What can artificial life offer ecology?

Jason Noble[1], Donna Clarke[2] and Rob Mills[1]

[1]Science and Engineering of Natural Systems group, School of Electronics and Computer Science, University of Southampton
[2]School of Biological Sciences, University of Southampton
jn2@ecs.soton.ac.uk

Artificial life is the simulation and synthesis of living systems, and ALife models show how interactions between simple entities give rise to complex effects. Ecology is the study of the distribution and abundance of organisms, and ecological modelling involves fitting a linear model to a large data set and using that model to identify key causal factors at work in a complex ecosystem. We are interested in whether the individual-based modelling approach of ALife can be usefully employed in ecology.

ALife models are "opaque thought experiments" (Di Paolo et al., 2000, Proc. ALife VII, p.497). They show that a phenomenon can arise from a given set of assumptions in cases where the implication is not clear from intuition alone: e.g., that spatial structure in a population can lead to altruistic behaviour. This type of modelling can be useful to ecology by showing the plausibility of a novel concept or process, which in turn suggests new natural experiments and new forms of data to collect. However, we argue that ALife models can go beyond this "proof of concept" role and serve as a direct account of data in the same way that statistical models do.

We focus on a typical problem from ecology: the effect of clearing powerline corridors through a forest on the local wildlife populations (Clarke et al., 2006, Wildlife Research, 33, p.615). The real data set in this case is complex and, of course, we don't know the true effects that underlie it. We therefore generated a fictional data set that reflects aspects of the original problem while allowing complete control over the simulated environment. The idea is to construct a test case for looking at the relative success of different modelling approaches. We know the true picture because we generated the data, but which modelling approach will get closer to the truth? The fitting of generalized linear models as is conventional in ecology, or the use of individual-based simulations as in ALife?

Statistical models are fitted using some variant of the method of maximum likelihood: given the data, which of the models in the family we're considering (e.g., a linear regression) makes the observed data most plausible? When dealing with simulations, however, it is difficult to establish that one model is a better fit to data than another. Simulations have many parameters, it may be difficult to determine a level of granularity at which the simulation output is supposed to "match" the data, and there will be no analytically tractable likelihood function. These problems are solved by the method of indirect inference (Gouriéroux et al., 1993, J. Applied Econometrics, 8, p.S85) in which an auxiliary model is fitted to both the real data and to the output from competing simulation models. The best simulation model is the one producing the closest match to the data in terms of fitted parameter values in the auxiliary model.

Using indirect inference with our fictional data set we demonstrate that ALife simulation models can be fitted to realistic ecological data, that they can out-compete standard statistical approaches, and that they can thus be used in ecology for more than just conceptual exploration.