Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

GEPOC ABM: A GENERIC AGENT-BASED POPULATION MODEL FOR AUSTRIA

Martin Bicher

Institute for Analysis and Scientific Computing
TU Wien
Wiedner Hauptstraße 8-10
1040 Vienna, AUSTRIA

Christoph Urach
Niki Popper

dwh Simulation Services
dwh GmbH
Neustiftgasse 57-59
1070 Vienna, AUSTRIA

ABSTRACT

Since 2015 researchers in Austrian health-care research project DEXHELPP (Decision Support for Health Policy and Planning) benefit from having access to a validated generic agent-based population model (GEPOC ABM) of Austria's population. This simulation model delivers a valid virtual image of Austria's population and is also able to make feasible prognoses. During the last years the model has been extended, remodeled and applied to several use-cases. We were able to add aspects like vaccination strategies, treatment pathways or spread of infectious diseases which underlines the flexibility of the implementation. Yet, a number of challenges have been identified, being the basis to contribute to the general discussion of population models. We will discuss evolving challenges according performance issues and present a newly implemented time-update approach. Thereafter we will discuss different parametrization concepts when adding a disease model. Finally we will present how we integrated GIS information based on Delauney Triangulation.

1 INTRODUCTION

With about 8.7 million inhabitants, 190 thousand emigrations and deaths and 260 thousand immigrants and births, Austria's total population fluctuated by about 2.2 percent in the course of 2016 (Statistik Austria 2016). This percentage is neither statistically high or low in comparison with other years or other countries, but it gives an idea about the total volume of population fluctuation and its potential impact on deducible numbers. It makes clear that any decision-support for policy making and planning can only be valid if it considers a model accounting for the underlying population dynamics.

Austrian research project DEXHELPP (Decision Support for Health Policy and Planning) poses a platform for collaboration of health-care stake holders, medical experts, modeling and simulation experts, statisticians, data scientists and visualization experts. By combining their skills they perform innovative, joint and data based research on all levels of the health system. With a wide range of integrated technologies they provide interactive tools for prognosis and decision support for policy making. In order to create a valid common founding for their decision-support tools research on population modeling and simulation is one of the most important research areas of this project:

GEPOC, short for Generic Population Concept, is a vital research part of DEXHELPP since 2014. It is founded on the idea that a related number of valid population models can be used as a basis for many different applied decision support models. In the first stage of the project, two structurally different population-models have been developed and validated: GEPOC SD and GEPOC ABM. The first one was developed using the method of system-dynamics (SD) and is (mathematically spoken) an ordinary differential equation model with several hundred coupled equations. The second model is a stochastic agent-based model (ABM). Both models have been validated using data from the Austrian Bureau of

Statistics (for details, see (Bicher et al. 2015)). Finally, in fall 2016, also a third population model was added to the collection in form of a partial differential equation (PDE) model (Bicher and Popper 2016)..

1.1 Introduction to GEPOC ABM

All mentioned population models have been sufficiently validated and are tested to produce equivalent results. In the next chapters, we will focus on the agent-based approach GEPOC ABM, as this model became the center of population based health-care research in DEXHELPP and has grown to a powerful and versatile simulation tool for any kind of population-based research problem in Austria. Hereby the coincidence of two important factors was responsible for this success:

- Intensive collaboration with health-care stakeholders provided the possibility for application of GEPOC ABM as a base model for many diverse health-care related research problems.
- Continued research on population modeling and continuous improvement of GEPOC ABM in collaboration with modeling and simulation experts from different institutions.

In this work we want to present the overall view on this versatile population model in detail for the first time. Besides giving a formal model definition we will emphasis on valuable lessons-learned from iteratively applying and improving the model. We will present interesting technical as well as model-theoretic challenges related to the model and its implementation and state our approaches to overcome them.

2 BASIC MODEL DEFINITION AND IMPLEMENTATION

As mentioned GEPOC ABM is an agent-based simulation model and has been validated to firstly, depict the status quo of Austria's population between 1991 and 2017 and secondly to make feasible prognoses matching the forecasts of the Austrian Bureau of Statistics (on the aggregate level). GEPOC ABM is defined via its initialization and its time-dynamics:

Initial Setup: Given a certain start date of the simulation an agent-based model with $N+1$ agents is initialized. The first N of them stand representative for the inhabitants of Austria and will be denoted as *person-agents* henceforth. Each person-agent is given a certain birth-date and (biological) sex. We will refer to them as female and male agents with a certain age. The remaining $N+1$-st agent will play the role of the government and will be denoted as *government-agent*.

Time Dynamics: The model is updated in not-necessarily equidistant time-steps which are defined a-priori. Each time-step consists of two parts:

In the first part all person-agents are iterated in random order. For each addressed agent, the model decides about death, emigration and birth of agents using an event-based strategy. First of all, random numbers decide about whether the addressed agent is scheduled to emigrate, die and/or (for female agents) had an offspring in the regarded time-step. For each action scheduled this way a uniformly distributed random number samples a date for the scheduled action and adds it to an event-list. After all possible events have been regarded the event-list is sorted and processed in correct order. Death and emigration events lead to a removal of the agent (skipping all further planned events) while the birth event leads to a newborn agent with correspondent birth-date added to the model. This strategy is sketched in Figure 1. After all person-agents have been iterated, the government agent generates a certain number of new person-agents (representing immigrants) and adds them to the model. This concludes one model time-step.

This model definition has changed from the original definition of GEPOC ABM ((Bicher et al. 2015)) at two points. Firstly, the original model was updated in equidistant time-steps. This small enhancement became relevant to satisfy the need to execute the model in monthly steps (which may take between 28 and 31 days). Secondly, the mechanism for agent-updates switched from a classic probability-based (markovian) to an event-based approach. We will discuss the benefits of this strategy in Section 2.2 and take a look at the implementation first.

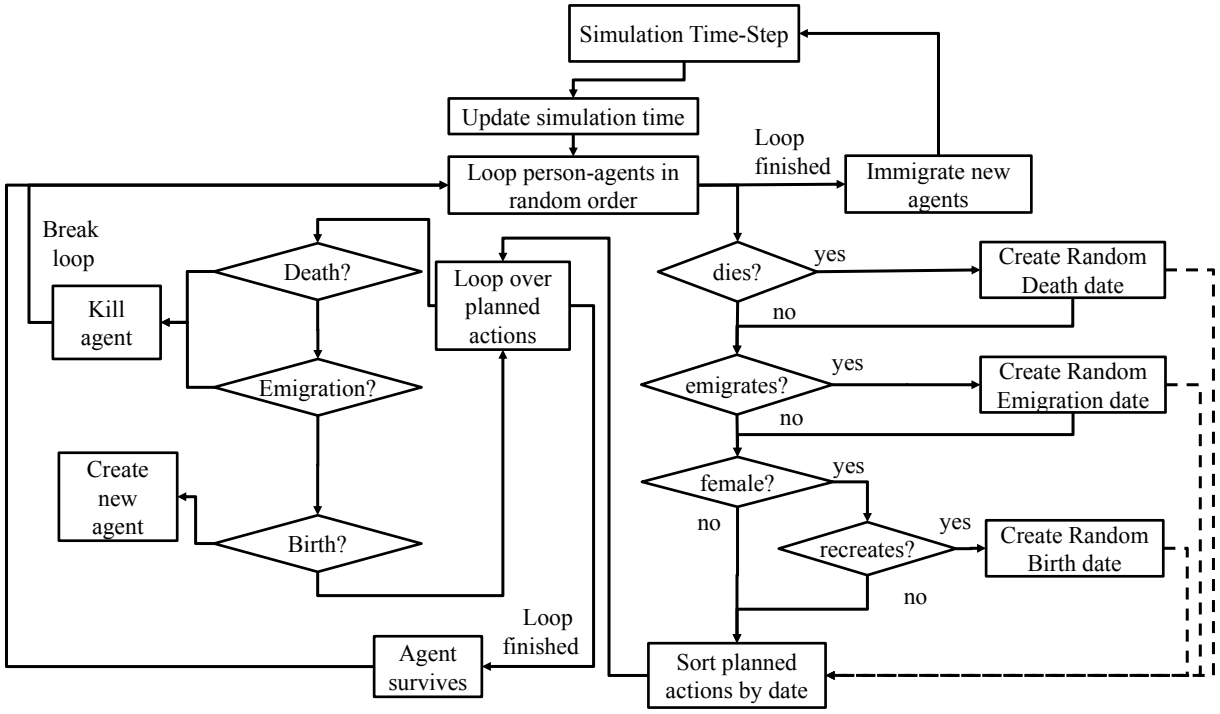

Figure 1: Discrete-event motivated strategy encapsulated in a basically time-discrete update of the person-agents in GEPOC ABM.

For our application we found it more useful to implement the model from the scratch than using existing ABM frameworks like Netlogo (Tisue and Wilensky 2004), Anylogic (Grigoryev 2012), Mesa (Masad and Kazil 2015), JADE (Bellifemine et al. 1999) or Mason (Luke et al. 2004). Neither of the mentioned was capable of 1) dealing with the high total number of required agents, 2) load and process all necessary parametrisation data (with reasonable preprocessing time) and 3) provide sufficient flexibility for all potential model extensions. Moreover, as we are dealing with very sensitive health-care data and research questions we wanted to stay in full control of all parts of the simulation and did not want to rely on often loosely documented 3rd party frameworks that work nicely for scientific applications, but reveal shortcomings and bugs when it comes to real-world applications.

We decided to implement the model using the (primarily) object-oriented programming language Python3. Firstly, most Python interpreters can be used free of charge and work platform independent which makes the model easily transferable. Secondly, Python programming requires the use of proper indentation making the code easily readable. Thirdly, millions of freely available Python packages provide high performance algorithms and interfaces to almost any known data format.

2.1 Code Performance

Although sub-packages like Numpy and SciPy provide highly efficient and vectorized algorithms to speed up computation times, Python (alike other dynamically typed, interpreted languages) is known to execute comparably slow. Therefore, execution of the simulation model with the full population of Austria (i.e. run the model with 8-9 million agents) is very time and memory consuming. To give a quick example, the execution of a 365day-time-step with 79000 agents takes a Intel® Core™i5-5200U processor about 2.02 sec without making use of multithreading. This number scales linearly with the number of agents and time-steps.

The easiest and most obvious solution to this problem is running the model with a reduced number of agents (i.e. one tenth or one hundredth of Austria's original population) instead. Afterwards the simulation results can easily be rescaled to the original size. This strategy was quickly approved to be valid from the modeling perspective: It is a direct consequence of the Law of Large Numbers that the aggregated simulation results with full population match the rescaled aggregated simulation results with reduced population. The only difference is the size of stochastic fluctuations which is proven to be larger when running the model with reduced number of agents (Note, that this result is not only valid for models without interaction as in this case, but also for a broad range of models with interaction. For more information see (Bicher 2017; Bicher and Popper 2015)). To compensate for the higher fluctuations with a downscaled population the simulation can be evaluated more often in Monte Carlo experiments, which increases computation time with a smaller extent.

Surprisingly, the described strategy encountered harsh opposition at decision-makers and its credibility was decreased. Discussing the model's internal logic its easier to communicate, that an agent poses for a statistical-representative of one real person instead of 10 or 100. Hence, we had to get it executable with the full population in reasonable time.

Besides standard means for code optimization two interesting technical measures have been implemented that finally improved performance of the code.

- The generation of new person-agents has a massive impact on the computation time due to sampling of multivariate random numbers with user-defined distribution functions. As this is needed extensively often when generating the initial model population a Markov-Chain Monte-Carlo (MCMC) sampling algorithm was applied for this purpose. We made use of the performant implementation of this algorithm in the *PyMC* package of Python3 (Patil et al. 2010).
- As many applications of GEPOC ABM did not make use of agent-agent contacts or did only require very local contacts (see Section 3) we used Python's native *subprocess* package to make the simulation model capable for multi-threading. Hereby, the initial population is split into a predefined number of parts which can be distributed among an arbitrary number of computation kernels. Hence, as long as it is sufficient that person agents have a very limited range of contact partners, GEPOC ABM can be executed fully parallelized.

Our current work in this area is focuses on improving the parallelization capabilities of GEPOC ABM to allow limited contacts between person agents in different threads comparable to (Collier et al. 2015). Summarizing, we learned the lesson, that performance is still an issue in population models. Strategies to cope with this, have to include not only methods to increase performance but also stakeholder interests.

2.2 Time-Update Strategy

To be fully versatile as a generic framework GEPOC ABM has to be capable of dealing with processes on different time scales. While e.g. infectious diseases like influenza spread in a few days or weeks is usually requires many years and decades to observe the impact of demographic changes on the health-care landscape.

The currently most prominent concept to overcome this problem is simulating the model in continuous time – i.e. using a discrete-event strategy (Buss and Al Rowaei 2010). Hereby, agents are emigrated and immigrated, die and are born at corresponding event dates which additionally schedule new future events. After each occurred event the simulation instantaneously skips to the next scheduled event and the model-time is enhanced. For the multi time-scale problem in GEPOC ABM this strategy would clearly be beneficial to a classic time-discrete update as the mechanism is independent of the observed time-scale and scope. Yet, we found two arguments why this type of update is not optimal for our applications (or at least requires further research).

- Finding the next event to occur is always related to a sorting problem. With N denoting the initial number of agents in the model the computational efforts of the ABM consists of iteratively executing the occurring events (resulting in a problem of $\mathcal{O}(N)$) and correctly inserting the newly scheduled events to the event list (e.g. using a standard divide-and-conquer algorithm with $\mathcal{O}(\log(N))$). Therefore, the total computational efforts of the model calculate to $\mathcal{O}(N\log(N))$ which is delicately larger than using a time-discrete strategy with $\mathcal{O}(N)$ effort. Though, there has been progress in reducing the computational efforts of continuous-time population models by using internal model logic (Reinhardt and Uhrmacher 2017; Warnke et al. 2016) they can never depend linearly on the number of agents. Hence, this kind of update strategy is significantly slower (at least as long as the model does not use agent-agent contacts).
- Discrete event update is known to cause difficulties if there exists a global interaction level. We explain this problem on a short example: Suppose, GEPOC ABM is used to investigate the effects of overpopulation. Therefore, the population density of the country is assumed to have a negative impact on the death rate. As the population density changes with every occurring event, it is impossible for a person-agent to correctly define its own death date in advance. The only solution to this problem would be, to re-sample all death dates of all agents whenever the population density changes. This leads to a massive overhead.

The second option to update ABMs is applying discrete time-steps: Instead of deciding *when* a specific event happens the model iterates through time asking *if* a specific event occurred in a regarded time-interval. Hereby so called transition probabilities are used. For the multi-scale problem in GEPOC ABM the simulation needs to be executable (and valid) with time-steps of arbitrary lengths. Hereby, two problems occur:

- Firstly, it is mathematically impossible to correctly transform transition probabilities from one to a different time-step length without changing the (expected) simulation outcome. This is exhaustively discussed in (Bicher 2017) and is best imagined by a simple gedankenexperiment: Say, a female agent has a probability p_t to give birth to a child during a time-interval with length t. Now, assume that the time-step length should be halved to $t/2$. Hence, we are looking for a rescaled probability $p_{t/2}$ so that two steps of the rescaled model lead to the same results as one step of the original one. Easily seen, this task is impossible to solve as (independent of the choice of $p_{t/2}$) the rescaled model makes it possible that two children are born after the regarded time-interval.
- Secondly, the occurrence of two or more events in one model time-step leads to causality problems. Especially in the case of population models it makes a crucial difference if an agent dies before it recreates, emigrates before it dies, recreates before it emigrates or vice versa. Hence, using discrete time-steps always requires additional model logic.

Consequently neither of the two time-update strategies is optimally suited for a generic population model. The proposed solution presented in the model definition and in Figure 1 can be interpreted as an event-based strategy embedded in a time-discrete update. On the global level, there is a time-step that manages the update of the *time* variable. For most transition probabilities we applied the approximation formula

$$p_{\Delta t'} = 1 - (1 - p_{\Delta t})^{\frac{\Delta t'}{\Delta t}} \tag{1}$$

to scale transition probabilities from one to a different time-step length ($\Delta t \to \Delta t'$). This formula is motivated from geometric distribution.

On the agent-level, the boolean-statement *that* something happens is linked to an event with occurrence time *when* it happens. Hereby, ordering of events is clear from the start and illogical event sequences are excluded. It is possible to e.g. hospitalize, treat and release an agent in just one model time-step automatically generating plausible hospitalization and release dates. Hence, as an additional benefit, it is not always necessary to use atomically small time-steps to investigate small time-scopes. Summarizing,

we learned the lesson, that there is no optimal time-update strategy for a *generic* population model. Event oriented concepts appear promising, but require further research.

3 APPLICATIONS AND MODEL EXTENSIONS

GEPOC ABM has already proven its flexibility as a basis model for population based research in various areas. Since its validation in 2015 GEPOC ABM has been used for several health-care related applications of which we specifically want to explain the three largest in detail.

Vaccination Rates: Eradication of measles and polio is one of many goals the World Health Association (WHO) is trying to achieve until year 2020. Hereby, besides other factors especially high vaccination numbers among the population play a key role. In case a high percentage (about 95% are estimated) of all inhabitants are vaccinated so-called herd-immunity effects will prevent potential epidemics from breaking out which, in the long run, leads to the full eradication of the disease. To stay in control about the progress every country is obliged to yearly report the percentage of vaccinated infants among their age-cohort – we will furthermore refer to this number as "vaccination rate" – to the WHO.

Though numbers of sold vaccination doses as well as age of their recipients are (quite) well known in Austria calculation of these rates for reporting reasons is not as simple as it seems. Due to fluctuations among the population primarily caused by high immigrant/refugee numbers a dynamic simulation model was used to correctly determine the vaccination rates and improve the formally used calculation method.

We extended GEPOC ABM to get an image about the current MMR (measles, mumps, rubella) and polio vaccination rates in Austria. According to availability of doses (gained from data about real sold doses) and the vaccination regimen each person agent is assigned vaccinations. With specifically calculated vaccination rates for regular immigrants and refugees the model fully considered the effects of a fluctuating population. The simulated numbers were reported by the Austrian Ministry of Health and Women's Affairs and can be accessed via the web-page of the WHO or in two short reports about the current situation in Austria (Bundesministerium für Gesundheit und Frauen 2017; Bundesministerium für Gesundheit und Frauen 2016). Besides giving access to a more precise calculation method GEPOC ABM additionally provides deeper insights into the dangers of measles outbreak. E.g. using accredited estimates for the chance that a vaccination successfully immunizes the recipient and people who were immunized by past illnesses we are additionally able to give information about the percentage and distribution of immune persons.

Re-hospitalization of Psychiatric Patients: Re-hospitalization rates of psychiatric patients are considered as a metric of quality of care. Yet, risk factors which enforce high percentages of re-hospitalized patients are still not fully understood and are a heavily researched area. In order to test the plausibility of several risk factors commonly believed by domain experts, and to compare different types of health service interventions in terms of differences in re-hospitalization outcomes, a simulation model was implemented.

GEPOC ABM was extended by several functionalities. First, person-agents were given a probability to visit mental hospitals and have a stay of several days during which they are diagnosed. Afterwards, every person-agent has a certain chance to become re-hospitalized again dependent on diagnosis, sex, age and other risk factors with were key objects of the investigation. Assuming that the chance depends on the mean-distance to the nearest hospital, person-agents were assigned a residence (NUTS3 region). Hereby, impact of infrastructural changes could were tested. Moreover, assuming that the chance depends on co-morbidities, diabetes mellitus was implemented as background disease. This way also the influence of our aging society was analyzed. More information about this model is found in (Zauner et al. 2017; Bicher et al. 2017).

Number, Severity and Diagnosis of Stroke Incidences: Implementation of stroke units in hospitals is a heavily discussed topic (Wilbacher 2005). On the one hand, these units are known to significantly decrease the risk of mortality and consequential damage in case of a stroke incident compared to regular hospital units (Barnett 2000). On the other hand, operation of these specialized units is expensive, especially

when not in use. Therefore, DEXHELPP started with rigorous analysis on the need for stroke treatment using a dynamic simulation model.

Person-agents in GEPOC ABM were extended by a chance to suffer from a stroke with a certain severity and a specific type (diagnosis). This chance is implemented to depend from the person-agent's age, sex and residence district as well as having had a previous stroke incident. Hereby, we were able to observe stroke-related parameters which (in Austria) cannot be accessed from data like the average number of stroke incidences per person or the total number of stroke-caused deaths. The model is not yet fully validated, but will contribute to improve services provided for stroke treatment by giving a very detailed picture of the need.

Motivated by these three applications a couple of **toolboxes** have been developed that can optionally be used to extend GEPOC ABM if needed. Hereby, certain parts that have been required for the case-studies and were deemed to have potential use in future applications were made reusable in a more generic form. We will present the two most interesting here.

3.1 Parametrization of Diseases via Incidence and Prevalence

Taking a closer look at the three applications presented above the experienced modeler will quickly observe that none of them relies on any contacts between person-agents (Note, that the first mentioned application modeled measles vaccinations and not measles infections). GEPOC ABM offers the possibility to implement contacts e.g. between persons/patients/hospitals/physicians, but the given research problems defined by our collaborating decision makers (e.g. Austrian Ministry of Health, Main Association of Social Insurances, Gesundheit Österreich GmbH) hardly required this functionality yet. Although we made use of contacts in smaller and more academic studies (patients $\leftrightarrow$ doctors in (Nowotny, K. 2018)), the three important applications presented earlier taught us that simulation-based research in Health Technology Assessment, Health System Research and Health Services Research does not necessarily rely on contacts or contact-networks. On the one hand, this can be considered as good news as GEPOC ABM can make full use of parallelization. On the other hand, the dynamics of the resulting models are scientifically less interesting. Causes for the lack of need in contact-based models in health-care applications can only be speculated. One possible reason might be that the impact of non-transmittable diseases (e.g. cardiovascular diseases, neurological diseases, chronic progressive diseases) on the health-care system is massive – even compared to infectious diseases.

For this reason we decided to implement a toolbox that makes it possible to quickly extend GEPOC ABM with a non-transmittable disease. We united the mechanism used for diabetes mellitus in the re-hospitalization module and the mechanism for stroke incidences in the last application to form one generically applicable model add-on. As diabetes is parametrized using prevalence data and stroke is parametrized using incidence data the generic module is capable for using both data of these epidemiological key figures. Hereby it is important to mention that the strategy only considers new cases and does not regard the recovery from the medical condition.

Incidence or to be precise the incidence rate is defined as a measure for the probability of at least one occurrence of a certain medical condition in the observed time-interval. An incidence rate of I per year implies that a person who does not show the regarded medical condition before has a probability of I to show the medical condition after one year. Often incidence rates are given as average number of persons showing the condition per 1000 or 10000 as it is easier to interpret.

Incidence rates can be used to extend GEPOC ABM in a very natural way. Every healthy person-agent schedules the "medical condition"-event in the course of the regarded time-step with a probability directly calculated from the incidence rate. In case GEPOC ABM is run with yearly steps, the incidence rate can be taken directly, otherwise it is rescaled using formula (1). Although incidences are sufficient to simulate new cases it is necessary to know about the prevalence at least for the initial setup of the person-agents. Hence, incidence rates alone are usually not sufficient to parametrize the model.

Prevalence is a measure for the total number of persons suffering from a specific medical condition and is usually given as a fraction of the total population. As for the incidence rate we often find this number described as number of cases per 1000, 10000 or 100000 persons to make it easier to depict.

In the contrast to incidence rates, the extension of GEPOC ABM using prevalences is not that natural. We found it most convenient to follow a two phase strategy. First, the model time-step is executed as defined in Section 2 (including immigration). Hereby, the total population P and the fraction F' of person-agents suffering from the medical condition are counted directly after execution of all agent-events. Thereafter, the known prevalence F of the medical condition is compared with F'. If data and model are valid, $F' < F$ should result as the number of cases is only reduced in the first phase (deaths, emigrations, recoveries). Hence, $(F - F')P$ describes the total number of person-agents that should suffer from the medical consideration according to the data, but do not show this behavior in the model so far. Therefore in phase two, $(F - F')P$ healthy person-agents are randomly picked from the agent population to start suffering from the medical condition. Easily seen, this strategy becomes more accurate the smaller the used time-step and the more prevalence data points are given. If the step-width of the model time-steps is chosen smaller than the time-resolution of the data it is useful to linearly interpolate the data points to avoid unsteady jumps of the prevalence in the model.

Clearly, in case of direct conflict the incidence strategy would be preferred as it is the more natural way parametrising a disease in an ABM. Yet, incidence data for diseases is usually harder to get. The strategy for parametrization of prevalence might seem unusual for an agent-based model, but gives perfect control about the total number of cases and has proven to be perfectly suited for simulation of chronic diseases like diabetes mellitus. Summarizing, we learned the lesson, that a lot of problems don't require agent-agent contacts. It is important to have the possibility, but its same important to get rid of it, if not needed or applicable.

3.2 Giving Agents a Place to Live

As seen in the Stroke and the Re-Hospitalization application of GEPOC ABM it is often necessary to extend the person-agents properties by a residence. One could mention this feature to be a necessary feature of population models in general, but turns out to be a massive overhead if not needed. We decided to generalize the findings of the two case studies that required agent residences in a generic *Geography* toolbox that samples residences to person-agents.

In the course of this development soon a couple of problems occurred. Firstly, the administrative landscape is permanently changing: Each year a couple of districts and municipalities are dissolved, joined or reassembled. A very prominent example for this is the former district "Wien Umgebung" which was split up into four neighbored districts in 2016. Secondly, different partitions of Austria are not always compatible. It happens quite often that smaller units are not uniquely contained in larger units. For example one quickly finds ZIP regions that belong to two or more different political districts. The administrative regions for health-care service ("Versorgungsregionen") even overlap with the Austrian federal states.

In order to develop a generic solution that works independently of the investigated partition of Austria we decided to sample residences in form of GIS coordinates. This method is beneficial compared to sampled regions as a coordinate is always linked to one unique region per investigated partition. This region may change with time if units are joined or separated, but can always be found as long as the GPS outline of the partition is known.

We implemented the following algorithm to sample a random GPS coordinate with respect to a given partition of Austria (equivalent to the one presented in Section 3.3.1 in (Gallagher et al. 2018)):

1. Sample a random region the person-agent is planned to live in according to a given distribution.
2. Sample a uniformly distributed point inside the region according to its GSP outline.

Hereby we worked hard to improve the performance of the latter part. Standard algorithms to sample a uniformly distributed coordinate in a given region are based on a rejection algorithm. I.e. a uniformly distributed point inside the bounding-box of the polygon (or to be precise multi-polygon) is sampled and accepted if it lies inside the regarded region. The strategy requires to check if the sampled point lies inside the polygon at least once which requires that many scalar multiplications as corner-points on the outline. It is particularly inefficient if shapes are not 0-connected (as the district of "Amstetten" seen in Figure 2), not 1-connected (as the district of "St Pölten Land" seen in Figure 2), or elongated and diagonally oriented.

Hence, we decided to use a different strategy based on the idea that there exists an explicit formula to calculate a uniformly distributed point inside a triangle. Given two independent uniformly distributed random numbers r_1 and r_2 between 0 and 1 and three points $A,B,C \in \mathbb{R}^2$ forming a triangle then

$$x := A(1-\sqrt{r_1}) + B(1-r_2)\sqrt{r_1} + Cr_2\sqrt{r_1} \tag{2}$$

is a uniformly distributed point inside $\triangle ABC$ (Osada et al. 2002). As we could not find a full proof for this statement in literature we added it to the Appendix section.

Using this formula our strategy states as follows.

2.a Perform a Constrained Delauney Triangulation (CDT) of the shape and calculate the areas of all resulting triangles. Note, that this has to be done only once for each region and can be reused.
2.b Pick one random triangle from the list of triangles weighted by their area.
2.c Pick a uniformly distributed point inside the triangle according to formula (2)

The concept of the CDT is visualized on the two aforementioned districts in Figure 2. Experiments showed that this version of the method is about ten times more efficient than the rejection algorithm. Figure 3 shows 100000 sampled residences according to a given distribution on municipality level (Austria is partitioned in about 2700 of them). Highly populated areas, especially the large cities Vienna, Graz, Linz, Salzburg and Innsbruck are well visible. Also the influence of the Alps which range from the south-west almost until Vienna in the north-east is very picturesque.

Although the sampling algorithm works nicely the Geography module of GEPOC ABM can not yet be considered a validated generic model extension so far especially due to a lack of parametrization data. First of all, joining and splitting of regions cause problems with standardized data storage and acquisition for parametrization of the module. Secondly, data availability for parametrization of internal migration of person-agents is unfortunately insufficient. We currently plan to include settlement information from the Global Human Settlement Project (Florczyk et al. 2016) to make population distribution even more realistic. Summarizing, we learned the lesson, that sampling of solely residential regions (Federal States, NUTS3 Regions, Political Districts,...) is not sustainable. We require sampled coordinates.

4 CONCLUSION

As seen in the three case studies GEPOC ABM has already proven its worth as a generic population base module for different health-care related research problems. Due to our close collaboration with decision makers we are able to continuously improve and extend the model to make it easier applicable and more flexible. Hereby we were taught valuable lessons about population modeling and modularity of simulation models which we shared in this work.

Still, there are many open questions which require further research. The parametrization of spatial aspects and hereby especially the internal migration involves data difficulties which we plan to solve in the next years. Also the usage of a large computation cluster for reduction of calculation times is planned very soon. Finally, we aim to apply the model for research problems apart from health-care to get additional insights.

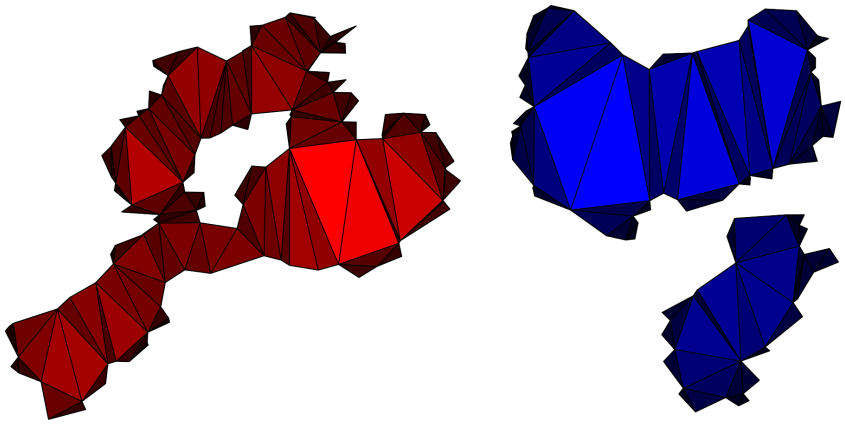

Figure 2: Constrained Delauney Triangulation of districts “St. Pölten Land” (left) and “Amstetten” (right) for GIS-coordinate sampling (status Jan 1st 2013). The colors of the triangles indicate their area.

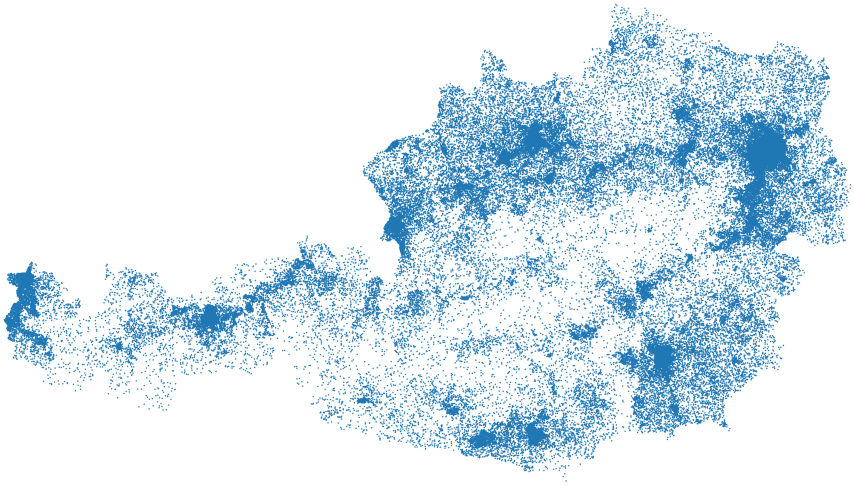

Figure 3: Sampled residences for 100000 agents according to distribution for municipalities (Jan 1st 2013).

REFERENCES

Barnett, H. J. M. 2000. “The Imperative to Develop Dedicated Stroke Centers”. *Journal of the American Medical Association* 283(23):3125.

Bellifemine, F., A. Poggi, and G. Rimassa. 1999. “JADE–A FIPA-compliant agent framework”. In *Proceedings of the Practical Applications of Intelligent Agents*, 97–108.

Bicher, M. 2017. *Classification of Microscopic Models with Respect to Aggregated System Behaviour*. Dissertation, Institute for Analysis and Scientific Computing, TU Wien, Vienna, Austria.

Bicher, M., B. Glock, F. Miksch, G. Schneckenreither, and N. Popper. 2015. “Definition, Validation and Comparison of Two Population Models for Austria”. In *Proceedings of 4th UBT Annual International Conference on Business,Technology and Innovation*, edited by E. Hajrizi, 174–179. Durres, Albania: UBT - Higher Education Institution.

Bicher, M., and N. Popper. 2015. “Spatial Effects in Stochastic Microscopic Models - Case Study and Analysis”. *IFAC-PapersOnLine* 48(1):153–158.

Bicher, M., and N. Popper. 2016. "Mean-Field Approximation of a Microscopic Population Model for Austria". In *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation*, 544–545. Oulu, Finland.

Bicher, M., C. Urach, G. Zauner, C. Rippinger, and N. Popper. 2017. "Calibration of a Stochastic Agent-Based Model for Re-Hospitalization Numbers of Psychatric Patients". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W.K.V. Chan et al., 12. Piscataway, New Jersey: IEEE.

Bundesministerium für Gesundheit und Frauen 2016. "Kurzbericht: Evaluierung der Masern - Durchimpfungsraten". Technical report, BMGF, Vienna, Austria.

Bundesministerium für Gesundheit und Frauen 2017. "Kurzbericht: Evaluierung der Polio-Durchimpfungsraten". Technical report, BMGF, Vienna, Austria.

Buss, A., and A. Al Rowaei. 2010. "A comparison of the accuracy of discrete event and discrete time". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson et al., 1468–1477: IEEE.

Collier, N., J. Ozik, and C. M. Macal. 2015. "Large-scale agent-based modeling with repast hpc: A case study in parallelizing an agent-based model". In *European Conference on Parallel Processing*, 454–465. Springer.

Florczyk, A. J., S. Ferri, V. Syrris, T. Kemper, M. Halkia, P. Soille, and M. Pesaresi. 2016. "A new European settlement map from optical remotely sensed data". *Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(5):1978–1992.

Gallagher, S., L. F. Richardson, S. L. Ventura, and W. F. Eddy. 2018. "SPEW: Synthetic Populations and Ecosystems of the World". *Journal of Computational and Graphical Statistics* 0(0):1–12.

Grigoryev, I. 2012. *AnyLogic 6 in three days: a quick course in simulation modeling*. Hampton, NJ: AnyLogic North America.

Luke, S., C. Cioffi-Revilla, L. Panait, and K. Sullivan. 2004. "Mason: A new multi-agent simulation toolkit". In *Proceedings of the 2004 swarmfest workshop*, Volume 8, 316–327. Michigan, USA.

Masad, D., and J. Kazil. 2015. "MESA: an agent-based modeling framework". In *14th PYTHON in Science Conference*, edited by K. Huff et al., 53–60.

Nowotny, K. 2018, June. "ECO - Landärzte gesucht: Immer mehr Orte ohne Ordination". TV documentary.

Osada, R., T. Funkhouser, B. Chazelle, and D. Dobkin. 2002. "Shape distributions". *ACM Transactions on Graphics (TOG)* 21(4):807–832.

Patil, A., D. Huard, and C. J. Fonnesbeck. 2010. "PyMC: Bayesian stochastic modelling in Python". *Journal of statistical software* 35(4):1.

Reinhardt, O., and A. M. Uhrmacher. 2017, April. "An Efficient Simulation Algorithm for Continuous-Time Agent-Based Linked Lives Models". In *ANSS 2017 Spring Simulation Multi-Conference*. Virginia Beach, Virginia.

Statistik Austria 2016. *Statistisches Jahrbuch Österreich 2016*. Verlag Österreich GmbH.

Tisue, S., and U. Wilensky. 2004. "NetLogo: A simple environment for modelling complexity". In *International Conference on Complex Systems*, Volume 21, 16–21. Boston, Massachusetts.

Warnke, T., O. Reinhardt, and A. M. Uhrmacher. 2016. "Population-based CTMCS and agent-based models". In *2016 Winter Simulation Conference (WSC)*, 1253–1264. Piscataway, New Jersey: IEEE.

Wilbacher, I. 2005. "Stroke Units - Österreich im Internationalen Vergleich". Technical report, HVB EBM.

Zauner, G., C. Urach, M. Bicher, N. Popper, and F. Endel. 2017. "Spatial psychiatric hospitalization modelling in an international setting - an agent based approach". In *Proceedings of the International Workshop on Innovative Simulation for Health Care 2017*. Barcelona, Spain. To Appear.

APPENDIX

Proof of statement (2).

Proof. Based on two independent uniform random numbers r_1, r_2 with common density

$$f_X : \mathbb{R}^2 \to \mathbb{R}^+ : \ (r_1, r_2)^T \mapsto 1$$

we define the transformation

$$\begin{aligned}\phi_{A,B,C} : \mathbb{R}^2 \to \mathbb{R}^2 : (r_1, r_2)^T \mapsto A(1-\sqrt{r_1}) + B(1-r_2)\sqrt{r_1} + Cr_2\sqrt{r_1} \\ = B + (A-B)(1-\sqrt{r_1}) + (C-B)r_2\sqrt{r_1}\end{aligned}$$

and aim to show that $\phi_{A,B,C}$ uniformly maps the unit square $[0,1]^2$ onto the triangle $\triangle ABC$. Firstly, we define $\phi_{A,B,C}$ as the conjunction of two separate mappings. With

$$\phi_0 : \mathbb{R}^2 \to \mathbb{R}^2 : (r_1, r_2)^T \mapsto \begin{pmatrix} (1-\sqrt{r_1}) \\ r_2\sqrt{r_1} \end{pmatrix}$$

we get

$$\phi_{A,B,C}(r_1, r_2) = B + ((A-B),(C-B))\phi_0(r_1, r_2).$$

Hereby an affine transformation is applied on the image of ϕ_0. As affine transformations (a) map triangles onto triangles and (b) conserve the uniformity of a distribution, it is sufficient to show that ϕ_0 maps r_1, r_2 onto the triangle $\triangle(1,0)(0,0)(0,1)$ and that this mapping conserves the uniformity.

The first statement is trivially fulfilled. To show the second, we apply the transformation formula for probability densities

$$f_{\phi_0}(y_1, y_2) = f_X({\phi_0}^{-1}(y_1, y_2)) \left| \det\left(J_{{\phi_0}^{-1}}(y_1, y_2) \right) \right|.$$

We calculate

$${\phi_0}^{-1}(y_1, y_2) = \begin{pmatrix} (1-y_1)^2 \\ \frac{y_2}{1-y_1} \end{pmatrix}, \text{ and } \quad J_{{\phi_0}^{-1}}(y_1, y_2) = \begin{pmatrix} -2(1-y_1) & \frac{y_2}{(1-y_1)^2} \\ 0 & \frac{1}{(1-y_1)} \end{pmatrix}.$$

Therefore,

$$f_{\phi_0}(y_1, y_2) \equiv 2 = \frac{1}{\text{Area}(\triangle(1,0)(0,0)(0,1))}$$

shows that the transformed density is (as well) constant. Therefore, the image of ϕ_0 and also the image of $\phi_{A,B,C}$ is uniformly distributed on the stated triangle proving (2). □

AUTHOR BIOGRAPHIES

MARTIN BICHER is research associate at the TU Wien and scientific employee at dwh Simulation Services GmbH. He finished his PhD in Technical Mathematics at TU Wien in Winter 2017. His doctoral thesis was about mean-field behaviour of microscopic models. Email address: martin.bicher@tuwien.ac.at.

CHRISTOPH URACH studied Technical Mathematics at TU Wien and specialised on Mathematical Modelling and Simulation in the field of HTA (Health Technology Assessment). He currently works at dwh simulation services in the department of health economics where he is developing applicable model structures for evaluation of health care interventions. He is also working on a PhD thesis supervised by Prof. Dr. Felix Breitenecker. Email address: christoph.urach@dwh.at.

NIKI POPPER is CEO of dwh - Simulation Services GmbH and research associate at TU Wien. He is responsible key-researcher of K-Project DEXHELPP and head of the corresponding association. His research focus lies on comparison of different modeling techniques. niki.popper@dexhelpp.at.