

# Modelling and Simulation of a SIR-type Epidemic with Cellular Automata and Ordinary Differential Equations – Definition ARGESIM Benchmark C17R

Florian Miksch<sup>1\*</sup>, C. Haim<sup>2</sup>, Günter Schneckenreither<sup>1,2</sup>, Nikolas Popper<sup>1,2</sup>

<sup>1</sup>dwh simulation services, Neustiftgasse 57-59, 1070 Vienna, Austria; \*[florian.miksch@dwh.at](mailto:florian.miksch@dwh.at)

<sup>2</sup>Inst. of Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria

Simulation Notes Europe SNE 25(1), 2015, 49 - 54  
DOI: 10.11128/sne.25.bn17r.10283  
Received: October 20, 2014; Revised February 10, 2015;  
Accepted: February 15, 2015;

**Abstract.** This Comparison investigates a classical population model for the spread of infection diseases (SIR ordinary differential equations model by Kermack and McKendrick) and an inhomogeneous spatial approach using cellular automata. An identification of parameters based on an abstract time discrete conceptual model is presented. The tasks of this comparison include the validation and analysis of this identification, an investigation on the impact of different spatial dynamics in the cellular automaton modelling approach and simulation scenarios for confining epidemic outbreaks that involve state-dependent interventions.

## Introduction

This comparison is a revision of the original ARGESIM Comparison 17 [1] and is targeted at the identification of the classical SIR-type differential equations model by Kermack and McKendrick [2] (which is a cumulative population model) with a microscopic individual based cellular automaton modelling approach [3].

For the purpose of a systematic identification of both approaches, a virtual individual based time-discrete population system with contact-induced SIR-characteristic spread of an infectious disease is presented. Based thereon the differential equations and cellular automaton models are derived and identified in an analytical fashion. Furthermore intervention scenarios for confining epidemic outbreaks are discussed.

## 1 System Definition

Let  $N$  be the number of individuals of the population. The population can neither be joined nor left by individuals, which means that  $N$  is a constant number. Each individual shall be in one of the states susceptible, infected or recovered. The system evolves by discrete steps of one time unit and the spread of the disease is characterised by contacts between individuals, transmission of the disease and recoveries.

Parameter	Description
$S_0$	initial number of susceptible
$I_0$	initial number of infected
$R_0$	initial number of recovered
$C$	contacts
$\alpha$	infection probability
$\beta$	recovery probability

**Table 1.** System parameters.

Each individual is assumed to have an average of  $C$  contacts per time step; these contacts always happen between two random individuals. Since the discrete time steps are atomic by definition, the order of contacts is irrelevant. However, in order to ensure that susceptible individuals cannot get infected and infect others simultaneously, the infection-states of the individuals change after all contacts have been processed according to the following paradigms:

- When a susceptible individual gets into contact with an infected individual, the susceptible individual becomes infected with probability  $\alpha$ . This probability applies for each contact separately.
- Infected individuals recover at the end of each time unit with probability  $\beta$ .
- Recovered individuals always remain recovered.

**Interventions.** In order to confine an epidemic, interventions might be applied. We define two different types of strategies ('soft' and 'hard') that can be applied when a certain critical threshold of infected individuals is reached or exceeded. The threshold is defined relative to the whole population as  $f_T N$ .

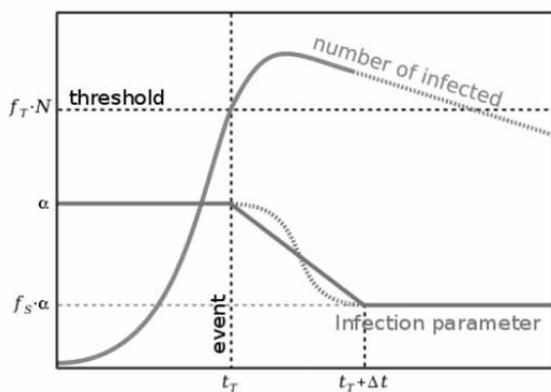
As a 'soft' strategy, the system parameters  $C$ ,  $\alpha$  or  $\beta$  are decreased to  $f_S \cdot C$  (or  $f_S \cdot \alpha$  or  $f_S \cdot \beta$ ) over a period of time  $\Delta t$ .  $0 \leq f_S \leq 1$  is called the reduction parameter. This can be either a linear decrease in the form  $f(x) = k \cdot x$  or a smooth step in the form  $f(x) = 3 \cdot x^2 - 2 \cdot x^3$ . Let  $r(t)$  be the function that describes the decrease from 1 to  $f_S$ , so that it can be multiplied with  $C$ ,  $\alpha$  or  $\beta$ , and let  $t_T$  be the time when the threshold is reached. For a linear decrease

$$r(t) = \left(1 - \frac{t - t_T}{\Delta t}\right) + f_S \frac{t - t_T}{\Delta t} \tag{1}$$

and for a smooth step it can be

$$r(t) = 1 - \left(3 \left(\frac{t - t_T}{\Delta t}\right)^2 - 2 \left(\frac{t - t_T}{\Delta t}\right)^3\right) \cdot (1 - f_S). \tag{2}$$

If  $\Delta t = 0$ , then the change is a discontinuous step. Figure 1 illustrates the idea and expected outcome of a 'soft' strategy.



**Figure 1.** Illustration of a 'soft' intervention. Once the number of infected reaches a critical threshold, the infection parameter decreases over a certain period of time.

'Hard' strategies involve the individuals directly. Representing a quarantine or vaccination strategy, susceptible or infected individuals, respectively, can become recovered. Such a 'hard' strategy is defined by choosing a fraction  $f_H$  of susceptible or infected individuals and immediately changing their state to 'recovered' when the threshold of infected individuals is reached. The individuals are chosen randomly among all possible individuals with the respective state, since further distinction is not possible with this system definition.

Table 2 lists all parameters that are relevant for interventions.

It could be that the threshold is reached more than once. This happens, for example, when the number of infected is growing, then it is reduced by an intervention but still keeps growing. The intervention strategy can be applied in two ways: either only once when the threshold is reached for the first time or every time it is reached.

Parameter	Description
$f_T$	Fraction that defines the threshold
$f_S$	Reduction parameter of a soft intervention
$\Delta t$	Duration of a soft intervention
$f_H$	Fraction parameter of a hard intervention

**Table 2.** Parameter of hard and soft interventions.

## 2 Differential Equations Model

The differential equation, which models the defined system, corresponds to the classical SIR epidemic model which was proposed by W. O. Kermack and A. G. McKendrick in 1926 [2].  $S'(t)$ ,  $I'(t)$  and  $R'(t)$  represent the change of susceptible, infected and resistant individuals. The amount of susceptible individuals that become infected is described as  $\gamma S(t)I(t)$ , where  $\gamma$  is referred to as infection rate. The amount of infected individuals that become resistant is described as  $\delta I(t)$ , and  $\delta$  is named recovery rate. Since the number of individuals in our system shall be constant, these growth terms yield the following system of ordinary differential equations (ODE) shown in (3).

$$\begin{aligned}
 S'(t) &= -\gamma \cdot S(t) \cdot I(t) \\
 I'(t) &= \gamma \cdot S(t) \cdot I(t) - \delta \cdot I(t) \\
 R'(t) &= \delta \cdot I(t)
 \end{aligned}
 \tag{3}$$

Before setting the ODE parameters, one needs to deal with another concern. An ODE  $A'(t) = -\xi A(t)$  with  $0 \leq \xi \leq 1$  represents a system where  $A$  is continuously decreased. However,  $\xi$  does not represent the amount of decrease within one time unit. If  $A$  should be decrease to  $\hat{\xi} \cdot A$  within one time unit, then  $\xi$  needs to be set as

$$\xi = \ln(1 - \hat{\xi}).
 \tag{4}$$

This is based on the fact that the general solution of the ODE is  $A(t) = A(0) \cdot e^{-\xi t}$ . The condition that  $A$  should be decreased leads to the equation  $A(t + 1) = (1 - \hat{\xi}) \cdot A(t)$ , and further results in the formula above.

Identification of the infection term  $\gamma \cdot S(t) \cdot I(t)$  takes a look at a single individual, which has  $C$  contacts per time unit in average. Among the contacts, it has  $C \cdot \frac{I}{N}$  contacts with infected individuals. Each contact causes a transmission with probability  $\alpha$ . The transmissions are statistically independent events. Hence, the infection probability per time unit is computed as the probability to get infected at least once, which is represented by the formula  $1 - (1 - \alpha)^{C \cdot \frac{I}{N}}$ . Considering Equation (4), the infections in the ODE are represented by  $-\ln\left(1 - 1 + (1 - \alpha)^{C \cdot \frac{I}{N}}\right) \cdot S$ , which can be rewritten as  $-I \cdot S \cdot \frac{C}{N} \cdot \ln(1 - \alpha)$ . Hence,  $\gamma$  is identified with  $-\frac{C}{N} \cdot \ln(1 - \alpha)$ .

Since the recovery rate  $\delta$  determines the fraction of infected individuals that recover during one time unit,  $\delta$  calculates as  $-\ln(1 - \beta)$ .

The identified parameters are summarised in Table 3.

Parameter	Identification
$S(0)$	$S_0$
$I(0)$	$I_0$
$R(0)$	$R_0$
$\gamma$	$-\frac{C}{N} \cdot \ln(1 - \alpha)$
$\delta$	$-\ln(1 - \beta)$

**Table 3.** Parameter identification of the differential equation model.

**Interventions.** In a ‘soft’ intervention strategy, the parameter  $\gamma$  or  $\delta$  in the ODE system needs to switch to a time dependent function  $\hat{\gamma}(t)$  or  $\hat{\delta}(t)$  when the threshold  $f_T N$  is reached at time  $t = t_T$ .

$\hat{\gamma}(t)$  or  $\hat{\delta}(t)$  calculates as the term in Table 3 where the desired parameter  $C, \alpha$  or  $\beta$  is replaced by  $C \cdot r(t)$  or  $\alpha \cdot r(t)$  or  $\beta \cdot r(t)$ , and  $r(t)$  corresponds to the function in Equation (1) or (2).

At time  $t = t_T + \Delta t$ , the ODE system switches back to Equation (3) where  $\gamma$  or  $\delta$  is replaced by  $f_S \gamma$  or  $f_S \delta$ .

In a ‘hard’ intervention strategy the ODE system abruptly changes  $S(t)$  or  $I(t)$  when the threshold is reached. This can be achieved using the delta distribution  $D(x)$  where  $D(0) = 1$  and  $D(x) = 0$  for  $x \neq 0$

For example, quarantining the fraction  $f_H$  of infected individuals when the threshold is reached, then the ODE can be rewritten as Equation (5).

$$\begin{aligned}
 S'(t) &= -\gamma \cdot S(t) \cdot I(t) \\
 I'(t) &= \gamma \cdot S(t) \cdot I(t) \\
 &\quad - \delta \cdot I(t) - D(I - I_T) \cdot f_H \cdot I(t) \\
 R'(t) &= \delta \cdot I(t) + D(I - I_T) \cdot I_Q
 \end{aligned}
 \tag{5}$$

### 3 Cellular Automaton Model

In the context of modelling and simulation Cellular Automata (CA) can be seen as a time- and space-discrete modelling approach. A CA consists of cells which are arranged on a regular grid and can hold different states [4]. Lattice Gas Cellular Automata (LGCA) are an extension of the concept of CA where particles move around these cells [5]. Especially with a hexagonal lattice, LGCA are used for simulating the movement of gas particles or fluids.

We will additionally allow particles to take one of the states susceptible, infected or recovered [1, 3, 6] in order to simulate the spatial spread of a SIR-type disease.

Accordingly, we assume that our cells are arranged on a two-dimensional hexagonal grid structure and represent a spatial segment. Each cell can hold at most six individuals. Each individual is in one of the three states susceptible, infected or recovered. Contacts happen pairwise between all individuals which are located in the same cell at the same time. To simulate a mixture of the individuals, they move around the cells in random directions (diffusion) or as defined by the FHP-I collision rules [5]:

- The position of an individual within a cell defines its moving direction (Figure 2).
- After the movement phase a collision phase (Figure 3) takes place. The FHP-I variant only defines special two and three particle collisions. All other collisions happen without any change of moving direction. When two individuals collide as in Figure 3, they are reflected clockwise or counter clockwise with probability 0.5. When three particles collide as pictured in Figure 3, then they are reflected clockwise.

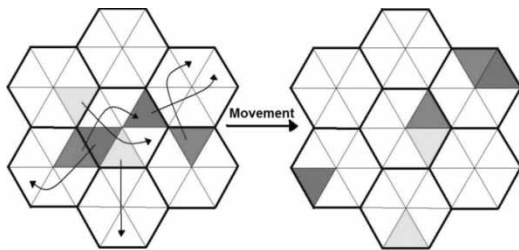


Figure 2. Schematic visualization of LGCA movement rules.

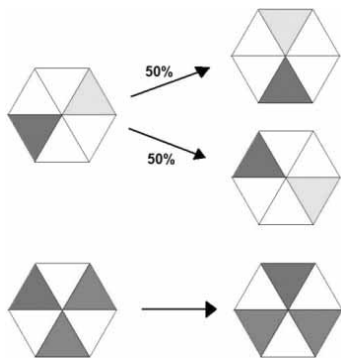


Figure 3. FHP-I collision rules.

When a susceptible individual meets an infected individual within a cell, it shall become infected with probability  $\varphi$ . An infected individual recovers with probability  $\phi$ .

The size of the LGCA plays an important role because it affects the density of particles and thus the number of contacts. For the sake of simplicity, we use a grid with  $width = length = n$  and accordingly  $n^2$  cells with six places each. Table 4 shows the parameters of the model with appropriate parameterisation. For a given number of individuals, the number of contacts depends on the size  $n$  of the LGCA.

The correct identification for  $n$  is crucial, but follows a simple calculation: Assuming a uniform distribution of the individuals, each of the six slots of a cell is occupied with the same probability. For a given individual there are 5 unoccupied slots in the same cell. Accordingly  $N - 1$  remaining individuals occupy  $6n^2 - 1$  remaining slots and the individual has an expected number of

$$C = 5 \cdot \frac{N - 1}{6n^2 - 1} \tag{6}$$

contacts within this cell. Adjusting  $n$ , which has to be an integer, to meet a given number of contacts leads to the identification in Table 4.

Parameter	Identification
$S(\mathbf{0})$	$S_0$
$I(\mathbf{0})$	$I_0$
$R(\mathbf{0})$	$R_0$
$\varphi$	$\alpha$
$\phi$	$\beta$
$n$	$\left\lceil \sqrt{\frac{5(N - 1) + C}{6C}} \right\rceil$

Table 4. Parameter identification of the cellular automaton model.

**Interventions.** Also for the CA approach intervention scenarios can be simulated. If the number of infected particles in the CA reaches the threshold, one of the strategies described in Section 1 can be applied.

Applying a ‘soft’ strategy is easy for  $\alpha$  and  $\beta$ . Then,  $\varphi$  or  $\phi$  are multiplied with  $r(t)$  in Equation (1) or (2) so that the  $f_5\varphi$  or  $f_5\phi$  are reached after  $\Delta t$ . Reducing  $C$  causes problems because it requires a change of  $n$ . First, changing  $n$  is very inaccurate and second, there is no instruction on how to enlarge or shrink the space in respect to the individuals that have positions on the grid. Thus, if a ‘soft’ strategy for  $C$  is desired, it should be performed very carefully.

‘Hard’ intervention strategies can be directly applied as stated in the system definition. Then the threshold is reached, a desired number of susceptible or infected individuals are randomly chosen, and immediately become recovered.

In contrast to the ODE model, individuals in the CA are distinguished by their spatial location on the lattice. Presumably, the selection of particular individuals/particles for changing their state can make a crucial difference. For testing purposes, it seems reasonable to deliberately violate the system definition and choose individuals with respect to their location.

## 4 Analytical Comparison

There is a strong analytical relation between the ODE approach and the CA model. For simplicity the parameter identification presented in this section neglects interventions.

**Infections.** The following calculation aims to estimate the number of new infections in a time step in the LGCA. Consider a susceptible individual in a cell (only susceptible individuals can get infected). Then there are altogether  $6n^2 - 1$  remaining slots in the LGCA, 5 remaining slots in the cell and  $I$  infected individuals. Define the probability of  $i$  slots in the cell being occupied by infected individuals as  $q_i$ . Under the assumption that the individuals are uniformly distributed, the number of infected individuals in this cell is distributed according to a hypergeometric distribution. The probabilities calculate as choosing  $i$  out of  $I$  infected individuals on 5 out of  $6n^2 - 1$  places:

$$q_i = \frac{\binom{5}{i} \binom{(6n^2 - 1) - 5}{I - i}}{\binom{6n^2 - 1}{I}}, \quad i = 0 \dots 5 \quad (7)$$

The expected value  $E$  of this hypergeometric distribution is

$$E = \sum_{i=0}^5 q_i i = I \frac{5}{6n^2 - 1}. \quad (8)$$

Using the identification in (6), the expected value can be written as

$$E = I \frac{C}{N - 1}. \quad (9)$$

With these preparations the actual infection probability of a susceptible individual can be calculated. If the cell is occupied by  $i$  infected individuals the probability for an infection of the susceptible individual is  $1 - (1 - \alpha)^i$ . Hence the expected probability for an infection is  $\sum_{i=0}^5 q_i (1 - (1 - \alpha)^i)$ . Considering the first two terms of the Taylor series expansion at  $\alpha = 0$  and the identification in (9) leads to the following approximation for this probability.

$$\begin{aligned} \sum_{i=0}^5 q_i (1 - (1 - \alpha)^i) &\approx \sum_{i=0}^5 q_i i \alpha = \\ &= \alpha \sum_{i=0}^5 q_i i = \alpha E = \alpha I \frac{C}{N - 1} \end{aligned} \quad (10)$$

Multiplying (10) with the total number of susceptible individuals leads to  $S \alpha I \frac{C}{N - 1}$  as an approximation for the expected total number of new infections for one time unit in the LGCA for small values of  $\alpha$ .

The term in the ODE for infections of one time unit per susceptible is computed in section 2 as  $1 - (1 - \alpha)^{\frac{C \cdot I}{N}}$ . In the term for the CA,  $\frac{I}{N - 1}$  can be approximated with  $\frac{I}{N}$ . Natural limitations are  $\frac{I}{N} \leq 1$  and  $C \leq 5$ . For small  $\alpha$  and the natural limitations,  $1 - (1 - \alpha)^{\frac{C \cdot I}{N}}$  is an approximation of  $\alpha \frac{C \cdot I}{N}$ .

**Recoveries.** An infected individual in the LGCA recovers during one time unit with probability  $\beta$ , hence the expected amount of infected individuals who regenerate in one time unit is  $\beta I$ . The same factor also occurs in the differential equation (3) and Table 3 as  $-\ln(1 - \beta)$ .

## 5 Tasks

In order to validate the analytical findings from Section 4 in an experimental fashion, both model approaches must be implemented in a simulation environment or as stand-alone programs. The primary output and point of comparison for both approaches is the evolution of the numbers  $S(t)$ ,  $I(t)$  and  $R(t)$ .

### 5.1 Task 1 – Model Comparison

We use the parameters from Table 5 as a starting point and perform a parameter analysis for  $I_0$ ,  $\alpha$  and  $\beta$  while keeping the overall population count ( $N = 10\,000$ ) constant. For the CA approach, FHP-I collision rules and a uniform distribution of the individuals as initial condition seem legit.

- Is it possible to identify parameter regions with similar behaviour in both modelling approaches? It is very likely that the reasons for qualitative and quantitative differences lie in the fact that the identification of the infection parameter is based on the assumption that the population is always uniformly distributed on the lattice and secondly that a Taylor approximation was used in (10).

- b) From Equation (10) we can see that the average probability of a susceptible individual to get infected depends on the contact rate  $C$  and the infection probability  $\alpha$ . We can analyse the trade-off between those two parameters by performing a parameter variation of  $\alpha$  and  $C$  leaving the product  $\alpha C$  constant. The dynamics of the CA imply different effects of variations in  $\alpha$  and  $C$  respectively. Note that, in particular, changing the contact rate  $C$  also involves changing the size  $n$  of the lattice and that the contact rate for the six-particle LGCA is limited.

Parameter	Value
$S_0$	9 500
$I_0$	500
$R_0$	0
$C$	4
$\alpha$	0.1
$\beta$	0.1

**Table 5.** Parameter set for Task 1.

## 5.2 Task 2 – Interventions

It is rather clear that different intervention strategies deliver different reactions of the system. Additionally for both modelling approaches the same intervention strategy can have different effects (compare Task 1).

- a) Based on a parameter set with similar behaviour in both modelling approaches (Task 1) choose a threshold  $f_T$  for the number of infected and compare different intervention strategies in both modelling approaches.
- b) What is the advantage of the spatial LGCA approach in combination with the ‘hard’ strategies? For example we can assume that vaccinating or quarantining individuals at the interface between regions with high and low infections reduces the spread of the disease. On the other hand, ‘hard’ interventions in the center of infected areas may have hardly any effect on the epidemic.

## 5.3 Task 3 – Spatial Inhomogeneity

A constantly homogeneous mixture of the population in the CA model can be achieved by introducing a third type of ‘movement’ rules: instead of moving to an adjacent cell, the particles jump to an arbitrary cell on the lattice (‘random movement’). Based on the findings from Task 1 we can choose a parameter setting for which the ODE approach and the original CA model deliver different qualitative behaviour. We can however postulate that the CA model with this new type of movement rules behaves identical to the ODE model. How can this be explained?

## References

- [1] Hötendorfer H, Popper N, Breitenecker F. *Temporal and Spatial Evolution of a SIR-type Epidemic – ARGES-IM Comparison C17 – Definition*, Simulation News Europe; 2004.
- [2] Kermack WO, McKendrick AG. *A Contribution to the Mathematical Theory of Epidemics*, Proc. R. Soc. Math. Phys. Eng. Sci.. 1927; 115(772): p 700–721.
- [3] Fuks H, Lawniczak AT. *Individual-based lattice model for spatial spread of epidemics*. Discrete Dynamics in Nature and Society. 2001;6: p 191–200.
- [4] Wolfram S. *A new kind of science*. Champaign, IL: Wolfram Media; 2002.
- [5] Wolf-Gladrow DA. *Lattice-Gas Cellular Automata and Lattice Boltzmann Models: An Introduction*, 1st ed. Berlin: Springer; 2000.
- [6] Yakowitz S, Gani J, Hayes R. *Cellular automaton modeling of epidemics*, Appl. Math. Comput.. 1990; 40(1): p 41–54.