NEURAL NETWORKING INCREASES THE EFFECTIVENESS OF NEAR INFRA-RED PREDICTIVE MODELS FOR THE SELECTION OF *ELDANA SACCHARINA* (LEPIDOPTERA: PYRALIDAE) RESISTANT SUGARCANE PHENOTYPES

RS RUTHERFORD, NA COETZEE AND MG KEEPING

South African Sugar Association Experiment Station, P/Bag X02, Mount Edgecombe, 4300

Abstract

Linear predictive models based on data acquired by near infra-red (NIR) spectrophotometry suggest that sugarcane nodal budscale extract components contribute towards resistance to the stalk borer *Eldana saccharina* Walker (Lepidoptera: Pyralidae). For this NIR application the possibility that nonlinear neural networked models would increase the effectiveness of NIR as a predictive tool was investigated.

Keywords: near infra-red, *Eldana saccharina*, sugarcane, host plant resistance, artificial neural network, PLS.

Introduction

Newly hatched *Eldana saccharina* Walker (Lepidoptera: Pyralidae) larvae spend about one week on the stalk surface before survivors attempt to bore into the stalk. The nodal bud is a favoured site of entry into the sugarcane stalk for *E. saccharina* and entry frequently follows some browsing of the protective budscale (Leslie, 1993). Observed differences in sugarcane resistance might therefore be partly explained by biochemical influences of the stalk surface on larval behaviour and survival. Progress towards the use of near infra-red (NIR) spectrophotometry as a rapid means of biochemically profiling internode surface wax and budscale extracts from resistant and susceptible sugarcane clones has been reported by Rutherford and van Staden (1996) and Rutherford (1997). Previously, standard linear NIR calibration methods were used. For this paper the possible contribution of the nodal bud is further investigated using the potential of neural networks to improve the predictive ability of NIR models.

The basis of artificial neural networks (ANNs)

Artificial neural networks (ANNs) can be used for a wide variety of learning tasks. They should be considered a mathematical tool, similar to linear regression analysis. According to Haykin (1994), an ANN is a "... massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles a brain in two respects: (i) knowledge is acquired by the network through a learning process; and (ii) interneuron connection strengths known as synaptic weights are used to store the knowledge." The key advantage of ANNs over regression analysis is that they use nonlinear mathematics and therefore can be used to model highly complex and nonlinear functions.

ANNs are typically organised in layers. Neurons in a layer are connected to neurons in adjacent layers. Patterns are presented to the network via the input layer, which communicates to one or more hidden layers where the actual processing is done. Each hidden layer of neurons has a transfer function, usually non-linear, assigned to it. The hidden layers then link to an output layer, where the answer is output as shown in Figure 1.

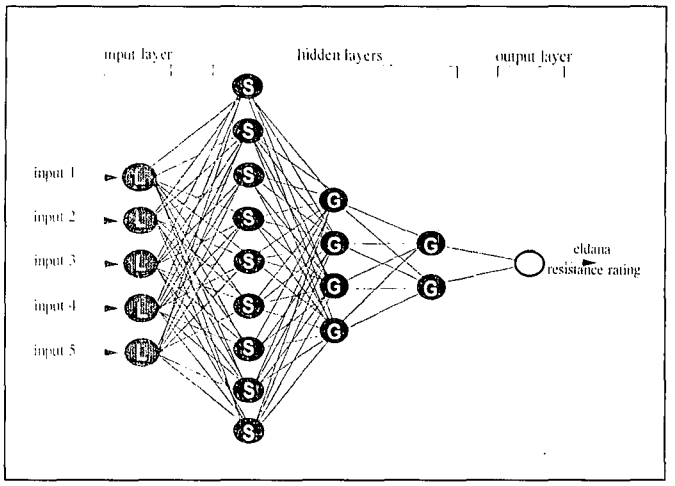

Figure 1. Architecture of an artifical neural network (L = linear transfer function, S = sigmoidal transfer function, G = gaussian transfer function).

Most ANNs contain some form of learning rule which modifies the weights of the connections according to the input patterns presented. In a sense, ANNs learn by example, as do their biological counterparts. The most common class of ANNs are called backpropagational neural networks (BPNNs). Back-propagation is an abbreviation for the backwards propagation of error. Errors are backpropagated through the network, and training algorithms adjust the synaptic connection weights in an attempt to drive the network's response error to a minimum. More simply, when a neural network is initially presented with a training pattern it makes a random 'guess' as to what the relationships with the output answer might be. Synaptic weights are assigned to each connection and the network then calculates

how far its answer was from the actual answer. Appropriate adjustments to its synaptic weights are made as the network cycles through all of the input patterns until training is complete.

BPNNs are the ultimate 'black boxes'. Apart from defining the general architecture of a network the user has no other role than to feed it input, watch it train and await the output. The final product of this activity is a trained network that provides no equations or coefficients defining a relationship (as in regression) beyond its own internal mathematics. The trained network is the final equation of the relationship.

NIR prediction accuracy (using linear models) generally decreases as a population of samples broadens (Westerhaus and Reeves, 1992). The purpose of this study was to compare calibrations produced by the standard linear methods with those obtained by neural networking for both narrow and broad calibration populations.

Materials and Methods

Sampling and NIR

Calibration sets A (60 clones), B, C and D (50 clones each) and one set of 30 clones of unknown resistance ratings (*E. saccharina* resistance field trial ERF1-93) were sampled separately, as indicated in Figure 2. Samples from ERF1-93 were collected from a trial site adjacent to calibration set B.

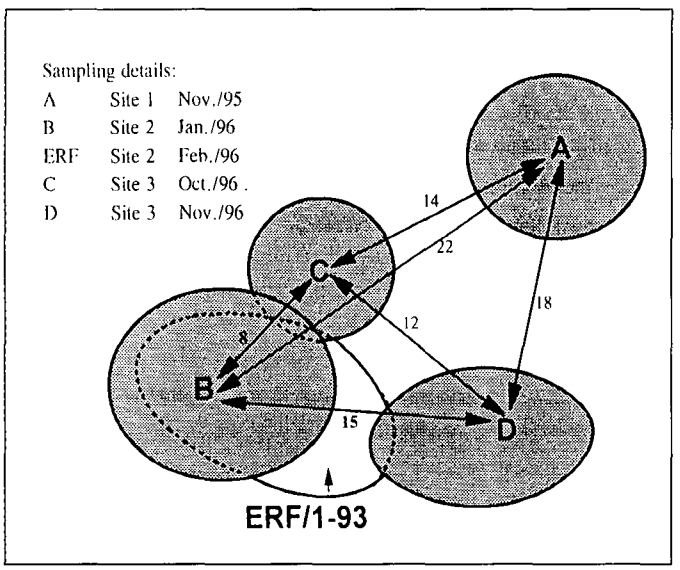

Figure 2. H (Mahanalobis) distances within and between five sets of budscale extract samples (math treatment: 1st derivative, segment 4 nm, gap 4 nm).

Budscales were extracted as described by Rutherford (1997). An NIRSystems 5000 scanning monochromator (NIRSystems, Silver Spring, MD) was used to acquire transmission spectra (0,1 cm light path) with near infra-red radiation from 1 100 to 2 500 nm for 1 mL budscale extracts. One spectral outlier was detected in set A using the -CENTER procedure of Infrasoft International (ISI) NIRS 3 software; i.e. its H (Mahanalobis) distance was greater than three from the mean spectrum of the set indicating that, for some reason, this spectrum did not 'belong' to the sample set. This sample was replaced. H distances between sets were also determined, using the -SELECT procedure of ISI NIRS 3 software.

Determination of resistance ratings

For calibration sets, resistance ratings were obtained by the Plant Breeding and Entomology Departments of the South African Sugar Association Experiment Station, as described by Nuss and Atkinson (1983). Ratings are expressed as low 1, 2 and 3 (resistant); intermediate 4, 5 and 6; and high 7, 8 and 9 (susceptible), based on comparison with several standard clones. Resistance ratings for ERF1-93 were determined in plant and first ratoons under field conditions (Keeping, 1997).

Narrow population calibration procedure

Set A was chosen as representative of a narrow population. In a similar approach, Borggaard and Rasmussen (1992) selected a one third subset of their samples to form a test set. Their calibration procedure was followed here:

- Set A was sorted in ascending order of resistance ratings from 1 to 9.
- Every nth spectrum (totalling 20) was selected for the test set where n was some integer from 1-5. The remaining 40 spectra were assembled into the calibration (training) set.
- A Linear PLS model was constructed on the calibration set. A total of six factors were calculated using the -CALIBRATE procedure of ISI NIRS 3 software. Stepwise multiple linear regression models using up to six individual wavelength data were also constructed.
- The test set was projected onto the PLS loadings (eigenvectors) found for the calibration set in step 3) to generate test set PLS scores for the first six factors.
- Linear PLS and stepwise multiple linear regression calibration models from step 3) were applied to the test set PLS factor scores and individual wavelength absorbance values respectively. A comparison was made between predicted and known ratings for the test set.
- Vesta Services (Winnetka, IL) Qnet™ ANN software package was utilised for training ANNs. PLS factor scores for the calibration set served as inputs to the network for training. Network design parameters (i.e. the number of network layers, input neurons and hidden neurons; and types of transfer function) were optimised by trial and error (Figure 1) with reference to the test set in order to avoid over-training. The above design parameters were varied until training set correlation coefficient maximisation coincided with test set correlation coefficient maximisation; i.e. no over-fitting occurred. A comparison was made between predicted and known ratings for the test set.

- The same network design was used to model individual wavelength data using the wavelengths selected by the stepwise multiple linear regression method (i.e. no further network optimisation was attempted). Again a comparison was made between predicted and known ratings for the test set.

Broad population calibration procedure

Spectra from sets B(45), C(1) and D(5), being at the most three H units from any spectrum in ERF1-93 (Figure 2), were assembled into a broad calibration population along with nine spectra from control clones present within ERF1-93 (total 60 spectra). Predictive models were again constructed using stepwise multiple linear regression of individual wavelength data, linear PLS, ANN on individual wavelength data, and ANN on PLS factor scores. These models were used to predict the resistance ratings of clones in trial ERF1-93. A comparison was made between predicted ratings and trial results.

Results and Discussion

Narrow population calibration

The resultant linear regression calibrations and trained ANNs were used to predict the ratings of the 20 test set spectra excluded from the calibration. A comparison was made between predicted and known ratings for the test set (Table 1).

Several regression statistics were used to optimise the number of wavelengths or PLS factors to be used in a calibration. Internal to the calibration the regression F ratio can be used as a test of significance. In the case of PLS linear regression, the mean squared error of cross validation (MSECV), where it begins to plateau at a minimum, indicates the optimal number of factors to use as far as internal validation is concerned. The test set coefficient of determination (r^2), regression F ratio and slope, obtained by regression of observed versus predicted ratings, will ideally be maximised; while the standard error of prediction (SEP) and regression intercept will ideally be minimised.

Table 1
Comparison of stepwise multiple linear regression on individual wavelength (l) data and linear PLS regression calibrations with non-linear ANN calibrations and their predictive ability in a narrow sample population.

MLR-λ, 40 spectra of set A calibration				MLR-λ, excluded 20 spectra of set A predicted				
No. λ (wavelengths)	r^2 (observed vs predicted)	F ratio		SEP	r^2 (predicted vs observed)	F ratio	Intercept	Slope
2	0,350	9,962		1,401*	0.444	**5,989**	1,110	0,711
3	0,507	**12,341**		1,634	0,444	3,727	0,742	**0,829**
4	0,562	11,227		**1,399**	**0,518**	3,493	**0,680**	0,824
5	0,588	9,705		1,599	0,437	1,863	0,812	0,780
6	0,645	9,993		1,542	0,441	1,446	0,857	0,777
ANN-λ , 40 spectra of set A calibration				**ANN-λ , excluded 20 spectra of set A predicted**				
No. λ	r^2	F ratio		SEP	r^2	F ratio	Intercept	Slope
2	0,523	**20,284**		1,473	0,400	5,000	0,960	0,686
3	0,524	13,210		1,292	0,577	**6,366**	0,133	0,860
4	0,536	10,108		**1,120**	**0,639**	5,753	0,198	0,850
5	0,545	8,145		1,365	0,540	2,817	**0,118**	0,844
6	0,604	8,389		1,342	0,556	2,296	0,153	0,857
Linear-PLS, 40 spectra of set A calibration				**Linear-PLS, excluded 20 spectra of set A predicted**				
No. factors	r^2	F ratio	MSECV	SEP	r^2	F ratio	Intercept	Slope
2	0,590	**26,622**	2,816	1,336	0,342	**3,898**	2,062	0,547
3	0,629	20,345	2,118	1,412	0,379	2,848	1,599	0,626
4	0,688	19,295	2,025	**1,364**	**0,461**	2,780	1,163	0,717
5	0,724	17,838	**1,924**	1,386	0,449	1,956	1,176	0,710
6	0,735	15,255	2,082	1,514	0,421	1,333	**1,041**	**0,732**
ANN-PLS, 40 spectra of set A calibration				**ANN-PLS, excluded 20 spectra of set A predicted**				
No. factors	r^2	F ratio		SEP	r^2	F ratio	Intercept	Slope
2	0,256	6,366		0,761	0,294	3,123	3,361	0,280
3	0,456	10,059		1,282	0,322	2,216	2,234	0,504
4	0,588	12,488		1,353	0,433	2,482	0,947	0,674
5	0,715	**17,060**		1,441	**0,569**	**3,168**	**-0,128**	**0,945**
6	0,740	15,654		1,423	0,518	1,970	0,503	0,843

*SEP = standard error of prediction

Using these criteria to optimise each calibration shows that three or four wavelength models might be selected in the case of MLR-λ and ANN-λ models, whereas four or five factor models might be selected in the case of linear-PLS and ANN-PLS. Although the calibrations continue to improve in terms of r^2 using six wavelengths or factors, data from the test set show that models are then over-fitted, in that predictive ability begins to decline.

ANN models showed greater predictive ability than did their linear counterparts. Using calculations made by Schenk and Westerhaus (1993), the magnitude of the r^2 improvement of ANN-λ over MLR-λ, shown in Table 1, would result in an increase from 60% of predicted clones classified in the correct resistance group (resistant, intermediate or susceptible) to 65% (Table 2).

Table 2
Relationship between r^2 and the probability (p) that a sample will be correctly classified out of three groups (e.g. resistant, intermediate, susceptible).

r^2	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
p that a sample will be predicted in the correct group of three	0,33	0,43	0,47	0,51	0,55	0,59	0,63	0,68	0,74	0,81
p that a sample will be predicted in the correct or adjacent group of three	0,66	0,86	0,89	0,92	0,94	0,96	0,97	0,99	1	1

Adapted from Schenk and Westerhaus (1993)

Broad population calibration

The H distances between different calibration sets are large enough such that any calibration based on one set cannot be expected to accurately predict the ratings of any other set (Figure 2). This means that calibrations capable of predicting unknowns from any field situation may require a significant development time in order to include sample sets from a number of growing seasons and environmental conditions in a broad population (Dardenne, 1996).

However, the presence of overlaps between ERF1-93 and sets B, C and D suggest that a broad calibration population constructed of spectra from these sets might be of value in predicting resistance in ERF1-93. The predictive ability in respect of ERF1-93 of calibrations made on a broad population selected from sets B, C and D (Figure 2) were compared with calibrations made on the narrow population represented by set A.

Not unexpectedly, both MLR-λ and linear-PLS calibrations on set A had virtually no predictive ability for ERF1-93 (see Figure 2; H distances). Linear-PLS was found to be better suited to the broad population calibration than MLR-λ, which again had virtually no predictive ability (Table 3). It is in the use of broad populations for calibration that prediction accuracy appears to be limited by linear regression techniques (Westerhaus and Reeves, 1992).

Over-training of the ANN was possible in the broad calibration population using ERF1-93 as the test set. This began when approximately 90% of the training iterations which would be required for maximum calibration r^2 were reached. Training was therefore stopped at the test set r^2 maximum. According to Fearn (1997), optimisation of a network architecture using one test set (in our case a 20 spectra subset of set A) can lead to adaptation to that test set. It is possible that some other network architecture would have been more optimal for use with ERF1-93.

It is notable that an ANN model on six PLS factors gives better results than a linear PLS model yields on seven factors, which were optimal for this method. However, the best ANN model was found by using seven PLS factors as input to the network (Table 3).

Table 3
Comparison of narrow and broad calibrations (60 spectra sets) and their predictive ability on ERF1-93 (30 spectra).

	Narrow (Set A) 60 spectra		Broad (composite B, C and D +9 ERF1-93 controls) 60 spectra	
Calibration method	Calibration r^2 (observed vs inputs)	Prediction* (ERF1-93) r^2 (predicted vs observed)	Calibration r^2 (observed vs inputs)	Prediction* (ERF1-93) r^2 (predicted vs observed)
7 MLR	0,572	–0,03	0,189	0,04
7 factors linear-PLS (optimal)	0,703	0,09	0,693	0,259
6 factors ANN-PLS			0,708	0,313
7 factors ANN-PLS (optimal)			0,731	0,396

*Predictions were compared with the mean of plant and first ratoon *E. saccharina* resistance trial results (ERF1-93).

The magnitude of the r^2 improvement of ANN-PLS over linear-PLS, shown in Table 3, would result in an increase from around 49% of predicted clones classified in the correct resistance group (resistant, intermediate or susceptible) to around 55% based on calculations from Schenk and Westerhaus (1993) (Table 2).

Nine of the 30 spectra from ERF1-93 fell in the area of no-overlap with sets B, C and D, each of which supplied calibration spectra (Figure 2). It can be expected that predictions for these spectra would not be as accurate as for those falling in the areas of overlap. Had a more suitable calibration set been available, a model with greater predictive ability towards ERF1-93 could have been produced.

ERF1-93 trial clones might now be considered to be knowns in respect of the actual resistance ratings derived from field exposure to *E. saccharina*. As such they begin to fill the space between sets B, C and D. Calibration sets will continue to be collected, from a number of growing seasons and environmental

conditions, until all areas of no-overlap have been covered (Figure 2). In this event it should be possible to extract a calibration set from a library of spectra, which will be capable of predicting the resistance ratings of new plant breeding material from any growing season or environmental condition.

ANNs provide an analytical alternative to conventional techniques which are often limited by strict assumptions of normality, linearity, variable independence, etc. One reason for this is that the multi-hidden layer construction used in most networks creates an increased cross-factoring of information and relationships. It can be appreciated that an ANN can capture many kinds of relationships and it allows the user to quickly and relatively easily model phenomena which may be difficult or impossible to explain otherwise.

A neural network calibration procedure is time consuming compared with standard linear methods. The reason for this is that the optimal numbers of network layers, input and hidden layer neurons, and optimal types of transfer function have to be found. Automated methods for evolving an optimal network are in development (Haykin, 1994). To properly evaluate the contribution of ANNs in the NIR application described here requires measures of the cost saving in correctly classifying (in three resistance groups) an additional 5-6% of clones, compared with the standard linear methods, and the costs incurred in developing this approach.

Recommendations

To optimise neural network-NIR predictive models for application to the plant breeding programme the following guidelines are proposed: (1) The technique should be applied to the single line stage (approximately 2 000 clones/annum); (2) continued sampling of calibration sets from different growing seasons and environmental conditions is necessary so that a global library of spectra can be achieved; (3) a set of nearest neighbours should be drawn from the library, on the basis of least H distance from spectra to be predicted, for use in any calibration; (4) around 30 control clones of known resistance should be dispersed throughout single line fields to contribute towards a test set, in order to optimise predictive models and networks (as has been described in this paper), and for addition to the calibration set.

Acknowledgments

Special thanks are due to the Plant Breeding and Entomology departments of the South African Sugar Association Experiment Station for the provision of stalk material and the determination of *E. saccharina* resistance ratings.

REFERENCES

Borggaard, C and Rasmussen, AJ (1992). Interpreting non-linear NIR-spectra using neural networks and PCA/PLS. *Proc 5th int NIRS Conference.* Oslo, Norway: 73-78.

Dardenne, P (1996). Stability of NIR spectroscopy equations. *NIR News.* 7(5): 8-9.

Fearn, T (1997). Validation. *NIR news.* 8(1): 7-8.

Haykin, S (1994). A Comprehensive Foundation. In: *Neural Networks.* p 2. Macmillan. New York.

Keeping, MG (1997). Field screening of South African sugarcane for resistance to the stem borer, *Eldana saccharina* Walker (Lepidoptera: Pyralidae). *Proc Joint Congr Entomol Soc Sthrn Afr and Afr Assoc Insect Scientists.* Stellenbosch : (in press).

Leslie, GW (1993). Dispersal behaviour of neonate larvae of the pyralid sugarcane borer *Eldana saccharina. Proc S Afr Sug Technol Ass* 67: 122-126.

Nuss, KJ and Atkinson, PRA (1983). Methods used to measure the susceptibility of sugarcane varieties to attack by *Eldana saccharina* Walker. *Proc S Afr Sug Technol Ass* 57: 92-94.

Rutherford, RS (1997). Prediction of resistance in sugarcane to the stalk borer *Eldana saccharina* Walker (Lepidoptera: Pyralidae) by near infrared spectroscopy on crude budscale extracts. The involvement of flavonoids. *J Chem Ecol* 23 : (in press).

Rutherford, RS and van Staden, J (1996). Towards a rapid near infrared technique for the prediction of resistance to the sugarcane borer *Eldana saccharina* Walker (Lepidoptera: Pyralidae) using stalk surface wax. *J Chem Ecol* 22: 681-694.

Schenk, JS and Westerhaus, MO (1993). *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy.* Monograph, Infrasoft International, Port Matilda, PA 16870.

Westerhaus, MO and Reeves, JB (1992). NIR calibrations of agricultural products with neural networks. *Proc 5th int NIRS Conference.* Oslo, Norway: 79-84.