dwh
technical solutions
simulation services

Agent-Based COVID-19 Simulation Model

Model Specification

May 29, 2020

Martin Bicher, Claire Rippinger, Dominik Brunmeir,
Christoph Urach, Niki Popper

Abstract

To simulate solely the spread of SARS-Cov-2 a variety of methods exists, of which many are probably more suited for prognoses than agent-based models. Yet, in order to to evaluate the impact of policies like tracing, individual-based models are required.

We developed an agent based simulation model to reproduce the current outbreak of COVID-19 in Austria that allows for exploratory analysis of tracing in different characteristics. Aim of this work is the presentation of this model and consequent evaluation and comparison of different policies: Can we achieve containment solely by successful tracing or do we need additional policies? How large is the impact of tracing, keeping in mind that a possibly large number of disease progressions are asymptomatic, yet infectious?

dwh GmbH
Neustiftgasse 57-59
1070 Vienna, Austria

1 Model Definition

We will explain our agent-based COVID-19 model based on the ODD (Overview, Design Concepts, Details) protocol by Volker Grimm et.al. [17, 18].

1.1 Overview

The modelling of the spread of the disease is based on the interplay of four modules.

1. Population. Altogether the agent-based COVID-19 model is based on the Generic Population Concept (GEPOC, see [13]), a generic stochastic agent-based population model of Austria, that validly depicts the current demographic as well as regional structure of the population on a microscopic level. The flexibility of this population model makes it possible to modify and extend it by almost arbitrary modules for simulation of population-focused research problems.
2. Contacts. In order to develop a basis for infectious contacts, we modified and adapted a contact model previously used for simulation of influenza spread. This model uses a distinction of contacts in different locations (households, schools, workplaces, leisure time) and is based on the POLYMOD study [24], a large survey for tracking social contact behaviour relevant to the spread of infectious diseases.
3. Disease. We implemented a module for the course of the disease that depicts the current pathway of COVID-19 patients starting from infection to recovery or death and linked it with the prior two modules.
4. Policies. Finally, we added a module for implementation of interventions, ranging from contact-reduction policies, hygienic measures and, in particular, contact tracing. This module is implemented in form of a timeline of events.

1.1.1 Purpose

The agent-based COVID-19 model aims to give ideas about the potential impact of certain policies and their combination on the spread of the disease, thus helping decision makers to correctly choose between possible policies by comparing the model outcomes with other important factors such as socioeconomic ones. In order to fulfil this target, it is relevant that the agent-based COVID-19 model validly depicts the current and near future distribution and state of disease progression of infected people and their forecasts. Yet this is not the key feature of this model.

In the following overview of the model, we will not state any parameter values to focus on the model concept. A full collection of model parameters including values, sources and justifications is found in the Appendix (Section 1.3.3).

1.1.2 Entities and State Variables

Each person-agent is a model for one inhabitant of the observed country/region. We describe state variables of a person-agent sorted by the corresponding module.

Population. Each person-agent contains the population specific state variables *sex*, *date of birth* ($\cong$ age) and *location*. The latter defines the person-agent's residence in form of latitude and longitude and uniquely maps to the agent's municipality, district and federal state.

Contacts. The person-agent features a couple of contact network specific properties. These include a *household* and might include a *workplace* or a *school*. We summarise these as so-called *locations* which stand for network nodes via which the person-agent has contacts with other agents. Assignment of person-agents to *locations* is based on distance of the agent's residence to the position of the *location*. Each day, an agent has a certain number of contacts within each of the *locations*, which essentially leads to spread of the disease. To model contacts apart from these places, every person-agent has an additional amount of leisure time contacts, which are sampled randomly based on a spatially-dependent distribution. The contact network is schematically displayed in Figure 1.

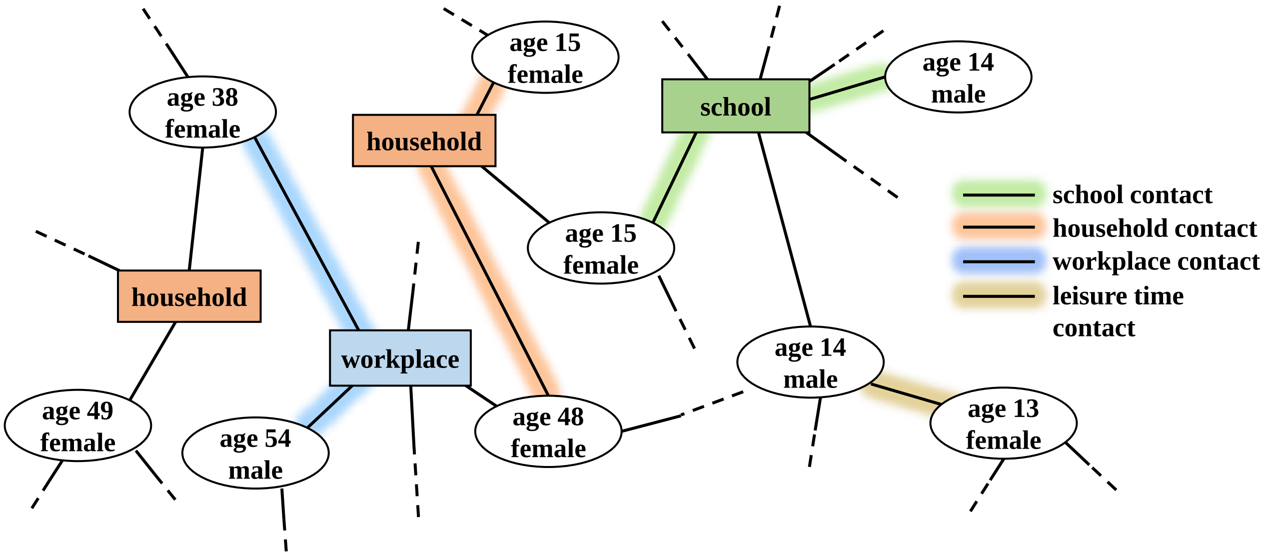

Figure 1: Contact network of agents in the agent-based COVID-19 model. Regular contacts between agents occur via locations (schools, workplaces and households, while random leisure time contacts extend the standard contact network.

Disease. In order to model the spread of the disease each person-agent has a couple of health states that display the current status of the agent. The most important ones are *infected*, *infectious*, *symptoms*, *hospitalised*, *critical*, *confirmed*, *severe*, *asymptomatic*, *home-quarantined* and *recovered*. These states can either be true or false, and multiple of them can be true at a time. They stand for certain points within the patient pathway of an infected person and enable or disable, respectively, certain person-agent actions. More on the influence of these state variables and how they change is described in Section 1.1.4.

Policies. Policies apply either to *locations* or to person-agent-behaviour directly and require additional agent properties. All *locations* except for households are defined *open* or *closed* which marks whether this place is available for having contacts. For person-agents the variable *preventive quarantine* is introduced to mark agents isolated due to tracing.

For the sake of simplicity of speech we furthermore address mentioned parameters as attributes for the corresponding agents. I.e. an agent with *infectious* set to true will be denoted as "*infectious* agent".

1.1.3 Scales

Unlike other agent-based models it is not possible to validly run the model with a smaller number of agents (e.g. one agent represents 10 or 100 persons in reality) as certain contact-network parameters do not scale this way (average school size,...). Consequently, one simulation run always uses agents according to the size and structure of the full population.

1.1.4 Process Overview and Scheduling

Like the underlying population model, the agent-based COVID-19 model can be interpreted as a hybrid between a time-discrete and a time-continuous (i.e. event-updated) agent-based model:

The overall simulation updates itself in daily time steps, wherein each step is split into three phases. In the first phase each agent is called once to plan what it aims to do in the course of this time step. In the second phase, each agent is, again, called once to execute all planned actions for this time step in the defined order. In the third step, a recorder-agent keeps track of all aggregated state variables.

On the microscopic scope, each person-agent is equipped with its own small discrete event simulator. In the mentioned planning phase, each agent schedules certain events for the future which may, but not necessarily must, be scheduled within the current global time step. In the second phase, the agent executes all events that are scheduled for the currently observed time interval, but leaves all events that exceed this scope untouched.

This strategy comes with the following benefits:

- In contrast to solely event-based ABMs, the event queue is distributed among all agents which massively increases the speed for sorting (a solely event-based ABM with millions of complex agents would not be executable is feasible time).
- Moreover, in contrast to solely event-based ABMs, usage of daily transition probabilities/rates instead of transition times is possible as well.
- In contrast to solely time-discrete ABMs, agents can operate beyond the scope of time steps and sample continuous time-intervals for their state-transitions.

We shortly describe all actions that are scheduled and executed by one person-agent within one time step sorted by the specified module.

Population. As briefly described in [13], agents trigger birth and death events always via time- and age-dependent probabilities that apply for the observed time step (i.e. the observed day). If one of these events triggers, the agent samples a random time instant within the current time step and schedules the event. Note that in contrast to the basic population model, immigration and emigration events are disabled in the agent-based COVID-19 model due to closed borders in reality.

Contacts. Also contact specific events are scheduled and executed within the scope of only one time step: First of all, the agent schedules a contact event with every other member of its household. Moreover, if such a location is present in the contact network, a certain number of workplace or school contacts, respectively, are scheduled and corresponding partners drawn randomly from the assigned location. Finally, a certain number of leisure time contacts are sampled and partners are drawn based on a region-specific distribution: a leisure time contact partner is drawn uniformly from the same municipality with probability p^l_m, from a different municipality within the same district with probability p^l_d, from a different district in the same federal state with probability p^l_f, and finally, from a different federal state with probability p^l_a.

As mentioned, some states limit the agents' capabilities of interaction. In specific, *quarantined* or *preventively quarantined* agents have no random leisure time contacts and no contacts at *work* or *school*. Furthermore, *hospitalised* agents do not even contact their household members.

Disease. First of all, it is important to mention that the model is not parametrised by a reproduction number R_0 or R_{eff}, but by a contact specific probability. Nevertheless, the agent-based model provides the opportunity to carry out estimates for R_0 and R_{eff} by its original definition: the average number of secondary infections of an infected agent. Hence, what comes as model input for many traditional SIR models becomes a model output for the agent-based COVID-19 model.

Anyway, in case of an aforementioned contact, *infectious* agents spread the disease to susceptible agents with a certain *infection probability* which triggers the start of the newly-infected agent's patient-pathway. This pathway describes the different states and stations an agent lives through while suffering from the COVID-19 disease and can be interpreted as a sequence of events of which each triggers the next one after a certain sampled duration.

We show this infection strategy in a state chart in Figure 2 and describe how to interpret this figure by explaining the initial steps in the pathway in more detail: As soon as a person-agent becomes infected, its *infected* state is set to *true* and a *latency period* is sampled according to a specific distribution. The corresponding "Infectious" event is scheduled for the sampled time instant in the future. As soon as this "Infectious" event is executed, the *infectious* parameter is set to *true* and a random number decides about whether the person will develop symptoms or not. This point marks the first branch in the patient's pathway and whether the "Symptoms Onset" event or the "Asymptomatic" Event is scheduled. The prior would be planned after a sampled time span corresponding to the difference between latency and incubation time, the latter would be triggered instantaneously. All other elements of the pathway follow analogously. The branches are evaluated with age-class-dependent probabilities (see Section 1.3.3).

Finally, it is important to clarify that the model does not specifically consider deceased agents as this is not within the scope of the model (see Section 1.1.1). Consequently, all agents recover from the disease and are assumed to be immunised thereafter.

Policies. Every policy is modelled as a global event occurring before the planning phase of any of the simulation time steps. Policies are timed-events that can be fed into the model as a timeline. The elements of this timeline may include the following policies:

- Close / Open Location. The model classifies a fraction of locations of a certain type as closed (except *household*) and makes them unavailable for contacts.
- Reduce contacts. Agents reduce the daily number of contacts within a certain location or in leisure time by a scaling factor. Leisure time contacts can be reduced dependent on the agent's age class, $[0, 18), [18, 65), 65+$.
- Increase awareness. Person-agents start to react more quickly to their symptoms and reduce their *reaction time.*
- Hygienic Measures. At specific locations the *infection probability* is reduced.
- Location Tracing. Locations are closed precautionarily, if one of its assigned person-agents becomes *confirmed.*
- Individual Tracing. A certain fraction of agents record their contacts. If the agent becomes confirmed, the traced contacts become isolated precautionarily.

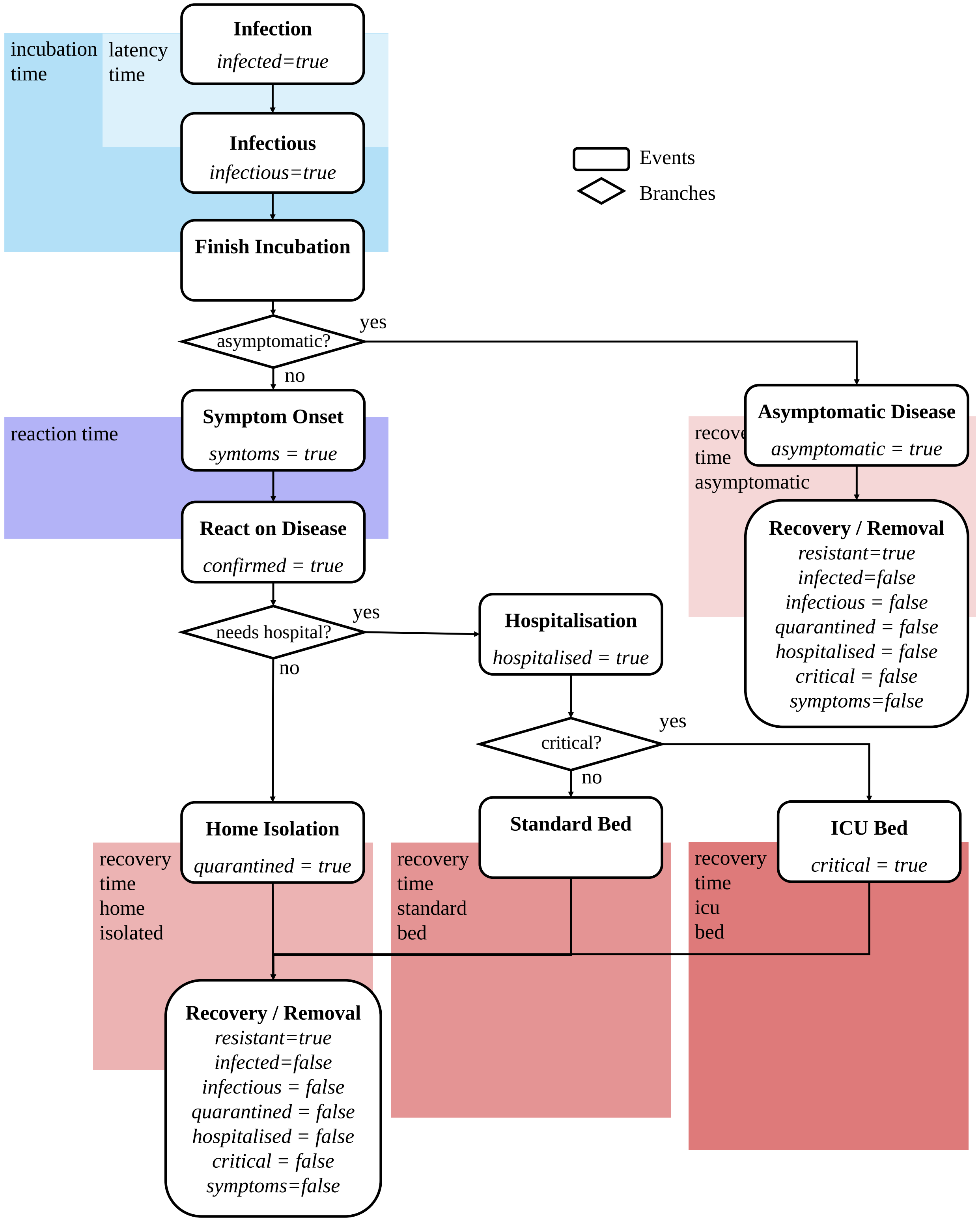

Figure 2: State chart of the patient pathway of a person-agent in the agent-based COVID-19 model. Only those state variables that are changed by the corresponding event are labelled, all others remain at the current value. The initial state of all infection-specific state variables is *false*.

1.2 Design Concept

1.2.1 Basic Principles.

In order to fulfill the modelling purpose, the agent-based model is designed as simple as possible yet depicting the most important features for evaluation of policies. Consequently lots of details within the pathway of an infected person and, in particular, lots of details within the personal daily routine are simplified to avoid indeterminable model parameters and unpredictable model dynamics.

1.2.2 Emergence.

In addition to the classic emergence of nonlinear epidemiological effects, analysis of the effects of interplay of different measures is one of the key objectives of the model. Hereby, seemingly unconnected policies like *school closure* and *contact reduction for the 65+* might lead to surprising effects when applied simultaneously. More generally speaking, the model displays that the individual effects of applied policies do not add up linearly.

1.2.3 Sensing.

Agents' perception of reality is one of the key problems of modelling COVID-19 as no agent is actually aware of its own disease and, more importantly, infectiousness until symptoms occur. Therefore, agent parameters can be distinguished into two sets: the ones the agent is aware of (e.g. *symptoms*, *hospitalised*), and the ones it is not (e.g. *infected*, *infectious*).

Interestingly, besides the individual perception of agents and the perception of an omniscient observer, there is also a third level of perception included into the model: the perception of the general public. While an individual agent knows about its symptoms, the public is not yet aware of this additional infected case, until the person-agent has reacted on the disease, has had itself tested and eventually becomes *confirmed*. Consequently, the levels of perception can be sorted with regards to their amount of knowledge:

$$\text{omniscient observer} > \text{person-agent} > \text{general public.}$$

1.2.4 Interaction.

Interaction between agents only occurs in form of contacts at *locations* or *leisure time*. The features provided by the underlying population model make it possible to investigate contacts on a very local level. As described before, *leisure time* contacts are weighted by their regionality, but also *school* and *workplace* contacts depict locality: Using specified latitude and longitude for locations, it is possible to assign person-agents with distance-dependent probabilities (see Section 1.3.1). Consequently, interactions between agents follow a spatially-continuous locally-biased contact network.

1.2.5 Stochasticity.

Basically all model processes, including the initialisation, contain sampling of random numbers. Therefore, Monte Carlo simulation is applied, results of runs are averaged and also their variability is assessed (see Section 2.1).

1.2.6 Observation.

As mentioned, a "recorder agent" takes care about tracking and aggregating the current status of the simulation. At the end of each global time step, all person-agents report to the recorder agent which furthermore keeps track of all necessary aggregated model outputs. This includes for example *confirmed active cases*, *confirmed cumulative cases*, *hospitalised agents*, *asymptomatic agents*, *pre-symptomatic agents*, *recovered agents*, *agents in a certain hospital*, or *average-number of contacts per infectious agent.* If required, numbers can also be tracked with respect to age, sex, regional level and/or location.

1.3 Details

Clearly, Section 1.1 could only outline the basic concepts of the model and left a lot of technical and modelling details necessary for a reproducible model definition open. In particular, this refers to the highly non-trivial initialisation process of the model. Hereby, two problems occur that require completely different approaches. The first problem considers the generation of the person-agents, locations and hospitals in the first place. The second problem deals with the initialisation of the status quo of the distribution of the disease states of the agents for the specified initial date.

1.3.1 Initialisation of Person-agents, Locations and Hospitals.

A lot of problems that deal with the sampling of the initial population have already been solved in the original GEPOC model [13]. In particular this refers to the delaunay-triangulation-based sampling method for locations. We apply this method to merge information from the national statistics institute and the global human settlement layer [15]. Consequently, besides initialisation of the disease states which is described in the next section, only new methods for location- and hospital-generation had to be implemented.

Schools are initialised based on known distributions w.r.t. average school size and number of pupils in total. A *school-sampler* iteratively generates schools with a random size/capacity (truncated normal distribution) until the sum of all capacities matches the known number of pupils in reality. Each school is furthermore sampled a position (latitude and longitude) analogous to the sampling for person-locations (see [13]). In a second step, schools are "filled" with person agents. Hereby, model agents with age between 6 and 18 are assigned to a random school within reach. This is done analogous to the sampling concept of leisure time contacts (see Section sec:process$_{o}verview)withprobabilitiesp^s_m, p^s_d, p^s_f$ and p^s_a for assigning the school in the same municipality, district, federal-state and country as the agent's residence. Clearly, the number of model agents in this age group is larger than the number of known pupils. Consequently, we force distribution of all 6 to 14 year old agents, and distribute as many 15 to 18 year old agents as possible. All remaining 15 to 18 year old agents are considered to be working.

Workplaces[1] are initialised analogously to schools. A *workplace-sampler* iteratively generates workplaces with size/capacity according to a discrete distribution (see Table 2). The sampler stops generating if the sum of all capacities matches $(a+b)(1-\alpha)$, whereas a denotes all model agents between 19 and 64, b denotes all agents between 15 and 18 that have not been yet assigned a school, and α denotes the current unemployment rate. Location sampling and "filling" works analogously to the *school-sampler* with different probabilities.

Hospitals are generated based on publicly available data. This includes capacities (beds, intensive-care units) as well as their location (latitude and longitude).

1.3.2 Initialisation of the Disease State

The spread of SARS-CoV-2 displays probably better than any other system, that the most dangerous enemy is the invisible one. While confirmed infected persons are detected and well known, they hardly contribute to the spread of the disease – they are already isolated properly, and most infections occur even before the onset of symptoms.

---

[1] Workplaces should not be confused with total companies. They rather represent the different teams where the members are in regular contact with each other.

Consequently, it is not possible to simply "start" the simulation with a certain number of confirmed cases, acquired for example from official internet sources. Valid values for pre-symptomatic (e.g. persons within latency and incubation period) and asymptomatic persons need to be acquired as well – yet, this number is hardly measurable in reality.

In order to solve this problem, a three stage concept, henceforth denoted as initialisation phase, was designed to generate a feasible initial state for a certain time t_0:

1. **Initialise-Simulation.** The agent-based COVID-19 model is set up with a small number of initially infected agents. This number corresponds to an estimated count of initial infection clusters in the country, but actually hardly influences the outcome. Furthermore, the agent-based simulation is run and interrupted by a state event, namely if the cumulative number of *confirmed agents* in the model is greater or equal to a specific value $C(t_{-1})$. Hereby, t_{-1} refers to a self chosen point in time and $C(t_{-1})$ to the reported number of positive tests in reality until t_{-1}. Hereby t_{-1} must be chosen properly so that the reported number of positive tests is large enough to be representative yet before implementation of any policies.

   As soon as the simulation is interrupted by the state-event, the timelines of simulation and reality are synced: t_{-1} in reality becomes t_{-1} in the simulation.

   The initialise-simulation is continued, considering all policies that have been implemented in reality, until, finally, t_0 is reached. Properly calibrated by a calibration routine (see Section 1.3.4), the initialise-simulation contains approximately the same cumulative number of *confirmed agents* as the corresponding reported number in the real system.

   The initialise-simulation is finished by exporting parts of the final state of the simulation. This refers to all households that contain either infected or recovered agents which are finally written into a file. Hereby, an initial population is generated that contains not only a valid approximation of the confirmed cases, but also a valid estimate for the unknown pre-symptomatic and asymptomatic persons, a correct distribution of their future planned events and a correct household distribution as well.

2. **Fine Tuning.** Even with best calibration routines (see Section 1.3.4) it is not possible to perfectly match the model output with the status quo in reality, in particular w.r.t. regional distribution. Therefore, a bootstrapping algorithm was implemented that corrects the small differences between the initialise-simulation output and the real data (confirmed cases, hospitalisation, intensive-care units and recoveries per region) to make sure, that the initial state of the actual simulation matches the current state precisely. This step can be omitted, if matching the current state precisely is not required.

3. **Load Households.** Finally, the actual simulation is initialised with the previously recorded and fine-tuned agents from the initialise-simulation. To be precise, this process does not only include agents themselves, but also the *households* these agents live in. Hereby, at least, the fundamental network structure from the initialise-simulation can be maintained.

1.3.3 Parametrisation

We finally state a list of used parameters and parameter-values including corresponding sources and/or justifications. They are found in Tables 1, 2, and 4.

Table 1: List of population specific parameters

parameter	description	value	source
birthrates, *deathrates*, *initial population*, *regional distribution*	parameters used by the underlying population model	see source	rates and population tables from Austrian National Statistics Institute [7]. Maps from the Global Human Settlement Project [5,15] and [4].

Table 2: List of contact specific parameters. Note that all parameter values are specified for the standard model without policies. The Γ-distribution is given as $\Gamma(k, \theta)$.

parameter	description	value	source
leisure time contacts per day	number of leisure time transmission-relevant contacts of an agent per day	$X \sim \Gamma(6.11, 1.0)$	based on the results of the POLYMOD study [24]
workplace contacts per day	number of transmission-relevant contacts at work (if assigned) of an agent per day	$X \sim \Gamma(5.28, 1.0)$	based on the results of the POLYMOD study [24]
school contacts per day	number of transmission-relevant contacts at school (if assigned) of an agent per day	$X \sim \Gamma(4.64, 1.0)$	based on the results of the POLYMOD study [24]
household sizes and structure	distribution of household sizes and structure	see source	distribution and structure from freely accessible tables for household statistics from the Austrian National Statistics Institute [7]
school sizes	The actual number of schools and pupils were gathered to calculate the average size of schools for the model. Based on this average, sizes for the simulation are sampled truncated normally.	$X \sim N(\mu, \sqrt{\mu})$, with $\mu = \frac{pupils}{schools}$ and $X > \mu/4$ a.s.	counts gathered from a publication of the Austrian National Statistics Institute [9]
workplace sizes	discrete distribution of workplace sizes	see source	gathered from a survey [10] by the Austrian National Statistics Institute
unemploy-ment rate	fraction of adults who are not assigned a workplace	10.4%	according to Austrian definition (AMS) gathered from the web-page of the city of Vienna [8]
regional distribution of leisure time contacts	leisure time contact partners are sampled based on locally biased distribution according to the regional structure: municipality, district and federal state	$[p^l_m, p^l_d, p^l_f, p^l_a] = [0.7, 0.2, 0.05, 0.05]$	estimates
regional distribution of schools	schools are assigned based on locally biased distribution according to the regional structure: municipality, district and federal state	$[p^s_m, p^s_d, p^s_f, p^s_a] = [0.66, 0.15, 0.13, 0.06]$	data from the Austrian Bureau of Statistics [7]
regional distribution of workplaces	workplaces are assigned based on locally biased distribution according to the regional structure: municipality, district and federal state	probabilities analogous to *regional distribution of schools* per federal-state	according to commuting behaviour data from the Austrian Bureau of Statistics (Pendlerinnen und Pendler, [7])

Table 3: List of disease specific parameters (1/2).

parameter	description	value	source
infection probability	probability that a contact between a susceptible and an infected agent leads to a transmission. Depending on the type of contact (household, school, work, leisure time) this parameter is scaled by a factor	$\alpha_{\text{household}} = 4.5 \cdot 0.057 \approx 0.257$, $\alpha_{\text{workplace}} = \alpha_{\text{school}} = \alpha_{\text{leisure time}} = 0.057$	calibrated based on the doubling rates in Austria before introduction of policies (see Section 1.3.4).
incubation time	time between infection and symptom on-set	scaled β distribution with $\min(X) = 2[d]$, $\max(X) = 14[d]$, $E(X) = 5.1[d]$	based on [21]
latency time	time between infection and infectivity	always incubation time minus $2[d]$	estimates by the Robert Koch Institute ([6], Section 8)
reaction time	time between symptom onset and reaction/testing of the agent which furthermore leads to its confirmation and isolation	$X \sim \text{Weibull}(4.29, 1.65)$	based on [19]

Table 4: List of disease specific parameters (2/2).

parameter	description	value	source
recovery time home quarantined	time between symptom onset and recovery for symptomatic persons in home isolation	$X \sim Tri(11, 18, 25)$	based on expert opinions and calibrated w.r.t. official recovery data
recovery time severe	time between symptom onset and recovery for symptomatic persons in normal hospital beds	$X \sim \text{Weibull}(18.0, 1.76)$	Cox regression on reported data by Austrian hospitals (Elektronisches Meldesystem by AGES [2])
recovery time critical	time between symptom onset and recovery for symptomatic persons in ICU beds	$X \sim \text{Weibull}(27.0, 3.0)$	Cox regression on reported data by Austrian hospitals (Elektronisches Meldesystem by AGES [2])
recovery time asymptomatic	time between symptom onset and recovery for asymptomatic persons	$Tri(1, 5, 7)$	estimated based on expert opinions
probability asymptomatic	probability of an asymptomatic disease progression	[92.5, 78.5, 46.2, 50.9, 40.6, 30.8, 50.6, 48.5, 18.3, 0.0]% for 10 year age-classes	Age distribution is gathered by comparison of Austria's age pyramid with the distribution of the confirmed cases. The overall probability is calibrated to 50% as stated by antibody tests in Iceland.
hospitalisation probability	age-dependent probability that a symptomatic patient requires hospitalisation	[1.0, 2.6, 2.7, 7.1, 10.3, 26.7, 48.3, 61.8, 53.3]% for 10 year age-classes	Distribution is based on comparison of the age distribution of confirmed cases with the age distribution of hospitalised cases (Elektronisches Meldesystem by AGES [2]). The overall probability is calibrated to Austrian hospital data.
icu probability	probability that a hospitalised agent becomes critical (needs an intensive care unit)	15.4%	calibrated for Austrian ICUs with data from the ministry of internal affairs
initialisation data	current epidemiological numbers for Austria for initialisation (see Section 1.3.2)	see source	initial data dumps updated on regular basis via Elektronisches Meldesystem by AGES [2]

1.3.4 Calibration

Clearly, there is no valid data available for direct parametrisation of the *infection probability* in case of a direct contact. First of all, this parameter is hardly measurable in reality and moreover strongly depends on the definition of "contact". Consequently, this parameter needs to be fitted in the course of a calibration loop.

The calibration experiment is set up as follows:

- We vary the parameter *infection probability* using a bisection algorithm.

- For each parameter value, the simulation, parametrised without any policies, is executed ten times (Monte Carlo simulation) and the results are averaged.
- The average time-series for the *cumulative confirmed cases* is observed and cropped to the beginning upswing of the epidemic curve, to be specific, all values between 200 and 3200. In this interval the growth of the curve can be considered as exponential.
- The cropped time-series is compared with the corresponding time-series of real measured data in Austria, specifically the confirmed numbers between March 10^{th} and 20^{th} 2020 (source EMS system, [2]).
- Both time-series are compared w.r.t. the average doubling time of the confirmed cases. The difference between the doubling times is taken as the calibration error for the bisection algorithm.

Note: As the sample standard deviation of each observable of the runs has been observed to be at most a fifth of the sample mean, the iteration number of nine for the Monte Carlo simulation has been considered to be sufficient for calibration purposes w.r.t. the ideas in [14, 20].

2 Model Implementation

Simulation of agent-based models like the agent-based COVID-19 model is a huge challenge with respect to computational performance. As the model cannot be scaled down, almost 9 Million interacting agents need to be included into the model in order to simulate the spread of the disease in Austria.

These high demands exclude most of the available libraries and software for agent-based modelling including AnyLogic [16], NetLogo [26], MESA [22], JADE [11] or Repast Simphony [25]. Most of these simulators cannot be used as their generic features for creating live visual output generates too much overheads.

Consequently, we decided to use our own agent-based simulation environment ABT (Agent-Based template, see [3]), developed in 2019 by dwh GmbH in cooperation with TU Wien. The environment is implemented in JAVA and specifically designed for supporting reproducible simulation of large-scale agent-based systems.

The next section contains more technical details about the implementation.

2.1 Technical Implementation Details

The implementation of the agent-based COVID-19 model uses JAVA 11 and applies the *UniformRandomProvider* random number generator (RNG) by Apache Commons [1]. This RNG implements a 64 bit version of the Mersenne Twister [23] and exceeds the standard RNG of JAVA, a simple Linear Congruential Generator, in both performance and quality.

The simulation itself is always executed in a Monte Carlo setting and several runs with different RNG seeds are averaged. Due to the huge number of agents, a Law-of-Large-Numbers-effect can be observed (similar to [12] Chapter 5.2), and the standard deviation of the model output is always comparably small. Consequently, Monte Carlo replication numbers of 10 to 20 are usually enough to estimate the mean sufficiently well (we apply the algorithms from [14, 20]).

3 Features and Limitations

Due to the highly flexible policy timeline, the model is capable of testing and combining lots of different policies in different characteristics at different times. Hence, it can easily depict almost any specified policy announced in reality, if estimates for the policy parameters are available.

The latter statement particularly refers to combination of policies: although the model correctly depicts the epidemiological impact of the combination of policies, the social impact needs to be parametrised manually. For instance, the causal relation between closed schools and intensified parent-children contacts needs to be parametrised and is not given by the model dynamics.

Unfortunately, as the model cannot be scaled down, a huge number of agents leads to long computation times, and the necessity of Monte Carlo simulation for flattening of stochastic results increases the time required to get simulation output even further. Consequently, the simulation's capabilities of dealing with multi-variate calibration problems are limited and the model is unhandy to generate short-time prognoses.

Finally, the model disregards "death" as a final disease state. However, this limitation should not be interpreted as a lack of a necessary property (the model would easily be capable of adding this feature), but originates from a conscious decision made at the beginning of the modelling process. The feature is irrelevant for our modelling purpose and, due to high medial interests of our work, we did not intend to contribute to the rising panic among the population due to death-count prognoses.

References

[1] Apache commons rng homepage. `https://commons.apache.org/proper/commons-rng/`. Accessed: 2020-04-17.

[2] Covid-19 information page by ages. `https://www.ages.at/en/wissen-aktuell/publikationen/epidemiologische-parameter-des-covid19-ausbruchs-oesterreich-2020/`. Accessed: 2020-04-08.

[3] dwh gmbh news entry for the abt simulation framework. `http://www.dwh.at/en/news/the-power-of-the-abt-simulation-framework/`. Accessed: 2020-04-17.

[4] Geo- and topojson files of municipalities, districts and states in austria by flooh perlot. `https://github.com/ginseng666/GeoJSON-TopoJSON-Austria`. Accessed: 2019-05-12.

[5] Global human settlement layer. `https://ghsl.jrc.ec.europa.eu/`. Accessed: 2020-04-08.

[6] Sars-cov-2 steckbrief zur coronavirus-krankheit-2019 (covid-19). `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html#doc13776792bodyText7`. Accessed: 2020-05-29.

[7] Statistik austria. `https://www.statistik.at/web_de/statistiken/index.html`. Accessed: 2020-04-08.

[8] Web-page of the city of vienna. `https://www.wien.gv.at/`. Accessed: 2020-03-01.

[9] Statistik Austria. Bildung in zahlen 2017/18-tabellenband. 2019.

[10] Wien Statistik Austria. *Arbeitsstättenzählung 2001*. Verlag Österreich, 2004.

[11] Fabio Bellifemine, Agostino Poggi, and Giovanni Rimassa. Jade–a fipa-compliant agent framework. In *Proceedings of PAAM*, volume 99, page 33. London, 1999.

[12] Martin Bicher. *Classification of Microscopic Models with Respect to Aggregated System Behaviour*. Dissertation, TU Wien, Vienna, Austria, November 2017.

[13] Martin Bicher, Christoph Urach, and Niki Popper. GEPOC ABM: A Generic Agent-Based Population Model for Austria. In *Proceedings of the 2018 Winter Simulation Conference*, pages 2656–2667, Gothenburg, Sweden, 2018. IEEE.

[14] Martin Bicher, Matthias Wastian, Dominik Brunmeir, Matthias Rößler, and Niki Popper. Review on Monte Carlo Simulation Stopping Rules: How Many Samples Are Really Enough? In *Proceedings of the 10th EUROSIM Congress on Modelling and Simulation*, Logrono, Spain, July 2019. In Print.

[15] Aneta Jadwiga Florczyk, Stefano Ferri, Vasileios Syrris, Thomas Kemper, Matina Halkia, Pierre Soille, and Martino Pesaresi. A new european settlement map from optical remotely sensed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1978–1992, 2016.

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

[17] Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K. Heinz, Geir Huse, Andreas Huth, Jane U. Jepsen, Christian Jørgensen, Wolf M. Mooij, Birgit Müller, Guy Pe'er, Cyril Piou, Steven F. Railsback, Andrew M. Robbins, Martha M. Robbins, Eva Rossmanith, Nadja Rüger, Espen Strand, Sami Souissi, Richard A. Stillman, Rune Vabø, Ute Visser, and Donald L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1):115–126, 2006.

[18] Volker Grimm, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. The ODD protocol: A review and first update. *Ecological Modelling*, 221(23):2760–2768, 2010.

[19] Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Fiona Sun, et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 2020.

[20] Juan Ignacio Latorre Jimenez. EUROSIM 2019 Abstract Volume. In *EUROSIM 2019 Abstract Volume*. ARGESIM, 2019.

[21] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 2020.

[22] David Masad and Jacqueline Kazil. Mesa: an agent-based modeling framework. In *14th PYTHON in Science Conference*, pages 53–60, 2015.

[23] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[24] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, and others. POLYMOD social contact data. 2017.

[25] Michael J North, Thomas R Howe, Nick T Collier, and Jerry R Vos. The repast simphony runtime system. In *Proceedings of the agent 2005 conference on generative social processes, models, and mechanisms*, volume 10, pages 13–15. Citeseer, 2005.

[26] S. Tisue and U. Wilensky. NetLogo: A simple environment for modelling complexity. pages 16–21, 2004.