Mathematical Biosciences 253 (2014) 63–71

Contents lists available at ScienceDirect

Mathematical Biosciences

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

On the derivation of approximations to cellular automata models and the assumption of independence

K.J. Davies *, J.E.F. Green, N.G. Bean, B.J. Binder, J.V. Ross

School of Mathematical Sciences, University of Adelaide, South Australia, Australia

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form 10 April 2014
Accepted 15 April 2014
Available online 24 April 2014

Keywords:
Cellular automata
Continuum approximations
Agent-based simulation
Motility and proliferation

ABSTRACT

Cellular automata are discrete agent-based models, generally used in cell-based applications. There is much interest in obtaining continuum models that describe the mean behaviour of the agents in these models. Previously, continuum models have been derived for agents undergoing motility and proliferation processes, however, these models only hold under restricted conditions. In order to narrow down the reason for these restrictions, we explore three possible sources of error in deriving the model. These sources are the choice of limiting arguments, the use of a discrete-time model as opposed to a continuous-time model and the assumption of independence between the state of sites. We present a rigorous analysis in order to gain a greater understanding of the significance of these three issues. By finding a limiting regime that accurately approximates the conservation equation for the cellular automata, we are able to conclude that the inaccuracy between our approximation and the cellular automata is completely based on the assumption of independence.

1. Introduction

Cellular automata (CA) are discrete agent-based mathematical models that allow for an individual agent's behaviour to depend upon the state of its neighbourhood. As such they are often an ideal tool for modelling discrete systems composed of interacting individuals. A discipline in which they have seen widespread use is cell biology, where they have been used, for example, to model and understand processes such as tissue and tumour growth [5,16,17,24,26], and wound healing [8,19]. We focus herein on CA models appropriate for such applications, in which the biological processes are cell *motility* and cell *proliferation*.

As the cell biological processes that we seek to understand are likely to be evolving continuously in time, we believe a continuous-time model to be most appropriate. However, the literature extensively considers discrete-time CA models [1,4–6,11,18,22,23], in particular when deriving approximations to the average behaviour of these processes, and so we begin our analysis with discrete-time CA models in Section 2 before considering continuous-time CA models in Section 3.

Of much interest from both a practical and theoretical perspective, is the derivation of approximations which capture the average behaviour of the CA. Such continuum models might allow for new insight and understanding of these important biological processes. It has been shown by numerical experiments that these continuum approximations are only valid under restrictive conditions on the probabilities of cell movement and proliferation (that is, where cell movement dominates cell proliferation) [22], limiting the range of scenarios and applications which may be considered [2,4,7,15,18,22].

A careful analysis of the development of existing continuum approximations is undertaken in this paper. This gives insight into the implicit assumptions regarding the magnitudes of the motility and proliferation probabilities which underlie their derivation, and shows how new approximations can be developed when these assumptions are relaxed.

We show that the assumption of independence between the state of different sites in the CA is a key issue with regards to the inaccuracy of existing continuum approximations, so long as proliferation is present; and that, when there is no proliferation, the approximation obtained by assuming independence is identical to that found when the independence assumption is relaxed. However, the earlier continuum models are shown to perform unexpectedly well in approximating the behaviour of the CA even when proliferation is included. We show that this is largely due to a fortuitous cancellation of errors.

* Corresponding author. Tel.: +61 8 8313 1606.
E-mail address: kale.davies@adelaide.edu.au (K.J. Davies).

http://dx.doi.org/10.1016/j.mbs.2014.04.004
0025-5564/

2. Discrete-time CA model

2.1. Defining the model

We begin by introducing our one-dimensional CA model. The model is a lattice-based system in which each site on the lattice is in one of two states: occupied or vacant. Each occupied site contains an agent whose behaviour is determined by some process. Two processes will be considered in our model, motility and proliferation.

We say that the cell size and hence the lattice spacing is Δx, and so the position of the ith lattice site is $x_i = i\Delta x$ for $i = 1, 2, \ldots, X$. For simplicity, we use periodic boundary conditions, resulting in a connection between site 1 and site X. We define a time step Δt and consider the state of the process at discrete times $t_k = k\Delta t$ for $k = 0, 1, \ldots$.

Consider an agent at site i on the lattice who is to undergo a motility or proliferation event. In both cases, an adjacent site is chosen uniformly at random, that is, site $i-1$ or site $i+1$. If the adjacent site is vacant then the event will be carried out, otherwise the event will be aborted [9]. In the case of motility, the agent will move from site i to the new site, resulting in site i becoming unoccupied and the new site becoming occupied. In the case of proliferation, both the new site and site i will become occupied.

There are many ways in which these events can and have been implemented, [3,12,22,23,25,26]. Each of these different implementations can produce different average results, although in most cases the differences in the average CA data are minor. For example, in many of these models, the order in which events take place in the model is arbitrarily chosen, unmotivated or in some cases not made clear. The importance of this will be discussed in greater detail in Section 3.1.

For the purposes of comparison, we use the following implementation as outlined in Simpson et al. [22]. We choose $2N(t_k)$ agents uniformly at random with replacement, where $N(t_k)$ is the number of agents in the system at time t_k. The first $N(t_k)$ agents are each given the opportunity to perform a motility event. The probability of each agent performing the event is P_m. If the motility event is aborted, this is still regarded as an event taking place (and similarly for proliferation events). The remaining $N(t_k)$ agents are then given the opportunity to perform a proliferation event, each with probability P_p. The time step is completed by moving from t_k to t_{k+1} and this procedure is repeated. A realisation of the model can be seen in Fig. 1.

In order to analyse the mean behaviour of the system described above we consider the ensemble average occupancy at position x_i at time t_k, denoted C_i^k. This value can be calculated numerically by averaging the occupancy over many realisations of the CA system. This allows us to easily validate the results obtained when deriving continuum approximations.

2.2. Deriving a continuum approximation

We aim to derive a partial differential equation (PDE) for the ensemble average occupancy. However, it is axiomatic that a continuous PDE model is only likely to provide a reasonable approximation of the discrete system when large numbers of agents are present, and the average occupancy of sites varies over length and time scales which are much larger than the agent size (Δx) and time step (Δt). In order to derive a continuum model, we must hence identify these characteristic macroscopic length and time scales for the system [14]. For the length scale, denoted L, it would be natural to take the size of the region of the domain initially occupied by cells whilst for the time scale, denoted T, it could naturally be the population doubling time or the average time taken for a cell to move a distance L.

We wish to approximate the ensemble average occupancy, C_i^k, of the CA model with the continuous function $C(x,t)$, such that $C_i^k \approx C(x_i, t_k)$, where $x_i = i\Delta x$ and $t_k = k\Delta t$, and desire that this provides a reasonable approximation when the occupancy varies over the macroscopic scales. We hence assume that the ratios of the micro- and macroscopic length and time scales are small, *i.e.*

$$\epsilon = \frac{\Delta x}{L} \ll 1, \quad \delta = \frac{\Delta t}{T} \ll 1,$$

and exploit this separation of scales to derive the PDE model.

We now consider the change in the ensemble average occupancy of site i, C_i^k, from time t_k to time t_{k+1}. Assuming that the state of each lattice site is independent of the state of every other lattice site, we obtain the following discrete conservation equation

$$\begin{aligned} C_i^{k+1} &= C_i^k + \left(\frac{P_m}{2} + \frac{P_p}{2}\right) C_{i-1}^k (1 - C_i^k) + \left(\frac{P_m}{2} + \frac{P_p}{2}\right)(1 - C_i^k) C_{i+1}^k \\ &\quad - \frac{P_m}{2}(1 - C_{i-1}^k) C_i^k - \frac{P_m}{2} C_i^k (1 - C_{i+1}^k) + \text{HoT} \\ &= C_i^k + \frac{P_m}{2}(C_{i-1}^k - 2C_i^k + C_{i+1}^k) + \frac{P_p}{2}(1 - C_i^k)(C_{i-1}^k + C_{i+1}^k) + \text{HoT}, \end{aligned} \tag{2.1}$$

where $i = 1, \ldots, X$ and HoT denotes higher order terms.

Eq. (2.1) says that the new average occupancy of each site will be the old average occupancy plus some terms that describe the change over that time step. It is derived by considering all different possible transitions into and out of site i. Consider, specifically, deriving the probability of an agent at site $i-1$ moving to site i. An agent has a probability of motility P_m and has probability of $1/2$ of moving in a given direction. Further, C_{i-1}^k is the probability of site $i-1$ being occupied at time t_k and $(1 - C_i^k)$ is the probability of site i being vacant at time t_k. The term $\frac{P_m}{2} C_{i-1}^k (1 - C_i^k)$ can be obtained by assuming independence between sites and taking the product of each of these probabilities.

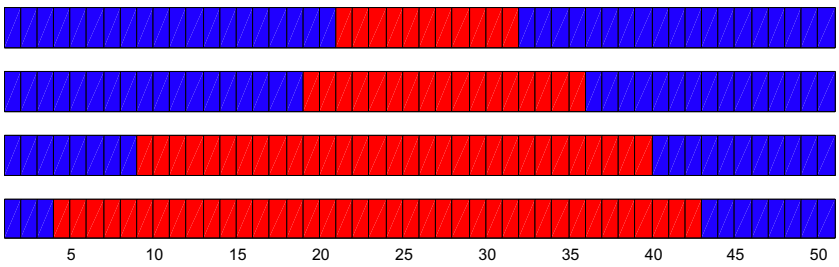

Fig. 1. A single realisation of the CA model progressing with time. Each row (from top to bottom) corresponds to $k = 0, 1000, 3000, 5000$ where k is the number of time steps since the start of the realisation. This process is purely proliferation, with $P_m = 0$ and $P_p = 1/200$. This realisation contains $X = 50$ sites of which 11 are initially occupied by agents.