

How does Space alter the Formulation of Evolutionary Models?

Peter Whigham¹ & Grant Dick²

¹Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
Phone: +64 3 479-7391 Fax: +64 3 479-8311
Email: pwhigham@infoscience.otago.ac.nz

²Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
Phone: +64 3 479-5282 Fax: +64 3 479-8311
Email: gdick@infoscience.otago.ac.nz

Presented at SIRC 2006 – The 18th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
November 6th-7th 2006

ABSTRACT

This short paper discusses how space is introduced within evolutionary models and reviews several approaches from the literature. In particular, the Moran process is used as one type of model that has been described in several ways when space is introduced. The evolutionary operations of parental selection, offspring placement, synchronous and asynchronous updating, survival and death timings are shown to be factors that may be interpreted in a number of ways once space is introduced into a model. One conclusion from this discussion is that the introduction of space requires a consideration of the time-step in models, directly through event driven processes or indirectly through the random selection of space prior to a local process.

Keywords and phrases: evolutionary model, neighbourhood, Moran process, selection, updating policy

1.0 INTRODUCTION

Evolutionary models are characterised by a number of operations including: selection of individuals as parents; breeding operations such as mutation and crossover; replacement strategies for offspring; formulation as a generational or steady state model; genotype representation of individuals; fitness evaluation; and life history concepts such as time to death and rate of reproduction. The introduction of space to an evolutionary model is generally formulated as a grid-based approach such as a cellular-automata model, or as a network representing the connections between locations or individuals.

This short paper has been motivated by recent results from the authors (Whigham & Dick, 2006) which compared two formulations of a graph-based evolutionary model of the Moran process. This work showed that the resulting properties of the model varied significantly based on how the evolutionary process was interpreted in terms of processes constrained by the spatial structure.

2.0 A BRIEF REVIEW OF EVOLUTIONARY PROCESS MODELS AND SPACE

The Moran process was originally formulated as a simple model of stochastic evolution and is often used for determining fixation probabilities from an introduced mutant to a well-mixed population. For the Moran process the probability of fixation ρ of a mutant allele with relative fitness r (where all other individuals in the population have fitness 1) for a finite population size N is given by (Lieberman et al. 2005):

$$\rho = \frac{\left(1 - \frac{1}{r}\right)}{\left(1 - \frac{1}{r^N}\right)} \quad (1)$$

However, Lieberman et al. (2005) showed that for a variety of spatial structures, defined by a graph, that this probability of fixation was amplified. This result was particularly significant given the previous theoretical results of Maruyama (1974) that showed certain properties were independent of spatial structure (such as probability of fixation in a spatially-structured Moran process). Clearly to reconcile these two opposing results it is necessary to consider the formulation of the spatial component of the models that have been used and to consider under what circumstances either of these models is appropriate. In addition, a number of other spatial evolutionary models will be introduced to show how other authors in the field of theoretical population biology consider the introduction and resulting constraints that spatial structure introduces.

2.1 The Moran Process on a Graph

The original Moran process was based on a well-mixed model which therefore ignored spatial structure. The approach of Maruyama (1974) assumed that spatial structure could be modeled by an island-based approach with migration to allow gene flow between islands. In particular, all individuals had the same negative exponential lifetime, and when an individual died they were replaced by an individual born from the same island (colony). The model was an overlapping generational model and each island was treated mathematically as a separate panmictic population. In contrast Lieberman et al. (2005) considered a population defined by a graph, where the vertices of the graph were occupied by each individual and the edges of the graph defined the local neighbourhood of each individual. The translation of the Moran process to this graph model assumed that selection of an individual was proportional to the fitness of the entire (spatially-distributed) population, however placement of the offspring from a selected parent was governed by the weightings on the graph from the parent vertex.

Whigham and Dick (2006) showed that by reinterpreting the use of space and selection the results of Lieberman et al. (2005) were altered and subsequently agreed with the general finding of Maruyama (1974). The change in model structure was a simple one: rather than selecting individuals based on their global proportional fitness as the first step in the model, a location was randomly selected and then, based on the deme of this location, a parent was selected using local proportional fitness. This change in the model had the effect of reducing the selection intensity of individuals with relative fitness r , and therefore the amplification of fixation probability noted by Lieberman did not occur. Clearly how spatial structure is interpreted with selection can dramatically alter the properties of the modeled system.

This raises the question of how space should be incorporated in an evolutionary model: although there are many possible ways in which space can change evolutionary behaviour, when we consider a biologically plausible model there should be some principles upon which space and evolutionary models are developed. This short paper will examine some of the possible scenarios that relate spatial structure and evolutionary processes to begin understanding which, if any, of these scenarios are most appropriate. The question of what is a biologically accurate model will not be directly addressed, since this is potentially too complex and difficult for a short discussion paper, however some theoretical models will be considered that hopefully begin to suggest how such plausible models should be formulated.

2.2 Selection and Space

In the following discussion we will consider a population of haploid individuals that produce offspring that are copies of the parent. This model is similar to the simple Moran process that is often used with biological problems. However, the conclusions are likely to be applicable to more complex individuals with multiple parents with crossover and mutation.

How may space be incorporated into selection? In a panmictic population model selection is often based on a proportional fitness selection scheme where the probability of any individual being selected is proportional to their relative fitness compared with the entire population. Since the population is well-mixed this approach is reasonable, although there are some situations where even this model does not appear biologically plausible. For example, if we consider one individual P_i that has a fitness which is much greater than all other individuals in the population, then P_i will be selected for breeding many more times than all other individuals. If the model does not consider the energy costs or an associated time to death involved in P_i breeding then it would appear that a single individual could fill the entire population without any other individual being involved. Clearly this type of extreme model does not appear valid, since no relatively weaker individuals contribute to the next generation. Of course, a tournament selection (such as 2-Tournament selection) would reduce the selection pressure between P_i and the remaining population and therefore weaker individuals would also contribute to the next generation. Since the tournament selection randomly selects a subset of the individuals prior to a local selection process this

model would appear more biologically plausible, allowing a gradual increase in P_i throughout the population, however this would take many generations to reach convergence. However, mathematical models such as those found in population genetics generally use fitness proportionate selection since this allows a straight-forward approach to modelling the propagation of genes in a population based on fitness.

The introduction of space to the population can be most simply considered in terms of a graph, where the vertices of the graph representing the locations in space where individuals reside (or alternatively can be considered to represent the individuals themselves), and the edges of the graph represent the local demes in which each individual interact. If every vertex of the graph has some path to all other vertices then the graph defines a population which is connected, however it is likely that locally-dominant individuals will take longer to propagate throughout the graph than the equivalent fully-connected (panmictic) graph.

Consider now a graph which is a long linear structure, where each vertex is only connected to one vertex either side. Define one individual with a very large relative fitness F_i who initially resides at one end of this linear spatial structure. If a global fitness proportionate select model is used for selection then the individuals at the opposite end of the graph will have a very low probability of ever being selected. In fact we could choose F_i such that the probability of any other individual being selected is below any particular threshold. This model leads to the conclusion that the individuals at the far end of the linear structure from the fit individual are never considered for breeding: it is as if time has stopped, since no breeding or interactions occur. Since this is due to an individual that is very distant from these locations, and therefore whose existence cannot be directly ascertained by these individuals, we obtain the somewhat implausible model that implies distantly connected individuals can change the dynamics of local interactions. Unless there is some form of information being transmitted across the space, independent of the individuals themselves, this does not appear to be a model that corresponds to a biologically-realistic system.

The obvious weakness involved with a global selection model in a distributed environment is that the progression of time, independent of the individuals, is ignored. There are, however, a number of simple approaches to alleviate this problem. For example, a spatially-explicit model using a form of the Moran process can incorporate death of an individual as an independent background process. Hence the models select an individual at random for death, and then use a local selection scheme, within the deme of the location which is now vacant, to determine the new offspring. This approach is used by Komarova (2006) to study the spatial dynamics in carcinogenesis. Here a Moran process is defined for a one-dimensional space, and at each time-step an individual is selected at random for death – the individual is replaced by a copy of a parent from the immediate neighbourhood, as shown in Figure 1. This parent is selected using a proportional fitness selection scheme, even though only two candidate parents (the local deme) are involved. This model clearly establishes a relationship to the relative fitness of an individual based on their local deme, rather than a global selection process.

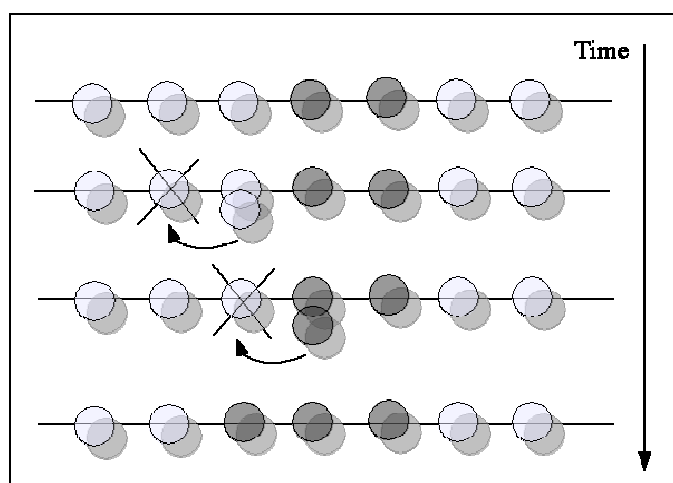


Figure 1. The Moran Process for a linear spatial structure (modified from Komarova (2006)). Individuals are randomly selected for death and replaced by one of their nearest neighbours, selected by proportional fitness.

In contrast there has also been work examining linear processes for somatic evolution, where the production of a new daughter cell forces all cells to move sideways (to the right) in the array – the finite length of the array means that the cell at the far right-hand side falls off the array and suffers apoptosis (Nowak et al. 2003). Since through the repeated right-shift of new cells the fixation rate of a mutation is independent of its relative fitness (since it must occur in the first cell position) the effect of selection was not critical to the model. However, since the model commences by selecting a cell for reproduction proportional to their fitness, and then divides this cell into two daughter cells, one that replaces the parent and one that takes the position to the right of the parent, this model does not correspond to the previous model in terms of the treatment of spatial structure.

A very different form of problem is described by Hauert and Doebeli (2004), where the emergence of cooperation in evolutionary game theory is considered for Prisoner's Dilemma and the snowdrift game. The introduction of spatial structure to Prisoner's Dilemma has been shown to promote cooperation between individuals, whereas for the snowdrift game spatial structure in game playing often reduces the proportion of cooperators over a wide range of possible scenarios. For these types of spatial models individuals interact only with their local neighbourhood, however two forms of updating are considered: synchronous across the lattice (i.e. a generational model) and asynchronous updates, which treats time in a more continuous manner. Their results show that for synchronous models the probability of cooperator-like behaviour decreases, whereas for asynchronous updating the modeled process is generally unaffected by spatial structure. Although in each case the spatial structure is observed during the process (there is no global measure used for determining how the process operates) the change in how time and space interact has a measurable effect on the resulting model.

A final example is drawn from the work of Doebeli and Dieckmann (2003), where they considered models of speciation associated with weak environmental gradients. Here a continuous model of space was described, with an environmental gradient in one dimension. Individuals were assigned birth rates, death rates and movement rates, with individuals reproducing at a fixed rate for asexual populations. For sexual models, once an individual had been identified as about to reproduce, a partner was selected based on their mating character and phenotypic distance in ecological or marker character (see Methods section of Doebeli and Dieckmann (2003)). The effect of space on this model altered the probability that any two individuals mated: the probability of mating reduced as the spatial distance between individuals increased. Although the model had some components of global decision making, such as the individual chosen for the next event was chosen with a probability proportional to the global properties of the system, the strength of competition determining the death rate depended on the number of individuals in the neighbourhood (a carrying-capacity) and their phenotypes. Hence local interactions were the main contribution to the behaviour and life history of each individual.

3.0 ARE SPACE AND TIME DISTINCT?

The previous examples highlight one significant concept – if global selection is used to select individuals for reproduction it is possible that spatially distant individuals can have their behaviour altered. Hence, although selection and reproduction are intimately related to time, most models that incorporate space allow time to operate at a uniform rate *independent of global properties*. For example, the majority of the previous models allow time to be equally shared amongst the model, either through an explicit sharing of events or through an unbiased selection of space as a starting point to determine selection and reproductive steps. This suggests that spatial structure and time cannot be treated independently, and that local processes cannot alter the processes of distant, but connected, regions.

The question of biologically plausible models has only been indirectly considered in this discussion. Future work will require a more detailed consideration of both theoretical and biologically observed processes with specific attention to the interaction of space, both locally and globally, and the timing of processes that occur because of spatial connectivity.

ACKNOWLEDGEMENTS

The authors would like to thank Karl Sigmund for highlighting the importance of this area.

REFERENCES

Doebeli, M. and Dieckmann, U. Speciation along environmental gradients. *Nature* **421** 259-264 (2003).

Hauert, C. and Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428** 643-646 (2004).

Komarova, N. Spatial Stochastic Models for Cancer Initiation and Progression, *Bulletin of Mathematical Biology*, **68** 1573-1599 (2006).

Lieberman, E. Hauert, C. and Nowak, M.A. Evolutionary dynamics on graphs. *Nature* **433** 312-316 (2005).

Maruyama, T. A simple proof that certain quantities are independent of the geographical structure of the population. *Theor. Popul. Biol.* **5** 148-154 (1974).

Nowak, M., Michor, F. and Iwasa, Y. The linear process of somatic evolution. *PNAS* **100(25)**, 14966-14969 (2003).

Whigham, P.A. and Dick, G. Evolutionary Dynamics on Graphs: The Moran Process. In *Simulated Evolution and Learning, Lecture Notes in Computer Science* **4247**, 1-8 Springer-Verlag (2006).