

Weather-based prediction of anthracnose severity using artificial neural network models

S. Chakraborty^{a*†}, R. Ghosh^b, M. Ghosh^b, C. D. Fernandes^c, M. J. Charchar^d and S. Kelemu^e

^aCommonwealth Scientific and Industrial Research Organization (CSIRO) Plant Industry, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia 4067, Queensland; ^bSchool of Information Technology and Mathematical Sciences (ITMS), University of Ballarat, PO Box 663, Victoria 3353, Australia; ^cEMBRAPA-CNPQC/Doctor student UNESP-FCA, Botucatu-SP, 18609-490 Brazil; ^dEMBRAPA-CPAC, 73301-970, Planaltina, DF, Brazil; and ^eCentro Internacional de Agricultura Tropical, Cali, Colombia

Data were collected and analysed from seven field sites in Australia, Brazil and Colombia on weather conditions and the severity of anthracnose disease of the tropical pasture legume *Stylosanthes scabra* caused by *Colletotrichum gloeosporioides*. Disease severity and weather data were analysed using artificial neural network (ANN) models developed using data from some or all field sites in Australia and/or South America to predict severity at other sites. Three series of models were developed using different weather summaries. Of these, ANN models with weather for the day of disease assessment and the previous 24 h period had the highest prediction success, and models trained on data from all sites within one continent correctly predicted disease severity in the other continent on more than 75% of days; the overall prediction error was 21.9% for the Australian and 22.1% for the South American model. Of the six cross-continent ANN models trained on pooled data for five sites from two continents to predict severity for the remaining sixth site, the model developed without data from Planaltina in Brazil was the most accurate, with >85% prediction success, and the model without Carimagua in Colombia was the least accurate, with only 54% success. In common with multiple regression models, moisture-related variables such as rain, leaf surface wetness and variables that influence moisture availability such as radiation and wind on the day of disease severity assessment or the day before assessment were the most important weather variables in all ANN models. A set of weights from the ANN models was used to calculate the overall risk of anthracnose for the various sites. Sites with high and low anthracnose risk are present in both continents, and weather conditions at centres of diversity in Brazil and Colombia do not appear to be more conducive than conditions in Australia to serious anthracnose development.

Keywords: anthracnose, *Colletotrichum gloeosporioides*, disease risk and severity, multiple linear regression analysis, *Stylosanthes scabra*

Introduction

Disease prediction is based on weather conditions under which a pathogen, when in contact with a susceptible host, can infect and become established (Hardwick, 1998). Weather-based forecasting systems reduce the cost of production by optimizing the timing and frequency of application of control measures and ensure operator, consumer and environmental safety by reducing chemical usage. A

major aim of many forecasting systems is to reduce fungicide use (Taylor *et al.*, 2003), and accurate prediction is important to synchronize the use of disease control measures to avoid crop loss. Forecasting systems with varying levels of sophistication and success are available for a large number of diseases affecting horticultural and field crops (Cambell & Madden, 1990).

In the case of anthracnose disease of the tropical pasture legume *Stylosanthes* spp., many promising varieties have been decimated by virulent strains of the pathogen *Colletotrichum gloeosporioides* in all parts of the world where this legume is grown for commercial use. In its centre of diversity in south and Central America, the extensive genetic variation in both host and pathogen populations (Chakraborty *et al.*, 2002; Weeds *et al.*, 2003) has meant

*To whom correspondence should be addressed.

†E-mail: Sukumar.Chakraborty@csiro.au

Accepted 12 March 2004

that many previously resistant varieties have been severely affected by new pathogen races and, so far, no commercial variety has escaped serious anthracnose damage (Miles & Grof, 1997). In Northern Australia over 500 000 ha of *S. humilis* pastures were destroyed by this pathogen in the 1970s. Although this highly susceptible species has since been replaced with more resistant species, the anthracnose pathogen continues to cause losses in both dry matter and seed yields (Davis *et al.*, 1987b). Nevertheless, *Stylosanthes* has been more successful as a forage legume in countries such as Australia and China, away from centres of host-pathogen diversity, due largely to a less-diverse pathogen population (Chakraborty *et al.*, 2002; Weeds *et al.*, 2003). Whether weather and other ecosystem- or farming system-related factors at centres of diversity might also be more conducive to serious anthracnose development has not been examined in detail. For instance, antagonistic bacteria that suppress *C. gloeosporioides* on leaf surfaces in the humid tropics of Peru do not have any effect on the pathogen in the savanna ecosystems (Lenné & Brown, 1991).

Colletotrichum gloeosporioides conidia are produced in a mucilaginous matrix (Louis & Cooke, 1985) that requires surface wetness for splash dispersal and subsequent infection (Fitt & McCartney, 1985). Under controlled environments anthracnose development is favoured by a leaf wetness period of 12 h or longer, and maximum severity is reached after 36 h wetness (Chakraborty *et al.*, 1990). Disease severity is unaffected by brief interruptions of 2–4 h wetness if RH is maintained over 85%. In control environments, severe anthracnose develops at 20–30°C if 24 h leaf wetness is maintained (Irwin *et al.*, 1984). Attempts to establish a quantitative relationship between weather and anthracnose infection from field studies have not always been successful. Davis *et al.* (1987a) did not find any obvious relationship between disease severity and weather, and infection could not be associated with any specific sequence of weather events. Using multiple linear regression (REG) analysis, Chakraborty & Billard (1995) showed that mean daily temperature and RH above a certain threshold was necessary for successful infection. However, the REG model was not equally effective in predicting infection events in the 2 years of this field study in Australia. Subsequently, another independent study has shown that the REG model does not adequately explain infection at a field site in Queensland, Australia (Pangga, 2002). These studies indicate that, although the REG model was able to explain anthracnose development at a particular field site for certain years, its ability to generalize across field sites and years was poor.

Other analytical approaches have been used to improve understanding of the role of weather in plant diseases, including anthracnose disease of *Stylosanthes*. These include the use of stochastic models (Shaw, 1994; Chakraborty & Smyth, 1995); fuzzy numbers (Scherf, 2000); and artificial neural networks (ANN), among others. In plant protection ANN have been used to model leaf wetness (Francl & Panigrahi, 1997), disease dynamics (Yang *et al.*, 1995), disease forecasting (De Wolf & Francl, 1997; Crisci *et al.*, 1998) and pesticides in soil (Yang *et al.*, 1997).

As part of an international collaboration, disease development and weather were monitored at selected field sites in Australia, Brazil and Colombia to improve knowledge of anthracnose epidemiology for effective disease management. Using data from Australian and/or South American field sites, the aims of this paper are to determine the usefulness of ANN models to predict anthracnose development based on prevailing weather conditions at other sites within or outside a continent, and to calculate the overall risk of anthracnose at field sites using a set of weights from the trained ANN models.

Materials and methods

Data on weather and anthracnose severity were obtained from field sites in Australia, Brazil and Colombia. The Australian data came from the CSIRO Pasture Research Station at Samford (27°22' S, 152°53' E) near Brisbane, and the Queensland Department of Primary Industries field stations at Southedge (17°0' S, 145°20' E) and Springmount (17°13' S, 145°18' E) in north Queensland. Samford is at the southern edge of the climatic zone for *Stylosanthes*, and Southedge and Springmount are typical of *Stylosanthes*-growing regions in Australia. The annual average rainfall for Springmount is 804 mm, Samford 1050 mm, and Southedge 1112 mm. The South American data were obtained from two sites each in Brazil and Colombia. The Centro de Pesquisa Agropecuária dos Cerrados research station of the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) at Planaltina (15°36' S, 47°42' W), near Brasília and the Centro Nacional de Pesquisa de Gado de Corte of EMBRAPA in Campo Grande (20°26' S, 54°42' W), Mato Grosso do Sul, were the Brazilian sites. Planaltina has a long history of *Stylosanthes* research and Campo Grande is at the heart of beef cattle country in Brazil. The two Colombian sites were located at the Centro Internacional de Agricultura Tropical (CIAT) field experiment stations in Carimagua (4°30' N, 71°19' W) and Caquetá (1°15' N, 75°41' W). Carimagua is a well established research station in the Colombian savanna with a long history of *Stylosanthes* research and development, and Caquetá is a relatively new site located within the Colombian Amazon basin. The annual average rainfall for Campo Grande is 1526 mm, Planaltina 1540 mm, Carimagua 2337 mm and Caquetá 3552 mm.

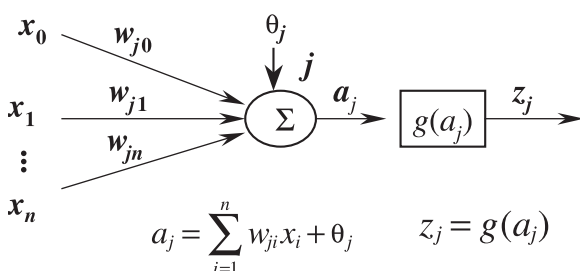
Three to five replicate plots, 5 × 5 m each, of the susceptible *S. scabra* cv. Fitzroy were established at each site by sowing scarified seeds in rows 1–2 m apart. Environmental variables were logged continuously at each site using an automatic weather station (Monitor Sensors, Queensland, Australia). Air temperature and RH sensors were housed in a meteorological screen, within 15–20 m of the field plots. Sensors for the duration of leaf wetness, canopy temperature and RH were mounted 20 cm above ground level in the canopy to avoid direct sunlight. Data on radiation, temperature, RH and wind speed were recorded hourly. Rainfall and leaf wetness data were recorded every 6 min when these events occurred. Ten weather attributes were considered in this work: maximum (RMX) and minimum (RMN) air RH; maximum (TMX) and minimum

(TMN) air temperature; sunshine hours (SUN); total radiation (RAD); rainfall (RAIN); hours of leaf wetness (LWP); wind daily total (WND) and gust (GST).

During each summer season anthracnose severity as a result of infection from naturally distributed inoculum was assessed on a randomly selected branch on five to 10 plants within a plot at approximately weekly intervals. Plants were visually examined and rated for disease severity using a 10-point scale: 0 = disease-free, 1 = 1–3% leaf area diseased, 2 = 4–6%, 3 = 7–12%, 4 = 13–25%, 5 = 26–50%, 6 = 51–75%, 7 = 76–87%, 8 = 88–94% and 9 = 95–100% (Chakraborty, 1990). Using the proportion of leaf area diseased, rating scores were either collapsed into six severity classes: disease-free; very low, <0.2; low, 0.2–0.4; medium, 0.4–0.6; high, 0.6–0.8; very high, >0.8, or transformed according to $\ln(\text{severity} + 1)$ before analysis. Weather and disease severity data for a total of 162 days used in this work comprised: 40 days from Samford (1987–92); 22 from Southedge (1989–97); 15 from Springmount (1989–91); 34 from Campo Grande (1995–98); 28 from Planaltina (1995–2000); 12 from Caquetá (1994–96); 11 from Carimagua (1994–96).

In the susceptible cv. Fitzroy the interval from inoculation to the first sign of symptom appearance is 3.9 days (Chakraborty *et al.*, 1988), and another 2–3 days elapse before at least 10% of typical sporulating lesions are clearly visible. To account for the lag in symptom expression, three series of models were developed to examine the influence of weather on disease severity. Weather data during the week prior to disease assessment were averaged in series 1 models; weather for the seventh day before disease assessment was used in series 2 models; and weather data on the day of disease assessment and the previous 24 h period were used in series 3 models.

Artificial neural networks are parallel computing systems made up of a large number of simple, highly interconnected processing elements called nodes or neurons that process information by their dynamic-state response to the external signals and can handle imprecise information. Important features of ANN are a set of processing units; an activation state for each unit equivalent to its output; connections between the units, generally defined by a weight w_{jk} that determines the effect that unit j has on unit k ; a propagation rule that determines the effective input of the unit; an activation function, a scalar-to-scalar function that transforms inputs to a unit; an external input (bias, offset), similar to a parameter estimate for each unit; and a method for information gathering, the so-called learning rule. A processing unit can be simply expressed as follows:



where x_0, x_1, \dots, x_n are inputs; $w_{j0}, w_{j1}, \dots, w_{jn}$ are weights; θ_j is the bias term; j is the unit; $g(a_j)$ is the activation function; and z_j is the output.

There are three types of unit (neuron): input units receive data from outside the network; output units send data out of the network; and hidden units, whose input and output remain within the network, perform internal computations to provide the nonlinearity that makes ANN powerful. Each unit can have one or more inputs but only one output. An input to a unit is either the data from outside the network, or the output of another unit, or its own output. Hidden and output units combine all values feeding into the unit using a combination function such as a linear combination function, which is then transformed by an activation function such as a sigmoid function. Output is compared with the target value to compute the value of the error function. A variety of optimization techniques exist that can tune the weights in order to find the minimum of the error function. More on the fundamentals of ANN and computational implementations can be found in e.g. Bishop (1995); Hertz *et al.* (1995).

Of the many variations of ANN, in this study a fully connected feed-forward network, also known as a multi-layered perceptron, has been used, and for simplicity a two-layer network structure (hidden and output layer) has been considered. One front node (8× Ultra Sparc III processors 750 MHz, 8 GB memory) and eight back nodes (each containing 8× Ultra Sparc III processors 900 MHz, 8 GB memory) in a cluster formation on an HPC supercomputer at Griffith University, Gold Coast, Queensland, Australia was used as the processing unit. Codes for training, testing and sensitivity analysis were written in MATLAB (The MathWorks Inc., Natick, MA, USA) and are provided in Appendix A.

As previously published weather-based models for anthracnose (Chakraborty & Billard, 1995) are based on multiple linear regression analysis, and many plant pathologists have used this approach to develop disease-forecasting models, regression models were developed in this study using each of the three weather summaries to provide reference points for the outputs of the ANN models. For REG, $\ln(\text{severity} + 1)$ -transformed severity data were analysed using the SPSS software (SPSS Inc., Chicago, IL, USA). As before (Chakraborty & Billard, 1995), a stepwise procedure was used for model selection where the condition for retaining a variable was set at $P < 0.05$. The stepwise method is a modification of the forward-selection technique where variables are added one by one to the model and the F statistic for the variable to be retained must be significant at the specified P level for entry. After a variable is added, the method checks all variables in the model and deletes any variable that does not produce an F statistic significant at the specified P level for retention. The process ends when none of the variables outside the model has an F statistic significant at the specified entry level.

For each of the three series, ANN and REG models were developed using data for all three Australian sites as the training set to predict the daily severity class for each

of the four South American sites. Similarly, data from the South American sites were used to predict daily severity at each of the Australian sites. These are referred to as the 'continent-wise models'. Second, data for all sites were pooled together and ANN and REG models were developed using data for six Australian and South American sites to predict the severity class for the remaining seventh site. This way each site was dropped in turn and the daily severity class of the dropped site was predicted using models developed from the six remaining sites. These are the 'cross-continent models'. The number of days when the models correctly predicted the actual severity class was used to measure classification success. Two types of error were used to compare the performance of the models: type I, where the model predicted a lower than actual severity class on a given day; and type II, where the prediction was higher than the actual severity class.

For the series of models with the lowest prediction errors, the relative importance of the various weather measures to the ANN models was determined using a sensitivity analysis where sensitivity was calculated from the set of weight vectors from the trained ANN.

To estimate the overall anthracnose risk factor for each site, the four most important weather attributes from the seven cross-continent ANN models were identified and their corresponding percentage sensitivities calculated. The class mean for each of the five severity classes was calculated for the four most important weather attributes as follows:

$$class_mean(j) = \frac{\sum_{i=1}^N value_i}{N}, \forall i \in class\ j \quad (2)$$

After finding the class means for each of the four significant attributes, the mean-square deviation (MSD) for all sites from every class was calculated using the following formula:

$$MSD(k,j) = \frac{\sum_{i=1}^N (value_i - class_mean(j))^2}{N}, \quad (3)$$

$$\forall i \in class\ j\ and\ k \in attribute$$

The combined weighted deviations from every class mean were calculated as follows:

$$WD(j) = \frac{\sum_{i=1}^4 (MSD(i,j) * Sensitivity(i))}{100} \quad (4)$$

and the reciprocal of weighted deviations was expressed as the overall anthracnose risk factor.

Results

Selection of ANN models

A rapid back-propagation learning algorithm (Battiti, 1992) was used in all ANN models and architectures with different numbers of hidden neurons explored to select the

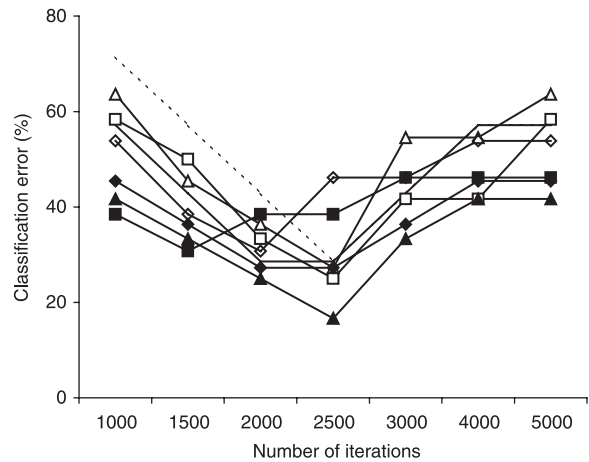


Figure 1 Changes in classification error in validation data sets with an increasing number of iterations for artificial neural network models: models were trained on data from Australian and South American sites to predict disease severity class at South American (dotted line) and Australian (solid line) sites, respectively; other models for Samford (Δ), Springmount (◇), Southedge (□), Campo Grande (◆), Planaltina (▲) and Carimagua (■) were trained on data for all sites except the site being used for prediction.

best architecture for the ANN models. The number of iterations was controlled to prevent the network from becoming overtrained. Overtraining was identified by an early stopping technique, determined as the point where the classification error in the validation data set started to increase with increasing iterations. The relationship between the number of iterations and the classification error in validation data sets is given in Fig. 1.

Four to five hidden neurons and a sigmoid activation function for both layers with 1500–2500 iterations offered the best architecture. When data from all Australian field sites were used in training, the ANN model was 60% accurate in predicting severity classes for the South American sites, and the percentage of error was 100% when predicting severity for the Caquetá site. Similarly, the model using data from all South American sites predicted severity classes for the Australian sites poorly (data not shown). *Stylosanthes* has only recently been introduced to Caquetá and no *C. gloeosporioides* inoculum or anthracnose symptoms were recorded at this site during the 1994–96 period. Therefore data from Caquetá were excluded, and all further analysis was conducted only on six Australian and South American sites. This increased the prediction success of the Australian model to >77% and that of the South American model to >76%. The classification errors for the various ANN models with three to 10 hidden neurons developed with or without data from the Caquetá site are given in Fig. 2.

Prediction errors of the three series of models

Of the three series of models developed to examine the influence of weather on disease severity, prediction errors

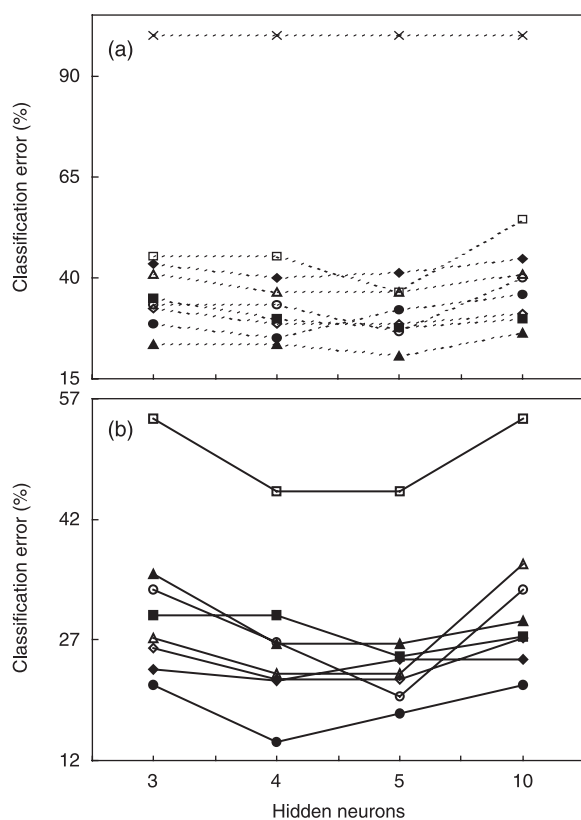


Figure 2 Classification error of artificial neural network models with an increasing number of hidden neurons: models were trained on data from Australian and South American sites to predict disease severity class at South American (\diamond) and Australian (\blacklozenge) sites, respectively; other models for Samford (\blacksquare), Southedge (\blacktriangle), Springmount (\circ), Campo Grande (\blacktriangle), Planaltina (\bullet), Carimagua (\square) and Caquetá (\times) were trained on data for all sites except the site being used for prediction. Broken and solid lines represent models developed with (a) and without (b) data from the Caquetá site, respectively.

were lowest for series 3 models, which used weather data on the day of disease assessment and the previous 24 h period; both continent-wise and cross-continent ANN models showed this consistent trend (Fig. 3). Prediction errors were generally similar for series 1 and 2 models that used either average weather during the week prior to disease assessment (26.6–81.8% for ANN; 29.4–81.8% for REG models) or weather for the seventh day before assessment (26.6–72.7% for ANN; 36–72.7% for REG models). Due to some missing weather data, the number of days with disease and weather data varied slightly for the three series of models. However, these differences were relatively small and may not have greatly influenced prediction errors (Fig. 3). All further work therefore, concentrated on series 3 models.

Prediction and sensitivity of ANN models

The two continent-wise ANN models, each trained on data from three field sites, correctly predicted disease severity class at field sites in the other continent on >77% of days; the overall prediction error was 21.9% for the

Australian and 22.1% for the South American model (Table 1). The prediction of the Australian ANN model was most accurate for the Carimagua site and least accurate for Planaltina. The South American model predicted severity at Southedge and Springmount with an accuracy >93%, but the prediction accuracy for the Samford site was <63%.

In the final series of ANN models, data from all Australian and South American sites were pooled and models were trained on data from five sites to predict the severity class for the remaining sixth site. Of the six cross-continent ANN models developed in this way, the model developed without the data from Planaltina was the most accurate, successfully predicting severity on >85% of days. The ANN model without the Carimagua data correctly predicted severity on only 54% of days at this site. The other four models were accurate on >73% of days (Table 2). The prediction errors of these models were considerably higher than for the continent-wise ANN models (Table 1) for all sites except Samford and Planaltina. The percentage of type I, II or total error was not generally influenced by the number of observations (days) in the training or test data sets in either continent-wise or cross-continent ANN models. The overall prediction error was between 14 and 45% for the cross-continent models developed with 110–139 observations (days) in the training data sets; this was between 4 and 37% for continent-wise models with either 73 or 77 observations in the training data set.

A sensitivity analysis of the continent-wise models showed that RAIN, LWP and RAD for the day of disease severity assessment and RAIN on the previous day are the four most important weather attributes in the Australian ANN model. RAIN and RAD on the day of severity assessment and RAIN and LWP on the previous day were also the four most significant input attributes in the South American ANN model, although RAIN and RAD were more important and LWP was less important in the South American than in the Australian model (Table 3). Similarly, RAIN, RAD and LWP on the day of severity assessment and/or the previous day were the most important weather attributes in each of the six cross-continent models, and RAIN during the day and/or the previous day was the single most important of all attributes in all models except the one trained on data that excluded Planaltina, where another moisture-related variable, LWP on the previous day, was more important than RAIN (Table 3).

Prediction and sensitivity of REG models

Overall, the REG models developed using data from all Australian or South American sites were not as effective as ANN models in predicting disease severity for sites in the other continent, and the prediction errors of the REG models were higher for two out of three sites in each continent (Table 1). The overall prediction error of 31.5% for the Australian REG model was higher than the ANN model, but the 20.8% error for the South American model was lower than the ANN model. When data for five Australian and South American sites were pooled to predict

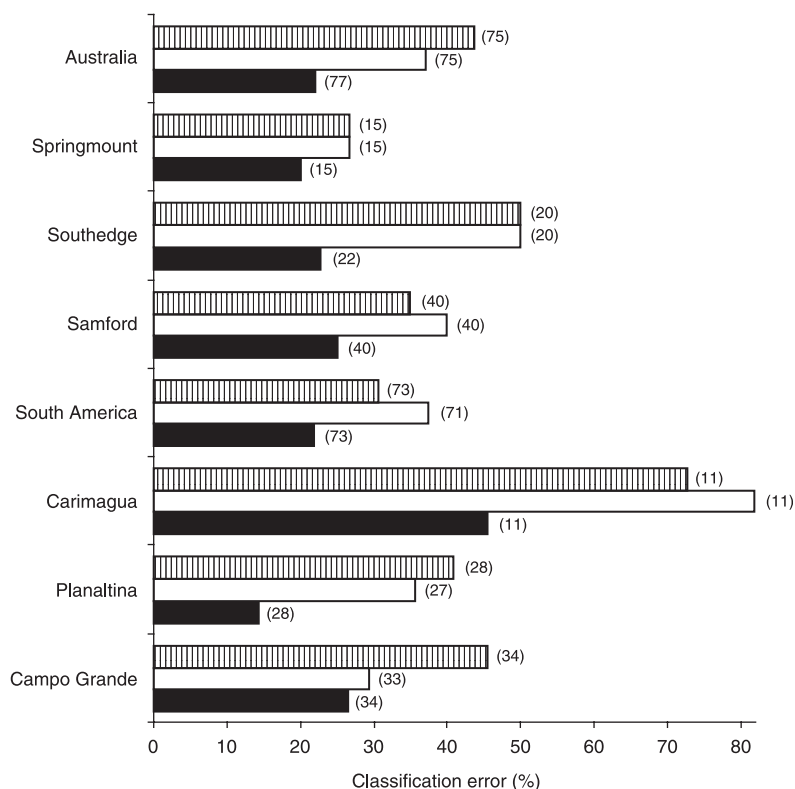


Figure 3 Classification error of artificial neural network models developed using weather data for the day of disease assessment and the previous 24 h period (black bars), average weather for 1 week prior to disease assessment (open bars) and weather on the seventh day before disease assessment (shaded bars). Models were trained on data from Australian and South American sites to predict disease severity class at South American and Australian sites, respectively; other models for Samford, Springmount, Southedge, Campo Grande, Planaltina and Carimagua were trained on data for all sites except the site being used for prediction. Number of days for which data are available given in parentheses.

Table 1 Classification error of artificial neural network (ANN) and multiple regression (REG) models developed using data from Australian or South American field sites as the training set to predict anthracnose severity class at sites in another country

Training data	Site	Type I error (%) ^a		Type II error (%) ^b		Total error (%)	
		ANN	REG	ANN	REG	ANN	REG
Australian sites (77) ^c	Campo Grande (34)	14.7 (5)	0	0	35.3 (12)	14.7 (5)	35.3 (12)
	Planaltina (28)	35.7 (10)	3.6 (1)	0	7.1 (2)	35.7 (10)	10.7 (3)
	Carimagua (11)	0	0	9.1 (1)	72.7 (8)	9.1 (1)	72.7 (8)
South American sites (73)	South America (73)	20.5 (15)	1.4 (1)	1.4 (1)	30.1 (22)	21.9 (16)	31.5 (23)
	Samford (40)	12.5 (5)	2.5 (1)	25 (10)	17.5 (7)	37.5 (15)	20.0 (8)
	Southedge (22)	4.5 (1)	9.1 (2)	0	9.1 (2)	4.5 (1)	18.2 (4)
	Springmount (15)	0	0	6.6 (1)	26.7 (4)	6.6 (1)	26.7 (4)
	Australia (77)	7.7 (6)	3.9 (3)	14.3 (11)	16.9 (13)	22.1 (17)	20.8 (16)

^aDays when predicted severity class was lower than actual severity class.

^bDays when predicted severity class was higher than actual severity class.

^cNumber of days for which data are available given in parentheses.

the severity of the sixth site, REG models always had higher error than the ANN models and the total prediction error was 1.3–2.1 times higher than ANN predictions (Table 2). As with ANN models, the prediction error of REG models was lower (10–72%) for the continent-wise models than for cross-continent models (26–100%); however, in contrast to ANN models, in most REG models the percentage of total error generally declined with increasing number of observations in the training and/or test data sets.

On a day-to-day basis the REG models tended to predict an average severity value, and did not follow

fluctuations in disease severity levels as closely as the ANN models. This was true for both continent-wise and cross-continent models. Data are shown for a continent-wise and two cross-continent ANN and REG models to illustrate this (Fig. 4). In general, the predicted severity from ANN models tracked the actual severity more closely than the predicted severity from the REG models at all sites.

Overall, all REG models had low to very low R^2 : 0.08 for the continent-wise model trained on Australian data, and 0.04 for the model trained on South American data. The R^2 of cross-continent models was between 0.16 and

Table 2 Classification error of artificial neural network (ANN) and multiple regression (REG) models developed using pooled data from six Australian and South American field sites as the training set to predict anthracnose severity class for the seventh site

Training data	Testing data	Type I error (%) ^a		Type II error (%) ^b		Total error (%)	
		ANN	REG	ANN	REG	ANN	REG
All except:							
Samford (110) ^c	Samford (40)	5 (2)	42.5 (17)	20 (8)	2.5 (1)	25 (10)	45 (18)
Southedge (128)	Southedge (22)	18.2 (4)	31.8 (7)	4.5 (1)	9.1 (2)	22.7 (5)	40.9 (9)
Springmount (135)	Springmount (15)	0	0	20 (3)	26.7 (4)	20 (3)	26.7 (4)
Campo Grande (116)	Campo Grande (34)	2.9 (1)	0	23.5 (8)	47.1 (16)	26.5 (9)	47.1 (16)
Planaltina (122)	Planaltina (28)	14.3 (4)	0	0	28.5 (8)	14.3 (4)	28.5 (8)
Carimagua (139)	Carimagua (11)	0	0	45.5 (5)	100 (11)	45.5 (5)	100 (11)

^aDays when predicted severity class was lower than actual severity class.

^bDays when predicted severity class was higher than actual severity class.

^cNumber of days for which data are available given in parentheses.

Table 3 Sensitivity of weather attributes^a on the day of disease severity assessment and for the previous 24 h period in artificial neural network models, where continent-wise models were trained on data from one continent to predict severity in another, and cross-continent models were trained on data from five Australian and South American sites to predict severity at the sixth site

Attributes	Sensitivity (%)							
	Continent-wise model		Cross-continent model					
	Australia	S. America	Samford	Southedge	Springmount	Campo Grande	Planaltina	Carimagua
Weather on day of disease severity assessment								
RMX	0.8	1.4	1.0	0.9	0.5	1.8	1.4	1.1
RMN	1.5	2.6	1.2	1.7	1.3	1.6	1.8	0.9
TMX	1.5	2.7	2.4	0.6	0.5	0.8	2.7	0.8
TMN	1.8	0.8	1.0	0.5	0.9	1.1	0.7	1.9
SUN	0.5	2.2	2.1	1.7	0.4	1.9	1.9	3.8
RAD	6.4^b	10.9	4.8	10.5	6.7	6.1	4.9	4.2
RAIN	23.3	35.8	35.4	52.8	40.8	23.1	11.6	26.5
LWP	26.6	1.3	3.1	1.6	8.3	11.6	4.5	10.8
WND	4.8	2.3	2.4	1.3	2.5	4.9	7.4	7.3
GST	2.0	2.3	3.4	1.3	1.1	4.5	2.7	3.1
Weather for previous 24 h								
RMX	1.4	2.4	1.0	0.6	0.8	2.0	1.2	0.9
RMN	1.8	1.9	1.6	0.8	0.6	1.4	0.8	1.9
TMX	0.9	1.4	1.0	2.7	0.3	1.3	1.3	1.5
TMN	0.7	2.2	0.4	0.4	0.7	1.8	1.3	0.3
SUN	0.9	2.1	0.8	0.3	0.7	2.9	1.8	2.5
RAD	4.7	2.8	4.2	4.8	2.4	11.5	6.9	9.0
RAIN	12.5	16.1	9.9	7.5	22.9	8.1	17.7	12.4
LWP	3.3	6.8	22.0	8.0	6.9	8.7	23.0	5.9
WND	3.4	1.1	1.7	1.6	1.3	2.7	4.2	4.1
GST	1.3	1.0	0.7	0.3	0.5	2.3	2.3	1.3

^aWeather attributes: maximum (RMX) and minimum (RMN) air RH; maximum (TMX) and minimum (TMN) air temperature; sunshine hours (SUN); total radiation (RAD); rainfall (RAIN); hours of leaf wetness (LWP); wind daily total (WND) and gust (GST).

^bSensitivity >5% in bold type.

0.18 (data not shown). All weather variables with >5% sensitivity in the ANN models were selected as significant variables in the REG models. The continent-wise models had fewer significant predictors (1–2) than the cross-continent models, with six to 10 weather variables in each (data not shown). As with ANN models, RAIN on the day of severity assessment was the single most important weather variable, and it was a significant term in all REG models except the continent-wise model trained on South American data. In addition to RAIN, RMX, WND, TMN

and LWP on the day of severity assessment, and WND on the previous day, were significant weather variables in all six cross-continent models.

Anthracnose risk

For each site, risk factors for each of the five severity classes and an overall risk were calculated. The Planaltina site showed the highest risk level for four out of the five severity classes, and Campo Grande and Samford were

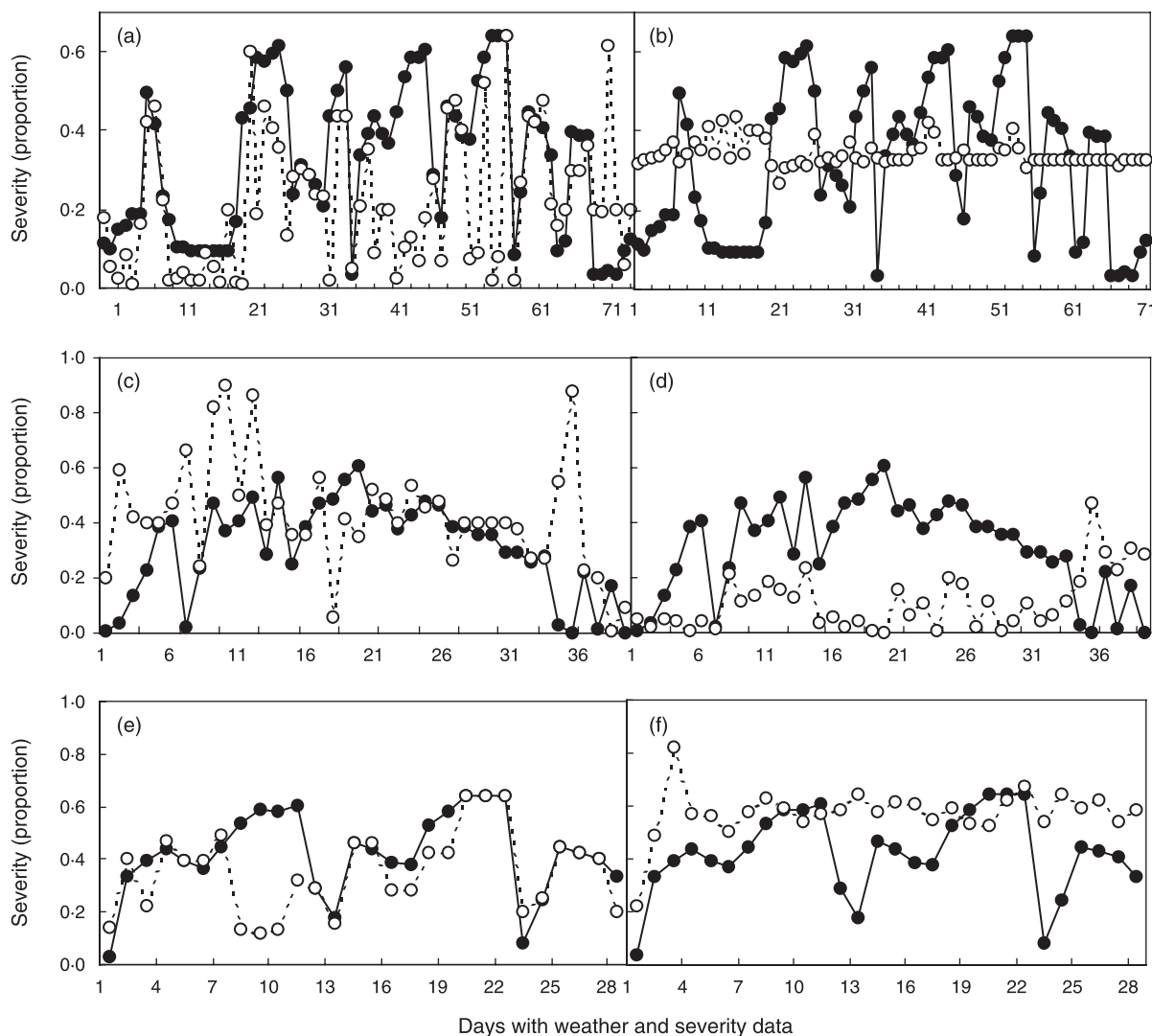


Figure 4 Actual (●) and predicted (○) mean disease severity at South American field sites predicted using continent-wise artificial neural network (ANN) model trained on Australian data (a); continent-wise multiple linear regression (REG) model trained on Australian data (b); cross-continent ANN (c) and REG (d) models trained on pooled data for five Australian and South American sites to predict disease severity for the Samford (Australia) site; and cross-continent ANN (e) and REG (f) models trained on pooled data for five South American sites to predict disease severity for the Planaltina (South America) site.

sites with the least risk for any severity class (data not shown). Overall, the highest possible anthracnose risk was at Planaltina (1), followed by Southedge (0.72), Campo Grande (0.50), Springmount (0.46), Carimagua (0.38), Caquetá (0.09) and Samford (0). Although no anthracnose was recorded at Caquetá, weather at this site is suitable for anthracnose development to be regarded as a low to moderate risk whereas, despite some disease development, weather at the Samford site is considered marginal for any serious risk of anthracnose development.

Discussion

This work is among a handful of examples applying ANN tools to developing weather-based prediction models for

plant diseases. While previous research has largely used data from the same site (De Wolf & Francl, 1997; Francl & Panigrahi, 1997; Chtioui *et al.*, 1999), the work reported here demonstrates that data from some or all field sites in Australia and/or South America can be used in ANN models to predict anthracnose severity at other sites. The series 3 ANN models trained on data from all sites within one continent correctly predicted disease severity in the other continent on >77% of days, and the prediction success of cross-continent ANN models trained on the pooled data for five sites from both continents to predict severity for the sixth site ranged from 54 to >85%. All ANN models consistently selected moisture-related variables such as RAIN, LWP and variables that influence moisture availability such as RAD and WND on the day

of disease severity assessment, or the day before assessment, as the most important weather variables influencing severity. This is one of a growing number of reports demonstrating the usefulness of ANN models for weather and plant disease prediction (De Wolf & Francl, 1997; Francl & Panigrahi, 1997; Chtioui *et al.*, 1999) and other related analyses (Hajmeer & Basheer, 2003).

In a previous work, an ANN model trained on data from a field site accurately predicted 81–87% of wheat tan spot (caused by *Pyrenophora tritici-repentis*) infection events for the same site (De Wolf & Francl, 1997). In the study reported here, ANN models were used to predict anthracnose severity for sites that were never used in model training/development. Training and testing sites were geographically distant and were often located in another country or continent. Despite this rigorous and stringent validation, accuracy of the various series 3 ANN model predictions was between 54 and 96%. Because of the stringent validation, ANN models developed from the multisite data are robust and useful over a broad range of field sites. This broad application also demonstrates the obvious underlying principle that the basic quantitative relationship between anthracnose and weather does not change from site to site within or outside a continent, given that an inoculum source of a virulent pathogen and a susceptible host are present. Further prediction accuracy in the models may be gained by considering other factors that influence anthracnose development, such as farming system, including naturally occurring biological control agents (Lenné & Brown, 1991) and spatial heterogeneity. Using data only from sites with similar characteristics may be useful. For instance, if data from the same site are used for both training and testing, the prediction accuracy of series 3 ANN models for Springmount and Carimagua increases to 100%. However, these models essentially become site-specific and lose application over a broad range of sites.

The fact that the series 3 models, developed using weather data for the day of disease assessment and the previous 24 h period, gave the highest prediction accuracy is an important finding. In a previous field study spanning 3 years, when healthy plants exposed to natural inoculum for a short period were maintained in either ambient or high RH (>95%), 42% more plants developed anthracnose in the high-RH environment than plants at ambient RH (Chakraborty & Billard, 1995). The number of lesions on each plant also increased following incubation at high RH. Pangga *et al.* (2004) have recently demonstrated a similar change in anthracnose severity when inoculated plants were transferred to different CO₂ environments. These studies highlight the importance of weather during both initial infection and disease development periods, and predictive models that include both components are likely to be more successful. Of the three series of models, by averaging the weather data the series 1 models have effectively removed important variation in the weather data; the series 2 models have considered weather only for infection events for a single day; while the series 3 models have considered weather for two consecutive days. Given the length of the incubation period of *C. gloeosporioides*

on cv. Fitzroy (Chakraborty *et al.*, 1988), the high prediction success of series 3 models probably represents long-term correlations within weather and may not explain a direct causal relationship between weather and anthracnose severity. Further studies are necessary to explore this.

The importance of moisture-related variables to anthracnose infection has previously been established from experiments in controlled environments (Irwin *et al.*, 1984; Chakraborty *et al.*, 1990). Other field-based research has similarly quantitatively linked moisture-related weather variables to detailed daily infection of *Stylosanthes* by the anthracnose pathogen. Of these, Chakraborty & Billard (1995) and Pangga (2002) used multiple linear regressions, and Chakraborty & Smyth (1995) used a stochastic logistic regression. Although none of these regression models was validated using data from an independent source, rainfall, relative humidity and temperature were important predictors in most models. In addition, radiation, evaporation and high temperature were significant terms in many models due to their influence on the extent and duration of free water on plant surfaces (Chakraborty & Smyth, 1995). In the current work, both ANN and REG models largely selected weather variables that were directly related to moisture availability and the duration of wetness, or variables such as radiation and wind that influence moisture and wetness.

One important contribution of this study has been the calculation of an overall risk for each site using weights from the ANN models. With the Australian Southedge site being only second to the Planaltina site for anthracnose risk, centres of diversity and utilization cannot be separated according to anthracnose risk, and sites with high and low anthracnose risk are present in both continents. Among the South American sites, risk of serious anthracnose was highest for Planaltina and lowest for Caquetá. With an annual average rainfall of 3553 mm, Caquetá receives twice as much rain as Planaltina (1540 mm), but the pathogen has not been detected at this site in the 2 years of this study. In contrast, despite some disease development, weather at Samford is considered marginal for anthracnose development to pose any serious risk. These clearly highlight the importance of proximity to an inoculum source. In Brazil and other countries there is spatial heterogeneity in the geographical distribution of both pathogenic races (Chakraborty *et al.*, 2002) and molecular haplotypes of *C. gloeosporioides* (Weeds *et al.*, 2003), and some races of this pathogen have significant advantages in terms of fitness components such as infection efficiency and fecundity (Chakraborty *et al.*, 1988; Chakraborty & Datta, 2003). Being a splash-dispersed pathogen, proximity to an inoculum source is important, and in field experiments d_{50} , the distance at which the number of spore dispersal decreases by half, of *C. gloeosporioides* spores is generally <10 m from a point source (Pangga, 2002). Interaction between *C. gloeosporioides* and antagonistic microorganisms in different ecosystems may be another important element of this risk (Lenné & Brown, 1991).

With standard back-propagation ANN models prediction is often more important than explanation, and model

construction is not easily understood (Frasconi *et al.*, 1993); this has created the perception of a 'black box' that has hindered the widespread use of these models. Predictions limited to local minima during gradient descent and optimization of the number of hidden layers, units, learning coefficient, momentum, etc. are among other weaknesses of ANN models (Frasconi *et al.*, 1993). In this study transparency has been increased and the explanatory power of the ANN models has been improved by using sensitivity analysis, where the contributions of input weather variables can be easily seen. The number of hidden neurons required for most of the models was 4–5, which is often greater than the number of important weather variables in the models. This suggests that the relationships are nonlinear, and hence could not be fitted well by the REG models, but hidden neurons in the ANN models intrinsically capture the nonlinearity. Given the widespread application to plant disease forecasting of multiple linear regression models, including for anthracnose (Chakraborty & Billard, 1995; Chakraborty & Smyth, 1995; Pangga, 2002), the REG models used in this work have provided reference points for the outputs of the ANN models. No attempt has been made to develop the best possible regression model, and certainly other variable selection procedures and techniques incorporating nonlinearity, including the use of multivariate adaptive regression spline (MARS model, Friedman, 1991) and stochasticity, can be used to improve the REG models. The application of ANN models for plant disease prediction shown in this and other work (De Wolf & Francl, 1997) makes them a useful tool for future forecasting models, and combining aspects of ANN and well established statistical tools (Specht, 1991; Chtioui *et al.*, 1999) may offer a more flexible option for the future.

Acknowledgements

We thank Ross Perrott, Arthur Kenji Akimoto and Roberto Monoel Gonçalves, Antonio Carlos Gomez, Jorge Badel, Raylee Trevorrow, Murray Anderson and Richard Ross for technical assistance. Financial assistance from the Australian Centre for International Agricultural Research, Co-operative Research Centre for Tropical Plant Pathology and the CSIRO is gratefully acknowledged.

References

- Battiti R, 1992. First and second order methods for learning: between steepest descent and Newton's method. *Neural Computation* **4**, 141–66.
- Bishop CM, 1995. *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press.
- Cambell CL, Madden LV, 1990. *Introduction to Plant Disease Epidemiology*. New York, USA: John Wiley & Sons.
- Chakraborty S, 1990. Expression of quantitative resistance to *Colletotrichum gloeosporioides* in *Stylosanthes scabra* at different inoculum concentrations and day–night temperatures. *Australian Journal of Agricultural Research* **41**, 89–100.
- Chakraborty S, Billard L, 1995. Quantitative relationships between *Colletotrichum gloeosporioides* infection of *Stylosanthes scabra* and weather factors under field conditions. *Plant Pathology* **44**, 63–72.
- Chakraborty S, Datta S, 2003. How will plant pathogens adapt to host plant resistance at elevated CO₂ under a changing climate? *New Phytologist* **159**, 733–42.
- Chakraborty S, Smyth GKS, 1995. A stochastic model incorporating the effect of weather conditions on anthracnose development in *Stylosanthes scabra*. *Journal of Phytopathology* **143**, 495–9.
- Chakraborty S, Cameron DF, Irwin JAG, Edye LA, 1988. Quantitatively expressed resistance to anthracnose (*Colletotrichum gloeosporioides*) in *Stylosanthes scabra*. *Plant Pathology* **37**, 529–37.
- Chakraborty S, McKay FJ, Ratcliff D, 1990. Anthracnose of *Stylosanthes scabra*: effect of leaf surface wetness on disease severity. *Plant Disease* **74**, 379–84.
- Chakraborty S, Fernandes CD, Charchar M, Jd'A, Thomas M, 2002. Pathogenic variation in *Colletotrichum gloeosporioides* infecting *Stylosanthes* spp. at its centre of diversity in Brazil. *Phytopathology* **92**, 553–62.
- Chtioui Y, Panigrahi S, Francl L, 1999. A generalized linear regression neural network and its application for leaf wetness prediction to forecast plant disease. *Chemometrics and Intelligent Laboratory Systems* **48**, 47–58.
- Crisi A, Moriondo M, Orlandini S, 1998. Analysis of downy and powdery mildew infection: modelling and neural network approach. In: *Proceedings of the 7th ICCTA – International Congress for Computer Technology in Agriculture*, 1998. Florence, Italy.
- Davis RD, Irwin JAG, Cameron DF, Shepherd RK, 1987a. Epidemiological studies on the anthracnose diseases of *Stylosanthes* spp. caused by *Colletotrichum gloeosporioides* in north Queensland and pathogenic specialization within the natural fungal populations. *Australian Journal of Agricultural Research* **38**, 1019–32.
- Davis RD, Irwin JAG, Shepherd RK, Cameron DF, 1987b. Yield losses caused by *Colletotrichum gloeosporioides* in three species of *Stylosanthes*. *Australian Journal of Agricultural Research* **27**, 67–72.
- De Wolf ED, Francl LJ, 1997. Neural networks that distinguish infection periods of wheat tan spot in an outdoor environment. *Phytopathology* **87**, 83–7.
- Fitt BDL, McCartney HA, 1985. Spore dispersal in splash droplets. In: Ayres PG, Boddy L, eds. *Water, Fungi and Plants*. Cambridge, UK: Cambridge University Press, 87–104.
- Francl LJ, Panigrahi S, 1997. Artificial neural network models of wheat leaf wetness. *Agricultural and Forest Meteorology* **88**, 57–65.
- Frasconi P, Gori M, Tesi A, 1993. Successes and failures of back-propagation: a theoretical investigation. In: Omidvar O, ed. *Progress in Neural Networks*. Toronto, Canada: Ablex Publishing.
- Friedman JH, 1991. Multivariate adaptive regression spline (with discussion). *Annals of Statistics* **19**, 1–141.
- Hajmeer M, Basheer I, 2003. Comparison of logistic regression and neural network-based classifiers for bacterial growth. *Food Microbiology* **20**, 53–5.
- Hardwick NV, 1998. Disease forecasting. In: Jones DG, ed. *The Epidemiology of Plant Diseases*. Dordrecht, the Netherlands: Kluwer Academic, 207–30.

- Hertz J, Krogh A, Palmer RG, 1995. *Introduction to the Theory of Neural Computation*. Reading, UK: Addison-Wesley.
- Irwin JAG, Cameron DF, Ratcliff D, 1984. Influence of environmental factors on the development of anthracnose diseases of *Stylosanthes* spp. *Australian Journal of Agricultural Research* 35, 473–8.
- Lenné JM, Brown AE, 1991. Factors influencing the germination of pathogenic isolates of *Colletotrichum gloeosporioides* on leaf surfaces of *Stylosanthes guianensis*. *Mycological Research* 95, 227–32.
- Louis I, Cooke RC, 1985. Conidial matrix and spore germination in some plant pathogens. *Transactions of the British Mycological Society* 84, 661–7.
- Miles JW, Grof B, 1997. Recent advances in studies of anthracnose of *Stylosanthes*. III. *Stylosanthes* breeding approaches in South America. *Tropical Grasslands* 31, 430–4.
- Pangga IB, 2002. Influence of climatic factors on plant architecture of *Stylosanthes scabra* and the epidemiology of the anthracnose disease. Brisbane, Australia: University of Queensland, PhD Thesis.
- Pangga IB, Chakraborty S, Room PM, Yates D, 2004. Resistance and canopy size in *Stylosanthes scabra* determine anthracnose severity at high CO₂. *Phytopathology* 94, 221–7.
- Scherm H, 2000. Simulating uncertainty in climate change models with fuzzy numbers. *Environmental Pollution* 108, 373–9.
- Shaw MW, 1994. Modelling stochastic processes in plant pathology. *Annual Review of Phytopathology* 32, 523–44.
- Specht DF, 1991. A general regression neural network. *IEEE Transactions on Neural Networks* 2, 568–72.
- Taylor MC, Hardwick NV, Bradshaw NJ, Hall AM, 2003. Relative performance of five forecasting schemes for potato late blight (*Phytophthora infestans*) I. Accuracy of infection warnings and reduction of unnecessary, theoretical, fungicide applications. *Crop Protection* 22, 275–83.
- Weeds PL, Chakraborty S, Fernandes CD, Charchar M, Jd'A, Ramesh CR, Kexian Y, Kelemu S, 2003. Genetic diversity in *Colletotrichum gloeosporioides* from *Stylosanthes* spp. at centres of origin and utilization. *Phytopathology* 93, 176–85.
- Yang XB, Batchelor WD, Tshanz AT, 1995. A neural network model to predict soybean rust. *Phytopathology* 85, 1172.
- Yang CC, Prasher SO, Sreekanth S, Patni NK, Masse L, 1997. An artificial neural network model for simulating pesticide concentrations in soil. *Transactions of the ASAE* 40, 1285–94.

Appendix A

```

Training code load 'inputpat.txt';
load 'outputpat.txt';
input = inputpat(:); %change this one for number of
training pattern output = outputpat(:);
p = input';
q = output';
outputsz = size(output,2);
hn = 4;

```

```

fid = fopen('config.txt','w');
fprintf(fid,'%d percentaged percentaged',size(input,2),
hn,size(output,2));
fclose(fid);
net = newff([0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1],
[hn,outputsz],{'logsig','logsig','trainrp'});
net = init(net);
net.trainParam.epochs = 2500; %total iteration number
net.trainParam.show = 100; %donot change this one
net.trainParam.lr = 0.2; %learning rate net.trainParam.mc
= 0.8; %momentum
net.trainParam.goal = 0.0000001; % error goal net =
train(net,p,q);
% matrix contains the output of the trained network;
w1 = net.IW{1,1};
w2 = net.LW{2,1};
b1 = net.b{1,1};
b2 = net.b{2,1};
save 'weighth.txt' w1 -ASCII;
save 'weighto.txt' w2 -ASCII;
save 'biash.txt' b1 -ASCII;
save 'biaso.txt' b2 -ASCII;
o = sim(net,p);
o = o';
save 'act.txt' o -ASCII;
testing code fid = fopen('config.txt','r');
inp1 = fscanf(fid,'%d');
inp = inp1(1);
hn = inp1(2);
outputsz = inp1(3);
fclose(fid);
net = newff([0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1],
[hn,outputsz],{'logsig','logsig','trainrp'});
load 'biash.txt';
b1 = biash;
load 'biaso.txt';
b2 = biaso;
load 'weighth.txt';
w1 = weighth;
load 'weighto.txt'; w2 = weighto;
load 'testin.txt';
actinput = testin;
p = actinput';
net.IW{1,1} = w1;
net.LW{2,1} = w2;
net.b{1,1} = b1;
net.b{2,1} = b2;
o = sim(net,p);
o = o';
save 'outtest.txt' o -ASCII;
Sensitivity code fid = fopen('config.txt','r');
inp1 = fscanf(fid,'%d');
inp = inp1(1);
hn = inp1(2);
outputsz = inp1(3);
fclose(fid);
load 'weighth.txt';
w1 = weighth;
for i = 1:size(w1,2),

```

```
temp = 0;
for j = 1:size(w1,1),
t = 0;
for k = 1:size(w1,2),
t = t + abs(w1(j,k));
end;
```

```
temp = temp + abs(w1(j,i))/t;
end;
temp = temp/size(w1,1);
b(i) = temp;
end;
save 'sensitivity.txt' b -ASCII;
```