Synthetic populations and microsimulations: applications in social science, celestial dynamics and beyond

Morgane Dumont

LISER, Avril 2023

Motivation

Simulating (statistical) populations largely used :

- Political decisions (health, ageing population, etc)
- Transport
- Epidemic simulations
- Space debris

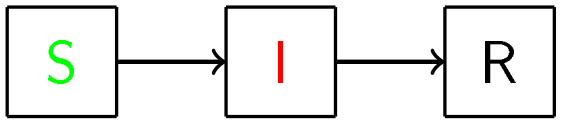
Complexity of the reality:

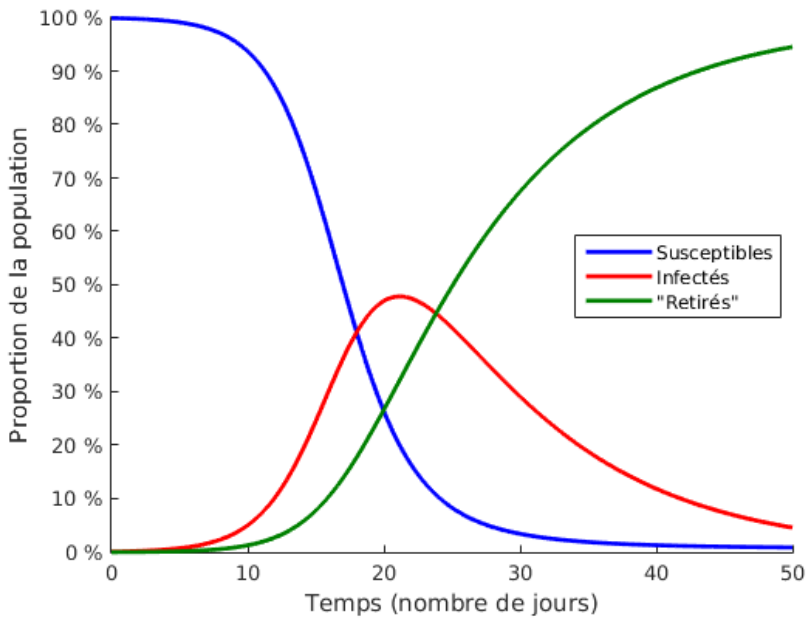
- Not totally predictable
- Need of individual behaviours
- Personal preferences
- Human >< computer

$\Rightarrow$ We can not predict exactly but estimate $\Rightarrow$ Statistic

Macrosimulation VS Microsimulation

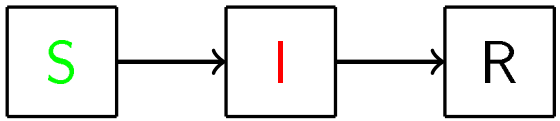
Macrosimulation

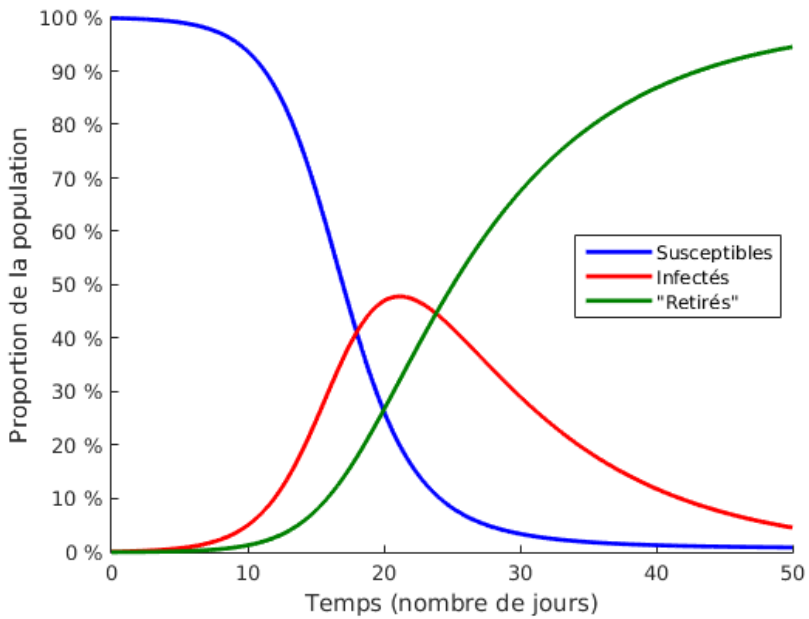




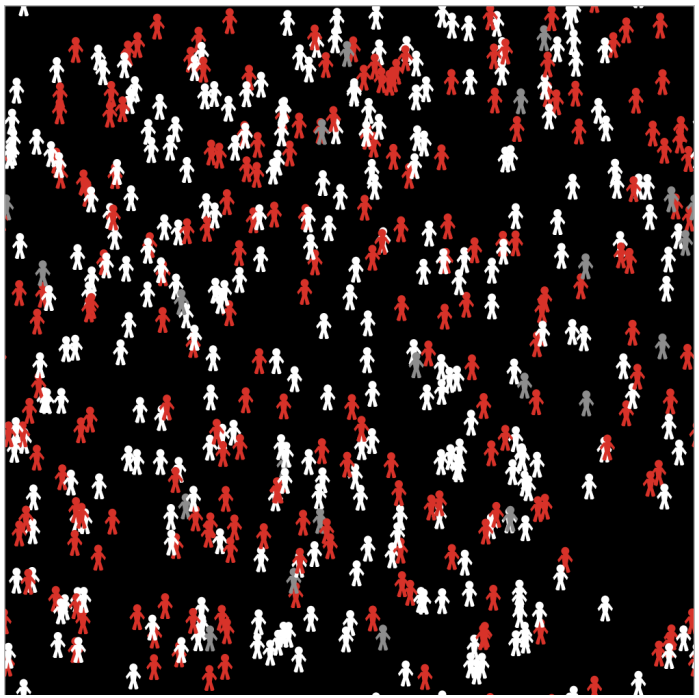
Macrosimulation VS Microsimulation

Macrosimulation





Microsimulation



Source : Paul Smaldino, Netlogo

Why microsimulations?

- Possibility to consider different levels of aggregation

Why microsimulations?

- Possibility to consider different levels of aggregation
- Compare scenarios at different levels (individual, household, municipality,...)

Why microsimulations?

- Possibility to consider different levels of aggregation
- Compare scenarios at different levels (individual, household, municipality,...)
- Individual outputs

Why microsimulations?

- Possibility to consider different levels of aggregation
- Compare scenarios at different levels (individual, household, municipality,...)
- Individual outputs
- Simulate scenarios depending on individual characteristics (ex: level of dependencies for elderly)

Why microsimulations?

- Possibility to consider different levels of aggregation
- Compare scenarios at different levels (individual, household, municipality,...)
- Individual outputs
- Simulate scenarios depending on individual characteristics (ex: level of dependencies for elderly)
- Possibility to "follow" the individuals (trajectories)

Why microsimulations?

- Possibility to consider different levels of aggregation
- Compare scenarios at different levels (individual, household, municipality,...)
- Individual outputs
- Simulate scenarios depending on individual characteristics (ex: level of dependencies for elderly)
- Possibility to "follow" the individuals (trajectories)

$\longrightarrow$ Flexibility

Difficulties

- Microsimulations are often stochastic

Difficulties

- Microsimulations are often stochastic
- Complexity (Need to model individually the possible events, changes etc)

Difficulties

- Microsimulations are often stochastic
- Complexity (Need to model individually the possible events, changes etc)
- Computationally intensive

Difficulties

- Microsimulations are often stochastic
- Complexity (Need to model individually the possible events, changes etc)
- Computationally intensive
- Discrete or continuous time

Difficulties

- Microsimulations are often stochastic
- Complexity (Need to model individually the possible events, changes etc)
- Computationally intensive
- Discrete or continuous time
- Need of input data ($\longrightarrow$ Synthetic populations)

Synthetic populations

Aim

Create a "virtual" static population that best matches the available data.

Synthetic populations

Aim

Create a "virtual" static population that best matches the available data.

Different methods exists depending on:

- the availability of data,

Synthetic populations

Aim

Create a "virtual" static population that best matches the available data.

Different methods exists depending on:

- the availability of data,
- the target structure of the population

Synthetic populations

Aim

Create a "virtual" static population that best matches the available data.

Different methods exists depending on:

- the availability of data,
- the target structure of the population

A simple example : Iterative Proportional Fitting:

- availability of a sample with the target structure
- availability of aggregated counts

Synthetic populations - Example

Sample

	Independent	Dependent	Total (sample)	Aggr. Count
< 60 years old	12	3	15	325
≥ 60 years old	9	23	22	125
Total (Sample)	21	26		
Aggr. Count	145	305		

Administrative data

Synthetic populations - Example

Sample

	Independent	Dependent	Total (sample)	Aggr. Count
< 60 years old	12	3	15	325
≥ 60 years old	9	23	22	125
Total (Sample)	21	26		
Aggr. Count	145	305		

Administrative data

$\longrightarrow$ Simple method but with some issues (zero cells, correlation?, integerisation,...)

Different types of synthetic populations

- Generating "only" independent individuals

Different types of synthetic populations

- Generating "only" independent individuals
- Generating individuals with different layers (for example households, schools,...).

Different types of synthetic populations

- Generating "only" independent individuals
- Generating individuals with different layers (for example households, schools,...).
- With or without samples

Different types of synthetic populations

- Generating "only" independent individuals
- Generating individuals with different layers (for example households, schools,...).
- With or without samples
- Synthetic reconstruction, combinatorial optimization,...

Context

Virtual Belgium In Health (program DGO6 WBHealth) :

- DEMO/ UCLouvain
- UNamur
- 2014-2017

Context

Virtual Belgium In Health (program DGO6 WBHealth) :

- DEMO/ UCLouvain
- UNamur
- 2014-2017

Objectives:

1. Create an evolving synthetic population

Context

Virtual Belgium In Health (program DGO6 WBHealth) :

- DEMO/ UCLouvain
- UNamur
- 2014-2017

Objectives:

1. Create an evolving synthetic population
2. Analyse the health needs of elderlies until 2030

Context

Virtual Belgium In Health (program DGO6 WBHealth) :

- DEMO/ UCLouvain
- UNamur
- 2014-2017

Objectives:

1. Create an evolving synthetic population
2. Analyse the health needs of elderlies until 2030
3. Insert it in a platform available for the "Agence pour une Vie de Qualité" (AVIQ)

Global process

1. Create a static initial Belgian population (grouped into households) for 2011 localised in municipalities

Global process

1. Create a static initial Belgian population (grouped into households) for 2011 localised in municipalities
2. Generate the spatial and temporal evolution of the population with a yearly time step until 2030

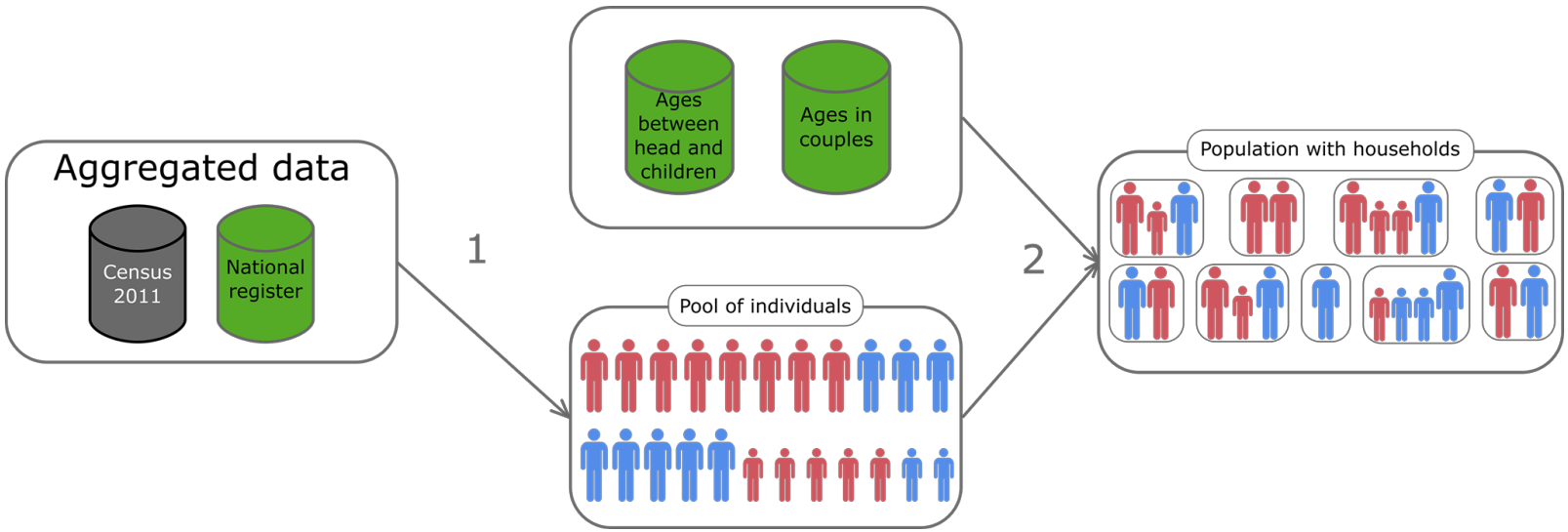
Global process

1. Create a static initial Belgian population (grouped into households) for 2011 localised in municipalities
2. Generate the spatial and temporal evolution of the population with a yearly time step until 2030
3. Quality/stability of the simulations

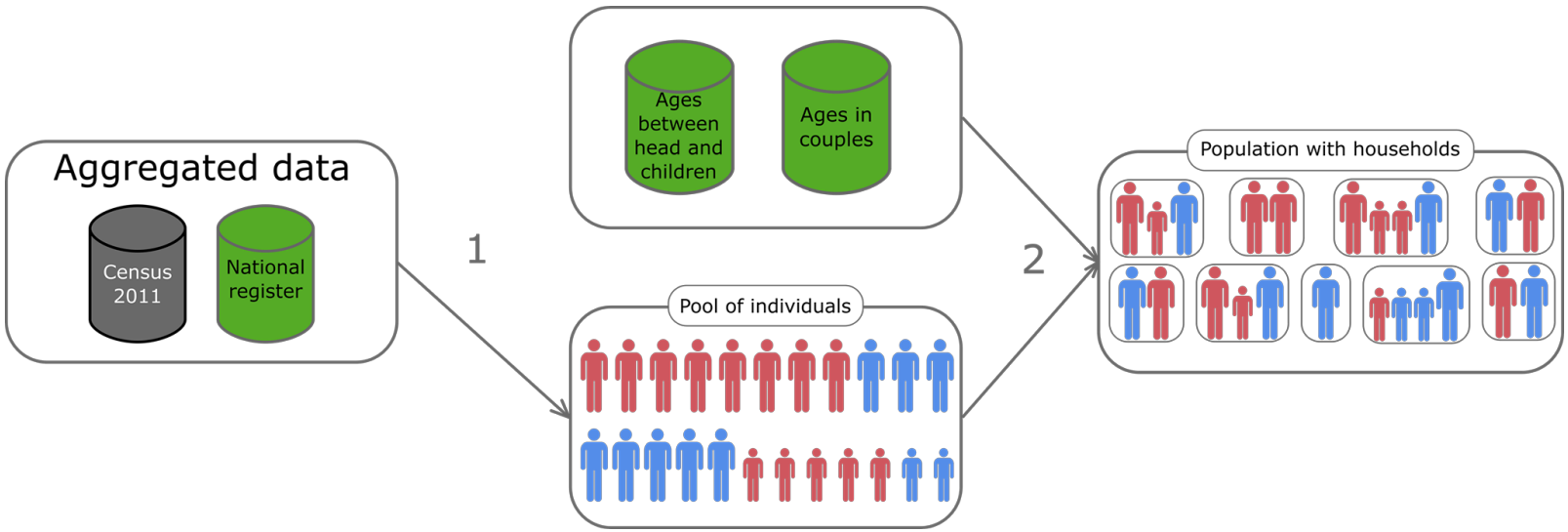
Global process

1. Create a static initial Belgian population (grouped into households) for 2011 localised in municipalities
2. Generate the spatial and temporal evolution of the population with a yearly time step until 2030
3. Quality/stability of the simulations
4. Addition of health data

Virtual Belgium In Health - synthetic static initial population per municipality

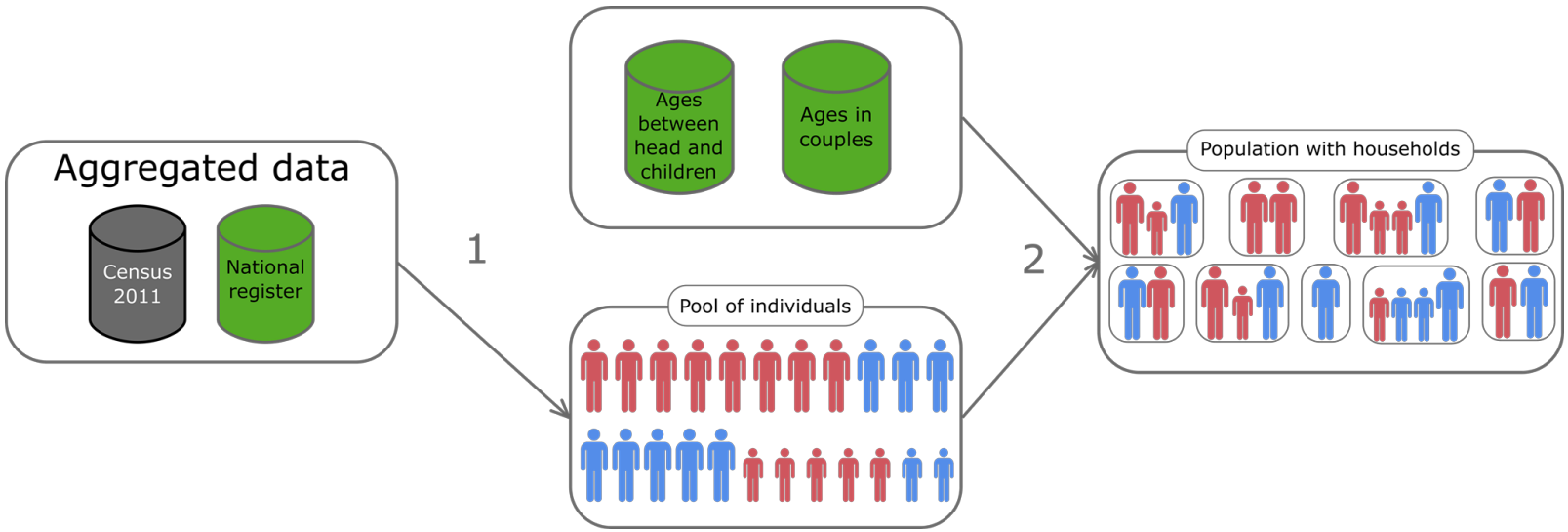


Virtual Belgium In Health - synthetic static initial population per municipality



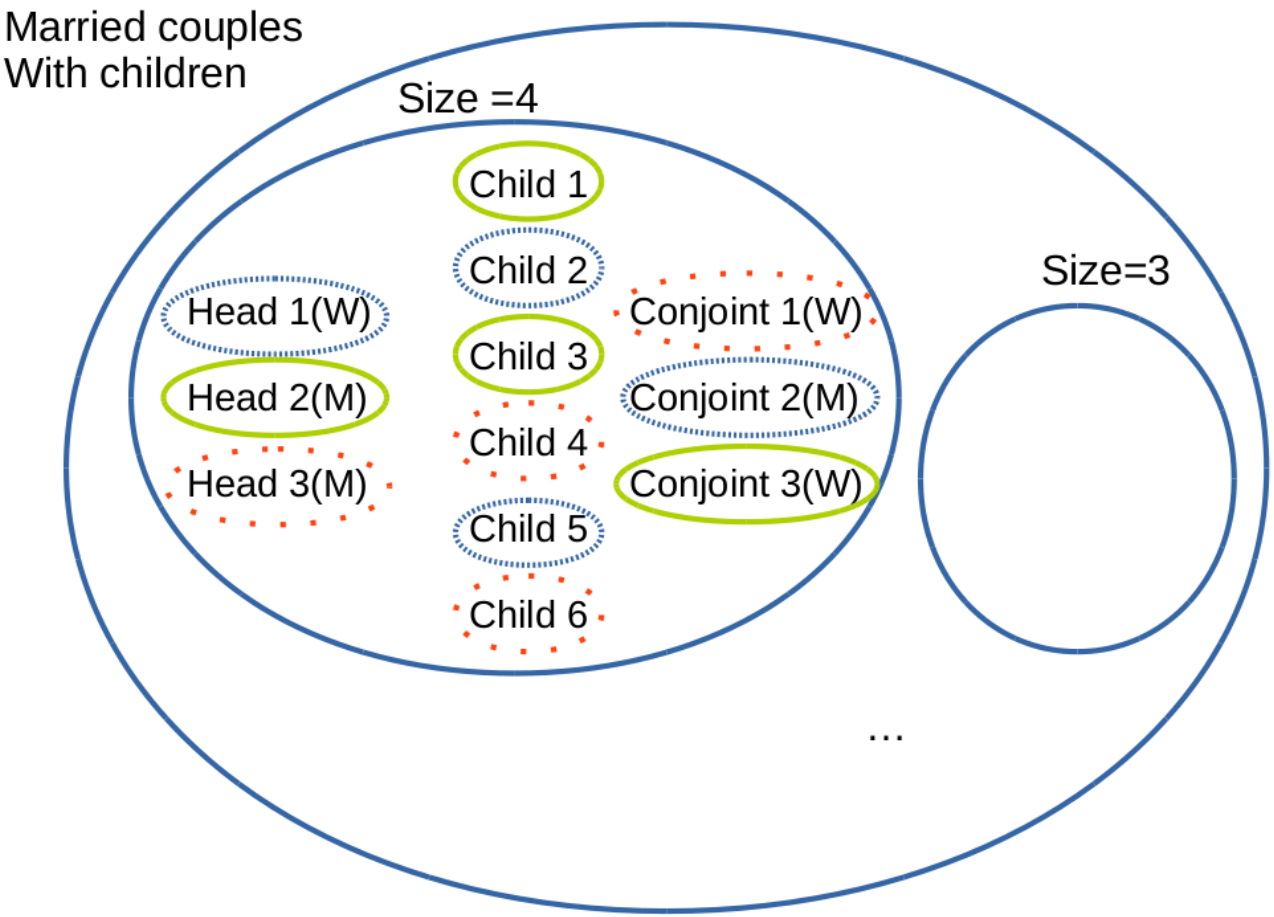
1. Iterative Proportional Fitting with final attributes:
   - municipality; gender; age ; civil status; education level; activity status
   - role in the household ; size of the household; type of household

Virtual Belgium In Health - synthetic static initial population per municipality

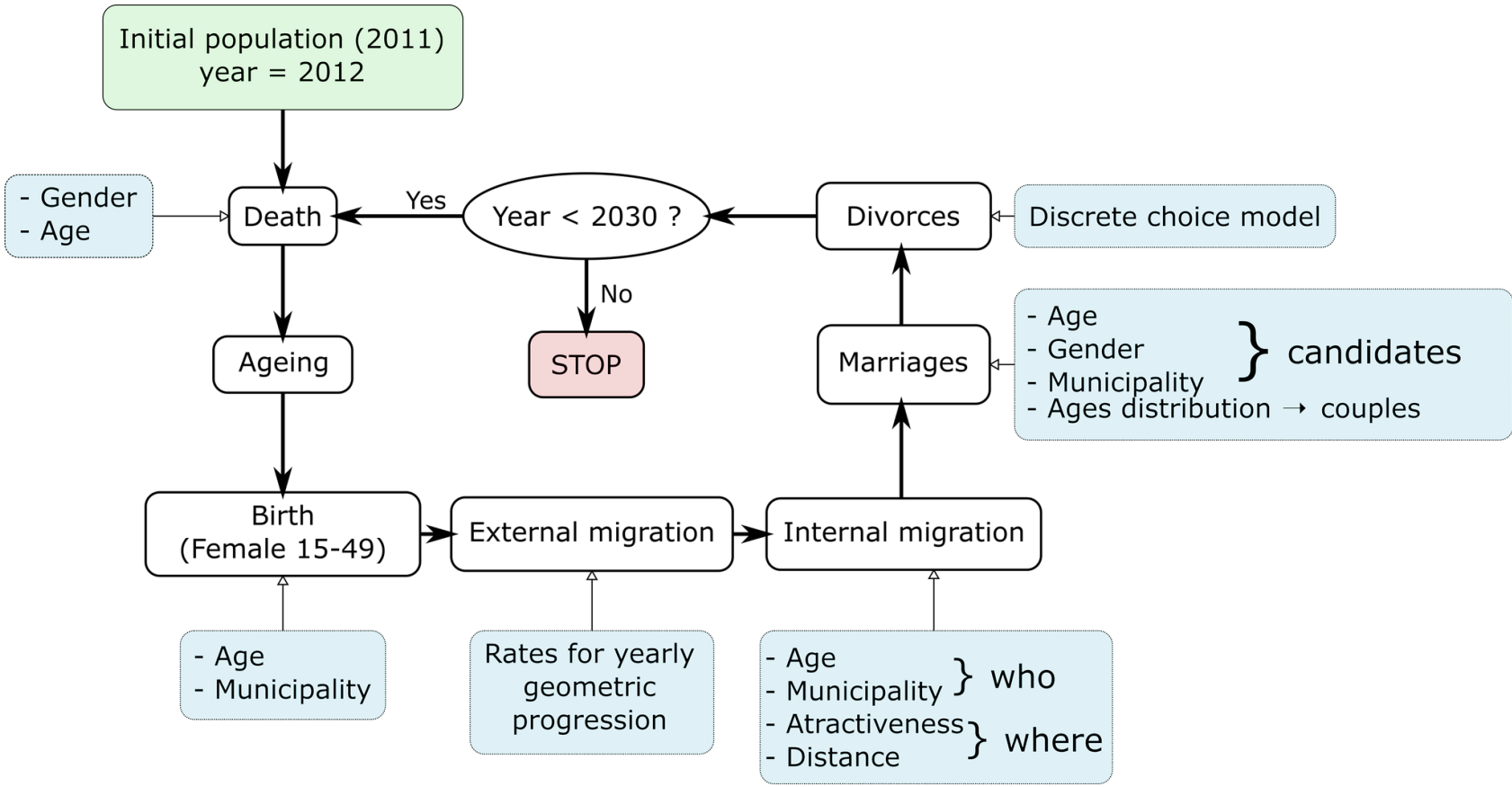


1. Iterative Proportional Fitting with final attributes:
   - municipality; gender; age ; civil status; education level; activity status
   - role in the household ; size of the household; type of household
2. Combinatorial Optimization

Grouping into households - Illustration



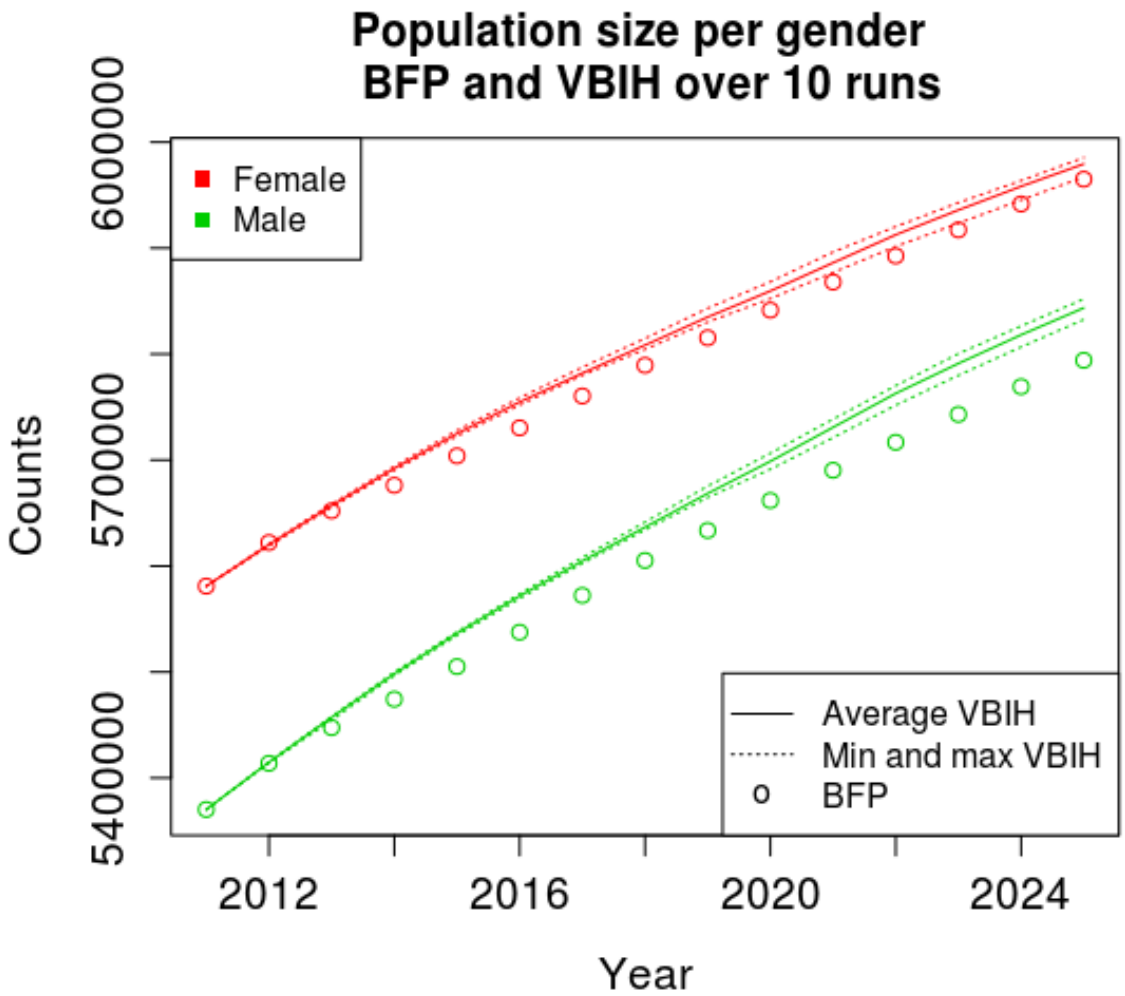
Virtual Belgium In Health : Dynamical Evolution



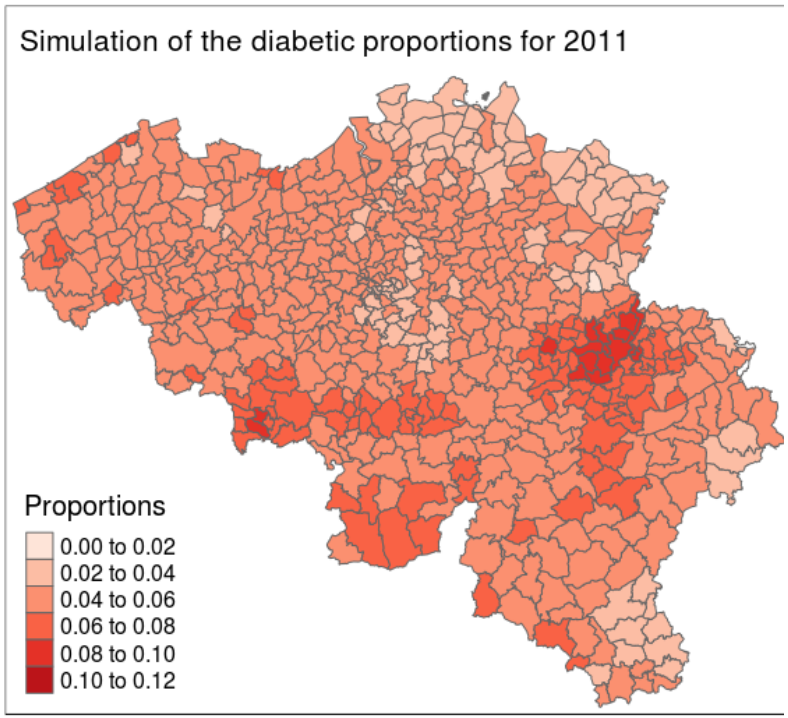
Order discussed with the group DEMO.

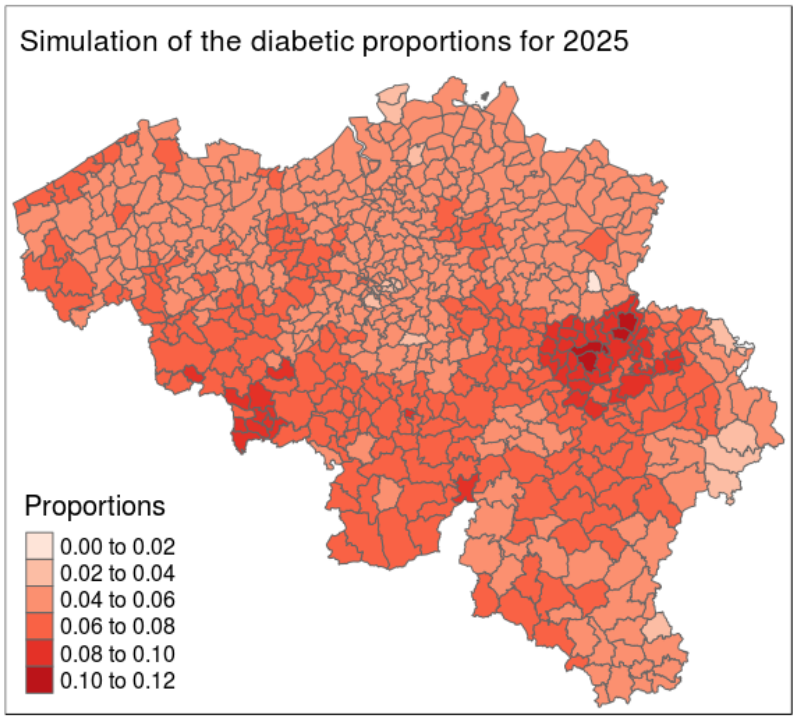
Challenge to launch the execution in parallel

Stability

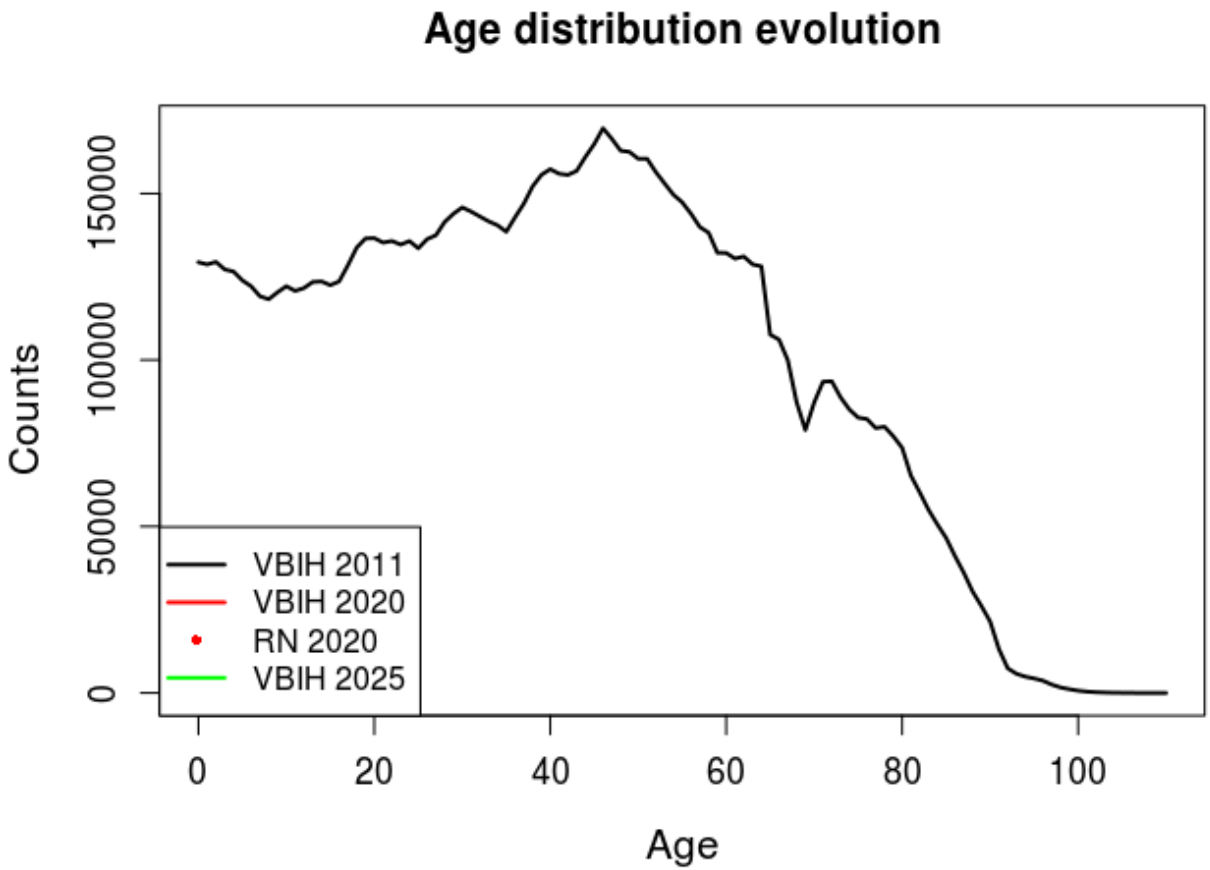


Addition of health data : example

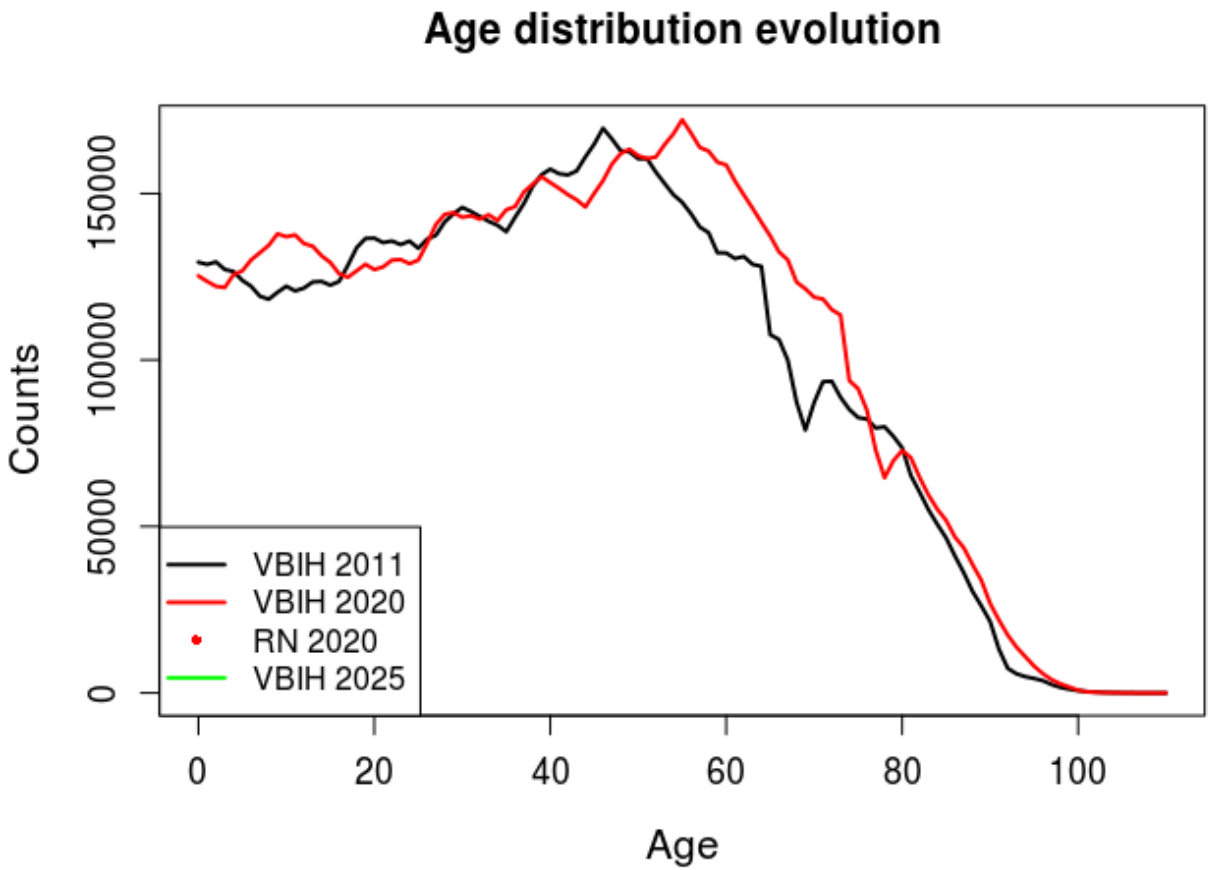




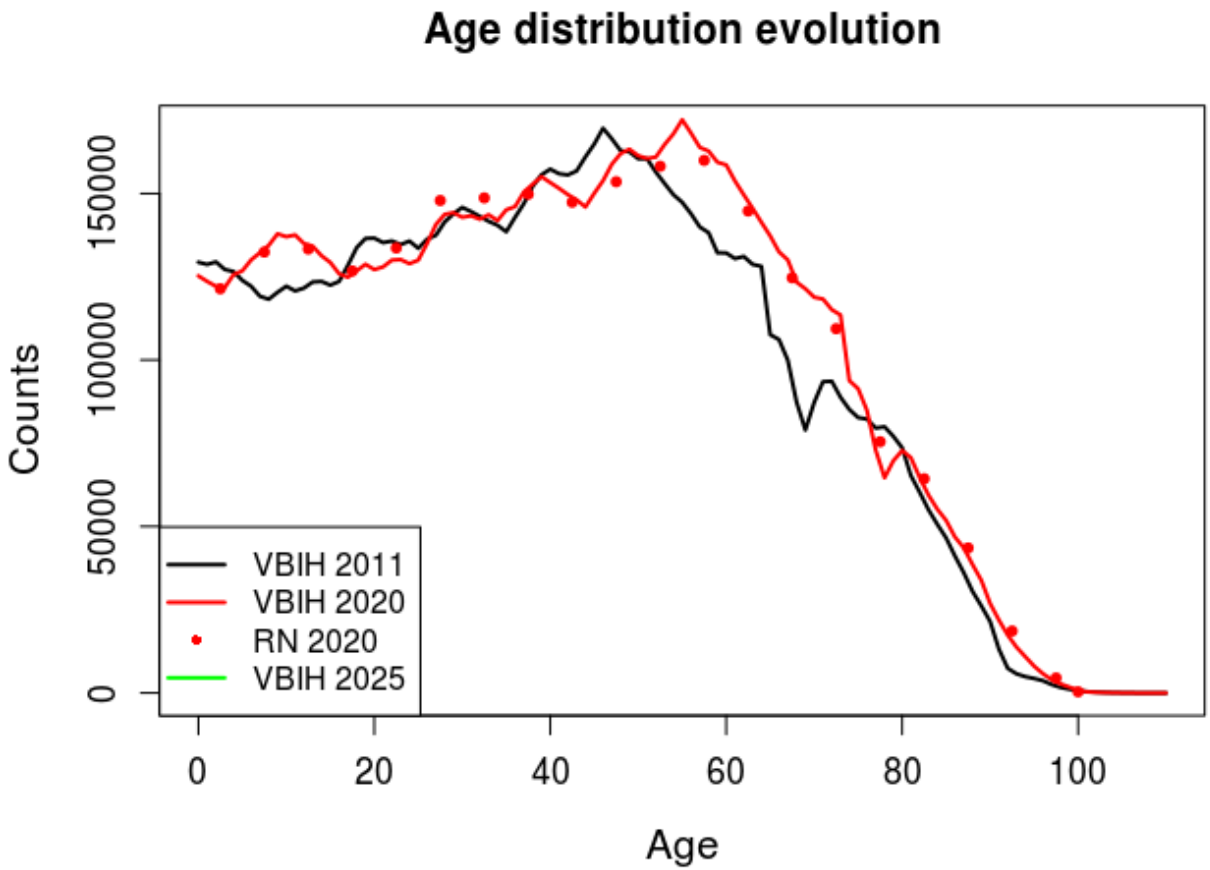
Future ageing population



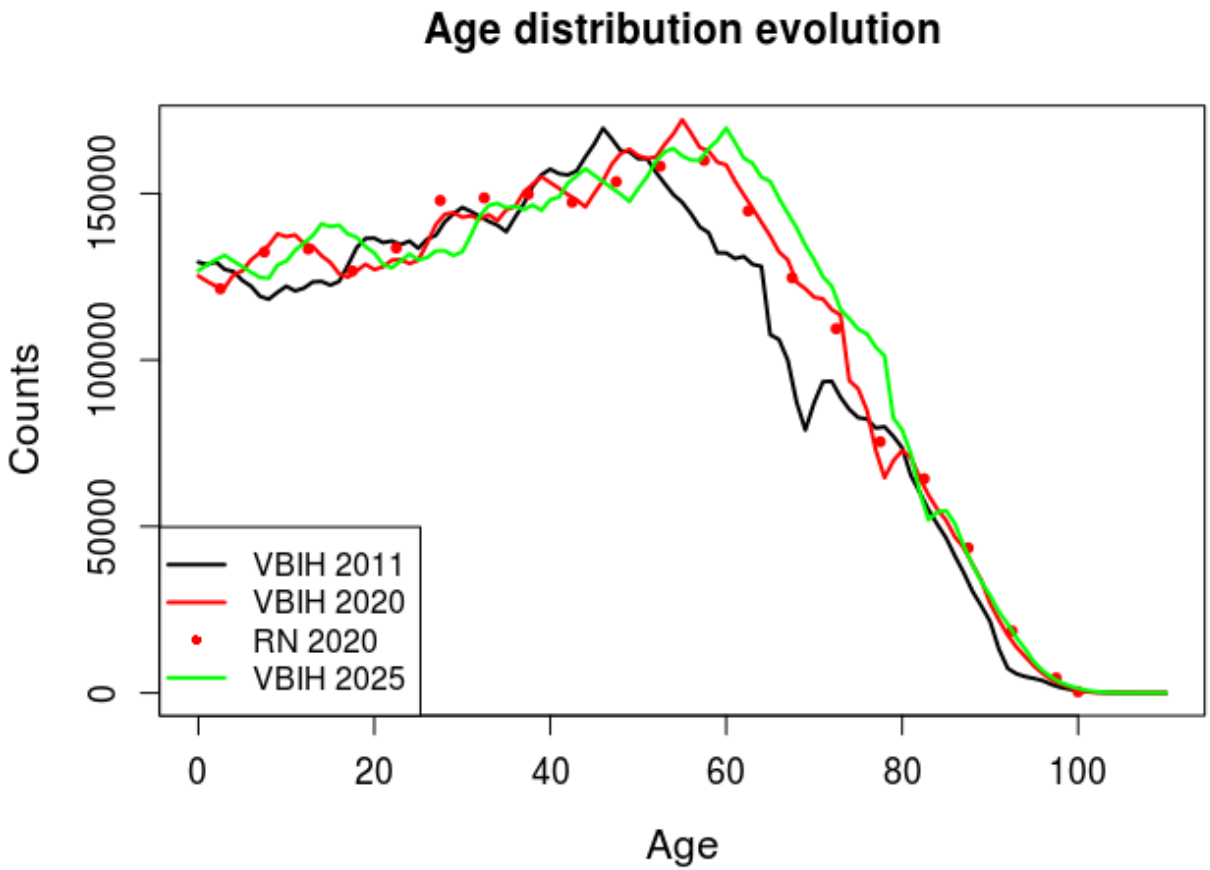
Future ageing population



Future ageing population



Future ageing population



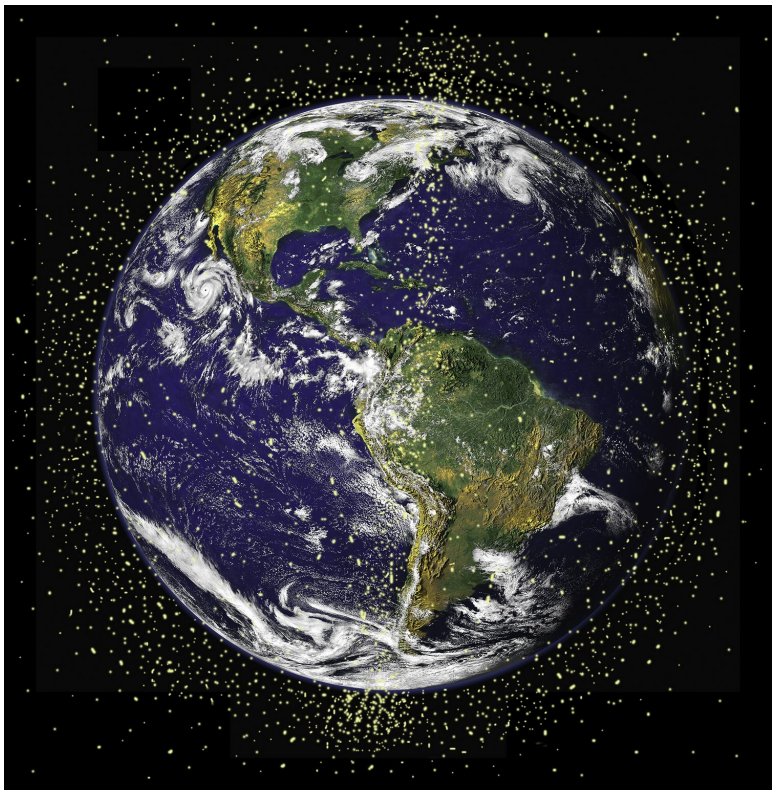
Conclusion for VBIH

A stochastic tool simulating:

- the population
- localised in municipalities
- grouped into households
- With a time step of one year

$\longrightarrow$ Good basis that can be used many different fields (health, transport, ...)

Space debris: Combination of methods



(Source:NASA)

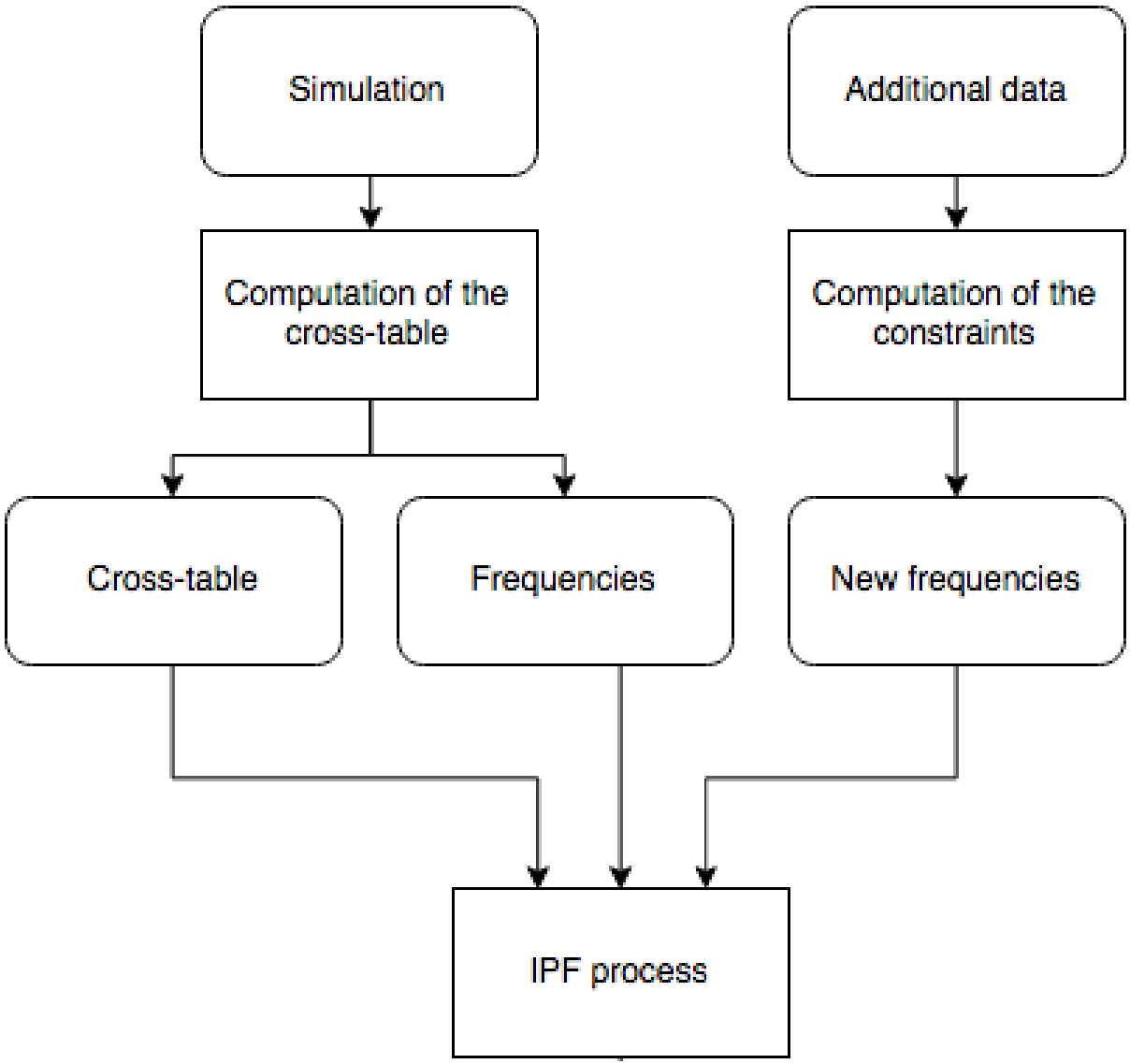
Initial point

Available :

- A break up model
- Observational data (only for biggest and brightest objects)

Petit, A., Casanova, D., Dumont, M., & Lemaitre, A. (2018) *Creation of a synthetic population of space debris to reduce discrepancies between simulation and observations.* Celest. Mech. & Dyn. Astr., 130(12), [79].

A synthetic population of space debris



Populations, space debris, what else?

All fields with "individuals", data and lack in data
$\longrightarrow$ synthetic populations.

Populations, space debris, what else?

All fields with "individuals", data and lack in data
$\longrightarrow$ synthetic populations.

All fields with "individuals", data and wishes to model individual scenarios
$\longrightarrow$ Microsimulation

Populations, space debris, what else?

All fields with "individuals", data and lack in data
$\longrightarrow$ synthetic populations.

All fields with "individuals", data and wishes to model individual scenarios
$\longrightarrow$ Microsimulation

Combinations micro/macrosimulation

Populations, space debris, what else?

All fields with "individuals", data and lack in data
$\longrightarrow$ synthetic populations.

All fields with "individuals", data and wishes to model individual scenarios
$\longrightarrow$ Microsimulation

Combinations micro/macrosimulation

Addition of a synthetic techniques to improve existing methods

Populations, space debris, what else?

All fields with "individuals", data and lack in data
$\longrightarrow$ synthetic populations.

All fields with "individuals", data and wishes to model individual scenarios
$\longrightarrow$ Microsimulation

Combinations micro/macrosimulation

Addition of a synthetic techniques to improve existing methods

BUT this is not always the best option of course.

My research focus

- Decision making : Health, pension, transport, ...
- Microsimulation
- Performance of methods
- Calibration of parameters
- Time definition in the model
- Modelling of modules
- Applications

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"
- Many different possible configurations

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"
- Many different possible configurations
- Testing the quality of the simulations?

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"
- Many different possible configurations
- Testing the quality of the simulations?
- So dependent on the goal and data available.

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"
- Many different possible configurations
- Testing the quality of the simulations?
- So dependent on the goal and data available.
- No universal solution $\longrightarrow$ trade-off

Take home message

Microsimulation and synthetic populations : promising tools for different topics.

With challenges/advantages :

- "Garbage in, Garbage out"
- Many different possible configurations
- Testing the quality of the simulations?
- So dependent on the goal and data available.
- No universal solution $\longrightarrow$ trade-off
- Privacy

Thank you for your attention.
Questions?