



EUROPEAN COMMISSION
European Research Area



Funded under Socio-economic Sciences & Humanities

Working paper 4.4

Demographic Model User Guide

Lorenzo Turci
Sophie Pennec
Laurent Toulemon
Arnaud Bringé
Elisabeth Morand
Rodolfo Baggio

INED

FP7-244557

Revision: 0

27/04/2012



Contents

- CHAPTER 1. OPERATING SYSTEM AND SOFTWARE REQUIREMENTS 7**

- CHAPTER 2. QUICK START (HOW TO RUN THE MODEL WITH THE DEFAULT SCENARIO) 9**

- CHAPTER 3. THEORETICAL DESCRIPTION OF THE MODEL 13**

- CHAPTER 4. INPUT FILES..... 16**
 - 4.1. Description of the initial population 16
 - 4.2. Links, IDs and imputations 17

- CHAPTER 5. PARAMETERS..... 18**
 - 5.1. Graphical user interface for the input parameters: an easy way to modify a unique parameter..... 18
 - 5.2. Modify a complete matrix of parameters : How to create a parameters .dat file..... 20

- CHAPTER 6. SCENARIO SETTINGS..... 24**

- CHAPTER 7. MODEL RUN AND OUTPUTS 28**
 - 7.1. Tables..... 32
 - i. Stocks 32
 - ii. Flows 32
 - 7.2. Output files 33

- CHAPTER 8. FURTHER DOCUMENTATION..... 36**

- ANNEX I. VARIABLES DESCRIPTION 38**
 - A. Qualitative variables used as parameter dimensions..... 38
 - B. Dimensions of quantitative variables used as parameter..... 40
 - C. Partitions of quantitative variables used as parameter dimensions 40

- ANNEX II. PARAMETER ESTIMATION..... 41**
 - A. Population 41

1. Initial population	41
2. Immigrants pool (from within France)	43
3. Immigrants pool (from abroad)	45
B. Estimation of transition probabilities and other parameters	45
1. Mortality	45
2. Fertility	45
3. Communal establishments.....	46
4. Complex Households.....	46
5. Emigration.....	46
6. Immigration.....	46
7. Labour market.....	47
8. Leaving the parental household.....	47
9. Union dissolution	47
10. Union formation.....	47
11. Marriage.....	47
 ANNEX III. MODULES AND PARAMETERS	 49
A. Order of the events in the simulation.....	49
B. Communal residence [BASE(Communal).DAT].....	49
C. Entry in complex household [BASE(ComplexEntry).DAT]	50
1. How to enter a complex household	50
2. List of parameters	50
D. Education [BASE(EDUCATION).DAT]	51
1. How Level of Education is Imputed.....	51
2. List of parameters	55
E. Emigration [BASE(EMIGRATION).DAT]	56
1. Principle of emigration computation	56
2. List of parameters	56
F. Fertility[BASE(FERTILITY).DAT].....	56
1. Principle of Fertility computation	56
2. List of parameters	56
G. Reference person [BASE(Hh_RefPerson).DAT].....	57
1. Definition	57
2. List of parameters	57
H. Initialization [BASE(INITIALIZATION).DAT].....	57
I. Immigration [BASE(IMMIGRATION).DAT]	57
1. Principle of Immigration computation.....	57
2. List of parameters:	58
J. Labour market [BASE(Labour).DAT].....	58

K. Leaving the parental home [Leaving]	58
L. Marriage[BASE(MARRIAGE).DAT]	58
M. Mortality [BASE(MORTALITY).DAT]	59
N. Output [BASE(OUTPUT).DAT]	59
O. Population sizes [BASE(PopulationSizes).DAT]	60
P. Union dissolution [BASE(UnionDissolution).DAT]	60
Q. Union formation [BASE(UnionFormation).DAT]	60
1. Principle of Union Formation	60
2. List of parameters	61
ANNEX IV. LIST OF GENERAL SCENARIO SETTINGS	62
LIST OF TABLES	63
LIST OF FIGURES	64

CHAPTER 1.

OPERATING SYSTEM AND SOFTWARE REQUIREMENTS

The simulation model requires machines running with Windows operating system, XP or higher. Users will be given a zipped archive containing two folders:

- *Model_and_parameters*, containing:
 - the model executable file, `Demo4.exe`;
 - the parameters (`.dat` files);
 - three `.csv` files containing the initial population and the two pools of potential immigrants;
 - the default scenario file `Base.scex`.
- *User_guide*, containing this guide.

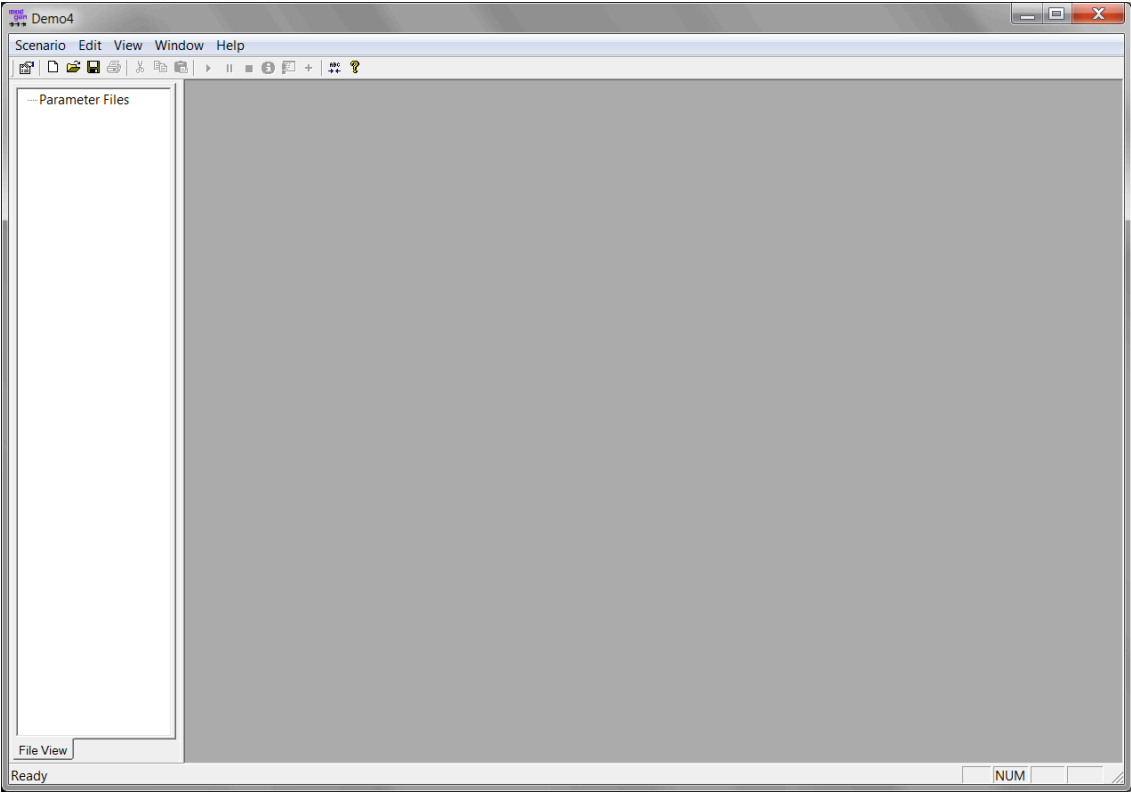
Users should also download, and install before running the model, the ModGen prerequisites. The prerequisites installer can be downloaded from <http://www.statcan.gc.ca/microsimulation/exe/modgen11-prereq.exe>.

CHAPTER 2.

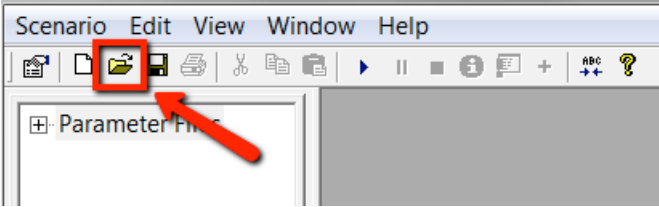
**QUICK START (HOW TO RUN THE MODEL WITH THE
DEFAULT SCENARIO)**

To run the model, follow these steps:

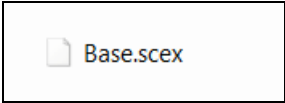
- 1. open the executable file `Demo4.exe`, in the folder `Model_and_parameters`.



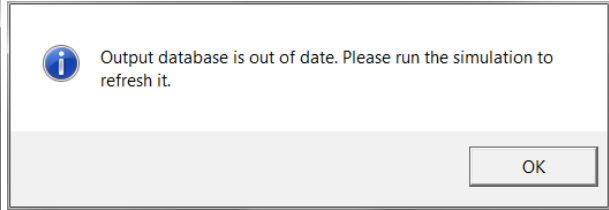
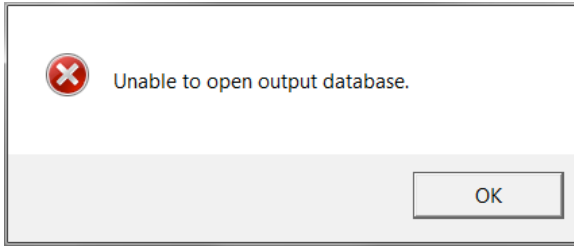
- 2. Click on the "open" icon (or select Scenario/Open (Ctrl+O)).



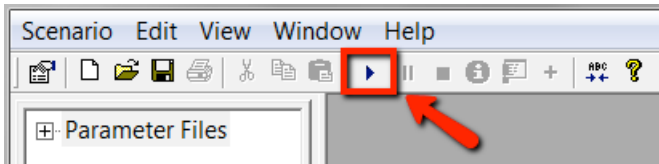
- 3. Select the only .scex file, `Base.scex`, which is the default scenario.



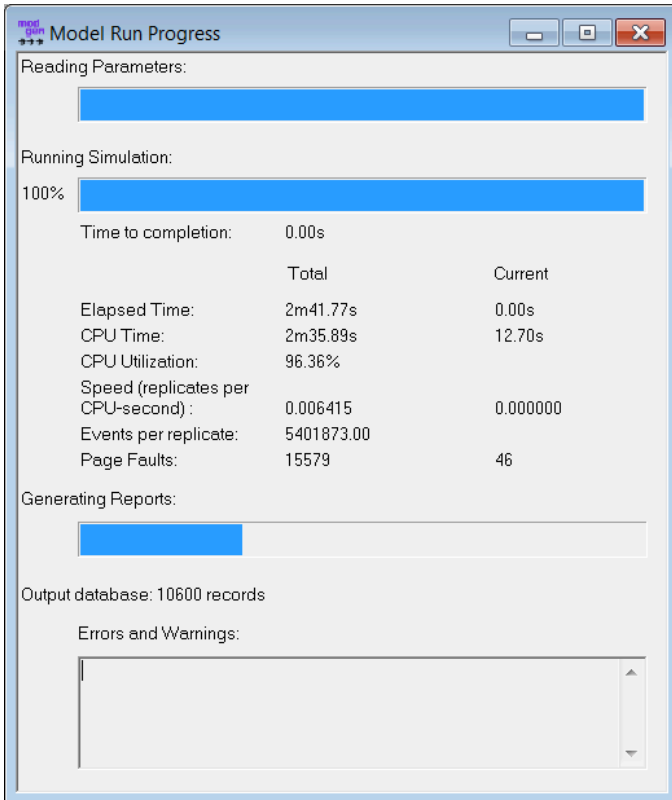
Don't be alarmed by the following warning messages, which might appear when you open a scenario:



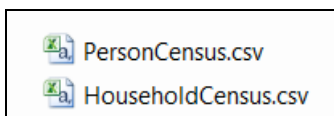
4. Run the simulation by clicking on the run/resume button (or by selecting Scenario/Run/resume).



5. Wait until the simulation ends.



6. At the end of the simulation, you can get the output .csv files from the folder where the executable has been run.



CHAPTER 3.

THEORETICAL DESCRIPTION OF THE MODEL

The proposed model is a closed cross-sectional discrete demographic model. The demographic model starts with a population made of persons and households at an initial point in time t_0 , and every member of the population ages from one year to the next, from t_0 to the end of the simulation period.

Annual probabilities. The ageing process depends on annual transition probabilities. These probabilities vary with several covariates (most often only sex and age). Some probabilities may vary with time, i.e. the transition probabilities are estimated for each year, based on existing projections. The users can choose to let the transition probabilities constant or to produce a set of probabilities for each year. Moreover, some events may depend not only on their probability but also on the features of other members of the population (e.g., agent-based model logic for the mating process).

Covariates. Individual probabilities are estimated as sex- and age- (or duration-)specific rates, and parameters indicating the relative risks. Some interactions can be included, but the model has to be as simple as possible: events already experienced are defining the current state of each individual. Transition probabilities (for each event) are defined for each individual, depending on sex, age and other covariates. A demographic covariate is included in the microsimulation model if it has an impact for other UrbanSim modules or if it is correlated with some of the transition probabilities within the demographic module. Some events may concern groups of individuals. Union formation, dissolution, and fertility processes are women-driven. Migrations are made by all the members of a household together. For each transition, a set of basic covariates is defined, as a minimum model. Additional covariates may be included in order to increase the relevance of the estimation.

As a matter of simplicity, age as a covariate used to estimate transition probabilities is increased every year, and remains constant over the year (period-cohort parallelograms in the Lexis diagram).

Social heterogeneity is accounted for with two covariates: *level of education*, divided in three groups, low (primary), middle (secondary) and high (university); *place of birth* (born in the region, born in the country, born abroad). For immigrants, *duration since arrival* can be used as a covariate. *Labour force participation* can take two values: "in the labour force" (working or unemployed) or "not in the labour force" (inactive or retired).

Individuals and households. Each individual belongs to one and only one household. One reference person is defined for each household and updated each year. For the time being, only the French definition of the reference person is used¹ (see <http://www.insee.fr/en/methodes/default.asp?page=definitions/pers-ref-menage-exp-prin-rrp.htm>). The demographic model does not include any information on dwellings. The location of households within dwellings will be part of the Location module.

¹ I.e.: if there is a couple in the household, the reference person is the man, otherwise the woman; otherwise, if there is no couple, the reference person is the oldest active person; otherwise, if there are no active persons, the reference person is the oldest one.

Order of events. The events are run in the following order:

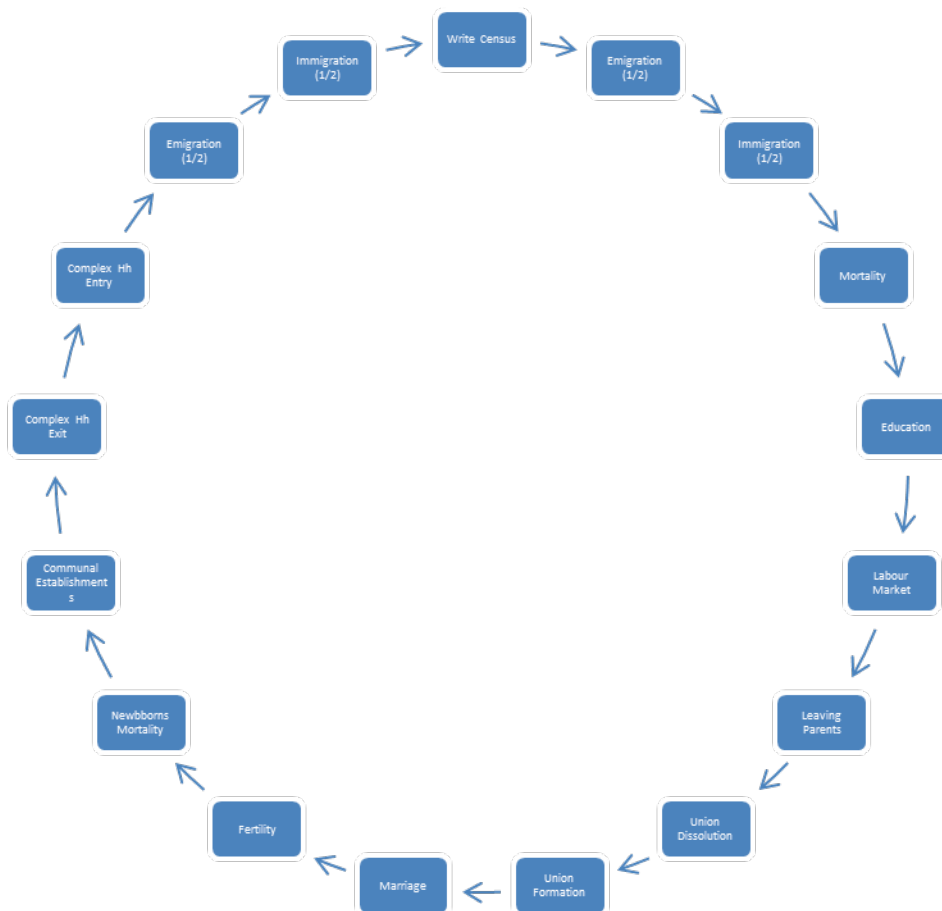


Figure 1 - Events Order

Household types.

Two types of "actors" are modelled: persons and households.

Persons are linked with other persons through 3 possible links: partner, mother and father.

Each person is also linked to one household (unless he/she lives in a communal establishment, in which case the agent is not linked to any household).

A "normal household" is defined as a household with either:

1. a single person;
2. a couple;
3. a couple with children;
4. a single parent with children.

A "complex household", instead, is a household where one or more persons, not partnered, join one of the four possible types of "normal households" e.g. not related adults living in the same dwelling or an old parent living with the family of his/her son/daughter².

People that are in communal establishments are not linked to any household.

Output. An overall population is produced each year, as if there was an annual census. Any data can then be extracted from this population, at the level of individuals or of households. The output of the demographic model can be at individual or household level.

Calibration. Some calibration can be made in order to post-stratify the population (individuals or households) by size or according to any variable, in order to get a population consistent with any exogenous constraint. This calibration process is not explicitly implemented in the demographic module. It must be done by trial and error, as there can be many ways to post-stratify a population. For instance, if the population growth is larger than expected, it can be made smaller, either directly through lower fertility, more severe mortality, less immigrants or more emigrants, or indirectly with lower union formation probabilities or larger union dissolution risks (if couples have a higher fertility than people living alone, or if people living as a couple emigrate less than people living by their own). Different calibrations may have different side-effects: a decline in fertility will lead to a small decrease in the number of households, at least in the long run, while an increase in union dissolutions will lead to less births, but also to an increase in the number of households, at least in the short run.

Programming language, interface. The demographic model has been written using the Modgen language. It is a language developed and maintained by Statistics Canada for creating microsimulation models. This language is basically a C++ pre-compiler. It has been used for 15 years and many different types of models have been developed using this language.

² For simplicity, just one couple can exist in each household. Therefore, a household with both parents living with the family of one of their children is not possible.

CHAPTER 4.

INPUT FILES

4.1. DESCRIPTION OF THE INITIAL POPULATION

The initial population and the two immigrants' pools are loaded from external .csv files, which need to be located in the same folder of the executable file. The names of the .csv files can be modified in the parameter file `Base(Initialization).dat`. The default names are: *InitialPopulation.csv*, *WithinPool.csv* and *AbroadPool*.

```
parameters {
  int   BaseYear = 1999; //EN Base Year
  file  InitialPopulationFile = "InitialPopulation.csv"; //EN File name of starting population
  file  WImmigPoolFile = "WithinPool.csv"; //EN File name of immigrants-from-within-the-country pool
  file  AbrImmigPoolFile = "AbroadPool.csv"; //EN File name of immigrants-from-abroad pool
};
```

Figure 2 - Input file names in `Base(Initialization).dat`

The rows represent individuals and the columns represent variables. *InitialPopulation.csv* needs to have 17 columns (i.e. variables), while the two pools files only need 15 columns: here's the list and order of the needed variables:

- Agent ID
- Household ID
- Sex
- Age
- Parity
- Age at end of studies
- Final level of education (0 = low, 1 = middle, 2 = high)
- Mother's level of education (0 = low, 1 = middle, 2 = high)
- Father's level of education (0 = low, 1 = middle, 2 = high)
- Mother's ID
- Father's ID
- Partner's ID
- Marital status (0 = never-in-union, 1 = separated from de-facto union, 2 = in-de-facto-union, 3 = married, 4 = divorced, 5 = widowed)
- Working status (0 = out of labour market, 1 = in labour market)
- Union duration, in years (for agents not in union, 0 should be entered for convenience)
- Communal residence (0 = private household, 1 = communal residence) [only for the initial population file *initialpopulation.csv*, not for the immigration pools]
- Place of birth (0 = simulated region, 1 = simulated country, 2 = abroad) [only for the initial population *initialpopulation.csv*, not for the immigration pools]
- Duration since arrival [only for the initial population *initialpopulation.csv*, not for the immigration pools]

Here's an example of the first few lines of file *InitialPopulation.csv*:

```
0,0,0,59,1,18,1,1,1,-1,-1,-1,5,1,0,0,1,-1
1,1,1,50,2,15,2,2,2,-1,-1,2,3,1,21,0,1,-1
2,1,0,46,6,15,0,0,0,-1,-1,1,3,0,21,0,1,-1
3,1,0,20,0,-1,0,0,2,2,1,-1,1,0,0,0,1,-1
4,1,0,13,0,-1,-1,0,2,2,1,-1,0,0,0,0,1,-1
5,2,1,45,6,15,1,1,1,-1,-1,6,2,1,21,0,1,37
```

Figure 3 - extract from *InitialPopulation.csv*

Missing values in the initialization files are only allowed for the following variables:

1. Mother's, father's and partner's IDs may be set to -1 if the agent is not linked respectively to his/her mother, his/her father or if the agent does not have a partner.
2. Level of education and age at end of studies may be set to -1: in this case, the model imputes the values of these variables as explained in the following paragraph.

4.2.LINKS, IDS AND IMPUTATIONS

The links between agents in the initial population and in the two immigrants' pools are created starting from the IDs. The initial population file must be ordered/sorted by households' ID and by persons' ID. For each of the three input files, both agents' and households' IDs must not have gaps and must start from 0. For the agents with no mother, father or partner in the initial population, the IDs of mother, father or partner respectively must be set to '-1'. Levels of education of both parents must be filled even if the parents are not in the initial population.

The model imputes a level of education and an age at end of studies for agents having these variables set to '-1'. The level of education is imputed on the basis of the level of education of the agent's parents (as for agents born in the simulation). The age at end of studies is imputed according to the agent's level of education for people younger than 15 years old or older than 30. Instead, for persons aged 15 to 30 (ends included), the imputation is performed using the user-modifiable parameter `dAesImpute`, which can be found in the file `Base(Education).dat`.

If an agent is loaded with the initial population and he/she lives in a communal establishment (CE), then if he/she is in a union, the union is broken and the link with his household is severed. People in CEs are not linked to any household and are only subject to ageing and to go through the education and labour force modules. All potential immigrants in the two pools are set to "private residence", i.e. immigrants cannot directly immigrate in communal establishments, but they can join a communal establishment after their arrival.

CHAPTER 5.

PARAMETERS

5.1. GRAPHICAL USER INTERFACE FOR THE INPUT PARAMETERS: AN EASY WAY TO MODIFY A UNIQUE PARAMETER

Before running the simulation, input parameters files can be opened by double-clicking on their name in the File View window on the left of the screen.

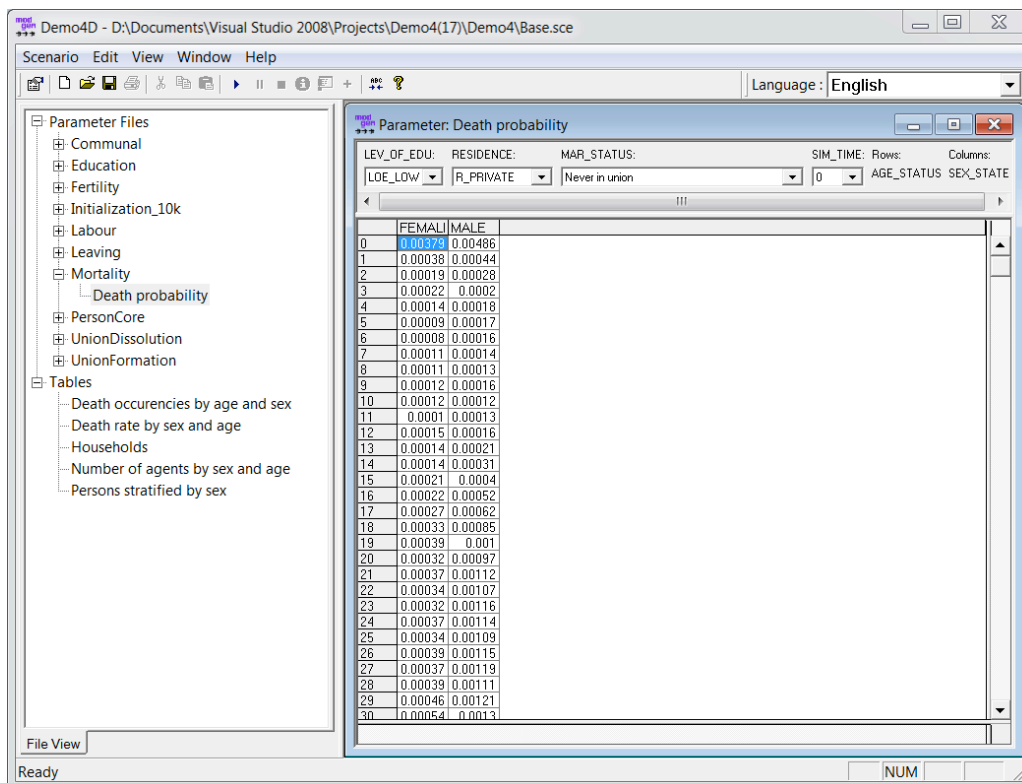


Figure 4 - Opening a parameter file

Parameters with more than two dimensions are shown with a drop-down menu for each additional dimension (see Figure 4). It is possible to modify a single cell by double-clicking on it. By right-clicking ("RC") on one or more selected cells, users can copy and paste figures from tab-delimited text files or from Excel tables. There are two possible ways to copy the parameters: a simple copy or a special copy. "RC" > **Copy Special...** allows to copy the whole current page or even to copy all pages, i.e. all dimensions.

Useful information is accessed by right-clicking on a parameter table and choosing **Properties**.

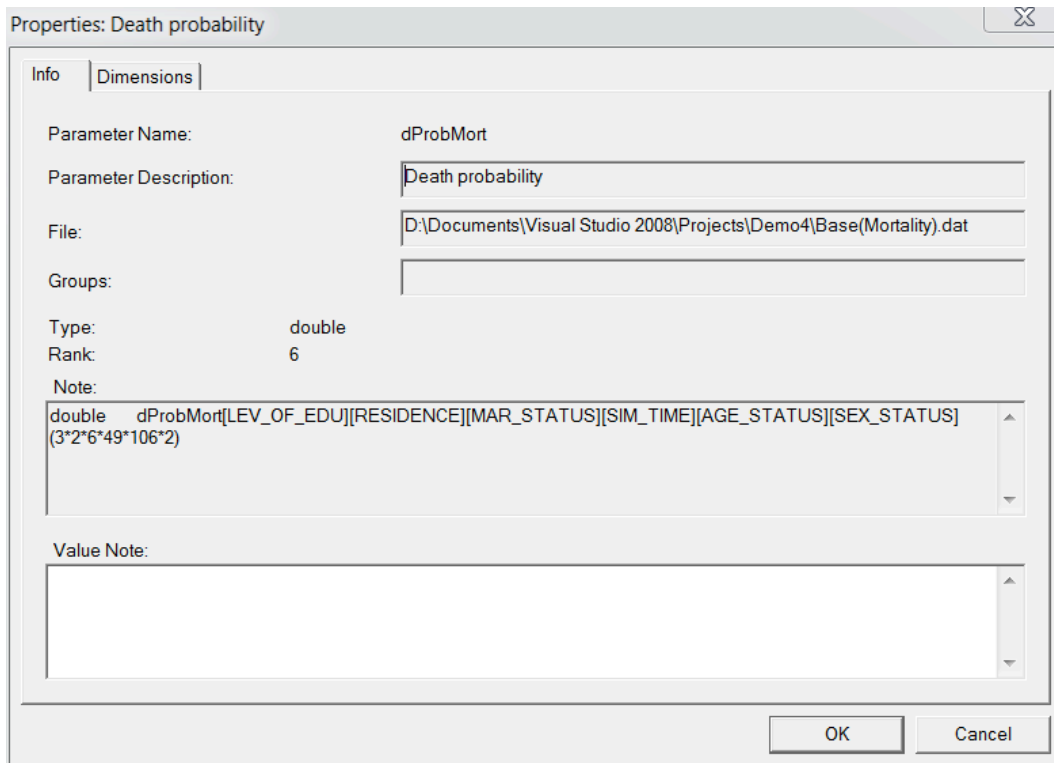


Figure 5 - Parameter properties - Info

The first tab of the Properties window displays the *type* of the parameter (in this case *double*), the number of dimensions (i.e. the *rank*) and the developer's note to this particular parameter. This tab also shows the full path of the file which contains the parameter.

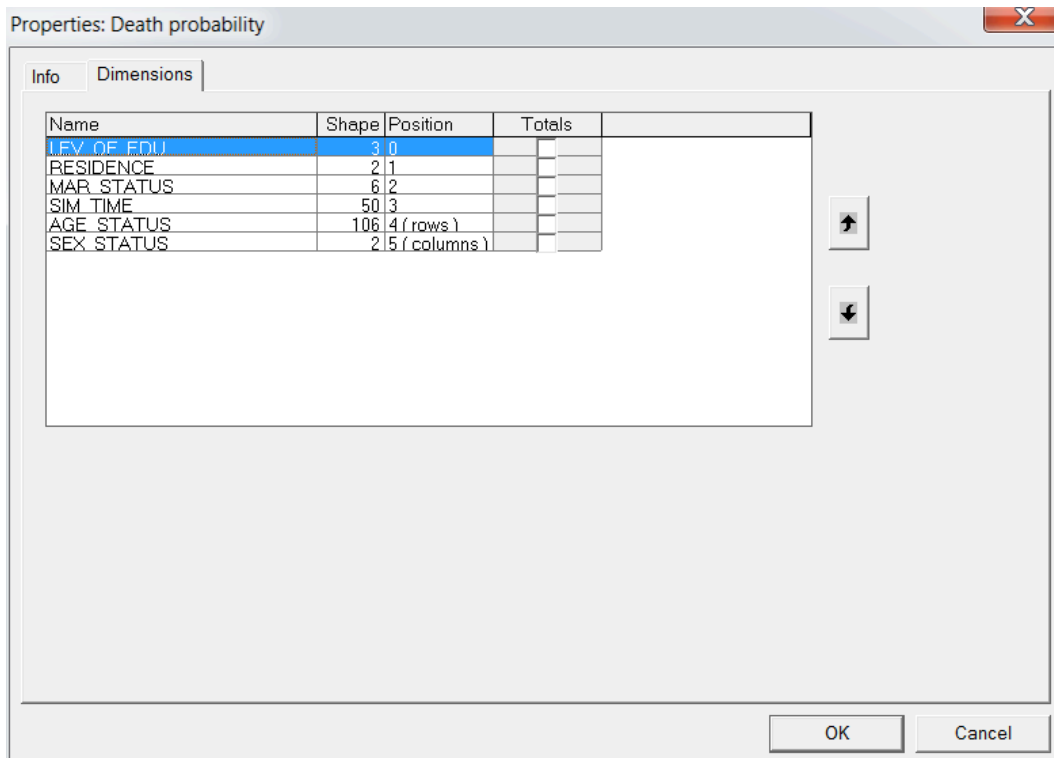


Figure 6 - Parameter properties - Dimensions

The second tab shows the dimensions of the parameter: their names, shapes (i.e. the number of different values each dimension can take) and positions. For parameters of more than two dimensions, the last two ones are the rows and columns of the parameter table. By ticking the appropriate box, it gives the totals on whichever dimension. Finally, the order of the dimensions can be changed using the two arrows on the right-hand side.

5.2. MODIFY A COMPLETE MATRIX OF PARAMETERS : HOW TO CREATE A PARAMETERS .DAT FILE

To create a new parameter file, without copying and pasting from Excel (which can be tedious for parameters with more than two dimensions), users can either modify the existing default parameter files or create a new flat file with a .dat extension and possibly a name following the rule `scenario_name(Module_name).dat`. The file must obey the following syntax rules:

```
parameters {
    type name[first dimension][second dimension]...[last dimension] = {
        X1, X2, ..., Xn,
    };
};
```

Figure 7 - Parameter .dat files syntax

where *type* is the type of the variable, *name* is the name of the parameter, *dimensions* are the names of the parameter dimensions and X_i are the values of the parameters. Note the comma also after the last value of the parameter.

The type and the name of parameters is then displayed in the Info tab of the Properties window (in the example shown in Figure 8, the parameter's type is "double" and its name is "dProbMort", while the name and order of the dimensions can be found in the Dimension tab (as shown in Figure 10, in the example the first dimension is called "LEV_OF_EDU", the second one "RESIDENCE" and so on). The *shape* of each dimension (i.e. the number of possible values the dimension can take, e.g. SEX_STATUS has a shape of 2, as it can either be FEMALE or MALE) is also shown in this tab. For the list of all variables of the model and available values, see Annex III. An easier way to change a parameter is to simply modify the existing parameters.

For multidimensional parameters, it is important that the values are entered in the right order. The program reads the parameters starting from the last dimension and proceeding to the second last dimension, and so on up to the first dimension. E.g. for a parameter having three dimensions, A, B and C, the first value applies to [A(0),B(0),C(0)], the second to [A(0),B(0),C(1)], up to [A(0),B(0),C(n_C)], then [A(0),B(1),C(0)]; [A(0),B(1),C(1)] and so on up to [A(n_A),B(n_B),C(n_C)], where n_A , n_B , and n_C are respectively the shape of A, B and C.

As an example, let's look at the death probability parameter:

```
parameters {
    //EN Death probability
    double dProbMort[LEV OF EDU][RESIDENCE][MAR STATUS][SIM TIME][SEX STATUS][AGE STATUS] = {
```

First of all, notice the presence of the heading keyword *parameters*. This keyword must appear at the beginning of all .dat files, followed by a brace "{". At the end of every .dat file, a closing brace followed by a colon must also appear: "};"

After this keyword, we see the name of the parameter preceded by *//EN* and on the following line the definition of the parameter. The parameter name, "dProbMort", is preceded by its type, "double", and it is followed by the list of dimension in brackets. The exact order of these dimensions must be respected. For the list of parameters, see Annex IV.

```
double dProbMort[LEV_OF_EDU][RESIDENCE][MAR_STATUS][SIM_TIME][SEX_STATUS][AGE_STATUS] = {
    (3){
        (2){
            (6){
                0.003257,0.000747,0.000256,0.000198,...
            },
        },
    },
};
```

Figure 8 - example of parameters file

Here, the first number, 0.003257 represents the probability of dying of a person with level of education (LEV_OF_EDU) equal to 0 (meaning low education level), residence (RESIDENCE) 0 (private residence), marital status (MAR_STATUS) 0 (single), simulation time (SIM_TIME) 0 (first year of simulation), sex (SEX_STATUS) 0 (female), age (AGE_STATUS) 0. The second number, 0.000747, represents the probability of dying of a person with all the above-mentioned variables equal to 0, but with age equal to 1. As users can check either in the parameter properties (Figure 10) or in Annex IV, AGE_STATUS's dimension is 106, as simulated ages go from 0 to 105, so we'll have 106 values for females in the first year, then 106 values for males in the first year, followed by 106 values for females in the second year and 106 values for males in the second year and so on.

Note the use of repeaters to replicate the entered probabilities for the dimensions for which we don't have any estimation. So we repeat everything 6 times, for all possible level of education, then again 2 times for all possible residence types, and finally other 3 times, for all possible level of educations. Pay attention to the peculiar syntax of repeaters. A tip: use them as much as possible as they simplify the understanding of parameter dimensions, and also save space on parameters' file size.

"Repeaters" are useful to repeat one or more numbers: if the number to be repeated is only one, then it is sufficient to put the desired number of repetition in parentheses just before the number to be repeated. For example, to repeat twice the number 1, write "(2) 1". Instead, to repeat more than one number, put the numbers to be repeated in curly brackets. For example, to repeat three times the numbers 1, 2 and 3, write "(3) {1, 2, 3}," which is equivalent to writing "1, 2, 3, 1, 2, 3, 1, 2, 3".

Repeaters are particularly useful to repeat probabilities for dimensions that didn't enter in the estimation.

Note that all numbers must be followed by a comma (even the last one).

Comments can be entered, preceded by a double slash (//).

The programme/model ignores spaces, comments and carriage returns (meaning that, all values for parameters could even be entered in a single line, provided they respect the syntax shown in Figure 7).

For parameters having more than one dimension (i.e. one value), values must be enclosed in curly brackets.

Level of education	Residence	Marital	Simulation year	Sex	Age	Death probability
Low	Private	Single	0	Female	0	0.003257
Low	Private	Single	0	Female	1	0.000747
Low	Private	Single	0	Female	2	0.000747
Low	Private	Single	0
Low	Private	Single	0	Female	105	1
Low	Private	Single	0	Male	0	0.004206
Low	Private	Single	0	Male	1	0.000882
Low	Private	Single
Low	Private	Single	0	Male	105	1
Low	Private	Single	1	Female	0	0.003119
Low	Private	Single
Low	Private	Single	49	Male	105	1
Low	Private	Separated	0	Female	0	0.003257
Low	Private
Low	Private	Widowed	49	Male	105	1
Low	Communal Est.	Single	0	Female	0	0.003257
Low
Low	Communal Est.	Widowed	49	Male	105	1
Middle	Private	Single	0	Female	0	0.003257
...
High	Communal Est.	Widowed	49	Male	105	1

Table 1 – Example of Parameter Table

For modules with more than one parameter in a single .dat file, parameter should be listed as follows:

```

parameters
{
    double dProbComplexEntry[AGE_STATUS][SEX_STATUS] = { //106*2
        1, 1, ...,
    };

    double dCxAgeDiff[AGE_CX_INT] = { //7
        1, 0.2, 0.4, 1, 0.6, 0.2, 0.1,
    };

    double dCxDsaDiff[DSA_PART] = { //7
        1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.25,
    };

    double dOutChildMult = 1.5;

    double dRelativesBoost = 2;

    int nComplexPool = 50;
};

```

Figure 9 - Content of file Base(ComplexEntry).dat

Once the input parameter file is created, users can check whether each parameter is read correctly by the program by double-clicking on the parameter name in the File View window on the left-hand side of the screen. If input and expected parameter dimensions don't correspond, an error occurs upon launching the simulation.

CHAPTER 6.

SCENARIO SETTINGS

To run the model, users have to open the executable file called `Demo4.exe`. They are then presented with the interface, which allows them to control many aspects of the simulation.

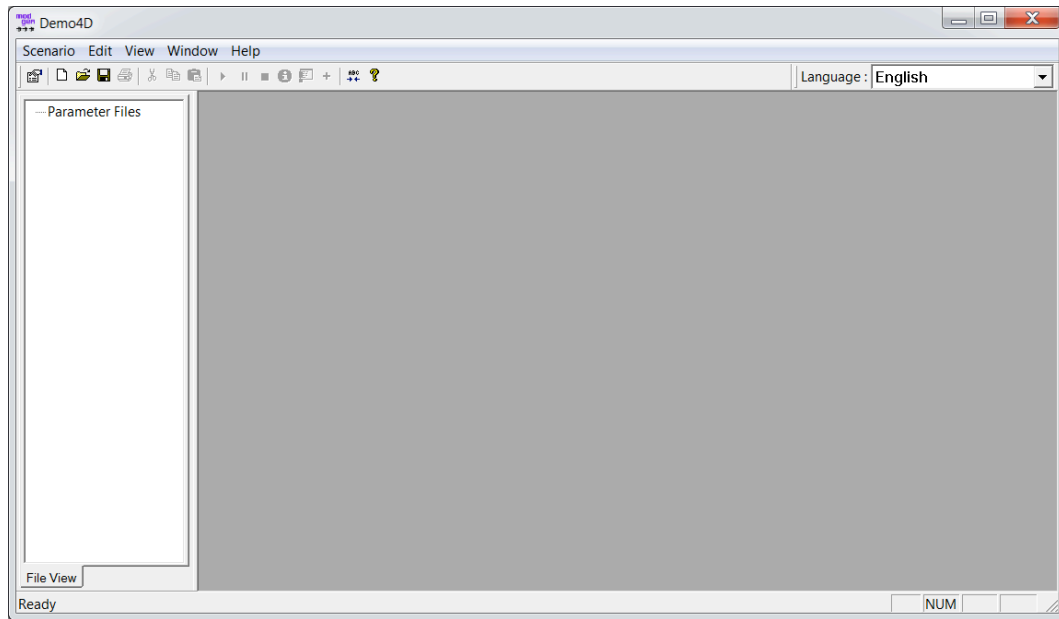


Figure 10 - Visual interface

The first step to run a simulation is to load or create a scenario (.scex) file. This simply defines the model inputs, outputs and some execution options.

From the **Scenario** menu, it is possible either to choose **New**, in order to start from scratch or **Open...** if an existing scenario is to be loaded. To load the default scenario, select **Open...** and then choose the only .scex file present in the folder, `Base.scex`.

Don't be alarmed by the following warning messages, which can occur when you open a scenario:

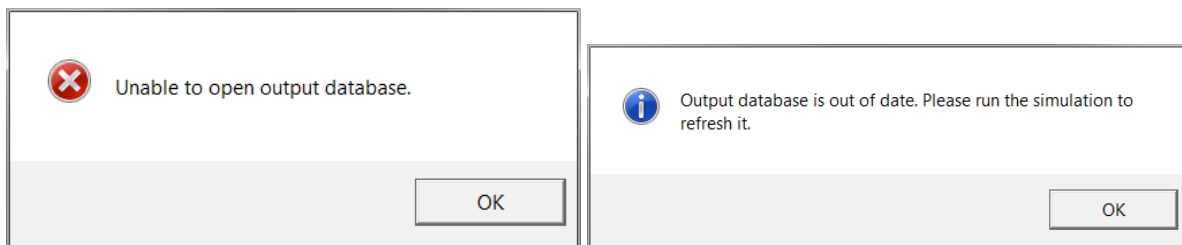


Figure 11 - Unimportant warning messages

In case one of them does appear, just ignore it, click OK and proceed.

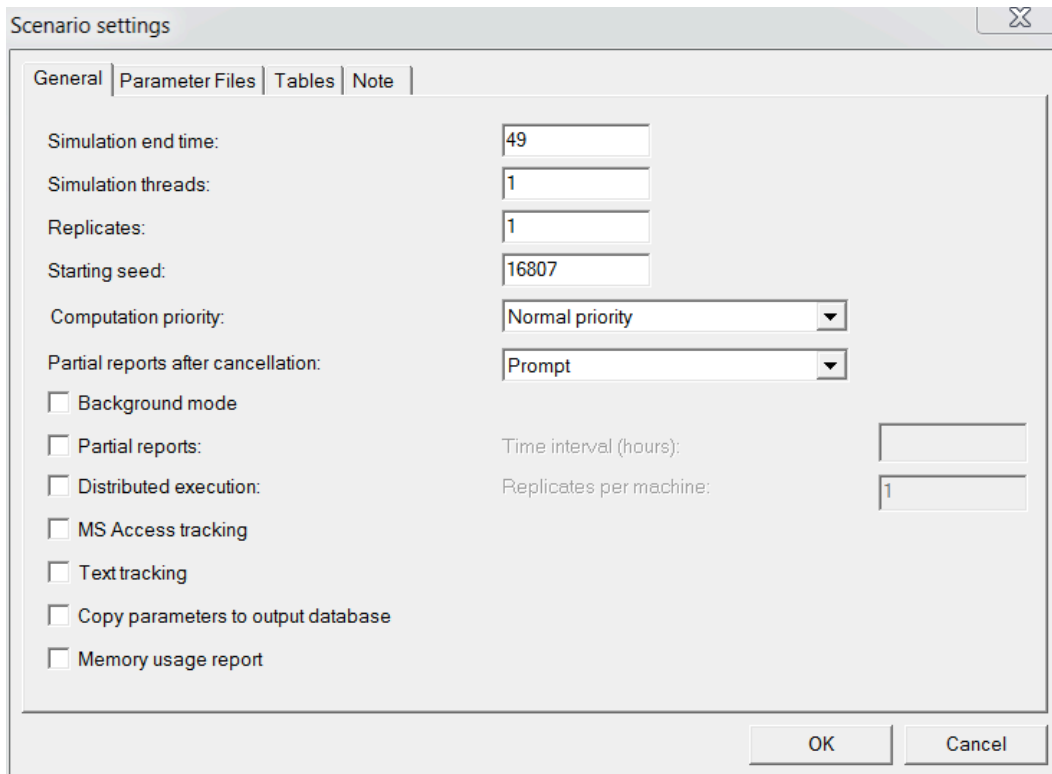


Figure 12 - Scenario settings - General

The Scenario Settings window, opened by choosing **Scenario > Settings**, or by pressing the first icon on the toolbar (🔧) is divided into four categories: General, Parameter Files, Tables and Notes.

The *General* tab sets the execution and output options.

Here we quickly review the most important options:

- ***Simulation end time***: this sets the end time of the simulation (the time starting at 0). Default value: 29. The maximum value is 49 (meaning 50 years of simulation).
- ***Replicates***: number of times the simulation is run.
- ***Starting seed***: random number generator(s) seed. As with all pseudo-random number generators, the results will always be identical for the same seed (provided the parameters are unchanged). Default value: 16807. The seed must be positive and not greater than 88046829567. It must also not be a multiple of 2147483648
- ***Computational priority***: settable to Low or Normal, the latter being the faster. This can also be set during the simulation (via **Scenario > Computational priority**).
- ***Partial reports after cancellation***: controls the generation of partial reports if the simulation is interrupted by the user. Available options: Yes; No; Prompt.
- ***Text tracking***: generate a text file named `scenario_name(trk).txt` that records every tracked state variable of every simulated agent. Not useful if you use the output file feature (see page 33 for more details on the output files).
- ***Copy parameter in output database***: include parameters in the exportable items. (If this option is ticked, the input parameters are printed in the export of the generated tables, generated either in an Excel or in a text file, via **Scenario > Export...**)

The *Parameter Files* tab allows users to set the input data files. These files must have the extension `.dat` and their names should be `scenario_name (module_name) .dat`³. (See page 20 for information on the format of these files).

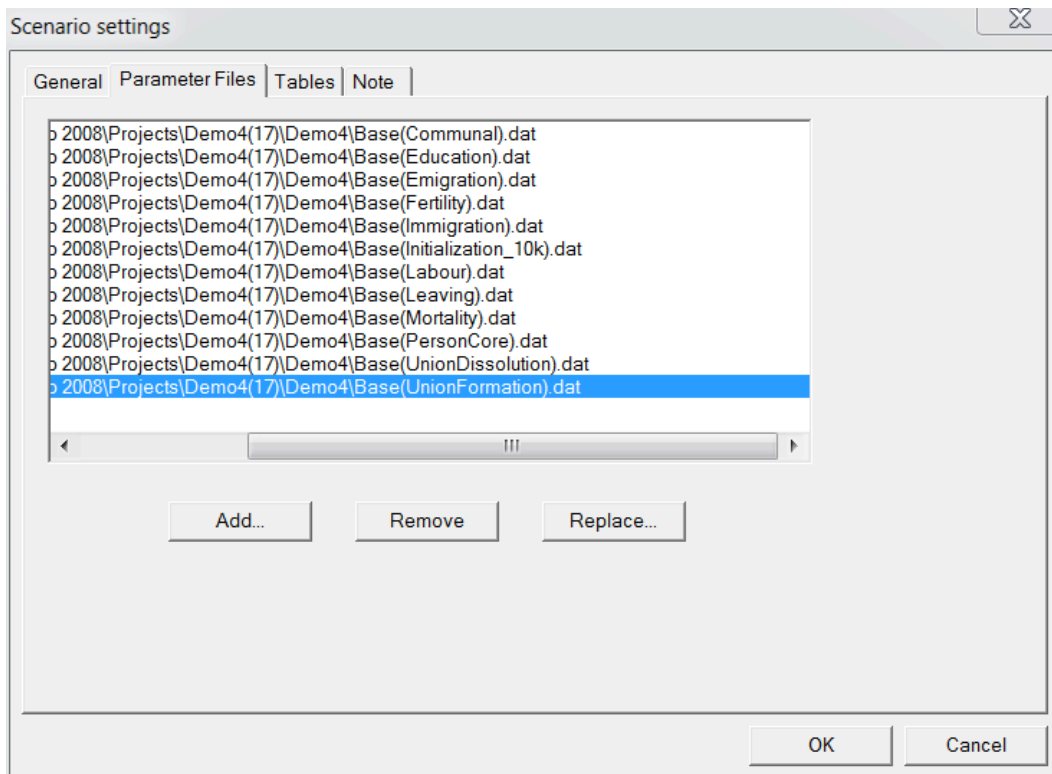


Figure 13 - Scenario settings - Parameter Files

The *Tables* tab lets the user control which tables are to be output at the end of the simulation. By default, all tables are output. The number of computed tables obviously has an impact on the speed of the simulation. For those interested in the output flat files, choose only the table "Simplest table", which just provides the overall number of actor initialised during the simulation. At least one table must be output and this is the easiest workaround to this constraint.

³ Where `scenario_name` is the name of the scenario and `module_name` is the name of the module which makes use of the parameter file. For example, the `.dat` file for the base scenario and the module Fertility should be called `Base (Fertility) .dat`.

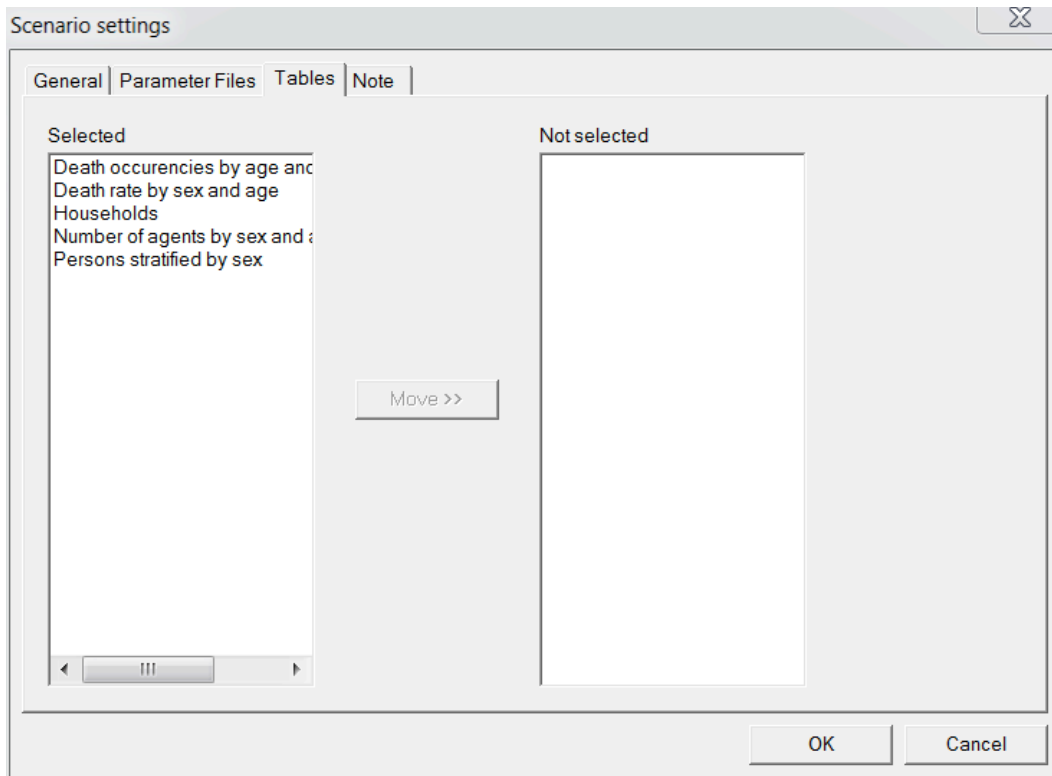


Figure 14 - Scenario settings - Tables

CHAPTER 7.

MODEL RUN AND OUTPUTS

To run the simulation, save the current scenario (**Scenario > Save**) and then run it (**Scenario > Run/resume**). The same commands are accessible from the menu bar:



There are three stages in the simulation process: reading of the parameters, simulation and generation of the reports. It is at the parameter reading step that an error occurs if a parameter is missing or if its structure (i.e. dimensions) is wrong.

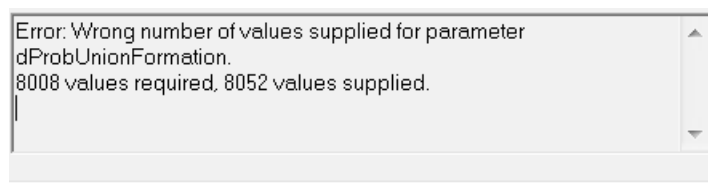



Figure 15 - Wrong dimension of parameter error - example

When the simulation is running, it is possible to monitor it via the Scenario menu or the toolbar . The simulation can be suspended or stopped. Information about the completed replicates and partial reports can be generated. Finally, the simulation end time can be extended.

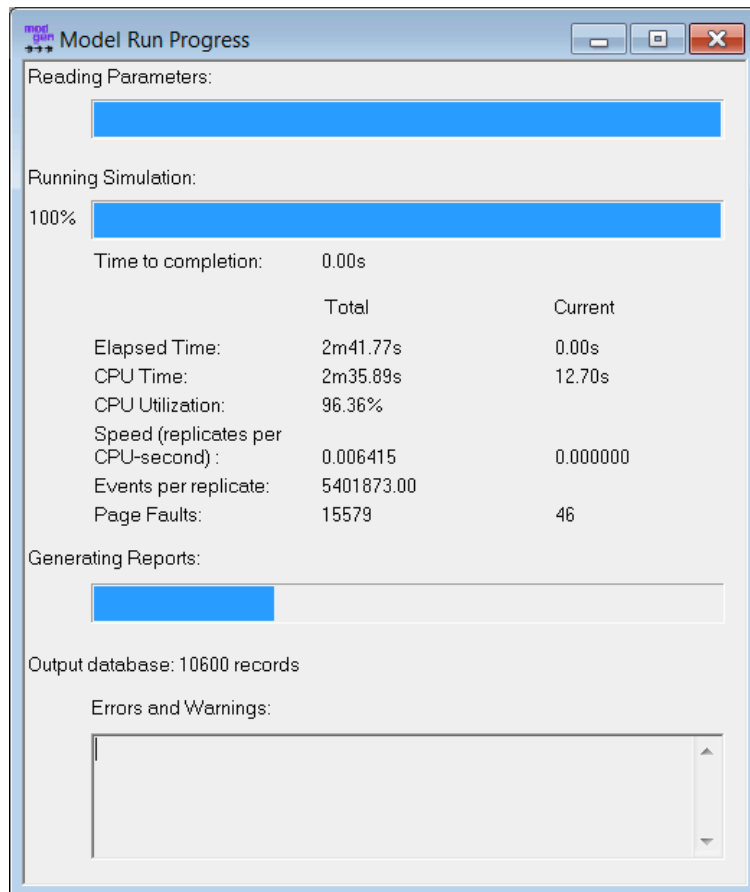


Figure 16 - Model Run Progress

Once the simulation step is over, the output tables are created.

Figure 17 shows an example of an output table. As for parameter input tables "RC" > **Copy Special...** allows users to copy all dimensions. "RC" > **Properties** (or alternatively View > Properties or Alt+Enter) opens the properties tabs. The first two tabs (Info and Dimensions) are equivalent to those described for the input parameter properties, while the "Decimals" and "Transformation" tabs are specific to output tables.

The Decimals tab controls the number of decimal figures shown in the table, while the Transformation tab shows standard error values and coefficient of variation values when more than one replicate has been simulated, the default output shown being the average value among all the simulated replicates.

Transformation: Value
Simulation year: Rows: Columns:
All Current integer age Agent's sex

	FEMALE	MALE
0	1897	13882
1	2040	2153
2	2226	2345
3	2430	2534
4	2620	2732
5	2788	2929
6	2937	3111
7	3073	3284
8	3232	3440
9	3389	3594
10	3517	3758
11	3650	3905
12	3798	4056
13	3940	4202
14	4072	4347
15	4210	4481
16	4333	4599
17	4455	4735
18	4581	4901
19	4733	5061
20	4917	5223
21	5083	5366
22	5278	5536
23	5467	5726
24	5685	5923

Figure 17 - Output table

Properties: Number of agents by sex and age

Info | Dimensions | Decimals | Transformation

- Value
- Standard error values
- Coefficient of variation values (%)
- Replicate 0
- Replicate 1

OK Cancel

Figure 18 - Table properties

It is also possible to export output tables along with standard errors, coefficients of variation and single replicates' output to a text file or to an Excel file with **Scenario > Export...** In the Output Contents window, users can choose which data, which tables and, if the proper option had been ticked in the Scenario Settings, which parameters to export.

Three different table styles are available:

- *one worksheet per table* creates an Excel worksheet for every complete table, by repeating the values of the leading dimensions;
- *one worksheet per table slice* creates a single worksheet for every two-dimensional table slice;
- *pivot table format* is to be chosen by users who want to create pivot tables within Excel with the output data: the *sparse* option exports only non-zero rows.

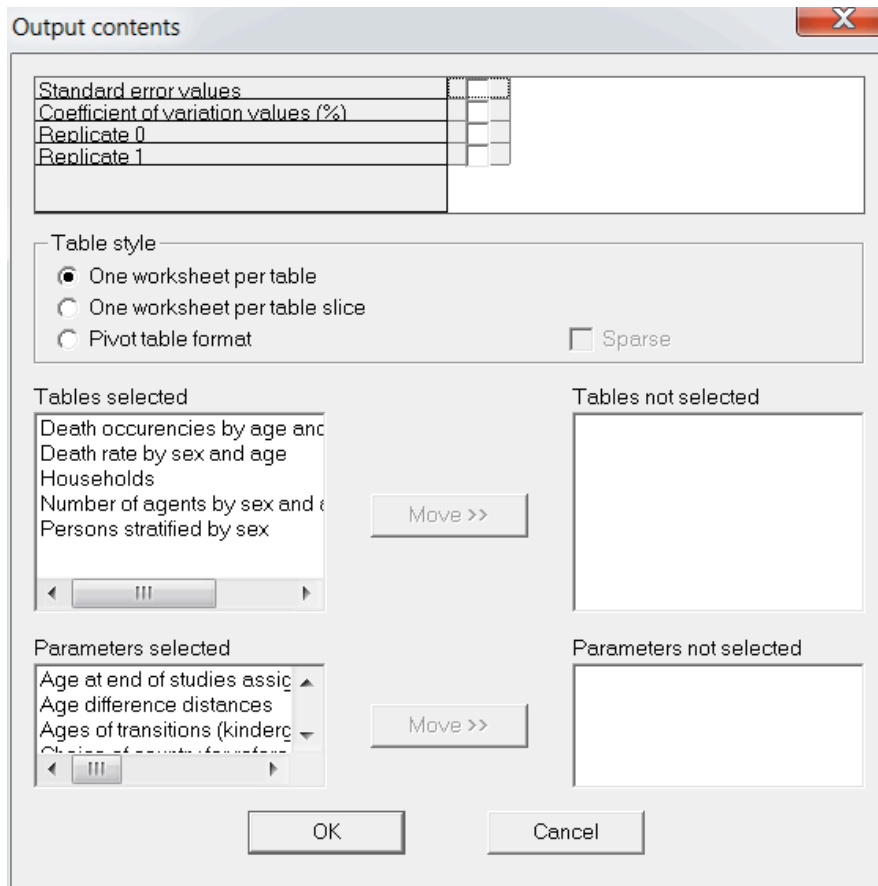


Figure 19 - Export to Excel settings

7.1. TABLES

In the current implementation, 29 tables can be computed:

i. Stocks

- Actors initialized (fastest table to compute, include ONLY this one when running simulation with large population or large number of replicates)
- Number of persons
- Number of persons by year and final level of education
- Number of persons in communal establishments
- Number of persons in de-facto union
- Number of persons in union
- Number of persons married

ii. Flows

- Births
- Deaths
- De-facto union disruption by death of partner
- De-facto union disruption not by death of partner
- Direct first marriages (women)
- Direct marriages
- Direct not-first marriages (women)
- Divorces
- Emigrants
- Emigrated couples
- Entries into communal establishments
- Exits from communal establishments
- First unions (Women)
- Immigrants from abroad
- Immigrants from within the country
- Immigrated couples
- Marriages

- Not-first unions (Women)
- Union disruptions by death of partner
- Union disruptions not by death of partner
- Union formations
- Widowing (i.e.: marriage disruption by death of partner)

NOTE: The model is unable to output stocks tables on the initial population. E.g. if the starting year of the simulation is 1999, the first year output in stocks tables will be 2000. Stocks table are output at the beginning of each year, starting from the second year. The only way to derive stocks tables on the initial population is to use the output .csv file at the first year.

Flows tables do not have this problem. E.g. if the starting year of the simulation is 1999, then the first year output in flows tables will be 1999 and it will show the number of events occurred during the year 1999.

7.2.OUTPUT FILES

Different types of census-like files are available for output in a comma-delimited .csv file⁴. It is no longer needed to use the default tracking .txt file.

Nine files can be output:

- `PersonCensus.csv`: yearly “census” of persons.
- `HouseholdCensus.csv`: yearly “census” of households.
- `AImmigPool.csv`: pool of the potential immigrants from abroad, persons.
- `AImmigHhPool.csv`: pool of the potential immigrants from abroad, households.
- `WImmigPool.csv`: pool of the potential immigrants from within the country, persons.
- `WImmigHhPool.csv`: pool of the potential immigrants from within the country, households.
- `EmigCensus.csv`: yearly “census” of emigrants, persons.
- `EmigHhCensus.csv`: yearly “census” of emigrants, households.
- `DeadCensus.csv`: yearly “census” of dead persons.

The variables output in `PersonCensus.csv` are:Year, ID, Household ID, Sex, Age, Mother’s ID, Father’s ID, Partner’s ID, Parity, Age of Youngest Child, Level of Education, Residence Type, Marital Status, Working Status, Union Duration, Education Enrolment (studying during the year) dummy, Grade, newborn dummy (born this year), just_union dummy (formed a union this year), imm_this_year dummy (immigrated this year), age_1_union (age at first union), age_1_marr (age at first marriage), just_comm (entered communal establishment this year), place_of_birth (place of birth: 0 = born in the simulated region; 1 = born in the simulated country, but not in the simulated region; 2 = born abroad).

⁴ In a previous version, the model needed to use the default tracking system to output files. This is no more the case with this new version.

The variables output in `HouseholdCensus.csv` are: Year, ID, Size (number of members), Couple dummy (1 if there is a couple in the household, 0 otherwise), Males (number of males), Females (number of females), Children (children of all ages), Children<25 (children younger than 25 years of age), Complex (number of isolated members), `age_young_child` (age of youngest child), `age_old_woman` (age of oldest woman), `age_old_man` (age of oldest man), newborn dummy (1 if a birth has occurred in the household this year), just_union dummy (1 if a union has occurred in the household this year), `ref_per_id` (ID of the reference person⁵)

The variables output in `AImmigPool.csv` and `WImmigPool.csv` are: ID, Household ID, Sex, Age, Mother's ID, Father's ID, Partner's ID, Parity, Level of Education, Age at End of Studies, Residence Type, Marital Status, Working Status, Union Duration, Education Enrolment (studying during the year) dummy, Grade.

The variables output in `AImmigHhPool.csv` and `WImmigHhPool.csv` are the same as `HouseholdCensus.csv`, but they don't include the Year, as, in the current release, the pool is static. They also don't include the reference person ID.

The variable output in `EmigCensus.csv` are: Year, ID, Household ID, Sex, Age, Mother's ID, Father's ID, Partner's ID, Parity, Level of Education, Residence Type, Marital Status, Working Status, Union Duration, Enrolled dummy, Grade, newborn dummy (born this year), just_union dummy (formed a union this year), `imm_this_year` dummy (immigrated this year).

The variables output in `EmigHhCensus.csv` are the same as `HouseholdCensus.csv`, but they don't include the reference person ID.

The variables output `DeadCensus.csv` are the same as `PersonCensus.csv`, but they don't include: Age of the youngest Child, `age_1_union`, `age_1_marr`, `just_comm`, `place_of_birth`.

All output files are sorted by year, if applicable, and ID. Rows represent individuals or households and columns represent variables.

⁵ At the time of writing, only the French rules are implemented: if there is a couple in the household, the reference person is the man, otherwise the woman; otherwise, if there is no couple, the reference person is the oldest active person; otherwise, if there are no active persons, the reference person is the oldest one.

These are the values of the "qualitative" variables in the .csv output files:

VALUES OF QUALITATIVE VARIABLES IN .CSV OUTPUT FILES		
SEX_STATUS		Sex
	0	Female
	1	Male
LEV_OF_EDU		Level of education
	0	Low
	1	Middle
	2	High
RESIDENCE		Type of residence
	0	Private
	1	In Communal Establishment
MAR_STATUS		Marital status
	0	Never in union
	1	Separated from common-law union
	2	In de-facto union
	3	Married and living with spouse
	4	Divorced and not living with a partner
	5	Widowed and not living with a partner
WORKING_STATUS		Working status
	0	Out of labour market
	1	In labour market
ENROLLMENT		Enrolment status (i.e studying or not)
	0	Not enrolled
	1	Enrolled
"GRADE"		School level
	0	Pre-kindergarten
	1	Kindergarten
	2	Primary school
	3	Secondary school
	4	Tertiary school
	5	Out of education

Table 2 – Values of qualitative variables in .csv output files

The names of the variables are printed on the first line of the files.

This is an example of the first lines of `PersonCensus.csv` file:

```

REP,YEAR,ID,HH_ID,SEX,AGE,MOTHER_ID,FATHER_ID,PARTNER_ID,PARITY,AGE_YOUNG_CHILD,LOE,RESIDE
NCE,MAR_STAT,WORK_STAT,UN_DUR,ENROLLED,GRADE,NEWBORN,JUST_UNION,IMM_THIS_YEAR,AGE_1_
UNION,AGE_1_MARR,JUST_COMM,

0,1999,2,125354,0,59,-1,-1,-1,1,0,1,0,5,1,0,0,5,0,0,0,-1,-1,0,
0,1999,3,125355,1,50,-1,-1,4,2,13,2,0,3,1,21,0,5,0,0,0,-1,-1,0,
...

```

Figure 20 - Example of output .csv file: first lines of PersonCensus.csv

CHAPTER 8.

FURTHER DOCUMENTATION

For further documentation on the ModGen visual interface, see <http://www.statcan.gc.ca/microsimulation/pdf/vis-guide-eng.pdf>.

I

ANNEX

Annex I.

VARIABLES DESCRIPTION

A. QUALITATIVE VARIABLES USED AS PARAMETER DIMENSIONS

Name	Index	
EDU_CLASS		School levels
	0	Kindergarten
	1	Primary school
	2	Secondary school
	3	Tertiary education
ENROLLMENT		Enrolment status (i.e studying or not)
	0	Not enrolled
	1	Enrolled
INIT_VARIABLES		Variables used in population initialization file
	0	ID
	1	Household ID
	2	Sex
	3	Age
	4	Parity (number of children)
	5	Age at end of studies
	6	Level of education
	7	Mother's level of education
	8	Father's level of education
	9	Mother's ID
	10	Father's ID
	11	Partner's ID
	12	Marital status
	13	Working status
	14	Union duration
	15	Residence type
	16	Place of birth
INIT_VARIABLES_IMM		Variables used to initialize the two immigrants pools
	0	ID
	1	Household ID
	2	Sex
	3	Age
	4	Parity (number of children)
	5	Age at end of studies
	6	Level of education
	7	Mother's level of education
	8	Father's level of education
	9	Mother's ID
	10	Father's ID
	11	Partner's ID
	12	Marital status
	13	Working status
	14	Union duration
LEV_OF_EDU		Level of education
	0	Low

	1	Middle
	2	High
LIVES_W_PARENTS		Dummy for “lives with parents”
	0	Doesn’t live with parents
	1	Lives with parents
MAR_STATUS		Marital status
	0	Never in union
	1	Separated from common-law union
	2	In de-facto union
	3	Married
	4	Divorced
	5	Widowed
NEWBORN		Dummy for “has a child aged zero”
	0	No
	1	Yes
PROB_VALUE		Probability-value dimension
	0	Probability
	1	Value
RED_MAR_STATUS		Not-in-union marital status
	0	Never in union
	1	Separated from de-facto union
	2	Divorced
	3	Widowed
RESIDENCE		Type of residence
	0	Private residence
	1	Communal establishment
SEX_STATUS		Sex
	0	Female
	1	Male
UNION_STATUS		Union status
	0	Not-in-union
	1	In union
UNION_TYPE		Type of union
	0	In de-facto union
	1	Married
WORKING_STATUS		Working status
	0	Out of labour market
	1	In labour market

B. DIMENSIONS OF QUANTITATIVE VARIABLES USED AS PARAMETER

Name	Range (ends included)	Meaning
ACTIVE_YEARS	16-70	Possible ages for entering into labour market
AGE_STATUS	0-105	Possible ages
ALL_POSSIBLE_AES	9-30	Possible ages at end of study for everybody
FERT_YEARS	15-49	Possible ages for having a child (women only)
NUM_AGE_PASS	0-2	Indices of possible ages at school transition
NUM_LOE	0-2	Indices of possible levels of education
PARITIES	0-10	Possible parities (number of children)
POSSIBLE_AES_KIDS	15-30	Possible ages at end of study for people born in the simulation
SIM_TIME	1999-2048	Simulation years, maximum range
UNION_AGE	15-105	Possible ages for entering into a union

C. PARTITIONS OF QUANTITATIVE VARIABLES USED AS PARAMETER DIMENSIONS

Name	Intervals	Meaning
AGE_CX_INT	[0,4],[5,14],[15,24],..., [45,54],[55+]	Age difference for complex household
AGE_ONE_YR_INT_10	[0,1],[1,2],[2,3],..., [9,10],[10+]	1-year age partition up to 10
AGE_ONE_YR_INT_18	[0,1],[1,2],[2,3],..., [17,18],[18+]	1-year age partition up to 18
AGE_TEN_YR_INT	[0,10],[10,20],..., [90,100],[100+]	10-year age partition
HH_NUM_CHILDREN_PART	[0,1],[1,2],[2,3],..., [9,10],[10+]	1-child household children partition up to 10
HH_SIZE_PART	[1,2],[2,3],..., [9,10],[10+]	1-member household size partition up to 10
PARITY_STATUS	[0,1],[1,2],..., [9,10],[10+]	1-child parity partition up to 10
UNION_DURATION	[0,1],[1,2],..., [9,10],[10,15],[15,20],... [55,60],[60+]	Union duration partition

Annex II.

PARAMETER ESTIMATION

To estimate the parameters needed for the demographic simulation of the Paris (Île-de-France) case study, we have relied upon the Family history survey ("Etude de l'Histoire Familiale") attached to the 1999 census, which counted 44330 respondents in IDF. Additional information on all the members of respondents' household was obtained by the census.

A. POPULATION

1. Initial population

The initial population contains 16 variables for each person: IDs of the person and of his/her household, mother, father and partner; sex; age; parity (i.e. number of children ever had); age at end of studies; final level of education; father's and mother's final levels of education; residence; marital status; employment status; union duration (set to 0 for people not in union).

Here is a brief description of these variables (sources of data used, potential imputation):

- **IDs:** census. Three variables have been used: household ID, family link (differentiating between male adults, female adults and children) and family type (out-of-family, principal, or secondary family). (variables num_men, lienf, numf)
- **Sex:** directly retrieved from the census. (variable sexex)
- **Age:** directly retrieved from census (anaix) and, if missing, complemented with the EHF survey (annais).
- **Parity:** EHF survey (nbenf3) for respondents and hotdeck imputation by sex and age for the other members of the household.
- **Level of education (LOE):**
 - For persons over-30 and for over-15s with a job, retrieved from census (nivetx).
 - For persons under-15, the LOE is set to -1 in the input files and imputed in the simulation on the basis of the parents' LOE (as for agents born in the simulation).
 - For those between 15 and 30 years old still in education, LOE is imputed from the distribution of LOE of persons aged 31 to 35 in 1999 who were still in education at each age. For instance, the level of education of a person aged 15, is based on the distribution of LOE of people aged 31 to 35 and still in education at age 15. Hotdeck imputation of missing values by sex and age is then performed.

Note: if the level of education of a person is set to -1, it is imputed by the model on the basis of the LOE of the person's parents, following the rules set in the user-modifiable input parameter `dLoe`.

- **Age at end of studies (AES):**
 - For the over-30s, retrieved from EHF survey (c09agetu) and hotdeck imputation of missing values by sex and age.
 - For the under-30s, set to -1 in order to be imputed in the simulation. In the simulation, the AES of under-15s is imputed from the person's own level of education (as for persons born in the simulation),
 - for those aged 15 to 30, AES is imputed from the user-modifiable parameter `dAesImpute`. The default values of this parameter are derived from the distribution of AES of people aged 31 to 35.

Note: if the age at end of studies of a person is set to -1, then, if the person is aged from 15 to 30 (ends included), his/her AES is imputed starting from the user-modifiable imputing rules parameter `dAesImpute`, otherwise it is imputed from the person's LOE (as for persons born in the simulation).

- **Marital status:** two variables of the census used: legal marital status at census (matr) and cohabitation mode (moco) to select people being in a union whatever their legal marital status. Married persons not living as a couple are set to divorced. In order to differentiate between singles who have never been in union and singles who have been in union (information not available in the census), a variable of the EHF survey is used, "union status" (c13cou). For non-respondents and missing values, an imputation by age is performed.
- **Employment status:** two variables from the census (indicator of working status (sitrav) and type of inactivity (inac)) and one variable from the EHF survey, "currently working" (c10act). People working or doing an internship or unemployed are set to "working". Students, retired and other not working persons are set to "not working". The remaining missing values (2083 cases, 1.99%) are imputed as "not working".
- **Union duration (for people currently in union):** 4 variables of the EHF survey ("year of first or unique union"; "year of latest union"; "year of first or unique marriage"; "year of latest marriage"). Estimated only for people in union. The remaining missing values (3308 cases, 6.03% of agents in union) are set to the mean union duration.
Note: union duration for agents not currently in union is set to 0⁶.
- **Residence:** based on the household population and the proportion living in communal establishments by sex and age, individuals are created by cloning from the household population the proper percentage of people by sex and age.
- **Place of birth:** two variables from the census: "place of birth" and "département of birth". Missing values (3674 cases, 3.54%) set to "born in France, but outside Île-de-France".
- **Duration since arrival:** census. Date of arrival in metropolitan France used (arrivx). Missing values (12310 cases, 53.61%) imputed using a GLM regression on sex, age and marital status. Set to -1 for people born in France.

Weighing: The census and EHF samples are weighted (poidsM5): each household has a specific weight. As the desired output network cannot be weighted, we adjusted the population to form a non-weighted one. Half of the households with the lowest weights (weight < 10% of distribution) were deleted, while those with the highest weights (weight > 80%) were duplicated or (weight > 95%) triplicated. The weight is inversely proportional to the probability for individuals of a particular kind to be interviewed. So, to create a non-weighted population similar to the real one, we need less individuals with low weight and more with high weight.

The initial population is composed of **124247** persons and **52630** households (average household size: 2.36 persons).

The exact sizes of the populations and pools of the model follow:

⁶ This cannot create any confusion with the "real" duration zero of people in union, as this variable is used only for persons in union.

	Size(individual)
Population	124247
Within-Country Immigrant Pool	25264
From-Abroad Immigrant Pool	1179

Table 3 – Population (number of people) and pool

2. Immigrants pool (from within France)

For internal migration data, we use the entry and exit rates estimated by Baccaïni (2001)⁷, especially Graphique V page 52, reproduced in Figure 21.

For people who were living in another region of France in 1990 (or at birth), or who lived abroad but without any information on their date of entry in France (e.g. people born in France), we estimated their probability to have entered in Île-de-France in 1990, 1991... 1998 as follows:

1. Based on rates of entry and exit (Figure 21, left), we estimate the probabilities $S(x,d)$, for a person entered at age $x-d$, to be still present at age x , for $d = 1$ to 9 (Figure 21, right);
2. We estimate for an individual of age x , the probabilities $pe(x,d)$ for each age at entry $x-d$, d from 0 to 9.
 - a. We first estimate the $e(x,d)$ the number of entries at age $x-d$ of people aged x in 1999. $e(x,5)$ is taken from the graph, and the corresponding $e(x,d)$ are obtained from a translation of this graph by $5-d$ (see Figure 22, left));
 - b. Then $pe(x,d)$ is estimated as (Figure 22, right):

$$pe(x,d) = \frac{e(x,d)}{\sum_{d=0}^9 e(x,d)}$$

- c. We thus estimate the products $pe(x,d) S(x,d)$ (Figure 23, left) and, finally, the distribution $P(x,d)$ of immigrants since 1990 by duration (Figure 23, right):

$$P(x,d) = \frac{pe(x,d) S(x,d)}{\sum_{d=0}^9 pe(x,d) S(x,d)}$$

The result is presented in Figure 24: the probability is declining up to age 10 (from 100% to 20%, is then larger between 12 and 22 years of age (above 25%) and almost constant (around 20% after age 30).

⁷ Baccaïni B., 2001, « Les migrations internes de France de 1990 à 1999: l'appel de l'Ouest », Économie et Statistique, 344, 39-79.

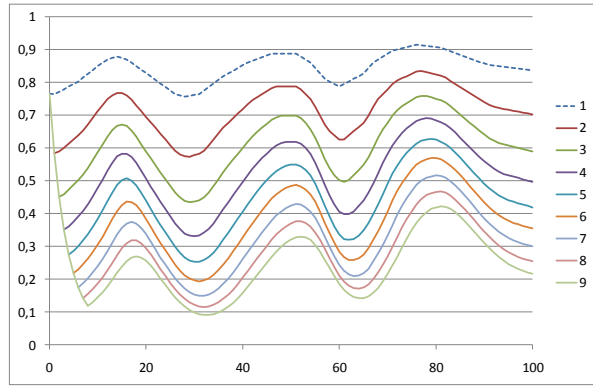
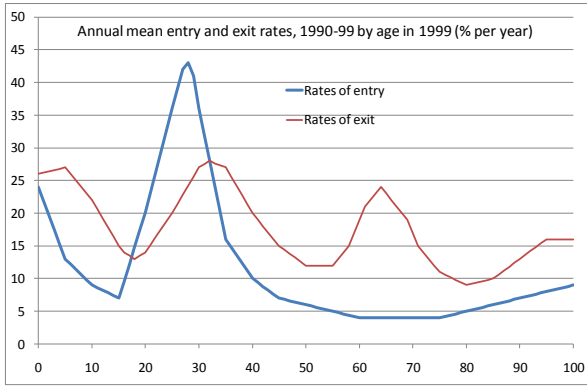


Figure 21 - Annual mean entry and exit rates, 1990-99, from Baccaïni (2001) and probabilities $S(x,d)$

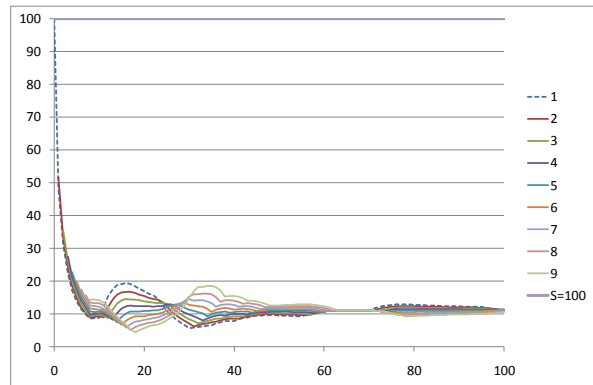
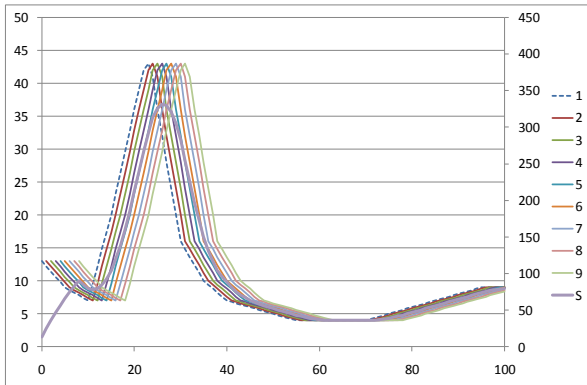


Figure 22 - Entries $e(x,d)$ and distribution of entries $pe(x,d)$

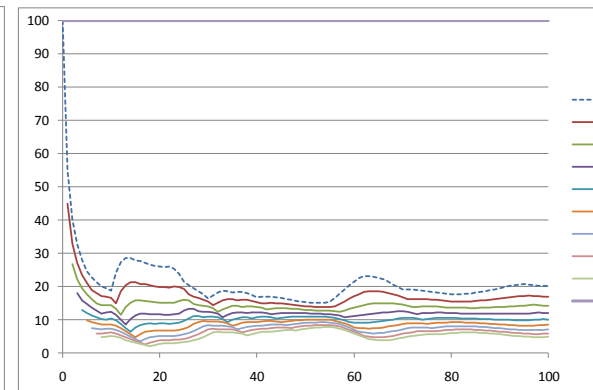
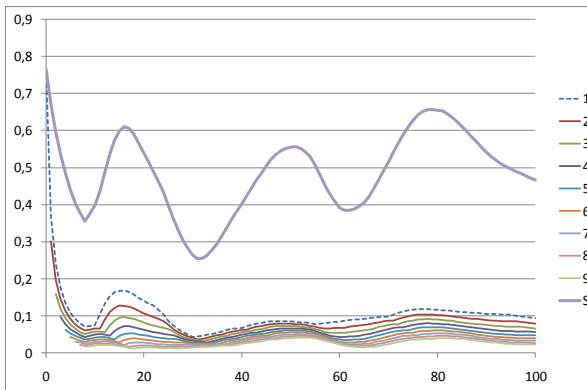


Figure 23 - Entries not followed by an exit $pe(x,d)S(x,d)$ and distribution of residents by duration $P(x,d)$

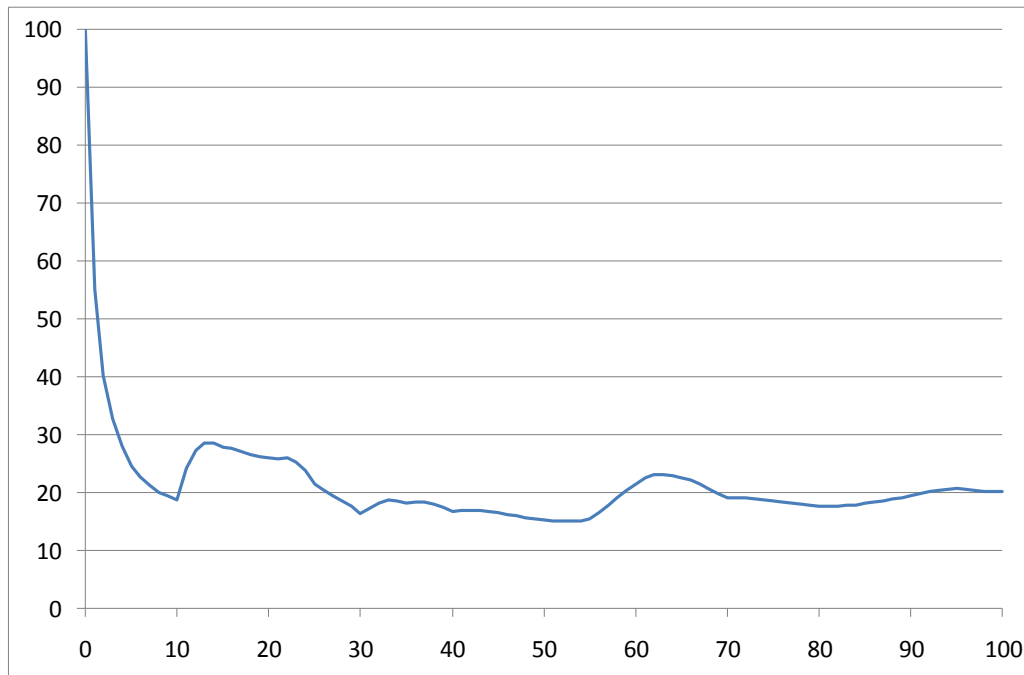


Figure 24 - P(x,1): Probability of being a recent immigrant by age

For each individual, the probability to be a recent immigrant from within France is thus estimated by $p = P(x,1)$.

An average probability is then given to each household, estimated as the average of the p of each household member.

A random selection of households based on the inclusion probabilities p is then made, and all the members of the selected households are included in the pool from where future immigrants are taken each year.

The pool is composed of 25264 persons and 10058 households (average household size = 2.53 persons).

3. Immigrants pool (from abroad)

For people born abroad, and who lived outside France at 1-1-1990, we use their date of entry in France (retrieved from the 1999 census), assuming that their first entry was the last one, and that it was in Île-de-France. Thus, a recent entry into France (1996, 1997, 1998 or 1999) means "recent immigrant"; 1995 or earlier means "not recent immigrant". The pool is composed of 1179 persons and 799 households (average household size = 1.48 persons).

B. ESTIMATION OF TRANSITION PROBABILITIES AND OTHER PARAMETERS

1. Mortality

Mortality rates forecasts by sex and age from 1999 to 2048, by J. Vallin et F. Meslé.

(http://www.ined.fr/cdrom_vallin_mesle/contenu.htm)

2. Fertility

Probability of having a child by sex, age and parity is estimated on the basis of a sub-sample of the 1999 EHF survey. In the simulation, only agents in union can have a baby: therefore, in order to obtain the probability of having a child given the fact of being in union, we divided the global probability by the ratio of agents in union by age, and censored the resulting probability at 0.5⁸.

3. Communal establishments

Transition probabilities to enter into communal establishment by age and sex were obtained using 1999 French census results. The probability of exiting from a communal household is set to 0.5 for everyone.

4. Complex Households

The probability to enter a complex household by sex and age was obtained also from the 1999 census.

5. Emigration

The proportion of people emigrating each year from Île-de-France is around 1.9%. With the target of keeping 1.9% as the average, we set the following emigration probabilities:

1. Households with at least a child: 2.62% if the youngest child is younger than 5; 1.6% otherwise.
2. Childless households with a couple: 1% if the eldest man is younger than 30; 2.5% if his age is in the interval [60,70); 1.6% otherwise.
3. Childless households without a couple: the probability of emigration of the household depends on the age of the eldest man, if a man is present, and otherwise on the age of the eldest woman. It is set to 1% if this age is less than 30; 2.5% if it is the interval [60,70) and 1.6% otherwise.

6. Immigration

Number of household immigrating each year from within the country set to 515; from abroad set to 468. It is possible to enter a different number of immigrating households for each simulation year.

	Model
Number of Households Immigrating from Within	515
Number of Households Immigrating from Abroad	468

Table 4 – Stock of immigration (one year)

These numbers were obtained from the 2008 census: the proportions of people living in another region in France and those living abroad five years before the interview were used. As households, and not individuals, immigrate, we had to take into account the average size of households in the two pools: 2.53 for the within-country pool and 1.48 for the from-abroad pool.

Data from 2008 census :

(<http://www.recensement.insee.fr/tableauxDetailles.action?zoneSearchField=ILE+DE+FRANCE&codeZone=11-REG&idTheme=10&idTableauDetaille=37&niveauDetail=2>).

⁸ To deal with the fact that, for young couples, high fertility coupled with low ratio of agents in union drove the probability to more than 1.

7. Labour market

The probabilities to enter/exit into/from the labour market by sex, age, and presence of a child under 3 years in the household were obtained from the 2008 French Labour Force Survey.

8. Leaving the parental household

Probability of leaving the parental household by age and sex, estimated from 1999 EHF survey.

9. Union dissolution

Probability of union break-up (for women) by type of union (de-facto union or marriage), age of the woman and age of youngest child, estimated from 1999 EHF survey for France and adjusted to Île-de-France.

10. Union formation

Probability of union formation (for women) by age of the woman and whether the woman lives with her parents estimated from 1999 EHF survey for France and adjusted to Île-de-France.

11. Marriage

Probability that an existing de-facto union turns into a marriage by age of the woman estimated from 1999 EHF survey for France and adjusted to Île-de-France.

Annex III.

MODULES AND PARAMETERS

The parameters data are grouped into .dat files associated to the different modules of the simulation. We present here the details of each parameter, grouped by modules.

A. ORDER OF THE EVENTS IN THE SIMULATION

1. Writing .csv files of initial person and household census; immigrant person and household pools from abroad and from within the country (just once at the beginning of the simulation)
2. Emigration (half)
3. Immigration (half)
4. Mortality
5. Education
6. Labour market participation
7. Leaving parental household
8. Union dissolution
9. Union formation
10. Marriage
11. Fertility
12. New-borns mortality
13. Communal establishments residence
14. Exit from complex households
15. Entry into complex households
16. Emigration (half)
17. Immigration (half)
18. Writing .csv files of person and household census
19. Birthday (increment age for everybody and duration variables where needed)
20. Output tables

Note:, Emigration and immigration are simulated in two times, half at the beginning and half at the end of each year. Thus, instead of applying half the rates of the events to those entering (leaving) the country and the potential emigrants (immigrants), we consider that half the migrants will face all the events for the whole year and the other half will not face any event. As mortality is simulated before fertility, we simulate infant mortality just after fertility.

B. COMMUNAL RESIDENCE [BASE(COMMUNAL).DAT]

Agents can enter into (and exit from) communal establishments (CE: such as nursing homes for elders, students' halls and so on). The model makes no distinction between the different types of CE. People in CEs are not linked to any household. In fact, when an agent is selected to enter a CE, the link with his/her household is severed. If the agent is in a union, the union is dissolved. The only modules applying to people in CEs are: mortality, education and labour market. That is, agents in CEs cannot emigrate, they cannot form a union (and therefore cannot procreate) and they cannot enter into complex households. Agents exiting from CEs are linked to a newly-created household and become subject to all the modules again.

- Probability of entry into a communal establishment.
(double **dProbCommEntry**[ENROLLMENT][MAR_STATUS][AGE_STATUS][SEX_STATUS])
It depends on enrolment status (whether the person is studying or not), marital status, age and sex.
- Probability of exiting from a communal establishment.
(double **dProbCommExit**[ENROLLMENT][MAR_STATUS][AGE_STATUS][SEX_STATUS])
It depends on enrolment status (whether the person is studying or not), marital status, age and sex.

C. ENTRY IN COMPLEX HOUSEHOLD [BASE(COMPLEXENTRY).DAT]

1. How to enter a complex household

Each year, every agent living alone has a probability to enter in a "complex household". The probability of joining another household is higher for those persons having one or more children living in the simulated area. The parameter called *dOutChildMult* (default value: 1.5) multiplies the baseline probabilities for these persons.

Once an agent is selected for entering into a complex household, a fixed number of household defined by the parameter *nComplexPool* (default value = 50) is randomly drawn from all the households having less than 16 members and for each of them a distance indicator is calculated. This "distance" depends on the age difference between the selected agent and the reference person of each household and on the size of each household. If the two persons have both been born abroad, the distance also depends on the difference between the two durations since arrival in the country.

The age difference is grouped in seven groups (0-4, 5-14, 15-24, 25-34, 35-44, 45-55, 55+) and for each of these age groups, the parameter *dCxAgeDiff* determines the weight given to each groups, ranging from 0 to 1. Also the difference of the durations since arrival in the country is grouped in seven groups (0, 1, 2, 3, 4, 5-9, 10+). For each of these age groups, the parameter *dCxDsaDiff* determines the weight. To these is summed the inverse of the size of the target household (so that it is more likely that small households are selected). This quantity (let's call it **q**) is computed for each household in the pool and then all of them are normalized in order to sum up to one and cumulated. A random number is drawn from a (0,1) uniform distribution and the household whose interval corresponds to this number is selected as household where the agent will move.

Additionally, for those household containing a child of the selected agent, an additional weight defined by *dRelativesBoost* is added to **q** (default value = 2).

2. List of parameters

- Probability of entering into a complex household.
(double **dProbComplexEntry**[AGE_STATUS][SEX_STATUS])
It depends on age and sex.
- Weight given to age difference in the matching process. It depends on the age difference between the selected agent and the reference person of the target household.
(double **dCxAgeDiff**[AGE_CX_INT])
Age intervals are fixed, weights are settable. The default values are:

Age difference	[0,4]	[5,14]	[15,24]	[25,34]	[35,44]	[45,55]	[55+]
Weight	1	0.2	0.4	1	0.6	0.2	0.1

- Weight given to difference in durations since arrival in the country. It depends on the duration-since-arrival difference between the selected agent and the reference person of the target household. It only kicks in if both persons are born abroad.

(double dCxDsaDiff[DSA_PART])

Age intervals are fixed, weights are settable. The default values are:

Dsa difference	0	1	2	3	4	[5,9]	[10+]
Weight	1	0.9	0.8	0.7	0.6	0.5	0.25

- Probability multiplier for lone parents having at least a child living in the simulated area
(double dOutChildMult)
It multiplies the baseline probabilities of joining another household (default value = 1.5)
- Additional weight given to household containing a child of the agent selected for joining another household (default value = 2)
(double dRelativesBoost)
- Size of the pool of household eligible for forming a complex household (default value = 50)
(int nComplexPool)

D. EDUCATION [BASE(EDUCATION).DAT]

1. How Level of Education is Imputed

In the initial population, before starting the simulation, the age at end of study for persons aged 15 to 30 is imputed. The default values of this parameter, called *dAesImpute*, are derived from the distribution of AES of people aged 30 to 35. The probabilities of exiting from education for this group at ages 15 to 30 were computed for each age. Table 5 displays the default values of this parameter. This table must be interpreted as follows: a student aged 15 will exit education at 15 with probability 0.081, at 16 with probability 0.091 and so on. Every line must sum up to 1.

		Age at end of studies															
		15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Age	15	0.081	0.091	0.095	0.159	0.083	0.104	0.069	0.078	0.080	0.060	0.040	0.024	0.013	0.011	0.004	0.009
	16	0	0.099	0.103	0.173	0.091	0.113	0.075	0.085	0.087	0.065	0.043	0.026	0.014	0.012	0.004	0.010
	17	0	0	0.115	0.192	0.101	0.125	0.083	0.094	0.096	0.073	0.048	0.029	0.016	0.013	0.005	0.011
	18	0	0	0	0.217	0.114	0.142	0.094	0.106	0.109	0.082	0.054	0.033	0.018	0.015	0.005	0.012
	19	0	0	0	0	0.145	0.181	0.120	0.136	0.139	0.105	0.069	0.042	0.023	0.019	0.007	0.016
	20	0	0	0	0	0	0.212	0.141	0.159	0.162	0.123	0.081	0.049	0.026	0.022	0.008	0.018
	21	0	0	0	0	0	0	0.178	0.202	0.206	0.155	0.102	0.062	0.033	0.028	0.010	0.023

22	0	0	0	0	0	0	0	0	0.246	0.250	0.189	0.124	0.075	0.041	0.034	0.012	0.028
23	0	0	0	0	0	0	0	0	0	0.332	0.251	0.165	0.100	0.054	0.045	0.016	0.037
24	0	0	0	0	0	0	0	0	0	0	0.375	0.247	0.149	0.081	0.067	0.024	0.056
25	0	0	0	0	0	0	0	0	0	0	0	0.395	0.239	0.129	0.108	0.039	0.090
26	0	0	0	0	0	0	0	0	0	0	0	0	0.396	0.214	0.178	0.064	0.148
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0.354	0.295	0.106	0.245
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.457	0.164	0.379
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.302	0.698
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000

Table 5 – dAes Impute parameter

So, the beginning of this parameter definition in the .dat file is:

```
double dAesImpute[POSSIBLE_AES_KIDS][POSSIBLE_AES_KIDS] = {
    0.081, 0.091, 0.095, ...
};
```

The education module imputes the level of education (LOE) of persons at birth, along with the age at end of studies and the ages of transition between the three different types of education levels (i.e. primary, secondary and tertiary education).

The level of education of a person depends on the level of education of his/her parents. The parameter *dLoe* controls this assignment. The default values follow:

Sex	Dad LOE	Mom LOE	Assigned LOE	Probability
Female	Low	Low	Low	0.8
Female	Low	Low	Middle	0.1
Female	Low	Low	High	0.1
Female	Low	Middle	Low	0.3
Female	Low	Middle	Middle	0.5
Female	Low	Middle	High	0.2
Female	Low	High	Low	0.1
Female	Low	High	Middle	0.6
Female	Low	High	High	0.3
Female	Middle	Low	Low	0.3
Female	Middle	Low	Middle	0.2
Female	Middle	Low	High	0.5

Female	Middle	Middle	Low	0.1
Female	Middle	Middle	Middle	0.7
Female	Middle	Middle	High	0.2
Female	Middle	High	Low	0.2
Female	Middle	High	Middle	0.3
Female	Middle	High	High	0.5
Female	High	Low	Low	0.1
Female	High	Low	Middle	0.6
Female	High	Low	High	0.3
Female	High	Middle	Low	0.2
Female	High	Middle	Middle	0.3
Female	High	Middle	High	0.5
Female	High	High	Low	0.1
Female	High	High	Middle	0.1
Female	High	High	High	0.8
Male	Low	Low	Low	0.8
Male	Low	Low	Middle	0.1
Male	Low	Low	High	0.1
Male	Low	Middle	Low	0.3
Male	Low	Middle	Middle	0.5
Male	Low	Middle	High	0.2
Male	Low	High	Low	0.1
Male	Low	High	Middle	0.6
Male	Low	High	High	0.3
Male	Middle	Low	Low	0.3
Male	Middle	Low	Middle	0.2
Male	Middle	Low	High	0.5
Male	Middle	Middle	Low	0.1
Male	Middle	Middle	Middle	0.7
Male	Middle	Middle	High	0.2
Male	Middle	High	Low	0.2
Male	Middle	High	Middle	0.3
Male	Middle	High	High	0.5
Male	High	Low	Low	0.1
Male	High	Low	Middle	0.6
Male	High	Low	High	0.3
Male	High	Middle	Low	0.2
Male	High	Middle	Middle	0.3
Male	High	Middle	High	0.5
Male	High	High	Low	0.1
Male	High	High	Middle	0.1
Male	High	High	High	0.8

Table 6 – dLoe parameter

The model cumulates the "Probability" values for each case, draws a random number from a uniform (0,1) distribution and compares it to the cumulated probability value corresponding to the sex of the agent and the level of education of his/her mother and father. For example, a woman with both parents having a low education level, will be assigned a low, middle or high level of education respectively with probabilities 0.8, 0.1 and 0.1.

Sex	Dad LOE	Mom LOE	Assigned LOE	Probability
Female	Low	Low	Low (0)	0.8
Female	Low	Low	Middle (1)	0.1
Female	Low	Low	High (2)	0.1

So, the beginning of the .dat file is:

```
double dLoe[SEX_STATUS][LEV_OF_EDU][LEV_OF_EDU][LEV_OF_EDU]= {
    0.8, 0.1, 0.1, 0.3, 0.5, 0.2 ...
};
```

Once the level of education is assigned, the age at end of studies (AES), which depends on the level of education, is determined. Three parameters define this assignment (*dAesLow*, *dAesMed*, *dAesHigh*). The possible ages at end of studies assigned in the model go from 15 to 30 years. Users can enter probabilities of exiting at each year. For example, the default values of *dAesLow* are:

Assigned AES	Probability
15	0
16	0.33333
17	0.33333
18	0.33333
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0

Table 7 – *dAesLow* parameter

Here, we see that agents with a low level of education are assigned an AES of 16, 17, or 18 with an equal probability (0.3333).

Therefore, in the .dat file, the beginning of *dAesLow* will be written as:

```
double dAesLow[SIM_TIME][SEX_STATUS][POSSIBLE_AES_KIDS] = {
    0, 0.3333, 0.3333, 0.3333, 0, 0, 0, ...
};
```

Note that in Table 7 the first two dimension are not shown. In fact, these probabilities can also vary with simulation time (50 values, from 0 to 49), in order to allow an evolution of the ages at end of studies by level of education, and with sex (female, male).

By default, agents with a middle LOE are assigned AES from 18 to 22 with equal probability (0.2) and those with a high LOE, AES from 21 to 29, always with equal probability (0.1111). Users can modify these probabilities: the only constraint is the possible AES range, fixed to the range from 15 to 30. Obviously, the sum of the probabilities for each sex must sum up to 1.

Finally, the ages of transition between school are assigned. Four school levels are envisaged (contained in the classification "EDU_CLASS": kindergarten, primary, secondary and tertiary education). The number of ages at transitions is fixed to three for every transition. Here's the content of the parameter in the education .dat file:

```
double dAgePass[EDU_CLASS][NUM_AGE_PASS][PROB_VALUE] = {
    0.1, 2, 0.8, 3, 0.1, 4,
    0.1, 5, 0.8, 6, 0.1, 7,
    0.1, 10, 0.8, 11, 0.1, 12,
    0.1, 18, 0.8, 19, 0.1, 20,
};
```

For each school level, this parameter contains three couples of values composed by the probability that the particular age will be assigned and the value of the age that will be assigned.

For example, the first line applies to the first school level, which is the entry into kindergarten: children will enter kindergarten at age 2 with a probability of 0.1, at age 3 with a probability of 0.8 and at age 4 with a probability of 4. The second line applies to entry into primary school, the third to the entry into secondary school and the third one to the entry into tertiary education.

2. List of parameters

- Probability of level of education assignment depending on parents' level of education
(**double dLoe[SEX_STATUS][LEV_OF_EDU][LEV_OF_EDU] [LEV_OF_EDU]**)
- Age at end of studies assignment probability, low level of education
(**double dAesLow[SIM_TIME][SEX_STATUS][POSSIBLE_AES_KIDS]**)
- Age at end of studies assignment probability, middle level of education
(**double dAesMed[SIM_TIME][SEX_STATUS][POSSIBLE_AES_KIDS]**)
- Age at end of studies assignment probability, high level of education
(**double dAesHigh[SIM_TIME][SEX_STATUS][POSSIBLE_AES_KIDS]**)
- Ages of transitions probabilities (kindergarten -> elementary -> secondary -> tertiary)
(**double dAgePass[EDU_CLASS][NUM_AGE_PASS][PROB_VALUE]**)

- Imputation probabilities of AES for persons aged 15 to 30. (see page 52 for its description)
(**double dAesImpute**[POSSIBLE_AES_KIDS][POSSIBLE_AES_KIDS])

E. EMIGRATION [BASE(EMIGRATION).DAT]

1. Principle of emigration computation

For the emigration module, households are divided into four groups and four different emigration probabilities are applied to them:

1. households with a couple and at least a child;
2. households with no couple and at least a child;
3. households with a couple but no child;
4. household with no couple and no child.

For the first two types of household, the probability can depend on the number of children in the household (from 1 to 15) and from the age of the youngest child (in intervals of one year up to 18 years old and for older than 18 years old: 0-1, 1-2, ..., 17-18, 18+)

For the last two types of household, the emigration probability depends on the number of household members and on the age of the eldest man, if present, or woman, if no man lives in the household.

2. List of parameters

- Probability of emigration for type-1 households
(**double dProbEmig_1**[HH_NUM_CHILDREN][AGE_ONE_YR_INT_18])
- Probability of emigration for type-2 households
(**double dProbEmig_2**[HH_NUM_CHILDREN][AGE_ONE_YR_INT_18])
- Probability of emigration for type-3 households
(**double dProbEmig_3**[AGE_TEN_YR_INT][HH_SIZE_PART])
- Probability of emigration for type-4 households
(**double dProbEmig_4**[AGE_TEN_YR_INT][HH_SIZE_PART])

F. FERTILITY [BASE(FERTILITY).DAT]

1. Principle of Fertility computation

Only women in union aged 15 to 49 years can give birth to a child. The probability of giving birth depends on the woman's level of education, age of youngest child, age and parity.

The age of the youngest child is partitioned in intervals of one year up to 10 years old and for older than 10 years old (0-1, 1-2, ..., 9-10, 10+), while fertile years ("FERT_YEARS") start at age 15 and end at year 49, ends included. Possible parities for mums ("PARITIES") range from 0 to 15.

2. List of parameters

- Probability of childbearing
(**double dProbFert**[LEV_OF_EDU][AGE_ONE_YR_INT_10][SIM_TIME][FERT_YEARS] [PARITIES])

- Probability that the newborn is a female (default = 0.4878)
(**double dProbFemale**)

G. REFERENCE PERSON [BASE(HH_REFPERSON).DAT]

1. Definition

The reference person is only used in the module ComplexEntry (see page 49).

Reference Person: for the time being, only the French definition of the reference person is implemented, i.e.: if there is a couple in the household, the reference person will be the man, otherwise the woman; otherwise, if there is no couple, the reference person will be the oldest active person; otherwise, if there are no active persons, the reference person will be the oldest one.

2. List of parameters

- Id of simulated country, to allow for country-specific definition of reference person in household (0: France, 1: Belgium, 2: Switzerland)
(**int CountryId**)

H. INITIALIZATION [BASE(INITIALIZATION).DAT]

These file simply contains the names of the .csv files used to load the initial population and the two immigrants' pools. Remember that these file must be located in the same folder of the executable file.

- Initial population .csv file name (default: "InitialPopulation.csv")
(**file InitialPopulationFile**)
- Name of .csv file for the pool of potential immigrant from within the country (default: "WithinPool.csv")
(**file WImmigPoolFile**)
- Name of .csv file for the pool of potential immigrant from abroad (default: "AbroadPool.csv")
(**file ABrImmigPoolFile**)

I. IMMIGRATION [BASE(IMMIGRATION).DAT]

1. Principle of Immigration computation

Immigration is defined by the number of households that immigrate each year. One parameter for each type of immigration pool (from other regions in the simulated country or from abroad) must be set. The immigrant household are drawn randomly from the corresponding immigration pool.

Note: the maximum number of households immigrating by year is given by the overall size of the immigration pools. The following table gives the size of each immigration pool.

Model

Within-Country Immigrant Pool

10132

From-Abroad Immigrant Pool

818

Table 8 – Size of immigrant pools (number of households)

2. List of parameters:

- Number of households immigrating from within the country, by simulation year. (**int nWMHouseholds[SIM_TIME]**)
Maximum value: 10132. (default = 515 for all years)
- Number of households immigrating from abroad, by simulation year. (**int nAMHousehold[SIM_TIME]**)
Maximum value: 818. (default = 468 for all years)

J. LABOUR MARKET [BASE(LABOUR).DAT]

This module simulates the entry into/exit from the labour market. Agents in the labour market are set as “active” agents, while agents not in the labour market as “inactive”.

- Probability of entering into the labour market (i.e. working and unemployed) (**double dProbLabEntry[UNION_STATUS][ENROLLMENT][LEV_OF_EDU][ALL_POSSIBLE_AES][ACTIVE_YEARS][HAS_BABY_CHILD][SEX_STATUS]**)
It depends on the agent’s union status (in union or not), enrolment status (studying or not), level of education, age at end of studies (fixed from 9 to 30), age (possible ages from 16 to 70), whether there is a child under 3 years in the household and sex of the person.
- Probability of exiting from the labour market (i.e. inactive and retired) (**double dProbLabExit[UNION_STATUS][ENROLLMENT][LEV_OF_EDU][ALL_POSSIBLE_AES][ACTIVE_YEARS][HAS_BABY_CHILD][SEX_STATUS]**)
It depends on the same covariates of dProbLabEntry.

K. LEAVING THE PARENTAL HOME [LEAVING]

This module simulates the exit from the parental household. If a person is selected for leaving the parental household, a new household is created and he/she is moved in it.

- Probability of leaving the parental household (**double dProbLeaving[UNION_STATUS][WORKING_STATUS][ENROLLMENT][AGE_STATUS][SEX_STATUS]**)
It depends on union status, working status, enrolment, age and sex.

L. MARRIAGE[BASE(MARRIAGE).DAT]

The marriage module applies only to persons already in de-facto union and it is women-driven. Direct marriages (from not-in-union to married) are simulated as persons who form a common-law union and get married in the same year: persons with a union_duration equal to 0 are chosen for marriage. A single parameter determines both direct marriages (for women with union_duration equal to zero) and de-facto unions turning into marriages (for women with union_duration greater than zero).

- Probability of marriage (all these are female's variables)
(double dProbMarriage[NEWBORN][LEV_OF_EDU][PARITY_STATUS][UNION_DURATION_PART][UNION_AGE])
It depends on: having or not a child aged zero ("newborn" = 1 or 0); level of education; number of children (0 to 10+); union duration; age (possible ages are from 15 to 105, ends included).

M. MORTALITY [BASE(MORTALITY).DAT]

This module simulates mortality. If a person dies, he/she is removed from the population and if he/she is in union the partner marital status is set to "separated" when in de-facto unions or "widowed" for legal marriages. A yearly "census" of deceased people can be output, by setting the parameter `produceDeadCensus` to 1 in `Base (Output) .dat`.

The death of one parent can produce orphans. These are the rules followed in case of orphanhood. If no parents are left in the household, then the child:

1. If a surviving parent exists in the simulated population, the child is moved to the household of the surviving parent.
2. If there is at least one person aged 18 and over in the household, then the eldest of these persons becomes the tutor of the child. The child is linked to the tutor through the "father" or "mother" link according to the sex of the tutor. The previous "father" or "mother" link is thus replaced.
3. If none of the previous situations applies, the child is moved to a newly-created household on his own until the `ComplexEntry` module is run. This module links the child to an existing household. If there is a couple in the household, both father and mother links are replaced to link make the couple the "new" parents of the child. If there is no couple, then the child is linked to the eldest person and, depending on the sex of this person, he/she becomes the new father/mother of the child, replacing the corresponding existing link.

N. OUTPUT [BASE(OUTPUT).DAT]

Contains options to turn on (1) or off (0) the production of output census-like comma-delimited .csv files.

- On/off switch for producing a yearly person census, called `PersonCensus.csv`
(int producePersonCensus)
- On/off switch for producing a yearly household census, called `Householdcensus.csv`
(int produceHouseholdCensus)
- On/off switch for producing a person census and a household census for the pool of potential immigrant from within the country, called respectively `WImmigPool.csv` and `WImmigHhPool.csv`
(int produceWImmigrantPool)
- On/off switch for producing a person census and a household census for the pool of potential immigrant from abroad, called respectively `AImmigPool.csv` and `AImmigHhPool.csv`
(int produceAImmigrantPool)
- On/off switch for producing a yearly census of emigrant people, called `EmigCensus.csv`
(int produceEmigrantCensus)

- On/off switch for producing a yearly census of dead people, called `DeadCensus.csv`
(**int produceDeadCensus**)

O. POPULATION SIZES [BASE(POPULATIONSIIZES).DAT]

- Size of the Starting population
(**int StartingPopulationSize**)
It must be set to the number of individuals to be simulated. It determines how many lines are read from the *InitialPopulation.csv* file.
NOTE: Users can, for testing purposes, simulate less than the maximum number of individuals, but, in this case, they must pay attention not to "divide" any household, i.e. they should make sure that all the members of the last household are entered, otherwise some linking error may occur.
- Size of pool of within-country immigrants
(**int nWMigrants**)
It must be set to the number of individuals in the within-country immigrants pool loaded through the *WithinPool.csv* file.
- Size of pool of from-abroad immigrants
(**int nAMigrants**)
It must be set to the number of individuals in the from-abroad immigrants pool, loaded through the *AbroadPool.csv* file.

P. UNION DISSOLUTION [BASE(UNIONDISSOLUTION).DAT]

Union dissolution, as well as union formation, is women-driven. So if a union is selected to end, the male partner is moved to a newly-created household. If the couple has children, they always stay with the mother in the initial household.

- Probability that an union ends (women-driven)
(**double dProbUnionDissolution[LEV_OF_EDU]][UNION_DURATION_PART][UNION_TYPE][UNION_AGE][AGE_ONE_YR_INT_10]**)
It depends on level of education, marital status, age of the youngest child grouped into one-year intervals up to age 10 ([0,1],[1,2],[2,3],..., [9,10],[10+]), union duration (grouped as [0,1],[1,2],..., [9,10],[10,15],[15,20],... [55,60],[60+]) and age of the woman (ages eligible for union dissolution, "UNION_AGE" range from 15 to 105, ends included).

Q. UNION FORMATION [BASE(UNIONFORMATION).DAT]

1. Principle of Union Formation

The union formation module contains a matching procedure similar to that used for the module `ComplexEntry` (see page 50). It is women-driven. All women aged from 15 to 105 and not currently in union are "at risk" of entering a union. The probability of entering into a union depends on marital status, parity and age.

Once a woman is selected for entering into a union, the matching process to find a partner within the population is run. A number of not-in-union men defined by the parameter *nUnionPool* (default value = 50) are randomly selected and for each potential couple a distance, *D*, which depends on differences in age and level of education, is computed. If both potential partners are born abroad, the distance also depends on the

difference in duration since arrival in the country, grouped in seven classes (0, 1, 2, 3, 4, 5-9, 10+): the weight given to each class is determined by the parameter *dCxDsaDiff*. This parameter is in the file *Base(ComplexEntry).dat*.

$$D = \frac{1}{1 + |AgeDiff|} + \frac{nEduWeight}{1 + EduDiff} + dCxDsaDiff$$

where *nEduWeight* and *dCxDsaDiff* are settable by users.

AgeDiff is the age difference between the two potential partners, while *EduDiff* is the difference between the two final level of education of the two potential partners (which are coded 0, 1, 2 for low, medium, high).

Then, all the distances are normalized in order to sum up to one and cumulated. A random draw from a uniform (0,1) distribution determines the chosen partner and a common-law union is formed.

If the woman does not live with her parents nor in a complex household, the man moves to her household. If the woman does not live alone, but the man does, it is the woman who moves to the man's household. If none of the partners live alone, a new household is created and both partners move in it.

2. List of parameters

- Probability of entering into a union
(**double dProbUnionFormation[RED_MAR_STATUS][PARITY_STATUS][UNION_AGE]**)
It depends on marital status ("reduced": meaning only no-union marital status are taken into account, i.e. "never", "separated", "divorced" and "widowed"), parity (0 to 10+) and age (ranging from 15 to 105, ends included).
- Size of the pool of potential matches
(**int nUnionPool**, default = 50).
- Weight of education difference
(**int nEduWeight**, default = 3).
- Weight given to difference in durations since arrival in the country. It depends on the duration-since-arrival difference between the selected agent and the reference person of the target household. It only kicks in if both persons are born abroad.
(**double dCxDsaDiff[DSA_PART]**)
Age intervals are fixed, weights are settable. The default values are:

Dsa difference	0	1	2	3	4	[5,9]	[10+]
Weight	1	0.9	0.8	0.7	0.6	0.5	0.25

This parameter is in the file *Base(ComplexEntry).dat*

Annex IV.

LIST OF GENERAL SCENARIO SETTINGS

- *Simulation end time*: sets the end time of the simulation (the time starting at 0).
- *Simulation threads*: sets the number of cores used in the simulation. This number must be positive and not greater than the number of replicates.
- *Replicates*: number of times the simulation is run. More replicates reduce the Monte Carlo error for simulated quantities, each replicates using a different seed for the random number generator. If the number of replicates is set to a number greater than 1, the values reported in the output tables will be the average value across the replicates. Coefficients of variation and standard errors are also computed.
- *Starting seed*: random number generator(s) seed. As with all pseudo-random number generators, the results will always be identical for the same seed (provided the parameters are unchanged).
- *Computational priority*: settable to Low or Normal, the latter being the faster. This can also be set during the simulation (via **Scenario > Computational priority**).
- *Partial reports after cancellation*: controls the generation of partial reports if the simulation is interrupted by the user. Available options: Yes; No; Prompt.
- *Background mode*: run the simulation only when computer is idle.
- *Partial reports*: stop every x hours to produce partial reports, x being set by the user.
- *Distributed execution*: enable execution on more than one machine, via the Executor application.
- *MS Access tracking*: generate a file named `scenario_name(trk).mdb` which tracks individual life histories and is accessible via the BioBrowser application. MS Access is required in order for BioBrowser to run.
- *Text tracking*: generate a text file named `scenario_name(trk).txt` which records every change to every state variable of every simulated agent.
- *Copy parameter in output database*: include parameters in the exportable items. (As we will see, it is possible to export the generated tables, and, if this option is checked, the input parameters, either to Excel or to a text file, via **Scenario > Export...**)
- *Memory usage report*: generate a memory usage report. It is not of interest for most users.

LIST OF TABLES

TABLE 1 – EXAMPLE OF PARAMETER TABLE	22
TABLE 2 – VALUES OF QUALITATIVE VARIABLES IN .CSV OUTPUT FILES	35
TABLE 3 – POPULATION (NUMBER OF PEOPLE) AND POOL	43
TABLE 4 – STOCK OF IMMIGRATION (ONE YEAR)	46
TABLE 5 – DAES IMPUTE PARAMETER	52
TABLE 6 – DLOE PARAMETER	53
TABLE 7 – DAESLOW PARAMETER	54
TABLE 8 – SIZE OF IMMIGRANT POOLS (NUMBER OF HOUSEHOLDS)	58

LIST OF FIGURES

FIGURE 1 - EVENTS ORDER	14
FIGURE 2 - INPUT FILE NAMES IN BASE(INITIALIZATION).DAT	16
FIGURE 3 - EXTRACT FROM INITIALPOPULATION.CSV	17
FIGURE 4 - OPENING A PARAMETER FILE	18
FIGURE 5 - PARAMETER PROPERTIES - INFO	19
FIGURE 6 - PARAMETER PROPERTIES - DIMENSIONS	19
FIGURE 7 - PARAMETER .DAT FILES SYNTAX	20
FIGURE 8 - EXAMPLE OF PARAMETERS FILE	21
FIGURE 9 - CONTENT OF FILE BASE(COMPLEXENTRY).DAT	22
FIGURE 10 - VISUAL INTERFACE	24
FIGURE 11 - UNIMPORTANT WARNING MESSAGES	24
FIGURE 12 - SCENARIO SETTINGS - GENERAL	25
FIGURE 13 - SCENARIO SETTINGS - PARAMETER FILES	26
FIGURE 14 - SCENARIO SETTINGS - TABLES	27
FIGURE 15 - WRONG DIMENSION OF PARAMETER ERROR - EXAMPLE	28
FIGURE 16 - MODEL RUN PROGRESS	29
FIGURE 17 - OUTPUT TABLE	30
FIGURE 18 - TABLE PROPERTIES	30
FIGURE 19 - EXPORT TO EXCEL SETTINGS	31
FIGURE 20 - EXAMPLE OF OUTPUT .CSV FILE: FIRST LINES OF PERSONCENSUS.CSV	35
FIGURE 21 - ANNUAL MEAN ENTRY AND EXIT RATES, 1990-99, FROM BACCAÏNI (2001) AND PROBABILITIES $S(X,D)$	44
FIGURE 22 - ENTRIES $E(X,D)$ AND DISTRIBUTION OF ENTRIES $PE(X,D)$	44
FIGURE 23 - ENTRIES NOT FOLLOWED BY AN EXIT $PE(X,D)S(X,D)$ AND DISTRIBUTION OF RESIDENTS BY DURATION $P(X,D)$	44
FIGURE 24 - $P(X,1)$: PROBABILITY OF BEING A RECENT IMMIGRANT BY AGE	45