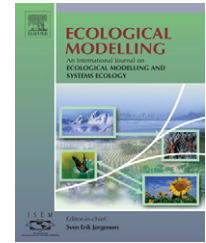


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

Sensitivity analysis and stability patterns of two-species pest models using artificial neural networks

Young-Seuk Park^{a,*}, Jorge Rabinovich^b, Sovan Lek^c

^a Department of Biology and The Korea Institute of Ornithology, Kyung Hee University, Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea

^b Centro de Estudios Parasitológicos y de Vectores (CEPAVE), Universidad Nacional de La Plata, Calle 2 No. 584, e/43 y 44, 1900 La Plata, Prov. de Buenos Aires, Argentina

^c EDB, CNRS-Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse cedex, France

ARTICLE INFO

Article history:

Received 15 October 2004

Received in revised form

30 January 2007

Accepted 30 January 2007

Keywords:

Pest insects

Stability

Host–parasite dynamics

Nezara viridula

Trichopoda giacomellii

Sirex noctilio

Pinus radiata

ABSTRACT

Sensitivity analysis is a critical step in mathematical modelling of ecological processes and it provides an idea of the response of the model dynamics to a variation in the values of some parameters. In analytic models, there are standard mathematical techniques for carrying out sensitivity analyses, but this is not so with simulation models, mainly due to the fact that their behaviour usually depends upon the interaction among different parameters, and so sensitivity analysis has to be carried out for all combinations of all parameters of interest. In this study, we explored the use of artificial neural networks (ANN) for sensitivity analysis of simulation models, as applied to simulation models of two-species pest populations: the parasitoid–host system *Nezara viridula*–*Trichopoda giacomellii*, *N. viridula* being a pest of soybean and the *Sirex noctilio*–*Pinus radiata* system, *S. noctilio* being a pest of pine plantations. We compare the ANN sensitivity analysis results with the ones of the Classification Trees (CT), Sobol and the stepwise multiple regression with standardized partial regression coefficients (SMR). The sensitivity analyses were carried out evaluating the simulation models' parameters effect on the stability behaviour of the simulation models. The ANN sensitivity analysis produced the same (or superior) results as the other two techniques (CT, Sobol and SMR), but showed additional advantages similar to those offered by sensitivity analyses of analytic models: partial derivatives were calculated to determine the contribution of each parameter of the simulation models to their stability behaviour. We conclude that ANN is adequate for simulation modelling sensitivity analysis with the additional advantage of evaluating the contribution of model parameters to the model's behaviour. Although, we used only two-species pest systems as an example, this approach may be applied in wide areas of pest management and population dynamics studies.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Sensitivity analysis is one of the several critical steps in mathematical modelling of ecological processes. In the case of computer simulation models, once a model has been

developed (conceptualized, structured and programmed), verified (no programming errors), subjected to parameterization (numerical estimation of parameters by some goodness of fit criterion of the model to the data) and validated (confirmed that it conforms satisfactorily with field and/or laboratory

* Corresponding author. Tel.: +82 2 961 0946; fax: +82 2 961 0244.

E-mail addresses: parkys@khu.ac.kr (Y.-S. Park), rabinovi@netverk.com.ar (J. Rabinovich), lek@cict.fr (S. Lek).

0304-3800/\$ – see front matter © 2007 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2007.01.021

data, with a data set different from the one used for parameterization), then one of the most frequent and useful additional steps to carry out is a sensitivity analysis.

The purpose of sensitivity analysis is to provide an idea of the response of the model dynamics to a variation in the values of some parameters. One or more outcomes of the model are selected (usually state variables or some statistical indicators) and their behaviour is evaluated for a plausible range of parameter values (McCallum, 2000). This analysis is usually carried out with respect to the main parameters (those suspect of having a strong effect due to conceptual or mathematical relationships with the behaviour of the model) or in relation to those parameters that, for some reason (cost or time constrains, interference with the behaviour of the ecological system or serious alteration and even destruction of the ecological system), could not be estimated in the field or in the laboratory.

When the ecological system has been represented by an analytic model, there are mathematical techniques that provide with adequate methods for carrying out the sensitivity analysis and the interpretation of its results. For example, if we have a simple linear model of the form $\hat{Y}_i = \hat{a}_1 X_{i1} + \hat{a}_2 X_{i2} + \dots + \hat{a}_m X_{im}$, where X_{ij} are the j independent variables for the i observations, \hat{a}_i are the estimated parameters and \hat{Y}_i is the model's predicted values, then $J_{ij} = \partial \hat{Y}_i / \partial a_j$, where $J = J_{ij}$ is called the Jacobian matrix, represents the sensitivity of the model's prediction to the parameter estimates (Hilborn and Walters, 1992). To calculate J , we have to be able to differentiate the model, which is not always possible, particularly when the model is a non-linear differential or difference equation, which sometimes requires some implicit differentiation of the equations (Kot, 2001).

Matters are not always simpler with simulation models. Due to the fact that the behaviour of a simulation model may vary with each parameter of interest and that there may be some degree (sometimes a very strong one) of interaction among the parameters, sensitivity analysis has to be carried out for all combinations of all parameters of interest. Even with the present-day calculation power of most computers this may demand a computer power not always available. If a model has 10 parameters, and we perform a sensitivity analysis assigning only 5 values to each parameter, the number of times the model will have to be run is 5^{10} ; even with a fast computer, that may execute the complete simulation model in 1 s, this would take 113 days (0.3 years) of computations on a 24-h a day basis. In these type of cases, simulation models suffer from what Bellman (1957) called (in the context of dynamic programming) the "curse of dimensionality".

In addition to the computer power problems, the results become difficult to analyze due to the mere size of the output, even if special programming techniques may help in screening and filtering the output for only particular or desirable types of results. There are some statistical tools available to analyze this kind of output (multiple regression, analysis of variance, principal components and Classification Trees, among others), that help making some sense from the appalling size of the sensitivity analysis results. However, few of these tools provide specific information of the type and degree of effect that each parameter produces on some selected model behaviour.

Artificial neural networks (ANNs) have been used as a tool in ecological modelling (for an introduction see Lek and Guégan, 1999). A multilayer perceptron (MLP) with a backpropagation learning algorithm, which is a supervised ANN, has been implemented in various applications (Lek and Guégan, 2000): patterning complex relationships (Lek et al., 1996; Tuma et al., 1996), predicting population and community development (Recknagel et al., 1997; Chon et al., 2000) and modelling habitat suitability (Paruelo and Tomasel, 1997; Özesmi and Özesmi, 1999). The explanatory power of the MLP has been criticized due to its black-box model approach, but now sensitivity analysis methods have been developed to identify the most influent variables in MLP models (Lek et al., 1996; Scardi and Harding, 1999; Dimopoulos et al., 1999). Although, the apparent complexity of ANNs was originally believed to limit our ability to gain explanatory insight into the prediction process, recent advancements (Olden and Jackson, 2002; Gevrey et al., 2003) have illustrated that this indeed is not the case and researchers now have the ability to identify individual and interacting contributions of the predictor variables in ANNs (Olden et al., 2004).

In this work, we apply for the first time the ANN methodology to the sensitivity analysis of simulation models of ecological population dynamic processes. We used two insect population simulation models and tested three sensitivity analysis tools and evaluated possible ANN comparative advantages.

2. Methods

We used two simulation models: (a) the population regulation of *Nezara viridula* (Hemiptera: Pentatomidae) by its parasitoid *Trichopoda giacomellii* (Diptera: Tachinidae), from now on referred to as the *Nv-Tg* model and (b) the population dynamics of a pest species (*Sirex noctilio*) (Hymenoptera: Siricidae) that affects pine plantations (mainly of *Pinus radiata*), from now on referred to as the *Sn-Pr* model. The former was programmed in FORTRAN77 and the latter in C language. The biological and ecological description of the *Nv-Tg* system is given in full details in Liljesthröm and Rabinovich (2004), and biological details about the wood wasp can be found in Ipinza and Molina (1991). After simulation of each model with different combinations of parameter values, model behaviours according to changes of parameters were patterned and predicted with model parameters using SOM and MLP, respectively. Sensitivity analyses of MLP models were carried out with partial derivatives algorithm. ANN algorithms were implemented in Matlab (The MathWorks 2001). Overall modelling procedure is given in Fig. 1. Below we provide a summary of each population dynamics simulation model and the ANN methodology as it was applied for sensitivity analysis.

2.1. The *N. viridula*-*T. giacomellii* (*Nv-Tg*) model

The simulation model is based on two Leslie matrices, one for the host and one for the parasite, with a time unit of 1 week, and populations expressed as individuals/m². Some model parameters were estimated in the laboratory and in the field, while other parameters were estimated by param-

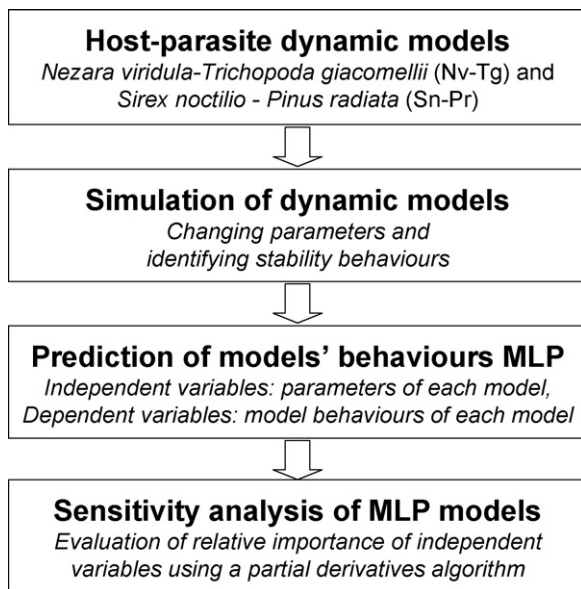


Fig. 1 – Modelling procedure with dynamic models and artificial neural networks. Parameters of each model were used as input variables in MLP models to predict model behaviours. Parameters and model behaviours are given in Tables 1 and 2 for Nv-Tg model and Sn-Pr model, respectively. Modelling procedures for two different models, Nv-Tg model and Sn-Pr model, were carried out independently.

eterization to 15 generations of field data of the host and the parasite. The stability analysis of the model showed four types of behaviour (Liljesthröm and Rabinovich, 2004): two unstable and two stable dynamics (the latter were typical limit cycles, although frequently with more than one peak per cycle). These four stability types of simulation model behaviour were used as a response indicator for sensitivity analysis of the Nv-Tg

model.

Three parameters and two population regulation mechanisms were considered to have major influences on stability, and varied with a wide range. The three parameters were: (i) the aggregated distribution of attacks, as described by the parameter k of the negative binomial distribution, (ii) the host survival in the first week of parasitized life (S_1), which is also an indication of parasitoid larval survival, because larvae cannot complete their development in a week or less and (iii) the host survival in the second week of parasitized life (S_2), which also measures the proportion of parasitoid larvae developing in more than 2 weeks. The two population regulation mechanisms were: (i) the castration effect (cease of reproduction of the parasitized hosts) and (ii) the differential selectivity for hosts by the parasitoid. These two population regulation mechanisms were implemented as on-off variables: 0 = absent and 1 = present. A description of the model's state (response) indicators used for sensitivity analysis as a function of different parameter/mechanism values are given in Table 1.

The degree of influence of different parameters and mechanisms on the behaviour of the model's state (response) indicators was evaluated by means of the Classification Trees method (Breiman et al., 1984), a multivariate technique similar to discriminant analysis but of a hierarchical nature and that is applied in a recursive way, and that has the advantage of combining a variety of types of predictor variables (categorical predictors, continuous predictors or any mix of them). For some type of ecological data, this method has proved to be as equally effective as a mixed effect analysis of variance (but far simpler) and more effective than linear regression (De'ath and Fabricius, 2000). As in the Nv-Tg model, stability behaviour was dependent of a mix of categorical (CAS and SEL) and continuous predictors (k , S_1 and S_2). The Classification Trees method was selected to determine the importance of various parameters on the stability behaviour of the Nv-Tg model. The software Statistica (Version 5.5A) was employed (StatSoft, 1999). The Classification

Table 1 – Description and values of the Nv-Tg simulation model parameters and regulation mechanisms used for sensitivity analysis (independent or predictor variables) and the stability behaviour types used as the simulation model's response indicators

| Parameters/output behaviour | Description | Range (min-max) (step) |
|-------------------------------------|--------------------------------------------------------------------------------------------|---------------------------------------------|
| Parameters (predictive variables) | | |
| Parameter S_1 | Host survival in the first week of parasitized life | 0.1-1 (0.1) |
| Parameter S_2 | Host survival in the second week of parasitized life | 0.1-1 (0.1) |
| Parameter k | Degree of aggregation of the distribution of attacks (as a negative binomial distribution) | 0.05, 0.1 and 0.2, and 0.5-2.3 (0.3) |
| Mechanism CAS (castration index) | Cease of reproduction of the parasitized hosts | 0-1 (1) |
| Mechanism SEL (selectivity index) | Differential selectivity for sex and/or stage of hosts by the parasitoid | 0-1 (1) |
| Model behaviour (output variables) | | |
| Unstable dynamics type 1 (Kodest-1) | Parasitoid and host growing without bound | Output variables (model's output behaviour) |
| Unstable dynamics type 2 (Kodest-2) | Extinction of the parasitoid and unbounded host population growth | |
| Stable dynamics type 1 (Kodest-3) | Stable cycles, both species reaching densities much higher than any observed field value | |
| Stable dynamics type 2 (Kodest-4) | Stable cycles, both species at densities in agreement with all observed field values | |

Trees analysis was applied using equal prior probabilities and equal misclassification costs; to test predictive accuracy 10 random samples for cross-validation from the learning sample were used, and applied them to predict class membership in the test sample. As stopping rule to control stopping tree splitting we used the rule called “pruning on misclassification errors”, which was applied with a minimum n of five. As in this application univariate splits were performed, the predictor variables (CAS, SEL, k , S_1 and S_2) can be ranked on a 0–100 scale in terms of their potential importance in accounting for responses on the dependent variable (details on how these rankings are calculated can be found in Breiman et al., 1984, pp. 146–150).

2.2. The *S. noctilio*–*P. radiata* (Sn–Pr) model

A spatially explicit individual-based simulation model of the wood wasp *S. noctilio* was developed (J. Aparicio, J. Corley and J. Rabinovich, personal communication, 2003) to explore whether outbreak densities may be reached without resorting to physical environmental factors. In this simulation model, the wasp population develops in a pine tree plantation, and each tree belongs to one of four categories: (i) healthy and unsuitable for oviposition, (ii) suitable for oviposition but not stressed, (iii) suitable for oviposition and stressed and (iv) dead. Tree growth or removal was not considered in the model. A squared area plantation was assumed, composed of $P \times P$ trees. Trees within a radius R of another tree is a short way for referring to all the $(2R + 1)^2$ first neighbours of that tree (R -neighbourhood).

Wasps are assumed to have a constant 1:1 male–female ratio. Only females were modelled, and it was assumed that all females are fertilized and capable of laying eggs. Each wasp visits τ trees during her life. The following is a brief description of how the visiting process was modelled. Because stressed trees are chemically “labelled”, they are detected with high probability. It was assumed that stressed trees within a radius R_s are detected with probability P_s . In such a case, the simulated wasps are moved to the position of the stressed tree where they lay N_e eggs. When more than one stressed tree is present in the neighbourhood, the wasp selects one at random. If there are no stressed trees in the neighbourhood the wasp selects between a long distance dispersal, moving to a tree chosen at random and a local dispersal, moving to a tree chosen at random in the R -neighbourhood.

The number of eggs laid by wasp i is a random number extracted from a Poisson distribution with mean H_i , which depends upon the tree’s category (e.g., in the case of unsuitable trees, eggs are laid with a (low) probability P_{e3}). About 60% of the eggs laid during a season emerge in following seasons while the rest emerge 1 year later. At the end of a given season, the simulated adult wasp population is set to zero. Inside trees the egg and larval mortality is assumed to be instantaneous and occurring at the end of the season in which egg laying took place. Egg and larval survival was set as a function of tree category. Wasps’ attacks induce changes in tree category: unsuitable trees which in a given season receive more than N_{th3} eggs become suitable for the next season; suitable trees receiving more than N_{th2} eggs become stressed; and stressed trees receiving more than N_{th1} eggs become dead.

The simulation model was executed with $P = 100$ (about 9 ha), for 200 years on a daily basis, and started with a randomly distributed small percentage (P_{ST_0}) of suitable trees (stressed) in a plantation of unsuitable (healthy) trees. A certain number of pioneer female wasps (S_0) was also randomly distributed in the plantation. Female replacement rate is greater than one only when the wasps find suitable trees for oviposition. Because it is assumed that suitable trees represent a scarce resource, population replacement rate is around one. In some cases, the wasp population goes to extinction; in other cases, the wasps attack some stressed trees that then become suitable for oviposition by neighbouring wasps with high probability, and a population outbreak takes place. Because of resource limitation ultimately the wasp population always results extinct.

The simulation model was executed 1000 times, and cases in which the wasp population went to extinction, with or without outbreaks, were registered. The following statistics were computed: (i) frequency of extinctions (F_{ext}), calculated as the number of runs that went extinct, with or without outbreaks, over the total number of runs, (ii) time to extinction, calculated as the average time from wasp introduction until population extinction in each run, with (T_{ext-wo}) outbreak, (iii) the same without (T_{ext-no}) outbreak, (iv) time to outbreak (T_{out}), calculated, for the runs that resulted in outbreaks, as the time from wasp introduction to the appearance of the first stressed tree and (v) value of outbreak (V_{out}), calculated as the average number of wasps at the peak time of the outbreak. These five statistics were used as output variables for sensitivity analysis as a function of different parameter. Their description and the parameter values used are given in Table 2. In addition, when calculating the averages of the time to extinction and the time to outbreak, their respective standard deviations (ST_{ext-wo} , ST_{ext-no} and ST_{out}) were also calculated.

Sensitivity analysis of the Sn–Pr dynamic simulation model results were analyzed with two procedures: a linear stepwise multiple regression technique and the Sobol method. A FORTRAN program was prepared that, in addition to the stepwise multiple regression procedure, it also calculated the standardized partial regression coefficients (regression coefficients expressed in standard deviation units). These standardized regression coefficients allow a straightforward arithmetic comparison of the relative importance of each independent variable in relation to the dependent variables. That is, if the standardized partial regression coefficient of one parameter is twice the value of the standardized partial regression coefficient of another parameter, the former is twice more important (“useful”) to predict the dependent variables than the latter. For brevity, we will refer to this method as the “stepwise” method.

For the Sobol method, we used the winding stairs sampling technique on the parameter space to reduce the number of simulations (Chan et al., 2000). The Sobol method for sensitivity analysis (Saltelli et al., 1999), measures the model’s sensitivity to the parameters by partitioning the total variance of the output variable Y in main effects and interaction effects among parameters. We calculated the first-order sensitivity index for the i th parameter (S_i), which measures the effect of parameter x_i on the output variable Y and the total sensitivity

Table 2 – Description and parameter values of the Sn-Pr population dynamics simulation model (predictor variables) and description the model's output behaviour (statistics of state (dependent) variables) used for sensitivity analysis

| Parameters/output behaviour | Description | Parameter change for stepwise: min-max (step) | Parameter change for the Sobol method (mean) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------|
| Parameters (predictive variables) | | | |
| PropA2 | Proportion of type 2 trees (suitable for oviposition but not stressed) | 0.01–0.05 (0.01) | 0.025 |
| P_LDF | Probability of long distance flight | 0.1–0.5 (0.1) | 0.25 |
| R | Radius of dispersal | 1–5 (1) | 3.0 |
| τ | Number of trees visited by a wasp during her lifetime | 8–12 (1) | 10 |
| MAX_NE | Maximum number of eggs that a tree can receive | 400–500 (100) | 450 |
| Egg load | Maximum egg number per female | 80–120 (40) | 100 |
| Model stability behaviour (output variables) | | | |
| F_{ext} | Frequency of extinction | Output variables (model's output behaviour) | |
| $T_{\text{ext-wo}}, ST_{\text{ext-wo}}$ | Average time to extinction with outbreaks, and standard deviation | | |
| $T_{\text{ext-no}}, ST_{\text{ext-no}}$ | Average time to extinction without outbreaks, and standard deviation | | |
| $T_{\text{out}}, ST_{\text{out}}$ | Average time to outbreak, and standard deviation | | |
| V_{out} | Average number of wasps at the peak time of the outbreak | | |
| The value of the last column represents the mean of a normal distribution, and the random value was generated assuming a 25% coefficient of variation around the corresponding mean, and then used in a combination order determined by the winding stairs procedure. | | | |

index (S_{Ti}), which takes into account the interactions between the i th parameter and the rest of the parameters. The total sensitivity index can be thought of as the expected fraction of variance that would be left if only the parameter x_i were to stay undetermined.

2.3. The artificial neural network (ANN) methodology

A multilayer perceptron (MLP) with a backpropagation algorithm was used as a non-linear predictor (Haykin, 1994) of population dynamics patterns. MLP is a supervised interactive learning algorithm designed to minimize the mean square error between the computed output of the network and the desired output; for a detailed description of the learning rules of MLP see Rumelhart et al. (1986), Kung (1993) and Lek and Guégan (2000). The network usually consists of three layers: an input layer, one or more hidden layers and an output layer. Each layer is composed of neurons, which are the computational units of MLP. It requires input vectors in the input layer, as well as target (or desired) values in the output layer corresponding to each input vector. The input layer contains neurons for the independent variables. We used five input neurons for the *Nv-Tg* system and four for the *Sn-Pr* system, i.e., one input neuron for each parameter. The output layer was composed by the neurons responsible for the production of the output variables to be predicted (i.e., four types of model dynamics for *Nv-Tg* model and five types of model behaviour for *Sn-Pr* model) (Tables 1 and 2). In this network, signals are

propagated from the input layer through the hidden layers to the output layer via the network connections. During the training phase, a comparison is made between the output values calculated by MLP and the expected values (the ones generated by the population simulation models), and the connection weights are modified in order to minimize the error of the response (difference between expected and calculated output values).

We used a hold-out cross-validation procedure to determine the performance of the MLP method: one half of the dataset was used for training the MLP model, one fourth of the dataset for validation and the last one fourth for testing. The dataset consisted of 4000 cases for *Nv-Tg* model using stepwise, and for the *Sn-Pr* model the dataset consisted of 15,625 cases using stepwise and 2400 cases using Sobol. The parameter values used in each combination are given in Tables 1 and 2.

The performance of the MLP models was evaluated using Cohen's Kappa statistics (Cohen, 1960) for *Nv-Tg* model (i.e., binary outputs) and the correlation coefficients between expected values generated by the population dynamics models and the calculated values of the MLP model for *Sn-Pr* model (i.e., continuous outputs).

2.4. Sensitivity analysis with MLP

After the learning process of the MLP models, sensitivity analysis was carried out to evaluate the contribution of each input variable (parameters of the population dynamic models) to the

Table 3 – Comparison of the sensitivity analysis of the *Nv-Tg* population model by the Classification Trees and the ANN (MLP) methods

| Predictor variable | Classification Trees | ANN (MLP) | | | | Average (S.E.) |
|--------------------|----------------------|-----------------------|----------|----------|----------|----------------|
| | | Kodest-1 ^a | Kodest-2 | Kodest-3 | Kodest-4 | |
| CAS | 45 | 0.12 | 0.16 | 0.25 | 1.02 | 0.39 (0.21) |
| SEL | 49 | 0.12 | 0.17 | 0.20 | 2.04 | 0.63 (0.47) |
| S ₁ | 41 | 1.92 | 6.32 | 8.06 | 51.12 | 16.86 (11.50) |
| S ₁ | 49 | 0.25 | 0.48 | 1.37 | 0.82 | 0.73 (0.24) |
| k | 100 | 97.59 | 92.87 | 90.12 | 44.99 | 81.39 (12.23) |

In both cases, the same matrix of $N=4000$ cases was used, the predicted variables being the four population stability patterns (Kodest-1 to Kodest-4) of the *Nv-Tg* population model (Table 1). In the Classification Trees method, the ranking of relative importance of each population model's parameter is from 0 (minimum importance) to 100 (maximum importance). In the ANN method, the values show the contribution (%) of the predictor variables in determining the population stability patterns and their average contribution (%) (standard error in parenthesis).

^a Types of model behaviour defined in Table 1.

output values of MLP. There are several ways to perform sensitivity analysis with MLP (Zurada et al., 1994; Dimopoulos et al., 1999). Gevrey et al. (2003) compared performances of several different methods to evaluate the relative contribution and/or the contribution profile of the input factors in MLP models. The algorithms compared were the 'partial derivatives' (PaD) method (Dimopoulos et al., 1995, 1999), the 'weights' method (Garson, 1991; Goh, 1995), the 'perturbation' method (Scardi and Harding, 1999), the 'profile' method (Lek et al., 1995, 1996) and the 'classical stepwise' regression method (Sung, 1998). Gevrey et al. (2003) showed that the PaD method was the most useful (followed by the profile method) to identify the degree of contribution of the input variables, and Olden et al. (2004) also showed that the PaD performed well, although it was only consistent in identifying the two most important variables in the network. We decided to use the PaD method because it is more coherent from a computational point of view.

The PaD method presents the output of the MLP models with respect to the input to obtain the profile of the variations of the output for small changes of one input variable (Dimopoulos et al., 1995, 1999; Gevrey et al., 2003). The formula for the partial derivatives (d_{ji}) is:

$$d_{ji} = S_j \sum_{h=1}^{n_h} w_{ho} I_{hj} (1 - I_{hj}) w_{ih} \quad (1)$$

where S_j is the derivative of the output neuron with respect to its input, I_{hj} the response of the h th hidden neuron, w_{ho} and w_{ih} are the weights between the output neuron and h th hidden neuron, and between the i th input neuron and the h th hidden neuron, respectively.

If the partial derivative is negative then, for each parameter being analyzed, the output variable will tend to decrease as the input parameter increases. Inversely, if the partial derivative is positive, the output variable will tend to increase as the input parameter increases. The relative contribution of input descriptors to the MLP output can be estimated as the sum of the squared partial derivatives (SSD) obtained for each input variable. The SSD values allow the classification of the variables according to their contribution to the output variable in the model, the input variable with the highest SSD value being the variable which most influences the output variable. The details of the MLP sensitivity analysis as applied

to ecological modelling can be obtained from Gevrey et al. (2003).

3. Results

3.1. The *N. viridula-T. giacomellii* (*Nv-Tg*) model

3.1.1. Prediction of population stability patterns

The Classification Trees analysis showed that, with an input-output matrix of 4000 cases, 207 cases were misclassified (5.2%), the worst misclassification occurring when 74 cases of stability Type 4 were classified as stability Type 3. The highest relative importance ranking in determining the model's stability behaviour corresponds to k , with an importance about twice the others (Table 3).

MLP applied to the *Nv-Tg* dynamic model showed high predictability of the population stability patterns with 95% or higher of correct predictions (Table 4). Cohen's Kappa showed very high values (range 0.86–0.94), indicating very high agreement between model dynamics types predicted by the MLP model and the corresponding expected types generated by the population dynamics models. The frequency histogram of error values showed that most error values lie around zero.

3.1.2. Influence of parameters

The negative binomial parameter k is the most important factor in determining the patterns of population dynamics stability behaviour (Kodest-1 to Kodest-4; see Table 1 for stability types). The contribution of k to behaviour Kodest-4 is relatively lower than to other behaviours, whereas S_1 is the highest contribution parameter to Kodest-4. Parameter S_1 also provides a relatively low contribution, and the rest of the parameters show a very low contribution.

As the parameter k showed the highest contribution to the model's stability behaviour, it was of interest to evaluate the response behaviour of the system in response to changes of k value (Fig. 2). Parameter k affects Kodest-1 and Kodest-3 negatively but the effect is positive on Kodest-2 and Kodest-4. Kodest-2 and Kodest-3 showed an inverse response to the changes of k values. It should be noted that the scales are very different for each stability behaviour, and they indicate the level of contribution of parameter k to each stability behaviour type.

Table 4 – Predictability of the population stability patterns (Kodest-1 to Kodest-4) of the *Nv-Tg* simulation model by the MLP procedure using the model's parameters as predictors

| Predicted stability type | Number of cases | | | % Correct answers | Cohen's Kappa |
|--------------------------|---------------------|-----------------------|-----------------|-------------------|---------------|
| | Predicted correctly | Predicted incorrectly | Total cases (N) | | |
| Kodest-1 | 153 | 14 | 167 | 91.62 | 0.9408 |
| Kodest-2 | 111 | 14 | 125 | 88.80 | 0.889 |
| Kodest-3 | 217 | 33 | 250 | 86.80 | 0.861 |
| Kodest-4 | 448 | 10 | 458 | 97.82 | 0.918 |
| Total | 929 | 71 | 1000 | | |

3.2. The *S. noctilio*-*P. radiata* (*Sn-Pr*) model

3.2.1. Prediction of population dynamics patterns

The results of the outbreak behaviour of the *Sn-Pr* model are expressed as correlation coefficients between the expected (simulation model) and the ANN calculated values as indicators of the predictability of the MLP model. The correlation coefficients ranged from 0.66 to 0.99 (0.97, 0.71, 0.73, 0.66 and 0.96, for output variables F_{ext} , T_{ext-no} , T_{ext-wo} , V_{out} , T_{out} , respectively), with the highest values related to "Frequency of extinction" (F_{ext}) ($p < 0.01$ for all variables).

3.2.2. Influence of parameters

Fig. 3 compares the contribution of the predictor variables to the each of the statistical indicators of model stability behaviours of the *Sn-Pr* population model, as result of the Sobol, the stepwise multiple regression and the ANN using MLP. We considered five output variables (time to outbreak peak, values at outbreak peak, time to extinction with outbreak, time to extinction without outbreak and frequency of extinction) and in some cases we also considered their standard deviations (for the ANN methodology). The contributions of input parameters for frequency of extinction were very similar for the three different methods (Fig. 3). However, the stepwise multiple regression method displayed relatively different patterns for time to extinction with outbreak, time to extinction without outbreak and time to Sirex peak (Fig. 3).

The Sobol and the MLP methods shows very similar contribution with $r = 0.921$ ($N = 30$, $p < 0.001$) of correlation coefficient in overall (all parameters and all output variables), although the differences were relatively large for values of Sirex peak. The agreement between the MLP and the stepwise methods was smaller, but still significant ($r = 0.8$, $N = 30$, $p < 0.01$). For all output variables, the contribution of tau (τ) (number of trees visited) was the highest in all three models. The contribution was also higher in the ANN method than the Sobol method. Input variables P.LDF, Max.NE and Egg load showed low contributions for all output variables in all three different methods.

Therefore, we looked into the response behaviour of the model according to changes of these two parameters (PropA2 and τ) that had the highest contribution to the model's stability behaviour. The response behaviour of the MLP method output as a function of changes in PropA2 had negative effects on "frequency of extinction", "mean time to extinction" and "standard deviation of time to extinction" (Fig. 4). However, its influence on these output variables decreased with higher values of PropA2, as displayed by partial derivatives being scattered around zeros at high values of PropA2. PropA2 had positive effects on "mean time to outbreak" and its standard deviation at low values; however, the influence was low (with relatively small negative effects) at high values, indicating that the increase of PropA2 resulted in a decrease in the output variables "frequency of extinction", "mean time to extinc-

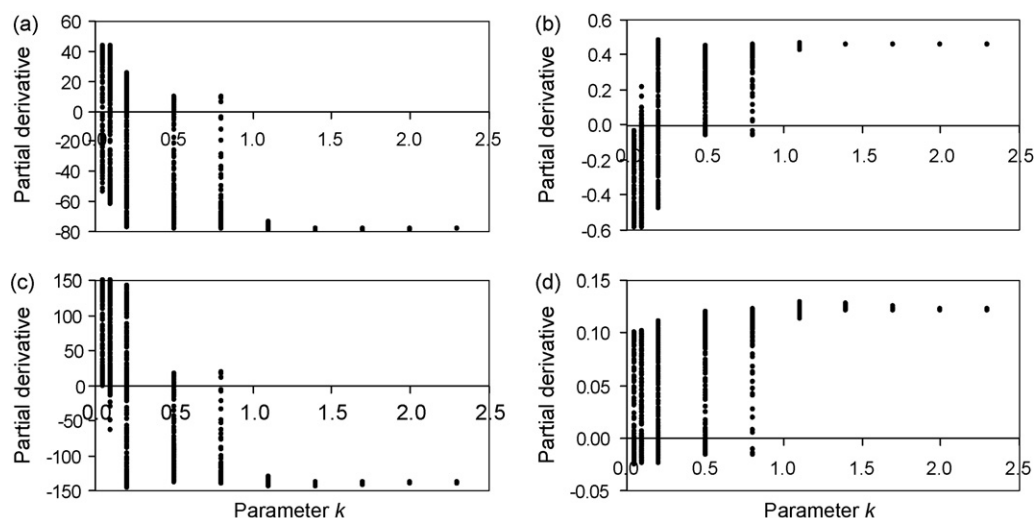


Fig. 2 – Profile of partial derivatives as a function of parameter k for each type of population stability behaviour of the *Nv-Tg* population simulation model. (a)–(d) Correspond to Kodest-1 to Kodest-4 stability types (see Table 1 for stability types).

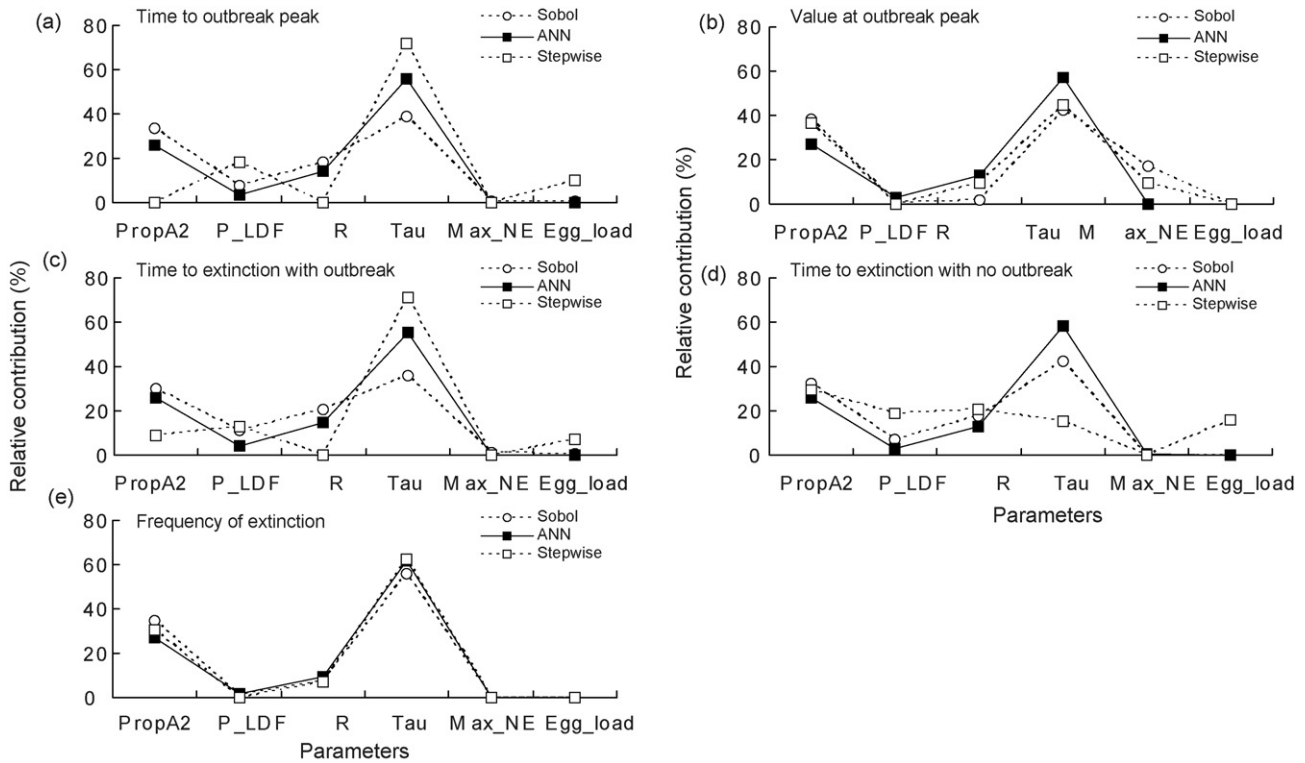


Fig. 3 – Comparison of the contribution of the predictor variables to the model stability behaviours of the Sn-Pr population model, as result of the Sobol, the stepwise multiple regression and the ANN (MLP) methods.

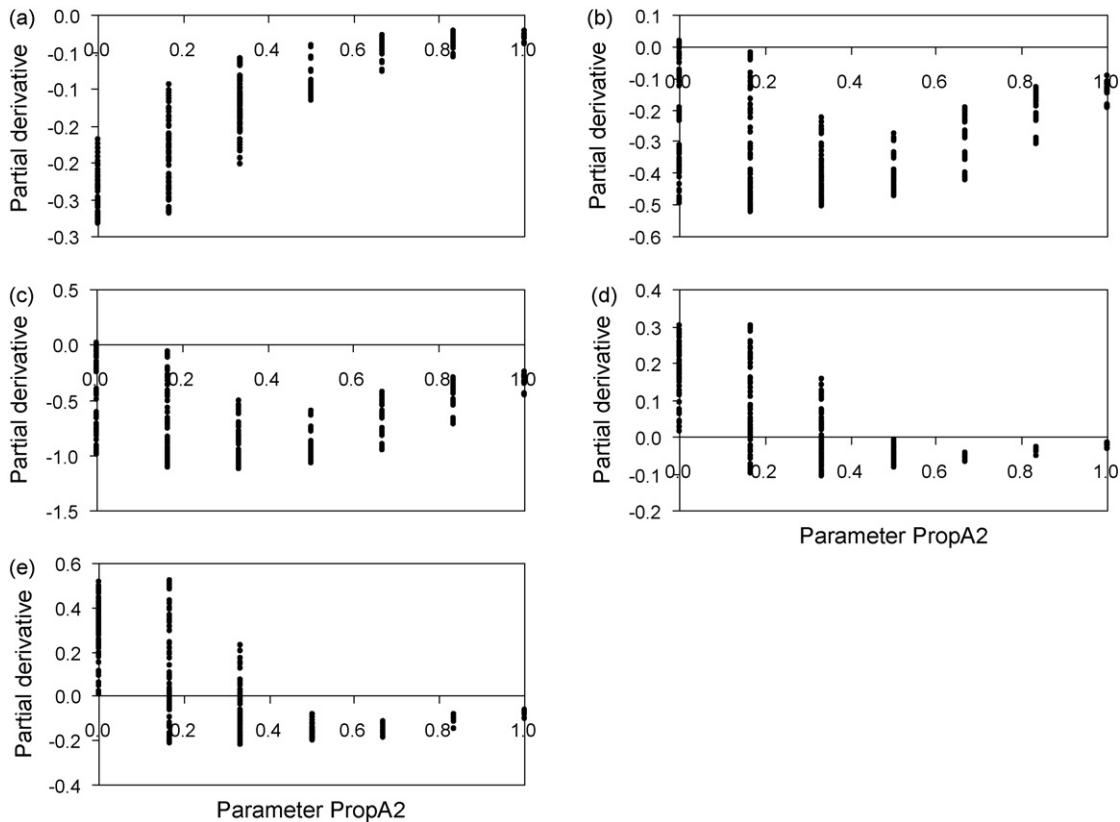


Fig. 4 – Partial derivatives as a function of the parameter “PropA2” (proportion of type 2 trees) on output variables of the Sn-Pr population simulation model. (a) “Frequency of extinction” (F_{ext}), (b) “Mean time to extinction” (T_{ext}), (c) “Std. dev. of time to extinction” (ST_{ext}), (d) “Mean time to outbreak” (T_{out}) and (e) “Std. dev. of time to outbreak” (ST_{out}).

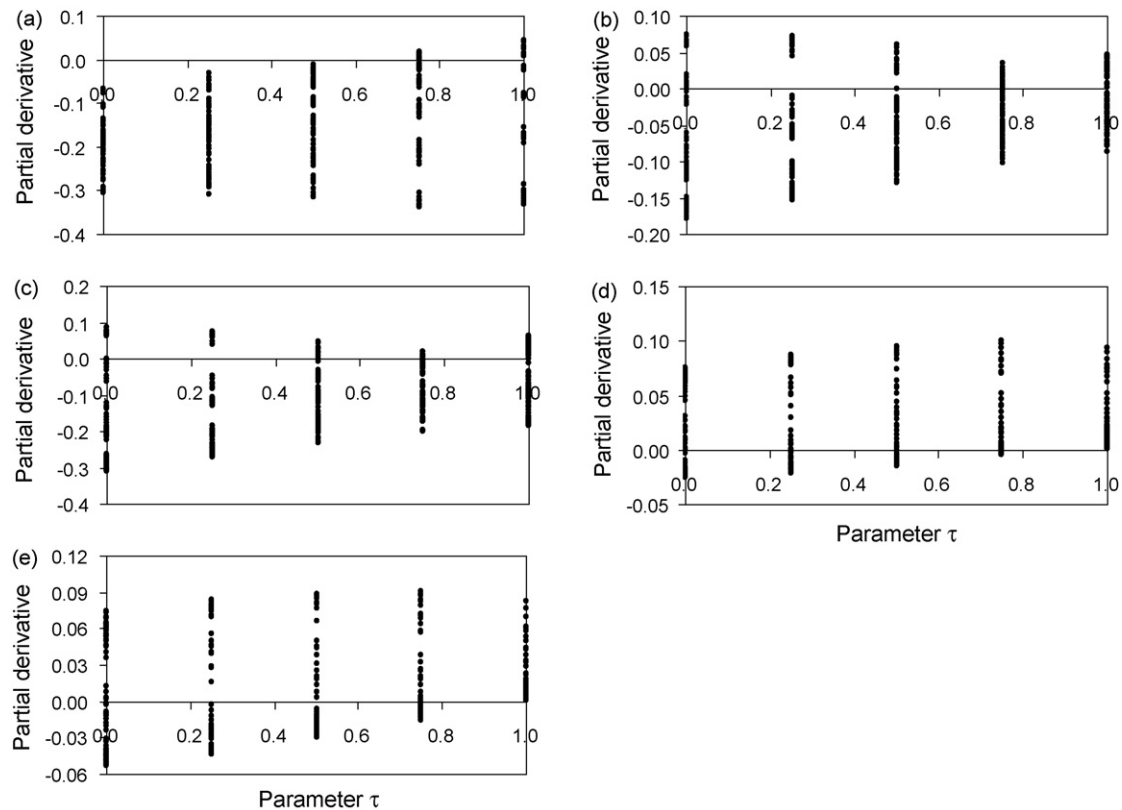


Fig. 5 – Partial derivatives as a function of the parameter τ (number of trees visited) on output variables of the *Sn-Pr* population simulation model. (a) “Frequency of extinction”, (b) “Mean time to extinction”, (c) “Std. dev. of time to extinction”, (d) “Mean time to outbreak” and (e) “Std. dev. of time to outbreak”.

tion” and “standard deviation of time to extinction”, while it resulted in an increase in “mean time to outbreak” and its standard deviation.

The partial derivatives response of the MLP procedure from changes in the parameter τ showed negative effects on the dynamic model’s output variables frequency of extinction, but the effects were relatively smaller at high τ values (Fig. 5). The responses of time to extinction (mean and standard deviation) were not strong, showing both positive and negative influences, with the response stabilizing as the values of τ increased. The parameters time to outbreak (mean and standard deviation) responded mainly positively, although large variations were observed.

4. Discussion and conclusions

Sensitivity analysis is a critical step in mathematical modelling of ecological processes and it provides an idea of the response of the model dynamics to a variation in the values of some parameters. In analytic models there are standard mathematical techniques for carrying out sensitivity analyses, but this is not so with simulation models, mainly due to the fact that their behaviour usually depends upon the interaction among different parameters. Therefore, sensitivity analysis in simulation models has to be carried out for all combinations of all parameters of interest. Due to these properties of the

simulation model, some statistical tools are used as indirect sensitivity analysis.

In this study, we explored the use of ANN for sensitivity analysis of simulation models, as applied to simulations models of two-species pest populations: the parasitoid–host system *N. viridula*–*T. giacomellii* (*Nv-Tg*), *N. viridula* being a pest of soybean and the *S. noctilio*–*P. radiata* (*Sn-Pr*) system, *S. noctilio* being a pest of pine plantations. The ANN (MLP) sensitivity analysis results of the *Nv-Tg* and the *Sn-Pr* simulation models agree quite well with the results based upon the other sensitivity analyses tested. In particular, the following aspects deserve to be emphasized.

In the case of the *Nv-Tg* simulation model, the variation of the contribution of each predictor variable (simulation model parameters and mechanisms) by the MLP model is relatively high because it was calculated from the contribution of a small number of outputs (i.e., in this case four outputs for dynamics stability types). By both methods (MLP and Classification Trees) the dominance of the aggregation of attacks by the parasitoid (the negative binomial parameter k) on the stability behaviour of the simulation model was evident. The other two parameters (S_1 and S_2) and the two biological mechanisms (CAS and SEL) played a lesser role by both methods, although the MLP model assigned a relatively higher importance to the parameter S_1 (the hosts’ survivorship in the first week of adult life). This makes biological sense because the value of S_1 is also used in the *Nv-Tg* simulation model as the variable deter-

mining the emergence of the parasitoid. In the case of the two biological mechanisms (CAS and SEL), the ANN methods had a high standard error because these are categorical variables.

The MLP model can offer additional information than the Classification Trees by providing an indication of the importance of each input variable upon stability behaviour of each simulation model. For example, it shows that the *Nv-Tg* simulation model parameter S_1 actually is second in importance after the parameter k , but only in relation to the stability behaviour type 4 (stable cycles with host and parasitoid species at densities in agreement with all observed field values). The contribution of S_1 on this type was 51.1%, whereas the effect was very low on other types ranged between 1.9% and 8.1%.

The ANN methodology can also predict additional features related to the sensitivity analysis, not present in the Classification Trees method. It provides, as the mathematical techniques available for carrying out the sensitivity analysis of analytic models do, partial derivatives of each simulation model's stability behaviour as function of the simulation model parameters and mechanisms. We carried out this procedure with the only dominant parameter (the negative binomial parameter k), and the results show a very high agreement with the dominant theory of the role of this parameter (an indicator of the aggregation of attacks by the parasitoid) on the stability behaviour of parasite–host systems. May (1978) considered the overall consequence of spatial heterogeneity as satisfactorily described by a negative binomial distribution, and Hassell (2000) claimed that homogeneity, in the sense of equal risk of being parasitized, results in a decrease in host survival as parasitoid density increases, leading to extinction of host–parasitoid systems. On the other hand, heterogeneity (e.g., aggregated distribution of parasitoid attacks among hosts) may result in a stabilization of the host–parasitoid system (Liljeström and Bernstein, 1990). This is so because decline in host survival with parasitoid density is more moderate than exponential and exerts a stabilizing influence (Hassell et al., 1991), due to less severe reductions in the host population following the build-up of parasitoid numbers (Chesson and Rosenzweig, 1991). The results of the MLP sensitivity analysis also agree with the field results that show that in the *N. viridula*–*T. giacomellii* system, the aggregated distribution of attacks fits well a negative binomial distribution with the aggregation parameter k estimated to lay between 0.2 and 0.8 (Liljeström, 1992), thus playing a dominant role in the stabilization of this host–parasitoid system. The partial derivative profiles of the ANN sensitivity analysis agree with the results of May (1978) in that the aggregated distribution of attacks among hosts – as represented by the negative binomial distribution parameter k – should lead to stability if $k < 1$.

In the case of the *Sn-Pr* simulation model, we compared the contribution of the predictor variables to the each of the statistical indicators of model behaviours between the stepwise linear multiple standardized regression method, the Sobol method and the ANN sensitivity analysis methodology (Fig. 3). We obtained a good agreement among three different methods. All of them show that the population simulation model parameter PropA2 (proportion of type 2 trees, i.e., suitable for oviposition but not stressed) is the dominant one in

determining the model's statistical behaviour, followed by the parameter τ (number of trees visited by a female wasp during her lifetime). The contributions of input parameters for frequency extinction were very similar for the three different methods. However, the stepwise multiple regression method displayed relatively different patterns for time to extinction with outbreak, time to extinction without outbreaks and time to Sirex peak.

The MLP sensitivity analysis methodology was able to show the specific contribution of the predictor variables (population dynamics simulation model parameters) to the each of the statistical indicators of the *Sn-Pr* population model's behaviour. Additionally, while the stepwise regression only shows the standardized coefficient values when an effect was statistically significant (and not deleted from the stepwise regression) the ANN method shows average values of the contribution for all combinations of predictor variables and model statistical behaviours of the *Sn-Pr* population model. A measure of the agreement between both sensitivity analysis methodologies was obtained by a simple linear correlation between the standardized coefficient values of the stepwise regression and the percent contribution of each parameter to the population model statistical behaviour as provided by the ANN methodology. The correlation coefficient was statistically significant ($r = 0.824$, $N = 13$, $p < 0.001$).

Another advantage of the MLP sensitivity analysis is that we can evaluate the model response behaviour against the changes of input variables, while it is not convenient in the Classification Trees, the Sobol and the stepwise multiple regression methods. We looked into the response behaviour of the model according to changes of parameters (PropA2 and τ in *Sn-Pr* model) through a partial derivatives profile method of the MLP sensitivity analysis.

In general, MLP results conformed well to the Classification Trees for the *Nv-Tg* model and the stepwise regression analysis for the *Sn-Pr* model. These classical statistical analysis techniques are not straightforward in handling non-linearity's, and do not incorporate causality in their models (Gevrey et al., 2003), while ANN has the capability to handle non-linear, complex ecological data and to incorporate causality (Lek and Guégan, 2000; Recknagel, 2003). Although ANN models are able to make very good predictions and are recognized as powerful tools (Skelton et al., 1995; Recknagel et al., 1997; Liong et al., 2000), at the beginning of their development they were considered as black-box approaches because of a lack of explanatory methods for relationships between input and output variables. Presently, many different algorithms have been developed to avoid the "black-box" flaw of ANNs, and now they can be used as sensitivity analysis tools to determine the contributions of the independent variables and the way they act on the dependent variable (Garson, 1991; Goh, 1995; Lek et al., 1996; Balls et al., 1996; Maier and Dandy, 1996; Scardi and Harding, 1999; Dimopoulos et al., 1995, 1999; Olden, 2003).

Our conclusion is that the MLP has high predictive power and is adequate for evaluating the contribution of model parameters to the model's behaviour. It has the advantage over the stepwise linear multiple regression in that it does not require that the simulation model behaves in a linear way. And although the Sobol method can also cope with non-linear

simulation models, it is a quite sophisticated method, and no standard analytical commercial package is still available. Of course the same can be said from the MLP method of the ANN approach, but in comparing the two (Sobol and ANN) the latter can determine the contributions of the independent variables and also the way they act on the dependent variable, while the former can only determine the relative contribution of the independent variables or parameters. Our results show that the MLP model is very efficient to predict the stability behaviour of pest population dynamics and to evaluate importance of parameters in the population dynamics. Although, we used only two-species pest systems as an example, this approach may be applied in wide areas of pest management and population dynamics studies.

REFERENCES

- Balls, G.-R., Palmer Brown, D., Sanders, G.-E., 1996. Investigating microclimatic influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks. *New Phytol.* 132, 271–280.
- Bellman, R.E., 1957. *Dynamic Programming*. Princeton University Press, Princeton.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth and Brooks/Cole Advanced Books and Software, Monterey, CA.
- Chan, K., Saltelli, A., Tarantola, S., 2000. Winding stairs: a sampling tool to compute sensitivity indices. *Stat. Comput.* 10, 187–196.
- Chesson, P.L., Rosenzweig, M., 1991. Behaviour, heterogeneity, and the dynamics of interacting species. *Ecology* 72, 1187–1195.
- Chon, T.-S., Park, Y.-S., Kim, J.-M., Lee, B.-Y., Chung, Y.-J., Kim, Y., 2000. Use of an artificial neural network to predict population dynamics of the forest-pest pine needle gall midge (Diptera: Cecidomyiida). *Environ. Entomol.* 29, 1208–1215.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* 20, 37–46.
- De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81, 3178–3192.
- Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A., Lek, S., 1999. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecol. Model.* 120, 157–165.
- Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Proces. Lett.* 2, 1–4.
- Garson, G.D., 1991. Interpreting neural network connection weights. *Artif. Intel. Expert* 6, 47–51.
- Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in Artificial Neural Network models. *Ecol. Model.* 160, 249–264.
- Goh, A.T.C., 1995. Back-propagation neural networks for modelling complex systems. *Artif. Intel. Eng.* 9, 143–151.
- Hassell, M.P., 2000. *The Spatial and Temporal Dynamics of Host-Parasitoids Interactions*. Oxford University Press, Oxford.
- Hassell, M.P., Pacala, S.W., May, R.M., Chesson, M.L., 1991. The persistence of host-parasitoid association in patchy environments. I. A general criterion. *Am. Natur.* 138, 568–583.
- Haykin, S., 1994. *Neural Networks*. Macmillan College Publishing Company, New York.
- Hilborn, R., Walters, C.J., 1992. *Quantitative Fisheries Stock Assessment. Choice, Dynamics and Uncertainty*. Chapman and Hall, London.
- Ipinza, R., Molina, M.P., 1991. Control integrado de *Sirex noctilio*. *Ciencia e Investigación Forestal* 5, 96–149.
- Kot, M., 2001. *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge.
- Kung, S.Y., 1993. *Digital Neural Networks*. Prentice Hall, Eaglewood Cliffs, New Jersey.
- Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling an introduction. *Ecol. Model.* 120, 65–73.
- Lek, S., Guégan, J.F. (Eds.), 2000. *Artificial Neuronal Networks: Application to Ecology and Evolution*. Springer, Berlin.
- Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., Moreau, J., 1995. Improved estimation, using neural networks, of the food consumption of fish populations. *Marine Freshwater Res.* 46, 1229–1236.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.* 90, 39–52.
- Liljeström, G., 1992. Distribución de los ataques del parasitoide *Trichopoda giacomellii* (Diptera: Tachinidae) sobre una población de *Nezara viridula* (Hemiptera: Pentatomidae). *Ecol. Austral* 2, 29–37.
- Liljeström, G., Rabinovich, J.E., 2004. Modelling biological control: the population regulation of *Nezara viridula* by the parasitoid *Trichopoda giacomellii*. *Ecol. Appl.* 14, 254–267.
- Liljeström, G., Bernstein, C., 1990. Density dependence and regulation in the system *Nezara viridula* (L.) (Hemiptera: Pentatomidae), host and *Trichopoda giacomellii* (Blanchard) (Diptera: Tachinidae), parasitoid. *Oecologia* 84, 45–52.
- Liong, S.Y., Lim, W.H., Paudyal, G.N., 2000. River stage forecasting in Bangladesh: neural network approach. *J. Comput. Civil Eng.* 14, 1–8.
- Maier, H.R., Dandy, G.C., 1996. The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* 32, 1013–1022.
- May, R.M., 1978. Host-parasitoid systems in patchy environments: a phenomenological model. *J. Anim. Ecol.* 47, 833–843.
- McCallum, H., 2000. *Population Parameters. Estimation for Ecological Models*. Blackwell Science, Oxford.
- Olden, J.D., 2003. A species-specific approach to modelling biological communities and its potential for conservation. *Conserv. Biol.* 17, 854–863.
- Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol. Model.* 178, 389–397.
- Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: understanding variable contributions in artificial neural networks. *Ecol. Model.* 154, 135–150.
- Özesmi, S., Özesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecol. Model.* 116, 15–31.
- Paruelo, J.M., Tomasel, F., 1997. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecol. Model.* 98, 173–186.
- Recknagel, F., 2003. *Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation*. Springer, Berlin.
- Recknagel, F., French, M., Harkonen, P., Yabunaka, K.-L., 1997. Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.* 96, 11–28.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. I: Foundations. MIT Press, Cambridge, pp. 318–362.
- Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model independent method for global sensitivity analysis of model output. *Technometrics* 41 (1), 39–56.

- Scardi, M., Harding, L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecol. Model.* 120, 213–223.
- Skelton, P.H., Cambray, J.A., Lombard, A., Benn, G.A., 1995. Patterns of distribution and conservation status of freshwater fishes in South Africa. *S. Afr. J. Zool.* 30, 71–81.
- StatSoft, Inc., 1999. *Statistica for Windows* (Computer Program Manual). Statsoft Inc., Tulsa.
- Sung, A.H., 1998. Ranking importance of input parameters of neural networks. *Expert Syst. Appl.* 15, 405–411.
- Tuma, A., Haasis, H.D., Rentz, O., 1996. A comparison of fuzzy expert systems, neural networks and neuro-fuzzy approaches controlling energy and material flows. *Ecol. Model.* 85, 93–98.
- Zurada, J.M., Malinowski, A., Cloete, L., 1994. Sensitivity analysis for minimization input data dimension for feedforward neural network. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 6, London, May 28–June 2. IEEE Press, pp. 447–450.