The MiniDemographicABM.jl (V 1.1)

Specification of a simplified agent-based demographic model of the UK

Atiyah Elsheikh

July 28, 2023

Abstract

This document presents adequate formal terminology for the mathematical specification of a simplified non-calibrated agent-based demographic model of the UK. Individuals of an initial population are subject to ageing, deaths, births, divorces and marriages. The main purpose of the model is to explore and exploit capabilities of the state-of-the-art Agents.jl Julia package [1]. Additionally, the model can serve as a base model to be adjusted to realistic large-scale socio-economics, pandemics or social interactions-based studies mainly within a demographic context. A specific simulation is progressed with a user-defined simulation fixed step size on a hourly, daily, weakly, monthly basis or even an arbitrary user-defined clock rate.

List of todos for model documentation:

- sufficient citations

1 Terminology

This section introduces the basic formal terminologies employed throughout the article for the specification of agent-based models and their simulation process.

1.1 Populations P, M and F

Given that $F(t) = F^t \equiv F$ ($M(t) = M^t \equiv M$) is the set of all females (males) in a given population $P(t) = P^t =\equiv P$ at an arbitrary time point t where

$$P = M \cup F \tag{1}$$

t is omitted for the sake of simplification. t is sometimes placed as a superscript, i.e. F^t, purely for algorithmic specification readability purpose. Similarly, individuals $p \in P(t)$ can be attributed with time, i.e. p^t, referring to that individual at a particular time point.

1.2 Population features $\mathcal{F}$

Every individual $p \in P$ is attributed by a set of features related to gender, location, age among others. The elementary population features f considered in the presented model are related to:

- age, e.g. particular age group e.g. neonates, children, teenagers, renters, thirties, etc.
- alive status, i.e. whether alive or dead
- gender, i.e. male or female
- location, e.g. sub-population of a particular town
- kinship status or relationship, e.g. father-ship, parents, orphans, divorcee, singles etc.

1.3 Featured sub-populations P_f , $f \in \bigcup \mathcal{F}$

Let P_f corresponds to the set of all individuals who satisfy a given feature f where

$$f(p \in P) = b \in \{true, false\} \tag{2}$$

That is

$$P_f = \{p \in P \text{ s.t. } f(p) = true\} \tag{3}$$

For example

- $M = P_{male}$
- $W_{married}$ corresponds to the set of all married women
- $P_{age>65}$ corresponds to all individuals of age older than 65

For the given set of features specified in Section 1.2, a subset of features

$$F' = \{f'_1, f'_2, \dots f'_m\} \;\subset F \tag{4}$$

is called a closed subset of elementary features, if the overall population constitutes of the union of the underlying elementary featured sub-populations, i.e.

$$P = P_{\mathcal{F}'} \equiv P_{f'_1 \cup f'_2 \cup \dots \cup f'_m} = P_{f'_1} \cup P_{f'_2} \cup \dots \cup P_{f'_m} \tag{5}$$

For example, male and female gender features constitute a closed set of elementary features.

1.4 Non-elementary features $\bigcup \mathcal{F}$

For a set of elementary features $\mathcal{F}$ (informally those which demands only one descriptive predicate), the set of all non-elementary features $\bigcup \mathcal{F}$ is defined as follows. A non-elementary feature

$$f^* \in \bigcup \mathcal{F} \;\text{ where }\; \mathcal{F} \subset \bigcup \mathcal{F}$$

can be recursively established from a finite number of arbitrary elementary features

$$f_i, f_j, f_k, ... \in \mathcal{F}$$

by

- union (e.g. $f_i \cup f_j$),
- intersection (e.g. $f_i \cap f_j$)
- negation (e.g. $\neg f_i$)
- exclusion or difference (e.g. $f_i - f_j$)

Formally, if

$$f^* = f_i \;o\; f_j \text{ where } o \in \{\cup, \cap, -\} \text{ and } f_i, f_j \in \bigcup \mathcal{F}$$

then

$$P_{f^*} = P_{f_i \;o\; f_j} = \{p \in P \;s.t.\; p \in P_{f_i} \;o\; P_{f_j}\} \tag{6}$$

Analogously,

$$P_{\neg f} = \{p \in P \;s.t.\; p \notin P_f\} \;\text{ with } f \in \bigcup \mathcal{F} \tag{7}$$

Negation operator s.a. ($\neg$) is beneficial for sub-population specification, e.g.

$$F_{married \;\cap\; \neg hasChildren} \tag{8}$$

corresponds to all married females without children. This sub-population can be equivalently described using the difference operator:

$$F_{married\ -\ hasChildren} \tag{9}$$

which entails to be a matter of style unless algorithmic execution details of the operators are assumed. For instance, it can be assumed that in intermediate computation the $-$ operator is operated directly on the set of married females rather than the set of all females.

Generally, any of the features f_i and f_j in Equation 6 can be either elementary or non-elementary and the definition is recursive allowing the construction of an arbitrary set of non-elementary features. For example, the sub-population

$$M_{divorced\ \cap\ hasChildren\ \cap\ age>45\ -\ hasSiblings} \tag{10}$$

corresponds to the set of all divorced men of age older than 45 who has no siblings but they have children. In order to improve readability, equation 10 can be re-written as:

$$M_{divorced}\ \cap\ M_{hasChildren}\ \cap\ M_{age>45}\ -\ M_{hasSiblings} \tag{11}$$

Both styles can be mixed together for readability purpose, cf. Section 7.3.

1.5 Composition operator $f(g)$

Another beneficial operator is the composition operator analogously defined as

$$P_{f(g)} = \{p \in P_f \text{ s.t. } g(p) = true\} \tag{12}$$

Based on the definition, the composition operator is not symmetric as the case with the intersection operator. For example,

$$M_{isAlive(isSibling)}\ \neq\ M_{isSibling(isAlive)}$$

The left hand side refers to the set of siblings of the alive male population, while the right hand side refers to alive siblings of the male populations. Moreover, the composition operator can be regarded to be more computationally efficient in comparison with the intersection operator[1].

The desired sub-population specification in the example given by equation 11 may not correspond to the desired specification. Namely, desired is to specify the

[1]In this work, the main purpose behind the composition operator mainly remains in the context of algorithmic specification rather than enforcing any implementation details regarding computational efficiency

alive divorced male population older than 45 years with alive children and alive siblings. In this case, the employment of the composition operator is relevant:

$$M_{isAlive(isDivorced)} \ \cap \ M_{|children(isAlive)|>0 \ \cap \ age>45 \ - \ |sibling(isAlive)|>0} \quad (13)$$

Nevertheless, to retain the desired simplicity, the previous equation can be rather rewritten as

$$M_{isAlive(isDivorced \ \cap \ hasAliveChildren \ \cap \ age>45 \ - \ hasAliveSibling)} \quad (14)$$

2 Temporal operators

This section introduces further operators, inspired by the field of temporal logic. These operators provide powerful capabilities for algorithmic specification of time-dependent complex phrases in a compact manner. This section is concerned with defining those operators employed within the context of the demonstrated example model described starting from Section 4. The demonstrated operators in this section shall be included in the set of non-elementary features $\bigcup \mathcal{F}$ defined in Equations 6 and 7.

2.1 just operator

A special operator is

$$just(P_f) \subseteq P_f \ , \ f \in \bigcup \mathcal{F}$$

standing for a featured subpopulation established by an event that has just occurred (in the current simulation iteration). So for instance,

$$P^{t+\Delta t}_{just(married)}$$

stands for those individuals who just got married in the current simulation iteration with a fixed step size Δt but they were not married in the previous iteration, i.e.

$$P^{t+\Delta t}_{just(married)} \ = \ P^{t+\Delta t}_{married} \ - \ P^{t}_{\neg married}$$

Formally,

$$P^{t+\Delta t}_{just(f)} = P^{t+\Delta t}_{f} \ - \ P^{t}_{\neg f} \quad (15)$$

The just operator provides capabilities for powerful specification when combined with the negation operator. For example,

$$P^{t+\Delta t}_{just(\neg married)}$$

stands for those who "just" got divorced or widowed.

2.2 pre operator

Another distinguishable operator is

$$pre(P_f) \ , \ f \in \bigcup \mathcal{F}$$

standing for "the previous iteration". So for instance,

$$P^{t+\Delta t}_{pre(married)}$$

stands for those individuals who were married (and not necessarily just got married) in the previous simulation iteration

$$P^{t+\Delta t}_{pre(married)} \ = \ P^{t}_{married}$$

Formally,

$$P^{t+\Delta t}_{pre(f)} = P^{t}_{f} \tag{16}$$

This operator may look unnecessary excessive, however cf. Section 7.3 as an example for the usefulness of the *pre* operator.

In this work, temporal operators is assumed to extend their applicability to individuals and their attributes. For instance

$$pre(location(p \in P)$$

stands for the location of a person in the previous iteration (which can be the same in the current iteration), e.g. cf. Section 7.3.

3 General form

3.1 Definitions

This article is concerned with formalizing an agent-based model simulation formally defined via the tuple

$$< \mathcal{M}, \alpha_{sim}, \mathcal{F}, \mathcal{M}^{t_0}, \mathcal{E} > \tag{17}$$

based on a demographic time-dependent model:

$$\mathcal{M} \equiv \mathcal{M}(t) \equiv \mathcal{M}^t \equiv \mathcal{M}(P, S, \alpha, D, t) \tag{18}$$

where

- $\underline{P = P(t)}$: a given population of agents (i.e. individuals) at time t evaluated via the model $\mathcal{M}(t)$

- $\underline{S = S(t) =< H(t), W >}$: the space on which individuals $p \in P$ are operating, i.e. the set of houses $H(t)$ distributed within the set of towns W, cf. Section 5 for further detailed insights

- $\underline{\alpha}$: time-independent model parameters, , cf. Section A.1

- $\underline{D(t)}$: input data integrated into the model as (possibly smoothed) input trajectories, cf. Section A.2

- $\underline{\alpha_{sim} = (\Delta t, t_0, t_{final}, \alpha_{meta})^T}$: simulation parameters including a fixed step size and final time-step after which simulation process stops

- $\underline{\alpha_{meta}}$: Implementation-dependent simulation parameters, e.g. simulation seed for random number generation

The rest of mathematical symbols are defined in the following subsections.

3.2 Featured sub-populations (via $\mathcal{M}_{f^*}$)

$\underline{\mathcal{F} = \{f_1, f_2, f_3, ...\}}$: a finite set of elementary features each distinguishes a featured sub-population

$$P_f(t) \subseteq P(t) \ , \ f \in \mathcal{F}$$

as defined in Equations 2 and 20 , cf. Section 1.3. Each featured sub-population $P_f(t)$ is evaluated by the submodel

$$f(\mathcal{M}) \equiv f(\mathcal{M}^t) \equiv \mathcal{M}_f^t = \mathcal{M}_f(P_f, S, \alpha, D, t) \tag{19}$$

evaluating or predicting the sub-population

$$f(P(t)) \equiv P_f(t) \ \text{ s.t. } \ \forall p \in P_f(t) \implies f(p) = true \tag{20}$$

Note that this definition extends to non-elementary features as well:

$$f^*(\mathcal{M}) \equiv \mathcal{M}_{f^*}(P_{f^*}, S, \alpha, D, t) \text{ for any } f^* \in \bigcup \mathcal{F} \tag{21}$$

Such non-elementary features, cf. Section 1.4, are used to distinguish sub-populations needed for describing the transient processes in agent-based modeling simulation process, cf. Section 7.

For a given closed set of elementary features as given in Equation 4, the overall population is the union of the elementary features, cf. Equation 5. In that case the comprehensive model $\mathcal{M}$ constitutes of the sum of its elementary featured submodels:

$$\mathcal{M} \equiv \sum_{f' \in \mathcal{F}'} \mathcal{M}'_f \tag{22}$$

3.3 Initial population and space (via $\mathcal{M}^{t_0}$)

$\underline{\mathcal{M}^{t_0}}$: a model that evaluates an initial space and a population at a proposed simulation start time t_0. The initial model also specifies featured sub-populations via:

$$\mathcal{M}_f^{t_0}, \quad \forall f \in \mathcal{F} \tag{23}$$

Consequently, both the corresponding initial population and featured sub-populations:

$$P(t_0) \text{ and } P_f(t_0), \quad \forall f \in \mathcal{F} \tag{24}$$

are specified as well as the initial space:

$$S(t_0) \ = \ < H(t_0) \ , \ W > \tag{25}$$

i.e., the distribution of an initial set of houses $H(t_0)$ within a set of towns W.

3.4 Events $\mathcal{E}$

$\underline{\mathcal{E} = \{e_1, e_2, e_3, ..., e_n\}}$: a finite set of events that transients a particular set of sub-populations evaluated by

$$\mathcal{M}_{f^*}(t), \quad \text{for some } f^* \in \bigcup \mathcal{F}$$

to modified sub-populations predicted by

$$\mathcal{M}_{f^*}(t + \Delta t)$$

Formally,

$$e(\mathcal{M}_{f^*}(t)) = \mathcal{M}_{f^*}(t + \Delta t) \text{ for some } f^* \in \bigcup \mathcal{F} \ , \ e \in \mathcal{E} \tag{26}$$

The appliance of all events transients the model to the next state

$$\prod_{i=1}^{n} e_i(\mathcal{M}(t)) = \mathcal{M}(t + \Delta t) \tag{27}$$

3.5 Single-clocked fixed-step simulation process

An agent-based simulation process follows the following pattern:

$$\sum_{t=t_0}^{t_{final}} \prod_{i=1}^{n} e_i(\mathcal{M}(t)) \tag{28}$$

Illustratively, the evolution of the population and its featured sub-populations is defined as a sequential application of the events transitions:

$$\mathcal{M}(t_0) \text{ evaluating } (P(t_0), P_f(t_0)) \; \forall f \in \mathcal{F} \overset{\mathcal{E}}{\Longrightarrow}$$

$$\mathcal{M}(t_0 + \Delta t) \text{ evaluating } (P(t_0 + \Delta t), P_f(t_0 + \Delta t)) \;\overset{\mathcal{E}}{\Longrightarrow}$$

$$\mathcal{M}(t_0 + 2\Delta t) \text{ evaluating } (P(t_0 + 2\Delta t), P_f(t_0 + 2\Delta t)) \;\overset{\mathcal{E}}{\Longrightarrow}$$

$$\ldots\ldots$$

$$\mathcal{M}(t_{final}) \text{ evaluating } (P(t_{final}), P_f(t_{final}))$$

4 Model example

In this and the following sections, a model example is introduced to demonstrate the descriptive capabilities of the proposed formal terminology.

4.1 Overview

The model is concerned with demographic agent-based model, a simplified demographic-only version of the lone parent model introduced in [2]. The presented model evolves an initial population of the UK through a combination of events listed in alphabetical order as follows:

- ageing, cf. Section 7.1
- births, cf. Section 7.2
- deaths, cf. Section 7.3
- divorces, cf. Section 7.4
- marriages, cf. Section 7.5

The population evolution follows Equation 28.

Establishing a mathematical model that corresponds to reality till the tiniest details is impossible. Therefore, initially a set of (potentially non-realistic) assumptions has to be made in order to simplify the model specification process. There are mainly two set of assumptions:

- population-based assumptions (to be labeled with P)
- space-based assumptions (to be labeled with S)

4.2 Population assumptions

The population assumptions are summarized as follows:

P. 1 There are no homeless individuals:

$$\text{if } p \in P^t_{isAlive} \implies house(p) \in H(t) \tag{29}$$

P. 2 (In- and out-) immigration is not included:

$$\text{if } p \in P^{t'}_{isAlive} \text{ where } t_0 < t' \leq t_{final} \implies \\ pre(town(p)) \in W \text{ and } town(p^{t'}) \in W \tag{30}$$

P. 3 Major demographic events s.a. world wars and pandemics are not considered

P. 4 Any two individuals living in a single house are either a 1-st degree relatives, step-parent, step-child, step-siblings or partners

P. 5 An exception to the previous assumption occurs when an orphan's oldest sibling is married

4.3 Space assumptions

Resoundingly from Section 3.1, the space S is composed of a tuple

$$S \equiv S(t) = < H(t) , W >$$

corresponding to the set of all houses $H(t)$ and towns W implying that:

S. 1 the space is not necessarily static and particularly the set of houses can vary along the simulation time span

S. 2 the set of towns is constant during a simulation, i.e. no town vanishes nor new ones get constructed

S. 3 each town $w \in W$ contains a dynamic set of of houses $H_w \equiv H_w(t)$

Furthermore,

S. 4 each house $h \in H(t)$ is located in one and only one town $w \in W$, i.e.

$$town(h) = w \in W$$

S. 5 the location of each house $h \in H_w$ is given in xy-coordinate of the town

$$location(h_{x,y} \in H_w) = (x, y)_w$$

S. 6 the houses within a town are uniformly distributed along the x and y axes

S. 7 a house never get demolished and remains always inhabitable

5 The space – detailed description

In the sake of comprehensive description of the space, further assumptions are listed as follows:

S. 8 The static set of towns of UK, cf. Assumption **S. 2**, are projected as a rectangular 12×8 grid with each point in the grid corresponding to a town

Formally, assuming that

$$location(w_{(x,y)}) = (x, y)$$

then

S. 9 the town $w_{(1,1)}$ corresponds to the north-est west-est town of UK whereas

S. 10 the town $w_{(12,8)}$ corresponds to the south-est east-est town of UK

S. 11 the distances between towns are commonly defined, e.g.

$$\text{manhattan-distance}(w_{(x_1,y_1)}, w_{(x_2,y_2)}) = \mid x_1 - x_2 \mid + \mid y_1 - y_2 \mid \tag{31}$$

The (initial) population and houses distribution within UK towns are approximated by am ad-hoc pre-given UK population density map. The map is projected as a rectangular matrix

$$M \in R^{12\times 8} \approx \begin{bmatrix} 0.0 & 0.1 & 0.2 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.2 & 0.2 & 0.3 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.2 & 0.3 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.2 & 1.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.0 & 0.2 & 0.2 & 0.4 & 0.0 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.0 & 0.3 & 0.8 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.6 & 0.8 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 1.0 & 0.8 & 0.6 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.2 & 1.0 & 0.6 & 0.3 & 0.4 \\ 0.0 & 0.0 & 0.5 & 0.7 & 0.5 & 1.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.4 & 0.6 & 1.0 & 1.0 & 0.0 \\ 0.0 & 0.2 & 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \tag{32}$$

It can be observed for instance that

- cells with density 0 (i.e. realistically, with very low-population density) don't correspond to inhabited towns
- the towns in UK are merged into 48 towns
- e.g. the center of the capital London spans the cells $(10, 6), (10, 7), (11, 6)$ and $(11, 7)$

S. 12 if an empty house h is demanded in a particular town $w \in W$, an empty house is randomly selected from the set of existing houses W_w in that town. If no empty house exists, a new empty house is established in conformance with assumption **S. 6**

S. 13 if an empty house h is demanded in an arbitrary town, a town is selected via a random weighted selection:

$$town(h) = random(W, M^T) \tag{33}$$

an empty house is selected or established according to the previous assumption

Further details on the initial set of houses is given in Section 6.6.

6 Model initialization $\mathcal{M}^{t_0}$

This section provides a detailed description of the initial model state as evaluated by M^{t_0} given that Section A demonstrates potential case studies specifying possible simulation parameter values for t_0, t_{final} and Δt. Further initialization assumptions are proposed by demand, distinguished by the labels **P0** for initial population assumptions or **S0** for initial space assumptions.

6.1 Initial population size $|P_w^{t_0}|$

The initial population size is given by the parameter $\alpha_{initialPop}$, cf. Section A.1. The matrix M, defined in Equation 32, provides a stochastic ad-hoc estimate of the initial population $P(t_0)$ distribution within the UK as well as the initial set of given houses $H(t_0)$. That is, the initial population size of a town $w \in W$ is approximated by

$$|P_w(t_0)| \approx \alpha_{initialPop} \times M_{y,x}/48 \quad \text{where} \quad \text{location}(w) = (x, y) \tag{34}$$

where 48 is the number of nonzero entries in M.

6.2 Gender $P^{t_0} = M^{t_0} \cup F^{t_0}$

The parameter $\alpha_{initialPop}$ specifies the size of initial population, cf. Section A for potential values. The gender ratio distribution is unrealistically specified via the following non-realistic assumption:

P. 6 An individual can be equally a male or a female[2]

[2] In reality, worldwide there is a tiny higher number of females births over males births

Gender assignment is established according to a uniform distribution, i.e.

$$Pr(isMale(p \in P^{t_0})) \approx 0.5 \tag{35}$$

6.3 Age distribution $P^{t_0} = \bigcup_r P^{t_0}_{age=r}$

The proposed non-negative age distribution of population individuals in years follows a normal distribution:

$$\frac{age(P^{t_0})}{N_{\Delta t}} \in \mathbb{Q}_+^{\alpha_{initialPop}} \;\; \propto \;\; \left| \lfloor \mathcal{N}(0, \frac{100}{4} \cdot N_{\Delta t}) \rfloor \right| \tag{36}$$

where $\mathbb{Q}_+$ stands for the set of positive rational numbers and $\mathcal{N}$ stands for a normal distribution with mean value 0 and standard deviation depending on

$$N_{\Delta t} \;\; = \;\; \begin{cases} \;\; \ldots \\ \;\; 12 & \text{if } \;\; \Delta t = month \\ 365 & \text{if } \;\; \Delta t = day \\ 365 \cdot 24 & \text{if } \;\; \Delta t = hour \\ \;\; \ldots \end{cases} \tag{37}$$

A possible outcome of the distribution of ages in an initial population of size 1,000,000 is shown in Figure 1.

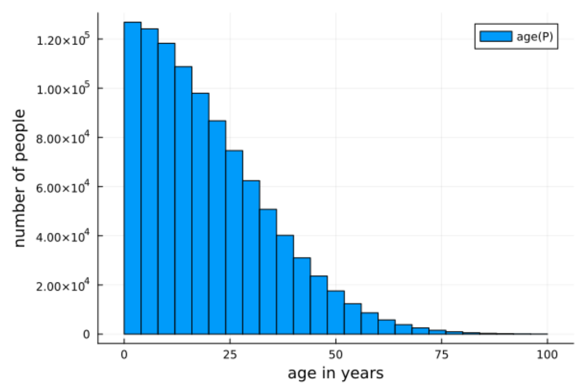

Figure 1: The distribution of ages in an initial population of size 1,000,000

6.4 Partnership $P^{t_0}_{isMarried} = M^{t_0}_{isMarried} \cup M^{t_0}_{partners}$

Initially the following population assumption is proposed:

P. 7 There is no grandpa or grandma for any individual in the initial relationship, i.e.

$$p \in P^{t_0}_{isChild} \equiv P^{t^0}_{age \leq 18} \implies \nexists q \in P^{t_0} \text{ s.t. } grandchild(q) = p \tag{38}$$

The ratio of married adults (males or females) is stochastically approximated according to

$$Pr(p \in P^{t_0}_{isAdult}) \approx \alpha_{startMarriedRate} \tag{39}$$

Before the partnership initialization, the following variables are initialized:

1. set $F_{isMarriageEligible} = F_{isMarEli} = F^{t_0}_{isAdult}$
2. set $n_{candidates} = max\left(\alpha_{maxNumMarrCand}\ ,\ \frac{|F_{isMarEli}|}{10}\right)$

For every male $m \in M^{t_0}_{isMarried}$ selected for marriage, his wife is selected according to following steps:

3. establish a random set of $n_{candidates}$ female candidates:

$$F_{candidates} = random(F_{isMarEli}, n_{candidates})$$

4. for every candidate, $f \in F_{isMarEli}$, evaluate a marriage weight function:

$$weight(m, f) \ = \ ageFactor(m, f) \tag{40}$$

where

$$ageFactor(m, f) = \begin{cases} 1/(age(m) - age(f) - 5 + 1) & \text{if } \ age(m) - age(f) \geq 5 \\ -1/(age(m) - age(f) + 2 - 1) & \text{if } \ age(m) - age(f) \leq -2 \\ 1 \quad \text{otherwise} \end{cases} \tag{41}$$

5. select a random female associated with the evaluated weights

$$f_{partner(m)} = weightedSample(F_{isMarEli}, W_m) \\ \text{where } \ W_m = \{w_i : w_i = weight(m, f_i)\ ,\ f_i \in F_{isMarEli}\} \tag{42}$$

6. $F_{isMarEli} = F_{isMarEli} - \{f_{partner(m)}\}$

6.5 Children and parents

The following assumptions are assumed only in the context of the initial population:

P0. 8 There is no orphan

P0. 9 There is no age difference restrictions among siblings, i.e. age difference can be less than 9 months

Children are assigned to married couples as parents in the following way. For any child $c \in P^{t_0}_{age<18}$, the set of potential fathers is established as follows:

$$\begin{aligned} M_{candidates} = \; \{ \; & m \in M_{isMarried} \text{ s.t.} \\ & min(age(m), age(wife(m)) \geq age(c) + 18 + \frac{9}{12} \text{ and} \\ & age(wife(m)) < 45 + age(c) \; \} \end{aligned} \tag{43}$$

out of which a random father is selected for the child:

$$\begin{aligned} father(c) =& random(M_{candidates}) \text{ and} \\ & mother(c) = wife(father(c)) \end{aligned}$$

6.6 Spatial distribution

The assignment of newly established houses to initial population considers the assumptions **S. 12** and **S. 13**. That is, the location of new houses in $H(t_0)$ is specified according to to Equation 33. Furthermore, the following assumption is proposed:

P0. 10 any house in $H(t_0)$ either occupied by a single person or a family

$$\begin{aligned} |occupants(h^{t_0})| > 1 \text{ with } p, q \in occupants(h^{t_0}) \text{ and } p \neq q \implies \\ p \in firstDegRelatives(q) \end{aligned}$$

Assignments of new houses to the initial population is conducted as follows:

$$\begin{aligned} & p^{t_0} \in P^{t_0}_{isSingle} \implies occupants(house(p)) = \{p\} \text{ otherwise} \\ & m^{t_0} \in M^{t_0}_{isMarried} \implies \\ & \quad occupants(house(m)) = \{m, wife(m)\} \cup children(m) \end{aligned} \tag{44}$$

Overall, all houses in $H(t_0)$ are occupied, i.e. $|occupants(h^{t_0})| \geq 1$.

7 Events

This section provides a compact algorithmic specification as a demonstration of the proposed terminology. The employed set of parameters and input data trajectories is given in Appendix A. The algorithmic specification makes use of rates and instantaneous probabilities conceptually reviewed in Appendix B.

The considered events are alphabetically listed in this section without enforcing a certain appliance order, except for the ageing event which should proceed any other events. That is, Equations 27 and 28 are constrained by setting

$$e_1 = ageing$$

The execution order of the events as well as the order of the agents subject to such events, whether sequential or random, remains an implementation detail. Nevertheless, since many of the events, specified in the following subsections, are following a random stochastic process, probably, the higher the resolution of the simulation becomes (e.g. weekly step-size instead of monthly, or daily instead of weekly), the less influential the execution order of the events becomes.

7.1 Ageing

Following the terminology introduced so far, ageing process of a population can be described as follows

$$ageing\left(P^{t}_{isAlive(age=a)}\right) = P^{t+\Delta t}_{isAlive(age=a+\Delta t)} \quad , \quad \forall a \in \{0, \Delta t, 2\Delta t, ...\} \tag{45}$$

The age of any individual as long as he remains alive is incremented by Δt for each simulation step. Furthermore, the following assumption is considered

P. 9 In case a teenager orphan becomes an adult and he/she is not the oldest sibling, the orphan gets re-allocated to an empty house within the same town, formally:

$$\begin{aligned} &ageing\left(P_{isAlive(age=18)\ \cap\ isOrphan\ \cap\ hasOlderSibling}\right) = \\ &\qquad P_{isAlive(age=18+\Delta t)\ \cap\ isOrphan\ \cap\ hasOlderSibling\ \cap\ livesAlone} \end{aligned} \tag{46}$$

Moreover,

$$\begin{aligned} &\text{If } p \in P^{t+\Delta t}_{isAlive(age=18)} \text{ and } pre(house(p)) \neq house(p) \implies \\ &\qquad town(p) = pre(town(p)) \end{aligned} \tag{47}$$

The re-allocation to an empty house is in conformance with assumption **S. 12**.

7.2 Births

For simplification purpose, from now on it is implicitly assumed (unless specified) that only the alive population is involved in event-based transition of population specification. Let the set of reproducible females be defined asi.e.

$$F_{reproducible} = F_{isMarried \,\cap\, age<45} \bigcap F_{youngestChild(age>1) \,\cup\, \neg hasChildren} \quad (48)$$

That is, the set of all married females in a reproducible age and either do not have children or those with youngest child older than one. The specification of a birth event demands enhancing the population-related assumptions as follows:

P. 7 a neonate's house is his mother house:

$$f \in F_{youngestChild(age=0)} \implies house(youngestChild(f)) = house(f) \quad (49)$$

P. 8 only a married female[3] gives birth

$$age(youngestChild(f \in F)) = 0 \implies isMarried(f) = true \quad (50)$$

P. 9 a married person is not a teenager

$$isMarried(p \in P) = true \implies age(p) \geq 18 \quad (51)$$

Assumptions **P. 8** and **P. 9** implies that only an adult person can become a parent. The birth event produces new children from reproducible females:

$$\begin{aligned} birth\left(F^{t}_{reproducible}\right) = & \\ & \left(F^{t+\Delta t}_{reproducible} - F^{t+\Delta t}_{just(reproducible)}\right) \bigcup \\ & F^{t+\Delta t}_{just(\neg reproducible)} \bigcup \\ & P^{t+\Delta t}_{age=0} \end{aligned} \quad (52)$$

As an illustration based on the *just* operator illustrated in Section 2.1

$$F_{just(reproudicble)} = F_{just(isMarried)} \cup F_{isMarried(youngestChild(age=1))} \quad (53)$$

and (given Assumption **P. 8**)

$$\begin{aligned} F_{just(\neg reproducible)} = & \\ & F_{youngestChild(age=0)} \cup F_{just(devorced)} \cup F_{isMarried(age=45)} \end{aligned} \quad (54)$$

The yearly-rate of births produced by the sub-population $F^{t}_{reproducible,(age=a)}$ i.e. reproducible females of age a years old with actual simulation time t, depends on the yearly-basis fertility rate data:

$$R_{birth,yearly}(F^{t}_{reproducible,(age=a)}) \propto D_{fertility}(a, currentYear(t)) \quad (55)$$

cf. fertility rate data in A.2. This implies that the instantaneous probability that a reproducible female $f \in F_{reproducible}(t)$ gives birth to a new individual $p \in P^{t+\Delta t}$ depends on $D_{fertility}(a, currentYear(t))$ and is given by Equation 73, cf. Appendix B.

[3]This was already assumed in the lone parent model and obviously the marriage concept needs to be re-defined in the context of realistic studies

7.3 Deaths

The death event transforms a given population of alive individuals as follows:

$$death\left(P^{t}_{isAlive}\right) \;=\; P^{t+\Delta t}_{isAlive-age=0} \bigcup P^{t+\Delta t}_{just(\neg isAlive)} \tag{56}$$

The first phrase in the right hand side stands for the alive population except neonates and the second stands for those who just became dead. The following simplification assumptions are considered:

P. 10 No adoption or parent re-assignment to orphans is established after their parents die

P. 11 Those who just became dead they leave their houses, i.e.

$$\text{if } p \in P^{t+\Delta t}_{just(\neg isAlive)} \;\text{ and }\; pre(house(p)) = h$$
$$\implies\; p \notin P_h \;\text{ and }\; house(p) = grave \tag{57}$$

The amount of population deaths depends on the yearly probability given by:

$$Pr_{death,yearly}(p \in P) = \alpha_{baseDieRate} + \begin{cases} \left(e^{\frac{age(p)}{\alpha_{maleAgeScaling}}}\right) \times \alpha_{maleAgeDieProb} & \text{if } isMale(p) \\ \left(e^{\frac{age(p)}{\alpha_{femaleAgeScaling}}}\right) \times \alpha_{femaleAgeDieProb} & \text{if } isFemale(p) \end{cases} \tag{58}$$

from which instantaneous probability of the death of an individual is derived as illustrated in Appendix B.

7.4 Divorces

The divorce event causes that a subset of married population becomes divorced:

$$divorce(M^{t}_{isMarried}) \;=\; M^{t+\Delta t}_{isMarried-just(isMarried)} \bigcup M_{just(\neg isMarried)} \tag{59}$$

The first phrase in the right hand side refers to the set of married individuals who remained married excluding those who just got married. The second phrase refers to the population subset who just got divorced in the current iteration. Note that it is sufficient to only apply the divorce event to either the male or female sub-populations. After divorce takes place, the housing is specified according to the following assumption:

P. 12 Any male who just got divorced moves to an empty house within the same town (in conformance with assumption **S. 12**):

$$location(m \in M_{just(isDivorced)}) = h \;\text{ and }\; pre(location(m)) = h' \implies |occupants(h)| = 1 \;\text{ and }\; town(h) = town(h') \tag{60}$$

The re-allocation to an empty house is in conformance with assumption **S. 12**. The amount of yearly divorces in married male populations depends on the yearly probability given by

$$Pr_{divorce,yearly}(m \in M^t_{isSingle}) \; = \; \alpha_{basicDivorceRate} \; \cdot \; D_{divorceModifierByDecade}(\lceil age(m)/10 \rceil) \quad (61)$$

That is, the instantaneous probability of a divorce event to a married man $m \in M_{isMarried}$ depends on $D_{divorceModifierByDecade}(\lceil age(m)/10 \rceil)$, cf. Equation 73.

7.5 Marriages

Similar to the divorce event, it is sufficient to apply the marriage event to a sub-population of single males. Assuming that

$$M_{isMarEli} \; = \; M_{isMarriageEligible} \; = \; M_{isSingle \, \cup \, age \geq 18} \quad (62)$$

the marriage event updates the state of few individuals within a sub-population to married males, formally:

$$marriage(M^t_{isMarEli}) \; = \; M^{t+\Delta t}_{isMarEli - just(isDivorced) - age=18} \bigcup M^{t+\Delta t}_{just(isMarried)} \quad (63)$$

The amount of yearly marriages is estimated by

$$Pr_{marraige,yearly}(m \in M^t_{isSingle}) \; = \; \alpha_{basicMaleMarriageRate} \; \cdot \; D_{maleMarriageModifierByDecade}(\lceil age(m)/10 \rceil) \quad (64)$$

from which simulation-relevant instantaneous probability is calculated as given in Equation 73. For an arbitrary just married male $m \in M^{t+\Delta t}_{just(married)}$, his partner was selected according to following steps:

- set $n_{candidates} = max\left(\alpha_{maxNumMarrCand} \; , \; \frac{|F_{isMarEli}|}{10}\right)$
- establish a random set of $n_{candidates}$ female candidates:

$$F_{candidates} = random(F_{isEliMar}, n_{candidates})$$

- for $m \in M_{isMarEli}, \; f \in F_{isMarEli}$, set a marriage weight function:

$$weight(m, f) \; = \; geoFactor(m, f) \; \cdot \; childrenFactor(m, f) \; \cdot \; ageFactor(m, f) \quad (65)$$

where

$$
\begin{aligned}
& geoFactor(m,f) = \\
& \quad 1/e^{(4\cdot \text{manhattan-distance}(town(m),town(f)))} && (66)\\
& childrenFactor(m,f) = \\
& \quad 1/e^{|children(m)|} \cdot 1/e^{|children(f)|} \cdot e^{|children(m)|\cdot|children(f)|} && (67)\\
& ageFactor(m,f) = \\
& \quad \begin{cases} 1/(age(m) - age(f) - 5 + 1) & \text{if } \ age(m) - age(f) \geq 5 \\ -1/(age(m) - age(f) + 2 - 1) & \text{if } \ age(m) - age(f) \leq -2 \\ 1 \quad \text{otherwise} \end{cases} \\
& && (68)
\end{aligned}
$$

- select a random female according to the weight function

$$
\begin{aligned}
& f^{t+\Delta t}_{partner(m)} = weightedSample(F^t_{isMarEli}, W_m) \\
& \qquad \text{where } \ W_m = \{w_i : w_i = weight(m, f_i) \ , \ f_i \in F_{isMarEli}\} \qquad (69)
\end{aligned}
$$

- $F_{candidates} = F_{candidates} - \{f_{partner(m)}\}$

Note that the just married male and his selected female don't belong to the set of marriage eligible population $P^{t+\Delta t}_{isMarEli}$ any more. The following assumption specifies the house of the new couple

P. 13 When two individuals get married, the wife and the occupants of actual house (i.e. children and non-adult orphan siblings) moves to the husband's house unless there are fewer occupants in his house. In the later case, the husband and the occupants of his house move to the wife's house.

Formally, suppose that $m \in P^{t+\Delta t}_{just(isMarried)}$ and $f^{t+\Delta t} = partner(m^{t+\Delta t})$, if

$$
\begin{aligned}
& |P_{house(m^t)}| \geq |P_{house(f^t)}| \text{ and } house(p^t) = house(f^t) \implies \\
& \qquad house(p^{t+\Delta t}) = house(m)
\end{aligned}
$$

Otherwise

$$
\begin{aligned}
& |P_{house(m^t)}| < |P_{house(f^t)}| \text{ and } house(p^t) = house(m^t) \implies \\
& \qquad house(p^{t+\Delta t}) = house(f) \qquad (70)
\end{aligned}
$$

acknowledgment

A Parameters and input data

A.1 Parameters

Model parameters

The following is a table of parameters employed for events specification. The values are set in an ad-hoc manner and they are not calibrated to actual data. Actual data is rather dependent on simulation parameters, e.g. the start and final simulation times.

α_x	Value	Usage
$basicDivorceRate$	0.06	Equation 61
$basicDeathRate$	0.0001	Equation 58
$basicMaleMarriageRate$	0.7	Equation 64
$femaleAgeDieRate$	0.00019	Equation 58
$femaleAgeScaling$	15.5	Equation 58
$initialPop$	10000	Section 6.2
$maleAgeDieRate$	0.00021	Equation 58
$maleAgeScaling$	14.0	Equation 58
$maxNumMarrCand$	100	Sections 6.4 & 7.5
$startMarriedRate$	0.8	Equation 39

The value of the initial population size is just an experimental value and can be selected from the set $\left\{10^4, 10^5, 10^6, 10^7\right\}$ to examine the performance of the implementation or to enable a realistic demographic simulation with actual population size.

Simulation parameters

In version 1.1 of the package MiniDemographicABM.jl [2], the following ad-hoc values of the simulation parameters are selected:

α_x	Value
t_0	2020
Δ_t	Daily
t_{final}	2030

It is beneficial in future to further propose several case studies with specific simulation parameter values for each case. This shall be hopefully accompanied in the documentation of the model provided as a pdf-file within the package.

A.2 Data

The data values are set as follows:

$$D_{divorceModifierByDecade} \in R^{16} \;= \\ (0, 1.0, 0.9, 0.5, 0.4, 0.2, 0.1, 0.03, 0.01, 0.001, 0.001, 0.001, 0, 0, 0, 0)^T$$

$$D_{maleMarriageModifierByDecade} \in R^{16} \;= \\ (0, 0.16, 0.5, 1.0, 0.8, 0.7, 0.66, 0.5, 0.4, 0.2, 0.1, 0.05, 0.01, 0, 0, 0)^T$$

In the archived Julia package MiniDemographicABM.jl Version 1.1 and originally taken from the lone parent model implemented in Python the fertility data

$$D_{fertility} \in R^{35\times 360} \;= \\ [d_{ij} : \text{ fertility rate of women of age } i - 16 \text{ in year} j - 1950]$$

This matrix reveals and forecast the fertility rate for woman of ages 17 till 51 between the years 1951 and 2050, cf. Figure 2.

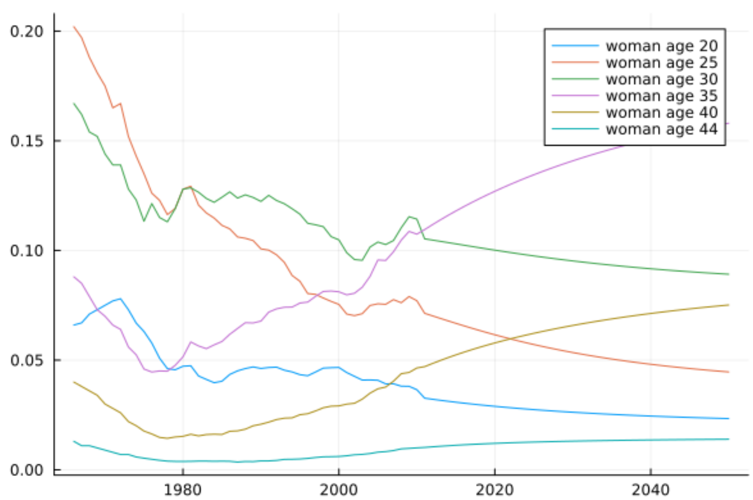

Figure 2: fertility rates of women of different age classes between the years 1966 and 2050

B Events rates and instantaneous probability

Pre-given data, e.g. mortality and fertility rates, are usually given in the form of finite rates (i.e. cumulative rate) normalized by sub-population length. In other words, the rate

$$R_{event,period(t,t+\Delta t)}(X^t)\ ,\ X^t \subseteq P^t$$

corresponds to the number of occurrences that a certain *event* within a sub-population X (e.g. marriage) takes place in the time range between $(t, t + \Delta t)$, e.g. a daily, weekly, monthly or yearly rate, etc. normalized by the sub-population length. That is, say if a pre-given typically yearly rate is given as input data:

$$R_{event,yearly}(X^t) \ = \ D_{event,yearly} \in R^{N\times M} \tag{71}$$

where M corresponds to a given number of years and N corresponds to the number of particular features of interest, for examples:

- $M = 100$ for mortality or fertility yearly rate data between the years 1921 and 2020
- $N = 28$ for fertility rate data for women of ages between 18 and 45 years old, i.e. $28 = 45 - 18 + 1$

The yearly probability that an event takes place for a particular individual $x^t \in X^t$ is:

$$Pr_{event,yearly}(x^t \in X^t) \ = \ D_{event,yearly}(yearsold(x^t), currentYear(t)) \tag{72}$$

Pre-given data in such typically yearly format desires adjustments in order to employ them within a single clocked agent-based-model simulation of a fixed step size typically smaller than a year. Namely, the occurrences of such events need to be estimated at equally-distant time points with the pre-given constant small simulation step size Δt. For example, if we have population of 1000 individuals with a (stochastic) monthly mortality rate of 0.05, then after

- one month (about) 50 individuals die with 950 left (in average)
- two months, about 902.5 individuals are left
- ...
- one year, 540 individuals are left resulting in a yearly finite rate of 0.46

Typically mortality rate in yearly forms of various age classes are given, but a daily or monthly estimate of the rates shall be applied within an agent-based simulation.

The desired simulation-adjusted probability is approximated by rather evaluating the desired rate per very short period regardless of the simulation step-size, assumed to be reasonably small (e.g. hourly, daily, weekly or monthly at maximum). Formally, the so called instantaneous probability is evaluated as follows:

$$Pr_{event,instantaneous}(x^t \in X^t, \Delta t) \;=\; -\frac{ln(1 - Pr_{event,yearly}(x^t))}{N_{\Delta t}} \tag{73}$$

where N_Δ is given as in Equation 37.

References

[1] George Datseris, Ali R. Vahdati, and Timothy C. DuBois. Agents.jl: A performant and feature-full agent-based modeling software of minimal code complexity. *SIMULATION*, 2022.

[2] Umberto Gostoli and Eric Silverman. Social and child care provision in kinship networks: An agent-based model. *PLOS ONE*, 15(12), December 2020.