Cellular Automata

11. 1 From the 'mean field' to the individual-based approach in Ecology

At the heart of much of the 20th century's ecological theory lies the assumption that individual organisms interact with each other *locally* in proportion to their *average abundance* across total space. The effect of neighbours over an individual is assumed proportional to their density averaged across a large spatial domain. This practice has been applied in physics and chemistry since long, examples being collisions between molecules in a well mixed-gas (in ecology molecules are likened to individuals) or the electrical field experienced by electrons within an atom – all electrons are assumed to be locally surrounded by the same electric field – called the "mean field". This is why sometimes in Ecology this assumption is also known as the mean-field assumption.

When is this mean-field assumption a good approximation ? We can think of a few examples. When organisms are highly mobile over the all space and we can hardly think of them as having lasting neighbours; when organisms interact with others over large enough neighborhoods, just as well as they interact locally. Another example is when physical forces exist that cause a strong mixing of organisms, as with plankton in the ocean.

Often conditions in the real world depart from these. A lack of mixing generates neighborhoods around individuals that deviate from spatial averages. Indeed, most of the time organisms are more likely to interact with their neighbours, and it is the density of these neighbours that matters, not the density averaged over some large spatial region. The presence of a nonrandom spatial pattern, for which there is plenty of evidence in nature, will lead to major departures from the mean-field assumption. When individuals are clustered, for example, most individuals within the cluster experience a density which is very different than the global average. Local experience can be quite different from the mean environment averaged across the entire habitat. This is true for animals and particularly true for plants in terrestrial habitats. These are relatively immobile in space and only interact with neighbours over restricted distances. Dispersal and propagules does less than one might expect to compensate for such immobility because most seeds do not travel far even when having structures that aid dispersal. Such departures from the mean field have different local effects on vital rates and, the resulting overall dynamics can be quite different from the one predicted by the mean-field assumption.

To construct spatio-temporal models that account for deviances from the mean-field assumption, researchers have to resort to alternatives to the conventional differential-equations-based models taught at undergraduate levels. Most popular among biologists are spatially explicit, *individual-based models*, implemented on computarised *cellular automata (CA)*. An explanation of these terms follows.

Individual-based models explicitly represent individuals one by one or, if in *sensu latu*, represent clusters of individuals within which mean-field applies. For each individual, the model explicitly considers probabilities of birth, death, growth, movement, and interaction with other individuals (both neighbours and non-neighbours). If the model is discrete in time, as is most often the case, these probabilities apply to a given time unit, the so-called model time-step. If the population is large, the total number of biological 'happenings' per time-step

is yet larger, and the entire process needs to be computarised in order to keep track of the fate of every individual in the population throughout time.

In individual-based models, typically each individual at each time step can be at one of a finite number of possible states of a so-called state-variable. An example of a state-variable is body size. Possible states of this variable are 'small', 'medium', 'large'. Age is another common state-variable. A very common state-variable in epidemiological individual-based models is the state of an individual in relation to a pathogenic agent. Examples of states could be 'susceptible to infection', 'infected', and 'immune to infection' (in analogy with the SIR model). In any of the above examples, individuals change state as time goes by, according to rules pre-established by the biologist. Most of the time, these rules have to do with time elapsed since an individual was born, interactions with other individuals, and random events.

A particular type of individual-based models is the one that explicitly takes into account the distribution of individuals over space. Space is viewed as a state-variable and all possible locations where individuals can be are states of this state-variable. At each time-step, individuals can either move between locations or remain immobile. In any case, each individual will be positioned in relation to others so that we can think of them as close neighbours, not-so-close neighbours, non-neighbours, etc. If the population under consideration is large, visualization of the relative position of every individual can be quite difficult, and that is where cellular automata become helpful.

11.2 Cellular Automata (CA)

Cellular automata modeling, also known by the term *grid-based* modeling, or *regular lattice* modeling, begins by subdividing the (bi-dimensional) area being modeled (the area occupied by an ecosystem, the cells in a tissue, a country, a city, etc.) into a grid (= regular lattice) of cells of some geometric shape (Fig 1), where each cell is occupied by an individual or groups of individuals. A cell may also be empty space, either because no individuals moved there, or because it is meant to represent space that cannot be occupied. When cells represent groups of individuals, their states may be described quantitatively (e.g. number of individuals, amount of biomass) or qualitatively (e.g. presence or absence of a species).

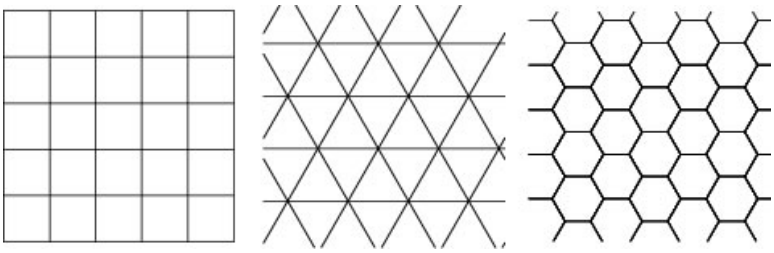


Fig. 1 Examples of grids used in CA's. A grid usually refers to the intersection of two or more sets of evenly-spaced parallel lines at particular angles to each other in a plane, yielding cells with different shapes and different neighborhoods.

Next, the biologist decides which state-variables will be used to describe each cell (age, sex, state in relation to a pathogen…), what are the possible states of each state-variable, the length of the time-step (1 hour, 1 day, 1 year …) used when running the model, and the rules for changes of the individual states during a time step. There are no recipes for designing a CA. The basic method is very simple, but the way it applies differs from model to

model. All decisions – namely, choosing the shape and number of cells, the possible states of a cell, the time step, and the rules for state change – must be adapted to the problem addressed by the model and to the available biological information concerning the phenomena under study. Almost no mathematics is needed for this type of modelling, at least at this level of approach. Thus the modeler's mathematical skill is less important than his/her biological understanding of the problem and the biological information available. Let us look into a bit more detail at each modeling step.

The first modeling step is to decide the number of cells in the grid. If each cell represents an individual, then the number of cells is equal to or greater than the population size itself. If cells represent clusters of individuals (packs, schools, towns…) and empty spaces, then there should be a close analogy between the number of cells and what is known about space topography and population distribution over space. Cell shape can be decided based upon the number of neighbours we want each cell to have and their relative positions (Fig. 1). In two-dimensional CA's, square cells are probably the most common choice. Whatever happens to a cell during a time step usually depends on the state of the neighbouring cells. We thus have to choose the neighborhood in which cells can interact. If the cell is a square, the simpler neighborhood consists of the 4 nearest neighbours, the so-called *von-Neumann neighborhood*.

Initial conditions and updating rules

Not only the boundary conditions and the neighborhood need to be specified, but also the initial conditions. Often initial conditions determine the subsequent evolution of the CA and must be carefully chosen. Initial conditions can be very special (constructed), or they can be generated randomly. For example, if the CA is meant to simulate a predator-prey interaction, initial conditions will be the initial number of individuals of both species and their spatial distribution. Predators and prey don't usually distribute randomly over space, so some previous knowledge of how they agglomerate and their average distances are advisable. In the extreme case where every prey is placed in the neighborhood of a predator, the prey population is likely to become depleted so fast that predators will then later starve and become extinct.

The change in state (or states, if there is more than one state-variable) of a cell is governed by the updating rules of the CA. It depends on the current state of the cell as well as on the states of its neighbours. CA models can be updated synchronously or asynchronously. In synchronous models, at each time step all cells are simultaneously updated according to the CA rules. In asynchronous models, at each step a single cell (or set of cells) is selected at random and changed according to the CA rules.

Updating rules are often probabilistic and correspond to events like birth, death, migration within the grid, short or long-distance contacts, etc. Migration is commonly simulated by moving a selected individual in a random direction, either by moving it to an empty cell or by swapping states with another occupied cell. The distance covered at each time-step is the modeler's choice, with greater distances usually meaning greater mixing among individuals. Details on updating rules obviously depend very much on the phenomena being modeled and are therefore better illustrated with specific examples.

11.3 Infectious disease spread

CA's can simulate the dynamics and spreading of infectious diseases in human populations. Suppose that an individual can be unambiguously classified in relation to the state-variable "how does he/she relate to a pathogenic microorganism" in three possible states: S= susceptible to get the infection, I= infected and infectious, R= recovered from infection and immune. Now consider the following possible changes of state and their rules:

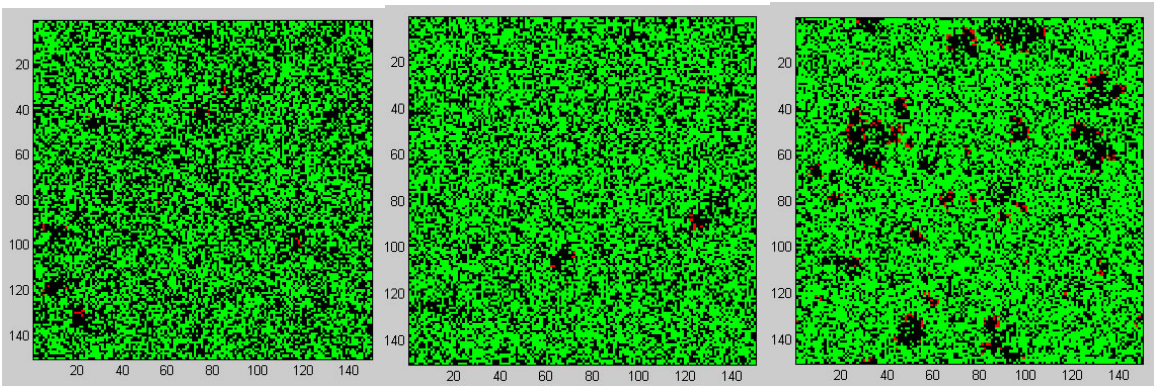
S → I: a susceptible becomes infected with probability q, if an infectious individual is within its neighborhood. A susceptible may also become infected if it is contacted by a non-neighbor, with (a lower) probability p.

I → R: an infectious individual recovers and becomes immune if a given number of time-steps has elapsed since infection took place.

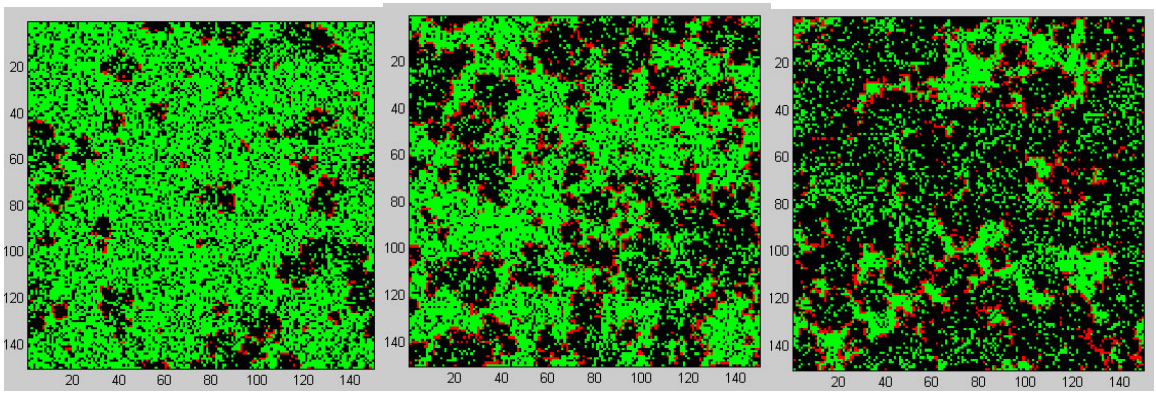
R → S: a recovered individual loses immunity and becomes susceptible again, if a given number of time steps has elapsed since recovery took place. This state change can also take place if an R cell is selected for a newborn to appear (see below), as newborns are considered susceptible to infection.

All the probabilities and number of time-steps mentioned above may be chosen to mimic a particular infectious disease. The state-variable and the rules could be made more complicated if appropriate. A common complication is to add an extra state, latency (L), where an individual is infected but not yet able to transmit the infection to others (infected but non-infectious). Thus a logical sequence of changes would be S→ L, L→I, I→R. Another possible complication would be to allow recovered individuals to become infected with a lower probability than susceptible individuals. This would mimic diseases where recovery from disease confers only partial immunity against re-infection, such as tuberculosis and influenza.

In addition to state changes taking place synchronously every time-step, there are also individuals being born and individuals dying. If the number of individuals in the population is to stay constant, the average number of newborns should equal the average number of deaths per time-step. A common procedure to generate newborns is to randomly select a cell in the R state and decide whether a newborn will take place there with a given birth probability. Deaths may take place by randomly selecting individuals of any state and "killing" them with a given probability too; alternatively, the CA may keep track of the individuals'age and deaths take place once individuals reach mean population longevity.



t = 15 t = 20 t = 25



t = 30 t = 35 t = 40

Figure 2. Snapshots of a CA that simulates the spread of an infectious disease, at selected time-steps (t). The CA is based on a grid with 150^2=22500 cells that represent individuals. These can be in 3 states: susceptible to the disease (green cells), infected (red), and immune (black). The disease begins spreading from a small number of infectious foci (t=15, 20, 25) and spreads in epidemic waves that sweep the entire population (t = 30, 35, 40).

Figure 2 illustrates the spreading of an infectious disease in a CA with 22500 cells that represent individuals. In this example, individuals do not move, but there are random contacts with non-neighbors with lower probability than with neighbors. These long-distance contacts play the same role as random movements of individuals.