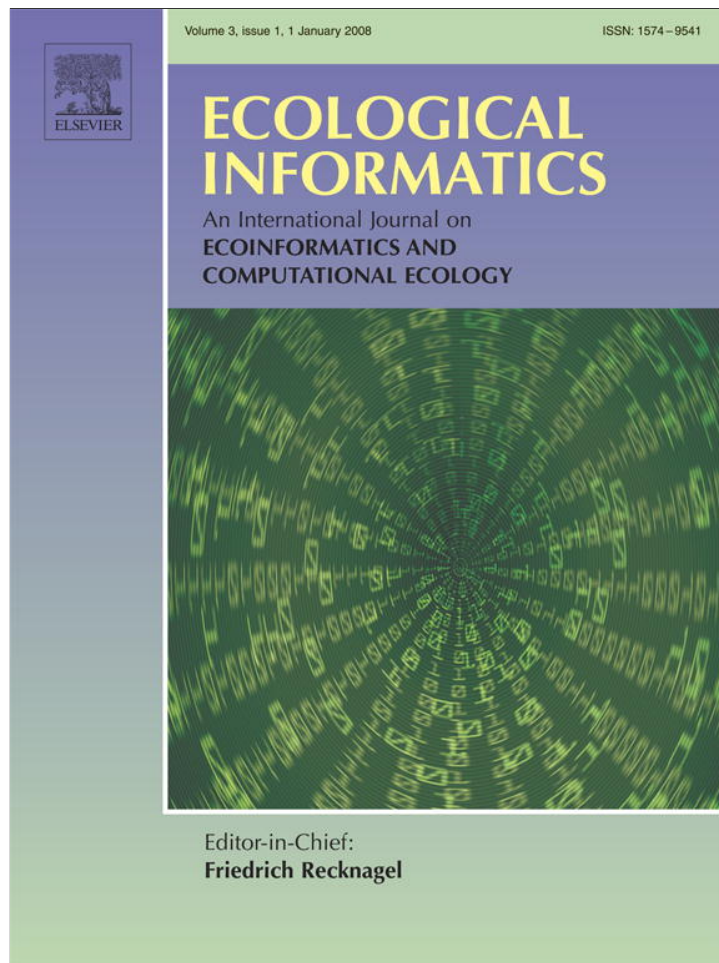


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

available at www.sciencedirect.comwww.elsevier.com/locate/ecolinf

Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species

Michael J. Watts^{a,*}, S.P. Worner^b

^aSchool of Biological Sciences, University of Sydney, NSW 2006, Australia

^bNational Centre for Advanced Bio-Protection Technologies, PO Box 84, Lincoln University, Lincoln 7647, New Zealand

ARTICLE INFO

Article history:

Received 22 December 2006

Received in revised form

20 June 2007

Accepted 22 June 2007

Keywords:

Artificial neural networks

Multi-layer perceptrons

Input contribution analysis

Knowledge discovery

ABSTRACT

The use of multi-layer perceptrons (MLP) to determine the relative significance of climatic variables to the establishment of insect pest species is described. Results show that the MLP are able to learn to accurately predict the establishment of a pest species within a specific geographic region. Analysis of the MLP yielded insights into the contribution of the individual input variables and allowed for the identification of those variables that were most significant in either encouraging or inhibiting establishment.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

The increasing movement of people and goods around the globe, along with the removal of trade barriers and increasing human influence means that as a consequence, many other species invade areas where they are not normally found. Such species threaten human health as well as productive and natural systems (Peterson and Vieglas, 2001).

Given that food is readily available, climate is considered to be the most important factor influencing the establishment of insect pests in new locations (Worner, 1988; Peacock et al., 2006). While the response of a species to different climatic factors can often be determined experimentally, such controlled experiments are expensive and time consuming (Araujo and Pearson, 2005). Additionally, when an exotic species is detected in a new region scientists are often required to carry out a pest risk analysis and advise agencies how to proceed, given little information. There is, therefore, a need to develop methods that can determine the most

important climatic factors influencing the establishment for individual species from existing climatic and species distribution data (Worner, 1994; Sutherst et al., 1995; Andersen et al., 2004; Andersen, 2005; Dentener et al., 2002; Park et al., 2003; Stynes, 2002; Sutherst, 2003).

Numerous statistical tools have been applied to the problem of predicting the establishment of species in regions where they are not normally found (Baker and MacCleod, 2004; Guisan and Zimmermann, 2000). These tools include regression techniques (Eyre et al., 2005; Lehmann et al., 2002; Soltic et al., 2004a,b), generalised additive models (Guisan et al., 2002; Zaniewski et al., 2002), stochastic simulations (Rossi et al., 1993), quantitative methodologies (Rafoss, 2003) and multivariate analysis (Peacock et al., 2006; Bobesberger, 2002). These traditional methods often have difficulties modelling ecological problems, because the relationships between variables in ecology are often non-linear (James and McCulloch, 1990).

An alternative technique that can be applied to ecological problems that involve complex but unknown relationships

* Corresponding author. Tel.: +61 2 9351 3898.

E-mail address: michael.watts@bio.usyd.edu.au (M.J. Watts).

between variables is artificial neural networks (ANN) (Crick, 1989; Haykin, 1994; Tarassenko, 1998; Reed and Marks, 1999). ANN have many advantages over statistical-based techniques. They can learn from existing data and therefore do not require an *a priori* model. If over-fitting is avoided they can generalise well, in other words, they can accurately classify data they have not been trained on. Additionally, ANN can model systems that involve multiple dependent variables and complex non-linear relationships between variables and outcomes. ANN can learn from noisy data (Kasabov, 1996). Also, the outputs of a type of ANN, the multi-layer perceptron (MLP) can be interpreted as probabilities. For example if an MLP is used to predict the presence of a species, the output of the MLP can be interpreted as the probability of that species being present. The widely used MLP is trained by the back-propagation of errors algorithm (Rumelhart et al., 1986). MLP have been successfully applied to several different problems in ecology (Kimes et al., 2000). For example Dimopoulos et al. (1999) used MLP to model the relationships between cities and the levels of contaminants in grasses. Brosse et al. (1999) used MLP to predict the abundance of freshwater fish based on environmental factors in a reservoir. Prediction of fish abundance was also the subject of a study by Lek et al. (1996), who modelled the abundance of brown trout, and of Laë et al. (1999), who predicted the fish yields of African lakes. The simple presence or absence of certain species of fish were also predicted based on environmental factors by Mastrorillo et al. (1997) and Joy and Death (2004). Jeong et al. (2006) modelled the presence of river phytoplankton in a South Korean river-catchment area, while the presence or absence of mayflies was predicted from environmental factors by Wagner et al. (2000). The distribution of more complex organisms have also been modelled by MLP: Mael et al. (1999) successfully used MLP to predict the presence of the Himalayan river bird *Rhyacornis fuliginosus*.

In ecological studies where ANN have been compared to traditional statistical models, the ANN have consistently outperformed the statistical models with respect to prediction accuracy (Brosse et al., 1999; Ibarra et al., 2003; Jeong et al., 2006; Laë et al., 1999; Lek et al., 1996; Mael et al., 1999; Mastrorillo et al., 1997; Wagner et al., 2000, 2006; Soltic et al., 2004a,b).

MLP are often used as a black box, where the model is constructed and used without any consideration of what the network has learned. This approach is perfectly acceptable if the goal is only to use the MLP to make predictions from data, but by opening up the black box and analysing what has been learned insights can be gained into how the MLP is making

Table 1 – Established species

Taxonomic name	Common name	Prevalence
<i>Myzus persicae</i>	Green peach aphid	234
<i>Brevicoryne brassicae</i>	Cabbage aphid	210
<i>Planococcus citri</i>	Citrus mealybug	220
<i>Aphis gossypii</i>	Cotton aphid	250
<i>Nezara viridula</i>	Green stink bug	222
<i>Erisoma lanigerum</i>	Woolly aphid	182

“Prevalence” lists the number of geographic regions in which each particular species was recorded as being present.

Table 2 – Top six non-established species

Name	Common name	Prevalence
<i>Sitophilus zeamais</i>	Greater grain weevil	127
<i>Drosophila melanogaster</i>	Common fruit fly	109
<i>Spodoptera exigua</i>	Beet armyworm	128
<i>Melanaphis sacchari</i>	Yellow sugarcane aphid	124
<i>Ceratitidis capitata</i>	Mediterranean fruit fly	103
<i>Chrysomphalus aonidum</i>	Red scale	174

Captions are the same as in Table 1.

these predictions, and what factors influence the predictions (Garson, 1991; Milne, 1995; Olden and Jackson, 2002; Gevrey et al., 2003; Recknagel et al., 2006).

The goal of this study was to open up the MLP black box and use it to identify the relative contribution of climate factors that influenced the establishment of some insect pest species.

This investigation consisted of two parts. First, MLP were trained on the relevant data and evaluated, and the contribution of each input variable analysed to determine which variables were the most significant. Second, the significance of the contribution of influential inputs was verified by a randomisation test.

2. Method

2.1. Data

The data set used in this study consisted of data of the climate for each of 459 geographic regions and the presence or absence in each of those regions of 844 insect pest species. The species presence data were sourced from the *Crop Protection Compendium* (CABI 2003). The climate data were compiled from Internet sites maintained by recognised meteorological organisations. None of this data contained any explicit information about the links between climate in a region and the presence or absence of any particular species.

The insect pest species were divided into two groups, those species that are recorded as being present in New Zealand, and those that are not. The non-established species were ordered according to the threat posed by the species, according to the method described in Watts and Worner (2006) and Worner and Gevrey (2006). The top six species of each set were selected as case studies for this paper and are listed in Table 1, for the species that are recorded as established in New Zealand, and Table 2 for the species that are not recorded as established in New Zealand.

Forty five climate variables were available and are listed in Table 3. To represent the range of each variable within a region the minimum, mean and maximum of each variable was calculated. There was therefore a total of one hundred and thirty five input variables describing the climate in each region. The data for each variable was linearly normalised to the range of zero to unity and comprised the input to the MLP models.

The data was randomly split into two major sets. The first, containing 80% of the data, was the training and test set, from which samples were randomly drawn to form training and test data sets for each trial. The second comprised the

Table 3 – Climate variable names and descriptions (units)

Variable	Description
TSum1	Temperature first month of summer (°C)
TSum2	Temperature second month of summer (°C)
TSum3	Temperature third month of summer (°C)
TAut1	Temperature first month of autumn (°C)
TAut2	Temperature second month of autumn (°C)
TAut3	Temperature third month of autumn (°C)
TWin1	Temperature first month of winter (°C)
TWin2	Temperature second month of winter (°C)
TWin3	Temperature third month of winter (°C)
TSpr1	Temperature first month of spring (°C)
TSpr2	Temperature second month of spring (°C)
TSpr3	Temperature third month of spring (°C)
TSummer	Summer temperature (°C)
TWinter	Winter temperature (°C)
Tannual	Annual temperature (°C)
RSumR1	Rainfall first month of summer (mm)
RSumR2	Rainfall second month of summer (mm)
RSumR3	Rainfall third month of summer (mm)
RAutR1	Rainfall first month of autumn (mm)
RAutR2	Rainfall second month of autumn (mm)
RAutR3	Rainfall third month of autumn (mm)
RWinR1	Rainfall first month of winter (mm)
RWinR2	Rainfall second month of winter (mm)
RWinR3	Rainfall third month of winter (mm)
RSprR1	Rainfall first month of spring (mm)
RSprR2	Rainfall second month of spring (mm)
RSprR3	Rainfall third month of spring (mm)
Rannual	Annual rainfall (mm)
PEannual	Annual potential evapotranspiration (mm per year)
AEannual	Annual actual evapotranspiration (mm per year)
Mi	Moisture index (mm)
AnnualDaylength	Length of day from sunrise to sunset (hours)
AD50mm	Annual soil moisture deficit at 50 mm (mm)
AS50mm	Annual soil moisture surplus at 50 mm (mm)
AD150mm	Annual soil moisture deficit at 150 mm (mm)
AS150mm	Annual soil moisture surplus at 150 mm (mm)
AD300mm	Annual soil moisture deficit at 300 mm (mm)
AS300mm	Annual soil moisture surplus at 300 mm (mm)
AD700mm	Annual soil moisture deficit at 700 mm (mm)
AS700mm	Annual soil moisture surplus at 700 mm (mm)
Im300	Soil moisture index at 300 mm depth (mm)
DD5	Degree-days 5 degrees (days)
DD15	Degree-days 15 degrees (days)
Thornthwaite value	Moisture index based on summation of monthly moisture values (mm)
Climate values	Most extreme minimum of all climate values (no units)

validation set, which was used to perform an independent evaluation of the prediction accuracy for each target species.

2.2. Training and evaluation of MLP

Standard three neuron-layer multi-layer perceptrons (MLP) were used in these experiments, and the learning algorithm used was unmodified back-propagation with momentum (Rumelhart et al., 1986). The parameters of the MLP and learning algorithm were three hidden neurons, a learning rate and momentum of 0.03 and 750 training epochs. These parameters were found via experimentation to yield the best balance of training and generalisation errors: more than three

hidden neurons consistently caused over-training, while less meant the MLP were unable to learn.

The method of training and evaluating the MLP (and also selecting the parameters) was similar to that suggested in Flexer (1996) and Prechelt (1996). A total of one thousand trials were performed over each species. For each trial, the training and test data set (consisting of 80% of the total data available) was randomly divided into a training set and a testing set in a two-thirds/one-third ratio. An MLP was then created with randomly initialised connection weights and trained over the training division. The accuracy of the MLP over the training division was then evaluated to determine how well the network had learned, and again over the testing division to determine how well the network generalised. Accuracy was measured using Cohen's Kappa statistic (Cohen, 1960). Whereas percentage accuracy is easily interpreted, it is also easily biased by unbalanced numbers of classes. That is, percentage correct may be misleadingly high when the data set in question has only a small number of examples from one class. Cohen's Kappa statistic takes the number of examples of each class into account and thus yields a less biased measure of accuracy than percentages.

For each trial the contributions of each input neuron to the output of the network was also determined. Many methods have been proposed for determining the importance of each of the input neurons of an MLP. These include the methods of Garson (1991), Milne (1995), Gevrey et al. (2003) and Olden and Jackson (2002). Since it was desirable to identify features that negatively contributed to establishment, and the methods of Milne (1995) and Gevrey et al. (2003) return unsigned values, these methods were rejected. Of the remaining methods, the work of Olden et al. (2004) has shown that the method of Olden and Jackson is the least biased, and it has also been previously used in ecological modelling applications (Joy and Death, 2004). Thus, this is the method that was selected.

At the completion of the one thousand trials, the MLP with the highest kappa over the test data (that is, the MLP with the best generalisation performance) was selected as the winner for that species. The accuracy of this winning network was then evaluated over the validation data set, so that an unbiased estimate of the generalisation capability of the MLP could be obtained.

2.3. Sensitivity analysis of inputs

A sensitivity analysis was also performed over each input variable of the winning network. This was carried out to illustrate the response of the network to variations of each variable so that the influence of strongly contributing inputs (as determined above) could be investigated.

The sensitivity analysis was performed by clamping each input, except the one being investigated, at its mean value for the data set. The values for the input being investigated were then varied across the range of zero to unity and the network recalled for each step.

2.4. Verification of variable importance

To verify these analyses, a randomisation test was used. The approach used here is similar to the procedure of null-model

Table 4 – Mean and standard deviation of accuracies (as Cohen's Kappa) for established species

Species	Train	Test	Validate
<i>M. persicae</i>	0.69/0.05	0.49/0.07	0.37
<i>B. brassicae</i>	0.68/0.05	0.49/0.07	0.36
<i>P. citri</i>	0.59/0.19	0.37/0.13	0.42
<i>A. gossypii</i>	0.60/0.07	0.38/0.08	0.31
<i>N. viridula</i>	0.63/0.18	0.41/0.13	0.39
<i>E. lanigerum</i>	0.75/0.04	0.55/0.07	0.48

“Train” is the accuracy over the training data set. “Test” is the accuracy over the test data set. “Validate” is the accuracy over the validation data set.

analysis of ecological data, such as that used for the analysis of species assemblages (Gotelli, 2000). In null-model analysis, a matrix of species presence and absence is randomly rearranged and compared to the original matrix. If there are no significant differences between a sufficient number of random (or null) matrices and the original, observed, matrix, then the species assemblages are said to be random. Whereas random matrices are created in null-model analysis, the importance of MLP input variables in this study were verified by evaluating the performance of randomly-selected sets of input variables, and comparing that performance with the accuracy of sets of input variables selected by the weightings assigned by input contribution analysis. The procedure comprises two main steps.

Firstly, MLP were constructed and trained on sub-sets of inputs that were selected according to the mean contributions of the inputs. The inputs were selected so that the sum of their absolute mean contributions was equivalent to predetermined percentages of the total absolute contribution of all inputs. In other words, the sum of the absolute contribution of all inputs was first calculated, and the inputs placed in descending order of mean absolute contribution. Inputs were then selected one at a time, in order, until the sum of the contributions of the selected inputs equalled or exceeded 10, 20, 30...90% of the sum of all contributions. Thus, the number of inputs that was selected varied for each species.

Secondly, MLP were constructed and trained on an identical number of randomly selected input variables, where variables were randomly selected without replacement. The null hypothesis in this verification procedure was that there would be no significant difference between the performance of MLP trained on randomly selected sub-sets of inputs, and the performance of MLP trained on sub-sets of inputs that were selected according to the results of the input contribution analysis. In other words, if there was no

Table 5 – Mean and standard deviation of accuracies (as Cohen's Kappa) for non-established species

Species	Train	Test	Validate
<i>S. zeamais</i>	0.44/0.28	0.26/0.18	0.40
<i>D. melanogaster</i>	0.44/0.14	0.18/0.10	0.22
<i>S. exigua</i>	0.54/0.23	0.33/0.16	0.60
<i>M. sacchari</i>	0.49/0.31	0.33/0.21	0.58
<i>C. capitata</i>	0.46/0.33	0.29/0.22	0.43
<i>C. aonidium</i>	0.37/0.29	0.19/0.17	0.51

Column labels are the same as in Table 4.

significant difference between the performance of the MLP trained using the selected inputs and the MLP using randomly selected inputs, then the method of selecting inputs could be said to be no better than selecting them randomly.

One thousand trials were performed for each species, where a trial involved creating and evaluating one MLP using inputs selected by contribution, and one MLP using randomly selected inputs.

3. Results

3.1. Initial results

The results for the initial experiments are presented in Table 4, for the established species, and Table 5 for the non-established species. The accuracies are presented as Cohen's Kappa statistic (Cohen, 1960).

These results show that the MLP were able to learn the relationships between the climate variables and the presence of the target species. For the established species, the performance over the training sets was uniformly high, with the lowest kappa 0.59 for *Planococcus citri*, and the highest 0.75 for *Erisoma lanigerum*. The standard deviation over the training kappas was also quite low for *Myzus persicae*, *Brevicoryne brassicae*, *Aphis gossypii* and *E. lanigerum*. This shows that the training kappas were consistently high, which indicates that the MLP were modelling species distribution quite well. The

Table 6 – Main contributing variables for established species

<i>B. brassicae</i>		<i>M. persicae</i>	
Variable	Contribution	Variable	Contribution
Max RSprR1	9.99/5.35	Max RSprR1	18.76/5.76
Max TAut1	9.67/3.31	Mean TAut1	9.85/3.52
Min RSprR3	-11.77/4.58	Min PEannual	-11.00/3.88
Min RAutR3	-13.28/3.98	Min DD15	-18.10/5.02
<i>P. citri</i>		<i>A. gossypii</i>	
Variable	Contribution	Variable	Contribution
Max RSprR1	15.47/5.79	Max TAut1	14.53/3.74
Max RSumR1	10.68/4.93	Max PEannual	13.78/4.57
Min RSumR2	-10.70/4.69	Min Climate Values	-11.10/5.46
Max TSpr2	-11.46/5.50	Min DD15	-14.08/5.50
<i>N. viridula</i>		<i>E. lanigerum</i>	
Variable	Contribution	Variable	Contribution
Max RSprR1	14.10/5.45	Max DD5	11.70/3.75
Max RAutR1	12.26/5.58	Max RWinR2	11.10/4.47
Mean Annual Day Length	-11.04/4.17	Mean DD15	-11.04/3.88
Min Annual Day Length	-17.33/4.89	Min DD15	-18.06/5.96

Variable names are the same as in Table 3. “Contribution” is a unit-less measure of input contribution to the activation of the output, where a positive contribution increases output activation and a negative contribution decreases output activation.

Table 7 – Main contributing variables for top non-established species

<i>S. zeamais</i>		<i>D. melanogaster</i>	
Variable	Contribution	Variable	Contribution
Max RWinR2	8.91/4.90	Max Annual Day Length	7.353.46
Max RWinR1	7.78/4.50	Max Im300	7.19/3.24
Max AEannual	-6.85/4.31	Max AEannual	-7.142/3.94
Mean RAutR3	-9.18/4.53	Min Climate values	-7.26/3.65
<i>S. exigua</i>		<i>M. sacchari</i>	
Variable	Contribution	Variable	Contribution
Max RSumR2	8.22/4.53	Max RSumR3	12.67/5.07
Max RSumR3	7.46/3.99	Mean RSumR3	11.56/4.72
Mean Climate values	-12.41/5.73	Min RSprR3	-11.06/5.38
Max AEannual	-13.09/6.16	Min DD15	-12.83/5.74
<i>C. capitata</i>		<i>C. aonidum</i>	
Variable	Contribution	Variable	Contribution
Min DD5	8.31/4.55	Mean TAut1	9.34/4.79
Max DD5	7.97/2.75	Mean TAut2	8.64/4.60
Min AEannual	-10.25/3.51	Mean AD50mm	-7.32/3.76
Mean AEannual	-12.67/4.05	Max Annual Day length	-7.65/3.55

standard deviation over the training kappas for *P. citri* and *Nezara viridula* was somewhat higher. The mean kappas for those species were slightly lower than the other species, which indicates that there was much more variation – and, therefore, greater difficulty – in modelling those species. The

performance over the test sets was also good, with a minimum of 0.37 for *P. citri* and a maximum of 0.55 for *E. lanigerum*. The slightly larger disparity between training and testing performance for *P. citri* and *A. gossypii* suggests that a small amount of over-training may have occurred. However, overall, the test accuracies were quite good. The true test of generalisation, however, was the accuracy over the validation data set. As described in Section 2.2, only the MLP with the highest test accuracy was evaluated with the validation data set. The validation kappas follow a similar range to the test kappas, with a minimum of 0.31 for *A. gossypii* to a maximum of 0.48 for *E. lanigerum*.

The training accuracies for the non-established species ranged from 0.37 for *Chrysomphalus aonidum* to 0.54 for *Spodoptera exigua*. This is, obviously, lower than the range for the established species. Given the lower prevalences of the non-established species, this is not surprising. The kappas over the testing sets ranged from 0.18 for *Drosophila melanogaster* to 0.33 for *S. exigua*. The relatively high training accuracy and low testing accuracy indicates that over-training occurred with *D. melanogaster*. This is further indicated by the low validation kappa of 0.22 for *D. melanogaster*, which was the smallest validation accuracy for these species. The highest validation accuracy was 0.60 for *S. exigua*, which is quite high indicating that the distribution of this species was modelled very well.

Overall, the results show that MLP are able to learn to predict the establishment of insect pest species from climate data.

The mean contribution of each input neuron across all one thousand trials were found and used to rank the inputs in order of importance. For brevity only the two most positive contributing variables and two most negative contributing variables are presented in this paper. The four most

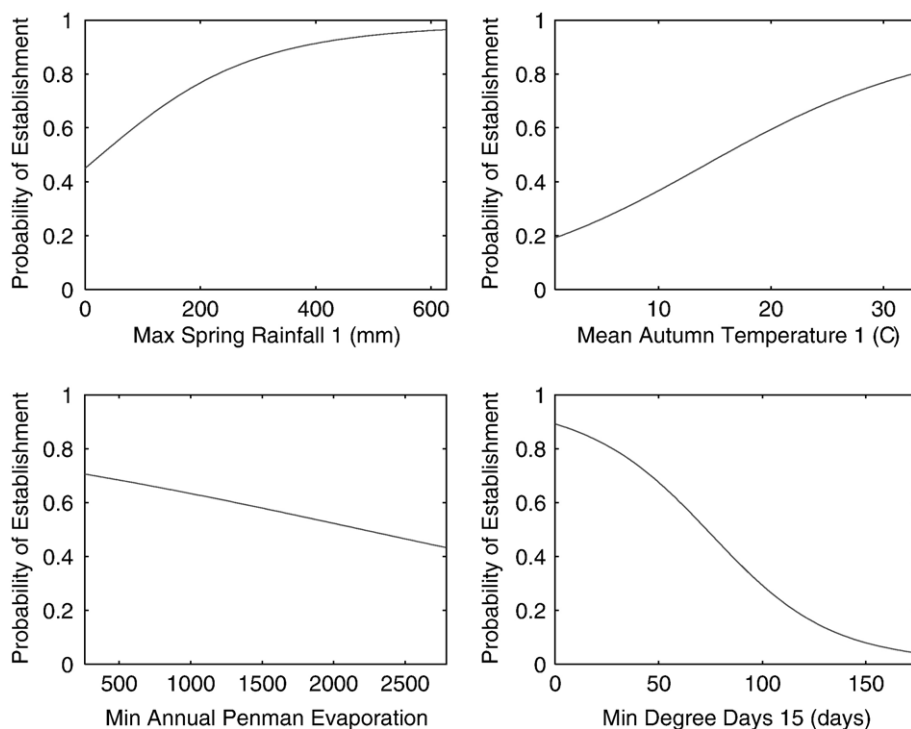


Fig. 1 – Sensitivity analysis of significant variables for *M. persicae*.

contributing variables for each of the established species are presented in Table 6. Table 6 shows that spring and summer rainfalls, along with autumn temperatures, have a large positive effect upon the establishment of these species. The four most contributing variables for each of the non-established species are presented in Table 7. Again, rainfall and temperature variables are predominant.

3.2. Sensitivity analysis of most contributing variables

While a sensitivity analysis was performed for each of the twelve species examined, for brevity the results of only four species are presented here. Two of these species, *M. persicae* and *E. lanigerum* are established in New Zealand, while the other two, *S. exigua* and *C. aonidum* are not.

Fig. 1 presents the results of the sensitivity analysis of the most significant variables for *M. persicae*. The variable “Min Degree Days 15”, which represents a minimum value for the number of degree-days above a threshold of 15 °C in a region, has the largest effect, ranging from 0.9 down to 0.05. This corresponds well to the large negative contribution assigned to that variable. Similarly, the strong increasing effect shown by the response to the “Max Spring Rainfall 1” variable matches the large positive contribution it was assigned. All other climate variables referred to in this discussion are further defined in Table 3.

The results of the sensitivity analysis of *E. lanigerum* are presented in Fig. 2. Both “Max Degree Days 5” and “Max Winter Rainfall”, which have positive contributions. *E. lanigerum* shows a strong increase in establishment probability with corresponding increases in the variable, while “Mean Degree Days 15” and “Min Degree Days 15” show corresponding negative contributions.

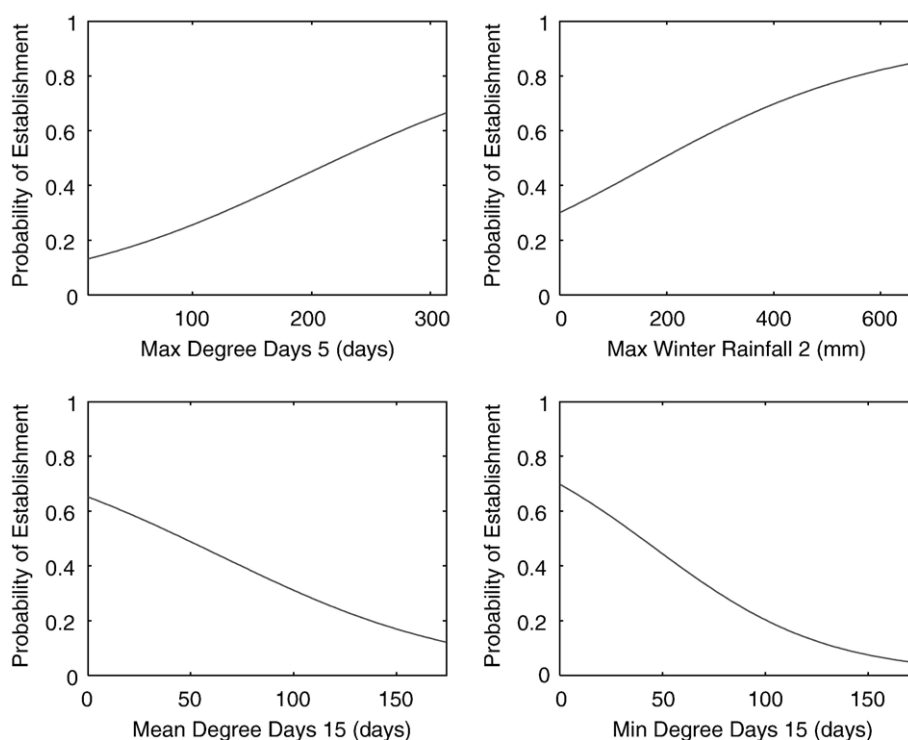


Fig. 2 – Sensitivity analysis of significant variables for *E. lanigerum*.

Fig. 3 displays the results of the sensitivity analysis of *S. exigua*. The effects of both “Max Summer Rainfall 2” and “Max Summer Rainfall 3” are very similar to one another, which reflects the similarity in their contribution (8.22 for “Max Summer Rainfall 2” versus 7.46 for “Max Summer Rainfall 3”). The effects of the two most negatively contributing variables, “Mean Climate Values” and “Max Annual Evapotranspiration” are also similar, and again the contributions are very similar.

The results for *C. aonidum* are presented in Fig. 4. Again, the responses to the top two variables (“Mean Autumn Temperature 1” and “Mean Autumn temperature 2”) are quite similar, and again the mean contributions for these two variables are quite close to one another (9.34 versus 8.64). It is interesting to note that while the mean contribution values for the two most negatively contributing variables “Mean AD50 mm” and “Max Annual Day Length” are quite similar to one another (–7.32 versus –7.65) the responses to changes in these variables are fairly dissimilar to one another. This may be due to the differences in the data used: obviously, no location in the world is going to have a maximum day length of zero hours. However, the slope of this plot does show that this variable has an effect, and is therefore significant.

In general the results of the sensitivity analysis have confirmed the importance of the variables as identified by contribution analysis. However, it has also shown that the values assigned by contribution analysis do not provide information on the relative response of the species to those variables, in other words, just how sensitive species establishment probability is to the value of the variable. Clearly contribution analysis is useful for identifying which variables are significant, whereas sensitivity analysis is useful for determining the behaviour of the model output in response to the significant variables.

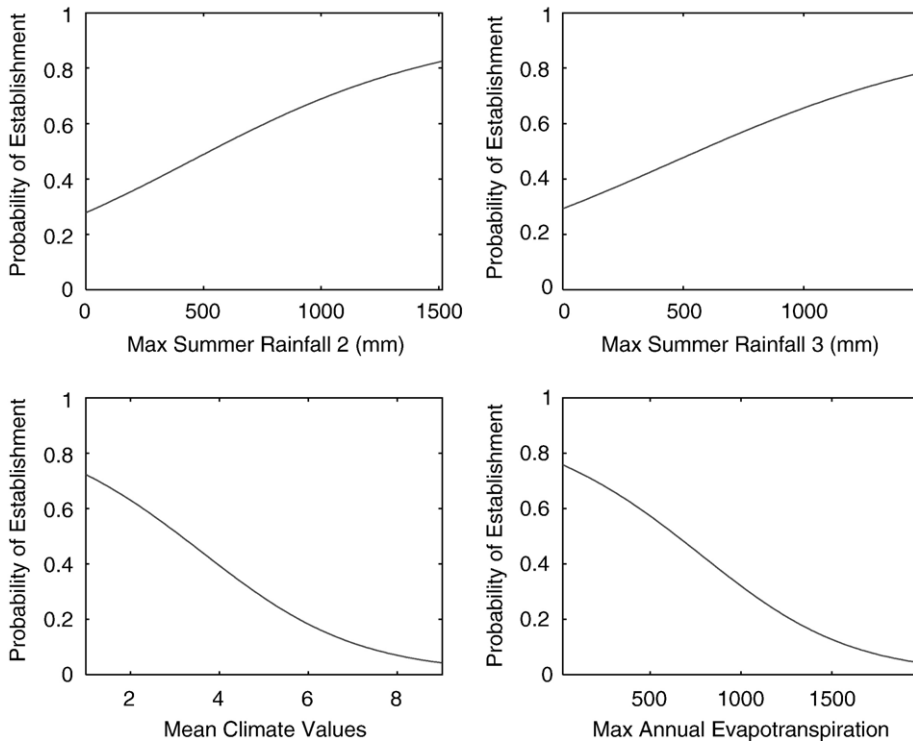


Fig. 3 – Sensitivity analysis of significant variables for *S. exigua*.

3.3. Verification of input significance

The accuracies of prediction as measured by Cohen's Kappa statistic are shown in Table 8 for the established species.

The accuracies for the MLP trained using selected sub-sets of inputs are consistently higher than those trained using randomly selected sub-sets of inputs. To determine if these differences were statistically significant, two-tailed t-tests

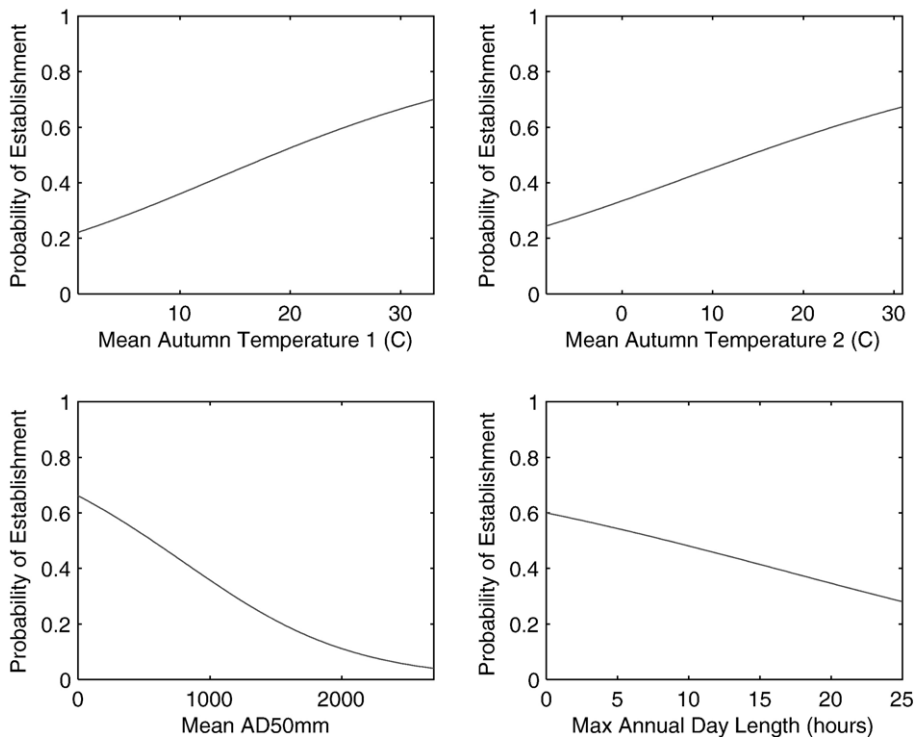


Fig. 4 – Sensitivity analysis of significant variables for *C. anidum*.

Table 8 – Mean and standard deviation of established species accuracies (as Cohen's Kappa) over training data sets

%	<i>B. brassicae</i>			<i>M. persicae</i>			<i>P. citri</i>		
	#	κ_S	κ_N	#	κ_S	κ_N	#	κ_S	κ_N
10	5	0.48/0.04	0.43/0.07*	4	0.57/0.03	0.39/0.09*	5	0.45/0.04	0.33/0.08*
20	10	0.56/0.04	0.52/0.05*	9	0.59/0.03	0.50/0.05*	11	0.55/0.06	0.44/0.07*
30	17	0.60/0.04	0.57/0.04*	16	0.59/0.04	0.55/0.05*	18	0.58/0.06	0.48/0.07*
40	25	0.63/0.04	0.59/0.04*	24	0.65/0.03	0.58/0.04*	25	0.57/0.11	0.50/0.09*
50	35	0.63/0.04	0.61/0.04*	33	0.66/0.03	0.60/0.04*	34	0.56/0.14	0.52/0.11*
60	45	0.64/0.04	0.62/0.04*	43	0.66/0.04	0.62/0.04*	44	0.59/0.14	0.54/0.12*
70	57	0.65/0.04	0.63/0.04*	54	0.67/0.04	0.65/0.04*	56	0.60/0.14	0.56/0.13*
80	72	0.67/0.04	0.61/0.04*	68	0.68/0.04	0.65/0.04*	71	0.60/0.15	0.56/0.15*
90	92	0.68/0.04	0.66/0.04*	86	0.68/0.04	0.66/0.04*	89	0.61/0.16	0.57/0.17*

%	<i>A. gossypii</i>			<i>N. viridula</i>			<i>E. lanigerum</i>		
	#	κ_S	κ_N	#	κ_S	κ_N	#	κ_S	κ_N
10	4	0.47/0.04	0.22/0.10*	4	0.47/0.04	0.40/0.07*	4	0.53/0.04	0.44/0.10*
20	9	0.49/0.04	0.36/0.07*	10	0.51/0.04	0.50/0.05*	9	0.64/0.04	0.55/0.06*
30	14	0.51/0.04	0.41/0.06*	16	0.56/0.04	0.53/0.06*	15	0.65/0.03	0.59/0.05*
40	21	0.52/0.04	0.46/0.05*	24	0.62/0.05	0.56/0.07*	22	0.68/0.03	0.62/0.04*
50	30	0.57/0.04	0.49/0.05*	32	0.64/0.05	0.58/0.09*	29	0.69/0.03	0.64/0.04*
60	39	0.58/0.04	0.52/0.05*	43	0.65/0.07	0.59/0.11*	38	0.71/0.03	0.66/0.04*
70	51	0.60/0.04	0.54/0.05*	55	0.66/0.06	0.60/0.12*	49	0.72/0.03	0.67/0.04*
80	64	0.60/0.04	0.56/0.05*	69	0.66/0.13	0.61/0.14*	66	0.73/0.03	0.69/0.04*
90	83	0.61/0.05	0.57/0.05*	89	0.66/0.14	0.62/0.16*	88	0.74/0.03	0.72/0.04*

“%” is the percentage contribution, “#” is the number of input features for that species, for that percentage of contribution. κ_S is the kappa for the MLP trained using sub-sets of inputs selected by contribution. κ_N is the kappa for the MLP trained using randomly selected sub-sets of inputs (the null model). Significant differences between κ_S and κ_N are denoted by “*”.

were carried out. In each case, the accuracy of the MLP trained using selected sub-sets of inputs was significantly higher ($p=0.001$) than the accuracies of those trained using randomly selected inputs.

To verify that the selection of the random input sets were not biased, additional tests were performed by calculating the degree of similarity between the randomly selected input sets and the input set selected by contribution. These similarities were then compared to the expected level of similarity (two-tailed t-tests, $p=0.001$). If the similarities did not differ significantly, then the sampling of the input set was assumed to be unbiased. The similarities for the established species are presented in Table 9. In each case, there were no significant differences between the actual and expected degrees of similarity.

The accuracies for the non-established species are presented in Table 10. Statistical tests revealed no significant differences for *S. zeamais*, *D. melanogaster* and *S. exigua*. Some differences were found for *M. sacchari*, *Ceratitis capitata* and *C. aonidum*.

The results for *M. sacchari* show that at the 90% level of contribution, there was no significant difference between the accuracies. Thus, the additional inputs that were selected for the 90% level of contribution did not add significantly to the model, and only the inputs selected up to the 80% level of contribution should be considered significant. That *M. sacchari* had the largest number of input variables selected at the 90% level of contribution is also noteworthy: this suggests that the inputs that were added at this point had very low contributions, and therefore did not add much to the network. The results for *C. capitata* and *C. aonidum* are less clear-cut. For *C. capitata* significant differences were found only at the 20, 30 and 40% levels of contribution. For *C. aonidum* differences were found

from the 10–40% levels of contribution, and at 60%. This suggests that only those inputs that were selected up to the 40% level of contribution should be considered significant for *C. aonidum*.

Table 9 – Mean and standard deviation of similarities between variable sets for established species

%	<i>B. brassicae</i>		<i>M. persicae</i>		<i>P. citri</i>	
	Expected	Actual	Expected	Actual	Expected	Actual
10	3.7	3.02/7.4	3.0	2.9/8.4	3.7	3.7/8.5
20	7.4	7.4/7.9	6.7	6.5/8.0	8.1	8.2/8.0
30	12.6	12.8/7.4	11.9	12.0/7.6	13.3	13.7/7.7
40	18.5	18.6/7.3	17.8	18.0/7.2	18.5	18.4/6.9
50	25.9	25.9/6.3	24.4	24.4/6.4	25.2	25.2/6.4
60	33.3	33.8/5.7	31.9	32.0/5.8	32.6	33.0/5.8
70	42.2	42.5/5.2	40.0	40.0/5.1	41.5	41.3/5.0
80	53.3	53.4/4.1	50.4	50.3/4.3	52.6	52.7/5.1
90	68.1	68.1/2.8	63.7	63.6/3.1	65.9	65.9/3.0

%	<i>A. gossypii</i>		<i>N. viridula</i>		<i>E. lanigerum</i>	
	Expected	Actual	Expected	Actual	Expected	Actual
10	3.0	3.1/8.7	3.0	2.9/8.3	3.0	2.6/7.6
20	6.7	6.4/8.2	7.4	7.5/8.2	6.7	6.7/8.3
30	10.4	10.2/7.9	11.9	11.4/7.3	11.1	11.3/7.8
40	15.6	15.6/7.2	17.8	17.8/7.0	16.3	16.2/7.0
50	22.2	22.2/6.2	23.7	23.9/6.7	21.5	21.2/6.6
60	28.9	28.7/6.0	31.9	31.7/6.1	28.1	28.4/6.3
70	37.8	38.0/5.4	40.7	40.6/5.2	36.3	36.1/5.5
80	47.4	47.3/4.6	51.1	51.1/5.2	48.9	49.1/4.4
90	61.5	61.6/3.3	65.9	65.9/2.9	65.2	65.1/3.0

“%” is the percentage contributions used to select the inputs, “Expected” is the expected percentage similarity, and “Actual” the actual percentage similarity.

Table 10 – Mean and standard deviation of non-established species accuracies (as Cohen's Kappa) over training data set

%	<i>S. zeamais</i>			<i>D. melanogaster</i>			<i>S. exigua</i>		
	#	κ_S	κ_N	#	κ_S	κ_N	#	κ_S	κ_N
10	5	0.17/0.10	0.15/0.15*	4	0.17/0.08	0.04/0.06*	4	0.34/0.08	0.23/0.15*
20	10	0.38/0.07	0.27/0.17*	9	0.26/0.08	0.13/0.09*	9	0.48/0.05	0.39/0.11*
30	17	0.44/0.07	0.33/0.20*	16	0.34/0.08	0.19/0.09*	15	0.53/0.06	0.43/0.12*
40	25	0.49/0.10	0.32/0.23*	23	0.35/0.10	0.24/0.09*	23	0.52/0.13	0.45/0.15*
50	34	0.51/0.15	0.34/0.24*	31	0.36/0.11	0.27/0.09*	32	0.54/0.13	0.46/0.16*
60	43	0.47/0.21	0.35/0.25*	41	0.39/0.10	0.30/0.09*	42	0.55/0.15	0.47/0.18*
70	54	0.47/0.024	0.37/0.25*	52	0.41/0.11	0.33/0.09*	54	0.55/0.17	0.48/0.19*
80	66	0.46/0.26	0.40/0.26*	68	0.41/0.13	0.36/0.10*	68	0.55/0.19	0.49/0.21*
90	85	0.46/0.27	0.41/0.27*	87	0.43/0.13	0.39/0.11*	88	0.55/0.20	0.50/0.22*

%	<i>M. sacchari</i>			<i>C. capitata</i>			<i>C. aonidium</i>		
	#	κ_S	κ_N	#	κ_S	κ_N	#	κ_S	κ_N
10	5	0.41/0.17	0.11/0.14*	6	0.09/0.08	0.09/0.14	5	0.26/0.08	0.20/0.13*
20	10	0.43/0.16	0.16/0.19*	12	0.45/0.16	0.20/0.21*	11	0.29/0.16	0.26/0.17*
30	16	0.46/0.20	0.19/0.22*	18	0.45/0.24	0.25/0.24*	17	0.32/0.20	0.24/0.21*
40	23	0.41/0.25	0.22/0.25*	26	0.36/0.29	0.27/0.27*	25	0.32/0.22	0.24/0.23*
50	31	0.44/0.27	0.23/0.27*	35	0.26/0.30	0.28/0.28*	34	0.29/0.24	0.23/0.24
60	41	0.41/0.30	0.27/0.28*	45	0.30/0.32	0.31/0.29	44	0.28/0.26	0.25/0.26*
70	53	0.41/0.32	0.31/0.30*	57	0.28/0.32	0.32/0.31	55	0.29/0.26	0.27/0.27
80	70	0.43/0.32	0.36/0.32*	74	0.34/0.33	0.34/0.31	68	0.30/0.27	0.28/0.28
90	91	0.44/0.33	0.41/0.32	87	0.40/0.33	0.41/0.32	87	0.31/0.29	0.31/0.28

Column labels are the same as in Table 8.

The similarities between the selected and random variables were also calculated for these species, and the results are presented in Table 11. There were no cases where the mean

level of similarity differed significantly from that which was expected by chance. This shows that the lack of significant differences in accuracies exhibited by *C. capitata* and *C. aonidium* were not due to biased sampling of features but to the data itself.

Table 11 – Mean and standard deviation of percentage similarities between variable sets for non-established species

%	<i>S. zeamais</i>		<i>D. melanogaster</i>		<i>S. exigua</i>	
	Expected	Actual	Expected	Actual	Expected	Actual
10	3.7	3.9/8.6	3.0	2.8/8.2	3.0	3.1/8.6
20	7.4	7.5/7.9	6.7	6.6/8.1	6.7	6.8/8.2
30	12.6	12.7/7.2	11.9	12.1/7.8	11.1	10.8/7.7
40	18.5	18.2/7.0	17.0	17.2/7.4	17.0	16.7/7.4
50	25.2	25.5/6.3	23.0	23.1/6.6	23.7	23.8/6.4
60	31.9	31.8/5.8	30.4	30.4/6.1	31.1	31.3/6.0
70	40.0	40.0/5.1	38.5	38.4/5.4	40.0	40.0/5.4
80	48.9	48.8/4.4	50.4	50.2/4.3	50.4	50.4/4.4
90	63.0	63.0/3.2	64.4	64.4/3.1	65.2	65.1/3.0

%	<i>M. sacchari</i>		<i>C. capitata</i>		<i>C. aonidium</i>	
	Expected	Actual	Expected	Actual	Expected	Actual
10	3.7	3.8/8.3	4.4	3.7/8.1	3.7	3.9/8.3
20	6.7	7.4/7.9	8.9	7.8/7.7	8.1	8.5/8.0
30	11.1	12.5/8.0	13.3	12.0/7.2	12.6	12.9/7.8
40	17.0	17.0/7.0	19.3	17.9/7.1	18.5	18.4/7.0
50	23.7	23.4/6.8	25.9	23.8/6.6	25.2	25.3/6.4
60	31.1	29.8/8.9	33.3	31.3/6.0	32.6	32.6/5.9
70	40.0	39.2/5.3	42.2	39.1/5.3	40.7	40.8/5.1
80	50.4	51.7/4.1	54.8	48.7/4.5	50.4	50.5/4.4
90	65.2	67.4/2.7	64.4	64.6/4.1	64.4	64.3/3.2

Column labels are the same as in Table 9.

4. Discussion and future work

It is clear from the results presented that MLP can learn to predict the establishment of pest species to a reasonable level of accuracy for many of the species studied here and to a high level for some.

The work presented in [Gevrey and Worner \(2006\)](#) employed back-propagation-trained MLP but only one of the species examined, *C. capitata*, was also examined here. Comparisons between the two studies is not appropriate as [Gevrey and Worner \(2006\)](#) used three-fold cross-validation to partition the data, while one-thousand random partitions were used here. However, the use of Cohen's Kappa as an accuracy measure in both studies indicate the accuracy of prediction is broadly comparable. While the training and testing kappas here are slightly lower, the validation accuracy (referred to as "Test" accuracy in [Gevrey and Worner \(2006\)](#)) is slightly higher. However, the relatively large standard deviations around the training and testing accuracies presented herein indicate that the accuracies are within the same range as that reported by [Gevrey and Worner \(2006\)](#). A comparison of the significance of the climatic variables used is also difficult. Firstly, in the [Gevrey and Worner \(2006\)](#) study, a different sub-set of climate variables were used, and secondly, the partial derivatives method of determining input significance was used. The

partial derivative method does not distinguish between positive and negative contributions (Dimopoulos et al., 1999). The variable “Annual evapotranspiration (AEannual)” is reported as being reasonably significant to *C. capitata*, which agrees with the results presented in this paper. However, while the randomisation-based verification of input significance that was carried out shows that the factors that were identified by contribution analysis are significant for most of the species studied, the results for *C. capitata*, *M. sacchari* and *C. aonidum* were less clear-cut, with the analysis indicating that some of the identified variables were not significant to the models created.

A more informative comparison with other work can be made with the results in Peacock et al. (2006). In this paper discriminant analysis was used to examine the contribution of each variable to the establishment of insect pest species. While no overall evaluation of the accuracy of prediction was made, the importance assigned to several of the variables do agree with those determined here for the species *M. persicae*, *B. brassicae*, *A. gossypii* and *C. aonidum*. The “Autumn temperature 1” variable was found to be significant for all four species, which agrees with the results here. Other variables in agreement were the “Spring rainfall 1”, “Spring rainfall 2” and “Autumn rainfall 3” for *B. brassicae*, “Penman annual evapotranspiration (PE annual)” for *A. gossypii* and “Autumn temperature 2” for *C. aonidum*.

Consideration of the influential variables from a biological perspective indicates that they certainly have biological significance. For example, rainfall during the first month of spring, and the temperature during the first month of autumn, are significant factors for the establishment of both *B. brassicae* and *M. persicae*, and the autumn temperature is significant to *A. gossypii*. These species are all aphids, and may be expected to have some environmental factors in common due to common elements of their life-cycles. During spring the aphids hatch from over-wintering eggs as their host plants undergo a spring flush. Relatively more rainfall during that time could potentially improve the development of the host plants and thus provide more bountiful food supplies for the aphids. During autumn the aphid species lay their eggs, which hibernate over winter and hatch in spring. Higher temperatures in autumn could potentially allow the aphids to survive to migrate back to their primary over-wintering hosts or allow another generation increasing the size of the over-wintering population.

A more extensive evaluation of the biological significance of the influential climate variables identified in this study is currently being carried out. While only abiotic factors have been considered here, abiotic and biotic factors are not independent. Present work is investigating the linkages between biotic factors, such as host plants, and the establishment of insect pests. These models will be combined with the models described here, and are expected to yield greater prediction accuracy.

5. Conclusion

This study has shown how MLP can be used to determine the importance of different climatic variables from many that may be influential for the establishment of several species of insect

pests. The results show that the MLP are able to learn the relationship between the climate within a geographic region and the establishment of pest species. Further sensitivity analysis of the trained MLP identified the main contributing variables with positive or negative effects on species establishment. While the most important features as expected were generally idiosyncratic to each species, there were also some similarities between species of the same taxon.

Acknowledgements

The authors wish to acknowledge the work of Muriel Gevrey and Joel Pitt who prepared the data used in this research. This study was funded by the Centre of Research Excellence, Bio-protection, at Lincoln University, New Zealand.

REFERENCES

- Andersen, M.C., 2005. Potential applications of population viability analysis to risk assessment for invasive species. *Human and Ecological Risk Assessment* 11, 1083–1095.
- Andersen, M.C., Adams, H., Hope, B., Powell, M., 2004. Risk assessment for invasive species. *Risk Analysis* 24 (4), 787–793.
- Araujo, M.B., Pearson, R.G., 2005. Equilibrium of species distributions with climate. *Ecography* 28 (5), 693–695.
- Baker, R., MacCleod, A., 2004. Pest risk assessments: tools, resources and key challenges. *Invasive Alien Species and the IPPC Framework*, pp. 106–132.
- Bobesberger, E.J., 2002. Multivariate techniques for estimating the risk of plant pest establishment in new environments. NAPPA PRA Symposium, Mexico, March 2002.
- Brosse, S., Guegan, J.-F., Tourenq, J.-N., Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling* 120, 299–311.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 3746.
- Crick, F., 1989. The recent excitement about neural networks. *Nature* 337, 129–132.
- Crop Protection Compendium—Global Module, 5th Edition. c CAB International, Wallingford, UK 2003.
- Dentener, P.R., Whiting, D.C., Connolly, P.G., 2002. *Thrips palmi* karny (*Thysanoptera: Thripidae*): could it survive in New Zealand? *New Zealand Plant Protection* 55, 18–24.
- 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). *Ecological Modelling* 120, 157–165.
- Eyre, M.D., Rushton, S.P., Luff, M.L., Telfer, M.G., 2005. Investigating the relationships between the distribution of British ground beetle species (Coleoptera, Carabidae) and temperature, precipitation and altitude. *Journal of Biogeography* 32, 973–983.
- Flexer, A., 1996. Statistical evaluation of neural network experiments: minimum requirements and current practice. In: Trapp, R. (Ed.), *Cybernetics and Systems '96, Proceedings of the 13th European Meeting on Cybernetics and Systems Research*. Austrian Society for Cybernetic Studies, pp. 1005–1008.
- Garson, G.D., 1991. Interpreting neural-network connection weights. *AI Expert* 6 (7), 47–51.
- Gevrey, M., Wornor, S.P., 2006. Prediction of global distribution of insect pest species in relation to climate using an ecological

- informatics method. *Environmental Entomology* 99 (3), 979–986.
- Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 160, 249–264.
- Gotelli, N.J., 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81 (9), 2606–2621.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147–186.
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157, 89–100.
- Haykin, S., 1994. *Neural Networks: A Comprehensive Foundation*. MacMillan Publishing Company.
- Ibarra, A.A., Gevrey, M., Park, Y.-S., Lim, P., Lek, S., 2003. Modelling the factors that influence fish guilds composition using a back-propagation network: assessment of metrics for indices of biotic integrity. *Ecological Modelling* 160, 281–290.
- James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review Ecological Systems* 21, 129–166.
- Jeong, K.-S., Kim, D.-K., Joo, G.-J., 2006. River phytoplankton prediction model by artificial neural network: model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system. *Ecological Informatics* 1, 235–245.
- Joy, M.K., Death, R.G., 2004. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology* 49, 1036–1052.
- Kasabov, N.K., 1996. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. MIT Press.
- Kimes, D.S., Nelson, R.F., Fifer, S.T., 2000. Predicting ecologically important vegetation variables from remotely sensed optical/radar data using neural networks. In: Lek, S., Guegan, J.-F. (Eds.), *Artificial Neuronal Networks: Application to Ecology and Evolution*.
- Laë, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural networks. *Ecological Modelling* 120, 325–335.
- Lehmann, A., Overton, J., Leathwick, J.J., 2002. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 157, 189–207.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling* 90, 39–52.
- Mael, S., Dias, J.-M., Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distribution: a case study with a Himalayan river bird. *Ecological Modelling* 120, 337–347.
- Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology* 38, 237–246.
- Milne, L.K., 1995. Feature selection with neural networks with contribution measures. *Proceedings of the Australian Conference on Artificial Intelligence AI'95*, Canberra.
- Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling* 154, 135–150.
- 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. *Ecological Modelling* 178, 389–397.
- Park, Y.-S., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecological Modelling* 160, 265–280.
- Peacock, L., Worner, S., Sedcole, R., 2006. Climate variables and their role in site discrimination of invasive insect species distributions. *Environmental Entomology* 35 (4), 958–963.
- Peterson, A.T., Vieglas, D.A., 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *Bioscience* 51 (5), 363–371.
- Prechelt, L., 1996. A quantitative study of experimental evaluations of neural network learning algorithms: current research practice. *Neural Networks* 9 (3), 457–462.
- Rafoss, T., 2003. Spatial stochastic simulation offers potential as a quantitative method for pest risk analysis. *Risk Analysis* 23 (4), 651–661.
- Recknagel, F., Cao, H., Kim, B., Takamura, N., Welk, A., 2006. Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecological Informatics* 1, 133–151.
- Reed, R.D., Marks, R.J., 1999. *Neural smithing*. MIT Press.
- Rossi, R.E., Borth, P.W., Tollefson, J.J., 1993. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. *Ecological Applications* 3 (4), 719–735.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Soltic, S., Pang, S., Peacock, L., Worner, S., 2004a. Evolving computation offers potential for estimation of pest establishment. *International Journal of Computers, Systems and Signals* 5 (2), 36–43.
- Soltic, S., Pang, S., Kasabov, N., Worner, S., Peacock, L., 2004b. Dynamic neuro-fuzzy inference and statistical models for risk analysis of pest insect establishment. *Lecture Notes in Computer Science* 3316, 971–976 (January).
- Stynes, B., 2002. Pest risk analysis: methods and approaches. *NAPPO PRA Symposium*, Puerto Vallarta, Mexico. March 2002.
- Sutherst, R.W., Maywald, G.F., Skarratt, B.D., 1995. Predicting insect distributions in a changed climate. In: Harrington, R., Stoke, N.E. (Eds.), *Insects in a changing environment*. Academic, London.
- Sutherst, R.W., 2003. Prediction of species geographical range. *Journal of Biogeography* 30, 805–816.
- Tarassenko, L., 1998. *A Guide to Neural Computing Applications*. Wiley, London.
- Wagner, R., Dapper, T., Schmidt, H.-H., 2000. The influence of environmental variables on the abundance of aquatic insects: a comparison of ordination and artificial neural networks. *Hydrobiologia* 422/423, 143–152.
- Wagner, R., Obach, M., Werner, H., Schmidt, H.-H., 2006. Artificial neural nets and abundance prediction of aquatic insects in small streams. *Ecological Informatics* 1, 423–430.
- Watts, M.J., Worner, S.P., 2006. Comparison of artificial neural networks models for predicting insect pest establishment. *International Journal of Information Technology* 12 (6), 35–42.
- Worner, S.P., 1988. Ecoclimatic assessment of potential establishment of exotic pests. *Journal of Economic Entomology* 81, 973–983.
- Worner, S.P., 1994. Predicting the establishment of exotic pests in relation to climate. In: Sharp, J.L., Hallman, G.J. (Eds.), *Quarantine Treatment for Pests of Food Plants*. Boulder Press, Westview, CO, pp. 11–31.
- Worner, S.P., Gevrey, M., 2006. Modelling global insect pest species assemblages to determine risk of invasion. *Journal of Applied Ecology* 43, 858–867.
- Zaniewski, A.E., Lehmann, A., Overton, J., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157, 261–280.