Expert Systems With Applications 48 (2016) 1–8

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Epidemiological modeling with a population density map-based cellular automata simulation system

A. Holko[a], M. Mędrek[b,a,*], Z. Pastuszak[b], K. Phusavat[c]

[a] The State School of Higher Education in Chelm, Pocztowa 54, 22–100 Chelm, Poland
[b] Maria Curie Sklodowska University, Plac Marii Curie-Sklodowskiej 5, 20–031 Lublin, Poland
[c] Department of Industrial Engineering, Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand

ARTICLE INFO

Keywords:
Cellular automata
SEIR-IBM model
Epidemic spread model
Disease spread model
Mathematical modeling
Square lattice

ABSTRACT

We present a new numerical, two-dimensional cellular automata framework for simulation of the spread of an infectious disease in a region with non-homogenous spatial population distribution. For the simulation the real map of the population density in Poland is used, where the sources of the infection are located. Presented model is a combination of SEIR and IBM models complemented with additional factors like variable population density, death, birth and some stochastic parameters to reflect the more realistic population dynamics. In proposed model the states of individuals are tracked through time like in IBM model and the evolution of the whole system is described by SEIR transition function which determines how cells interact with their neighbours, influencing global behavior of the system. Presented model requires less complicated input than IBM models and is less expensive computationally. We explore influenza as the contagious disease in our map-based simulation. The results of the simulation show the spreading-rate of the disease and can be used to describe possible actions for preventing pandemic.

1. Introduction

Epidemiology is an interdisciplinary field, in which many specializations like medicine, social sciences or economics are important components of the discussion. In our paper we tried to describe how epidemiology can be complemented with computational techniques combined with empirical data.

Traditionally, differential equation models have been used to describe the spreading of a contagious disease (Murray, 1993). Usually, an epidemic model belongs to one of the following types: *SIR, SIS, SEIR* or *SEIRS*, where particular letters denote one of the separate groups of the whole population (*S*usceptible, *I*nfective, *E*xposed and *R*ecovered), and the time evolution of these groups is modeled by the set of differential equations (Liu, Jin, & Liu, 2006; Milne, Fermanis, & Johnston, 2008; Pfeifer et al., 2008). However this mathematical approach has some serious drawbacks. It neglects external infections due to traveling individuals. It does not include variable susceptibility of individuals, and complex boundary and initial conditions (Achmed & Agiza, 1998).

Cellular automata (CA) can overcome these above drawbacks. CA is specified by lattice of cells which states can reflect initial and boundary conditions and the evolution in time of the CA can include the dynamics of the individuals. Initially CA was a model used to analyze the dynamics of micro-scale particle, such as thermodynamic rule, spin-glass model, etc. Wolfram showed that CA models are often based on rules that are simpler than complex mathematical equations and produce comparable results (Wolfram, 1983), therefore, CA may serve as a framework which can be applied for study complex natural phenomena in a way that is conceptually clearer and more realistic than conventional mathematical systems (Itami, 1994). Since then, CA has been also used by several researchers as an alternative method of modeling macro-scale phenomena, e.g. epidemics and building spatial and time discrete models of dynamical systems in which the contagious disease can spread (White & Sanchez, 2007; White, del Rey, & Sanchez, 2009). Alternatively to the above applications, CA can also be used to imitate the dynamics of similar, but not natural phenomena e.g. spreading of infectious diseases through internet: Peng, Wang, and Yu has used CA for building a model describing the propagation of worms in smartphones, where worms are self-replicating computer viruses, which can propagate through computer networks without any human intervention (Peng et al., 2013).

Most numerical models are based on the use of two-dimensional CA, where each cell stands for a square portion of the environment. Simulation region in two-dimensional CA consists of an array

* Corresponding author at: Maria Curie Sklodowska University, Plac Marii Curie-Sklodowskiej 5, 20-031 Lublin, Poland. Tel.: +48504136711.
E-mail addresses: arkadiusz.holko@gmail.com (A. Holko), marek.medrek@umcs.pl (M. Mędrek), z.pastuszak@umcs.lublin.pl (Z. Pastuszak), fengkkp@ku.ac.th (K. Phusavat).

http://dx.doi.org/10.1016/j.eswa.2015.08.018
0957-4174/

Nomenclature

a	Fixed length in days of the exposed state (E)
b	Fixed length in days of the infective state (I)
c	Number of columns
c_v	Variation coefficient of the infection probability
C	Cellular space
$d(ij)$	Index of the closest dense cell at the Moore distance greater than one
E	Number of exposed individuals which are infected but not yet infectious
$E^t_{ij\|d}$	Number of exposed individuals in the (i, j)–th cell at the time t in the d–th day of this stage
f	Local transition function
i, i_0	Cell's row index
I	Number of infective individuals which are capable of transmitting the disease
$I^t_{ij\|d}$	Number of infective individuals in the (i, j)–th cell at the time t in the d–th day of this stage
j, j_0	Cell's column index
n^t_{ij}	Vector of numbers of individuals who move out of the (i, j)–th cell at the time t
N^t_{ij}	Number of individuals in the (i, j)–th cell at the time t
$N^t_{ij\to xy}$	Number of individuals commuting from the cell (i, j) to the cell (x, y) at the time t
Q	Finite set of states
p^t_{ij}	Probability that a susceptible individual in the (i, j)–th cell at the time t will be infected
r	Number of rows
R	Number of recovered individuals which are permanently immune
R^t_{ij}	Number of recovered individuals in the (i, j)–th cell at the time t
R_0	Basic reproduction number
s^t_{ij}	Vector representing (i, j)-th cell's state at the time t
$s^t_{ij\to xy}$	Vector of individuals commuting from the cell (i, j) to the cell (x, y) at the time t
S	Number of susceptible individuals which are able to contract the disease
S^t_{ij}	Number of susceptible individuals in the (i, j)-th cell at the time t
t	Time in days
V	Function mapping the cell to its neighboring cells
V^*	Updated (considering cells without population) function mapping the cell to its neighboring cells
β	Contact rate between individuals in a cell
δ	$1/\delta$ is the mean latent period for the disease (in SEIR model)
γ	$1/\gamma$ is the mean infectious period (in SEIR model)
μ	Birth and death rate (in SEIR model)
μ_b	New births per one individual per time step
μ_d	Natural deaths per one individual per time step
μ_{vm}	Mortality rate per time step, i.e. probability of a death of an infected individual
ϕ_h	Fraction of a healthy population commuting outside of their cell
ϕ_s	Fraction of an infected population commuting outside of their cell
ϕ_c	Fraction of commuters commuting outside of their neighborhood. There is an assumption that they commute to the closest dense cell (a cell with population above 50, 000) at the Moore distance greater than one

(lattice) of identical cells, which are endowed with the state (*S, I, E, R*) that changes with discrete steps of time according to a rule called transition function (White et al., 2009). The evolution of the state of a particular cell in CA depends on the states of the neighboring cells and these local interactions determine the spreading of the disease in the system. Usually it is supposed that the distribution of the population is homogeneous which means that all cells in the lattice have the same population and the whole size of the population is constant (White et al., 2009). Moreover, commonly used models assume that individuals are not mobile and the epidemic spread occurs across individuals fixed in their positions (Moreno, Gmez, & Pacheco, 2003; Yang & Wang, 2007).

Another approach presented Situngkir (2004) who used map-based simulation and simplified data of transportation through sea for modeling influenza disease in Indonesia. Jin and Liu (2006) used CA for simulations in spatial heteregenous hosts mixing and the natural birth rate and he compared the results with homogenous cases.

Many of existing models employ individual based model (IBM) in CA environment. In the model by Lopez, Burguerner, and Giovanini (2014) each individual of the population is represented by a cell in the lattice of CA . This way of modeling an epidemic situation allows to individually define the characteristic of each individual but it does not reflect the real density of the population. In the model by Sirakoulis, Karafyllidis, and Thanailakis (2000) the state of the cell is obtained from the fraction of the number of individuals which are susceptible, infected, or recovered from the disease . This approach shows the epidemic propagation during the population movement within the cells but does not specify anything about inhomogeneous and mixing of susceptible, infective, and recovered accordingly. Similar model extended for patchy population has been given by Athithan, Shukla, and Biradar (2014) who introduced heterogeneity and population movement into the model by Sirakoulis et al., (2000).

Our model combines presented solutions and introduces several new improvements. This model allows us to capture the individual heterogeneity as well as a realistic behavior of individual contacts. Presented solution is based on inhomogeneous distribution across lattice - we constructed simulation region according to the real population density map of Poland and the accuracy of map projection can be easily changed by changing the lattice resolution. In addition to heterogeneity we introduce new spatial and stochastic parameters to reflect the real processes related to population dynamics. In our model we approximate all other possible types of infection (infections in a cell, death and birth rate, mortality rate) by a semi-empirical mathematical parameters, which is an extension of models from Jin and Liu (2006) Situngkir (2004) White and Sanchez (2007) White et al. (2009). Additionally, we model the possibility of infection of an individual due to mobility of other infective individuals, which is in agreement with the model by Min-Hua, Duan-Ming, Gui-Jun, Yan-Ping, and Xin-Yu (2008), but we introduce daily commutes of individuals with variable ratio of commuters depending on whether they're healthy or infected. The state of the cell is obtained from the fraction of the number of individuals which are susceptible, infected, or recovered from the disease.

Consequently we present novel, complete SEIR model to simulate the epidemic spread based on CA. It is expected that incorporating several improvements i.e. real density of the population, parameters reflected the dynamics of population and different types of infection in one numerical system, we can get the results consistent with the behavior of a real epidemic.

2. Disease modeling frameworks

One of the most popular models of the infectious diseases is the classical *SEIR* model (Aron & Schwartz, 1984). In this model, the whole population is divided into four compartments which describe separated groups of individuals: susceptible which are able to contract

the disease (denoted by S), exposed which are infected but not yet infectious (denoted by E), infective which are capable of transmitting the disease (marked by I) and recovered which are permanently immune (denoted by R). The letters represent the number of individuals in each compartment at a particular time, and the whole population size N is the sum of above fractional groups, i.e.

$$S + E + I + R = N. \tag{1}$$

It is assumed that the population size N described by Eq.(1) is constant.

The time evolution of the population compartments in the *SEIR* model is described by four nonlinear differential equations (2)–(5) (Pfeifer et al., 2008), namely

$$\frac{dS}{dt} = \mu(N - S) - \beta(t)\frac{I}{N}S, \tag{2}$$

$$\frac{dE}{dt} = \beta(t)\frac{I}{N}S - (\mu + \delta)E, \tag{3}$$

$$\frac{dI}{dt} = \delta E - (\mu + \gamma)I, \tag{4}$$

$$\frac{dR}{dt} = \gamma I - \mu R. \tag{5}$$

Here, time is denoted by t, μ is both birth and death rate, $1/\delta$ is the mean latent period for the disease, $1/\gamma$ is the mean infectious period and β is the contact rate which denotes the probability of getting the disease in a contact between a susceptible and an infectious individuals. Above parameters define the basic reproduction number (R_0) by the following equation

$$R_0 = \frac{\delta}{\mu + \delta}\frac{\beta}{\mu + \gamma}. \tag{6}$$

Eq. (6) determines whether or not an infectious disease can spread through a population. Basic reproduction number is the number of cases generated by one case on average over the course of its infectious period: when $R_0 < 1$ ($R_0 > 1$) the infection will die out (will spread), respectively.

SEIR model is the population level model, which keeps track of the total numbers of individuals in each compartment, but does not explicitly model individuals and their specific behaviors. This model assumes that population is large enough that the effect of deviations from the mean can be ignored when considering the infection process at the population level (Sannetspiel, 2009). When using the SEIR model it is not possible to integrate geographical data and demographic interactions into the model (Pfeifer et al., 2008). Other way to model the spreading of infection are individual-based models (IBM), which model the state of each member of the population. This approach is computationally expensive and the complexity grows with the population size (Sannetspiel, 2009).

In this paper we describe the results of the numerical simulations of the CA, which is a combination of the SEIR and IBM models and introduces some stochastic and spatial parameters, which allow to analyze the dynamic of the disease also for real geographical and demographic data. Fig. 1 presents the component diagram of our modeling framework as integrated CA and IBM systems. Our model includes the key characteristics of the individuals which are diffused into CA model, which initiate and control transitions in the modeled system. In our framework, the characteristics of each individual (i.e. state of an individual - Fig. 2, interactions with other individuals due to daily commutes, probability of death of an infected individual) are tracked through time like in IBM model. Evolution of the system is described by simulation loops which proceed individual by individual like in IBM model and next cell by cell like in SEIR model.

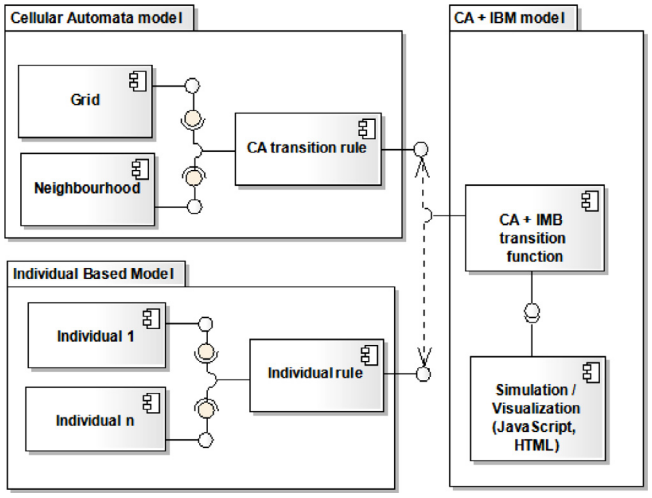

Fig. 1. Component diagram of the combined CA and IBM model.

3. Epidemic spread modeling based on cellular automata

Two–dimensional CA is a discrete dynamical system consisting of a regular lattice of objects called cells. All cells change their state simultaneously at every time step, using local transition function. More formally a CA is defined by a 4–tuple (C, Q, V, f), where C is a cellular space defined as

$$C = \{(i, j) : 0 \le i < r \wedge 0 \le j < c\}. \tag{7}$$

In Eq.(7) (i, j) represent position in ith row and jth column, r is the row count and c is the column count; Q is a finite set of states; V is a function mapping the cell to its neighboring cells (Delorme, 1999; del Rey, White, & Sánchez, 2006). In our work we use the Moore neighborhood with a distance of one. Therefore, function V for (i_0, j_0)-th cell is defined as

$$V_{i_0,j_0} = \{(i, j) \in C \setminus \{(i_0, j_0)\} : |i - i_0| \le 1 \ \wedge \ |j - j_0| \le 1\}. \tag{8}$$

Finally, f is a local transition function, such that $s_{ij}^t = f(i, j, s_{1,1}^{t-1}, \ldots, s_{r,c}^{t-1}) \in Q$, where s_{ij}^t is an (i, j)-th cell's state at the time t.

As the simulation's target we chose the density map of Poland. In our model each cell of the lattice relates to a geographical region of Poland (Fig. 3). We use a square lattice consisting of 36 rows and columns, therefore each cell represents a square area of Poland with approximate dimensions of 18 × 18 km (Fig. 3). The state of an (i, j)-th cell at the time t can be expressed by a $(a + b + 2)$-dimensional vector

$$s_{ij}^t = (S_{ij}^t, E_{ij|1}^t, E_{ij|2}^t, \ldots, E_{ij|a}^t, I_{ij|1}^t, I_{ij|2}^t, \ldots, I_{ij|b}^t, R_{ij}^t), \tag{9}$$

where S_{ij}^t and R_{ij}^t are respectively numbers of susceptible and recovered individuals and $E_{ij|d}^t$ and $I_{ij|d}^t$ are the numbers of exposed and infective individuals in a dth day of the given stage of an infection. Moreover, a and b are lengths of exposed and infective stages in days, which implies that the model can be used only for fixed-length diseases. We also define N_{ij}^t to be the sum of elements of s_{ij}^t, i.e. the population of the cell. Since elements of s_{ij}^t are represented in a computer memory, set of possible states Q is finite. Fig. 2 shows all possible state transitions of the individuals in our model with lengths of the E (a) and I states (b).

Distribution of the population is inhomogeneous among the different cells and can change over time due to births and deaths. Values of N_{ij}^0 for all cells are based on real-world Polish population counts from NASA (CIESIN, Food, Programme, & de Agricultura Tropical, 2005). Cells located on the Baltic Sea or out of the border of Poland have the population set to zero and we assume that these

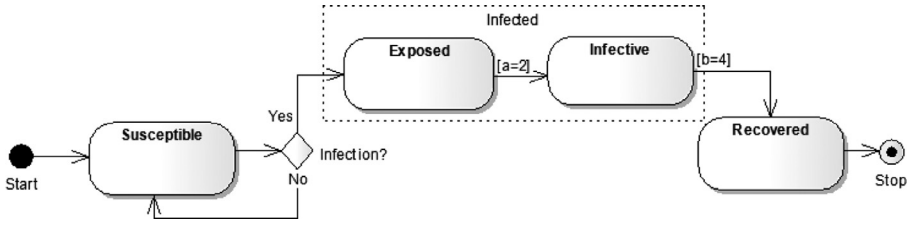

Fig. 2. State machine diagram of the possible state transitions of the individuals in modeled system.

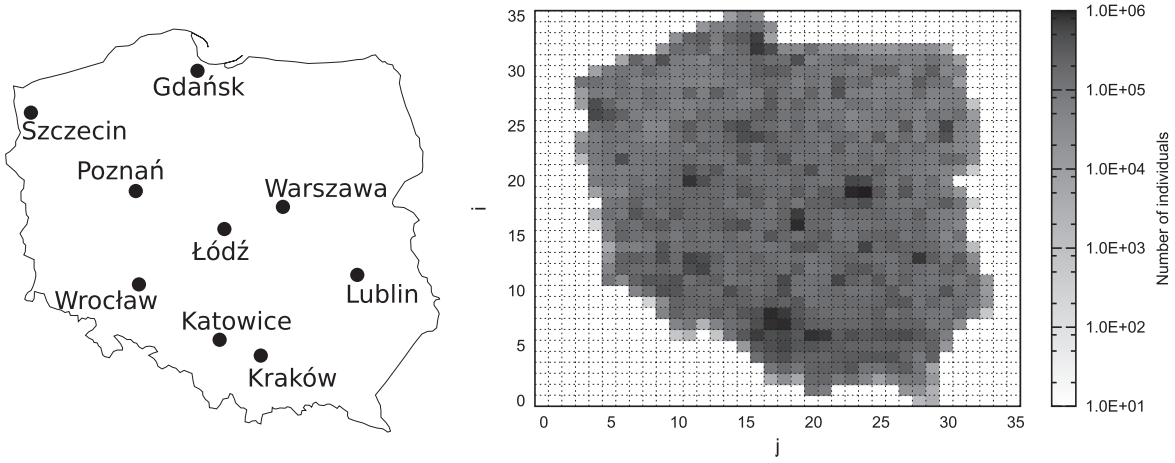

Fig. 3. Visualization of the region used in simulation: the contour of Poland (left-hand panel) and contour map of the population density (right- hand panel).

cells are not available for settlement and exclude them from neighborhoods in the updated neighborhood function

$$V^*_{i_0,j_0} = \{(i,j) \in V_{i_0,j_0} : N^0_{ij} > 0\}\,. \tag{10}$$

The model includes daily commutes, which are simulated as follows. Every commuting individual starts each iteration step (each day) in his source cell. Then, the individual goes to work or school by moving to his destination cell, where he interacts with other individuals. At the end of the iteration step he goes back to his source cell. Let n^t_{ij} denote a vector of numbers of individuals who move out of the cell at the time t

$$n^t_{ij} = \left(\phi_h S^t_{ij}, \phi_s E^t_{ij|1}, \ldots, \phi_s E^t_{ij|a}, \phi_s I^t_{ij|1}, \ldots, \phi_s I^t_{ij|b}, \phi_h R^t_{ij}\right), \tag{11}$$

where $\phi_h\,(\phi_s)$ is a fraction of the healthy (infected) population commuting outside of their cell. Since we have to keep track of commuting individuals during a step of the iteration, we use

$$s^t_{ij\to xy} = (S^t_{ij\to xy}, E^t_{ij\to xy|1}, \ldots, E^t_{ij\to xy|a}, I^t_{ij\to xy|1}, \ldots, I^t_{ij\to xy|b}, R^t_{ij\to xy}), \tag{12}$$

to denote the state of individuals commuting from the cell (i,j) to the cell (x,y) at the time t. Also, let $N^t_{ij\to xy}$ denote the sum of elements of $s^t_{ij\to xy}$.

Here we describe the transition of the CA from the time step t to the time step $t+1$. For the cell $(i,j) \in C$, such that $N^t_{ij} = 0$, state does not change and there are no moves between cells, leading to the following equations

$$s^{t+1}_{ij} = s^t_{ij}, \tag{13}$$

$$\forall_{(x,y)\in C}\; s^t_{ij\to xy} = 0 \cdot n^t_{ij}\,. \tag{14}$$

For the cell $(i,j) \in C$, such that $N^t_{ij} > 0$, we have

$$s^t_{ij\to xy} = \begin{cases} (1-\phi_c)\cdot \dfrac{n^t_{ij}}{|V^*_{ij}|}, & (x,y) \in V^*_{ij}, \\ \phi_c \cdot n^t_{ij}, & (x,y) = d(ij), \\ 0 \cdot n^t_{ij}, & \text{otherwise}, \end{cases} \tag{15}$$

where $d(ij)$ is the index of the closest dense cell at the Moore distance greater than one[1] and ϕ_c is a fraction of commuters commuting outside of their neighborhood. There is an assumption in our model that they commute to the closest dense cell (a cell with population above 50,000) at the Moore distance greater than 1. Probability that a susceptible individual will be infected, p^t_{ij}, is defined as

$$p^t_{ij} = \begin{cases} 0, & q^t_{ij} < 0, \\ 1, & q^t_{ij} > 1, \\ q^t_{ij}, & \text{otherwise}, \end{cases} \tag{16}$$

$$q^t_{ij} = \text{rnd}\left(1 - \exp \times\left(-\beta \frac{\sum_{k=1}^{b} I^t_{ij|k} + \sum_{(x,y)\in C}\sum_{k=1}^{b}\left(I^t_{xy\to ij|k} - I^t_{ij\to xy|k}\right)}{N^t_{ij} + \sum_{(x,y)\in C}\left(N^t_{xy\to ij} - N^t_{ij\to xy}\right)}\right), c_v\right), \tag{17}$$

where $\text{rnd}\,(\mu, c_v)$ is a Gaussian random number generator with the mean μ and the coefficient of variation c_v. As in the SEIR model and other CA–IBM based models probability of the infection is based on the fraction of infective individuals (see Eq. (17)). Similarly to a model by Sun, Jin, Song, Chakraborty, and Li (2011) we use Gaussian noise to incorporate randomness . The transition function of the model is

[1] We use breadth-first search to find the closest dense cell. If there are two or more dense cells at the same distance, then there is no guarantee which cell will be chosen.

given by the set of equations for $k \in \{n \in \mathbb{Z} : 2 \leq n \leq a\}$ and $l \in \{n \in \mathbb{Z} : 2 \leq n \leq b\}$

$$S_{ij}^{t+1} = (1 - \mu_d + \mu_b)\left((1 - p_{ij}^t)(S_{ij}^t - \sum_{(x,y)\in C} S_{ij\to xy}^t) + \sum_{(x,y)\in C}(1 - p_{xy}^t)S_{ij\to xy}^t\right), \quad (18)$$

$$E_{ij|1}^{t+1} = (1 - \mu_d + \mu_b)\left(p_{ij}^t\left(S_{ij}^t - \sum_{(x,y)\in C} S_{ij\to xy}^t\right) + \sum_{(x,y)\in C} p_{xy}^t S_{ij\to xy}^t\right), \quad (19)$$

$$E_{ij|k}^{t+1} = (1 - \mu_d + \mu_b - \mu_{vm})E_{ij|k-1}^t, \quad (20)$$

$$I_{ij|1}^{t+1} = (1 - \mu_d + \mu_b - \mu_{vm})E_{ij|a}^t, \quad (21)$$

$$I_{ij|l}^{t+1} = (1 - \mu_d + \mu_b - \mu_{vm})I_{ij|l-1}^t, \quad (22)$$

$$R_{ij}^{t+1} = (1 - \mu_d + \mu_b)R_{ij}^t + (1 - \mu_d + \mu_b - \mu_{vm})I_{ij|b}^t, \quad (23)$$

where μ_b (μ_d) denotes new births (natural deaths) per one individual per time step; μ_{vm} denotes mortality rate per time step, i.e. probability of a death of an infected individual.

Eq. (19) describes new infections where individuals in the cell are infected with the probability p_{ij}^t. The individuals commuted out of the cell are infected with the probability for a cell they commuted to. Eqs. (20)–(23) describe progress of the disease in already infected individuals. Furthermore, Eqs. (18)–(23) include simulation of natural deaths, natural births and deaths due to the virus morbidity.

In this work, following an approach presented by Chowell, Hengartner, Castillo-Chavez, Fenimore, and Hyman (2004) reproduction number R_0 is computed empirically in preliminary simulations for each set of parameters . In these simulations a is set to the same value as b. At $t = a$ simulation is stopped. At this moment all infections come from the first generation of infection, i.e. from individuals that were infected at $t = 0$. Therefore reproduction number is given by the equation

$$R_0 = \frac{\sum_{(i,j)\in C}\sum_{k=1}^{a} E_{ij|k}^a}{\sum_{(i,j)\in C} I_{ij|1}^0}. \quad (24)$$

4. Numerical results and discussions

In this section we consider the numerical results of epidemic spreading in the two-dimensional CA model. To verify the rationality of the proposed model we implemented a JavaScript and HTML 5 simulator, so all simulations can be parametrized and performed in a web browser. The geometry and resolution of the simulation region are described in Section 3 and the time evolution is defined by the set of Eqs. (18)–(23). For comparative purposes all simulations were performed for the same initial configuration of infected individuals, which was generated by a random selection of 15 cells with the population above 50,000 of inhabitants and moving in these cells a group of 20 individuals to the infective state. The threshold of 50,000 of inhabitants in infected cells at $t = 0$ was chosen to reflect the stronger likelihood of location of the disease's sources in bigger cities which group more mobile individuals. Fig. 4 shows the initial configuration of infected individuals at $t = 0$.

Our simulations were performed for an epidemic of influenza, thus we set the parameters describing the characteristics periods of the states E and I to $a = 2$ and $b = 4$, respectively (Lloyd, 2001)

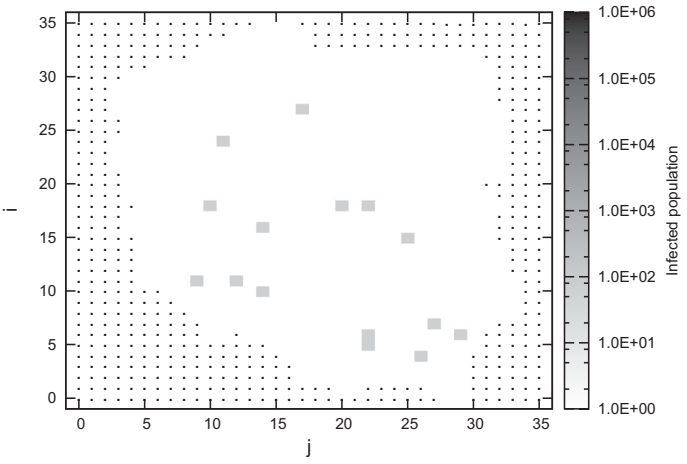

Fig. 4. Distribution of infected individuals in the initial configuration, at $t = 0$.

(Fig. 2). Moreover, using the data provided by Central Statistical Office of Poland we set the parameters of birth rate $\mu_b = 2.77 \times 10^{-5}$ and death rate $\mu_d = 2.7 \times 10^{-5}$(Central Statistical Office, 2013). For μ_{vm} we used the information from 2009 flu epidemic in England, where reported mortality rate was equal to 0.026% (Donaldson, Rutter, Ellis, Greaves, Mytton, Pebody, & Yardley, 2009); for our simulations we assumed that this probability of death is distributed uniformly throughout all days of the infection. Based on the statistical data for Poland (Central Statistical Office, 2011) we determined the values of the parameters $\phi_h = 5.5 \times 10^{-2}$ and $\phi_c = 0.23$. We set also $\phi_s = 1.3 \times 10^{-2}$, which is our estimate because real-world data for this parameter is not available.

First, we consider three values of the coefficient of variation of the probability of infection $c_v \in \{0, 0.5, 1\}$ and perform for it the calculation of the basic reproduction number R_0 as a function of β. Fig. 5 presents the ensemble averaged relationship between the reproduction number R_0 and the contact rate parameter β. The value of R_0 grows with β and the standard deviations of the results is larger for greater value of c_v, which is a consequence of stronger randomness and corresponds with our assumptions. There is also a discernible dependence between R_0 and c_v: the reproduction number R_0 increases for higher values of c_v (right panel of Fig. 5).

Next, we perform detailed simulations for chosen values of $\beta \in \{0.2, 0.4, 0.6\}$, corresponding to values of R_0 in the most realistic range $R_0 \in <0.6, 2.46>$.

Fig. 6 presents the results of the epidemic disease spreading for particular (non-averaged) simulations with $\beta = 0.6$. For non-random simulation (left-hand column) we can observe that at time $t = 40$ infection spreads mainly around places where it has started. This spread is boosted in the neighborhoods of cells containing more people, which correspond to big cities (compare left panel of Fig. 3). At $t = 80$ epidemic is close to its peak - it is easily visible on the bottom row of Fig. 7, that the number of infective individuals reaches its maximum for time $t \approx 70$. Next, we can see that the infection started to collapse in cells that were highly infected at $t = 40$ and finally, at $t = 120$ infection's wave moved to previously non–infected regions. Middle panel of Fig. 6 illustrates the weak random case with $c_v = 0.5$. Evolution of the infection for short periods of time, $t < 60$, resembles the results for non-random case with $c_v = 0$; upper row in middle column of Fig. 6 shows the results for $t = 40$. For longer time, $t = 80$ and $t = 120$, differences are more noticeable - we can observe small spots of infections in cells located far from the main epidemic's wave. Overall, for stronger randomness evolution of the infection become more chaotic and the random effects are more noticeable with growing c_v - middle and right columns of Fig. 6.

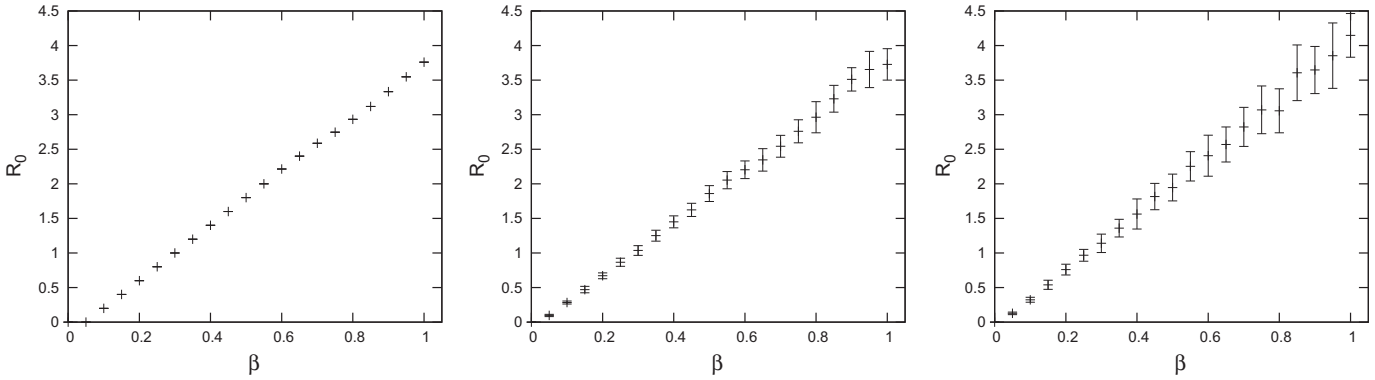

Fig. 5. Ensemble averaged values of R_0 (over 10 runs) for $c_v = 0$ (left-hand panel), $c_v = 0.5$ (central panel), $c_v = 1$ (right-hand panel). Error bars represent one standard deviation of the results.

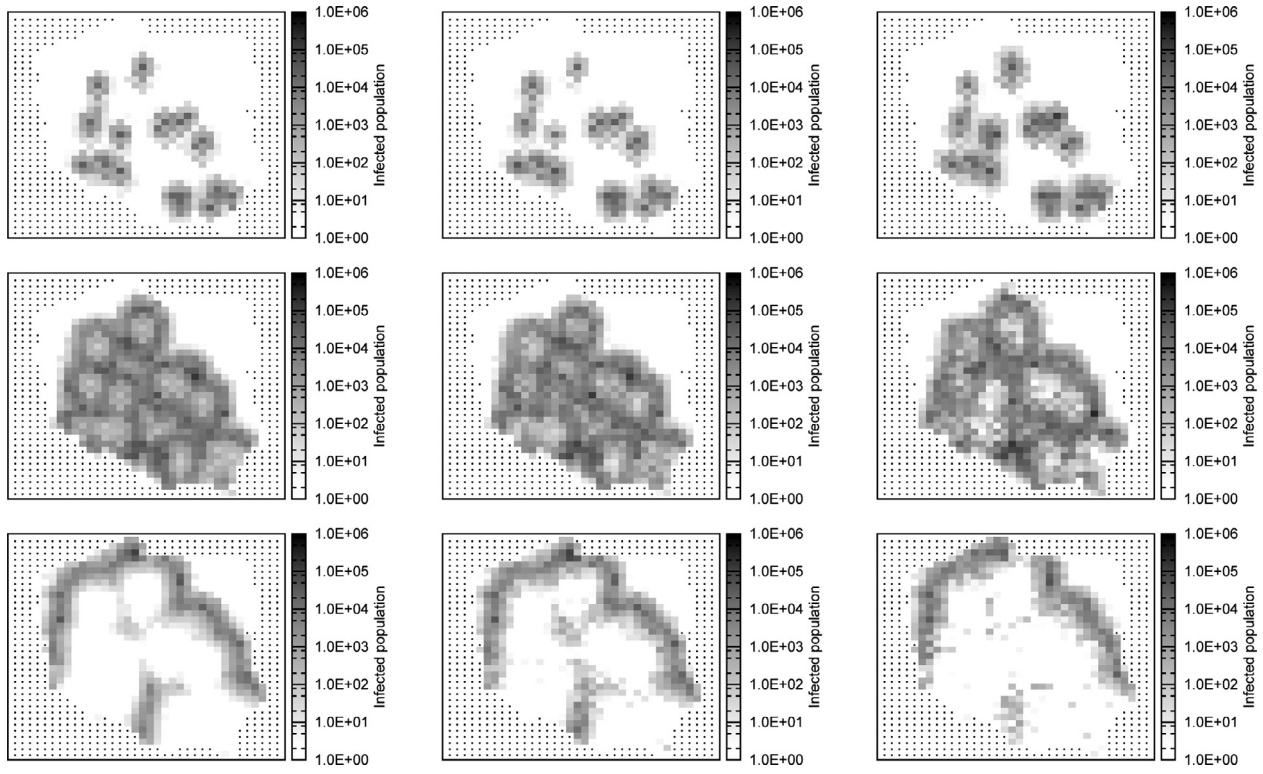

Fig. 6. Evolution of the infection for $\beta = 0.6$ and $c_v = 0$ (left-hand column), $c_v = 0.5$ (central column), $c_v = 1$ (right-hand column) at three different instants: $t = 40$, 80 and 120.

Fig. 7 shows time history of the number of individuals in states $E + I$, S and R for ensemble averaged results of simulations performed for all values of β analyzed until now. The upper row of Fig. 7 shows that for small values of the contact rate parameter (β) the infection cannot spread and the number of individuals susceptible (S) does not change with time regardless of the strength of random effects (c_v). Middle and bottom rows of Fig. 7 reveal that the time history of the number exposed and infective ($E + I$) depend significantly on the choice of contact rate. For higher values of β the infection spread quicker and reaches its maximum for $t \approx 130$ for the case with $\beta = 0.4$, while for $\beta = 0.6$ the peak of the infection is located at $t \approx 70$.

We can easily compare the number of infected individuals as the function of time for different values of c_v and β - the results of Fig. 7 are used to draw Fig. 8 which displays the number of infected individuals as a function of time. The results confirm that higher values of contact rate β speed up and broaden the disease. Similarly, stronger effects of randomness c_v cause the increase of the number of infected individuals and this effect agree with the relationship between the number of recovered individuals R and contact rate β on middle and bottom row of Fig. 7. The characteristic step on the falling edge of the time history of the infected population on Fig. 8 is an artificial effect caused by the geometry of the simulation region and cut-off of the simulation at the borders.

Fig. 9 displays the parameters describing the dynamic of the disease (left panel – the contact rate, middle panel – time of the peak of the epidemic and right panel – maximum number of infected at one instant of time) as a function of basic reproduction number R_0. Simulation results show that spread of the infection strongly depends on R_0. For $R_0 < 1$ total number of infected people was less than 5×10^3, whereas for $R_0 = 1$ (obtained for $c_v = 0$ and $\beta = 0.3$) it was equal to 7.9×10^6. Results of the simulation show the exponential relationship between basic reproduction number and the attack rate (cumulative

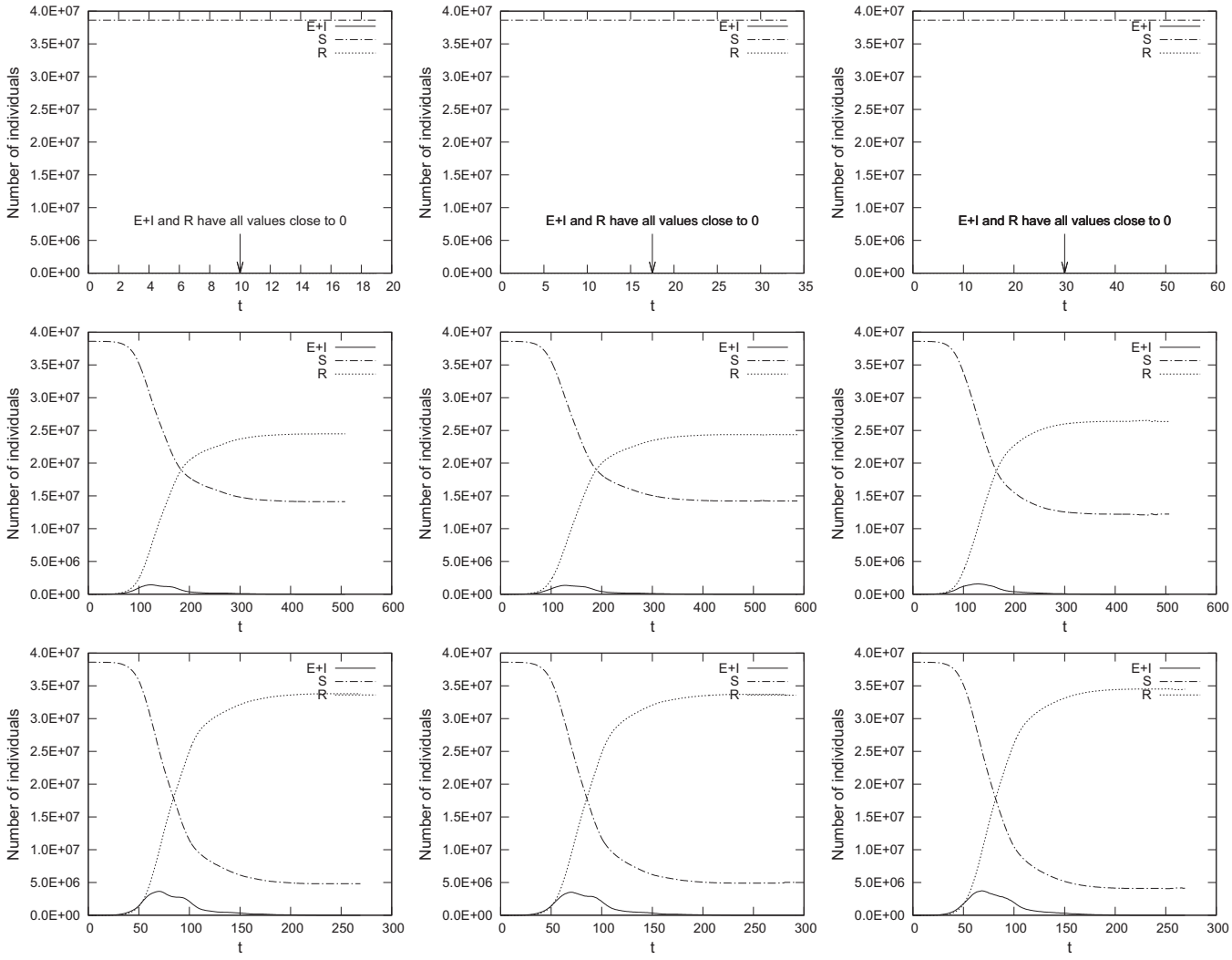

Fig. 7. Average simulation results from 10 runs for $c_v = 0$ (left-hand column), $c_v = 0.5$ (central column), $c_v = 1$ (right-hand column) and $\beta = 0.2$ (upper row), $\beta = 0.4$ (middle row), $\beta = 0.6$ (lower row).

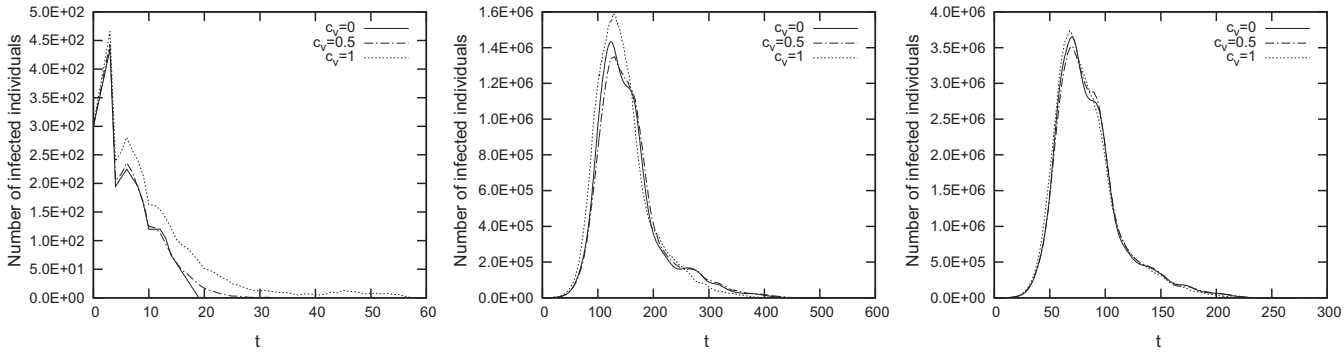

Fig. 8. Average evolution of the infected population from 10 runs for $\beta = 0.2$ (left-hand panel), $\beta = 0.4$ (central panel), $\beta = 0.6$ (right-hand panel).

number of infected individuals during epidemic) and the dependence of the maximal number of currently infected on R_0 is almost linear. We did not discover realistic epidemic scenarios for attack rates lower than 0.4. In these cases epidemic either collapsed after short time (see e.g. results for $\beta = 0.2$ on Fig. 8) or spread very slowly, especially for $R_0 \approx 1$ (see middle panel of Fig. 9).

5. Conclusions

A new mathematical model for simulating the spread of epidemics is introduced in this paper. It is based on the two–dimensional cellular automata, in which each cell represents a part of the land with real population's count from Poland. Each cell's population is

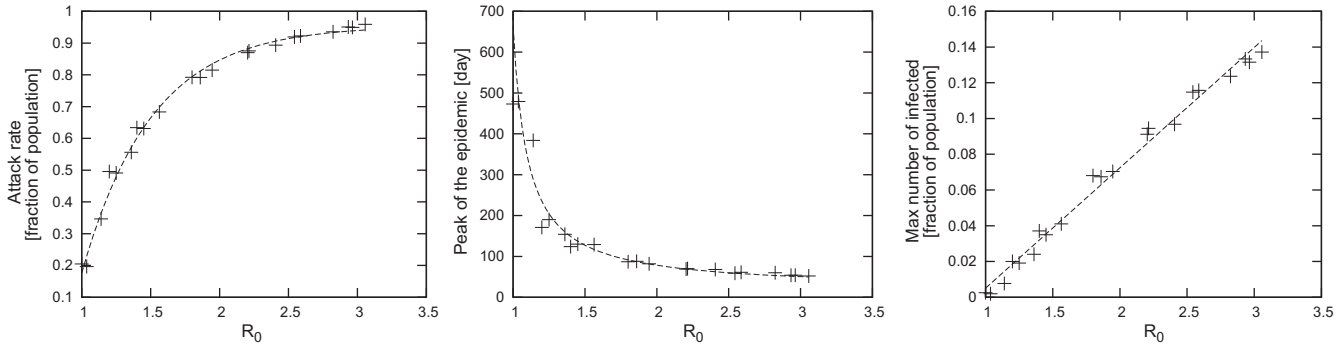

Fig. 9. Relationships between R_0 and: the attack rate (left-hand panel); the day at which number of infected individuals was the highest (central panel); the maximal number of currently infected individuals (right-hand panel). Cases for $R_0 < 1$ were omitted because the attack rate for them was close to 0.

divided into four groups: susceptible, exposed, infective and recovered, therefore the model can be classified as a SEIR-type model. We present the first CA-based model that includes daily commutes of individuals with variable ratio of commuters depending on whether they're healthy or infected. We include commutes outside of the nearest neighborhood. For Poland we assume that individuals commute to the closest cell with population above 50,000, since such regions should be more economically advanced.

Performed simulation leads to similar results as pure IBM models, even though it requires less complicated input and is less expensive computationally. Computational techniques used in the presented approach save time of the simulation and allow it to extend the scope of simulation to quickly adapt them to a selected real environment.

In the current form of the model travels across the country are not simulated. Also the age structure of the population used in the simulation is uniform, therefore, it does not reflect impact of the age on susceptibility to disease. We have not been able to compare the simulation results to the behavior of real influenza outbreaks in Poland because data from them had not been available. From 1918 flu pandemic only mortality rates from Wrocław had been published (Lubinski, 1923). More recent flu epidemics were supressed before reaching major parts of the population.

Further work aiming at enriching the model needs to be done. Since probability of the infection can depend on the age of an individual, population should be grouped by age. Moreover, probability of the infection in some diseases can be seasonal and change over time but viral mutation is currently not incorporated into the model. A future study could use the model to assess effectiveness of vaccination in preventing epidemics.

The simulation framework implemented in JavaSrcript//HTML 5 can be accessed online at

http://zasoby.umcs.lublin.pl/marek.medrek/public_html/epidemic-simulation-master/ .

Acknowledgments

We would like to thank I. Karafyllidis and A. M. del Rey for the provision of their papers.

References

Achmed, E., & Agiza, H. (1998). On modeling epidemics including latency, incubation and variable susceptibility. *Physica A, 253*, 347–352.

Aron, J. L., & Schwartz, I. B. (1984). Seasonality and period-doubling bifurcations in an epidemic model. *Journal of Theoretical Biology, 110*(4), 665–679.

Athithan, S., Shukla, V., & Biradar, S. (2014). Dynamic cellular automata based epidemic spread model for population in patches with movement. *Journal of Computational Environmental Sciences, 2014*.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., & Hyman, J. (2004). The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. *Journal of Theoretical Biology, 229*(1), 119–126.

CIESIN, C. U., Food, U. N., Programme, A. & de Agricultura Tropical, C. I.(2005). Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-count. Accessed: 06.07.13.

Delorme, M. (1999). An introduction to cellular automata. *Mathematics and its application, 460*, 5–50.

Donaldson, L. J., Rutter, P. D., Ellis, B. M., Greaves, F. E., Mytton, O. T., Pebody, R. G., & Yardley, I. E. (2009). Mortality from pandemic a/h1n1 2009 influenza in england: public health surveillance study. *BMJ: British Medical Journal, 339*.

Itami, R. (1994). Smiulating spatial dynamics: cellular automata theory. *Landscape and Urban Planning, 30*.

Jin, Z., & Liu, Q. (2006). A cellular automata model of epidemics of a heterogeneous susceptibility. *Chinese Physics, 15*, 1248–1256.

Liu, Q. X., Jin, Z., & Liu, M. X. (2006). Spatial organization and evolution period of the epidemic model using cellular automata. *Physical Review E, 74*(3), 031110.

Lloyd, A. L. (2001). Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical Population Biology, 60*(1), 59–71.

Lopez, L., Burguerner, G., & Giovanini, L. (2014). Addressing population heterogeneity and distribution in epidemics models using a cellular automata approach. *BMC Research Notes, 7*.

Lubinski, H. (1923). Statistische betrachtungen zur grippepandemie in breslau 1918–22. *Zentralbl Bakteriol Parasitenkd Infektionskrankheiten, 1924*(91), 372–383.

Central Statistical Office(2011). Commuting to work in 2010 r.on the basis of bael. http://www.stat.gov.pl/gus/5840_12281_PLK_HTML.htm. Accessed: 06.13.14.

Central Statistical Office(2013). Basic information about polish demographic development until 2012. http://www.stat.gov.pl/cps/rde/xbcr/gus/L_podst_inf_o_rozwoju_dem_pl_do_2012.pdf. Accessed: 06.13.13.

Milne, G., Fermanis, C., & Johnston, P. (2008). A mobility model for classical swine fever in feral pig populations. *Veterinary Research, 39*(6), 53.

Min-Hua, H., Duan-Ming, Z., Gui-Jun, P., Yan-Ping, Y., & Xin-Yu, T. (2008). Spatiotemporal characteristic of epidemic spreading in mobile individuals. *Chinese Physics Letters, 25*.

Moreno, Y., Gmez, J. B., & Pacheco, A. F. (2003). Epidemic incidence in correlated complex networks. *Physical Review E, 68*.

Murray, J. (1993). *Mathematical Biology*. Springer-Verlag.

Peng, S., Wang, G., & Yu, S. (2013). Modeling the dynamics of worm propagation using two-dimensional cellular automata in smartphones. *Journal of Computer and System Sciences, 79*, 586–595.

Pfeifer, B., Kugler, K., Tejada, M. M., Baumgartner, C., Seger, M., Osl, M., et al. (2008). A cellular automaton framework for infectious disease spread simulation. *The Open Medical Informatics Journal, 2*, 70.

del Rey, A. M., White, S. H., & Sánchez, G. R. (2006). A model based on cellular automata to simulate epidemic diseases. In *Cellular automata* (pp. 304–310). Springer.

Sannetspiel, L. (2009). *The Geographic Spread of Infectious Diseases*. Princeton University Press.

Sirakoulis, G. C., Karafyllidis, I., & Thanailakis, A. (2000). A cellular automaton model for the effects of population movement and vaccination on epidemic propagation. *Ecological Modelling, 133*(3), 209–223.

Situngkir, H. (2004). Epidemiology through cellular automata, case of study: avian influenza in indonesia. *Working Paper WPF200*.

Sun, G.-Q., Jin, Z., Song, L.-P., Chakraborty, A., & Li, B.-L. (2011). Phase transition in spatial epidemics using cellular automata with noise. *Ecological research, 26*(2), 333–340.

White, S. H., del Rey, A. M., & Sanchez, G. R. (2007). Modeling epidemics using cellular automata. *Applied Mathematics and Computation, 186*(1), 193–202.

White, S. H., del Rey, A. M., & Sanchez, G. R. (2009). Using cellular automata to simulate epidemic diseases. *Applied Mathematical Sciences, 3*(20), 959–968.

Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of Modern Physics, 55*, 601–644.

Yang, R., & Wang, e. B. H. (2007). Epidemic spreading on heterogeneous networks with identical infectivity. *Physics Letters A, 364*.