TRANSPARENT NEURAL NETWORK DATA MODELLING

C.M.Roadknight

A thesis submitted in partial fulfilment of the requirements of
the Nottingham Trent University for the degree of Doctor of Philosophy

This research programme was carried out in the
Department of Computing,
Faculty of Engineering and Computing,
The Nottingham Trent University,
Burton street, Nottingham, NG1 4BU.

May 2000.

Contents

Acknowledgements

I would like to thank the UK Department of the Environment for financial support of the work in this project (Project number PECD 7/12/145) and Dr J. J. Colls, University of Nottingham for supplying raw data for training of the network described in chapter 2.

Special thanks must go to my supervisors Dominic Palmer-Brown, David Al-Dabass and Gina Mills. And colleagues Joanne Benton, Chris Goodwin, Stu Barker, John Tepper and Graham Ball.

Abbreviations

AOT40 - dose accumulated over a threshold of 40 ppb

ppb - parts per billion

ppb.h - accumulated hourly ppb

VPD - vapour pressure deficit

ANN - artificial neural network

CNAPS - connected network of adaptive processors

CAPing - correlated activity pruning

7 hour mean - mean ozone concentration over a 7 hour period

PCA - Principal Component Analysis

MARS - Multiple Adaptive Regression Splines

UN/ECE - United Nations Economic Commission for Europe.

ICP - International Cooperative Programme

Abstract

The research set out in this thesis was carried out with the aim of making the adoption of Neural Networks for real world problem solving more likely. It attempts to guide the reader in methods of application and provide novel tools for successful adoption. The testing ground for this thesis is a biological problem, but the findings of the research are applicable to any real world problem where the number and complexity of causative agents make deciding their actions complex.

The use of ANNs as predictors of natural phenomena is an important application but equally important is any resulting explanation of the heuristics a network uses to achieve this prediction. The novel methods of equation synthesis and correlated activation pruning (CAPing) are introduced and used to extract meaning from a trained ANN. Equation synthesis involves the incremental increase in the number of connections of the trained ANN used until satisfactory prediction is achieved. CAPing involves the identification of nodes that have similar effects on the desired output. Comparison of the inputs to these nodes can lead to useful dependancy relationships. Several useful generalisations have been made in this project by using these methods. Generalisations have been made using ANNs, equation synthesis and CAPing. For example, the temporal dynamics of an ozone exposure are evidenced as being more important than the quantity of the ozone exposure and light levels are shown to be the most important modifying factor in the injury process.

When the concentration of ground level ozone reaches significant levels, severe and economically important damage can occur to agricultural crop plants. Environmental modifying factors affect the expression of this injury. The use of artificial neural networks (ANNs) as tools for ozone damage prediction is investigated in this project and novel approaches to extracting meaning from these networks are examined.

Two sets of biological results have been investigated which have allowed injury development in semi-natural and natural environments to be modelled with satisfactory

predictive success. In the first instance, it was possible, given seven days of data for ozone and climatic conditions, to predict if further injury will develop on the leaves during these 7 days and, with less accuracy, the amount of leaf area effected. For the second model, a neural network can predict accurately if leaf injury will be expressed on the following day when given detailed data for the levels of ozone and light on the 5 preceding days.

ANNs have been further used to create a set of rules by which the onset of injury can be discriminated, the performance of these rules is only slightly inferior to the ANNs and their explicit nature makes them valuable.

Biological experimentation has been carried out to confirm any generalisations and add to the clarification process. This final step in the research sequence completes a cycle that ends with more data being generated that can be used for further network training.

The effect of Ozone, and its modifying factors, on yield is modelled using ANNs. The structures of these yield predicting models are explained. The importance of the nature of the ozone episode and modifying factors is analysed

The methods discussed in this thesis could be applied to any suitably complex and multivariate data set. A degree of transparency, not generally expected from neural network approaches, would be apparent. Therefore the theories would be suitable for both neural network practitioners looking to add more transparency to their modelling and for the traditional data modeller looking for alternative techniques for their complex data set but fearful of ANNs traditional 'black box' downfall.

Chapter 1. Introduction

1.1. Problem Specification.

This research uses biological data-sets and attempts to address a real world biological problem but is fundamentally an attempt to develop Neural Network modelling techniques and create novel methods that improve the success rate for application of ANNs . So the problem addressed is:

'How can ANNs be efficiently applied to real world problem solving domains, with all their associated noise and non-linearities. And what methods can be derived to aid this application."

There are some key characteristics about the data sets used in this research that will be found in many other types of data, the results of the research are therefore not bound solely to the single problem space, but to any problem space that exhibits some or all of a set of key attributes. The key attributes for suitable data sets are:

- Cause - Effect relationships that are, at most, only partially understood. If a good model of the relationships existed there would be no 'problem space'!
- Relationships are non-linear. Other methods (discussed later) are more applicable and transparent when the relationships are linear.
- Complexity. If the data set is multivariate and contains multiple dependencies and interdependencies then the first 2 points are more likely to be true.
- A lack of confidence that input variables are independent.
- Knowledge of what makes up the complete set of causative variables unknown.
- Sparsely populated regions in the data.
- Noisy inputs. Empirical methods of data modelling rely on noise free data, as they seldom generalise well. Noise could also come from human or computer

error associated with the data collection.

- Sparse outlying regions on the data.

The development and evaluation of the theories exposed in this thesis was carried out using the real-world pollution modelling case study; for any reader of this thesis to understand this, the biological data testing scenario needs to be thoroughly introduced. The effects of environmental factors and man made pollutants on agricultural crops are, to some extent, known (Heck *et al*, 1988), but this is mainly through single factor, dose response experiments. What is largely unknown is the effect of differing levels of multiple factors, such as humidity, temperature and ground level ozone concentrations. Many crops suffer leaf damage and reduced yields due to ozone related injury (Adams and Crocker, 1989). This damage can have both a qualitative and quantitative effect on crops. A decreased yield of a crop is obviously less valuable but a crop of poorer quality is also less valuable. The ability to predict levels of injury based on climate and pollution measurements would be a desirable asset but, more importantly, the analysis of any artificial neural network with the ability to predict injury could yield important information on the modifying factors that influence ozone injury.

The process of ozone synthesis is itself a complex one. The complexity comes mainly from the equilibrium states of various chemicals in the reaction, and not from the sunlight induced ozone production itself (fig 1.1), hence the ozone levels are related to other variables.

1.2. Existing Approaches to Equating Ozone Levels to Crop Injury

Some basic observations of the effects of ground level ozone levels on plants were published more than 30 years ago (Mensa et al, 1963, Otto and Daines 1969), the fact that the true impact of ozone on plants is still largely unknown is testament to the complexity of the effect. The injury causing effects of ozone and the influence of microclimate conditions have been studied in depth (Bull, 1991). Some examples of the injury caused by ozone are shown in plates 1-3. Mortensen (1992) gave linear dose response curves showing increasing leaf injury with decreasing vapour pressure deficit (VPD). This and other work (Kersteins *et al*, 1992; Van Pul & Jacobs, 1993) point to stomatal based sensitivity to ozone. This arises because stomata are by far the biggest entry point for ozone into the plant. Anything that influences stomatal conductance will therefore affect the amount of damage an ozone episode causes. Light, VPD, temperature and ozone concentrations all affect stomatal conductance (Amiro *et al* 1985; Aben *et al*, 1990; Moldau *et al*, 1990). Meteorological conditions also affect the uptake of ozone (Pearson and Mansfield, 1993). Therefore, ground level (tropospheric) ozone dose and the amount of ozone taken in by the leaf are not the same.

PLATE 1. Ozone injury on trifoliate leaf of subterranean clover (Left) compared to untreated control (right). Source: Sweden.

PLATE 2. Ozone injured clover leaves. Source: Austria.

PLATE 3. Cold stress in clover. The brown pigmentation on the central leaf is **not** caused by ozone. Source: Belgium.

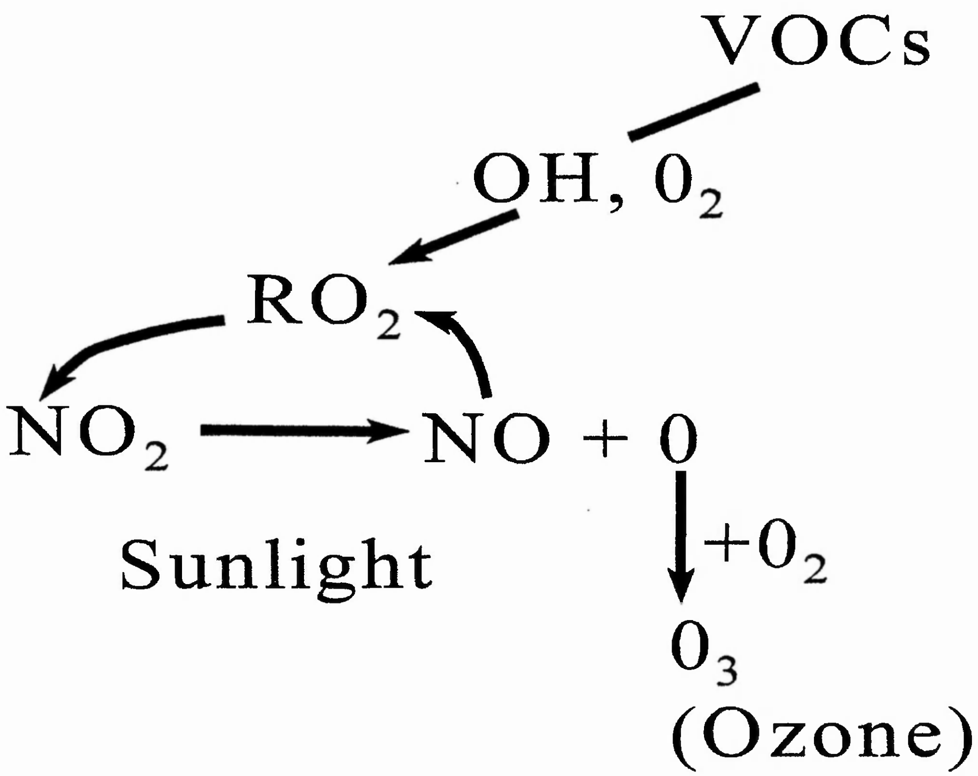


Figure 1.1. The ozone formation process. Showing a simplified version of the reaction scheme. (VOC = volatile hydrocarbons, $R0_2$ = peroxy radicals, NO_2 = nitrogen dioxide, NO = nitric oxide)

Another reason for the non-linear relationship between ozone dose and resulting injury is that once ozone has entered the leaf the amount of damage it causes is modified by biochemical factors (eg. photosynthetic rate and antioxidant level, Guzy and Heath, 1994). To simplify statistical complexity, many studies fix some of the micro-climatic conditions or ignore their effects thereby making any results fragile and dependant on a tight set of conditions (eg. Schut 1985). Any insight gained by using outdoor crops in tightly monitored (but not controlled) conditions will be of more value because of the robustness gained by using crops in diverse and natural situations.

Linear methods were attempted, in an attempt to establish critical levels for injury. The lack of clear dividing lines or step functions in figures 1.2-1.5 show how difficult it is to establish these critical levels when the relationship between ozone and injury is effected greatly, in a non-linear way by other factors. Figure 1.5 shows the 3 day and 5 day levels of ozone, as AOT40s, preceding injury. It is clear that an obvious threshold is not present. Figure 1.2 shows how injury occurs at a wide range of humidity and temperature levels. Figures 1.3 and 1.4 show how a possible 2 phase humidity influencing range could be proposed (an AOT40 of 250 at VPDs of less than 1.5 and an AOT40 of 500 at VPDs of greater than 1.5), but the wide spread of the points shows how 'loose' this proposal would be.

Linear methods of modelling have been attempted for various other biological problems (Cullan 1985, Fry 1993) including attempts to model the effect of ozone using regression analysis (Bender *et al*, 1990). These methods have proved unsatisfactory due to their inability to generalise to non-linear models.

There has been some success in applying neural networks to this problem (Balls *et al*, 1995; Balls *et al*, 1996) and similar problems (Simpson *et al*, 1992). Basic analysis of weights of the network have proved effective at pointing to important variables in the leaf damage equation. This work used results from indoor, closed-chamber experiments and while this is of importance, correlation with networks developed using wide ranging field data is an essential aim.

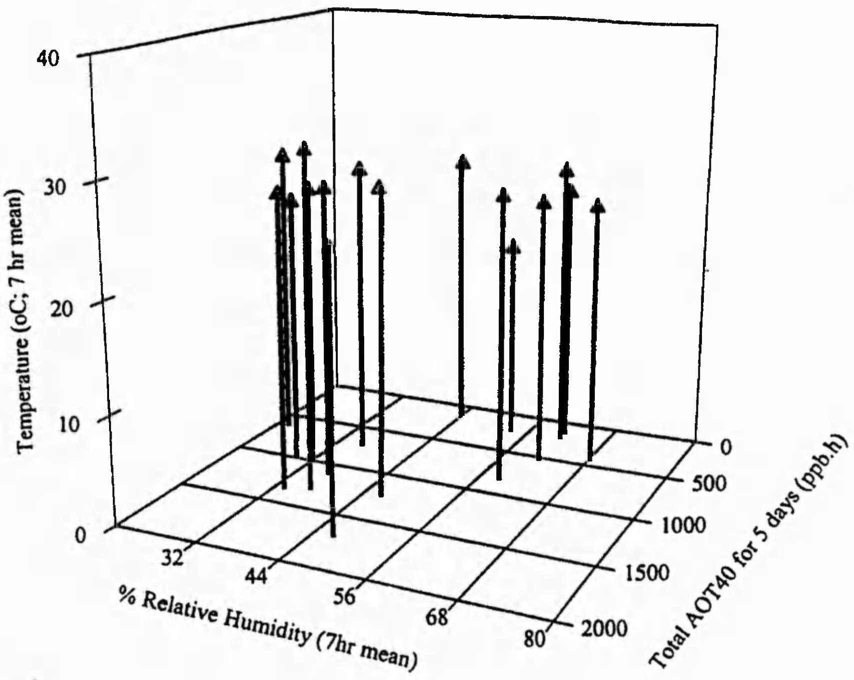


Figure 1·2 Total AOT40 (when global radiation exceeds 50 Wm^{-2}), mean % relative humidity (RH) and mean temperature (0930-1630 h) during the 5 days before injury on clover species

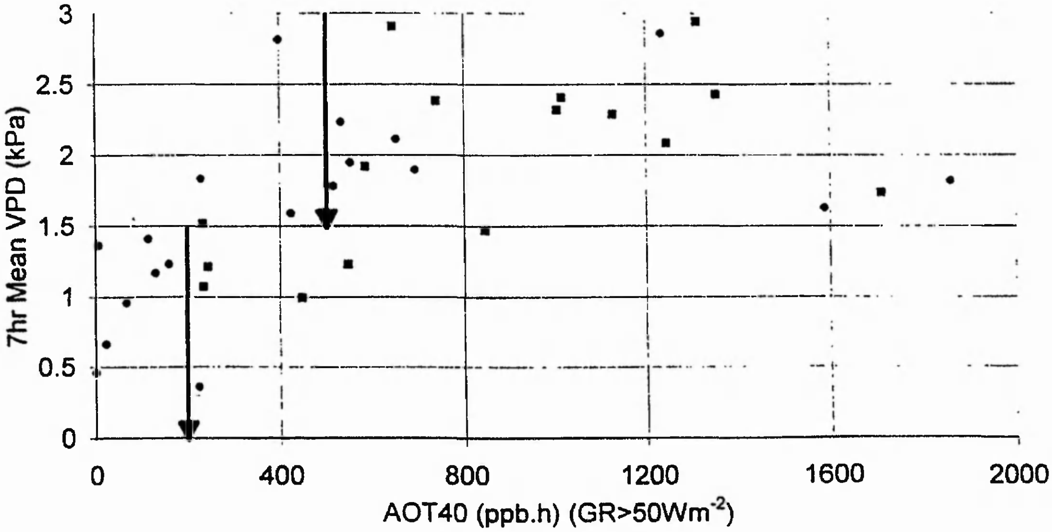


Figure 1·3 Total ozone (AOT40 when GR exceeds 50 Wm^{-2}) and the mean vapour pressure deficit (VPD) (0930-1630 h) during the 5 days preceding the presence (■) and absence (●) of injury. Data are from 1995.

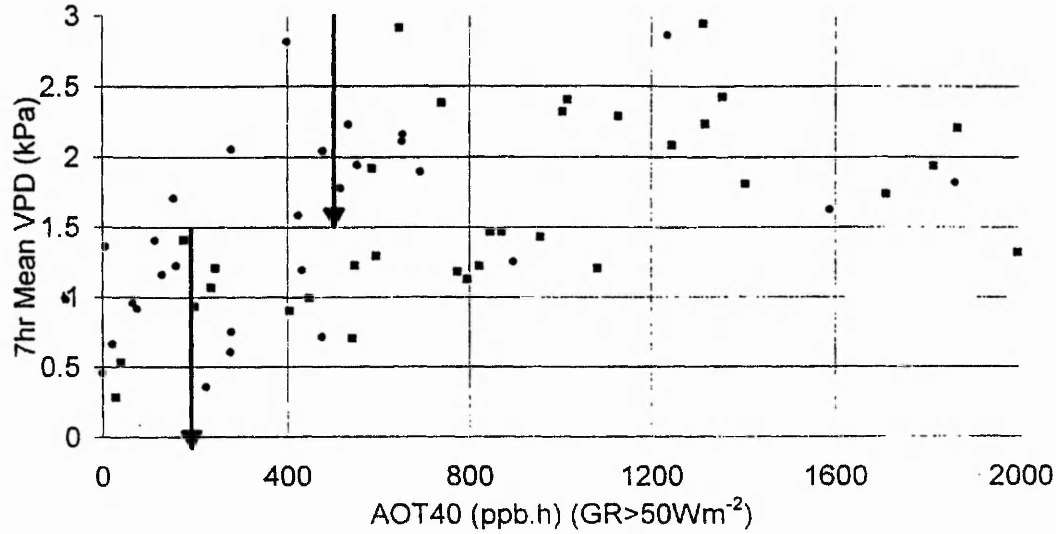
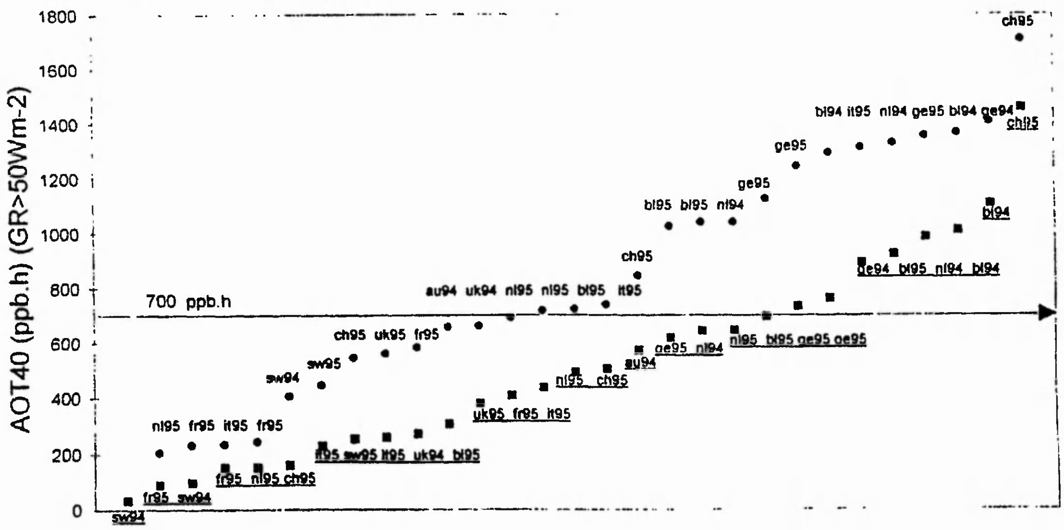


Figure 1.4 Total ozone (AOT40 when GR exceeds 50 Wm^{-2}) and the mean vapour pressure deficit (VPD) (0930-1630 h) during the 5 days preceding the presence (■) and absence (●) of injury. Data are from 1995 and 1996.



(■) (●)

Figure 1.5 The sum of the AOT40s (when global radiation exceeds 50 Wm^{-2}) for the 3 and 5 days preceding the onset of ozone injury on clover species at ICP-Crops experimental sites in 1994 and 1995.

Complex statistical Meta-analysis has been applied to the area of Biological Sciences (Hedges and Olkin, 1985). These derive confidence measures based on sample sizes, respective variances and distribution. These methods always involve some estimation and for this reason are dependent on the expert knowledge of the analyser. There are also a multitude of different ways of combining statistics each yielding different results, meta-analysis was therefore deemed to be not suitable to this problem.

1.3. Artificial Neural Networks (ANNs)

This project will apply advanced computing techniques to vast databases of information on beans (*Phaseolus vulgaris* cv. Lit) and clover (*Trifolium subterraneum* and *T. alexandrinum*). Artificial Neural Networks (ANN's) are a simplified simulation of a biological nervous system made up of a collection of connected neuron units. ANNs have been shown to cope well with data containing non-linear inter-relations and to model non-linear relationships (eg. Funahashi, 1989). While ANNs have outperformed many statistical techniques for theoretical problem solving, their application to real world problems is the true test of their usefulness. They have been successfully applied to some areas, for example, the fields of medicine (Burke et al, 1995. Orr, 1995), birth weight prediction (Lapeer *et al*, 1995) and finance (Davalo & Niam, 1990; Tan *et al* 1995) receive much attention. For an interesting selection of current applications see the July 1997 edition of Transactions on Neural Networks, this is a special issue on everyday applications and contains 13 application papers within it.

1.3.1 The Perceptron.

For each input vector X there is an output vector Y. The target output can be a binary or a real valued vector. The simplest ANN for the former example is the Perceptron, the simplest for the latter is a linear regression unit.

Developed in the early 1960's by Rosenblatt (Rosenblatt 1962), the Perceptron maps a real vector x into binary output y. The output is computed by summing all the weighted inputs (x_1w_1, x_2w_2....x_nw), if this exceeds [theta] then y =1 else y = 0. This splits the input space into two. So for each case the network checks to see if y correctly predicted. If 1 is predicted for an output that should be 0 then the weights for active input lines are decreased and the threshold [theta] is increased. If 0 is predicted for an output that should be 1 then the weights for active input lines should be increased and the threshold decreased.

The vector form of the learning algorithm is as follows:

change in weights = (y actual - y predicted)x

Figure 1.6 shows a processing unit from an ANN.

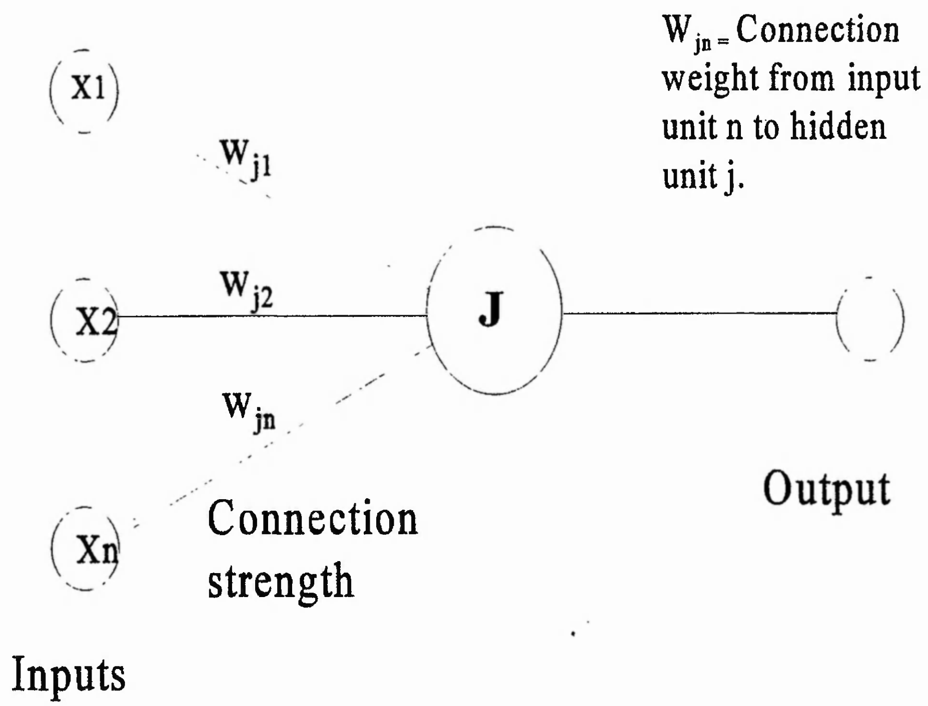


Figure 1.6 Schematic processing unit from ANN.

This is the building block for the main class of ANN's. The processing unit performs a weighted sum on the inputs and uses a non-linear function, F, to compute the output. The most useful function to use can depend on the date but a sigmoid function is widely used:

$$O_j = \frac{1}{(1 + e^{-S_j})}$$

where O_j = output and S_j = weighted sum of inputs

1.3.2 Feed Forward Neural Networks.

The essential idea of a feed-forward ANN is that each neuron outputs a smoothly rising function of the sum of its weighted inputs, eg. F(a*w1+b*w2+c*w3). The weighted sum in the brackets also equals the scalar product of the data and weight vectors, d.w, which in turn equals D*W*cos(angle between d and w), where d is (a,b,c) and w is (w1, w2, w3). This is worth knowing because it demonstrates that the neuron is effectively detecting the feature w.

When networks are built using three layers, the middle layer is called 'the hidden layer'. The benefit of more than three layers is questionable since theoretically three layers are sufficient to model (or at least approximate) arbitrary many to one mappings between inputs and outputs. This is a result that comes from Kolmogorov's Theorem (Kurkova. 1992) and as such it assumes we have activation functions for the neurons that are not just smooth and continuous, but also appropriate to the underlying 'model' function of the data. This is an important proviso, although the results of many authors working with MLPs over many years, including the author of this thesis, tends towards the optimistic view that simple monotonic functions like the sigmoid or the tan(h) are widely applicable.

Each unit within the hidden layer may act as a feature detector, responding to features[1] appearing within the input data. This neural network structure is usually called a multi-layer perceptron (MLP). The MLP architecture is the most popular in real world applications. Each layer is fully connected to the next.

1.3.3 Supervised Training.

The connection weights of a neural network need to be discovered for a correct solution to any problem and this is called training. There are many approaches to training an ANN, either supervised or unsupervised. Both supervised and unsupervised learning are valuable in intelligent data analysis. Where the interpretation of a set of (training) data is known, it is appropriate to use supervised learning; whereas if there are no available interpretations for the data, supervised learning cannot be used and unsupervised learning can be useful. Many unsupervised learning methods are either analogous to or equivalent to clustering (ART1, ART2, some Kohonen nets).

Supervised learning involves presenting the network with target answers as well as inputs so the network learns by example. One algorithm used to adjust network weights correctly is back-propagation (Rumelhart *et al* 1986). This involves presenting input data to an ANN and comparing the output from the network with the desired output and adjusting the weights to minimise the error. Types of back-propagation are used in many neural network systems. For example, recurrent neural networks have feedback connections but the input and output patterns change with time. This is

[1] Features could be repeatable patterns in the data, or distinguishing characteristics. The features on a human's face act as a set of discriminaters that enable each person to be visually individual: the shape of the eyes, size of a nose, roundness of a face.

back-propagation through time. Here the network is expanded over time. Simple recurrent networks (Elman 1990) provide an approximation of this dynamic network by truncating the error signal after each time step. This method has proved useful for temporal problems such as speech recognition.

1.3.4 Unsupervised Training.

Unsupervised learning does not need desired output as it classifies input data into categories according to their similarity. In most cases the output targets are the same as the inputs, therefore unsupervised networks usually compress complex information from the inputs. They are also very useful for data visualisation (Ripley 1993, 1994). The most common use of unsupervised networks is for cluster analysis, the most common learning algorithm used is Kohonen's Learning Vector Quantisization (LVQ). In LVQ, each of the hidden units corresponds to a cluster centre and the error function is the sum of squared distances between each training case and the nearest centre. The error function is the same as that used in standard statistical k^{th} mean cluster analysis.

Kohonen's self-organising feature maps (Kohonen 1995) combine competitive learning with dimensionality reduction. This is done by smoothing the cluster with respect to a predetermined grid.

1.4 CNAPS

CNAPS stands for "Connected Network of Adaptive ProcessorS" and is a complete parallel computer system based on a custom 64 processor chip. This system is ideal for neural network applications because of its increased speed in handling arrays. The purchase of this equipment was funded jointly by the DOE contract and the University with the intention of supporting this project.

All active processors in the CNAPS can execute the same instruction simultaneously, thus performing an operation on an entire block of data in one step. This is known as a SIMD architecture (Single Instruction, Multiple Data). In our own laboratory tests the performance was typically 20 times faster than existing workstations (in 1997) at executing ANNs of the size used in this research. The comparative performance of the CNAPS improves as the network architecture draws nearer 64 decision units (hidden + output). Parallelisation means that these 64 sets of calculations can be handled in parallel.

Other software (both self-programmed and commercially available) was used for diagnostic tasks and cross-validation. A basic C program was written that carried out simple back prop neural networks, while this was slower than the CNAPs it allowed more investigation of the inside workings of the ANN. Software called 'Neuroshell' (http://www.wardsystems.com/) was also used, as a comparison and fault checker for the other 2 software systems.

1.5. Available Data

Data on ozone related injury and the accompanying levels for a variety of pollutants and climatic factors was available from 2 sources:

- Open topped chamber experiments carried out at the university of Nottingham experimental farm at Sutton Bonnington during the summers of 1989 and 1990. These provided very complete data for green bean (*Phaseolus vulgaris*).

- Outdoor experiments carried out at many sites around Europe as part of UN/ECE ICP-Crops[2] programme. The crops used within these experiments included green bean and several species of clover. These databases contain hourly readings for various environmental and pollutant factors and regular assessments of development and injury of the crop. Fuller details of this data are given in later sections.

1.6. Aims of Project

The stages of the proposed research plan were as follows:

- Use the data available as a case study for investigating ANN methods of data analysis.
- Build a provisional ANN model using part of a data set and assess its accuracy.
- Analyse the model's performance to identify strengths and weaknesses, of both the model and the data set, which can then be used to create better models.
- Design a set of models that optimally cover the full range of possible data, thus providing a flexible and accurate approach.
- Conduct biological crop cultivation experiments to provide additional data to extend and fully test the models.
- Test wider applications of this approach by using data from an internationally conducted experiment in which a variety of crops are exposed to ambient air pollution at sites in 20 European counties (Sanders *et al*, 1993).
- Research and develop methods for simplifying the ANN models to increase their interpretability while retaining their effectiveness.
- Use the simplified models, in the form of equations, to interpret the biological significance of the data and to identify situations leading to pollution based plant damage.

---

2 United Nations Economic Commission for Europe. International Cooperative Programme on effects of air pollution and other stresses on crops and non-wood plants.

This report will show how all these aims have been achieved. Some of the aims yield results that encourage the undertaking of further work, which is detailed in chapter 10.

Chapter 2. Existing Techniques.

2.1. Comparative Statistical Techniques

For the performance of ANNs to be assessed the assessor must have some knowledge of relevant traditional statistical techniques (Linear models, projection pursuit, decision trees) and recent, or rediscovered statistical advances (MARS, Bayesian statistics). There follows a brief survey of these techniques.

2.1.1 Linear Methods.

This is a general title for simple modelling techniques such as linear regression and linear descriminent analysis (Steel and Torrie 1980). A regression or classification surface is a plane (that is why the term is "linear"), which is a combination of the available predictors (which may be non-linear functions of the original data). Because they are flexible and straightforward, these methods are widely used within other techniques. Their failings are well known and implicit, sometimes cause-effect relationships don't fit onto a line or plane.

2.1.2 Projection Pursuit Analysis.

This involves the visual examination of data using histograms, scatter-plots and rotating 3-D plots. It can often reveal structure that seemed hidden from automated induction algorithms. Included in this is the grand tour strategy, where data is rotated smoothly through all 2-D views, allowing the analyst to discover interesting perspectives (Press and Lee 1996). However, it is difficult to describe or assign meaningful equations to 'interestingness'. Conversion of visually significant observations into numerical statistics is a fundamental problem for much feature discovery research. Expressing the data in such a way as to show features only gives a qualitative pointer to a relationship. Only by representing that feature in a statistically valid way can relationships be measured and modelled.

2.1.3 Decision Trees.

Decision trees recursively divide the space into different regions and allow great local responsivity (Lim *et al* 1997). The method is very flexible and, in practice, seems to make up for the "crude" basis function (binary decisions) with its self explanatory nature. The main problem is that decision trees use process data at a rate exponential with depth, so to uncover complex structure, extensive data is required. Also, general rules are seldom found and the problem space is usually divided up too extensively.

Rule induction and decision trees are related approaches to discovering logical patterns within data sets. Although rules and decision trees may seem similar at first, they are in fact quite different both in terms of the information they discover from databases and in terms of their behaviour on new data items. Decision trees may be viewed as a simplistic approach to rule discovery. Their weakness is in the crudeness of their decision functions and ad hoc rule generation methods. There is nothing a decision tree could classify that an ANN couldn't classify.

approximation of a Bayesian approach.

Bayesian statistics is based on a different view of what it means to learn from data, in which probability is used to represent uncertainty about the relationship being learned. Bayesian methods can be used for neural network training. Our prior opinions about what the true relationship might be can be expressed in a probability distribution over the network weights that define this relationship *before* training is begun. After the data is looked at (or after our program looks at the data), our revised opinions are captured by a distribution over network weights. Network weights that seemed plausible before, but which don't match the data very well, will now be seen as being much less likely, while the probability for values of the weights that do fit the data well will have increased (Neal 1996).

The result of conventional network training is a single set of weights that can be used to make predictions of generalised outputs from a set of inputs. In contrast, the result of Bayesian training is a posterior distribution over network weights. If the inputs of the network are set to the values for some new case, this posterior distribution will give rise to a distribution over the outputs of the network, which is known as the predictive distribution for this new case. If a single-valued prediction is needed, one might use the mean of the predictive distribution, but the full predictive distribution also tells you how uncertain this prediction is.

MacKay (MacKay 1995) proposes that 'Bayesian probability theory provides a unifying framework for data modelling'. The aims are common to most modelling approaches, to find models that are well-matched to the data, and to use these models to make optimal predictions. Neural network learning is interpreted as an inference of the most probable parameters for the model, given the training data. The search in model space can also then be treated as an inference problem, in which relative probability of alternative models are inferred, given the data. Gaussian approximations for

2.1.4 Adaptive Splines.

Multiple Adaptive Regression Splines (MARS) use recursive partitioning to locate product spline basis functions of adjustable degree, rather than constants. This results in a smooth adaptive function approximation, rather than just steps or plateaus in regression tree (Freidman 1991). This method is fairly well known for non-linear regression of highly dimensional data in the statistics community. The input space is carved up into several (overlapping) regions in which splines are fit. The fit is built using first a constructive phase which introduces input regions and splines followed by a pruning phase. The final model has the form of a sum of products of univariate splines; it is a continuous function (with continuous derivatives) and is additive in the sets of variables allowed to interact. While this method overcomes some of the failings of decisions trees, its applicability is limited due to some of the overly complex implementation algorithms.

2.1.5 Bayesian Statistics.

For any two events X and Y, the probability of both X and Y being true is clearly equal to the probability of {X is true} times the probability of {Y is true given that X is true}. This can be written as $\Pr\{X \text{ and } Y\} = \Pr\{X\}\ \Pr\{Y|X\}$, from which it follows that:

$$\Pr\{Y|X\} = \frac{\Pr\{X \text{ and } Y\}}{\Pr\{X\}}$$

This is Bayes' theorem. This is an important statistical method of prediction in its own right but is also becoming an option for incorporation into the training of ANNs. Conventional ANN training methods using error minimisation can be expressed as an implementation of these methods for controlling, comparing and using adaptive networks are evaluated in MacKay 1995.

The Bayesian approach is powerful. However, in this thesis, the focus of interest is on optimising, interpreting and simplifying individual networks. The best model is not necessarily the mean case, since a slightly less likely case may be simpler and may produce a more plausible interpretation of the data.

2.2 ANN optimisation

Neural Networks are far from being an 'off the shelf' solution and how expertly they are applied can make the difference between success and failure. The correct ANN architecture must be chosen, but this may not be intuitive and may involve a series of trial and error steps. Even when a suitable architecture is decided, the application still involves many possible improvement steps to achieve the best, or optimum, performance.

2.2.1 Data pre-processing

This is possibly the most important, but most overlooked stage in ANN application. Neural networks have a reputation as universal feature approximators, performing non-linear functional mappings between groups of variables. It is often assumed that an ANN will take your raw input data and map it directly to the desired output parameters. In practice, this is only true for a simple example, for problems involving any depth of complexity this approach will give very poor results.

To explain the reasons for this, let us take a theoretical mapping problem that a neural network may be applied to. Suppose we have an image recognition task, a digital image of 256 by 256 pixels is to be categorised into one of N sets. If there was no data pre-processing then the activation of each of the 65536 pixels could be taken as an input requiring 65537 weights (inputs + bias) to each hidden node. For all these weights to be accurately trained a very large training set would be needed to ensure generalisation, this coupled with the possibly large number of epochs needed might make the problem too computationally demanding, leading to insufficient training and poor results. Pre-processing could take many forms; maybe the intensity of some of the pixels is directly correlated, meaning one pixel would suffice for all correlated pixels; maybe some pixels are always the same intensity for all examples (pixel 1 may always be black); maybe groups of pixels (2X2, 3X3) could be aggregated to give one intensity but reduced number of connections (4, 9 times). There are many example of how data can be pre-processed (eg chapter 8, [BIS98]) and there is more discussion on this later in the thesis.

Another form of pre-processing is linear transformations or encodings of the data. Different types of input data might differ by several orders of magnitude (ie. One input may have a range 0- 100 another might have a range 0-1). This is another example of carrying out a mapping in pre-processing that the ANN could carry out during training to help improve performance. Benefits also include added transparency of the trained weights due to the inputs being of similar range, also random initialisation of the weights is made more appropriate. Data can also be rescaled to give similar standard deviations, this is important for efficient use of the network weight range. For example, if 99% of the input data is >0.9<1.0 and 1 % is >0<0.9 it would make sense to rescale the data so as to make use of the full data range 0-1 as there is usually only a finite degree of granularity for the network weights to be expressed in.

The handling of missing data should also be carried out during preprocessing. Real world data often has missing points that were not collected (Little 1992)). All data points have to be presented to the network, if nothing is presented a value of 0 may be assumed, this may have a major effect on performance. (eg, 0 may represent 0 degrees, freezing, this will have major effects on any biological system and missing points should not be assumed to be 0!). There are several options for the handling of missing data. If only a small proportion of the data set has missing points then these patterns can be discarded (the reason for the missing data must be independent for this to be true). The missing point may be replaced with the mean for all other points, or if the data is temporal by the average of the post and preceding points. But the most accurate method is to model the input data distribution first so that the missing values for the missing points can be predicted (Ahmad and Tresp 1993).

So far we have considered data points as continuous variables. Its is sometimes more suitable to discrete-ise the data. If the data is of a polarised nature (ie, it is either high or low) then converting all the high values to 1 and all the low values to 0 will significantly reduce the computational task of the ANN. There are also cases when a continuous value may be forced on very discontinuous data (ie representing colour as a value 0-1 forces a linear value set on a non-linear range, it is therefore far better to represent a colour by the amount of the 3 primary colours that will make it up.).

2.2.2 Training

Training a Neural Network can require many iterative steps. It is essential that data points from the widest range of circumstances are present. In effect we are teaching the network and we must present it with examples of all the possible types of patterns it will be exposed too. This does not mean that a network must see all the patterns it will be tested on, this would just involve memorisation, but enough data to allow the network to generalise for unseen patterns must be presented. The goal of this training is to model the processes that generate the results observed, not to memorise exact patterns of inputs, thereby fitting any noise in the data set (real world data usually has a large amount of noise present). One of the variables a neural network practitioner has in his search for an optimal model is the number of hidden units; given enough hidden units an ANN will tend to train its weights to memorise the correct output for each pattern presented; given too few and the network will not have the complexity to model all the non-linear behaviour present in the problem (Bishop 1995).

The goal of training (from the ANN's perspective at least) is often to minimise the error generated, but the exact nature of this 'error' can be an important factor in the training process. The most common error function is the sum of the square of the error on the output units (Webb *et al*, 1988). The squaring has 2 roles, firstly it ensures that the value is always positive (though this could equally be achieved by taking the absolute value of the error), and secondly it exaggerates 'bigger' errors[3]. The error function should be carefully chosen, depending on the ANN goal. A more suitably trained network might be achieved by using the sum of the cubed error, absolute error, thresholded error[4] or even the error minus a constant.

Several training options are available that aid the traversing of the search space. Firstly there is the learning rate, this is a value that is applied to weight update value. In the gradient descent learning algorithm the weights of the network begin as random values,

---

3 Without squaring an error of 0.4 is twice that of an error of 0.2, but after squaring the errors become 0.16 and 0.04, a four-fold difference.

4 If certain errors are acceptable, ie. less than 10%, then an error that is within this tolerance can be ignored and treated as 0.

the error is calculated for each pattern and a weight correction filtered back through the network, a smaller learning rate corrects the weights less. The learning rate affects both the time taken to reach a minimum and the quality of that minimum. Finding which learning rate is optimum for the problem space is another case of trial and error. In some cases it helps to make the learning rate dependent on individual pattern errors. This is similar to using an absolute cubic error function (Tepper *et al*, 1995. Zhang *et al*, 2000). Because the ANN starts with a set of random values for the weights, it is usually worthwhile training the network many times, as each session will start with a different set of random weights, the value of these will affect the final weights and therefore the final generalisation performance. The learning rate can also be varied during the training run.

Another technique used to aid the search for a global minima is momentum (Plaut *et al*, 86), this adds an inertia to the search through weight space. A momentum value of between 0 and 1 is applied to the learning rate, but its effect is linked to the recent performance in error minimisation. The learning rate is increased during successful periods to speed up training and is reduced at more delicate periods in the training run. Though momentum can sometimes speed up learning it must be remembered that it is yet another variable to be optimised.

A decision must also be made as to when to stop the training of the network, this could be based on number of iterations, time, error function value (or change in error value) or on some comparison with a validation step. Generalisation can be compromised if training is carried out for too long, but this should not be so if the architecture is optimally designed.

Other factors that may affect the training could be the order that the patterns are presented to the ANN, and if the weights are updated after each pattern or after a number of patterns.

Given perfect training and sufficient plasticity in the network architecture, it can be shown that a neural net classifier produces the maximum likelihood classification of the inputs (Bishop 1995). In general we cannot expect ideal or sufficient training data in the real world, and the training of a neural net by gradient descent methods of error minimisation is as much heuristic as it is probabilistic.

The link with Bayesian methods has been extended in recent years, with many researchers applying Bayesian methods to training and using ANNs (Bishop 95). As yet, this work is not ideal for the purposes of interpreting ANNs, because it involves combining the results of many nets, or sampled weight distributions. The direct, deterministic relationships that can be established in a single, optimised network architecture are more helpful in terms of intelligent data analysis.

2.2.3 Testing

Once a network has been trained it must be tested, in much the same way as students are taught a range of subjects then given an exam containing questions they have not necessarily seen before. The importance of testing is to find out if the network has discovered the underlying mechanisms and features that cause the outputs observed, as opposed to having memorised a set of relationships.

Choosing a suitable set of patterns to make up the test set can be a difficult task. One popular method is to take a random selection of patterns from the training set and make these the test set. Sometimes this may be the only 'safe' option if knowledge of the data is limited, but sometimes a more discriminating approach may be more prudent, this is especially the case for small data sets. For example, suppose the ANN was set the task of identifying 5 different types of tanks from their engine noise. For each tank there are

20 samples of engine noise from different positions, over different terrain etc. This would enable you to select a representative test set, so the ANN could be trained using 15 samples of each tank and tested on 5 samples from each tank. Random selection of training set may yield the same sets but could also yield a test set with no samples for tank a, or containing all 20 samples for tank b. In the first case the ANN would not be tested on its ability to recognise tank a, in the second case it would be attempting to recognise tank b in the test set without ever being trained on it.

Another decision must be made as to the ratio of test set to training set, this is especially important in cases where data is limited. Use too many of the patterns for testing and you may not have enough for accurate training, use too few and the effectiveness of the test may be compromised. The right ratio will again depend on how much knowledge of the data is available.

There is another type of testing which is sometimes called validation. This assumes that if a test set is tried but found to prove the network has not suitably generalised it is then incorporated into the training set and retrained. Eventually a test set will show that the ANN has generalised well enough to predict the results of the test set with desired accuracy, a validation set can then be used to improve the sureity of the testing. This approach is fine when datasets from the problem space are plentiful or artificially generated, but in real world problems datasets may be scarce or take a long time to generate. The work in this research falls into the latter catagory, for some of the very complex, multi-dimensional problems less than a hundred datasets were available[5]. In this situation the most efficient use of a limited supply of data becomes of high priority and without sufficient patterns in the training and test set any training would be of too poor a quality to even justify its presentation to a validation set. It was therefore decided that no validation sets could be justified and the results of the networks on the

---

5 Consider here that one data pattern for an ozone experiment may take several months to generate and the value of each data point becomes apparent.

training and test sets would express their ability to solve a problem.

2.2.4 Improving Performance

Once initial results are achieved it is an essential part of the problem solving process to optimise all available metrics so best performance is achieved, given the range of options available this is not a trivial task.

Cross - Validation.

Initial trials of training and testing might allude to the fact that there is an insufficient amount of data for sufficient generalisation, if this is the case then methods such as cross-correlation and the k-fold method might be used. Cross-validation is a method for estimating error based on resampling (Weiss and Kulikowski 1991). In k-fold cross-validation, you divide the data into k subsets of (approximately) equal size. You train the net k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the error. If k equals the sample size, this is called "leave-one-out" cross-validation. "Leave-v-out" is a more elaborate and expensive version of cross-validation that involves leaving out all possible subsets of v cases. The merits of using cross-validation for small data-sets is demonstrated by Goutte (1997). Cross-validation can be used simply to estimate the generalization error of a given model, or it can be used for model selection by choosing one of several models that has the smallest estimated generalization error. For example, you might use cross-validation to choose the number of hidden units, or you could use cross-validation to choose a subset of the inputs (subset selection). A subset that contains all relevant inputs will be called a "good" subset, while the subset that contains all relevant inputs but no others will be called the "best" subset.

Leave-one-out cross-validation often works well for estimating generalization error for continuous error functions such as the mean squared error, but it may perform poorly for discontinuous error functions such as the number of misclassified cases. In the latter case, k-fold cross-validation is preferred. But if k gets too small, the error estimate is pessimistically biased because of the difference in training-set size between the full-sample analysis and the cross-validation analyses.

Data Gathering

Once results have been achieved with the data set available, more informed analysis might be possible as to what data would be beneficial to the training. If the model generalises poorly to certain types of pattern, then further data of that type should be collected and used. Take a hypothetical speech recognition application, the ANN may not identify the speech of old females as well as other groups, so more samples from this subset should be taken and used for further training. This procedure sometimes involves a semantic assessment of the results to identify examples of this type, and is usually iterative. So in our example, once we have added our old female data we would then look at the results and see if a new weakness had arisen. This cycle of feedback is essential to process of optimisation, but is only as good as the data used, if you have no examples of old females you will not know you have a weakness for identifying their speech, so the data must be as global as possible.

Prior Knowledge

The researcher may be able to optimise the network structure by taking into account some proven facts about the data. For the speech recognition problem, it might be a fact that speech within a certain frequency range has no bearing on the word, so inputs for changes at these ranges can be safely ignored. This is fine if the facts are irrefutable and non-conditional, but if this is not the case then large areas of the problem space will be ignored that might contain the best solution. This may be the case if simple rules of thumb are applied to remove inputs. This becomes even more dangerous if the ANN is being used as a research tool, for pollution modelling for instance, one might assume that oxygen is not a pollutant so should be removed from the list of possible influencers in the model, but if it has a modifying effect on the impact of another pollutant then it cannot be ignored. Use of prior knowledge can reduce the search space effectively but great care must be taken to retain as much of an ANN's "clean sheet" approach as required, in order that it is still an effective tool for discovering the key relationships in the data.

2.3. Discussion.

This chapter firstly discussed existing techniques for modelling complex behaviour. None of the methods discussed were wholly suitable for both the types of problems that were to be tackled. They were either too linear, or could only be applied to a rigid domain, or just inferior to ANNs! The fact that many of the methods could be simulated using an ANN (but not vice versa) points to the flexibility of ANNs and encourages their use.

The second part discussed accepted methodology for applying ANNs to a problem. Correctly applying an ANN is more difficult than is often assumed, with many variables and metrics of evaluation making for a very subjective optimisation problem. Improved accuracy in the design and application stage will mean better performance in the 'meaning generation' and analysis stages.

Chapter 3. Demystifying ANNs

Mathematical explanations are simple to generate from an ANN but their non-linear complexity makes interpreting them, for non-trivial ANN's, difficult. Interpretations of ANNs are essential if functioning ANNs are to be adopted as modelling tools for real world problems. Anecdotal evidence (its accuracy at predicting a subset of the data) of an ANN's performance is usually cited as a performance indicator. When someone boards a plane for the first time, they want to trust the principles of fixed wing flight as well as statistics about flight safety. Also when a plane crashes, investigators need to know the mechanisms of flight and not just safety statistics. ANNs will never be fully adopted until they cease to be seen as a black box prediction machine (Figure 3.1).

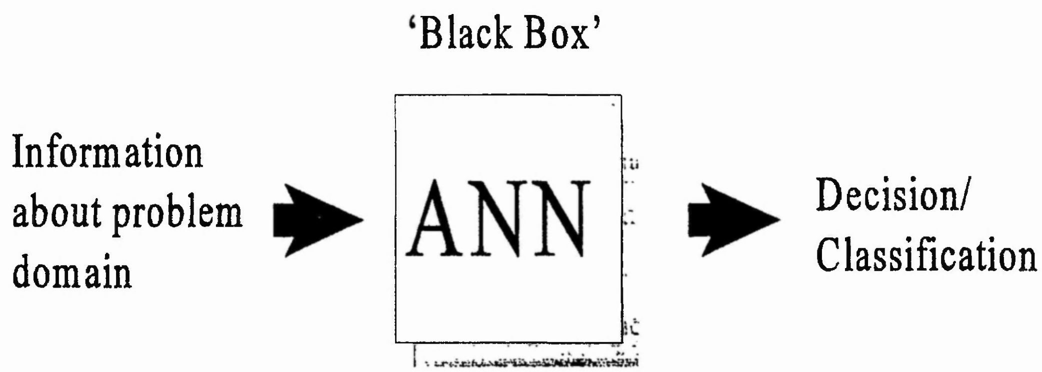


Figure 3.1 ANNs as decision makers

In order to analyse and interpret data, it is necessary to find the appropriate structure or topology for the data model, and in the case of a neural network, this implies finding the optimal architecture.

3.1 Simple methods of interpretation

Occam's razor says "If you have two theories which both explain the observed facts then you should use the simplest until more evidence comes along". This can be applied to interpretation by saying it is always best to try simple methods of interpretation, before complex ones (Garson 1991).

3.1.1 Linear analysis of weights.

There are a few simple approaches to understanding relative, causal relationships in an ANN. It is sensible to remove any connections with a weight of zero, but these weights are extremely unlikely, even if the input was a set of random numbers. It is sensible to remove any inputs that are constant in all input patterns (these can be represented by the bias unit). Correlations between two inputs throughout the training set mean that the inputs will have identical effects, so one can be removed. These trivial preprocessing techniques seldom reduce the complexity of an ANN because instances of these redundancy are rare.

MLPs can solve complicated, non-linear problems and yet the importance of inputs can be illustrated with simple functions. The following non-linear function is used as an example of the uses and problems of this linear approach.

An MLP with three inputs (X1, X2, X3), a single hidden layer with 5 tanh hidden units (H1, H2, H3, H4, H5) and an output Y is used. The weights of the MLP are given in table 1.

	H1	H2	H3	H4	H5	Y	
Bias	25	25	150	150	0	-0.1	
X1	100	-100	0	0	0		
X2	0	0	100	-100	0		
X3	0	0	0	0	1		
H1						0.1	
H2						0.1	
H3						0.1	
H4						0.1	
H5						1	

Table 1.

The output function of this simple MLP can be written as:

$$Y = 0.1*\tanh(100*(X1+0.25)) - 0.1*\tanh(100*(X1-0.25)) +$$
$$0.1*\tanh(100*(X2+1.5)) - 0.1*\tanh(100*(X2-1.5)) +$$
$$+ \tanh(X3)$$

Y is therefore the sum of three functions contained within the above equation. Each of these functions uses only one input. This is an additive model and is the easiest case because each input is independent of all other inputs, so each input can be assessed in isolation in one dimension. If inputs are distributed over the interval 3 to -3 the effects of different inputs can be visualised by the plot in figure 3.2.

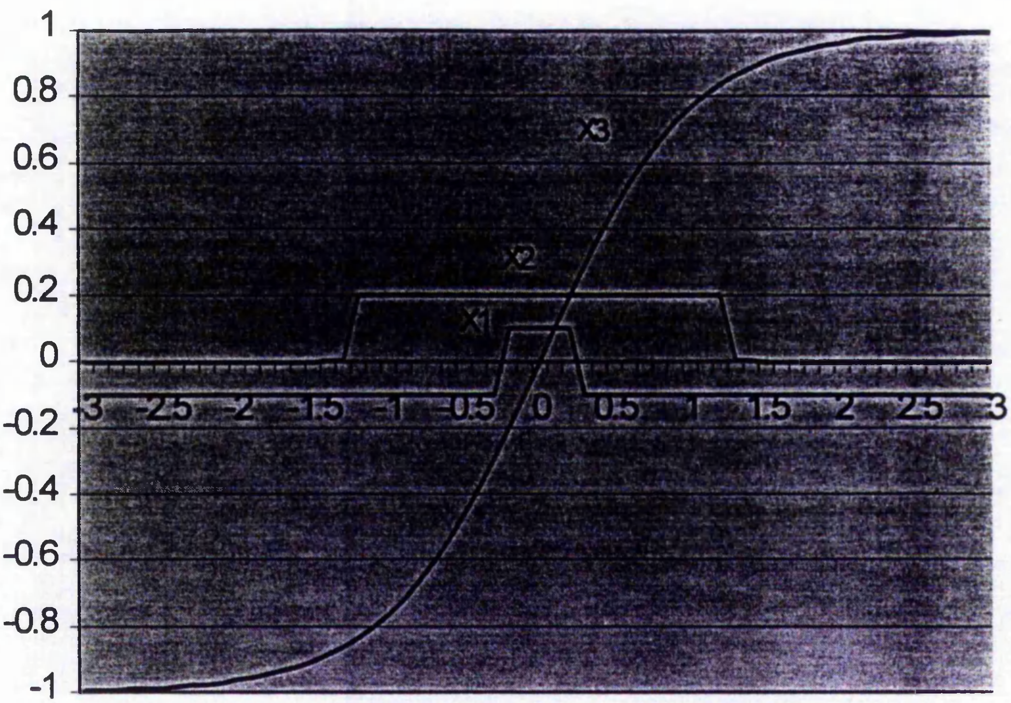


Figure 3.2. Effect of different inputs values on an additive MLP.

From this graph it is apparent that X1 is the least important input only influencing the output to a small degree over a narrow range. X2 affects Y over a larger range and to a greater degree. X3 has a much greater effect on the output over all input ranges.

We can now look at the weights again to see why comparing the weights of a MLP, to infer the relative importances of the model inputs is a poor approach. The size of the input to hidden unit weights does not automatically mean that the input has a huge effect on the output. The squashing function of hidden units limit that effect. The size of this weight is related more to the abruptness of change than the size of the change.

If just the input-to-hidden weights were considered for our simple example MLP the sum of the absolute input weights for X1,X2 and X3 respectively would be 200, 200 and 1. This would suggest that X1 and X2 were of equal importance and that X3 is much less important that X1 and X2. Both of these assumptions are clearly incorrect by examination of graph 3.2.

An alternative approach is to combine input-to-hidden weights and hidden-to-output weights, one possible formula would be to multiply the Input-hidden and hidden-output weights. For our simple example this would give an importance for X1, X2 and X3 of 20,20 and 1 respectively. This suggests the same incorrect conclusions as just looking at input to hidden weights. Using the weights from all the layers still does not take into account the squashing function of the hidden units, to incorporate this function involves knowledge of the exact input values and bias's, incorporating these complexities just returns us to our original, complex equation.

3.1.2 Partial Derivatives and Differences.

It has been shown that interpreting the weights in a linear way is flawed. An approach that is capable of capturing more information about non-linearities is to monitor the change in output with respect to changes in the input. The gradient of this change in output can be calculated (partial derivatives), each giving a rate of change of the output for a particular input. This relationship between changes in input and changes in output seems a more attractive method of allocating importance, but it too is vulnerable to giving misleading results. If an input exists as a discrete set of values (ie. binary, 1 or 0), then taking partial derivatives over small ranges of input space will be unsafe. The inputs must be continuous, with a non-deterministic distribution, because the derivative is calculated over very small changes in the input which can, for example, not be representative in a boolean input range.

Another approach is to vary one input while fixing all others at their average. This too can lead to incorrect conclusions. For our example network the partial derivative at the input mean (0) for each input X1, X2 and X3 is 0, 0 and 1 respectively. This is a misleading as X1 and X2 are not completely unimportant.

It is clear that partial derivatives must be sampled over the complete range of the input space. This generates a large number of values which must be reduced into a representative value. One possible way to do this would be to take an average, this is one useful approach but large positive and negative values may cancel each other out. In our example MLP the average partial derivatives for X1, X2 and X3 are 0, 0 and 0.33 respectively, still giving no importance to X1 and X2 and no difference between the effect of X1 and X2. One method to overcome this is to take the absolute values of partial derivatives, this gives a good measure of the importance of inputs but gives no information on the direction of that influence over the input space. It can be seen in our example MLP that X1 and X2 have the same mean and absolute partial derivative. The absolute partial derivative for X1 and X2 is 0.5 which is larger than the absolute partial derivative for X3 at 0.33 which also suggests an incorrect conclusion in terms of relative importance.

If the partial derivative only provides local information, then a means of providing more far reaching information is required. One method proposed for this is to take the difference in output over a range of inputs. If the following formula is used:

$$\text{Difference1} = f(X1+h, X2, X3) - f(X1, X2, X3) \text{ where h is the interval;}$$

then average or absolute values for each set of differences will be indicative of the importance of each input for simple networks. An even safer approach is examine differences for a range of h values.

The problem with this system is that the importance is not yet described as a single metric. An average of absolute differences is one solution, this seems to work for our simple multi-layer perceptron (MLP) example giving importance measurements for X1, X2 and X3 of 0.03, 0.099 and 0.919. This may not work for more complex networks where more inter-related, non-linear features have been modelled or where the inputs are discrete.

3.1.3 Relative Importance and Interdependency.

A fundamental question in ANNs and intelligent data analysis is how to identify and separate the key inputs from those that are merely dependent. Later in this thesis we attempt to tackle the problem of modelling some of the relationships between key features in the data set by correlating hidden neuron activation profiles. But firstly, it is important to consider the individual inputs. Initially, as a matter of course, modellers correlate the input variables to identify linear relationships. Once they have been removed, leaving linearly independent inputs, we still need to understand which are the most important inputs.

A time consuming but simple method for judging importance is to remove inputs from a neural network, retrain the network and compare respective training errors. If an input is essential for modelling a complex function then removing it will cause a large increase in the training and/or testing error. It may be useful to approximate this effect using a Hessian matrix. (Hassibi and Stork, 1993). In cases where there are many more training cases than network weights, a Hessian matrix approximates the change in error function more efficiently than cpu intensive retraining. This should be avoided if the number of training patterns does not vastly exceed the number of weights as the approximation may be poor.

Particularly if these methods are automated they can be invaluable. One approach is to calculate the relative removal index (RRI) for the inputs. To do this, one input is removed at a time to test its importance, which is measured as the increase in error. The relative removal index for each input is the ratio of the increase for the input to the sum of the increases for each of the inputs removed separately.

All inputs whose RRIs are not very small are important inputs, but where the RRI is approximately zero, or very small, there is a chance the input can be removed. The index will not in general be zero because even with an optimised network, the training set is finite and imperfect. (Some RRIs may be negative, indicating the improvement of the model by removing the input.)

In the simple case of just two inputs having very low RRIs, one or possibly both can be removed. The more general case to be considered is where there are several low RRIs and there may be dependencies between one and a group of other inputs taken together in addition to one or more one-to-one relationships. RRIs will be small for all of the group as well as the single input related to the group. In this case RRIs can be calculated for pair removals of combinations drawn from the low single RRIs. Any pair containing a critical input with one from a dependent group will give a significant index. By identifying non-zero pair removals the critical inputs can be identified and the rest of the lone RRI inputs removed.

These methods are limited in the sense that they do not attempt to model dependencies, only to identify them. The next section details some approaches to assigning importance to inputs, or combinations of inputs, for more complex neural network structures, by interpreting the topology of the network.

3.2 Advanced Methods of Interpretation.

The above methods are in some sense intuitive and are an application of simple maths and common sense. Other methods required much more thought and testing to become viable approaches, this section details those approaches, some of which are still active areas of research.

3.2.1 Linearised Equation Synthesis.

Synthesising the simplest functional form of the ANN is a powerful means of providing interpretations of the data, but since it relies on the particular F's (neuronal activation functions) in the ANN, it is not necessarily in the simplest possible form; because we do not know the choice of F's in the first place were necessarily optimal for the data concerned. For example, the optimum may be a combination of function types, such as gaussians and tan(h)'s. Although the tan(h) or sigmoid are general purpose, their exclusive use can mean that extra weighted hidden neuron inputs or even additional hidden neurons are necessary, compared to the optimal solution that would be obtained if it were known that a certain mix of functions should ideally be used.

Given the probable hidden non-linear redundancy, due to non-optimal function selection, linearising the F's can yield significant simplifications, by allowing weighted inputs on the same variables into different neurons to be combined over certain ranges of values, for which the neuron has a nearly straight line response. This approach is called piecewise linear modelling (commonly used to model the response of transistors in amplifiers) of the neurons and involves converting non-linear functions into straight line segments, for example, an approximation to the sigmoid, $F(x) = 1/(1+\exp(-x))$, is $f(-2.5 < x < 2.5) = 0.5 + 0.2x$, $f(x<-2.5)=0$, and $f(x>2.5)=1$. This is shown graphically in figure 3.3.

For an accurate interpretation we can generate a set of linear equations for given ranges of weighted sums of parameters, x. This will not be useful in all cases, since it may produce too many equations, but in cases where there are only a few hidden neurons it is generally effective, and presents a model that is a linear piecewise version of a non-linear one, and that would not have been possible to find by traditional linear or non-linear methods of regression.

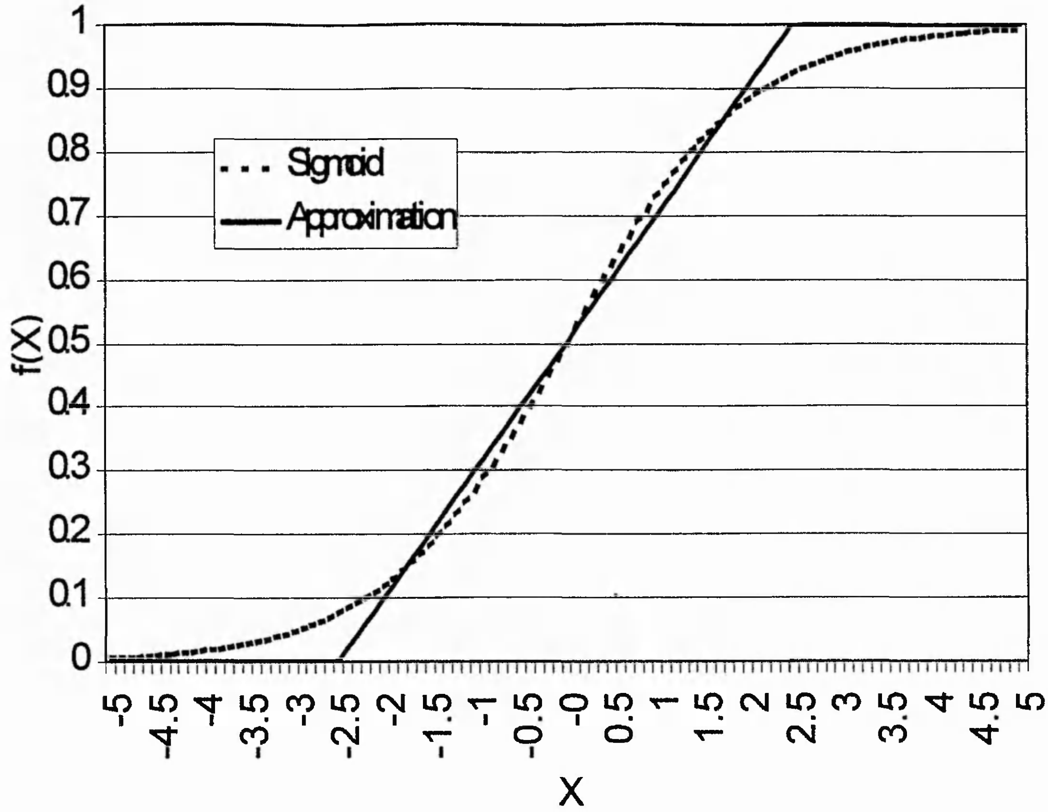


Figure 3.3. Approximation of a Sigmoid function

3.2.2 Boolean Rule Extraction.

ANNs trained using back-propagation as a training algorithm are the most established of all neural approaches; it is therefore most important that interpretation techniques be developed for this area. Much of the work is this area has involved extracting Boolean rules at the decision unit level of a trained ANN (Wiersma et al 1995, Craven and Shavlik, 1995).

The approach of Craven and Shalik is to initially search for a set of weights containing a connection of sufficient positive value to exceed any threshold regardless of the other connections. These connections are then written as a rule. Once all single connections have been analysed, combinations of connections are analysed to see if they have this over-riding effect on a thresholded unit. The pseudo-code for this process is given below:

For each hidden unit and output unit:

> *Extract up to Sp subsets of the positively weighted incoming links for which the summed weight is greater than the bias for the unit.*
>
> *For each element p of the Sp subsets:*
>
> > *Search for a set Sn of a set of negative attributes so that the summed weights of p plus the summed weights of N-n (where N is the set of all negative attributes and n is an element of Sn) exceed the threshold of a unit.*
> >
> > *With each element n of Sn set, form a rule: 'if p and Not n, then the concept designated by the unit.*

One major problem with this approach is that the solution time for finding all rules is exponentially related to the number of links to each unit, therefore limiting its application to small ANNs. Any ceilings or restrictions imposed to reduce the amount of computation cause a degradation in rule quality.

An interesting method for overcoming this problem is proposed in Wiersma *et al* 1995. A notion of input quality is used, input fields that have a low value for a variable called 'Rating' are ignored, therefore reducing the search space. This is done by rating all input fields for how much a 0 to 1 change of an input field can maximally change an output field.

3.2.3 Fuzzy rules.

Running parallel to the development of boolean rule extraction is the extraction of fuzzy rules from a trained ANN (Tickle et al 1994), these are often called neurofuzzy systems. Neurofuzzy systems normally comprise of 3 elements.

1. A mechanism for inserting expert knowledge into a an ANN in the form of fuzzy rules.
2. Training of the ANN
3. Identification and extraction of knowledge as a set of membership functions.

In an example of this Masuoka *et al* (1990) used a decompositional approach to produce a set of initial fuzzy rules from experts in the chosen domain. Small, three layer ANN's were used to represent membership functions. A rule net is then devised that represents the fuzzy operations on the input variables. This rule net is then pruned to generate the required number of rules. There is however, no procedure for yielding the optimum number of rules.

Another approach that falls between both types of rule extraction is proposed in a paper by Tickle *et al* (1994). This is called DEDEC and attempts to produce a robust methodology for decoding information contained within an ANN trained to solve any problem. A central part of this process attempts to extract symbolic rules form an ANN by looking at a subset of cases. The rule extraction process attempts to minimise the information required to distinguish differing patterns of input data. Cases are generated by presenting a set of inputs to the ANN and monitoring the resulting output. Cases are also ranked in order of importance based on the size of the weights for each input as an indication of their impact on the output. DEDEC has been successfully tried on ANNs trained to solve classical optimisation benchmarks like the Monk problem and real world problems like mushroom identification and generates believable rules on how ANNs solve problems. However, the use of weight values as importance indicators, as discussed earlier is to be treated with caution.

3.2.4 Meaning extraction from non-standard ANNs.

Most meaning extraction algorithms are devised to be used on back-propagation trained, feed-forward neural networks. This makes sense as this is the most widely applied network architecture. Problems arise when meaning extraction is required for ANN's that are of more unusual structure. One of these is Time-Delay Neural networks (TDNN). The following section shows an approach for extracting meaning from non-conventional ANN's.

3.2.5 Network Unfolding.

Effective rule extraction algorithms may be available for a simple feed forward ANN but what if the network that needs analysing is not a simple feed forward ANN? One approach would be to preprocess the network weights, producing a conventional feed forward network. Very little work has been done on this but one example of this a network unfolding algorithm developed by Lin and Dayhoff (1996) that "unfolds" a time-delay neural network.

Input nodes are first unfolded, this involves duplicating new nodes and assigning existing weights and time delays to the new nodes. A re-adjustment of input values is then carried out to remove time delays. The same two processes is then carried out on the hidden layer nodes.

3.3 Discussion.

A range of methods have been proposed that all enable some visualisation of the meaning associated to the weights of an ANN. While they all have some worth, none of them gave enough accuracy or applicability to use 'off the shelf'. The methods pf linear analysis and linearised equation synthesis came closest to fulfilling the requirements of this project and (in spirit) were used as starting point for some of the methods presented later in this report.

Chapter 4. Predicting the Magnitude of Complex Real-World Events.

As an initial task we apply neural networks to the task of predicting the amount of a certain event, we can also attempt to see if the confidence in this prediction can be used to determine if the event will happen at all. One example could be asking an ANN how much rain there will be tomorrow. The magnitude of the output is firstly used to decide the question "How much rain will there be tomorrow?", but the magnitude could then be used to decide if it is going to rain at all. Breaking it down into a simple rule set:

1. Magnitude of event A is proportional to the ANN output.
2. Outputs of below X can be deemed 'No Event'.

This chapter describes an initial approach to the modelling of ozone induced injury, to predicting the magnitude of this injury. The success or failure of this initial application would tell us if our data sets were suitable and sufficient. The data used was collected at Nottingham University.

4.1. Data Preparation

Data for *P.vulgaris* was collected for the 1989 and 1990 growing season, for two cultivars, Lit and Nerina. The plants were grown in open top chambers which allowed the modification of pollutant levels surrounding a plant while keeping it in a more natural outdoor position (Sanders *et al* 1990). The bean plant was grown in eight microclimatic situations:

1. Ambient air (AA). (No additions)

2. Non Filtered (NF) air pumped into the open top chamber.

3. Charcoal Filtered (CF) air pumped into the open top chamber.

4. NF + 8 ppb Ozone added for 7 hours per day

5. NF +16 ppb Ozone " " " " " "

6. NF +24 ppb Ozone " " " " " "

7. NF +32ppb Ozone " " " " " "

8. NF +40ppb Ozone " " " " " "

This simulated a range of ozone conditions in semi-natural conditions. The data was in the form of hourly readings of a number of pollutants and environmental parameters (4.1). The most significant of these were the parameters: ozone, solar radiation, sulphur dioxide, nitric oxide, nitrogen dioxide, air temperature and humidity. This data was put into the form of a set of vectors with inputs and outputs for the neural networks. The data was formatted as seven day blocks. The output data was the change in ozone induced damage from day 0 to day 7 and the change in number of leaves. The leaf damage was assessed by scoring the amount of leaf damage on the third leaf. The leaves were initially scored on a nonlinear basis (Plate 4), this can be converted to a more linear percentage damage scale using the damage score to damage % relationship on 1. The readings of each parameter, for each day, were compressed into a daily maximum and a daily sum (fig 4.2).

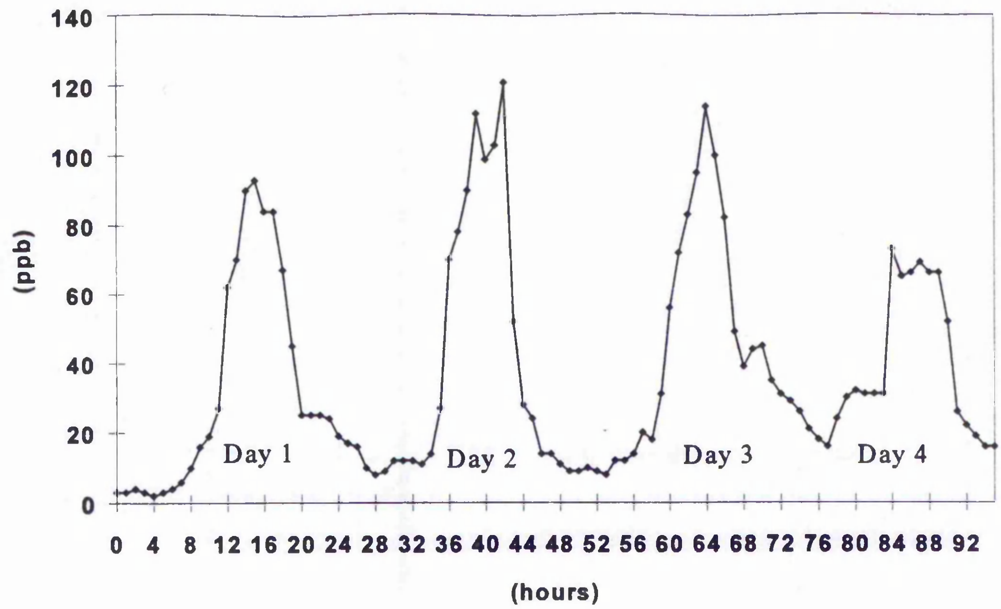


Figure 4.1 Hourly ozone levels in non-filtered air with 40ppb additions of ozone

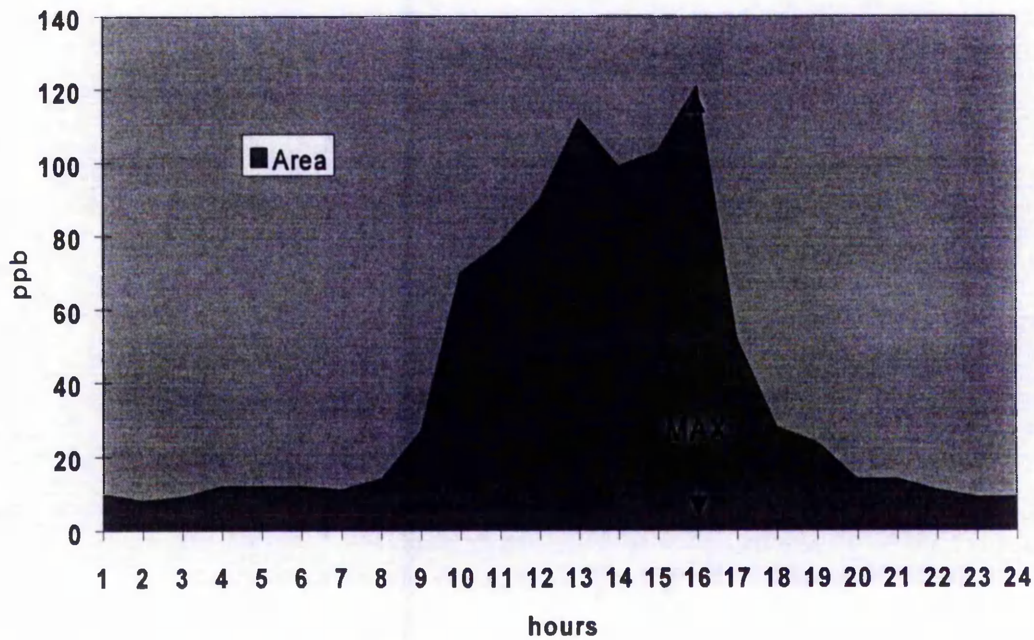


Figure 4.2. One Day of Ozone Readings

This compresses 24 hourly readings into two statistics that represent the conditions. A sum and a maximum were calculated for each parameter, for each day. Thus 98 network inputs were needed (7*7*2). Four extra inputs were also available from the source data:

1. Level of ozone damage at day 0, as this might be important to the network in assessing how much further damage can occur.

2. Number of leaves at day 0, as the network was initially asked to predict the change in the number of leaves. This gives another indication of the growth stage.

3. Growth stage (eg.18 = pair of simple unfolded leaves, 38 = all flowers open, at day 0.

4. Growth stage at day 7.

This method of information extraction produced 72, 7 day data sets with which training and testing could be carried out. Damage outputs of zero were common amongst these sets (54 out of 72), this is because many plants experienced no change in damage over a seven day period. These 72 data sets were separated into a training set and a test set in a random 3:1 ratio.

Plate 4 Ozone Damage of *Phaseolus Vulgaris* (green bean) Leaves

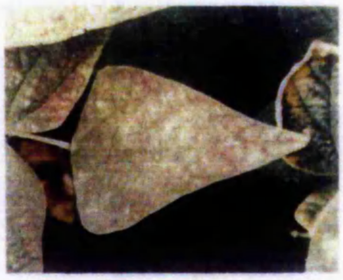
KEY FOR O3 SYMPTOMS

0 (0%)

1 (1–3%)

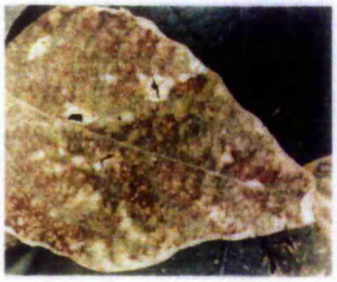
2 (4–10%)

3 (11–25%)



4 (26–50%)

5 (51–75%)



6 (>75% & Dead)

April 1990, Gina Sanders

4.2. Network Architecture

A three-layer back-error propagation network was used throughout the experiments. The CNAPs software allows for the varying of a range of parameters that affect the training of the network. Learning rate, momentum, derivative offset and number of hidden units were all varied until the network both trained and tested satisfactorily. Although there were 102 possible inputs to this damage predicting ANN, it was thought that may be more useful not to use all of them. Firstly, if an ANN can learn successfully without a subset of the input variables then it can be concluded that these variables are not of a big enough importance in the development of injury. Secondly, an input may give a very good indication to the output but not for suitable reasons. The following is an extreme example of this and can be used to help understand the problem. If the desired output (eg. % change in leaf injury) was given as an input along with the rest of the inputs, the ANN would soon learn that the best input for predicting change in leaf damage over 7 days is change in leaf damage over 7 days but this would not be making use of the available, useful information. Rationalisation of the network format was required to implement both of these guidelines, but initially the ANN was trained with all of the available inputs.

4.3. Equation Synthesis

Once an ANN has been successfully applied the solution manifests itself as a set of activation function parameters and connection weights. In all but the simplest of networks these have no apparent meaning. These weights and functions can be used to create an equation but while this will have more meaning to a scientist, it will only be comprehensible for trivial networks.

To synthesise useful equations from non-trivial networks some rationalisation is required to keep the size of the equation manageable. This is achieved by removing connections of low weight and testing the remaining, partially connected network. The algorithm for equation synthesis is as follows:

1. For all hidden nodes

> **Set required threshold for weights from hidden units to output unit**
>
> **Remove all hidden units below this threshold**

2. For all input nodes

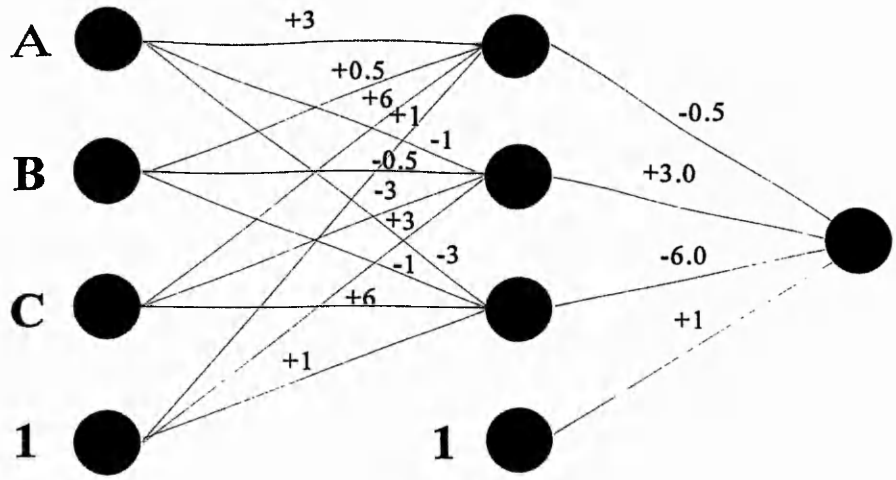
> **Set required threshold for weights from input units to hidden units**
>
> **Remove all inputs to hidden units below this threshold**

3. Create equation by matching input nodes, approximate signed weights to source of input node value.

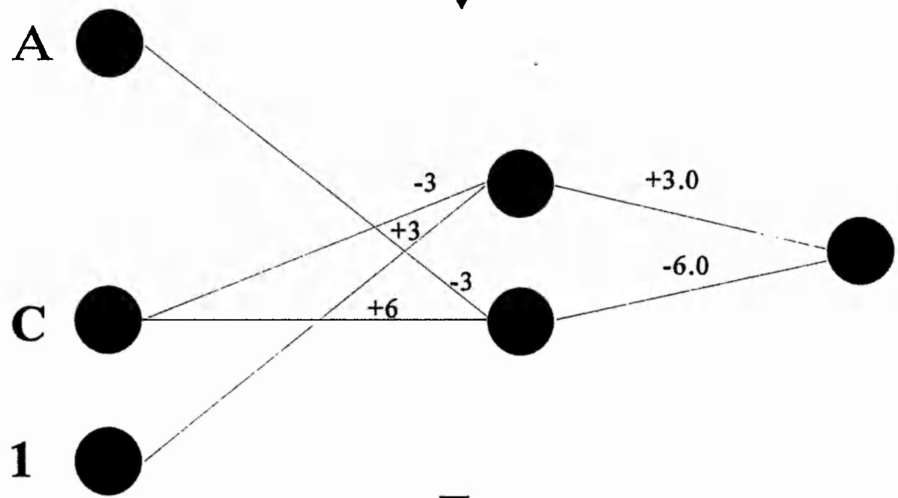
4. Test synthesised equation using appropriate partially connected ANN

5. Repeat with lower thresholds until partially connected ANN is sufficiently accurate

Initially, thresholds are set sufficiently high as to only include the most important input to the most important hidden unit. This was invariably too limited to solve the problem so the thresholds are reduced until the required accuracy is achieved. The resulting equation is a simplification of the network which acted as a simple model that could be extended to include more detailed information. This algorithm produces equations which are a simplified representation of the ANN, preserving the principle characteristics of the learnt model. The equation synthesis process can be represented by the following schematic diagram (fig. 4.3).



▼
Remove weak connections
▼



▼
Convert remaining weights to equations
▼

$$F(1-C)-2F(2C-A)$$

▼
Simplify
▼

$$F(1-C)+2F(A-2C)-2$$

Figure 4.3 A Schematic Diagram of the Equation Extraction Process

The final simplification step includes the transposition:

$$-F(-x)=F(x)-1$$

The relationship between F(x) and -F(-x) can be seen in the figure 4.4

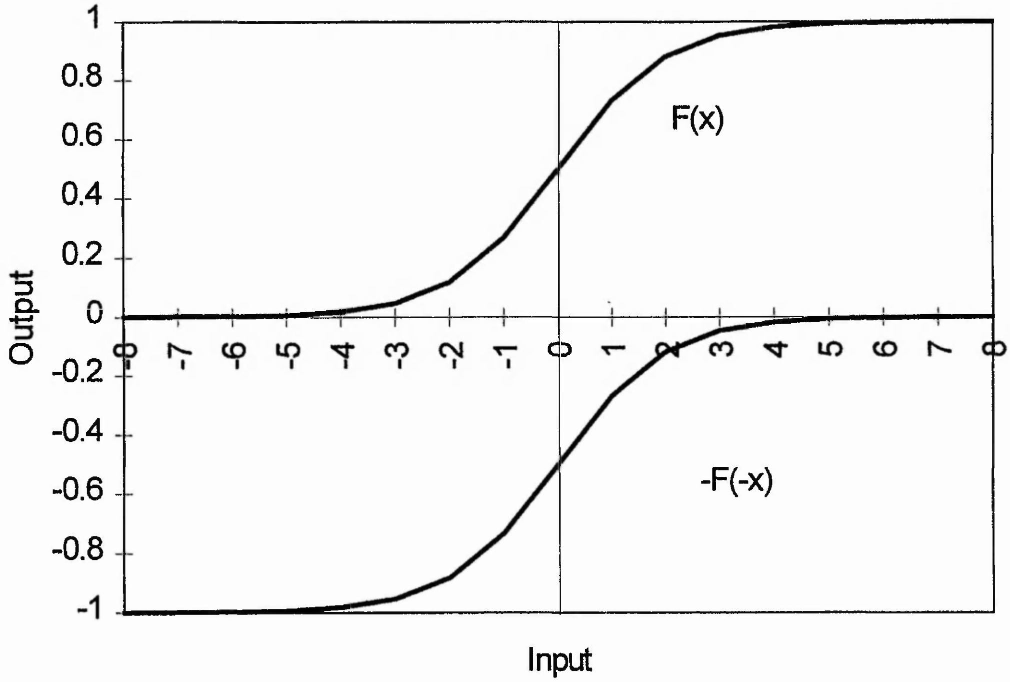


Figure 4.4. Graphical display of why -F(-x) = F(x) -1

The performance of partially connected networks was usually worse than the fully connected network, but the relationship between number of connections, described by terms in an equation, and performance was not linear (fig 4.5).

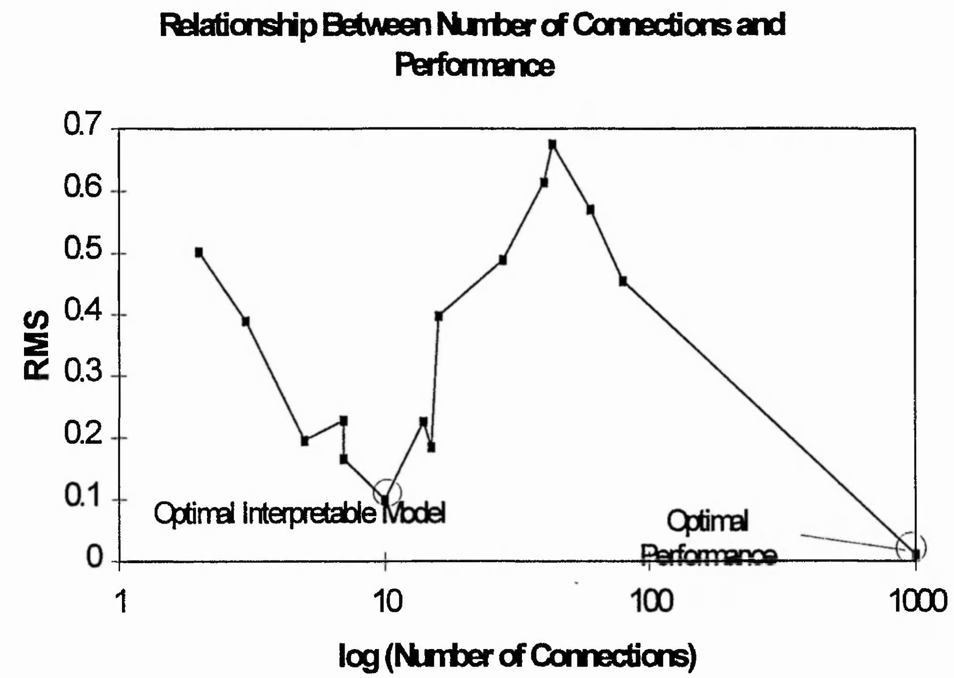


Figure 4.5. Non-linear relationship between degree of connectivity and performance.

This shows that increasing the number of terms in the equation did not necessarily increase the accuracy of predictions the associated ANN would make. One reason for this may be that beyond an essential number of generalising terms, the introduction of new inputs only brings in nodes that learnt pattern specific noise. It also underlines the importance of seeking a minimal model, ie. the simplest accurate model possible, since any unnecessary terms are extraneous and serve to both obfuscate the interpretations of the model and to increase the error.

4.4. Network Performances

Neural network performance evaluation and optimisation is a time consuming process, so it is important to set precise training and testing criterial and strive to achieve those goals. Performance needs to be evaluated in a careful manner, as it can be subjective, with the same network giving good or bad results depending on what results are being looked at (ie. Average error might be satisfactory but one sample might be completely misclassified.)

4.4.1 Use of all available data as inputs

The idea of one network providing prediction for two events (leaf damage and leaf number) was appealing since the events may be related and leaf number will be influenced by the health of the plant. Giving the network two outputs to predict meant the network optimised to the joint goal of minimising error for the combined outputs and this allowed poorer performance on the leaf damage prediction if it was suitably countered by a better performance in the leaf number prediction, or vice-versa. Satisfactory prediction was attained for Lit when the network was presented with 18 unseen patterns (fig 4.6) though there is certainly some error in this initial result.

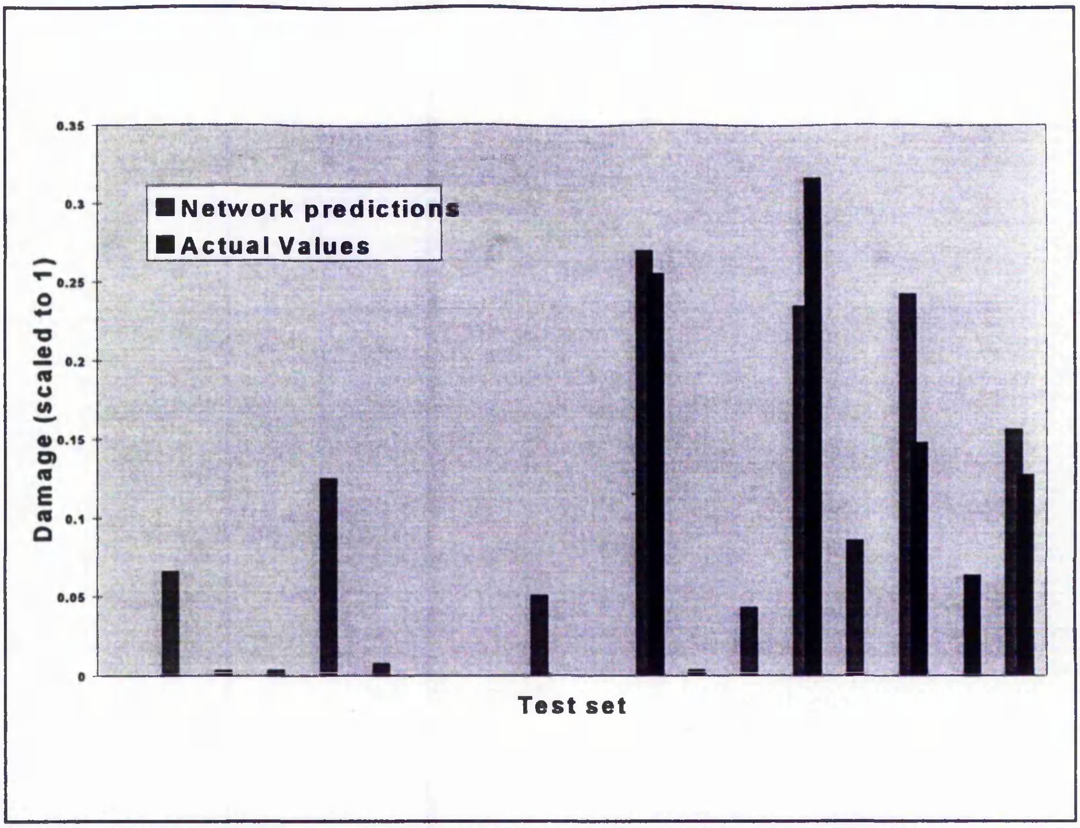


Figure 4.6. Lit. Network predictions for increase in injury for the test set

The network could predict, within a 10% error, 12 of the 13 occurrences of no change in leaf damage. At this point very little optimisation of the network model, such as locating the optimum derivative offset value, was attempted.

4.4.2 Effect of Differing Cultivars

Data was available for two cultivars, Lit and Nerina. Initially the network was much more accurate when predicting changes in injury for Lit than for Nerina. This was due to there being fewer vectors with positive outputs in the Nerina data set than in the Lit data set (Nerina is more resistant to ozone induced damage with the onset of ozone damage occurring at a later date than Lit (fig 4.7)).

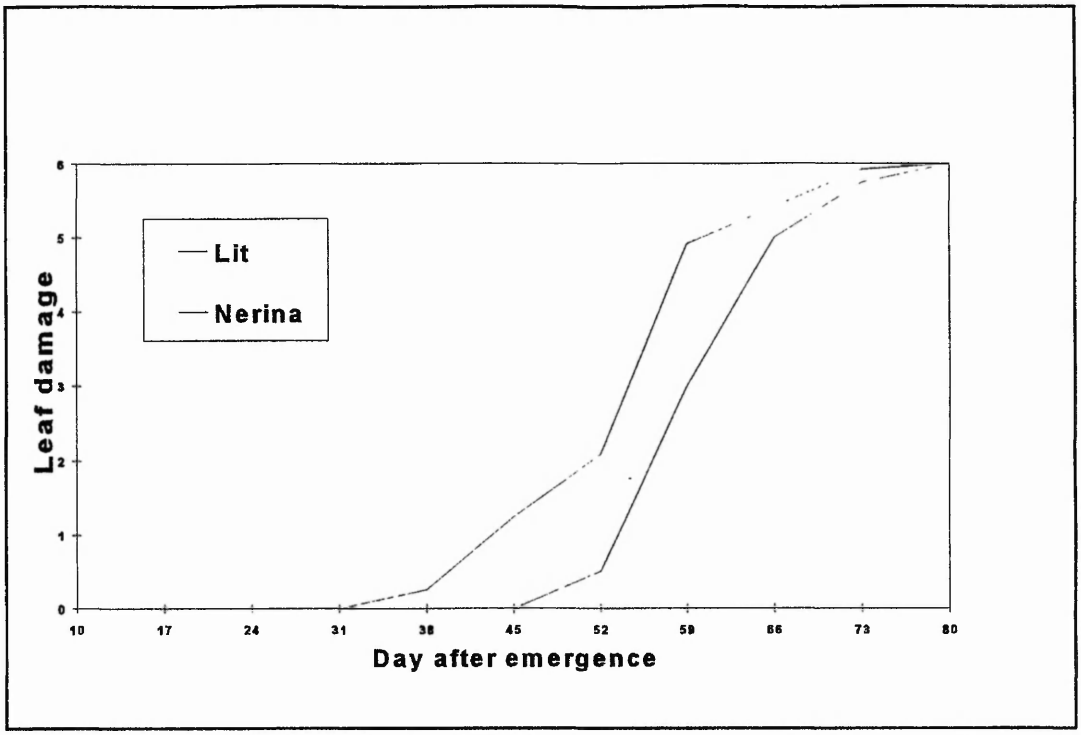


Figure 4.7. Injury expression for identical ozone doses for two bean cultivars

In an attempt to improve performance, the number of vectors with non-zero outputs was increased to a ratio of 50:50 by replicating suitable training set vectors. This was to make these vectors more important in order to discourage the network output from tending too often towards 0, which was happening. This was probably a result of 0 being the correct output for most vectors. Small improvements in the testing results for Nerina were noted, but not for Lit. This pre-processing is most important when the frequency of non-zero outputs is low and it will therefore be used only when appropriate. At this point, research into predicting damage changes in Nerina was dropped in favour of the more susceptible Lit cultivar.

Preliminary attempts were made to synthesise a simplified form of the complex equations which represent the trained network. The weights of the most successful network were taken and equation synthesis rules applied to find significant sets of inputs.

The network with the best performance yielded the following equation for prediction of ozone related leaf damage:

$$\begin{aligned}&2F[((2 * leaf\ damage\ on\ day0) - sumRad@d1 + sumOz@d2 \\ &+ sumOz@d4 + sumOz@d5 + \sum_{day5}^{day2} maxOz)] \\ &+ F[(leaf\ number\ on\ day0 - (maxNO@d3 + (2 * maxSO2@d3) + SumNO@d3))]\end{aligned}$$

F = a linear function between 0 and 1

The terms are of the form sumOz@d5 = sum of 24 hourly readings of ozone levels during day 5 and maxNO@d3 = maximum reading over the 24 hours of day 3. This equation contains the primary inputs for 2 hidden units, contained within an activation function (**F**), which is approximated as a straight line between 0 and 1, with a value of 0.5 at F(0).

The first part of the equation (the first F function) is derived from the hidden node with the largest weight and indicates a positive link between leaf damage and several days of higher than normal ozone levels (ie. more ozone, more damage). It also relies heavily on the level of damage at day 0 as a positive indicator of an increase in damage. The second line is derived from the hidden node with the second largest weight. It is more reliant on non ozone pollutants, pointing to a negative relationship (ie more nitric oxide, less damage). This may be linked to the following chemical equilibrium:

$$O_3+NO \rightleftarrows O_2+NO_2$$

So high levels of ozone may push the equilibrium to the right, turning ozone into oxygen but also removing nitric oxide from the atmosphere. This assumes the level of NO is governed by the equilibrium, whilst the level of ozone is primarily determined by other factors, such as sunlight and pollution.

This attachment of meaning to sections of the equation is an important aspect of the project. To have faith in a neural network's output, some of the rules behind its performance must be known. This is also essential in order to gain new insights into the mechanisms that give rise to the model.

4.4.3. Network Rationalisation

At this stage the network structure was changed in order to simplify the network's learning task by asking it to predict only the increase in the level of leaf damage, by far the most important event. Prediction of leaf number had been accurate but given that inputs such as growth stage were included, this was not surprising.

The influential role of the amount of leaf damage at day one was apparent in its high importance in equations generated. On one hand it might be important for the network to know the amount of leaf damage to decide if, and how much more, damage will occur and it should learn from training that leaf damage is a finite parameter e.g. a leaf with 90% damage cannot get a further 20%. But if the network is just using the fact that leaves with damage invariably get worse because they remain in a high ozone environment then it will be learning the experimental protocol and not forming rules on pollution-induced leaf damage. Two modifications were made to improve the integrity of the model:

a. Removing the input representing level of leaf damage at day zero. Results from this proved that satisfactory performance could be attained without the leaf damage input, although this was not as good as the results with the leaf damage at day zero included. It was also important to remove any data sets that began the 7 day period with severe (>50%) leaf injury, if the network was no longer given existing damage as an input it could predict more injury to already heavily damage leaves, predicting injury to areas of leaves that were already necrotic.

b. Implementing improvements gained by other changes to allow the network to learn more efficiently. One such change was the representation of the output as a percentage as opposed to an arbitrary scale.

The resulting predictive results were good for the training (fig 4.8) and test set (fig 4.9).

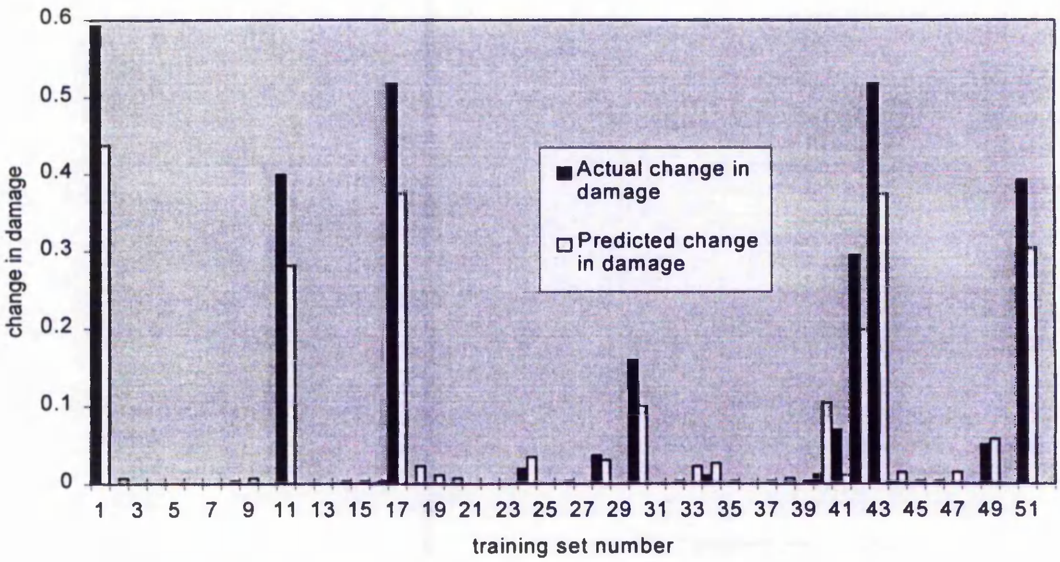


Figure 4.8. Training set damage prediction for modified network

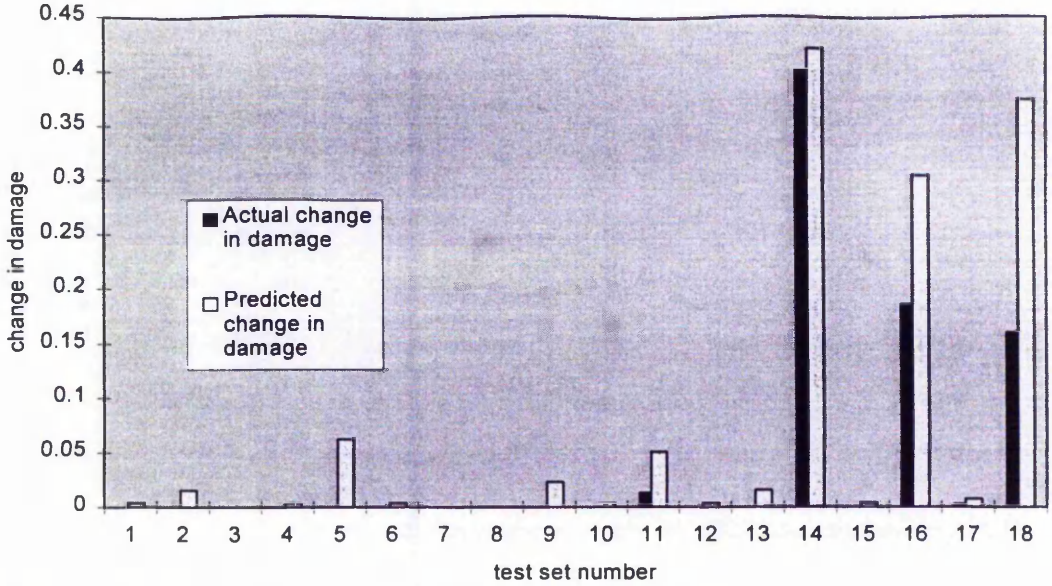


Figure 4.9. Test set damage prediction for modified network

These predictive results include a few quantitative errors:

a. The trained ANN tended to underestimate the amount of damage, though not by an important amount. The reason for this could be the disproportionate number of '0 output' patterns in the training set. During training, errors on the underestimating side have a lesser impact on the total error so may be more tolerated if the overall error is minimised. Replication of non-negative output patterns did not reduce this problem..

b. Predicted damage in the test set was overestimated for patterns 16 and 18. This was due to the nature of the damage in these two 7 day data sets. Some damage had already occurred on these leaves and the predicted damage may have been more accurate if the leaves were undamaged at the beginning of the 7 days. So at the beginning of the 7 days only about 70% of the leaf was available to be damaged. This problem could be eased by modifying the output to take into account the level of damage already present on the leaves. This could be achieved by changing the output from 'Change in percentage of leaf area damaged over 7 days ' to 'Change in percentage of available leaf area damaged over 7 days'.

The weights of this network were used to synthesise an equation:

Change in percentage of leaf area damaged over 7 days =

F [D2(SumOz+MaxOz) - 2*(D1SumRad) - D1MaxRad - D2MaxRad + D3(SumOz+MaxOz) + D4(SumOz+MaxOz) +D5 (SumOz+MaxOz) - D6MaxRad + D7(MaxOz - SumOz)] +

F [Leaf Age] - 1

The terms in this equation are environmental or pollutant factor levels eg. Age of plant at day 0 (Leaf Age), highest ozone reading on day 3 (D3(MaxOz)), sum of hourly readings of solar radiation during day 7 (D7(SumRad).

This illustrates the importance of ozone in the production of injury, and also the greater effect of ozone associated with low light levels. One possible reason for this is the effect of light on stomatal opening. Very high light levels cause stomatal closure, which allows less ozone to pass beyond the protective epidermal layer, whereas lower levels of light can increase stomatal opening. The equation also shows a general trend that older leaves are injured more than younger leaves. This observation is well documented (Guzy and Heath 1994, Karlsson *et al* 1995) and the equation can be written heuristically as follows:

Change in percentage of leaf area damaged over 7 days =

F[Ozone - Light] + F[Leaf age]

Since the equation includes only three of the possible 7 inputs, it shows that a smaller network can be used. Several smaller networks were tested using combinations of these three input types. Although a small degradation in performance was witnessed (From an RMS of 0.06 to 0.1), an ozone only network was capable of predicting damage changes for unseen test patterns to a reasonable accuracy. The equation synthesised from one of these networks showed the relative merits of using sum or maximum ozone to predict damage changes.

F [D1 MaxOz + D3 Max Oz - (2*D5SumOz) + D6 MaxOz] +

F [D4 MaxOz - D3 SumOz + D4 MaxOz +D5MaxOz - D6 SumOz + D7MaxOz - D7 SumOz]

A good simplification of this found equation is:

Change in percentage of leaf area damaged over 7 days =

F[Max Ozone - Sum Ozone] = F[sharpness of daily peaks]

Since only ozone factors are presented to the network, some of them are likely to assume a negative value. Here the daily sum ozone levels take the negative value. So if two days have the same sum ozone levels (fig. 4.10) and one day has a higher maximum ozone level (solid line) than another (dotted line) this will lead to a greater increase in damage. This agrees with findings by other researchers (Musselman *et al* 1986, Karlsson *et al* 1995).

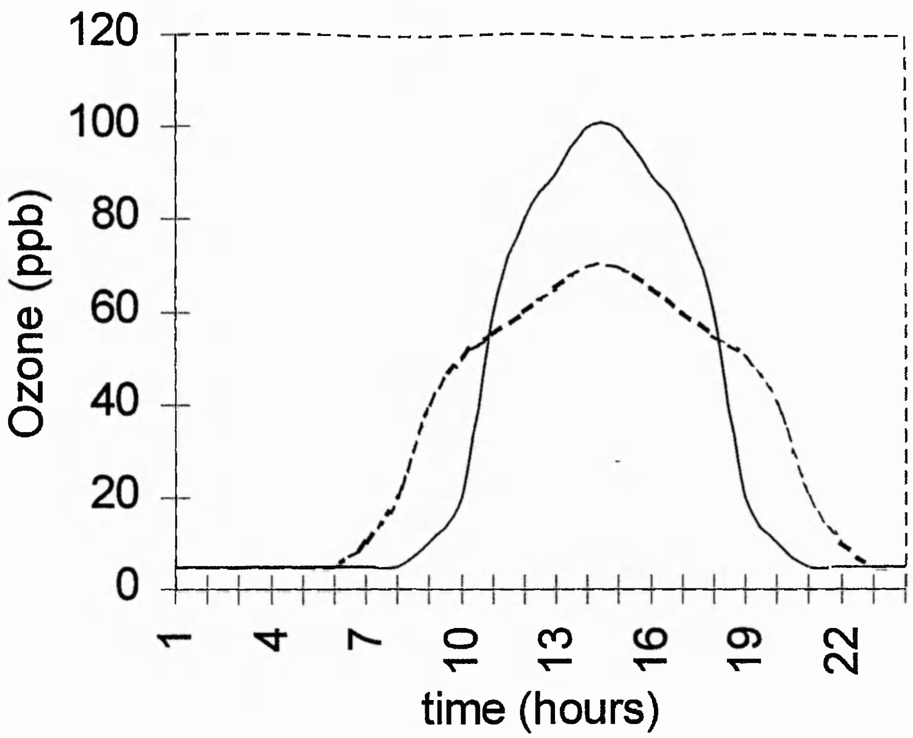


Figure 4.10 A comparison of two ozone episodes with identical ppb.h

4.5. Discussion

The modelling work carried out so far shows that meaningful equations synthesised from trained ANNs can point to important, previously unnoticed, factors in the data. These factors, or combination of factors would otherwise go unnoticed due to the high volume of multivariate data that has to be examined. This is a fundamental concept for any application of ANNs to any real world problem space. Real world relationships are seldom simple linear 'cause and effect' ones.

The primary or secondary nature and the interrelatedness of the pollutants is a key to the problems associated with damage prediction. Looking for causes of effects involves analysis of multidimensional causal relationships. A summary table is provided with some of the results and the relevance of these results (Table 2.1)

While early results made use of some of the primary agents of ozone pollution (eg. NO_2) this seemed over complex and therefore an unlikely approach for an ANN to take. Modifications to the inputs making up vectors in the training set (change in representation of desired output for arbitrary to percentage, and the removal of unimportant or misleading data) produced networks that yielded more meaningful equations.

The appearance of solar radiation in these equations is not unexpected but the fact that it is shown to be a negative indicator of damage, in the presence of ozone needs explaining. When an equation contains the combination of:

"Ozone level factors - solar radiation level factors = Injury"

This is saying that conditions of relatively high ozone and relatively low light are more damaging than both high ozone and high light and low ozone and low light. There are shown to be two important roles of solar radiation in the expression of ozone related leaf injury.

1. As an important catalyst for the production of ozone from primary pollutants (Photochemical Review Group (PORG) 1993). The networks would use this as a positive indicator of damage if ozone levels were not available but as they are available there is no need to use light in this way if ozone levels are already included within the equation.

Architecture	Test Set Performance	Equation summary	Points answered	Points Raised
Dual output	RMS=0.054 Max error = 0.125	Initial Damage + ozone + (leaf age +other pollutants) = amount of injury	Dual output prediction feasible.	Dual output architecture may not be best method for predicting injury
Lit vs Nerina	Nerina{RMS =0.08. Max error =0.23} Lit{RMS = 0.06, Max error =0.126}		More difficult to predict injury on more resistant cultivars, because of fewer cases of injury	
No 'Damage Day1'	RMS= 0.06, Max error = 0.213	(Ozone -Light) + leaf age = amount of injury	Good performance could be achieved without 'damage day 1'	Could equal performance be achieved using only oz,day and light.
No 'other pollutants'	RMS= 0.05, Max error = 0.15	Ozone -Light = amount of injury	Other pollutants are non-essential for the prediction of injury	Would an 'ozone only' network predict to a sufficient standard.
Ozone only	RMS= 0.06, Max error= 0.17	Max - Sum = injury	Acceptable performance using just ozone parameters	

Table 4.1 Summary of ANN results.

2. As an important factor in controlling the dimensions of stomatal aperture (Kappen and Haeger, 1991). As ozone enters the leaf through the stomata then their aperture

is of obvious importance. At high levels of light stomata may close to reduce the loss of water by transpiration which inadvertently forms a barrier to the influx of ozone.

It has been shown that an increases in ozone concentrations from day to day is a more important indicator of an increase in ozone damage than is the ozone dose *per se*. This may be explained by the biological defence mechanisms used by plants to protect themselves against oxidative attack. Stomatal conductance and anti-oxidative chemical production are two methods used by plants to reduce the harmful effects of an ozone episode but both require time to work, so sudden increases in ozone levels may have a more damaging effect than steady levels. This has important implications in the prevention of ozone damage:

1. If there was some chemical way of smoothing the peaks and troughs of ozone episodes this may be effectual in reducing levels of crop damage. This could be in the macro-chemical approach of smoothing the causes of ozone increase (eg. Smoothing of traffic flow peaks)

2. There may be a biological approach to smoothing the ozone peaks. Strains of crop with smaller stomatal apatures would restrict the flow of ozone into the leaf as the levels increase. So changes experienced inside the leaf would be less sudden (Figure 4.11). This figure is a hypothesised representation of the effect of smaller stomata apature but it is based on the simple assumption that smaller diameter apatures allow less gas flow/mixing so the rate of change inside the leaf would be dependant on apature size. It shows that when peak increases outside the leaf are 80 ppb/hr increases inside are reduced to less than 80ppb. With this theoretical ozone episode the maximum ozone levels are the same inside the leaf as outside, but if the peak ozone levels were for short periods smaller stomatal size may mean that external maximums are never reached.

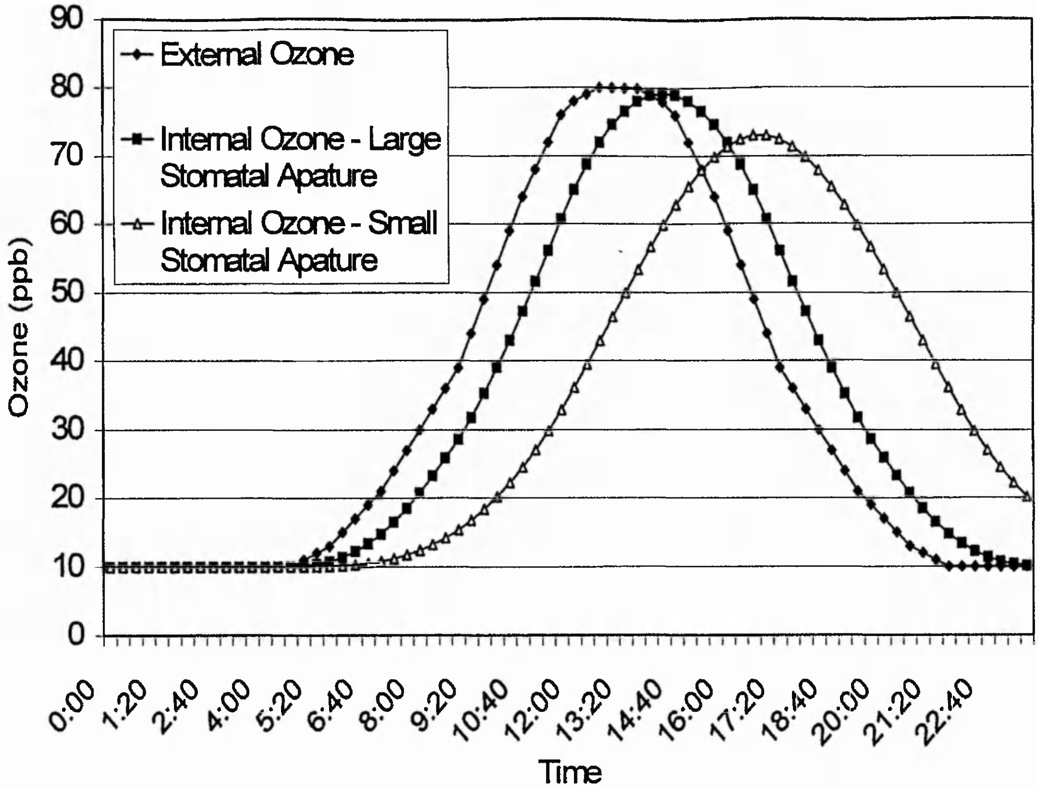


Figure 4.11. Theoretical effect of smaller apature stomata on change in internal ozone levels.

While the careful selection and formatting of the data is time consuming and involves some degree of trial and error, it has proved to be essential to make results more accurate and more meaningful. While ANN's are excellent pattern recognisors the temptation to just throw data at them and expect accurate predictions must be resisted. Thought must be given to the inputs and equation synthesis allows for the informed removal of less important inputs. The compact equations generated by the incremental nature of equation synthesis help in the structuring of the input data set. This feedback of results is a continued theme throughout this thesis and is an example of how meaning extracted is not just valuable for the facts it gives you about the system but also as a guide for further research.

Chapter 5. Predicting the Occurrence of an Event.

By definition, events happen! Given a fixed time period an event will or will not happen. Bookmakers give odds on these events happening (ie. Alien landing this year, 10 000-1, or a white Christmas, 8-1). The events given are very difficult to predict, the latter due to the fractal nature of weather (ie the butterfly effect), the former due to the lack of reliable data in its favour. In a deterministic world everything is predictable, a dice roll is predictable given infinitely accurate measurements of all the variables involved (angle of throw, trajectory, velocity etc). The 2 main problems involved in predicting 'everything' are:

1. Perfect measurements. Real world prediction problems are usually sensitive to initial conditions, tiny changes in these initial variables can have major changes in the resulting outcome (ie. The butterfly effect).

2. Choosing the correct variables. For each event, every variable involved in the outcome must be included in any predictive equation. Knowing these without knowing the mechanism of the cause/effect relationship may be impossible; knowing the mechanism of effect precisely would mean no modelling was necessary.

In this chapter the problem of predicting an event is addressed, using our pollution modelling paradigm to supply the desired cause/effect relationship.

5.1. Introduction to ICP-Crops

The ICP-crops was established in 1988 under the LRTAP convention. Its purpose is to quantify the impact of ozone on non-wood species in the UN/ECE area. Since 1992, the ICP-crops has been co-ordinated from the Nottingham Trent University. In previous years the data had been analysed in a more conventional, statistical way. A provisional short-term critical level of ozone for the development of visible injury was set (Fuhrer and Achermann, 1994) and modified (Benton *et al* 1996a). The first critical level was a level 1 approach, assigning a single value to protect the most sensitive known receptor (Sanders *et al*, 1995) and was set at an accumulated ozone dose above a threshold of 40ppb (AOT40) of 700 ppb.h accumulated during daylight hours (global radiation ≥ 50 Wm^{-2}) over three consecutive days. The modification of this (Benton *et al*, 1996a) is closer to a level 2 approach, considering factors causing variation in the sensitivity of the individual receptors, in particular VPD and relative humidity levels. The modification of the initial critical level was made because data collected during 1995 showed that injury sometimes occurs before the 3 day, so a 700 ppb.h level was reached(Benton *et al* 1996a). This was probably due to the modifying factors that influenced plant responses to ozone prior to or during the ozone episode, such as climatic conditions. It has also been described how injury did not always occur after the proposed critical level of an AOT40 exceeding 700 ppb.h over 3 days (Benton *et al* 1996a). The following work led to the revised definition of the short term critical level for injury and is published by Benton *et al* 1996b.

5.2. Data Preparation and Analysis

The data used to train and test ANNs was collected at many European sites as part of the ICP-Crops programme. Countries that participated in 1994 and/or 1995 were

Austria, Belgium, Denmark, Finland, France, Germany, Hungary, Italy, Latvia, The Netherlands, Poland, Russian Federation, Slovenia, Spain, Sweden, Switzerland and The United Kingdom. Each participant conducts a series of coordinated experiments each year using a standard protocol (UN/ECE 1996) and seeds supplied by NTU. This data collection is ongoing, this means that any networks developed and tested using 1 or 2 years values can be further modified or reinforced by adding further yearly data. The training with data from many European sites over many years makes for exceptionally robust predictive networks. Initially this data was suitable for the prediction of the date of the onset of injury using ozone levels as predictors but climatic data was added to the model as it became available.

The hourly means for parameters were compressed into meaningful daily indicators: 7 hour mean, daily maximum, accumulation over threshold 40ppb (AOT40) and/or 24 hour mean. Ozone-only networks gave information on the relative merits of the ozone indicators, and on the relative importance of the days preceding injury. Addition of climatic data helped define the effects of modifying factors on ozone injury expression.

In 1994 and 1995, participants of the UN/ECE ICP-Crops programme were asked to note and record when injury was first observed on clover (*Trifolium subterraneum* and *T. repens*) and bean (*Phaseolus vulgaris*). Clover results were used initially as they provided the largest set of information and also clover is a good indicator of ozone injury (Karlsson *et al* 1995, Becker *et al*, 1989)). The ICP-crops data from 1994 was examined and the 3 ozone indicators for the 3 or 5 day periods prior to injury expression were used, along with 'days after emergence/harvest', as inputs for a network.

In an effort to visualise and understand the data that the ANN would be working on the variation in ozone level preceding injury can be plotted (fig. 5.1a and 5.1b). It can be seen that ozone levels vary quite considerably during the duration of an experiment. As weather conditions have a very important effect on Ozone level, this daily variation is perhaps not too surprising. Expression of leaf injury is also presented on these graphs either implicitly (fig 5.1a) or explicitly (fig 5.1b)

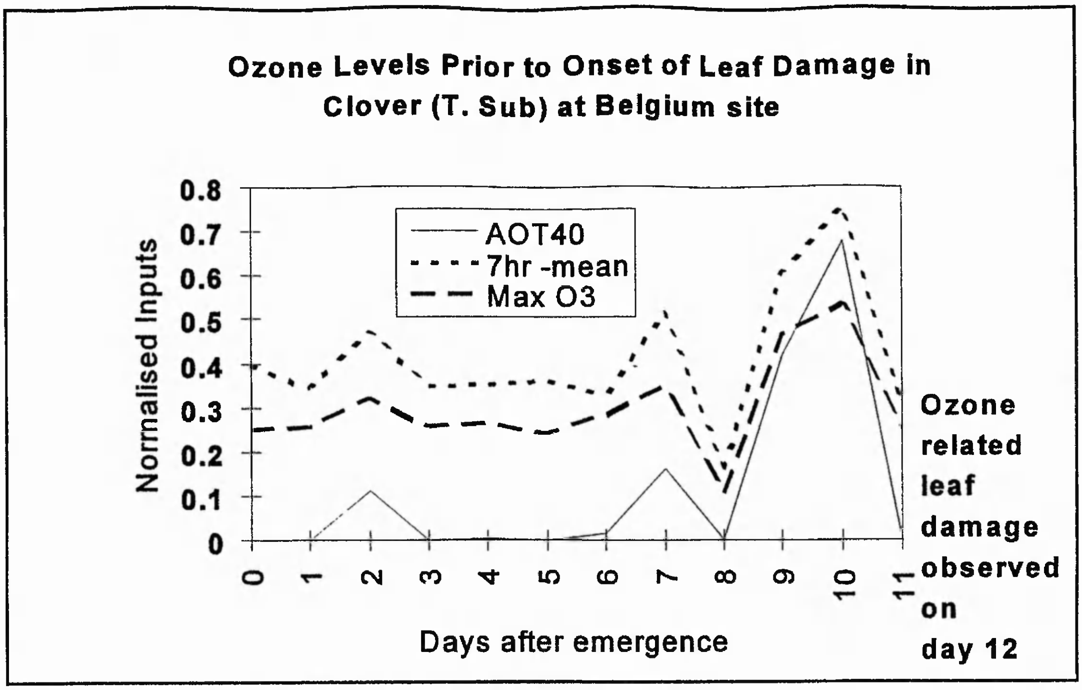


Figure 5.1a. Ozone levels preceding injury

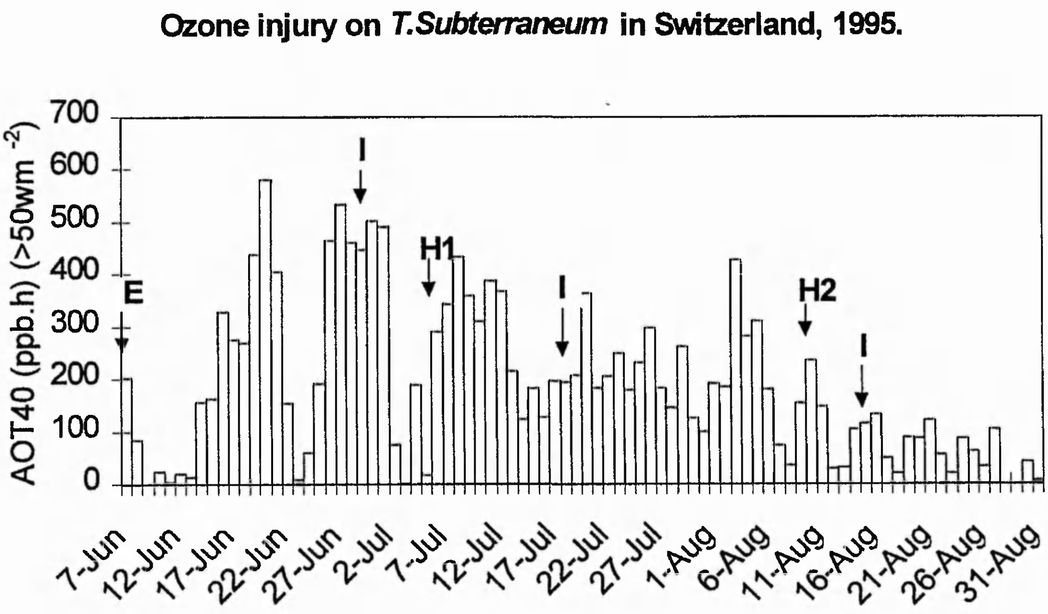


Figure 5.1b AOT40 levels over entire growth period. E = emergence, H = harvest, I = injury.

5.3. Network Architecture

The architecture of the simplest of these ANNs is shown (fig. 5.2).

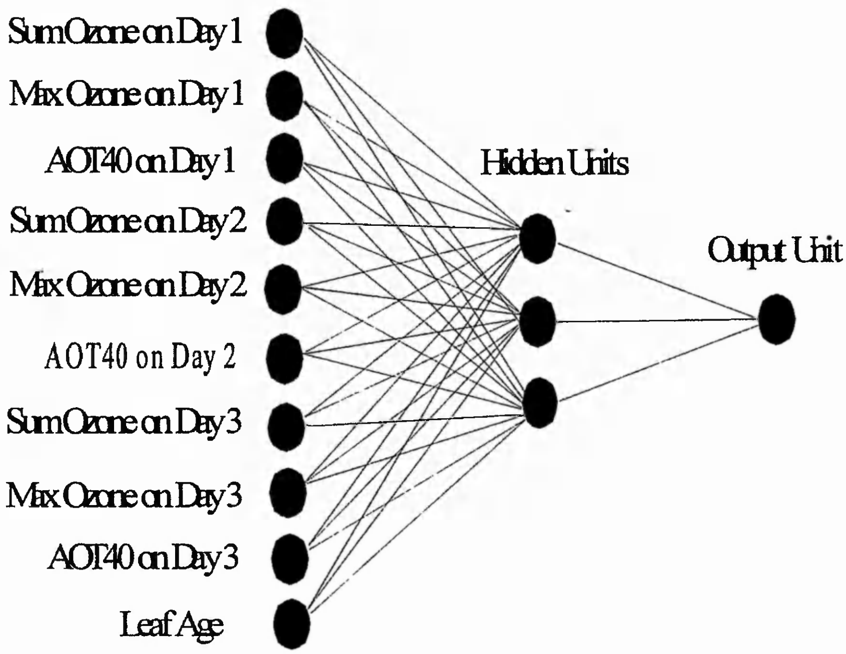


Figure 5.2 Architecture of basic ANN

This ANN used has 10 input units, each deriving its value from ozone levels or the age of the leaves in days. The 3 hidden units are feature detectors, differences in weight distribution on the connections mean that each node receives the data with different emphases due to the different weights. The output node takes the results of these nodes and gives a sum output, a decision on the question asked. More advanced networks contained inputs for environmental factors (relative humidity, temperature and solar radiation). These factors are known to modify the damaging effect of ozone (Mortensen 1992), but the extent of this modification is unclear.

5.4. Network Performances

Judging a network's performance, when the desired output is a binary event, is easier than most evaluation scenarios, but still leaves some room for the modeller to add value to his results. Deciding what category borderline outputs fall into must be applied fairly to all outputs. Examples of output to evaluation rules include:

Greater than 0.5 = true less than 0.5 = false.

Greater than 0.666667 = true, less than 0.33333 = false, otherwise count as 'don't know'.

A threshold could also be chosen that equally divides all outputs (ie, a value of 0.56 might make 50% of the outputs true and 50% false).

There is also some choice as to what the desired output should be for a true or a false. 0.9 and 0.1 are the most usual used, these 'less than unity' figures give the network to correct any over fitting of the data.

The test case pollution impact modelling network tried to predict the onset of leaf injury. A degree of success was achieved in predicting the onset of injury. The ability of an ANN to predict the desired output is contained in the weights of the connections. Equations were synthesised from these weights using the previously described algorithm.

From the weights of the ANN defined in figure 5.2, the following equation was derived.

Occurrence of injury =

F[Day2(7hmean + Max) - Day3(AOT40+Max0zone)] +

F[Day2(MaxOzone-AOT40) - Day1(AOT40 + 7hmean) + day3(7hmean)] - 1

The terms in this equation are pollutant factor levels eg. Mean of hourly readings of ozone levels between 10am and 5pm during day 2 is **Day2(7hmean).**

This equation contains the primary inputs for 2 hidden units, contained within an activation function **(F)**, which can be approximated as a straight line between 0 and 1, with a value of 0.5 at F(0).

The first, and most influential, node shows that a fall from high to low levels of ozone precedes the onset of injury. This is apparent because ozone levels on day three are used in a negative way , ie higher levels of ozone give a lower output. Ideal conditions for a positive prediction of injury are therefore high levels of ozone on day 2 and low levels on day 3. This equation can be further simplified to:

Occurrence of injury = F[rise in ozone levels] + F[fall in ozone levels]

This conclusion was confirmed by equations synthesised from ANN trained on a larger training set that included data from the 1995 growing season. A visual examination of many graphs of ozone levels and injury agree with this conclusion (see figures 5.1a and 5.1b)

This finding was further investigated with greenhouse experiments which revealed new insights into the expression of injury in clover leaves and is discussed later.

An uneven profile of ozone that caused damage is shown below (Figure 5.3a):

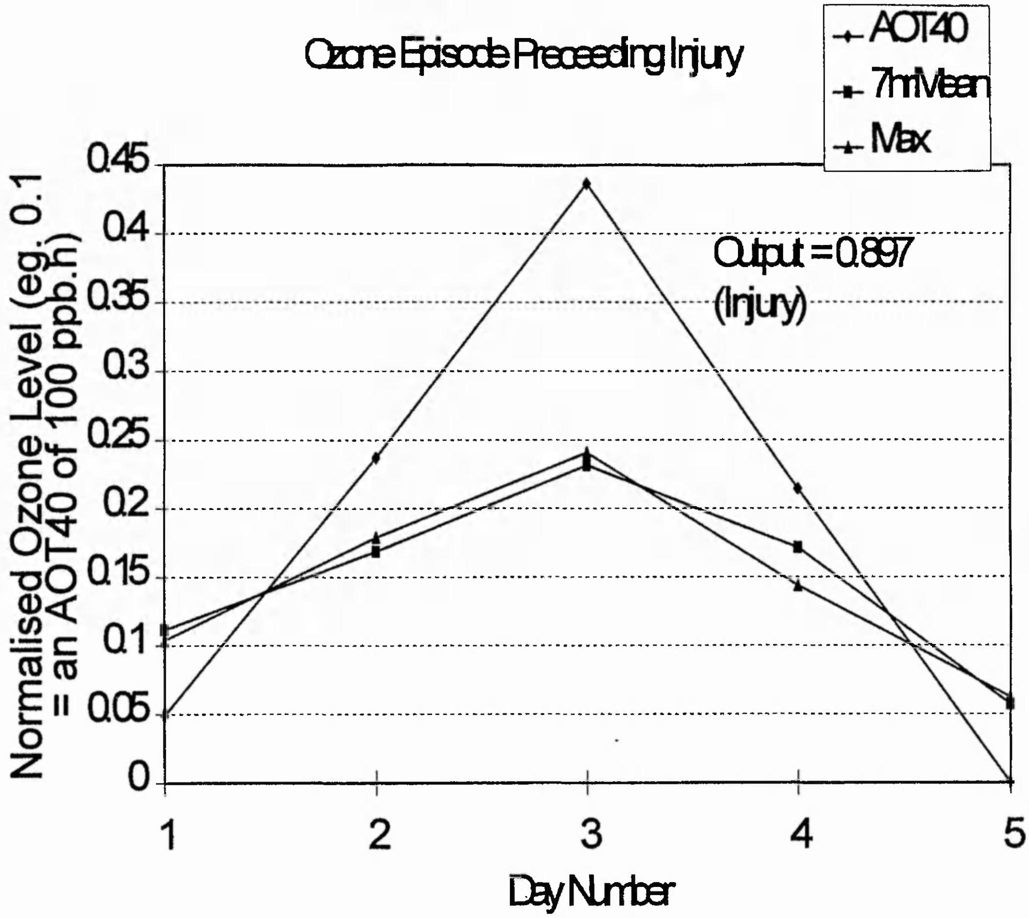


Figure 5.3a Ozone exposure preceding injury.

A 5 day profile of a damage causing ozone exposure was averaged and then entered into the trained network (Figure 5.3b). The resulting output was less than 0.5, indicating that the ANN thought that no injury would occur after this theoretical exposure.

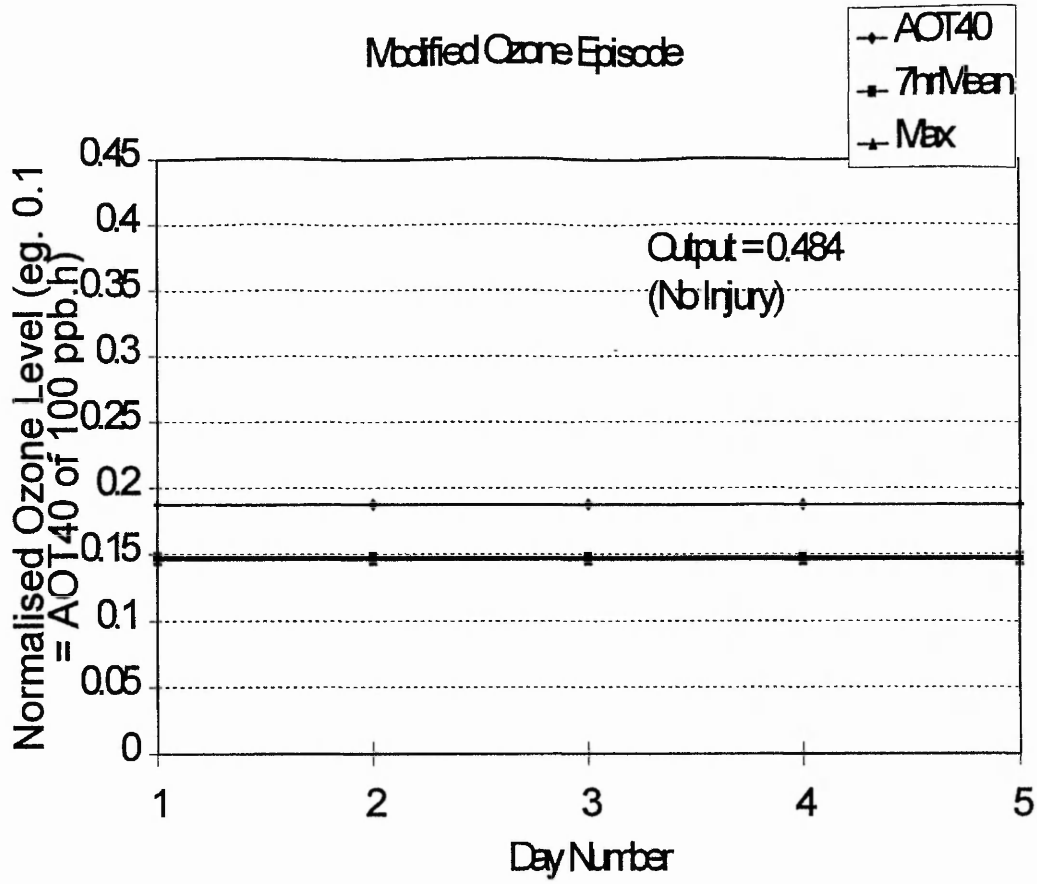


Figure 5.3b. Same net ozone exposure averaged over 5 days.

It can be concluded that the 'peakiness' of an ozone exposure is an important factor in damage development.

The 1995 data was comprehensive enough to train networks to predict injury using climatic data as well as ozone. The inputs of the ANN accepted the following parameters for each of the five days preceding injury: Ozone (AOT40, 7 hour mean, daily maximum), Temperature (24 hour mean, 7 hour mean, daily maximum), light (24 hour mean, 7 hour mean, daily maximum), solar radiation (24 hour mean, 7 hour mean, daily maximum) and number of days after emergence. This network trained to a high predictive accuracy (fig. 5.4) and generalised enough to be able to accurately predict onset of injury for a small test set of inputs (fig. 5.5).

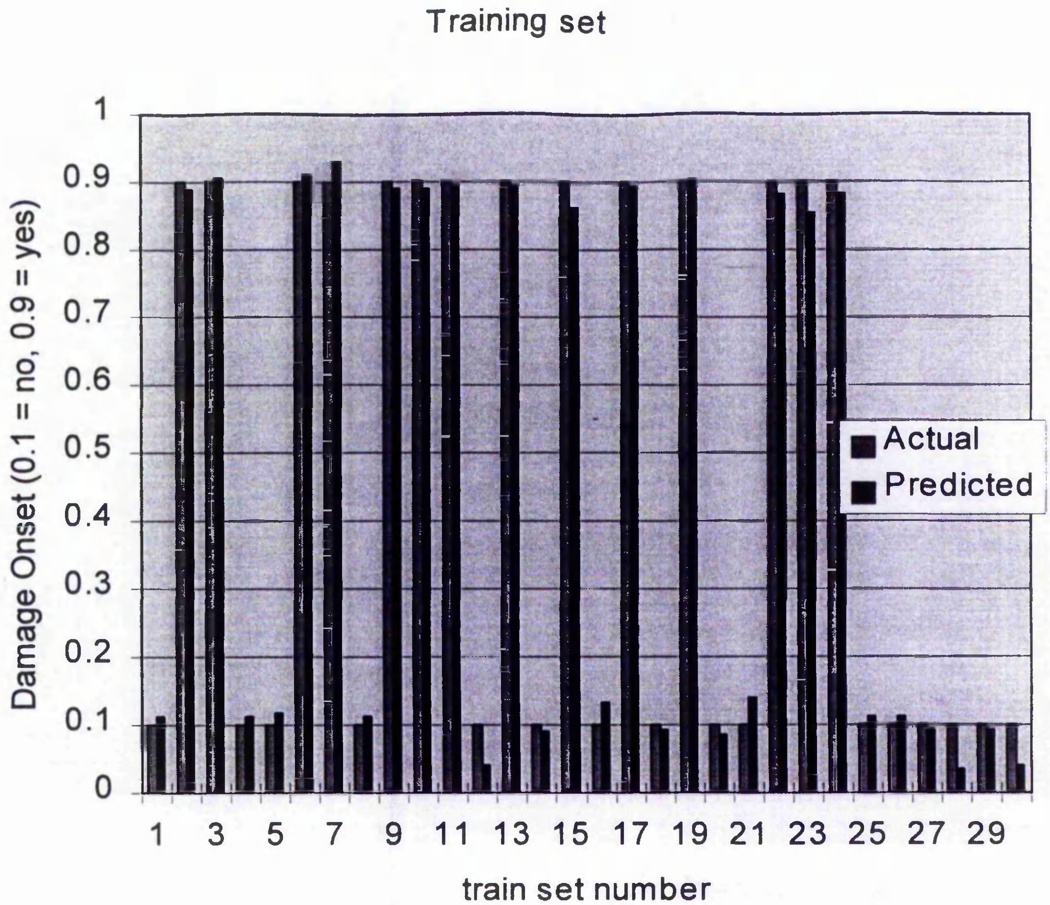


Figure 5.4. Training accuracy of ANN trained on ozone plus climatic data

In these graphs the output from the ANN is on the Y axis and is compared with the desired output. If injury was expressed after the 5 days presented to the network then the desired outputs 0.9; if injury did not occur then the desired output is 0.1. The network gives an output that is trained to be as close as possible to the desired output. It is true to say that the predicted output increases with the ANN's increasing confidence that injury will be expressed.

An equation can be synthesised from this ANN and although it is a considerable simplification of the 204 weighted connections of the whole network, it remains both long and multifaceted:

Occurrence of injury =

F[Day2(MaxLight-AOT40) -Day1(7hrmean RH) - Day3(AOT40 + 7hrmeanO3) - Day4(AOT40) +Day5(24hrmeanLight+Maxlight)]

+F[Day3(MaxOzone) - Day1(7hrmeanRH) - Day 4(7hrmeanRH+MaxRH) - Day5(AOT40 +7hrmeanOzone +7hrmean RH)]

+F[Day2(24hourRH+24hourLight)+Day3(AOT40+7hrmeanOzone+MaxOzone) - Day5(AOT40) - Day1(MaxRh)]

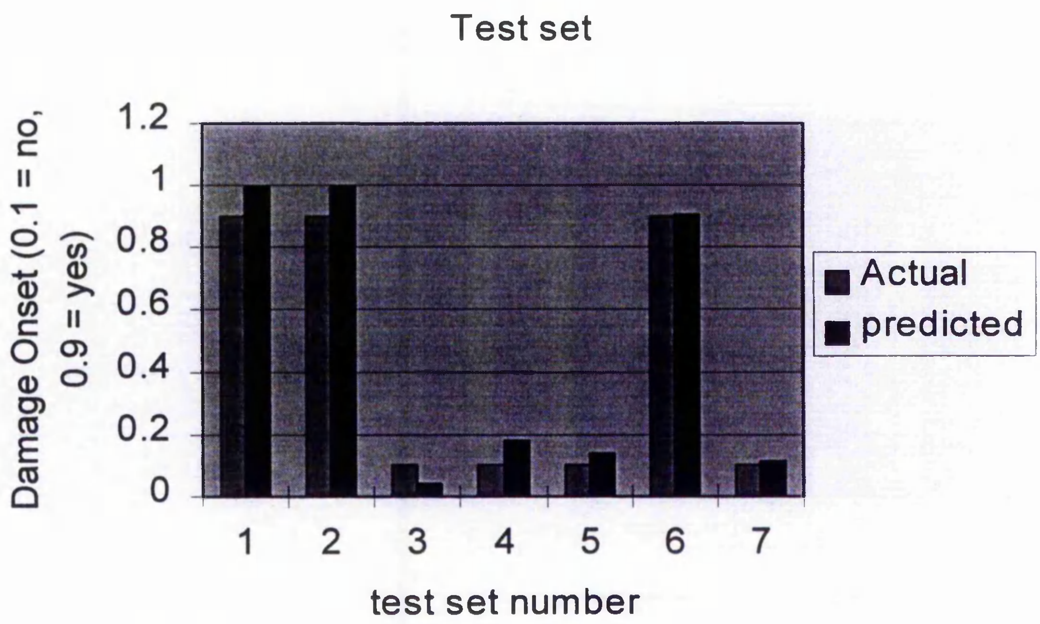


Figure 5.5. Testing accuracy of ANN trained on ozone plus climatic data

This equation excludes temperature from any role as a modifying factor and points to influential roles for humidity and light levels. This can be seen by the appearance of these factors, alongside ozone. They appear as both positive and negative contributors to the output suggesting a complex influence. It still appears that a fall in ozone levels precedes onset of necrosis by one day. The ozone humidity relationship remains a complex one even within this simplified equation. This is demonstrated when the network is presented with a range of artificially generated ozone and humidity combinations (fig. 5.6). The non-linearity of the data is sufficiently represented by the undulating curves of the model's response to a range of input values. A three dimensional output enforced some limitations on the complexity of this exercise, identical values for each of the days and each of the three parameter variables were used to test the networks predicted output for 'ozone' and 'RH' conditions. If the training data was averaged this way its range would be seen as quite limited (fig. 5.7) so the 3D surface in figure 5.6 involves some extrapolation by the ANN as well as interpolation.

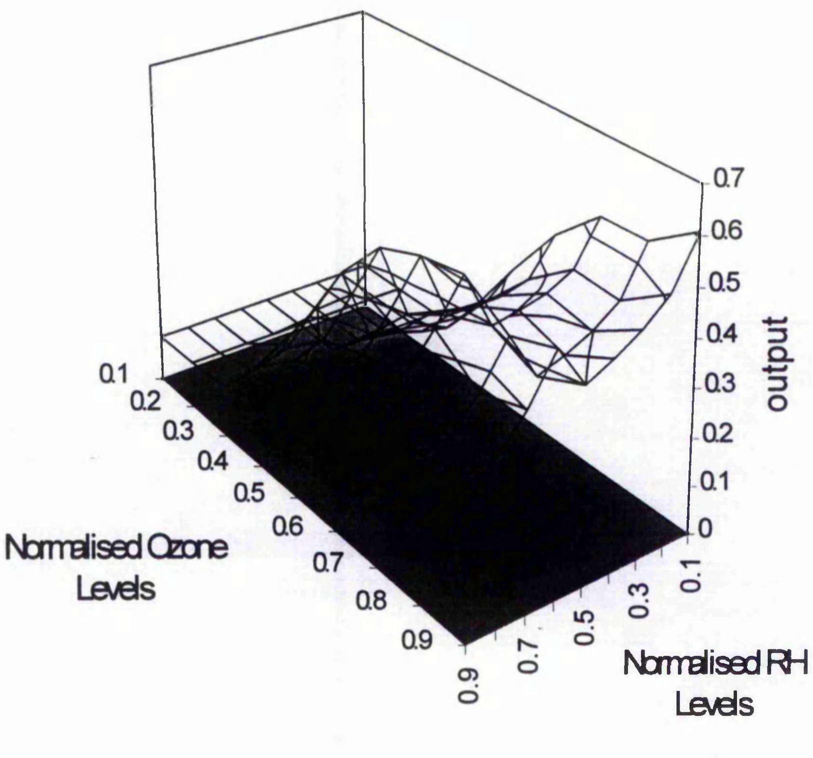


Figure 5.6 Response of a Trained Network to a Complete Range of Inputs

An overall impression of the relative importance of the different factors can be gained from a simple analysis of the combined weights that correspond to each input to the model. Results from this are shown in table 5.1. Network 1 trained with ozone, global radiation, VPD and day number indicates that ozone (AOT40, 7 hr mean, daily maximum), global radiation (daily maximum) and VPD (7 hr mean) are important inputs for injury development as the weights have values of 1.92, 1.69, 1.17, 1.43 and 1.12 respectively. Network 2 also gave significant importance to these parameters but a lower importance to relative humidity and temperature, VPD may appear less significant as both are included in a VPD calculation. Although this method of meaning extraction can be criticised because it attempt to linearise a non linear model, the results do concur with the only other work in this field (Balls *et al* 1995, Balls *et al* 1996) in which a similar hierarchy of importance was found.

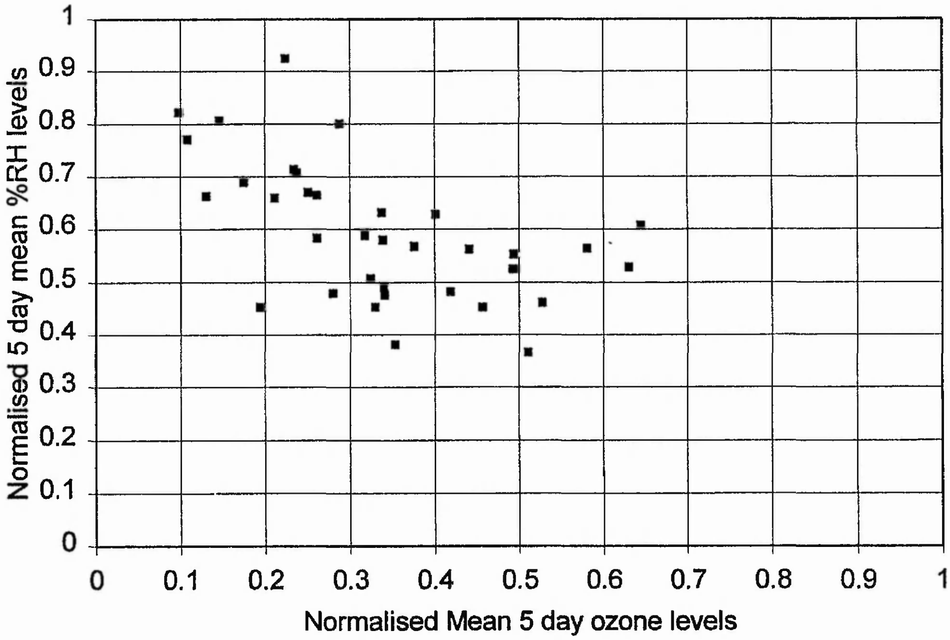


Figure 5.7 Range of training data when averaged.

	Network 1	Network 2
Parameter	Relative Importance	Relative Importance
Ozone (AOT40)	1.92	2.58
Ozone (7 hr mean)	1.69	1.86
Ozone (daily maximum)	1.17	1.44
RH (24hr mean)		0.75
RH (7 hr mean)		1.86
RH (daily maximum)		1.44
Global radiation (24 hr mean)	0.88	1.23
Global radiation (7 hr mean)	1.03	1.44
Global radiation (daily maximum)	1.43	1.87
Day number	0.62	0.8
Temperature (24 hr mean)		1.11
Temperature (7 hr mean)		1.12
Temperature (daily maximum)		1.29
VPD (24hr mean)	0.75	0.65
VPD (7hr mean)	1.12	1.36
VPD (daily maximum)	0.94	1.22

Table 5.1. Relative importance of factors for two damage predicting ANNs

5.5. Activation Based Grouping

A theoretical example of how equation synthesis and hidden unit activation monitoring helps to distinguish leaf injury scenarios is given below (fig 5.8).

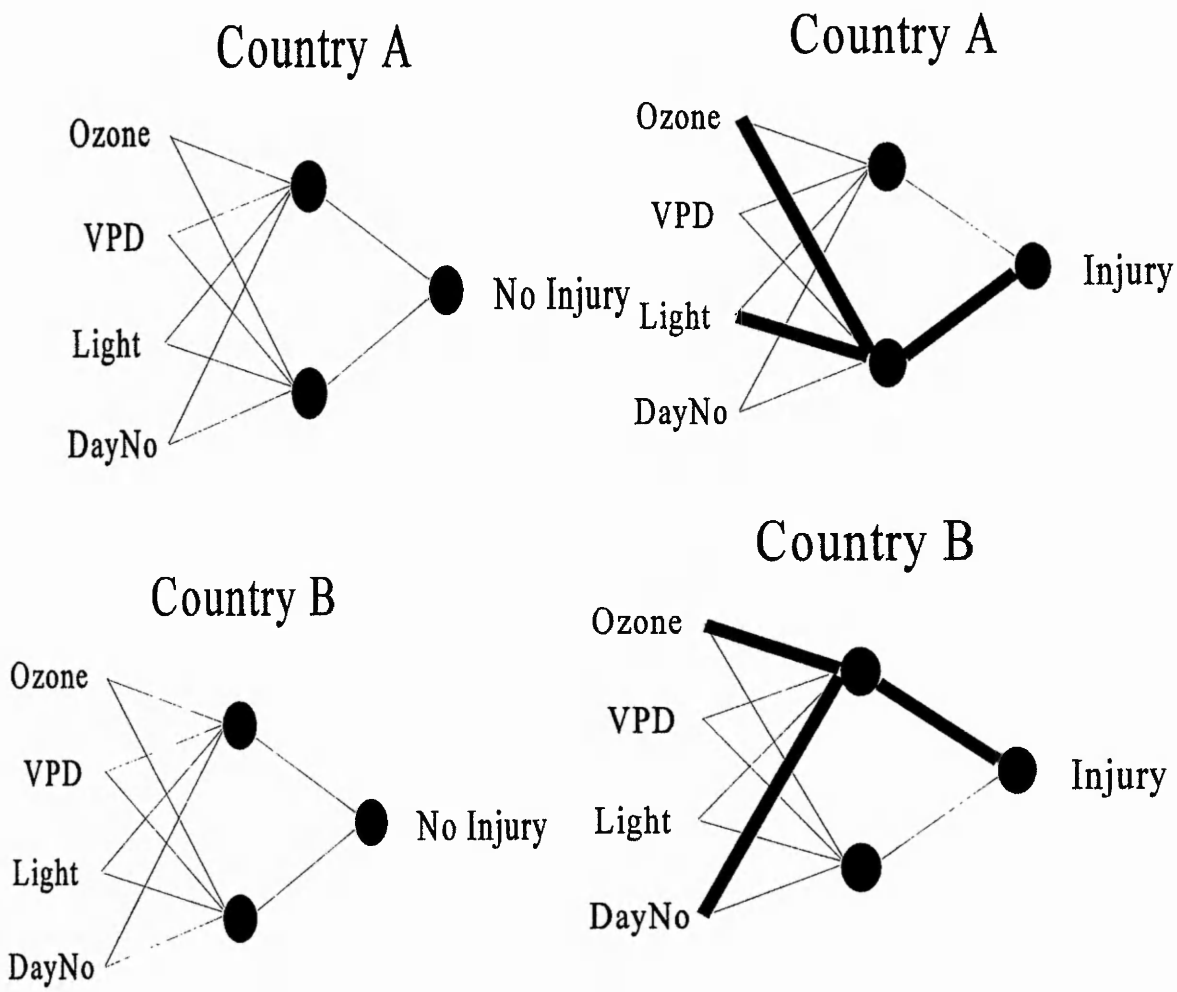


Figure 5.8. Grouping by pathway tracing.

This has been carried out on some real data. Two examples are shown below. The hidden unit which discriminates between injury and no injury is shaded (table 5.2).

HU1 Activation	HU2 Activation	HU3 Activation	Bias	Summed Activations	Desired	Actual Output	
-6.80	**4.26**	**-0.64**	**0.83**	**-2.36**	0.1	0.09	Sweden95
-2.53	**4.54**	**-1.06**	**0.83**	**1.78**	0.9	0.86	Sweden95
-1.44	**4.66**	**-5.76**	**0.83**	**-1.71**	0.1	0.15	Germany96
-1.47	**4.88**	**-2.65**	**0.83**	**1.58**	0.9	0.83	Germany96

Table 5.2

Each hidden unit has an associated equation, generated by equation synthesis. Country/year combinations can be matched to the relevant unit and equation (table 5.3)

	Hidden Unit 1	Hidden Unit 2	Hidden Unit 3
Country, year	SW95 GER(Branch.)95 FR95 {IT95} FIN95 IT96 CH96	{IT95) {CH96}	CH95 {IT95} GER (Branch.)96 GER (Geissen)96 UK96

Table 5.3. Grouping of countries based on their hidden unit activations.

At first, the usefulness of this hidden unit based discrimination may not be obvious, but it must be accepted that the grouping is not arbitrary. It may therefore be possible to divide countries into their subgroups then train separate networks especially for that subgroup. This is another method of fine tuning an ANN for transparency and efficiency. Tight examination of similarities between conditions at grouped countries could lead to a better understanding of the processes involved in the damage forming process. This is equally applicable to other problem domains, different patients may be grouped based on which hidden units their precise set of conditions activate etc.

5.6. Critical Level Extrapolation

In order to use the ANN model to estimate critical levels the ozone data for a 5 day, non-injury causing period, can be modified to simulate increasingly larger ozone episodes (figure 5.9).

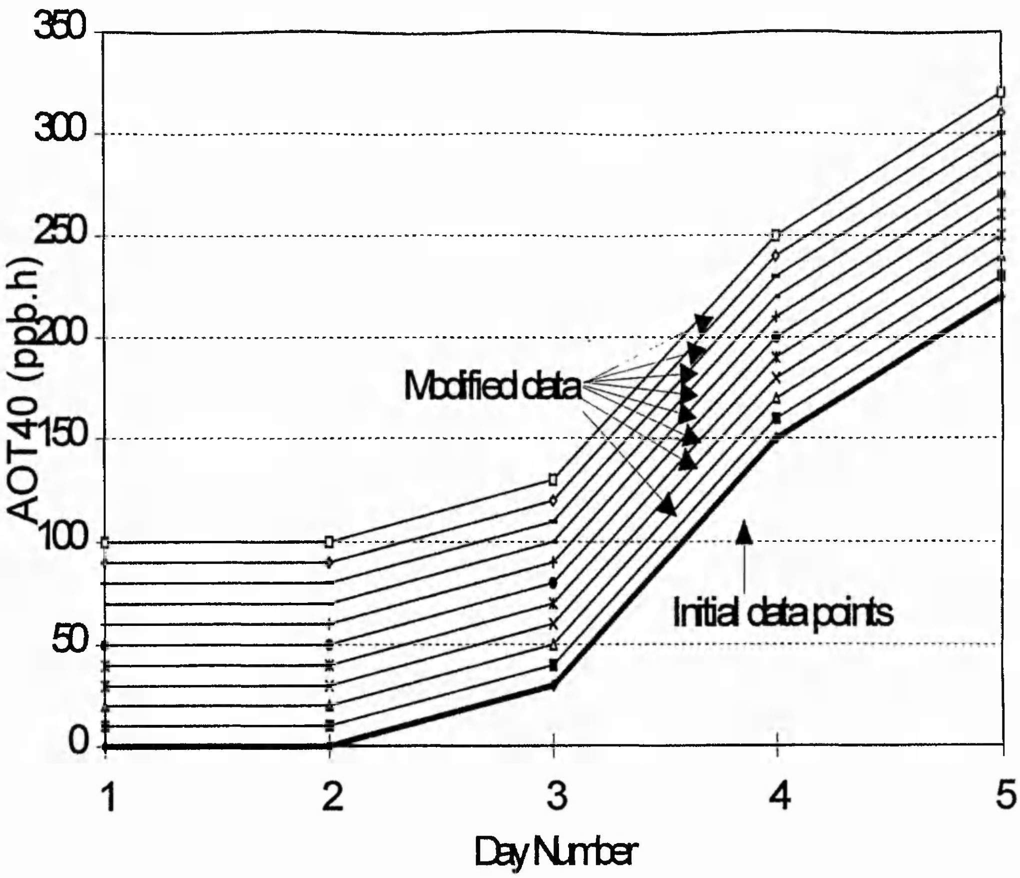


Figure 5.9. Original Ozone exposure and progressive increases presented to the network.

This means that an episode that failed to cause injury can be scaled up in a linear manner and the resulting output from the ANN observed. The results show the effect on the ANN output of increasing ozone level inputs (Figure 5.10). An AOT40 threshold of approximately 680ppb.h could be proposed from this extrapolation, this is in line with critical levels set by UN/ECE (Benton *et al* 1996a, Benton *et al* 1996b).

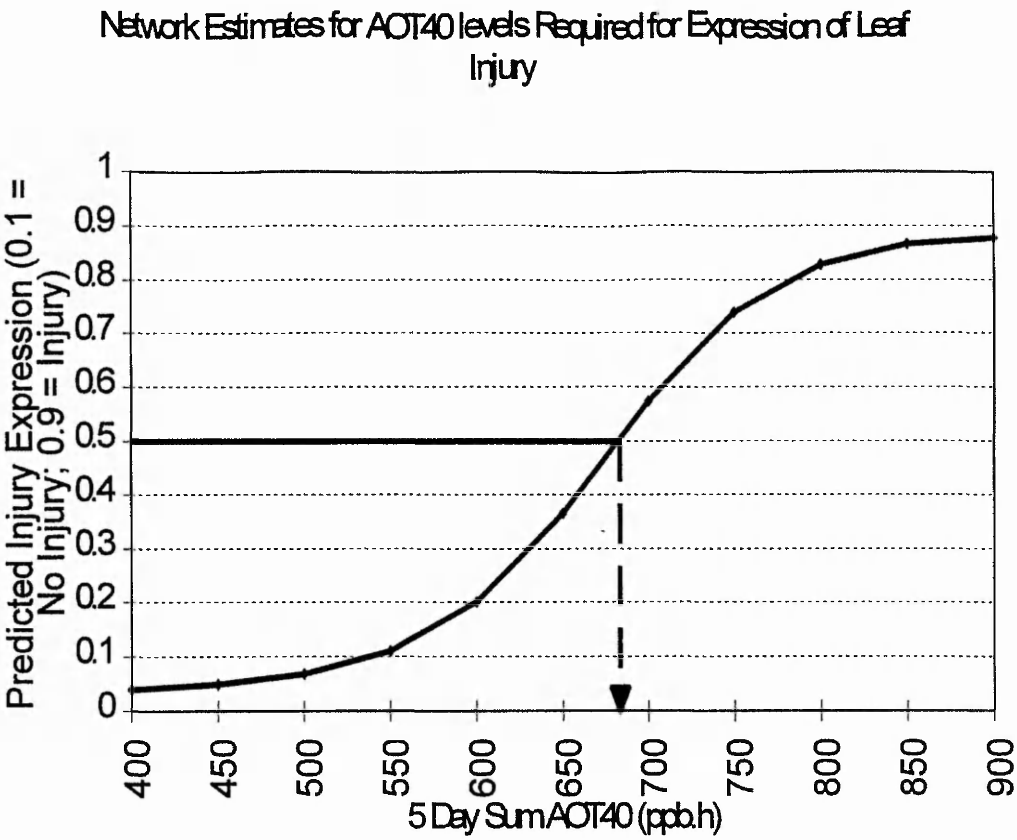


Figure 5.10. Effect of progressive increases in ozone dose on network output.

The climatic data for one site was used so this critical level is specific to the climatic conditions and ozone profiles at the site. Ozone periods from other sites were modified and applied in this way and gave similar output curves.

5.7. Closed Chamber Experiments

5.7.1 Introduction

An experimental protocol has been devised to test the hypothesis put forward by many of the equations derived from ANNs that successfully predict onset of injury, the hypothesis being:

"A fall in ozone levels is required for the expression of injury"

This hypothesis may look counter intuitive as high ozone episodes are required for leaf injury but this hypothesis is more concerned with the dynamics of an ozone episode

5.7.2 Materials and Methods

Clover seeds (*Trifolium subterraneum*) were sown in 3.5 inch plant pots containing J A Bowers seed and potting compost. Plants were raised under glasshouse conditions with natural daylight and supplementary lighting (300 - 800 μmol m^{-1} s^{-1} PPFD, 60 to 80% relative humidity (RH), at temperatures of between 20 and 35°C day and 9 and 15°C night and regular watering. Seedlings were thinned to 3 per pot and the experiment started once all the plants had at least three leaves. 8 pots were exposed to ozone for 7 hours per day, using a greenhouse-based closed chamber exposure system, for each of the different time periods (Table 2.) by removing 8 pots per day.

Plant Group	Day1	Day2	Day3	Day4	Day5
1	Ozone	AA	AA	AA	AA
2	Ozone	Ozone	AA	AA	AA
3	Ozone	Ozone	Ozone	AA	AA
4	Ozone	Ozone	Ozone	Ozone	AA
5	Ozone	Ozone	Ozone	Ozone	Ozone
6	AA	AA	AA	AA	AA

Ozone conditions in closed chambers.

	Day1	Day2	Day3	Day4	Day5
7hmean	57.1	65.7	59.8	39.5	41.9
Max	82.0	89.8	79.8	66.4	75.3

Table 5.2. Exposure conditions for hypothesis testing

AA = Ambient air, where Ozone levels remained below 10 ppb

The leaves on these plants were examined on day five for ozone injury, both quantitatively and qualitatively.

5.7.3 Results

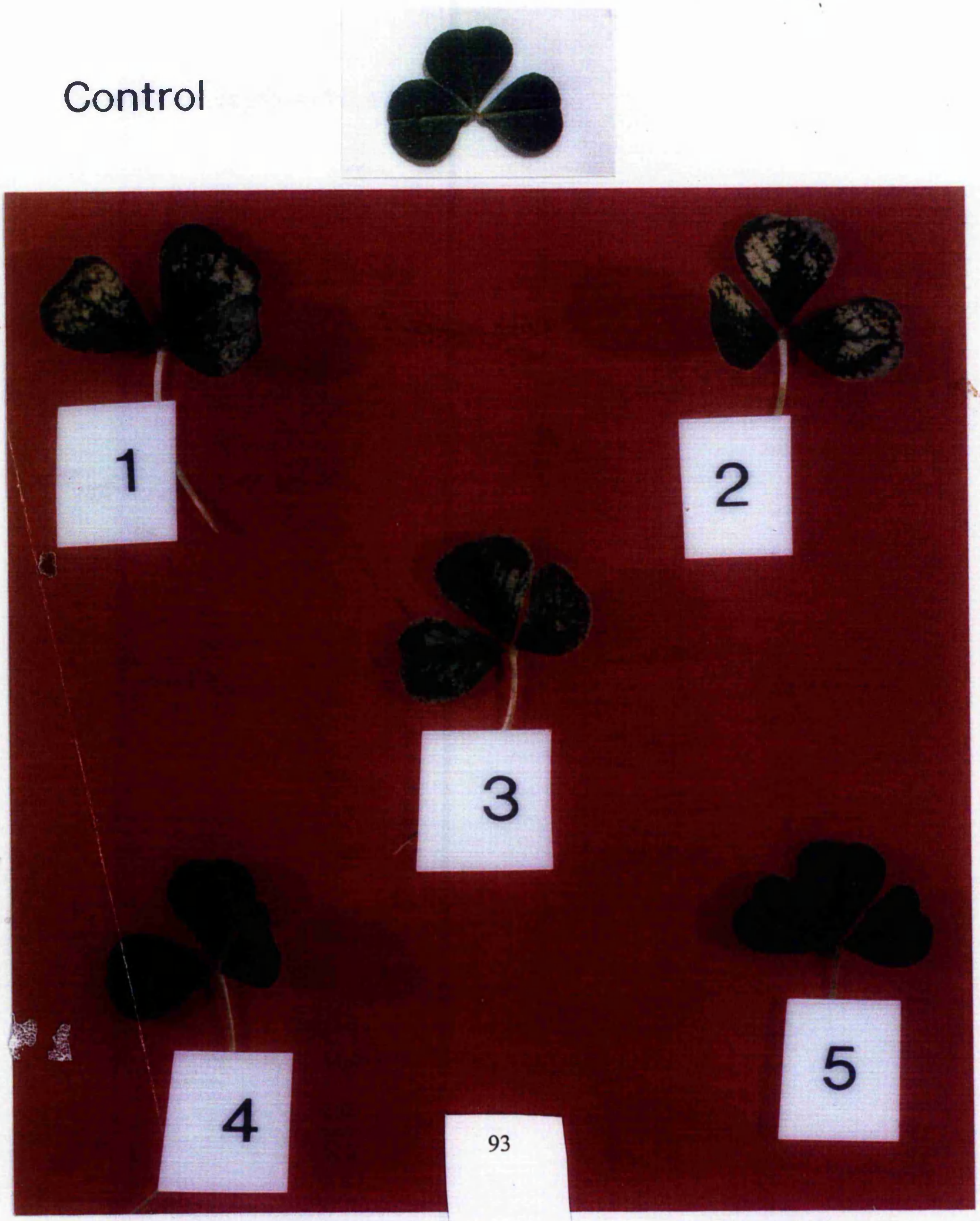
It was apparent that more classic, necrotic ozone injury was found on the plants from groups 1 and 2. (Plate 5) Plants from groups 3 to 5 had larger areas of what could be described as green injury. Moist, semi-translucent areas that covered the same part of the leaf as classic necrotic injury. The ICP-Crops protocol is based around the observation of necrotic injury so this type of injury would not have been deemed suitable to register as onset of injury. The expression of necrotic injury on clover leaves at lower levels (<100ppb) does not appear directly after exposure but is preceded by an area of moist leaf inflammation.

5.7.4 Discussion of Closed Chamber Experiment

Although it is possible that at high doses the expression of necrotic injury is immediate, this is not the case at lower, more natural levels. This inflammation is part of the plant's response to the ozone episode but its duration may be dependent on the ozone levels and the climatic conditions. It may also be true that the transition to necrosis is not the only outcome for an inflamed area and that a return to a healthy condition may be possible under certain conditions. A new hypothesis is proposed that:

"A fall in ozone levels is required for the onset of NECROTIC Injury"

Plate 5 Leaves from groups 1 to 5 on day 5, plus a healthy leaf from a plant not exposed to ozone. See table 1 for ozone treatments.



5.8. Discussion

This chapter shows the several stages of research from structuring of raw data to hypothesis generation and then hypothesis testing. The most difficult of these stages is traditionally hypothesis generation, usually involving extensive analysis of data to find trends or dependencies. The use of neural networks and equation synthesis proves a useful aid in this difficult area. The neural network provides the means to model all the data and equation synthesis gives useful indicators to features within the data. The problem of finding features within data increases as the amount of data increases but this is far more apparent when conventional statistical techniques are used than when ANN's are used.

The problem approached here was a difficult one for several reasons, the following are some of these:

1. The onset of leaf injury is an all or nothing event, the leaf is either injured or not injured, but the early stages of injury are difficult to distinguish to all but the keenest of observers.

2. The fact that the data came from many sites meant different assessors at each site so the consistency of monitoring cannot be guaranteed.

3. While the data available was very time consuming to collect and came from several European sites, it did not amount to a comprehensive set of conditions.

It is visible from these results that the prediction of the onset of injury, binary decision from such noisy and sparse data is achievable using ANN's. It is doubtful, though this has not been tested, that any other technique could have been applied to this particular problem and given such useful results.

The suggestion that a fall in ozone may bring on the expression of leaf injury is still being studied using closed chamber experiments. It is apparent that ozone damages leaves at two or more rates; at very high (approximately 150 ppb for 7 hours) levels its effects are nearly instant, causing widespread necrotic damage to all parts of the leaf in less than 24 hours but at lower levels (40 - 100ppb) injury development is more gradual. The plant seems capable, at least in the short term, to negate or delay any resulting injury. It is also thought that climatic conditions have an important effect on this injury development, and vapour pressure deficit may be particularly important in influencing the development of necrotic injury.

Chapter 6. Correlated Activation Pruning (CAPing)

6.1. Pruning

The generalisation ability of an ANN is dependent on its architecture. An ANN with the correct architecture will learn the predictive task presented by the training set but also gather enough general rules to correctly predict outputs for unseen test set examples. To obtain this optimum network architecture it is often necessary to apply a labourious 'trial and error' approach. One approach to achieving optimum network architecture in a more intelligent way is pruning. Weight pruning is the most researched area, this involves the removal of connections based on the value of the connecting weights, these can be divided into two groups.

1. Sensitivity Calculation.

Once the full sized network is trained the sensitivity of the error function to zeroing of a weight is estimated and weights with a low impact are removed (Karin 1990).

2. Penalty-Term Methods

This involves the introduction of a new cost function to enforce weights of small magnitude to converge to zero during training, they can then be removed with no effect (Weigend *et al* 1991).

Another method proposes the use of a genetic algorithm (GA) to reduce the number of connections of a fully trained network (Hancock 1992). The GA decides which combination of weights provides the most efficient network, the network is then pruned and retrained.

Sietsma and Dow (1988) describe an interactive pruning method that uses several heuristics to identify units that fail to contribute to the solution and therefore can be removed with no degradation in performance. This approach removes units with constant outputs over all the training patterns as these are not participating in the solution. Also, units with identical or opposite activations for all patterns can be combined. The approach in this paper is rudimentary and a more comprehensive and mathematically robust approach would be much more useful.

All these methods propose to benefit from the learning benefits of larger networks (faster learning and more features approximated) while reducing the amount of overtraining within these networks.

6.2. CAPing Theory

This section sets out to explain the mathematical basis for capping, an will enable the reader to implement CAPing on their own ANNs if required.

6.2.1. Equations

Each hidden unit within a three layer neural network produces an activation when a set of inputs is presented to it. If two hidden units produce outputs that are 100% correlated for the entire training set then the activations for these two hidden units can be merged, with no loss of performance. Two hidden units with correlated activities can be simplified into one hidden unit using equations 1 - 4.

$$Signal' = Signal_X + Signal_Y$$

Where $Signal_X$ and $Signal_Y$ are the profiles of hidden node activations for all training patterns for hidden nodes X and Y, and where Signal' is the required activation profile for a hidden node replacing X and Y.

IF $Signal_X = Signal_Y$ **Then** $Signal' = Signal_Y + Signal_Y$

If Signal X correlates positively with Signal Y then

$$W_Y' = W_Y + \left(\frac{\sigma_X}{\sigma_Y} * W_Y\right) \quad \textbf{(1)} \qquad \textbf{And} \qquad Bias' = Bias + \left(\mu_X - \left(\frac{\sigma_X}{\sigma_Y} * \mu_Y\right)\right) \quad \textbf{(2)}$$

W_Y' = new hidden unit to output weight

Bias = original bias weight

Bias' = new bias weight

μ_X, σ_X = mean and standard deviation of activations for hidden unit with lowest weighting

W_Y, μ_Y, σ_Y = weight, mean and standard deviation of activations for hidden unit with highest weighting

If Signal X correlates negatively with Signal Y then

$$W'_Y = W_Y - \left(\frac{\sigma_X}{\sigma_Y} * W_Y\right) \quad (3) \qquad \text{And} \qquad Bias' = Bias + \left(\mu_X + \left(\frac{\sigma_X}{\sigma_Y} * \mu_Y\right) \quad (4)$$

The equations for merging positively correlated units are reached using the following derivation.

Zero means by subtracting the mean of signal from signal (Fig. 6.1).

$$S_X = Signal_X - \mu_X \qquad\qquad S_Y = Signal_Y - \mu_Y$$

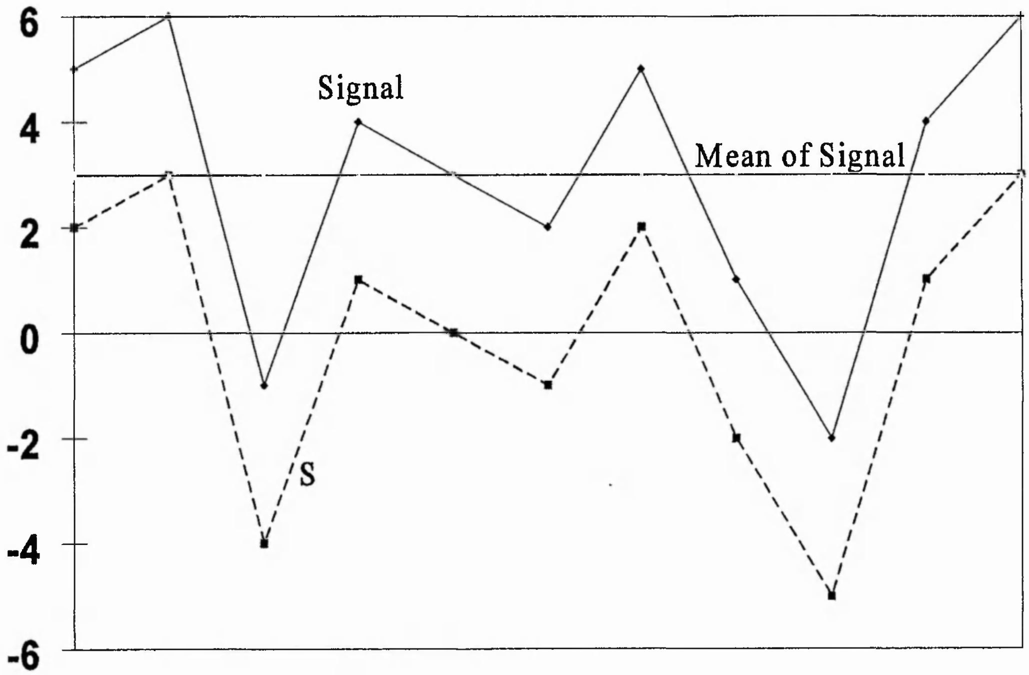


Figure 6.1. Signal Modification

thus

$$\frac{\sigma S_X}{\sigma S_Y} = \frac{\sqrt{\Sigma S_X^2}}{\sqrt{\Sigma S_Y^2}}$$

also since

$$\rho_{X,Y} \cong 1$$

$$Signal_X \cong K_1 Signal_Y + K_2$$

$$S_X \cong K_3 S_Y + K_4$$

$$\mu S_X = \mu S_Y = 0$$

And since $$\mu S_X = \mu(K_3 S_Y + K_4) \quad = K_3 * \mu S_Y + K_4 = 0$$

$$\therefore K_4 = 0$$

$$\therefore S_X \cong K_3 S_Y$$

Now $$\frac{\sigma S_Y}{\sigma S_X} = \frac{\sqrt{\Sigma S_Y^2}}{\sqrt{\Sigma (S_Y * K_3)^2}} = \frac{1}{K_3}$$

And so, $$S_X = K_3 * S_Y = \frac{\sigma_X}{\sigma_Y} * S_Y$$

$$Signal_Y^{\prime} = Signal_X + Signal_Y \qquad S_X + S_Y + \mu_X + \mu_Y = (K_3 * S_Y) + S_Y + \mu_X + \mu_Y \qquad = \frac{\sigma S_X}{\sigma S_Y} * S_Y + S_Y + \mu_X + \mu_Y$$

$$\therefore \qquad (W_Y * H_Y) = S_Y + \mu_Y = Signal_Y$$

W_Y' can be written so that $$(W_Y^{\prime} * H_Y) = S_Y + \mu_Y + S_X$$

And $$Bias^{\prime} = Bias + \mu_X$$

In which case, $$(W_Y^{\prime} * H_Y) = (W_Y * H_Y) + (\frac{\sigma_X}{\sigma_Y} * S_Y) = (W_Y * H_Y) + (\frac{\sigma_X}{\sigma_Y} * ((W_Y * H_Y) - \mu_Y)$$

The Bias' can therefore be modified by adding the constant:

$$-\frac{\sigma_X}{\sigma_Y}*\mu_Y$$

To give

$$Bias^{/}=Bias+\mu_X-\frac{\sigma_X}{\sigma_V}*\mu_Y \qquad \textbf{(1)}$$

Leaving

$$(W_Y^{/}*H_Y)=(H_Y*W_Y)+\frac{\sigma_X}{\sigma_Y}*(W_Y*H_Y)$$

$$\therefore W_Y^{/}=W_Y+\frac{\sigma_X}{\sigma_Y}*W_Y \qquad \textbf{(2)}$$

W_X, H_X = weight and output of activations for hidden unit with lowest weighting

H_X = output for hidden unit with highest weighting

σS_X = standard deviation of activations, offset to a mean of 0, for hidden unit with lowest weighting

σS_Y = standard deviation of activations, offset to a mean of 0, for hidden unit with highest weighting

$\rho_{X,Y}$ = correlation coefficient for $Signal_X$ and $Signal_Y$

For the activations of 2 hidden units to be 100% correlated their output for each pattern does not need to be identical. For example the following number series (table 4.1) all correlate with each other with a correlation coefficient of 1 or -1. This is because they could all be identical (1) or mirror images along the x axis (-1) if their offset and gain were modified.

Series X	Series Y	Series Z
0.2	0.6	-0.4
0.4	0.8	-0.8
0.8	1.2	-1.6
0.4	0.8	-0.8

Table 6.1. Three number series that correlate with a correlation coefficient of 1 or -1

The correlation coefficient is derived using equation 5.

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x, \sigma_y} \quad (5)$$

where:

$$Cov(X,Y) = \frac{1}{n} \sum_{(i-1)}^{n} (x_i - \mu_x)(y_i - \mu_y)$$

and

$$-1 \leq \rho_{xy} \leq 1$$

Correlation coefficients of 1 or -1 are not usual for activations from any two hidden units of an ANN, but high correlations do exist in networks that use more hidden units than are required to learn a problem. Two sets of activations are more highly correlated as the correlation coefficient gets nearer to 1 or -1. Error can be introduced into the networks performance if the correlation coefficient is not equal to 1 or -1 but less error will be introduced with closer correlations.

Correlated activations can occur for several reasons:

1. Similar weights from input units to hidden units.

2. Each of the two hidden units detect a different, co-dependent feature. For example, if high ozone levels always occur when there is a north wind there is no need to detect both north winds and ozone.

3. If two factors have the same effect. For example, if high ozone levels or low humidity levels both cause the same injury there is no need to detect both of these factors.

6.2.2. Benefits of CAPing

1. Speed Up.

An ANN can be trained in fewer epochs if more hidden units are used than are required. Parallel processing allows the addition of more hidden units without an accompanying increase in time to train an epoch. Therefore an ANN could be trained quicker using more hidden units and then capped to a near optimum format (Fig. 6.2)

2. Analysis of Correlated units.

The weights of the input layer to hidden unit connections for merged units can be looked at, by comparing these weights, either directly or by synthesising equations, allowing an understanding of related events to be made.

3. Network Optimisation.

Trial and error is a common approach to network optimisation. CAPing of a network may reach a near optimum structure in one training session.

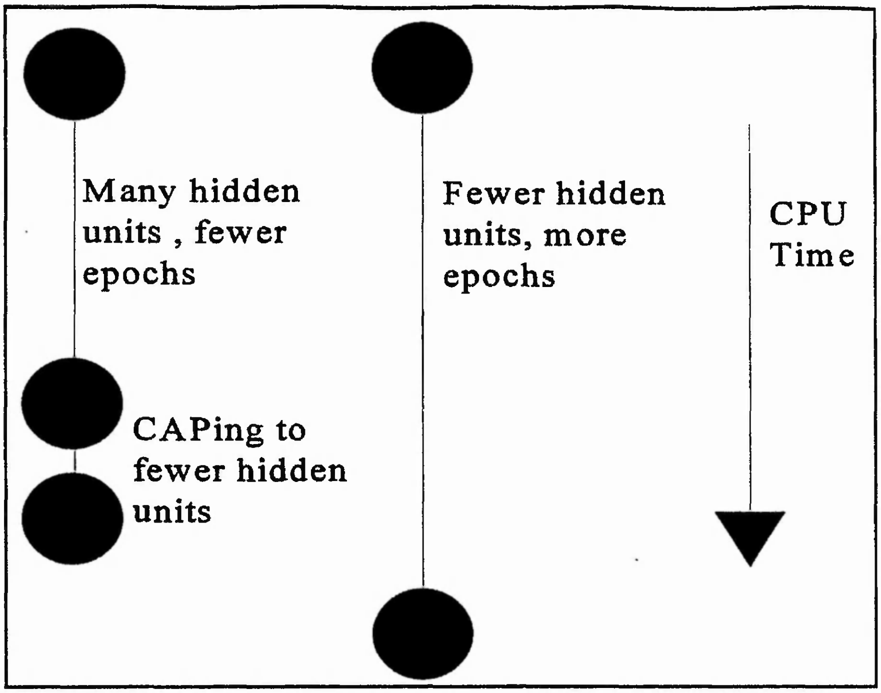


Figure 6.2. A schematic diagram of CAPing Benefits

6.3 CAPing in Practice

To test if the theoretical promise of the CAPing algorithm could be transferred to the optimising of working areas of research, the algorithm was applied to two such areas; curve fitting and ozone damage prediction.

6.3.1. Curve Fitting

Training an ANN to learn a complex, multiple variable equation is an accepted method of analysing network performances, this can appear as a curve fitting exercise (eg. Ripley, 1995) or as boolean equation solving (eg. Wiersma *et al*, 1995).

In this case, a 4 variable equation was used and random numbers were used as variables to generate an output, X. Constants were required to keep X within an appropriate range. The equation used was:

$$\frac{A-(B*C)+(\frac{B}{(D+1)})+K}{J}=X \qquad \textbf{(6)}$$

Where A,B,C,D are variable and K and J are constants

A three-layer back propagation network was then trained using 500 data sets, each consisting of 4 inputs (A,B,C,D) and an output {X (a,b,c,d)}. The level of accuracy to be achieved was set at an RMS of <0.03 and a maximum error of <0.1. The minimum number of hidden units needed to learn this was found, by exhaustive trial, to be four. A network was then trained using 20 hidden units, many more than were required. This took fewer epochs to reach the required accuracy and, because of the parallel architecture used, each epoch took a similar training time as a 4 hidden unit network. The CAPing equations were then used to recursively merge hidden units, with the most highly correlated units merged first. The activations of two such units and the activations of the combined unit are shown in figure 6.3.

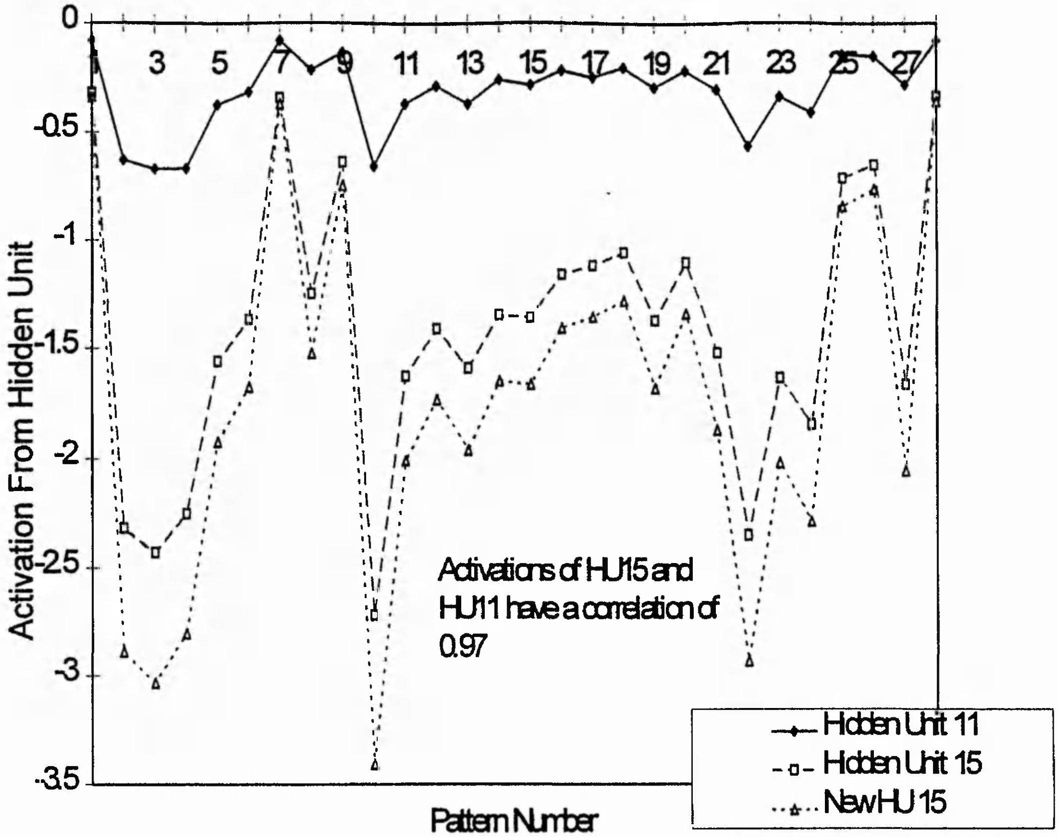


Figure 6.3. The effects on activations of merging two correlated hidden units

Hidden unit 11 was merged into hidden unit 15 to produce a new weight from hidden unit 15 and a new bias into the output unit. The effect of changing the weight from hidden unit 15 and the bias is shown.

Very little error was introduced until units with correlation coefficients of less than 0.8 were used (Fig. 6.4), the correlation coefficients are given above the error. Merging to even fewer hidden units was possible if some post-CAPing retraining was carried out, this retraining needed approximately 1/10 of the epochs of training from random weights.

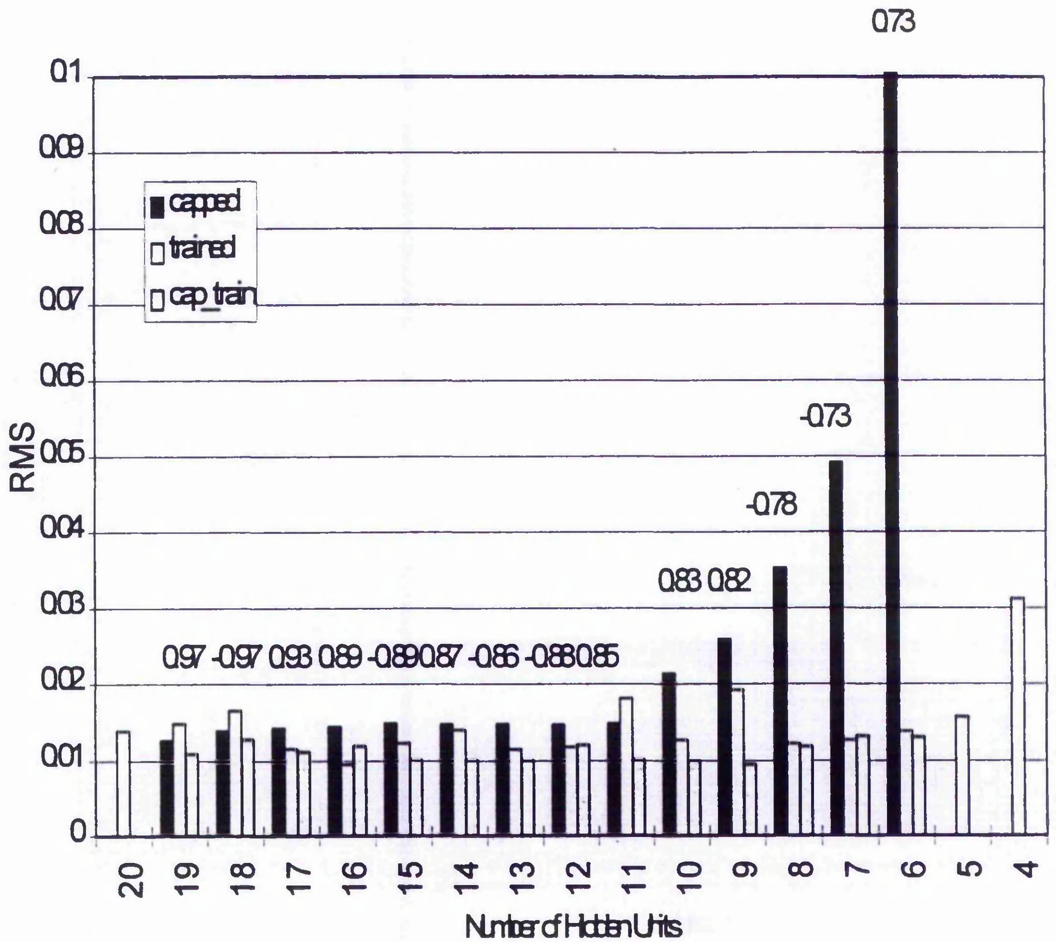


Figure 6.4. The effect of merging less correlated units.

Analysis of the weights from the input units of the merged hidden units showed some sets of weights to be highly correlated but other sets to be non-correlated (Table 6.2). This shows how CAPing removes several types of redundancy.

Weight from:	to hidden unit 1	to hidden unit 2	to hidden unit 3	to hidden unit 4
Variable A	0.2756	0.04419	-1.4898	-1.4746
Variable B	-0.573	-0.7546	-0.6228	-1.6257
Variable C	-0.2666	-0.4243	-0.8889	-1.1646
Variable D	-0.9353	-1.0815	-1.0183	-1.2954
	Correlation 1,2	0.9991	Correlation 3,4	-0.1099

Table 6.2. Input to hidden unit weights for pairs of correlated hidden units

6.3.2. CAPing for Crop Damage Modelling

Various models for predicting onset of leaf injury have already been discussed in this report (chapter 3). The CAPing algorithm was applied to three such models:

Model 1: Inputs - 5 day levels of ozone (O_3), relative humidity (rh), temperature, global radiation and day number. Output - Binary representation about occurrence of injury on following day.

Model 2: As model 1 but relative humidity and temperature computed into vapour pressure deficit (VPD).

Model 3: As model 2 but with weekly variables not daily.

In all cases it has been shown that networks with 3 or 4 hidden units can both learn to solve the problem presented by the training set (eg. Fig 6.5) and also generalise enough to correctly classify patterns in the test set (eg. Fig 6.6).

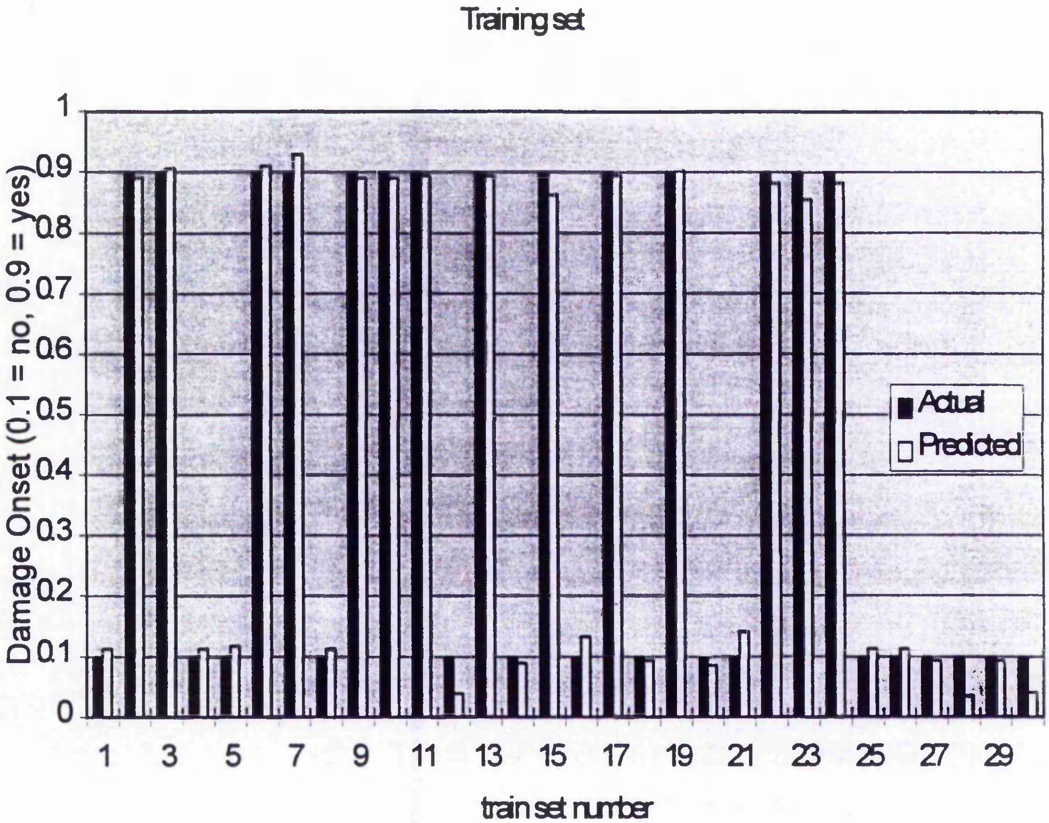


Figure 6.5. Predictive performance of model 1 on training set

These networks were then trained with excess hidden units and CAPing was carried out recursively until the CAPed network no longer correctly predicted all outputs in the test and training set. Model 1 could be capped from 10 to 7 hidden units and model 2 could be capped from 10 to 8 hidden units and model 3 could be capped from 14 hidden units to 6.

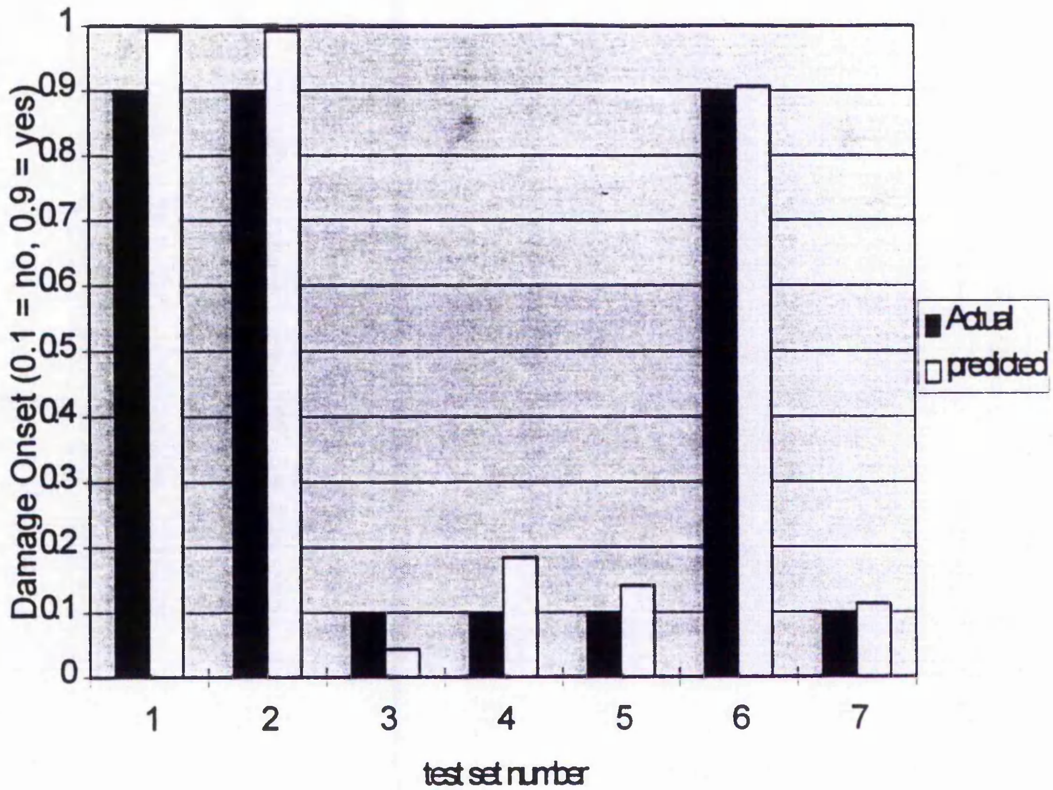


Figure 6.6. Predictive performance of model 1 on test set

Although less hidden units could be merged than the curve fitting example, some CAPing was at least possible. The most highly weighted connections between input unit to hidden unit were looked at for hidden units that were merged during the CAPing procedure. Some produced similar activations because the weights were similarly prioritised. Other pairs of hidden units with correlated activations had no obvious similarities in the input unit to hidden unit weights. In model 3, weekly variables, the following equations were synthesised from two hidden units that showed closely correlated activities.

F{7hr mean light + Max light - (2*AOT40)}   and

F{7hr mean Ozone + Max Ozone - MaxLight}

It can be seen that two hidden units that use maximum light differently still produce correlated activations in conjunction with differing modifiers

CAPing offers, in one iteration, a means to compress ANNs. This could be used for several reasons and is a useful tool in the use or optimisation of ANNs.

6.4. Discussion.

CAPing offers many benefits to the neural network practitioner. Reducing the size of an ANN without significantly reducing its performance can only lead to more transparency. This is aided by the fact that CAPing is very objective whereby simple rules are followed that don't claim to reduce every network, only those where too many hidden nodes are used. This objectivity is crucial as there are already enough moveable goalposts in a researchers drive for optimum performance.

Chapter 7. Advanced Quantitative Neural Network Models.

In chapters 4 and 5 methods of predicting events and the degree of these events are introduced, using onset of ozone injury and amount of injury as our test cases. Now we attempt a more complex prediction problem, the resulting effects of an event. Again many biological situation could be used as example. Using the rainfall prediction problem, an extrapolation to the impact on rivers and lakes could be made. From a modelling point of view, this chapter aims to supply methods for assessing the impact on secondary effects and the example of Ozone's effect on yield is used.

Two approaches are taken to discovering the effects on yield of ozone exposure. ANNs are used to model the effect of the ozone exposure profile on yield reduction and the effect of modifying factors on yield reduction is investigated. As part of the ICP-Crops programme researchers grew two strains of *Trifolium Repens*, one is ozone resistant the other is ozone susceptible (Heagle *et al* 1994). The two clones were grown from cuttings in controlled conditions at most of the participating ICP-Crops European sites (see ICP-Crops protocol. UN/ECE. 1996.). Harvests were taken at 28 day intervals and the dry weights recorded, by comparing the yields of these two strains, the effects of ozone can be estimated.

7.1 Biological Modifiers of Yield

The profile of an ozone exposure has been shown to have an important effect on development of leaf injury (Balls *et al* 1996a, Benton *et al* 1996b, Roadknight *et al*

1997a, Musselman *et al* 1986, Karlsson *et al* 1995), a volatile ozone exposure with both high a low levels of ozone has a more damaging effect than a stable ozone exposures. So if two days have the same sum ozone levels (fig. 5.3) and one day has a higher maximum ozone level (solid line) than another (dotted line) this will lead to a greater increase in damage. The effect that differing exposures have on yield of a crop is less clear. The relationship between crop damage and yield is also unclear.

7.2 Existing models of yield reducing effects of pollutants

Some models have been developed by Pleijel (eg Pleijel *et al* 1995) in WASP, also Jurg Fuhrers' wheat dose response model (Fuhrer and Achermann 1994) is looking at the yield of wheat in response to exposure to ozone. Bender 1990 also developed a non-linear model of yield dose response.

7.3 ANN Approach

So far ANN's have been used to make mainly qualitative assessments of the impact of ozone doses, given the levels of other influencing factors. Looking for yield changes starts to bring in some quantitative factors. Fuller details of quantitative approaches are given in chapter 8, here when look at yield changes we look for the network to predict yield to a given degree of error (+- 10%)

7.3.1 ANN method

The 28 day growing period between harvests can be broken up into 7 day periods and the ozone exposure during that period can be defined in 3 parameters, AOT40, Mean of daily maximum ozone and mean of daily 7hr mean ozone. Simple one hidden unit network used to model yield reduction due to ozone.

An ANN was trained to predict the reduction in yield using these inputs (figure 7.1).

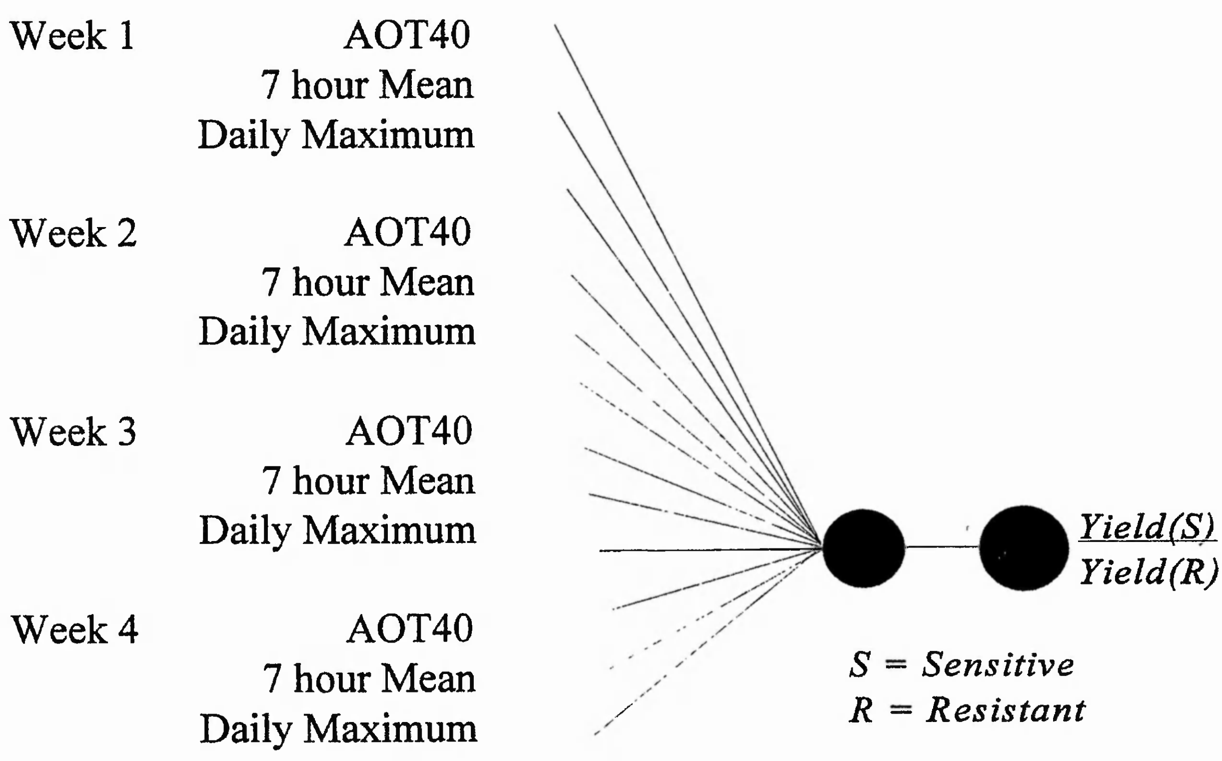


Figure 7.1 Architecture of a yield predicting ANN.

The ratio of sensitive to resistant gives us a level of control, as we are not attempting to predict the gross yield (which would be effected by sunlight, moisture etc) but the impact of ozone on yield.

7.3.2 ANN performance (ozone + climatic parameters).

It is to be expected that an ANN with only one hidden node should have difficulty modelling a complex, non-linear problem. The single hidden node will be forced, by back-propagation, to model every feature within the data set. If there are multiple non-linear features (as is suspected) the one node will be forced to compromise. This is less of a problem if only a subset of the entire data set is available for analysis. Performance is also more difficult to assess if the desired output is continuous. Both of these are true in our trial. A training and test set of 10 and 7 were used, this is far to few for meaningful conclusions to be made from any models developed but still useful as a modelling exercise.

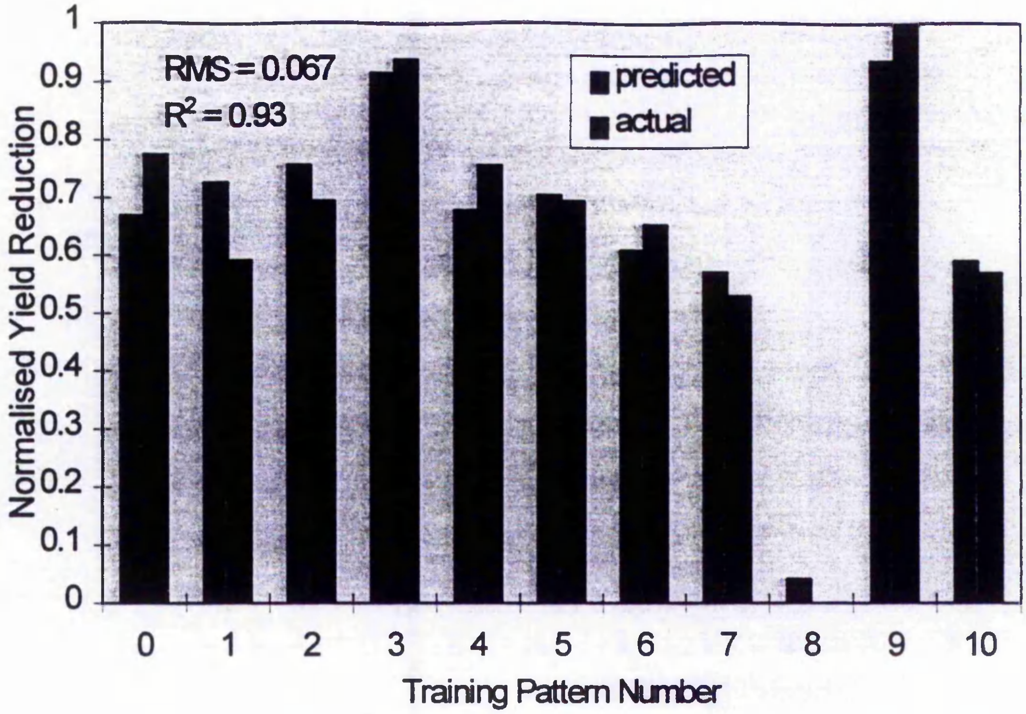


Figure 7.2. Accuracy of prediction for training set of ANN trained to predict yield reduction.

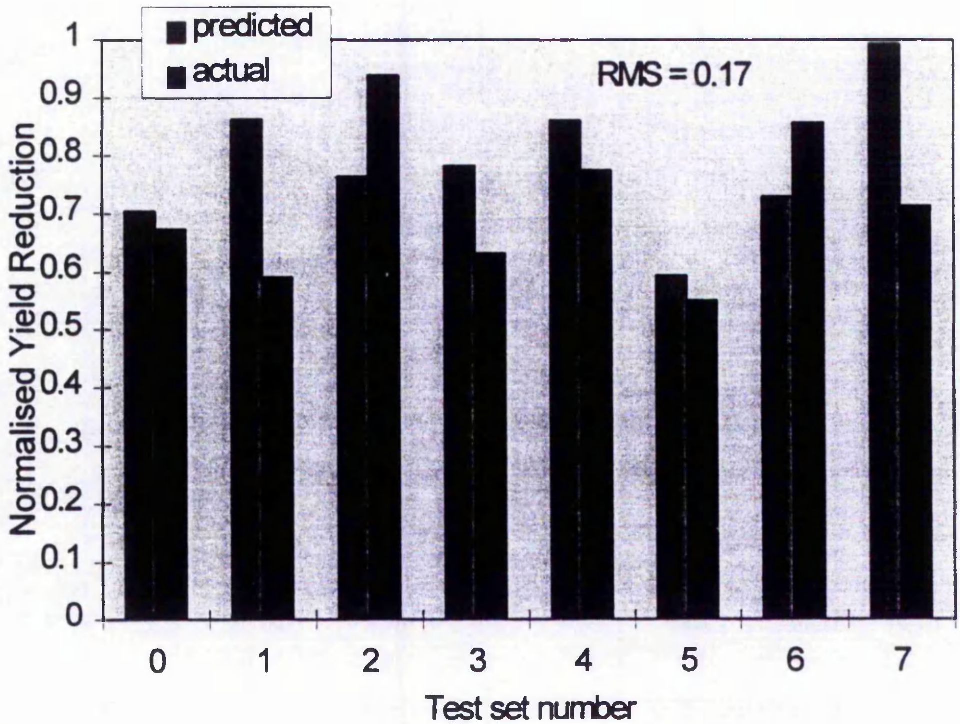


Figure 7.3. Performance of yield predicting ANN on test set.

This network modelled similar data to that used in chapter 4 but this network can predict yield reduction to an R^2 of 0.9. (S/R vs AOT40 gives an R^2 of ~0.6). These results are very encouraging, given the high degree of accuracy displayed for both the test set and the training set. This was without environmental factors such as humidity and temperature, and without any data about other pollutants. This suggests that for yield prediction ozone levels offer a significant amount of information, and that yield may not be that closely linked to the display of injury on a plant.

7.3.3 Analysis of Yield predicting ANNs

Analysis of this network gives some useful indications to the yield reducing features of an ozone episode. Weight analysis of the ANN suggests that:

- Higher ozone levels during the Second 2 weeks have a larger negative effect on yield than during the final two weeks. (Week 3>Week 4 >Week 1>Week 2)

- 7 hour mean ozone has a lesser impact on yield reduction than AOT40 or daily maximum.

The higher impact of ozone during the latter stages of the growth period may be due to the increased leaf area, the ozone can therefore the larger area of effect. The second point concurs with earlier research in this thesis with the 7 hour mean being consistently shown to be a less important indicator of ozone induced leaf damage than maximum daily ozone levels and AOT40.

7.3.4 Addition of modifying factors.

Environmental and climatic factors have a modifying effect on ozone injury (Mortensen 1992, Kersteins *et al*, 1992; Van Pul & Jacobs, 1993) but their effect on yield is less well known. The effects of humidity and, to a lesser extent, temperature are known to be particularly important. A measure of these variables is recorded by ICP-Crops participants for the 28 day period between harvests, along with ozone levels. These parameters can be used as inputs to a simple ANN, the desired output being yield of sensitive crop/yield of resistant crop. The structure of a trained network that best predicted these outputs ($R^2 = 0.7$) is given in figure 7.4. The training performance of the network is shown in figure 7.5.

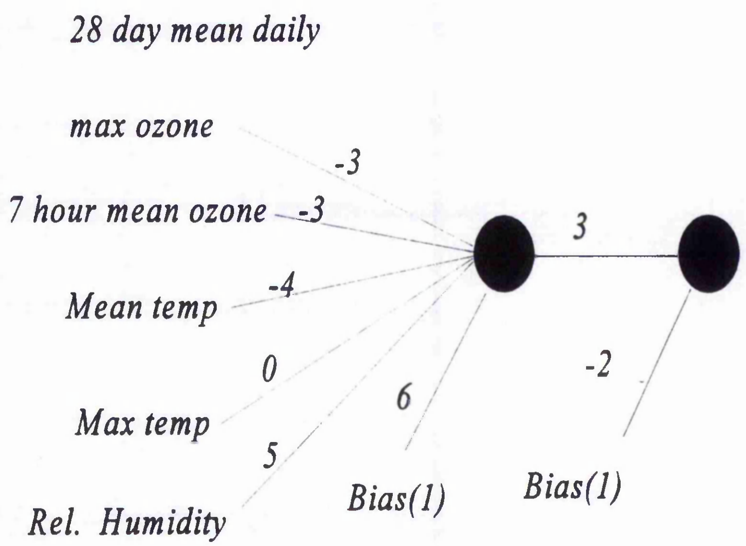


Figure 7.4. Structure of yield predicting ANN, with inputs for modifying factors.

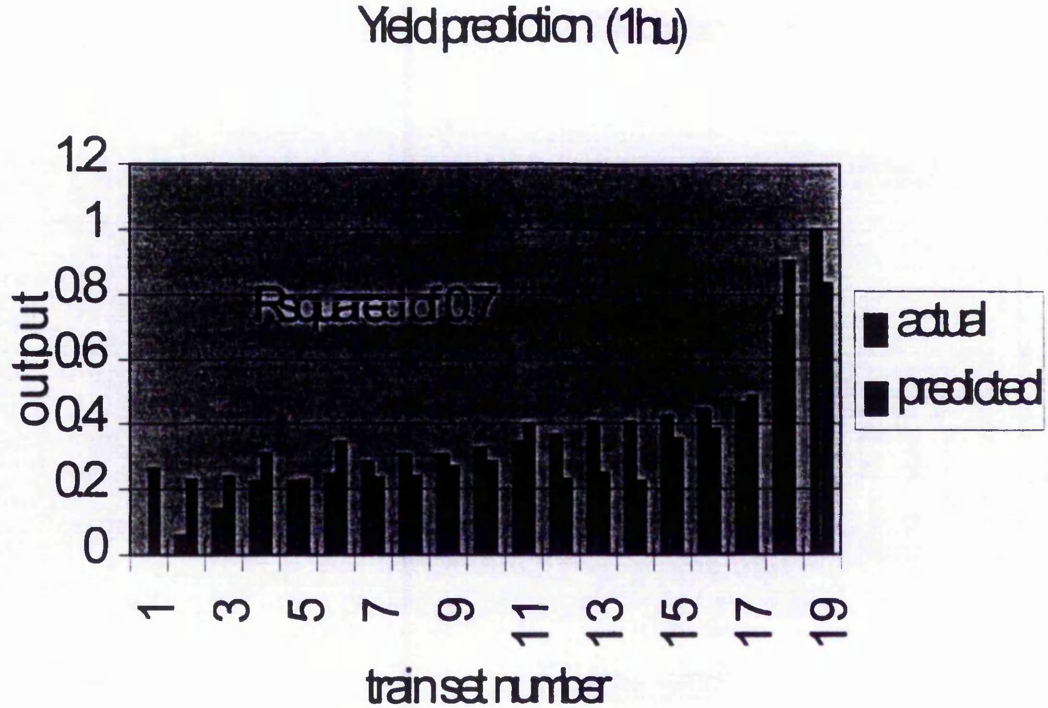


Figure 7.5. Performance of yield predicting ANN.

The weights of this network point to negative influences on the S/R for max ozone, 7hr mean ozone and mean temperature and a positive effect for humidity. In simpler terms, the first three parameters reduce yields and the final one increases yield.

It will been apparent to the reader that this network only has 1 hidden unit. Networks with more hidden unit were tried, but using the methods of CAPing and equation synthesis it became apparent that only 1 hidden unit was needed, suggesting a relatively linear cause-effect relationship was at work.

7.4 Conclusions

The small amount of data makes any conclusions drawn from this section of research tentative. What has been shown is that if a simple ANN accurately models a cause effect relationship, then useful pointers to the relationship can be revealed by analysis of the network weights, which in this case is relatively straightforward since there is only 1 hidden unit

The given example results in an implication that humidity has a beneficial effect of yield whereas ozone and (less explainable) temperature have a reducing effect on yield. The linked nature of temperature and ozone levels may responsible for the temperature inclusion in the harmful effects group. There is a correlation between ozone levels and temperature (see fig. 7.3), but when data from many countries is analysed this was not significant enough to enable the removal of one of the input sets. This difference is evidenced when some extracted equations show conflicting effects for ozone and temperature. Its is also obvious that the basic effect of temperature is non-linear (very high OR very low temperatures will kill the plant!)

It is apparent that if a secondary effect is an approximately linear result of a primary effect then a variable suitable for the primary effect can be used to predict the secondary effect. Although being sure of avoiding non-linearities is a very 'hit or miss' affair. Even if there is a straightforward relationship along the known data-points, there may be seemingly insignificant areas of the data that have a non-linear relationship.

Chapter 8. Hidden Unit Analysis for Discovery of Important Thresholds.

In this thesis, the behaviour of the hidden units has been an important focus. This is rightly so, in CAPing they provide the only good indicator of complex redundancy within a problem solving ANN. This chapter discusses the usefulness of analysing these activations for generating thresholds.

So far the problems of modelling ozone related leaf injury have largely been constrained to qualitative predictions[6]. Now methods to give the researcher an idea on quantities of damage are examined, whether that be in the changes in yield, leaf area effected or giving quantitative rules of thumb for leaf injury expression.

8.1 Problems involved in quantifying the effects of inputs to an ANN

Some of the mathematical problems associated with apportioning meaning to weights of an ANN have already been discussed in section 5. The non-linear nature of an ANN construction means that any assumption that can be made (using an appropriate algorithm) must be tested for validity.

---

6 Chapter 4 has some approaches that predict increase in damage, but is fundamentally about predicting 'if' there will be an increase in damage, not how much that increase will be.

8.2 Techniques used.

In this chapter, mostly diagnostic methods are used. States are analysed both visually then quantitatively, initial conditions can then be varied or new conditions added, the resulting effects on activations being the diagnostic tool. This is made possible by equation synthesis and CAPing enabling the accurate evaluation of the purpose of individual hidden nodes. For instance, 'sliding' variables and examining their effect gives an idea of a particular variable's impact, and the nature (non-linear, linear, thresholded) of its affect on the network's output. Some fixing of variables is sometimes required to enable one variable to be analysed at a time, but realistic values for secondary influencers are used.

The activations need to be put into relative and absolute context, chapter 3 showed how misleading making straightforward assumptions about activations can be. The activations from on hidden node must be compared with other activations from the same node when presented with different inputs. Also, the importance of the node relative to the other nodes in the network must be gauged.

8.3 Quantifying the effect of ozone on the onset of leaf injury.

Each activation at each node within an ANN has a non-linear effect on the final output, this is due to the inherent non-linearity of the network. It is therefore difficult to take the activation of nodes and make extrapolations from these. The methods proposed in the earlier chapters of this thesis enable a minimised network to be created and for the salient rules by which that network performs its task to be alluded to. These methods are not specifically for the case study used. Taking a minimised network with a known set of equations generated by equation extraction, it was possible to look at the activations occurring at one node. This is possible because the most important node can be taken, or the node with strong equational bias towards factors of interest. In our example the most important node of the onset of leaf damage predicting ANN was taken. As can be seen from chapter 4, the weights for one node showed up as being particularly important in terms of ozone levels, whereas the other 2 were more concerned with modifying factors. This approach could be applicable to any ANN ie. Identifying how the model is divided up over the nodes. All levels for the three ozone related variables (AOT40, Maximum ozone levels and 7hr mean) were passed through the ANN, with their associated modifying factor levels. The results of this are shown in figure 8.1. It is encouraging to see a clear cut off point for all three variables at which it can be deemed that the highest level has been reached that did not cause a significant activation, these levels are shown within the figure. Notice how the activities vary between 0 and 8 but to the right of the line they are always above a value of 7.

This speculative approach can be applied to any ANN, and just involves looking for distinct boundaries within a set of activations and then making statements about the position of these boundaries. This is a useful method for finding general thresholds around which activation patterns change considerably.

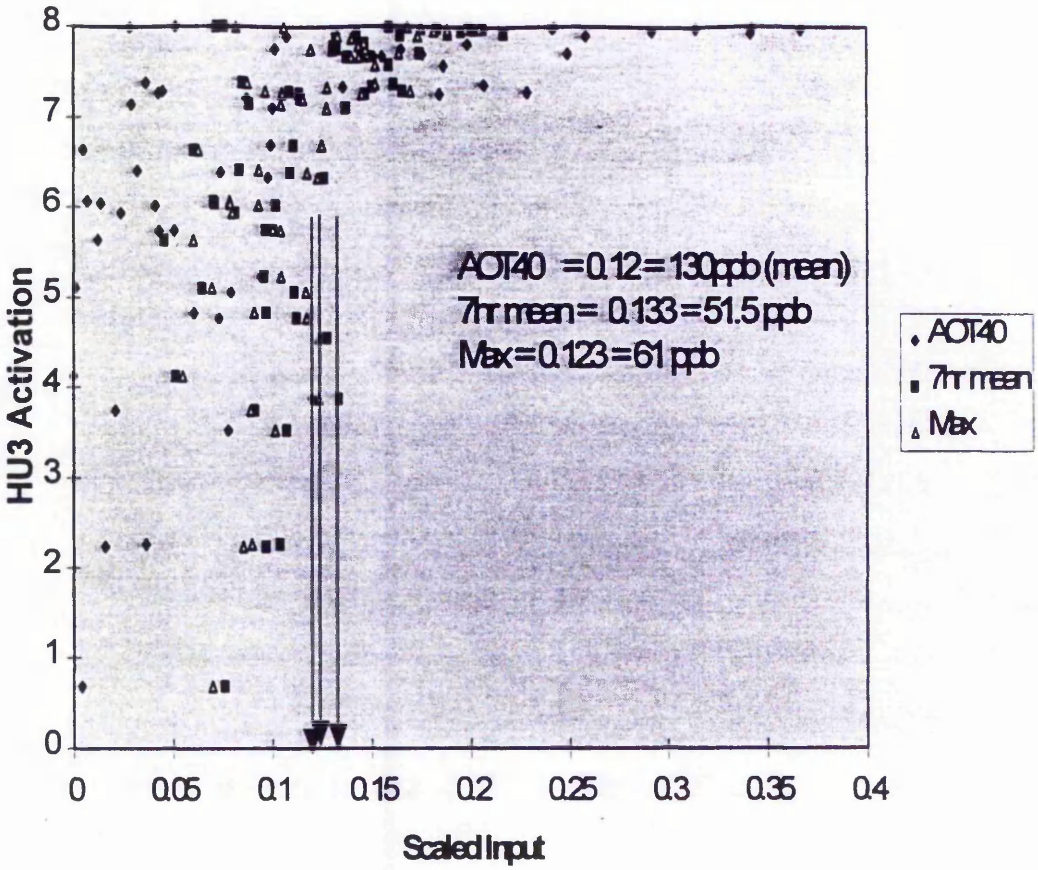


Figure 8.1. Activation levels

8.4 Quantifying the effect of ozone on yield.

With any neural network, the inputs will represent a range of states or conditions. All data sets have extremes. For ozone the extremes could be 'no ozone' therefore no injury and 'enough ozone to effect complete yield loss'. The points between thesee 2 extremes can be complex and non-linear, but can be examined and any trends evaluated. With the ozone injury example, non-linearity could arise from interactions with other chemicals or conditions, but a simpler, one parameter possibility also exists. As ozone levels increase various thresholding points may be reached in relation to yield, for example, up to level T_1 ozone levels may not be sufficient to do any damage to the leaf; then as the ozone levels move up from level T_1 to level T_2 the plant may repair any damage caused or replace damaged matter with fresh matter; then beyond level T_2 damage may increase proportionally to exposure until the entire plant is killed.

In an attempt to quantify the effect different ozone and humidity concentrations have on yield, different levels for ozone input were entered into the network at two realistic humidity settings. The output from this trial is displayed in figure 8.2.

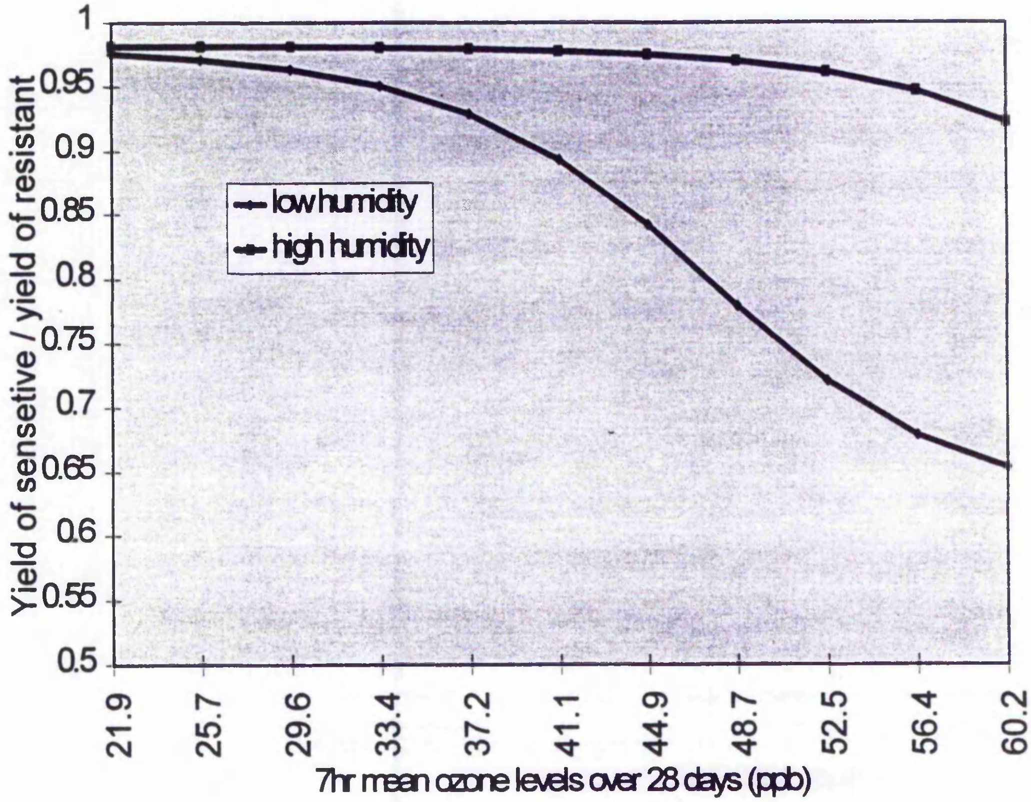


Figure 8.2 Effect of different humidities on the ozone - yield relationship.

Conventional wisdom says that ozone damage is more likely to occur at higher humidity levels. This is not, at first sight, reflected in the graph. This is where the distinction between 'damage' and 'yield' must be made. One possible reason for this result is that the network was trained on crops grown outdoors where high humidity (and the weather conditions it signal) may be more beneficial to a crop for growth reasons and outweigh the accompanying increase in ozone damage. This is where outdoor experiments suffer statistical comparison problems that closed chamber experiments do not. In a closed chamber environment where the only variable change are humidity and ozone, a reversal of the two yield curves may be observed as the effect of the ozone damage is taken in isolation.

8.5 Conclusions.

These results show that once networks have been optimised and some meaning given to each hidden unit, the behaviour of an ANN can be explained Anyone thinking of implementing an ANN for any problem solving task needs to be able to take sections of an ANN and explain what this part is doing, this transparency is one of the most important factors for ANN adoption. Making the ANN believable should be of as high a priority as making it predict well, as both are equally important if the ANN solution is to be 'sold' to an implementor. Even for the purposes of self testing and evaluation, generating some meaningful rules of thumb, that make up part of the ANN's functional capabilities, can be very useful.

Chapter 9. Hybrid Methods for Search Space Reduction

9.1 Introduction

If irrelevant or redundant data is presented to an ANN its task of generalisation is made more difficult as its clean sheet approach treats all inputs equally. It is argued that it should be able to apportion weights correctly so to ignore irrelevant or redundant inputs, which is true, but while it is computing these weights it could be working on more important tasks. If too much time is wasted handling un-needed data, the required generalisation task may become unfeasible to compute.

9.1.1 The curse of dimensionality

In modelling or problem solving the search space is as big as all the available models or solutions. An ANN tries to map the input space to output space, to do this it needs to represent every part of the input space if it is to be an accurate mapping device. As the number of input space dimensions increases so does the difficulty of mapping, this is the 'curse of dimensionality' (Bellman 1961). The growth of the search space increases exponentially with each new input that is added to the model. This is made more dramatic if the inputs can take continuous values rather than binary ones[7]. The curse of dimensionality means that networks with irrelevant inputs behave less well than

[7] Binary inputs expand exponentially in a 2^X way, whereas a continuous input with N possible values between 0 and 1 expands in an N^X manner.

networks only consisting of relevant inputs, because an ANN will model irrelevance in addition to the relevant features.

9.1.2 Analysis of previous approaches

There are many ways of reducing the search space for a given problem. Removing inputs is one way. An input can be removed if its removal has little or no effect on the network training and generalisation performance. Removing an irrelevant input may even improve a networks performance. Some preprocessing algorithms reduce the search space, for instance, principal component analysis (PCA) is sometimes used to convert raw data into variables that are not correlated with each other (Khoshgoftaar *et al.* 1997). Another way to reduce the search space would be to focus different types of inputs into relevant channels. For example if you wanted to model the greenhouse temperature you would not use external light levels in a negative way. By not allowing a network to explore such regions you reduce the search space. The downside to this is that it removes the networks' clean sheet approach which is one of an ANN's strengths.

9.2 Dual Simplified Networks

The approach of 'divide and conquer' is a useful technique in many realms of problem solving. Here two distinct, but important, aims of the network are solved separately, then the results combined to give a final overall result. By doing this the search space can be effectively reduced but care must be taken to retain enough complexity in key areas to enable problem solving.

9.2.1 Network description.

It has been stated several times that ozone injury was more complex than the levels of ozone, but equation synthesis and CAPing pointed to quite distinct roles for ozone levels over time and modifying factors, this could be seen by the how different hidden units. So an architecture was constructed that consisted of two smaller ANNs, the first looking at ozone inputs over time, the second looking carefully at the levels of other pollutants in less detail (fig. 9.1). This vastly reduces the search space, but does limit the possibilities for the detection of features involving the both ozone AND other factors. The performance of the network would tell us if these features were an important part of the underlying behaviour of the system.

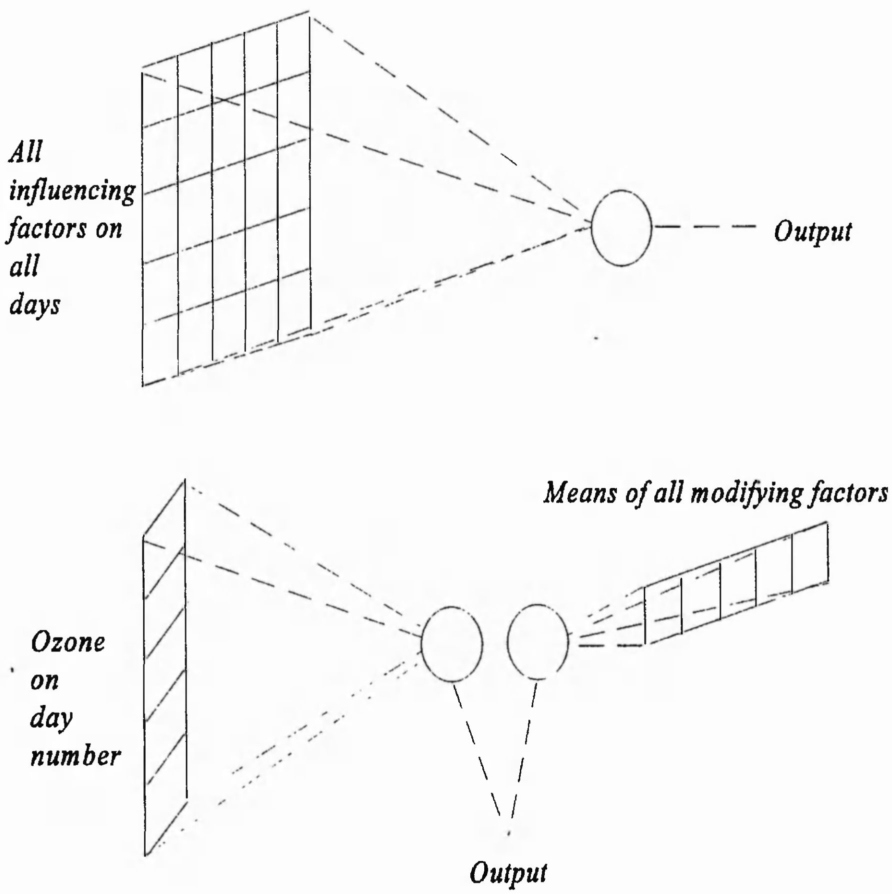


Figure 9.1. Design for a dual simplified ANN.

9.2.2 Performance of dual networks

If two networks are trained to model the same problem using different input parameters, their predictions must be combined in a logical way. There were several options available to carry out this dual comparison. The outputs could be averaged and the average value checked to see if it was above or below 0.5. A logical operation could be carried out (eg. or, and, exclusive or) to decide if the predicted output. The former approach was used, taking the mean of the two outputs generated by the sub-networks (figure 9.2). It is interesting to note that more mis-classifying was observed in the outputs from each individual sub-network than when the outputs were combined. Training results were less than perfect, in a binary threshold sense at least, with 7 errors in the 57 patterns.

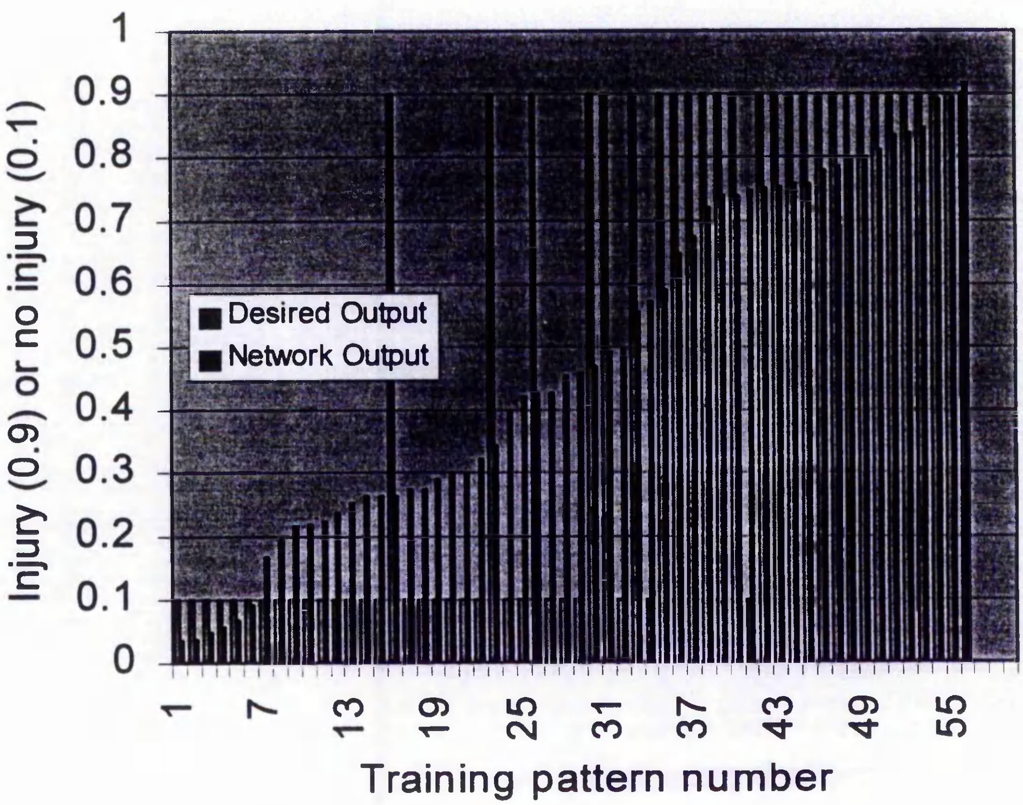


Figure 9.2. Training set performance for dual network.

In the testing arena, performance was similar with 2 of the patterns getting mis-classified (fig.9.3).

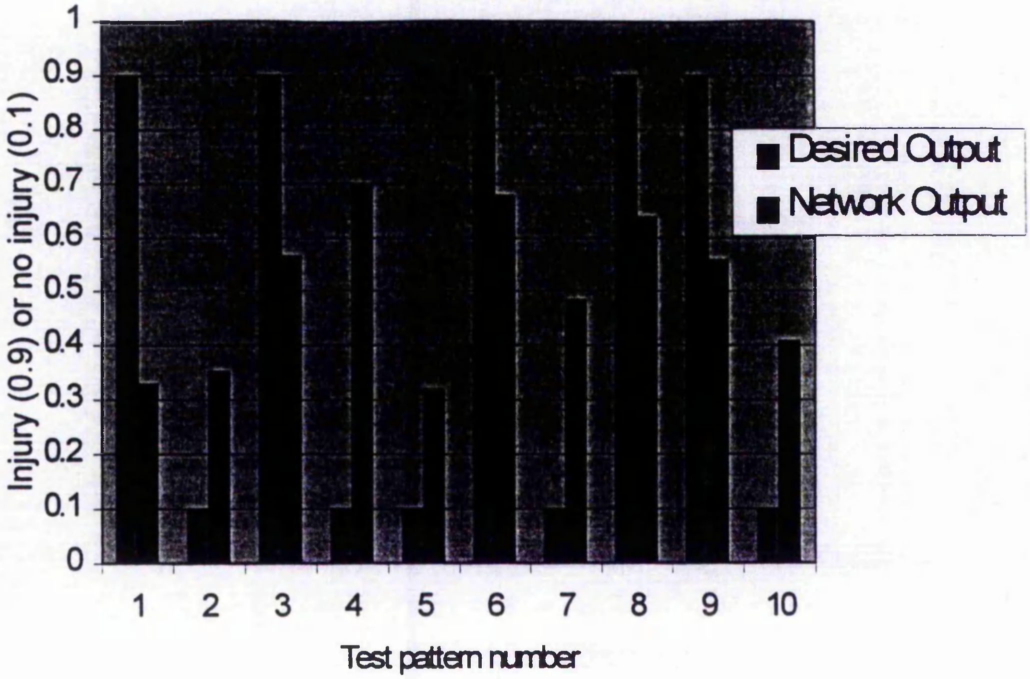


Figure 9.3. Testing performance of dual network.

More successful training and testing results could be achieved if a more generous merging technique is used. Doing a binary 'or' on the results from both networks gives a 100% classification performance (all the outputs for all patterns when merged, passed at least one of the two classifying thresholds). This method is only useful for assessing grades of performance, as no validation on unknown outputs could be achieved using this 'or' approach.

9.2.3 Analysis of dual networks

There seems to be some degree of predictive power possible using a dual network approach, suggesting that the roles of ozone and other factors have a linear relationship, at least to some extent. This also aids meaning extraction, due to the clear cut nature of each component. Both training and testing performances, and therefore generalisation, is poorer than a full single ANN. This suggests some generalising ability may require some cross variable, non-linear relationship. This approach looks promising and would benefit from more research into best methods for combination of results and a robust data set partitioning algorithm.

9.3 Preprocessing

Pre-processing takes into account any procedure carried out on the data. The simplest form of pre-processing could be removing obviously irrelevant inputs. Though at first this might sound like common sense, even this pre-processing is fraught with danger. For example, if you had a network predicting weather patterns you might choose to ignore days of the week as an input (thereby assuming this is irrelevant), but the influence of weekday work practices can make an important impact on temperatures. It is therefore very important to be sure any input that is removed from the network is irrelevant.

Another form of pre-processing that has been used throughout this research is re-normalising of the input values. Data is not usually normally distributed, if the data points are too skewed it is useful to modify the data. Suppose there were 5 data point for one input: 0.6, 0.7, 0.75, 0.8, 0.9. Firstly, by subtracting 0.25 from all values a mean of 0.5 could be achieved, giving: 0.35,0.45, 0.5, 0.55, 0.65. This range only uses 30% of the allowed 0 - 1 range (0.35-0.65) so by applying the equation $X^1 = X - (0.5-X)$ as set of:

0.05, 0.35, 0.5, 0.65, 0.95.

This range both uses the full possible input range and retains a mean of 0.5.

It might also be useful to modify the data into binary inputs, so using the above data set, and a threshold of >0.5 a data set of 0,0,0,1,1. could be formed. This is an arbitrary threshold but sometimes a threshold might have some meaning (ie. The light reading that divides night and day).

9.3.1 Why Pre-process data?

Pre-processing of data can take several forms but the fundamental idea is to reduce the volume of data without losing any of the information contained within. If irrelevant inputs can be removed, or aggregated linearly with other inputs, this reduces the size of the generalisation problem for the ANN.

9.3.2 Principal Component Analysis(PCA)

The aim of PCA is to redraw the axis for data of n dimensions so that points lie as close as possible to the axes. The derive principal components can express a large proportion of total variance of data with a smaller number of variables.

Suppose you have samples located in problem space. If you could simultaneously envision all variables then there would be little need for ordination methods. However, with more than three dimensions, help is usually needed. What PCA does is take a cloud of data points, and rotates it such that the maximum variability is visible. Another way of saying this is that it identifies your most important gradients.

Let us take a hypothetical example where you have measured three different variables, X1, X2, and X3. In this example, suppose it is possible (though it might be difficult) to tell that X1 and X2 are related to each other, and it is less clear whether X3 is related to either X1 or X2. The task is then to determine whether there is/are a hidden factor(s) or component(s) along which our samples vary.

The first stage in rotating the data cloud is to standardize the data by subtracting the mean and dividing by the standard deviation. Thus, the centroid of the whole data set is zero.

The relative location of points remains the same. Principal Components Analysis chooses the first PCA axis as that line that goes through the centroid, but also minimizes the square of the distance of each point to that line. Thus, in some sense, the line is as close to all of the data as possible. Equivalently, the line goes through the maximum variation in the data.

The second PCA axis also must go through the centroid, and also goes through the maximum variation in the data, but with a certain constraint: It must be completely uncorrelated (i.e. at right angles, or "orthogonal") to PCA axis 1.

We have only plotted two PCA Axes. However, there exist three axes in the data set. The third axes is not plotted as it ends up with a diagram that is as complicated as the one we start with.

How do we determine how many axes are worth interpreting? Ultimately, this is left up to the reasons for the investigation. But a big hint can be found with the eigenvalues. Every axis has an eigenvalue (also called latent root) associated with it, and they are ranked from the highest to the lowest. These are related to the amount of variation explained by the axis. The sum of the eigenvalues is 3, which is also the number of variables. It is usually typical to express the eigenvalues as a percentage of the total, so suppose we had the following values:

PCA Axis 1: 63%

PCA Axis 2: 33%

PCA Axis 3: 4%

Our first axis explained or "extracted" almost 2/3 of the variation in the entire data set, and the second axis explained almost all of the remaining variation. Axis 3 only explained a trivial amount, and might not be worth interpreting.

PCA is extremely useful when we expect variables to be linearly (or even monotonically) related to each other. Unfortunately, we rarely encounter such a situation in nature. But as a form of preprocessing, drastically simplifying the range of input variables, it seemed an interesting procedure to try, with respect to the theme of this research.

9.3.3 Performance of PCA Pre-processed ANN.

An automated PCA program generated Eigen vectors from the 96 available inputs only 7 were above a 1% importance threshold (table 9.1), so the composite values for these vectors were used. A 100% training prediction result was achieved (fig 9.4) , but not without considerable fine tuning and with a large amount of residual error in the actual values (see pattern 29, actual 0.1, predicted 0.48).

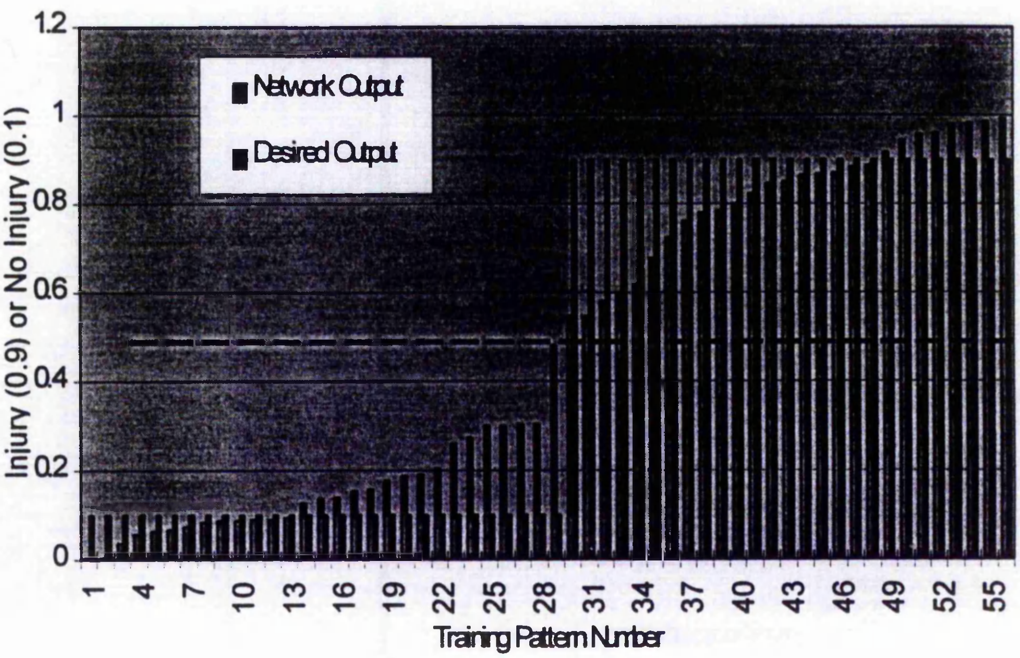


Figure 9.4. Training set for PCA preprocessed network.

Component Number	Eigenvalue	Cum Variance	Percent	Cum Percent
1	11.2397	11.2397	0.3038	0.3038
2	8.7395	19.9792	0.2362	0.54
3	4.2106	24.1898	0.1138	0.6538
4	3.3675	27.5573	0.091	0.7448
5	1.9739	29.5312	0.0533	0.7981
6	1.5499	31.0811	0.0419	0.84
7	1.3644	32.4455	0.0369	0.8769
8	0.772	33.2175	0.0209	0.8978
9	0.6921	33.9097	0.0187	0.9165
10	0.5717	34.4814	0.0155	0.9319
11	0.4823	34.9637	0.013	0.945
12	0.4005	35.3642	0.0108	0.9558
13	0.2853	35.6495	0.0077	0.9635
14	0.2706	35.9201	0.0073	0.9708
15	0.1708	36.0909	0.0046	0.9754
16	0.1379	36.2287	0.0037	0.9792
17	0.1165	36.3452	0.0031	0.9823
18	0.103	36.4482	0.0028	0.9851
19	0.0899	36.5381	0.0024	0.9875
20	0.0778	36.6159	0.0021	0.9896
21	0.0586	36.6745	0.0016	0.9912
22	0.0503	36.7248	0.0014	0.9926
23	0.0418	36.7667	0.0011	0.9937
24	0.0344	36.8011	0.0009	0.9946
25	0.0318	36.8329	0.0009	0.9955
26	0.0303	36.8631	0.0008	0.9963
27	0.0249	36.888	0.0007	0.997
28	0.0219	36.9099	0.0006	0.9976
29	0.0207	36.9306	0.0006	0.9981
30	0.0167	36.9474	0.0005	0.9986
31	0.0139	36.9613	0.0004	0.999
32	0.0111	36.9724	0.0003	0.9993
33	0.0096	36.9819	0.0003	0.9995
34	0.0068	36.9887	0.0002	0.9997
35	0.005	36.9937	0.0001	0.9998
36	0.0036	36.9973	0.0001	0.9999
37	0.0027	37	0.0001	1

Table 9.1. Principal components for injury predicting network.

Testing of the network yielded 1 incorrect classification and several borderline ones (fig 9.5). This performance is worse than a straight 96 input ANN, possibly due to some removal of important data in the Eigen vector threshold. If too many Eigen vectors had been used the benefit of pre-processing would have been negated. Even if performance had been improved, the problem off explaining how the ANN worked, using CAPing and/or equation synthesis would have been made more difficult with the added combinatorial step.

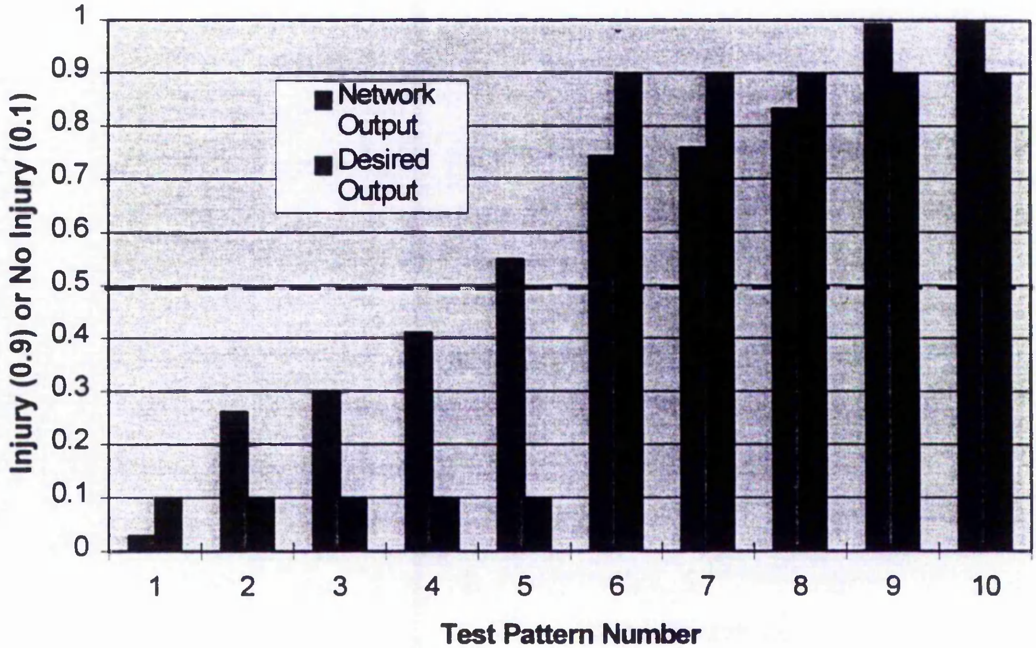


Figure 9.5. Test sets for PCA preprocessed network.

9.4 Rule Based Approaches

A rule based approach is really a more decipherable, semantic embodiment of an ANN. A rule set is applied to a vector within the set in a sequential and logistic way, until the pattern is classified. Rule based approaches are appealing because they are very readable, but it can be argued that anything that is classifiable using a rule tree is solvable using an ANN and the ANN will be a better generaliser.

9.4.1 What are rule based approaches

Real world data contains an abundance of features, some of which have analytical uses. A rule based system uses discriminating features to classify sets of data into useful subsets. It usually requires an expert to decide the rules and the resulting system is sometimes called an expert system (Liebowitz 1988). There are several approaches for automatic selection of rules, the most popular of which is the ID3 algorithm (Dietterich *et al* 1995). ID3, which stands for ``Induction of Decision Trees", is a supervised learning system which constructs classification rules in the form of a decision tree. It takes a set of objects, the training set, as input, and builds the decision tree by partitioning the training set. Attributes are chosen to split the set, and a tree is built for each subset, until all members of the subsets belong to the same class.

The main benefit of a symbolic approach to modelling or problem solving is the inherent self description of the solution, this is in contrast to the non-symbolic, black-box appearance of an ANN. ANNs have a very good record in solving non-linear modelling problems (eg. Funahashi, 1989), much better than simple rule based systems, this is due to the non-linear nature of their structure and their automated learning algorithms. Their adoption into mainstream engineering is less impressive.

9.4.2 Basic Rules Applied to Ozone Data.

Basic rules are intuitive statements that are well known in pollution modelling communities. One such statement is "Ozone can cause the onset of injury if cumulative AOT40 levels are greater than X ppb over the 5 preceding days". Rules of this type were tested but yielded prediction performance levels of less than 60%, even when the value for X was varied to an optimal level. This is a very poor result given that a guess would achieve a performance level of 50%. This shows again how difficult and complex the interactions are that cause ozone injury, when ozone is most definitely the injuring factor

9.4.3 ANN derived rules

It has been shown in this report that equations synthesised from ANNs can be successfully used to format a minimal ANNs It was hoped that the contents of these equations could also be converted into a set of rules to discriminate pollution and climatic conditions with differing effects. The following equation was taken:

Occurrence of injury =

F[Day2(MaxLight-AOT40) -Day1(7hrmean RH) - Day3(AOT40 + 7hrmeanO3) - Day4(AOT40) +Day5(24hrmeanLight+Maxlight)]

+F[Day3(MaxOzone) - Day1(7hrmeanRH) - Day 4(7hrmeanRH+MaxRH) - Day5(AOT40 +7hrmeanOzone +7hrmean RH)]

+F[Day2(24hourRH+24hourLight)+Day3(AOT40+7hrmeanOzone+MaxOzone) - Day5(AOT40) - Day1(MaxRh)]

These rules were then applied directly to the data set. Due to the complex nature of the data set it was not expected that these coarse rules would achieve the same performance as the complete (or even minimal) network but the performance was still disappointing at between 60 and 70%. The loss of the non-linear-nature of an ANN is shown to have a large effect on performance. It can be concluded that the data set is full of non-linear affecters that cannot be described by the strict interpretation of the ANNs key weights.

A selection of 'soft' rules were applied to the data sets with the aim of assessing the performance of more general guidelines in predicting the onset of ozone related leaf injury. The results of weight analysis and equation synthesis were examined to generate these rules, but knowledge of plant physiology was also used.

The rules were:

Mean Rule 1 - If age + AOT40 > average then injury occurs.

Mean Rule 2 - If 7hOzone-(aot40+24h Light+7hVPD) > average then injury occurs.

Mean Rule 3 - If 7h light+ maxlight+ 24hr VPD

Ozone Only Rule 1- If ratio of ozone levels on day 4 to other 4 days > average then injury occurs.

Ozone Only Rule 2 - If 7hr mean >Y1 and Max ozone > Y2 the injury occurs.

The rules generated differed from the strict ANN created rules in that there was some human influence and some modification of the rules to improve performance. For example, the thresholds are specifically chosen to optimise performance, giving the correct proportion of injury/no-injury outputs. Table 9.2 shows the comparative performance. They show that adding some expert knowledge and optimising the values based on performance brings the performance up to 70% plus. This is approaching the performance of various ANNs but is still a significant amount off.

These performances are inferior to ANNs but do show a degree of useful predictive ability. A combination of rules, using 'and' and 'or' comparisons, offers the best rule based approach. This gives some credence to the ANN approach and the rules generated from it. It would be interesting to see which decision making process field researchers would use given the choice of the 'black box' ANN or the self descriptive rule trees, but this is moving into psychological research and is therefore beyond the bounds of this research.

mean ANN	Ozone ANN	both ANN	mean rule1	mean rule2	mean rule3	combine means	oz rule1	oz rule2	oz and mean comb
85%	90%	95%	60.6%	59.1%	62.1%	68.1%	66.7%	63.6%	72.7%

Table 9.1 Performance of various rule based approaches.

9.4.4 Rules as a form of preprocessing

It was considered possible to carry out some pre-processing of the data sets to eliminate some of the patterns that were easily classified. It was hypothesised that an ANN trained on the remaining data would be more focussed on the more complex interaction involved in the development of ozone related leaf injury. Any meaning or equations derived from these networks would contain less obvious indicators and would possibly give clearer indication to the mechanisms involved in ozone injury.

Rules were generated using expert knowledge and deductions made from equations derived from neural networks. These were four rules that were applied to the data set of five day mean levels of ozone, VPD, solar radiation and age (Fig 9.6).

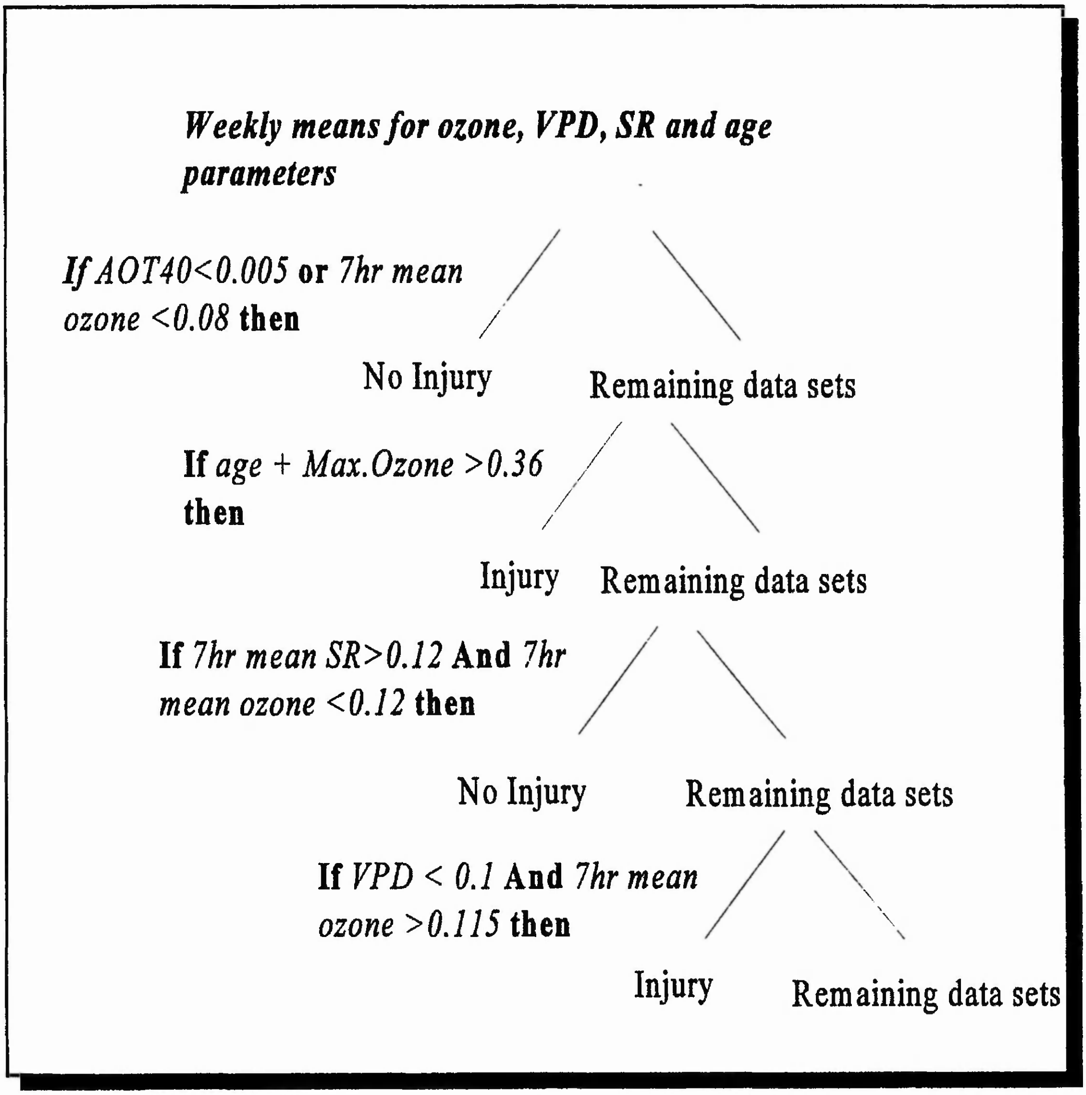


Figure 9.6. Decision tree for pre-processing leaf damage data.

Figure 9.6 has normalised values for it's rule thresholds, these convert as follows:

Note: See below for conversion of normalised values:

Normalised AOT40	= 0.005	= 6ppb.h
Normalised 7hr mean ozone	= 0.08	= 30ppb
Normalised 7hr mean Solar radiation	= 0.12	= 530 WM^{-2}
Normalised 7hr mean VPD	= 0.1	= 7kP

This rule tree removed 23 of the 65 available data sets by correctly predicting the injury onset state. The remaining 43 data sets were divided into 33 training patterns and 10 test patterns.

The two ANN models (daily ozone, 5 day means) were trained (fig 9.7) and tested (fig 9.8). It was clear that although the rule-tree pre-processing does not impair the training of a network, ~90% correctly classified (in fact training was achieved with fewer hidden units), it does have an effect on how well a neural network can generalise to an unseen test set with a classification rate of 20%. The poor generalisation points to a need for the more simple indicators within a training set for good generalisation. These clear indicator sets were removed by the rule pre-processing and thus were not available for the network to use .

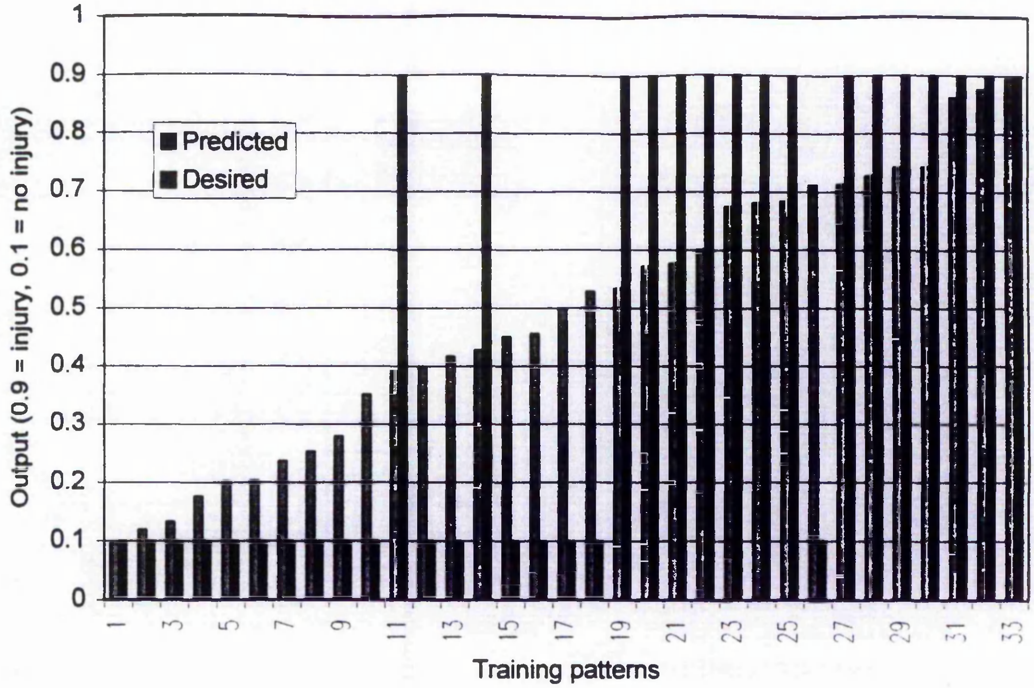


Figure 9.7. Performance of rule tree pre-processed network on training set.

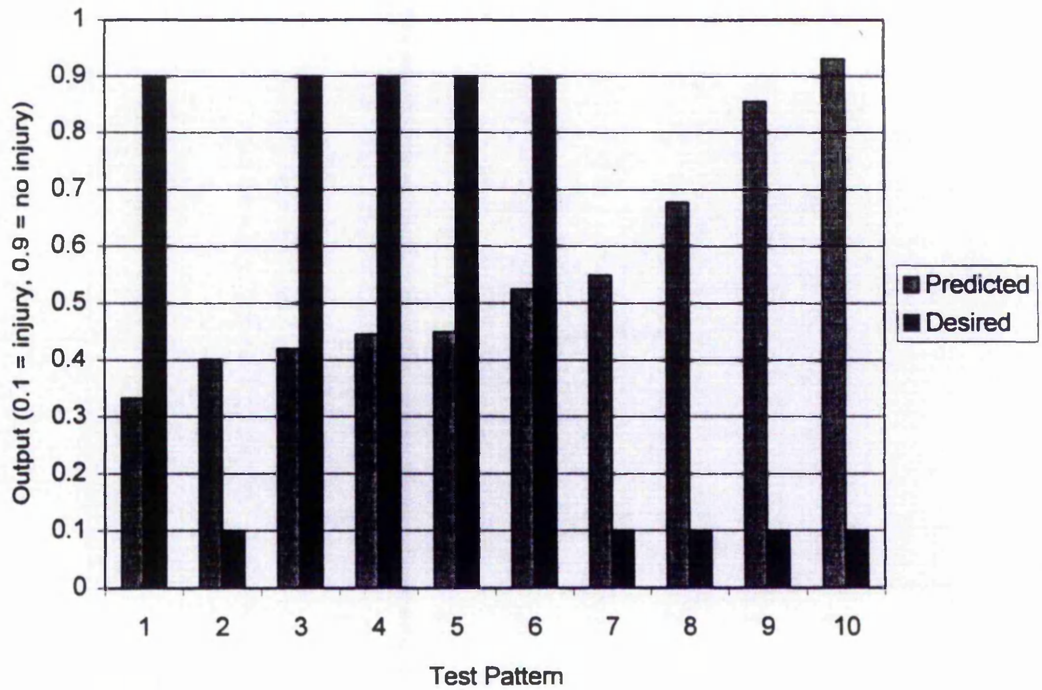


Figure 9.8. Performance of rule tree pre-processed network on test set.

9.5 Discussion.

Rule based systems are very attractive to people in charge of managing and administering systems. They offer humanly parsable sentences that, when applied correctly, can give a satisfactory performance. This chapter tried to splice some rule-based methods and pre-processing into the ANN methodology in an attempt to make the predictions more transparent and possibly better. The attempts made have little effect on either goal. Inserting PCA into the modelling makes the process *less* transparent with little impact on performance. Rule based methods, whether ANN derived or not, are unable to offer an adequate performance. Adopting a rule based system must, therefore, be assumed as victory for transparency over performance. The work in this thesis is intended to make this short-sighted decision a little more difficult to make by making an ANN's functionality more transparent.

10. Conclusions and Future Work

The aims of this project can be summerised as:

1. To take complex data sets of agricultural crop data and apply ANNs in an effort to predict levels of injury and yield reduction in the crop.

2. To develop novel methods for simplifying any resulting networks.

3. To investigate methods of generating a heuristic explanation for the resulting network weight values.

4. To study the possibilities of hybridising multiple techniques to produce the optimum balance of predictive power and transparency of solution provision.

The following sections draw conclusions on what has been learned in the course of this research.

10.1. Design and implementation of an ANN based predictive model.

A unique data set was made available that conformed to all the requirements needed for the tasks set out for this research. The cause-effect relationship was not fully understood, there were multivariate and non-linear factors involved, input variables had complex interdependencies and manual data collection was used (proberbly introducing error and noise). This data set was one of climatic and pollution conditions and the resulting lifespan of an associated crop.

Performance in predicting onset of leaf injury and degree of yield reduction, based on pollutant and environmental conditions, was good. Injury predictions for a randomly selected test set were within acceptable bounds. The complex, multivariate nature of pollutant/environment interactions were more suitable to the non-linear, error minimising nature of ANNs than to other available methods. Its was also possible, by varying key inputs, to identify critical levels of pollutants that when present in the right conditions lead to injury or yield loss.

10.2. Investigation of weight minimisation techniques.

The most successful weight minimising (or pruning) technique developed was CAPing (Roadknight *et al.* 1997c). By removing nodes, the ANN is made simpler. This has two main benefits; it makes the ANN more interpretable; and it makes further training (or retraining) more likely to succeed in a shorter time. CAPing works well because it looks for redundancy in the network activations, an unambiguous point at which to make comparisons.

10.3. Development of simple rule extraction techniques.

The rule extraction techniques were in the form of a simple thresholded application of a set of heuristics (Roadknight *et al* 1995, Roadknight *et al* 1997a, Roadknight *et al* 1997b). This approach may not always be successful, but the degree of success can be easily checked by retrospective analysis of the extracted rules performance. The success rate can be improved by ensuring the ANN only includes 'important' inputs and connections, algorithms such as capping and careful preprocessing help ensure that this is the case. These simple rules have 2 roles, firstly, they may be sufficient to apply as a stand alone technique for prediction and secondly they allow external assessment by non-ANN specialists. The weights of an effective ANN are often incomprehensible, leading to a resistance to their adoption, if a degree of perspicuity can be achieved, an ANN solution is more likely to be adopted.

10.4. Hybrid approaches to creating interpretable neural networks.

Several attempts were made to decompose the task set for the neural network, these included adding a stage of principal component analysis (PCA), adding a rule based processing stage, and breaking the ANN approach up into several steps. In theory this would make the ANN step simpler and more tractable. This was partially successful. Some degradation of performance was perceived when an intuitive rule based pre-processing step was added, this was also the case when the ANN was broken down into 2 simpler ones acting on a subset of the data. A PCA step failed to make the ANN step more tractable due to its creation of multivariate Eigen vectors. The investigation into these hybrid methods was far from comprehensive and these techniques may ultimately lead to more compact ANN weight vectors.

10.5. Further work on the same data.

Firstly, any new data should be passed through the trained ANN, these would act as a new test set. If the ANN failed to correctly identify the onset of damage, or approximate yield reduction two theories could be hypothesised.

1. The model's generated were influenced by time, some unused time dependent variable influences plant pathology in a way that differs year on year. This can only be corrected by identifying the time dependent factor.

2. The model was insufficiently trained, depending on the quality of the original data set, and the complexity of the problem, more data sets may reveal new facets of the problem that were not modelled in the original analysis.

10.6. Further work on methods.

Further work on the hybrid methods would be desirable. It seems clear that any method that can take the favourable elements from several data modelling techniques must be profitable, finding the right balance of these techniques is the only obstacle.

10.7. Summary.

ANN's are central to this research in that they are the tool used to generalise complex, multi-dimensional, non-linear cause-effect problems. This is a novel use of an existing A.I technique. Having established that existing techniques were inadequate, novel techniques (eg. CAPing) are devised to both optimise the application of ANN's and also to explain the hidden mechanisms involved in a trained neural net.

This work has shown that ANN's can be successfully applied to the realm of pollution impact modelling, the success of this is due, in part, to the degree of complex, partially non-linear interactions found in real world biological systems. Useful diagnostic and optimising methods have also been applied and evaluated. ANN's have been shown to generalise good solutions to problems that occupy too large a search space for a brute force search, so the results shown were not unexpected, but the methods shown for extracting heuristics are an important bonus.

Chapter 11. References

Aben JMM, Janssen-Jurkovicova M and Adema EH. 1990. Effects of low level ozone exposure under ambient conditions on photosynthesis and stomatal control of *Vicia faba* L. Plant Cell and Environment 13, 463-469

Adams RM and Crocker TD. 1989. The agricultural economics of environmental change - some lessons from air-pollution. Journal of Environmental Management. 26, 295-307.

Ahmad S and Tresp V. 1993. Some solutions to the missing feature problem in vision. In Hanson SJ, Cowen JD and Giles CL (Eds). Advances in Neural Information Processing Systems, Volume 5, pp. 393-400. San Mateo, CA: Morgan Kaufman.

Amiro BD, Gillespie TJ and Thurtell GW. 1985. Injury response of *Phaseolus vulgaris* to ozone flux density. Atmospheric Environment 18:6, 1207-1215.

Andrews R, Diederich J and Tickle AB. 1995. A survey and critique of techniques for extracting rules from trained Artificial Neural Networks. Knowledge Based Systems Vol 8 (6, December) p. 373-389.

Balls GR, Palmer-Brown D, Cobb AH and Sanders GE. 1995. Towards unravelling the complex interactions between microclimate, ozone dose and ozone injury in clover. Journal of Water, Air and Soil Pollution. 85, 1467 - 1472.

Balls GR, Palmer-Brown D, & Sanders GE. 1996. Investigating microclimate influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks. New Phytologist, 132, 271 -280

Becker K, Saurer M, Egger A and Fuhrer J. 1989. Sensetivity of white clover to ambient ozone in Switzerland. New Phytologist 122, 235-243.

Bender J, Weigel H-J and Jager H-J. 1990. Regression analysis to describe yield and metabolic responses of bean (Phaseolus Vulgaris) to chronic ozone stress. Angew. Botanik 64, 329-343.

Bellman, R. 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press.

Benton J, Fuhrer J, Gimeno BS, Skarby L, Balls G, Palmer-Brown D, Roadknight C & Sanders G. 1996a. ICP-Crops and critical levels of ozone for injury development. Exceedences of Critical Loads and Levels. In Exceedences of Critical Loads and Levels. Eds. M. Knoflacher, J. Schneider and G. Soja. Umweltbundesamt (Federal Environment Agency) Wein, Austria. 97-112.

Benton J, Fuhrer J, Gimeno BS, Skarby L, Balls G, Roadknight C & Sanders-Mills G. 1996b. The critical level of ozone for visible injury on crops and natural vegetation (ICP-Crops). In: Critical Levels of Ozone in Europe: Testing and Finalising the Concepts. UN ECE Workshop Report. Eds. L Karenlampi and L Skarby. University of Kuopio, Finland. 44 - 57.

Bishop C M. 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Bull KR. 1991. The critical loads/levels approach to gaseous pollutant emission control. Environmental Pollution 69, 105-123.

Burke HB, Hoang A and Rosen DB. 1995. Survival function estimates in cancer using artificial neural networks. Proceedings of World Congress on Neural Networks. Vol II. p.748-749.

Craven MW and Shavlik JW. 1995. Extracting Tree-Structured Representations of Trained Networks. In Advances in Neural Information Processing Systems 8

Cullan MR. 1985. Linear models in biology. Ellis Horwood series in mathematics and its applications. Ellis Horwood, Chichester, England.

Davalo E, Niam P. 1990. Neural Networks. Macmillan Press. p. 111-112.

Dietterich TG, Hild H and Bakiri G. 1995 A Comparison of ID3 and Backpropagation for English Text-to-Speech Mapping. Machine Learning, 18(1), 51-80.

Dillon T, Arabshahi P and Marks RJ. 1997. IEEE transactions on neural networks: special issue on everyday applications of neural networks. (ISBN 1045-9227)

Elman JL. 1990. Finding structures in time. Cognitive Science. Vol.14 p179-211

Friedman, J. H. 1991. Multivariate adaptive regression splines (with discussion). Annals of Statistics, 19, 1-141 (March).

Fry JC. 1993. Biological data analysis. A practical approach. Oxford University Press.

Fuhrer J & Achermann B (eds). 1994. Critical levels for ozone; a UN/ECE workshop report. Swiss Federal Research Station for Agricultural Chemistry and Environmental hygiene CH-3097 Liebefeld-Bern, Switzerland. No 16

Funahashi K. 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks, 2, 183 - 192.

Garson G. 1991. "Interpretting Neural Network Connection weights." AI Expert. April 1991.

Goutte, C. 1997, "Note on free lunches and cross-validation," Neural Computation, 9, 1211-1215, ftp://eivind.imm.dtu.dk/dist/1997/goutte.nflcv.ps.gz.

Guzy MR and Heath RL. 1994. Responses to ozone of varieties of common bean (*Phaseolus vulgaris* L) New Phytologist 124, 617-625.

Hancock PJB. 1992. Pruning neural nets by genetic algorithm. Artificial Neural Networks, (2) p991-994.

Hassibi B and Stork D.G. 1993. "Second order derivatives for network pruning: Optimal Brain Surgeon" Advances in Neural Information Processing 5, 164-171. Morgan Kaufman.

Heagle AS, Miller JE and Sherrill DE. 1994. A white clover system to estimate effects of tropospheric ozone on plants. Journal of Environmental Quality. 23, 613-621.

Heck WW, Taylor O. and Tingley DT. (eds) 1988. Assessment of Crop Loss from Air Pollutants. Elsevier Applied Science, New York.

Hedges L and Olkin I. 1985. Statistical Methods for Meta-Analysis. Academic Press Inc

Kappen L and Haeger S. 1991. Stomatal responses of *Tradescantia albiflora* to changing air humidity in light and darkness. Journal of Experimental Botony. 42. 239 -253

Karin ED. 1990. A simple procedure for pruning back-propagation trained neural networks. IEEE Trans. Neural Networks. vol.1 no.2. p239-242.

Karlsson G, Sellden G, Skarby L and Pleijel H. 1995. Clover as an indicator plant for phytotoxicozone concentrations: visible injury in relation to species, leaf age and exposure dynamics. New Phytologist. 129, 355-365.

Kersteins G, Federholzner R and Lendzian KJ. 1992. Dry deposition and cuticular uptake of pollutant gasses. Agriculture, Ecosystems and Environment. 42, 239-253.

Khoshgoftaar TM, Allen EB, Hudepohl JP and Aud SJ. 1997. Application of neural networks to software quality modelling of a very large telecommunications system. IEEE Trans. Neural Networks, vol. 8 no. 4, 902-909

Kohonen T. 1995. Self-Organising Maps. Berlin: Springer Verlag.

Kurková V. 1992. Kolmogorov's theorem and multilayer neural networks, Neural Networks 5: 501-506,

Lapeer RJA, Dalton KJ, Prger RW, Forstrom JJ, Selbmann HK and Derom R. 1995. Application of neural networks to the ranking of perinatal variables influencing birthweight. Scandinavian Journal of Clinicla Laboratory Investigation. 55, Suppl. 222: 83-93

Liebowitz J. 1988. Introduction to Expert Systems. Mitchell Publishing, Inc 1988 (ISBN: 0-394-39141-1).

Lim T-S, Loh, W-Y, and Shih, Y-S. 1997. An empirical comparison of decision trees and other classification methods. Technical Report 979, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

http://www.stat.wisc.edu/~limt/compare.ps

Lin D and Dayhoff J. 1996. Network Unfolding Algorithm and Complexity Analysis. http://www.isr.umd.edu/TechReports/ISR/1995/TR_95-6/TR_95-6.phtml

Little R J A. 1992. Regression with missing X's: a review. Journal of the american statistical association 87 (420), 1227-1237.

MacKay, D. J. C. 1995 Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks, available at ftp://wol.ra.phy.cam.ac.uk/pub/www/mackay/network.ps.gz.

Masuoka R, Watanabe N, Kawamura A, Owade Y and Asakawa K. 1990. Neurofuzzy systems - fuzzy inference using a structured neural network. Proceedings of the International Conference on Fuzzy Logic and Neural networks. Iizuka, Japan p173-177.

Mensa HA, Haggestad HE, Street OE and Jeffrey RN. 1963. Response of plants to air pollutants. I. Effects of ozone on tobacco plants preconditioned to light and temperature. Plant Physiology 38, 605-609

Moldau H, Sober J, Sober A. 1990. Differential sensitivity of stomata and mesophyll to sudden exposure of bean shoots to ozone. Photosynthetica. 24, 446-458.

Mortensen LM. 1992. Effects of ozone concentration on growth of tomato at various light, air humidity and carbon dioxide levels. Scientia Horticulturea. 16, 17-24

Musselman RC, Heurta PM, McCool PM and Oshima RJ. 1986. Response of beans to simulated ambient and uniform ozone distributions with equal peak concentrations. J. Amer. Soc. Horticultural Sci. Vol 111. Pp.470-473.

Neal, R. M. 1996. Bayesian Learning for Neural Networks, New York: Springer-Verlag, ISBN 0-387-94724-8.

Orr RK. 1995. Use of probabilistic neural networks to predict mortality following cardiac surgery. Proceedings of World Congress on Neural Networks. Vol II. p. 754-757.

Otto HW and Daines RH. 1969. Plant injury by air pollutants: Influence of humidity on stomatal apetures and plant responses to ozone. Science. 163, 1209-1210.

Pearson DB and Mansfield TA. 1993. Interacting effects of ozone and water stress on stomatal resistance of beech (*fagus sylvatica L*). New Phytologist 123, 351-358.

Plaut D, Nowlan S and Hinton GE. 1986. Experiments on learning by back propagation. Technical Report CMU-CS-86-126, Dept of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Pleijel H, Wallin G, Karlsson G, Skarby L and Sellden G. 1995. Gradients of ozone at a forest site and over a field crop - consequences for the AOT40 critical level. Journal of Water, Air and Soil Pollution. 85. 2033-2038

PORG. 1993. Ozone in the United Kingdom. Third Report of the United Kingdom Photochemical Review Group.

Press, W.A. and Lee, C.W. 1996 V1 receptive fields reflect the statistical structure of natural scenes: A projection pursuit analysis. In: CNS 96 Proceedings, ed. by J. Bower, Plenum Press, NY, pp. 143-149.

Rosenblatt F. 1962. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. New York: Sparten Books.

Rumelhart DE and McClelland JL. 1986. Parallel Distributed Processing. The MIT Press, Cambridge, Mass. USA.

Sanders GE, Turnbull ND, Clark AG and Colls JJ. 1990. The growth and development of *Vica faba* L. in filtered and unfiltered open-top chambers. New Phytologist, 116, 27-78

Sanders GE, Balls G and Booth C. 1993. Ozone critical levels for agricultural crops-Analysis and interpretation of the results from the UN-ECE International Cooperative Programme for Crops. p. 58-72

Sanders GE, Skarby L, Ashmore MR and Fuhrer J. 1995. Establishing critical levels for the effects of air pollution on vegetation. Journal of Water, Soil and Air pollution. **85**, 189 - 200

Schut HE. 1985. Models for the physiological effects of a short ozone exposure on plants. Ecological Modelling 30, 175-207.

Sietsma J & Dow RJF. 1988. Neural net pruning - Why and how. Prc. IEEE Int. Conf. Neural Networks. Vol 1. p. 325-333.

Simpson R, Williams R, Ellis R and Culverhouse PF. 1992. Biological pattern recognition by neural networks. Marine ecology progress series 79, 303-308.

Steel R and Torrie JH. 1980. Principles and Procedures of Statistics. A Biometrical Approach, 2nd Edition, McGraw-Hill, Inc. New York.

Tan H, Prokhorov DV and Wunsch DC. 1995. Probabilistic and time-delay neural network techniques for conservative short-term stock trend prediction. Proceedings of World Congress on Neural Networks. Vol II. p. 44-47.

Tepper J, Powell H, Palmer-Brown D 1995. Integrating Symbolic and Subsymbolic Architectures for Parsing Arithmetic Expressions and Natural Language Sentences" Proceedings of 3rd SNN Neural Network Symposium, Nijmegen, Sept 1995, pp 81-84, Eds Bert Kappen and Stan Gielen, ISBN 3-540-19992-6.

Tickle AB, Orlowski M and Diederich J. 1994. DEDEC: decision detection by rule extraction from neural networks. QUT NRC.

UN/ECE. 1996. Critical Levels for Ozone in Europe. In Proc. Critical Levels Workshop. Kuopio. Finland. April 15-17

Van Pul WAJ, Jacobs AFG. 1993. The conductance of a maize crop and the underlying soil to ozone and other various environmental conditions. Boundary Layer Meteorology. 69, 83- 99

Webb A R, Lowe D and Bedworth MD. 1988. A comparison of non-linear optimisation strategies for feed-forward adaptive layered networks. RSRE Memorandum

4157, Royal Signals and Radar Establishment, St. Andrew's Road, Malvern, UK.

Weigend AS, Rumelhart DE and Huberman BA. 1991. Generelization by weight elimination with applications to forecasting. In Advances in Neural Information Processing (3). Lippmann R, Moody J and Touretzky D. Eds. p. 875-882.

Weiss SM and Kulikowski CA. 1991. Computer Systems that Learn. Morgan Kaufmann.

Wiersma FR, Poel M and Oudshoff AM. 1995. The BB neural network rule extraction method. Proceedings of 3rd annual SNN symposium on neural networks (eds. Kappen B and Gielen S) Springer-Verlag. 69-73.

Zhang S, Powell H, Palmer-Brown D 2000. Keyword Extraction from Stemming and Sense Information by Neural Networks. Proceedings of Int Conf on Artificial Intelligence 2000 (IC-AI'2000).

1996b. The critical level of ozone for visible injury on crops and natural vegitation (ICP-Crops). In: Critical Levels of Ozone in Europe: Testing and Finalising the Concepts. UN ECE Workshop Report. Eds. L Karenlampi and L Skarby. University of Kuopio, Finland. 44 - 57.

Benton J, Fuhrer J, Gimeno BS, Skarby L, Ball G, Palmer-Brown D, Roadknight C & Mills G. 1999. An international cooperative programme indicates the widespread occurrence of ozone injury on crops. Agriculture, Ecosystems and Environment. 1508 1-12.

Appendix 1. Copies of authored published papers.

Learning the Equations of Data

By C. M. Roadknight[1], D. Palmer-Brown[1] and G. E. Sanders[2].

1. Parallel Research Group, Department of Computing, The Nottingham Trent University,
Burton Street, Nottingham NG1 4BU.

2. Department of Life Sciences, The Nottingham Trent University,
Clifton Lane, Nottingham NG11 8NS.

The use of artificial neural networks (ANNs) for data modelling is now established [1, 2]. An extension to this is presented which enables the characteristic equations of complex datasets to be learned.

The effects of environmental factors and man made pollutants on agricultural crops is well known [3], but this is mainly through single factor, dose response experiments. What is largely unknown is the effect of differing levels of multiple factors, such as humidity, temperature and ground level ozone concentrations. Standard statistical techniques for prediction and decision making cope badly with the non-linearity and small sample sizes inherent in this problem. However, prediction can be achieved using an ANN approach. An ANN strategy was chosen because the predictive powers of ANN's are well documented and suitable for this problem [4].

Considerable success has been achieved in predicting damage with this method, demonstrating that the ANN is modelling the environmental and pollution effects. A 3-layer network is trained using a standard back propagation algorithm [5] on a selection of seven-day datasets and then tested on datasets that it has not been trained on. The input layer of the network consists of nodes associated with daily environmental or pollution related factors for each of the seven days during the growth period of a bean plant. Other inputs are encodings of the condition of the plant (number of leaves, level of damage at day 0). This amounts to over a hundred input units. The output for the network is the change in percentage area of leaf damage over the seven days.

The network is trained to a point where it can correctly predict the change in the level of damage over the 7 days in all 54 of the training datasets and it can also predict the level of damage, to a good accuracy (RMS 0.0053), for all 18 of the test datasets (fig 1).

In addition to predicting damage levels, the model is required to provide an interpretation of the interaction of many environmental factors. An algorithm has been devised that simplifies the weights of the ANN into a reasonably concise equation (typically 10-20 terms). In general this equation contains a mixture of both biologically understood rules and less understood indicators of damage. The former give credence to the equation, the

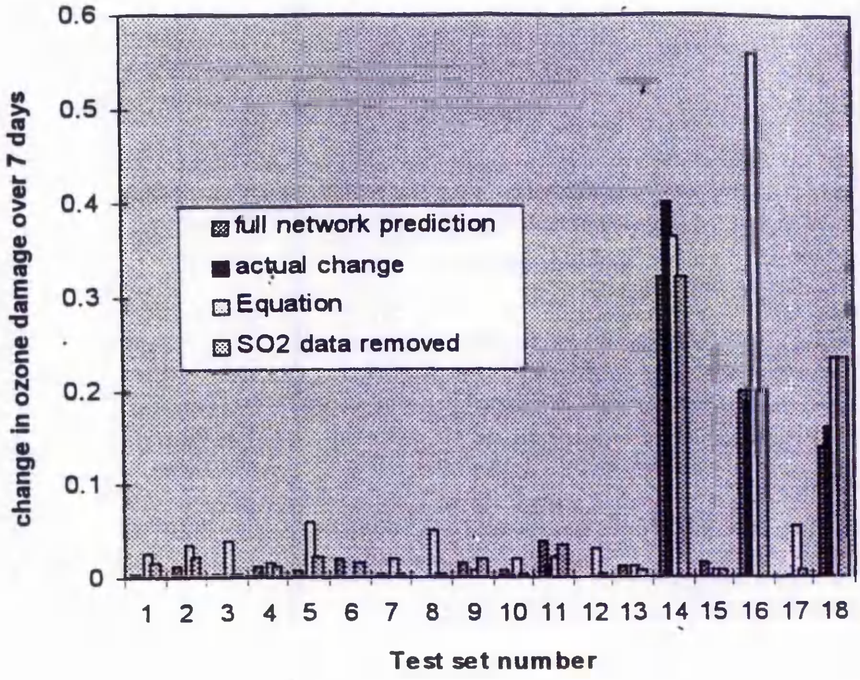


Figure 1

latter can then become the focus of biological experiments. The equation extractor works by building combinations of linear, piecewise models of the predominant neurons in the network. This is performed to any specified level of model complexity.

For the initial data, the following equation was extracted:

F[InitialLeafNumber - MaxSO2@d3 - (2*InitialLeafDamage)] +
F[InitialLeafNumber +InitialLeafDamage - (SumSO2@d2 + MaxSO2@d2 + MaxSO2@d3)]+
F[SumSO2@d1+SumNO2@d2-(MaxSO2@d3+ MaxSO2@d6+ MaxNO2@d6+ SumNO2@d5 + SumRad@d7+MaxNO@d7+InitialLeafDamage)]

The terms in this equation are environmental or pollutant factor levels eg. Number of leaves on plant at day 0 (InitialLeafNumber), highest sulphur dioxide reading on day 3 (MaxSO2@d3), sum of hourly readings of solar radiation during day 7 (SumRad@d7).

This equation contains the primary inputs for 3 hidden units, contained within an activation function (F), which is approximated as a straight line between 0 and 1, with a value of 0.5 at F(0).

The algorithm for network processing and equation extraction is as follows:

For the chosen maximum number of hidden nodes
  Set required importance level (H) for weights from hidden units to output unit
  Find all hidden units above H
For the chosen maximum number of input nodes
  Set required importance level (I) for weights from input units to hidden units
  Find all inputs to chosen hidden units above I
Create equation by matching input nodes, approximate weights and sign to source of input node value.

At present the number of terms in the equation (10 to 20) and the number of hidden units contributing to the equation (eg 2 or 3) are decided before extraction. The validity of the equation is tested by running a minimal network consisting only of the weights used to generate the equation. Even though this minimal ANN may contain as few as 1% of the connections of its predecessor, it performs the predictions to a similar standard (fig 1); correctly classifying the 3 large (>0.1) changes, the 2 small changes (>0,<0.01) and the 13 zero changes. The test set 16 prediction is overestimated by this network but it is still classified correctly as large (>0.1). The minimised network can achieve this good level of prediction because it is based on the key fetures of the dataset, as learned by the full network. This suggests that these simple equations can be used as predictive tools and as indicators and interpreters of the source of crop damage.

It is possible for the network to use secondary indicators which obscure the model being sought. One of the best performing equations for predicting ozone related leaf injury contained no ozone level inputs and contained many sulphur dioxide (SO_2) inputs and this seemed biologically improbable. There are three possible explanations:
a. Lower levels of SO_2 are a greater cause of ozone related injury than ozone. This may arise if sulphur dioxide offers a protection from ozone damage. This seemed unlikely
b. Less SO_2 indicates higher levels of ozone. If this is true it is necessary to find out why the ANN used a secondary indicator and not the factor itself.
c. Lower levels of SO_2 indicates higher levels of two or more factors, so sulphur dioxide might fall as a consequence of an increase in ozone levels and an increase in temperature levels, for example. So the network would increase the weight on the one multi factor indicator in preference to the several factors themselves.

To test this, a network was trained without all SO_2 related inputs. There are many possible models that an ANN can learn from one application. Their relevence depends on which aspects of the complete dataset are presented to the network. This ANN trains to a similar standard as it did with sulphur dioxide (fig 1), showing that the role of SO_2 is as a secondary indicator. The equation extracted from this trained ANN is as follows:

3*F[D2(SumOz)+D3(SumOz+MaxOz)-D1(SumRad+MaxRad)-2*(D6(MaxRad))+(4*InitialLeafDamage)]
+F[3*InitialLeafNumber - D3(MaxNo)]
+ F[3*InitialLeafNumber - 4*InitialLeafDamage]

The profiles of the sulphur dioxide were of small but significant difference whereas the differences between levels of ozone were much greater. The network weights needed to be larger to amplify the differences in the inputs to account for this, so SO_2 appeared more important in the equation. This can be overcome by normalising the input data, to between 0 and 1 but also to a universal mean and universal standard deviation.

In summary, this paper has reported how a method has been devised for extracting working equations from ANN's, so that their 'black box' nature is overcome and greater trust can be placed in their results.

References

[1] Ripley, B.D. Neural networks and related methods of classification. Journal of the Royal Statistical Society series B 56, 1994, pp. 409-456.

[2] Weiss, S.M. and Kulikowski, C.A. Computer Systems that Learn. Morgan Kaufmann. 1991

[3] Heck, W.W., Taylor, O.C. and Tingley, D.T. (eds). Assessment of Crop Loss from Air Pollutants. Elsevier Applied Science, New York. 1988

[4] Weigend, A.S., Huberman, B.A. and Rumelhart, D.E. Predicting the future: a connectionist approach. Int. Journal of Neural Systems, vol. 1, 1990, pp.193-209

[5] Rumelhart, D.E., Hinton, G.E., & Williams, R. J. Learning internal representations by error propagation. In D.E. Rumelhart, & J.L. McClelland (Eds), Parallel distibuted processing: Exploration in the microstructure of cognition (pp. 318-362). 1986. Cambridge, MA: MIT Press

IEEE TRANSACTIONS ON

NEURAL NETWORKS

A PUBLICATION OF THE IEEE NEURAL NETWORKS COUNCIL

JULY 1997 VOLUME 8 NUMBER 4 ITNNEP (ISSN 1045–9227)

SPECIAL ISSUE ON EVERYDAY APPLICATIONS OF NEURAL NETWORKS

GUEST EDITORIAL

PAPERS

ANNOUNCEMENTS

Modeling Complex Environmental Data

Chris M. Roadknight, Graham R. Balls, Gina E. Mills, and Dominic Palmer-Brown

***Abstract*— Artificial Neural Networks (ANN's) are used to model the interactions that occur between ozone pollution, climatic conditions, and the sensitivity of crops and other plants to ozone. A number of generic methods for analysis and modeling are presented. These methods are applicable to the modeling and analysis of any data where an effect (in this case damage to plants) is caused by a number of variables that have a nonlinear influence. Multilayer perceptron ANN's are used to model data from a number of sources and analysis of the trained optimized models determines the accuracy of the model's predictions. The models are sufficiently general and accurate to be employed as decision support systems by United Nations Economic Commission for Europe (UNECE) in determining the critical acceptable levels of ozone in Europe. Comparison is made of the accuracy of predictions for a number of modeling approaches. It is shown that the ANN approach is more accurate than other methods and that the use of principal components analysis on the inputs can improve the model. The validation of the models relies on more than simply an error measure on the test data. The relative importance of the causal agents in the model is established in the first instance by summing absolute weight values. This indicates whether the model is consistent with domain knowledge. The application of a range of conditions to the model then allows predictions to be made about the nonlinear influences of the individual principal inputs and of combinations of two inputs viewed as a three-dimensional graph. Equations are synthesized from the ANN to represent the model in an explicit mathematical form. Models are formed with essential parameters and other inputs are added as necessary, in order of decreasing priority, until an acceptable error level is reached. Secondary indicators substituting for primary indicators with which they are strongly correlated can be removed. From the synthesized equations both known and novel aspects of the process modeled can be identified. Known effects validate the model. Novel effects form the basis of hypotheses which can then be tested.**

***Index Terms*— Climate, crop damage, environmental data, equation synthesis, modeling, ozone, pollution, prediction, statistics.**

I. Introduction

THE methods described within this paper are intended to be applicable to any data set which involves complex nonlinear interactions between a number of causal agents and their influence upon processes. In this case the methods are applied to environmental data which is of this nature. The artificial neural network (ANN)-based models described in this paper are currently used in decision support for determinin the critical levels of ozone for visible injury to occur unde various microclimatic conditions. The setting of critical level is part of the UNECE International Cooperative Progran on effects of air pollution and other stresses on crops an nonwood plants (ICP Crops) for which the Nottingham Tren University, U.K., is the coordination center. The remit of th program is also to clarify the models to define the length c one damaging episode of ozone, to identify the importance c episode dynamics on the injury response, and identify whic climatic parameters have the greatest importance in influencin the ozone response. The equations generated as part of th modeling are currently used to examine the influence c climate in the different climatic zones of Europe. The method used in this program are also applicable to a wide range c applications of a similar nature.

Manuscript received November 20, 1996; revised March 5, 1997. C. M. Roadknight is supported by the U.K. Department of the Environment. This work was supported by the U.K. Department of the Environment Project PECD 7/12/145.

C. M. Roadknight and D. Palmer-Brown are with the Novel Architectures Group, Department of Computing, The Nottingham Trent University, Nottingham NG1 4BU U.K.

G. R. Balls and G. E. Mills are with the Department of Life Sciences, The Nottingham Trent University, Nottingham NG11 8NS U.K.

Publisher Item Identifier S 1045-9227(97)04909-6.

A. Problem Specifications

The effects of individual environmental factors and ma made pollutants on agricultural crops are to some extent know [1], but this is mainly through single factor, dose respons experiments. What is largely unknown is the effect of differin levels of multiple factors, such as humidity, temperature, ar ground-level ozone concentrations. Many crops suffer le damage and reduced yields due to ozone related injury [2 This damage can have both a qualitative and quantitative effe on the crop. A decreased yield of a crop is obviously le valuable but a crop of poorer quality is also less valuabl The ability to predict levels of injury based on climate ar pollution measurements is a desirable asset. Equally importa is the analysis of any artificial neural network with the abili to predict injury. This provides important information on tl modifying factors that influence ozone injury.

B. Existing Approaches

The injury causing effects of ozone and the influence microclimate conditions have been widely studied [3]–[5 These and related studies have shown that light, humidit temperature, and ozone concentrations all affect the mec anisms for ozone uptake of the plant [6]–[8] and therefo the uptake of atmospheric ozone. Ground level (tropospheri ozone concentration and the amount of ozone taken in l the leaf vary according to conditions. Another reason f the complexity and nonlinearity of the relationship betwe ozone dose and resulting injury is that once ozone has entere the leaf, the amount of damage it causes is modified l biochemical factors (e.g., photosynthetic rate and antioxida level, [9]). To simplify statistical complexity, many attemp to model these interactions fix some of the microclimat conditions or ignore their effects, thereby making any results dependant on a tight set of conditions [10], [11].

C. Statistics

Statistical meta-analysis has been applied in the area of biological sciences [12]. It derives confidence measures based on sample sizes, respective variances, and distribution. These methods, however, always involve some estimation of confidence limits or inherent data noise and for this reason are dependent on the expert knowledge of the analyzer. There are also a multitude of different ways of combining statistics each yielding different results. Standard statistical techniques used on their own, such as multiple regression analysis and principal components analysis, also fail to adequately model complex nonlinear biological phenomena because they attempt to linearize a nonlinear problem (see Section III-C). The most important advantages of ANN's over purely statistical approaches for this application are their ability to model nonlinear data and their nonreliance on previously assumed equations.

D. Rule Extraction

The knowledge that an ANN gains about a problem domain is encoded in the weights assigned to the connections of the ANN. Because the knowledge acquired is represented by a set of numbers, it is difficult to interpret in all but the most trivial of networks. The ANN acts as a "black box," taking in and giving out information.

Attempts have been made to dissect ANN's into a set of rules, for a full review of these methods see [13] and [14]. These methods usually aim for a set of if ... then conjunctive rules based on a "decompositional" approach at the individual unit level.

Demystifying of an ANN is desirable for many reasons. Systems that declare knowledge explicitly are adopted more freely; a rule base generated from an ANN is sometimes sufficient for accurate modeling [15], [16]; and in some cases the ability to explain how a solution is arrived at is essential (i.e., controlling temperature regulation in a nuclear power station).

This paper discusses two methods of extracting meaning from a trained network, relative importance measurements [17], [18], and equation synthesis [19]. The results from these methods are then used to implement smaller models, closer to the minimal model required for accurate prediction. These approaches are defined and discussed later in the text.

II. Prediction of Ozone Damage from Pollutant and Climatic Conditions

A. Data Sources

Data on ozone related injury and the accompanying levels for a variety of pollutants and climatic factors come from three sources:

1) Open topped chamber experiments carried out at the university of Nottingham experimental farm at Sutton Bonnington during the summers of 1989 and 1990. These provide very complete data for the green bean (*Phaseolus vulgaris*).
2) Outdoor experiments carried out at many sites around Europe as part of the UNECE International Cooperative Program on effects of air pollution and other stresses on crops and nonwood plants (ICP crops). The crops used within these experiments included green bean and several species of clover. These databases contain hourly readings for various environmental and pollutant factors and regular assessments of development and injury of the crop.
3) Closed-chamber experiments on subterranean clover carried out at the Nottingham Trent University at controlled ozone concentrations. This database contains information on the mean and summed microclimatic conditions for each day of exposure, mean ozone concentrations during exposure, ozone doses received on individual days of exposure, the length of exposure, and the percentage visible injury produced for these conditions.

B. Open-Top Chamber (OTC) Data

Data for the French bean was collected for the 1989 and 1990 growing seasons for two cultivars, Lit and Nerina. The plants were grown in open top chambers which allowed the modification of pollutant levels surrounding the plants while maintaining the fluctuations due to ambient climatic conditions [20]. The bean plants were grown in eight pollutant treatments, ranging from clean air with no pollutants to additions of 40 parts per billion (ppb) of ozone to ambient air. This simulated a range of ozone conditions in near-natural climatic conditions. The data was in the form of hourly readings of a number of pollutants (for example ozone in Fig. 1) and environmental parameters. The most significant of these were the parameters: ozone, sunlight, sulphur dioxide, nitric oxide, nitrogen dioxide, air temperature, and humidity.

This data was put into the form of a set of vectors with inputs and outputs for the neural networks. The data was formatted as seven-day blocks. The output data was the change in ozone induced damage from day zero to day seven and the change in number of leaves. The readings of each parameter, for each day, were compressed into a daily maximum and a daily sum (Fig. 2).

This compresses 24 hourly readings into two statistics which represented the conditions. Four extra inputs were also available from the source data:

1) Level of ozone damage at day zero, as this might be important to the network in assessing how much further damage can occur.
2) Number of leaves at day zero, as the network was initially asked to predict the change in the number of leaves. This gives another indication of the growth stage.
3) Growth stage at day zero.
4) Growth stage at day seven.

The total number of possible network inputs is 102. The database consists of 72 seven-day data sets with which training and testing could be carried out. These 72 data sets were

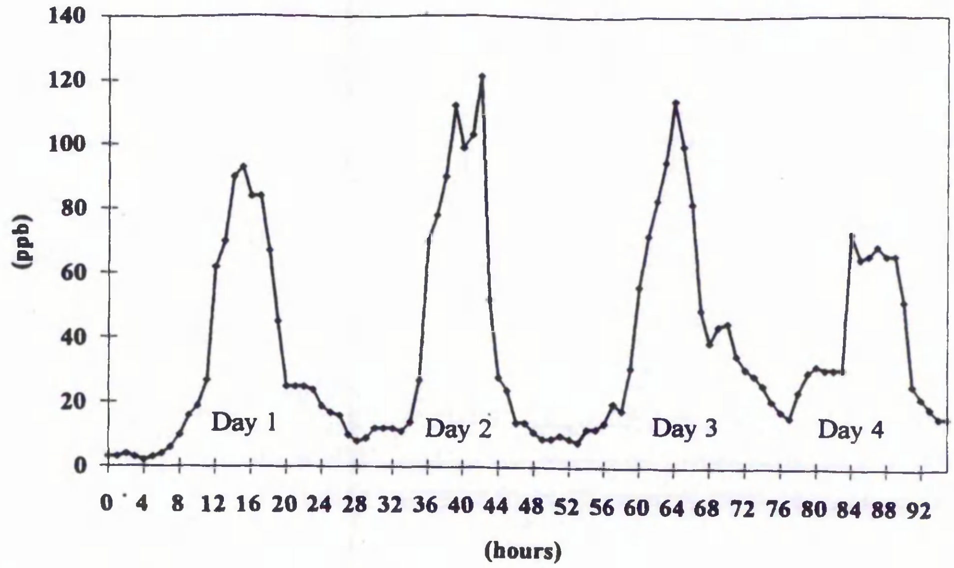


Fig. 1. Hourly ozone levels in nonfiltered air with 40 ppb additions of ozone.

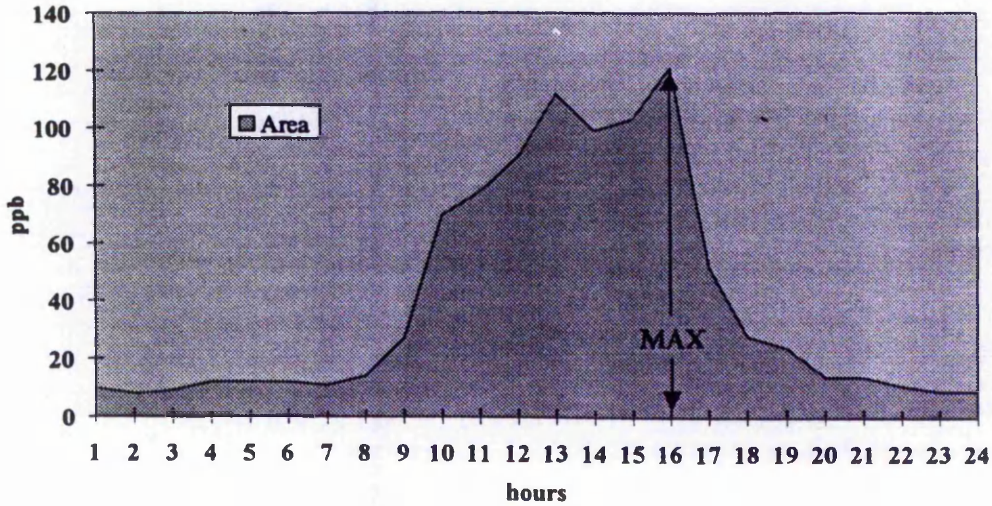


Fig. 2. One day of ozone readings.

separated into a training set and a test set of 54 and 18, respectively.

C. European Outdoor Data

In 1994 and 1995, participants of the UNECE ICP Crops program were asked to note and record when injury was first observed on white and subterranean clover.

The data used to train and test the ICP-Crops ANN's was collected at many European sites as part of the ICP Crops program. Countries that participated in 1994 and/or 1995 were Austria, Belgium, Denmark, Finland, France, Germany, Hungary, Italy, Latvia, The Netherlands, Poland, Russian Federation, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. Each participant conducts a series of coordinated experiments each year using a standard protocol (e.g., [21]) and seeds supplied by Nottingham Trent University. This data collection is ongoing; this means that any networks developed and tested using one or two years' values can be fur modified or reinforced by adding yearly data. The trai with data from many European sites over many years ma for exceptionally robust predictive networks.

The hourly means of parameters were compressed meaningful daily indicators: 7-h mean, daily maximum, a mulation over threshold 40 ppb (AOT40), and/or 24-h me

D. Closed-Chamber Data

Data were collected from closed-chamber exposures ov period of two years, generating 404 data points. These were used to determine an ozone dose response relationsh threshold above which ozone injury occurred and to investi the nature of the influences of microclimate, leaf age, the variation of dose with time upon the dose respons subterranean clover. The ozone dose response relation showed a large amount of scatter (Fig. 3), caused by additi

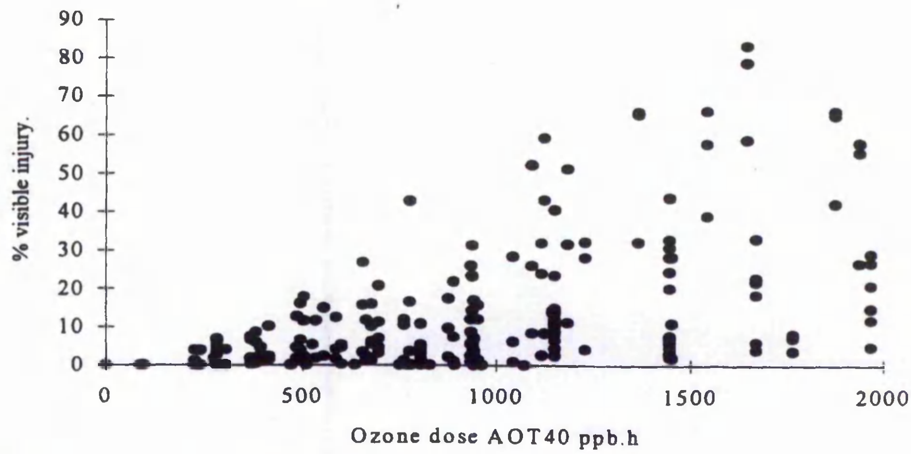


Fig. 3. Visible injury response to ozone for the closed-chamber data.

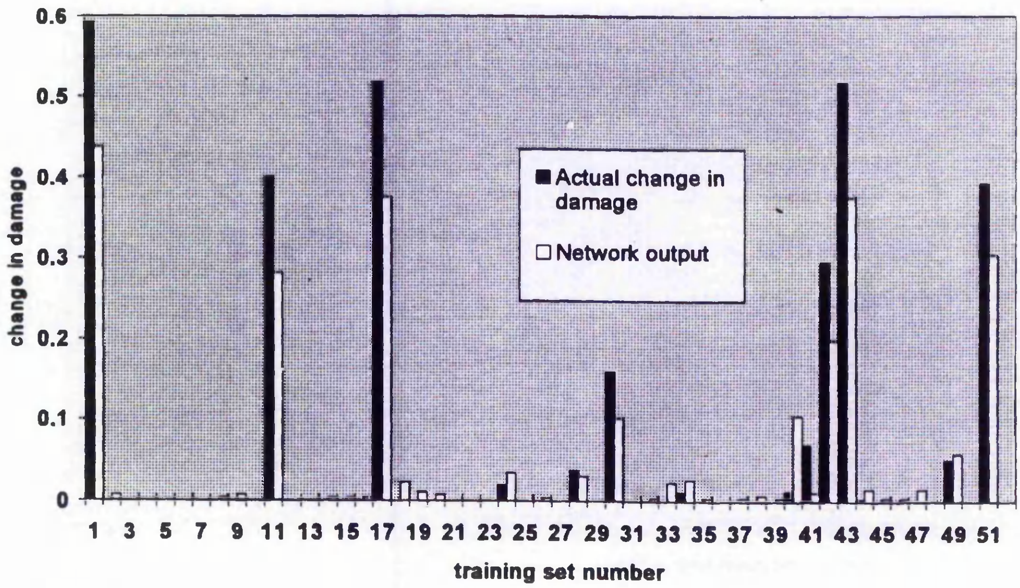


Fig. 4. Training set performance.

influences upon the plants response to ozone. The database was used to create an ANN model of the interactions occurring [17]. A test data set comprising 25% of the whole data (101 points) was randomly extracted.

E. OTC Results

The number of hidden units was minimized in the network. With this limited training and test set, the optimal test performance is achieved before the training error is minimized. This overtraining is an inevitable consequence of a data base that is too small. However, as can be seen from Figs. 4 and 5, on the limited data set available, the network performed to a reasonable accuracy. It has in all training and test cases correctly identified the presence or absence of significant damage.

The trained ANN tended to underestimate the amount of damage. The reason for this could be the disproportionate number of "zero output" patterns in the training set.

F. ICP Results

The ICP data was comprehensive enough to train networ to predict injury using climatic data as well as ozone [2 The inputs of the ANN accepted the following parameters each of the five days preceding injury: ozone (AOT40, 7 mean, daily maximum), Temperature (24-h mean, 7-h mea daily maximum), light (24-h mean, 7-h mean, daily ma mum), solar radiation (24-h mean, 7-h mean, daily maximu and plant age. This network trained to a high accuracy (Fig. and generalized enough to be able to predict onset of inju for a small test set of inputs (Fig. 7).

G. Closed-Chamber Results

Training was stopped when the error of the networks outp had stabilized. The model was retrained using a range hidden units until an optimum performance was achieved the training data. The performance of the model was assess

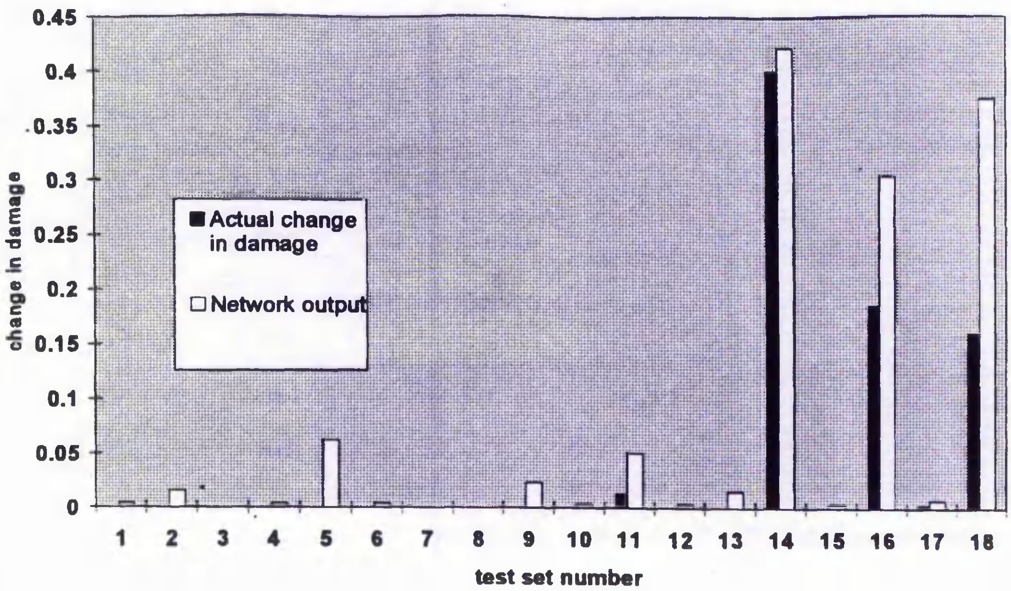


Fig. 5. Test set performance.

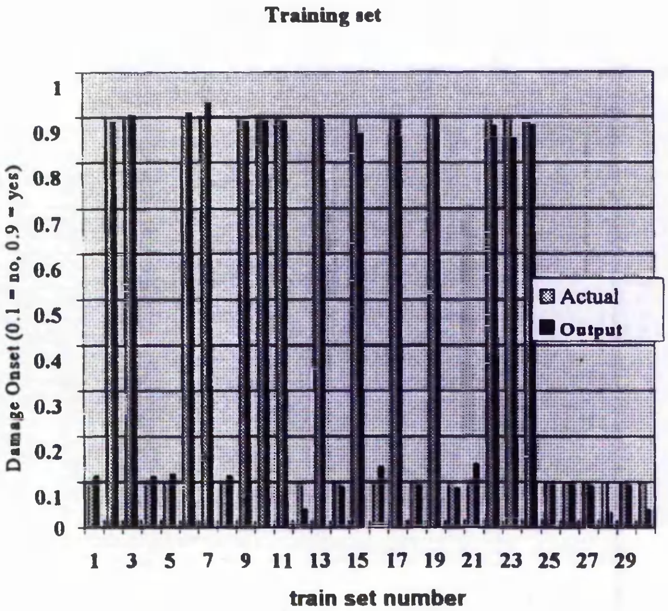


Fig. 6. Training accuracy of ANN trained on ozone plus climatic data.

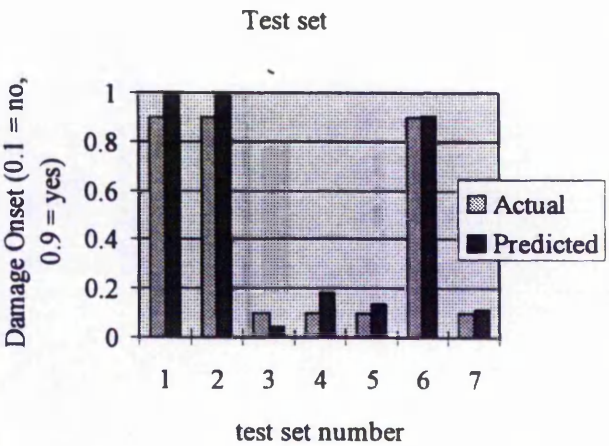


Fig. 7. Testing accuracy of ANN trained on ozone plus climatic data.

by regression of the predictions of the model with the actual percentage injury data to produce an r^2 value representing the accuracy of the model's predictions. Predictions of the optimized model had an r^2 value of 0.797. This value indicated that the model had performed well and was able to make reasonably accurate predictions of a complex system.

III. Analysis of Trained Networks

An overall impression of the relative importance of the different factors can be gained from a simple analysis of the combined weights that correspond to each input to the model. Results from this are shown in Table I. Network 1 trained with ozone, global radiation, humidity, and day number indicates that ozone (AOT40, seven-h mean, daily maximum), global radiation (daily maximum), and humidity (seven-h mean) are important inputs for injury development as the weights have values of 1.92, 1.69, 1.17, 1.43, and 1.12, respectively. Network 2 also gave significant importance to these parameters but a lower importance to humidity and temperature.

Analysis of the weightings of the closed-chamber trained network are presented in Table II. Although this method of meaning extraction can be criticized because it attempts to linearise a nonlinear model, the results do concur between the three models created (Tables I and II) in which a similar hierarchy of importance was found. Clearly the ozone dose received is of the greatest importance followed by light and relative humidity. Additionally, the order in which the individual days of exposure are ranked is similar, with the ozone dose being of greatest importance on the first and third day of exposure. Similar patterns were also seen in the importance of light on the individual days of exposure.

TABLE I
RELATIVE IMPORTANCE OF FACTORS FOR TWO DAMAGED PREDICTING ANN'S

	Network 1	Network 2
Parameter	Relative Importance	Relative Importance
Ozone (AOT40)	1.92	2.58
Ozone (7 hr mean)	1.69	1.86
Ozone (daily maximum)	1.17	1.44
Global radiation (24 hr mean)	0.88	1.23
Global radiation (7 hr mean)	1.03	1.44
Global radiation (daily maximum)	1.43	1.87
Day number	0.62	0.8
Temperature (24 hr mean)		1.11
Temperature (7 hr mean)		1.12
Temperature (daily maximum)		1.29
Humidity (24hr mean)	0.75	0.65
Humidity (7hr mean)	1.12	1.36
Humidity (daily maximum)	0.94	1.22

TABLE II
RELATIVE IMPORTANCE VALUES FROM THE CLOSED CHAMBER MODEL

Method.	R^2 for predictions on the test data.
The Basic ANN Model (with random number).	0.734
Simple linear regression.	0.544*
Multiple linear regression.	0.591
Non-linear regression	0.459*
PCA with LSR.	0.726
PCA with ANN	0.770

The closed-chamber model was further analyzed to make dose response predictions and climatic response predictions for individual days of exposure. Predictions were made by presenting the trained model with a range of conditions while keeping others constant (taking into account correlations occurring between parameters). The first analysis involved keeping the microclimatic conditions constant while varying the ozone dose received on individual days of exposure (the other days were kept constant). This served to determine the nature of the influence of the ozone dose received on individual days of exposure. The results of this testing are presented in Fig. 8.

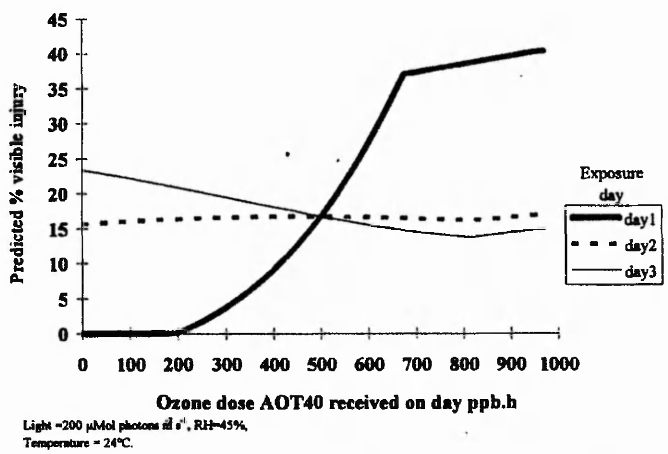


Fig. 8. Predictions of the closed-chamber model for the influence of ozone doses received on individual days of exposure.

This analysis has clearly indicated that ozone has a large positive effect on the first day of exposure, has very little influence on the second day of exposure and has a slight negative influence on the third day of exposure. Influences such as these could not have been detected using conventional statistical analysis.

Further analysis of the closed-chamber model was carried out to determine the influence of climatic factors on the extent of injury. In this case covarying microclimate values (covarying as they would be in the closed chambers) were applied to the trained network at a constant ozone dose of 500

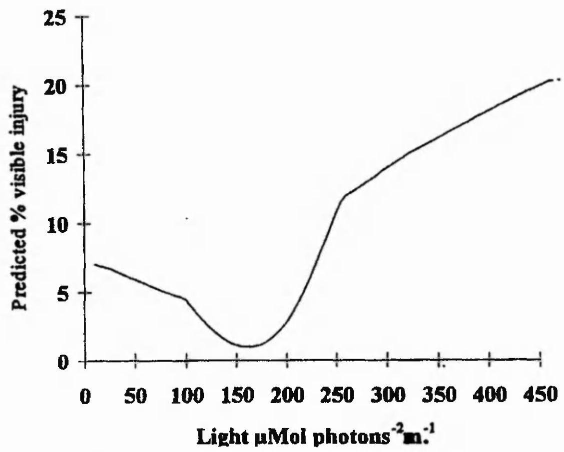


Fig. 9. Predictions of the influence of light on injury.

ppb·h AOT40. Predicted values were generated and plotted in Fig. 9.

Analysis of the network in this way indicates a nonlinear influence. At low light levels an increase in light causes a decrease in injury. However, at higher light levels an increase in light causes an increase in injury. Again these interactions would not have been detected using conventional statistical analysis techniques.

A. Equation Synthesis

Once an ANN has been successfully applied the solution manifests itself as a set of activation function parameters and connection weights. These weights and functions can be used to create an equation, but while this will have more meaning to a scientist it will only be comprehensible for trivial networks.

To synthesise useful equations from nontrivial networks some rationalization is required to keep the size and form of the equation manageable. This is achieved by removing connections with low weights and testing the remaining partially connected network. The algorithm for equation synthesis is as follows.

1) For all hidden nodes:
   Set required threshold for weights from hidden units to output unit.
   Remove all hidden units below this threshold.
2) For all input nodes:
   Set required threshold for weights from input units to hidden units.
   Remove all inputs to hidden units below this threshold.
3) Create equation by matching input nodes, approximate weights, and sign to the source of input node value.
4) Test synthesized equation using appropriate partially connected ANN.
5) Repeat with lower thresholds until partially connected ANN is sufficiently accurate.

Initially, thresholds were set sufficiently high as to only include the most important input to the most important hidden unit. This was invariably too limited to solve the problem so the thresholds were reduced until the required accuracy is achieved. The resulting equation was a simplification of the network which acted as a simple model that could be extended to include more detailed information. This algorithm produced equations which are a simplified representation of the ANN, preserving the principal characteristics of the learned model.

B. Equation Synthesis and Minimal Model Method

Network architectures were minimized by exclusion of inputs, weights, and nodes that did not appear in resulting equations. The reduced architectures are validated on the test data. Conclusions about the nature of ozone injury can also be drawn from equation content.

The performance of partially connected networks was usually worse than the fully connected network, but the relationship between number of connections, described by terms in an equation, and performance was not linear (Fig. 10). This shows that increasing the number of terms in the equation did not necessarily increase the accuracy of predictions the associated ANN would make. One reason for this may be that beyond an essential number of generalizing terms, the introduction of new inputs only brings in nodes that learned pattern specific noise. It also underlines the importance of seeking a minimal model, i.e., the simplest accurate model possible, since any unnecessary terms are extraneous and serve to both obfuscate the interpretations of the model and to increase the error.

When all the available data types were used to train a network a long equation was generated that suggested an importance for sulphur dioxide and not ozone, in the formation of ozone injury. This was investigated and it was discovered that at the data collection site there was a strong correlation between ozone and sulphur dioxide concentrations, the network was therefore using sulphur dioxide as a secondary indicator of ozone injury. This shows how a scientific explanation of a network can be used to validate network architecture. This secondary indicator was not desirable, so the sulphur dioxide was removed from the inputs. The weights of a network trained with all of the remaining data were used to synthesize an equation

$$\begin{aligned}&\text{Change in percentage of leaf area damaged over seven days}\\&= F[D2(\text{SumOz}+\text{MaxOz}) - 2*(D1\text{SumRad})\\&\quad - D1\text{MaxRad} - D2\text{MaxRad} + D3(\text{SumOz}+\text{MaxOz})\\&\quad + D4(\text{SumOz}+\text{MaxOz}) + D5(\text{SumOz}+\text{MaxOz})\\&\quad - D6\text{MaxRad} + D7(\text{MaxOz}-\text{SumOz})] + F[\text{Leaf Age}].\end{aligned}$$

The terms in this equation are environmental or pollutant factor levels, e.g., age of plant at day 0 = (leaf age), highest ozone reading on day 3 = ($D3$(MaxOz)), sum of hourly readings of solar radiation during day 7 = ($D7$(SumRad)).

This equation contains the primary inputs for two hidden units, contained within an activation function (F).

This illustrates the importance of ozone in the production of injury, and also the greater effect of ozone associated with low light levels. One possible reason for this is the effect of light on stomatal opening. Very high light levels, encountered under field conditions, cause stomatal closure, which allows less ozone to pass beyond the protective epidermal layer

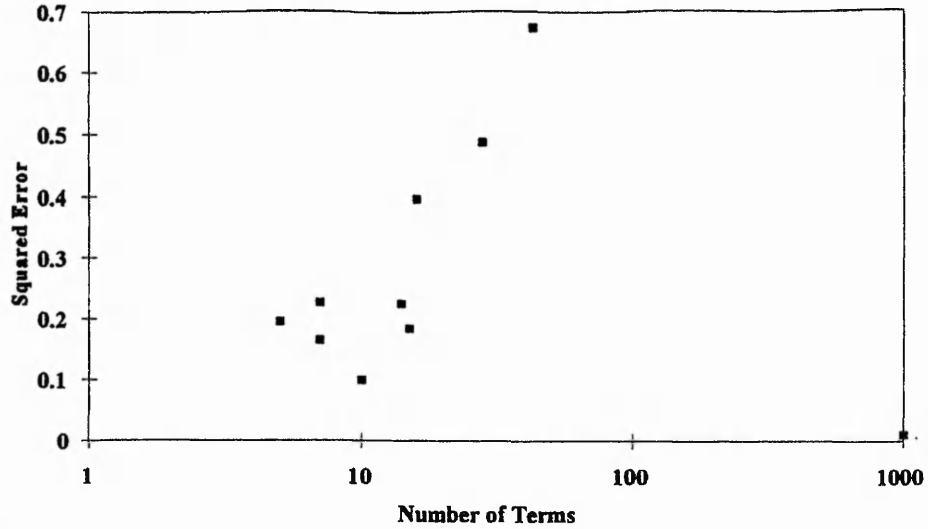


Fig. 10. The nonlinear relationship between number of equation terms and error.

whereas lower levels of light can increase stomatal opening. The equation also shows a general trend that older leaves are injured more than younger leaves. This observation is well documented [23]. The equation can be summarily written as

Change in percentage of leaf area damaged over seven days
$= F$[Ozone − Light] $+ F$[Leaf age].

Since the equation includes only three of the possible seven inputs, it shows that a smaller network can be used. Several smaller networks were tested using combinations of these three input types. Although a small degradation in performance was witnessed (from a root mean square of 0.06 to 0.1), a network consisting only of ozone inputs was capable of predicting damage changes for unseen test patterns to a reasonable accuracy. The equation synthesized from this network showed the relative merits of using sum or maximum ozone to predict damage changes

Change in percentage of leaf area damaged over seven days
$= F[D1\text{MaxOz} + D3\text{MaxOz} - (2 * D5\text{SumOz})$
$+ D6\text{MaxOz}] + F[D4\text{MaxOz} - D3\text{SumOz} + D4\text{MaxOz}$
$+ D5\text{MaxOz} - D6\text{SumOz} + D7\text{MaxOz} - D7\text{SumOz}]$.

A good simplification of this found equation is

Change in percentage of leaf area damaged over seven days
$= F$[MaxOzone − SumOzone]
$= F$[sharpness of daily peaks].

Since only ozone factors are presented to the network, some of them are likely to assume a negative value. Here the daily sum ozone levels take the negative value. So if two days have the same sum ozone levels (Fig. 11) and one day has a higher maximum ozone level (solid line) than another (dotted line) this will lead to a greater increase in damage. This agrees with findings by other researchers [23], [24].

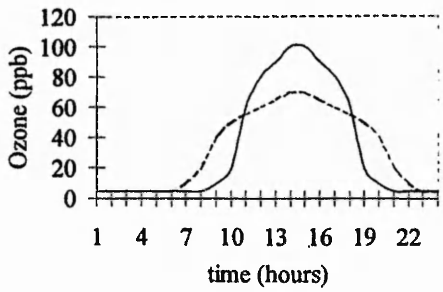


Fig. 11. A comparison of two ozone episodes with identical ppb·hours.

From the weights of an onset of injury predicting ANN, the following equation was derived:

Occurrence of injury
$= F$[Day2(7h mean + Max) − Day3(AOT40
+ MaxOzone)] $+ F$[Day2(MaxOzone − AOT40)
− Day1(AOT40 + 7h mean) + day3(7h mean)] − 1.

The terms in this equation are pollutant factor levels, e.g., Mean of hourly readings of ozone levels between 10 am and 5 pm during day two is Day2(7-h mean).

The first, and most influential, node appears to show that a fall from high to low levels of ozone precedes the onset of injury. This is apparent because ozone levels on day three are used in a negative way, i.e., higher levels of ozone give a lower output. Ideal conditions for a positive prediction of injury are therefore high levels of ozone on day two and low levels on day three. This equation can be further simplified to

Occurrence of injury
$= F$[rise in ozone levels] $+ F$[fall in ozone levels].

This conclusion was confirmed by equations synthesized from ANN trained on a larger training set that included data from the 1995 growing season.

The generation of simple equations from complex ANN's enables the generation of hypotheses. These hypotheses can then be tested by carrying out the relevant biological experiments. The final equation points to a link between falls in ozone levels and the onset of leaf injury, this hypothesis is currently being tested in greenhouse experiments.

C. Comparison of ANN and Statistical Methods

Comparative analysis was carried out on closed-chamber data containing seven-h mean values for light, temperature, relative humidity, the leaf position, and the extent of visible injury. The data was limited to single day exposures. An additional random number input was added to the model to include extra nonlinearity to the data. The data set was divided into a training data set and a test data set. The test data was extracted randomly and comprised 25% of the total data set. The remainder was used as the training data set. A number of statistical approaches were used in addition to ANN's. These included simple linear regression, nonlinear regression, multiple linear regression, and principal components analysis combined with least square regression.

A simple closed-chamber ANN model was trained and optimized using the training data set as described earlier. When training was completed the accuracy of the model's predictions for the test data set were determined, generating an r^2 value for the predictions of the network. An r^2 of 0.734 was produced for the model.

Simple linear regression analysis was also carried between ozone dose AOT40, light, temperature, relative humidity, and the extent of injury. Regressions were repeated for each leaf position. R^2 values were calculated for each leaf, for each parameter. The best fit was produced for the ozone dose AOT40, which had an r^2 value of 0.544. A similar approach was used in the nonlinear regression analysis where regressions were repeated using logarithmic, power, and exponential functions again r^2 values were calculated for each leaf, for each parameter, for each function. Again the best fit for the model was produced for the ozone dose AOT40 with the logarithmic function, having an r^2 value of 0.322.

Multiple linear regression analysis was also carried out on the training data using the statmost statistical analysis package. The independent variables used were the same inputs used for the ANN model. Analysis of the data generated an equation which was the applied to the test data. The values produced by the equation and the actual values were regressed to generate an r^2 value for the predictions. The equation generated by this analysis approach produced an r^2 value of 0.591 when compared with the actual data.

Principal components analysis was carried out on the training data using the Unistat statistical package. The principal components generated were then regressed against the extent of injury to generate a least squares regression model. These principal components were then applied to the least squares regression model to generate predicted injury values for the test data. Regression analysis of the predicted and actual values was carried out to determine the accuracy of the model and calculate an r^2 value. This method showed the best performance of all the conventional statistical techniques producing an r^2 value of 0.726.

Finally the principal components of the training data were used to create an ANN model. The test principal components were then applied to the model to generate predictions. These predictions were then regressed with the actual data to generate an r^2 value. This modeling approach produced an r^2 value of 0.770

A summary of the r^2 values produced for each analysis method used is presented in Table III.

TABLE III
SUMMARY OF ACCURACY OF VARIOUS ANALYSIS APPROACHES

Parameter	Relative importance
Ozone dose AOT40 on day1	9.945327
Ozone dose AOT40 on day3	7.927049
Leaf Position	7.874798
Sum of Light on day3	7.206108
Sum of Light on day1	6.839422
7 hour mean %RH	5.99886
7 hour mean Light	5.676245
Total AOT40 for all days	5.546995
Sum of Light on day2	5.189877
Sum of %RH on day3	4.938922
Sum of temperature on day2	4.390927
Ozone dose AOT40 on day2	4.229449
Sum of %RH on day2	4.180871
7hour Mean temperature	4.063485
Sum of temperature on day1	4.040356
Sum of temperature on day3	4.01259
Sum of %RH on day1	3.618812

IV. CONCLUSION

The approach applied to the domain of plant response to pollution is transferable to other nonlinear domains. Given sufficient training data, an accurate model can be produced using a modeling strategy that incorporates MLP's and backpropagation. The results obtained compare favorably with alternative statistical methods of comparable complexity. The best model is not usually created in a single training session because it is necessary to ensure that all important parameters are represented and that none of them are substituted for by secondary indicators which happen to be correlated with them. The optimal model is created by starting with essential inputs and incrementally adding only those noncritical parameters that are required to minimize the error.

In order to provide an understanding and validation of the learnt model, it is necessary to elucidate the model by calculating importance factors and synthesizing equations of the principal combined factors in the ANN's structure. This helps to provide a clear interpretation of the model which supports scientific enquiry; hypotheses can be formed and tested from the equations (e.g., ozone fall before necrosis) and the network response can be used to visualize the primary influences (e.g., U-shaped curve, Fig. 9).

A. Current Uses

The information generated from the trained ANN's is used in the development of protocols for international pollution control by the UNECE [25]. In these protocols, pollution control is implemented in locations where the critical level for a pollutant is exceeded. The critical level for a pollutant is defined as the concentrations of the pollutant in the atmosphere above which direct adverse effects on receptors, such as plants, ecosystems, or materials, may occur according to present knowledge [26]. A critical-levels-based protocol is being developed which will commit countries to reducing the concentration of the precursors of ozone formation. This requires a scientific understanding of the effects of ozone on vegetation. In addition to assisting in the determination of critical levels of atmospheric pollution, the ANN modeling of the effects of ozone exposure profile and modifying climatic factors provide useful guidance for pollution impact scientists. This ANN work, funded by the U.K. Department of Environment, is also being used in an assessment of the scale of the ozone problem in the UK.

Trained ANN's are being used as a teaching aid, the models based on the closed-chamber data have been embedded into Microsoft Excel. This allows input values to be varied within the spread sheet to generate a predicted injury value. Once embedded into the spreadsheet students can carry out virtual experiments to look at the influences of various parameters on the ozone dose response. Normally these experiments would take a great deal of time to perform, well out of the scope of a normal teaching practical session. This method allows the student to investigate the interactions and draw conclusions based on predicted values. Because the ANN approach produces more accurate predictions, the model is more realistic than other equation-based models.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. J. Colls, University of Nottingham, for supplying raw data for training of the OTC network.

REFERENCES

[1] W. W. Heck, O. Taylor, and D. T. Tingley, Eds., *Assessment of Crop Loss from Air Pollutants*. New York: Elsevier, 1988.

[2] H. Pleijel, L. Skärby, G. Wallin, and G. Selldén, "A process-oriented explanation of the nonlinear relationship between grain yield of wheat and ozone exposure," *New Phytologist*, vol. 131, pp. 241–246, 1995.

[3] L. M. Mortensen, "Effects of ozone concentration on growth of tomato at various light, air humidity, and carbon dioxide levels," *Scientia Horticulturea*, vol. 16, pp. 17–24, 1992.

[4] G. Kersteins, R. Federholzner, and K. J. Lendzian, "Dry deposition and cuticular uptake of pollutant gasses," *Agriculture, Ecosystems, and Environment*, vol. 42, pp, 239–253, 1992.

[5] W. A. J. Van Pul and A. F. G. Jacobs, "The conductance of a maize crop and the underlying soil to ozone and other various environmental conditions," *Boundary Layer Meteorology*, vol. 69, pp. 83–99, 1993.

[6] B. D. Amiro, T. J. Gillespie, and G. W. Thurtell, "Injury response of *Phaseolus Vulgaris* to ozone flux density," *Atmospheric Environment*, vol. 18, no. 6, pp. 1207–1215, 1985.

[7] J. M. M. Aben, M. Janssen-Jurkovicova, and E. H. Adema, "Effects of low-level ozone exposure under ambient conditions on photosynthesis and stomatal control of *Vicia faba* L," *Plant Cell and Environment*, vol. 13, pp. 463–469, 1990.

[8] H. Moldau, J. Sober, and A. Sober, "Differential sensitivity of stomata and mesophyll to sudden exposure of bean shoots to ozone," *Photosynthetica*, vol. 24, pp. 446–458, 1990.

[9] M. R. Guzy and R. L. Heath, "Responses to ozone of varieties of common bean (*Phaseolus vulgaris* L)," *New Phytologist*, vol. 124, pp. 617–625, 1994.

[10] M. J. Kropff, "Modeling short-term effects of sulphur dioxide. I. A model for the flux of SO_2 into leaves and effects on photosynthesis," *Netherlands J. Plant Pathology*, vol. 95, pp. 195–213, 1989.

[11] H. E. Schut, "Models for the physiological effects of a short ozone exposure on plants," *Ecological Modeling*, vol. 30, pp. 175–207, 1985.

[12] L. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*. New York: Academic, 1985.

[13] L. Fu, *Neural Networks in Computer Intelligence*. New York: McGraw-Hill, 1994, pp. 351–369.

[14] R. Andrews, J. Diederich, and A. B. Tickle, "A survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge Based Syst.*, vol. 8, pp. 373–389, Dec. 6, 1995.

[15] G. Towell and J. Shavlik, "The extraction of refined rules from knowledge based neural networks," *Machine Learning*, vol. 131, pp. 71–101, 1993.

[16] S. Sestito and T. Dillon, *Automatic Knowledge Acquisition*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[17] G. R. Balls, D. Palmer-Brown, A. H. Cobb, and G. E. Sanders, "Toward unravelling the complex interactions between microclimate, ozone dose, and ozone injury in clover," *J. Water, Air, and Soil Pollution*, vol. 85, pp. 1467–1472, 1995.

[18] G. R. Balls, D. Palmer-Brown, and G. E. Sanders, "Investigating microclimate influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks," *New Phytologist*, vol. 132, pp. 271–280, 1996.

[19] C. M. Roadknight, D. Palmer-Brown, and G. E. Sanders, "Learning the equations of data," in *Proc. 3rd Annu. SNN Symp. Neural Networks*, B. Kappen and S. Gielen, Eds. Berlin: Springer-Verlag, 1995, pp. 253–257.

[20] G. E. Sanders, N. D. Turnbull, A. G. Clark, and J. J. Colls, "The growth and development of *Vica Faba* L. in filtered and unfiltered open-top chambers," *New Phytologist*, vol. 116, pp. 27–78, 1990.

[21] UNECE, "Critical levels for ozone in Europe: Testing and finalizing the concepts," in *Proc. Critical Levels Wkshp.*, Kuopio, Finland, Apr. 15–17, 1996.

[22] J. Benton, J. Fuhrer, B. S. Gimeno, L. Skarby, G. Balls, C. M. Roadknight, and G. Sanders-Mills, "The critical level of ozone for visible injury on crops and natural vegetation (ICP crops)," in *Critical Levels of Ozone in Europe: Testing and Finalizing the Concepts*, UNECE Wkshp. Rep., L. Karenlampi and L. Skarby, Eds., Univ. Kuopio, Finland, pp. 44–57.

[23] G. Karlsson, G. Sellden, L. Skarby, and H. Pleijel, "Clover as an indicator plant for phytotoxic ozone concentrations: Visible injury in relation to species, leaf age, and exposure dynamics," *New Phytologist*, vol. 129, pp. 355–365, 1995.

[24] R. C. Musselman, A. J. Heurta, P. M. McCool, and R. J. Oshima, "Response of beans to simulated ambient and uniform ozone distributions with equal peak concentrations," *J. Amer. Soc. Horticultural Sci.*, vol. 111, pp. 470–473, 1986.

[25] J. Benton, J. Fuhrer, B. S. Gimeno, L. Skarby, G. Balls, D. Palmer-Brown, C. Roadknight, and G. Sanders, "ICP crops and critical levels of ozone for injury development," in *In Exceedences of Critical Loads and Levels*, M. Knoflacher, J. Schneider, and G. Soja, Eds., Umweltbundesamt, Federal Environment Agency, Wein, Austria, pp. 97–112, 1996.

[26] K. R. Bull, "The critical loads/levels approach to gaseous pollutant emission control," *Environmental Pollution*, vol. 69, pp. 105–123.

Chris M. Roadknight received the B.Sc. degree in applied biological sciences from Manchester Polytechnic (now Manchester Metropolitan University) in 1992 and is currently working toward the Ph.D. degree in neural-network models of ozone-induced crop damage at Nottingham Trent University, U.K.

Graham R. Balls received the Ph.D. degree from the Nottingham Trent University, U.K., in 1996.

He is currently working as a Postdoctoral Research Fellow at the same University. His research interests include the analysis of complex environmental data using artificial neural networks.

Gina E. Mills received the Ph.D. degree in plant physiology from the then Trent Polytechnic (now Nottingham Trent University), U.K.

She is a Senior Lecturer in Plant Sciences at Nottingham Trent University. Her interests include the effects of environmental stress, including air pollution, on vegetation.

Dominic Palmer-Brown received the Ph.D. degree in computer science with adaptive resonance classifiers from the University of Nottingham, U.K., in 1992.

Since then he has been a Senior Lecturer in computing at the Nottingham Trent University. His primary research interests include data modeling and pattern recognition with artificial neural networks.

The Analysis of Artificial Neural Network Data Models

C.M. Roadknight[1], D. Palmer-Brown[1] and G.E. Mills[2]

[1] Department of Computing, The Nottingham trent University
Burton Street, Nottingham NG1 4BU
[2] Department of Life Sciences, The Nottingham Trent University
Clifton Lane, Nottingham NG11 8NS

Abstract. Artificial neural networks are good non-linear function approximators but their multi-layer, non-linear form gives little immediate indication of the features they have learnt. Several methods are put forward in this paper that reduce the complexity of the network or give simplified equations that are easier to interpret. Relative weight analysis and equation synthesis are summarised while correlated activity pruning is introduced and explained in detail. The former techniques use the weights of a trained network to assign importance to inputs or groups of inputs. The latter algorithm reduces complexity of a network by merging hidden units that have correlated activations. This procedure also allows the relationship between detected features to be evaluated. Data from pollutant impact studies are used but the techniques developed are applicable to many scientific data modelling environments.

1 Introduction

1.1 The data to be analysed

Artificial neural networks (ANN's) are used at Nottingham Trent University for modelling the complex climatic and pollutant interactions that damage crops in this and other European countries [1, 2]. ANN's are trained using daily climatic and pollutant conditions to predict development of crop injury or yield reduction. These effects have traditionally been modelled using standard statistical techniques such as dose response curves [3]. While the techniques covered in this paper were developed and tested using data from this domain, they are applicable to any 3 layer, feed forward networks.

1.2 ANN's for data analysis

ANN's have been shown to cope well with data containing non-linearity and to model non-linear relationships [4]. While ANNs have outperformed many statistical techniques for theoretical problem solving, their application to real world problems is the true test of their usefulness. They have been successfully

X. Liu, P. Cohen, M. Berthold (Eds.): "Advances in Intelligent Data Analysis" (IDA-97)
LNCS 1280, pp. 337–346, 1997.

applied in a number of areas, including the fields of medicine [5, 6] and finance [7, 8]. In this application of ANN's, the extent to which the training data is of sufficient quantity and quality is detected by R.M.S error measures and R^2 coefficients on randomly selected test data.

1.3 Current rule extraction and optimisation techniques

The connection weights of a trained ANN contain all the information required to carry out the desired task (prediction, classification etc.), but because the knowledge acquired is represented in a complex, non-linear form, it is difficult to interpret in all but the most trivial of networks. The ANN acts as a 'Black Box' taking in and giving out information with no explanation of the relationship between them.

Efforts have been made to dissect ANNs into a set of rules [9, 10]. These methods usually aim for a set of conjunctive rules based on a 'decomposisional' approach at the individual unit level.

There are several reasons why the demystifying of an ANN is desirable:

- The ability to explain how a solution is arrived at is essential for any safety critical' systems (ie. Controlling temperature regulation in a nuclear power station).

- Symbolic AI systems (ie. Expert Systems) declare their knowledge explicitly. This ability to explain their decisions can mean inferior systems are chosen ahead of more accurate ANN.

- Rules synthesised from an ANN could be used to construct a knowledge base for an expert system, the most difficult and time consuming part of building an Expert System [11].

- When small data sets are used, extracting meaning can allow an expert to decide under what conditions the ANN will not be able to generalise.

- There is evidence that rules extracted from the network sometimes give more accurate solutions than the source ANN [12] though this is not usually the case.

Transforming the network into a form that is more interpretable, however, is of maximum value if the network architecture has been optimised. Pruning can help to achieve optimum network architecture. Such methods benefit from the learning advantages of larger networks in terms of speed and aquisition of features while reducing the amount of overtraining or memorisation within these networks.

Weight pruning is most commonly used. This involves the removal of connections based on the value of the connecting weights and these methods can be divided into two groups:

- Sensitivity Calculation. Once the full sized network is trained the sensitivity of the error function to zeroing of a weight is estimated and weights with a low impact are removed [13].

- Penalty-Term Methods. These involve the introduction of a new cost function to enforce weights of small magnitude to converge to zero during training, so can then be removed with no effect [14].

2 Basic Approaches to Weight Analysis

In this paper, two approaches are taken to explain the rules by which an ANN makes its decisions; relative weight analysis [15] [16] and equation synthesis [17] [18].

2.1 Relative weight analysis

The weights of any ANN can be followed from input to output, the product of each route giving an indication to the relative influence of the chosen input. This can only be a coarse measure of an input's affect due to the non linear nature of ANN neuron activation functions.

This method can be used to prioritise environmental and pollutant factors that affect crop injury [16]. For example, a weight analysis from a network trained to predict the onset of injury in clover (*Trifolium subterraneum* and *T. repens*) showed 4 variables to have the following total effect on the output (development of leaf injury) (table 1).

2.2 Equation Synthesis

Once an ANN has been successfully applied, the solution manifests itself as a set of activation functions and connection weights, but the full network equation will only be comprehensible for trivial networks.

To synthesise useful equations from non-trivial networks some rationalisation is required to keep the size and form of the equation manageable. This is achieved by removing all connections with relatively low weights and testing the remaining, partially connected network. More connections may be included until this minimised network performs to an acceptable standard [17] [18]. It is important that the model is found in this way and not by recursively removing

Table 1. Impact of environmental parameters on leaf injury

Variable	Relative Influence on Output
Ozone	12.8
Vapour Pressure Deficit	4.6
Plant Age	3.8
Solar Radiation	-2.20

small weighted connections from a complete network. The relationship between number of connections and network accuracy is a non-linear one so the simplest accurate model will not in general be found by succesive removal of connections.

The performance of partially connected networks was usually worse than the fully connected network, but the relationship between number of connections, described by terms in an equation, and performance was non-linear. This shows that increasing the number of terms in the equation did not necessarily increase the accuracy of predictions the associated ANN would make. One reason for this may be that beyond an essential number of generalising terms, the introduction of new inputs only brings in nodes that learnt pattern specific noise. It also underlines the importance of seeking a minimal model ie. the simplest accurate model possible, since any unnecessary terms are extraneous and serve to both obfuscate the interpretations of the model and to increase the error.

Once the minimal network has been found, it can be written in the form of an equation. From the weights of a network trained to predict whether plant injury would follow a three day ozone episode (0 errors from 37 training and test patterns), the following equation was derived:

Occurrence of injury =
F[Day2(7hmean + Max) - Day3(AOT40+MaxOzone)] +
F[Day2(MaxOzone-AOT40) - Day1(AOT40 + 7hmean)
+ day3(7hmean)] - 1

The terms in this equation are pollutant factor levels eg. Mean of hourly readings of ozone levels between 10am and 5pm during day 2 is Day2(7hmean). *Where* AOT40 *= accumulated ozone above 40 parts per billion.*

This equation contains the primary inputs for 2 hidden units, contained within the activation function (F), which can be approximated as a straight line between 0 and 1, for a linearised model.

The partially connected network described by this equation performed equally well in terms of number of erroneous outputs. The first, and most influential, node appears to show that a fall from high to low levels of ozone precedes the onset of injury. This is apparent because ozone levels on day three are used in a negative way, ie higher levels of ozone give a lower output. Ideal conditions for a positive prediction of injury are therefore high levels of ozone on day 2 and low levels on day 3. This equation can be further simplified to:

Occurrence of injury =
F[rise in ozone levels] + F[fall in ozone levels]

3 Correlated Activity Pruning

The generalisation ability of an Artificial Neural Network (ANN) is dependent on its architecture. An ANN with the correct architecture will learn the task presented by the training set but also assimilate rules that are general enough to correctly predict outputs for unseen test set examples. More importantly for intelligent data analysis, the weights of a smaller ANN will be easier to interpret and the smallest net gives the most general rules for interpreting the data.

There are many possible networks that will model any given data. The process of finding the minimal network establishes a model that is a good approximation to the canonical model that would properly represent an accurate analysis of all available data.

Sietsma and Dow [19] describe an interactive pruning method that uses several heuristics to identify units that fail to contribute to the solution and therefore can be removed with no degradation in performance. This approach removes units with constant outputs over all the training patterns as these are not participating in the solution. Also, units with identical or opposite activations for all patterns can be combined. The approach to merging hidden units detailed in Sietsma and Dow's paper is useful, however it only covers perfectly correlated, binary activations.

The method presented here generalises correlated activity pruning to all real valued positively and negatively, highly correlated activation sets from hidden neurons.

3.1 CAPing Equations

Each hidden unit within a three layer neural network produces an activation when a set of inputs is presented to it. If two hidden units produce output signals that are 100 percent correlated for the entire training set then the activations for these two hidden units can be merged, with no loss of performance. Two hidden units with correlated activities can be simplified into one hidden unit by recalculating the weight vector from the remaining hidden unit to the output node using equations 4 or 5. A brief derivation of these equations is also shown. Where *Signal'* is the sum of activation profiles from 2 hidden nodes, X and Y.

$$If Signal_X = Signal_Y \quad then \quad Signal' = Signal_y + Signal_y \tag{1}$$

If Signal X and Signal Y correlate with a correlation coefficient (ρ) of 1 then

$$Signal' = Signal_y + \frac{\sigma_x}{\sigma_y}(Signal_y - \mu_y) + \mu_x \tag{2}$$

$$= Signal_y(1 + \frac{\sigma_x}{\sigma_y}) + (\mu_x - (\frac{\sigma_x}{\sigma_y} * \mu_y)) \tag{3}$$

$$therefore, W' = W_y + (\frac{\sigma_x}{\sigma_y} * W_y) and Bias' = Bias + (\mu_x - (\frac{\sigma_x}{\sigma_y} * \mu_y)) \tag{4}$$

If Signal X and Signal Y correlate with (ρ) of -1 then

$$W' = W_y - (\frac{\sigma_x}{\sigma_y} * W_y) and Bias' = Bias + (\mu_x + (\frac{\sigma_x}{\sigma_y} * \mu_y)) \tag{5}$$

Bias' = new bias weight. W' = new hidden unit to output weight
W_x, μ_x and σ_x = output weight, mean and standard deviation of activations for

hidden unit with lowest weighting
W_y, μ_y and σ_y = output weight, mean and standard deviation of activations for hidden unit with highest weighting

The equation for the correlation coefficient used is: $\rho_{xy} = \frac{Cov(x,y)}{\sigma_x, \sigma_y}$

Where $-1 \leq \rho_{xy} \leq 1$ and $Cov(x,y) = \frac{1}{n}\sum_{i-1}^{n}(x_i - \mu_x)(y_i - \mu_y)$

Correlation coefficients of 1 or -1 are not common for activations from any two hidden units of an ANN, but high correlations do exist in networks that use more hidden units than are required to learn a problem. Error can be introduced into the networks performance if the correlation coefficient is not equal to 1 or -1, but the closer the correlation the smaller the error.

Closely correlated activations can occur for several reasons:
1. Similar weights from input units to hidden units.
2. Each of the two hidden units detecting a different, co-dependent feature. For example, if high values always occur for input X when there are high (or low) values for input Y.
3. If two factors have the same effect. For example, if high values for input X have the same effect on the output as high (or low) values for input Y.

3.2 Benefits of CAPing

Analysis of Correlated units. The weights of the input layer to hidden unit connections for merged units are equivalent in terms of their influence on the network's output. They therefore represent eqivalent features. Identifying these equivalencies allows for a greater understanding of related data events.
Network Optimisation. Trial and error is a common approach to network optimisation. CAPing of a network can reach a near optimum structure in one training session.
Speed Up. An ANN can be trained in fewer epochs if more hidden units are used than are required. Parallel processing allows the addition of more hidden units without an accompanying increase in training time per epoch. Therefore an ANN can be trained quickly using more hidden units and then capped to a near optimum architecture.

3.3 CAPing in Practice : A Curve Fitting Example

Training an ANN to learn a complex, multiple variable equation is an accepted method of analysing network performances. This is performed as a curve fitting exercise (eg [21]) or as boolean equation solving (eg. [20]).

In this case, a 4 variable equation is used and random numbers are used as variables to generate an output, x. The equation is:

$$x = \frac{a - (b * c) + (\frac{b}{(d+1)}) + K}{J} \tag{6}$$

In equation 6 a,b,c,d are variables in the range [0,1]. J and K are set so that x then varies in the range [0,1].
A three-layer back propagation network is then trained using 500 data sets, each consisting of 4 inputs (A,B,C,D) and an output X(a,b,c,d). The level of accuracy to be achieved is set to an RMS of less than 0.03 and a maximum error of less than 0.1. The minimum number of hidden units needed to learn this can be found, by exhaustive trial, to be five. A network can then be trained using 20 hidden units, many more than are required. This takes fewer epochs to reach the required accuracy and, because of the parallel architecture used (a 64 processor SIMD machine), each epoch takes a similar training time as a 4 hidden unit network. The CAPing equations can then be used to recursively merge hidden units, with the most highly correlated units merged first. The activations of two such units and the activations of the combined unit are shown in figure 1.

Hidden unit 11 was merged into hidden unit 15 to produce a new weight from hidden unit 15 and a new bias into the output unit. The effect of changing the weight from hidden unit 15 and the bias is shown in figure 1.

Very little error is introduced until units with correlation coefficients of less than 0.8 are used. The correlation coefficients are given above the error in figure 2. Merging to even fewer hidden units is possible if some post-CAPing retraining is carried out, this retraining needs approximately 10 percent of the epochs of training from random weights.

Analysis of the weights from the input units of the merged hidden units shows some sets of weights to be highly correlated but other sets to be non-correlated (Table 2). This illustrates how CAPing removes two types of redundancy. Firstly when weight vectors represent similar features, and secondly differing features that have an equivalent effect on the network's output.

Table 2. Input to hidden unit weights for pairs of weight vectors with highly correlated hidden unit activation profiles.

Weight from:	to hidden unit 1	to hidden unit 2	to hidden unit 3	to hidden unit 4
Variable A	0.2756	0.04419	-1.4898	-1.4746
Variable B	-0.573	-0.7546	-0.6228	-1.6257
Variable C	-0.2666	-0.4243	-0.8889	-1.1646
Variable D	-0.9353	-1.0815	-1.0183	-1.2954
Comment	*Correlation coefficient for weights to 1 and 2 = 0.9991*		*Correlation coefficient for weights to 3 and 4 = -0.1099*	

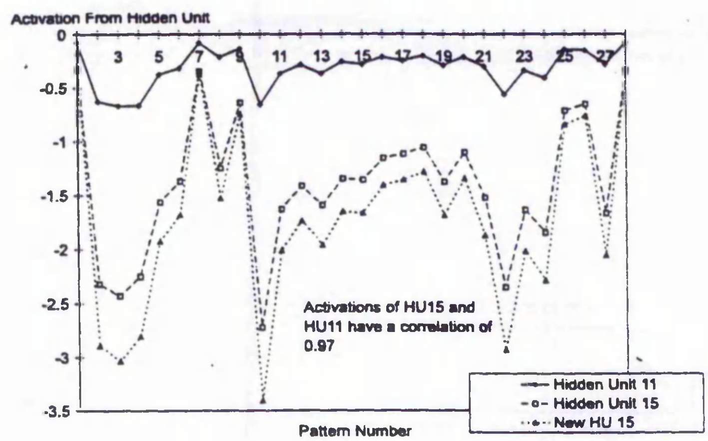


Fig. 1. The effects on activations of merging two correlated hidden units.

3.4 Real CAPing example: modelling pollution effects

ANNs have been applied to the problem of predicting leaf damage to crop plants ([1, 2]). ANNs with more hidden units than were required could be CAPed down to architectures with a near optimal number of hidden units. An examination of the weights of hidden units with correlated activations showed that similar weight redundancy' was the main form of redundancy removed, but there were also some cases of non-correlated weights giving correlated activations.

For example, in one model involving weekly pollutant and climate variables, the following equations were synthesised from two hidden units and were found to be correlated:

F[7hr mean light + Max light - (2*AOT40)] and
F[7hr mean Ozone + Max Ozone - MaxLight]

Both of these units detect the same situation in different ways ie. when average ozone and light levels are fairly high (but not very high). Such situations can be very damaging to plants since stomatal aperture is large and ozone can easily enter the plant.

4 Conclusions

While ANN's are capable of processing complex data, they can only be used for intelligent data analysis if their functionality can be explained. Relative weight

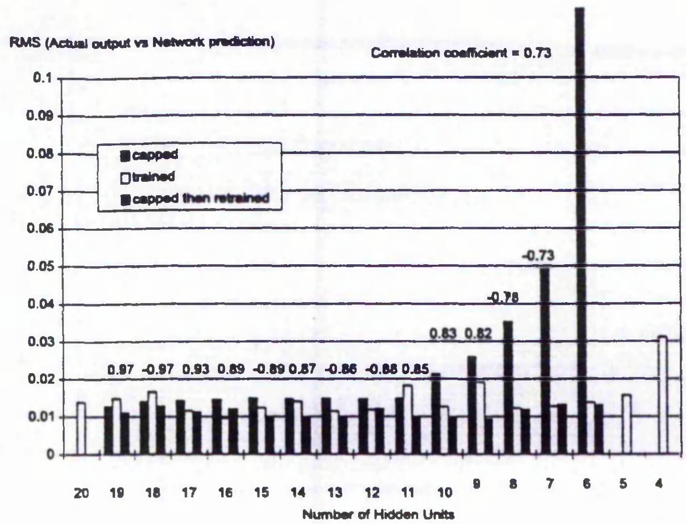


Fig. 2. The effect of merging hidden units with less correlated activations.

analysis and equation synthesis support this process. CAPing can be a valuable tool in minimising network architecture by detecting redundancy at the level of hidden unit activations. The comparison of hidden units with correlated activations also leads to a model of the relationships between causal agents.

5 Acknowledgements

We would like to thank the UK Department of the Environment for financial support of the work in this project (Project number PECD 7/12/145).

References

1. Benton J, Fuhrer J, Gimeno BS, Skarby L, Balls G, Palmer-Brown D, Roadknight C and Sanders G.1996. ICP-Crops and critical levels of ozone for injury development. Exceedences of Critical Loads and Levels. In Exceedences of Critical Loads and Levels. Eds. M. Knoflacher, J. Schneider and G. Soja. Umweltbundesamt (Federal Environment Agency) Wein, Austria. 97-112.
2. Benton J, Fuhrer J, Gimeno BS, Skarby L, Balls G, Roadknight C and Sanders-Mills G. 1996. The critical level of ozone for visible injury on crops and natural vegetation (ICP-Crops). In: Critical Levels of Ozone in Europe: Testing and Finalising the Concepts. UN ECE Workshop Report. Eds. L Karenlampi and L Skarby. University of Kuopio, Finland. 44 - 57.

3. Heck, WW, Taylor O. and Tingley DT. (eds) 1988. Assessment of Crop Loss from Air Pollutants. Elsevier Applied Science, New York.
4. Funahashi K. 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks, 2, 183 - 192.
5. Burke HB, Hoang A and Rosen DB. 1995. Survival function estimates in cancer using artificial neural networks. Proceedings of WCNN. Vol II. p.748-749.
6. Orr RK. 1995. Use of probabilistic neural networks to predict mortality following cardiac surgery. Proceedings of WCNN. Vol II. p. 754-757.
7. Davalo E, Niam P. 1990. Neural Networks. Macmillan Press. p. 111-112.
8. Tan H, Prokhorov DV and Wunsch DC. 1995. Probabilistic and time-delay neural network techniques for conservative short-term stock trend prediction. Proceedings of World Congress on Neural Networks. Vol II. p. 44-47.
9. Fu L. 1994. Neural Networks in Computer Intelligence. McGraw-Hill International Editions. p.351-369.
10. Andrews R, Diederich J and Tickle AB. 1995. A survey and critique of techniques for extracting rules from trained Artificial Neural Networks. Knowledge Based Systems Vol 8 (6, December) p. 373-389.
11. Sestito S and Dillon T. 1994. Automatic knowledge acquisition. Prentice Hall.
12. Towell G and Shavlik J. 1993. The extraction of refined rules from knowledge based neural networks. Machine Learning Vol 131, p71-101.
13. Karin ED. 1990. A simple procedure for pruning back-propagation trained neural networks. I.E. Trans. Neural Networks, vol.1 no.2. p239-242.
14. Weigend AS, Rumelhart DE and Huberman BA. 1991. Generalization by weight elimination with applications to forecasting. In Advances in Neural Information Processing (3). Lippmann R, Moody J and Touretzky D. Eds. p. 875-882.
15. Balls GR, Palmer-Brown D, Cobb AH and Sanders GE. 1995. Towards unravelling the complex interactions between microclimate, ozone dose and ozone injury in clover. Journal of Water, Air and Soil Pollution. 85, 1467 - 1472.
16. Balls GR, Palmer-Brown D, and Sanders GE. 1996. Investigating microclimate influences on ozone injury in clover (Trifolium subterraneum) using artificial neural networks. New Phytologist, 132, 271 -280
17. Roadknight CM, Palmer-Brown D and Sanders GE. 1995. Learning the equations of data. Proceedings of 3rd annual SNN symposium on neural networks (eds. Kappen B and Gielen S) Springer-Verlag. 253-257.
18. Roadknight CM, Balls GR, Sanders GE and Palmer-Brown D. Modelling complex environmental data. IEEE Transactions on Neural Networks: Special edition on everyday applications. (In Press July 1997.)
19. Sietsma J and Dow RJF. 1988. Neural net pruning - Why and how. Prc. IEEE Int. Conf. Neural Networks. Vol 1. p. 325-333.
20. Wiersma FR, Poel M and Oudshoff AM. 1995. The BB neural network rule extraction method. Proceedings of 3rd annual SNN symposium on neural networks (eds. Kappen B and Gielen S) Springer-Verlag. 69-73.
21. Ripley BD. 1995. Statistical ideas for selecting network architectures. Proceedings of 3rd annual SNN symposium on neural networks (eds. Kappen B and Gielen S) Springer-Verlag. 183-190.

Simulati

Institute
Universit
e

Abstract

Analysis of sim
ever, their com
tions. A promi
Existing appro
pret. We prese
simulation dat
plex model fun
allows the ana
help to unders
world token b

1 Introdu

The use of m
ence for the c
the complexit
plexity of the
a highly time
simulations a
extremely co
A commo
which is less
task of analy
analyze the p
generation of
being analyze
Of partic
observations
experiments
to find some
mathematica

X. Liu, P. Co
LNCS 1280, p

Correlated Activity Pruning (CAPing)

By C. M. Roadknight[1], D. Palmer-Brown[1] and G. E. Mills[2].
1. Novel Architectures Group, Department of Computing, The Nottingham Trent University, Burton Street, Nottingham NG1 4BU.
2. Department of Life Sciences, The Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS.

Abstract.
The generalisation ability of an Artificial Neural Network (ANN) is dependent on its architecture. An ANN with the correct architecture will learn the task presented by the training set but also acquire rules that are general enough to correctly predict outputs for unseen test set examples. To obtain this optimum network architecture it is often necessary to apply a labourious 'trial and error' approach. One approach that helps to achieve optimum network architecture in a more intelligent way is pruning. Such methods benefit from the learning advantages of larger networks while reducing the amount of overtraining or memorisation within these networks. Sietsma and Dow (1988) describe an interactive pruning method that uses several heuristics to identify units that fail to contribute to the solution and therefore can be removed with no degradation in performance. This approach removes units with constant outputs over all the training patterns as these are not participating in the solution. Also, units with identical or opposite activations for all patterns can be combined. The approach to merging hidden units detailed in Sietsma and Dow's paper is useful, however, it only covers perfectly correlated, binary activations.

The method presented here generalises correlated activity pruning (CAPing) to all real valued positively and negatively, highly correlated activation sets from hidden neurons. There are several positive results to be gained by CAPing. Firstly, a speed up of the training and optimisation process can be achieved. Secondly, the weights of correlated units can be analysed and relationships between correlated features detected.

CAPing is initially applied to a theoretical curve fitting problem. Very little error is introduced until units with correlation coefficients of less than 0.8 are used. Analysis of the weights from the input units of the merged hidden units showed some sets of weights to be highly correlated but other sets to be non-correlated. This shows how CAPing removes two types of redundancy. Firstly when weight vectors represent similar features, and secondly differing features that occur for the same input patterns.

ANN's have been applied to the problem of predicting leaf damage to crop plants (Roadknight *et al* 1995, Balls *et al* 1996). ANN's trained to predict onset of leaf damage, with more hidden units than were required, were CAPed down to architectures with a near optimal number of hidden units.

Acknowledgements.
We would like to thank the UK Department of the Environment for financial support of the work in this project (Project number PECD 7/12/145)

References.

Balls GR, Palmer-Brown D, & Sanders GE. 1996. Investigating microclimate influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks. New Phytologist, 132, 271 -280

Roadknight CM, Palmer-Brown D & Sanders GE. 1995. Learning the equations of data. Proceedings of 3rd annual SNN symposium on neural networks (eds. Kappen B and Gielen S) Springer-Verlag. 253-257.

Sietsma J & Dow RJF. 1988. Neural net pruning - Why and how. Prc. IEEE Int. Conf. Neural Networks. Vol 1. p. 325-333.

ICP-CROPS AND CRITICAL LEVELS OF OZONE FOR INJURY DEVELOPMENT

J. Benton[1]., J. Fuhrer[2]., B.S Gimeno[3]., L. Skärby[4]., G. Balls[1]., D. Palmer-Brown[5]., C. Roadknight[5] and G. Sanders[1]

[1]) The ICP-Crops Coordination Centre, Department of Life Sciences, The Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, UK.
[2]) FAC, CH-3097, Liebefeld-Bern, Switzerland.
[3]) CIEMAT-Ima 3B, Avda. Complutense 22, Madrid 28040, Spain.
[4]) IVL, Box 47086, S-40258 Göteborg, Sweden.
[5]) Department of Computing, The Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK.

Abstract
Studies by the UN/ECE ICP-Crops in 1994 showed that visible injury to crops caused by ozone pollution was widespread across Europe indicating the extent of exceedance of the provisional short-term critical level (an AOT40 of 700 ppb.h accumulated during the daylight hours over three consecutive days). Injury was seen on a wide range of crops such as clover, bean, soybean, tomato and watermelon at ICP-Crops sites throughout Europe with the exception of Finland. Mean daily maximum ozone concentrations for individual clover experiments ranged from 40 ppb in Finland to 84 ppb in Switzerland. Examination of the daily ozone concentrations before the onset of injury indicated that injury usually developed following episodes where concentrations exceeded 700 ppb.h. However, in some cases, injury did not develop and possible explanations for this are being considered by the ICP-Crops. Analysis of ICP-Crops data using artificial neural networks (ANNs) indicated that lower ozone concentrations, following a high ozone episode, coincided with the appearance of visible injury. It is believed that ANN analysis of more comprehensive data sets obtained by the ICP-Crops in 1995 will improve the short-term critical level for visible injury (Level I and Level II).

Introduction

The role of the United Nations Economic Commission for Europe International Cooperative Programme on effects of air pollution and other stresses on crops and non-wood plants (UN/ECE ICP-Crops) is to monitor and document the effects of ambient air pollution on crop and more recently, non-wood plant species, throughout Europe. Since its initiation, 22 Parties of the Convention on Long-range Transboundary Air Pollution (LRTAP Convention) have participated in the programme. A large database now exists for ozone pollution and the visible injury and yield/biomass losses that it causes in sensitive crop species. The crops of primary interest to the ICP-Crops are from the *Leguminosae* family and include *Trifolium subterraneum* L. (subterranean clover), *Trifolium repens* L. (white clover) and *Phaseolus vulgaris* L. (green bean). Other species being studied include *Glycine max* (soybean), *Lycopersicon esculentum* (tomato), *Citrullus lanatus* (watermelon), *Nicotiana tabacum* (tobacco), *Malva sylvestris* L. (common mallow), *Cirsium arvense* (L.) scop. (creeping thistle) and *Plantago major* L. (great plantain). Characteristic injury symptoms such as bronzing, chlorosis and necrosis of leaf foliage have been widely reported throughout Europe on these and other ozone sensitive species (Sanders and Benton, 1995).

Two critical levels for ozone were established at the UN/ECE Critical Levels for Ozone Workshop held in Bern Switzerland (Fuhrer and Achermann, 1994) . These were (i) a long-term critical level which would cause reductions in biomass or yield if exceeded and (ii) a short-term critical level which, when exceeded, would cause visible injury to develop (Sanders *et al*, 1994).

Long-term critical level	-	an AOT40* of 5300 ppb.h accumulated during daylight hours over a three month period.
Short-term critical level	-	an AOT40 of 700 ppb.h accumulated during daylight hours over 3 consecutive days.

* accumulation of the hourly means of ozone where the mean exceeds 40 ppb.

The long-term critical level for ozone was derived from analysis of spring wheat (*Triticum aestivum*) data (Fuhrer, 1994) which were published as part of the results of the European Open-Top Chamber Project (EOTC) (Jäger *et al*, 1993). The provisional short-term critical level of ozone for visible injury was established using data collected from ambient air field observations by the ICP-Crops and The Netherlands Biomonitoring Network.

It is recognised that the use of a single critical level for a gaseous pollutant (Level I) is rather simplistic as it does not consider variations in the sensitivity of individual receptors to pollutants or the influence of environmental factors such as relative humidity and temperature on receptor response (Level II). Thus, it has been acknowledged that further research is required to improve the assessment of the geographical distribution of critical level exceedances in Europe (Sanders *et al*, in press) and that critical levels for ozone need to be defined which take environmental conditions into consideration (Gimeno *et al*, in press).

This paper presents some of the data from the 1994 ICP-Crops experiments which demonstrate the extent of exceedance of the short-term critical level for visible injury. In addition, the methodology and preliminary results are described of a computer modelling investigation into the factors which may modify the response of a receptor to ozone and the onset of visible injury.

Experimental methodology of the ICP-Crops

Each year, participants conduct a series of coordinated experiments across Europe to determine the effects of ambient ozone on crops. The Coordination Centre provides a comprehensive experimental protocol, seed and other equipment to ensure that experiments are standardised as far as possible (UN/ECE, 1994a). Participating countries in 1994 were Austria, Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Latvia, The Netherlands, Poland, Russian Federation, Spain, Sweden, Switzerland and the United Kingdom.

Ambient air experiments to determine the extent of visible injury on sensitive crops:
T.subterraneum cv. Geraldton (Subterranean clover) and T. repens cv. Menna (white clover) :
Forty pots containing three *T.subterraneum* or *T. repens* clover seedlings were used for each experiment at each experimental site. Twenty pots were treated with 100 ml 150ppm ethylene diurea (EDU) at emergence of the first trifoliate leaf and then after at regular intervals in order to establish alterations in biomass due to ozone exposure (data not presented). EDU is a chemical which is effective at protecting plants from ozone injury and can be applied to the foliage or to the roots (Carnahan *et al*, 1978). For example, a soil application of EDU to *Phaseolus vulgaris* was shown to provide complete protection against visible injury caused by exposure to ozone (Bergmann *et al*, 1993). The remaining twenty pots were treated with 100 ml of distilled water and used to determine the extent of visible injury. All pots were placed outside in ambient conditions. The pots were watered by a self-watering system placed below ground. Plants were examined daily for the appearance of visible injury and assessed 28 and 56 days after emergence of the first trifoliate leaf for the number of ozone injured leaves and the total number of leaves. From these data, the percentage of injured leaves could be calculated per pot. The leaves were removed from the plants at 28 days and their dry weight determined (data not shown). Plants were harvested again 56 days after emergence and the total plant dry weight determined (data not shown).

Throughout the growing period, climatic conditions (temperature, relative humidity, daylight hours, rainfall) and pollutant levels (ozone, sulphur dioxide and oxides of nitrogen) were recorded continuously.

P. vulgaris cv. Lit, cv. Stella, cv. Groffy (green bean):

Twenty bean plants of each cultivar (one per pot) were exposed to ambient conditions for approximately 100 days. Plants were assessed daily for the onset of ozone injury and the percentage of ozone injured leaves determined. Pollution and climate conditions were monitored throughout the growing season.

G. max cv. Ceresia (soybean), L. esculentum cv. Tiny Tim (tomato), B. napus cv. Comet (oilseed rape) and Phaseolus vulgaris cv. Lit (green bean):

Ten plants (one per pot) were placed in ambient ozone conditions for 56 days (tomato and soybean), 70 days (green bean) or 96 days (oilseed rape) and assessed daily for the appearance of ozone injury. Pollution and climatic conditions were monitored throughout the experiment.

C. lanatus cv. Sugar Baby (watermelon), N. tabacum cv. Bel-W3, Bel-C, Bel-B (tobacco):

Extensive field surveys were conducted in Valencian and Catalan Autonomous Communities in eastern Spain to evaluate the extent of ozone injury on tobacco and watermelon plants. In total, 20 sites were used in the tobacco survey and several commercial fields in Valencia and Tarragona were visited to determine the amount of injury on watermelon.

Analysis of data by the use of artificial neural networks (ANN's):

The theory of artificial neural networks:

Over the last decade, significant advances have been made in pattern recognition and computational learning and this includes the development of artificial neural networks (ANNs). ANNs perform functions which are similar to some brain functions (Kothari and Heekuck, 1993). They do not rely on a program containing step by step procedures in order to function but actually 'learn' by studying patterns within data sets which is very much like the process of the human brain. They can also deal with 'noisy' or incomplete images or patterns which is another key feature of the brain (Kothari and Heekuck, 1993). Similar to biological systems, neural networks consist of many, connected, simple processing elements called neurons or nodes (Figure 1). They are ideal to use for analysis of data obtained from experiments where many parameters have an influence on the response. For example, the extent of ozone injury may be modified by factors such as relative humidity, temperature and photosynthetically active radiation (PAR) as these influence stomatal conductance (Kappen and Haeger, 1991) and thus the flux of ozone into the plant (Kersteins *et al*, 1992).

There are several different forms of ANN and those used in this work learn the data by a process of error minimisation. Initially the network must be trained with numerous items of data (the training set) enabling it to detect patterns within the data. The input layer receives data which represents the causal agents of an effect whereas the neurons in the output layer contain data which represents the 'actual' effect. From the input data, the network calculates a predicted output value and compares this with the 'actual' value entered into the output layer. The network then makes modifications which relate the predicted output to the 'actual' output. It does this by continually modifying the strength of the connections between the input and output neurons until the learning process is complete and the network can generate output which is a sufficiently accurate and meaningful interpretation of the input for all the training examples. The network can be tested by entering new data (independently selected) in to the input neurons and then comparing the predicted output response with an 'actual' response. In addition, the strength of the connections of the network may be analysed to indicate the relative importance of parameters entered into the input neurons.

The artificial neural network approach has been used in many scientific situations. For example, ANNs have been used to model complex data sets obtained from closed chamber experiments investigating the effects of different microclimates, ozone doses and exposure times on the extent of injury using *T. subterraneum* (Balls *et al*, in press). This approach has also been used in the taxonomic identification of plankton (*Ceratium* spp) (Simpson *et al*, 1992).

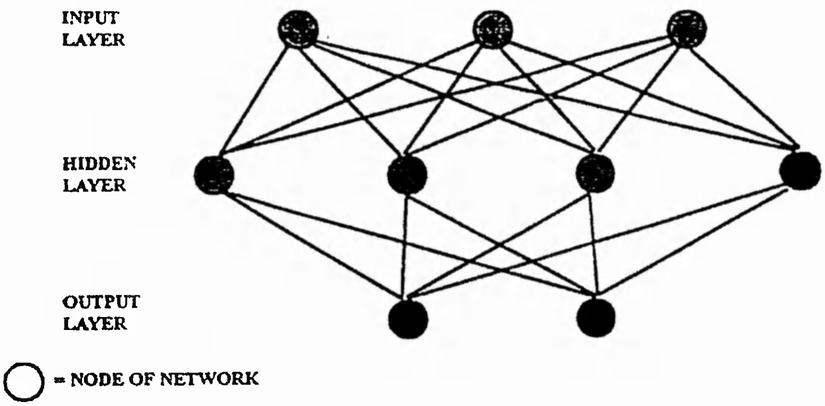


Figure 1. A simplified model of a neural network structure (adapted from Kothari and Heekuck, 1993).

Analysis of the ICP-Crops data using ANNs

Analysis of the ICP-Crops data requires intensive processing and thus a 64 processor parallel computer system (CNAPS) has been employed for all ANN computations. This system is ideal for neural network applications because of its increased speed in handling sets of data where each data element is operated in a similar way. All of the active processors in the CNAPS can execute the same instruction simultaneously thus, performing an operation on an entire block of data in one step. This is known as a SIMD architecture (Single Instruction, Multiple Data).

ANN models were developed using data from sites where the date of onset of ozone injury, hourly ozone concentrations, and climatic conditions were recorded for *P. vulgaris*, *T. repens* and *T. subterraneum* experiments. These were Austria, Belgium, Italy, The Netherlands, Sweden, Switzerland and the United Kingdom. ANN's were used to elucidate why injury occurred after one 3 or 5 day ozone episode but not after another. The ozone parameters entered into the network were the 7 hour mean, the daylight AOT40 and the daily maximum ozone concentration. The day after emergence on which injury occurred and whether injury occurred before or after the first harvest were also entered. All values were normalised so that the largest value for each input was 1 and the lowest value for each input was 0. The intermediate values were linearly transformed. The 3 or 5 day period before the onset of ozone injury was given an output of 0.9 and the 3 or 5 day period preceding the absence of injury was given a value of 0.1 (it is better not to operate at the extremes of the range). The 3- and 5-day networks were trained using 24 and 18 complete data sets respectively. These data sets contain data for both the input and output neurons. Following training, a network configuration was discovered that would predict whether leaf injury would appear or not given certain inputs. The 3- and 5-day models were tested with 8 and 6 further data sets respectively, previously unknown to the network, to see whether the model could correctly predict whether injury would occur.

Results and discussion

The extent and onset of visible injury on crops throughout Europe.

Visible injury due to ambient ozone conditions was widespread during the 1994 growing season (Figure 2) indicating that the short-term critical level for visible injury was exceeded widely across Europe. A wide range of ozone concentrations were recorded at ICP-Crops sites in 1994. For example, the mean daily maximum ozone concentrations measured during the first clover (*T.subterraneum* or *T. repens*) experiment (usually June and July) in Finland and Switzerland were 40 ppb and 84 ppb respectively. The range of mean daily maximum ozone concentrations at *P.vulgaris* sites was from 40 ppb in Finland to 70 ppb in Italy (Naples). Many crop species were susceptible to ambient ozone pollution (Table 1). The variability in response emphasises the need for a clearly defined short-term critical level for visible injury for each species or group of species (Level II).

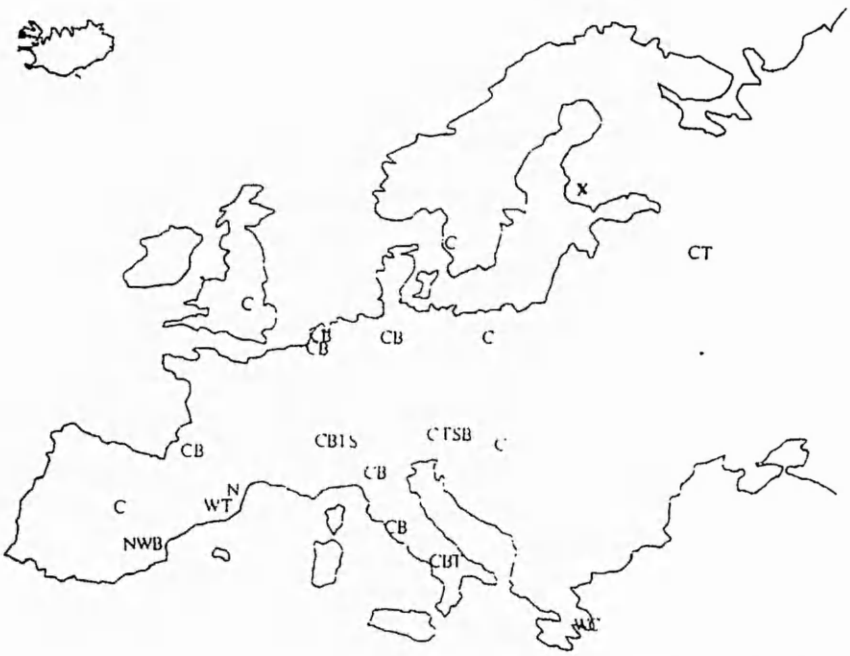


Figure 2. Map showing the ICP-Crops sites where visible injury was seen on crop species in 1994 (C - clover; B - beans, T - tomato; S - soybean, N - tobacco and W - watermelon, X - no injury).

Table 1. Ozone injury on different crops species from selected ICP-Crop sites (*T.subterraneum* data is from the 2nd harvest of the first experiment)

Country	Crop species	% injured leaves (per pot or plant)	Growing period AOT40 (ppb.h)
Austria	*P. vulgaris cv. Lit* *T. subterraneum* *G. max* *L. esculentum*	19.5 7.4 11.6 16.6	13129 6434 13129 13129
The Netherlands	*T. subterraneum*	7.5	4604
Sweden (Ostad)	*T.subterraneum*	45.5	3064
Switzerland	*T.subterraneum* *P. vulgaris cv. Lit* *G.max*	34.5 32.1 2.0	8012 5125 15950

The short-term critical level for visible injury was investigated further using data from sites where the onset date of visible injury was recorded and where ozone concentrations were monitored hourly. Daily AOT40 values for the growing season of the crop were plotted and compared with the date that injury was first identified (for example, Fig. 3). Thus, it was possible to see if injury occurred after a three day episode where the AOT40 was equal to, or greater than, 700 ppb.h. Visible injury was first seen on *T. subterraneum* clover in Belgium on July 4th (experiment 1) and 25th (experiment 2) after three days where ozone concentrations were greater than 700 ppb.h (Fig. 3a). Similarly, injury occurred on July 5th and Aug 1st after high three day ozone episodes at the experimental site in The Netherlands (Fig.3b). However, injury also occurred on September 20th in The Netherlands even though the AOT40 did not exceed 700 ppb.h in the preceding 3 days. This latter effect was also seen at a site in Austria (Fig. 3c) and in Sweden (Fig. 3d) and this variability suggests that the provisional short-term critical level for clover may require revision following further data analysis To substantiate this, open top chamber studies in Sweden by Pihl Karlsson *et al* (1995) have shown that 50% of clover leaves are injured before the AOT40 is exceeded and it has been suggested by results from the 'Clover Sweden' project that an AOT 'cut off' value of 40 ppb is too high for clover grown in Swedish conditions (Pihl Karlsson *et al*, in press). In contrast to this, there was a large ozone episode at the end of June in the United Kingdom (Fig. 4) but injury failed to develop on *T. subterraneum* until July 13th after the 1st harvest.

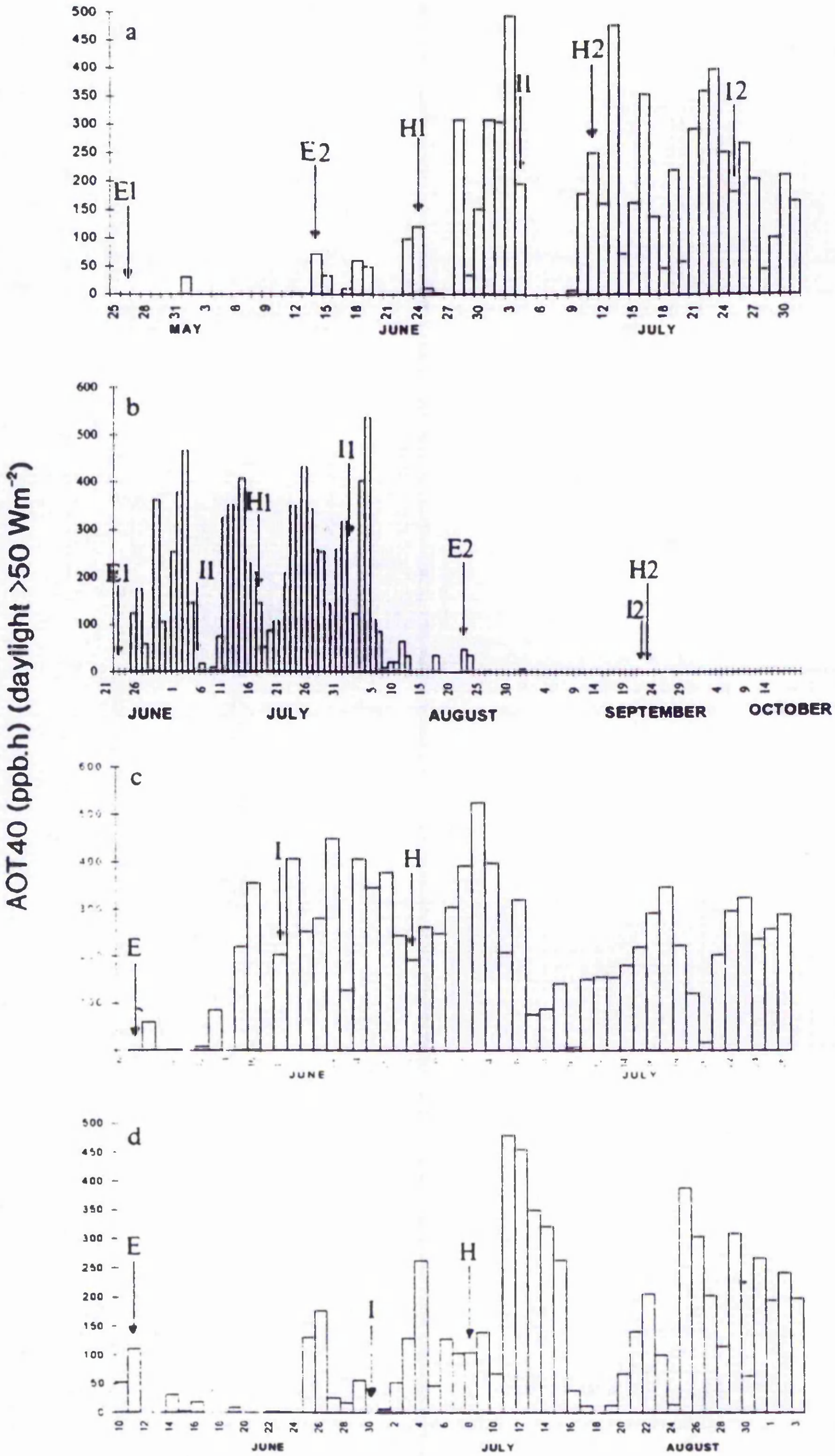


Figure 3. Daily AOT40 values and the onset of injury on *T. subterraneum* clover in (a) Belgium and (b) The Netherlands (c) Austria and (d) Sweden) (E - emergence of 1st trifoliate ; I - injury and H - harvest; 1 - experiment number 1; 2 - experiment number 2).

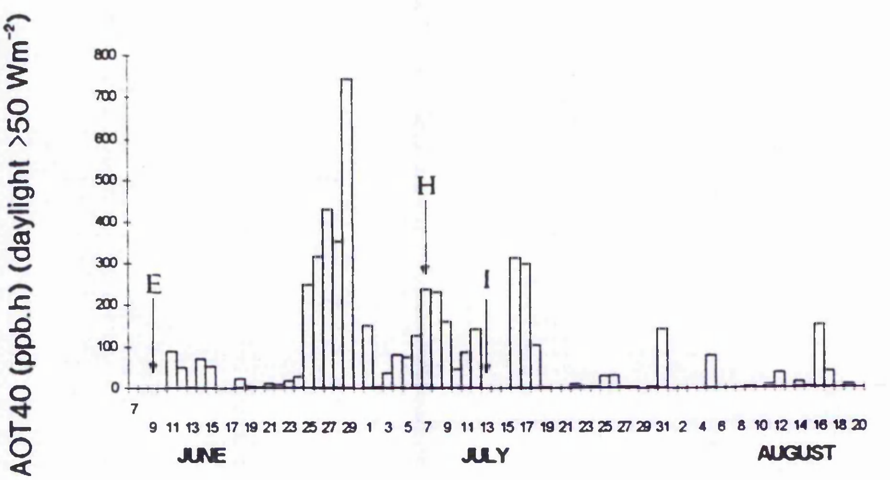


Figure 4. Daily AOT40 values and the onset of ozone injury on *T. subterraneum* clover in The United Kingdom (E - emergence of 1st trifoliate, I - injury and H - harvest)

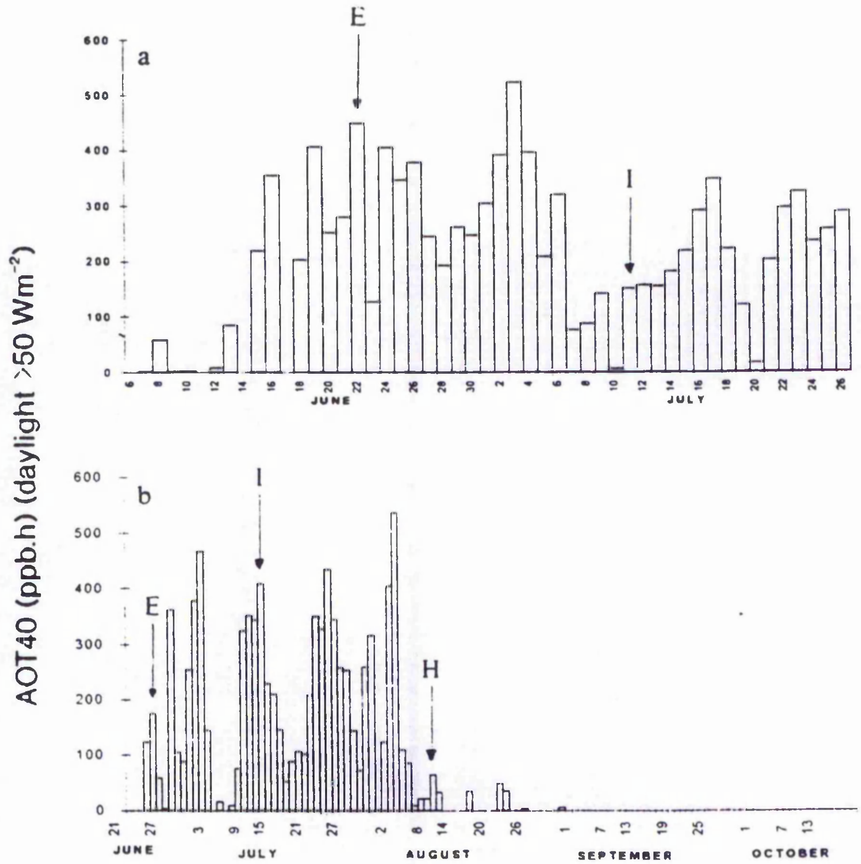


Figure 5. Daily AOT40 values and the onset of ozone injury on *P. vulgaris* foliage in (a) Austria and (b) The Netherlands (E - emergence of 1st trifoliate, I - injury and H - harvest)

Injury occurred on *P. vulgaris* foliage in Austria (Fig. 5a) and The Netherlands (Fig. 5b) after episodes where the AOT40 was above 700 ppb.h for 3 consecutive days although it is unclear which 3 days were responsible for causing injury. In summary, the data from the ambient air experiments of the ICP-Crops provides some support for the use of a short-term critical level for visible injury of 700 ppb.h. However, the inconsistencies in the data indicate that this value needs substantiating.

Analysis of data using ANNs:

The complexity of the information gained from the ICP-Crops ambient air experiments necessitated the use of a novel approach to data processing and modelling as two dimensional statistical methods are believed to be ineffective. ANNs were chosen for data analysis because it is thought these will determine the influence of multiple modifying factors such as climate on the critical levels of ozone for visible injury. In addition, equations from the networks may be extracted to justify and interpret the behaviour of the network. These equations may be used to predict critical levels in different physical and pollution climates since they constitute an explicit, fully-interpretable model of the data which can be related to scientific processes which are already understood or that are actively being researched. An example of equation extraction can be given for the networks which had been successfully trained, as they correctly predicted that injury did/did not occur for nearly all the test data sets and provided some useful information regarding the onset of injury. The analysis indicates that the sequence of ozone concentrations during an episode may be important to the onset of injury. The network trained with the 3-day data indicated that a day with low ozone concentration coincided with the appearance of ozone injury on the following day. For example, equation 1 has been extracted from the network and is a simplified form of the equation that represents the most important hidden nodes of the trained 3-day network. It indicates that ozone concentrations on day 2 and 3 are more important than those on day 1 and the negative sign at day 3 indicates the lower ozone concentration on the day before injury development. This was also indicated by part of the equation extracted from the 5-day network as day 5 (AOT40) is multiplied by -3 (Equation 2). This suggests that a lower ozone concentration may serve to accelerate the onset of ozone injury. The 5-day network also showed the positive influence of day number on the onset of injury confirming that injury observation occurs on older plants (Equation 2).

Equation 1: day 2 (7h mean + max ozone) - day 3 (AOT40 + max ozone)

Equation 2: F(day 1 (AOT40 - 7h mean) + day 2 (AOT40) + day 4 (7h mean) - 3*day 5 (AOT40))

+ F(sum (no. of days after emergence for days 1 to 5) - day 2 (7h mean))

(F is a function with a minimum value 0 and maximum value 1 which increases with its input value).

These effects can be related to the results shown by Figure 3. For instance, injury to *T. subterraneum* clover in Belgium (Fig. 3a) and The Netherlands (Fig 3b) occurred following a day when the ozone concentration was lower than on previous days although this effect was not quite as clear with the *P.vulgaris* data (Fig. 5).

Results of the analysis of ICP-Crops data by ANNs has provided some useful information regarding the short-term critical level for visible injury. It has been indicated that day number, which can be related to leaf and plant age, influences the onset of injury and this is in agreement with the findings of others. For example, open-top chamber studies by Pihl Karlsson *et al* (1995) have shown that *T. subterraneum* leaves of up to seven days old can withstand injury even at high ozone concentrations. It is believed that the leaf is most ozone sensitive following the period of maximum expansion rate but before the period of maximum surface to volume ratio and full expansion involving the formation of the secondary cell wall and lignification. In addition, visible injury was more apparent on *C. lanatus* (watermelon) after the flowering and fruiting growth stage rather than in the early stages of plant development (Gimeno *et al*, 1995; UN/ECE, 1994b) which clearly indicates that younger plants are less sensitive to ozone pollution.

The network predicted that the timing and magnitude of ozone concentrations may be important for injury development as a day with a lower ozone concentration seemed to coincide with the expression of ozone injury. This effect has also been indicated by studies using closed chambers and *T. subterraneum* (Balls *et al*, in press). Data analysis using an ANN (Neuroshell 2, Ward Systems Group Inc., USA) gave predictions that plant sensitivity decreased as the time of ozone exposure increased and this was consistent with 'actual' experimental data.

Further studies incorporating the use of ANNs have shown that other factors are important for the onset of injury on clover plants and these need careful consideration when establishing and applying the Level II concept of the short-term critical level. ANN predicted injury and actual injury indicated that clover plants were more sensitive to ozone at lower PAR, lower temperatures and at higher humidities. For example, plants were more sensitive to ozone at 100μmol m^{-2} s^{-1} PAR, 20°C and 55% RH. as opposed to 350μmol m^{-2} s^{-1} PAR, 28°C and 28% RH (UN/ECE, 1995). This effect can be partly related to the stomatal response to humidity. Stomata close in response to increased transpiration and a reduction in guard cell turgor (Mott and Parkhurst, 1991) but if the humidity is high then the tendency is for the stomatal pore to remain open which allows the passage of ozone into the leaf tissue. Similar results were shown by Mortensen (1992) using growth chambers. The shoot dry weight of ozone treated *L.esculentum* was reduced to a greater extent at 90% RH than at 70% RH and ozone injury was reduced at higher light intensities. Thus, the results obtained from closed chamber experiments can be used to indicate how environmental conditions influence the extent of ozone injury seen in the field. Evidence for this was obtained by field surveys of *N.tabacum* grown in eastern Spain. Ozone injury was more prevalent in the coastal regions of Catalunya which had a higher relative humidity than regions further inland (Gimeno *et al*, in press).

Following the 1995 experimental season, much more data will be available from the ICP-Crops which will be used to verify the results and theories described above. As well as recording the actual date of onset of injury, participants have assessed the level of injury. This means that an actual value can be entered into the network and not just numbers corresponding to whether injury did or did not occur. Daily values for relative humidity and temperature and PAR will be entered into the network for the days preceding the onset of injury in addition to other parameters describing daily ozone concentrations. This will determine the relative importance of environmental factors as well as ozone concentrations for the development of ozone injury. In addition, further species have been examined and assessed for injury at regular intervals. These results will provide further information regarding 'stocks at risk' from ozone pollution and for the derivation of the short-term critical level for visible injury (Level I and Level II) and this information can be further used to verify the predictions made on exceedance maps.

Conclusion

Ozone injury was widespread across Europe in 1994 providing evidence that the critical level for visible injury was exceeded. Investigation of the AOT40 profiles for the days before the onset of injury indicated that, although injury usually developed after an AOT40 of 700 ppb.h, this was not always the case. Artificial neural networks have indicated that the plants may be responding to the pattern of daily ozone concentrations. A day with low ozone concentration may be necessary before the appearance of visible injury and the response to individual episodes may depend upon the previous ozone history of the plants. It is believed that following analysis of data from the 1995 season, the ICP-Crops will be in a position to provide a more 'robust' Level I short-term critical level together with suggestions for Level II.

Acknowledgements

We would like to acknowledge all the participants of the ICP-Crops and The Department of the Environment (UK) for funding the coordination of the ICP-Crops experiments (contract number EPG/1/3/13).

References

Balls GR., Palmer-Brown D., Cobb AH and Sanders GE. Towards unravelling the complex interactions between microclimate, ozone dose and ozone injury in clover. In press. *Journal of Water, Air and Soil Pollution.*

Bergmann E., Weigel H-J and Jäger H-J (1993) Mode of action of EDU on two bean varieties of different O_3 susceptibility. In: Effects of Air Pollution on Agricultural Crops in Europe. Results of the European Open-Top Chambers Project. Jäger H-J., Unsworth M., De Temmerman L and Mathy P (eds). CEC Air Pollution Research Report 46, Brussels.

Carnahan JE., Jenner EL and Wat EKW (1978) Prevention of ozone injury in plants by a new protectant chemical. *Phytopathology*, **68**, 1225-1229.

Fuhrer J (1994) The critical level for ozone to protect agricultural crops - an assessment of data from European open-top chamber experiments. In: Critical Levels for Ozone; a UN/ECE Workshop Report. Fuhrer J and Achermann B. (eds) Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH - 3097 Liebefeld-Bern, Switzerland, No. 16, 42 - 57.

Fuhrer J and Achermann B (eds) Critical Levels for Ozone; a UN/ECE Workshop Report. Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH - 3097 Liebefeld-Bern, Switzerland. No. 16, 1994.

Gimeno BS., Peñuelas J., Porcuna JL and Reinert RA. Biomonitoring ozone phytotoxicity in Eastern Spain. In press. *Journal of Water, Soil and Air Pollution.*

Gimeno BS., Bermejo V., Mendoza M and Sánchez S. Ozone effects in agricultural crops at the Ebro Delta in 1995. Poster presentation at the 'Exceedances of Critical Levels and Loads' Workshop, Vienna, November 22 - 24, 1995.

Jäger HJ., Unsworth M., De Temmerman L., Mathy P (eds). Effects of Air Pollution on Agricultural Crops In Europe. Results of the European Open-Top Chambers Project. CEC Air Pollution Research Report 46, Brussels. 1993.

Kappen L and Haeger S (1991) Stomatal responses of *Tradescantia albiflora* to changing air humidity in light and darkness. *Journal of Experimental Botany*, **42**, 979 - 986

Kothari SC and Heekuck OH (1993) Neural Networks for Pattern Recognition. In: Advances in Computers, **37**, 119 - 166.

Kersteins K., Federholzner R and Lendzian KJ (1992) Dry deposition and cuticular uptake of pollutant gases. *Agriculture, Ecosystems and Environment*, **42**, 239 - 253.

Mortensen LM (1992) Effects of ozone concentration on growth of tomato at various light, air humidity and carbon dioxide levels. *Scientia Horticulturae*, **49**, 17 - 24.

Mott KA and Parkhurst DF (1991) Stomatal responses to humidity in air and helox. *Plant Cell and Environment*, **14**; 509 - 515.

Pihl Karlsson G., Selldén G., Skärby L and Pleijel H (1995) Clover as an indicator plant for phytotoxic ozone concentrations: visible injury in relation to species, leaf age and exposure dynamics. *New Phytologist*, **129**, 355-365.

Pihl Karlsson G., Pleijel H., Sild E., Danielsson H., Selldén G., Ericson L and Skärby L. Clover Sweden - A national three year study of the effects of tropospheric ozone on *Trifolium subterraneum* L. In press. *Journal of Water, Air and Soil Pollution.*

Sanders G., Balls G and Booth C (1994) Ozone critical levels for agricultural crops - Analysis and interpretation of the results form the UN/ECE International Cooperative Programme for Crops. In: Critical Levels for Ozone; a UN/ECE Workshop Report. Fuhrer J and Achermann B. (eds) Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH - 3097 Liebefeld-Bern, Switzerland, 58 - 72.

Sanders G and Benton J (1995) Ozone Pollution and Plant Responses in Europe - an Illustrated Guide. The ICP-Crops Coordination Centre, The Nottingham Trent University, United Kingdom.

Sanders GE., Skärby L., Ashmore MR and Fuhrer J. Establishing critical levels for the effects of air pollution on vegetation. In press. *Journal of Water, Air and Soil Pollution.*

Simpson R., Williams R., Ellis R and Culverhouse PF (1992) Biological pattern recognition by neural networks. Marine Ecology Progress Series. **79**, 303-308.

UN/ECE (1994a) The ICP-Crops Experimental Protocol. The ICP-Crops Coordination Centre. The Nottingham Trent University, Nottingham, United Kingdom.

UN/ECE (1994b) Annual Report of the ICP-Crops - Activities performed in Spain. CIEMAT, CREAF. Servicio de Sanidad Vegetal Generalitat Valenciana, Spain.

UN/ECE (1995) Progress Report for the ICP-Crops. The ICP-Crops Coordination Centre, The Nottingham Trent University, Nottingham, United Kingdom.

COPY

The Critical Level of Ozone for Visible Injury on Crops and Natural Vegetation (ICP-Crops)

J Benton[a]; J Fuhrer[b]; BS Gimeno[c]; L Skärby[d]; D Palmer-Brown[e]; C Roadknight[e] and G Sanders-Mills[a]

[a]The ICP-Crops Coordination Centre, Department of Life Sciences, The Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, United Kingdom.
[b]IUL, Institute for Environmental Protection and Agriculture, CH-3097 Liebefeld-Bern, Switzerland.
[c]Ciemat-Ima 3B, Avda. Complutense 22, Madrid 28040, Spain.
[d]IVL, Box 47086, S-40258 Göteborg, Sweden.
[e]Department of Computing, The Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, United Kingdom.

Abstract

Data from the UN/ECE ICP-Crops experiments with crop and natural species have shown that exceedance of the provisional short-term critical level of ozone for visible injury is widespread throughout Europe. However, data from the 1994/95 clover (*Trifolium subterraneum* and *T. repens*) experiments have shown that injury occurs before the critical level is exceeded at many experimental sites. Thus, further analysis of the climate and pollutant conditions in the 5 days preceding injury has been done. Three dimensional graphs of AOT40 (global radiation ≥ 50 Wm^{-2}), percentage relative humidity and either global radiation or temperature indicated that visible injury occurred when AOT40 was less than 500 ppb.h and relative humidity was greater than 50%. Injury was also observed if the AOT40 for the 5 days was above 500 ppb.h and the relative humidity less than 50%. Furthermore, if mean VPD was above 1.5 kPa, injury did not occur until the AOT40 was greater than 500 ppb.h. If VPD was below 1.5 kPa injury occurred at an AOT40 below 500 ppb.h. Artificial neural networks (ANNs) indicated that the most important inputs for injury development were ozone concentration (AOT40, 7 hr mean and mean daily maximum), VPD (7 hr mean) and global radiation (mean daily maximum). Furthermore, ANN analysis indicated that a day with low ozone concentration, following an ozone episode, coincided with injury expression. Injury did not develop on the leaves of *Malva sylvestris* plants until two months after emergence even though the critical level was exceeded. These results suggest that the provisional short-term critical level requires modification and therefore two critical levels for injury are recommended which consider VPD conditions.

1 Introduction

Following the UN/ECE Critical Levels For Ozone Workshop in Bern (1993) (Fuhrer and Achermann, 1994), a provisional short-term critical level of ozone for the development of visible injury on crops was set at an AOT40* of 700 ppb.h accumulated during daylight hours (global radiation ≥50Wm^{-2}) over three consecutive days. This critical level was established using data obtained from The Netherlands Biomonitoring Network and data collected from ambient air field experiments by the UN/ECE ICP-Crops**. This approach is recognised as being Level I as it assigns a single value to protect the most sensitive known receptor (Sanders *et al*, 1995). It does not consider factors causing variation in the sensitivity of individual receptors which are included in a Level II approach. Data presented from the 1994 ICP-Crops experiments at the Workshop on Exceedances of Critical Loads and Levels (Vienna, 1995) showed that injury occurred before the AOT40 exceeded 700 ppb.h. This was possibly because of factors which influence plant responses to ozone such as climatic conditions during an episode. It was also described how injury did not always occur following a 3 day ozone episode where the AOT40 had exceeded 700 ppb.h (Benton *et al*, 1995). This latter effect was partially explained by analysis of 1994 ICP-Crops data using artificial neural networks (ANNs). This analysis suggested a day with low ozone concentration, following a high ozone episode, coincided with the onset of injury. Thus, it is feasible that if ozone concentrations remain high, injury may not develop despite exceedance of the critical level or that a decrease in ozone concentration may accelerate the onset of injury. At this workshop, it was concluded that further work was necessary in order to improve the Level I short-term critical level and to enable suggestions to be made for a Level II approach (Benton *et al*, 1995).

This paper describes the results from ICP-Crops ambient air experiments conducted in 1994/1995 including the extent of visible injury on crops and natural vegetation throughout Europe. Furthermore, the effects of climatic factors on the short-term critical level for injury are described with direct reference to ICP-Crops data. The results are discussed in relation to plant and climate interactions and how these may influence the flux of ozone into the plant. From the results and discussions, suggestions are made for two short-term critical levels of ozone for visible injury.

* accumulation of the hourly means of ozone where the mean exceeds 40 parts per billion

** United Nations/Economic Commission for Europe International Cooperative Programme on effects of air pollution and other stresses on crops and non-wood plants

2 Procedure

21 General

Participants conducted a series of coordinated experiments across Europe in order to determine the effects of ambient ozone concentrations on crop and natural species. The Coordination Centre provided an experimental protocol (UN/ECE, 1994; 1995), seed and other equipment to ensure that experiments were standardised as far as possible. Participating countries in 1994 and 1995 were Austria, Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Latvia, The Netherlands, Poland, Russian Federation, Slovenia, Spain, Sweden, Switzerland and the United Kingdom (Note: all countries did not participate in each year).

22 Ambient air experiments to determine the extent of visible injury on sensitive species *Trifolium subterraneum* cv. Geraldton (subterranean clover) and *Trifolium repens* cv. Menna (white clover): At each experimental site, forty pots containing three *T.subterraneum* or *T. repens* clover seedlings were used for each experiment. Twenty pots were treated with 100 ml of 150 ppm ethylene diurea (EDU) at emergence of the first trifoliate leaf and thereafter at 14 day intervals in order to establish alterations in biomass due to ozone exposure (data not presented). EDU is a chemical which is effective at protecting plants from ozone injury and can be applied to the foliage or to the roots (Carnahan *et al*, 1978). The remaining twenty pots were treated with 100 ml of distilled water for determining the extent of visible injury and all pots were placed outside in ambient conditions. The pots were watered by a self-watering system placed below ground. Plants were examined daily and the date when injury was first observed before the first and after each consecutive harvest was recorded.

Other crop and natural species

Ten plants (one per pot) were raised in the glasshouse and placed outside in ambient ozone conditions after the emergence of the first true leaf for 56 days (*Lycopersicon esculentum* (Tomato); *Glycine max* (Soybean): *Malva sylvestris* (Common Mallow); *Cirsium arvense* (Creeping Thistle); *Plantago major* (Plantain); 70 days *(Phaseolus vulgaris* (bean) and 84 days *Helianthus* sp (sunflower). Plants were observed daily for the appearance of visible injury.

23 Monitoring of climate and pollutants

Throughout the growing period of all species, climatic conditions (temperature, percentage relative humidity, global radiation, rainfall) and pollutant levels (ozone, sulphur dioxide and oxides of nitrogen) were recorded continuously.

24 Analysis of data

Daily AOT40 values (where global radiation $\geq$ 50 Wm^{-2}) were plotted to show the onset of injury in relation to ozone episodes and the developmental stage of the plant. The ozone concentrations (AOT40) were summed for the three and five day period preceding the onset of ozone injury for each clover experiment at each site. This would define if there were specific areas within Europe where injury was observed before an AOT40 of 700 ppb.h was exceeded. In addition, three dimensional graphs of total AOT40, mean % relative humidity (0901 - 1659 h***) and either mean global radiation or temperature (0901 - 1659 h) for the 3 and 5 day period before the onset or absence of injury were drawn using SlideWrite (version 2.1). A five day period was used in addition to a three day period because of the uncertainty surrounding which individual daily ozone concentrations were actually responsible for causing injury. Furthermore, vapour pressure deficit (VPD) was calculated for the 5 days preceding injury or the absence of injury.

*** values used were a 7 hr mean from within this time period

25 Data analysis by artificial neural networks (ANNs)

Over the last decade, significant advances have been made in pattern recognition and computational learning which includes the development of artificial neural networks (ANNs). Reviews of the theory and uses of ANNs are given by Kothari and Heekuck (1993), Benton *et al* (1995) and Balls *et al* (1996). ANNs were chosen for data analysis because they can determine the influence of multiple modifying factors such as climate on the critical levels of ozone for visible injury. ANNs consist of a series of connected processing elements called neurons or nodes and these receive (i) input values and (ii) output values. For example, the input values could be climatic and ozone parameters and the output values could be the presence or absence of injury (Figure 1). The weightings of the interconnections between the nodes of the network can be analysed for each input factor to indicate the relative importance of the input parameters (Balls *et al*, 1996). In addition, equations from the networks may be extracted to justify and interpret the behaviour of the network (Roadknight *et al*, 1995). These equations may be used to predict injury expression in different physical and pollution climates since they constitute an explicit, fully-interpretable model of the data which can be related to scientific processes which are already understood or that are actively being researched.

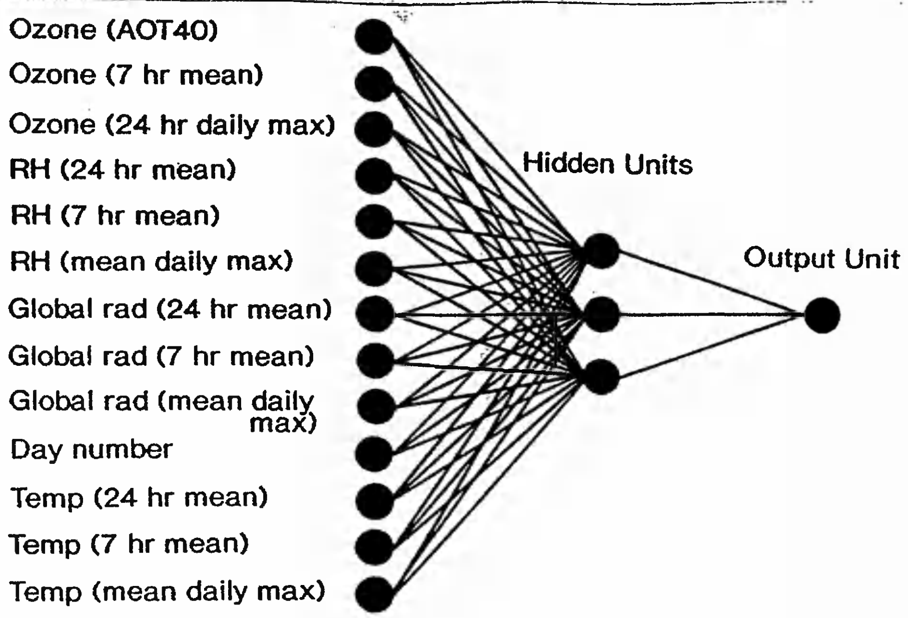


Figure 1. The structure of an artificial neural network

Three ANN models were developed using 5 day data from sites where the date of onset of ozone injury, hourly ozone concentrations and climatic conditions were recorded for *T. repens* and *T. subterraneum* experiments. These were Austria, Belgium, France, Germany, Italy, The Netherlands, Sweden, Switzerland and the United Kingdom. All three models were to explain why injury occurred after one 5 day ozone episode but not after another but the second and third models were to also show which climatic factors were important in the 5 days before injury.

Network 1
The input parameters entered into the first network were ozone concentrations (7 hr mean (0901 - 1659h), AOT40 (global radiation $\geq$ 50 Wm^{-2}) and mean daily maximum) for 5 days before injury/no injury.

Network 2
Ozone parameters (as above) *plus* global radiation and VPD (24 hr mean, 7 hr mean (0901 - 1659 h), mean daily maximum) for 5 days before injury/no injury were inputs for the second network.

Network 3
As Network 2 *plus* inputs for relative humidity and temperature (24 hr mean, 7 hr mean (0901 - 1659 h), mean daily maximum).

The number of days from emergence to injury was also entered into all networks. The output parameter for each network was whether injury did/did not occur.

The networks were initially trained using complete data sets which contained data for both the input and output neurons. Following training, a network configuration was discovered that would predict whether leaf injury would appear given certain inputs. Further data sets, previously unknown to the network, were used to test the model. Equations and the weightings of the interconnections for each input factor were extracted from the networks in order to explain their behaviour and to indicate which input parameters were important for injury development.

3 Results

31 The occurrence of visible injury on crops and natural species in 1995

Visible injury due to ambient ozone concentrations was widespread across Europe in 1995 both on crop and natural species (Figure 2). Injury was detected in fifteen countries and as in 1994, injury was not seen in Finland (experiments in Denmark were with open top chambers). In Finland, ozone concentrations were low in comparison to those at sites where injury developed. For example, the AOT40 values in Finland (Jokioinen), Germany (Braunschweig) and Italy (Milan) for the first clover experiment (May to August) were 264, 7844 and 14028 ppb.h respectively. Characteristic ozone injury (bronzing, chlorosis and necrosis) was seen on a range of crop species including *T. subterraneum*, *T. repens*, *T. alexandrinum*, *P. vulgaris*, *G. max*, *L. esculentum*, *C. lanatus* and on natural species including *M. sylvestris* and *C. arvense*.

32 Daily AOT40 values and the onset of injury

Analysis of the 1995 data revealed that visible injury occurred on *Trifolium* species following the exceedance of the provisional short term critical level. For example, in Belgium (Figure 3), ozone injury was observed on *T. subterraneum* on two occasions after the AOT40 had exceeded 700 ppb.h. However, towards the end of the growing season, injury also occurred before the AOT40 had exceeded 700 ppb.h in the 3 days before injury. Figure 4 shows the onset of visible injury on *Malva sylvestris* (Common Mallow) in Switzerland. Injury did not develop on this species until two months after the beginning of the experimental season despite exposure to high ozone concentrations. This effect was also seen on *Malva sylvestris* in Belgium (data not shown) which suggests that injury does not occur in the early stages of the life cycle of this species.

(a) (b)

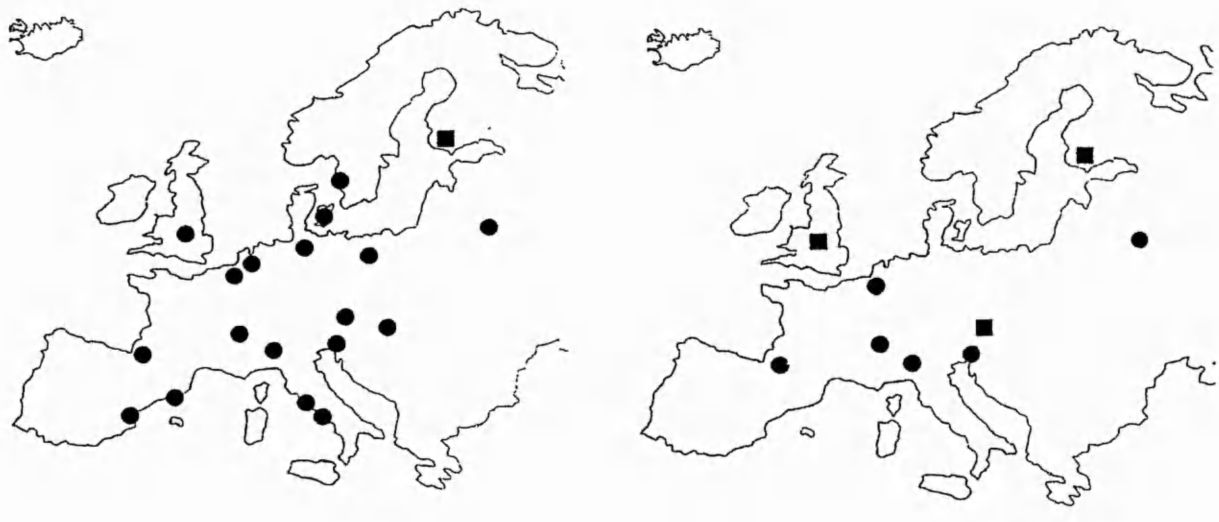


Figure 2. The presence (●) and absence (■) of visible injury on (a) crop and (b) natural species grown at ICP-Crops experimental sites in 1995. The range of crop and natural species on which injury was detected includes *T. subterraneum*, *T. repens*, *T. alexandrinum*, *P. vulgaris*, *G. max*, *L. esculentum*, *C. lanatus*, *M. sylvestris* and *C. arvense*. (Note: injury was not detected on all species at all sites).

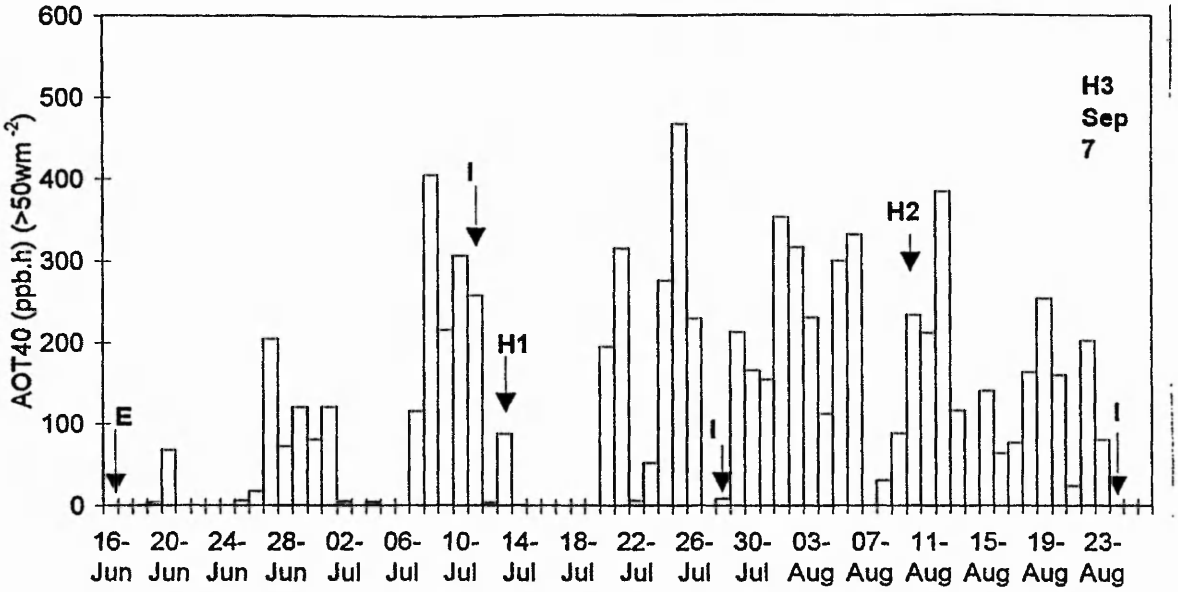


Figure 3. Daily AOT40 values (global radiation ≥ 50 Wm^{-2}) and the onset of ozone injury on *T. subterraneum* at an ICP-Crops experimental site in Belgium (E - emergence; I - injury; H1 - first harvest; H2 - second harvest; H3 - third harvest).

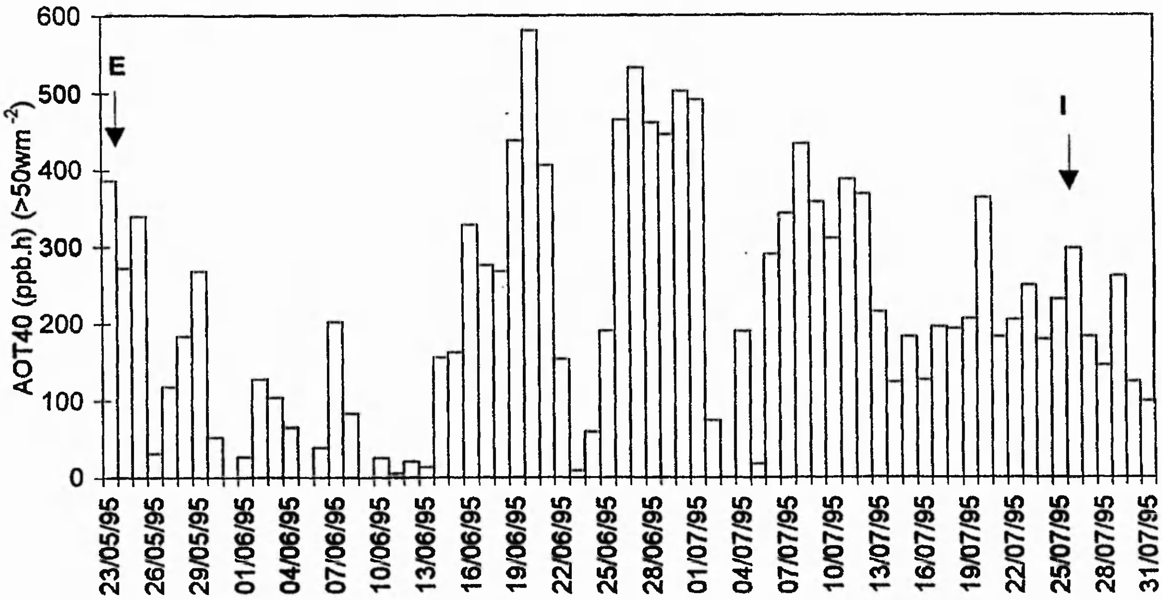


Figure 4. Daily AOT40 values (global radiation ≥ 50 Wm^{-2}) and the onset of ozone injury on *M. sylvestris* at an ICP-Crops experimental site in Switzerland (E - emergence; I - injury)

Calculating the sum of the daily AOT40s for the 3 and 5 days before the onset of injury clearly shows that injury occurred at approximately 75% and 50% of experimental sites respectively before the short-term critical level was exceeded (Figure 5). For example, at sites in Austria, Belgium, France, Germany, Italy, The Netherlands, Sweden, Switzerland and the United Kingdom. This shows that injury does not occur before the critical level is exceeded in one area of Europe.

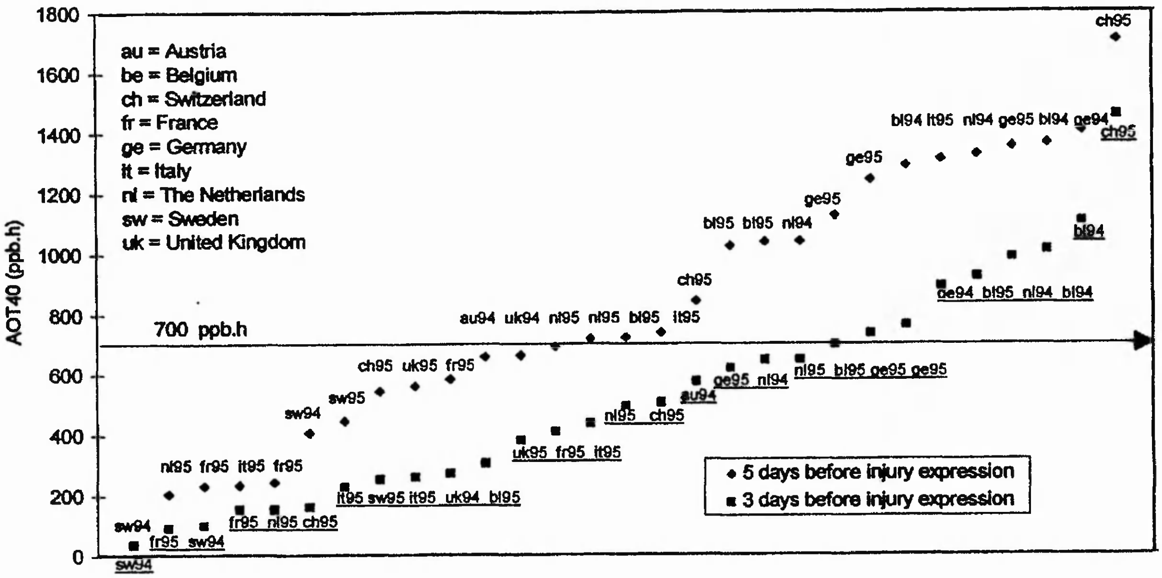


Figure 5. The sum of AOT40s (global radiation ≥ 50 Wm^{-2}) for the 3 and 5 days preceding the onset of ozone injury on *Trifolium* species at ICP-Crops experimental sites in 1994 and 1995

33 Injury development, ozone concentrations and climatic conditions

Climate data (temperature, relative humidity, global radiation and VPD) and ozone concentrations were analysed for the 3 and 5 days preceding ozone injury. Figure 6 and 7 show that data for the 5 days preceding injury fall into two groups, this being more apparent with 5 day than 3 day data (data not shown). These groups of data suggest that injury occurs if (i) the total AOT40 for the 5 days is < 500 ppb.h when the mean relative humidity (0901 - 1659 h) is > 50% and (ii) if the AOT40 is > 500 ppb.h when relative humidity is < 50%. There was not a clear trend between AOT40 and either global radiation (Figure 6) or temperature (Figure 7) and injury development. Two sets of conditions did not cause injury and these were (i) relative humidity > 50% and AOT40 > 900 ppb.h and (ii) relative humidity < 50% and AOT40 < 500 ppb.h and this correlates with the data shown in Figure 8. This figure shows that high relative humidity and high ozone concentrations were not recorded at any of the experimental sites in 1995 but these conditions are unlikely to occur in ambient situations. Figure 8 also shows quite clearly that injury does not occur if AOT40 and relative humidity are less than 500 ppb.h and 50% respectively and at relative humidity > 50% if the AOT40 is lower than 200 ppb.h.

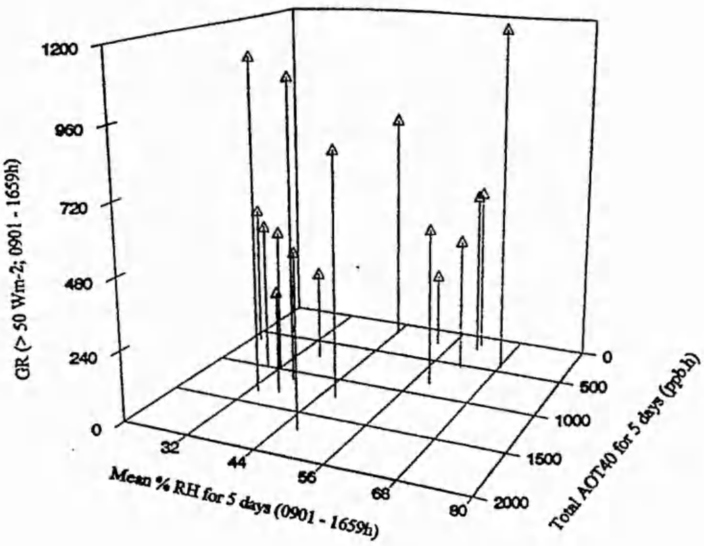


Figure 6. AOT40 (Global radiation ≥ 50 Wm^{-2}), %relative humidity and global radiation (0901 - 1659h) during 5 days before injury on *Trifolium* species

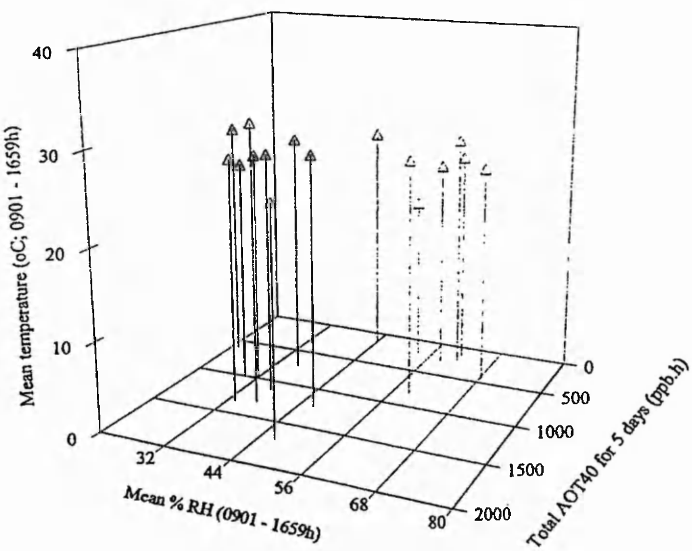


Figure 7. AOT40 (Global radiation ≥ 50 Wm^{-2}), %relative humidity and temperature (0901 - 1659h) during 5 days before injury on *Trifolium* species

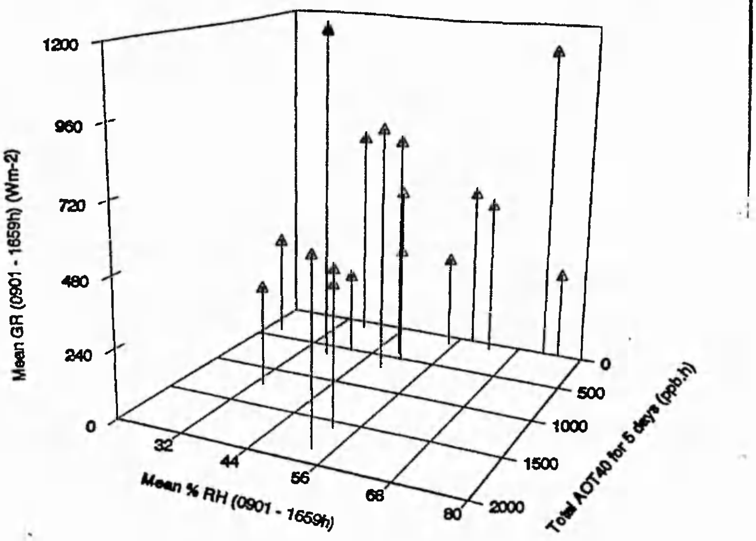


Figure 8. AOT40 (global radiation ≥ 50 Wm^{-2}), mean % relative humidity and global radiation (0901 - 1659 h) for the 5 day period which did not cause injury

Temperature and relative humidity data was used to calculate mean VPD values for the 5 days preceding/not preceding injury and these data were plotted against the total AOT40 for the 5 days (Figure 9). If the VPD is greater than 1.5 kPa, injury does not appear to occur unless the AOT40 for the 5 days is greater than 500 ppb.h. However, if the VPD is less than 1.5 kPa, visible injury can occur below an AOT40 of 500 ppb.h. It is apparent that injury is absent when the 5 day AOT40 is below 200 ppb.h even if the VPD is less than 1.5 kPa. At some sites, injury did not occur even if ozone concentrations and VPD were conducive for injury development but this may have been due to the influence of other climatic factors.

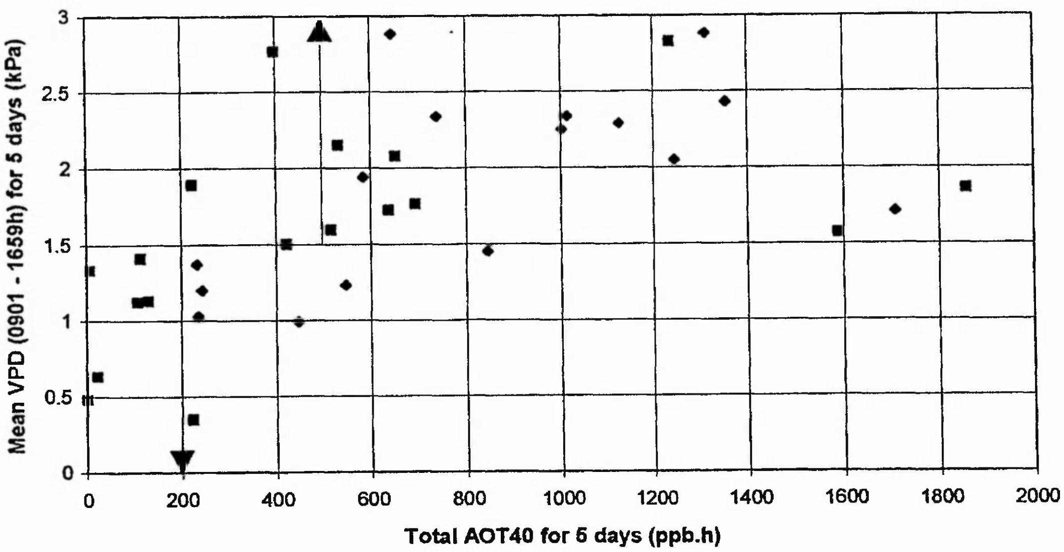


Figure 9. Sum of the ozone concentrations (AOT40) (global radiation ≥ 50 Wm^{-2}) and mean VPD (0901 - 1659 h) during the 5 days preceding the presence or absence of injury on *Trifolium* species (injury (◆) and no injury (■)).

34 Analysis by artificial neural networks

Ozone concentrations for each of the 5 days preceding injury were entered into the network and the following equation extracted from Network 1 which represents the most important hidden units:

$$F\,[2*(AOT40_4 + 7\text{ hr mean}_4 + \max_4) - 3*AOT40_2 - 2*AOT40_3 - 2*AOT40_5] + ..$$

$$F[(2*AOT40)_2 + (3*7\text{hr Oz})_2 + (2*Max)_2 - (AOT40_3 + 7\text{ hr mean}_3 + \max_3) - (3*AOT40_5) +$$

$$(2*7\text{ hr Oz}_5) + MaxOz_5)]$$

Subscript values represent day number

As in 1994, this equation suggests that a day with low ozone concentration coincides with the observation of injury. This is denoted by the negative sign associated with the AOT40 on Day 5. The equation also indicates that a peak ozone concentration in one (or two) of the five days may be responsible for causing injury as ozone concentration is associated with a positive sign on day 4 and day 2 but a negative sign on the remaining days. If this is the case, using the 5 days before the observation of injury should encompass the ozone episode which is responsible for causing damage. Extracting the weightings of the interconnections from the networks trained with ozone, climatic conditions and day number shows which parameters are important for the onset of injury. Network 2 trained with ozone, global radiation, VPD and day number indicates that ozone (AOT40, 7 hr mean and mean daily maximum), global radiation (mean daily maximum) and VPD (7 hr mean) are important inputs for injury development (Table 1) as the weightings have values of 1.92,1.69,1.17,1.43 and 1.12 respectively. Network 3 also gave importance to these parameters but a lower importance to relative humidity and temperature (Table 1). Both networks indicated that VPD (7 hr mean) was more important than a 24 hr mean and the mean daily maximum.

4 Discussion

Results from the ICP-Crops ambient air experimental programme clearly show that visible injury on crop and natural species is widespread throughout Europe. This indicates that the short-term critical level is being widely exceeded. However, at many sites, injury was observed before the AOT40 exceeded 700 ppb.h (Figure 5) which implies that the short-term critical level requires modification. Three dimensional graphs of AOT40, % relative humidity and either global radiation or temperature suggest that relative humidity is an important climatic factor influencing the onset of ozone injury. These effects were more apparent with the 5 day than the 3 day data. Two sets of conditions were conducive for injury; high ozone concentrations/low relative humidity and low ozone concentrations/high humidity (Figure 6 and 7). Furthermore, the data suggests that ozone injury will not occur at both low relative humidity and low ozone concentrations. There did not appear to be a clear trend between AOT40 and either global radiation (Figure 6) or temperature (Figure 7) and the onset of injury. More importantly, it was clear that VPD had a influence on the onset of injury. If the mean VPD exceeds 1.5 kPa then the AOT40 over 5 days needs to be above 500 ppb.h before injury is observed but if it is less than 1.5 kPa, injury can occur below a 5 day AOT40 of 500 ppb.h (Figure 9). It was also clear that injury did not occur when the 5 day AOT40 was less than 200 ppb.h. ANN analysis of ICP-Crops data indicated that ozone concentration, VPD and global radiation are important factors for determining the onset of ozone injury.

Table 1. The relative importance of input parameters of the neural network trained with ozone, climatic parameters and day number

	Network 2	Network 3
Parameter	**Relative importance (arbitrary units)**	
Ozone (AOT40)	1.92	2.58
Ozone (7 hr mean)	1.69	1.86
Ozone (mean daily max)	1.17	1.44
RH (24 hr mean)		0.75
RH (7 hr mean)		0.68
RH (mean daily max)		1.21
Global radiation (24 hr mean)	0.88	1.23
Global radiation (7 hr mean)	1.03	1.44
Global radiation (mean daily max)	1.43	1.87
Day number	0.62	0.80
Temperature (24 hr mean)		1.11
Temperature (7 hr mean)		1.12
Temperature (mean daily max)		1.29
VPD (24 hr mean)	0.75	0.65
VPD (7 hr mean)	1.12	1.36
VPD (mean daily maximum)	0.94	1.22

The absence of a clear relationship between ozone concentration and injury development may be explained by plant and climate interactions, or alternatively, by atmospheric conductivity influencing the ozone dose absorbed by plants grown in ambient air (Grünhage and Jager, 1994). It is well established that stomata respond to changes in ambient humidity (Aphalo and Jarvis, 1991; Kearns and Assmann, 1993, Gutierrez *et al*, 1994). Stomata close in response to a decrease in ambient humidity (Kearns and Assmann, 1993) or to an increase in VPD which causes a decrease in stomatal conductance (Grantz and Meinzer, 1990). A decrease in stomatal conductance will reduce ozone flux into the leaf and the occurrence of injury. The results obtained in this study can be partly explained by the influence of ambient humidity or VPD on stomatal conductance. If the VPD is above 1.5 kPa and AOT40 is less than 500 ppb.h, injury may not occur because of a reduction in stomatal conductance and so sufficient ozone does not enter the plant to cause injury. However, if the AOT40 is less than 500 ppb.h and the VPD is less than 1.5 kPa, stomatal conductance may rise and so increasing the flux of ozone into the plant which causes injury despite ambient ozone concentrations being low. When VPD is above 1.5 kPa and AOT40 is above 500 ppb.h, the high ambient ozone concentrations could enable adequate ozone to enter the leaf to cause injury even though stomatal conductance may be reduced. Alternatively, stomatal conductance may increase in response to an increase in global radiation. Studies with sugarcane by Grantz and Meinzer (1990) have shown that stomatal closing, in response to a increase in VPD, is counteracted by opening responses to light. For example, as VPD increases, stomatal conductance decreases but increases again in response to a rise in light intensities. Thus, it is hypothesised that, at high ozone concentrations, stomata conductance may increase in response to an increase in global radiation (usually associated with high ozone levels) allowing sufficient ozone to enter the leaf and cause injury. However, the influence of global radiation on the onset of visible injury is not clear from the data presented here but this hypothesis is substantiated by ANN analysis of ICP-Crops data which indicated that VPD and global radiation were both important parameters for injury development. This was also found by ANN analysis of data obtained from closed chamber experiments with *T. subterraneum* (Balls *et al*, 1996). Analysis of the weightings from the model created with this data showed that VPD and PAR had a strong influence on plant response to ozone.

The influence of relative humidity on ozone phytotoxicity has previously been described. For example, Mortensen (1992) described how the height of *Lycopersicon esculentum* plants was significantly reduced as the relative humidity increased from 70% to 90%. Balls *et al* (1996) reported how the injury observed on *T. subterraneum* increased as VPD deceased and attributed this to an influence on stomatal conductance and Gimeno *et al* (1995a) described how ozone injury on tobacco cultivars in Spain was more prevalent in coastal areas which experience higher relative humidity.

Analysis of ICP-Crops data by ANNs has provided further useful information in addition to indicating the importance of climatic factors associated with injury development. It was implied that day number, which can be related to leaf and plant age, influenced the onset of injury. This is in agreement with the results of other studies by Pihl Karlsson *et al* (1995) and Gimeno *et al* (1995b) and has been described by Benton *et al* (1995). It is generally accepted that older, mature leaves/plants are more prone to ozone injury than young, immature plants. This is substantiated by ICP-Crops experiments with *M. sylvestris*. Injury did not occur on this species until the plants were approximately 2 months old despite high ozone episodes.

The network trained with just ozone data predicted that the timing and magnitude of ozone concentrations may be important for injury development as a day with a lower ozone concentration seemed to coincide with the expression of ozone injury. This implication could be of importance because it suggests that injury will not occur unless ozone concentrations decrease or that lower ozone concentrations may accelerate the onset of injury. ANN analysis of data obtained using *T. subterraneum* and closed chambers gave predictions that plant sensitivity decreased as the time of ozone exposure increased which was consistent with experimental data (Balls *et al*, 1995). Further work is being conducted at the ICP-Crops Coordination Centre to verify the importance of a decrease in ozone concentration before injury development.

It is possible that variability in the height at which ozone concentrations are measured may account for some of the low AOT40 values causing injury. Ozone concentration at the sites in Figure 4 were measured from 0.1 to 4 m. Ozone concentrations measured at 0.1 m would probably be lower than if measured at 4 m as it has been shown that ozone concentrations increase as measurement height above the crop increases (Pleijel *et al*, 1995). However, injury occurred before the AOT40 exceeded 700 ppb.h even at sites where the ozone concentrations were measured at 3 and 4 m.

From the data presented here, it is clear that climatic factors, plant age and the length and magnitude of ozone episodes may all influence the short-term critical level of ozone for visible injury. At present, insufficient data is available to integrate plant age and the magnitude of ozone episodes into a revised short-term critical level. Thus, the recommendations for a new critical level are based on the influence of climatic factors.

5 Recommendations

Analysis of 1994 and 1995 ICP-Crops data from ambient air experiments has provided important information regarding the short-term critical level for visible injury. It is quite apparent that an AOT40 of 700 ppb.h over three consecutive days is not suitable to use as injury occurs before this has been exceeded at many sites. From the information presented here, two new short-term critical levels could be established which consider the influence of VPD on ozone flux into the plant. These are:

(i) an AOT40 of 500 ppb.h over five days (global radiation ≥ 50 Wm^{-2}) when mean VPD (0901 - 1659 h) exceeds 1.5 kPa

(ii) an AOT40 of 200 ppb.h over five days (global radiation ≥ 50 Wm^{-2}) when mean VPD (0901 - 1659 h) is below 1.5 kPa

It is anticipated that these suggestions will be substantiated by analysis of data obtained from the 1996 experimental season of the ICP-Crops.

6 Acknowledgements

We would like to acknowledge all the participants of the ICP-Crops and The Department of the Environment (UK) for funding the coordination of the ICP-Crops experiments (contract number EPG/1/3/13).

7 References

Aphalo, P.J & Jarvis P.G (1991) Do stomata respond to relative humidity? Plant, Cell and Environment. 14, 127 -132.

Balls G.R., Palmer-Brown D., Cobb A.H & Sanders G.E. (1995) Towards unravelling the complex interactions between microclimate, ozone dose and ozone injury in clover. Journal of Water, Air and Soil Pollution. 85, 1467 - 1472.

Balls G.R., Palmer-Brown D & Sanders G.E (1996) Investigating microclimatic influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks. New Phytologist, 132, 271-280.

Benton J., Fuhrer J., Gimeno B.S., Skärby L., Balls G., Palmer-Brown D., Roadknight C & Sanders G. (1995) ICP-Crops and critical levels of ozone for injury development. Exceedances of Critical Loads and Levels (Vienna, November 22 - 24th)

Carnahan J.E., Jenner E.L & Wat E.K.W (1978) Prevention of ozone injury in plants by a new protectant chemical. Phytopathology, 68, 1225-1229.

Fuhrer J & Achermann B (eds) Critical Levels for Ozone; a UN/ECE Workshop Report. Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH - 3097 Liebefeld-Bern, Switzerland. No. 16, 1994.

Gimeno B.S., Peñuelas J., Porcuna J.L & Reinert R.A (1995a) Biomonitoring ozone phytotoxicity in Eastern Spain. Journal of Water, Air and Soil Pollution. 85, 1521 - 1526.

Gimeno BS., Bermejo V., Mendoza M & Sánchez S. Ozone effects in agricultural crops at the Ebro Delta in 1995. (1995b) Poster presentation.at the 'Exceedances of Critical Levels and Loads' Workshop, Vienna, November 22 - 24.

Grantz D.A & Meinzer F.C (1990) Stomatal response to humidity in a sugarcane field: simultaneous porometric and micrometerorological measurements. Plant, Cell and Environment. 13, 27 - 37.

Grünhage L & Jäger H.-J (1994) Influence of the atmospheric conductivity on the ozone exposure of plants under ambient conditions: considerations for establishing ozone standards to protect vegetation. Environmental Pollution 85, 125-129.

Gutierrez M.V., Meinzer F.C & Grantz D.A (1994) Regulation of transpiration in coffee hedgerows: covariation of environmental variables and apparent responses of stomata to wind and humidity. Plant, Cell and Environment. 17, 1305 - 1313.

Kearns E.V & Assmann S.M (1993) The guard cell-environment connection. Plant Physiology 102, 711 - 715

Kothari S.C & Heekuck O.H (1993) Neural Networks for Pattern Recognition. Advances in Computers, 37, 119 - 166.

Mortensen L.M (1992) Effects of ozone concentration on growth of tomato at various light, air humidity and carbon dioxide levels. *Scientia Horticulturae*, 49, 17 - 24.

Pihl Karlsson G.,Selldén G., Skärby L & Pleijel H (1995) Clover as an indicator plant for phytotoxic ozone concentrations: visible injury in relation to species, leaf age and exposure dynamics. New Phytologist, 129, 355-365.

Pleijel H., Wallin G., Karlsson., Skärby & Sellden G (1995) Gradients of ozone at a forest site and over a field crop - consequences for the AOT40 concept of critical level. Journal of Water, Air and Soil Pollution. 85, 2033 - 2038.

Roadknight C.M., Palmer Brown D & Sanders G.E. Learning the equations of data. Proceedings of the 3rd Annual SNN Symposium on Neural Networks (eds Kappen B and Gielen S) Springer-Verlag. 253. 1995.

Sanders G.E., Skärby L., Ashmore M.R & Fuhrer J. (1995) Establishing critical levels for the effects of air pollution on vegetation. Journal of Water, Air and Soil Pollution. 85, 189 - 200.

UN/ECE (1994) The ICP-Crops Experimental Protocol. The ICP-Crops Coordination Centre. The Nottingham Trent University, Nottingham, United Kingdom.

UN/ECE (1995) The ICP-Crops Experimental Protocol. The ICP-Crops Coordination Centre. The Nottingham Trent University, Nottingham, United Kingdom.

[redacted]

ELSEVIER

Agriculture, Ecosystems and Environment 1508 (1999) 1–12

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

An international cooperative programme indicates the widespread occurrence of ozone injury on crops

J. Benton [a], J. Fuhrer [b], B.S. Gimeno [c], L. Skärby [d], D. Palmer-Brown [e], G. Ball [e], C. Roadknight [e], G. Mills [f,*]

[a] Department of Life Sciences, The Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, UK
[b] Institute for Environmental Protection and Agriculture (IUL) Liebefeld, CH-3003 Bern, Switzerland
[c] CIEMAT-DIAE 3B, Avda. Complutense 22, Madrid 28040, Spain
[d] IVL, Box 47086, S-40258 Göteborg, Sweden
[e] Department of Computing, The Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK
[f] Institute of Terrestrial Ecology-Bangor Research Unit, Deiniol Road, Bangor, Gwynedd, LL57 2UP, UK

Received 17 November 1998; received in revised form 5 July 1999; accepted 30 July 1999

Abstract

The UN/ECE ICP-Vegetation[1] routinely investigates the effects of ambient ozone pollution on crops throughout Europe. Each year, a series of co-ordinated ambient air experiments are conducted over a large area of Europe and a range of crop species are observed for the occurrence of injury following ozone episodes. In 1995 and 1996, ozone injury was observed at sites throughout Europe from United Kingdom (Nottingham) to the Russian Federation (Moscow) and from Sweden (Östad) to Italy (Naples). The only site participating in the ICP-Crops where it was not observed was that at Finland (Jokioinen). Injury was identified on subterranean and white clover, French bean, soybean, tomato, and watermelon at one or more sites. Injury was also detected in gardens and on crops growing in commercial fields. Two short-term critical levels which incorporate ozone dose and air saturation vapour pressure deficit (VPD) were derived from the 1995 data. These were (i) an AOT40[2] of 200 ppb.h over 5 days when mean VPD (0930–1630 h) is below 1.5 kPa and (ii) an AOT40 of 500 ppb.h over 5 days when mean VPD (0930–1630 h) is above 1.5 kPa. In general, the 1996 data supported these critical levels although injury did occur on two occasions when the AOT40 was less than 50 ppb.h, and the VPD was less than 0.6 kPa. Thus, ICP-Vegetation experiments have shown that ozone injury can occur over much of Europe and that plants are most at risk in conditions of high atmospheric humidity.

Keywords: Critical levels; Crops; Injury; Ozone; Europe

1. Introduction

Over recent decades, photochemical ozone pollution has been identified as a potential threat to crops in rural areas (Photochemical Oxidants Review Group, PORG, 1997). This secondary pollutant forms from a chemical reaction involving oxides of nitrogen and volatile organic compounds (mainly from vehicle ex-

* Corresponding author. Tel.: +44-1248-370045; fax: +44-1248-355365
E-mail address: g.mills@ite.ac.uk (G. Mills)

[1] United Nations Economic Commission for Europe International Cooperative Programme on effects of air pollution and other stresses on crops and non-wood plants.

[2] The sum of the difference between the hourly concentration (ppb) and 40 ppb (when the concentration exceeds 40 ppb) for the hours when global radiation (GR) exceeds 50 Wm^{-2}.

0167-8809/99/$ – see front matter
PII: S0167-8809(99)00107-3

haust fumes and industrial processes) in the presence of sunlight. Annual average concentrations of ozone have doubled during the last 100 years (Voltz and Kley, 1988). Ozone episodes, where the concentration exceeds 60 ppb (parts per billion) by volume commonly occur during the spring and summer when crops are actively growing (PORG, 1997). The main entry route into plants is via the stomata; ozone subsequently reacts with cell wall and membrane components (Kangasjärvi et al., 1994) resulting in the formation of reduced oxygen species such as hydroxyl and superoxide radicals and hydrogen peroxide which are highly reactive with biological molecules. Consequently, membrane integrity is disturbed, thus modifying cell permeability (Heath, 1987) and osmotic pressure, membrane potentials and the activity of membrane-bound enzymes such as ATPases are affected (Dominy and Heath, 1985). Membrane disruption, leading to cell death, causes chlorotic flecking, necrosis and bronzing on foliage. This has important implications for crops grown for their foliage as symptoms of visible injury may affect market value. In addition to causing visible injury, ozone exposure can modify physiological processes such as carbon dioxide assimilation which influences plant productivity. In 1983, a European Open-Top Chamber programme was initiated to establish if ambient concentrations of atmospheric pollutants in Europe were sufficient to cause loss of crop yield or induce visible injury (CEC, 1993). The majority of studies using *Triticum aestivum* (spring wheat) showed that yield was reduced with increasing levels of ozone indicating that ambient ozone concentrations in Europe were sufficient to reduce the grain yield of this crop (Skärby et al., 1993).

As with other major air pollutants, such as sulphur dioxide, the precursors of ozone can be transported over long distances and thus present a transboundary air pollution problem. In 1983, the United Nations/Economic Commission for Europe Convention on Long-Range Transboundary Air Pollution (UN/ECE LRTAP Convention) came into force to address this problem and an effects-orientated approach was adopted for establishing successful pollutant abatement policy (UN/ECE, 1996a). This included establishing critical levels for sulphur, reduced/oxidised nitrogen and ozone i.e. the concentrations of pollutants in the atmosphere above which direct adverse effects on receptors, such as plants, ecosystems or materials may occur (Bull, 1991). Five International Cooperative Programmes (ICPs) and a Mapping Programme operate under the Convention's Working Group on Effects to assess and monitor air pollution effects.

The UN/ECE ICP-Vegetation (International Cooperative Programme on effects of air pollution and other stresses on crops and non-wood plants) came into force in 1988 and involves research in 15 European countries. The programme aims to establish the crop and non-wood plant species at risk from ozone pollution by determining those species which show symptoms of visible injury and/or a reduction in biomass or yield following exposure to ambient ozone. This information, in conjunction with known ambient air ozone concentrations and climatic conditions, can be used to establish critical levels for ozone and to identify areas of Europe at risk from ozone pollution.

Two short-term critical levels of ozone for visible injury were set at the UN/ECE Workshop on Critical Levels for Ozone in Europe-Testing and Finalising the Concepts (Kuopio, Finland, 1996) following analysis of the 1995 ICP-Vegetation experimental data (Benton et al., 1996). These were an AOT40 of 200 ppb.h accumulated over 5 days when mean VPD (0930–1630 h) was less than 1.5 kPa, and an AOT40 of 500 ppb.h when mean VPD exceeded 1.5 kPa. The ozone dose parameter, AOT40, is calculated as the sum of the difference between the hourly concentration (ppb) and 40 ppb (when the concentration exceeds 40 ppb) for the hours when global radiation (GR) exceeds 50 Wm^{-2}. AOT40 is used because it enables data from different experiments to be combined, and it allows for a linear relationship between ozone exposure and yield and it considers the effects of cumulative exposures to concentrations above a cut-off of 40 ppb (Fuhrer and Achermann, 1994). Concentrations below this cut-off value are strongly influenced by natural background processes in the northern hemisphere (UN/ECE, 1996a).

Vapour pressure deficit was included in the definition for the short-term critical levels because it is one of the main factors influencing stomatal conductance and thus the flux of ozone into the plant. For example, when VPD is high, conductance is reduced (Grantz and Meinzer, 1990) and the flux of ozone into the leaf is restricted. In these conditions, higher ambient ozone concentrations (or a higher AOT40) are required before an adequate dose is absorbed to cause injury. However, when VPD is lower, stomatal conductance increases which boosts the flux of ozone into the leaf. Thus, it is possible that the absorbed dose is sufficient to cause injury even though ambient ozone concentrations are low. Indeed, it has been suggested that higher yield reductions or injury levels occur when ozone concentrations are moderate (Grünhage and Jäger, 1994; Krupa et al., 1995) because high ozone concentrations often coincide with conditions which are not conducive to ozone uptake, for example, high VPD. Gimeno et al. (1995) described how ozone injury on tobacco cultivars in Spain was more prevalent in coastal areas but decreased at sites further inland. It was suggested that the high relative humidity at the coastal sites could favour ozone phytotoxicity. Similarly, Balls et al. (1996) reported how the level of injury observed on subterranean clover increased as VPD decreased and attributed this to an influence on stomatal conductance.

This paper describes the ozone concentrations in 1995 and 1996 across the ICP-Crops experimental sites and the geographical extent of visible injury on selected species grown throughout Europe. The crops were chosen because of their commercial importance in Europe, and because previous studies with white clover (Horsman et al., 1981; Becker et al., 1989), subterranean clover (Horsman et al., 1981), bean (Tonneijck, 1983; Guzy and Heath, 1993), watermelon (Fernandez-Bayon et al., 1993), tomato (Lorenzini et al., 1984) and soybean (Salleras et al., 1989) had shown that these species were ozone-sensitive. The data from the 1996 experiments have been used to validate the short-term critical levels for visible injury that were set at the Kuopio Workshop. ICP-Vegetation experiments that monitored the effects of ozone on biomass are reported by Ball et al. (1998).

2. Materials and methods

2.1. Experimental sites and crop species

In 1995 and 1996, experiments designed to investigate the onset of ozone-induced visible injury were conducted at sites in 15 countries across Europe (UN/ECE, 1995; UN/ECE, 1996b). Table 1 gives details of the sites involved and species grown are shown in. The most northerly and southerly sites were Finland-Jokioinen (60°47′N) and Italy-Naples (40°48′N), respectively. Participation in the programme is on a voluntary basis. Hence there has been no effort to influence the location of sites, nor their representation of ozone conditions in Europe. Nevertheless, the spread of the sites across Europe has allowed for plant responses to a wide range of ozone and physical climatic conditions to be determined.

At each experimental site, one or more of the following crop species were grown: French bean (*Phaseolus vulgaris* L. cv. 'Lit' and 'Groffy'); subterranean and white clover (*Trifolium subterraneum* L. cv. 'Geraldton' and *T. repens* L. cv. 'Menna'); tomato (*Lycopersicon esculentum* Miller. cv. 'Tiny Tim'); soybean (*Glycine max* (L.) Merr. cv. 'Ceresia') and watermelon (*Citrullus lanatus* (Thunb.) Matsum. and Nakai cv. 'Sugar baby'). The seeds were purchased from commercial suppliers in the UK (French bean, subterranean and white clover), Spain (tomato and watermelon), and Austria (soybean), and then distributed to all participants in the programme.

2.2. Cultivation

2.2.1. Subterranean and white clover

Twenty 15 cm diameter plastic plant pots were filled with an appropriate soil mixture for the site (typically a mixture of peat, sand and soil). Two 20 cm glass-fibre wicks (Vitrulan Textilglas GmBH, Bayern, Germany) were inserted vertically into the growing medium of each pot. One end of each wick protruded through a hole in the pot base and the other end was 2 cm below the soil surface. Nine clover seeds were sown in each pot to a depth of 0.5 cm with the soil depth being 1 cm below the pot rim. Seedlings were raised in a glasshouse, thinned to three per pot when the first leaf had emerged (approximately 14 day after sowing) and placed outside in ambient conditions once the first trifoliate leaf had unfolded. Pots were placed 25 cm apart in trenches (to reduce the influence of temperature) with the pot rim being 5 cm above ground level. The pots were positioned on a second pot containing water in which the protruding wicks were placed. This formed a self-watering system. The experiment was repeated twice at each site in each year. Typically, exposure of the first set of plants started in mid-June and

Table 1
The location of ICP-Crops experimental sites and crop species grown at each site in 1995 and/or 1996

Country-site	Altitude (m.a.s.l)[a]	Coordinates	Crop species grown
Austria (Seibersdorf)	190	47°58′N, 16°30′E	Subterranean clover, white clover, bean, soybean
Belgium (Tervuren)	80	50°49′N, 04°31′E	Subterranean clover, white clover, bean soybean
Finland (Jokioinen)	100	60°47′N, 23°28′E	White clover
France (Pau)	120	43°18′N, 00°22′W	White clover, bean
Germany (Braunschweig)	85	52°15′N, 10°30′E	White clover, bean
Germany (Essen)	60	51°22′N, 06°55′E	White clover
Germany (Trier)	129	49°46′N, 06°39′E	White clover
Germany (Giessen)	190	50°32′N, 08°41′E	White clover
Hungary (Keszthely)	n.a.[b]	46°47′N, 17°16′E	Bean
Italy (Milan)	120	45°28′N, 09°12′E	White clover, soybean
Italy (Rome)	n.a.	41°53′N, 12°03′E	White clover, bean
Italy (Naples)	20	40°48′N, 14°20′E	Bean
Netherlands (Westmaas)	−0.5	51°47′N, 04°27′E	Subterranean clover and bean
Netherlands (Schipluiden)	−2.0	51°59′N, 04°16′E	Sbterranean clover and bean
Netherlands (Zegveld)	−1.5	52°08′N, 04°50′E	Subterranean clover and bean
Netherlands (Wageningen)	7.0	51°58′N, 05°38′E	Subterranean clover and bean
Poland (Kornik)	n.a.	52°15′N, 17°06′E	White clover, bean
Russian Federation Moscow	n.a.	55°45′N, 37°42′E	Subterranean and white clover, bean
Slovenia (Ljubljana)	250	46°03′N, 14°30′E	White clover, bean
Slovenia (Kovk)	780	46°08′N, 15°06′E	White clover, bean
Slovenia (Zavodnje)	800	46°25′N, 15°01′E	White clover
Spain (Ebro Delta)	0	40°75′N, 00°45′E	Tomato, bean, watermelon
Spain (Pamplona)	n.a.	42°49′N, 01°39′W	White clover, bean
Spain (Begur)	190	41°55′N, 03°15′E	Bean
Spain (Veciana)	725	41°40′N, 01°30′E	Bean
Sweden (Östad)	60	57°54′N, 12°24′E	Subterranean and white clover
Switzerland (Cadenazzo)	200	46°10′N, 08°56′E	Subterranean, white clover, bean, soybean
UK (Nottingham)	47	52°53′N, 01°11′W	Subterranean and white clover, bean

[a]m.a.s.l.: Metres above sea level.
[b]n.a.: Not available.

the second set of plants a month later. At 28 and 56 day after exposure to ambient air, the dry weight of leaves/petioles and flowers from each plant was determined and at 84 day, the entire above-ground biomass (including stolons) for each pot was determined (data not presented).

2.2.2. *French bean cv. Lit and Groffy, and soybean*

Twenty (for bean experiment) and 10 (for soybean) 20 cm diameter plastic plant pots were filled with soil mixture and three 30 cm wicks inserted as described above. Four bean or soybean seeds were sown in each pot to a depth of 2 cm and raised in the glasshouse. The seedlings were thinned to one per pot and placed in ambient conditions when the first true leaf had unfolded. The pots were placed in holes and the plants watered by the self-watering system already described.

2.2.3. *Experiments at the Ebro Delta (North East Spain)*

French bean cv. Lit, tomato and watermelón were grown using a modified experimental protocol described by Gimeno et al. (1996).

2.3. *Assessment of ozone injury*

Photographs of injury on all species were distributed to participants (UN/ECE, 1994; Sanders and Benton, 1995) in order to confirm that the symptoms recorded were those of ozone injury.

2.3.1. *Subterranean and white clover*

Plants were observed daily and the date on which injury was observed for the first time before each harvest was recorded. Injury was always seen on the upper surface and appeared as chlorotic flecking on white clover but as necrotic lesions on subterranean clover.

The plants were also assessed at 28, 56 and 84 day after exposure to ambient conditions when the total number of leaves and the number of ozone injured leaves per pot were determined.

2.3.2. *French bean cv. Lit and Groffy, and soybean*

Both species were observed daily for the appearance of ozone injury. Typical symptoms included necrosis and bronzing on the upper leaf surface. Bean plants were also assessed when 50% of plants had flowers (flowering) and when the skin of the pods was flat (green harvest) for the number of ozone injured leaves and the total number of leaves per plant. Assessments made at the Ebro Delta site in Spain of French bean, tomato and watermelon are described by Gimeno et al. (1996).

2.3.3. *Surveys of gardens/commercial fields*

Nearby gardens and commercial fields of crops were surveyed for visible injury after injury had been detected on the experimental plants at the field sites at Belgium (Tervuren), France (Pau), Spain (Ebro Delta), and Switzerland (Cadenazzo).

2.3.4. *Measurement of ambient ozone concentrations and climatic conditions*

Ambient ozone concentrations and climatic conditions were measured at the sites throughout the experimental season. Ozone concentrations were measured at a height of 3 m by UV photometric methods; standard calibration procedures were used at each site. The concentrations presented in this paper are expressed as a mean daily maximum (ppb) and as AOT40. The AOT40 for the 5 days preceding the onset of visible injury on white clover was calculated for data obtained in both 1995 and 1996. The climatic parameters measured during the experimental season included mean daily temperature (°C), percent relative humidity (RH) and global radiation (GR; Wm^{-2}). The 7 h mean (0930–1630 h, GMT) air saturation vapour pressure deficit (VPD) for the 5 day period preceding injury was calculated from these data.

Table 2
Mean daily maximum ozone concentrations and AOT40 (for hours with GR $\geq 50\,Wm^{-2}$) at ICP-Crops experimental sites in Europe from 21st June to 2nd September 1995 inclusive. This represented the longest time period of ozone measurement common to all sites listed in this table

Country-site	Mean daily maximum (ppb)	Total AOT40 (ppb. h)
Austria (Seibersdorf)	64	12900
Belgium (Tervuren)	55	7988
Finland (Jokioinen)	35	290
France (Pau)	52	3893
Germany (Braunschweig)	57	8241
Italy (Milan)	68	11781
Netherlands (Westmaas)	57	7465
Slovenia (Zavodnje)	60	16800
Sweden (Östad)	41	1734
Switzerland (Cadenazzo)	72	15665
UK (Nottingham)	67	8842

3. Results

3.1. *Pollution climate and the occurrence of visible injury*

The range of ambient ozone concentrations experienced across Europe was demonstrated by the data recorded at experimental sites in 1995 (Table 2). For example, at Finland (Jokioinen) and Sweden (Östad) the mean daily maximum ozone concentrations were 35 and 41 ppb, respectively whereas at sites in central Europe, the mean daily maximum ozone concentrations were typically between 50 and 60 ppb for the same period. Concentrations were even higher in the southern European sites (Austria, Italy and Switzerland) where the mean daily maximum ozone concentrations were 64, 68 and 72 ppb, respectively. Variation in the mean daily maximum ozone concentration was reflected in the AOT40 values for each site with the total AOT40 for the period of June 21st to Sept 2nd was 290 ppb.h at Finland (Jokioinen), whereas at Switzerland (Cadenazzo) it was 15665 ppb.h. The cut-off value of 40 ppb was exceeded on only 15 days at the site in Finland but at sites in France and Switzerland, 40 ppb was exceeded on most days although the extent of exceedance was much greater in Switzerland than in France (Fig. 1). These data clearly indicate the variation in ozone pollution climate throughout Europe in 1995.

Injury occurred on clover species at all sites throughout Europe in 1995 and 1996 with the excep-

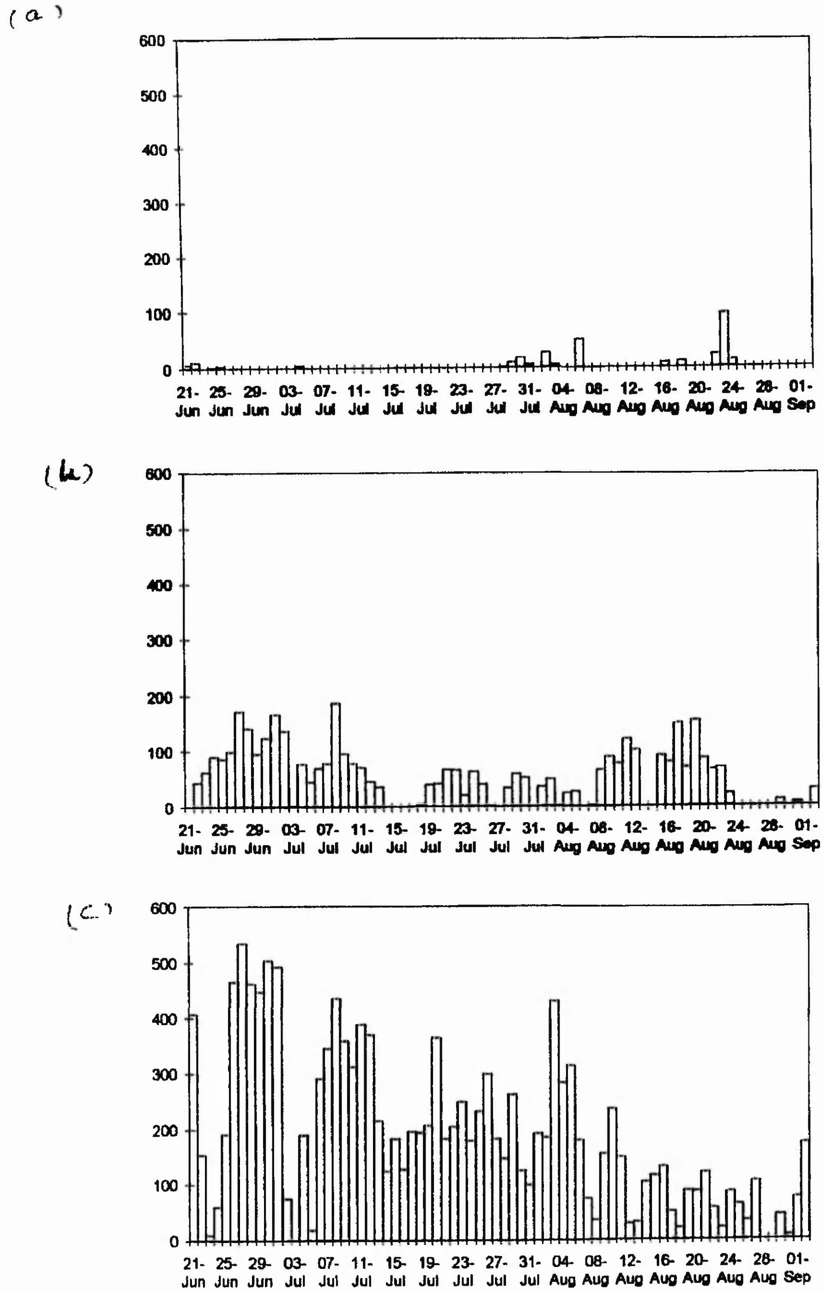


Fig. 1. Daily AOT40 values (ppb.h) for ozone (for hours with $GR \geq 50\,Wm^{-2}$) at ICP-Crops experimental sites from 21st June to 2nd September 1995 (inclusive) in (a) Finland (b) France and (c) Switzerland. This represented the longest time period of ozone measurement common to all sites listed in Table 2.

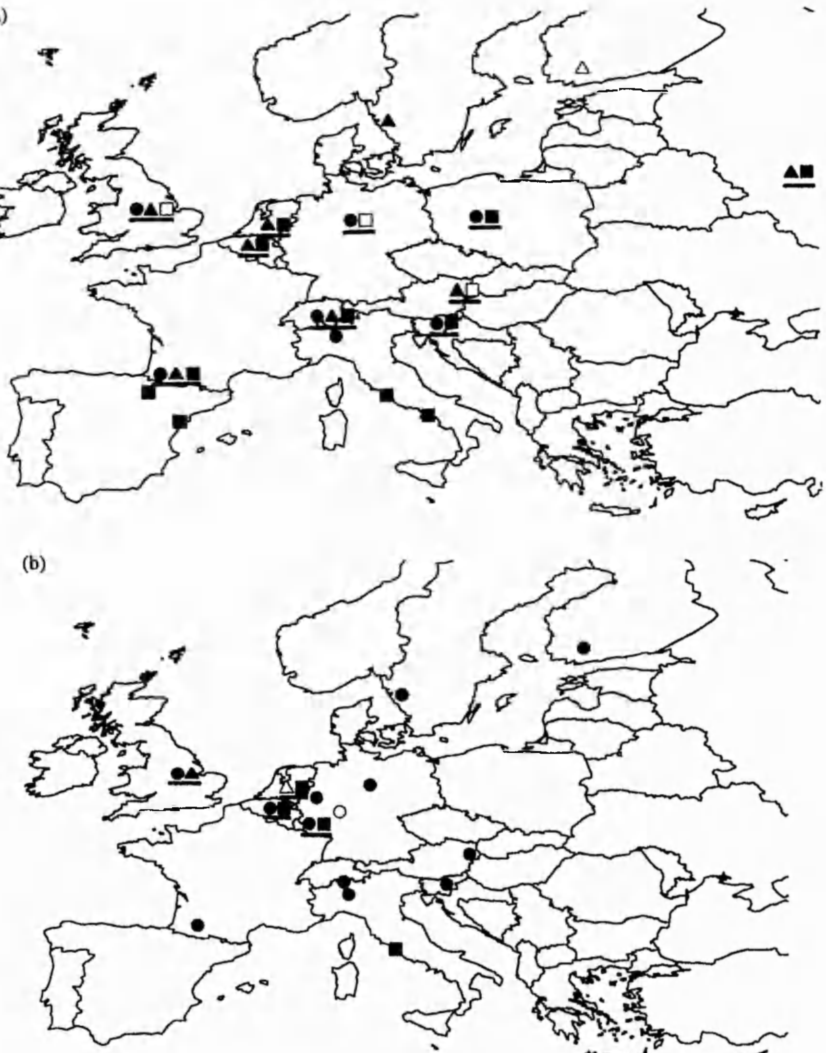


Fig. 2. Ozone injury at ICP-Crops experimental sites in (a) 1995 and (b) 1996. Key to symbols: (●) injury on white clover, (○) white clover grown, but injury not detected, (▲) injury on subterranean clover, (△) subterranean clover grown, but injury not detected, (■) injury on French bean cv. Lit, (□) French bean grown, but injury not detected. Where more than one symbol is present at a site, the symbols are underlined to avoid confusion with symbols for adjacent sites. The national boundaries shown on the maps do not reflect the official view of the United Nations or any subsidiary body.

tion of the site at Finland (Jokioinen) in 1995 (Fig. 2). Table 3 shows the extent of ozone injury to white and subterranean plants recorded at the first and second harvest in 1995. The highest levels of injury (>10% of injured leaves per pot) were seen at sites in Austria, Belgium, Germany, the Netherlands, Poland, Sweden and Switzerland, whereas lower injury levels (<10% of injured leaves per pot) were recorded at sites in France, Italy, The Russian Federation, Slovenia and the United Kingdom. The mean total leaf number per pot differed considerably among sites and could be attributed to climatic variation. For example, at the second harvest at Germany (Braunschweig) and Italy (Milan), the mean number of leaves per pot of white

Table 3
The percentage of ozone injured leaves at harvest 1 and 2 on white and subterranean clover grown throughout Europe in 1995 (values are a mean per pot ± SD where $n = 20$ and are from experiment 1). AOT40 calculated for hours with GR $\geq 50\,\mathrm{W\,m^{-2}}$

Country-site	Date harvest 1[a] Date harvest 2	Total number of leaves	Number of ozone injured leaves	% injured leaves	AOT40[b] (ppb.h)
White clover					
France (Pau)	3 July	61.6 ± 15	1.6 ± 1.8	2.6	n.a.[c]
	31 July	96.4 ± 30.5	9.0 ± 6.2	9.3	n.a.
Germany (Braunschweig)	21 July	106.1 ± 12.6	15.5 ± 10.2	15	2889
	16 August	226.7 ± 60.9	6.5 ± 9.2	2.9	3307
Italy (Milan)	5 July	52 ± 19.6	0.05 ± 0.2	0.1	4943
	7 August	47.5 ± 22.8	3.8 ± 5.1	7.9	6848
Poland (Kornik)	11 July	52.5 ± 12	5.5 ± 4.6	10.5	n.a.
Slovenia (Zavodnje)	10 July	100.8 ± 4.2	4.2 ± 3.8	4.2	n.a.
	2 August	384.9 ± 97.1	18.9 ± 11.2	4.9	n.a.
UK (Nottingham)	24 July	38.1 ± 8.1	0	0	1645
	21 August	45.4 ± 8.4	0.06	0.1	5764
Subterranean clover					
Austria (Seibersdorf)	5 July	90.3 ± 15.1	0.16 ± 0.5	0.18	3568
	31 July	287.3 ± 105.3	66 ± 32.9	23	6325
Belgium (Tervuren)	13 July	86.3 ± 12.1	12.4 ± 9.6	14.4	1997
	10 August	217.2 ± 43	63.5 ± 21.9	29.2	4062
Netherlands (Westmaas)	18 July	173.7 ± 34.1	40.4 ± 15.3	23.3	2662
	15 August	177.6 ± 35	27.6 ± 7.9	15.5	5690
Netherlands (Schipluiden)	15 August	152.5 ± 23.2	47.4 ± 14.1	31.1	4067
	12 September	302 ± 40.4	46.3 ± 6.7	15.3	1391
Netherlands (Wageningen)	18 July	168 ± 19.7	34.9 ± 12.8	20.8	3558
	15 August	185.1 ± 31.1	22.4 ± 10.8	12.1	4770
Netherlands (Zegveld)	18 July	190.3 ± 16.2	28.2 ± 14.3	14.8	2270
	15 August	211.2 ± 31.7	40.5 ± 12.3	19.2	3649
Russian Federation (Moscow)	11 August	13.9 ± 1	0.4 ± 0.7	2.5	n.a.
	8 September	37 ± 4	0.5 ± 0.7	1	n.a.
Sweden (Östad)	19 July	67.3 ± 14.2	1.3 ± 3.1	1.9	378
	16 August	146.9 ± 36.6	27.7 ± 13	18.8	954
Switzerland (Cadenazzo)	6 July	139 ± 21	62 ± 12	44.6	6564
	10 August	209 ± 83	66 ± 37	31.6	8335

[a] Harvest 1–28 day after exposure; harvest 2–56 day after exposure to ambient conditions.
[b] AOT40 for 28 day period.
[c] n.a.: Data not available.

clover was approximately 227 and 48, respectively. Injury was also observed on the leaves of French bean cv. Lit and Groffy (Table 4) at both the flowering and green harvest growth stages. However, injury to this species was not as widespread as injury to clover species (Fig. 2). For example, injury was not observed on bean at either Germany (Braunschweig) or the United Kingdom (Nottingham) in 1995. Highest levels of injury (>10% of leaves injured) were recorded at sites at Austria (Seibersdorf), Belgium (Tervuren), France (Pau), The Netherlands (Schipluiden), Poland (Kornik) and Slovenia (Kovk). Injury was also seen on soybean at Switzerland (Cadenazzo) and Austria (Seibersdorf), and on bean, tomato and watermelon at the Ebro Delta site in Spain (Gimeno et al., 1995). Ozone injury was observed during and after flowering on beans and when the flowers were already formed on watermelon. Injury was also recorded on plants grown in gardens and commercial fields in addition to those grown for experimental purposes (Table 5). For example in 1995 and 1996, injury was recorded on watermelon, soybean, French bean, potato, maize, wheat and butter bean.

The differences in ozone concentrations at each site were reflected to a certain extent in the occurrence and amount of visible injury. For example, injury did not occur on the subterranean clover grown for ICP-Vegetation experiments at Finland (Jokioinen)

Table 4
The number of ozone injured leaves on bean cv. Lit and cv. Groffy plants grown throughout Europe in 1995 (values are a mean per plant ± SD where $n = 20$)

Country-site	Growth stage	Total number of leaves	Number of ozone injured leaves	%
Austria (Seibersdorf)	FL[b]	16.9 ± 2.1	1.7 ± 0.7	10.1
	GH[c]	18.5 ± 4	3.7 ± 1.4	19.7
Belgium (Tervuren)	FL	22.8 ± 3.3	2.8 ± 0.7	12.1
	GH	33.8 ± 5	14 ± 4.1	41.5
France (Pau)	GH	19 ± 3	2.9 ± 1.3	15
Hungary (Keszthely[a])	FL	13.7	4.2	30.6
Italy (Naples[a])	FL	17.1 ± 3.6	1.8 ± 1.5	10.51
Italy (Rome)	FL	8.9 ± 1.6	0.2 ± 0.4	2.3
Netherlands (Westmass)	GH	28.2 ± 4.8	1.4 ± 2.3	4.6
Netherlands (Wageningen)	GH	30.6 ± 3.7	1.4 ± 1.8	4.6
Netherlands (Schipluiden)	GH	28.3 ± 2.3	5.2 ± 4.2	18.3
Netherlands (Zegveld)	GH	30.1 ± 2.4	2.1 ± 1.7	6.6
Poland (Kornik)	FL	13 ± 2	2.6 ± 0.9	20
Russian Federation (Moscow)	FL	4.4 ± 0.6	0.3 ± 0.4	5.8
Slovenia (Kovk)	FL	14.8 ± 4.1	4.8 ± 1.8	32.5
United Kingdom	FL	24.5 ± 5.1	0	0

[a]Groffy.
[b]FL: Flowering (when 50% of plants have flowers).
[c]GH: Green harvest (when pods are still flat).

Table 5
Crops which showed ozone injury in gardens/commercial fields in 1995 and 1996

Crop	1995	1996	Country
Bean (*Phaseolus vulgaris*)	√		Belgium (Tervuren)
Soybean (*Glycine max*)	√		France (Pau)
Watermelon (*Citrullus lanatus*)	√	√	Spain (Ebro Delta)
Potato (*Solanum tuberosum*)	√	√	Belgium (Tervuren)
	√	√	Switzerland
Maize (*Zea mays*)	√		Belgium (Tervuren)
Wheat (*Triticum aestivum*)		√	Belgium (Tervuren)
Butter bean (*Phaseolus lunatus*)		√	Belgium (Tervuren)

in 1995 where ozone concentrations were lower than in the rest of Europe (Fig. 2, Tables 2 and 3). The high ozone concentrations at Switzerland (Cadenazzo) injured 45 and 32% of subterranean clover leaves per pot (Table 3). The AOT40 values for the 28 day period before each harvest were 6564 and 8335 ppb.h, respectively. However, at Sweden (Östad), where the ozone concentrations were lower, 19% of subterranean clover leaves were injured at the second harvest following a 28 day period of 954 ppb.h and in Italy (Milan), where the AOT40 for 28 day was 6848 ppb.h, injury at the second harvest was approximately 7.9%. This indicated that there was not a definite relationship between long-term exposure and the extent of injury development.

3.2. Validation of the short-term critical level of ozone for visible injury

The short-term critical levels defined in the introduction to this paper were set from the 1995 data shown in Fig. 3. Because of the episodic nature of ozone pollution, AOT40 and mean VPD for the 5 days before injury expression proved to be the best fitting parameters (Benton et al., 1996). Plotting the 1996 data on the same axes showed that, with the exception of three outliers, the original critical levels held up to validation with data from a year in which VPD at the ICP-Vegetation sites was generally lower.

4. Discussion

Ozone concentrations differed at sites across Europe and were generally lower in the Scandinavian countries and higher in central and southern Europe in 1995 and 1996. ICP-Vegetation experiments showed that the ozone concentrations were sufficient to induce visible injury on two clover species and bean at the sites in southern and central Europe, and most of those in northern Europe. Injury was also detected on eight other species growing at or near some of the experimental sites. The only site where injury was not

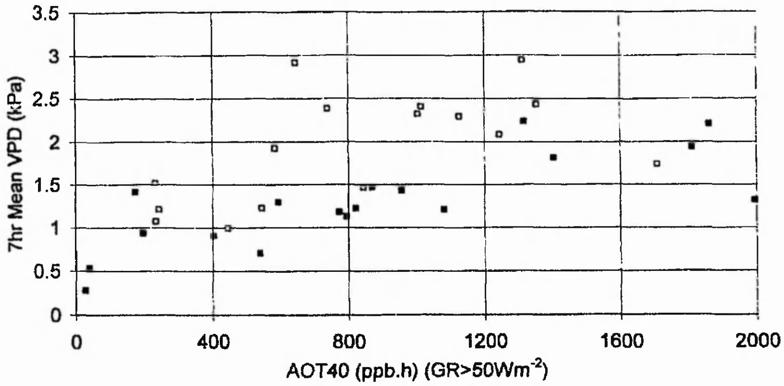


Fig. 3. The AOT40 (for hours with GR $\geq$ 50 Wm^{-2}) and the mean vapour pressure deficit (0930–1630 h) during the 5 days preceding the presence of injury on *Trifolium* species in 1995 (□) and 1996 (■).

observed on any of the species monitored in either year, was that at Finland (Jokioinen) where ozone concentrations were lower than at any other site in the study. Thus, the ICP-Vegetation programme has documented the widespread occurrence of damaging ozone episodes in Europe, and has indicated that a range of crops are potentially at risk from ozone pollution.

The extent of injury to clover and French bean varied among sites and no clear relationship existed between ozone dose in the 28 day period before harvest and the extent of injury on clover at harvest. This may be because injury usually develops following short-term ozone exposure rather than after exposure to a long-term average concentration (Sanders et al., 1994, 1995). Pihl-Karlsson et al. (1995) showed that injury to subterranean clover was greater following a short period with high ozone than following a longer period with lower ozone concentration. However, Amiro et al. (1984) found that ozone concentration and length of ozone exposure were not sufficient to explain the onset of injury on *P. vulgaris*, and Tonneijck (1994) suggested that ozone-climate interactions were important.

The two critical levels for visible injury (defined in the introduction) were set from the 1995 data and included terms for VPD and ozone (Benton et al., 1996). In general, data from the 1996 ICP-Crops experiments supported these critical levels although injury did occur on three occasions before the 5 day AOT40 exceeded 200 ppb.h, when VPD was below 1.5 kPa. Thus, it appears that when VPD is very low, injury can occur following a 5 day AOT40 as low as 27 ppb.h. Similar findings have been reported by Tonneijck and Van Dijk (1997) following studies with subterranean clover. Therefore, it may be necessary to refine these critical levels to address the occurrence of injury when VPDs are small. Other climatic factors that modify the outcome of ozone exposure may become important in these conditions. For example, temperature and solar radiation influence stomatal conductance, and windspeed, through an influence on the laminar boundary layer and atmospheric resistance determines the actual ozone dose available for absorption by the plant (Grünhage and Jäger, 1994, 1996).

Data from the ICP-Vegetation experiments have shown that injury occurs on a range of crop species throughout Europe following exposure to ambient ozone episodes. Injury is especially likely when ozone episodes coincide with periods of high atmospheric humidity, and can occur at relatively low ozone concentrations. Surveys of commercial fields have shown that injury also occurs on crops growing under normal agronomic practices. Thus, several crops are at risk from ozone pollution. The critical levels for visible injury based on 1995 ICP-Vegetation data held for 83% of incidences of injury in 1996. Thus, a combination of VPD and accumulated ozone dose provides a good indicator of the likelihood of injury. However, the presence of outliers suggests that the critical levels need further refinement, possibly by the inclusion of other climatic factors that modify ozone flux into the plant.

Acknowledgements

The authors wish to thank the Department of the Environment, Transport and the Regions (UK) for funding the co-ordination of the ICP-Vegetation (Contract numbers PECD 7/12/145 and EPG 1/3/13) and all participants of the ICP-Vegetation for their valuable support and contributions.

References

Amiro, B.D., Gillespie, J.T., Thurtell, G.W., 1984. Injury response of *Phaseolus vulgaris* to ozone flux density. Atmos. Environ. 18, 1207–1215.

Ball, G.R., Benton, J., Palmer-Brown, D., Fuhrer, J., Skärby, L., Gimeno, B.S., Mills, G.E., 1998. Identifying factors which modify the effects of ambient ozone on white clover (*Trifolium repens* L.) in Europe. Environ. Pollut., in press.

Balls, G.R., Palmer-Brown, D., Sanders, G.E., 1996. Investigating microclimatic influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural networks. New Phytol. 132, 271–280.

Becker, K., Saurer, M., Egger, A., Fuhrer, J., 1989. Sensitivity of white clover to ambient ozone in Switzerland. New Phytol. 112, 235–243.

Benton, J., Fuhrer, J., Gimeno, B.S., Skärby, L., Palmer-Brown, D., Ball, G., Roadknight, C., Sanders-Mills, G.E., 1996. The critical level of ozone for visible injury on crops and natural vegetation (ICP-Crops). In: Kärenlampi, L., Skärby, L. (Eds.), Critical Levels for Ozone in Europe: Testing and Finalising the Concepts. UN-ECE Workshop Report, University of Kuopio, Department of Ecology and Environmental Science, pp. 44–57.

Bull, K.R., 1991. The critical loads/levels approach to gaseous pollutant emission control. Environ. Pollut. 69, 105–123.

CEC, 1993. Results of the European open-top chamber project. Air Pollution Research Report 46. Publication number Eur 14975, EN of the Commission of the European Communities, p. 618.

Dominy, P.J., Heath, R.L., 1985. Inhibition of the K^+-stimulated ATPase of the plasmalemma of Pinto bean leaves by ozone. Plant Physiol. 77, 43–45.

Fernandez-Bayon, J.M., Barnes, J.D., Ollerenshaw, J.H., Davison, A.W., 1993. Physiological effects of ozone on cultivars of watermelon (*Citrullus lanatus*) and muskmelon (*Cucumis melo*) widely grown in Spain. Environ. Pollut. 81, 199–206.

Fuhrer, J., Achermann, B., 1994. Critical Levels for Ozone: a UN/ECE Workshop Report. Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH-3097 Liebefeld-Bern, Switzerland, No. 16, p. 328.

Gimeno, B.S., Mendoza, M., Sánchez, S., Bermejo, V., 1996. Assessment of EDU protection from ozone exposure on three horticultural crops. In: Knoflacher, M., Schneider, J., Soja, G. (Eds.), Exceedance of Critical Loads and Levels. Umweltbundesamt, Federal Environment Agency, Wien, pp. 123–135.

Gimeno, B.S., Penuelas, J., Porcuna, J.L., Reinert, R.A., 1995. Biomonitoring ozone phytotoxicity in Eastern Spain. Water Air Soil Pollut. 85, 1521–1526.

Grantz, D.A., Meinzer, F.C., 1990. Stomatal response to humidity in a sugarcane field: simultaneous porometric and micrometeorological measurements. Plant Cell Environ. 13, 27–37.

Grünhage, L., Jäger, H.-J., 1994. Influence of the atmosphere conductivity on the ozone exposure of plants under ambient conditions: considerations for establishing ozone standards to protect vegetation. Environ. Pollut. 85, 125–129.

Grünhage, L., Jäger, H.-J., 1996. Critical levels for ozone, ozone exposure potentials of the atmosphere or critical absorbed doses for ozone: a general discussion. In: Kärenlampi, L., Skärby, L. (Eds.), Critical Levels for Ozone in Europe: Testing and Finalising the Concepts. UN-ECE Workshop Report, University of Kuopio, Department of Ecology and Environmental Science, pp. 151–168.

Guzy, M.R., Heath, R.L., 1993. Responses to ozone of varieties of common bean (*Phaseolus vulgaris* L.). New Phytol. 124, 617–625.

Heath, R.L., 1987. The biochemistry of ozone attack on the plasma membrane of the plant cells. In: Saunders, J.A., Kosak-Channing, L., Conn, E.E. (Eds.), Recent Advances in Phytochemistry. Phytochemical Effects of Environmental Compounds 21. Plenum Press, New York, pp. 29–54.

Horsman, D.C., Nicholls, A.O., Calder, D.M., 1981. Effects of chronic ozone exposure on the growth of *Trifolium subterraneum* and *Trifolium repens*. Aust. J. Plant Physiol. 8, 405–408.

Kangasjärvi, J., Talvinen, J., Utriainen, M., Karjalainen, R., 1994. Plant defence systems induced by ozone. Plant Cell Environ. 17, 783–794.

Krupa, S.V., Grünhage, L., Jäger, H.-J., Nosel, M., Manning, W.J., Legge, A.H., Hanewald, K., 1995. Ambient ozone (O_3) and adverse crop response: a unified view of cause and effect. Environ. Pollut. 87, 119–126.

Lorenzini, G., Triolo, E., Materazzi, A., 1984. Evidence of visible injury to crop species by ozone in Italy. Rivista Ortoflorofrutticoltora Italiana 68, 81–84.

Pihl-Karlsson, G., Selldén, G., Skärby, L., Pleijel, H., 1995. Leaf age and exposure dynamics. New Phytol. 129, 355–365.

PORG, 1997. Ozone in the United Kingdom. Fourth Report of the United Kingdom Photochemical Oxidants Review Group. Prepared for the Department of the Environment, London, pp. 234.

Sanders, G.E., Balls, G.R., Booth, C.E., 1994. Ozone critical levels for agricultural crops — analysis and interpretation of the results form the UN/ECE International Cooperative Programme for crops. In: Fuhrer, J., Achermann, B. (Eds.), Critical Levels for Ozone a UN/ECE Workshop Report. Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene CH-3097 Liebefeld-Bern, Switzerland, No. 16, pp. 58–72.

Sanders, G., Benton. J., 1995. Ozone Pollution and Plant Responses in Europe — an Illustrated Guide. The Nottingham Trent University, UK, p. 15.

Sanders, G.E., Skärby, L., Ashmore, M.R., Fuhrer, J., 1995. Establishing critical levels for the effects of air pollution on vegetation. Water Air Soil Pollut. 85, 189–200.

Salleras, J.M., Gimeno, B.S., Bermejo, V., Ochoa, M.J., Tarruel, A., 1989. Evolución del ozono y de la sintomatología de sus efectos sobre sandías y otros cultivos en el Delta del Ebro durante. Fruticultura Profesional 26, 127–136.

Skärby, L., Selldén, G., Mortensen, L., Bender, J., Jones, M., Trappeniers, M., Wenzel, A., Fuhrer, J., 1993. Responses of cereals exposed in open-top chambers to air pollutants. In: Jäger, H.J., Unsworth, M.H., De Temmerman, L., Mathy, P. (Eds.), Effects of Air Pollution on Agricultural Crops in Europe, CEC Air Pollution Research Reports 46, pp. 241–259.

Tonneijck, A.E.G., 1983. Foliar injury responses of 24 bean cultivars (*Phaseolus vulgaris*) to various concentrations of ozone. Neth. J. Plant Pathol. 89, 99–104.

Tonneijck, A.E.G., 1994. Use of several plant species as indicators of ambient ozone: exposure-response relationships. In: Fuhrer, J., Achermann, B. (Eds.), Critical Levels for Ozone. FAC Schriftenreihe No. 16, Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene, Liebefeld-Bern, pp. 288–292.

Tonneijck, A.E.G., Van Dijk, C.J., 1997. Assessing effects of ambient ozone on injury and growth of *Trifolium subterraneum* at four rural sites in The Netherlands with ethylene diurea (EDU). Agric. Eco. Environ. 65, 79–88.

UN/ECE, 1994. The ICP-Crops Experimental Protocol. The Department of Life Sciences, The Nottingham Trent University, Nottingham, UK.

UN/ECE, 1995. The ICP-Crops Experimental Protocol. The Department of Life Sciences, The Nottingham Trent University, Nottingham, UK.

UN/ECE, 1996a. Effects of Nitrogen and Ozone. Report prepared by the International Cooperative Programmes and the Mapping Programme under the Working Group on Effects, NIVA, Norway.

UN/ECE, 1996b. The ICP-Crops Experimental Protocol. The Department of Life Sciences, The Nottingham Trent University, Nottingham, UK.

Voltz, A., Kley, D., 1988. Evaluation of the Montsouris series of ozone measurements made in the nineteenth century. Nature 332, 240–242.