

A Norm-based Comparison of Approaches in ARGESIM Benchmark C17R 'Modelling and Simulation of a SIR-type Epidemic with Cellular Automata and Ordinary Differential Equations' using MATLAB

Stephan Reichl^{1*}, Florian Miksch², Günter Schneckentreiter^{1,2}

¹Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria; *reichl.stephan@gmail.com

²dwh simulation services, Neustiftgasse 57-59, 1070 Vienna, Austria

SNE 27(1), 2017, 53-59, DOI: 10.11128/sne.27.bn17.10368
Received: May 10, 2016, Revised: December 20, 2017
Accepted: January 30, 2017
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna,
ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

Abstract. This model comparison quantifies the difference between the results of a population model, using ordinary differential equations, and an inhomogeneous spatial approach, using a lattice gas cellular automaton. Both approaches describe a classical SIR-epidemic. The system definition of the epidemic, model specification and parameter identification were already formulated by Miksch et al [1]. For the sake of quantifiable comparisons a norm was introduced. The tasks from the revised ARGESIM Comparison 17 [1] were completed and the results were analysed and visualised. Every result was interpreted according to the knowledge of the modelling approaches and the setup of the experiment at hand.

Introduction

This model comparison follows the outlined tasks of the revised ARGESIM Comparison 17 [1] and is targeted on quantifying the difference of SIR-type epidemic simulations, which were either modelled by classical ordinary differential equations (ODEs) or a lattice gas cellular automaton (LGCA). Both modelling approaches use the same system definition and therefore the same basis for parameter calculation. From now on we will use the term "population group" to describe any of the following cohorts: susceptible, infected or recovered individuals.

Although both models are strictly defined in the re-

vised Comparison 17[1] some changes or specifications had to be made:

- **Periodic boundary conditions** were used in the LGCA as follows: when an individual passes through any boundary of the lattice, it reappears on the opposite side of the lattice at the same position inside the cell and therefore with the same movement direction as before.
- The presented formula for the **calculation of the dimension** n of the LGCA's lattice yields a number, which is too small for certain populations where the contact rate C is chosen to be 5. This should be the maximum in an LGCA with hexagonal cells. The solution to this problem is to simply round up instead of down, as it is shown in the following.

$$n = \left\lceil \sqrt{\frac{5(N-1) + C}{6C}} \right\rceil \quad (1)$$

- The specification of **hard interventions with the LGCA model** was extended by the following step: Since the LGCA approach operates on discrete values the fraction of the affected population group had to be rounded down after the multiplication with the factor f_H .
- In the **ODE model** we were confronted with a **limitation** on the variation spectrum of the probability rates α and β . If we choose the value 1 for either of them then we get an numerically undefined term ($\log(0)$) when we try to carry out the parameter identification as instructed.

1 Implementation

The simulations were performed using a MacBook Air (11 inch, End 2010) running Mac OSX 10.11.6 and the Software MATLAB R2015b. It should be pointed out that the results of the LGCA vary in every iteration, because of the usage of probability principles. The realization of the ODE model always yields exactly the same results (provided that the same ODE-solver is used). To ensure exact reproducibility a large number of iterations with the LGCA model would have been necessary. Then the average values of the infected individuals at every time step would have been used in the comparisons, but this would have gone beyond the scope of this paper. Here we used one round of simulations, after a lot of test runs, to generate the results.

1.1 Differential Equations Model

The system of differential equations was solved using the MATLAB ODE-solver `ode45`, which is used to solve systems of nonstiff differential equations. Furthermore we only calculated the solutions for integer time steps starting with 1, to ensure comparability between the continuous ODE-model and the discrete LGCA-model. The solution contained the number of individuals for each population group at every integer time step.

1.2 Cellular Automaton

The model was implemented as described in the revised comparison [1] with one exception, which was previously explained resulting in equation 1 for the dimension of the lattice.

The initial conditions, represented by the individuals of every population group at time $t = 0$, were uniformly distributed in the cells of the lattice, which was implemented as a matrix $M \in \mathbb{N}^{n \times n \times 6}$. This was realized by randomly permuting a vector, with the length equal to the number of free cells ($6n^2$), filled with the integers from 1 to $6n^2$. Then it was possible to take the first x values of that vector as indexes for the individuals of one population group, where x is its initial condition.

After initializing the LGCA the movement phase is implemented according to the rules, which were defined in the revised comparison [1] and extended by the boundary condition presented in the Introduction.

Then the collision phase is realized by using the FHP-I collision rules. Followed by that, the infection- and recovery-phase is implemented. In this phase we simulate the contacts between individuals in the same hexagonal-cell. With every contact between an infected individual and a susceptible one, there is a certain possibility of an infection. Additionally for an infected individual there is the possibility to recover in every single time step. In order to assert synchronicity, the infection and recovery takes place in a buffered copy of the whole lattice.

Functions of probability are used twice in the implementation: when we determine the distribution of the initial conditions on the lattice (this happens only once per simulation) and every time when there is the possibility of an infection or recovery of an individual.

At the end of every iteration/time step, we sum up all the population groups and save them in association with the according point in time.

1.3 Results

The result of an experiment always consists of a difference-determining value, which will be introduced in the next section, and five plots: each modelling approach with the corresponding SIR-curves, one which showed the infected curves of both simulations and the 1-Norm and 2-Norm of the difference in infected individuals per time step. For every experiment or variation of parameters a MATLAB-file was generated and the simulations took place with the help of a script which loaded the parameter file and generated the result document automatically for every experiment. As a result of this procedure it was ensured that in every experiment both approaches used the same base-parameters.

2 Comparison Norm

It is always useful to have one value which determines the difference between two functions. We decided that the curve of infected individuals best describes an epidemic over the course of time. For further analysis and quantifiable comparability we introduced a value to determine the exact difference between the two curves, because it often happens that the curves are very similar but offset by a certain time.

We came to the conclusion that the best way to quantify and ensure comparability between experiments is to determine the area of the absolute difference between

both curves of infected individuals divided by the total population (N). This can formally be written as

$$\frac{\int_0^T |y(t)_{ode} - y(t)_{ca}| dt}{N} \quad (2)$$

Since the comparison takes place under discrete circumstances with equally spaced time steps we have to simplify equation 2 with the help of the trapezoid rule to

$$\frac{\sum_{t=0}^T (|y(t)_{ode} - y(t)_{ca}| + |y(t+1)_{ode} - y(t+1)_{ca}|)}{2N} \quad (3)$$

In MATLAB this was implemented by calculating the absolute difference at every discrete point in time (1-Norm) and the use of the function `trapz()`.

3 Parameters & Figures

The parameters of Table 1 were used to conduct different experiments by varying only one at a time.

S_0	I_0	C	α	β	T_{end}
9500	500	4	0.1	0.1	100

Table 1: Default configuration of the available parameters.

The population group Recovered was always initialized with zero individuals ($R_0 = 0$) and the total population (N) was set constant with 10000 individuals. With this data pool as setup, both model approaches were simulated and compared. The results are different plots and a value for the comparison of the difference between the two approaches in different experiments, as described before. In Figure 1, 2 and 3 we present the different kinds of plots that were used for every experiment with the parameters of Table 1. The calculation for the comparison norm for these parameters results in

$$d(ODE, LGCA) = 1.4844 \quad (4)$$

In further consequence we will not display these three plots for every parameter configuration because of lack of room, readability and comparability. Therefore bar charts will be used with the help of the comparison norm, which was introduced earlier in equation 3.

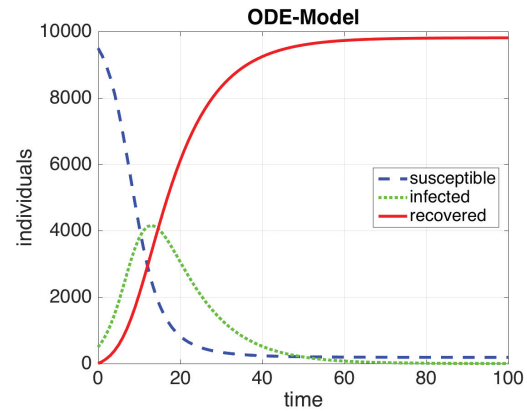


Figure 1: The epidemic over time with the ODE-model.

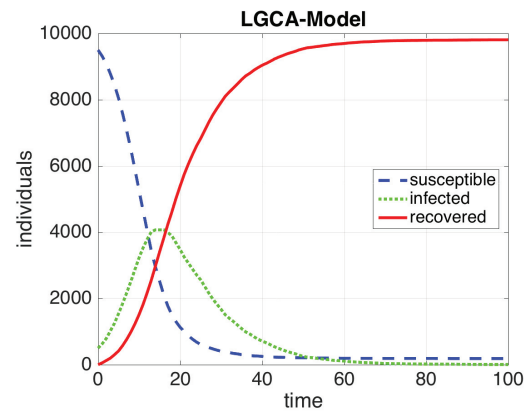


Figure 2: The epidemic over time with the LGCA-model.

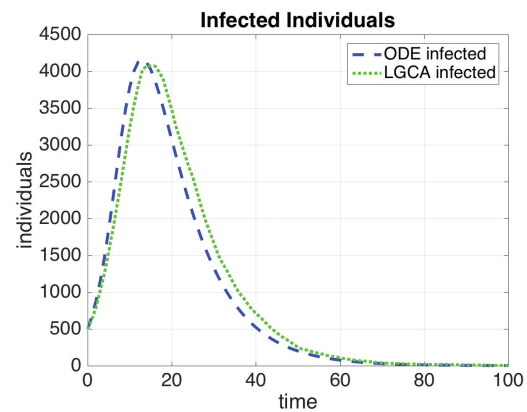


Figure 3: The Infected individuals over time for each model.

In the following we will always present the used parameters, which deviate from the default values from Table 1, as tables and the results as bar plots.

4 Task 1) Parameter Variation

In the first part of task 1 simulations with different variations of one parameter at a time were performed. The values of the following parameters were varied: initial number of infected individuals I_0 (and therefore S_0), the average of contacts per time step C , the infection probability α and the recovery probability β . The different variations of every parameter are displayed in Table 2, 3, 4 and 5 respectively. In the second part only the contact rate C was varied while the product of $C\alpha$ had to remain constant as it can be seen in Table 6.

4.1 Task 1) a) - Variation of I_0

	1	2	3	4	5	6	7
I_0	5	50	250	500	750	1000	5000

Table 2: Parameter variation of I_0 .

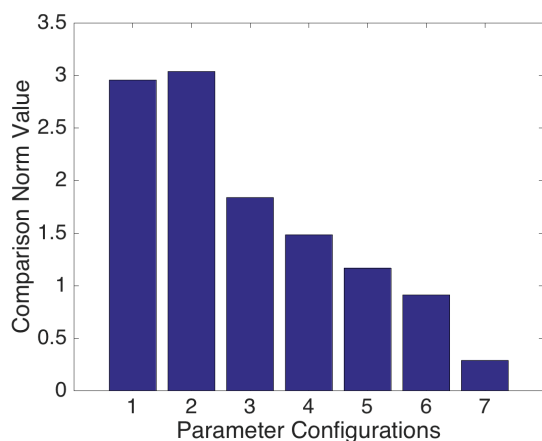


Figure 4: Results of the variation of I_0 .

4.2 Task 1) a) - Variation of C

C	1	2	3	4	5
-----	---	---	---	---	---

Table 3: Parameter variation of C .

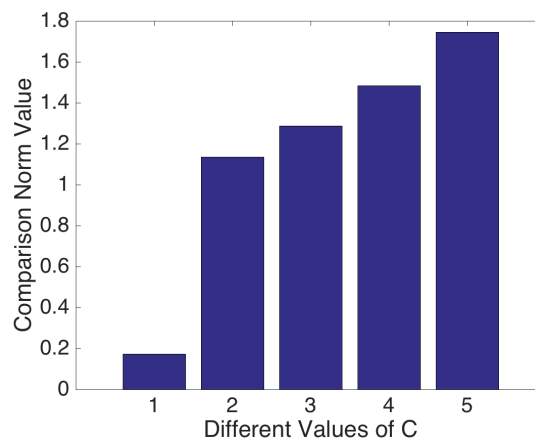


Figure 5: Results of the variation of C .

4.3 Task 1) a) - Variation of α

	1	2	3	4	5	6	7
$\alpha \times 10^2$	1	5	7.5	10	20	50	99.99

Table 4: Parameter variation of α .

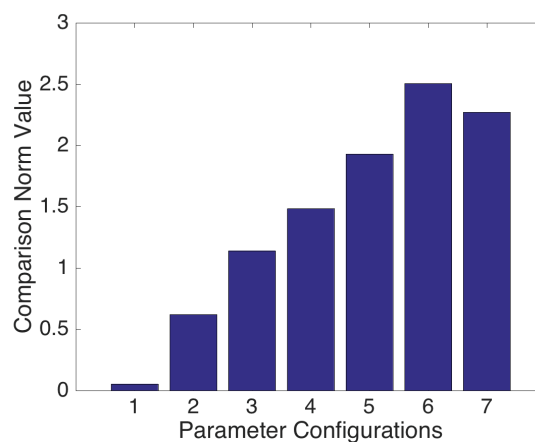


Figure 6: Results of the variation of α .

4.4 Task 1) a) - Variation of β

	1	2	3	4	5	6	7
β	0.001	0.01	0.05	0.1	0.2	0.25	0.5

Table 5: Parameter variation of β .

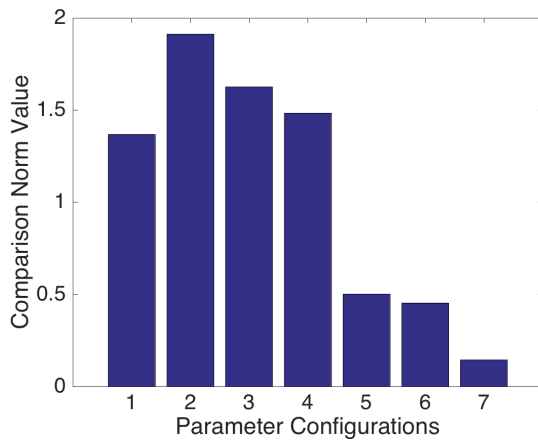


Figure 7: Results of the variation of β .

4.5 Task 1) a) - Interpretation

Looking at the results, the following hypotheses can be formulated

- We observe in Figure 4 that more infected individuals at the start result in smaller differences.
- Figure 5 indicates that the more contacts happen, the greater the difference in the results get.
- We can see in Figure 6 that greater infection probability results in larger norm values.
- The opposite seems to be true for the recovery probability as illustrated in Figure 7. Greater recovery probability results in smaller norm values.

The first and last point can probably be explained by the fact that the configurations, which yield small norm values, compensate for the spatial component in the LGCA model. Furthermore the recovery phase does not take into account where a individual is positioned and therefore the spatial component of the LGCA has no impact on it. The second and third hypothesis may arise from the fact that the movement of individuals is modelled differently in the two approaches.

4.6 Task 1) b) - Variation of C with $C\alpha = const$

	1	2	3	4	5
C	1	2	3	4	5
α	0.4	0.2	0.1333	0.1	0.08

Table 6: Parameter variation of C and α with $C\alpha = const$.

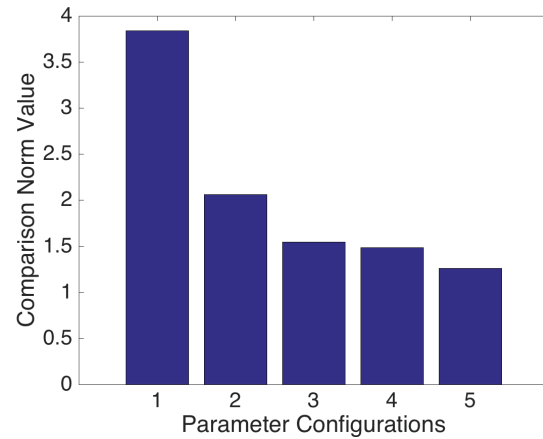


Figure 8: Results of the variation of C and α with $C\alpha = const$.

4.7 Task 1) b) - Interpretation

In Figure 8 it can be observed that for $C \geq 2$ the norm values drop below 2 but never 1. We want to point out that if we compare Figure 5 and Figure 8 and the tables, respectively we have to come to the conclusion that the adapted infection probability α is not causing any improvement compared to the situation with constant α . This should not be surprising as we have already seen in Figure 6 that the norm increases if α is increased.

5 Task 2) Interventions

The interventions were implemented with 60 time steps, because they cause an earlier end of an epidemic. In the **Soft Interventions** we modelled the decrease (within $\Delta t = 10$) of the intervened parameter (α) with a smooth step, instead of a linear one, because it is more realistic and the difference was smaller than 0.1. As **Hard Intervention** strategy we chose to vaccinate susceptible individuals ($f_H S$). Both interventions start when a certain threshold ($f_T N = I$) is reached. We used the

default configuration from Table 1 and only varied the intervention specific parameters as stated in Table 7 and 8.

5.1 Task 2) a) - Soft Intervention

	1	2	3	4	5
f_T	0.1	0.2	0.3	0.2	0.2
f_S	0.5	0.5	0.5	0.1	0.25

Table 7: Soft Intervention parameter variation of f_T and f_S .

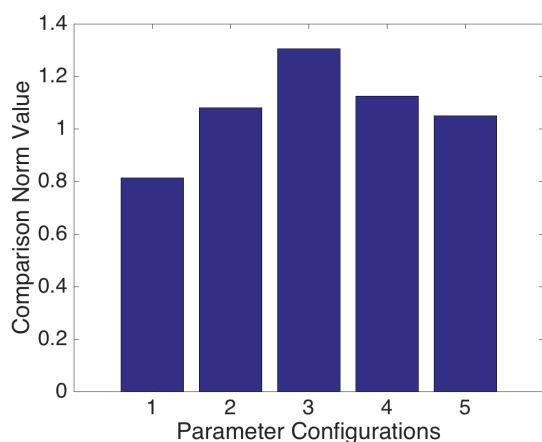


Figure 9: Soft Intervention parameter variation of f_T and f_S .

5.2 Task 2) b) - Hard Intervention

	1	2	3	4	5
f_T	0.1	0.2	0.3	0.2	0.2
f_H	0.1	0.1	0.1	0.25	0.5

Table 8: Hard Intervention parameter variation of f_T and f_H .

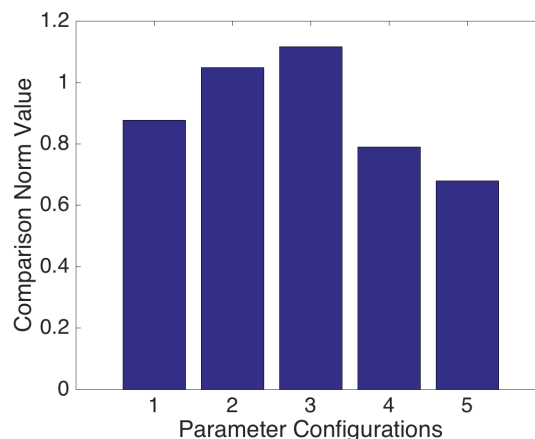


Figure 10: Hard Intervention parameter variation of f_T and f_H .

5.3 Task 2) Interpretation

Looking at Figure 9 and Figure 10 we can see that the norm values of all the experiments with both kinds of interventions always lie between 0.6 and 1.4. Therefore it is safe to say that by varying the intervention parameters no parameter regions could be detected where the models do not behave very similar. Often the difference between the results was caused by the lacking behind of the LGCA approach because after reaching the defined threshold ($f_T N = I$), which triggers the intervention, actions took place in the next time step and not at that moment, as it is realised in the ODE model.

6 Task 3) Spatial Inhomogeneity

We replaced the implemented movement rules of the LGCA with random movements to achieve a constant homogeneous mixture of the population. After every time step the individuals were randomly distributed on the lattice. We chose four bad and four good results of previous experiments and their parameters (as shown in Table 9), to see if the new rule changes the outcome. The bold values in table 9 indicate deviations from Table 1.

	S_0	I_0	C	α	β	T_{end}
1	9500	500	4	0.9999	0.1	100
2	9500	500	1	0.4	0.1	100
3	9950	50	4	0.1	0.1	100
4	9500	500	2	0.2	0.1	100
5	9500	500	1	0.1	0.1	100
6	9500	500	4	0.01	0.1	100
7	9500	500	4	0.1	0.5	100
8	5000	5000	4	0.1	0.1	100

Table 9: Parameters for Spatial Inhomogeneity Comparison.

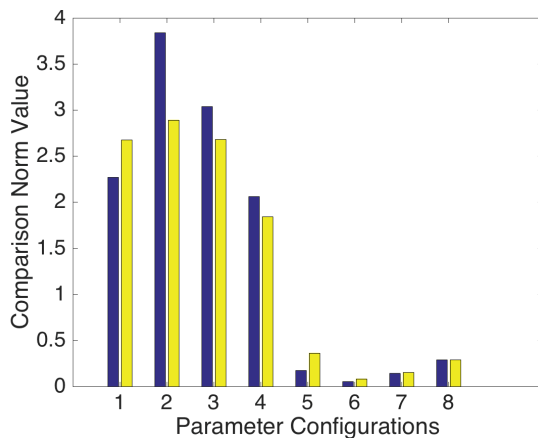


Figure 11: Results with (blue) and without (yellow) Spatial Inhomogeneity.

6.1 Task 3) - Interpretation

As we can see, in Figure 11, nearly every scenario produced better results with the random movement. Only configuration 1 and 5 experienced a worsening in the result. These are probably outlier. In the results of configuration 6,7 and 8, where the results were already good, no significant change was noticed.

References

[1] Miksch F, Haim C, Schneckenreither G, Popper N. Modelling and Simulation of a SIR-type Epidemic with Cellular Automata and Ordinary Differential Equations - Definition ARGESIM Benchmark C17R. *Simulation Notes Europe*. 2015;25(1):49–54.