Tracking Uncertainty in a Spatially Explicit Susceptible-Infected Epidemic Model

Jan M. Baetens* and Bernard De Baets

KERMIT, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, Gent, Belgium
{jan.baetens,bernard.debaets}@ugent.be

Abstract. In this paper we conceive an interval-valued continuous cellular automaton for describing the spatio-temporal dynamics of an epidemic, in which the magnitude of the initial outbreak and/or the epidemic properties are only imprecisely known. In contrast to well-establish-ed approaches that rely on probability distributions for keeping track of the uncertainty in spatio-temporal models, we resort to an interval representation of uncertainty. Such an approach lowers the amount of computing power that is needed to run model simulations, and reduces the need for data that are indispensable for constructing the probability distributions upon which other paradigms are based.

Keywords: continuous cellular automaton, epidemic spread, imprecision, uncertainty.

1 Introduction

As a consequence of their rigorous formulation of macroscopic phenomena, as well as their rich history which can be traced back to the development of modern calculus during the 17th and 18th century, and during which their efficacy has been proven manifold, ordinary differential equations (ODEs) are generally resorted to for describing (a)biological processes, as illustrated extensively in the work of Murray [16], whereas partial differential equations (PDEs) are mostly employed if one is not merely interested in the process' temporal dynamics but also in the spatial patterns it generates, such as the spread of an epidemic [17]. Further, in order to cope with the variability inherent to natural processes, researchers have resorted to stochastic DEs [20], fuzzy DEs [11,19], and to massive Monte Carlo (MC) simulations [26] in the hope that the simulation results obtained through a model based upon one of these approaches would agree to a larger extent with the described process than the outcome of their deterministic counterparts do. Yet, each of these paradigms suffers from a serious drawback. More specifically, stochastic DEs are difficult to solve analytically, or require advanced numerical techniques in order to find an approximate solution, the theory on fuzzy ODEs, and, especially fuzzy PDEs is still maturing, while much computing time and effort is needed to perform MC simulations.

* Corresponding author.

S. Bandini et al. (Eds.): ACRI 2010, LNCS 6350, pp. 95–105, 2010.

To overcome these barriers, we propose an interval-valued continuous cellular automaton (ICCA) for describing epidemic spread if there is imprecision involved about the magnitude of the initial outbreak or the epidemic's characteristics. In essence, an ICCA can be regarded as a continuous CA (CCA) – also known as a coupled-map lattice – formulated by Kaneko [9], in which a cell's state is represented by an interval in $\mathbb{R}$, and not longer by a single real value.

A short overview of the mathematical preliminaries that are essential for a clear understanding of this paper is given in Section 2. In the third section we introduce the ICCA that can be used to describe epidemic spread if there is imprecision involved in the magnitude of the initial outbreak or the epidemic's characteristics. The former is addressed in the first part of the final section, while the latter is investigated more closely in the second part of this paper's final section.

2 Preliminaries

For the sake of clarity we state the definition of an ICCA on an arbitrary tessellation of a 2-dimensional Euclidean space. This paradigm constitutes an extension to the CCA paradigm since the states of the spatial entities are represented by an interval-valued in $\mathbb{R}$, while it also entails an extension to the classical CA paradigm conceptualized by von Neumann [27] since it allows irregular tessellations of $\mathbb{R}^2$.

Definition 1. *(Interval-valued continuous cellular automaton)*
An interval-valued continuous cellular automaton (ICCA) $\mathcal{C}$ can be represented as a sextuple

$$\mathcal{C} = \langle \mathcal{T}, S, s, s_0, N, \Phi \rangle \ ,$$

where

(i) *$\mathcal{T}$ is a countably infinite tessellation of a 2-dimensional Euclidean space $\mathbb{R}^2$, consisting of cells c_j, $j \in \mathbb{N}$.*
(ii) *S is an infinite set of intervals, where*

$$S \subseteq [\mathbb{R}] = \{[y_1, y_2] \mid y_1 < y_2 \wedge y_1, y_2 \in \mathbb{R}\} \ .$$

(iii) *The output function $s : \mathcal{T} \times \mathbb{N} \to S$ yields the state value of cell c_j at the t-th discrete time step, i.e. $s(c_j, t) = [s_1(c_j, t), s_2(c_j, t)]$.*
(iv) *The function $s_0 : \mathcal{T} \to S$ assigns to every cell c_j an initial state, i.e. $s(c_j, 0) = s_0(c_j)$.*
(v) *The neighborhood function $N : \mathcal{T} \to \bigcup_{p=1}^{\infty} \mathcal{T}^p$ maps every cell c_j to a finite sequence $N(c_j) = (c_{j_k})_{k=1}^{|N(c_j)|}$, consisting of $|N(c_j)|$ distinct cells c_{j_k}.*
(vi) *$\Phi = (\phi_j)_{j \in \mathbb{N}}$ is a family of functions*

$$\phi_j : S^{|N(c_j)|} \to S \, ,$$

each ϕ_j governing the dynamics of cell c_j, i.e.

$$s(c_j, t+1) = \phi_j\big(\tilde{s}(N(c_j), t)\big)\,,$$

where $\tilde{s}(N(c_j), t) = \big(s(c_{j_k}, t)\big)_{k=1}^{|N(c_j)|}$.

In the framework of this paper, we define N in such a way that $N(c_j)$ yields the Moore neighborhood of c_j, consisting of those cells $c_k \in \mathcal{T}$ that share either a vertex or a line segment with c_j. Sticking to this neighborhood function, it becomes straightforward to map $\mathcal{T}$ on a undirected graph $G(V, E)$, with vertex set $V = \mathcal{T}$, while E represents the edge set of G, containing an edge between c_j and c_k if $c_k \in N(c_j)$. Furthermore, in the remainder of this paper we restrict to the family of ICCA for which ϕ_j is the same for all $c_j \in \mathcal{T}$.

Definition 2. *(Homogeneous interval-valued continuous cellular automaton) A homogeneous interval-valued continuous cellular automaton (ICCA) is an ICCA fulfilling premises (i)-(v) of Definition 1 and for which there exists a* $\Theta : \bigcup_{k\in\mathbb{N}} S^k \to S$ *such that*

$$s(c_j, t+1) = \Theta\big(\tilde{s}\,(N(c_j), t)\big)\,.$$

Essentially, the construction of a homogeneous ICCA is less intricate than the composition its generalized counterpart given by Definition 1 since only one function Θ should be chosen that governs the dynamics of every $c_j \in \mathcal{T}$. Actually, most studied CA, such as rule 30 or the Game of Life [8], belong to this CA family.

3 A Spatially Explicit Model for Describing Epidemic Spread

3.1 The Model

The rich variety of CCA- and CA-based models that has been developed during the last decade for describing various spatial biological phenomena such as epidemics [6,14,28], population dynamics [3,5], tumor growth [13,23,24], biofilm development [21,22] and many other phenomena [10,25] is illustrative for the suitability of such models to mimic complex bioprocesses.

In a forthcoming work, Baetens and De Baets [2] propose a generalized CCA for modelling various biological processes that are traditionally described by means of PDEs. In this paper we focus on an epidemic sweeping through a geographical region, and which involves only non-reproducing susceptible and infected individuals. The spatio-temporal dynamics of such an epidemic can be captured by the following set of difference equations

$$\begin{cases} H(c_j, t+1) = H(c_j, t) - H(c_j, t) \displaystyle\sum_{c_k \in N(c_j)} w_{jk}\, F\left(\mathbf{U}_j, d_{jk}\right) U(c_k, t) \\ \\ U(c_j, t+1) = U(c_j, t) + H(c_j, t) \displaystyle\sum_{c_k \in N(c_j)} w_{jk}\, F\left(\mathbf{U}_j, d_{jk}\right) U(c_k, t) \end{cases} \tag{1}$$

where $H(c_j,t)$, resp. $U(c_j,t)$, represent the fraction of susceptible (healthy), resp. infected (unhealthy) individuals within polygon c_j at the t-th time step such that $H(c_j,t)+U(c_j,t)=1$, at all t, and for all c_j, F is a function describing the effect of landscape and connectivity characteristics, embodied in $\mathbf{U}_j$, on the epidemic, d_{jk} is the distance measured on a graph between the polygons c_j and c_k. Further, w_{jk} is a weighing factor, representing the influence of every $c_k \in N(c_j)$ in the determination of $H(c_j,t+1)$. A brief analysis of Eq. (1) shows that it has two fixed points, namely $\left(H_j^*,U_j^*\right)=(0,1)$ and $\left(H_j^*,U_j^*\right)=(1,0)$. Clearly, this model may be regarded as a discrete analog of a PDE-based SI-model, such as described in [17].

Taking into account that $H(c_j,t)+U(c_j,t)=1$, at all t, and for all c_j, we observe that the epidemic's dynamics can be tracked by considering only one of the system's equations. Further, by assuming that the region is spatially homogeneous, meaning that F does not depend on $\mathbf{U}_j$, we can reduce Eq. (1) to

$$U(c_j,t+1)=U(c_j,t)+H(c_j,t)\sum_{c_k\in N(c_j)} w_{jk}\,H(d_{jk})\,U(c_k,t)\,, \tag{2}$$

where we introduced the function H, for which

$$H(d_{jk})=\begin{cases}\nu_0, \text{ if } d_{jk}=0,\\ \nu_1, \text{ if } d_{jk}=1,\end{cases} \tag{3}$$

with ν_0 and ν_1 quantifications of the epidemic's virulence. These measures have to be chosen such that

$$\sum_{c_k\in N(c_j)} w_{jk}\,H(d_{jk})\leq 1, \quad \forall\, c_j\,, \tag{4}$$

assuring that $0\leq U(c_j,t)\leq 1$, at all t, and for all c_j. Finally, we put $w_{jk}=\frac{1}{8}$ for all j,k and $j\neq k$ and $w_{jk}=1$ if $j=k$. Consequently, c_j's eight nearest neighbours influence $U(c_j,t+1)$ to the same degree.

3.2 Incorporating Uncertainty in the Proposed Model

Clearly, the above outlined model is deterministic since it yields exactly the same simulation result if its parameters and the initial condition from which it evolves are unchanged. In order to turn it into a stochastic model that is capable of grasping the variability inherent to natural process, commonly, it is presumed that $H(c_j,0)$ and $U(c_j,0)$, or the model's parameters follow a prescribed type of probability distribution, and the model is simulated through extensive MC simulations. Unfortunately, the latter require much computing time and effort, whereas a thorough construction of the aforementioned distributions demands a considerable amount of spatial data, which, mostly, cannot be collected easily.

For that reason, we propose to characterize uncertainty by using an interval representation of the variables and parameters in Eq. (2). More specifically, in the remainder of this paper $H(c_j,t)$ and $U(c_j,t)$ are considered intervals

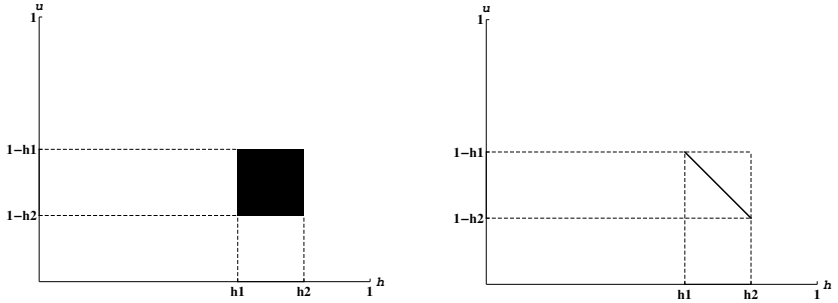

Fig. 1. Non-interactive (a) and interactive (b) intervals

in $[0, 1]$, such that we can write $H(c_j, t) = [h_1(c_j, t), h_2(c_j, t)]$ and $U(c_j, t) = [u_1(c_j, t), u_2(c_j, t)]$. Since the notion of uncertainty is for most researchers inextricably bound up with probability distributions, we will refer to an interval characterization of uncertainty as imprecision.

Seen the absence of derivatives or advanced mathematical functions in Eq. (2), it is relatively straightforward to evolve the system's spatio-temporal dynamics by means of basic interval arithmetic [15]. However, there is one pitfall that complicates the calculations, and inevitably leads to faulty conclusions if discarded. More precisely, one has to bear in mind the coupling between $H(c_j, t)$ and $U(c_j, t)$ through the condition $H(c_j, t) + U(c_j, t) = 1$, at all t, and for all c_j, which makes that $H(c_j, t)$ and $U(c_j, t)$ cannot take values independently of each other, so they can be termed interactive variables [7]. Hence, mathematical operators may not act on all couples in $H(c_j, t) \times U(c_j, t)$ (Fig. 1(a)), but only on the couples contained in

$$\{l\,(h_1(c_j, t), u_2(c_j, t)) + (1 - l)\,(h_2(c_j, t), u_1(c_j, t)) \mid l \in [0, 1]\}\,, \tag{5}$$

such as depicted in Figure 1(b).

In view of the existing interactivity, we can then write

$$H(c_j, t) + U(c_j, t) = [h_1(c_j, t) + u_2(c_j, t), h_2(c_j, t) + u_1(c_j, t)]\,, \tag{6}$$

and, analogously,

$$H(c_j, t) \cdot U(c_j, t) = [\min\left(h_2(c_j, t) \cdot u_1(c_j, t), h_1(c_j, t) \cdot u_2(c_j, t)\right), \\ \max\left(h_2(c_j, t) \cdot u_1(c_j, t), h_1(c_j, t) \cdot u_2(c_j, t)\right)]\,. \tag{7}$$

4 Simulation Study

In this section two sources of imprecision in Eq. (2) are examined more closely. The first one concerns the initial condition $U(c_j, 0)$ that is necessary to iteratively solve Eq. (2), and which can be deduced from spatial epidemiological data

that are becoming increasingly available as indicated by Beale [4]. Nonetheless, we must be aware of the imprecision that can be present in the outbreak data, as illustrated only recently by the outbreak of H1N1 [12]. Analogously, the parameters in Eq. (3) might only be known imprecisely. This is regarded as the second source of imprecision. All simulations reported in this section were performed in Mathematica 7.0 (Wolfram Research, Inc.) on a desktop PC with an Intel Dual Quad Core 3.16 GHz processor. Although a square tessellation consisting of 101×101 polygons was used in this paper, the described simulations could easily be performed when an irregular tessellation is employed. Such an irregular tessellation seamlessly complies with the spatio-temporal data in vector format [1], which are commonly available through geographical information systems and can contribute considerably to a more accurate description of bioprocesses. No boundary conditions had to be imposed since we employed differentiated neighborhood structures along the tessellation's boundaries. As such, the use of periodic boundary conditions, which is rather questionable if one wants to simulate an epidemic over a given geographical extent, is avoided.

4.1 Imprecise Initial Conditions

Often only imprecise information is available on the magnitude of an epidemic during its initial stage. This kind of imprecision can be incorporated easily in the model (Eq. (2)), by choosing $U(c_m, 0)$ an interval in $[0, 1]$, where c_m represents the polygon in which the epidemic broke out. In practice, the choice of an appropriate interval should be based upon expert opinions, though, in order to exemplify the ability of the formerly described discrete modeling paradigm to incorporate imprecision it suffices to adopt an arbitrary initial condition such as

$$U(c_j, 0) = \begin{cases} [0.2, 0.4] & , \text{ if } j = m\,, \\ [0, 0] & , \text{ else}\,. \end{cases} \tag{8}$$

Further, we assume that reliable information is available on the virulence of an epidemic, which allows us to assess $\nu_0 = 0.5$ and $\nu_1 = 0.5$, meaning that the spread of an infection in a polygon c_j can be equally attributed to infected individuals living in c_j as to infected individuals residing in c_j's neighborhood $N(c_j)$.

Figure 2 shows the center of $U(c_j, t)$ at two, five, ten and fifteen time steps after an epidemic outbreak occurred in the polygon c_m, as well as the length of the interval $U(c_j, t)$, denoted $|U(c_j, t)|$, at the same number of time steps. The former gives information on the expected proportion of infected individuals in every c_j, while the latter quantifies the imprecision that is related to this proportion. For reasons of clarity, we limited the depicted spatial extent of this figure to polygons through which the epidemic sweeps during the considered simulation period. This figure clearly shows that the imprecision originating from the imprecisely known proportion of infected individuals at $t = 0$ in the polygon c_m, propagates circularly like the epidemic wavefront. Since Figure 2(f) clearly shows that $|U(c_j, t)| \to 0$ as t increases, we may conclude that the ICCA

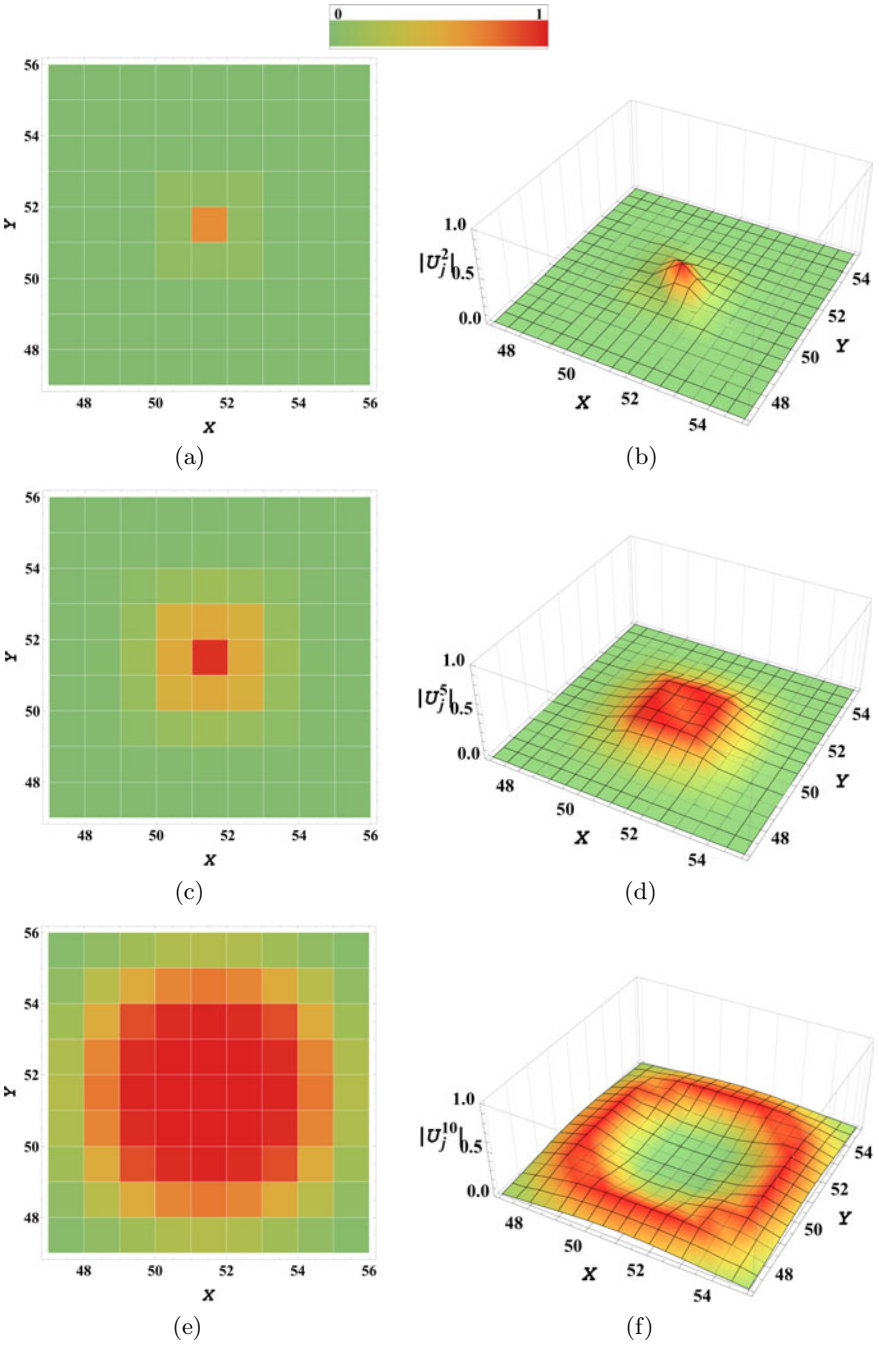

Fig. 2. Proportion of infected individuals, more precisely, the center of $U(c_j, t)$ (a,c,e), and the length of the interval $U(c_j, t)$, denoted $|U(c_j, t)|$ (b,d,f), two (a-b), five (c-d) and ten (e-f) time steps after an epidemic broke out in the center polygon c_m of a square tessellation with an initial magnitude given by Eq. (8)

evolves towards its fixed point $\left(H_j^*, U_j^*\right) = (0, 1)$, notwithstanding we imposed an imprecise initial condition. It should be stressed that this intuitive tendency would not have been observed if the formerly described interactivity between the model's variables was discarded.

4.2 Imprecise Epidemic Properties

In this section we consider the model given by Eq. (2) with imprecise initial conditions given by Eq. (8), but in addition we assume that also ν_0 is only known imprecisely. The imprecision related to this parameter can be taken into account by representing it as an interval in $[0, 1]$. For that purpose, we choose $\nu_0 = [0.2, 0.5]$. Figure 3 visualizes the length of the interval $U(c_j, t)$, denoted $|U(c_j, t)|$ two, five and ten time steps after an epidemic struck c_m. Comparing Figs. 2(b), 2(d) and 2(f) on the one hand, and Fig. 3 on the other hand, one clearly sees that $U(c_j, t)$ is considerably larger when both the initial condition and ν_0 are only imprecisely known. We verified that the maximum attainable interval length spanned the entire unit interval in polygons more distant from c_m as the wavefront propagates, which can be attributed to the successive non-interactive multiplication of ν_0 and $U(c_j, t)$. Nevertheless, $U(c_j, t)$ tended to a crisp number as $t \to \infty$ since the ICCA evolves towards its fixed point $(0, 1)$.

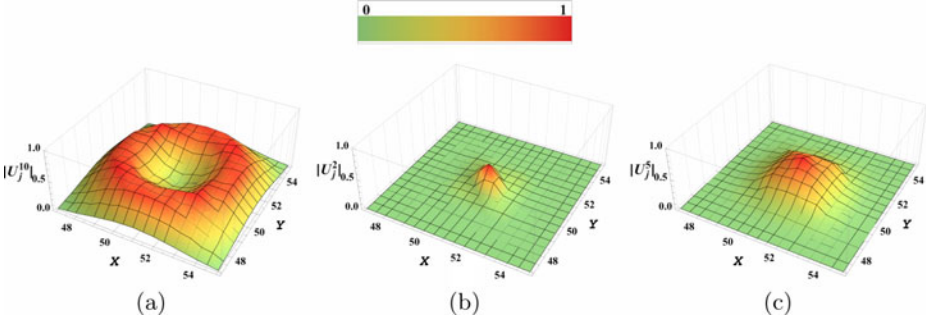

Fig. 3. Length of the interval $U(c_j, t)$ two (a), five (b) and ten (c) time steps after an epidemic broke out in the center polygon c_m of a square tessellation with an initial magnitude given by Eq. (8), and $\nu_0 = [0.2, 0.5]$

5 Discussion

Notwithstanding it was shown in the previous section that the proposed ICCA provides a means to deal with the imprecise nature of an epidemic in terms of the size of its initial magnitude and its properties, we must emphasize that the proposed paradigm is still to be improved in such a way that the quantities enclosed in a given interval are assigned a possibility with which they occur. Then, the impreciseness would no longer be represented by an interval that

merely encloses all possible values, but by a so-called fuzzy interval as depicted for illustration in Fig. 4. Unavoidably, this brings with it a complication of the calculations involved that then should be done in the light of Zadeh's [29] extension principle making that approach computationally less efficient than an ICCA. Naturally, an ICCA is trivially efficient since it merely requires two parallel model simulation, one for each interval limit, whereas a multiple of them would be required by MC methods. Off course, the additional computational effort enables to treat uncertainty in a much more informative way. Yet, by relying on fuzzy intervals one could combine an efficient numerical recipe, which would still not demand as many model simulations as needed for MC methods since Nguyen's [18] theorem can be invoked, with a model output bearing a much higher information degree.

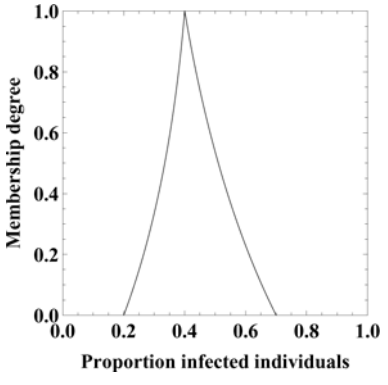

Fig. 4. An exemplary fuzzy interval in which every quantity in $[0, 1]$ is assigned a grade of membership to the set of infected individuals

6 Conclusions

In this paper we showed that imprecise information, described by means of intervals, can be used easily within a spatially explicit epidemic spread model that is based upon the continuous CA paradigm. More precisely, we demonstrated that uncertainty arising from both imprecise initial conditions or an epidemic's properties can be taken into account straightforwardly. The presented modeling framework is perfectly suited to cope with the growing importance and availability of spatio-temporal data. In forthcoming work, we will extend the presented model to cover also recovered individuals such that it can serve as a full-fledged alternative to PDE-based models. Besides, instead of representing imprecision by means of intervals, fuzzy numbers could be employed if there is information on the possibility with which every element in the interval occurs.

References

1. Baetens, J., De Baets, B.: Cellular automata on irregular tessellations. Chaos, Solitons Fractals (2010) (submitted)
2. Baetens, J., De Baets, B.: A generalized coupled-map lattice to model biological phenomena. Mathematical Biology (2010)(submitted)
3. Baltzer, H., Braun, P., Köhler, W.: Cellular automata models for vegetation dynamics. Ecol. Modell. 107, 113–125 (1998)
4. Beale, L., Abellan, J., Hodgson, S., Jarup, L.: Methodologic issues and approaches to spatial epidemiology. Environ. Health Perspect. 116, 1105–1110 (2008)
5. Dewdney, A.: Sharks and fish wage an ecological war on the toroidal planet. Sci. Am. 251, 14–22 (1984)
6. Doran, R., Laffan, S.: Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible-infected-recovered cellular automata model. Prev. Vet. Med. 70, 133–152 (2005)
7. Dubois, D., Prade, H.: Possibility Theory: an Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
8. Gardner, M.: Mathematical games: The fantastic combinations of John Conway's new solitaire game 'Life'. Scientific American 223, 120–123 (1971)
9. Kaneko, K. (ed.): Theory and Applications of Coupled Map Lattices. John Wiley & Sons Ltd., Chichester (1993)
10. Kier, L., Seybold, P., Cheng, C.: Modelling Chemical Systems using Cellular Automata. Springer, Dordrecht (2005)
11. Lakshmikantham, V., Mohapatra, R.: Theory of fuzzy differential equations and inclusions. In: Agarwal, R., O'Regan, D. (eds.). Series in Mathematical Analysis and Applications, vol. 6. Taylor & Francis, New York (2003)
12. Lipsitch, M., Riley, S., Cauchemez, S., Ghani, A.C., Ferguson, N.M.: Ferguson: Managing and reducing uncertainty in an emerging influenza pandemic. N. Engl. J. Med. 361, 112–115 (2009)
13. Mallet, D., De Pillis, L.: A cellular automata model of tumor-immune system interactions. J. Theor. Biol. 239, 334–350 (2006)
14. Milne, J., Fu, S.: Epidemic modelling using cellular automata. In: Proc. ACAL 2003, Canberra, pp. 43–57 (December 2003)
15. Moore, R.: Interval Analysis. Prentice Hall, Englewood Cliffs (1966)
16. Murray, J. (ed.): Mathematical Biology: I. An Introduction, 2nd edn. Springer, Berlin (1993)
17. Murray, J. (ed.): Mathematical Biology: II. Spatial Models and Biomedical Applications, 3rd edn. Springer, Berlin (2007)
18. Nguyen, H.: A note on the extension principle for fuzzy sets. J. Math. Anal. Appl. 64, 369–380 (1978)
19. Oberguggenberger, M., Pittschmann, S.: Differential equations with fuzzy parameters. Math. Comput. Modell. Dyn. Syst. 5, 181–202 (1999)
20. Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer, Berlin (2003)
21. Picioreanu, C., van Loosdrecht, M., Heijnen, J.: Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. Biotechnol. and Bioeng. 58, 101–116 (1998)
22. Pizarro, G., Griffeath, D., Noguera, D.: Quantitative cellular automaton model for biofilms. J. Environ. Eng. 127, 782–789 (2001)

23. Preziosi, L.: Cancer Modelling and Simulation. Chapman & Hall, Boca Raton (2003)
24. Qi, A., Zheng, X., Du, C., An, B.: A cellular automaton model of cancerous growth. J. Theor. Biol. 161, 1–12 (1993)
25. Schiff, J.: Cellular Automata: A Discrete View of the World. John Wiley & Sons Ltd., Chichester (2008)
26. Ulam, S.: The monte carlo method. J. Am. Stat. Ass. 44, 335–341 (1949)
27. von Neumann, J., Burks, A.: Theory of Self-Reproducing Automata. University of Illinois Press, Champaign (1966)
28. White, S., del Rey, A., Sanchez, G.: Modeling epidemics using cellular automata. Appl. Math. Comput. 186, 193–202 (2007)
29. Zadeh, L.: Fuzzy sets. Inf. Control 8, 338–353 (1965)