Parameter Estimation in Complex Plankton Models using the Boundary Eigenvalue Nudging – Genetic Algorithm (BENGA) Method

R.A. Cropp[a], M.L. Bates[a], D.W. Hawker[a] and J. Norbury[b]

[a] Griffith School of Environment, Griffith University, Gold Coast, Queensland.
[b] Mathematical Institute, University of Oxford, Andrew Wiles Building, ROQ, Woodstock Road, Oxford.
Email: r.cropp@griffith.edu.au

Abstract: The analysis of trophically complex mathematical ecosystem models is typically carried out using numerical techniques because it is considered that the number and nonlinear nature of the equations involved makes the use of analytic techniques virtually impossible. In particular, building such models is a notoriously difficult task; most competing populations in many ecosystem models collapse (i.e. have multiple spurious extinctions). This is a realisation *in silico* of the Principle of Competitive Exclusion (Gause 1932, 1934) that stipulates that only as many competing populations can coexist as there are resources to support them. The incongruity of Gause's laboratory findings with the natural world led Hutchinson (1961) to propose the Paradox of the Plankton. This paradox articulated the observation that many populations of plankton coexisted in the ocean in apparent contradiction of Gause's Principle and the predictions of the simple theoretical models of the day. Whilst many solutions to the Paradox of the Plankton of varying robustness have been proposed (for example, Huisman and Weissing 1999, Schippers et al. 2001, Cropp and Norbury 2012b) the parameterization of complex ecosystem models remains a significant practical challenge (Cropp and Norbury 2013) that applied modelers need to overcome (for example, Gabric et al. 2008).

The conservative normal (CN) framework articulates a number of ecological axioms that govern ecosystems by exploiting the properties of systems that are written in Kolmogorov form. Previous work has shown that trophically simple models developed within the CN framework are mathematically tractable, simplifying analysis. By exploiting the properties of Kolmogorov ecological systems it is possible to ensure models have particular properties, such as all populations remaining extant, into an ecological model. Here we demonstrate the usefulness of analytical results known for models of Kolmogorov form to construct a trophically complex ecosystem model. We also show that the properties of Kolmogorov ecological systems can be exploited to create ecosystem models with structural coexistence, that is, models for which no population goes extinct for all realistic (i.e. positive) parameter sets. We utilize this property in conjunction the key attribute of Kolmogorov systems, that closed form analytical expressions for the eigenvalues associated with populations that are zero at an equilibrium point are trivially obtained, to provide a computationally efficient method for the refinement of model parameters.

Genetic Algorithms (GAs) are commonly used to estimate parameter sets that allow ecological models to reproduce observed data or theoretical objectives (Kristensen et al. 2003). It is a well-known property of GAs and other optimization approaches that convergence slows as the dimension of the solution space increases (Fournier et al. 2011), that is, as the number of parameters required to be estimated increases. We combine the properties of Kolmogorov systems with a GA to increase the rate of convergence by preconditioning the problem to a region of the solution space where desirable solutions reside. The method can be used to precondition parameter values used in standard optimization techniques, such as genetic algorithms, to significantly improve convergence towards a target equilibrium state. We use an equilibrium state community composition as this is an observed property of systems with structural coexistence, and provides a useful and easily derived goal function in the absence of measured data. However, the method is equally applicable to any chosen point in the solution space of a model.

Keywords: Boundary eigenvalue nudging, genetic algorithm, Kolmogorov system, parameter estimation

1.   INTRODUCTION

The development of complex ecosystem models with application to climate change, biomagnification of pollutants through food webs and fisheries management has highlighted a fundamental problem in ecosystem modelling. The highly non-linear nature of trophic interactions, the paucity of data and the lack of clearly articulated and widely accepted governing laws (analogous to those found in fields like thermodynamics) make the development of such ecosystem models challenging. Trophically complex models of realistic systems have a large number of parameters and many are often poorly constrained due to the paucity of observations. Therefore, parameters are chosen to fit a model to, for instance, an observed population biomass. A common tool that is used for parameter optimisation in non-linear systems is a genetic algorithm (GA). As their name suggests, GAs optimise a solution by breeding and mutating parameter sets; "fit" parameter sets (individuals), that produce model solutions close to the desired solution, have a higher probability of reproducing than "unfit" parameter sets, that produce solutions dissimilar to the desired solution. For many models, the parameter values that provide stable solutions where all populations remain extant are only a very small subset of all possible values (e.g. Cropp and Norbury 2012b). It is often assumed that because models with a high fidelity to natural systems are often trophically complex, they are too complicated for any tractable mathematical analysis.

Much of the analysis of model ecosystems is concerned with finding equilibrium points and determining their stability. In the context of ecological modeling, an equilibrium point is a point in the ecospace where all derivatives with respect to time are zero. The ecospace is a subset of the state space that contains ecologically valid solutions to the model equations. (Note that a negative population may be part of the mathematical state space, but is exterior to the ecospace.) An internal equilibrium point is a point where the system does not change in time and all populations are greater than zero, whilst a boundary equilibrium point is an equilibrium point on a face or vertex (boundary) of the ecospace, that is, where one or more populations are extinct. Finding the eigenvalues of the Jacobian matrix, sometimes called the community matrix, at the equilibrium point reveals the local stability of each equilibrium point. The Jacobian matrix is a linear approximation to the non-linear equations that govern the ecosystem.

Many, if not all plankton ecosystem models applied to studies of biological processes in the ocean, are of Kolmogorov form (Kolmogorov 1936):

$$\dot{x}_i = x_i f_i\left(x_1, x_2, \cdots, x_n\right), \quad i = 1, 2, \cdots, n \ , \tag{1}$$

where x_i is the biomass of the ith population measured in the common currency of the mass of the limiting nutrient. We refer to the f_i as life functions, which contain information on the trophic interactions of each population and determine whether they grow or die. Kolmogorov form provides models with a substantial improvement in the mathematical tractability of finding analytical forms for exactly half of the eigenvalues of the model. These are the eigenvalues of the boundary equilibrium points that relate to the populations that are extinct. These eigenvalues are obtained trivially in Kolmogorov systems since they are the life function of populations x_j evaluated at the extinction point:

$$\lambda_j = f_j\big|_{x_j^*=0} \ . \tag{2}$$

Here the superscripted asterisk denotes the value of the population at an equilibrium point. These boundary eigenvalues are often referred to as "competition eigenvalues" because they are the eigenvalues that determine whether a population that is competing for resources with existing populations can invade. We have utilized this property of Kolmogorov systems to devise a methodology to efficiently pre-condition parameter sets and better constrain the parameter space to optimise convergence of a model solution to a target solution. Once the mathematical forms for the various trophic interactions have been chosen the particular solutions of the model are dependent on the choice of parameter values. It is often difficult to derive values for these parameters from observations or laboratory experiments and thus the parameters are often very poorly constrained. The parameters are thus usually found by optimisation techniques to fit the model solutions to quantities that are easier to derive from observations (e.g. biomass). Finding parameter sets that allow the model to reproduce observed data requires some sort of optimisation routine, such as a GA, which typically require many model evaluations. Here we describe a new method that exploits the properties of Kolmogorov systems to significantly improve the convergence to a desired model solution and reduce the computational time required to calibrate the model.

The classical approach to building ecosystem models begins at the species level, where priority is given to reproducing observed species population trajectories. A community level approach, on the other hand, attempts to reproduce community level patterns (for example, rank abundance curves: Record et al. 2013, Cropp and Norbury 2014)). We use a community level approach here by placing the emphasis of our parameter refinement efforts on reproducing a community trophic pyramid structure. The model we develop to demonstrate the methodology has fourteen functional groups, from primary producers to apex predators. It represents plants (autotrophs), animals that eat plants (herbivorous heterotrophs), animals that eat animals that eat plants (carnivorous heterotrophs), and animals that eat animals that don't eat plants (also carnivorous heterotrophs, including top predators), and therefore notionally encompasses all trophic levels. Our overarching aim in this study is to build a trophically complex model with known properties (i.e. all populations coexist) and for the equilibrium solution of the model to have a particular community structure. To impart the coexistence property, we develop our model in the conservative normal (CN) framework (described below) and use particular mathematical forms for trophic interactions that ensure coexistence. To obtain a particular community structure requires an appropriate set of parameters to be chosen. Our new, computationally efficient, method for parameter refinement, which we call Boundary Eigenvalue Nudging (BEN), exploits the properties of the competition eigenvalues. As our method does not operate on the entire parameter space of the model we couple it with a Genetic Algorithm (GA) to complete the optimization. The advantage of the coupled (BENGA) system is that the BEN preconditions the parameter space for the GA and overcomes one of the deficiencies of GAs, that convergence slows dramatically as the dimension of the parameter space increases (Mitchell 1997).

2. MODEL AND METHODS

The model is constructed so that it satisfies the rules for a CN system (Cropp and Norbury 2009b, a, 2012a, b, 2013), although one strict inequality is relaxed as shown below. The CN framework is a set of ecological axioms that define basic properties that characterise ecological systems, and from which all properties of the system can be derived. This is analogous to, for example, the laws of thermodynamics. These rules formalise basic ecological concepts, principally that populations must have resources to consume in order to survive, and that these resources are identifiable and finite. A population x_i that has maximal resources R_i must be able to grow ($f_i|_{R_i=1} > 0$), and a population that has no resources must die ($f_i|_{R_i=0} < 0$). (Note that our system is scaled so that all populations and resources can only take values between zero and one.) To ensure coexistence of all populations, we relax the strict inequality of the zero resource condition ($f_i|_{R_i=0} \leq 0$). This renders our model ecologically unrealistic, in the sense that no population can go extinct, but provides a model that is exceedingly useful for the current state of biogeochemical modeling in the ocean, where the emphasis is currently on the influence of plankton on climate. When this feedback is sufficiently well understood for the plankton modeling community to move to the next step, investigating the influence of climate change on plankton, the relaxation of the zero resource condition would be inappropriate for those fragile populations.

2.1. The Model

Constructing a model that implements the relaxed zero resource condition $f_i|_{R_i=0} \leq 0$ can be simply done by choosing Michaelis-Menten uptake for all autotroph uptake terms, choosing Holling Type III terms for all predation interactions, and choosing nonlinear mortality for all populations (although other choices are possible). We construct a model with five phytoplankton populations (equivalently, functional types) (P_{1-5}), four zooplankton populations (Z_{1-4}), three small fish populations (F_{1-3}) and two large fish populations (G_{1-2}). Each predator population feeds on all prey in the trophic level immediately below it, while all phytoplankton compete for the single limiting nutrient (N, Fig 1). The model is defined by the equations:

$$\dot{P_i} = P_i \left(\frac{\mu_i N}{N + \kappa_i} - \sum_{j=1}^{4} \frac{\varphi_{i,j} P_i Z_j}{P_i^2 + \kappa_{i,j}} - \sigma_{P_i} P_i \right), \quad i = 1, 2, \cdots, 5 , \tag{3}$$

$$\dot{Z_j} = Z_j \left(\sum_{i=1}^{5} \frac{\varphi_{i,j} \left(1 - \psi_{i,j}\right) P_i^2}{P_i^2 + \kappa_{i,j}} - \sum_{k=1}^{3} \frac{\varphi_{i,j} Z_j F_k}{Z_j^2 + \kappa_{j,k}} - \sigma_{Z_j} Z_j \right), \quad j = 1, 2, \cdots, 4 , \tag{4}$$
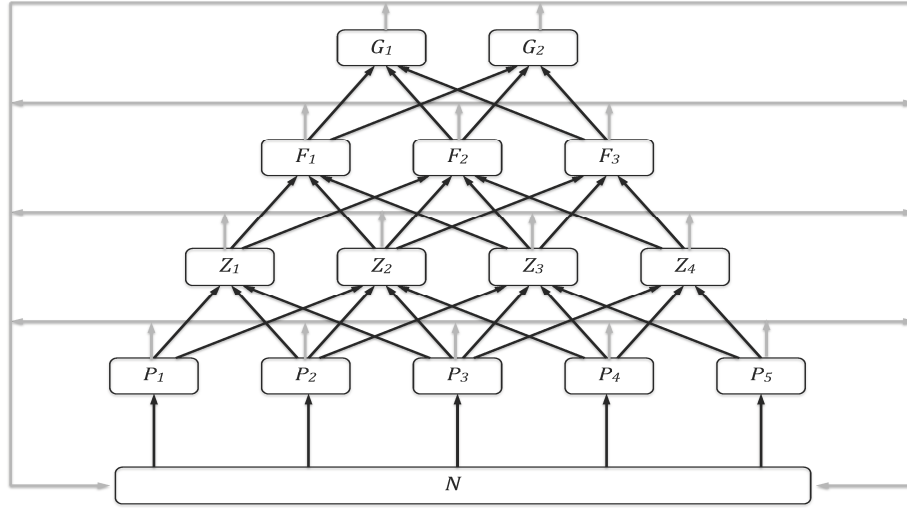
$$\dot{F}_k = F_k \left(\sum_{j=1}^{4} \frac{\varphi_{j,k}\left(1-\psi_{j,k}\right)Z_j^2}{Z_j^2 + \kappa_{j,k}} - \sum_{l=1}^{2} \frac{\varphi_{k,l}F_k^2 G_l}{F_k^2 + \kappa_{k,l}} - \sigma_{F_k} F_k \right), \quad k=1,2,3 \; , \tag{5}$$

$$\dot{G}_l = G_l \left(\sum_{k=1}^{3} \frac{\varphi_{k,l}\left(1-\psi_{k,l}\right)F_k^2}{F_k^2 + \kappa_{k,l}} - \sigma_{G_l} G_l \right), \quad l=1,2 \; , \tag{6}$$

with conservation of mass ($\sum_{i=1}^{5} P_i + \sum_{j=1}^{4} Z_j + \sum_{k=1}^{3} F_k + \sum_{l=1}^{2} G_l + N = 1$) providing an equation describing the dynamics of the inorganic nutrient pool:

$$\dot{N} = -\sum_{i=1}^{5} \dot{P}_i - \sum_{j=1}^{4} \dot{Z}_i - \sum_{k=1}^{3} \dot{F}_i - \sum_{l=1}^{2} \dot{G}_i \; . \tag{7}$$

The model has 138 parameters that need to be estimated to enable the model to reproduce observed data, of which 124 appear in the boundary eigenvalues and hence may be pre-conditioned by the boundary eigenvalue nudging (BEN) method.

Figure 1. Schematic of the plankton ecosystem model. Black arrows indicate flows of mass due to nutrient uptake or predation (for clarity not all these arrows are shown) and grey arrows indicate flows of mass due to implicit bacterial remineralisation of excreted or deceased biological material.

2.2. The Boundary Eigenvalue Nudging (BEN) Method

An n population Kolmogorov system has 2^n sets of equations that define the locations of its equilibrium points. Each of these points has n eigenvalues, so the system has $n2^n$ eigenvalues (ignoring the detail that some sets of equations may define more than one location), $n2^{n-1}$ of which are (trivially) known analytically. The BEN method utilizes the n competition eigenvalues of the n boundary equilibrium points where a single population is zero. We have designed the model to have structural coexistence, a property that is ensured by using non-vanishing terms for all growth processes and vanishing terms for all loss processes (Cropp and Norbury 2012b). Vanishing terms are those that go to zero as the population of interest goes to zero and include terms such as Holling Type III grazing and nonlinear mortality. Non-vanishing terms do not have this property and include terms such as Holling (Types I, II and III) and Michaelis-Menten growth. Structural coexistence ensures that all the ecologically realistic equilibrium points of the model exist in the ecospace. Each of these n competition eigenvalues is given by the life function of the population evaluated at the boundary equilibrium point at which it is zero. For our example model, we note that, due to the structure of the equations (3)-(6), these eigenvalues are composed only of the (positive) growth terms of each life function. We note that this approach, which has a solid theoretical basis, is a preferable solution to the problem of *in silico* competitive exclusion than some of the computational approaches such as using "just so" initial conditions and/or vulnerability functions (for example, Fournier et al. 2011).

The BEN method proceeds by integrating the model to steady state. (Note that substantial numerical analysis of models with structural coexistence has suggested that having a stable interior equilibrium point is a property of such models (Cropp and Norbury 2012b, 2013, Bates et al. 2015) but we have yet to prove this. Having a stable steady state is no more crucial to the success of the BEN method than it is to any other calibration method, however, it does shorten the integration times required.) The magnitude of the difference between the model estimate and the desired solution is calculated for each population:

$$D_i = \left| x_i^* - x_i^D \right| \, , \tag{8}$$

where x_i^* is the model solution and x_i^D is the desired solution. The population with the largest D_i is identified and its key parameters selected for modification. There are many ways of choosing the parameters to be adjusted to improve the model estimates – the simplest is perhaps just to adjust the growth rate of the population up or down depending on whether it is less or greater than its desired population level. Here, we utilize a similarly simple principle: if the population is larger than desired we reduce its maximum growth rate (μ_i in the case of phytoplankton, $\varphi_{i,j}$ in the case of all other populations). The convergence of the method may be improved by further identifying the population that it interacts with that is most wrong in the opposite direction. For example, if a phytoplankton population is too large, we identify the zooplankton population that is too small, and increase the rate of its grazing on the phytoplankton that is too large. Similarly, if a predator population is, for example, too small, we increase its grazing rate on the prey that is too large.

There are also a number of ways to estimate how well each model integration fits the data, and in common with other optimization methods, the choice of these will influence how the method converges and what parameters it adjusts. We have chosen the simple metric of the absolute difference for this example, which has the desirable characteristic of optimizing the system from the lowest trophic levels up. The overall convergence of the method is tracked with the Euclidean Error of each integration m:

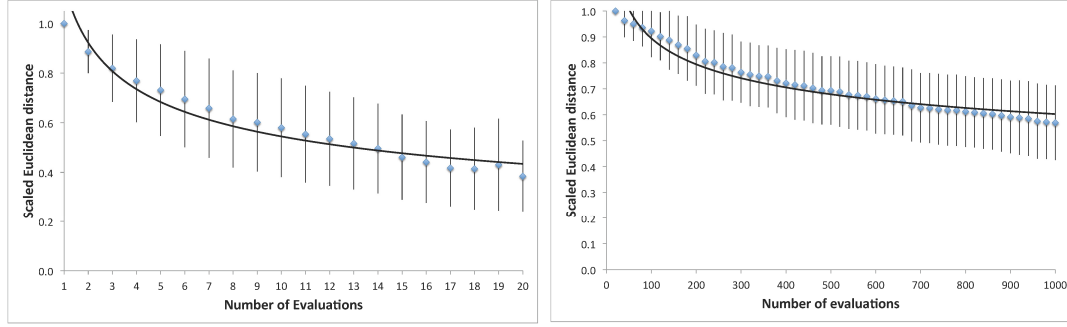$$E_m = \sqrt{\sum_{i=1}^{n} \left(x_i^* - x_i^D \right)^2} \, . \tag{9}$$

The form we have shown optimizes from the lowest trophic levels of the food web up, as its goal function considers only absolute error. Other optimization strategies may be implemented by choosing more sophisticated goal functions. The BEN method was implemented for 10 runs each of 20 model evaluations to collect data on its efficacy and convergence. The final state of each BEN run was used as the initial state for an associated GA optimization to improve the parameter estimates, this time including the 14 mortality parameters not included in the BEN process. This coupling of the BEN and GA methods gives us the Boundary Eigenvalues Nudging – Genetic Algorithm (BENGA) method.

2.3. The Genetic Algorithm (GA) Method

The GA used the final state of its associated BEN run as its initial state and a parameter space of ±50% of the BEN result was defined as the search space. One individual of the initial population was assigned to values of the final BEN result and the other 19 individuals were randomly generated from the search space. One generation of the GA was therefore equivalent in model evaluations to one run of the BEN. The GA was run for 50 generations of a population of 20 individuals (i.e. 1,000 model evaluations). It used 10 bits to represent each parameter (so each individual was represented by a binary string 1,380 bits long) and implemented an initial mutation probability of 0.01 for any bit, a final mutation probability of 0.005, an initial cross-over probability of 0.75 for any individual and a final cross-over probability of 0.975. The probabilities were scaled from the initial to the final values using a power law. Reproduction was implemented using a Monte-Carlo method to select the breeding population. The best individual from the preceding generation was always retained so that the GA best fit, unlike the BEN best fit, never got worse.

3. RESULTS

The convergence of the BEN and GA methods, as a function of the total number of model evaluations is shown in Fig 2. Each of the BEN runs performed 20 model evaluations, while each run of the GA performed 1,000 model evaluations (i.e. the BEN part is equivalent to one generation (of 50) of the GA part). The parameter spaces searched were the same. The data for each run was scaled by its initial value prior to calculating the average and standard deviation of the 10 runs.

Figure 2. Average convergence of 10 runs of the BEN (left) and the GA (right). Error bars are $\pm$ one standard deviation. The curves of best fit are given in (10) and (11).
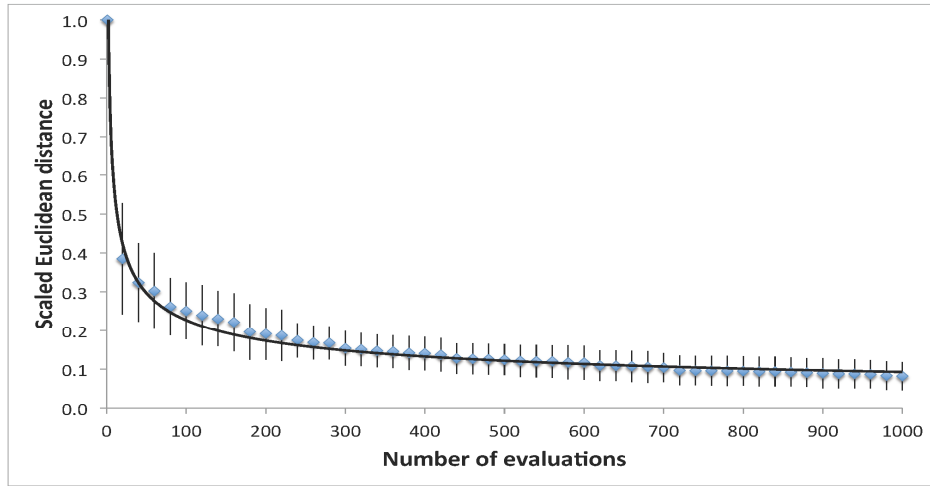
The curve of best fit through the average scaled Euclidean distance for the BEN is:

$$\overline{E}_B = 1.16n^{-0.33}, \qquad R_B^2 = 0.94 \ , \tag{10}$$

while the curve of best fit through the average scaled Euclidean distance for the GA is:

$$\overline{E}_G = 1.97n^{-0.17}, \qquad R_G^2 = 0.93 \ , \tag{11}$$

where $\overline{E}_M$ is the average scaled Euclidean distance between the model estimate and the desired solution for method M and n is the number of model evaluations.

Figure 3. Average convergence of 10 runs of the BENGA. The best starting population for each GA run was the completion point of the associated BEN run. The remainder of the initial population was created a parameter space defined by $\pm 50\%$ of the BEN solution. Legend as for Fig 2.

It is clear from these results that the BEN initially converges more rapidly than the GA and hence is useful to precondition the GA's parameter search space. However, the BEN does not operate on the entire parameter set, and cannot be used independently. Further, the BEN only operates on a very small subset of its parameters at each evaluation (those associated with the worst production terms) and only optimizes one to several parameters per model evaluation to improve the fit of one or two populations. The GA in contrast operates on the whole parameter set to improve the fit of all populations simultaneously. Depending on the probabilities used for mutating and crossing-over particular alleles, which must be optimized for each problem, the GA can optimize many parameters at each model evaluation. The real benefit then lies in combining the two methods. When the methods are combined the benefits of the BENGA are clearly evident (Fig 3). The curve of best fit through the average Euclidean distance between the model estimate and the desired solution for the BENGA is:

$$\overline{E}_{BENGA} = 1.36n^{-0.39}, \qquad R_{BENGA}^2 = 0.97 \ . \tag{12}$$

This relationship predicts that 1,000 model evaluations by the BENGA will result in the Euclidean distance reducing to about 10% of its initial value. In contrast, equation (11) predicts that 1,000 model evaluations by

the GA alone will result in the Euclidean distance reducing to only 60% of its initial value, approaching an order of magnitude difference in convergence. Although these estimates of the rates of convergence are dependent on the number evaluations considered, and for large numbers of evaluations the GA's rate of convergence will dominate, the BENGA evidences a more rapid convergence than the GA.

4.   DISCUSSION AND CONCLUSIONS

The analytical properties of Kolmogorov systems, that allow simple and explicit formulae for the eigenvalues of boundary equilibrium points associated with populations that are identically zero at these points facilitates the development of a new parameter optimization method. The BEN method does not operate on all parameters, nor optimize all populations simultaneously as some other methods do, and cannot optimize parameter sets independently. However, the GA does operate on all parameters and optimizes all populations simultaneously, but GA convergence slows dramatically as the dimension of the parameter space increases (Mitchell 1997). The BEN is useful to efficiently precondition the GA's parameter set by constraining its range. The BENGA evidences more rapid convergence to an optimum solution than either method independently. Suitable choice of the error metric allows the convergence to prioritise different parts of the dynamical system.

REFERENCES

Bates, M., S. M. Bengtson Nash, D. W. Hawker, J. Norbury, J. S. Stark, and R. A. Cropp. 2015. Construction of a trophically complex near-shore Antarctic food web model using the Conservative Normal framework with structural coexistence. Journal of Marine Systems **145**:1-14.

Cropp, R., and J. Norbury. 2014. Comment on "The paradox of the 'paradox of the plankton'" by Record et al. ICES Journal of Marine Science **71**:293-295.

Cropp, R. A., and J. Norbury. 2009a. Parameterizing plankton functional type models: insights from a dynamical systems perspective. Journal of Plankton Research **31**:939-963.

Cropp, R. A., and J. Norbury. 2009b. Simple predator-prey interactions control dynamics in a plankton foodweb model. Ecological Modelling **220**:1552-1565.

Cropp, R. A., and J. Norbury. 2012a. Constructing Ecologies. Journal of Theoretical Biology **294**:1-8.

Cropp, R. A., and J. Norbury. 2012b. The mechanisms of coexistence and competitive exclusion in complex plankton ecosystem models. Ecosystems **15**:200-212.

Cropp, R. A., and J. Norbury. 2013. Modelling plankton ecosystems and the Library of Lotka. Journal of Marine Systems **125**:3-13.

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert. 2011. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software:doi:10.1080/10556788.10552011.

Gabric, A. J., P. A. Matrai, R. P. Kiene, R. Cropp, J. W. H. Dacey, G. R. DiTullio, R. G. Najjar, R. Simó, D. A. Toole, D. A. delValle, and D. Slezak. 2008. Factors determining the vertical profile of dimethylsulfide in the Sargasso Sea during summer. Deep Sea Research Part II: Topical Studies in Oceanography **55**:1505-1518.

Gause, G. F. 1932. Experimental studies on the struggle for existence. Journal of Experimental Biology **9**:389-402.

Gause, G. F. 1934. The Struggle for Existence. Williams and Wilkins, Baltimore.

Huisman, J., and F. J. Weissing. 1999. Biodiversity Of Plankton By Species Oscillations And Chaos. Nature **402**:407-410.

Hutchinson, G. E. 1961. The paradox of the plankton. American Naturalist **95**:137-145.

Kolmogorov, A. N. 1936. Sulla Teoria di Volterra della Lotta per l'Esisttenza. Giornale dell' Instituto Italiano Degli Attuari **7**:74-80.

Kristensen, N., A. J. Gabric, R. D. Braddock, and R. A. Cropp. 2003. Is maximizing resilience compatible with established ecological goal functions? Ecological Modelling **169**:61-71.

Mitchell, M. 1997. An introduction to genetic algorithms. MIT Press, Cambridge, Massachusetts.

Record, N. R., A. J. Pershing, and F. Maps. 2013. The paradox of "the paradox of the plankton". ICES Journal of Marine Science.

Schippers, P., A. M. Verschoor, M. Vos, and W. M. Mooij. 2001. Does "supersaturated coexistence" resolve the "paradox of the plankton"? Ecology Letters **4**:404-407.