COMMENT AND REPLY COMMENT

Reproducibility in simulation experiments: Comment on Munguia et al. (2010)

Jean-Sébastien Lauzon-Guay[1],*, **Devin A. Lyons[2]**

[1]Institut Maurice-Lamontagne, Fisheries and Oceans Canada, Mont-Joli, Quebec G5H 3Z4, Canada
[2]Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

ABSTRACT: In a recent article, Munguia et al. (2010; Mar Ecol Prog Ser 413:229–240) presented a spatially explicit mathematical model of competition among 4 sessile species of benthic invertebrates. We argue that, based on the description of the model, it is impossible to repeat the analysis without making numerous guesses and assumptions. Furthermore, the results do not correspond to the model as it is described. We highlight elements of the model formulation and description relating to dispersal, colonization, and competition that, if clarified, would help the reader to understand, evaluate, and extend the results of this study.

KEY WORDS: Priority effect · Competition · Dispersal · Simulation model

Introduction

There has recently been increased interest in promoting and encouraging scientific clarity to insure that published research is reproducible (Donoho 2010, Mesirov 2010) and that all data be made available to readers (Baggerly 2010, Whitlock et al. 2010). The scientific imperatives of repeatability and reproducibility are especially difficult to achieve in ecology, because ecological phenomena depend on biotic and abiotic factors that vary in space and time (Ellison 2010). One exception is ecological research performed *in silico*, via computer simulations.

An example can be found in a recent Theme Section on 'Threshold dynamics in marine coastal systems' (Osman et al. 2010). Munguia et al. (2010) presented a model examining the interplay between disturbance, dispersal, and competition in determining the distribution and abundance of 4 species within a benthic metacommunity in Long Island Sound, USA. The model was aimed at elucidating the role of priority effects on species dominance — a subject of great interest, as many

gaps remain in our understanding of community dynamics in spatially structured environments — but the article shares a difficulty with many simulation studies: the presentation of the methods is inadequate for the reader to understand precisely what was done and replicate the modelling exercise. The model contains assumptions that appear to be inconsistent with ecological principles or with the community that the model represents. We ask the authors to clarify their work so that their findings can be properly assessed and built upon.

Model summary

Munguia et al. (2010) presented 'A simulation model … to generate community patterns during succession' (p. 231). The model tracks 4 species of benthic sessile invertebrates in 4 habitats (or communities). The model is seemingly governed by 2 equations that dictate recruitment and dispersal. Their Eq. (1) defines the transition probability T_{ij} for

each species i in each community j:

$$T_{ij} = b_i - \left[\frac{b_i \times d_i}{\sum \text{distances to } j} \right] \qquad (1)$$

where b_i and d_i are respectively the birth and dispersal rates of species i. Their Eq. (2) provides the core working of the model:

$$N_{ij}(t+1) = N_{ij}(t) + T_{ij} \times N_{ij}(t) \left(\frac{K - N_{ij}(t) - \sum_{s \neq i} (N_j(t)_s \times \alpha_s)}{K} \right) \qquad (2)$$

where $N_{ij}(t)$ is the abundance of species i in community j at time t, K is the carrying capacity of a community, α_s represents the competitive ability of species s, and '$N_j(t)_s$ is the sum of species abundances in community j excluding species i' (Munguia et al. 2010, p. 232–233). There may be a typographical error here, as it is unlikely that $N_j(t)_s$ is the 'sum of species abundances'; we think that it may be the abundance of species s in community j. Otherwise, the sum in Eq. (2) would not make sense.

At each time step, each population had a 5% chance of 10% mortality. A total of 3 disturbance scenarios were simulated: (1) without disturbance; (2) at every 4th time step, 1 of the 4 communities was randomly selected to suffer an additional 70% mortality; (3) all communities suffered an additional 70% mortality. Each model run lasted 200 time steps, with 4 time steps intended to represent the passage of 1 yr. Most of the results centre on the occurrence of dominance events, defined as a species reaching an abundance of 50 ind. (50% of the carrying capacity).

Dispersal, connectivity and distances between communities

Our first concern centres on dispersal and distances between communities. Munguia et al. (2010) describe the model as spatially explicit, with distances between the 4 communities defined for each model run. However, we are unable to determine the actual distances between communities in the model. We are told that 'distance is a positive integer with a ceiling of 60' (p. 232) and that distances between communities are changed slightly (<5%) between runs. An average value (Distance units) for each habitat is given in Munguia et al. (2010, their Fig. 1b); distance *to what* is not explicitly defined, but it seems logical to conclude that this value is the average distance to the other 3 communities. A clearer description of the spatial layout of the communities, which is bound to affect population dynamics, is needed to understand or extend the results of the model.

The influence of distance between communities on population dynamics in the model is linked to dispersal. The model weights the dispersal rate by the distance between the community j and all other communities to ensure that species with a '… high dispersal rate … would have a relatively low self-seeding ability' (Munguia et al. 2010, p. 232). We find Eq. (1) rather counter-intuitive for 3 reasons:

(1) Colonization success (excepting self-seeding) is expected to increase with dispersal ability. However, according to Eq. (1), the proportion of propagules reaching a community j (T_{ij}) decreases with the dispersal ability of that species, regardless of the source of the propagules. For a constant birth rate, a species with low dispersal ability will have more propagules reaching all communities than a species with a high dispersal rate (T_{ij} decreases with d_i).

(2) Dispersal between 2 communities is expected to decrease with the distance between them. However, according to Eq. (1), dispersal appears not to depend explicitly on pair-wise distances, but rather on the overall isolation of a given community. Moreover, dispersal to a community apparently increases, rather than decreases, with isolation.

(3) The number of propagules of a given species arriving at a community is expected to depend on the abundance of that species in the source habitats. According to our reading, abundance of Species i at Community j is a function only of its abundance at Community j at the previous time step, and not on its abundance in the other communities. In this case, there is confusion as to whether N_{ij} represents an element of the vector N (corresponding to Species i in Community j) or the entire vector N (see Munguia et al. 2010, their Appendix 1) and whether T_{ij} is meant to represent an element or the entire matrix T. However, this interpretation is not consistent with how N_{ij} and T_{ij} are defined in the text. Whatever the case, we are left to assume that Eq. (2) is not really representative of how the model actually works, that there is some connectivity between communities, and that the authors use the matrix T and vector N and use 'matrix multiplication … to simulate changes in each habitat…' (Munguia et al. 2010, p. 232). This, however, leads to further complications.

Matrix multiplication

In addition to not understanding the link between Eq. (2) and the matrix presented in the appendix, we also do not understand how the matrix operates. If we multiply matrix T by vector N, we obtain the abundance of each species in each community at the next time step, such that the abundance of Species A in Community 1 is:

$$A_1(t+1) = A_1(t)a_1 + A_2(t)a_2 + A_3(t)a_3 + A_4(t)a_4 \qquad (3)$$

where a represents elements of Matrix T. The abundance of Species A in Community 2 is:

$$A_2(t+1) = A_1(t)b_1 + A_2(t)b_2 + A_3(t)b_3 + A_4(t)b_4 \qquad (4)$$

According to the Appendix caption (Munguia et al. 2010, their Appendix 1), each element of Matrix T and Vector N is represented by a letter (A to D) representing a species and a number (1 to 4) representing a community. It appears that the abundance of Species A in Community 1 is related to parameters pertaining to Species A (or a), while the abundance of Species A in Community 2 is related to parameters pertaining to Species B (or b). This is nonsensical, and we assume that Matrix T or its description in their Appendix 1 is erroneous.

Local population dynamics: a numerical example

Apart from the apparent irregularities concerning dispersal among communities, the local population dynamics are also problematic. For the sake of simplicity, we assume that *Mytilus* has a population abundance of 5 in one community (initial conditions used by Munguia et al. 2010) and that all other communities are empty of all species. Using the birth (100) and dispersal (0.8) rates for *Mytilus* in Munguia et al. (2010, their Table 1), and assuming that the community of interest is on average 30 distance units from all other communities (i.e. sum distances to $j = 91$, a community is 1 distance unit away from itself) we can calculate T_{ij} for *Mytilus* from Eq. (1) as 99.12. If we input this into Eq. (2), we obtain a population size of *Mytilus* of 475.82 individuals at Community j after the first time step. It is important to note that this abundance is more than 4 times greater than the carrying capacity of the community ($K = 100$) and that *Mytilus* would more than qualify as a species that has achieved dominance in that community. The criteria for dominance given in the paper is an abundance >50% of carrying capacity (i.e. 50 ind.). Given the rapid growth of *Mytilus* after the first time step and its immediate dominance of the community, it is surprising to find that *Mytilus* only achieves dominance in <10% of the runs, and takes more than 1 time step to do so (Munguia et al. 2010; their Figs. 2a & 3a). Even if we include the variability used by Munguia et al. (2010) — which involves multiplying the birth rate by a random number between 0 and 1 at each time step — *Mytilus* would still achieve dominance after 1 time step whenever the birth rate is >10.09 (i.e. 89.91% of the time). It appears that there is a discrepancy between the model (as it is presented) and the results presented by Munguia et al. (2010).

Competition

The model uses a discretized Lotka-Volterra type equation (Munguia et al. 2010, their Eq. 2), with a carrying capacity K and parameters α_s that control competitive interaction between species. Species with large α values are stronger competitors and are able to '… maintain or displace individuals from a particular local habitat' (Munguia et al. 2010, their Table 1). In the Lotka-Volterra equations, the competition coefficient of a species is a function of the species being competed against and it is expressed in units of the other species. In this model, the α values appear to be fixed for a species. As a result, the implicit assumption is that each species has the same effect on all of its competitors. Such a simplifying assumption may be acceptable in a generic model, but we imagine that such a situation is rare in nature, and unlikely for the benthic species of Long Island Sound. In addition, the values chosen for competitive ability of each species was apparently based on its ability to maintain space and overgrow other species. In the model however, competition appears to take place during recruitment. Settling in unoccupied space, overgrowing previously settled competitors, and resisting overgrowth are very different processes and there is no indication that these are correlated.

The costs and benefits of choosing fixed α values are debatable, but the choice of the actual values is critical to the results. Munguia et al. (2010, p. 233) state that, 'Competition affected N_{ij}, depending on the competitive strength and abundance of coexisting species; there was no explicit intraspecific competition taking place in the model'. This suggests that the abundance of species i should not be taken into account in the density dependence term of Eq. (2), despite the fact that abundance (N_{ij}) is subtracted from K (which implies that $\alpha_{ii} = 1$). The actual meaning of the sentence is therefore unclear. However, the choice of parameter values suggests that intraspecific competition is actually stronger than interspecific competition. All α values are <1, indicating that individuals of a species have a smaller effect on other species than on members of their own species. In such a case, no species can prevent another species from recruiting in a community, even if the previously established species is at carrying capacity. Weak interspecific competition promotes long-term coexistence at the local and metacommunity levels, and prevents competitive exclusion. As a result, it is logical and inevitable that 'priority effects are not creating competitive exclusion' (Munguia et al. 2010, p. 236) in this model. Although Munguia et al. (2010) suggest that biodiversity is maintained by 'the combined effects of dispersal ability and disturbance' (p. 236), it appears that the relative strength of interspecific and intraspecific competition in this model may be equally — or perhaps more — important.

Conclusions

We were unable to reproduce the model or results of Munguia et al. (2010), after making several guesses as to what was done. We do not imply that the authors' findings are erroneous. Rather, the description and formulation of the model make it difficult to understand, and impossible to build upon. The subject matter of Munguia et al. (2010) is important. We hope that the authors will clarify their work so that it can be extended by others. As many journals offer authors the opportunity to publish data, figures, computer code, and detailed description of methods as online supplementary material, we should all make the required information available to the reader, to ensure that our research is understood and used.

LITERATURE CITED

Baggerly K (2010) Disclose all data in publications. Nature 467:401

Donoho DL (2010) An invitation to reproducible computational research. Biostatistics 11:385–388

Ellison AM (2010) Repeatability and transparency in ecological research. Ecology 91:2536–2539

Mesirov JP (2010) Accessible reproducible research. Science 327:415–416

Munguia P, Osman RW, Hamilton J, Whitlach RB, Zajac RN (2010) Modeling priority effects and species dominance in Long Island Sound benthic communities. Mar Ecol Prog Ser 413:229–240   Retracted

Osman RW, Munguia P, Zajac RN (eds) (2010) Threshold dynamics in marine coastal systems. Mar Ecol Prog Ser 413:185–304

Whitlock MC, McPeek MA, Rausher MD, Rieseberg L, Moore AJ (2010) Data archiving. Am Nat 175:145–146

Clarification of model parameters:
Reply to Lauzon-Guay & Lyons (2011)

P. Munguia[1,]* J. Hamilton[2], R. W. Osman[3], R. B. Whitlatch[2], R. N. Zajac[4]

[1]Marine Science Institute, The University of Texas at Austin, 750 Channel View Dr, Port Aransas, Texas 78373, USA
[2]Department of Marine Sciences, University of Connecticut, Groton, Connecticut 06340, USA
[3]Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, Maryland 21037-0028, USA
[4]Department of Biology and Environmental Science, University of New Haven, West Haven, Connecticut 06516 USA

Lauzon-Guay & Lyons (2011) have highlighted modeling problems in Munguia et al. (2010), making us aware of errors in the model code. We regret these errors and will present a corrected version of the original paper with an improved model and novel results.

*Email: munguia@mail.utexas.edu

LITERATURE CITED

Lauzon-Guay JS, Lyons DA (2011) Reproducibility in simulation experiments: comment on Munguia et al. (2010). Mar Ecol Prog Ser 432:299–302

Munguia P, Osman RW, Hamilton J, Whitlach RB, Zajac RN (2010) Modeling priority effects and species dominance in Long Island Sound benthic communities. Mar Ecol Prog Ser 413:229–240   Retracted