A geographic automata system for modelling disease outbreaks in wild and unfenced animal populations

Shawn W. Laffan[1], Eugene Lubarsky[1], Michael P Ward[2], Linda D. Highfield[2]

[1]School of Biological, Earth and Environmental Sciences,
University of New South Wales,
Sydney, NSW, 2052, Australia.
Telephone: +61 2 9385 8093
Fax: +61 2 9385 1558
Email: Shawn.Laffan@unsw.edu.au

[2]Department of Veterinary Integrative Biosciences,
College of Veterinary Medicine & Biomedical Sciences,
Texas A&M University
College Station, TX 77845-4458, USA

1 Introduction

Knowledge of the potential spread of disease through animal populations is critical for planning for and management of a possible outbreak. While considerable effort has been spent in modelling the spread via livestock movements (e.g. Keeling et al., 2001, Morris et al., 2001, Bates et al., 2003), the threat from wild and unfenced animal populations has received comparatively little attention. This is despite their representing a considerable proportion of the susceptible animal population in many regions around the world, for a variety of diseases. One key example is cloven-hoofed animals, for which wild and unfenced populations can form reservoirs of foot-and-mouth disease or exacerbate the extent of an outbreak (Ward et al. 2007). Disease management in the presence of such populations represents a considerable challenge, so predictions are needed for any such effects. The Sirca geographic automata model has been developed for this purpose.

Disease outbreaks are inherently spatial phenomena, with non-linear dynamics based on the effects of spatial interactions between susceptible animals or groups of animals, and how they change over time. These characteristics mean that artificial life models like cellular automata (CA) are a good approach to modelling disease outbreaks, particularly when combined with a state based Susceptible-Infected-Recovered framework (Doran and Laffan, 2005). In these cases, the repeated application of simple rules based on neighbouring population densities allows the prediction of the possible spatio-temporal dynamics of an outbreak.

The regular lattice structure of the CA makes them computationally efficient, as it is a simple process to determine what the spatial neighbours are. However, this structure is also a fundamental limitation of the approach, as susceptible populations will rarely be distributed along an equally spaced lattice. Geographic automata (GA) do not suffer from this limitation. GA are a generalisation of cellular automata, having the same rules about interactions and transitions between states remain as in 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. In this case, the Sirca model uses a point representation.

The application of the model to potential FMD outbreaks in deer, pigs and cattle has been described by Ward et al. (2007). Here were concentrate on the model structure and implementation.

2 The model

2.1 Structure

The model is built using an object oriented geographic automata framework. It is developed in the Perl programming language and can be accessed as a series of library functions or through a graphical user interface.

The model represents susceptible populations at two levels. The fundamental unit is the "Group", representing for example a herd or flock. "Populations" consist of a set of groups of the same type or that have similar behaviour, for example cattle and sheep. This also allows the separation and different treatment of different types of groups in a population, for example sows with piglets can be treated separately from boars.

Each group in the model is in one of four states: susceptible, latent, infectious, or recovered. The initial change in state from susceptible to latent is determined through interactions with infectious groups, with the period in each subsequent state randomly assigned along a user specified range, with recovered groups returning to the susceptible state at the end of the specified period. Mortality is assigned randomly between an upper and a lower distribution functions.

The transmission of disease from an infectious to a susceptible group depends on the interactions between infectious groups and susceptible groups, with the probability of an interaction determined by geographic distances and weights regulating the interactions between different populations. The interaction probabilities between groups from different populations are adjusted by a set of user defined weights to represent factors such as avoidance behaviour or occupation of different landscape elements. 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 in model results (Van Niel and Laffan 2003), and in this case we use the Mersenne Twister algorithm (Matsumoto and Nishimura 1998).

2.2 Computational efficiency

The non-lattice data structure used in the GA approach results in increased computational requirements. We use two main approaches to improve the GA speed, largely based on a memory-speed trade-off.

The determination of geographic neighbours is considerably more computationally intensive for geographic automata than for cellular automata. In the naïve case one must assess all other groups in a population to determine if they are neighbours of a particular group. In this case, speed is achieved using a single level hash-based spatial indexing scheme, similar to that described by Bithell and MacMillan (2006). The lattice structure of the index allows easy calculation of the relevant index neighbours in which to search for neighbouring groups. Further speed is achieved by caching the neighbours for each group once they have been determined.

Temporal changes in the model, for example state changes and mortality (natural and culling), are handled using an events scheduler. This considerably reduces the

computational effort spent finding groups due for change at a time step, as one merely runs all events that are scheduled for that time step. These events can also be run for any level of the model structure, from an individual group through to the entire collection of populations. This approach is equally applicable to conventional CAs.

2.3 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, and GIS compatible data formats. 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 culling, 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.
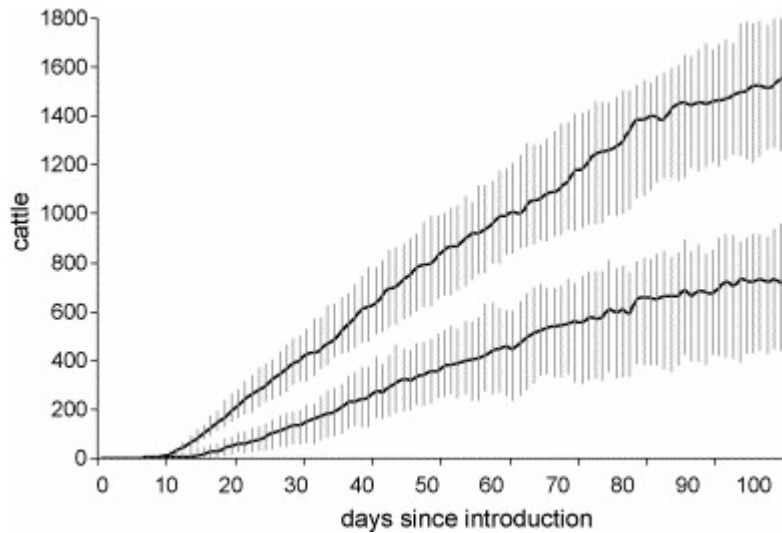

Figure 1. Epidemic curves of predicted median number of cattle in southern Texas, USA, following foot-and-mouth disease virus introduction in feral pig (lower curve) or wild deer (upper curve) populations, using 100 Sirca model runs. Ninety percent prediction intervals are indicated. Source: Ward et al. (2007).
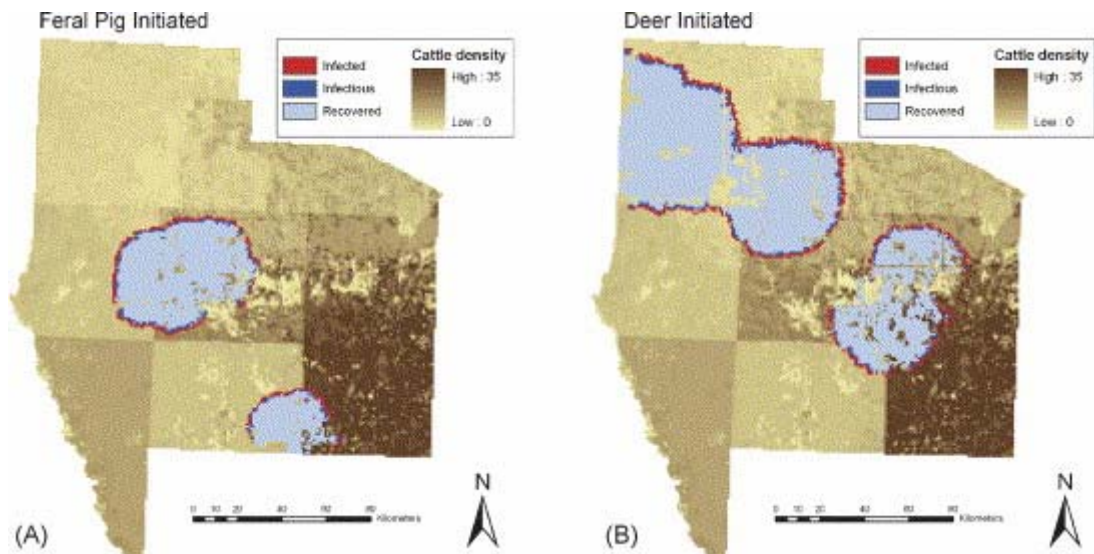
Figure 2. Snapshots of the spatial distribution of infected cattle populations in southern Texas, USA at day 100 following simulated incursions of foot-and-mouth disease virus in feral pigs (left) or wild deer (right) at five sites, using the Sirca model. Source: Ward et al. (2007).

3 Summary

The geographic automata approach contains 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.

4 Acknowledgements

5 References

Bates TW, Thurmond MC and Carpenter TE, 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 WD, 2006, Escape from the cell: Spatially explicit modelling with and without grids. Ecological Modelling, 200(1-2), 59-78.

Doran RJ and Laffan SW, 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 MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Cornell SJ, Kappey J, Wilesmith J and Grenfell BT, 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 RS, Wilesmith JW, Stern MW, Sanson RL and Stevenson MA, Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain. *Veterinary Record*, 149, 137–144.

Ward MP, Laffan SW and Highfield LD, 2007, The potential role of wild and feral animals as reservoirs of foot-and-mouth disease. *Preventive Veterinary Medicine*, 80, 9-23.

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