# Comparison of differential equations and cellular automata for epidemic simulation

**Florian Miksch[1], Clemens Haim[2], Günter Schneckenreither[1], Felix Breitenecker[2]**

[1]*dwh Simulation Services, Vienna, Austria*
[2] *Vienna University of technology, Institute for Analysis and Scientific Computing, Vienna, Austria*
*Email: florian.miksch@dwh.at*

*Running through the modeling process, the choice of an appropriate modeling technique is one of the first and most fundamental questions. This crucial point might be responsible for success or failure. The right choice is almost never unique and depends strongly on the questions one wants to answer. This paper compares two modeling techniques coming from completely different point of views: ordinary differential equations and cellular automata. For comparison a simple epidemic spread is modeled and simulated with the two approaches and the results are investigated on both an experimental and analytical level. The idea is defining an underlying system that has to be modeled. This allows a standardized model parameterization and comparable representation of the results. It turns out that for many settings both models behave similarly and can be considered as describing the same system correctly. After reading this paper, one should be aware of the differences and similarities of both techniques especially when applied on epidemic spread and should know about the different model properties.*

## 1 Introduction

This paper should carve out the connections and differences of two modelling techniques for simulation of infectious disease propagation. For this aim a well-defined system is given which has to be simulated using differential equations and a cellular automaton [1]. The system describes a simple SIR-type epidemic, based on the ideas of Kermack and McKendrick [2]. The idea is to simulate the system with both approaches to find out differences and similarities.

## 2 Definition of the System

The system describes the spread of a SIR-type disease. Given is a population of N individuals. Each individual is in one of the states susceptible, infected and recovered. The population is constant for the whole simulation which means that no individuals can enter or leave the system. The system evolves by discrete time steps of unit one.
Describing the system, the spread of the disease is given by contacts, transmissions and recoveries. Each individual has in average C contacts per time step; contacts always happen between two random individuals. Time steps cannot be split up by definition, hence the order of contacts within a time step is irrelevant.
State changes of individuals happen after the contacts and apply for the following time unit. When an infected individual has a contact with a susceptible individual, the susceptible individual gets infected with probability α. This probability applies for each contact separately. Infected people recover at the end of each time unit with probability β. This strict procedure ensures that susceptible individuals cannot get infected and infect others within the same time unit. Recovered individuals remain in this state for the rest of the time.
Together with the six system parameters in table 1, the system is completely defined.
The task is simulating the system and track the number of susceptible, infected and recovered individuals in each time unit, given as time-dependent function S(t), I(t) and R(t).

Table 1. The parameters of the system

| Parameter | Description |
|---|---|
| S0 | number of susceptible individuals in time 0 |
| I0 | number of infected individuals in time 0 |
| R0 | number of recovered individuals in time 0 |
| C | contacts |
| α | infection probability |
| β | recovery probability |

## 3 Modeling with differential equations

The differential equations model generally follows the well-known SIR model which was proposed first by W. O. Kermack and A. G. McKendrick in 1926 [2]. It is shown in equation (1).

$$\begin{aligned} S'(t) &= -\gamma \cdot S(t) \cdot I(t) \\ I'(t) &= \gamma \cdot S(t) \cdot I(t) - \delta \cdot I(t) \\ R'(t) &= \delta \cdot I(t) \end{aligned} \tag{1}$$

If every individual has C contacts in average, in the whole system $\frac{C \cdot N}{2}$ pairwise contacts happen. Division by 2 is necessary because a pairwise contact always affects two individuals. The probability that a random pairwise contact happens between an infected and a susceptible

individual is $\frac{I(t)}{N}\frac{S(t)}{N-1}+\frac{S(t)}{N}\frac{I(t)}{N-1}=2\frac{S(t)\cdot I(t)}{N\cdot(N-1)}$. Hence, the expected number of contacts between susceptible and infected individuals in one time unit is $\frac{C\cdot N}{2}\cdot 2\frac{S(t)\cdot I(t)}{N\cdot(N-1)}=\frac{C}{N-1}S(t)I(t)$, and $\alpha\frac{C}{N-1}S(t)I(t)$ susceptible individuals expectedly get infected. Thus, we can identify γ with $\alpha\frac{C}{N-1}$. δ is simply the fraction of infected individuals that recover and thus equals β.

Table 2. The parameters of the ODE model

| Parameter | Identification |
|---|---|
| S(0) | S0 |
| I(0) | I0 |
| R(0) | R0 |
| γ | $\alpha\frac{C}{N-1}$ |
| δ | β |

## 4 Modelling with a Lattice Gas Cellular Automaton

Cellular Automata (CA) are a time- and space-discrete modelling approach. A CA consists of cells which are placed on a regular grid and can hold different states [3]. Lattice Gas Cellular Automata (LGCA) are further developed CA where particles move around these cells [4]. With a hexagonal grid it can be used to simulate the movements of gas particles or fluids. Another possible extension is to let the particles be in different states and thus simulate the spread of an infectious disease [1] [5]. This technique will be used here.

The LGCA consist of cells which are placed on a 2-dimensional hexagonal grid and can hold at most six individuals. Individuals are in one of the three states susceptible, infected and resistant. Contacts happen pairwise between all individuals which are in the same cell in a discrete time step. To simulate a mixture of the individuals, they move around the cells as defined by the FHP-I rules [4]: The position of the individuals in the cells, define the direction in which the individuals move (figure 1). After this movement phase there is the collision phase (figure 2). The FHP-I variant only allows two and three individual collisions. When two individuals collide as in image 2, they are reflected clockwise or counterclockwise with probability 0.5. When three particles collide then they are reflected clockwise.

When a susceptible individual meets an infected individual, it becomes infected with probability $\alpha_c$ at the end of the time unit. An infected individual recovers with probability $\beta_c$.

The size of the LGCA plays an important role because it affects the density of particles and thus the number of contacts. To keep it simple, only square LGCAs with width=length=n are used here. Hence, such an LGCA consists of n² cells with six places each. Table 3 shows the parameters of the model with appropriate parameterization. Having a given number of individuals, the number of contacts depends on the size of the LGCA.

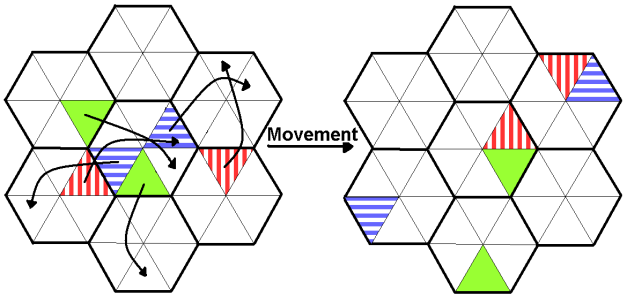



Figure 1: Movement of particles in the FHP-I model

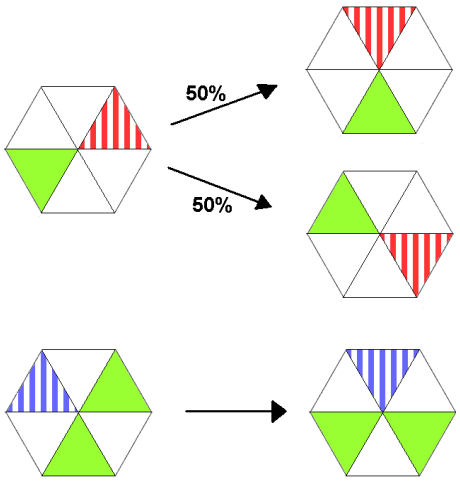



Figure 2: Collisions of particles in an FHP-I model

The correct identification for n is crucial. It follows a simple calculation: Assuming equally distributed individuals, each place in a cell is occupied with the same probability. A given individual is in a cell with 5 other places, thus (N-1) individuals and (6n ²-1) places are remaining, hence the individual has an average of

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

contacts within this cell. Adjusting n, which has to be an integer, to meet a given number of contacts leads to the term in table 3.

Table 3. The parameters of the LGCA model

| Parameter | Identification |
|---|---|
| S(0) | S0 |
| I(0) | I0 |
| R(0) | R0 |
| αc | α |
| βc | β |
| n | $\left\lceil\sqrt{\frac{5\cdot(N-1)+C}{6\cdot C}}\right\rceil$ |

## 5 Analytical comparison

**Infections**

There is a strong analytical relation between both models. 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 because only susceptible individuals can get infected. Then there are $(6n^2 - 1)$ remaining places in the LGCA, 5 remaining places in the cell and $I$ infected individuals. Define the probability of $i$ places in the cell being taken by infected individuals as $q_i$. Under the assumption that the individuals are uniformly distributed, the number of infected individuals in this cell follows a hypergeometric distribution. It calculates by choosing $i$ out of $I$ infected individuals on 5 out of $(6n^2 - 1)$ places:

$$q_i = \frac{\binom{5}{i}\binom{(6n^2-1)-5}{I-i}}{\binom{6n^2-1}{I}}, i = 0 \dots 5 \quad (3)$$

The expected value $E$ of this hypergeometric distribution is

$$E = \sum_{i=0}^{5} q_i i = I\frac{5}{6n^2-1} \quad (4)$$

Using the shown identity (2), the expected value can be rewritten as

$$E = I\frac{C}{N-1} \quad (5)$$

With these preparations the actual infection probability of the 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 Taylor series expansion for $\alpha = 0$, we get an approximation for this probability. This approximation can be rewritten using the identity (5).

$$\sum_{i=0}^{5} q_i(1-(1-\alpha)^i) \approx \sum_{i=0}^{5} q_i i\alpha = \alpha\sum_{i=0}^{5} q_i i = \alpha E = \alpha I\frac{C}{N-1} \quad (6)$$

Multiplying this with the total amount 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.

It turns out that this factor is already used in the differential equation (1) as $\gamma \cdot S(t) \cdot I(t)$, where $\gamma$ is identified with $\alpha\frac{C}{N-1}$.

**Recoveries**

One infected individual in the LGCA recovers with probability β, hence the expected amount of infected individuals who regenerate in one time unit is $\beta I$. This factor also occurs in the differential equation.

## 6 Results

In the part of the analytical comparison, some implicit and explicit assumptions have been made to analytically compare both models. One can ask whether these assumptions are sufficient to justify these analyses or under which circumstances the analyses are not valid. We will investigate these questions in the following section by comparing both methods on an experimental level. Looking at figure 3 to observe the difference under special parameter settings, we find out that they are partly small and partly enormous.

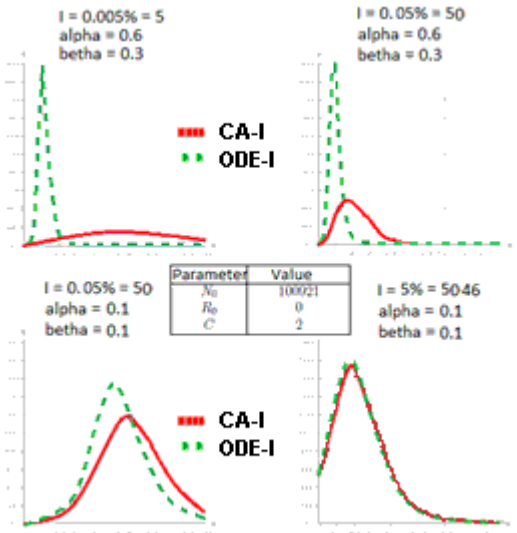



Figure 3: Comparison of the ODE model and the LGCA

There are two 'meta' reasons and the two more relevant structural reasons contributing to this difference.

**Meta contributions:**

A) One has to be aware that the ODE System is by definition a continuous model and the LGCA model is discrete. The continuity in the number of persons seems to be no problem, but the continuity in time is one. In figure 4 one can observe that the discretized ODE-model (explicit Euler method) with time unit 1 has a significantly different quantitative behavior compared to the continuous one (Runge-Kutta-4/5).

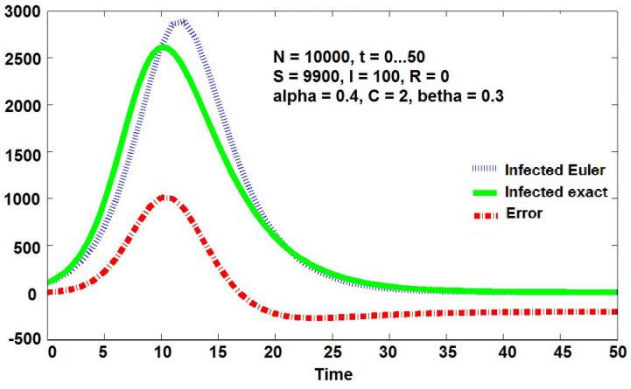


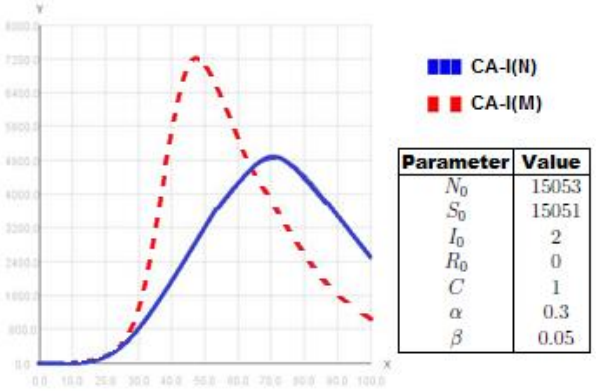

Figure 4: Descrete and continuous model.

B) One has to take into account that the LGCA model is probabilistic, so the results vary between simulations. For large populations the effect of probabilistic

variations is negligible. If the probabilistic behavior for small populations is not favored, one could take the average over a number of simulations.

**Structural contributions:**

C) The ODE-System satisfies the requirement of mixing the three groups of the population completely homogenous. In section 5 we assumed the same behavior for the LGCA, but this assumption is not always sufficiently fulfilled. In figure 5, a simulation of the above defined LGCA (CA-I(N)) and a modified LGCA (CA-I(M)) is performed with the parameters below. The modified LGCA redistributes the individuals in each time step uniformly on the grid. One can observe a big difference between both models. This is due to the small number of initially infected individuals. The normal LGCA is not able to spread the infected individuals fast enough, so after some time areas of high density of infected individuals occur in the LGCA. In the modified LGCA a homogenous mixing of the population is enforced so the infection spreads much faster. If the number of infected individuals initially is large enough this problem is almost not relevant because the LGCA is able to conserve the homogenous distribution.



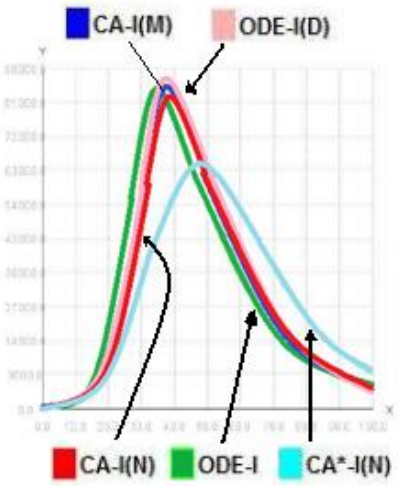

| Parameter | Value |
|---|---|
| $N_0$ | 15053 |
| $S_0$ | 15051 |
| $I_0$ | 2 |
| $R_0$ | 0 |
| $C$ | 1 |
| $\alpha$ | 0.3 |
| $\beta$ | 0.05 |

Figure 5: Normal :LGCA and mixing LGCA.

D) Another contribution to the difference is that the approximation in (6) is good for small values of the infection rate. Moreover, the value of both, the contact number C and the infection probability α is relevant in the LGCA whereas in the ODE-model only the product of the quantities count. The effect of this relation can be seen in figure 6. In curve CA-I(N) the parameters for α and C are α = 0.3 and C = 1 so that the product of them remains the same.

## 7 Discussion

Definition of an underlying, simplified system is a promising approach for comparison of two models. Here, it allows setup and parameterization of two structurally different models. Comparison is only possible for results, hence it is a crucial question whether the results are representing the same situation.




| Parameter | Value |
|---|---|
| $N_0$ | 158995 |
| $S_0$ | 102215 |
| $I_0$ | 159 |
| $R_0$ | 0 |
| $C$ | 4 |
| $\alpha$ | 0.075 |
| $\beta$ | 0.05 |

Figure 6 visualizes the different contributions to the 'error'. The green curve is the ODE-model and the red one is the LGCA-model. The pink curve is the discretisized ODE-model considered in A, the dark blue curve is the modified LGCA from C. The light blue curve is explained in D. For large populations, sufficiently large infected individuals initially and small infection probability the models are not only qualitatively but also quantitatively the same.

The task was comparing a time-continuous non-spatial model with a time-discrete spatial model for epidemic simulation. They are two incompatibilities by definition. However, the question is whether these approaches are able to simulate the same epidemics answering the same questions. This comparison is part of the validation process for epidemic models and helps to understand the impact of those diverging modelling methods.

The comparison shows that results of both models are similar in many cases. This means that the spatial structure in the LGCA and the homogeneous time in the ODE have a minor impact on the results for many settings. This compliance is also expectable by the analytical comparison. However, divergence is obvious whenever the approximation in equation (6) is bad. Hence, the approximation gives a hint for getting an idea when similar and diverging results can be expected.

## Literature


[1] Fukś, H., and A.T. Lawniczak. "Individual-based Lattice Model for Spatial Spread of Epidemics" no. 6. Discrete Dynamics in Nature and Society (2001): 191–200.

[2] Kermack, W. O., and A. G. McKendrick. "A Contribution to the Mathematical Theory of Epidemics." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 115, no. 772 (1927): 700–721.

[3] Wolfram, Stephen. A New Kind of Science. Champaign, IL: Wolfram Media, 2002.

[4] Wolf-Gladrow, Dieter A. Lattice-Gas Cellular Automata and Lattice Boltzmann Models: An Introduction. 1st ed. Lecture Notes in Math. 1725. Berlin: Springer, 2000.

[5] Yakowitz, S., J. Gani, and R. Hayes. "Cellular Automaton Modeling of Epidemics." Applied Mathematics and Computation 40, no. 1 (November 1990): 41–54.