Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Modelling complex households and communities—Example of Demo4: Demographic module of the SustainCity project

Sophie Pennec, Élisabeth Morand and Laurent Toulemon[1]

Abstract

As part of the European SustainCity project, a microsimulation model of individuals and households was created to simulate the population of various European cities. The aim of the project was to combine several transportation and land-use microsimulation models (land-use modelling), add on a dynamic population module and apply these microsimulation approaches to three geographic areas of Europe (the Île-de-France region and the Brussels and Zurich agglomerations).

Keywords: Dynamic microsimulation, population

1. Introduction

1.1 The SustainCity project

The aim of the European SustainCity project (www.sustaincity.org) is to combine several transportation and land-use microsimulation models (land-use modelling), such as UrbanSim, MATSim and METROPOLIS; add on a dynamic population module to create UrbanSimE, adapted to European cities; and apply these microsimulation approaches to three geographic areas of Europe (the Île-de-France region and the Brussels and Zurich agglomerations).

The presentation deals with Demo4, the demographic model. The model was conceived as an autonomous dynamic module within the SustainCity project and can therefore be used for strictly demographic applications.

1.2 Microsimulations

Unlike more traditional simulation models, microsimulation models are characterized primarily by the fact that they operate at the individual level and not at an aggregate level. Therefore, the macro-level results arise from the sum of the behaviours of the micro-individuals. These micro-behaviours are based on probabilities of transition from one state to another, as dictated by individual characteristics and by random selection that determines whether events occur or not.

In the late 1950s, to avoid the errors associated with simulation models that use aggregates and to better reflect the complexity of reality, Orcutt (1957) proposed a form of modelling of "various sorts of interacting units which receive inputs and generate outputs," inputs being "anything which enters into, acts upon, or is taken into account of, by the unit" and outputs "anything which stems from, or is generated by, the unit." His main goal was to better simulate social policy variants and, in particular, the diversity of individual circumstances. With this type of model, it is possible to determine whether policies are actually having an impact on the people for whom they are intended, whether there are any counterproductive effects (for example, a benefit that results in the loss of other benefits or increases deductions and unintentionally worsens the circumstances of some people), and who the winners and

1. Sophie Pennec, Institut national d'études démographiques (INED), 133 boulevard Davout, 75980 Paris Cedex 20, France, pennec@ined.fr; Élisabeth Morand, INED, elisabeth.morand@ined.fr; Laurent Toulemon, INED, toulemon@ined.fr.

losers from a policy change would be. This form of modelling enables us to produce results in the form of a distribution rather than an average.

The same logic applies when we look at the distribution of households and individuals for tracking purposes. We need to know not only the average characteristics of households but also their distribution for various characteristics inferred from those of household members.

This paper provides a description of the demographic model constructed as part of the SustainCity project—in particular, it describes the choices we made to address the constraints of the project. We also describe the simulation of collective households and complex households, two types of households that are often ignored in models.

2. The demographic model

2.1 Characteristics of the model

The demographic model for the SustainCity project was developed under a number of constraints: it had to be compatible with the UrbanSim land-use model; it had to be simple enough that it could be used by the other teams in the project and flexible enough to accommodate constraints in terms of data from the other partners in the Zurich and Brussels agglomerations, after calibration on the Paris region of Île-de-France; and it had to be free and based on open-source software.

We wanted the model to be independent of the UrbanSim platform, which is written in the Python programming language. There were two reasons for this choice: it made it easier to test and modify the model, and we wanted to be able to use the model subsequently for purposes unrelated to UrbanSim.

We selected the Modgen language developed by Statistics Canada because it provides a stable, regularly updated programming environment for microsimulation models. It allows us to create a user-friendly interface, which we felt was very convenient; project partners could use the model, including inputting the initial simulation data and changing the simulation parameters, without having any knowledge of computers.

The demographic model for SustainCity:

- Is a cross-sectional model. It uses an initial population which is representative of the population being studied and makes this sample evolve over time so as to have population states at various points in time.
- Is a closed model. The size and structure of the population change only as a result of events that affect members of the population. People enter the population through birth or immigration and leave it through death or emigration. In addition, relationships between household members are explicitly created and preserved during the simulation. As a result, the formation of unions takes place within the population, and a sort of marriage market has to be created to form couples.
- Is a discrete-step model. The simulation step is annual, and, therefore, we use annual transitions to simulate behaviours.
- Is a model that simulates both individuals and households. Changes in the structure of a household result from events affecting household members. Departures from existing households generate new households (e.g., children who leave the parental home, or separation of partners), and the disappearance of a household follows the departure of all of its members to another household through unions, or the death of a person living alone.

2.2 Simulated events

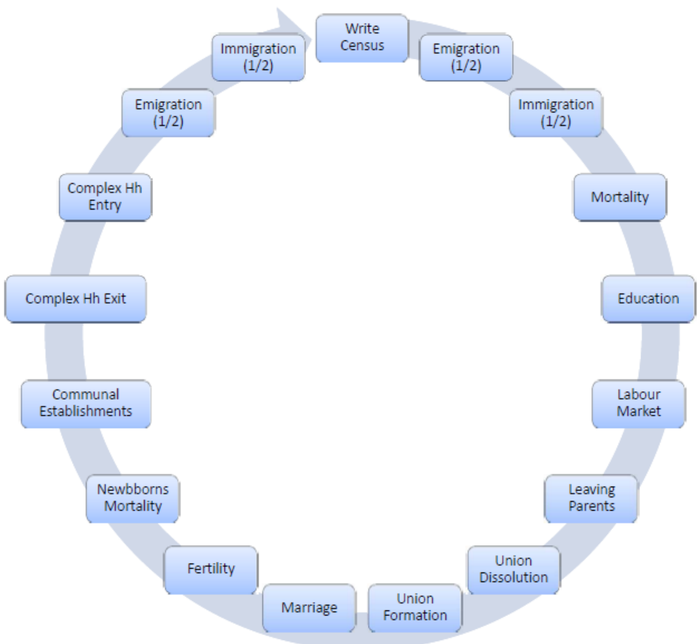
The model simulates two types of actors: persons and households. Each person in the sample belongs to only one household (except persons who live in a collective household, who are not associated with any specific household). The model creates three forms of relationships between persons: partner, mother and father. It defines a household reference person. Population aging—that is, all the changes that each person or household can undergo—depends on annual discrete-step transitions. Age is the only variable that evolves deterministically. Probabilities vary depending on various personal characteristics, which for some events are limited to sex and age. Some characteristics vary over time and therefore are estimated for every year of the simulation on the basis of known changes. Users of the model can choose to keep the probabilities constant over time or incorporate estimates for every year. In addition, some events depend not only on the probability of experiencing them but also on events affecting other people (widowhood results from the death of a partner) or on characteristics of other people (as when two people form a union).

Even though microsimulation models use Markov chains and transitions based on the current situation, this potential limitation is generally overcome by using biographical characteristics and variables containing duration information. Hence, this type of model takes into account the past behaviour of people and some of the variables of interest (for example, the probability of having a child depends on how many children a woman has had previously and on the age of the last child, and this makes it possible to control the length of time between births).

The choice of simulated events reflects two constraints. One is having the necessary variables to simulate the demographic evolution of people and households; the other has to do with the availability of data for all regions to be simulated in the SustainCity project.

Two variables take account of social heterogeneity, and each one is divided into three classes: level of education (low [elementary], medium [secondary] and high [university]) and place of birth (in the region, in another region of the country, and in another country). For migrants, we can use length of time since arrival as a complementary variable. Labour-market participation is binary: in the labour market (active or unemployed), or not in the labour market (student, inactive or retired).

Figure 2.2-1
Demographic model: Order of simulation of events



Some events involving more than one person are women-driven in this model; this is the case for the formation of unions and for fertility. For union formation, for example, we determine which women, among those who are not in a union, will form a union (using a probability of union formation that varies by age and status in the household, and a stochastic selection to determine which of the eligible women will actually form a union). Then, we create a marriage market to match these women with a husband from the set of men who are not in a union, using constrained optimization techniques. We select a set of potential husbands, calculate the similarity of the women with each of these potential husbands in terms of differences in age, level of education and place of birth, and we select a husband at random on the basis of the similarities.

Migrations take place at the household level; all household members immigrate or emigrate together. Immigrations reflect the observed profile of recent migrants in the region.

Events are simulated in the order shown in Figure 2.2-1. We developed this order to reduce the effect of concurrent events. This limitation, associated with the use of discrete transitions, is mitigated by carefully choosing the order of events. For example, we simulate union dissolutions first, then union formations; this means that people who experience the dissolution of a union in a particular year can form another union the same year, whereas the dissolution of a union can occur only in the year following the formation of that union.

3. Complex households and collective households

In addition to simple households made up of one family (a couple or an adult with or without children), the model also includes complex households and atypical households.

3.1 Complex households

A household is said to be 'complex' if it does not consist exclusively of a nuclear family. This includes all of the different modes of cohabitation outside the nuclear family or in addition to the nuclear family: multi-generation households, households consisting of a family and another related or unrelated person (an elderly parent, another family member, or a renter), households composed of unrelated people sharing rental accommodations, and so on.

In the model, we consider a complex household to be

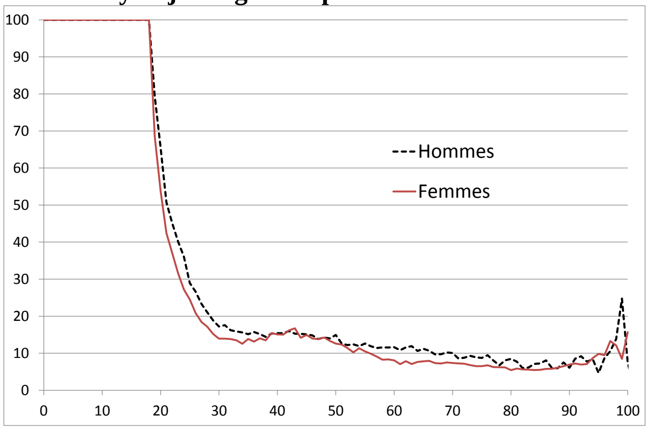
- a household consisting of isolated people with no family ties
- a household consisting of a nuclear family and other people who may or may not be related to some members of the nuclear family (a parent, a nephew or a student renting a room as a boarder).

To simplify the simulation, we decided that there could be only one couple in a household. Hence, there are no cases in which two parents are living with one of their children who is also in a couple. This is not such a strong assumption, since very few households consist of two couples or even two nuclear families.

To simulate complex households, we assigned to every person living alone a probability of joining an existing household (Figure 3.1-1). This probability is derived from an estimate by age and sex based on the census of population data. The probability is increased for people with children in the region being simulated. Thus, we obtain a set of people likely to join another household. Matching with an existing household is carried out in much the same way as matching between spouses. We select 50 potential households at random, and calculate a similarity indicator based on the age difference between the lone person and the household reference person (two types of differences are considered preferable: less than 5 years and between 25 and 35 years, to simulate young people sharing rental accommodations and multi-generation households, respectively) and the size of the 'recipient' household. We select the pair consisting of the person living alone and the recipient household with the highest similarity index.

The probability of leaving a complex household is estimated at 0.5, but it is possible to rejoin a complex household in the same year and thus remain in a complex household. A person who leaves a complex household forms a new one-person household.

Figure 3.1-1
Probability of joining a complex household



3.2 Collective households

The French national statistical office (INSEE) defines a community, or a collective household, as a complex of residential premises that falls under the same managing authority and whose residents usually share a common mode of living. This covers retirement homes, forensic facilities, boarding schools, university residences, hospitals, penitentiaries, religious communities and military establishments.

In 2011, 2.5% of the people enumerated in the census, 1.3 million people, were living in communities. About 30% of them were living in retirement homes, 20% in medical–social facilities, 30% in boarding schools or university residences, 8% in workers' hostels, 4% in military establishments, 4% in penitentiaries, and 2% in religious communities or some other type of community.

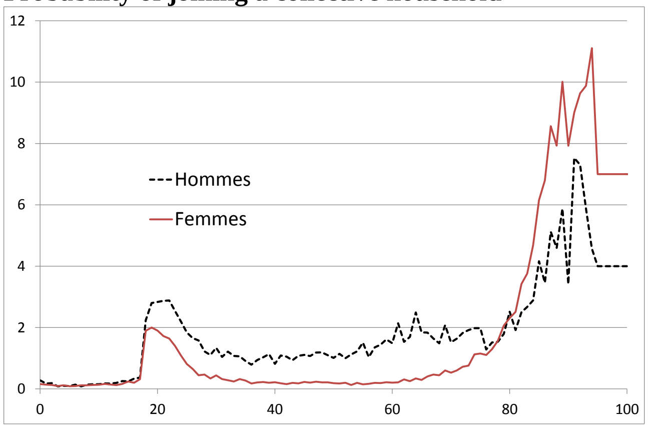
The model treats all forms of collective households as one household type. Age is a strong determinant of the type of collective dwelling (retirement homes for the elderly, boarding schools for children, university residences for young adults). People living in collective households are not attached to any particular household, and joining a collective household destroys the connection with the previous household. For the sake of simplicity, we deemed people in collective households to be without a partner (if a person was in a union before joining a collective household, the union is considered dissolved).

The probability of joining a collective household is estimated using census data. On the basis of an age and sex profile, a person who is in a couple has a lower probability of joining (multiplied by 0.3), while a person who is not in a couple has double the probability. The probability does not depend on the level of education or on whether a person is a student or not.

People in collective households can die, change their level of education, change their labour-market status and leave the collective household. They cannot form a union with a partner, have children, emigrate or directly join a complex household.

As in the case of complex households, the probability of leaving a collective household is estimated at 0.5, and a person can leave and rejoin a collective household in the same year, which amounts to remaining in a collective household. After age 90, the probability of leaving declines—at higher ages, remaining in a collective household is particularly frequent, and the rate of joining a collective household is also very high. A person who leaves a collective household lives alone in a new one-person household.

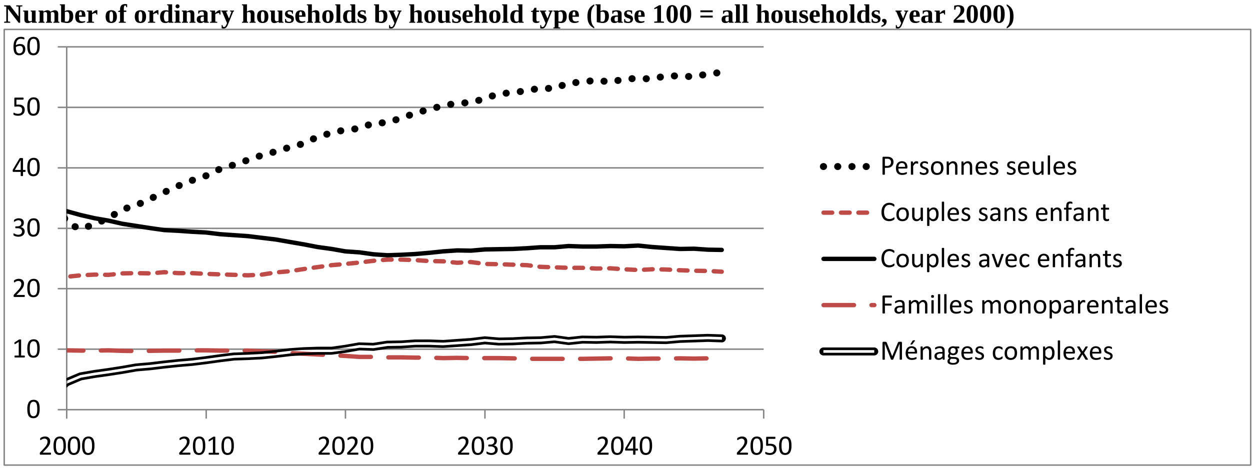
Figure 3.2-1
Probability of joining a collective household



4. Some results

Transition probabilities are mostly estimated from the survey "Étude de l'histoire familiale" which was conducted as part of the 1999 population census, and which contains a wealth of retrospective information on a large sample (Cassan, Héran and Toulemon 2000; Lefèvre and Filhon 2005). The first result from the model relates to the distribution of households by size and type. In the case of the Île-de-France region, the number of households is growing faster than the population, because of the decline in average household size. One-person households show the largest increase (Figure 4-1): their number almost doubles between 2000 and 2050. A second category of households, complex households, increases in number, while family households (couples with or without children and lone-parent households) remain virtually unchanged.

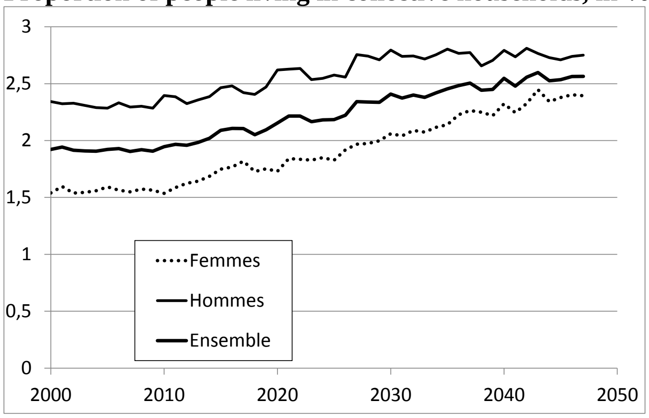
Figure 4-1
Number of ordinary households by household type (base 100 = all households, year 2000)



The increase in the number of complex households is mainly attributable to the growing propensity of young people to share rental accommodations after leaving their parents' home and before living as a couple. Hence, complex households are usually composed of people who share a dwelling but who report that they are not living as a couple. The increase is particularly pronounced at the beginning of the period, because of the delay in forming the first union.

The change in the number of collective households has another explanation (Figure 4-2). It is primarily population aging that accounts for the increasing proportion of people living in collective dwellings. The proportion of the population living in collective households increases throughout the simulation period, and the increase is particularly pronounced for women.

Figure 4-2
Proportion of people living in collective households, in %, by year



5. Conclusion

The results of the demographic module made it possible for the SustainCity project to support the projections of households in European cities with a microsimulation of individual behaviours (Turci et al. 2010; Turci et al. 2011; Turci et al. 2012). The collection of detailed data on people living in collective households will validate the observed changes. The demographic module is very sophisticated and performs detailed modelling of the demographic behaviours underlying the changes in the number and structure of households.

A future step will be to estimate the probability of leaving complex households and collective households. Current estimates that are based on very high exit probabilities lead to consistent cross-sectional estimates, and they are useful for estimating entry probabilities based on census data. However, they produce individual histories with too many entries and exits. The annual census data (Lefranc 2010), which, since 2011, contain information about residence one year before the census, will be very useful for estimating the movements of individuals between households more precisely.

References

Cassan, F., F. Héran, and L. Toulemon. 2000. “Study of family history. France’s 1999 family survey.” Annual issue, *Courrier des statistiques, English series* 6: 7–19.

Lefèvre, C., and A. Filhon, eds. 2005. *Histoires de familles, histoires familiales. Les résultats de l’enquête Famille de 1999*. Les Cahiers de l’INED, no. 156. Paris: INED.

Lefranc, C. 2010. “La diffusion des résultats du recensement rénové de la population.” *Courrier des statistiques* 129: 1–6.

Turci, L., A. Bringé, E. Morand, S. Pennec, L. Toulemon, R. Baggio, and F. Billari. 2010. *Provisional Demographic Outline*. SustainCity Working Paper no. 4.1. Paris: INED. Available at: http://www.sustaincity.org/publications.

Turci, L., S. Pennec, L. Toulemon, A. Bringé, and E. Morand. 2011. *Demo4: SustainCity’s Demographic Model*. Paper presented at the CEPS/INSTEAD, MiDal Final Conference, Esch-sur-Alzette, November 28 to 30, 2011.

Turci, L., S. Pennec, L. Toulemon, A. Bringé, E. Morand, and R. Baggio. 2012. *Demographic Model User Guide*. SustainCity Working Paper no. 4.4. Paris: INED. Available at: http://www.sustaincity.org/publications.