Report Cover Page

ACERA Project
07/04
Title
Simplifying models of alien biota
Author(s) / Address (es)
John Hearne, RMIT University
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)
Final Report

Summary

A new method has been developed for testing the sensitivity of the response functions used in dynamic models. The method has been illustrated with a simple model involving an agricultural product and the biological control of a pest by a parasitoid population.

A model found in the recent literature and involving a biological control problem in Australia was then chosen for further work. The most insensitive functions were identified as candidates for simplification. It turned out that there was limited opportunity for simplification in this particular model. It was, however, possible to show that for one particular response function identified by the method, a simpler formulation performed sufficiently well.

A literature review of methods for model simplification has been conducted and some recommendations are made as to which methods which best suit dynamic models of alien biota. Algorithms for some of these methods are given for easy implementation.

ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	ACERA / AMSI SAC Approval:	Date:

Simplifying models of alien biota: 0704

John Hearne; RMIT University

Final Report

April 2009

Acknowledgements

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and Australian Research Centre for Urban Ecology (ARCUE).

Disclaimer

Table of contents

List of Tables

List of Figures

1. Executive Summary

Many models dealing with biosecurity problems are large and complex. This makes it difficult or even impossible to analyse the system thoroughly and hence their validity might be called into question. Much of the complexity is often unnecessary but identification of redundant variables, parameters, and unnecessary complexity usually cannot be achieved during model formulation.

In this project attention was given to identifying the most insensitive functions used in a model. These functions become candidates for simplification and hence the possibility of achieving a reduction in model complexity. A new pragmatic approach was proposed to identify the relative sensitivity of model functions. This approach was illustrated on a simple stochastic model involving the biological control of an agricultural pest. The method was then tested in a case study on a published model that looked at the possibilities of the biological control of cat’s claw creeper in Australia. Although it later transpired that this model was not a good candidate for model reduction it was possible to show that some simplification could be achieved using the proposed method.

There are a number of other methods for dealing with aspects of model simplification. A literature review of these methods was conducted and some recommendations are made as to which methods best suit dynamic models of alien biota. Algorithms for some of these methods are given for easy implementation. All this is included in a separate report.

2. Introduction

There are many modelling software packages available today. Their ease of use has made it possible to formulate large complex models with limited mathematical ability. This has led to their widespread use to evaluate candidate management strategies aimed *inter alia* at minimising damage, or risk of damage, from alien or pest biota. Some Australian examples include those dealing with prickly acacia (Kriticos et al, 2003), rice blast disease (V. Lanoiselet *et al,* 2002), striped rust (White et al, 2004) and cat's claw creeper (Raghu *et al*, 2007).

Large complex models comprise effects such as those due to competition and parasitism along with standard processes such as production or growth. It is often difficult during the model formulation phase to identify which interactions might be critical to model output so there is a tendency towards greater complexity rather than parsimony.

Large complex models have a number of drawbacks: Large data requirements can make them difficult to calibrate (Scholten et al. 1998; Toal et al. 2000; Jorgensen et al. 2002) and validate (Scholten et al. 1998; Homann et al., 2000; Bugmann, 2001). It is difficult to draw conclusions about the underlying causes of system behaviour (Scholten et al., 1998) and hence generate confidence in the model amongst potential users of the results; The models might be over-parameterised and hence have poor predictive performance; Computational considerations are a disincentive to thoroughly testing a model and testing alternative formulations; Extending the scope of a model (eg from local to landscape) can become computationally prohibitive; More importantly, they do not necessarily perform better than simpler models (Hakanson, 1995; Fulton, 2001; Fulton et al., 2003, 2004). Indeed, the sensitivity of the model output to the input tends to increase with complexity (Snowling and Kramer, 2001; Lindenschmidt, 2006). With all the above problems possible, a reduction in model complexity has clear advantages.

The problem of reducing the complexity of dynamic models is widely covered in the engineering literature. For some reviews see Hirata (1987) and Antoulas et al (2001). Most of this relates to linear systems. Even in the context of complex ecological models the most common model structures examined are the linear (O'Neill and Rust, 1979; Cale and Odell, 1980) and Lotka-Volterra forms (O'Neill and Rust, 1979; Luckyanov et al,. 1983). A further problem encountered with some of these methods is that the reduced systems contain some state variables and parameters that have no biological interpretation.

Fulton (2001) managed to reduce the number of state variables in the Port Phillip Bay Ecosystem Model by aggregating variables. This success was dependent on a combination of system knowledge, trial and error, and considerable computational effort. No general method arises from this work. More automated methods have been developed for specific objectives (Pinnegar et al., 2005. Cox, et al, 2006, and Lawrie and Hearne, 2007 and 2008) but they require further development for other objectives.

The aim of this project was to find methods for simplifying large complex models that are designed to find control strategies for alien biota. In order to identify components of a model that are candidates for simplification it is necessary to identify the relative sensitivity of these components. Furthermore, a method was needed for a functional sensitivity analysis that is easy to implement. Such a method was proposed and first tested on a simple model for illustrative purposes. The method was then applied as a case study on a model found in the literature. This model was formulated to investigate a control strategy for cat's claw creeper.

3. Methodology

The objectives of this project were the following:

- To develop automated methods for reducing the complexity of existing dynamic models of invasive alien biota without loss of biological interpretation or decision-making capability.
- To develop methods for determining the sensitivity of functions used in a model
- Hence, to propose and illustrate a process, comprising a suite of methods, for simplifying models, thus making them amenable to more thorough strategy and sensitivity analysis, easier for managers to understand, and more suitable for expanding to larger spatial and temporal scales.

The second objective was addressed first. A new method for analysing the sensitivity of response functions used in ecosystem models was proposed.

Consider the following system of difference equations:

$$x_i(t+\Delta t) = x_i(t) + f_i(\underline{x}, t, \underline{\alpha})\Delta t,\ i = 1,2,\dots,n, \tag{1}$$

where $\underline{x}$ is the state vector and $\underline{\alpha}$ a vector of parameters. In ecological models the functions f_i comprise terms representing the rates of the system such as birth rate and death rate. Thus we can write:

$$f_i(\underline{x}, t, \underline{\alpha}) = \sum_j r_{ij}(\underline{x}, t, \underline{\alpha}) \tag{2}$$

A basic parameter sensitivity analysis involves changing the values of the parameters $\underline{\alpha}$ in the functions $r_i(\underline{x}, t, \underline{\alpha})$ by a small amount, one at a time, and observing the change it produces in the output. A possible pragmatic approach for function SA is to multiply each function or rate by a parameter with a nominal value of one. These parameters can then be perturbed as is done in a parameter SA. This should yield some indication as to which rates are the most sensitive. This method was tried with some success by Lawrie and Hearne (2007). But it does not yield any information about the sensitivity of the model output to changes in the shape of the functions. In some cases changing the parameters in the functions $r_i(\underline{x}, t, \underline{\alpha})$ will change the shape of the functions to some extent but it does so in very restrictive ways. Clearly other changes to the shape of the functions are possible. In ecological models there is frequently uncertainty about the appropriateness of the functional form chosen for the model. The choice of functional form will affect the model solutions and hence possibly the conclusions drawn from these results. It is therefore important to obtain a better idea of how sensitive the models results are to changes in the functions $r_i(\underline{x}, t, \underline{\alpha})$.

The simplest approach towards this end, going beyond the method mentioned above, is to multiply, one at a time, each rate r_{ij} by the following function which comprises a product of triangular-shaped functions:

$$H_{ij}(\underline{x}, \underline{p}, \underline{m}) = \prod_{i=1}^{n} h_i(x_i, p_{ij}, m_{ij}),$$

(3)

where

$$h_i(x_i,p_i,m_i) = \begin{cases} 1 + m_{ij}(x_i - c)/(p_{ij} - c) & \text{if } a_i \le x_i \le b_i \\ 1 & \text{otherwise} \end{cases}$$

$$\text{and } c = \begin{bmatrix} a_i \text{ if } x_i \le p_{ij}, \\ b_i \text{ if } x_i > p_{ij}. \end{bmatrix}$$

(4)

Note that m_{ij} can be negative or positive. Reasonable choices of a_i and b_i are the respective minimum and maximum values of the corresponding state variable x_i over the solution interval $[a_i, b_i]$.

Changes in the functions can now be effected by changes in the parameters corresponding to the position and magnitude of the peak in the respective h-functions. The sensitivity analysis proceeds by testing the effect of such changes in the function for each rate on some decision criteria. This is best described with the illustrative example given in Hearne (submitted) involving an agricultural product, a pest and a parasitoidal biological control agent

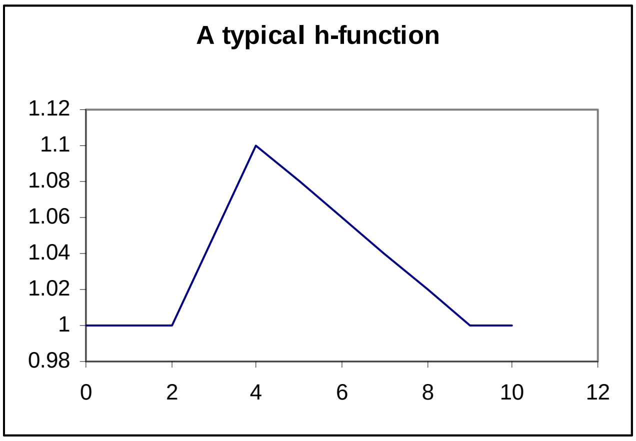

Figure 2: An example of an h-function. Here a=2, b=9, m=0.1 and p=4.

The method can be automated and is easy to implement. It does, however, require optimisation software. While such software is readily available for most platforms this is not the case for some modelling software systems such as DYMEX.

To meet objectives 1 and 2 we had hoped to use a DYMEX model. We were unable to secure permission to access the code of the DYMEX engine. We then commenced the task of programming the SPANDX model in another language. We found that the paper on this model did not provide sufficient information to reproduce the model. We were unable to get responses from the authors at this stage presumably due to other commitments that they had. We then searched the literature again looking for a suitable model that involved alien biota in an Australian context and was completely described so that we could reproduce it. We were amazed at how many models had incomplete descriptions although this lack of information only became apparent when preparing the published model for our own software. Eventually, we had some success with a model developed by Raghu et al. [2007] on predicting the risk involved to a non-target species when biological control is used to manage an invasive plant species (cat's claw creeper).

The function sensitivity method was applied to this model. The automated method identified the least sensitive functions and hence candidates for simplification. It turned out that this model was not a good candidate for model simplification. Nevertheless one of the response functions was simplified and the simplification shown not to affect the results significantly. Thus it was possible to illustrate our new method but in a limited way. More details of the method and the results can be found in Bennett and Hearne (submitted).

4. Issues

It was unfortunate that a large proportion of the budget was spent on a postdoctoral fellow who subsequently resigned after several months before any product was achieved. There was a severe limit on what could be achieved after that. Nevertheless, to some extent at least the objectives and deliverables have all been met.

The original intention was to illustrate the methods developed on the model SPANDX. We received a DYMEX version of this model but had trouble running it and producing similar results to those published from this model. It was also hard to manipulate the model beyond standard analyses such as parameter sensitivity analysis. We then hoped to access the DYMEX source code to increase our manipulative abilities. We were unable to secure permission to access the DYMEX source code. We tried to rectify this by re-writing the model in a programming language. When attempting to do this it became apparent that despite having the DYMEX software, the model, and a publication comprising a description of the model with some data and results, there were not enough details to reproduce the model in another format. An investigation of other models in the literature revealed similar missing details. This lack of detail to enable another to reproduce the results published should be a source of grave concern to scientists.

5. Recommendations

There are well-developed methods for undertaking a parameter sensitivity analysis of a model. And the importance of this has not gone unnoticed. It is rare to find any published model that has not undergone such an analysis is some form or another. Equally, if not more important, is an analysis of the effect of the functional forms used in models yet this rarely done. Although further development is needed the method proposed and illustrated in this project provides a reasonably simple and automated way for functional sensitivity analysis.

Functional sensitivity analysis plays a similar role to parameter sensitivity analysis in that it identifies where resources should be focussed to reduce uncertainty in model results. It has the further advantage in that by identifying functions that are in some sense ‘insensitive’ the analysis identifies candidates for simplification.

The advantages of simpler models are well-known. A literature review of methods for model simplification has been undertaken and recommendations made as to which methods to use. The recommendations are based on ease of implementation. This has all been included in a separate report (Bennet and Hearne, 2009).

6. References

Anatoulos, A.C., Sorensen, D.C., Gugercin,S. 2001. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, 280:193-219.

Bennet, JC & Hearne, JW. (2009) A Guide to Model Simplification. Report presented to ACERA.

Bennet, JC & Hearne, JW ‘Sensitivity analysis of model functions’, submitted

Bugmann, H. A review of forest gap models. *Climate Change*, 51:259-305.

Cale, W.G.Jr, Odell, P.L., 1980. Behaviour of aggregate state variables in ecosystem models. Mathematical Biosciences, 49, 121-137.

Cox, G.M., Gibbons, J.M., Wood, A.T.A, Craigon,J.,Ramsden,S.J., Crout, N.M.J. 2006. Towards the systematic simplification of mechanistic models. Ecol. Model. 198:24-246.

Fulton, E.A., 2001. The effects of model structure and complexity on the behaviour and performance of marine ecosystem models. Ph.D. Thesis, School of Zoology, University of Tasmania, Hobart, Australia.

Fulton, E.A., Smith, A.D.M., Johnson, C.R., 2003. Mortality and predation in ecosystem models: is it important how these are expressed? Ecol. Model. 169, 157-178.

Fulton, E.A., Parslow, J.S., Smith, A.D.M., Johnson, C.R., 2004. Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. Ecol. Model. 173, 371-406.

Hakanson, L., 1995. Optimal size of predictive models. Ecol. Model. 78, 195-204.

Hearne, J., Haque, E. & Stacey, A. (2007). 'A Technique for the Sensitivity Analysis of Functions in Relation to Decision-Making Objectives' *Land, Water and Environmental Management: Integrated Systems for Sustainability,* Oxley, L and Kulasiri, D (ed.), MODSIM 2007 International Congress on Modelling and Simulation, Canterbury, New Zealand

Hearne, JW ‘An automated method for extending sensitivity analysis to response functions’, submitted.

Hirata,H. 1987. Modelling and analysis of ecological systems: the large-scale system viewpoint. Int. J. Systems Sci. 18:1839-1855.

Homann, P.S., McKane, R.B., Sollins, P., 2000. Below-ground processes in forest-ecosystem biogeochemical simulation models. Forest Ecology and Management, 138 (1-3), 3-18.

Jorgensen, S.E., 1976. A eutrophication model for a lake. Ecol. Model. 2, 147-165.

Jorgensen, S.E., Ray, S., Berec, L., Straskraba, M., 2002. Improved calibration of a eutrophication model by use of the size variation due to succession. Ecol. Model. 153, 269-277.

Lawrie, JS & Hearne JW, 2007. "Reducing Model Complexity Via Output Sensitivity", Ecological Modelling, 207:137-144

Lawrie, Jock & Hearne, John 2008, ‘A method for aggregating state variables in large ecosystem models’ in Mathematics and Computers in Simulation 79:368–378

Lindenschmidt, K.E., 2006. The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. Ecol. Model. 190 (1-2), 72-86.

.
.Lukyanov, N.K., Svirezhev, Yu.M., Voronkova, O.V., 1983. Aggregation of variables
.in simulation models of water ecosystems. Ecol. Model. 18, 235-240.
.
.Murray, A., Parslow, J., 1997. Port Phillip Bay Integrated Model: Final Report. Technical
.Report No. 44. Port Phillip Bay Environmental Study. CSIRO, Canberra, Australia.
.
.Murray, A., Parslow, J., 1999. The analysis of alternative formulations in a simple model
.of a coastal ecosystem. Ecol. Model. 119 (2-3), 149-166.
.

. O'Neill, R.V., Rust, B., 1979. Aggregation error in ecological models. Ecol. Model. 7, 91-105.
.

Raghu S., Dhileepan K., Scanlan J. C., (2007) Predicting risk and benefit a priori in biological control of invasive plant species: A systems modelling approach, *Ecological Modelling* 208: 247-262.
.

. Pinnegar, J.K., Blanchard, J.L., Mackinson, S., Scott, R.D., Duplisea, D.E., 2005. Aggregation and removal of weak-links in food-web models: system stability and recovery from disturbance. Ecol. Model. 184, 229-248.
.

. Scholten, H., Van Der Tol, M.W., Smaal, A.C., 1998. Models or measurements. Meet.of the Int. Counc. for the Exploration of the Sea, Cascais, Portugal, 16-19 Sep. ICES, Copenhagen, Denmark.
.

. Snowling, S.D., Kramer, J.R., 2001. Evaluating modelling uncertainty for model selection. Ecol. Model. 138 (1-3), 17-30.12
.

7. Appendix

Manuscripts arising out of this project

Bennet, JC & Hearne, JW. (2009) A Guide to Model Simplification. Report presented to ACERA.

Bennet, JC & Hearne, JW 'Sensitivity analysis of model functions', submitted

Hearne, J., Haque, E. & Stacey, A. (2007). 'A Technique for the Sensitivity Analysis of Functions in Relation to Decision-Making Objectives' *Land, Water and Environmental Management: Integrated Systems for Sustainability*, Oxley, L and Kulasiri, D (ed.), MODSIM 2007 International Congress on Modelling and Simulation, Canterbury, New Zealand

Hearne, JW 'An automated method for extending sensitivity analysis to response functions', submitted.

A guide to model simplification

J. C. Bennett* J. W. Hearne†

1 Introduction

There is extensive literature on the subject of model simplification. Due to the advances in computational technology modellers are employing increasingly detailed models to describe a dynamical system. The added complexity of models requires an increase in storage requirements and results in an increased computation time. Even with the increases in computational speed, modellers are often forced to employ simplification techniques to obtain a more tractable model.

A desirable feature for a model simplification technique is little or no requirement of model information a priori. Such a method is then applicable to any dynamical systems model and can be implemented automatically to determine a simpler model. Another desirable property is computational speed. It can be computationally restrictive to perform a sensitivity analysis on a complex model. If the simplification technique is applied then it may be possible to perform a sensitivity analysis on the reduced model. A further desirable objective is to produce a reduced model that is still physically interpretable.

Many of the techniques throughout the literature require specific model structures and hence are not widely applicable in a general sense to any large complex model. However other techniques require no model information *a priori* and may be implemented in a systematic way. Another limitation is the requirement to have access to the model code. This becomes an issue when large scale systems are constructed in a package, for example DYMEX. There is no access at a user level to the system code and hence the methods described here can not be applied.

*School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia. mailto:james.cameron.bennett@rmit.edu.au

†School of Mathematical and Geospatial Sciences, Royal Melbourne Institute of Technology, Melbourne, Australia. mailto:john.hearne@rmit.edu.au

This guide gives a brief description of some of the methods commonly employed for model simplification with some insight as to the most easily implementable and applicable to *any* dynamical systems model.

2 Model simplification techniques

Cox et al. [6] use a systematic approach to model simplification by replacing state variables with their nominal mean values over the simulation, i.e. the average values of the state variables in the original model simulation. The process begins with identifying possible variables for replacement by individually replacing them with their mean values. If the replacement did not more than double the residual sum of squares with respect to the original data set then this variable was a candidate for replacement. Once these replacement candidates are identified, all combinations are tested against the original parameterisation data set to obtain alternate model formulations. This is a systematic method and easily implementable, however, it is computationally exhaustive particularly if the process identifies many replacement variables and if the original model is computationally expensive.

Lawrie and Hearne [11] introduce two methods for model simplification. The first method: Advanced Rate Elimination Method (AREM) identifies rates that are small and subsequently eliminates them. The second method: Variable Simplification Method (VSM) identifies state variables that may be approximated by a constant and replaces them. AREM involves measuring the sensitivity of each of the models diagnostics to the rates contained in the model. The disadvantage of this method is that the model has to be run the amount of rates that are present in the model which can be computationally expensive. However, Lawrie and Hearne [11] direct the reader to a simpler method [10] for an initial step before the more comprehensive AREM is carried out to improve computation time. The Variable Simplification Method identifies variables in the system that may be approximated by a constant, i.e. the derivative may be set to zero. It is very easy to implement and requires no system knowledge a priori.

Martin [14] introduces a method of simplifying a component of complex process-based models using canonical nonlinear S-system modelling. A drawback of this method is the need to specify interactions in the model a priori, requiring model knowledge. Also the parameters and expression have no physical interpretation. Therefore, as far as an easily employable scheme requiring no system knowledge a priori and obtaining physically interpretable simpler models, this method is too restrictive.

Lawrie and Hearne [12] employ a method of aggregating variables based on a least squares criterion to a model of the nitrogen cycle in Port Phillip Bay. They introduce a method to systematically select variables for aggregation based on a modified technique of Proper Orthogonal Decomposition (POD) [16, 18] which they call Partial Proper Orthogonal Decomposition (PPOD). The basic idea of POD is minimising the Euclidean distances between an original data set and an approximation to that data set, i.e. the line of best fit. In multi-dimensions the best fit is a hyperplane. After the hyperplane of best fit is identified from the simulated data, both the data and the time derivative vectors are projected onto the plane. The result is a reduction in model complexity. Lawrie and Hearne's [12] PPOD method may be used to enhance system knowledge or dispel possible theories based on the aggregates generated. In some cases the choice of a model aggregate leads to a model which is physically unrealistic and hence the authors stress that the method should be used as a guide and verified via simulation. However, the simplified model obtained was physically interpretable and PPOD was successful in its application to the Port Phillip Bay model with a reduction of 5–7 (of the 29 present in the full model) state variables.

Asgharbeygi et al. [2] approach model revision by generating a series of alternate models based on an initial model. Their algorithm searches model space finding appropriate models based on user-specified constraints that control which processes of the model are fixed, may be removed or allow changes to their parameters. The algorithm also searches user defined generic processes that may be added to the initial model and each revised model is fitted to observed data. This method requires system knowledge and the possible inclusion of extra processes requires a degree of effort in the initial step. This method is a useful tool for model validation and is particularly useful in the development of an unfinished model.

Kooi et al. [8], Auger and Bravo de la Parra [4] and Auger et al. [5] use singular perturbations to aggregate variables. These methods require the assumptions that equilibria and limit cycles exist which in general for large scale models often don't exist at all. Hence as far as adapting to any dynamical systems model this method is too restrictive. Singular perturbations have been used to omit state variables. This involves omitting the state variables by replacing them with constants based on differing time scales [13, 17]. Due to the complexity of modern models this method is difficult to apply.

Antoulas et al. [1] provide an overview of model simplification methods. These methods are split into two main categories which are SVD (Singular Value Decomposition) based and moment matching based methods. They consider seven model order reduction algorithms. These are Balanced Model Reduction, Approximate Balanced Reduction, Singular Perturbation Method, Hankel Norm Approximation, Arnoldi Procedure, Lanczos Procedure and Rational Krylov Method. Of these methods Antoulas et al. [1] find that Balanced Model Reduction and Approximate Balanced Reduction perform the best out of these methods when the whole frequency range was considered and the Approximate Balanced Reduction method requires less computational effort and less storage requirements. However these methods are applied to linear systems hence for the highly nonlinear complex models that exist (particularly in ecological modelling) these methods are insufficient. The SVD based nonlinear systems method for model reduction is Proper Orthogonal Projection (POD) [16, 18].

A different approach to addressing the problem of model complexity is the idea of automated processes for model construction [3, 7, 9]. Džeroski and Todorovski [7] construct their model based on data and domain specific knowledge. They search the space of feasible model structures using the grammar-based equation discovery method Lagramge and select the one that fits the best based on the measured data. Todorovski and Džeroski [19] apply their modelling framework to real world modelling tasks and are able to produce a comprehensible model that best fits the measured data. This method is in contrast to the intension of this guide since we are assuming the complex model already exists and requires simplification.

3 Recommendations

In this Section we make some recommendations for the process of simplifying any large scale model. Therefore methods that do not require system knowledge a priori are required. The intention of this guide is to provide modellers a quick reference as the most easily implementable and simple methods for reducing the complexity of *any* large complex model. In what follows we will give a basic algorithm for the selected models.

3.1 The Advanced Rate Elimination Method

For simplicity and comparability we adopt the terminology used in Lawrie and Hearne [11] for the AREM, VSM and PPOD method in what follows. Consider a model of the form

$$\dot{x} = f(x, u, t; p); \quad y = y(x) \tag{1}$$

where $x(t) \in \mathbf{R}^n$, for some n, is the state vector at time t, p is a parameter vector, $y \in \mathbf{R}^d$, for some d, is a vector of diagnostics of the model and u is the input vector. The initial state is $x(0) = x_0$ and the time domain is $0 \leqslant t \leqslant T$ for some time T.

In what follows, the set keep_m is the list of rates f_{jk} to be kept under input $u_m(t)$, $m = 1, \ldots, s$ and u is the set of all inputs. The set keep is the set of rates to be kept under all inputs u. The algorithm for the AREM is carried out for each model diagnostic y_i and is as follows:

1. Replace f_{jk} with $p_{jk}f_{jk}$, where f_{jk} is the k^{th} rate of the j^{th} derivative, p_{jk} is a parameter and α_j is the number of rates in f_j.
2. Set keep={}, $m = 1$;
3. Set $j = 1$, $\text{keep}_m = \{\}$;
4. Set $k = 1$;
5. Run the model with $p_{jk} = 1.01$ (i.e. $\Delta p = 0.01$)
6. Calculate $s_{ijk} = (\hat{y}_i - y_i) / \Delta p_{jk} y_i$ which is a normalised approximation of $\partial y_i / \partial p_{jk}$.
7. **if** $|s_{ijk}| > tol$ **then**
   Add f_{jk} to keep_m, set $k = k + 1$;
   else
   Set $k = k + 1$;
   end if
8. **if** $k > \alpha_j$ **then**
   Set $j = j + 1$, go to step 9;
   else
   Return to step 5;
   end if
9. **if** $j > n$ **then**
   Set $m = m + 1$;
   else
   Return to step 4;
   end if
10. **if** $m > s$ **then**
    Set keep=$\cup_m \text{keep}_m$, break;
    else
    Return to step 3;
    end if

3.2 The Variable Simplification Method

In what follows, the set simp_m is the list of variables to be made constant under input $u_m(t)$, $m = 1, \ldots, s$ and u is the set of all inputs. The set simp is the set of variables are made constant in the reduced model under all inputs u. The algorithm for the VSM is carried out for each model diagnostic y_i and is as follows:

1. Replace f_j with $p_j f_j$, where f_j is the j^{th} derivative and p_j is a parameter.

2. Set simp={}, $m = 1$;

3. Set $j = 1$;

4. Run the model with $p_j = 1.01$ (i.e. $\Delta p = 0.01$)

5. Set $\text{simp}_m = \{\}$;

6. Calculate $s_{ij} = (\hat{y}_i - y_i)/\Delta p_j y_i$ which is a normalised approximation of $\partial y_i / \partial p_j$.

7. **if** $|s_{ij}| < tol$ **then**
   Add x_j to simp_m, set $j = j + 1$;
   else
   Set $j = j + 1$;
   end if

8. **if** $j > n$ **then**
   Set $m = m + 1$;
   else
   Return to step 4;
   end if

9. **if** $m > s$ **then**
   Set simp=$\cap_m \text{simp}_m$, break;
   else
   Return to step 3;
   end if

3.3 Partial Proper Orthogonal Decomposition

The method of Partial Proper Orthogonal Decomposition (PPOD) is a modification of Proper Orthogonal Decomposition (POD). A reduced model after carrying out POD (see Rathinham and Petzold [16] for details) is

$$\dot{\hat{x}}(t) = Pf(\hat{x}(t), u(t), t) \tag{2}$$

where $\hat{x}(t) \in S$, where S is a k-dimensional ($k < n$) affine space and n is the dimension of the full model state vector. POD minimises the Euclidean distance between a vector x and an approximation $\hat{x}$ and hence typically focusses on the components with the largest variance. PPOD reduces this focus by projecting the normalised version of the state variables:

$$\frac{x_i - \bar{x}_i}{\sigma_i}, \tag{3}$$

where $\bar{x}_i$ is the mean and σ_i is the standard deviation of x_i. The difference in the projection matrix:

$$P = \rho^T \rho, \tag{4}$$

is that it is not constructed from the eigenvalues of the covariance matrix (as in POD) but the eigenvalues of the correlation matrix. Hence ρ is a $k \times n$ matrix where the rows correspond to the k eigenvectors relating to the k largest eigenvalues of the correlation matrix. Since POD increases the complexity of the approximated derivatives [12], PPOD projects mutually exclusive subsets of the state variables. The final modification that PPOD makes to POD is that as well as the aggregates being projected, their sums are also projected. Hence the projection is applied to $(x_{i1}, \ldots, x_{i\alpha}, x_{i1} + \ldots + x_{i\alpha})$, where α is the number of variables in the aggregate. The aggregation of the x_{ij} is validated by the size of

$$E_{\text{agg}} = \frac{\int_0^T \left(\frac{d}{dt}\,\hat{\text{agg}} - \frac{d}{dt}\,\text{agg}\right)^2 dt}{\int_0^T \left(\frac{d}{dt}\,\text{agg}\right)^2 dt}, \tag{5}$$

where

$$\frac{d}{dt}\,\hat{\text{agg}} = P_{\alpha+1} \cdot f_{\text{agg}}(x^{\text{agg}}, \text{agg}, u, t), \tag{6}$$

and x^{agg} is the vector of the x_i that are not in the aggregate agg. $P_{\alpha+1}$ corresponds to the last row of the projection matrix P. If the x_i and $\hat{x}_i$ are close, and if agg and $\hat{\text{agg}}$ are close then (5) will be small.

In the algorithm to follow agg_{jm} denotes the list of variables that are aggregation candidates with x_j under the input $u_m(t)$, and agg_j is the list of reduced model candidate variables for aggregation with x_j. For a more detailed explanation of PPOD refer to Lawrie [10] and Lawrie and Hearne [12]. The algorithm for choosing the aggregates is as follows:

1. Set $m = 1$;

2. Run the model;

3. Set $j = 1$;

4. Set $\text{agg}_{jm} = \{\}$, $\text{agg}_j = \{\}$;
5. $\forall x_i \notin \text{agg}_{jm}$, calculate E^i_{agg}, where $\text{agg} = \sum_A x_\alpha$, and $A = \cup\{\{x_i\}, \{x_j\}, \text{agg}_{jm}\}$;
6. Find k such that $E^k_{\text{agg}} = \text{Min}(E^i_{\text{agg}})$.
7. **if** $E^k_{\text{agg}} \leqslant tol$ **then**
   Add x_k to agg_m, Return to Step 5;
   else
   if $j = n$ **then**
   Set $m = m + 1$;
   else
   Set $j = j + 1$, Return to Step 4;
   end if
   end if
8. **if** $m > s$ **then**
   Set $\text{agg}_j = \cap_m \text{agg}_{jm}$, break;
   else
   Return to step 2;
   end if

If there is a set of indices $J = \{i_1, \ldots, i_\alpha\}$ such that $\text{agg}_{ia} = \text{agg}_{ib}, \forall a, b \leqslant \alpha$, then the aggregate $\{x_{i1} + \cdots + x_{i\alpha}\}$ is substituted for $\{x_{ij}\}$, $ij \in J$ in the reduced model.

3.4 The Variable Replacement Method

Consider a model of the form:

$$y_j = f(I_j, \theta) + \epsilon_j, \quad j = 1, \ldots, n, \tag{7}$$

where n is the number of variables, y_j is the response, I_j is the input variables, θ is the parameter vector, and ϵ_j are normally distributed independent random error terms with mean zero and variance σ^2.

The algorithm implemented by Cox et al. [6] is as follows:

1. Run the model and calculate the mean of each variable over the simulation;
2. Replace each variable separately with its corresponding mean value and calculate the residual sum of squares, i.e., calculate:

$$\sum_{j=1}^{n} (y_j - f(I_j, \theta))^2; \tag{8}$$

3. If a variables replacement does not more than double the residual sum of squares then it is a replacement candidate;

4. For all combinations of replacement candidates generate new models based on a specified selection criterion (see Cox et al [6] for details) and estimate the adjustable parameters using the Marquart procedure [15].

This method is very simple but may be computationally restrictive particularly when a lot of replacement candidates are identified.

3.5 Discussion

The methods described above are chosen for their simplicity and their ability to be applied to any dynamical system. They are all automated processes that do not require any system knowledge *a priori*. PPOD is the most complicated of the methods described here. The method of Cox et al. [6] is very simple and leads to a physically realistic model. However, it may be computationally expensive if a lot of variable replacement candidates are identified. The methods described in Lawrie and Hearne [11] namely AREM and VSM, were proven to be effective on the complex Port Phillip Bay model they considered. AREM may be computationally expensive if the model contains a lot of rates. However, Lawrie and Hearne [11] suggest an initial screening with a simpler method [10] before carrying out AREM to improve computation time. The VSM is simple and easy to implement and is faster than AREM, however, unlike Cox et al.[6] VSM does not consider different combinations of approximated state variables.

For its simplicity, the authors recommend Cox et al.'s [6] method. However, if this proves to be too computationally expensive, the authors recommend Lawrie and Hearne's [11] AREM (with the possible implementation of the simpler method described by Lawrie [10]) or VSM. For more advanced model simplifiers, Lawrie and Hearne's [12] modification of POD, namely PPOD, is very effective at enhancing system knowledge. These methods may all be applied with no system knowledge. For PPOD the resulting aggregations should be verified by simulation to discount unrealistic simpler models.

3.6 Acknowledgements

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA project 07/04). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and Australian Research Centre for Urban Ecology (ARCUE).

References

[1] Antoulas A. C., Sorensen D. C., Gugercin S., A survey of reduction methods for large-scale systems, *Contemporary Mathematics*, Vol. 280, 2001, pp. 193–219.

[2] Asgharbeygi N., Langley P, Bay S., Arrigo K., Inductive revision of quantitative process models, *Ecological Modelling*, Vol. 194, 2006, pp. 70–79.

[3] Atanasova N.,Todorovski L., Džeroski S.,Kompare B., Constructing a library of domain knowledge for automated modelling of aquatic ecosystems, *Ecological Modelling*, Vol. 194, 2006, pp. 14–36.

[4] Auger P., Bravo de la Parra R., Methods of aggregation of variables in population dynmaics, *Comptes Rendus de l'Académie des Sciences Série III, Sciences de la Vie*, Vol. 323, 2000, pp. 665–674.

[5] Auger P., Charles S., Viala M., Poggiale J. C., Aggregation and emergence in ecological modelling: integration of ecological levels, *Ecological Modelling*, Vol. 127, 2000, pp. 11–20.

[6] Cox G. M., Gibbons J. M., Wood A. T. A., Craigon J., Ramsden S. J., Crout N. M. J., Towards the systematic simplification of mechanistic models, *Ecological Modelling*, Vol. 198, 2006, pp. 240–246.

[7] Džeroski S., Todorovski L., Learning population dynamics models from data and domain knowledge, *Ecological Modelling*, Vol. 170, 2003, pp. 129–140.

[8] Kooi B. W., Poggiale J. C. Auger P., Aggregation methods in food chains, *Mathematical and Computer Modelling*, Vol. 27(4), 2000, pp. 109–120.

[9] Langley P., Shiran O., Shrager, Todorovski L., Pohorille A., Constructing explanatory process models from biological data and knowledge, *Artificial Intelligence in Medicine*, Vol. 37, 2006, pp. 191–201.

[10] Lawrie J., Hearne J. W., Reducing model Complexity of Large Ecosystem Models, Ph.D. Thesis, School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia, 2006.

[11] Lawrie J., Hearne J. W., Reducing model complexity via output sensitivity, *Ecological Modelling*, Vol. 207, 2007, pp. 137–144.

[12] Lawrie J., Hearne J. W., A method for aggregating state variables in large ecosystem models, *Mathematics and Computers in Simulation* Vol. 79, 2008, pp. 368–378.

[13] Mangel M., Aggregation and fishery dynamics: multiple time scales, times to extinction, and random environments, *Ecological Modelling*, Vol. 15, 1982, pp. 191–209.

[14] Martin P. G., The use of canonical S-system modelling for condensation of complex dynamic models, *Ecological Modelling*, Vol. 103, 1997, pp. 43–70.

[15] Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, 2007.

[16] Rathinam M., Petzold L. R., A new look at proper orthogonal decomposition, *SIAM J. Numer. Anal.*, Vol. 41(5), 2003, pp. 1893–1925.

[17] Rinaldi S., Scheffer M., Geometric analysis of ecological models with slow and fast processes, *Ecosystems*, Vol. 3, 2000, pp. 507–521.

[18] Rowley C. W., Colonius T., Murray R. M., Model reduction for compressible flows using POD and Galerkin projection, *Physica D: Nonlinear Phenomena*, Vol. 189(1–2), 2004, pp. 115–129.

[19] Todorovski L., Džeroski S., Integrating knowledge-driven and data-driven approaches to modelling, *Ecological Modelling*, Vol. 194, 2006, pp. 3–13.

A Further reading

The area of model simplification as mentioned earlier is extremely vast. This guide is buy no means exhaustive and some suggested reading is contained below. The above methods were highlighted due to their applicability to any dynamical systems model and their simplicity.

1. Antoulas A. C., A new result on passivity preserving model reduction, *Systems and Control Letters*, Vol. 54, 2005, pp. 361–374.

2. Bai Z., Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, *Applied Numerical Mathematics*, Vol. 43, 2002, pp. 9–44.

3. Bond B., Daniel L., Parameterized Model Order Reduction of Nonlinear Dynamical Systems, *Proceedings of the IEEE Conference on Computer-Aided Design*, November, 2005.

4. Butcher E. A., Lu R., Order reduction of structural dynamic systems with static piecewise linear nonlinearities, *Nonlinear Dynamics*, Vol. 49, 2007, pp. 375–399.

5. Feng L., Review of model order reduction methods for numerical simulation of nonlinear circuits, *Applied Mathematics and Computation*, Vol. 167, 2005, pp. 576–591.

6. Gugercin S., Antoulas A. C., A survey of model reduction by balanced truncation and some new results, *Int. Journal of Control*, Vol. 77, 2004, pp. 748–766.

7. Gugercin S., Antoulas A. C., Model reduction of large-scale systems by least squares, *Linear Algebra and its Applications*, Vol. 415, 2006, pp. 290–321.

8. Lall S., Krysl P., Marsden J. E., Structure-preserving model reduction for mechanical systems, *Physica D*, Vol. 184, 2003, pp. 304–318.

9. Lall S., Marsden J. E., Glavaški S., A subspace approach to balanced truncation for model reduction of nonlinear control systems, *Int. Journal of Robust and Nonlinear Control*, Vol. 12, 2002, pp. 519–535.

10. Leibfritz F., Volkwein S., Reduced order output feedback control design for PDE systems using proper orthogonal decomposition and nonlinear semidefinite programming, *Linear Algebra and its Applications*, Vol. 415, 2006, pp. 542–575.

11. Pinnegar J. K., Blanchard J. L., Mackinson S., Scott R. D., Duplisea D. E., Aggregation and removal of weak-links in food-web models: system stability and recovery from disturbance, *Ecological Modelling*, Vol. 184, 2005, pp. 229–248.

12. Rudnyi E. B., Korvink J. G., Model order reduction for large scale finite element engineering models, *Proceedings of the European Conference on Computational Fluid Dynamics*, Delft, The Netherlands, 2006.

13. Subbey S., Howell D., Bogstad B., Asnes M., Reducing fisheries model complexity using a finite Fourier series reparameterization, *Fisheries Research*, Vol. 84(3), 2007, pp. 390–394.

14. Sun C., Hahn J., Reduction of stable differential-algebraic equation systems via projections and system identification, *Journal of Process Control*, Vol. 15, 2005, pp. 639–650.

15. Zhang Y., Henson M. A., Kevrekidis Y. G., Nonlinear model reduction for dynamic analysis of cell population models, *Chemical Engineering Science*, Vol. 58, 2003, pp. 429–445.

Sensitivity analysis of model functions

J. C. Bennett* J. W. Hearne†

Abstract

Numerous papers deal with the sensitivity analysis of parameters and initial values of dynamic system models. Much less attention has been given to the sensitivity analysis of the functions used in these models. In this paper we investigate an automated process for function sensitivity analysis. We test four different function perturbations comprising constant, triangular, sinusoidal and quadratic disturbances. The methods are used to rank the sensitivity of the functions. The most sensitive functions, as in parameter sensitivity, require more care to reduce uncertainty in the model results. The least sensitive functions are candidates for model simplification. The ideas are illustrated using a model from the literature.

1 Introduction

Model simplification has received a great deal of attention recently, particularly in ecological modelling. A brief overview of the methods commonly employed may be found in Cariboni et al [1]. One of the fundamental questions ecological modellers must ask when constructing models is: What level of complexity is necessary to display the phenomena we wish to model [10]? This paper is focused on a key element that contributes towards an understanding of model characteristics, namely function sensitivity. By how much can a function be changed without changing the model conclusions? We focus, for illustration, on a model posed by Raghu et al. [13] on predicting the risk involved to a non-target species when biological control is used to manage an invasive plant species.

*School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia. mailto:james.cameron.bennett@rmit.edu.au

†School of Mathematical and Geospatial Sciences, Royal Melbourne Institute of Technology, Melbourne, Australia. mailto:john.hearne@rmit.edu.au

Raghu et al. [13] model the risk in the introduction of biological control to manage an invasive plant species. The chrysomelid beetle *Charidotis auroguttata* is introduced to manage the spread of *Macfadyena unguis-cati* (cat's claw creeper - CCC) in Australia and calculate the risk to the non-target species *Myoporum boninense australe* (MBA). The risk is calculated to be the initial leaf amount of the non-target species *Myoporum boninense australe* divided by the amount at the end of the simulation. The focus throughout this paper will be on function sensitivity as opposed to the model itself.

Sensitivity analysis on parameters and initial values for a model has been investigated widely [2, 5, 6, 15] (among many others). Less attention, however, has been paid to the sensitivity of the functions in a model. Where this has been addressed ad hoc techniques were used rather than an automated process [7, 8, 16]. In this paper, we explore and further develop a method suggested by Hearne et al. [9] for function sensitivity and apply it to a real world model. We perturb the model functions with constant, triangular, sinusoidal and quadratic function disturbances. We determine to what extent the functions can be perturbed before it leads to a change in the conclusions drawn from the model results. If the model conclusions are unaffected by a relatively large change in a particular function then this function could be regarded as insensitive. Similar to parameter sensitivity analysis this process provides a ranking of functions from the most sensitive to the least sensitive.

Modelling a real world process can often involve a highly complex model containing many state variables and parameters, therefore requiring the need for more data than for a simpler model [14]. An increase in complexity often results in the model being restricted to a specific range of scenarios, however, a simpler model may only provide a *trend* as opposed to a true solution for a range of scenarios. The choice of a simpler model has the trade-off of an increase in error but results in a decrease in sensitivity. Conversely, a more complex model produces a smaller error but increases the sensitivity [12, 14]. If a complex model can be reduced where the changes in decision variables are smaller than a prescribed tolerance then this reduces the chance of over-parameterisation [3].

Cox et al. [3] propose a systematic way of simplifying a complex model. By simulation they produce reduced model formulations by replacing state variables with constants. An exhaustive search is carried out to identify possible variables for replacement. The variable is assumed a candidate for replacement if it does not more than double the residual sum of squares when replaced by its mean value. This method however does not investigate the effect of function variations.

The technique employed in this paper may be modified for any model to gain an understanding of which model equations are sensitive to changes.

2 The problem

Consider a dynamic system model of the following form:

$$x_i(t+\Delta t) = x_i(t) + f_i(\boldsymbol{x}, t, \boldsymbol{\alpha})\Delta t, \quad i = 1, \ldots, n, \tag{1}$$

where t is discrete time, $\boldsymbol{x}$ is the state vector and $\boldsymbol{\alpha}$ is a vector of parameters. Suppose that the model will be used to make a decision regarding some action to be taken on the system. Suppose that this decision depends on whether

$$q(\boldsymbol{x}(t_{\max})) \geqslant \xi, \tag{2}$$

where q is some output and ξ is a threshold value. For example, q may be the amount of crop present at the end of a growing season and ξ may be the amount required for a good harvest. In what follows we will investigate perturbations in the model functions and determine if the model decision is changed to ascertain if the function is sensitive or not. We will be considering a system of discrete equations similar to those posed by Raghu et al. [13]. To determine the effect of the change in the shape of the model function we introduce a series of simple function disturbances. The model equations will be given briefly in the next Section.

As mentioned earlier, we will be considering constant, triangular, sinusoidal and quadratic disturbances to determine an easy effective way to carry out function sensitivity. The disturbances are given by

$$H(\boldsymbol{x}, \boldsymbol{\rho}) = \prod_{i=1}^{n}(1 + h_i(x_i, \boldsymbol{\rho})), \tag{3}$$

where $\boldsymbol{\rho}$ is the vector of minimisation parameters. The shape functions h_i are

$$\begin{aligned}
&\text{Constant:} && h_i(c_i) = c_i, \\
&\text{Triangular:} && h_i(x_i, m_i, p_i) = m_i \tfrac{(x_i-\nu)}{(p_i-\nu)}, \quad \nu = \begin{cases} x_{\min} & \text{if } x \leqslant p_i \\ x_{\max} & \text{if } x > p_i \end{cases} \\
&\text{Sinusoidal :} && h_i(x_i, a_i, \omega_i, \phi_i) = a_i \sin(\omega_i x_i + \phi_i), \\
&\text{Quadratic :} && h_i(x_i, \beta, \gamma, \delta) = \beta_i x_i^2 + \gamma_i x_i + \delta_i,
\end{aligned} \tag{4}$$

where a_i, ω_i, ϕ_i, c_i, m_i, p_i, ν, β_i, γ_i and δ_i are the minimisation parameters. If the model function is independent of a variable then the corresponding disturbance in that variable is zero, i.e., $h_i(x_i, \boldsymbol{\rho}) \equiv 0$ if $f_i \not\equiv f_i(x_i)$.

We consider the sensitivity of each function by multiplying $f_i(\boldsymbol{x}, t, \boldsymbol{\alpha})$ by the disturbance, i.e.,

$$g_i(\boldsymbol{x}, t, \boldsymbol{\alpha}, \boldsymbol{\rho}) = H(\boldsymbol{x}, \boldsymbol{\rho}) f_i(\boldsymbol{x}, t, \boldsymbol{\alpha}) \tag{5}$$

and then we solve the following minimisation problem:

$$\min_{\boldsymbol{\rho}} \left\| \frac{g_i - f_i}{f_i} \right\|_\infty, \tag{6}$$

subject to some model decision failing, where $\| \ \|_\infty$ is the infinity or the maximum norm. Note if $H \equiv 1$ which occurs when $h_i \equiv 0, \ \forall i$ then we retain the original model.

Hence the problem is: Find the set $(\boldsymbol{\rho}^*)$ that minimises the objective function (6) such that some decision (2) is changed. In other words, what is the smallest function disturbance that causes the model to change some decision criterion. If the objective function (6) is small then the function is sensitive to change, i.e. a small change in the function has led to a change in decision. If the value is large then the function is not sensitive to change and has the potential to be replaced by a simpler function.

The desired outcome is an easy to implement process to determine model function sensitivity to shape change. The test functions (4) are chosen for their simplicity. The constant function disturbance is the simplest but does not account for function shape change, only magnitude and is the more usual approach.

3 The model

As mentioned earlier the model we are using to test function sensitivity on is that of Raghu et al. [13] who developed their model based on results obtained from prior quarantine studies [4]. The model equations are given below with the parameter values used in the model contained in Table 1. A description of the parameters is contained in Raghu et al. [13]. The change in the population of eggs is given by

$$\mathrm{E_C}(t+1) - \mathrm{E_C}(t) = k1_\mathrm{C} \times \mathrm{A_C}(t) - \mathrm{Ehatch_C} - \mathrm{Edeath_C}, \tag{7}$$

$$\mathrm{E_M}(t+1) - \mathrm{E_M}(t) = k1_\mathrm{M} \times \mathrm{ABr_M}(t) - \mathrm{Ehatch_M} - \mathrm{Edeath_M}, \tag{8}$$

where

$$\mathrm{Ehatch}_i = \mathrm{E}_i(t)\frac{k2_i}{k3_i}, \quad \mathrm{Edeath}_i = \mathrm{E}_i(t)\frac{1-k2_i}{k3_i}, \quad i = \mathrm{C, M}, \tag{9}$$

Table 1: Table of model parameter values.

Parameter	i =CCC		i =MBA
$k1_i$	0.2		0.081
$k2_i$	0.95		0.9501
$k3_i$	9		90
$k4_i$	0.998		0.964
$k5_i$	32		59
$k6$		Set by user	
$k7_L$		Table 2	
$k7_A$		Table 3	
$k8_i$	80		5
$k9_i$	0.04		0.023
$k10_i$	100		10
$k11$			0.001
$k12_i$	0.015		0.003
$k13_i$	0.003		0.00015
$k14_i$	0.003		0.0003
$k15_i$	0.011		0.031
$k16$			50

and $\mathrm{A_C}$ and $\mathrm{ABr_M}$ are the populations of breeding adults for CCC and MBA, respectively. Note that in this model the time step is equal to 1 day, i.e. $\Delta t = 1$. Eggs are laid only between day 1 and day 220 with the further restriction that for adults maturing or moving from CCC to MBA there is a delay of 50 days until they can lay eggs. This delay is described later.

The change in the population of larvae is given by

$$\mathrm{L_C}(t{+}1){-}\mathrm{L_C}(t) = \mathrm{Ehatch_C}{-}\mathrm{Ldeath_C}{-}\mathrm{Lmatur_C}{-}\mathrm{LCtoM}{-}\mathrm{Lxsdeath_C}, \quad (10)$$

and

$$\mathrm{L_M}(t+1) - \mathrm{L_M}(t) = \mathrm{Ehatch_M} - \mathrm{Ldeath_M} - \mathrm{Lmatur_M} + \mathrm{LCtoM} - \mathrm{Lxsdeath_M}, \quad (11)$$

where

$$\mathrm{Lmatur}_i = \mathrm{L}_i(t)\frac{k4_i}{k5_i}, \quad \mathrm{Ldeath}_i = \mathrm{L}_i(t)\frac{1-k4_i}{k5_i}, \quad i = \mathrm{C, M}. \quad (12)$$

The model allows larval movement from CCC to MBA, LCtoM, which is dependent on the dynamic maximum, given by

if $\mathrm{L_C}(t) < \mathrm{Ldynmax_C}$ **then**

LCtoM=0
else
LCtoM=min(Lspillover, $\mathrm{L_C}(t)$)
end if

where Lspillover is given by

$$\mathrm{Lspillover} = \max(0, k6 \times k7_L \times (\mathrm{L_C}(t) + \mathrm{L_M}(t)) - \mathrm{L_M}(t), \tag{13}$$

and the dynamical maximum or larvae on CCC, $\mathrm{Ldynmax_C}$, is given by

$$\mathrm{Ldynmax_C} = \mathrm{Leaf_C}(t) \times k8_\mathrm{C}, \tag{14}$$

where $\mathrm{Leaf_C}$ is the amount of CCC leaf present. The value of the parameter $k7_\mathrm{L}$ is dependent on the dynamic maximum proportion of larvae on MBA. The dynamic maximum proportion of larvae on MBA is given by

if $\mathrm{Ldynmax_C} + \mathrm{Ldynmax_M} = 0$ **then**
$\mathrm{LdynpropM} = 0.00001$
else
$\mathrm{LdynpropM} = \mathrm{Ldynmax_M}/(\mathrm{Ldynmax_C} + \mathrm{Ldynmax_M})$
end if

where the dynamical maximum of larvae on MBA, $\mathrm{Ldynmax_M}$, is given by

$$\mathrm{Ldynmax_M} = \mathrm{Leaf_M}(t) \times k8_\mathrm{M}, \tag{15}$$

where $\mathrm{Leaf_M}$ is the amount of MBA leaf present. Therefore $k7_\mathrm{L}$ is determined by linear interpolation of the values contained in Table 2. When the dynamic maximum of larvae on CCC is exceeded the excess die, where the excess is defined to be larvae that is not transferring from CCC to MBA, maturing or dying naturally, i.e.,

if $\mathrm{L_C}(t) < \mathrm{Ldynmax_C}$ **then**
$\mathrm{LCtoM} = 0, \mathrm{Lxsdeath_C} = 0$
else
$\mathrm{LCtoM} = \min(\mathrm{L_C}(t), \mathrm{Lspillover}), \mathrm{Lxsdeath_C} = \mathrm{L_C}(t) - \mathrm{Ldynmax_C} - \mathrm{Lmatur_C} - \mathrm{LCtoM} - \mathrm{Ldeath_C}$
end if

The excess death for the larvae on MBA is defined as

if $\mathrm{L_M}(t) > \mathrm{Ldynmax_M}$ **then**
$\mathrm{Lxsdeath_M} = \mathrm{L_M}(t) - \mathrm{Ldynmax_M}$
else
$\mathrm{Lxsdeath_M} = 0$

Table 2: Table of values for parameter $k7_{\mathrm{L}}$

LdynpropM	$k7_{\mathrm{L}}$
0	0
0.1	0
0.2	0.01
0.3	0.025
0.4	0.05
0.5	0.1
0.6	0.145
0.7	0.215
0.8	0.365
0.9	0.605
1.0	1.0

end if

The change in the population of adults on CCC is given by

$$\mathrm{A_C}(t+1) - \mathrm{A_C}(t) = \mathrm{Lmatur_C} - \mathrm{Adeath_C} - \mathrm{ACtoM} - \mathrm{Axsdeath_C}, \quad (16)$$

where

$$\mathrm{Adeath_C} = \mathrm{A_C}(t) \times k15_{\mathrm{C}}. \quad (17)$$

Adult spillover is similar to larval spillover. If the population of adults on CCC exceeds the dynamic maximum then there is adult excess mortality and transfer onto MBA, i.e.,

if $\mathrm{A_C}(t) < \mathrm{Adynmax_C}$ **then**

$\mathrm{ACtoM} = 0, \mathrm{Axsdeath_C} = 0$

else

$\mathrm{ACtoM} = \min(\mathrm{A_C}(t), \mathrm{Aspillover}), \mathrm{Axsdeath_C} = \mathrm{A_C}(t) - \mathrm{Adynmax_C} - \mathrm{ACtoM} - \mathrm{Adeath_C}$

end if

where

$$\mathrm{Aspillover} = \max(0, k7_{\mathrm{A}} \times (\mathrm{A_C}(t) + \mathrm{A_M}(t)) - \mathrm{A_M}(t), \quad (18)$$

and the dynamical maximum of adults on CCC, $\mathrm{Adynmax_C}$, is given by

$$\mathrm{Adynmax_C} = \mathrm{Leaf_C}(t) \times k10_{\mathrm{C}}, \quad (19)$$

The value of the parameter $k7_{\mathrm{A}}$ is dependent on the dynamic maximum proportion of adults on MBA. The dynamic maximum proportion of adults on MBA is given by

Table 3: Table of values for parameter $k7_{\mathrm{A}}$

$AdynpropM$	$k7_{\mathrm{A}}$
0	0
0.1	0
0.2	0.01
0.3	0.015
0.4	0.035
0.5	0.05
0.6	0.1
0.7	0.135
0.8	0.21
0.9	0.35
1.0	1.0

if $\mathrm{Adynmax_C} + \mathrm{Adynmax_M} = 0$ **then**
  $\mathrm{AdynpropM} = 0.00001$
else
  $\mathrm{AdynpropM} = \mathrm{Adynmax_M}/(\mathrm{Adynmax_C} + \mathrm{Adynmax_M})$
end if

where the dynamical maximum of adults on MBA, $\mathrm{Adynmax_M}$, is given by

$$\mathrm{Adynmax_M} = \mathrm{Leaf_M}(t) \times k10_{\mathrm{M}}. \tag{20}$$

Therefore $k7_{\mathrm{A}}$ is determined by linear interpolation of the values contained in Table 3. The adults on MBA are made up of an initial pre-reproductive stage and two main life stages (non-breeding and breeding). The total population of adults on MBA is given by

$$\mathrm{A_M}(t+1) = \mathrm{Apre_M}(t+1) + \mathrm{Anbr_M}(t+1) + \mathrm{Abr_M}(t+1), \tag{21}$$

where $\mathrm{Apre_M}(t)$ is the pre-reproductive stage, $\mathrm{Anbr_M}(t)$ is the non-breeding stage (due to the delay in breeding) and $\mathrm{Abr_M}(t)$ is the breeding stage. The change in the pre-reproductive stage $\mathrm{Apre_M}(t)$ is given by

$$\mathrm{Apre_M}(t+1) - \mathrm{Apre_M}(t) = \mathrm{Lmatur_M} + \mathrm{ACtoM} - \mathrm{Adevelop_M}, \tag{22}$$

where

$$\mathrm{Adevelop_M} = \mathrm{Apre_M}(t). \tag{23}$$

Therefore the pre-reproductive stage is an instantaneous transfer to the non-breeding stage. As mentioned earlier there is a delay of 50 days before the adults maturing or transferring onto MBA can lay eggs. The change in the population of non-breeding adults is given by

$$\mathrm{Anbr}_{\mathrm{M}}(t+1)-\mathrm{Anbr}_{\mathrm{M}}(t) = \mathrm{Adevelop}_{\mathrm{M}}-\mathrm{Anbrdeath}_{\mathrm{M}}-\mathrm{Apre}_{\mathrm{M}}(t-k16)(1-k11)^{k16}, \tag{24}$$

where

$$\mathrm{Anbrdeath}_{\mathrm{M}} = \mathrm{Anbr}_{\mathrm{M}}(t) \times k15_{\mathrm{M}}. \tag{25}$$

Upon maturing the adults enter the breeding adult stage. The change in the population of breeding adults is given by

$$\mathrm{Abr}_{\mathrm{M}}(t+1)-\mathrm{Abr}_{\mathrm{M}}(t) = \mathrm{Apre}_{\mathrm{M}}(t-k16)(1-k11)^{k16}-\mathrm{Adeath}_{\mathrm{M}}-\mathrm{Axsdeath}_{\mathrm{M}}, \tag{26}$$

where

$$\mathrm{Adeath}_{\mathrm{M}} = \mathrm{Abr}_{\mathrm{M}}(t) \times k15_{\mathrm{M}}. \tag{27}$$

The adult excess mortality is given by

if $\mathrm{A}_{\mathrm{M}}(t) - \mathrm{Adeath}_{\mathrm{M}} > \mathrm{Adynmax}_{\mathrm{M}}$ **then**
  $\mathrm{Axsdeath}_{\mathrm{M}} = \mathrm{A}_{\mathrm{M}}(t) - \mathrm{Adynmax}_{\mathrm{M}} - \mathrm{Adeath}_{\mathrm{M}}$
else
  $\mathrm{Axsdeath}_{\mathrm{M}} = 0$
end if

where

$$\mathrm{Adynmax}_{\mathrm{M}} = \mathrm{Leaf}_{\mathrm{M}}(t) \times k10_{\mathrm{M}}. \tag{28}$$

The change in leaf amount present is given by

$$\mathrm{Leaf}_{\mathrm{C}}(t+1) - \mathrm{Leaf}_{\mathrm{C}}(t) = \mathrm{Leaf}_{\mathrm{C}}(t) \times k12_{\mathrm{C}} - \mathrm{Leafrem}_{\mathrm{C}}, \tag{29}$$

and

$$\mathrm{Leaf}_{\mathrm{M}}(t+1) - \mathrm{Leaf}_{\mathrm{M}}(t) = \mathrm{Leaf}_{\mathrm{M}}(t) \times k12_{\mathrm{M}} - \mathrm{Leafrem}_{\mathrm{M}}, \tag{30}$$

where $\mathrm{Leafrem}_i, \quad i = \mathrm{C}, \mathrm{M}$ is the amount of leaf removed from larval and adult damage, with the restriction that a maximum of 90% of leaf can be removed per time step for CCC and 95% for MBA. This is given by

$$\mathrm{Leafrem}_{\mathrm{C}} = \min\left(0.9 \times \mathrm{Leaf}_{\mathrm{C}}(t), \mathrm{Ldamage}_{\mathrm{C}} + \mathrm{Adamage}_{\mathrm{C}}\right) \tag{31}$$

and

$$\mathrm{Leafrem}_{\mathrm{M}} = \min\left(0.95 \times \mathrm{Leaf}_{\mathrm{M}}(t), \mathrm{Ldamage}_{\mathrm{M}} + \mathrm{Adamage}_{\mathrm{M}}\right) \tag{32}$$

where

$$\mathrm{Ldamage}_i = \mathrm{L}_i(t) \times k13_i, \quad i = \mathrm{C}, \mathrm{M}, \tag{33}$$

and

$$\text{Adamage}_i = \text{A}_i(t) \times k14_i, \quad i = \text{C, M}. \tag{34}$$

The model has the further restriction that the populations are assumed non-negative. To test our method on the model the parameter values displayed in Table 1 were used where there is 75% proximity ($k6$), 25% target patch and the initial release density of 100 *Charidotis auroguttata* per plant.

The decision criterion is related to the risk value to the non-target species, MBA. The risk is calculated as the ratio of the initial leaf abundance to the final leaf abundance of the non-target species, MBA, for the simulation period, i.e.,

$$\text{Risk} = \frac{\text{Leaf}_\text{M}(250)}{\text{Leaf}_\text{M}(0)}. \tag{35}$$

If this ratio is greater than or equal to 1 then the foliage is unable to replenish and the release of *Charidotis auroguttata* is deemed an unacceptable risk [13]. As a precautionary measure suppose that we require the risk to be less than 0.8. The nominal model solution produced a risk value of approximately 0.72 which is 10% below the precautionary value. Hence based on this result the decision is positive to introduce *Charidotis auroguttata*.

We perform the functional sensitivity on the state Equations (7, 8, 10, 11, 16, 22, 24, 26, 29, 30). Hence we test the function sensitivity of the 10 functions f_i which are the right hand sides of these equations and are denoted f_{EC}, f_{EM}, f_{LC}, f_{LM}, f_{AC}, f_{ApreM}, f_{AnbrM}, f_{AbrM}, f_{LeafC} and f_{LeafM}, respectively.

The functions contained in the model are dependant on up to six state variables which increases the space of the minimisation. Table 4 shows which state variables are present for each function. Note that some of the variables are present only for some of the simulation period depending on the dynamics of the model. When they are not present, the corresponding disturbance h_i is zero.

We consider the sensitivity of each model function by multiplying $f_i(\boldsymbol{x}, t, \boldsymbol{\alpha})$ by the disturbance, i.e.,

$$g_i(\boldsymbol{x}, t, \boldsymbol{\alpha}, \boldsymbol{\rho}) = H(\boldsymbol{x}, \boldsymbol{\rho}) f_i(\boldsymbol{x}, t, \boldsymbol{\alpha}) \tag{36}$$

and then we solve the following constrained minimisation problem:

$$\min_{\boldsymbol{\rho}} \left\| \frac{g_i - f_i}{f_i} \right\|_\infty, \tag{37}$$

s.t.

$$\text{Risk} > 0.8. \tag{38}$$

Hence the problem is: Find the minimum distortion of the original functions required for the risk to violate the precautionary value (0.8). If this distortion (that is the objective function (37)) is small then the function is sensitive to change. If the distortion is large then the function is not sensitive to change and has the potential to be replaced by a simpler function.

The functions contained in this model are reasonably simple in a nonlinear sense but contain a significant degree of coupling. The model was chosen to portray the techniques on a simple realistic model that is readily available in the literature. The method may be expanded to most ecological models.

4 Results and Discussion

The results of the minimisation are displayed in Table 5 which shows the effect of each function shape. In each case f_{AC} is very sensitive to change. Therefore a small change in this functions results in an increase in risk to the precautionary value 0.8. On the other hand, from Table 5, the function which is least sensitive (apart from the constant function) to change is f_{EM}. This function needs to be multiplied by a large disturbance relative to f_{AC} before there is an increase in risk to the precautionary value. Therefore, we may conclude that the shape of the function in this case is not critical to achieving the objective of the model. The sinusoidal and quadratic disturbances are

Table 4: Dependent variables for each function. A "✓" means the function depends on the variable.

	$\mathrm{E_C}$	$\mathrm{E_M}$	$\mathrm{L_C}$	$\mathrm{L_M}$	$\mathrm{A_C}$	$\mathrm{Apre_M}$	$\mathrm{Anbr_M}$	$\mathrm{Abr_M}$	$\mathrm{Leaf_C}$	$\mathrm{Leaf_M}$
f_{EC}	✓				✓					
f_{EM}		✓						✓		
f_{LC}	✓		✓	✓					✓	
f_{LM}		✓	✓	✓					✓	✓
f_{AC}			✓		✓	✓	✓	✓	✓	
f_{ApreM}				✓	✓	✓	✓	✓	✓	
f_{AnbrM}						✓	✓			
f_{AbrM}						✓	✓	✓		✓
f_{LeafC}			✓		✓				✓	
f_{LeafM}				✓		✓	✓	✓		✓

Table 5: Table of objective value, relative insensitivity and sensitivity rank for each function disturbance.

Function	**Constant**			**Triangle**		
	Obj.	Rel. Insens.	Rank	Obj.	Rel. Insens.	Rank
fEC	1.6174	20.22	7	1.1079	8.13	7
fEM	1.0771	13.46	6	1.6515	12.13	10
fLC	0.8029	10.04	5	1.1915	8.75	9
fLM	35.46	443.25	9	1.0832	7.95	5
fAC	0.1231	1.54	2	0.1362	1.00	1
fApreM	254.67	3183.38	10	1.0793	7.92	6
fAnbrM	0.5647	7.06	4	0.5727	4.20	4
fAbrM	62.64	783.00	8	1.1436	8.40	8
fLeafC	0.08	1.00	1	0.2911	2.14	2
fLeafM	0.2452	3.07	3	0.3459	2.54	3

Function	**Sinusoidal**			**Quadratic**		
	Obj.	Rel. Insens.	Rank	Obj.	Rel. Insens.	Rank
fEC	0.4637	3.87	4	0.6647	4.30	4
fEM	1.5669	13.07	10	1.8205	11.78	10
fLC	0.5416	4.52	5	0.8360	5.41	5
fLM	1.00000006	8.34	7	1.0000006	6.47	7
fAC	0.1199	1.00	1	0.1546	1.00	1
fApreM	1.00001	8.34	8	1.1081	7.17	9
fAnbrM	0.4276	3.57	3	0.4457	2.88	3
fAbrM	1.00002	8.34	9	1.00009	6.47	8
fLeafC	1	8.34	6	1	6.47	6
fLeafM	0.3359	2.80	2	0.3374	2.18	2

restrictive compared to the triangular shape functions. There is an imposition that the sinusoidal function disturbance may only contain at most one period. This most likely accounts for the similarity in results between the sinusoidal and quadratic perturbations. The choice of a sinusoidal function disturbance has the drawback that it causes a non-local disturbance, i.e. the disturbance is not able to simulate a disturbance around a point exclusively (it has an effect elsewhere also). The sinusoidal and quadratic disturbances (compared to the triangle disturbance) have the drawback of an increase in the parameter search space. The triangular functions only require two parameters per state variable: the point and magnitude of the disturbance, whereas the sinusoidal and quadratic require three parameters. Therefore, the triangular shape disturbances are chosen for the ability to identify the point at which maximum disturbance occurs and their simplicity.

The results of the triangular minimisation from Table 5 indicate the function f_{EM} to be the least sensitive. This function is already very simple so we will consider the next least sensitive function f_{LC} which is contained in Equation (10). The only part of this equation that contains any nonlinearity is the LCtoM term. Since its relative insensitivity we expect to be able to simplify this equation. The nonlinearity in the LCtoM term arises trough the variation in parameter $k7_{\mathrm{A}}$ with the dynamic maximum proportion of larvae on MBA, LdynpropM. If the choice is made for a direct relationship between $k7_{\mathrm{A}}$ and LdynpropM, i.e. $k7_{\mathrm{A}} = \mathrm{LdynpropM}$ then the nonlinearity is removed from this function. This was carried out and yielded an increase in risk of approximately only 3.7%. Hence the model decision is unchanged and has been simplified.

The minimisation carried out for the constant function disturbance identifies same most sensitive functions. There is a difference for the rest of the order though. The values for the constant disturbances $h_i = c_i$ for the functions f_{LM}, f_{ApreM} and f_{AbrM} are significantly larger to cause an increase in risk to 0.8. The triangular, sinusoidal and quadratic disturbances did not need such a dramatic effect as the constant function for these equations. Hence this shows a relatively small disturbance in the right position causes the same increase in risk as a large constant disturbance which indicates the importance of the shape of the functions. The large constant disturbances cause unrealistic model behaviour whereas the function disturbances result in plausible solutions.

For argument sake, let us carry out a standard sensitivity analysis, similar to that carried out by Lawrie and Hearne [11], by perturbing each function by 10%, i.e., the constant function disturbance contained in Equation (4), $h_i(c_i) = \pm 0.1$. Hence there is no minimisation procedure performed. This is carried out for each function separately and the increase in risk is found. The results are shown in Table 6 and a representative plot for the sensitivity is shown in Figure 1.

From Table 6 we can see the first four most sensitive functions are the same as those for the function disturbances. The function f_{ApreM} is the least sensitive whereas for the function disturbances it is relatively sensitive to changes. The most insensitive function, f_{EM}, for the function disturbances is approximately 1.5 times more insensitive to changes in the shape of the function. Hence the standard sensitivity analysis identifies the most sensitive functions, however it doesn't identify the functions that are least sensitive to the function shape changes. Therefore possible candidates for function simplification have been identified. Figure 1 shows the difference in risk

Table 6: Table of change in risk and sensitivity rank for each constant function disturbance.

Function	Risk_Risk$_{\mathrm{Nom}}$	Rank
f_{EC}	0.0074	6
f_{EM}	0.0105	5
f_{LC}	0.0035	9
f_{LM}	0.0045	8
f_{AC}	0.0548	2
f_{ApreM}	0.0024	10
f_{AnbrM}	0.0126	4
f_{AbrM}	0.0071	7
f_{LeafC}	0.1078	1
f_{LeafM}	0.0295	3

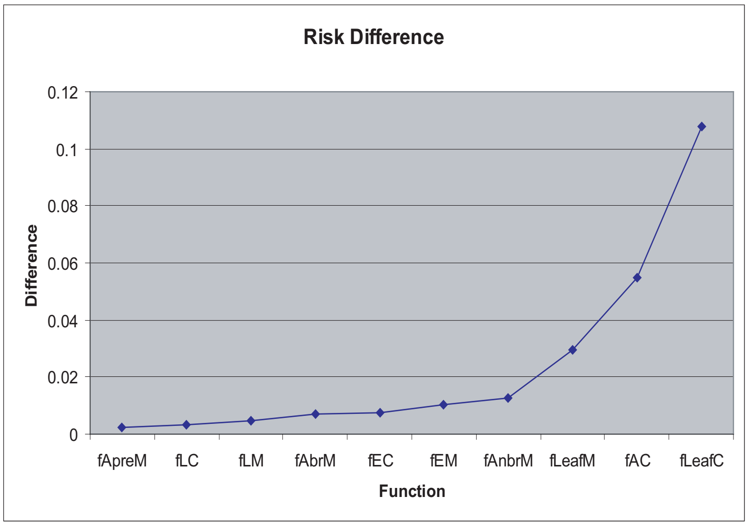

Figure 1: Constant function sensitivity. Larger difference in risk corresponds to a greater sensitivity.

(between the perturbed and nominal model). As a quick analysis tool to identify the sensitive functions the straight forward constant change in this model yielded the same results as the more complex function optimisations. Using Figure 1 we can see that the sensitive functions f_{LeafM}, f_{AC} and f_{LeafC} are significantly more sensitive than the remaining functions. Hence the process identifies quickly and easily which functions the modeller has to pay close attention to.

An example of the sort of model change that the small change in f_{AC} causes is displayed in Figure 2. This corresponds to the adult population of *Charidotis auroguttata* on the non-target species MBA for the nominal model compared to the population when the disturbance in the function f_{AC} is present. The perturbation to f_{AC} causes an increase in the adult population

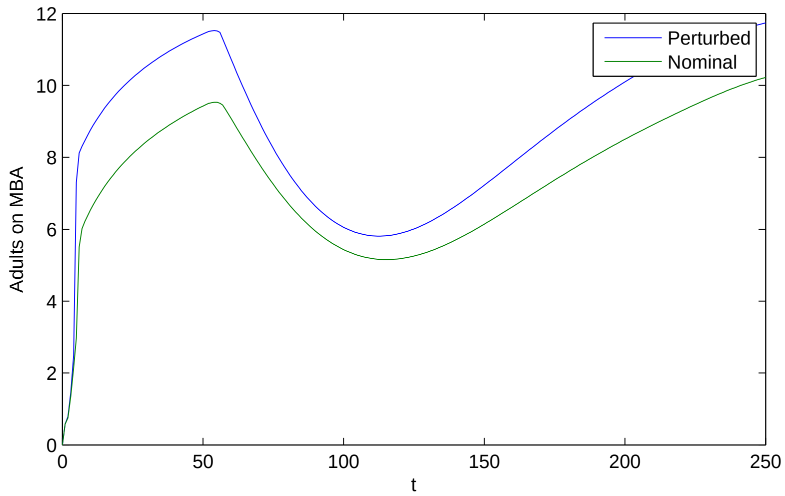

Figure 2: Disturbance to adult population on MBA with (small) f_{AC} perturbation.

on MBA which is due to an increased spillover. Hence this increases the risk to MBA since the increased population consume a greater amount of leaf matter.

For completeness, we introduce stochasticity into the model to assess whether a little uncertainty changes the sensitivity of the functions. The incubation and development time for eggs and larvae on each leaf were distributed normally by 10% about their means. Hence the parameters $k3_i$ and $k5_i$, $i = C, M$ become stochastic parameters. This may correspond to, for example, a variation between simulation periods in temperature. The standard constant perturbation was carried out and we obtained very similar results as in Table 6. Hence, in this model, the introduction of stochasticity yields the same sensitive functions as the deterministic model.

5 Conclusion

In this paper we addressed the sensitivity of model functions. Under the influence of function disturbances the model functions react differently. In

Section 4 we found that in particular the function relating to the change in the adult population on CCC, i.e., f_{AC} was the most sensitive with the least sensitive, f_{LM}, requiring a function disturbance approximately 12.13 times larger (for the triangular disturbance) to increase the risk to MBA to 0.8.

The results of the straight forward variation in the constant function and then measuring the change in the decision is a quick way to identify the sensitive functions since no minimisation is necessary. However this does not account for the sensitivity in the shape of the functions. The results in Section 4 show that when the minimisation is carried out for the constant function disturbance, some functions require an unrealistic change due to the magnitude needed to increase the risk to the precautionary value 0.8. The function minimisation requires a comparatively small change which highlights the effect of changing a functions shape. For the triangle function disturbance in particular, the point at which the change in the shape of the function, which causes the increase in risk, is identified.

The same technique may be applied to highly nonlinear functions occurring in an ecological model to identify insensitivity. The required specifications for this technique for any model is to identify the functions and their dependent variables and then carry out the minimisation procedure such that some essential criterion of the model fails using triangular shape disturbances. This identifies which functions (if any) are sensitive to change in their shape. The technique applied here may be applied to a wide range of problems.

6 Acknowledgements

This work was funded by the Australian Centre of Excellence for Risk Analysis (ACERA project 07/04).

References

[1] Cariboni J., Gatelli D., Liska R., Saltelli A., The role of sensitivity analysis in ecological modelling, *Ecological Modelling* Vol. 203, 2007, pp. 167–182.

[2] Cossarini G., Solidoro C., Global sensitivity analysis of a trophodynamic model of the Gulf of Trieste, *Ecological Modelling* Vol. 212, 2008, pp. 16–27.

[3] Cox G. M., Gibbons J. M., Wood A. T. A., Craigon J., Ramsden S. J., Crout N. M. J., Towards the systematic simplification of mechanistic models, *Ecological Modelling* Vol. 198, 2006, pp. 240–246.

[4] Dhileepan K., Treviño M., Donnelly G. P., Raghu S., Risk to non-target plants from *Charidotis auroguttata* (Chrysomelidae: Coleoptera), a potential biocontrol agent for cat's claw creeper *Macfadyena unguis-cati* (Bignoniaceae) in Australia, *Biological Control* Vol. 32, 2005, pp. 450–460.

[5] Dubus I. G., Brown C. D., Beulke, S., Sensitivity analyses for four pesticide leaching models, *Pest Management Science* Vol. 59, 2003, pp. 962–982.

[6] Fieberg J., Jenkins K. J., Assessing uncertainty in ecological systems using global sensitivity analyses: a case example of simulated wolf reintroduction effects on elk, *Ecological Modelling*, Vol. 187(2–3), 2005, pp. 259–280.

[7] Fulton E. A., Parslow J. S., Smith A. D. M., Johnson C. R., Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance, *Ecological Modelling*, Vol. 173, 2004, pp. 371–406.

[8] Fulton E. A., Smith A. D. M., Johnson C. R., Mortality and predation in ecosystem models: is it important how these are expressed?, *Ecological Modelling*, Vol. 169, 2003, pp. 157–178.

[9] Hearne J. W., Haque E., Stacey A. J., A Technique for the Sensitivity Analysis of Functions in Relation to Decision-making Objectives, MODSIM 07, December 2007.

[10] Kimmins J. P. (Hamish), Blanco J. A., Seely B., Welham C., Scoullar K., Complexity in modelling forest ecosystems: How much is enough?, *Forest Ecology and Management* Vol. 256, 2008, pp. 1646–1658.

[11] Lawrie J., Hearne J. W., Reducing model complexity via output sensitivity, *Ecological Modelling* Vol. 207, 2007, pp. 137–144.

[12] Lindenschmidt K. -E., The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling, *Ecological Modelling* Vol. 190, 2006, pp. 72–86.

[13] Raghu S., Dhileepan K., Scanlan J. C., Predicting risk and benefit a priori in biological control of invasive plant species: A systems modelling approach, *Ecological Modelling* Vol. 208, 2007, pp. 247–262.

[14] Snowling S. D., Kramer J. R., Evaluating modelling uncertainty for model selection, *Ecological Modelling* Vol. 138, 2001, pp. 17–30.

[15] Zaldivar J. M., Campolongo F., An application of sensitivity analysis to fish population dynamics. In: Saltelli A., Chan K., Scott E. M. (Eds.), Sensitivity Analysis, Wiley Series in Probability and Statistics, 2000.

[16] Zhou X., Peng C., Dang Q., Sun J., Wu H., Hua D., Simulating carbon exchange in Canadian Boreal forests I. Model structure, validation, and sensitivity analysis, *Ecological Modelling* Vol. 219, 2008, pp. 287–299.

An Automated Method for Extending Sensitivity Analysis to Response Functions

John W Hearne
School of Mathematical and Geospatial Sciences, RMIT University, GPO Box 2476v, Melbourne, Vic 3001, Australia.
Email: john.hearne@rmit.edu.au

ABSTRACT

Parameter sensitivity analysis is a relatively well-developed field compared with function sensitivity. How sensitive are model conclusions to the choice of functions used in the right hand side of difference or differential equation models? Most work in this area has been scenario-based where alternative functions are tested. This requires knowledge of plausible alternatives and is usually restricted to well-known functions. In this work, a method is proposed for examining the sensitivity of a model's results and conclusions to changes in the shape of the functions. This is done in an automated way without the need to specify alternative functional forms.

The question addressed is: what is the least amount of change required in a function to produce a significant change in some performance measure? If the changes in functions need to be large in some sense, to cause significant changes in a performance measure, then there is less need to focus attention on getting the model functions correct. Furthermore, such a function might be a candidate for replacement by a simpler form. This might offer possibilities for model simplification in large models. The proposed method is illustrated on a small ecosystem model. This analysis identified a model function that was very robust in terms of its effect on the critical value in the decision criterion. In fact it led to further experimentation that revealed that a state variable could be replaced by a constant thus simplifying the model. The insight gained from the proposed form of function sensitivity analysis is at least as important as the corresponding knowledge of the effect of uncertainty in parameter values.

Keywords

Parameter sensitivity analysis, function sensitivity analysis, ecosystem models, automated method, model simplification, model uncertainty.

INTRODUCTION

A parameter sensitivity analysis is considered to be so important in any modelling activity that it has become a routine exercise that is expected of any modelling project. A typical example of many can be found in a recent paper by Xenakis et al (2008). It is surprising then that such little attention has been given to the equally important task of performing a sensitivity analysis of the various functions used in a model. A more general need than just parameter sensitivity analysis was noted by Quade (1968) in the early days of modern computer models. He expressed the view that: "A good system study will include sensitivity tests on the assumptions in order to find out which ones really affect the outcome and to what extent. This enables the analyst to determine where further investigation of assumptions is needed".

There are two common approaches that have been used to investigate the sensitivity of the functions used in a model. Some consider alternate functions for some components of a model by manually changing a model equation (eg Fulton et al 2003). Another method is to use functions whose shape changes with the set of parameter values used. For example, Boukal et al (2007) use a general predator response function which by changing just one parameter can take many forms including both Holling type II and type III functional responses. Both these approaches are effectively limiting the analysis to a few functional forms. Wood and Thomas (1999) consider the effect of two different response functions in a model. Both these response functions fit empirical data equally well. They then construct a space of functions that is a weighted sum of the two functions with the weight treated as a variable parameter. The importance of considering different functional forms is well-illustrated by their analysis. The model they analyse is shown to be 'super-sensitive' to the function used.

Walker *et al* (2003) recently noted the increasing requirement to articulate uncertainty, when working at the interface of science and management, in model-based decision support. The main purpose of a sensitivity analysis of a model used in decision support should be to determine the extent to which the decisions or policies based on model results are robust with respect to the uncertainty in the model. Where a model result is used to make some decision what is relevant is whether changes to parameter values or model functions would lead to a different decision being made. This was exemplified by the practical approach adopted by Ford and Gardiner (1979). Their work was a noteworthy early exception to straight parameter sensitivity analysis (SA). They convened a workshop of public and private leaders where the group was presented with model forecasts and asked to decide on a policy. Changes were then made to the model and the group was presented with the new forecasts. Based on those forecasts, the group was again asked to vote on the policy. If the policy decision was unchanged the model could be regarded as insensitive to the changes. So it might well be that some change in a parameter value or function causes a very large change in a state variable but if this does not alter the decision of the policy-making body then in practical terms the model is insensitive.

Uncertainty includes uncertainty in parameter values, model inputs, and the functions in a model. Most analyses of models do include a sensitivity anlaysis (SA) of the parameters and a scenario analysis of the inputs. Alternative functions within a model, however, are only sometimes tested and in an *ad hoc* manner. Little has been done to perform a SA of *functions* in the automated way that is done with parameters. This paper takes some tentative steps towards addressing this problem.

METHOD

Consider the following system of difference equations:

$$x_i(t+\Delta t) = x_i(t) + f_i(\underline{x}, t, \underline{\alpha})\Delta t,\ i = 1, 2, \dots, n, \tag{1}$$

where $\underline{x}$ is the state vector and $\underline{\alpha}$ a vector of parameters. In ecological models the functions f_i comprise terms representing the rates of the system such as birth rate and death rate. Thus we can write:

$$f_i(\underline{x}, t, \underline{\alpha}) = \sum_j r_{ij}(\underline{x}, t, \underline{\alpha}) \tag{2}$$

A basic parameter sensitivity analysis involves changing the values of the parameters $\underline{\alpha}$ in the functions $r_i(\underline{x}, t, \underline{\alpha})$ by a small amount, one at a time, and observing the change it produces in the output. A possible pragmatic approach for function SA is to multiply each function or rate by a parameter with a nominal value of one. These parameters can then be perturbed as is done in a parameter SA. This should yield some indication as to which rates are the most sensitive. This method was tried with some success by Lawrie and Hearne (2007). But it does not yield any information about the sensitivity of the model output to changes in the shape of the functions. In some cases changing the parameters in the functions $r_i(\underline{x}, t, \underline{\alpha})$ will change the shape of the functions to some extent but it does so in very restrictive ways. Clearly other changes to the shape of the functions are possible. In ecological models there is frequently uncertainty about the appropriateness of the functional form chosen for the model. The choice of functional form will affect the model solutions and hence possibly the conclusions drawn from these results. It is therefore important to obtain a better idea of how sensitive the models results are to changes in the functions $r_i(\underline{x}, t, \underline{\alpha})$.

The simplest approach towards this end, going beyond the method mentioned above, is to multiply, one at a time, each rate r_{ij} by the following function which comprises a product of triangular-shaped functions:

$$H_{ij}(\underline{x}, \underline{p}, \underline{m}) = \prod_{i=1}^{n} h_i(x_i, p_{ij}, m_{ij}), \tag{3}$$

where

$$h_i(x_i, p_i, m_i) = \begin{cases} 1 + m_{ij}(x_i - c)/(p_{ij} - c) & \text{if } a_i \le x_i \le b_i \\ 1 & \text{otherwise} \end{cases}$$
$$\text{and } c = \begin{bmatrix} a_i \text{ if } x_i \le p_{ij}, \\ b_i \text{ if } x_i > p_{ij}. \end{bmatrix} \tag{4}$$

Note that m_{ij} can be negative or positive. Reasonable choices of a_i and b_i are the respective minimum and maximum values of the corresponding state variable x_i over the solution interval $[a_i, b_i]$.

The sensitivity analysis proceeds by testing the effect of changes in the function for each rate on some decision criteria. This is best described with an illustrative example.

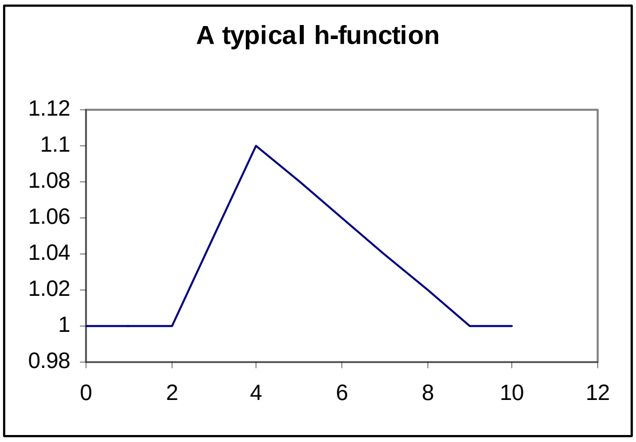

Figure 1: An example of an h-function. Here a=2, b=9, m=0.1 and p=4.

ILLUSTRATIVE EXAMPLE

An agricultural product X will be ready for harvesting in T (=12) months time. A pest species Y consumes X at a certain rate depending on the density of X. The damage caused by Y is unacceptable and two means of controlling the pest have been proposed: (1) biological control through the introduction of a parasitoid Z and (2) chemical control. The first method is much cheaper and also more desirable from environmental considerations but there is more confidence in the efficacy of chemical control. To facilitate making a decision, a model of the system with Z has been formulated. The aim of the model is to answer the following question:

Will the introduction of the parasitoid population Z ensure that the pest Y is sufficiently controlled to ensure that the biomass of X achieves a minimum level at harvest time T? In particular will the 10th percentile of X be above a threshold value V (=60)?

Let x_1, x_2, and x_3 denote the population levels of X, Y, and Z, respectively. The model is given by the system of equations (1) (see also 2) with the following RHS functions:

$$\begin{aligned}
f_1 &= r_{11} + r_{12} = g_1 x_1 (1 - x_1 / k) - \frac{\alpha_1 x_2 x_1}{(x_1 + \alpha_2)}, \\
f_2 &= r_{21} + r_{22} = \frac{g_2 x_2 x_1}{(x_1 + \alpha_2)} - \frac{\beta_1 x_3 x_2}{(x_2 + \beta_2)}, \\
f_3 &= r_{31} + r_{32} = \frac{g_3 x_3 x_2}{(x_2 + \beta_2)} - \gamma_1 x_3 \exp(\gamma_2 x_3).
\end{aligned} \tag{5}$$

Initial and parameter values are

$$\begin{aligned}
&x_1(0) = 40,\ x_2(0) = 12,\ x_3(0) = 6 \\
&g_1 = 0.8,\ g_2 = 0.25,\ g_3 = 0.1 \\
&\alpha_1 = 0.5,\ \alpha_2 = 20,\ \beta_1 = 0.2,\ \beta_2 = 5 \\
&\gamma_1 = 0.001,\ \gamma_2 = \ln(2)/30 \\
&k = 100 + N(0,10)
\end{aligned} \tag{6}$$

where *N(0,10)* is a normally distributed random number with mean 0 and standard deviation 10.

That the introduction of Z is effective can be seen by comparing the two graphs with and without the parasitoid population Z in Figure 2. These are solutions of the deterministic model with *k* held constant at 100. Further analysis was undertaken by performing 1000 simulations of the stochastic model. These solutions indicated that at the final time T, X would have a mean of approximately 66 and a 10th percentile of nearly 62 (>V). This suggests that the decision can be made tentatively to go for option (1), biological control.

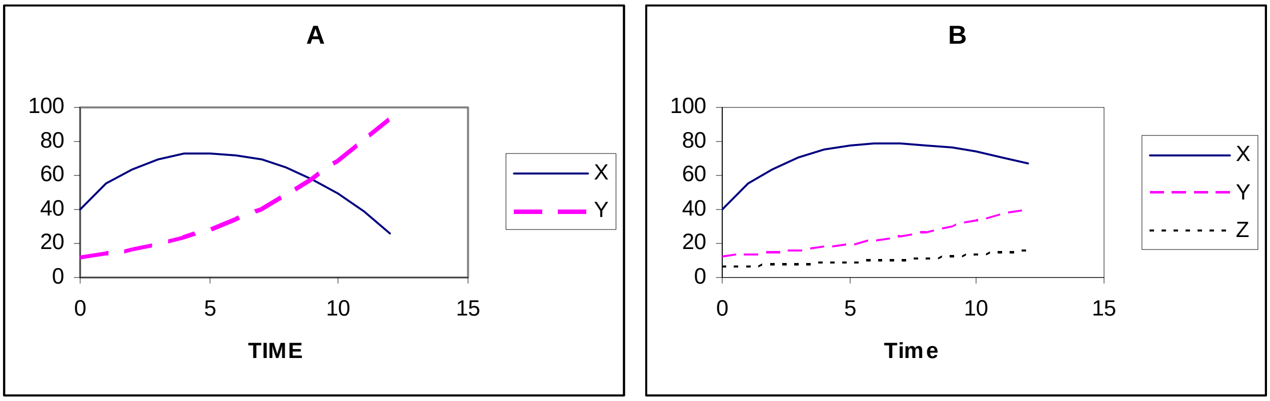

Figure 2: (A) Without biological control the crop X drops below the threshold at harvest. (B) The introduction of the parasitoid Z controls the pest Y sufficiently to enable the crop X to remain above the threshold value (60) at harvest time.

Normally at this point SA of parameters and initial values would be undertaken and possibly some experimenting with alternative model formulations. As the purpose of this project is to go beyond that, we assume that all parameter and initial values are perfectly known. The question then remains whether uncertainty relating to the functions $r_{ij}(\underline{x},t,\underline{\alpha})$ of the model will could change the results to the extent that a different conclusion would be reached from that reached with the nominal functions. In particular, we are interested in the following:

What is the smallest distortion to the functions that will lead to the violation of the decision criterion? The decision criterion being that the 10th percentile of population X lies above the threshold V at the final time and hence that the first option for control will be the preferred one?

If the criterion is violated with only small changes to the functions then one might conclude that the decision is sensitive to the choice of model functions. On the other hand if the function requires large changes before violating the decision criterion then one might conclude that to some extent the particular choice of functions used in the model is not critical.
We now formulate the mathematical problem to answer this question.

METHOD

Problem P1

Consider, one at a time, a change to each function $r_{ij}(\underline{x},t,\underline{\alpha})$, where the changed function is given by $H(\underline{x},\underline{p},\underline{m})r_{ij}(\underline{x},t,\underline{\alpha})$ (see equation (3) for definition of $H(\underline{x},\underline{p},\underline{m})$). If the functions $r_{ij}(\underline{x},t,\underline{\alpha})$ are independent of x_i then $h_{ij}(x_i,p_{ij},m_{ij})$ is set to one which means that it has no effect on the system. For the ij^{th} rate this means:

Find $(\underline{p}^*,\underline{m}^*)$, the solution to the constrained minimization problem:

$$\min_{\underline{p},\underline{m}} \underline{m}^2$$

constrained by the condition that the 10th percentile of $x_1(T) \le V$.

This is the smallest change of $r_{ij}(\underline{x},t,\underline{\alpha})$ which no longer ensures with sufficient probability that the harvest has minimum biomass greater than 60 units.

Effectively P1 means that we find the position in state space where the model is most sensitive to changes in the functions $r_{ij}(\underline{x},t,\underline{\alpha})$. Moreover, we can determine if increasing the function or decreasing the function at that point produces the greatest change in the output measure – through the sign of $\underline{m}$. This means that regardless of position or direction (increase or

decrease) any smaller change in the function will ensure that the final level of population X is acceptable, and hence robust to the decision.
Problem P1 is solved for cases corresponding to each of the six rates in the model. Some care is required here. The rates r_{12} and r_{21} differ only by a constant. They both represent the rate of predation by Y on X and hence comprise the same response function. The function sensitivity method described above should therefore be applied simultaneously to both rates. In effect this yields information about the sensitivity of the results to the choice of response function. A similar argument holds for r_{22} and r_{31}. There are thus four functions to consider for the SA.

RESULTS

The smallest magnitudes of the h-functions required to violate the decision criterion is shown in Table 1. It is seen that relative to the other functions the predation response function in r_{12} and r_{21} is the most sensitive.

The last term in the difference equation r_{32} (see equation 4) can be distorted by large amounts without the population X dropping below the decision threshold.
The most sensitive point in the shape of each function is also given in Table 1. Note that a perturbation of the same magnitude at any other point than that given in the table will satisfy the decision criterion. For example, the response function in r_{12} and r_{21} is most sensitive when the peak occurs at (x_1 , x_2)=(67,16). For $m_{i,}$ the magnitudes of the peaks in the h-functions given in Table 1, the 10th percentile for the final biomass value for X is fractionally below the threshold value of 60. Performing the simulations again with the same disturbance magnitude but at the point (x_1 , x_2)=(50,25) yields a 10th percentile value of 62 – above the threshold.
As the final rate in the parasitoid population Z is insensitive to large changes it is useful to see whether this function can be replaced by a simpler function. It soon becomes apparent that this is the case and a little further thought and experimentation reveals that the parasitoid can control the pest sufficiently well even if the population Z is held constant at its initial value. The method has thus first identified a function that can be simplified to a constant. This has further led to a conclusion that the state variable itself can be replaced by a conservative constant.

Function	m_1	m_2	m_3	Magnitude $\lvert \underline{m} \rvert$	x_1	x_2	x_3
r_{11}	-0.4	---	---	0.4	72	---	---
r_{12} and r_{21}	0.06	0.08	---	0.1	67	16	---
r_{22} and r_{31}	---	-0.12	-0.19	0.22	---	26	7

Table 1: Results of the function SA. The function r_{32} is not shown as it was robust to all reasonable perturbations. Column 5 gives an indication of the magnitude of the changes made to the functions. The last three columns contain the point where the function is most sensitive.

Conclusion

While parameter sensitivity analysis is expected of all models an analysis of the functions used in a model is performed less frequently. Even where some investigation along these lines has been undertaken a very restrictive class of functions has been considered. Part of the reason for this is probably due to the lack of an easy-to-implement automated approach to such an analysis. A method has been proposed here which is easy to implement provided the modeller has access to optimisation software. Such software is readily available for most platforms including spreadsheets and can handle both deterministic and stochastic models. Although not all of the infinitely many alternative functional forms is covered by the proposed method it does considerably extend the function space analysed.
Similar to a standard parameter sensitivity analysis the proposed method identifies the most sensitive functions. Further attention can then be given to improving confidence in these sensitive functions. Functions that are relatively insensitive to some objective criterion become candidates for simplification.
The proposed method has been tested on a relatively small model but the method can be applied equally well to a large ecosystem model. As each *rate* function in a model is usually dependent on only a few state variables the optimisation procedure is not overly cumbersome.

Acknowledgments

Funding for this project from the Australian Centre of Excellence for Risk Analysis (ACERA project 07/04) is gratefully acknowledged.

References

Boukal, D.S., Sabelis, M.W., Berec,L., 2007. How predator functional responses and Allee effects in prey affect the paradox of enrichment and population collapses. Theoretical Population Biology 72:136-147.

Ford, A. and Gardiner, P. C. (1979), New measure of sensitivity for social system simulation-models, *IEEE Trans. on Systems Man and Cybernetics*, 9(3),105-114.

Forrester, J. W. (1969), Urban Dynamics, MIT Press.

Fulton, E A , Parslow, J S, Smith A D M, Johnson, C R 2003. Mortality and predation in ecosystem models: is it important how these are expressed? Ecol. Model. 169:157-178

Lawrie, J. and Hearne, J. W. (2007), Reducing model complexity via output sensitivity, *Ecological Modelling*, 207, 137-144.
Quade, E. S. (1968), By Way of Summary, in ed. Quade, E. S. and Boucher, W. I., Systems Analysis and Policy Planning: Applications in Defense, Elsevier, New York.

Walker, W. E., Harremoes, P., Rotmans, J., Van Der Sluijs, J. P., Van Asselt, M. B. A., Janssen, P., and Krayer Von Krauss, M. P. (2003), Defining uncertainty- a conceptual basis for uncertainty management in model-based decision support, *Integrated Assessment*, 4(1), 5-17.

Wood, S.N., Thomas, M.B.,1999. Super-sensitivity to structure in biological models. Proc. R. Roc. Lond. B 266:565-570.

Xenakis, G.,Ray, D., Mencuccini, M., 2008. Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. Ecol.Model. 219:1-16.