VirtualBelgium: a simulation platform for the Belgian population and an application to mobility

Johan Barthélemy, Eric Cornelis, Laurie Hollaert and Philippe L. Toint*

February 28, 2013

Abstract

The VirtualBelgium project aims at developing understanding of the evolution of the Belgian population using agent-based simulations and considering various aspects of this evolution such as demographics, residential choices, activity patterns, mobility, etc. This simulation is based on a validated synthetic population consisting of approximately 10,000,000 individuals and 4,350,000 households localized in the 589 municipalities of Belgium (LAU2 level).

The mobility behaviour is simulated using an activity-based approach in which the travel demand is derived from the activities performed by the individuals. This approach has the advantage of reflecting the scheduling process of activities in time and space. Consequently, an agenda is assigned to the individual agents using a 3-steps procedure: generation of a set of activity chains for each individual type, activity chains assignment to each individual and choice of activity's duration and localization.

Travel behaviour of an individual changes over time, *i.e.* the set of feasible activity chains patterns evolves together with his/her socio-demographic characteristics. It is therefore interesting to have an evolution process for the population to forecast future travel demand.

The final outcome is an integrated micro-simulator which is, to the authors' knowledge, the first attempt to simulate the evolution and mobility behaviours of a population of more than 10.000.000 agents and not only a sample of the population of interest. Note that this work is still in progress: the paper will only present the current state of VirtualBelgium: the synthetic agents' characteristics, the general design of an activity-based model and the first steps regarding the population evolution together with their preliminary results.

Keywords Micro-simulation, agent based simulation, activity chains, transport demand forecasting, synthetic population evolution

1 Introduction

Activity-based models is a class of travel demand forecasting model originally based on the works by Hägerstrand (1970) and Chapin (1974) proposed as an alternative to the classical four-stages, trip-based, models for travel demand forecasting, whose drawbacks have been identified in previous works (Dickey (1983), Domencich and McFadden (1975), Spear (1977), Oppenheim (1995)). Activity-based approaches relies on the paradigm that people travel to carry out activities they need or wish to perform. Such models reflect the scheduling of activities in time and space and the sequence of activities becomes the relevant unit of analysis. Activity scheduling problems have been discussed in the literature since the 1970s. They are nowadays widely accepted and continue to attract a lot of attention.

The activity-based models can be categorized in at least four families. The first two are discrete choice models (Adler and Ben-Akiva (1979), Bhat and Koppelman (1999)) and mathematical programming (Gan and Recker (2008)). Nevertheless, the former approach may requires an extremely large choice set in order to capture every feasible mobility pattern, while the latter may not be tractable as the decision processes' formulation may be extremely complex. Structural equation modelling is also an alternative for activity-based modelling. We refer the reader to (Golob 2003) for a review of works using this approach. Finally, since the advent of high performance computing it has also been noted that *"micro-simulation ... is drawing attention as*

*Namur Research Center for Complex Systems (NAXYS), University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: {johan.barthelemy}{eric.cornelis}{laurie.hollaert}{philippe.toint}@fundp.ac.be

a new approach to travel demand forecasting" (Miller 1996). This allowed the creation and use of massive multi-agents simulations in order to reproduce the behaviours of a complex system such as the mobility behaviours of a large population (Kitamura, Chen and Pendyala (1997)), leading to the development of powerful open source simulation systems such as MatSim (see http://www.matsim.org, accessed on February, 2013), used by Meister et al. (2010) for a large-scale agent-based for travel demand forecasting, and Urbansim (Waddell (2002)). Several operational micro-simulators such as ALBATROSS (Arentze and Timmermans (2000)), ILUTE (Salvini and Miller (2005)) and SAMS (Kitamura et al. (1996)) are currently in use. See (Goran 2001) for a review and comparison of various micro-simulators and discrete choice models for activity-based modelling.

The approach taken in this work is the micro-simulation of the Belgian population's mobility behaviours: this paper details the design of a nationwide agent-based micro-simulator named VirtualBelgium. The agents are derived from a validated synthetic population previously generated (see Barthelemy and Toint (2012) for a complete description of the synthetic population generator). The proposed activity scheduling model is a three steps procedure: first, a set of feasible activity chains patterns is generated for every agent type; a chain is then randomly drawn and assigned to every individual agent of the simulation; and every activities' characteristics of the chain are determined. Finally, the outputs can be processed using MatSim for traffic assignment and visually represent the agents' journeys. VirtualBelgium's activity-based models rely on data extracted from the Mobel national mobility surveys conducted in Belgium (Hubert and Toint (2002)).

As travel behaviour of an individual changes over time, *i.e.* the set of feasible activity chains patterns evolves together with his socio-demographic characteristics. It is therefore interesting to have an evolution process for the population to forecast future travel demand, resulting in an integrated micro-simulator such as the ILUTE model.

The remainder of this paper is organized as follows. Section 2 describes VirtualBelgium's agents, data sources and activity chains possibly performed by the agents. In Section 3, we detail a method to assign an activity chain to an individual agent. We next present in Section 4 the preliminary results of the proposed methodology applied to VirtualBelgium. Dynamic evolution of the agents is discussed in Section 5. Concluding remarks and future perspectives are finally discussed in Section 6. Note that the simulator is still under development, and only preliminary results will be exposed.

2 VirtualBelgium: a multi-agent simulation for Belgium

VirtualBelgium is a research project aiming to simulate the mobility behaviours and evolution of the Belgian population using a multi-agents approach, based on the Repast HPC framework (Collier and North (2012)). The agents of interest consists of a population $\mathcal{P} = (\mathcal{I}, \mathcal{H})$ of approximatively 10.000.000 individuals $\in \mathcal{I}$ gathered in 4.350.000 households $\in \mathcal{H}$ and localized in 589 municipalities. As no disaggregate data fully describing it was available, a consistent synthetic population for Belgium was generated using a sample-free generator (see (Barthelemy and Toint 2012) for a detailed description of the algorithm). Agent's initial attributes, which significantly influence travel behaviour (Avery (2011), Hubert and Toint (2002), Cornelis et al. (2012)), are described in Tables 2.1 and 2.2. This work being focused on agent's mobility behaviours, their evolution processes will not be discussed in this paper.

Attribute	**Values**
Gender	male; female
Age class	0-5; 6-17; 18-39; 40-59; 60+
Age	an integer from 0 to 110
Activity	student; active; inactive
Education level	primary; high school; higher education; none
Driving license ownership	yes; no

Table 2.1: Individuals' characteristics

Attribute	Values
Type	single man alone single woman alone single man with children (and other adults) single woman with children (and other adults) couple without children (and other adults) couple with children (and other adults)
Number of children	0 to 5
Number of other adults	0 to 2 (mate not included)

Table 2.2: Households' characteristics

2.1 Data source, activity chains and general assumptions

Activity chains data used by VirtualBelgium is derived from the 2001 Mobel mobility survey conducted in Belgium. This survey highlighted 12 base activities:

1. pick up or drop someone
2. staying home
3. work related visit
4. work
5. school
6. eating outside
7. shopping
8. personal reason
9. visiting relatives
10. walk-around
11. leisure activity
12. other activity

Each activity is also characterized by a localization, i.e. a node of Belgium's road network, extracted from OpenStreetMap (Haklay and Weber 2008), and a duration.

An activity chain is then a sequence of these base activities. It is assumed that each activity chains begins and end at the individual's home. Moreover the total duration of an activity chains must be less than 24 hours. These assumptions seems fairly acceptable for a large majority of the population of interest. These concepts are formally described in definitions 2.1 and 2.2. A total number of 10.000 different activity chains patterns have been extracted from Mobel. Note that individual below 5 years old (included) are discarded as it is assumed that they always travel with their relatives and they don't have proper activity chains.

Definition 2.1 (Activity) *An activity α performed by an individual is a triplet $(\alpha^p, \alpha^l, \alpha^d)$ where*

- α^p = *the purpose ;*
- α^l = *the localization ;*
- α^d = *the duration ;*

of the activity.

Definition 2.2 (Activity chain) *An activity chain α of size n is a sequence $\alpha_1, \cdots, \alpha_n$ of activities such that*

$$\sum_i \alpha_i^d \leq 24 \ h \qquad \text{and} \qquad \alpha_1^l = \alpha_n^l.$$

3 Activity chains generation and assignment

This section presents a methodology assigning an activity chain to every individual of the simulation. We start by outlining its main steps before the more formal description.

Assume that a set of feasible activity-chains for each individual type is available. The affectation method consists of 2-steps procedure applied to every individual:

1. Random draw of an activity chain α from the appropriate set of activity-chains.

2. Computation of the localization and the duration $\forall\ a \in \alpha$.

The generation of activity-chains patterns for each individual type and the affectation steps are described and illustrated in the next subsections.

3.1 Preliminary Step : Generation of activity chains pattern

Assume that an individual is characterized by a vector of m attributes $V = (V_1, \ldots, V_m)$, whose components may take a discrete set of values and let's denote by $\mathcal{T_I}$, $\mathcal{A}_i$ and n_i the set of all individual type, the set of activity chain patterns that could be extracted from the data $\forall i \in \mathcal{T_I}$ and the size of $\mathcal{A}_i$ respectively.

Note that, depending on the data, for a subset $\mathcal{T_J} \subset \mathcal{T_I}$, n_j may be lower than a desired minimal number of activity chains $t\ \forall j \in \mathcal{T_J}$. The first step is then to add activity chains patterns to the associated $\mathcal{A}_j$ such that the constraint

$$n_j \geq t \qquad \forall j \in \mathcal{T_J} \tag{3.1}$$

yields.

Given an individual class j, the idea is to add to $\mathcal{A}_j$ the activity chains patterns of A_k, where k is as close as possible to j. Formally, the missing patterns are drawn from the ones associated with the individual type's belonging to $\mathcal{N}_j^l$ $(l = 1, \ldots, m)$, a l-neighbourhood of j generated by shifting l attributes' value of j. $\mathcal{N}_j^{l+1}$ is generated by modifying the same l attributes considered by $\mathcal{N}_j^k$ and an additional one.

Note that only shifts between two contiguous modality are allowed. For instance consider an attribute U whose modalities are u_1, u_2, u_3 and u_4, where

$$u_1 \leq u_2 \leq u_3 \leq u_4. \tag{3.2}$$

If an individual type is initially characterised by u_2, then the shift is either u_1 or u_3. Note that these shifts can only be applied to numerical or ordinal variables.

In the conducted experiments, the threshold t has been set to 5. As one can observe in Figure 3.1, out of 192 individual types, 116 problematic classes were identified in the raw data and at most a 3-neighbourhood was required to satisfy the constraint. The $\mathcal{N}_i^l$ $(l = 1, 2, 3)$ are generated by sequentially modifying the following attributes:

1. gender;
2. gender and age class;
3. gender, age class and education level.

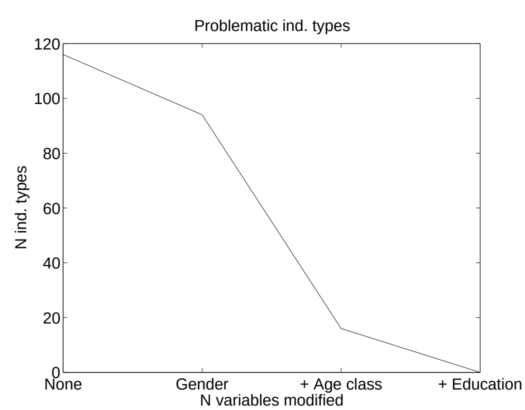

Figure 3.1: Numbers of problematic individual class with respect to the neighbourhood's level.

3.2 Activity chains assignments

A set of activity chains patterns being available for each individual type $i \in \mathcal{T}_{\mathcal{I}}$, this step assigns a chain to every individual agent of $\mathcal{I}$. This is done by randomly drawing an activity chain $\alpha \in \mathcal{A}_|$, where j is the type of the individual considered.

The random draws are weighted accordingly to the data extracted from Mobel. For instance, Table 3.3 illustrates the set $\mathcal{A}$ containing the feasible activity chains and their respective weights for a woman between 18 and 39 years old with a higher education degree, still studying and without a driving licence.

Pattern	**Weight**
2 5 10 2	0.272
2 2 2 9 2	1.025
2 5 2	0.913
2 5 2 10 2	0.412
2 5 6 5 2	0.412
2 11 2	0.284

Table 3.3: Weighted $\mathcal{A}$ for a given individual agent (Mobel).

3.3 Determination of an activity's localization

This subsection details how the localisation of an activity is determined inside the road network of Belgium. This network is extracted from OpenStreetMap (Haklay and Weber, 2008) and consists of roughly 262.000 nodes and 830.000 links. The reader can refer to Figure 3.2 for an illustration of the road network of the city of Namur.

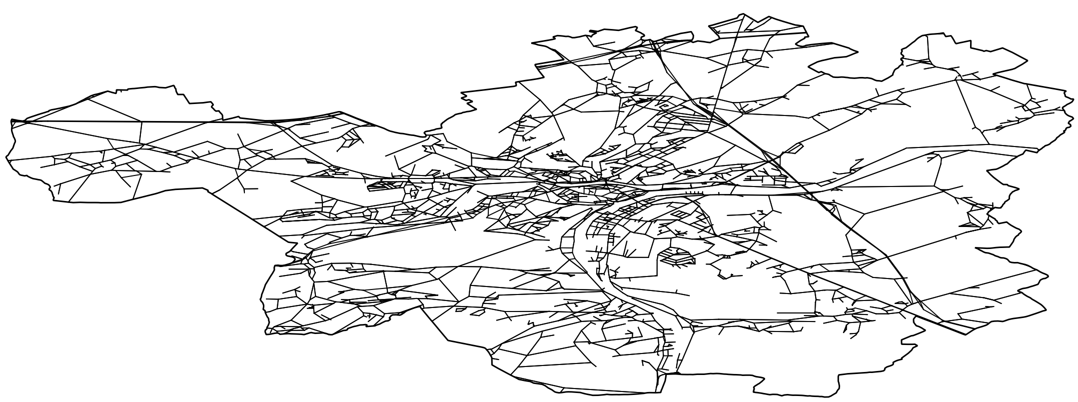

Figure 3.2: Road network of Namur.

3.3.1 Household's house localization

Each household and its constituent members are already located in one of the 589 municipality, *i.e.* at the LAU2 level (Barthelemy and Toint, 2012). Nevertheless, as the goal is to locate an activity at a network's node level, and no data at a more disaggregate level is available, the first part of the process consists of assigning each household to a node of their municipality's road network. The node, randomly drawn following a discrete uniform distribution (in order to preserve the population density), will be referred as the household's house.

3.3.2 Activity localization

Now that each individual has a house, it is possible to localize each of their activity in the network. These activities will also take place at a node of the network, which is determined as follow:

1. A distance d is randomly drawn;

2. A set of nodes at distance d is generated using a fast Dijkstra algorithm with a *minimum binary heap* data structure;

3. Finally a node is drawn from the set generated at previous step.

The remainder of this subsection will focus on steps one and three.

Since this simulation involves a huge number of agents, the number of randoms draws can be a critical bottleneck. The generation speed of random draws becomes an essential issue to address and requires an efficient approach. Two techniques were investigated: the first one draws from the empirical cumulative function while the second one consists of building a model that fits the data. The model is a mixture distribution $f(x)$ made of 3 independent components, *i.e.*

$$f(x) = \sum_{i=1}^{3} w_i p_i(x; \mu_i, \sigma_i)$$

where w_i is the weight associated with the component p_i such that $w_i \geq 0$ and $\sum_i w_i = 1$. The components considered in this work follow Log-Normal distribution of parameter (μ, σ) where μ is the mean and θ the standard error of the distribution. Experiments showed that the second method is up to 3 times faster than the first one. For a detailed description of such mixture distribution, see (McLachlan and Peel 2000). It is important to note that each mixture distribution fitted to the data in the next subsection is statistically similar to the empirical distribution according the well-known Kolmogorov-Smirnov goodness-of-fit test (Massey (1951)).

3.4 Computation of an activity's duration

We now provide some comments on the determination of an activity's duration, for which random draws from mixture distribution are also applied.

3.4.1 House departure time

The first step taken by an agent is to leave its home in order to perform the first activity of the day. Regarding the activity type to be performed, the time departure distribution vary (see Figure 3.4), and the departure time is randomly draw accordingly to the corresponding distribution.

3.4.2 Duration

The same process used for the computation of an activity's duration is applied. Indeed, the duration of a particular activity is randomly draw accordingly to its type from the corresponding distribution (see Figure 3.5).

4 Application on VirtualBelgium: results

Up to now, only basic experiments have been conducted in order to validate the code of VirtualBelgium. The activity-based model have been successfully applied to the synthetic population of Namur, and the results have been processed with Matsim. The simulation involved 100.000 agents and the running time was less than 1 hour using a cluster of four Six-Core Intel Xeon X5650 and 24 GB of RAM. Figure 4.6 illustrate a snapshot of the beginning of the morning peak on the Namur's road network.

Simulation runs with the whole synthetic population and the complete Belgian road network are still under test, but the first experiments are promising.

5 Agents' dynamic evolution

As mentioned earlier, the set of feasible activity chain pattern for a given individual evolves together with his/her socio-demographic characteristics. Therefore it is interesting to implement an evolution process for the population of interest in order to forecast the travel demand in the future as well as the socio-demographic evolution of Belgium. For instance, an individual agent should have the capability to get older, to give birth to new agents, to die, to move out, to find a mate, to divorce, etc... The goal is to obtain a cycle of life as illustrated in Figure 5.7.

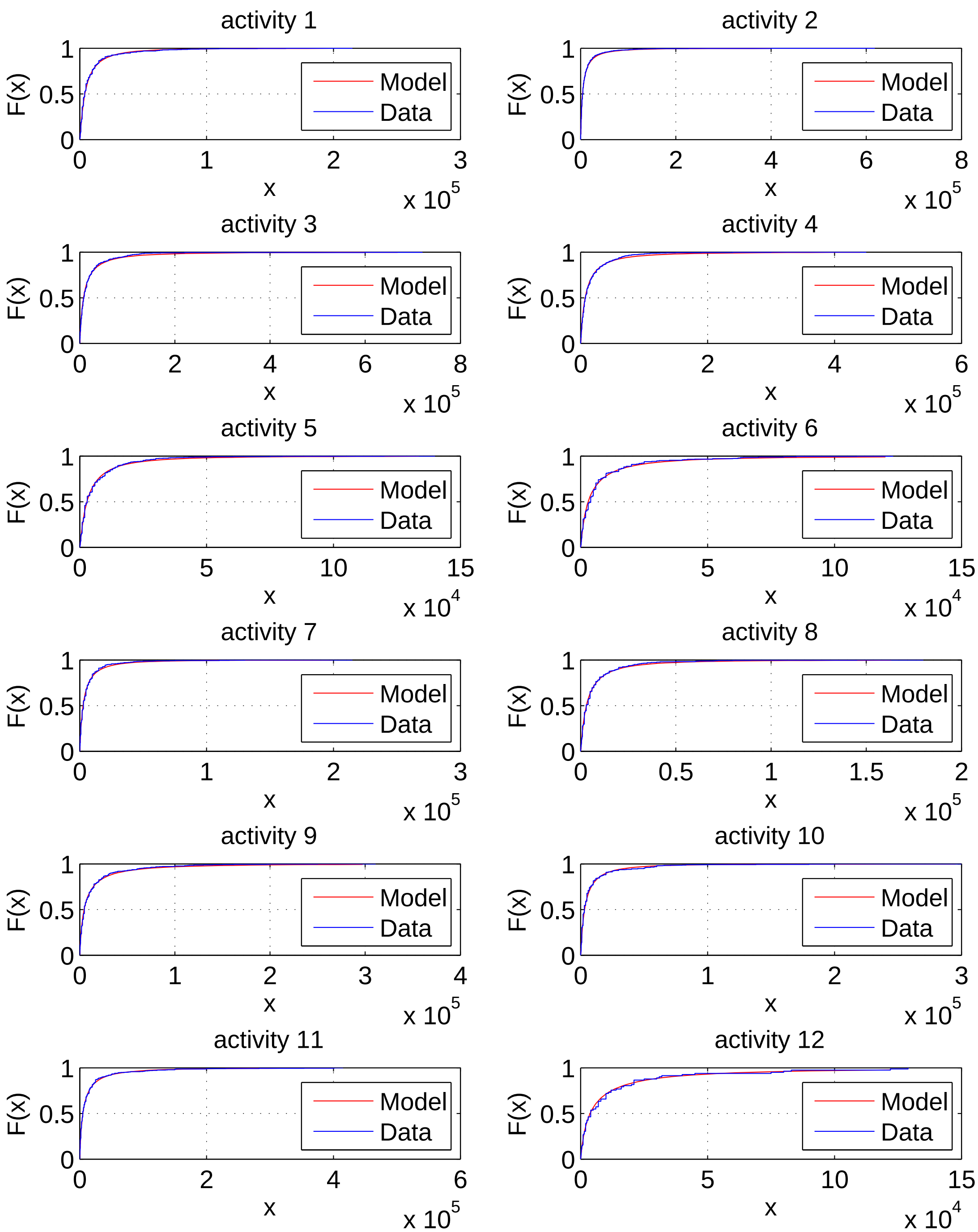

Figure 3.3: Distance: empirical and estimated cumulative distribution functions by activity type (x unit : meters).

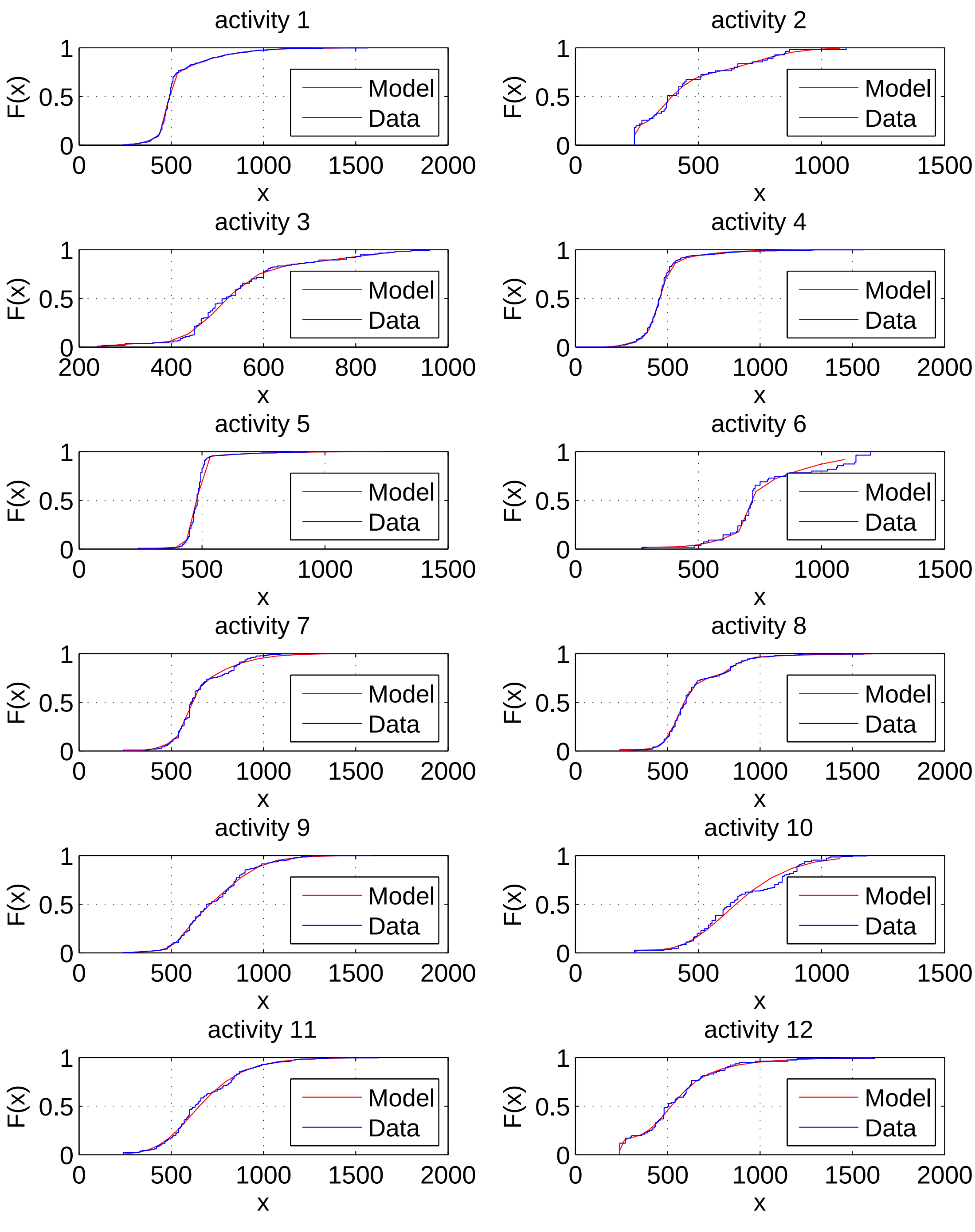

Figure 3.4: House departure time: empirical and estimated cumulative distribution functions by first activity type (x unit : minutes after midnight).

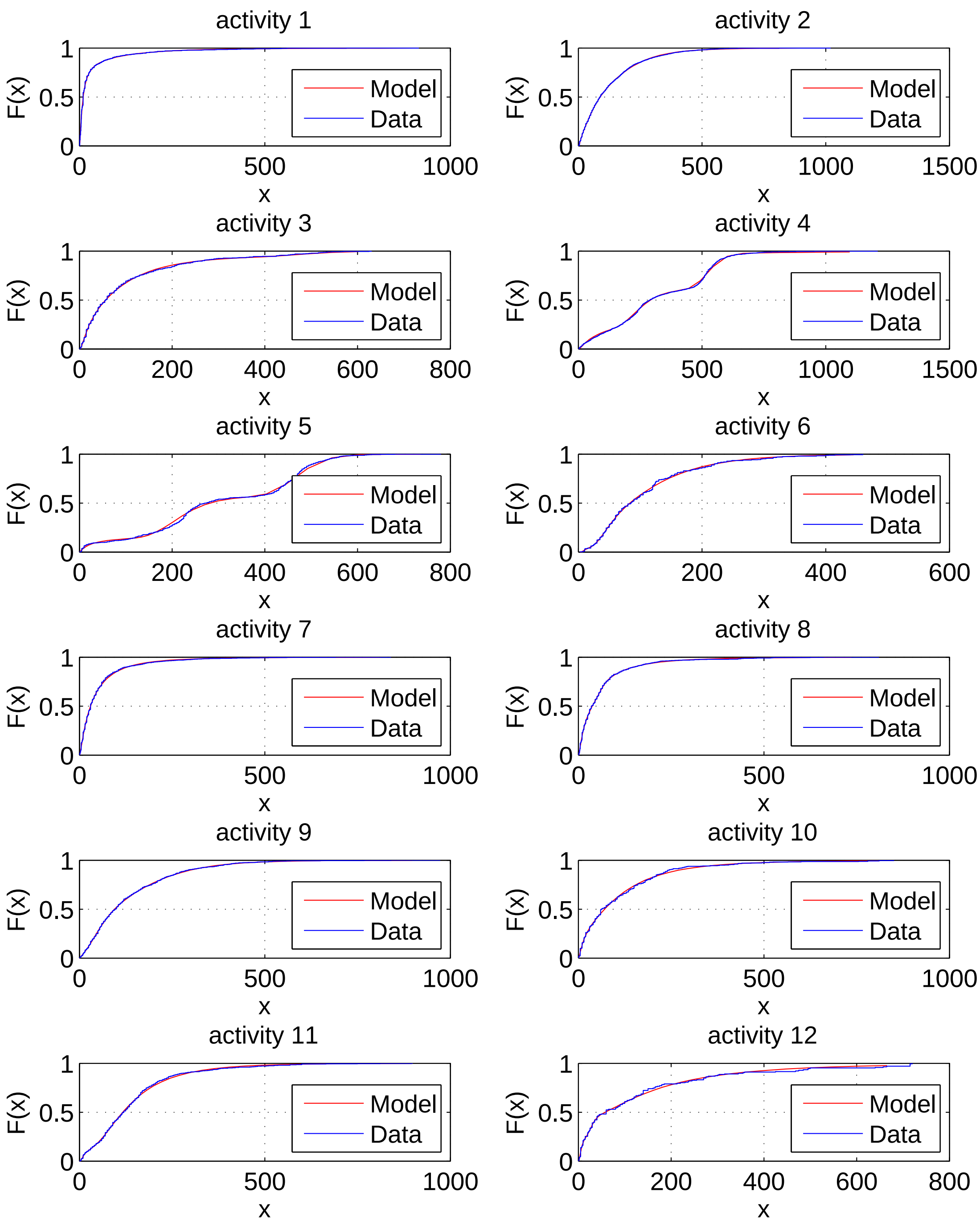

Figure 3.5: Activity duration: empirical and estimated cumulative distribution functions by activity type (x unit: minutes).

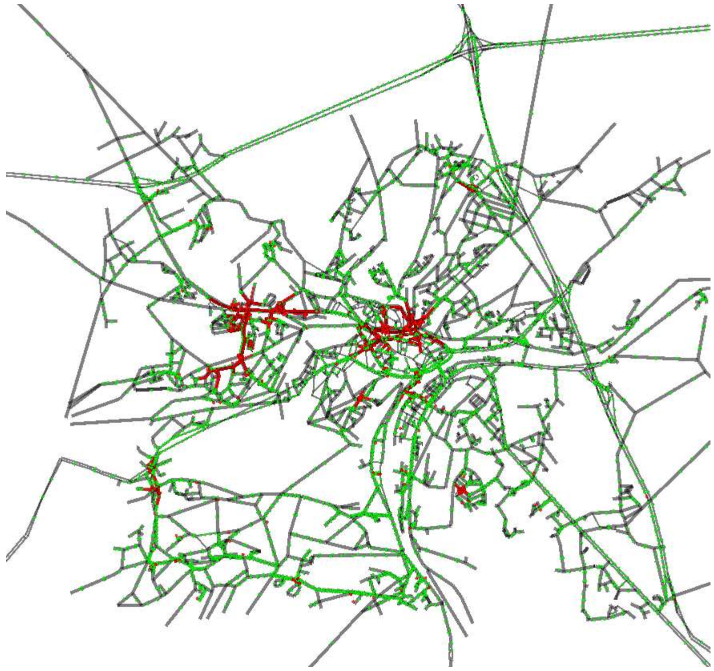

Figure 4.6: Snapshot of a Matsim output. Red agents are stuck in a traffic jam.

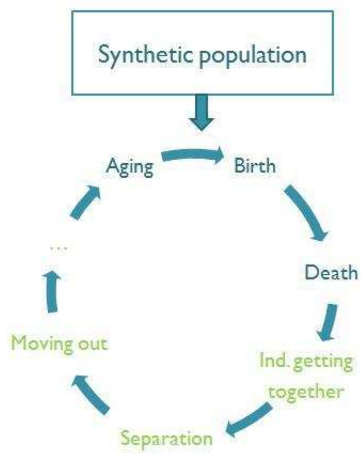

Figure 5.7: Life cycle (blue: already implemented, green: to be implemented).

Only simple models for ageing, dying and giving birth have been implemented up to now in VirtualBelgium. Note that the death is processed after the births in order to capture infant mortality. The data used by these models originate from the *Directorate-general Statistics and Economic information* of the Belgian Government and are provided for the base year 2001. Preliminary results for the city of Namur are shown in figures 5.8 to 5.11. One can easily observe that the total number of agents is decreasing. One possible explanation is that the immigration is not taken into account. Moreover, we only considered women being head of household or mate of a household head as potential mother: as we did not have taken yet into account a moving out process, the number of potential mother is decreasing, which causes a decrease of births.

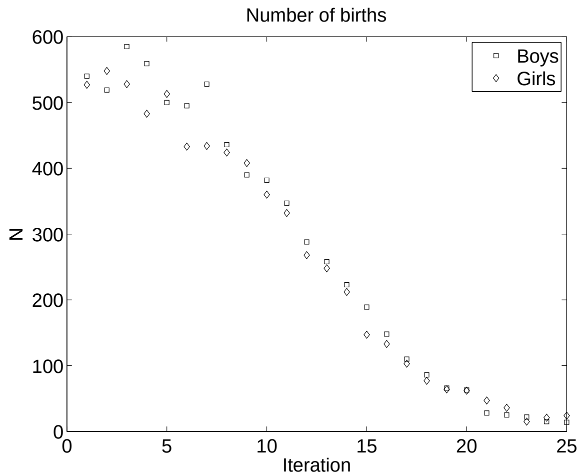

Figure 5.8: Evolution of the number of births (2001-2026, city of Namur).

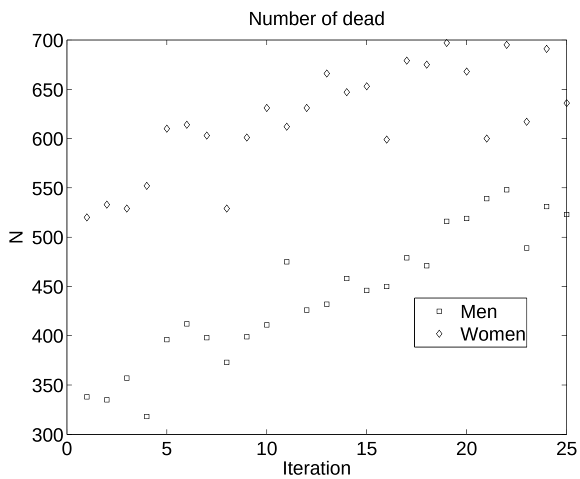

Figure 5.9: Evolution of the number of deaths (2001-2026, city of Namur).

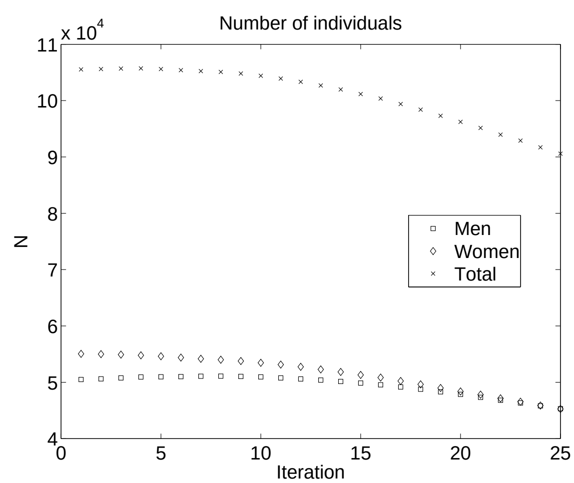

Figure 5.10: Evolution of the total number of agents (2001-2026, city of Namur).

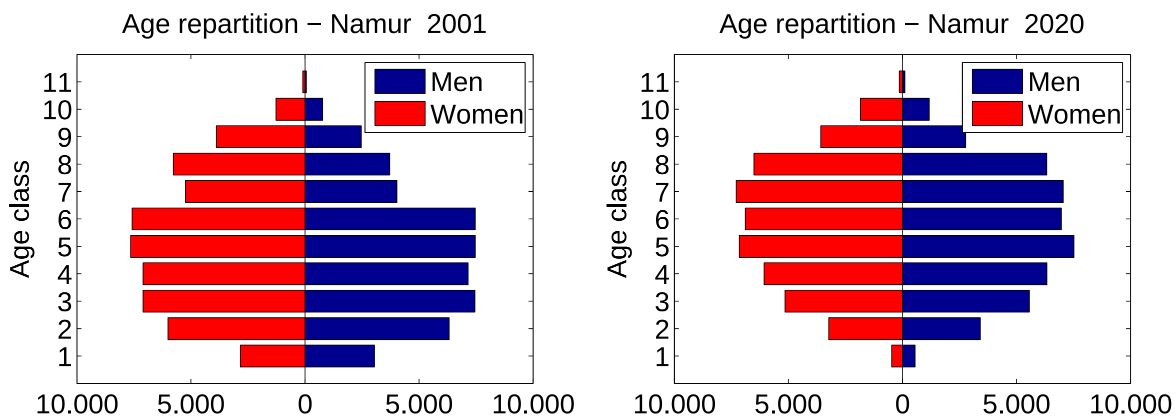

Figure 5.11: 2001 (left) and 2006 (right) age pyramid for the city of Namur. Age classes are 0-5; 6-15; 16-25; 26-35; 36-45; 46-55; 56-65; 66-75; 76-85; 86-95; 96+.

6 Conclusions and future work

This paper detailed the current status of VirtualBelgium, a large agent-based micro-simulation designed to replicate the mobility behaviour of the Belgian population and its evolution. At the moment, only simple models are implemented and act as a proof of concept, showing that simulating more than 10.000.000 agents is nowadays feasible.

Each agent have been assigned a sequence of activities, which are spatially and temporally localised. Nevertheless the model still requires major improvement in order to become fully operational. In the short term, the following improvement will be considered :

- trying other implementation of the Dijkstra algorithm using Fibonacci and n-ary heaps to improve the execution speed;
- investigating mode choice (public transportation, car, walking);
- adding more data for destination choice (job and service indicators, number of schools, ...).

As for the population socio-demographic evolution, enhancement are required as well. The first step will be to model the wedding and divorce processes, which is not a trivial work as it implies individual and household agents interacting with each others. Our efforts will then focus on the departure of individuals from their initial household. Finally, residential mobility will be implemented following the MOBLOC model already developed by Pauly, Cornelis and Walle (2011).

Acknowledgments

The authors wish to thanks Frédéric Wautelet for its help to set up the simulation on a high performance computing facility. Numerical simulations were made on the local computing resources at the Namur Center for Complex Systems (NAXYS).

References

T. Adler and M. Ben-Akiva. A theoretical and empirical model of trip chaining behavior. *Transportation Research B*, **13**(3), 477–500, 1979.

T. A. Arentze and H. J. P. Timmermans. Albatross: A learning-based transportation oriented simulation system. Technical report, European Institute of Retailing and Services Studies. Eindhoven, The Netherlands, 2000.

L. Avery. *National Travel Survey: 2010*. National Travel Survey. Department for Transport, 2011.

J. Barthelemy and Ph. L. Toint. Synthetic population generation without a sample. *Transportation Science*, 2012.

C. R. Bhat and F. S. Koppelman. Activity-based modeling of travel demand. *in* R. W. Hall, ed., 'Handbook of Transportation Science', pp. 35–61, Dordrecht, The Netherlands, 1999. Kluwer Academic Publishers.

F.S. Chapin. *Human activity patterns in the city: Thing people do in time and space*, Vol. 13 of *Wiley series in urban research*. J. Wiley and Sons, 1974.

N.T. Collier and M. North. Parallel agent-based simulation with repast for high performance computing. *SIMULATION*, 2012.

E. Cornelis, M. Hubert, Ph. Huynen, K. Lebrun, G. Patriarche, A. De Witte, L. Creemers, K. Declercq, D. Janseens, M. Castaigne, L. Hollaert, and F. Walle. La mobilité en belgique en 2010 : résultats de l'enquête beldam. Technical report, SPF Mobilité et Transports and BELSPO, Brussels, Belgium, 2012.

J.W. Dickey. *Metropolitan Transportation Planning*. McGraw-Hill, New York, USA, 2 edn, 1983.

T. Domencich and D. McFadden. *Urban Travel Demand: A Behavioural Analysis*. North Holland, Amsterdam, The Netherlands, 1975.

L.P. Gan and W. Recker. A mathematical programming formulation of the household activity rescheduling problem. *Transportation Research Part B: Methodological*, **42**(6), 571 – 606, 2008.

T. F. Golob. Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, **37**(1), 1 – 25, 2003.

J. Goran. Activity based travel demand modelling - a literature study. Technical report, Danmarks Transport-Forskning, 2001.

T. Hägerstrand. What about people in regional science. *Papers of the Regional Science*, **4**(1), 6–21, 1970.

M. M. Haklay and P. Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, **7**(4), 12–18, 2008.

J.-P. Hubert and Ph. L. Toint. *La mobilité quotidienne des Belges*. Number 1 *in* 'Mobilité et Transports'. Presses Universitaires de Namur, Namur, Belgium, 2002.

R. Kitamura, C. Chen, and R. M. Pendyala. Generation of synthetic daily activity-travel patterns. *Transportation Research Record*, (1607), 154–162, 1997.

R. Kitamura, E. Pas, C. Lula, T. k. Lawton, and P. Benson. The sequenced activity mobility simulator (SAMS): an integrated approac to modelling transportation, land use and air quality. *Transportation*, **23**, 267–291, 1996.

F. J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, **46**(253), 68–78, 1951.

G. McLachlan and D. Peel. *Finite Mixture Models*. J. Wiley and Sons, Chichester, England, 2000.

K. Meister, M. Balmer, F. Ciari, A. Horni, M. Rieser, R. A. Waraich, and K.W. Axhausen. Large-scale agent-based travel demand optimization applied to switzerland, including mode choice. paper presented at the 12th World Conference on Transportation Research, July 2010.

E. Miller. Microsimulation and activity-based forecasting. *in* T. T. Institute, ed., 'Activity-Based Travel Forecasting Conference: Recommendations, and Compendium of Papers', pp. 151—-172. Travel Model Improvement Program, US Department of Transportation, US Environmental Protection Agency, June 1996.

N. Oppenheim. *Urban Travel: From Individual Choices to General Equilibrium*. J. Wiley and Sons, New York, 1995.

X. Pauly, E. Cornelis, and F. Walle. Mobloc: links between residential mobility and accessibility. *in* 'Proceedings of the BIVEC-GIBET Transport Research Day 2011', pp. 126–137, 2011.

P. Salvini and E. J. Miller. Ilute: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, **5**, 217–234, 2005.

B.D. Spear. *Application of new travel demand forecasting techniques to transportation planning: a study of individual choice models*. Dept. of Transportation, Federal Highway Administration, Office of Highway Planning, Urban Planning Division, 1977.

P. Waddell. Urbansim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of the American Planning Association*, **3**(3), 297–314, 2002.