A HYBRID MODELLING APPROACH FOR INFLUENZA EPIDEMICS BASED ON CELLULAR AUTOMATA AND AGENT BASED TECHNIQUES

Štefan Emrich[1], Felix Breitenecker[1], Günther Zauner[2], Nikolas Popper[2]
Michael Gyimesi[3]

[1] Institute for Analysis and Scientific Computing,
Vienna University of Technology,
Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria
[2] 'die Drahtwarenhandlung – Simulation Services',
Neustiftgasse 57-59, 1070 Vienna, Austria
[3] Hauptverband der Öst. Sozialversicherungsträger, Abt Gesundheitsökonomie,
Kundmanngasse, 1030 Vienna, Austria

stefan.emrich@gmx.net (*Štefan Emrich*)

Abstract

The classical tools for modeling epidemics are ODEs and PDEs, but they have some shortcomings. It is extremely complex to expand the model onto heterogeneous populations (age-groups, sex, infection risk, etc.) or to add time dependend effects (e.g. daily routines or vaccination after infection). Thanks to electronic data processing there is also much more data available nowadays than 100 or even 30 years ago, making it possible to create new kinds of models. Cellular automata (CA) and agent-based (AB) computing allow a different approach of the problem. CA by definition introduce spatial structures to a model. AB systems even take it a step further, they are based on agents that may hold various characteristics (age, sex, membership of a certain family or neighborhood, etc.). Because of their distinct strengths and drawbacks a combination of both methods seems promising and therefore was implemented for the simulation of influenza outbreaks in urban populations. An AB framework was created allowing the population to be heterogeneous, within it CA are introduced for timeconsuming computation of small units such as schools or neighborhoods. The framework allows it to steer the daily routine of the population and control its movement from one unit to another. The results of the model are fairly interesting as they show a behavior similar to real influenza patterns.
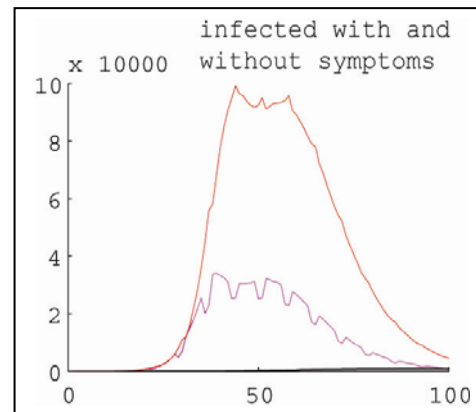
Keywords: Epidemic Spread, SIR, SIRS, Agent-based modeling, AnyLogic.

Presenting Author's biography

Štefan Emrich was born on June 30th 1981 in Vienna, Austria. After high school he went on to study technical mathematics at the Vienna University of Technology, where he graduated in June 2007 at the Institute for Analysis and Scientific Computing. Currently he is continuing his studies towards a PhD.

1   Development of Epidemic Modeling

The wish to understand and forecast the spread of diseases (among the human population) is a very old one. It took a long time before the mathematic methods, as well as the bio-chemical understanding of the processes involved was developed far enough to be able to (at least) partially fulfill this wish. Classically methods based on systems of differential equations, ordinary differential equations (ODEs) and partial differential equations (PDEs), are used for this purpose (see [1]).

The problem of these approaches is that they are based on homogeneous populations – in terms of individuals as well as in terms of their spatial distribution. If it would be possible to create models which pay respect to heterogeneities, more realistic simulations would become possible.

In the last 60 years two interesting techniques that do take into account spatial distributions and/or heterogeneous populations have been developed. The first one being cellular automata (CA) – in the middle of the last century – and the second one being agent-based (AB) systems also referred to as multi agent systems (MAS), during the 1990ies. In the following we will take a brief look at both techniques and compare them to localize their weaknesses and strengths.

2   CA vs. MAS

2.1   Cellular Automata

Cellular automata and multi-agent systems are both classified as "bottom-up" approach. Both methods describe (complex) systems only by defining local interaction rules. Theses interaction rules within CA are very strict and usually quite simple. Generally classical cellular automata can be described by four points (adopted from [2]):

- *Cell geometry* – CA consist of equal (geometrical) cells which are arranged in a (regular) lattice.

- *Neighborhood definition* – A cell neighborhood is defined for the automaton. The neighborhood determines which cells are influencing each other. This neighborhood is valid for all cells during the whole runtime.

- *Cell states* – A finite number of (discrete) states is defined. Every cell can assume one of those states. The state of the cell is subject to change. The transition between states is determined by the transition rules. The states of the cells are updated synchronously (after discrete time steps) for all cells.

- *Transition rules* – These rules describe how the cells change its states. The rules only depend on the state of the neighboring cells and on the state of the cell itself. One set of transition rules is valid for the whole CA over the whole runtime.

2.2   Agent-based Systems

A comparably strict and exact definition for agent-based systems does not exist. Depending on the author many possible definitions can be found, but AB modeling is lacking a common standard, although this is not a big problem.

One can define multi agent system by starting with the agent itself, as being a computer system situated in an environment. It further has the capabilities to flexibly and autonomously act in this environment in order to reach (predefined) objectives/goals. This requires specifying the following three terms more detailed:

- *Situated* in our case means that the agent is interacting with its environment, it is capable to receive input from the surrounding (e.g. via sensors) and can also manipulate its environment to some extent.

- The definition of *autonomy* needs to be handled with care (we are talking about a pre-programmed computer system). It is satisfactory if the agent can reach decisions without (human) interaction.

- *Flexibility* is required in multiple ways. Firstly one demands that the system is operating and acting in reasonable time. Secondly the agents are not to be solely reactive but goal-oriented or in the best case anticipating. And thirdly agents may have the capability to communicate or interact with other agents and/or real humans.

Multi agent systems consist – as the name indicates – of a number of such agents. Further we want MAS to have following characteristics:

- agents have a limited point of view (incomplete information and/or problem solving capabilities),

- the absence of global system control,

- decentralized data and

- asynchronous computation of the agents.

This definition is adopted from [3]. And with these two definitions we can now start a comparison of the methods.

2.3 Comparison

The smallest unit of CA, the cell, is by definition fixed and may hold discrete states. Whereas agents in MAS are very flexible and may hold several characteristic features (e.g. height, color, diameter, health status, speed, pressure, e.g.).

Time is being processed in CA in a very discrete way. All cell states are updated simultaneously. Between these updates nothing changes. AB systems may use either discrete or virtually continuous time. Changes may occur at any given time.

The spatial structure of cellular automata is defined by its lattice and usually symmetrical (see [4] for an exception). Agent-based systems on the other hand do not require any (spatial) structure. Agents may or may not be spatially bound, and any desired structure can be introduced.

The cells of CA do have a fixed range of influence – the neighborhood. Whereas agents within MAS have a very variable sphere of influence which does not need to be connected to spatial neighborhoods.

Concluding we can generally say, that AB systems are more flexible than cellular automata. One could even say that agent-based systems are an extension of CA (see [5]). On the other hand this extension and freedom when using MAS comes at a price – a trade off becomes necessary. The hardware demands of AB systems are considerably higher than those of CA. Thus the time consumed by simulation of large models increases when using MAS, whereas CA are computationally extremely efficient and can also take advantage of parallel computing which is apparently not possible for MAS. On the other hand it is possible to introduce interesting behavior to CA models by loosening the tight rules (see [6]).

This is making a combination of MAS and CA interesting in which one could take advantage of the flexibility of AB modeling and also achieve computational efficiency comparable to those of CA models.

3 Considerations and Model Setup

In order to simulate within a realistic heterogeneous population one needs to populate the model with diverse agents and take their social interactions into account. To start with we can divide the daily routine into three parts, beginning with the work / educational part. The second part being spare time, in which an individual goes shopping, visits friends, relaxes, etc. and finally time spent at home. This is of course a very simplified daily routine, but sufficient to start with.

Further one needs to identify the demographic structure of the population. Based on the number of people per age group and sizes of households, schools and workplaces one can start to set up the model.

For simulation the agents are initialized according to the given demographic structure and then assigned to households and workplaces / schools. Then the daily routine is applied repeatedly to the whole population (see Fig. 1). The temporal partition is visualized in Fig. 2.
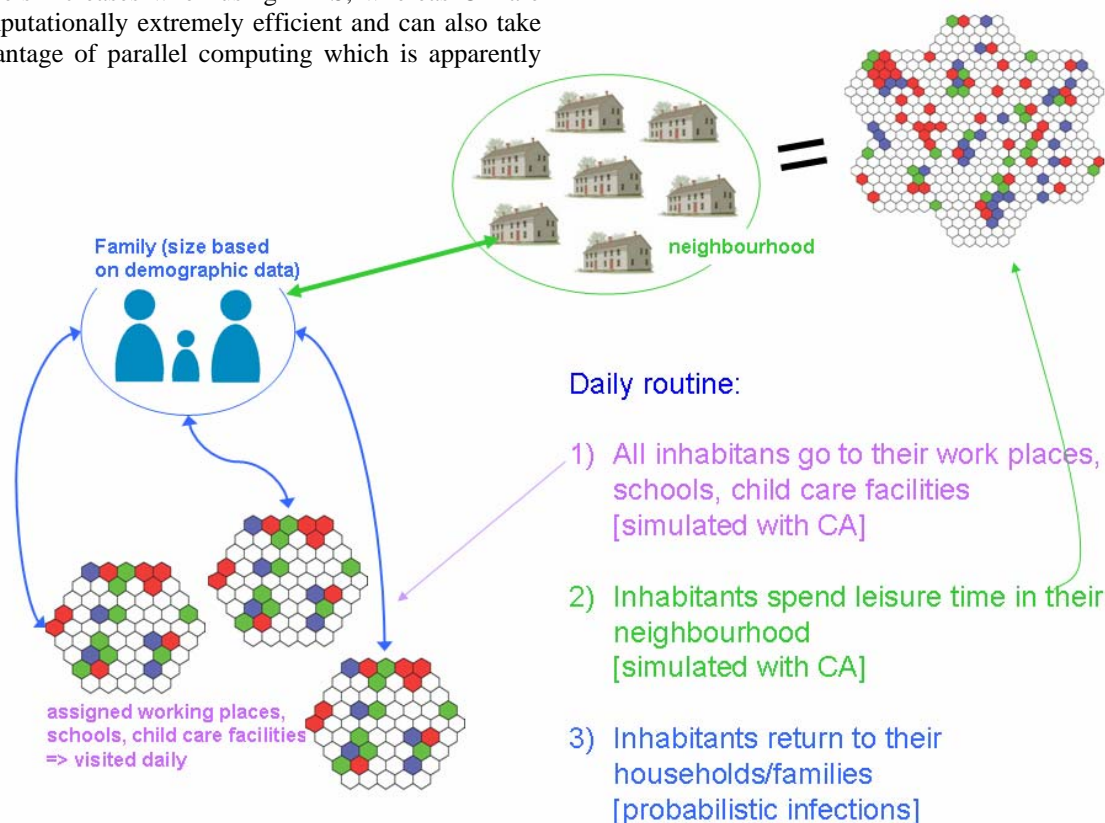

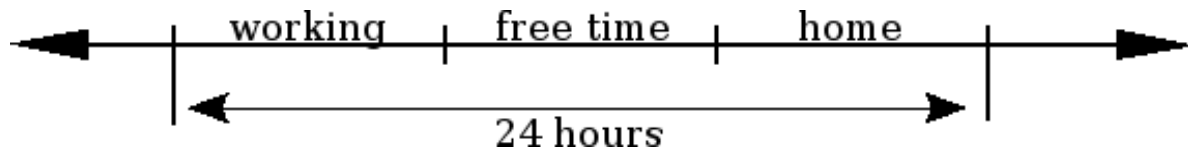Fig.1 Scheme of used model structure

Fig.2 Temporal partition of agent's whereabout

The model was implemented in MATLAB. The population is realized as agents. All agents have ten assigned variables (see Table 1 for a description). These variables are stored in an agent look-up matrix. This is very convenient and efficient.

Since MATLAB is optimized for matrix operations all necessary read/write operations are executed fast. In addition this way of implementing the agents allows for easy enlargement in case that another variable/characteristic is to be added to the agent or system. This lookup table is stored as a global variable and thus accessible to all subroutines of the simulation program.

Table 1 Description of "agent look-up matrix"

Row	Attribute
1	Key – unique ID of agent
2	Health status
3	Age of agent
4	Household ID
5	Neighborhood ID
6	Workplace ID (age specific)
7	Time stamp
8	Infection probability
9	Control variable (if symptomatic)
10	Help variable

In order to achieve good runtime the work places, schools and neighborhood (spare time environments) are implemented as lattice-gas cellular automata (LGCA) of type FHP-I (see [2] for further information). Within those the agents move and are at risk of infection at contact with infected agents with respective probabilities. In households the number of agents is very small and contact can be taken as granted. Therefore households are modeled by simple probabilistic routines.

These cellular automata are initialized by the main routine of the model one after another. The first workplace, school or child care facility is being called and the cellular automata filled with all respective workers / children / pupil. Within the CA the agents loose all their properties except for their unique ID-key. This key becomes the content of particle within the LGCA. If any infections/deaths happen during the runtime of the CA the change is written directly into the affected agent's column of the look-up table.

This procedure is repeated for all workplaces/schools/child care facilities. After these are finished the spare-time neighborhoods are processed in the same way. This implementation has several advantages. First of all it is necessary to program only one routine for cellular automata. The same one is used for every unit of the model, the dimensions are simply adjusted to the number of agents within and to the desired density of the CA-population. Such it would be possible to assign different densities (space per individual) to different "work places" (e.g. workplaces and schools). Secondly all agents that are not inside the currently processed automata are not exposed to the models time. By this at the end of the day all agents did spend the same time within cellular automata even though they were most of the time just "parked".

Of course this model largely uses simplifications, for example the ways from one unit to another are not modeled. Different working times of agents are neglected as well as certain "hot spots" for the spread of diseases such as hospitals. The implementation of the neighborhoods is also a simplification. All agents are put into their neighborhood every day with the same other agents. Although this can be interpreted as pool of personal contacts (starting from friends, over neighbors up to doctors and lawyers) that one meets.

4 Results

One of the most difficult parts in producing good models is the proper fitting. In this case this was especially difficult, since to the knowledge of the author no data on disease transmission of influence exists. Even if such data would exist it would still be (almost) impossible to use it for the model. This is explained by the fact that the influenza virus is a constantly mutating virus and every mutation has different infectious behavior. Thus the illness attack rates used in the model, though based on various different sources, remain estimates of the author.

In addition real data would be necessary for proper parameterization of the model. This data exists but unfortunately was not yet available in the needed way. Still the model results are very interesting, as they show several effects that can be traced back to the spatially and demographically inhomogeneous population.
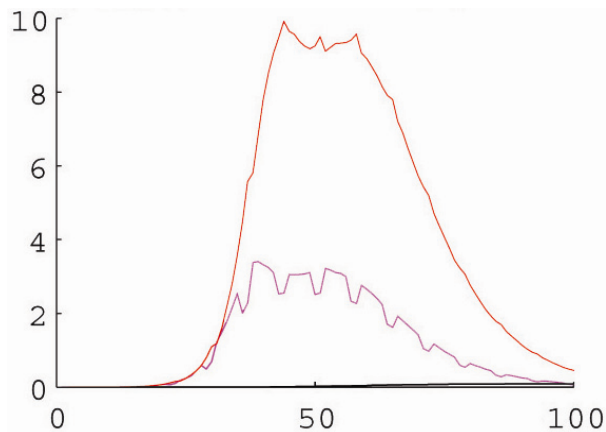
Fig.3 Numbers of infected individuals that are
not yet symptomatic (purple, lower curve)
and with symptoms (red, higher curve)

In Fig. 3 (y-axis: population in 10.000 individuals, x-axis: days of simulation) the result of an epidemic outbreak within a population of one million inhabitants is displayed (the graphic only shows the curves of infected individuals). The simulation was run over a period of 100 days, which corresponds to the observed usual duration of influenza epidemics. The jags of the lower curve are explained by the programmed weekend behavior – On weekends the population does not go to their workplaces respectively schools. The agents spend that time as "spare time" in their respective neighborhoods. If compared with the influenza curve for Germany of the season 2004/05 (see Fig.4) the similarity of the patterns becomes obvious.
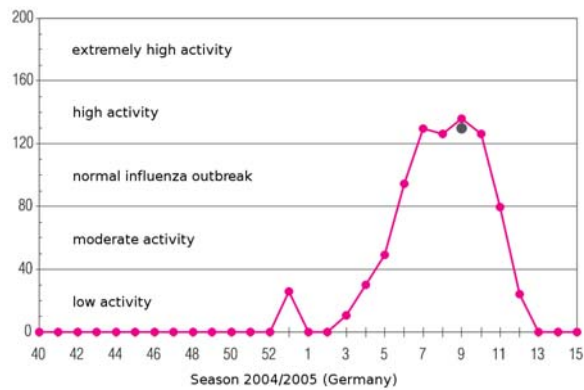

Fig.4 Influenza level for Germany (2004/05)

Unlike in classic ODE models the pattern described is not a bell shaped curve. The disease does reach a maximum level which is maintained for a few weeks. If one does take a look at the different age groups within the population more interesting things can be observed (see Figures 5 through 10). The figures do show the ratio of infected individuals (agents) within an age-pool at given time during the simulation (y-axis: ratio, x-axis: days of simulation).
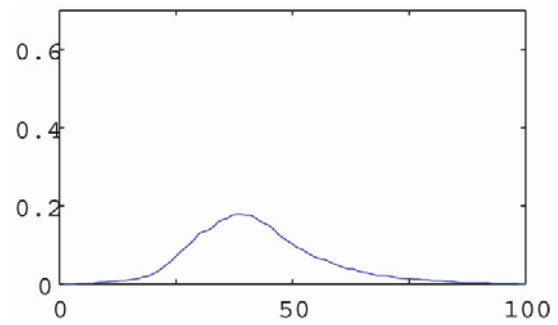

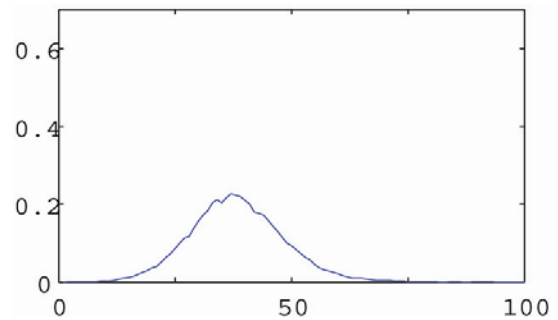Fig.5 Ratio of infected agents in sub-population of age group "infants"

Fig.6 Ratio of infected agents in sub-population of age group "kinder garden"
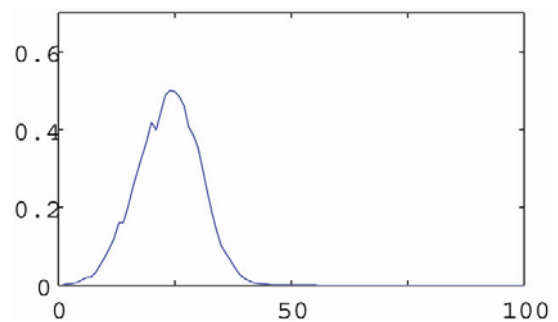

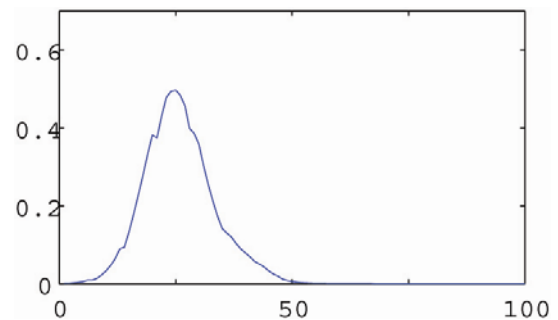Fig.7 Ratio of infected agents in sub-population of age group "elementary school"

Fig.8 Ratio of infected agents in sub-population of age group "high school"

Aside from the varying height of the peak, which is explained by the different illness attack rates (infection probabilities) the time when the peak is reached is quite interesting. It is obvious, that the age groups of elementary kids and high school children are much faster infected than the other age groups. The peaks of the respective graphs are reached after

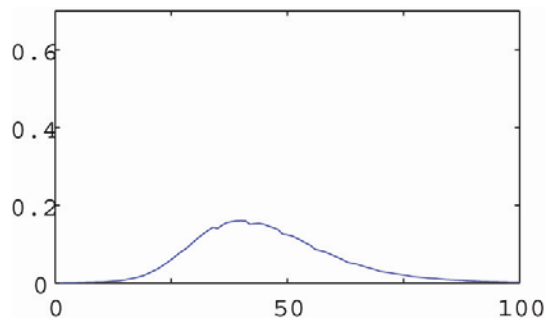about 25 days, whereas the rest of the sub-population reaches their peaks after 40 days of the simulation.

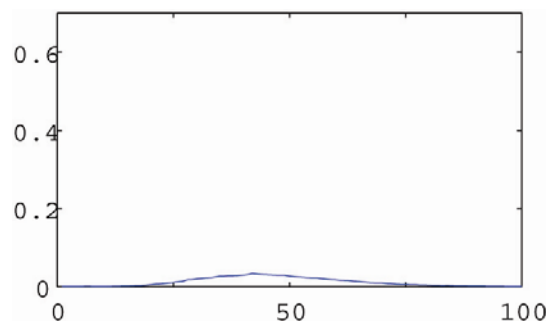Fig.9 Ratio of infected agents in sub-population of age group "adults"

Fig.10 Ratio of infected agents in sub-population of age group "senior citizens"

One possible interpretation is that children in schools are the perfect distributors for the influenza epidemic. The virus is spread in schools and brought home and into the neighborhoods where adults become infected. Infants and senior citizens, who stay at home in this model, are of course exposed to a lower risk of getting infected.

It would be extremely interesting to validate this kind of temporal correlations with real data. If the implications of the model are right it could be used to experiment with different strategies on how closure of schools / childcare facilities affects the spread of influenza epidemics. This can eventually lead to cost saving possibilities to reduce the effects of such an epidemic.

5   Conclusion

Judging by the first results produced with the hybrid model set up by combining cellular automata with techniques of agent based modeling the way seems right. It was possible to create a model for the simulation of epidemic spread within a heterogeneous population with spatial structure.

Although the model was not fitted according to real data, and thus the results can only be understood as being purely academic, the patterns observed seem very promising. If one takes into account the huge amount of possibilities for further refinement of the model (e.g. addition of hospitals, a better solution for infants and senior citizens, as well as taking into account jobless people, etc.) then the potential becomes quite astonishing.

It seems likely that good results can be achieved with this model structure given the proper data and further expansion of the model. As stated in section 3 such an enlargement is easy to implement with the agent based framework and modular set up of the model itself.

6   References

[1] Baylei Norman T. J. *The Mathematical Theory of Infectious Diseases and its Applications.* Charles Griffing & Company LTD, 1975.

[2] Wolf-Gladrow D. A. *Lattice-Gas Cellular Automata and Lattice Boltzmann Models.* Springer, 2000

[3] Jennings, Sycara, Wooldridge. *A Roadmap of Agent Research and Development (Autonomous Agents and Multi-Agent Systems).* Kluwer Academic Publishers. 1998

[4] Hegselmann, Flache. *Do Irregular Grids make a Difference? Relaxing the Spatial Regularity Assumption in Cellular Models of Social Dynamics.* JASS Journal of Artificial Societies and Social Simulation, 2001

[5] Lustick J. *Agentbased Modelling of Collective Identity: Testing Constructivist Theory.* JASS, 2000

[6] Roli, Zambonelli. *On the Emergence of Macro Spatial Structures in Dissipative Cellular Automata, and its Implications for Agentbased Distributed Computing.* http://www.sci.unich.it./aroli/pubs/dca.pdf