

Subpopulation Agents Emerge from Individual Agents in Metapopulation Simulations

H.J. Stolk^{1,2}, J. Hanan² and M.P. Zalucki³

¹ Dione Complex Systems, Bangor, Gwynedd, United Kingdom

² Advanced Computational Modelling Centre & ARC Centre for Complex Systems, The University of Queensland, Brisbane, Queensland, Australia (for H.J. Stolk until beginning 2005)

³ School of Integrative Biology, The University of Queensland, Brisbane, Queensland, Australia
Email: Henk@DioneComplexSystems.com

Keywords: *Emergent models, multi-agent simulation, ecological modelling, meta-population dynamics*

EXTENDED ABSTRACT

Complex systems consist of interacting entities or components. Different levels can be distinguished in a complex system, from the bottom level with the most fine-grained entities, through intermediate levels of composite entities, up to the level of the whole system. Properties and interactions of entities at a lower or micro-level give rise to properties and behaviour of more course-grained entities at a higher or macro-level, but the latter are not obviously predictable from the former. At the highest level, there is one macro-level entity consisting of the whole system. This phenomenon of macro-level properties and behaviour arising out of micro-level properties and behaviour, without being immediately apparent from them, is known as emergence, a fundamental property of complex systems.

An example of a complex system in ecology is a population of organisms subdivided into several subpopulations. Such a system is often studied as a metapopulation using a modelling approach that addresses the particular characteristics of populations in patchy environments. Metapopulation theory is based on the idea of a 'population' of local populations.

The new Emergent Models methodology uses multi-agent simulations to study emergence in complex systems. Agents model components at different levels of a complex system. Emergent models reveal properties and behaviour of higher-level agents as emerging from properties and behaviour of lower-level agents.

Metapopulation models can be related to individual-based models. Ways of linking individual-based models with higher level metapopulation models are examined here by relating levels in multi-agent simulations using the Emergent Models methodology. Metapopulation dynamics distinguishes three levels in an

ecosystem, with organisms at the lowest level, subpopulations of organisms at an intermediate level, and the total population consisting of subpopulations at the highest level. Thus, emergent models of subpopulation agents could be discovered from properties and behaviour of organism agents, and an emergent model of the total population could be discovered from properties and behaviour of subpopulation agents.

Deriving higher-level emergent models from lower-level models was examined in ecosystem simulation, modelling a metapopulation with interacting monarch butterflies (*Danaus plexippus*) and their milkweed (*Asclepias*) host plants as a complex system with individual organisms at the micro-level and butterfly subpopulations around plant patches at the macro-level. Genetic programming algorithms, complemented with linear and non-linear regression techniques, were used to construct interacting macro-level patch subpopulation agents, together composing a metapopulation model, from micro-level simulations of individual butterflies interacting with plants.

The experiments have demonstrated how, in a multi-agent simulation of a complex system, macro-level group agents can derive their behaviour from micro-level agents. Once that is done, simulations can be run with group agents instead of individual agents, yielding similar results. A complex system can thus be understood at the level of subsystem agents, enabling integration with other higher-level agents such as human managers. Further, group agent simulations use much larger space and time scales than individual level simulations, enabling huge performance gains. This makes possible numerous simulation experiments that would not be possible with micro-level simulations because of prohibitive time constraints.

1. INTRODUCTION

Complex systems consist of interacting entities or components. Generally, a number of different levels can be distinguished in a complex system, from the bottom level with the most fine-grained entities, through intermediate levels of composite entities, up to the level of the whole system. Properties and interactions of entities at a lower or micro-level give rise to properties and behaviour of more coarse-grained entities at a higher or macro-level, but the latter are not obviously predictable from the former. At the highest level, there is one macro-level entity consisting of the whole system. This phenomenon of macro-level properties and behaviour arising out of micro-level properties and behaviour, without being immediately apparent from them, is known as emergence, a fundamental property of complex systems (see e.g. Bar-Yam 1997).

Scientific research has often sought to understand one level of reality in terms of another 'lower' level. This is illustrated by thermodynamics, describing properties of matter in terms of macroscopic parameters, and statistical mechanics, seeking to explain the laws of thermodynamics through the microscopic application of Newton's laws (Bar-Yam 1997, p. 58-95).

1.1. Metapopulation Models

An example of a complex system in ecology is a population of organisms subdivided into several subpopulations. Such a system is often studied as a metapopulation using a modelling approach that addresses the particular characteristics of populations in patchy environments (e.g. Hanski & Gilpin 1997; Hanski & Gaggiotti 2004). Metapopulation theory is based on the idea of a 'population' of local populations and this idea is captured by a category of models called stochastic patch occupancy models (Ovaskainen & Hanski 2004). These models are based on the assumptions that a landscape consists of discrete patches of breeding habitat surrounded by land that is unsuitable for the species under study, and that each habitat patch has only two possible states: occupied and unoccupied. A stochastic patch occupancy model describes how colonisation and extinction rates or probabilities of patches being occupied depend on the structure and present occupancy pattern of the landscape (Ovaskainen & Hanski 2004, p. 77, 83).

Metapopulation models are often used in the context of conservation policy and tend to focus on colonisation and extinction effects. Although the metapopulation approach is useful, models should

be developed to describe levels of local populations rather than only colonisation and extinction events. Attempts to take into account more details of local populations have been made by formulating so-called structured metapopulation models (e.g. Gyllenberg *et al.* 1997).

1.2. Emergent Models

To elucidate the relationships between components of complex systems and the emerging system level behaviour, verbal or mathematical analyses are typically used. Examples in population biology and ecology are described in Okubo and Levin (2001). More recently computer simulation has become available as an additional tool for complex systems studies. The metapopulation models introduced in Section 1.1 are mostly mathematical models, but computer simulation is also used to simulate such populations.

Both mathematical modelling and computer simulation as tools for studying complex systems and emergence have strengths and limitations. A strength of mathematical models is that they lead to general and clearly understandable results. A limitation is that many simplifying assumptions have to be made to keep the analysis tractable. Computer simulations, on the other hand, can incorporate any desired complexity and be carried out without many of the simplifying assumptions necessary for mathematical modelling. However, their results are less general and can be hard to understand, if the simulation model is complex.

Therefore, a general methodology is needed to derive macro-level properties and behaviour from individual micro-level properties and behaviour in complex system simulations, combining the strengths and avoiding the limitations of both mathematical modelling and of computer simulation. The Emergent Models methodology developed by Stolk (2005) and described by Stolk *et al.* (2003), which is outlined in Section 2, is such a methodology.

1.3. Emergent Metapopulation Models

Metapopulation models can be related to individual-based models. For example, the colonisation rate of empty patches in patch occupancy metapopulation models decreases with distance, while in individual-based models the same effect is achieved by assuming high migration mortality between patches or poor searching ability, and the two modelling approaches can be linked by determining parameters of a patch occupancy model using data from individual-based simulations (Ovaskainen &

Hanski 2004, p. 99). Ways of linking individual-based models with higher level metapopulation models are examined here by relating levels in multi-agent simulations using the Emergent Models methodology.

Metapopulation dynamics distinguishes three levels in an ecosystem, with organisms at the lowest level, subpopulations of organisms at an intermediate level, and the total population consisting of subpopulations at the highest level. Thus, emergent models of subpopulation agents could be discovered from properties and behaviour of organism agents, and an emergent model of the total population could be discovered from properties and behaviour of subpopulation agents.

Potential applications would include estimating the potential for survival, or the expected time to extinction, of a population as a function of landscape characteristics; and relating individual-based and metapopulation-level models with a view to defining weed or other pest control strategies involving macro-level decision making. These kinds of application areas could be supported much more effectively by using simulations with subpopulation agents (with agent models derived from previous individual-based simulations) to predict total population behaviour, than by using individual-based simulations directly.

2. THE EMERGENT MODELS METHODOLOGY

In order to combine the strengths of the mathematical and simulation approaches to scientific discovery, while avoiding their respective limitations, the Emergent Models methodology uses computer simulations to study how models of macro-level properties and behaviour of a complex system emerge from the properties and behaviour of the micro-level components of the system. This methodology consists of building multi-agent simulations (see e.g. Ferber 1999), with agents at different levels modelling micro-level and macro-level components of a complex system. Statistics produced by micro-level simulations are used to discover models describing properties and behaviour of macro-level components. Models discovered in this way are models of the properties and behaviour of higher-level agents emerging from properties and behaviour of lower-level agents, or emergent models.

Various methods can be used to discover emergent macro-level models from micro-level simulations. If the variables of the model equations are already

known from theoretical considerations and only unknown parameters need to be estimated, standard linear or non-linear regression techniques can be used. If the important variables also have to be discovered, more advanced techniques are needed, such as evolutionary algorithms or other machine learning methods.

In the present work we use *genetic programming* (see e.g. Koza *et al.* 1999) to discover macro-level emergent models, as it is an all-purpose method with sufficient flexibility to be applicable to many interesting cases. Genetic programming is an extension of genetic algorithms (see e.g. Holland 1975) and uses evolutionary mechanisms such as mutation and recombination to obtain a good computer program or mathematical function for solving a given problem in a population of possible solutions. A genetic programming search for solutions of a problem starts with an initial population of functions composed of operators and terminals appropriate to the problem. The operators are frequently merely standard arithmetic and logical operations. The terminals typically include the external inputs to the program or function as variables, and may also include constants and zero-argument functions. During the search, individuals representing possible combinations of operators and terminals are mutated and recombined, until a good solution is obtained. More details about genetic programming are given by Stolk and Hanan (2007).

Genetic programming can be applied to many problems, as it performs a search based on trial and error, randomly mutating and recombining building blocks of possible solutions to obtain a best solution to a problem. The building blocks of solutions are provided by the programmer and can be defined in any desired way. We complement genetic programming with linear and non-linear regression in two ways. First, we use regression to estimate unknown parameters of those equations for which we make theoretical assumptions about the important variables. Second, after having determined the important variables of other equations with genetic programming, we use regression to improve the estimates of these equations' parameters.

3. EXAMPLE: MONARCH BUTTERFLIES AND MILKWEED

A fairly simple system to study metapopulation dynamics consists of monarch butterflies (*Danaus plexippus*) and milkweed plants (*Asclepias*). Milkweed is the only host plant of the monarch butterfly for egg laying (Zalucki 1983). In

addition, milkweed grows in patches, so the butterflies tend to concentrate in and around milkweed patches, giving rise to a system where metapopulation dynamics are easily studied. The present work is meant as a demonstration of the Emergent Models methodology and the monarch data are used to make a fairly realistic model of individual butterflies, but no attempt is made to model a real ecosystem.

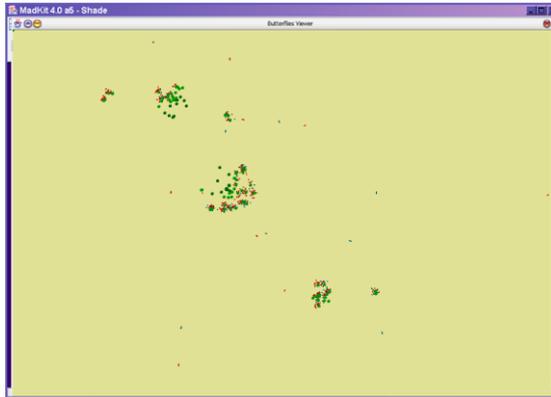


Figure 1. A screenshot of a micro-level simulation of monarch butterflies (small dots) and milkweed plants (large green and red dots).

Figure 1 shows a simulation of monarch butterflies interacting with milkweed patches. The butterflies fly a distance d at each simulation step, while turning an angle θ (picked at random). If butterflies observe the presence of milkweed plants, they decrease d and increase the average value of θ , as described by Zalucki (1983). In addition, butterflies lay eggs on milkweed plants. The eggs may hatch after some time, eventually producing new butterflies with a probability determined by mortality of eggs and immature individuals. It is assumed that surviving immature individuals appear as adult butterflies 28 days after egg-laying and can then leave their patch. They die with a probability which is determined by their environment and is lower in milkweed patches than in the non-milkweed area outside the patches.

We have carried out computational experiments to derive a macro-level multi-agent model from simulations of individual butterflies interacting with milkweed plants. The macro-level agents simulate metapopulation dynamics at the level of subpopulations. A subpopulation is defined as the population of butterflies around one milkweed patch. Macro-level agents are patch subpopulations, as well as dispersal agents simulating the dispersal process.

We subsequently used the macro-level agents in a multi-agent simulation, hypothesising that simulations with macro-level agents would yield results similar to those of simulations with individual insect agents, but with better insight in macro-level processes. This is important for example for defining intervention strategies related to policies of conservation or pest control. In addition, the macro-level simulations can be run with a much better performance than the full individual-based simulations.

The macro-level agents we use are patch subpopulation agents observing variables such as their own patch population, patch area, patch radius, distances to other patches and emigrants of other patches. The population of the area outside the patches is considered an agent as well. The subpopulation agents use their observations to update their own population through appearances of new butterflies, deaths, immigration and emigration; and to produce emigrants.

Emergent models are derived by providing to the genetic programming algorithm a set of variables and operators to be used for finding functional relationships between observations and acts. The algorithm will not only find particular functions, but also select the relevant variables to be observed by the patch subpopulation agents.

The Emergent Models methodology is flexible in the sense that constraints can be imposed on the algorithm to find only models with characteristics specified by the researcher. In the present research the models to be discovered were constrained to models similar to existing metapopulation models. For some equations we use traditional approaches, defining their variables and structure by *a priori* reasoning, only leaving some unknown parameters to be determined by statistical curve fitting methods such as linear or non-linear regression. This makes possible an intuitive demonstration in line with contemporary research in ecology, especially metapopulation dynamics. However, we do not limit ourselves to curve-fitting methods and use our approach to search a space including different models and not only different parameters. In contrast to defining metapopulation models based on intuition, the Emergent Models methodology determines what multi-agent model is consistent with given assumptions about individual behaviour. In the present research the multi-agent model is inspired by metapopulation dynamics, but the approach is equally well applicable to problems not satisfying the assumptions of metapopulation dynamics.

It was attempted to find a macro-level multi-agent model describing the behaviour of each of n patch agents and of the non-patch agent. The behaviour of each of n patch agents i ($i = 1, \dots, n$) was hypothesised to be described by equations of the form:

$$B_i = F_{B_i}(P_i) \quad (1)$$

$$M_i = F_{M_i}(P_i) \quad (2)$$

$$E_i = F_{E_i}(P_i) \quad (3)$$

$$I_i = F_{I_i}(E_j) \quad \text{for } i \neq j \in \{0, \dots, n\} \quad (4)$$

The behaviour of the non-patch agent was hypothesised to be described by:

$$B_0 = F_{B_0}(P_0) \quad (5)$$

$$M_0 = F_{M_0}(P_0) \quad (6)$$

$$E_0 = F_{E_0}(I_0) \quad (7)$$

$$I_0 = F_{I_0}(E_j) \quad \text{for } i \neq j \in \{0, \dots, n\} \quad (8)$$

where, for $i \in \{0, \dots, n\}$, P_i is the population of patch i at time t ; B_i , M_i , I_i and E_i are numbers of appearances, deaths, immigrants and emigrants of patch i between times t and $t + 1$, and F_{B_i} , F_{M_i} , F_{E_i} and F_{I_i} are the functions to be found.

Thus, appearances and deaths of a patch depend on the patch's population, emigration depends on its population and area, and immigration depends on its area and dispersal from a subset of other patches. Only a subset of other patches is assumed to be important for immigration of a given patch: patches outside this subset are not important, for example because they are too far away or because they are 'hidden' from the patch under consideration by other patches in between.

This general problem can be solved using Emergent Models by providing a genetic programming algorithm with possible operators and parameters as building blocks for the functions to be found. The methodology is very flexible and can be adapted easily to existing or assumed knowledge about the problem to be solved by imposing appropriate constraints on the genetic programming algorithm to allow different forms for the functions to be found. In the present case the following assumptions were made to constrain the functions.

Numbers of appearances in a patch are explained by a constant appearance rate β_i and an additive constant γ_i to take account of appearances due to

eggs present in addition to adult individuals, so equations (1) become

$$B_i = \beta_i P_i + \gamma_i \quad \text{for } i \in \{1, \dots, n\} \quad (9)$$

No appearances take place outside patches, so equation (5) becomes

$$B_0 = 0 \quad (10)$$

Numbers of deaths in every subpopulation are explained by a constant death rate μ_i of a subpopulation, so equations (2) and (6) become

$$M_i = \mu_i P_i \quad \text{for } i \in \{0, \dots, n\} \quad (11)$$

Emigration of a patch depends on the square root of population, because emigration happens on the edges of the patch areas, so, if population is roughly proportional to patch area, emigration is expected to be roughly proportional to patch circumference and thus to the square root of patch area or population and equations (3) become

$$E_i = \tau_i \sqrt{P_i} \quad \text{for } i \in \{1, \dots, n\} \quad (12)$$

Emigration of the non-patch area depends linearly on immigration into the non-patch area and equation (7) becomes

$$E_0 = \vartheta I_0 \quad (13)$$

Equations of these types can be fitted to observed data, or in our case simulated data, with standard linear and non-linear regression techniques. This was done with results as shown in Table 1, where equation (14) is the same as equation (10), equations (15) – (20) result from fitting equations (9), equations (21) – (27) result from fitting equations (11), equation (35) results from fitting equation (13) and equations (36) – (41) result from fitting equations (12).

Immigration of the non-patch area was, as a result of the simulation design, equal to total emigration of all patches, so equation (8) becomes equation (28). Immigration of each patch was assumed to be potentially dependent on emigration of all other patch and non-patch subpopulation agents. To determine which other subpopulation agents contributed most to immigration of a given subpopulation agent, genetic programming was used to find that for each patch immigration was best explained by a linear function of emigration of the non-patch area. Subsequently, linear regression was used to calibrate the results, shown in equations (29) – (34) that realise equations (4).

Table 1. Emergent model of subpopulation agents' behaviour derived from an individual-based simulation.

$B_0 = 0$ (14)	$M_0 = 0.981P_0$ (21)	$I_0 = \sum_{i=1}^6 E_i$ (28)	$E_0 = 0.286I_0$ (35)
$B_1 = 0.0102P_1 + 8.103$ (15)	$M_1 = 0.00102P_1$ (22)	$I_1 = 0.280E_0$ (29)	$E_1 = 0.607\sqrt{P_1}$ (36)
$B_2 = 0.0148P_2 + 4.938$ (16)	$M_2 = 0.00221P_2$ (23)	$I_2 = 0.325E_0$ (30)	$E_2 = 0.719\sqrt{P_2}$ (37)
$B_3 = 0.0331P_3 + 0.794$ (17)	$M_3 = 0.00419P_3$ (24)	$I_3 = 0.167E_0$ (31)	$E_3 = 0.730\sqrt{P_3}$ (38)
$B_4 = 0.0601P_4 + 0.260$ (18)	$M_4 = 0.00606P_4$ (25)	$I_4 = 0.0859E_0$ (32)	$E_4 = 0.713\sqrt{P_4}$ (39)
$B_5 = 0.0643P_5 + 0.642$ (19)	$M_5 = 0.00803P_5$ (26)	$I_5 = 0.136E_0$ (33)	$E_5 = 0.947\sqrt{P_5}$ (40)
$B_6 = 0$ (20)	$M_6 = 0$ (27)	$I_6 = 0.00694E_0$ (34)	$E_6 = 0$ (41)

As an example of data used and results, micro-level simulation data and a plot of the discovered macro-level function for I_2 , described by equation (30), are shown in Figure 2.

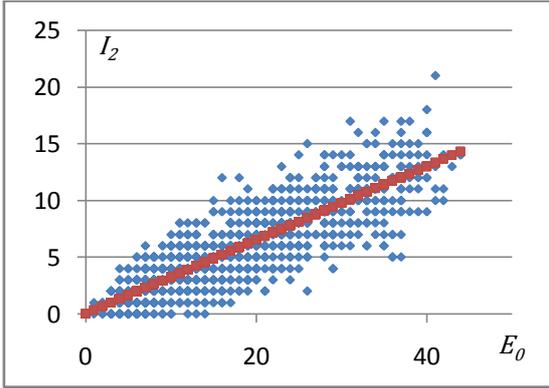


Figure 2. Micro-level simulation data (blue) and macro-level function (red) for $I_2 = f(E_0)$.

In this model patch agents have an autonomous behaviour in the form of natural population increase. They interact with other patch agents by providing emigrants to them and by taking immigrants from them. Immigrants are a subset of dispersed emigrants from all other patch agents. It is not obvious from the outset which other patch agents are relevant for immigration in a particular patch. The genetic programming algorithm can use emigrants from any patch, but does not have to do so and will retain only those patch agents in the solution for a given patch that are relevant for obtaining a good immigration equation for that patch. Thus, the problem is partly to find optimal parameters for the equations, but also to determine the structure of each patch's immigration function

by finding the subset of other patches contributing to a patch's immigration.

Figure 3 is an example of simulation results obtained with the emergent macro-model. It shows the population of one patch from the beginning of the simulation until the end at hour 1012. The blue line shows the population generated by the micro-level simulation, which has a time step of one minute. The red lines show the population generated by the macro-model, starting at the micro-simulation population levels at hours 600, 700 and 800 respectively.

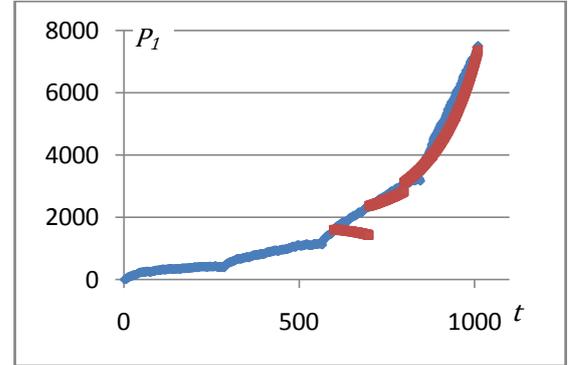


Figure 3. Micro-level simulation (blue) and macro-level simulations (red) of $P_1 = f(t)$.

4. CONCLUSIONS

This example demonstrates how, in a three level system, intermediate level group agents can derive their behaviour from micro-level agents. In order to focus on the simulation's multi-level character in this example and to keep its implementation as simple as possible, the derivation of the subpopulation agents' macro-level behaviour is

based on data generated by only one micro-level simulation. A logical next step is to run a representative number of micro-level simulations with different initial conditions. The derived macro-level behaviour would then be an average behaviour for all possible micro-level simulations, making the group agent simulation results more robust.

In conclusion, with the Emergent Models methodology one can find emergent macro-models corresponding to micro-models of ecological systems. A three level system has been described consisting at the micro-level of insects interacting with plants, at an intermediate level of subpopulations of insects corresponding to plant patches, and at the macro-level of the entire system of insect population and plant patches. Once a model of subpopulation behaviour has been derived, "subpopulation-based" simulations can be run instead of individual-based simulations to obtain essentially the same results for the behaviour of the whole system or population.

In this way, understanding of behaviour at the macro-level can be increased. In addition, simulation performance can be improved dramatically, at least after a "learning phase", during which the macro-level behaviour is discovered by the genetic programming algorithm. A macro-model can be used to conduct numerous experiments that would not be possible with micro-level simulations. For example, we could add a probability distribution of random deviations to the macro-model and then use the macro-model for Monte Carlo simulations to study the behaviour of the model with different initial conditions, potentially solving policy problems such as the optimal patch structure for maintaining a population.

The definition of the most appropriate macro-level agents depends on the application of the model. When the purpose of the modelling exercise is decision support, macro-level agents can be defined in terms of the entities decision makers perceive to be interacting with. Decision makers could be modelled as agents interacting with subpopulation agents rather than individual insects, for example. Multi-agent simulations at the macro-level could then be used in a natural way for evaluation of different policy scenarios.

5. REFERENCES

Bar-Yam, Y. (1997), *Dynamics of Complex Systems*, Perseus Books, Reading, MA.

DeAngelis, D.L., and L.J. Gross (eds) (1992), *Individual-Based Models and Approaches in Ecology: Populations, Communities and Ecosystems*, Chapman & Hall, New York.

Ferber, J. (1999), *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*, Addison-Wesley, Harlow.

Gyllenberg, M., I. Hanski, and A. Hastings (1997), *Structured metapopulation models*, In Hanski & Gaggiotti (2004).

Hanski, I., and M.E. Gilpin (eds) (1997), *Metapopulation Biology: Ecology, Genetics, and Evolution*, Academic Press, San Diego, CA.

Hanski, I., and O.E. Gaggiotti (eds) (2004), *Ecology, Genetics, and Evolution of Metapopulations*, Elsevier, Burlington, MA.

Koza, J.R., F.H. Bennet III, D. Andre, and M.A. Keane (1999), *Genetic Programming III: Darwinian Invention and Problem Solving*, Morgan Kaufmann, San Francisco, CA.

Okubo, A., and S.A. Levin (eds) (2001), *Diffusion and Ecological Problems: Modern Perspectives*, 2nd ed, Springer, New York.

Ovaskainen, O., and I. Hanski (2004), *Metapopulation dynamics in highly fragmented landscapes*, In Hanski & Gaggiotti (2004).

Stolk, H.J. (2005), *Emergent Models in Hierarchical and Distributed Simulation of Complex Systems*, PhD Thesis, University of Queensland, Brisbane, Australia.

Stolk, H., K. Gates, and J. Hanan (2003), *Discovery of emergent natural laws by hierarchical multi-agent systems*, paper presented at IEEE/WIC International Conference on Intelligent Agent Technology, Halifax, Canada, October.

Stolk, H.J., and J. Hanan (2007), *Discovering genetic network models in *Pisum sativum**, paper presented at MODSIM07, Christchurch, New Zealand, December.

Zalucki, M.P. (1983), *Simulation of movement and egg laying in *Danaus plexippus* (Lepidoptera: Nymphalidae)*, *Resources in Population Ecology*, 25, 353-365.