D_j f(d_{i,j}, \beta)} \\ \forall j : B_j = \dfrac{1}{\sum_{i \leq n} A_i O_i f(d_{i,j}, \beta)} \\ with\ O_i = F_{i.}\ et\ D_j = F_{.j} \\ f(d_{i,j}, \beta) = e^{\beta * d_{i,j}} \text{ or } {d_{i,j}}^{\beta} \text{ or } e^{\beta_1 * d_{i,j}} * {d_{i,j}}^{\beta_2} \end{cases}$$

Besides the respect of margins, the case of Belgium also means taking account of regional belonging and related linguistic specificities (Dujardin, 2001).

Introduction of border effect

It is easy enough to take into account known border in the model by introducing additional parameters. In the Belgian case, where three regions exist, it could be written as:

$$\widetilde{F_{ij}} = A_i B_j O_i D_j f(d_{i,j}, \beta, \gamma)\ with\ f(d_{i,j}, \beta, \gamma) = e^{\beta_1 * d_{i,j}} * {d_{i,j}}^{\beta_2} * {\gamma_1}^{FB} * {\gamma_2}^{FW} * {\gamma_3}^{BW}$$

where FB, FW and BW are binary variables equal to 1 if the flow crosses the border between Flanders and Brussels (respectively Flanders and Wallonia, Brussels and Wallonia) and 0 otherwise.

These barrier parameters measure the eventual brake (or accelerating) of flows from one area to another: γ < 1 indicates a braking and conversely, γ > 1 indicates an intensification of flows.

Model estimating

Given these preliminary choices, the model can then be estimated in two main ways: either by optimization (under constraints) or by a statistical method.

Estimation by optimization

The estimation by optimization has the advantage of not assuming any hypothesis on the flows distribution and thus allows introducing additional explanatory variables. However, it does not provide test statistics.

In this case, the criterion to minimize is the weighted chi-square which is written as:

$$\sum_{i\leq n, j\leq n}\left(\frac{\widehat{F_{ij}}-F_{ij}}{\sqrt{\widehat{F_{ij}}}}\right)^2$$

Calculating A_i and B_j

β being a parameter of braking, it is necessarily negative. It is put = -1 for this first step and by iterations, we calculate

$$\begin{cases} B_j^{(k-1)} = \dfrac{1}{\sum_{i\leq n} A_i^{(k-1)} O_i f\left(d_{i,j}, \beta^{(0)}\right)} \\ A_i^{(k)} = \dfrac{1}{\sum_{j\leq n} B_j^{(k-1)} D_j f\left(d_{i,j}, \beta^{(0)}\right)} \end{cases}$$

The iterative process continues until the norms $\left\|A^{(k)} - A^{(k-1)}\right\|$ and $\left\|B^{(k)} - B^{(k-1)}\right\|$ do not exceed a fixed epsilon.

From these first approximations, β is estimated by minimizing:

$$\sum_{i\leq n, j\leq n}\left(\frac{\widehat{F_{ij}}-F_{ij}}{\sqrt{\widehat{F_{ij}}}}\right)^2.$$

During the minimization, the parameters A_i and B_j are recalculated at each step.

We can calculate an adjustment factor R^2 by comparing the model residuals versus the independence model: This model does not involve any parameter but simply respects the conservation of marginal flows: $\widehat{F_{i,j}^*} = \frac{F_{i.} * F_{.j}}{F_{..}}$

Then, $R^2 = 1 - \dfrac{\sum_{i\leq n, j\leq n}\left(\frac{\widehat{F_{ij}}-F_{ij}}{\sqrt{\widehat{F_{ij}}}}\right)^2}{\sum_{i\leq n, j\leq n}\left(\frac{\widehat{F_{i,j}^*}-F_{ij}}{\sqrt{\widehat{F_{i,j}^*}}}\right)^2}$

Statistical estimation

The equations of the model in its simplest form, can be rewritten as:

$ln(\widehat{F_{ij}}) = \ln(A_i) + \ln(B_j) + \ln(F_{.j}) + \ln(F_{i.}) + \beta\ln(d_{i,j})$ which suggests a linear model. However, this method is not appropriate for several reasons: among them, we include the non-compliance with the conservation of margins (and therefore of total flows), failure to take into account the zero flows (since we minimize the least squares: $\min\left(\sum\left(\ln(F_{ij}) - \ln(\widehat{F_{ij}})\right)^2\right)$, or the assumption of normality of errors that is not verified.

Finally, F_{ij} are measures of counts (ie integers) and this suggests a continuous statistical distribution such as that of Poisson. Instead of choosing a linear or log-linear regression, it is preferable to turn to a Poisson regression (Flowerdew R., Aitkin M., 1982).

Model assessment

The choice of the *ad hoc* model is based on the comparison of several models. In our case, four models were tested, varying the type of the distance decay and the possible addition of a barrier/regional belonging effect:

1. Pareto model: $\ln(\widehat{F_{ij}}) = \ln(A_i) + \ln(B_j) + \ln(F_{.j}) + \ln(F_{i.}) + \beta\ln(d_{i,j})$
2. Exponential model: $\ln(\widehat{F_{ij}}) = \ln(A_i) + \ln(B_j) + \ln(F_{.j}) + \ln(F_{i.}) + \beta d_{i,j}$
3. Mixed model: $\ln(\widehat{F_{ij}}) = \ln(A_i) + \ln(B_j) + \ln(F_{.j}) + \ln(F_{i.}) + \beta_1 d_{i,j} + \beta_2\ln(d_{i,j})$
4. Mixed model with barrier effect:

$$\ln(\widehat{F_{ij}}) = \ln(A_i) + \ln(B_j) + \ln(F_{.j}) + \ln(F_{i.}) + \beta_1 d_{i,j} + \beta_2\ln(d_{i,j}) + FB\ln(\gamma_1) + FW\ln(\gamma_2) + WB\ln(\gamma_3)$$

To compare these four models, three goodness of fit statistics were selected, namely:

1. Deviance: $\frac{\sum|F_{ij} - \widehat{F_{ij}}|}{\sum F_{ij}}$

2. X²: $\sum\left(\frac{F_{ij} - \widehat{F_{ij}}}{\sqrt{\widehat{F_{ij}}}}\right)^2$

3. AED (absolute entropy difference):

$|H_p - H_0|$ with $H_p = -\sum p_{ij}ln(p_{ij})$, $H_0 = -\sum q_{ij}ln(q_{ij})$, $p_{ij} = \frac{\widehat{F_{ij}}}{F_{..}}$, $q_{ij} = \frac{F_{ij}}{F_{..}}$

Additionally, two modelling procedure were used and compared for each model: the GENMOD procedure from SAS software, on the one hand, and DAL-POULAIN procedure, developed for the project MOBLOC by L. Dal (GéDAP, UCL), on the other. These two procedures differ in the choice of estimators (optimization procedure for DAL-POULAIN and statistical method for GENMOD, see Table IV)

Table IV: Goodness of fit statistics and parameters

		GENMOD (SAS)	MOBLOC (DAL-POULAIN)
Pareto model	X^2	2,452,802.736	2,172,747.563
	AED	0.224	0.622
	Deviance	1,902,006.870	2,052,654.522
	Scale	2.663	2.506
	β	-2.147	-1.959
Exponential model	X^2	5.9699698E14	16,165,603.353
	AED	0.382	1.545
	Deviance	3,242,955.556	6,345,416.024
	Scale	41553.670	6.837
	β	-0.128	-0.057
Mixed model	X^2	5,218,387.423	1,816,077.712
	AED	0.167	0.553
	Deviance	1,424,267.213	1,675,262.105
	Scale	3.885	2.292
	β	[-1.5464; -0.0364]	[-1.6427; -0.0169]
Mixed I model with barrier	X^2	4,283,521.443	1,576,916.186
	AED	0.145	0.477
	Deviance	1,232,247.132	1,434,702.069
	Scale	3.519	2.136
	β	[-1.5659; -0.0317; 0.0341; -1.4617; 0.7864]	[-1.6643 ; -0.0143 ; 1.0153 ; 0.3391; 1.9152]

Considering the results of this comparison, we chose to use the mixed model (to estimate correctly both the short and long distance flows) with the regional belonging effect (to take account of the strong polarization of Brussels and brakes between Flanders and Wallonia). We also retain the DAL-POULAIN procedure insofar as it does not involve any assumption on the distribution of flows.

Residual analysis

At the end of this stage of modelling the gravity flow distribution, we obtain an acceptable fit with the reference matrix (ESE 2001). However, some residuals exist, notably through the overestimation of long-distance intercity flows (Figure 11).

The residuals are calculated as follow:

$$R_{ij} = \frac{\left(F_{ij} - \widehat{F_{ij}}\right)}{\sqrt{\widehat{F_{ij}}}}$$

The main residuals of our model consist in the overestimating of long distance trips, especially to Brussels Region. Furthermore, in the Brussels Region some underestimating of intra-regional flows appears, highlighting thus the local employment dynamics in the capital.

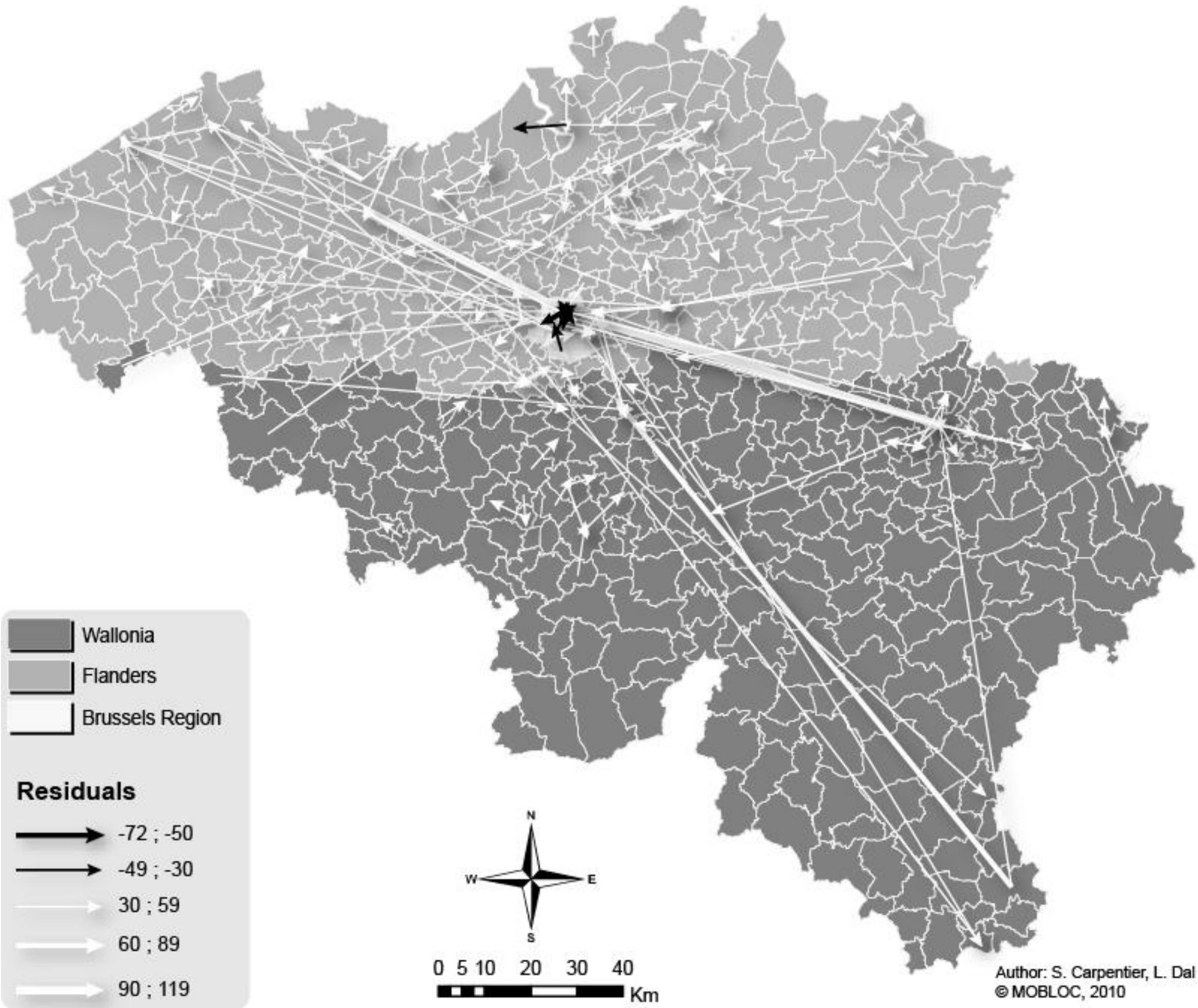

Figure 11: Inter-municipalities residuals

Considering, in a second step, the intra-urban residuals (Figure 12), it appears that while the global fitting appears to be quite good, local specificities lead to some under and over estimates of intra-urban journeys to work. At a regional scale, Wallonian intra-municipalities flows are underestimated while Flanders and Brussels region trips are more often overestimated. Moreover, the major under and over-estimatings mainly affect the biggest cities.

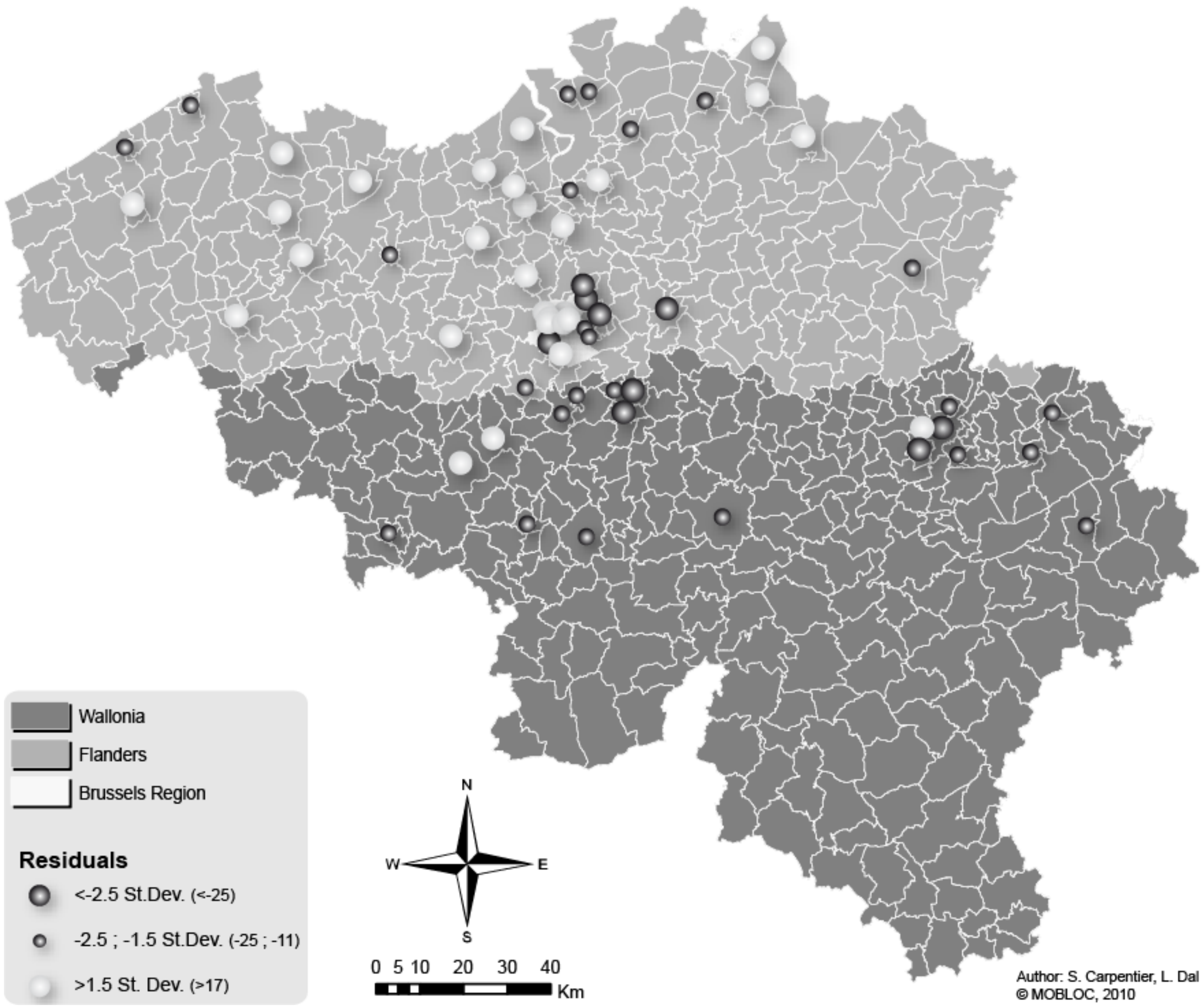

Figure 12: Intra-municipalities residuals

Conclusion

The gravity modelling of journeys to work distribution in the Belgian case implies to take additional parameters into account, namely the regional belonging effect. At the end of this step, a quite well fitted model was built in order to provide the best accuracy possible regarding observed data. It seems now acceptable to use such a model for the forecasting of daily trips in order to provide an estimation of flows for the accessibility model.

However, some methodological improvements could be foreseen in order to enhance the accuracy of the model. The challenge is then to better model the long distance trips while keeping good fit for the shorter ones. It could be achieved by taking into account additional parameters such as wages level or unemployment rates depending on the availability of such data at the municipality level.

D. Propensity to move model

In this section, let us first expose recent trends of residential migration in Belgium during the past decades and present a set of explanatory variables to study this event. In other words through this point we will lay the emphasis on the basic but crucial elements for explaining and modelling residential migration. Then we will describe our researches on modelling the propensity to move throughout different steps which lead to the building of a propensity to move model. Finally, main findings will be highlighted.

Before going deeper in methodologies it should be mentioned that we could access a database (*National Register* - **NR**) describing "states" at successive 1st of January at an individual level from 2001 till 2006. That means that we cannot observe any real events but the ones deducted from transitions between two consecutive first of January.

A second source of data is the national census (Socio-Economic Survey – **ESE 2001**) achieved in October 2001 giving only one shot information. This remark about the lack of series data is important for the prospective step as we will see later.

Moreover, it must be clearly defined that by residential migration we only consider migrations occurring between two municipalities since the data does not allow investigating migration within a same municipality.

Residential migration: a concise overview

First of all, we studied the evolution of residential migrations between municipalities for the last decades. Figures highlight two main trends. On one hand residential migration is rising. Since 1988, figures in residential migrations have strongly increased, growing from 379,000 in 1988 to 481,000 in 2004 and even 532,000 in 2009 (Statbel, SPF Economie, PME, Classes moyennes et Energie, Direction générale Statistique et Information économique). The calculation of an index which corrects the effects of differences in age structure (age structure of the Belgian population has been getting older for the last decades), shows that the propensity to move has increased at all ages from 0 to 65.

On the other hand we observe that the periurbanization process is ongoing and getting larger on the Belgian territory whilst young households are looking for cheaper housing and building plot. Different factors can be exposed to explain these trends. Among them, we should mention more frequent household transformations that increase residential mobility and the "deadlock" of ancient periurban areas that force in particular the younger ones who leave their parent's home to go farther and farther away (Eggerickx et al., 2008).

Pre-selection of variables for the propensity model

Let us recall that given the available data, we defined that a residential migration occurs for an individual when his/her residential place (municipality) is different between two successive 1st of January. So migration inside the same municipality or multiple migrations in one calendar year are not detectable in the datasets since we only have the records of the municipalities for the place of residence at each 1st of January.
The measured "risk" reflects thus the fact that the residential municipality at a 1st of January is different from one year to the next one. This kind of "event" is quite rare since it only concerns 4% of the Belgian population each year.

The explanatory variables were selected according to a literature review (e.g. Debrand and Taffin, 2005; or Henley, 1998) and to their availability through the National Register or the Socio-Economic Survey of 2001. We are dealing with individual characteristics as well as housing and area of residence characteristics.

As *individual characteristics* we selected age, gender, nationality, household type, number of individuals per household, position in the household (i.e. the link with the head of household), the highest education level successfully completed, the activity status and type of activity, and if a migration occurred the year before.

As *housing characteristics* we considered the type of housing (house, flat…) and housing tenure type.

As regards *area of residence characteristic*, we took into account the urban/rural profile of the municipality (downtown, rest of the city, old and recent periurban areas, rural areas) built by Van der Haegen et al. (1996).

In order to explore what could be the most relevant covariates we built some contingency tables with these explanatory variables and the binary variable "Has an individual moved? Yes or no". We will not expose the results in details but just indicate which categories of people have the highest (or the lowest) propensity to move for each covariate we selected (Table V).

Table V: Categories having the highest or lowest propensity to move regarding available explanatory variables

Variables	Categories having the *highest* propensity to move	Categories having the *lowest* propensity to move
Age	19-29	55-74
Gender	Men	Women
Nationality	Congolese RDC	Turk
Household type	Cohabitants without children	Married couple without children
Number of people per household	1 individual	4 individuals
Position in the household	Other (than head, partner or child)	Partner
Previous migration (the year before)	Migrated the year before	Did not migrate the year before
Place of residence	Agglomeration (downtown excluded)	Small cities
Highest education level successfully completed	High education	Primary or no instruction
Activity status	People looking for a job	Retired
Housing tenure type	Renters in the private sector	Owners
Type of housing	Apartments	4 front houses

On the basis of the previous observations and of practical considerations we selected only a group of these variables. We could indeed not keep the whole set for the following reasons:

- On one hand, we recommended the propensity to move model to focus mainly on individual characteristics; so it excluded *de facto* the urban/rural profile of the area of residence and the type of housing.
- On the other hand, we found some limits provided by the data for the upcoming steps. Indeed, in the second phase of this project, we planned to achieve forecast for municipalities or group of municipalities. So we already need to know which variables will be available for extrapolation. The problem is that extrapolation needs individual data at at least two dates, and yet for some explanatory variables such as education level, occupation status, or housing tenure type, we know that we only can have data for a single time i.e. October 2001, from the 2001 Socioeconomic Survey.

We thus had to consider which are the most essential variables from the 2001 Socioeconomic Survey to be retained for the model. Here are explained the most crucial choices:

Education level versus activity status

Since the highest successfully completed education level and the activity status are strongly associated, we decided to keep only one from these two variables. We finally chose the education level for two main reasons:

Firstly, the way activity status and activity types have been built in the ESE2001 survey is not satisfying: the variable "activity status" contains modalities of responses that are not exclusive: one can be student and at the same time in activity; so we wondered how people answered choosing one rather than the other category. As regards "activity types", the main critic is that in the 2001 census a distinction is made between workers and employees in the private sector, but not in the public sector so that we cannot compare the same positions in the two sectors.
Secondly, education level is a "capital" that one cannot lose contrary to activity status; hence, it is a variable more stable in time and should be easier to project in a near future. In comparison, occupation status or activity types might change from one year to the other according to the conjuncture or socioeconomic policy. Therefore, these variables are more volatile and then would be more difficult to manage for forecast.
In this context and given that education level is generally as good as activity status as a proxy for the socioeconomic profile of individual we preferred selecting education level.

Housing tenure type

We have also kept in the set of variables the housing tenure type because it is rather essential to analyze residential migration: this variable largely discriminates people in migration depending on whether they are tenant or owner, or whether they are tenant in public or private market. For instance the ratio of migrating people is more than 3 times higher for tenants in a private dwelling than for owners.
Extracted from the Socio-Economic Survey of 2001, housing tenure type at individual level is only available at one date (October 2001). However trends provided by the National Institute of Statistics shows that this is stable along time, especially in Belgium where ownership is traditionally high: between 1991 and 2001, the percentage of owners increased from 65.4 to 65.9%, and forecast concerning potential impact of an economic uncertainty on becoming owners do not seem significant on the Belgian market (Vanneste et al., 2007). In this context, it appears that this variable would be stable during time, and not too complex to forecast in a near future.
However an additional disadvantage for the housing tenure type is that it is provided at the household level. That means that if the head of household answers that he/she owns his/her housing then each household's member is as well considered as owner. This makes problems especially if we are interested in leaving parents-household phenomenon because if a young adult lives with his parents' owners in 2001, he is then also considered as owner. As we have only data at one time, once he leaves his parents' home he will still be considered as owner whatever the new housing. Despite these reluctances, we have included this essential variable and tried to deal with its drawbacks.

Here are described the currently pre-selected variables and their modalities:

- *Age* (years over January, the 1st) - 6 modalities: 0-18, 19-29, 30-44, 45-54, 55-74, 75 and more.
- *Gender* - 2 modalities: Male, Female
- *Nationality* - 8 modalities: Belgian, Congolese, Moroccan, Turk, citizens from the 12 new EU members, citizens from the 15th former EU members and citizens from West European countries non EU members (Switzerland, Norway, Iceland,..), rest of East Europe (Balkans, Byelorussia, Russia, Ukraine), rest of the world.
- *Household Types* - 8 modalities: Married couple without children, Married couple with children, cohabiting couple without children, cohabiting couple with children, single-household, one-parent family, collectives and households "others" (with more than one family kernel).
- *Position in the household* – 4 modalities: head of household, spouse (wife, husband, partner), child, others. The position is determined in relation to the head of household.
- *Number of people per household* – 5 or 6 modalities: 1, 2, 3, 4 or more, or 1, 2, 3, 4, 5 or more. The total amount of modalities depends on household types and except for single-household and couple without children, the last class is open.
- *Highest education level successfully completed* – 5 modalities: no instruction or primary, secondary school (inferior), secondary school (superior), higher education, no answer.
- *Housing tenure type* - 5 modalities: owner, private tenant, social tenant, tenant of a free housing, no answer.
- *Residential migration during previous year* – 2 modalities: yes, no.

Methodological choices to build the model

The technique used is a binary logistic regression. The original idea was to use discrete choice model for the residential migration model. For this purpose, it was planned to use BIOGEME (software developed by Michel Bierlaire from the EPFL, Lausanne). Since we faced some problems with the size of the database and given that, for the propensity model (first part of the migration model) we had a binary dependant variable and therefore considered to choose a logit function, we decided to make a binary logistic regression rather than a discrete choice model. Nevertheless discrete choice modelling remains the technique we used for the localization model because it offers more possibilities when there are a lot of alternatives.

There are different ways to conduct a logistic regression. At a time, our ambition for the propensity model was to deal with panel data methods. Using a five-year period database, we first explored how to build a model on the basis of the years 2001 to 2005. For this purpose, we built a person-year database: it means five observations per individual. But at the end, this method had some drawbacks for our case so that we finally opted for a simple binary regression model.

We had two main constraints when building this model:

Annual step for the migration and for the planned projections

When we started to build the model for propensity to move, the question arose about how to express the temporal unit of the dependent variable (and explanatory variables). As we had information at each 1st of January it seemed obvious that the most useful and simplest way to study migrations would be based on an annual step, which would estimate residential migration within one year, i.e. between two consecutive 1st of January.

Take advantage of the dynamic dimension: past and anticipation effects.

As we already mentioned, the available database is a sequence of states for the overall Belgian population during the 5-year period 2001-2005.
By convention in demography, when we refer to a "period", it means a period in years of time. For instance, the 2001-2005 period is a 5-year period in year of time; i.e. from 1st January 2001 till 31st December 2005. However, in this section and in particular here below we also use the terminology of the "transition period" which refers to the year between 2 firsts of January, e.g. the transition period 2002-2003. This results from the use of observed differences of states between two consecutive 1st of January.
Data from the National Register provide the demographic profile of each Belgian people each year (while data from the 2001 Socioeconomic Survey supply some other variables, but only for October 2001). Therefore we want to take advantage of the dynamic dimension conveyed by this database.
Indeed, it seems obvious that past information about the life of an individual can help to explain his/her migration behaviour. For example a change in the household's structure or size (which can reflect events such as separations, births or unions) can increase (or decrease) the probability for change in residential municipality. At the same time we also want to test effects from anticipation.

For instance, testing whether an observed migration for a couple in the transition period Y to Y+1 may be explained by the fact that this couple has a baby in the transition period Y+1 to Y+2. They could have anticipated the birth by looking for a bigger housing the year before.

While building the model which has to explain migration between years Y and Y+1, we looked for a suitable way to include this dynamic dimension. Rather than simply use the state variables at the first of January of the years Y-2, Y-1, Y, Y+1 and so on, we decided to use the state variables only at time Y and to create transition variables which describe variation in the state variables between two successive 1st of January.

These transition variables have been created for variables from the National Register i.e. household structure and size, position in the household and nationality. But for anticipation we have only considered household transitions (variables related to household). Regarding the Socio-Economic Survey of 2001 variables, as they are only available in October 2001, it was not possible to create transition variables.

These transformed variables are helpful to add in the model a "temporal depth" of one or more years back or forward. From a statistical point of view, transition variables act as a substitute of a sequence of temporal successive states and have the great advantage to handle multicolinearities or autocorrelation, which could occur when using these states as explanatory variables in the model. Moreover, they make easier the interpretation of the results since transition variables implicitly suppose an event (although not observed).

At a last point, it should be said that since dynamic database allows catching unobserved heterogeneity, we add in the model an autoregressive dimension which is the dependent variable at the previous year (did an individual move the year before?).

Transformation of variables

Before creating transition variables, we studied and applied specific treatments for intrinsically highly associated variables such as household types and number of people per household. Indeed, e.g. single households have all one person living, while married or cohabitant couples with no children are all composed of two members. In other cases, the association between household type and household size is not univocal so that we wanted to keep information from both variables. As a consequence, we created synthetic variables based on these two variables describing the household type and the number of people living in it.

Another variable transformation was the setting of transition variables. As already described, such a variable substitutes for 2 successive state variables. For each individual, if these two successive state variables are identical then the value of the transition variable is "no change". If not, we first concatenated the two successive states and then we grouped these categories in less numerous modalities that could explain the migration behaviour.

For instance naturalization variable 0102 was built from two variables: nationality in 2001 (more precisely at 1st January) and in 2002. If this nationality has changed we distinguished two cases: if it was a nationality change towards the Belgian or the UE citizenship, then we consider it as a change (i.e. naturalization); but if it was a change towards another extra-European nationality, then we do not. Here, becoming Belgian or UE citizenship has been only seen as a change because becoming Belgian or a European citizen might encourage a long stay in Belgium and the adoption of local residential migration behaviour.

In the same way, we redefined a lot of transitions, especially for the synthetic variable (household type and size).

Sample design

The question to use a sample rather than the entire base arose when we faced some problems with the size of the database and the calculation time. As there are more than 10 million citizens in Belgium, we have about the same number of observations, which is more than necessary. We may indeed reasonably reduce the size of the sample without losing quality of information for the model.

The sample design was also a constraint and needed technical adaptations. Since the model deals with residential migration within the country we worked with an enclosed population which was people who were residents in Belgium at the considered years.

Furthermore, including the "temporal depth" - of one or two years in the past and one year after the reference year - implied to add the condition of being resident in Belgium during at least 3 consecutive years (or 4 consecutive 1st of January).

Furthermore as residential migration is a rare event (4% per year), we wanted to increase its frequency in order to improve the quality of the coefficients of regression (Allison, (1999). This is the reason why we oversampled the migrants. As a consequence the sample is based on a stratified random sampling in the overall Belgian population. Stratification has been done according to the dependent variable: for one year observed, we selected all people having changed from municipalities and, for the non-migrants, we realized a simple random withdraw. The total number of drawn individuals arises 2 millions and the percentage of annual migrants comes around 15%. It should be mentioned that we realized several sampling depending on the years of the studied migration. If we study the migration between the firsts of January of the years Y and Y+1, then we use the sample in which migrants in the transition period from Y to Y+1 are oversampled.

Finally from each sample, we created two subsamples:

- the first one is composed of 70% of the individuals and is used for the model calibration,
- the second one with the rest 30% to test and validate the models.

The propensity model is based at the individual level. If we agree that the decision to migrate can be individual, it could also be the decision of the household, or at least of the couple.
At first we wanted to integrate the household and individual units in order to take into account both units of decision. However we realized that drawing a sample on household units with a temporal depth was hardly impossible because the numerous household transformations do not allow the follow-up of every individual in the sample. Indeed, between two consecutive years one household can be divided into two households; and one individual can change from a household to another. It is thus difficult to build a model with households as units and to study their migration behaviour with a temporal depth. Hence, we chose to limit the propensity to move model at individual level. However we included explanatory variables related to household characteristics and positions in the household that could help to take indirectly this second unit into account.

Model(s)

Given that the available data covers the five-year period from begin 2001 to end 2005 (1st January 2001… 2006), we finally concentrated on two transition periods: 2002-2003 and 2003-2004. The reasons for this choice are:

- First of all, since variables from the Socio-Economic Survey of 2001 (education level, occupation status, housing tenure type …) are only available at that time (in our individual database), the closest data from the National Register (household structure, nationality …) are thus those from the first of January 2002; that is why we first decide to study 2002-2003.

- But we also wanted to introduce, as explanatory variable, transformations in the household structure occurred in the past of the individuals (such as marriage, divorce, birth …). Unfortunately, the available data from the National Register only covers 2001-2005, so the only previous period that we could use to explain the period 2002-2003 was 2001-2002. In order to test the improvement of the quality of the model by adding older information, we decided to also study the transition period 2003-2004. We have to notice that the data from the Socio-Economic Survey of 2001 are less contemporary and so less relevant in the model for 2003-2004 than in the one for 2002-2003; that is why we have not chosen, for example, 2005-2006. If we want to estimate the propensity to move between 2005 and 2006, information from 2001 start to be out of date. So we chose to concentrate on migration between 2002 and 2003 and between 2003 and 2004. This implied that we fixed the maximum "temporal depth" at two years back because we had no information before 2001 from the National Register.

- Finally, the purpose (of studying two different periods) was also to determine if the results (parameters) from the models for 2002-2003 and 2003-2004 were comparable in time and led to the same interpretations. It was actually important that the model would be independent from the transition period it is based on, since it had to be used to predict the migration behaviour in the future (population forecasting).

To sum up, we tested different sets of variables in order to:

- compare migration model for two different transition periods : 2002-2003 and 2003-2004,
- test the effects/advantages of adding older (delay) or future (anticipation) information.

The Table VI below summarizes the five models tested and the variables selected (with their source) in each case:

Table VI: Explanatory variables of the tested models for the propensity to move

Dependant Variable	Models	Explanatory Variables (with sources)
Migration between 2002 and 2003	Model 1	Age in 2002 (NR)
		Gender (NR)
		Nationality in 2002 (NR)
		Type and size of the household in 2002 (NR)
		Evolution of the type of the household between 2001 and 2002 (NR)
		Evolution of the type of the household between 2002 and 2003 (NR)
		Link with the household head in 2002 (NR)
		Evolution of the link with the household head between 2001 and 2002 (NR)
		Evolution of the link with the household head between 2002 and 2003 (NR)
		Migration between 2001 and 2002 (NR)
		Education level successfully completed (ESE 2001)
		Housing tenure type (ESE 2001)
	Model 2	Age in 2002 (NR)
		Gender (NR)
		Nationality in 2002 (NR)
		Type and size of the household in 2002 (NR)
		Evolution of the type of the household between 2001 and 2002 (NR)
		Evolution of the type of the household between 2002 and 2003 (NR)
		Evolution of the type of the household between 2003 and 2004 (NR)
		Link with the household head in 2002 (NR)
		Evolution of the link with the household head between 2001 and 2002 (NR)
		Evolution of the link with the household head between 2002 and 2003 (NR)
		Evolution of the link with the household head between 2003 and 2004 (NR)
		Migration between 2001 and 2002 (NR)
		Education level successfully completed (ESE 2001)
		Housing tenure type (ESE 2001)

Dependant Variable	Models	Explanatory Variables (with sources)
Migration between 2003 and 2004	Model 3	Age in 2003 (NR)
		Gender (NR)
		Nationality in 2003 (NR)
		Naturalization between 2002 and 2003 (NR)
		Naturalization between 2003 and 2004 (NR)
		Type and size of the household in 2003 (NR)
		Evolution of the type of the household between 2002 and 2003 (NR)
		Evolution of the type of the household between 2003 and 2004 (NR)
		Link with the household head in 2003 (NR)
		Evolution of the link with the household head between 2002 and 2003 (NR)
		Evolution of the link with the household head between 2003 and 2004 (NR)
		Migration between 2002 and 2003 (NR)
		Education level successfully completed (ESE 2001)
		Housing tenure type (ESE 2001)
	Model 4	Age in 2003 (NR)
		Gender (NR)
		Nationality in 2003 (NR)
		Naturalization between 2001 and 2002 (NR)
		Naturalization between 2002 and 2003 (NR)
		Type and size of the household in 2003 (NR)
		Evolution of the type of the household between 2001 and 2002 (NR)
		Evolution of the type of the household between 2002 and 2003 (NR)
		Evolution of the type of the household between 2003 and 2004 (NR)
		Link with the household head in 2003 (NR)
		Evolution of the link with the household head between 2001 and 2002 (NR)
		Evolution of the link with the household head between 2002 and 2003 (NR)
		Evolution of the link with the household head between 2003 and 2004 (NR)
		Migration between 2001 and 2002 (NR)
		Migration between 2002 and 2003 (NR)
		Education level successfully completed (ESE 2001)
		Housing tenure type (ESE 2001)
	Model 5	Age in 2003 (NR)
		Gender (NR)
		Nationality in 2003 (NR)
		Naturalization between 2001 and 2002 (NR)
		Naturalization between 2002 and 2003 (NR)
		Naturalization between 2003 and 2004 (NR)
		Type and size of the household in 2003 (NR)
		Evolution of the type of the household between 2001 and 2002 (NR)
		Evolution of the type of the household between 2002 and 2003 (NR)
		Evolution of the type of the household between 2003 and 2004 (NR)
		Evolution of the type of the household between 2004 and 2005 (NR)
		Link with the household head in 2003 (NR)
		Evolution of the link with the household head between 2001 and 2002 (NR)
		Evolution of the link with the household head between 2002 and 2003 (NR)
		Evolution of the link with the household head between 2003 and 2004 (NR)
		Evolution of the link with the household head between 2004 and 2005 (NR)
		Migration between 2001 and 2002 (NR)
		Migration between 2002 and 2003 (NR)
		Education level successfully completed (ESE 2001)
		Housing tenure type (ESE 2001)

We used the stepwise procedure (Allison, 1999) selecting one by one the variables which are significantly related to the explanatory variable conditionally to the previously entered variables.

This procedure can also eventually remove one of these covariates thereafter, if it has lost its significance due to the addition of another variable.

In our models, all of the explanatory variables removed from the regression because of lack of significance concerned naturalization:

- naturalizations between 2001 and 2002 and between 2002 and 2003 for models 1 and 2;
- naturalization between 2003 and 2004 for models 4 and 5;

Then we have compared these models thanks to some global criteria such as AIC and log likelihood. We present here below (Table VII) the different likelihood ratios of our models:

Table VII: Tests on the likelihood ratio of the 5 tested models

Tests on the Likelihood Ratios			
	Khi-2	***ddl***	***Pr > Khi-2***
Model 1	502377,9	215	< 0,0001
Model 2	505456,7	301	< 0,0001
Model 3	511072,5	219	< 0,0001
Model 4	511247,3	306	< 0,0001
Model 5	506794,7	394	< 0,0001

The differences between the tested models are not large. From these data we can assess that all our models have a good calibration: indeed, all khi-2 tests on likelihood ratios are extremely significant.

Another way to assess models performance relies on the discrimination: does the model well predict the behaviour of the individuals? Is somebody moving also predicted as a mover? Does the model well separate the migrants from the non-migrants? If we have a look at the Figure 13 here below, we can see that principal diagnostics about discrimination show little amelioration between models 1 and 2 and from models 3 to 5. All four statistics have to present high scores if the models are well fitted.

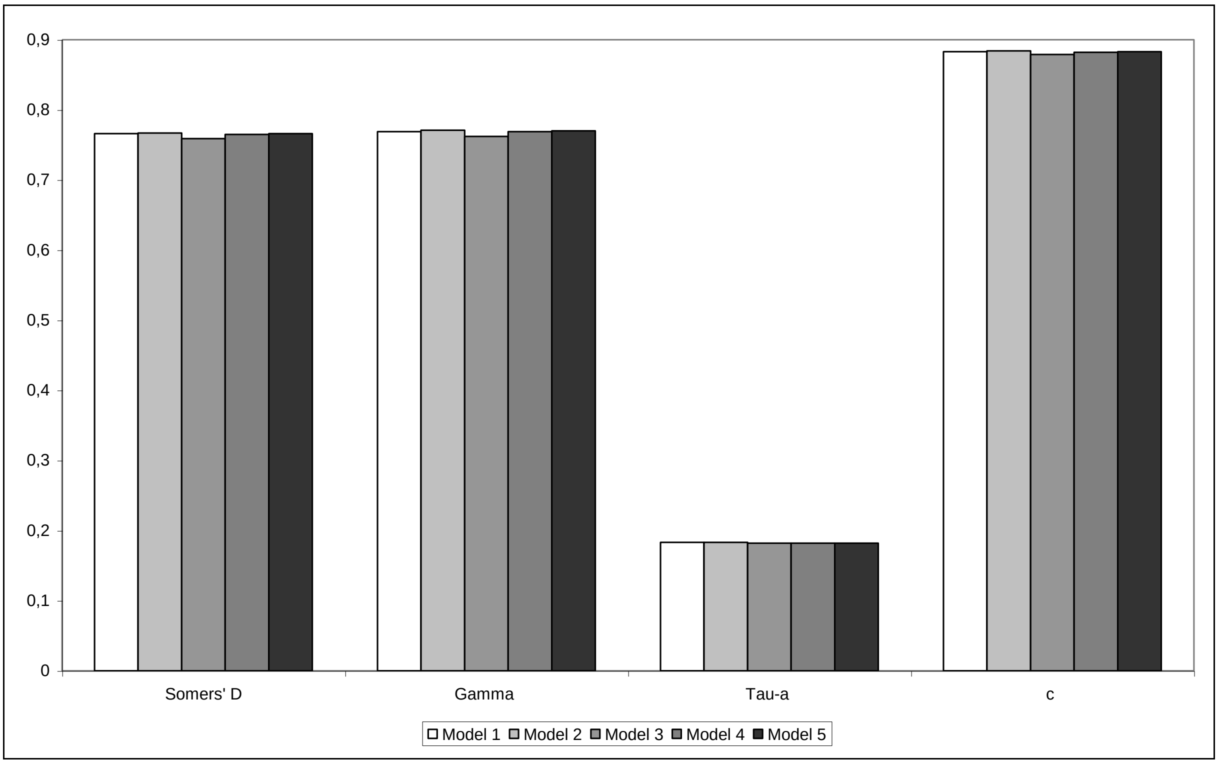

Figure 13: Comparison of the 5 models of propensity to move according to various diagnostics (regarding to the predictive power)

The fact that model 3 seems to be worse than the first 2 models (regarding the predictive powers) can be explained by the loss of relevance of covariates coming from the census.

The statistic *c* is an approximation of the area under the ROC Curve which also provides a measure of discrimination. As a general rule, when this area is between 0.8 and 0.9 one can consider the model have an excellent discrimination (Hosmer and Lemeshow, 2000). In our case values of c are close to 0.9 which means our models are really discriminatory.

As a conclusion about the models, we remark that the 5 models are really good. As the first one seems to be already more than satisfactory, we decided that this one could be used in the next steps of our project.

We give here below (Table VIII) the four most explanatory covariates of this first model (in a decreasing order) as they also appear in the following table for the first model:

- evolution of the type and size of the household between Y and Y+1
- housing tenure type
- evolution of the link with the household head between Y and Y+1
- age in Y (six classes)

Table VIII: p-values of the explanatory variables of the propensity to move model

Are the Covariates of Model 1 Significant? (Migration between 2002 and 2003)			
Covariates	***ddl***	***Khi-2***	***pvalue***
Evolution of the type of the household between 2002 and 2003	74	59530,5	<,0001
Housing tenure type	4	48683,1	<,0001
Evolution of the link with the household head between 2002 and 2003	12	32138,8	<,0001
Age in 2002	5	12677,9	<,0001
Type and size of the household in 2002	17	4079,0	<,0001
Migration between 2001 and 2002	1	3563,2	<,0001
Evolution of the type of the household between 2001 and 2002	74	2946,7	<,0001
Link with the household head in 2002	3	2458,2	<,0001
Highest education level successfully completed	5	2384,8	<,0001
Evolution of the link with the household head between 2001 and 2002	12	590,0	<,0001
Nationality in 2002	7	218,0	<,0001
Gender	1	34,2	<,0001

The most significant variable is clearly the simultaneous transition of the household structure (type and size). Compared to the reference category (no change in the household structure between the two consecutive years), all the odd ratios are positive which means that any change in the household structure is correlated with an increasing probability to move (everything else being equal). The importance of this increase depends on the household structure change. Amongst the most significant, let us point out the people whose household structure were "married couple with children" and became "unmarried couple with/without children", "one-parent family" or "isolate". They have from 20 to 30 times more likelihood to move than people whose household structure did not change.

Let us remember that family events are univocal in the data. For example, such people could be a child who left parents' home, or member of a couple which split up, or etc.

This can be related to another significant variable: the change of relation to the household head. For example, from "being a child" status to a "head" or "spouse" status (or inversely) is related to a high increase of the probability to move (between 11 and 16 times more than in the case of no change).

The second most significant variable for predicting the propensity to migrate is the housing tenure type. Tenants in the private sector are 4 times more likely to move than owners.

Age class also appears amongst the four most significant variables. The less residentially mobile are 75 years old and more while the most mobile are the youngest (less than 18 years). The probability to move decreases with age (everything else being equal).

All these results underline that the propensity to move is linked to the life course of individuals, and more particularly their family trajectories. These are well observable through the age and transition of the household structure covariates. To sum up, the transitions leading to move are break-up situations, new family-units compositions and leaving parents' home. One can observe that the more stable situation concerns people who are in a married couple (with or without children) ; this situation is often associated with an owner status, which is another factor to stay in the same municipality of residence. In other words, the likely evolutions of family situations marked by the rise of less stable households (cohabitation situations, one-parent family…) will still generate higher propensity to move rates in the coming years.

E. Localization model

Preliminary analysis

Contrary to the propensity to move model (where we only focus on individual characteristics to determine if people are going to leave their places of residence), we considered also municipalities' attributes to model the choice of a new residential municipality.

A way to approach the residential migration is to study the distances between the new and the former municipality. The Figure 14 illustrates this distance (between municipal centroids as we do not have more accurate localization in the database) for a sample of 100,000 people having moved between 01/01/2001 and 01/01/2002.

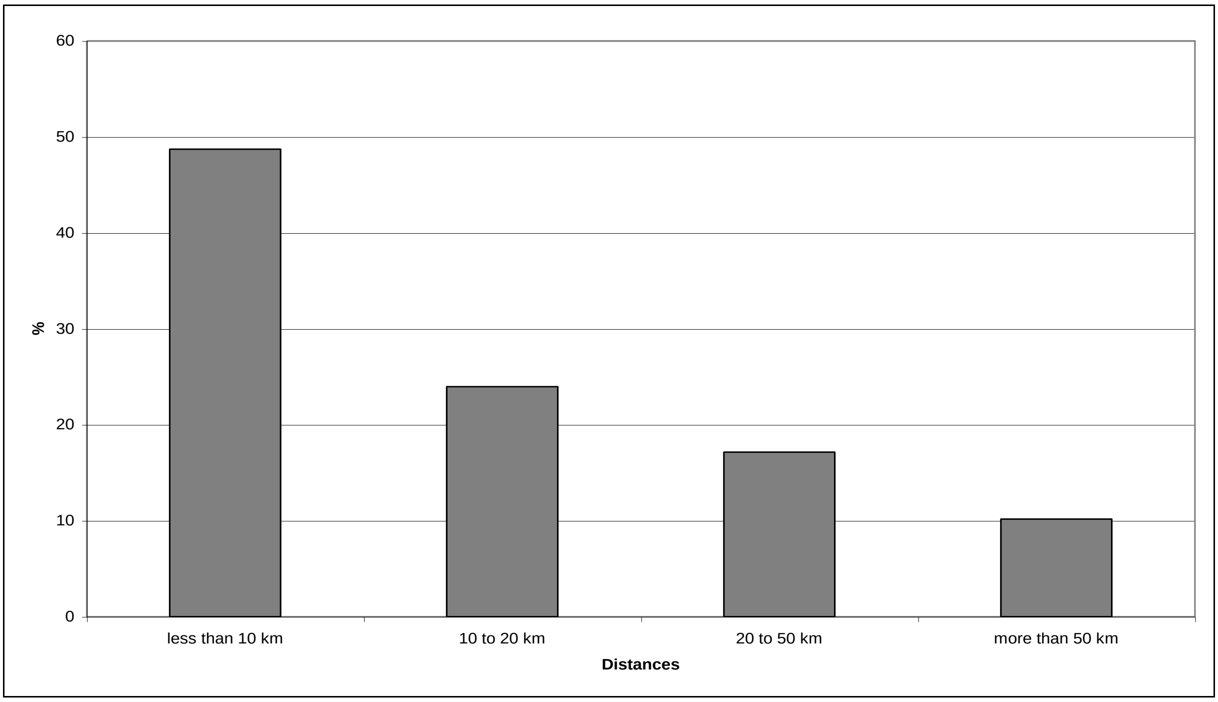

Figure 14: Intercommunal distances when a change of residence municipality occurs between 2001 and 2002 (sample of 100,000 heads of household)

One can observe that residential moves in Belgium are often short distance: for almost one in every two people changing their residential municipality, the intercommunal distance is less than 10 kilometers, whereas only 10 percents of the movers decide to go living 50 kilometers or more away from their former municipality of residence.

Other preliminary analyses were done with the covariates available for the modelling. For each of them, we estimated how each can bring explanations (Figure 15). To do so, we estimated models with only one explanatory variable (age of the individuals, distances between municipalities…) under BIOGEME (Bierlaire, M., 2003) and we present here below (Table IX) the results. Models are sorted by adjusted rho-square values. They were run with a sample of 60 municipalities and considered residential migration between 01/01/2001 and 01/01/2002 for head of household.

Among the available variables, the less explanatory variable taken independently is the property prices indicator. This variable takes into account the average prices for houses and flats in the municipality. On the contrary, the most explanatory one concerns the distance between the municipality where the individual lived in 2001 and the municipality where he settled down in 2002.

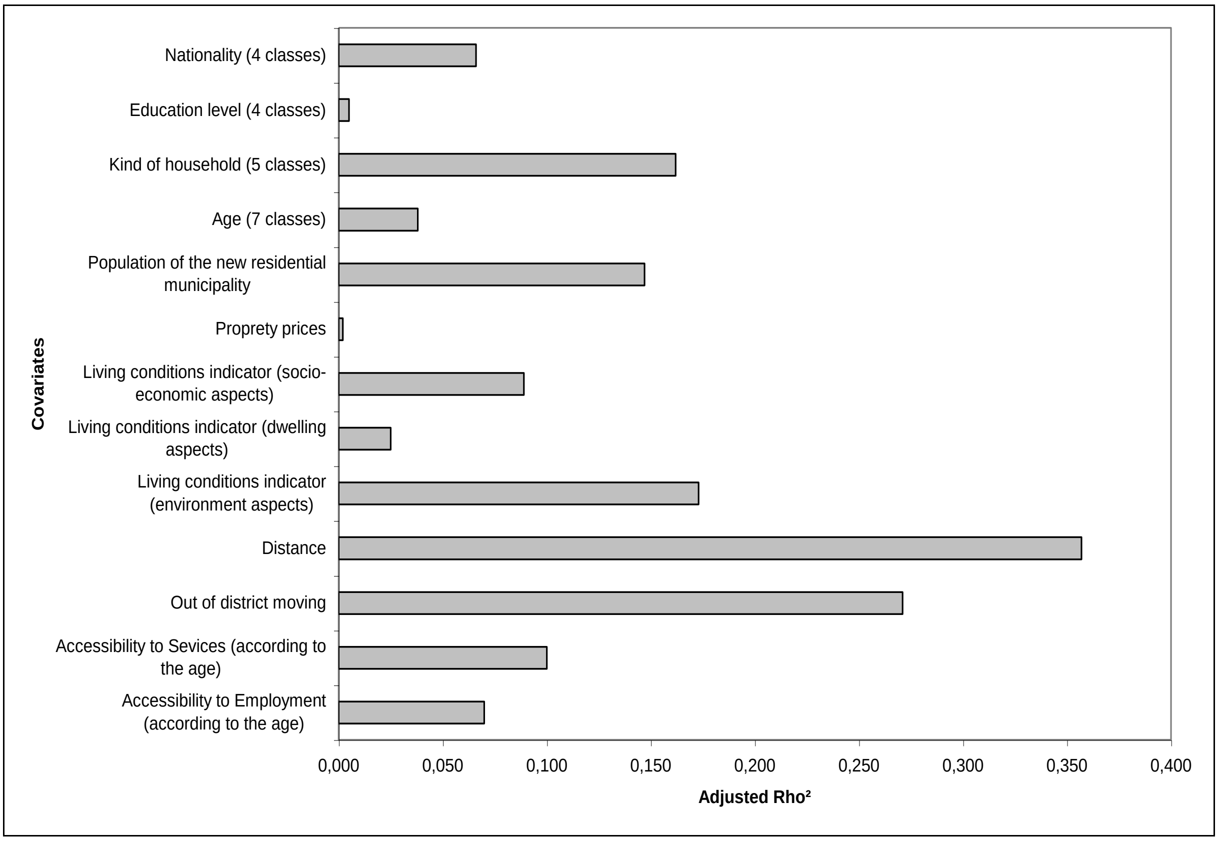

Figure 15: Adjusted Rho² of each covariate taken as only explanatory variable

Remark that β values can be interpret as follows: if β is higher than 0, it means that the attractiveness of the municipality increases (its utility raises) whereas if β is lower than 0 the attractiveness of the municipality decreases (its utility goes down).

Table IX: Parameter of each covariate modality (covariate being taken as only explanatory variable - let us point out that the only significativeless parameter (at α = 0.05) in this table is the one related to nationality of border countries, although its p-value is 0.10)

Covaraites	**β values**
Proprety prices indicator	-0.580
Education : other	5.77
Education : secondary school (inferior)	-4.82
Education : secondary school (superior)	-6.65
Education : higher education	7.19
Living conditions indicator (dwelling aspects)	3.73
Age : 0 to 18	-36.3
Age : 19 to 29	46.4
Age : 30 to 44	-13.5
Age : 45 to 54	-27.6
Age : 55 to 64	30.5
Age : 65 to 74	33.4
Age : 75 and more	53.8
Nationality : other	22.9
Nationality : belgian	-5.86
Nationality : border countries	*2.09*
Nationality : European Union	51.3
Accessibility indicator to employment for active people	0.629
Accessibility indicator to employment for non-active people	0.507
Living conditions indicator (socio-economic aspects)	6.06
Accessibility indicator to services for active people	0.0803
Accessibility indicator to sevices for non-active people	0.0649
Population of the new residential municipality	0.746
Kind of household : other	20.0
Kind of household : couple (married or not) with children	-10.7
Kind of household : couple (married or not) without children	13.7
Kind of household : single-household	38.4
Kind of household : one-parent family	34.1
Living conditions indicator (environment aspects)	8.23
Staying inside the same district	3.66
Distances between the new residential municpality and the former one	-0.0665

In general, municipality variables seem to bring more explanatory power than individuals ones.

Methodology

The purpose of the localization model is to determine which "new" municipality will be chosen by people who decided to migrate (between two consecutive years as studied in the propensity to move modelling). To do so the model will take into account the covariates which significantly play a role in this choice and their relative weight in the decision process.

Discrete choice method

The objective for Mobloc and for this part of the project is to understand processes of residential migration. The model may not be a black box that gives a new situation

but it has to describe some internal mechanisms of decision, to decompose the decision in some interpretable parameters, so that, it could be possible to act on some parameters for simulating the results of different changes/evolutions of behaviour or the effects of some policies.
Even if the macroscopic methods are more common and are useful for some purposes, our interest was however more compatible with micro-simulations. Therefore discrete choice method (Train, 2003) was chosen for modelling as well as possible the behaviour of people moving from a municipality to another. This technique consists in determining the utility of each alternative and then computing the probability of choice of each alternative. Nevertheless resorting to such models was notably ambitious because some aspects of the project and the available data would lead us to difficult discussion and decision.
To sum up the way discrete choice methods work, we can simplify things as follows: each agent of decision (in our case, each person leaving his housing to settle down in another municipality), evaluates the utilities of each alternative (which are all the other 588 Belgian municipalities in the frame of this project). The estimation supposes that a subject chooses the alternative with the highest utility. So only the difference between utilities is important, not the utility itself.
The utility of each alternative is a combination of explanatory variables weighted by some parameters (being estimated in a calibration phase). As we will develop below, these variables can be related to the unit of decision (age, type of household…) or to the alternatives (property prices in the municipality, distances between the alternative and the previous municipality...). The formulation of the utility can be summarized as shown on Figure 16:

$$U_{in} = C_i + \sum_j PInd_{ij} \times VarInd_{jn} + \sum_k PMun_k \times VarMun_k + \varepsilon_{in}$$

Figure 16: Formulation of the utility in discret choice methods

With:

- U_{in}: utility of the alternative i for the individual n
- C_i: alternative specific constant
- $PInd_{ij}$: the parameters for individual covariate j and alternative i
- $VarInd_{jn}$: covariates related to the individuals
- PMunk: parameters related to the alternatives (municipalities) characteristics
- $VarMun_k$: covariates related to the municipalities
- ε_{in}: random term (to model unobserved effects)

Let us note that the utility includes a random term which allows the model not to be deterministic and takes unmeasurable or unknown effects into account.

Different hypotheses can be made for the relation between the utilities and the choice probabilities. The most-used model is the logit formulation. Another possibility was to use a probit formulation which is less restrictive regarding the necessary assumptions but substantially more time consuming so that we choose the logit formulation.

Concerning the choice of the software for calibrating the model, we opted for BIOGEME (Bierlaire, 2003). This software is more flexible in the utility formulation than many other ones. Moreover the last versions of BIOGEME include Biosim, a package designed to perform simulations from the estimated model.

The choice of the discrete choice method came quite soon and easily in the project but some points of the model structure required long discussion and many trials before leading to an acceptable model. This phase of the project clearly benefited from the multidisciplinary nature of the research team. In particular the definition and choice of explanatory variables and the choices made to reflect the dynamic nature (time depth) of migration process were very enlightening.

We summarize here below one of the main considerations we faced:

Unit of decision (individual VS household)

As explained before, the propensity to move model was built with individual as unit/agent of decision. This choice was quite natural as the members of a household can have different behaviours. For example, if a child moves from his parents' house, only one individual from the household decides to move.

For the localization model, we would intuitively like to consider a group of individuals. Indeed, if some members of a household move to a new municipality, they will move together and "choose together" the same municipality. The unit/entity of decision should be this (new) entity/household.

Nevertheless, we cannot take the initial household for this unit because the composition of the households changes between two consecutive years.

The solution we adopted was to work with all the individuals who moved and were household head in their new municipality. It thus includes individuals who remain the household head (e.g. the whole household moves) and individuals who become household head (e.g. he/she moves alone…).

This unit "head of household" is characterised by some household variables (as the household type and size) so that the household information can also be used for the localization decision as it would intuitively be the case. A disadvantage of working this way is that we only consider that, for example, the level of education of the household head influences the decision and we should therefore suppose that this one is representative of the whole household. However the lack of other types of data prevented us to act another way.

Number of alternatives

The main challenge for this model is the number of alternatives. As the common space level defined for Mobloc is the municipalities, the choice set is composed of the 589 Belgian municipalities (although for a given individual, his original municipality is not available in the choice set, leading to 588 alternatives for the model).

So many alternatives in the choice decision is unusual in the literature. The theory does not set a limit in the number of alternatives. Nevertheless, we could suppose (and actually faced) that some complications could appear such as:

- computing problems due to the amount of necessary data,
- difficulty in the validation of some hypotheses, such as the independence of the alternatives (IIA).

Sampling of alternatives

In order to reduce the time of computation, it is possible to limit the number of alternatives that are considered as available for each unit of decision during the estimation process. It means that, for each unit (household head), we select randomly only some of the alternatives. Then, the algorithm compares the utility of the chosen alternative (observed in reality) with the utilities of the selected alternatives for this unit. (The selected alternatives are not the same for all the household heads).

From the theory (Ben Akiva and Lerman, 1985) we know that such a process does not change the value of the estimated global parameters if the hypotheses are respected (IIA, for example). Nevertheless, the precision is reduced and the confidence interval is larger.

Working this way allowed us to reduce the computing time. We performed estimations with 10 and 588 alternatives and results showed that we could gain substantial time for the estimation process using sampling procedure.

Model structure

There are different ways to express the link between the utilities and the choice probabilities. Different forms of discrete choice models exist: logit, probit, and more sophisticated structures. For example, the nested logit model takes into account similarities between alternatives: it considers that some alternatives belong to a same group by the way of additional parameters.

In our case, the nested logit was a promising way to explore. With so many alternatives, we could suppose that some similarities between municipalities would not be explained by the variables. We considered two sources of such similarities: the geographic proximity and the municipality type (rural, urban, etc.).

No automatic method exists to form the nests of alternatives; the question on how to group the 589 alternatives was thus also a big issue. Two main ideas were explored:

- geographic proximity,
- municipality type (rural, urban, etc.) : on basis of Van der Haegen typology or Van Hecke classification,
- both of them.

Variables

Another issue is related to the variables to be taken into account in the model formulation. In our case, here are the main available variables:

- Individual variables
  - age, education level, nationality, household type.

Source : National Register coupled with the Socio-Economic Survey (ESE2001).

- Communal variables :
  - Living condition indicator(s), property price, intercommunal distance.

  Source : previous project results.
  - Potential accessibilities (services, job opportunities).

  Source : accessibility model.

Utility formulation

Another issue regarding the number of alternatives concerns the formulation of the utility. The utility term for a variable can take different forms. We can try to estimate a model with a generic parameter for a variable. It supposes that this variable influences all the alternatives on the same way. Another method is to use alternative specific parameters, which means that this variable has a different parameter for each alternative. This can improve the model if the estimated parameters are significantly different from each others. Intermediate forms may also be considered.

In our case, performing tests with alternative specific parameters for a variable means that 589(588) parameters should be estimated (per variable). Working this way would see the number of parameters exploding. Moreover, most of these would be insignificant or inaccurate, especially for the alternatives (municipalities) for which we do not have many observations (small municipalities). (It would require a lot more of data).

In our modelling, we can give as an example the case of the age covariate: we decomposed this variable in 7 classes (less than 18, 19 to 29, 30 to 44, 45 to 54, 55 to 64, 65 to 74 and 75 and more); so if we kept the same way of formulating as what is usually achieved, 4,123 parameters (7 * 589) would have had to be estimated for the consideration of only this covariate in our utility. Problems would rapidly occur in the estimation process if working like that (unidentifiable model and/or too high time running).

To avoid this difficulty, we decided to use a generic parameter for all the alternatives. To come back to the example of the age, we reduced so the number of parameters to 7.

As explained before, it was unrealistic to use alternative specific parameters and we decided to use generic parameters. Such a technique reduces the number of parameters to estimate but it makes the individual variables non-explanatory. Each of the 589 alternatives would indeed have the same utility term for the variables of this kind (e.g. the value of the age variable would be the same whichever municipality is considered). As the choice lies on the difference between alternatives utilities, this term may not be explanatory (the model could not distinguish amongst alternatives relatively to such a variable) and lead to unidentifiable models.

In order to get different values for the explanatory variables and to avoid problem of constancy between the alternatives, we had recourse to a "contextualization approach": for categorical individual variable; the utility term of an individual variable is computed as the product of a class specific parameter, a binary variable indicating whether an individual belongs to this class and the class proportion in the municipality (alternative) (e.g. for the "less than 18" age class, the utility term is the product of the associated parameter (giving the weight of this variable in the utility), a binary variable having 1 as value if and only if the decider is less than 18 years old and a number (between 0 and 1) giving the part of people less than 18 years in the population of the considered municipality). Formally, it implies one term per class, but for an individual every but one of these terms are null: the term of the class that the individual belongs to. So, for an individual j, the utility term for such a variable concerning alternative i is actually the product of a class specific parameter with the class (the individual belongs to) proportion in the municipality.

For example, for the individual j and the alternative i, utility terms for the 7 age classes are written as follows:

$$U_{i,j} = ... + \sum_{k=1}^{7} P_{age\_class\_k} * IsInClass(k) * \mathrm{Pr}op_{class\_k,alt\_i} + ...$$

where

- $P_{age\_class\_k}$ is the parameter for the age class k,
- IsInClass(k) is a binary variable indicating if the individual i is in class k,
- $\mathrm{Prop}_{class\_k,\ alt\_i}$ is the proportion of the class k in the alternative (municipality) i

For all individuals in class k=r, this expression reduces in:

$$U_{i,j} = ... + P_{age\_class\_r} * \mathrm{Pr}op_{class\_r,alt\_i} + ...$$

The interpretation of the parameters is less evident but can be explained this way : a positive parameter would indicate that people having the same profile tend to settle down in a same place while a negative parameter would represent a trend to a dispersion of the concerned people.

Link with former residence municipality

We mainly tested two ways to take into account the link with the former residence municipality:

- the intercommunal distances (distances between the communal centroids),
- a change of district via a binary variable (1 for municipalities in the same districts as the former one).

Accessibility indicators

We tested different forms for the utility terms related to the accessibility indicators.

- a unique parameter,
- distinct parameters
  - ✓ for each age class (7 classes),
  - ✓ for active and non active people (2 parameters per indicator).

The idea of distinct parameters was to try to detect different influences of accessibility according to individual characteristics such as age class. For example, we can suppose that the job opportunities indicator could be less explanatory for elderly people.

Practical estimation

Even with sampling of alternatives, estimation remains time consuming. With BIOGEME, only data reading was already memory space and time consuming.

Therefore, in order to spare time, we first worked with a sub-set of data and alternatives to test several model structures. Then, we performed estimation on bigger datasets and all the alternatives for the models that appeared the best ones with the partial data.

The way we chose this subset of alternatives was not random. We wanted to have a representative subset of Belgian municipalities, i.e. municipalities from different types (rural, urban, etc.) and from different zones. We thus fixed 5 zones (Brussels Region, two in Flanders, two in Wallonia) and the proportion of municipality types we wanted for each zone (according to the real proportion).

As a summary, the Table X below summarizes the main process in the model building.

Table X: Practical estimation of the localization model

1. Tests with 60 municipalities	1.1. Sampling or not
	1.2. Structure : logit or nested (different ways to build the nests)
	1.3. Variables : selection and forms (contextualisation...)
2. Estimation with 589 municipalities	

Models and results

Goodness-of-fit measures

A discrete choice model simulation gives for each individual probabilities of choosing each alternative. Comparing the observed choice with the predicted one is not relevant for discrete choice model. Indeed, the model do not predict "the choice" of an individual but gives probabilities for each alternative to be chosen by this individual.

In order to measure how well a model fits the data, Train (2002) recommends the use of the likelihood ratio index.

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

where LL(β) is the value of the log-likelihood function at the estimated parameters and LL(0) is its value when all the parameters are set equal to zero (i.e. a purely random model).

Its value is comprised between 0 and 1. A value of 0 means that the model is not better than no model (more exactly than a purely random one) . A value of 1 means that the model predicts perfectly each sampled decision maker's choice.

The index ρ looks similar to R^2 used in regression but its interpretation is completely different and values between 0 and 1 of ρ have no intuitive interpretation. So, the index can only be used to compare models built on the same data and with the same set of alternatives. In this case, the higher is the index ρ, the better is the model.

Another possibility to compare model is the hypothesis testing. It can be used to test whether two parameters are significantly different from each other or if a unique parameter is sufficient.

Model results

Here are exposed the main conclusions of the models testing and the results of the finally retained model.

Sampling of alternatives

Models were tested without and with sampling of alternatives (10% of the alternatives used for estimation). As expected, the confidence intervals are larger for estimation with sampling which use less information to estimate the parameters. But the parameters values were not quite different, i.e. the confidence intervals (95%) "have a common part of range". Generally, the values with sampling belong to the confidence interval of the parameter without sampling.

Model structure

The retained model is a nested model with 4 nests based on the communal typology (and not on the geographic proximity).

The four nests regroups the municipalities into:

- agglomerations,
- suburbs,
- commuters municipalities,
- rural municipalities.

according to the Van der Haegen typology (1996).

Let us note that the geographic proximity (other type of nests that was tested) is more efficiently taken into account thanks to the distance variable (see further). More sophisticated nests (crossing both geographic and typological criteria) lead to model with many unsignificant parameters.

Covariates

The most significant variable is the distance from the former municipality (distance between the two centroids). The negative sign of the distance parameter expresses that the utility of a municipality decreases with the distance from the former municipality (see Table XI). As explained in the preliminary analysis, (intermunicipal) residential moves are mostly short distance moves.

Table XI: Parameters of the retained model for the residential localization (calibration on all municipalities without sampling)

Parameter	Value	Std err	t-test	p-value
Accessibility indicator to employment for active people	-0.241	0.00573	-42.09	0.00
Accessibility indicator to employment for non-active people	-0.318	0.0134	-23.66	0.00
Distances between the new residential municpality and the former one	-0.0663	0.000385	-172.28	0.00
Living conditions indicator (environment aspects)	0.949	0.0483	19.66	0.00
Living conditions indicator (dwelling aspects)	-1.67	0.0699	-23.84	0.00
Living conditions indicator (socio-economic aspects)	3.57	0.0639	55.93	0.00
Living conditions indicator (services aspects)	-1.73	0.0757	-22.90	0.00
Property prices indicator	0.407	0.0262	15.55	0.00
Population of the new residential municipality	0.506	0.00534	94.85	0.00
Age : 0 to 18	-4.27	1.77	-2.41	0.02
Age : 19 to 29	13.1	0.289	45.20	0.00
Age : 30 to 44	4.59	0.365	12.58	0.00
Age : 45 to 54	0.480	0.887	0.54	0.59
Age : 55 to 64	19.8	0.967	20.43	0.00
Age : 65 to 74	10.8	1.07	10.09	0.00
Age : 75 and more	1.30	1.36	0.96	0.34
Education : other	2.22	0.168	13.23	0.00
Education : secondary school (inferior)	5.45	0.292	18.63	0.00
Education : secondary school (superior)	4.41	0.300	14.73	0.00
Education : higher education	4.19	0.109	38.31	0.00
Kind of household : other	0.234	1.09	0.21	0.83
Kind of household : couple (married or not) with children	2.36	0.128	18.36	0.00
Kind of household : couple (married or not) without children	2.78	0.256	10.89	0.00
Kind of household : single-household	8.11	0.152	53.42	0.00
Kind of household : one-parent family	4.03	0.248	16.27	0.00
Nationality : other	5.94	0.278	21.38	0.00
Nationality : belgian	2.71	0.0745	36.46	0.00
Nationality : border countries	4.23	0.362	11.68	0.00
Nationality : European Union	4.45	0.312	14.28	0.00

Regarding the accessibilities variables, only one of the two accessibility variables was kept. The accessibility to services indicator and the accessibility to employment indicator are indeed highly correlated (r=0.89). The decision to keep the accessibility to employment indicator came from the consideration that the accessibility to services indicator is more complicated to make evolve. Moreover, attractiveness of services can be taken into account via one of the four components of the living conditions indicator (services aspects).

In the utility formulation, the best results for the accessibility to employment indicator are obtained with specific parameters for active and non active individuals. People were considered as active between 19 and 64 year old, and non active outside of these limits.

Considering these two distinct parameters lead to a better model than considering a unique parameter for employment indicator. Introducing one parameter per age class (7 classes) does not improve significantly the quality of the model.
Both parameters for accessibility to employment indicator are negative.

Regarding the other municipal characteristics:

- the four components of the living condition indicator are significant : Note that these indicators have a range between 0 and 1 and the 1 value is the less interesting so that the parameters must be carefully interpreted;
- the average property price at municipal level is also significant.

Regarding the individual characteristics, let us recall that these variables have been defined as contextualisation variables, so that a positive sign means that individuals tend to settle down in the municipalities where they find similar individuals. In the retained model, most of these parameters are significant. Amongst them, only the parameter for 0-18 age class is negative. All other parameters are positive (for age classes, household types, nationalities and levels of education). From this point of view, the model reflects well the behaviour of tending to live with one's own socio-economic group.

3. POLICY SUPPORT

The project's objective is to investigate the links between long-term residential choice, accessibility and daily mobility, with the ambition to provide better understanding of the behaviour of Belgian households regarding these issues. In particular, the respective importance of several classes of factors is crucial for the establishment of land-use and accessibility policies. The mixing of long-term decisions such as housing and short-term ones such as daily mobility turns out to be a challenging issue.

The project developed three main models: an accessibility model, a model of the propensity of an individual to move his/her residence and a model of localization choice for a new residence.

The first of these models (accessibility) involves a specific assignment procedure to compute travel times matrices from OD matrices at the municipal level; the flows are assigned on a simplified road network that was built to this purpose. The model outputs both travel time in free-flow for non-peak hour and travel times for peak hours, taking congestion into account. Finally, a few relevant accessibility indicators were also produced; they allow the characterization and comparison of municipalities and we feel that this has direct consequences on municipality management.

The model describing the propensity of the Belgian individuals to move their residence incorporates a number of explanatory factors at the individual and household levels. Not unexpectedly, the parameters which turned out to dominate the individual's choice of changing residence are changes in the household structure, change in the position in the household and age class.
This reinforced the idea that societal trends (as opposed to material infrastructure evolution) is crucial to explain internal migration within a country. In particular, population aging and the growth in importance of smaller household structures may present specific challenges in urban planning and land-use in general during the forthcoming years.

Finally, the residential localization model is central to the design of suitable land-use regulations at the regional level. Remarkably, our analysis indicates that the dominant factors are, by decreasing level of importance, the distance between the previous residence and the new one, the perceived quality of life in the new municipality of residence, the household structure, and, in fourth position, the accessibility of the new municipality of residence. Accessibility is therefore less important than expected at the start of the project.

A simple interpretation of the results obtained in the course of the MOBLOC project is that internal migration within the country is less determined by infrastructural factors (within which accessibility is a important example) than factors related to societal life in a more general sense: household structure and its evolution, closeness to one's relations, age class and quality of life score indeed higher than purely transport related factors in our results.

We believe that these conclusions are important for any prospective view on the development of land-use and transportation systems. They will be discussed in the new regional prospective study group (SRP) which is being established under the leadership of the Institut Destrée and the Institut Wallon d'Evaluation, de Prospective et de Statistiques (IWEPS). We also intend to dissiminate those conclusions more widely, via scientific publications but also aiming the municipality managers and the general public.

4. DISSEMINATION AND VALORISATION

In this section, we list different research projects which rely on the outcomes of the MOBLOC project our use its main results, but we can also make reference to different informal dissemination in the frame of contacts abroad (e.g. meeting focusing on the residential localizations at the LET, Lyon, in January 2009) or coming seminar in the frame of the NAXYS seminars (Namur Center for Complex Systems).

Three first projects are presented here after: as explained, they can be seen as prolongations of the MOBLOC Project.

A. MOEBIUS - Mobilities, Environment, Behaviours Integrated in Urban Simulation

CEPS/INSTEAD coordinates a research project (MOEBIUS) funded by the FNR (Fonds National de la Recherche Luxembourg) that deals with some of the issues addressed in the MOBLOC project. The outcomes of our project (and the expertise derived of our research) will be helpful for the several partners involved in this MOEBIUS project to reach the objectives planed. A brief description of this new project follows here below.

Luxembourg emerges as a very attractive cross-border regional metropolis leading to increasing residential migration (topic modelled in the MOBLOC project) and longer commutes. Empirical evidence shows that current urbanisation trends toward suburban and more remote periurban areas favour urban sprawl and car dependence. In Luxembourg, the awareness of the role of spatial planning for the implementation of sustainable development has led to the developing of a planning concept, called IVL, involving several Ministries. It promotes an integrated spatial development which joins number of assumptions of the New Urbanism Theory. The main objective of the MOEBIUS project is then to assess the sustainability of this planning scenario (IVL / "New Urbanism") comparing to the current urban sprawl.

Further understanding the social and environmental impacts of both the residential mobility and the daily mobility of households is at the core of our research project. We will deal with the complexity of this issue by using spatial simulation tools that are hybrids from agent-based and cellular-automata. This simulation platform prototype will integrate land use, residential mobility (well studied in the frame of the MOBLOC project) and daily mobility, within the cross-border metropolitan area of Luxembourg. Because of its complexity, this platform will be devoted to a particular population (i.e. workers employed in the Grand-Duchy and living inside or outside the country) and involve expertise from different disciplines (urban planning, geography, psychology, economics, mathematics, and computer sciences).

The development of a small, simple and transparent model is planed: this model will include new theoretical components, by using belief theory, residential choices and determinants of daily commuting specific to Luxembourg.

The aim is to simulate (i) the future urbanisation (based on demographic forecast and residential choice modelling) in the commuting area of Luxembourg, and (ii) the future daily mobility (commuting pattern and travel mode choice) for different planning scenarios. When assessing those scenarios, particular attention will be put on how they can provide a good trade-off between, economic growth (via the provision of attractive and affordable living places) and environmental and social sustainability (modal split, land take, accessibility…).

Because the simulation model is spatially explicit, the specificities of Luxembourg can be taken into account, e.g. the urban network and the cross-border setting. In addition, unique datasets are accessible in the frame of this project (e.g. the cross-borders transport & mobility survey (2010) or Social Security longitudinal data) and building on the empirical knowledge based on the MOBILLUX outcomes is possible (FNR project on the link between daily and residential mobilities in Luxembourg).

B. SimBelgium - NAXYS

The substantial amount of work invested during the MOBLOC project into the design and calibration of models for internal migration of individuals and households within Belgium has resulted in the creation of a new research project in the framework of NAXYS, the Namur Center for Complex Systems. After a short description of the NAXYS interests, we briefly describe this projet below.

NAXYS is a new research center at FUNDP (established in the fall 2010) whose purpose is the analysis, modelling and general scientific approach of complex systems in society and nature. It has research interests in several complementary directions, including natural chaotic systems (in celestial mechanics and elsewhere), large networks, opinion propagation, weather forecasting and, more importantly in this context, spatio-temporal dynamics of large populations. It is truly multi-disciplinary group, combining mathematicians, economists, geographers, computer scientists, engineers, etc. Its initial activity (which can be found on http://www.fundp.ac.be/en/sci/naxys) reflects its scientific dynamism.

The NAXYS project which is extending the work of MOBLOC is (temporarily) named "SimBelgium" and aims at producing a detailed microsimulation of the behaviour of the Belgian population with a special focus on the interaction of internal migration (the MOBLOC input is crucial here), population dynamics and evolution of the participation to the job market.

This microsimulation is based on the availability of a virtual population for the complete Belgian territory, whose (synthetic) individuals belong to (synthetic) households, themselves localized in the 589 Belgian municipalities. The technique of using virtual individuals is necessary in order to avoid privacy issues in the research program. Daily activity chains (obtained from the MOBEL and, when available, BELDAM projects) are associated with each individual. This in turns is intended to allow a dynamic simulation of the aggregate effect of the
combined activity demand.

The research group associated with this project includes participants of the MOBLOC projects, but extends well beyond this group. In particular, direct collaborations have been established with the FUNDP Department of Geography in order to secure additional expertise in the spatial analysis of internal migration and of the housing market and their dynamics.

The project has already started and has, in the past three months, established a first version of the population dynamics over a period of 30 years. The association of activity patterns with individuals and households is currently proceeding. The incorporation of the results obtained by the MOBLOC project is expected to follow shortly. Of particular interest in this context are the MOBLOC models for the propensity of individuals to migrate within the country and the localization model at the municipality level. Because a large part of the methodology used in the new project is based on modelling the many individuals' choices by using tools derived from the random utility theory, and because this approach has also been central in the MOBLOC project, the integration of the results of the latter in the former is expected to proceed reasonably smoothly.

It is very clear that this new ambitious project would not have been possible without the significant contribution of the MOBLOC research.

This project is also at the basis of several discussions that have taken place in the context of the SRP (Service Régional de Prospective), a working group established in the Région Wallonne under the leadership on the Institut Destrée and Institut Wallon d'Evaluation, de Prospective et de Statistiques (IWEPS).

C. Population and households forecasting of Belgian municipalities

The aim of this project is to provide population and households forecasting across municipalities of Belgium. It will be multi-state projections.
A special method taking into account the problem of "small numbers" will be developed.
This project will provide for each municipality, in 2020, the number of population, population distribution by age and sex and household characteristics.
Several scenarios will be considered and the assumptions of internal migration trends will be based on the results of the propensity to move model developed in the project MOBLOC

Finally, we must also mention that links were established with the team in charge of the SIMBAD projet at LET (Laboratoire d'Économie des Transports), Lyon (F). Their goal is to apply a LUTI (land Use and Transport Integrated) model to the Lyon agglomeration. These researchers are quite interested in the methodologies developed and used for this MOBLOC project. Therefore they have invited Eric CORNELIS to become member of their accompying committee.

5. PUBLICATIONS

KLEIN S., OMRANI H., CARPENTIER S., GERBER P., VANDENBULCKE G., Validation d'un modèle d'accessibilité par recoupement de données multi-sources. Application aux communes de Belgique, 2ème Journée de Recherche « Mobilité, Transport et Logistique » (MTL 2010), Lyon, Mercredi 23 juin 2010 (copy in annex)

PAULY X., CORNELIS E., WALLE F., MOBLOC : links between residential mobility and accessibility, in "Proceedings of BIVEC Transport Research Day 2011", E. CORNELIS ed., pp. 126-137, University Press, 2011

BAHRI A., DAL L., CORNELIS E., EGGERICKX T., PAULY X., WALLE F., Propensity to move and residential localization choice in Belgium, modelling of the individual behaviours, Namur Center for Complex Systems (NAXYS) Report, 2011 (in preparation)

DAL L., CARPENTIER S., CORNELIS E., EGGERICKX T., KLEIN S., GERBER P., PAULY X., WALLE F., TOINT P., Interactions between accessibility and residential choices in Belgium: modelling at a municipal level, Namur Center for Complex Systems (NAXYS) Report, 2011 (in preparation)

6. ACKNOWLEDGMENTS

In order to perform our calibrations for the accessibility model, the national network was necessary. This network was provided by the National Ministry of Mobility and Transports (SPF MT – Service Public Fédéral Mobilité et Transports).
Concerning the residential choices, numerous variables were necessary for the calibration of models. Those were available thanks to by the National Register and Statbel (Service Public Fédéral Economie, P.M.E., Classes moyennes et Energie, Direction générale Statistique et Information économique).

7. REFERENCES

ALLISON P., 1999, Logistic Regression Using SAS: Theory and Application, SAS Press

ANDAN O., POCHET P., ROUTHIER J.-L., SCHEOU B., 1999, Stratégies résidentielles des ménages et mobilité domicile-travail, Programme de recherche Transports Terrestres 1996-2000, Direction de la Recherche et des Affaires Scientifiques et Techniques

BARTON-ASCHMAN ASSOCIATES, Inc.; CAMBRIDGE SYSTEMATICS, Inc., 1997, Model Validation and Reasonableness Checking Manual, 118 p.

BEN AKIVA M., LERMAN S., 1985, Discrete choice analysis, MIT Press

BERTRAND M., DUPONT V., GUÉRIN-PACE F., 2003, Espaces de vie. Une revue des concepts et applications, Paris, INED, Documents .de travail, Paris

BIERLAIRE M., 2003, BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland,

BLACK W. R., 1998. Sustainability of Transport. In: Hoyle, B. S., Knowles R., (Eds), Modern Transport Geography, 2nd Edition. Chichester: Wiley, 337-351.

BLIJIE B., 2005, The impact of accessibility on residential choice: empirical results of a discrete choice model, Delft University of Technology, 45TH Congress of the European Regional Science Association, Vrije Universiteit Amsterdam

BRANSTON, D., 1976, Link capacity functions : a review. Transportation Research 10, 223-236

BRÜCK L., MAIRY N., HALLEUX J.-M., MERENNE-SCHOUMAKER B., SAVENBERG S., VAN HECKE E., 2001, Les comportements résidentiels des ménages face à la problématique du développement durable

BUREAU OF PUBLIC ROADS,1964, Traffic Assignment Manual. Urban Planning Division, US Department of Commerce, Washington D.C.

CORNÉLIS E., BAHRI A., EGGERICKX T., CARPENTIER S., KLEIN S., GERBER P., PAULY X., WALLE F., TOINT P., 2009. Mobilities and long term location choices in Belgium "MOBLOC", Final report for the Belgian Science Policy, Phase 1. 47 p.

COURGEAU D., BACCAINI B., 1997,Analyse multi-niveaux en sciences sociales, Population (French Edition), 52e Année, No. 4, Nouvelles approches méthodologiques en sciences sociales, pp. 831-863.

DEBRAND T., TAFFIN C., 2005, Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans, Economie et statistique n° 381-382, pp. 125-146

DELAUNAY D., 2001, L'inscription dans l'espace des biographies individuelles. Communication présentée à la XXIVe Conférence Générale de l'IUSSP, Salvador (Brésil). (téléchargée sur le site de l'IUSSP).

DES ROSIERS F., THÉRIAULT M., KESTENS Y. and VILLENEUVE P., 2002, Landscaping attributes and house prices: looking at interactions between property features and homeowners' profiles, Paper presented at the 18th Annual ARES Conference, Naples, Florida.

DES ROSIERS F., THÉRIAULT M., VILLENEUVE P-Y. and KESTENS Y., 2001, Landscaping and house values : an empirical investigation, Laval University, Paper submitted for presentation at the 17th Annual ARES Conference, Coeur D'Alene, Idaho

DIJKSTRA E. W., 1959, A note on two problems in connexion with graphs. Numerische Mathematik, 1, p. 269–271

DUJARDIN C., 2001. Effet de frontière et interaction spatiale. Les migrations alternantes et la frontière linguistique en Belgique. L'espace géographique 01(4): 307-320.

EGGERICKX T., DEBUISSON M., DAL L., 2008, Les tendances récentes des migrations entre les arrondissements, In: Perspectives de Populations 2007-2060, Bureau Fédéral du Plan / DGSIE ed(s), Planning Paper, 105, p. 35-38

EGGERICKX T., SANDERSON J.-P., BAHRI A., 2010, Transformation des ménages et ségrégation sociale de l'espace. Une analyse de la précarisation des parcours résidentiels en Belgique (1991-2005), Dynamiques de pauvretés et vulnérabilités. Mesures et processus explicatifs en démographie et en sciences sociales, Bruno Masquelier, Thierry Eggerickx, ed(s), Chaire Quetelet 2007, Presses Universitaires de Louvain, Louvain-la-Neuve, pp. 187-222

EGGERICKX T., HERMIA J.-P., SURKIJN J., WILLAERT D., 2011 (à paraître) (avec la collaboration de DAL L., POULAIN M. et SANDERSON J.-P.), Les migrations internes en Belgique, Monographie 2 de l'Enquête socioéconomique de 2001, DGSIE, Bruxelles, 200 p.

ENAUX C., 1997, Essai de modélisation spatio-temporelle des flux de déplacements de travail. Exemples de la région urbaine strasbourgeoise de 1975 à 1990, Thèse de doctorat, Université Louis Pasteur, Strasbourg, 265 p. (+ 1 volume d'annexes 107 p. + 3 disquettes)

FLOWERDEW R., AITKIN M., 1982. A method of fitting the gravity model based on the Poisson distribution, Journal of Regional Science, 22(2): 191-202.

FLOWERDEW R., 1991. Poisson regression modelling of migration. In Congdon P., Stillwell J., Migrations models : macro and micro approaches, London: Behaven Press, 92-112.

FOOT D., 1981, Operational Urban Models : an Introduction. London: Methuen & Co Publishing Company, 230 p.

GOODWIN, P., 1995, Car dependence. Transport Policy, 2(3), p. 151–152.

HALLEUX J.-M., 2001, Évolution des organisations urbaines et mobilités quotidiennes : espace de référence et analyse des processus, L'Espace géographique, Vol. 30, n°1, pp. 67-80

HENLEY A., 1998, Residential mobility, housing equity and the labour market, The Economic Journal, n° 108, pp. 27-414

HOSMER D. W., LEMESHOW S., 2000, Apllied Logistic Regression, 2nd Edition,

HUBERT J.-P., TOINT P., 2002. La mobilité quotidienne des Belges, Presses Universitaires de Namur, 352p.

IRWIN M., BLANCHARD T., TOLBERT C., NUCCI A., LYSON T., 2004, Pourquoi certains ne migrent pas : L'impact du contexte local sur la sédentarité aux Etats-Unis, Population (French Edition), 59e Année, No. 5, pp. 653-680.
JOLY I. ,2005, Décomposition de l'hypothèse de constance des budgets temps de transport, in Mobilités et temporalités Montulet B et al.(dir), Facultés Universitaires Saint Louis, Bruxelles, 129-150

KNUDSEN D., FOTHERINGHAM S., 1986. Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling, International Regional Science Review; 10(2): 127-147, doi:10.1177/016001768601000203.

MURAT C.H., 2010. Sample size needed for calibrating trip distribution and behavior of the gravity model. Journal of Transport Geography, 18(1): 183-190 doi:10.1016/j.jtrangeo.2009.05.013.

NELDER A., WEDDERBURN R., 1972. Generalized linear models, Journal of the Royal Statistical Society. Series A (General), 135(3): 370-384.

NIJKAMP P., RIETVELD P., SALOMON I., 1990. Barriers in spatial interactions and communications. A conceptual exploration. The Annals of Regional Science 24, 237-252, doi:10.1007/BF01580472.

PUSHKAR A. O., Modelling household residential search processes: methodology and preliminary results of an original survey, University of Toronto

REYMOND H., CAUVIN C., KLEINSCHMAGER R., (coord.), 1998, L'espace géographique des villes - Pour une synergie multistrates, Anthropos, collection Villes, Paris, 557 p.

ROMANI J., SURIN'ACH J., ARTI M., 2003, Are commuting and residential mobility decisions simultaneous?: The case of Catalonia, Spain, Regional Studies, Vol. 37.8, pp. 813–826

SHANNON, C. E., 1948, The mathematical theory of communication. Bell System Technical Journal 27: 379-423 ; 623-56

TAYLOR P., 1983, Distance decay in spatial interaction, CATMOG - Concepts and Techniques in Modern Geography, n°2, 35 p.

TRAIN K. E., 2003, Discrete Choice Methods with Simulation, Cambridge University Press

VAN DER HAEGEN H., VAN HECKE E., JUCHTMANS G., 1996, Les régions urbaines belges en 1991, Études Statistiques, n° 104

VAN HECKE E., 1998, Actualisation de la hiérarchie urbaine en Belgique, Bulletin du Crédit Communal, vol. 205, n° 3, p. 45-76.

VANDENBULCKE G., STEENBERGHEN T., THOMAS I., 2009, Mapping accessibility in Belgium: a tool for land-use and transport planning? Journal of Transport Geography, p. 39-53

VANNESTE D.., THOMAS I., GOOSSENS L., 2007, Woning en woonomgeving in België, Sociaal-economische enquête 2001 , SPF Politique scientifique & SPF Economie – Direction générale Statistique

WARDROP J.G., 1952, Some theoretical aspects of road traffic research, Proceedings of the Institute of Civil Engineers. Pt. II, vol. 1, p. 325-378

WILLEKENS F., DE BEER J., VAN DER GAAG N., 2005, MicMac - From demographic to biographic forecasting, paper prepared for presentation at the Joint Eurostat-ECE Work Session on Demographic Projections, September 21-23, Vienna.

WILSON A., 1971. A family of spatial interaction models, and associated developments, Environment and Planning A, 1971(3): 1-32.

ZAHAVI Y., 1979, The 'UMOT' Project. Report prepared for the U.S. Department of Transportation and the Ministry of Transport of Federal Republic Of Germany, 267 p.

ZAHAVI Y., TALVITIE A.,1980, Regularities in Travel Time and Money Expenditure, in : Transportation Research Record, 750, 13-19.