Prediction of Relative Humidity in Australia Using a Simple Back-Propagation Neural Network

Sean A. Skilton
Cooperative Research Centre for Tropical Pest Management, and
Department of Computer Science
The University of Queensland
QLD 4072 Australia
E-mail: S.Skilton@ctpm.uq.edu.au

Abstract This paper discusses the application of neural network modelling techniques to the problem of estimating relative humidity. A three layered back-propagation network was developed to predict relative humidity across Australia using inputs such as maximum and minimum temperature, rainfall, longitude and latitude. Comparisons were made between the values predicted by the network and those expected. The neural network is being utilised to estimate relative humidity values for Australia under various enhanced greenhouse effect climate change scenarios.

1. INTRODUCTION

Modelling of insect populations requires a variety of meteorological data inputs. Many insect models utilise relative humidity as an important value. To predict population dynamics under various enhanced greenhouse scenarios it is necessary to utilise meteorological data based on the output of one or more Global Circulation Models (GCM). Due to the large amount of the data output the various atmospheric research groups choose to only save values that are deemed important. At present relative humidity is not included in the set of stored outputs. It was therefore necessary to attempt to model relative humidity based solely upon the parameters available, including GCM predicted and greenhouse invariant parameters. This paper describes the development of a simple back-propagation neural network to model the relative humidity system.

1.1 Artificial Neural Networks

Traditional modelling is based on the application of algorithms. Neural networks, on the other hand, offer a non-algorithmic method of modelling systems. Rather than depending on strict algorithms, neural networks utilise a process of learning.

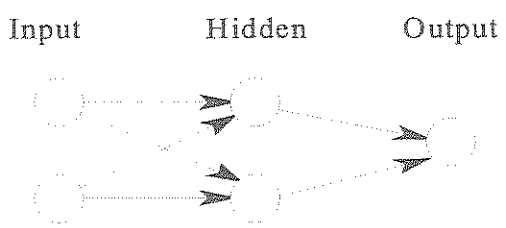

Figure 1. A simple three layer feed forward artificial neural network.

Based upon the brain, neural networks utilise a system of interconnecting nodes (neurons). The most commonly utilised network is known as the Multi-Layer Perceptron (MLP). An MLP consists of an input layer, one or more hidden layers and an output layer. Each layer consists of one or more nodes. In a simple feed-forward network each node in each layer is connected to all the nodes in the next layer. For example a node in the input layer is connected to all the nodes in the first hidden layer.

Neural networks are not only based upon the brain in topology terms, but also in basic inter-node (neuron) communication. Each linkage has an associated weight. This weight is used as a simple scaling factor to be applied to the stimulation value. The weighted inputs are combined with a bias value in the node then transformed by a threshold function. The output of this threshold function forms the input, with another weight, for next node. Neural networks are uniquely identified by several items: The network topology, weights, biases, combining and threshold functions.

As with the brain, neural networks learn by example. A neural network operates in one of two states, training or recall. During the training state a set of known inputs and their corresponding outputs are presented. One of the more common learning algorithms called back-propagation uses a system of automatic weight adjustment. Inputs are presented, fed forward through the network and output value(s) determined. These values are compared to the known values in the training set and adjustments are made to the weights and biases so that the error is minimised.

Once a neural network is trained it can be used in recall mode. Recall mode operates in exactly the same way as learning mode without the back-propagation step. Input nodes are stimulated and activations are fed forward through the network to the output nodes.

2. DATA SETS

To train the neural network to predict relative humidity it was necessary to select a suitable data set. The GCM data (see section 2.3) that was available included change per degree global warming for temperature and rainfall on a 50km grid across Australia. This effectively limited the inputs of the neural network. The data set chosen for training had to at least provide these parameters. In addition the data set needed to be representative to maximise the generalisation power of the trained network.

2.1 MetAccess

MetAccess is a database of real meteorological data recordings for several thousand stations around Australia. It is compiled from Australian Bureau of Meteorology data by the CSIRO Division of Crops and Pastures. This data set suffers from two short comings: uneven distribution of recording stations, and a small number of stations with long runs of the relevant parameters.

The distribution of meterological recording centres is not even. The vast majority of centres lie in the populated eastern seaboard areas of the continent, and there are very few sites distributed across the outback areas. As a result, the data set is not representative of the entire system and reduces the chances that the trained neural network will generalise well. Additionally, only a fraction of the sites include long term recordings of relative humidity and the other relevant variables. The development of a suitable training set would be made extremely difficult utilising the real data; another source of training data is needed.

2.2 Esoclim

Esoclim is a system of spline models of Australian climate (Hutchinson [1989]). Given the latitude, longitude and elevation of a point, Esoclim will provide long term average values for sixteen different climate variables. Relative humidity, although not explicitly output by Esoclim, can be calculated from other surfaces that are provided (see equations 1-3). Since Esoclim is a series of models it is possible to generate climate data for locations evenly distributed across the continent.

$$ED = 6.1078 \times 10^{\left(\frac{TD \times A}{TD + B}\right)} \tag{1}$$

$$ES = 6.1078 \times 10^{\left(\frac{TDB \times A}{TDB + B}\right)} \tag{2}$$

$$RH = 100 \times \frac{ED}{ES} \tag{3}$$

In these equations ED is Vapour Pressure, ES is Saturated Vapour Pressure, RH is Relative Humidity, TD is Dewpoint Temperature, TDB is Drybulb Temperature, A = 7.5 and B = 237.3.

Due to the benefits of better representation and larger data set size the Esoclim system was used to develop a training set for the neural network.

2.3 Climate Change Data

The CSIRO Division of Atmospheric Research provided climate change data based upon the output of five separate Global Circulation Models (GCM). Change data per degree global warming for temperature and rainfall under various enhanced greenhouse scenarios were provided on a seasonal basis. The data consisted of change data in five year increments from 1990 to 2100 for four separate scenarios: hot/wet, hot/dry, cold/wet and cold/dry on a 50km grid across Australia.

3. PREPARATION OF TRAINING SET

Esoclim was utilised to produce a set of climate surfaces for locations across Australia on the same 50km grid as those in the climate change data. A Digital Elevation Model (DEM) was employed to determine elevations at the chosen locations. The longitude, latitude and elevation tuples were then used in Esoclim to build climate data. Relative humidity values were derived from the Esoclim surfaces as described in section 2.2. The Esoclim and relative humidity surfaces then provided the training data for the neural network.

A random ⅔ selection of the locations were included in the training set. The random selection scheme ensured that there would be locations not included in the training set available for testing the generalisation of the trained network. It also ensured that the distribution of these testing points was relatively evenly distributed across the continent.

4. INPUT & OUTPUT ENCODING

The learning algorithm used in neural networks is very powerful. Experience has shown that networks learn in a more efficient way if the input nodes have only small absolute maximum values. It was therefore necessary to perform some preprocessing of the inputs used in the training set. A simple normalisation technique was employed to ensure input and output values remained within the range 0 to 1. Valid ranges were calculated for each of the variables, then equation 4 was applied to perform the transformation.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

This normalisation transformation was applied to all inputs and outputs used by the neural network.

5. NETWORK TOPOLOGY

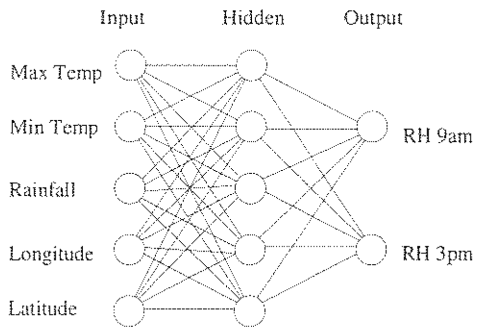

Figure 2. Topology of the neural network used to predict relative humidity.

All simulations were based upon a three layer back-propagation network with 5 hidden nodes and two output nodes, relative humidity at 9am and 3pm. Input attributes were varied to achieve the best model. Initially I utilised maximum and minimum temperature and rainfall, however longitude and latitude were added later to achieve better results (section 6). Figure 2 shows the topology of the network finally used.

6. THE SIMULATIONS

A series of simulations were run to develop the best network for the simulation of relative humidity. Separate training sets for each month were used to train separate networks. Initially a simple network was employed with only maximum and minimum temperature and rainfall as inputs. This series of simulations produced mixed results. Approximately half of the months produced results with errors less than 15 units from the correct value, eg. January (Figure 3a). However, many months included larger errors (up to 30 units from the expected values). The worst month, September is shown as Figure 3b. It was found that the errors followed spatial patterns. The upper Western Australia and Northern Territory regions were areas of high error in several monthly networks. This clumping of errors suggested that the addition of spatial data to the training set to improve the generalisation capabilities.

Longitude and latitude were added to the training set and a new set of simulations were run. Testing of the trained networks showed an improvement over the previous results. All months now had errors of less than 10 units from the correct value. In fact the vast majority of monthly networks had errors of less than 5. January and September have been included for comparison with the previous series of simulations (Figures 3c & d).

7. APPLICATION OF THE TRAINED NETWORK

The monthly neural networks were used to predict relative humidity for various enhanced greenhouse scenarios. Esoclim was employed to provide base data on the same grid as the climate change data from CSIRO Division of Atmospheric Research (section 2.3). The changes were then applied to the base data to produce predicted values for maximum and minimum temperature and rainfall. Using the climate change scenarios provided climate projections for the four different scenarios up to the year 2100 were produced. The trained neural networks were then used to provide the relative humidity estimates from the predicted temperature and rainfall values.

The predicted climate data generated is currently being used within the Cooperative Research Centre for Tropical Pest Management in the modelling of various insects.

8. ACKNOWLEDGEMENTS

PlaNet version 5.6, produced by Yoshiro Miyata, Computer Science Department, University of Colorado, was used as the artificial neural network simulation package.
ARC/INFO version 7.03 was used to visualise errors.
Thanks to Neil White, Tania Yonow and Guy Smith for assistance in both my research and the preparation of this paper.

9. REFERENCES

Hutchinson, M.F., A new objective method for spatial interpolation of meterological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun, CSIRO Division of Water Resources Technical Memo, 89/5, 95-104, 1989.

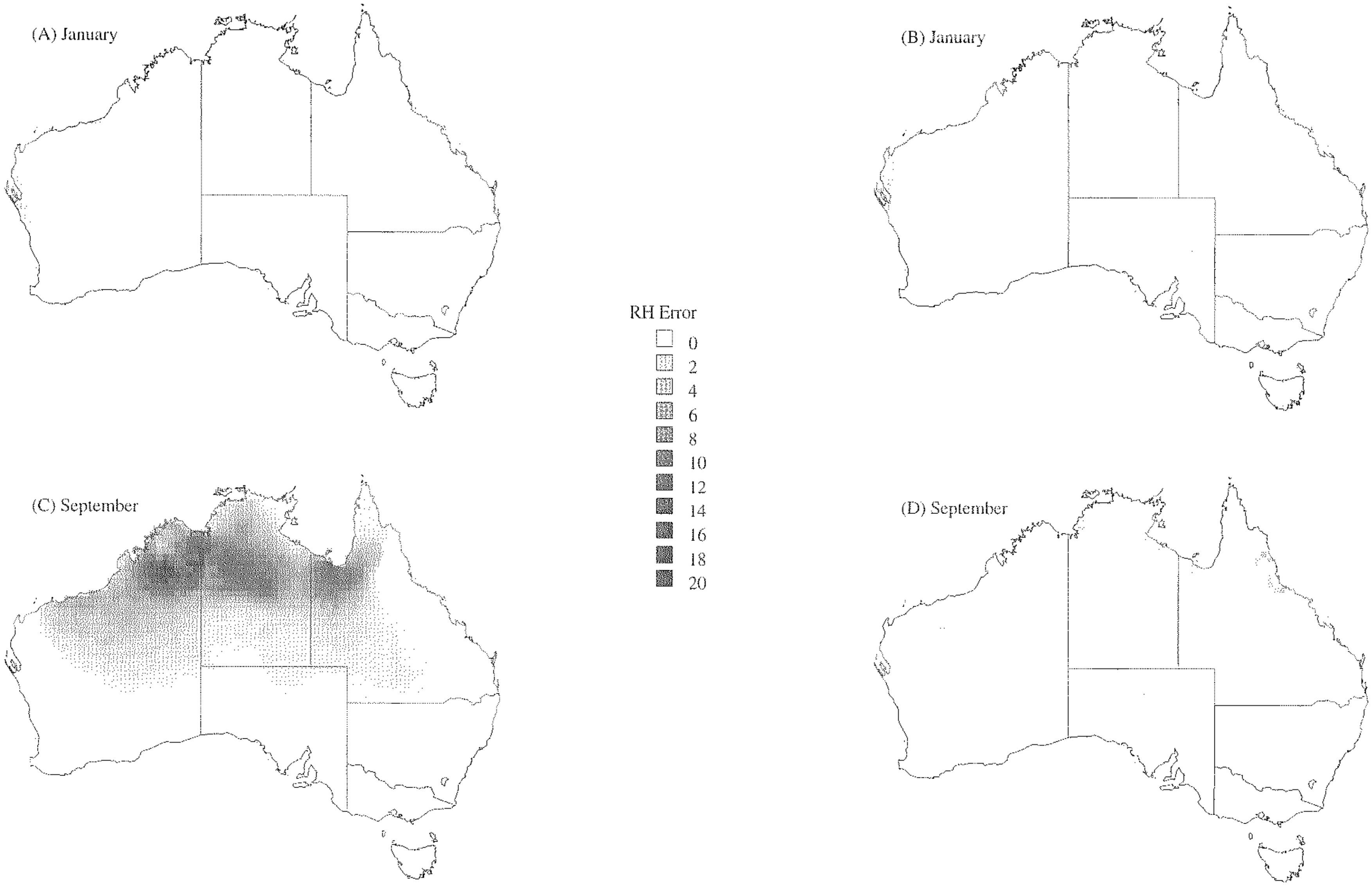

Figure 3. Errors in the prediction of relative humidity. A and C were produced by a neural network using only maximum and minimum temperature and rainfall as inputs. B and D were produced by a network which additionally used longitude and latitude as inputs. January provided the best results, September the worst in both sets of simulations.