A Geographic Automata System for Modelling Disease Outbreaks

S.W. Laffan [1], E. Lubarsky [1], M.P. Ward [2] and L.D. Highfield [2]

[1] School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia.
[2] Department of Veterinary Integrative Biosciences, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, Texas, USA
Email: Shawn.Laffan@unsw.edu.au

Keywords: *Geographic automata, Spatial modelling, Susceptible-Infected-Recovered models, disease mapping*

EXTENDED ABSTRACT

Management response to contagious disease outbreaks requires rapid application of appropriate control measures (for example, quarantine and vaccination), but with little current or empirical data about how the disease may spread. This means that modelling is one of the few available options for making important decisions, and geographical modelling is a key approach.

While there are many models that deal with disease spread within livestock populations, there are few that deal explicitly with widely dispersed animal populations such as unfenced livestock and wild and feral animals. It is for this purpose that we have developed a geographic automata model, Sirca, to simulate possible outbreaks of disease in wild and unfenced animal populations.

The application of the model to potential FMD outbreaks in deer, pigs and cattle in a region of Texas, USA, has been described by Ward et al. (2007). In this paper we describe the structure and implementation of the model.

The model is implemented using a geographic automata (GA) framework to enable a better representation of the geographic distribution of animal populations than the lattice structure enforced by cellular automata (CA). Beyond this, the model has a number of features that enable flexibility and speed. First, the user is able to define arbitrarily complex neighbourhoods to determine the possible interactions between infectious and susceptible animal groups. Second, it uses a spatial index approach to reduce the computational burden of finding possible neighbours. Third, it uses an events register to both schedule and track changes in model state over time. It is worth noting that many of these advantages could be equally applied to a CA approach, reflecting the fact that GA are generalisations of CA.

The model has already been used for FMD scenarios, but can be applied to any disease outbreak which has geographic dispersal (eg. Classical Swine Fever, Avian Influenza).

1. INTRODUCTION

Management response to contagious disease outbreaks requires rapid application of appropriate control measures (for example, quarantine and vaccination), but with little current or empirical data about how the disease may spread. This means that modelling is one of the few available options for making important decisions, and geographical modelling is a key approach.

While considerable effort has been spent in modelling animal disease spread via livestock movements (e.g. Keeling et al., 2001, Morris et al., 2001, Bates et al., 2003), the disease threat from wild animal (feral and native) and unfenced livestock populations that might form disease reservoirs has received comparatively little attention. In many regions around the world, such uncontrolled animal populations can represent a considerable proportion of the entire susceptible animal population for a variety of diseases. One key example is foot and mouth disease (FMD). Susceptible populations include cattle, sheep, deer and feral pigs. Animal systems that include large and interconnected susceptible populations exist, particularly in Australia and the USA (Doran and Laffan 2005, Ward et al. 2007). Such populations can exacerbate outbreaks, form reservoirs of disease, and might allow FMD to become endemic within countries previously free of disease.

Disease outbreaks are inherently spatial phenomena. Non-linear dynamics exist as a result of the effects of spatial interactions between susceptible animals or groups of animals, and how these relationships change over time. These characteristics make artificial life models, for example cellular automata (CA), an attractive approach to modelling disease outbreaks, particularly when combined with a state based Susceptible-Infected-Recovered framework (Doran and Laffan, 2005). CA are models of physical systems that incorporate spatial relationships, treating space and time as discrete units. The repeated application of simple transmission rules based on neighbouring population densities allows modelling of the spatio-temporal dynamics of an outbreak from one or more initiation locations. In addition, management actions occurring at any time step can be incorporated into this modelling framework.

The regular lattice structure of the CA makes the spatial component of the model computationally efficient, as it is a simple process to determine the geographic neighbours for disease transmission. Such models are readily developed as extensions to GIS software using raster data structures (eg Doran and Laffan, 2005). However, the representation of population distributions using a lattice structure is also a fundamental limitation of the approach. It is an extremely rare animal population that is distributed along an equally spaced lattice or grid, especially across the regional extents over which outbreaks occur (eg. Sutmoller 2000). Geographic automata (GA) do not suffer from this representational limitation. GA are a generalisation of CA, having the same rules about interactions and transitions between disease states as an equivalent CA, but the underlying data structure is not constrained to a regular lattice. Instead, it can be any geographic data structure used to represent discrete objects.

To this end, we have developed a geographic automata model, Sirca, to simulate possible outbreaks of disease in wild and unfenced animal populations. The name Sirca reflects its origins as a Susceptible-Infected-Recovered CA. Subsequent developments to be a Susceptible-Latent-Infectious-Recovered GA lacked the requisite acronymic melody.

The application of the model to potential FMD outbreaks in deer, pigs and cattle in a region of Texas, USA, has been described by Ward et al. (2007). In this paper we describe the structure and implementation of the model. We give examples developed using the same data set as Ward et al. (2007), where pig, deer and cattle population data were estimated using habitat modelling approaches.

2. THE MODEL

The Sirca model is implemented as a series of object oriented libraries written in the Perl programming language. It can be accessed through a graphical user interface (GUI, written in Python) or as a series of library functions called from a script.

The model represents the geographic distribution of susceptible animals in a hierarchy of three levels: *Groups*, *Populations* and *Landscapes*.

The *Group* is the fundamental unit in the model. It represents a functional group of animals, for example a herd of cattle or a flock of sheep. At a minimum, each group contains data about its location, current disease state and population size; however, other model parameters can be stored for each individual group when such variation is needed. Each group also has a unique identifier based on its geographic coordinates.

Each group in the model is in one of four disease states: susceptible, latent infected, infectious, or recovered (resistant). The initial change in state from susceptible to latent is determined through interactions with infectious groups, with the period in each subsequent state assigned randomly within user specified ranges. Recovered groups return to the susceptible state at the end of the specified period of disease resistance.

Groups are currently represented using a point data structure, but polygon and polyline could be used given development of appropriate routines to determine their geographic neighbours. These are standard GIS operations, and so do not represent a major challenge.

Populations consist of a set of geographically distributed groups of the same type, for example cattle herds and sheep flocks. This also allows separate modelling of different types of groups within a population, for example disease spread in sows with piglets can be modelled separately from boars. The disease state of each group within the population is tracked.

The *Landscape* level consists of a set of populations, and allows control over interactions between groups from different populations.

Most parameters can be stored at any level of the model, with the higher levels providing defaults if they are not specified at the lower levels. This allows geographic variation of these parameters. For example, the distance over which groups will interact with other groups, or the state transition periods, can be specified for some groups where it is known, and taken as a default value for others.

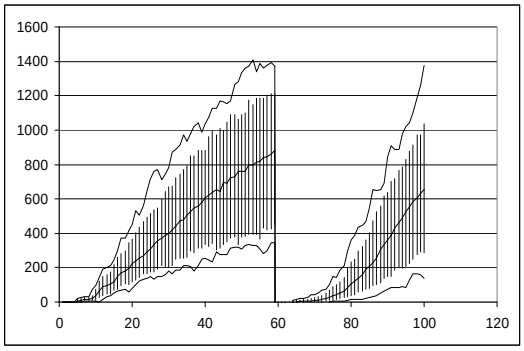

Figure 1. Epidemic curves of predicted median number of cattle in southern Texas, USA, following foot-and-mouth disease virus introduction in feral pig populations, using 100 Sirca model runs. There is a complete cull of cattle at day 60. Ninety percent prediction intervals are indicated using the vertical bars.

2.1. Model outputs

The model outputs a series of different data sets for individual model runs and series of model runs. These include epidemic curves (fig. 1), snap-shots (fig. 2) and animations of the model state changes. The changes in the model states and other properties are output in GIS compatible formats.

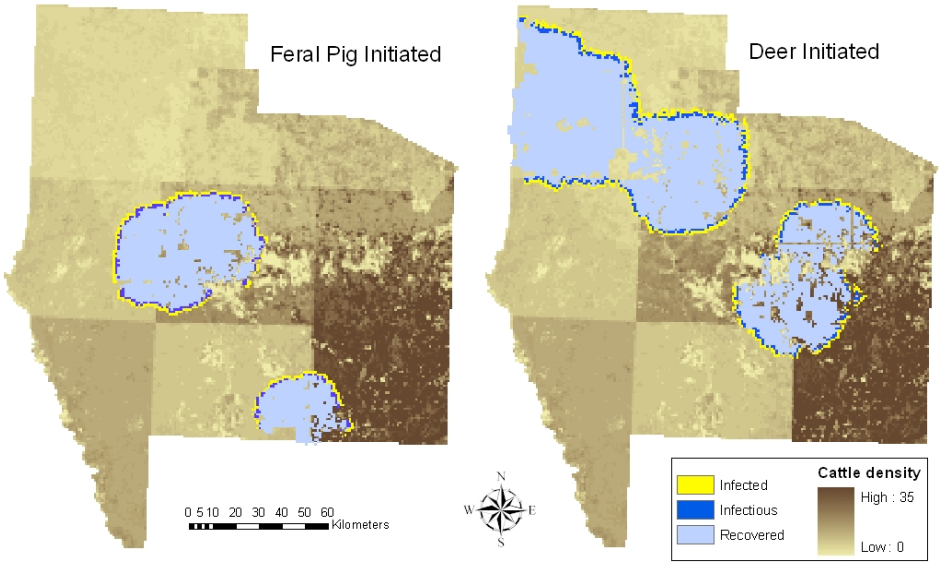

Figure 2. Snapshots of the spatial distribution of infected cattle populations in southern Texas, USA at day 100 following simulated incursions of foot-and-mouth disease virus in feral pigs (left) or wild deer (right) at five sites, using the Sirca model.

This last aspect allows for further spatio-temporal analysis using geocomputational tools, for example to characterise locations that will act as disease reservoirs in the event of incomplete animal depopulation, that represent locations where outbreaks have a higher chance of becoming established, or to derive rules that allow extrapolation of any trends in the predictions based on environmental correlates.

2.2. Disease transmission

The transmission of disease from an infectious to a susceptible group depends on the interactions between infectious groups and susceptible groups. Interactions occur between groups within and between each population in a landscape.

The probability of an interaction, and thus transmission, is determined by the relative densities of the groups, modified by a spatial kernel. Additional control is exercised by specifying the possible number of interactions at each time step.

Density is calculated as a linear function between user-defined minima and maxima. Group sizes below the minimum are considered to be too small for an interaction to occur, while those exceeding the maximum have sufficient animals that the interaction probability is maximised (Doran and Laffan, 2005). Functions other than linear could be used if justified.

An interaction occurs when a randomly generated value is below the interaction probability. The quality of the random number generators algorithm can have a serious impact on model results (Van Niel and Laffan 2003), and for this reason we use the Mersenne Twister algorithm (Matsumoto and Nishimura 1998). It has an extremely long period length before the sequence repeats ($2^{19,937}$-1) and good spectral properties where correlation structures within the random number sequence are very small.

2.3. Neighbours

The definition of which groups are neighbours is fundamental to the model. This is done using a set of user defined functions of arbitrary complexity. These are currently implemented using Perl syntax, primarily based on Euclidean distances along the X and Y axes, but allowing access to all other model parameters including time.

While complex for the novice user, this approach allows the neighbourhoods to change over time, for example through seasonal changes in animal behaviour, mortality, or due to the physical effects of a disease such as lameness.

Support for neighbourhood matrices is planned for a later development.

2.4. Mortality

Mortality at the group level is assigned randomly between an upper and a lower value using piecewise linear functions. This allows the definition of minimum and maximum rates of mortality during any period, with the piecewise linear functions allowing relatively complex functions if needed. In most cases a triangular distribution will be used where there is a single peak mortality at some timestep during the period the group is in the infectious state.

3. MODEL SPEED

While the GA approach has considerable advantages in terms of representing reality, it also incurs some computing speed penalties due to the more complex calculations required.

As a baseline, and as with any such model, the processing speed depends largely on the complexity of the input data sets and the number of interactions that must be assessed. For example, the models described in Ward et al. (2007), at 100 iterations each, take between one and two minutes on a Dell XPS-1210 with a 2.2 GHz dual core CPU and 2 GB of RAM. This analysis involved approximately 25,000 groups each of pig, deer and cattle, with five initiation locations. Speed is a direct consequence of the number of interactions that must be assessed, so will obviously decrease as larger outbreaks or more initiation locations are assessed.

Three approaches are used to improve speed in the Sirca model. As an indication of the relationship between GA and CA, it should also be noted that all three features have application to the standard CA approach. First, the number of interactions to be computed is reduced by using a "push" approach, where only neighbours of infectious groups are evaluated for disease transmission. Second, the model uses a spatial index to reduce the search times for possible neighbours. Third, an events register is used to control temporal changes. Approaches two and three are described below.

3.1. Spatial index

The determination of geographic neighbours is considerably more computationally intensive for GA than for CA. In the naïve case one must assess

all other groups in a population to determine if they are neighbours of a particular group. Speed is achieved in the Sirca model by using a single level spatial indexing scheme, similar to that described by Bithell and MacMillan (2006). The lattice structure of the spatial index allows easy calculation of the relevant elements in which to search for neighbouring groups. Further speed is achieved by caching the neighbours for each group once they have been determined.

Other index structures, such as quadtrees, could be used. These are very efficient when the problem involves determining the nearest neighbour. However, the arbitrary complexity of the neighbourhood definitions in Sirca counteracts the effectiveness of such hierarchical schemes since one must typically assess all terminal nodes in the tree to determine which index elements contain possible neighbours.

3.2. Events register

Temporal changes in the model, for example state changes (eg. latent to infectious) and mortality (natural and disease control), are handled using an events register indexed by time step, with one register stored for each population.

Two types of events are stored. Global events are applied to a set of groups, with examples including depopulation, random initiation of disease outbreaks, and long distance wind-borne spread. Group events apply to individual groups and include such events as mortality and transitions between states following infection.

The events scheduling system avoids the need to iterate over all groups to find those scheduled for changes. This also has the advantage that the events acts as a register of the changes over the temporal duration of the model run, allowing later reconstruction of the model without the computational overhead of recalculating neighbours, mortality and the like.

4. SUMMARY

The geographic automata approach contains nearly all the advantages of cellular automata for disease modelling, with the additional advantage that the underlying animal populations are not constrained to occur on a regular lattice. While the computational complexity of the GA approach makes them slower than an equivalent CA, the truer representation of the population surfaces is a considerable benefit. The speed penalties of the more complex data structure can be reduced considerably using simple computational approaches.

The object oriented structure of the code libraries means that other geographic data structures, such as lines and polygons, can be used if appropriate neighbourhood functions are developed. This would allow the incorporation of other important components of disease outbreaks, for example livestock transport.

The model has already been used for FMD scenarios, but can be applied to any disease outbreak which has geographic dispersal (eg. Classical Swine Fever, Avian Influenza).

5. ACKNOWLEDGEMENTS

This research was funded by a UNSW Faculty Research Grants Program award to Laffan.

6. REFERENCES

Bates, T.W., Thurmond, M.C. and Carpenter, T.E., (2001) Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. *American Journal of Veterinary Research*, 62 1121–1129.

Bithell, M. and MacMillan, W.D. (2006) Escape from the cell: Spatially explicit modelling with and without grids. *Ecological Modelling*, 200, 59–78.

Doran, R.J. and Laffan, S.W. (2005) 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. *Preventive Veterinary Medicine*, 70, 133–152.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. and Grenfell, B.T. (2001) Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294, 813–817.

Matsumoto, M. and Nishimura, T. (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computational Simulation*, 8, 3–30.

Morris, R.S., Wilesmith, J.W., Stern, M.W., Sanson, R.L. and Stevenson, M.A. (2001) Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain. *Veterinary Record*, 149, 137–144.

Sutmoller, P., Thomson, G., Hargreaves, S., Foggin, C. and Anderson, E. (2000) The foot and mouth disease risk posed by African buffalo within wildlife conservancies to the cattle industry in Zimbabwe. *Preventive Veterinary Medicine*. 44, 43–60.

Van Neil, K. and Laffan, S.W. (2003) Gambling with randomness: the use of pseudo-random number generators in GIS. *International Journal of Geographical Information Science*. 17 49–68.

Ward, M.P., Laffan, S.W. and Highfield, L.D. (2007) The potential role of wild and feral animals as reservoirs of foot-and-mouth disease. *Preventive Veterinary Medicine*, 80, 9–23.