PREDICTING AIR TEMPERATURE FOR FROST WARNING USING ARTIFICIAL NEURAL NETWORKS

by

ABHISHEK JAIN

(Under the Direction of R.W. McClendon)

ABSTRACT

One of the most important factors in crop growth is weather, therefore, weather forecasting is vital for agricultural production decision making. For crops such as peaches and blueberries, low temperatures during the bloom period can result in crop damage. Thus frost forecasts are important to provide a warning to farmers, who can then take appropriate actions to minimize damage to their crop. However there are no local short-term frost forecasting systems available at the moment. The complex and non-linear nature of the relationships between various meteorological factors cannot be easily modeled. The goal of this research was to develop a decision support system using Artificial Neural Networks (ANNs) to forecast temperatures in hourly increments from one to twelve hours for any location in south Georgia region, for which, current weather data was available.

INDEX WORDS: Artificial Neural Networks, Frost Prediction, Temperature Forecast, Air Temperature, Weather Data Network, Decision Support System.

PREDICTING AIR TEMPERATURE FOR FROST WARNING USING ARTIFICIAL NEURAL NETWORKS

by

ABHISHEK JAIN,

B.E., The University of Sydney, Australia, 2001

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2003

PREDICTING AIR TEMPERATURE FOR FROST WARNING USING ARTIFICIAL NEURAL NETWORKS

by

ABHISHEK JAIN

Major Professor:	Dr. Ron W. McClendon
Committee:	Dr. Gerrit Hoogenboom Dr. Khaled M. Rasheed

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2003

DEDICATION

I dedicate this thesis to my late mother Mrs. Savita Jain, my father Dr. Ashok Kumar Jain and to my grandmother Mrs. Yagya Devi Tripathi.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the kind guidance of Dr. McClendon, Dr. Hoogenboom and Dr. Rasheed. I thank Dr. McClendon for introducing me to this problem. Without his knowledge guidance and support this research would not have been possible. And without his patient reading and re-reading of my drafts and making suggestions, the thesis would have been atrocious to read. I would like to thank Dr. Hoogenboom for providing much valued perspective from the domain expert point of view and keeping the research firmly oriented towards the end goal. I also thank Dr. Rasheed for being very patient and allowing me to impose on his time at the very last moments.

I would also like to thank my grandmother (Mrs. Yagya Devi Tripathi), my father (Dr. Ashok Kumar Jain), my sister (Mrs. Garima Sharma) and my aunt (Mrs. Anita Kaushik) for giving me encouragement and believing in me.

I would also like to thank my friends Mayukh Dass, Ramyaa, Jason Schlachter, John Burke, Sarah Witzig and Astrid Glende for the support and encouragement they gave me.

And last but not least my girlfriend Ms. Erin Leigh Foret who lifted my spirits when the chips were down and acted like my therapist in my darkest moments. And even though there were times that I was not there for her, she was always there for me. Thank you.

This work was funded in part by the USDA Federal Crop Insurance Corporation Risk Management Agency and by State and Federal Funds allocated to Georgia Agricultural Experiment Stations Hatch projects GEO00895 and GEO00877

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Weather is one of the most important factors in agricultural production, especially in rain-fed agricultural production systems. In such systems, up to 90% of the variability in the yield can be attributed to weather differences (Hoogenboom, 2000c). Fruit crops such as blueberries and peaches are particularly susceptible to low temperatures. Temperatures near but above freezing might slow plant growth and development but such conditions do not cause significant damage if the plants are not exposed to them for a long time. However, once the temperatures go below freezing, the plants are easily damaged, with the severity of damage being determined by the duration of low temperature as well as the temperature itself. Factors such as the type of plant, variety, stage of development, amount of leaf cover and wind speed, also determine the severity of the damage (Tyson et al., 2002).

During the spring of 2002 a large area in blueberry and peach production was destroyed in South Georgia due to an unusually severe and unexpected late frost. Farmers can provide some frost protection by using wind machines that induce air movement, through heating the air by using orchard heaters or through irrigation. Irrigation is the most widely practiced frost protection measure for southeastern crops including peaches and blueberries. It works by forming a layer of ice that keeps the temperature of the flower near freezing, preventing it from dropping to lower temperatures. Farmers need information about when to start irrigation, as the process has to be commenced before the temperature drops to freezing. In addition to the expected low temperatures, farmers also need information about local wind speed, dew point or vapor pressure deficit to determine the point to initiate the freeze protection measures (Hoogenboom and McClendon, 2002). Thus there is a need for accurate local weather information and short-term weather forecasts.

Traditionally the role of providing weather forecasts has been the responsibility of the National Weather Service (NWS). However, changes in the laws have made it so that NWS no longer provides data for agricultural applications. The NWS collects data from urban centers, thus the data is not useful for rural areas where farming is mostly done. In response to this need, University of Georgia initiated the Georgia Automated Environmental Monitoring Network (AEMN) (Hoogenboom, 1996, 2000a, 2000b, Hoogenboom et al., 2000). This is a network of over 50 automated weather stations that are mainly located in the remote areas of Georgia. These weather stations collect temperature, relative humidity, soil temperature at 2 cm depth, 5 cm depth, 10 cm depth & 20 cm depth, wind speed, wind direction, solar radiation, vapor pressure deficit and soil moisture every second. The averages, or totals depending on the variable, are calculated for every fifteen minutes and stored in the data logger. In addition, daily summaries are also calculated. The data are downloaded to a central computer located in Griffin. The AEMN program has a website (www.Georgiaweather.net) that disseminates this information as well as simple calculators that can dynamically calculate degree days, chilling hours or water balance for management of irrigation (Georgiev and Hoogenboom, 1998, 1999, Hoogenboom et al. 1998). The web page has proved to be very popular; however it does not have a forecasting component.

This thesis outlines the research that was carried out to create the ANNs that would eventually serve as the engine for the DSS. The broad aims of this thesis were 1) To create

ANNs that would be able to predict short term temperatures for a given location when trained with historical weather data from the same location and 2) To create ANNs that could predict short term temperatures for a location that did not have sufficient or lacked historical weather data.

ANNs have been used to predict frost formation in the past. A study was conducted by Robinson and Mort (1996) to develop an ANN based system that would predict overnight frost formation in Sicily, Italy. They chose to cast the problem as one of classification as freezing or non freezing rather than predicting the temperature. They primarily experimented with one- and two-hidden layer architectures. In their models the inputs consisted of previous days' minimum and maximum temperatures, cloud cover, maximum wind speed and direction, humidity, wind speed and wind direction at 1900 hrs. The variables listed were tested for their usefulness to the ANN. The number of previous days required were also tested for their usefulness to the ANN. They chose gray code instead of continuous value variables to encode the variables and the output. The ANN developed in their study would classify the weather conditions that would lead to frost by outputting a set of binary values that were then interpreted by the same laws that were used in the training phase. The output of the ANN was thus a classification of whether the temperatures would be freezing or non-freezing in the next 24 hours. A frost event was defined as any temperature below $1^{o}C$. Data from the months of January and February were used as these were the two months that the frost is formed in the geographical area of their interest, i.e. Sicily. They reported that the best ANN predicted two false alarms and one failure over a course of a 50-day model evaluation set. This study is limited by the fact that it only classifies a given 24 hour period as a non-freezing or freezing period. It does not provide predicted temperatures during the next 24 hours or a prediction as to when the freeze event will occur.

Li et al. (2004) conducted a study in which they developed ANNs that estimated historical daily solar radiation and minimum and maximum air temperatures. The estimation was used for prior temperatures and solar radiation for locations that had missing observations for these variables or locations that did not have a weather station. They used weather data from other neighboring locations to estimate the values of these variables. The inputs used were dependent on what was the variable that was being estimated. For example they outlined using the straight line distance between the adjoining locations and the location for which the variable needs to be estimated, change in elevation of the two locations and the daily minimum temperature at the adjoining locations to estimate the daily minimum temperature. They replaced the daily minimum temperature input variables by daily maximum temperature input variables when they wished to estimate the daily maximum temperature at the location of interest. The authors reported that the results generated by the ANN were better than other statistical techniques.

In chapter 1 the problem is introduced and the overall goal of the research is stated along with some background information. This chapter will also provide information about the organization of the thesis and contents of chapter two, three, four and five.

Chapter 2 describes the objectives, methodology, results and preliminary conclusions dealing with development of location specific models (i.e. objective 1). The discussion would be restricted to the location specific models of Fort Valley, Alma and Blairsville.

In chapter 3 we describe the development of a general model for temperature prediction (i.e. objective 2). We develop a general model that uses data from many locations for training and which can then be used to predict temperature of any location.

In chapter 4 we summarize thesis research and provide conclusions about the results in the previous chapters as well as future research.

CHAPTER 2

FROST PREDICTION USING ARTIFICIAL NEURAL NETWORKS: A TEMPERATURE PREDICTION APPROACH

Jain, A., R.W. McClendon and G. Hoogenboom, to be submitted to The Transactions of the ASAE

ABSTRACT

The goal of this research was to develop Artificial Neural Networks (ANNs) to forecast air temperatures in hourly increments from one to twelve hours. ANNS were developed for predicting temperature for a location in one hour increments beginning from one hour to twelve hours in the future. Weather data for model development and evaluation for three locations in Georgia (Fort Valley, Blairsville, and Alma) were obtained from the Georgia Automated Environmental Monitoring Network (AEMN). The data consist of observations of various meteorological variables such as air temperature, relative humidity, wind speed, rainfall and solar radiation. Important inputs were determined by developing ANNs that used them in various permutations as inputs and their affect on the accuracy of the predictions. The optimal duration of prior data for the input variables was determined by experimenting with inputs that used two, four and six hours of prior data for the variables. It was found that temperature, relative humidity, solar radiation, and wind speed were important in generating temperature forecasts. The optimal duration of prior data varied from two to six hours depending on the length of the period into the future the ANN was predicting. The MAE varied depending on the period of predictions. The range in MAE for predicting temperature one hour and twelve hours in the future was 0.6^{o}C and 2.4^{o}C for Fort Valley, 0.7^{o}C and 3.0^{o}C for Blairsville, and 0.6^{o}C and 2.6^{o}C for Alma, Georgia.

INTRODUCTION

One of the most important factors that influences agricultural production is weather. In rainfed agricultural production systems, up to 90% of the variability in the yield can be attributed to weather differences (Hoogenboom, 2000a). Fruit crops such as blueberries and peaches are particularly susceptible to low air temperatures. Temperatures near but above freezing might slow plant growth and development but such conditions do not typically cause significant damage if the plants are not exposed to them for a long time. However, once the temperature drops below freezing, the plants are easily damaged, with the severity of damage being determined by the duration of low temperature as well as the temperature itself. Factors such as the type of plant, variety, stage of development, amount of leaf cover and wind speed also contribute to the severity of the damage (Okie et al., 1998 Tyson et al., 2002 and Powell and Himelrick, 2003).

As an example, during the spring of 2002, a large area in blueberry and peach production in South Georgia was destroyed due to an unusually severe and unexpected late frost. In Fort Valley the last freezing temperatures were recorded as low -1.18^{o}C on March 23, 2002. Farmers can provide some frost protection by using wind machines that induce air movement, by heating the air using orchard heaters, or by irrigating. Irrigation is the most widely practiced frost protection measure for southeastern U.S. crops including peaches and blueberries. It results in the formation of a layer of ice that keeps the temperature of the flower near freezing, preventing it from dropping to below freezing temperatures. Farmers need information about when to start irrigation, as the process has to be commenced before the temperature drops to freezing. In addition to the expected low temperatures, farmers also need information about local wind speed, dew point or vapor pressure deficit to determine the point to initiate the freeze protection measures. Thus there is a need for accurate local weather information and short-term weather forecasts.

Traditionally the role of providing weather forecasts has been the responsibility of the National Weather Service (NWS). However, due to changes in the laws, the NWS no longer provides data for agricultural applications. The NWS collects data from urban centers, thus the data are less useful for rural areas where farming is mostly done. In response to this need, the University of Georgia initiated the Georgia Automated Environmental Monitoring Network (AEMN) (Hoogenboom, 1996; 2000a; 2000b; Hoogenboom et al., 2000). This is a network of over 57 automated weather stations that are mainly located in the rural areas of Georgia (www.Georgiaweather.net). These weather stations collect data every second on the following variables: air temperature, relative humidity, soil temperature at 5 cm, 10 cm and 20 cm depths, wind speed, wind direction, solar radiation, vapor pressure deficit and soil moisture. The averages, or totals depending on the variable, are calculated every fifteen minutes and stored in the data logger. In addition, daily summaries are also calculated. The 15-minute and daily summary data are downloaded automatically to a central computer located in Griffin. The website disseminates this information as well as simple calculators that can dynamically provide degree days, chilling hours or water balance for management of irrigation (Georgiev and Hoogenboom, 1998; 1999; Hoogenboom et al., 1998). The web page has proved to be popular, however it does not have a forecasting component.

An Artificial Neural Network (ANN) is a computational intelligence technique that mimics the behavior of neurons in the brain. The basic components of an ANN are its nodes or neurons and the connections between the nodes. A node is primarily a computational unit. It receives inputs, calculates a weighted sum and presents the sum to an activation function. The nodes are generally arranged in layers. In the back propagation ANN architecture, the input layer nodes receive the inputs and pass the results of their computations to the nodes in the hidden layer by means of the connections. The hidden nodes sum these weighted values as inputs, calculate an activation level and pass the results to the output nodes.

For an ANN to be useful it should be able to learn or capture the complex relationships between inputs and outputs. This is done by searching for an optimal set of the weights of the connections between the nodes. It is achieved by first sending one set of inputs in the feed forward mode through the ANN. The error between the ANN output and the expected output is calculated. The error is then used to adjust the weights of the connections by using the method of gradient descent. The data that are used for this process is called the training data. The process is repeated until the error on another set of data called the testing data reaches a minimum. The testing data set are used in the feed forward mode only and are not used to adjust the weights. The training data and testing data comprise the model development data set. Once the training is complete, the ANN is then used with a separate model evaluation data set to determine its accuracy.

ANNs have been used in several studies to estimate meteorological variables. Elizondo et al. (1994) developed an ANN to estimate daily solar radiation. In this study, daily meteorological data such as minimum and maximum air temperature, precipitation, day length and clear sky radiation were the inputs. Felker and Han, (1997) used a radial basis function ANN to estimate daily soil water evaporation with daily average relative air humidity, air temperature, wind speed and soil water content in a cactus field serving as inputs. Bruton et al. (2000) utilized ANNs to estimate daily pan evaporation. Daily rainfall, temperature, relative humidity, solar radiation and wind speed were the inputs. The results generated by the ANNs compared favorably against other numerical methods. Mehuys et al. (1997) created ANNs to simulate daily fluctuations of soil temperature at various depths. Their approach worked well compared to a more traditional modeling approach.

Cook and Wolfe (1991) developed an ANN that predicted long term temperatures. Their ANN predicted monthly average air temperature for one month, two months and three months in advance. The inputs consisted of monthly average daily maximum temperatures from the previous eleven months and the present month, precipitation in the current month, relative monthly humidity at four times per day for current month and monthly average daily minimum temperature. Their ANN tended to over predict temperatures earlier months and under predict later months. Abdel-Aal and Elhadidy (1994) conducted a study to estimate daily minimum temperature one and three day(s) in advance. The input variables included maximum air temperature, minimum air temperature, mean air temperature, maximum wind speed, mean wind speed, minimum wind speed, wind speed vector, wind direction, maximum barometric pressure, mean barometric pressure, minimum barometric pressure, maximum relative humidity, mean relative humidity, minimum relative humidity, prevailing wind direction, number of hours in the prevailing wind direction, direction of maximum wind, mean global solar radiation. Their model resulted in 92.8% of the predictions to have an MAE of less than 3^{o}C.

Robinson and Mort (1996) developed an ANN based system to predict overnight frost formation. The problem was cast as one of classification rather than predicting the temperature. The output of the ANN was a classification of temperatures as freezing or non-freezing for the next 24 hours. Input variables were previous days' minimum and maximum temperatures, cloud cover, maximum wind speed and direction, humidity, wind speed and wind direction at 1900 hrs. The ANN would classify the weather conditions that would lead to frost by outputting a set of binary values that were then interpreted along the same laws that were used in the training phase. They defined a frost event as any temperature below 1^{o}C. The best ANN predicted two false alarms (false positive) and one failure (false negative) over a course of a 50-day validation set.

Blennow and Persson (1998) used a Geographic Information System (GIS) and a stepwise linear regression model to find out where frost could occur over an area covering 7.5 km^2 of forested and heterogeneous region in Sweden. Their model was able to explain between 87% and 88% of the air temperature variations in two separate validation sets.

The overall goal of the research reported herein is to develop an ANN based Decision Support System (DSS) for predicting the occurrence of frost by predicting hourly temperatures during the subsequent twelve hour period. The specific objectives of this study are to determine (1) the most important inputs needed for the temperature predictions, (2) the duration of prior weather data needed for each period of prediction, (3) the best ANN architecture for a given set of inputs and (4) if the ANNs developed for a specific location can forecast temperatures at other locations.

MATERIALS AND METHODS

NeuroShell™ software (Ward System Group Inc., Frederic, MD, 1993) was used to develop the ANNs in this study. In a preliminary study the accuracy of four types of architectures were compared. The ANN architectures included standard back propagation with one hidden layer and three hidden layers, jump connection and Ward ANNs. The back propagation ANN as shown in Figure 2.1 has three layers: input layer, hidden layer and output layer. The functions of these layers are as previously outlined. A three hidden layer ANN as shown in Figure 2.2 has three hidden layers instead of one. A jump connection ANN as shown in Figure 2.3 also has three hidden layers but the connections are made from the input layer to first, second and third hidden layer individually as well as to the output layer directly. The first hidden layer is also connected not only to the second hidden layer but also to the third and the output layer. Similarly the second hidden layer is connected to the third hidden and the output layer. The third input layer is connected only to the output layer. A Ward ANN as shown in Figure 2.4 is a back propagation ANN with three slabs of hidden nodes in a single hidden layer that have different activation functions. The output and input layers also have different activation functions. The assumption is that the various activation functions will capture relationships among variables that would be missed if a single activation function was used.

From a preliminary study it was found that results of the Ward ANN were slightly more accurate than the other ANN architectures described above. Thus for all subsequent model development only the Ward ANNs were considered. The ANNs developed for this preliminary study used inputs which were the current values of the temperature, relative humidity, wind speed, solar radiation, rainfall and time of day. Time of day was converted to a cyclic variable. Hourly weather data for Fort Valley from 1998 to 2001 were used for model development and 2002 data were used for model evaluation.

For all experiments herein the number of hidden nodes was varied to determine the preferred number for each input configuration considered. The best architecture for each input variable configuration was taken as the standard for that configuration i.e. each input parameter configuration had a different hidden node configuration. The activation functions used were linear for the input layer, Gaussian, tanh and Gaussian complement for the slabs in the hidden layer and logistic for the output layer. All these functions were the default functions for the Ward ANNs.

For the purpose of this study weather data were obtained from the Georgia AEMN. The data originated from the fruit producing areas of Georgia, including sites in peach producing areas (Fort Valley and Blairsville) as well as a blueberry producing area (Alma). The data from each location were divided into a model development set and model evaluation set. Model development set was subdivided into training and testing set with 60% of model development data used as training data and the remaining 40% used as testing data. The data from the training set were used in the iterative search to determine the optimal weights of connections between the nodes, while the data from the testing set were used in a feed forward mode only to determine when to stop the training. Data prior to 2001 were used for model development and data from years 2001 and 2002 were used for model evaluation. Model evaluation data were used in a feed forward mode to determine the final accuracy of the model and selecting the best architecture. A final evaluation of the models was performed with data for 2003 since the 2001 and 2002 data, while not directly used for model development, were used to select the best configuration of hidden nodes. Only data from the months of January through to April were used in the study. It is during this period that air temperatures vary between freezing and non-freezing and the crops are susceptible to freeze conditions.

Data for the years beginning in 1996, or earlier for some locations were in a 15-minute format. Prior to this time the data were recorded in an hourly format. Both formats were used in this study. Data with 15-minute resolution were available from late 1996 to the present and as such there are presently eight years of data available. The hourly data were available from when the particular weather station was set until early 1996. There are four to five years of data available in this format for the three locations considered. Obviously, data in the 15-minute resolution format can be converted into hourly format. Since part of the 1996 data was in the hourly resolution format and the remaining was in 15 minute resolution, it can be counted for both the resolutions. Thus eleven to twelve years of hourly resolution data, depending on the location, were available while eight years of 15 minute resolution data were available.

The weather data consisted of observations of temperature, relative humidity, wind speed, solar radiation and rainfall. The current values of these variables were used as inputs along with corresponding prior values of these variables. The change in value of the weather variables from the prior values to current value (Δ) were calculated and also used as inputs. For example if current temperature is t_0 and the temperature one hour ago was t_1 (one hour of prior data) then the Δ value for this variable and for this duration was $t_0 - t_1$. Time of day was also included as an input using a periodic input variable approach to capture the cyclic nature of the hours of day. We used four input variables for time with variables corresponding to 2400 hrs, 0600 hrs, 1200 hrs and 1800 hrs. For example the 1200 hr variable has a value of 1 at 1200 hrs and a value of 0 for period 1800 to 0600 hrs. Thus an intermediate hour such as 1500 hrs would have value of 0.5 for the 1200 hr variable, 0.5 for the 1800 hr variable and 0 for 0600 and 2400 hr variables.

To identify which weather variables are important, experiments were conducted with the hourly data from Fort Valley. The duration of prior data was held constant at four hours and the period of prediction was kept constant at four hours. The first set of experiment was performed to determine the best architecture when only temperature and its related inputs i.e. four hours of prior temperatures as well as their respective Δ values were used along with time of day. This set of inputs was called the core set of input variables. It was assumed that the current and prior air temperatures as well as time of day were important input variables. Subsequently, experiments were conducted for determining the best architecture when two weather variables are used as inputs. For this set of experiments the core set of inputs was used in conjunction with each of the other weather variables (relative humidity, solar radiation, rainfall and wind speed) one at a time to serve as the inputs. These experiments were then ranked based on the performance of their best architectures. The combination ranked first then became the new core set of inputs and subsequent experiments were conducted by adding a third weather variable. This process was carried out until the most important input variables were determined in order of importance.

Once the most important inputs were determined, the next step was to determine the duration of prior data needed for each period of prediction. In a preliminary study it was determined that prior weather data durations greater than six hours were not beneficial to the model accuracy. Thus prior data up to six hours prior to the current time were considered. Experiments were conducted with two hours, four hours and six hours of prior data to predict temperature one to twelve hours in the future in hourly increments. The hourly format weather data, from Fort Valley were used for these experiments.

Data were available in two formats (15-minute format and hourly format). To determine which format gave the most accurate predictions, experiments were conducted with durations of two, four and six hours of prior data for the 15-minute resolution data set. The prior data itself was in 15-minute increments. At the end of this stage, the best ANNs for individual periods of predictions were determined. Using these results, models were developed for Alma and Blairsville.

To determine if models developed for a particular location could be used to make predictions for other locations experiments were conducted in which model evaluation data from one location were used as inputs for a model that had been developed with data from another location. For example, data from Blairsville (specifically data from the years 2001 and 2002) were the input for the model developed for Fort Valley. The data were used in the feed forward mode only and the accuracy was compared to the accuracy of a model developed for Blairsville.

For all these analyses, the measure of accuracy was the mean absolute error (MAE) calculated on the model evaluation set. The exception was the portion of the study during which the most important input variable were being determined. The MAE, for that portion of the study, was calculated for those temperatures in the model evaluation set, which were observed to be below 5^{o}C. This was done to insure that while determining the most important weather variables, the emphasis was on temperatures of interest, i.e. temperatures that were close to 0^{o}C

RESULTS AND DISCUSSION

When using only temperature and its associated Δ values as the only inputs, it was found that the lowest MAE was 1.41^{o}C (Table 2.1) when predicting four hours in the future. Each of the remaining weather variables and associated Δ values were included one at a time to determine the second most important input variable. When the relative humidity variables were added to the temperature variables it was found that, the MAE improved to 1.19^{o}C (Table 2.1). When wind speed was the second input variable the MAE was 1.34^{o}C (Table 2.1). With solar radiation as the second variable the MAE was 1.35^{o}C and with rainfall it was 1.44^{o}C (Table 2.1). On the basis of these results, temperature and relative humidity variables were now considered to be the core set of input variables. Subsequently, wind speed, solar radiation and rainfall variables were added one at a time to the core set of input variables. It was found that adding wind speed variables to the core set of input variables improved the MAE to 1.15^{o}C (Table 2.1). Adding solar radiation variables improved the MAE to 1.17^{o}C compared to 1.19^{o}C when using only temperature and relative humidity (Table 2.1). Adding rainfall variables degraded the MAE to 1.21^{o}C. Since adding wind speed improved the accuracy of the models, wind speed was added to the core set of input variables. Again, the two remaining variables solar radiation and rainfall were added to these new core set of input variables. The addition of solar radiation improved the MAE to 1.12^{o}C compared to the 1.15^{o}C of the previous core set (Table 2.1). However the addition of rainfall variables to the core set of input variables increased the MAE to 1.16^{o}C (Table 2.1). Thus solar radiation was added to the core set of input variables. Finally only rainfall remained and it was added to the new core set of input variables and the MAE increased to 1.14^{o}C (Table 2.1). Using this approach it was thus determined that the order of importance of the weather variable inputs considered was temperature, relative humidity, wind speed and solar radiation. The addition of rainfall slightly reduced the accuracy of the ANNs and it was thus excluded from all subsequent model development.

ANN models were then developed using weather data with a 15-minute resolution for predicting temperatures in hourly increments starting from one hour to twelve hours in the future. The preferred duration of prior data was determined. Models were developed with two hours, four hours and six hours of prior data as inputs. It was found that the two hours of prior data produced the highest accuracy for one hour predictions with MAE of 0.59^{o}C (Table 2.2) and six hours of prior data for predicting twelve hours in the future with MAE of 2.47^{o}C (Table 2.2). For the twelve hour predictions there was slight variation in accuracy between the three durations considered. Therefore two hours of prior data were used for the development of all models utilizing the 15 minute data format. The results for the two prediction periods are shown in Table 2.3.

It was also determined that individual ANNs that predicted the temperature for a given period (single output node) were slightly more accurate that a single ANN that predicted all twelve of periods (twelve output nodes) as shown in Table 2.4. The MAE associated with period of prediction of one hour was 0.67^{o}C and a twelve hour period of prediction was 2.63^{o}C. The errors for each prediction period are slightly larger than the errors for individual models for each corresponding periods as shown in Table 2.3.

It was observed that when using data in the hourly format the duration of prior data needed depended on the period of prediction. Experiments were conducted for one and twelve hour periods of prediction to determine the duration of prior data (two, four and six hours) with lowest MAE. It was found that for one hour period of prediction, two hours of prior data had the lowest MAE and six hours of prior data had the lowest MAE for the twelve hour period of prediction. Subsequently experiments were conducted for intermediate periods of prediction. Models that predicted temperature, one, two and three hours in the future had the lowest MAE associated with two hours of prior data, while the rest of the models had the highest accuracy with six hours of prior data. The results for the lowest MAEs are shown in Table 2.5. The hourly format models had slightly smaller MAEs in comparison to the corresponding models that used 15-minute format (Table 2.3). It is likely that the hourly resolution weather data produced better results than the 15-minute resolution weather data due to the additional years of data available. With the 15-minute format, inputs had a higher resolution but fewer years of data available. Due to constraints in the Ward System software in terms of the number of patterns allowed in model development data set, this data were further limited that it could only include data from 1998 until 2000, while in the hourly format the number of years of data were greater but the resolution of the inputs was lower. It is likely that the models that used hourly format data were more robust as they were trained with the additional years of weather data. Based on these results, the hourly resolution data format was used for all subsequent model development.

Also as expected, the accuracy of the ANNs decreased as the prediction period increased (Table 2.5). The MAE for predicting the temperature one hour in the future was 0.56^{o}C using the hourly format and was 2.36^{o}C for predicting temperature twelve hours in the future. A plot of predicted temperatures for a one hour period of prediction vs. the observed temperatures using the hourly data format for Fort Valley, GA, is shown in Figure 2.5. A linear regression line fit to this data gives an R^2 of 0.989. A similar plot of predicted temperature for a twelve hour period is shown in Figure 2.6. As expected, the scatter is greater for the 12 hour period of prediction and the R^2 was 0.818.

ANN models were then developed using data for Alma and Blairsville GA, using the architecture and approach determined from the study using Fort Valley, GA, data. The models were developed using hourly format data for Blairsville from 1992 to 2000 and evaluated with the 2001 and 2002 data. As shown in Table 2.6, for Blairsville, the MAE for the evaluation data set for predicting temperature one hour in the future for the evaluation data set is 0.65°C and for predicting twelve hours in the future is 2.96°C. The MAEs for the other prediction periods varied between these two values. The MAEs for Blairsville were higher than errors for the corresponding ANNs developed for Fort Valley. Blairsville is located in the north Georgia mountains and experiences a wider range of temperatures than Fort Valley, which is in the Coastal Plain. For Alma, the models were developed using hourly format data from Alma from 1993 to 2000 and evaluated on 2001-2002 data. For Alma, MAE for the evaluation data set for predicting temperature one hour in the future was 0.56°C and for the prediction period of twelve hours was 2.6°C (Table 2.6). The MAEs for the other prediction periods varied between these two values. The results are similar to the results for Fort Valley. Although these two locations are approximately 100 miles (161 km) apart, they are both located in the Coastal Plain.

Once the ANN models for the three locations were developed, the final evaluation data set for 2003 from the three locations was input to their respective models to determine the overall accuracy of the models. The MAE for the prediction periods of one and twelve hours for Fort Valley was 0.54°C and 2.41°C respectively (Table 2.7), which is very similar to the MAEs of

0.56^{o}C and 2.36^{o}C that were obtained for the 2001 and 2002 data (Table 2.5). Similarly the MAEs associated with predicting temperatures one and twelve hours in the future for Alma for 2003 are found to be 0.53^{o}C and 2.44^{o}C respectively (Table 2.7) which was slightly lower than the MAEs of 0.56^{o}C and 2.6^{o}C that were found for the 2001 and 2002 data from Alma (Table 2.6). The 2003 data from Blairsville also demonstrated similar behavior with the MAEs associated with predicting temperature one and twelve hours in the future being 0.62^{o}C and 2.98^{o}C, respectively (Table 2.7). These MAEs were very similar to the MAEs that were obtained for the 2001 and 2002 data from Blairsville, which were 0.65^{o}C and 2.96^{o}C for predicting temperatures one and twelve hours in the future, respectively (Table 2.6).

The performance of the ANNs for a prediction periods of one hour and twelve hours were used to gauge the robustness of the models for predicting temperatures at other locations. When Alma models used Alma data for 2001 and 2002, the MAEs associated with the one and twelve hours period of prediction models were 0.56^{o}C and 2.60^{o}C, respectively (Table 2.8). When Blairsville models used Alma data as input, the MAEs were 0.83^{o}C and 3.37^{o}C for one and twelve hour period of prediction, respectively (Table 2.8). The Fort Valley models, when presented with the same data, had MAEs of 0.61^{o}C and 2.71^{o}C for the one and twelve hour periods of predictions models, respectively (Table 2.8). The accuracy of the Alma models was the highest as these models were trained with data from Alma. The accuracy of the one hour and twelve hours period of prediction models developed for Fort Valley and Alma was comparable while that of Blairsville models was higher. It is likely that Alma and Fort Valley results were comparable because both Alma and Fort Valley are located in South Georgia in the Coastal Plain and as such experience similar weather patterns. However, Blairsville is situated in the North Georgia mountains in an area that experiences different weather conditions than South Georgia.

The Blairsville models when presented with 2001 and 2002 weather data from Blairsville resulted in MAEs of 0.65°C and 2.96°C for ANNs predicting temperatures one and twelve hours in the future respectively (Table 2.8). When Alma ANNs used the same Blairsville data as input, the MAEs were 0.75°C and 3.27°C for predicting temperatures one and twelve hours in the future (Table 2.8). Fort Valley ANNs that predict temperatures one and twelve hours in the future, when presented with the Blairsville data had MAEs of 0.75°C and 3.28°C, respectively (Table 2.8). The accuracy of the twelve hour models for Fort Valley and Alma models was almost identical, while that of the one hour model was the same. However they were both worse than the accuracy of the original Blairsville model. Again the difference can be attributed to different weather phenomena experienced in Blairsville compared to Fort Valley and Alma.

Similarly when Fort Valley ANNs that predict temperature one and twelve hours future, used the 2001 and 2002 data from Fort Valley as inputs, the MAEs were 0.56°C and 2.36°C, respectively (Table 2.8). For the same period of predictions (one and twelve hours) the Alma models when presented with the Fort Valley data resulted in MAEs of 0.58°C and 2.44°C respectively (Table 2.8). When Blairsville models used the Fort Valley data as inputs the resulting MAEs were 0.73°C and 2.98°C for ANNs predicting temperatures one and twelve hours in the future, respectively (Table 2.8). The accuracy of the Alma and Fort Valley models was comparable, with the accuracy of the Fort Valley models being higher. The accuracy of the Blairsville models was lower than the accuracy of Fort Valley and Alma models. Alma and Fort Valley are in the same region of the state and hence experience similar weather patterns, which could explain why the MAEs of the corresponding models for the two locations are comparable. Hence it is seen that models developed for a particular location can predict temperature at another location with slightly lower accuracy as long as the other location is near. However when these models predict temperatures for locations that are in different regions then the accuracy of the predictions is low.

SIMULATION

The overall goal of this research project was to develop ANN models for predicting temperature which could be incorporated into a real time frost warning system. As such, all twelve ANNs predicting hourly temperatures for a given location were used to generate a simulated temperature forecast for the subsequent twelve hours. The outputs from the twelve networks that were developed were combined into a Decision Support System for frost prediction. The data for the simulation were selected from periods which included a freeze event. A Decision Point (DP) is defined as that point in time ‘t’ when all networks are presented with the latest values of the input variables. The first decision point (DP1) was selected to be approximately ten hours prior to a freeze event. The second and third decision points (DP2 and DP3) were selected to be approximately six and two hours prior to a freeze event, respectively.

As an example a simulation using data from Fort Valley started at 1800 hours on January 21, 2001 and ended at 1400 hours on January 22, 2001 (Figure 2.7). The first freezing temperature was ten hours after DP1. The twelve ANNs were then used to predict temperatures for the subsequent twelve hours and they are shown in Figure 2.7. The ANNs using the data from DP1 predicted the temperatures with reasonable accuracy. The ANNs predicted the first freeze event and hour earlier and then predicted the temperature to rise above $0^{o}C$ slightly for one hour and then drop to below $0^{o}C$ again (Figure 2.7). Still the temperature predicted when the first freezing temperature actually occurred is very close to the observed value. The observed temperature at that point is $-0.05^{o}C$ while the predicted temperature is $0.40^{o}C$ (Figure 2.7). The

ANNs predict the first freeze temperature as -0.25°C, one hour prior to the actual occurrence of frost. The observed temperature at this point is 0.71°C (Figure 2.7).

For the DP2 (which is at 2200 hours on January 21), the ANNs used current observed data at the time of DP2. For this simulation the second DP was arbitrarily chosen to be four hours later. In actual implementation, the DSS could be accessed as often as every fifteen minutes. From DP2 the predicted temperature just hovered close to 0°C. At the time of the first freeze predicted temperature was 0.11°C and the observed temperature was -0.05°C (Figure 2.7).

Similarly for DP3 (which is at 0200 hours on January 22), the ANNs used the current observed data at DP3 (eight hours after DP1). Predictions of the ANNs using data of DP3 are more accurate than the ANNs of previous decision points. The first freeze temperature is correctly predicted. The ANNs predict a temperature of -0.16°C at 0400 hours on January 22 while the actual temperature at that time was -0.05°C (Figure 2.7).

When the DP from which predictions are made was close to a temperature, the accuracy of the predictions in general increased. In this case the first observed freeze temperature was -0.05°C at 0400 hours on January 22, and while ANNs using data of DP1 predicted a temperature of 0.40°C, the ANNs using data of subsequent DPs predicted values that were much closer to the observed (0.11°C by ANNs using data of DP2 and -0.16°C by ANNs using data of DP3). A similar trend was apparent for the temperature observed nine hours after DP1 (one hour prior to the first observed below 0°C temperature). In this case the observed temperature is 0.71°C. ANNs using data of DP1 predicted a temperature of -0.25°C while the ANNs using data of DP2 and DP3 predicted a temperature of 0.34°C and 0.25°C, respectively. It was found that the temperature predictions could also be used for estimating the duration of a particular freeze event by determining the first temperature predicted above 0°C after predicting the freeze temperature.

In this case while the scope of predictions for DP1 ended when the temperature conditions were still below 0^{o}C, ANNs using data of DP2 and DP3 were in a position to predict when the freeze conditions would end. ANNs using data of DP2 predicted the end of the freeze conditions one hour prematurely by predicting the first non-freezing temperature after predicting a freezing temperature at 0800 hours on January 22, whereas the first non-freeze temperature occurred at 0900 hours on January 22. The same was observed for ANNs using data of DP3; however, the magnitude of the over estimation of the temperature is less than that of the ANNs using data of DP2 (ANNs using data of DP2 predicted the first non-freeze temperature as 1.12^{o}C and ANNs using data of DP3 predicted 0.38^{o}C while the observed temperature at that time was -1.13^{o}C. The accuracy of the predictions of the ANNs increased as the decision points were closer to a temperature event.

The results of a similar simulation using data from Alma from 1700 hours on January 7, 2002 to 1300 hrs on January 2002 are shown in Figure 2.8. DP1 for this simulation was at 1700 hours on Janaury 7, 2002. It was ten hours before the first freeze temperature which occurred at 0300 hours on January 8 (Figure 2.8). The freeze event was randomly selected and then data prior to this event was used to run the simulation. The ANNs using current observed data of DP1 failed to predict any freeze temperatures. In contrast, the ANNs using data of DP2 (at 2100 hours on January 7) successfully predicted most of the freeze temperatures. It was observed though that the ANNs predicted the first freezing temperature two hours prior, then predicted the subsequent temperature to rise above 0^{o}C and subsequently drop again to below 0^{o}C. The second time the ANNs predicted the temperature to drop below 0^{o}C is when the actual freezing temperature was observed. The first observed freeze temperature was -1.07^{o}C and the ANNs predicted a temperature of -0.98^{o}C (Figure 2.8). ANNs using current observed data at DP3 (at 0100 hours on

January 8) did not actually predict a below 0°C temperature for the first freeze temperature but predicted all below 0°C temperatures subsequently.

For this simulation, ANNs using data of DP1 and DP2 could not be used to predict the end of the freeze event as their scope ended before the freeze event was over. The scope of prediction for any DP is twelve hours after the DP. In this case the freeze event lasted longer than twelve hours after the DPs. However ANNs using data of DP3 were in a position to predict the end of the freeze. They accurately predicted the end of the freeze event. The first non-freezing temperature was 2.16°C and occurred at 1000 hours on January 8. ANNs using data of DP3 predicted 0.56°C for that time (Figure 2.8). None of the decision points accurately predicted the severity of the freeze event. The lowest temperature observed during the simulation was -4.1°C while the lowest temperatures predicted by the ANNs using data from DP2 was -1.54°C and the lowest temperature predicted by DP3 was 1.51°C.

A simulation was carried out using data from Blairsville. Figure 2.9 represents the simulation using data from Blairsville starting at 1900 hours on January 21, 2001 and ending at 1500 hours on January 22, 2001. DP1 was at 1900 hours on January 21, twelve hours prior to the first freezing temperature that occurred at 0700 hours on January 22. From this DP only the first freeze temperature could have been predicted and the ANNs using the current observed data from DP1 predicted it correctly. The first freeze temperature was observed to be -0.16°C and the ANNs predicted a temperature of -1.92°C (Figure 2.9). ANNs using data of DP2 (at 2300 hours on January 21) overestimate the severity and duration of the freeze events. The ANNs predicted the start of the freeze event two hours prematurely at 0500 hours on January 22. They predicted the lowest temperature to be -2.63°C while the lowest observed temperature was -1.16°C (Figure 2.9). The ANNs using data from DP2 accurately predicted the end of freeze. The first non-freeze temperature was observed at 1100 hours on January 22 with temperature of 0.92^{o}C. The ANNs predicted the first non-freeze temperature to be 0.89^{o}C at the same time (Figure 2.9). ANNs using data of DP3 (at 0300 hours on January 22) hovered very close to the 0^{o}C temperature. The first below 0^{o}C temperature was observed to be -0.16^{o}C and the ANNs predicted the temperature to be 0.06^{o}C at the time. ANNs using data of DP3 accurately predicted the time at which the freeze would end. They predicted their first non-freeze temperatures to be 1.15^{o}C (Figure 2.9). In this case the duration of the freeze itself was small (the freeze lasted for only about four hours with temperatures at the beginning and the end of the freeze hovering just below 0^{o}C) which could be an explanation for the ambiguity vis-à-vis the temperature predictions from DP3.

As was stated previously South Georgia experienced an unexpected late and severe frost in 2002. For example, in Fort Valley the last freezing temperatures were recorded as low -1.18^{o}C on March 23, 2002 and in Alma the last freeze temperatures were recorded as -1.39^{o}C on March 6, 2002. Both the locations are in the south Georgia region. Thus an analysis was conducted on predictions of the ANNs for the year 2002 for these two locations that are in the region. In the year 2002, Fort Valley experienced seventeen frost events of which the model failed to predict only one. Alma experienced twelve frost events and all were correctly predicted by the model.

SUMMARY AND CONCLUSION

ANN models were developed to predict hourly temperatures in hourly increments for three locations in Georgia. Experiments were conducted to determine the important weather variable inputs, which included temperature, relative humidity, solar activity and wind speed. It was found that the duration of prior data needed depended on the period of prediction as well as the location. For the prediction periods one to three hours for Fort Valley the duration of prior weather data with the smallest MAE was two hours and for other prediction periods it was six hours. In case of Alma and Blairsville, for prediction periods one and two hours the smallest MAEs were associated with two hours of prior data and for other prediction periods they were with six hours of prior data. The Ward ANNs produced the highest accuracy: the optimal number of hidden nodes varied between 5 hidden nodes per slab to 45 hidden nodes per slab depending on location and period of prediction. When an ANN model developed for a particular location was used to predict temperatures for another location, the accuracy was less than when using a model developed for the original location.

In a simulation of the proposed DSS for Fort Valley, Alma and Blairsville, the system predicted both the temperatures and the duration of freezing temperature conditions reasonably well. Future research will focus on developing a general ANN temperature predictor based on data from multiple locations. Such a general model is needed because a location may not have historical weather data to be able to develop a model.

ACKNOWLEDGEMENTS

This work was funded in part by the USDA Federal Crop Insurance Corporation Risk Management Agency and by State and Federal Funds allocated to Georgia Agricultural Experiment Stations Hatch projects GEO00877 and GEO00895.

REFERENCES

Abdelaal, R.E. and M. A. Elhadidy. 1994. A Machine Learning Approach To Modeling And Forecasting The Minimum Temperature At Dhahran, Saudi-Arabia, *Energy 19(7): 739-749*.

Blennow, K. and P. Persson. 1998. Modeling Local-Scale Frost Variations Using Mobile Temperature Measurements With a GIS, *Agricultural and Forest Meteorology 89: 59-71*.

Bruton, J.M., G. Hoogenboom and R.W. McClendon. 2000. Estimating Daily Pan Evaporation With Artificial Neural Networks, *Transactions of the ASAE 43(2): 491-496*.

Cook, D.F. and M.L. Wolfe. 1991. A Back Propagation Neural Network to Predict Average Air Temperature, *AI Applications 5(1): 40-46*.

Elizondo, D., G. Hoogenboom and R.W. McClendon. 1994. Development Of a Neural Network To Predict Daily Solar Radiation, *Agricultural and Forest Meteorology 71: 115-132*.

Felker, P. and H. Han. 1997. Estimation Of Daily Soil Water Evaporation Using An Artificial Neural Network*, Journal of Arid Environments 37:251-260*.

Georgiev, G. and G. Hoogenboom. 1998. Crop growth and yield estimation using current weather forecasts and weather observations, *p 69-72*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Georgiev, G. A. and G. Hoogenboom. 1999. Near real-time agricultural simulations on the web, *Simulation 73(1): 22-28*.

Hoogenboom, G. 1996. The Georgia Automated Environmental Monitoring Network, In: *Preprints of the 22nd Conf. On Agricultural and Forest Meteorology 343-346*, Boston, Mass., American Meteorological Society.

Hoogenboom, G. 2000a. Contribution of agrometeorlogy to the simulation of crop production and its applications, *Agricultural and Forest Meteorology 103(1-2): 137-157*.

Hoogenboom, G. 2000b. The Georgia Automated Environmental Monitoring Network, *p. 24-25*, In: *Preprints of the 24th Conference On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. Georgiev, D. Clarke, D.D. Gresham and G. Harbers. 1998. Internet Tools for delivery of weather data and applications, *p. 251-253*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. A. Georgiev, and D.D. Gresham. 2000. Development of weather based products for agricultural and environmental applications, *p. 66-67*, In: *Preprints of the 24th Conf. On Agricultural and Forest Meteorology*, Boston, Mass., American Meteorological Society.

Mehuys, G.R., N.K. Patni, S.O. Prasher and C.C. Yang. 1997. Application of Artificial Neural Networks for Simulation of Soil Temperature, *Transactions of the ASAE 40(3): 649-656*.

Mort, N. and C. Robinson. 1996. A Neural Network System For The Protection Of Citrus Crops From Frost Damage, *Computer and Electronics in Agriculture 16(3): 177-187*.

Okie W.R., G.L. Reighard, W.C. Newall Jr., C.J. Graham, D.J. Werner, A.A. Powell, G. Krewer and T.G. Beckman. 1998. Spring freeze damage to the 1996 peach and nectarine crop in the southeastern United States, *HortTechanology 8(3): 381-385.*

Powell, A.A. and D.G. Himelrick. 2003. Principles of freeze protection for fruit crops, http://www.aces.edu/dept/peaches/freeze.html (Verified November 18, 2003).

Tyson, AW. , R.E. Sneed, C.M. Mainland, K.B. Perry, E.B. Polling, D.C. Sanders and R Unrath. 2002. Frost/Freeze protection by Sprinkler Irrigation, www.smallfruits.org/Weather/frost_freeze.htm (Verified December 4,2003)

Ward System Group, Frederic, MD. 1993. Manual of NeuroShell 2. Frederic, MD: Ward Systems Group.

Table 2.1: Four hour temperature prediction for various combinations of inputs for Fort Valley, GA

Temperature[a]	Rel. Humidity[a]	Wind[a]	Solar Radiation[a]	Rainfall[a]	MAE* °C
X					1.41
X	X				1.19
X		X			1.34
X			X		1.35
X				X	1.44
X	X	X			1.15
X	X		X		1.17
X	X			X	1.21
X	X	X	X		1.12
X	X	X		X	1.16
X	X	X	X	X	1.14

* - The MAE was calculated only for observed temperatures below 5°C in the model evaluation set (data from 2001-02)

[a] – For any given term there are 4 prior data values (corresponding to 4 hours of prior data corresponding to 4 input nodes), 4 Δ values (corresponding to another 4 input nodes) and current value (corresponding to one more input nodes)

X – Denotes which terms were used

Table 2.2: Accuracy of one hour and twelve hour temperature prediction accuracies for various prior data durations, Fort Valley, GA, 15 minute data format

Duration for Prior Data (hours)	Prediction Period (hours)	MAE* (°C)
2	1	0.59
4	1	0.64
6	1	0.66
2	12	2.49
4	12	2.48
6	12	2.47

* - Model Evaluation data set, 2001-02 weather data

Table 2.3: The best results for each period of prediction for Fort Valley, GA

Prediction Period (hours)	Duration of Prior Data (hours)	MAE (^oC)
1	2	0.59
2	2	0.94
3	2	1.21
4	2	1.41
5	2	1.6
6	2	1.77
7	2	1.91
8	2	2.05
9	2	2.18
10	2	2.27
11	2	2.4
12	2	2.49

* - Model Evaluation data set, 2001-02 weather data in 15-minute format

Table 2.4: The best results when using only one model for all periods of prediction for Fort Valley, GA

Prediction Period (hours)	Duration of Prior Data (hours)	MAE* (^oC)
1	2	0.67
2	2	0.98
3	2	1.43
4	2	1.45
5	2	1.64
6	2	1.81
7	2	1.93
8	2	2.10
9	2	2.22
10	2	2.35
11	2	2.52
12	2	2.63

* - Model Evaluation data set, 2001-02 weather data in 15-minute format

Table 2.5: The best results for each period of prediction for Fort Valley, GA

Prediction Period (hours)	Duration of Prior Data (hours)	MAE* (^{o}C)
1	2	0.56
2	2	0.9
3	2	1.17
4	6	1.37
5	6	1.55
6	6	1.7
7	6	1.83
8	6	1.96
9	6	2.09
10	6	2.18
11	6	2.32
12	6	2.36

* - Model Evaluation data set, 2001-02 weather data in hourly format

Table 2.6: The best results for each period of prediction

Prediction Period (hours)	Alma		Blairsville	
	Duration of Prior Data (hours)	MAE* (^{o}C)	Duration of Prior Data (hours)	MAE* (^{o}C)
1	2	0.56	2	0.65
2	2	0.97	2	1.07
3	6	1.28	6	1.42
4	6	1.52	6	1.7
5	6	1.71	6	1.94
6	6	1.89	6	2.14
7	6	2.05	6	2.31
8	6	2.19	6	2.5
9	6	2.32	6	2.64
10	6	2.4	6	2.78
11	6	2.49	6	2.86
12	6	2.6	6	2.96

* - Model Evaluation data set, 2001-02 weather data in hourly format

Table 2.7: The best results for each period of predictions for final model evaluation data set, 2003 weather data in hourly format

Prediction Period (hours)	MAE (°C)		
	Alma	Blairsville	Fort Valley
1	0.53	0.62	0.54
2	0.93	1.01	0.88
3	1.2	1.33	1.16
4	1.44	1.63	1.36
5	1.58	1.9	1.56
6	1.77	2.13	1.7
7	1.87	2.31	1.87
8	2.05	2.47	2.0
9	2.15	2.6	2.13
10	2.25	2.79	2.24
11	2.32	2.88	2.4
12	2.44	2.98	2.41

Table 2.8: The accuracy of an ANNs developed for Fort Valley, Blairsville and Alma, and evaluated for weather data from the three locations

Evaluation Data	Prediction Period (hours)	Models		
		Alma (MAE* °C)	Blairsville (MAE* °C)	Fort Valley (MAE* °C)
Alma	1	**0.56**	0.83	0.61
	12	**2.60**	3.37	2.71
Blairsville	1	0.75	**0.65**	0.75
	12	3.27	**2.96**	3.28
Fort Valley	1	0.58	0.73	**0.56**
	12	2.44	2.98	**2.36**

* - Model Evaluation data set, 2001-02 weather data in hourly format

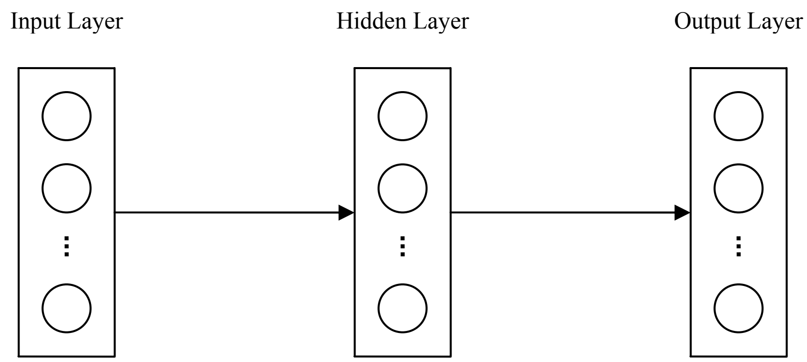

Figure 2.1: A Schematic representation of a Single Hidden Layer ANN.

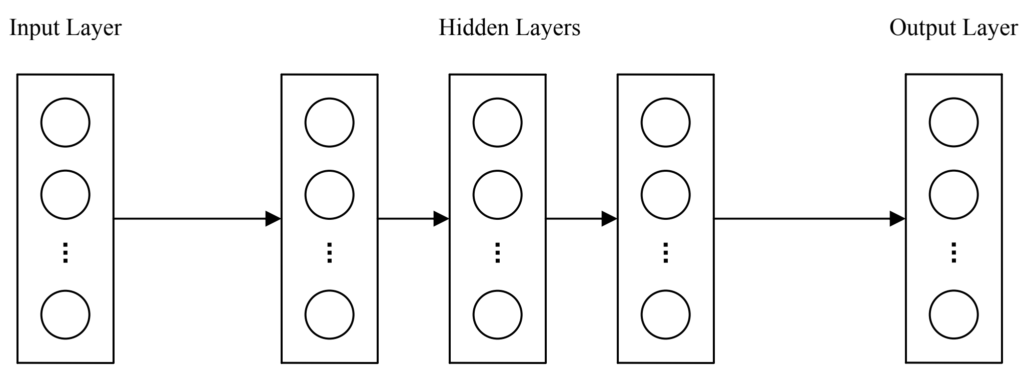

Figure 2.2: A Schematic representation of a Three Hidden Layer ANN.

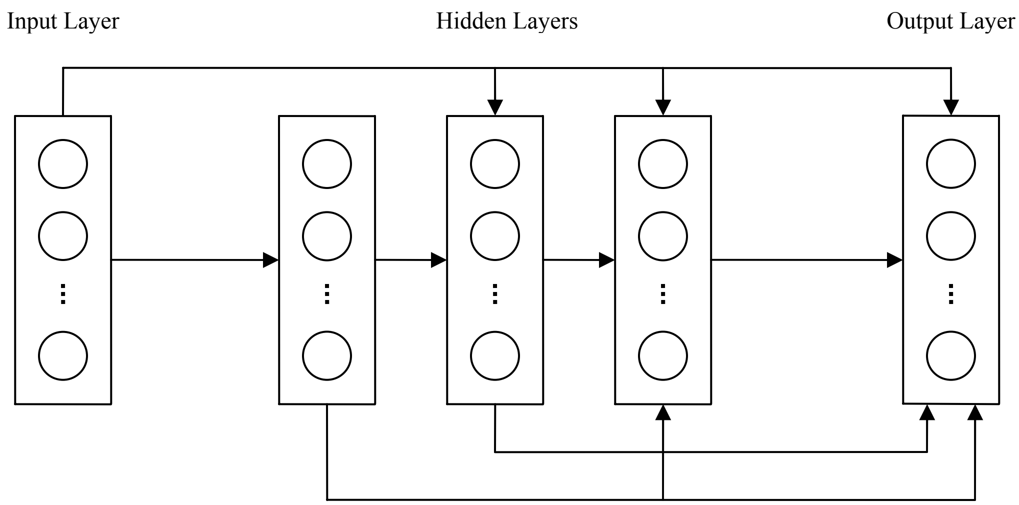

Figure 2.3: A Schematic representation of a Jump Connection ANN.

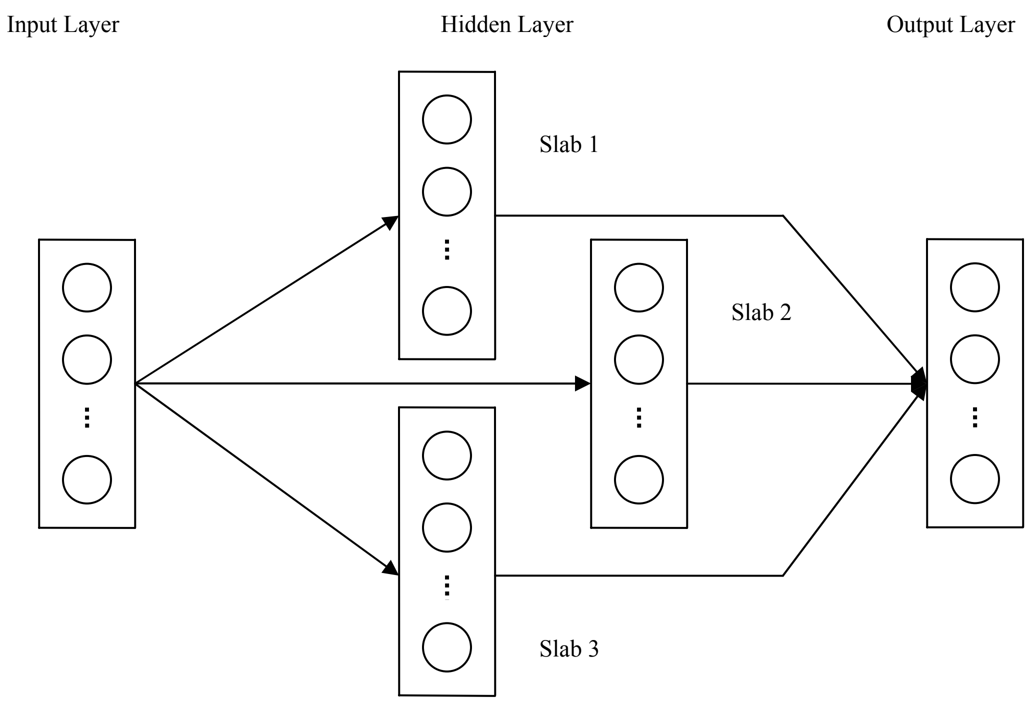

Figure 2.4: A Schematic representation of the Ward ANN.

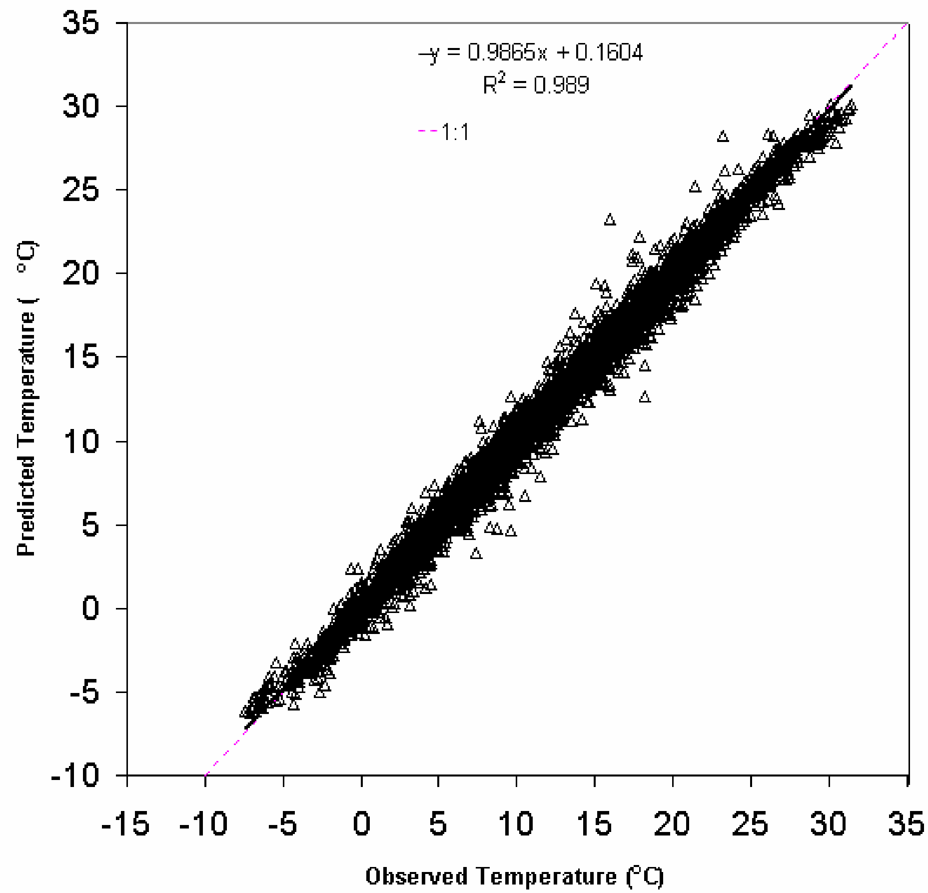

Figure 2.5: Predicted and observed one-hour temperature predictions for Fort Valley for 2001 and 2002.

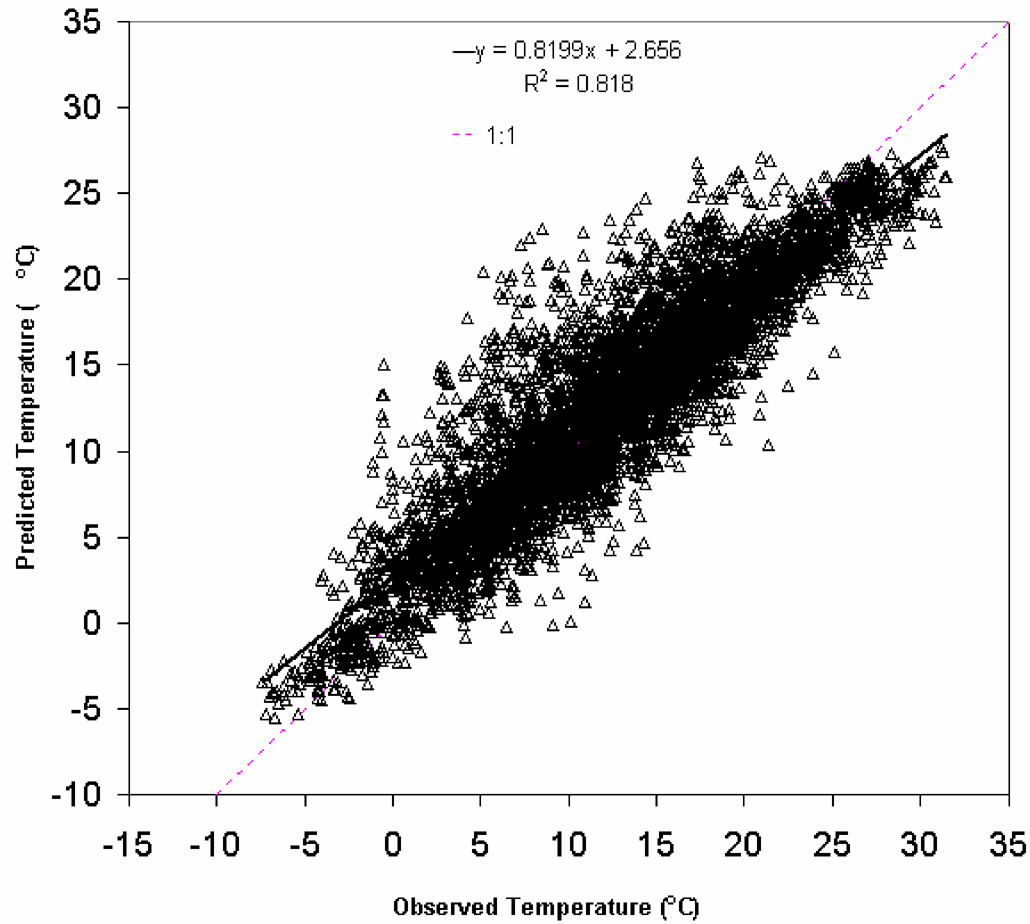

Figure 2.6: Predicted and observed 12 hour temperature predictions for Fort Valley for 2001 and 2002.

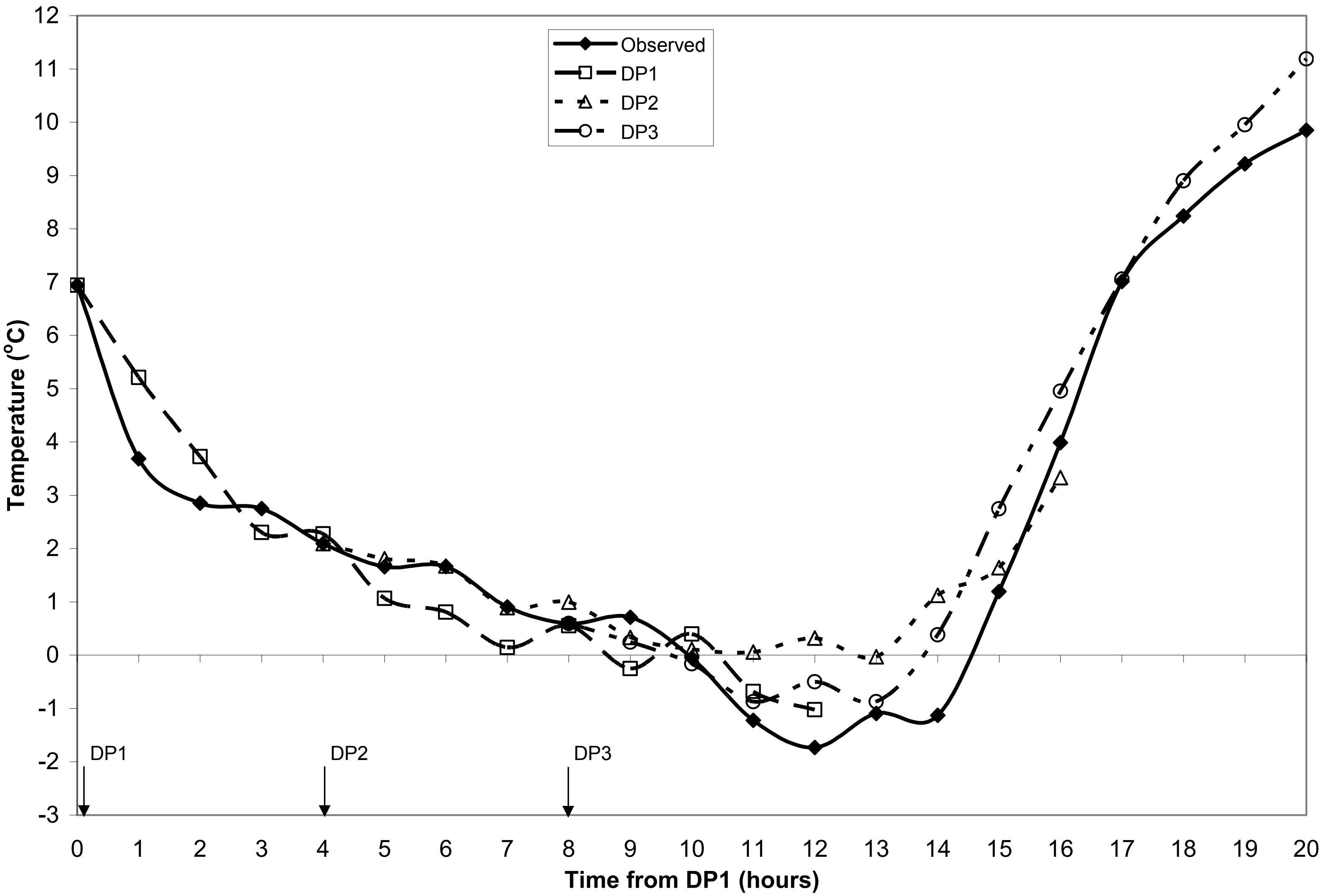

Figure 2.7: Observed and predicted temperatures, 1800 hrs on January 21, 2001 to 1400 hrs on January 22, 2001, Fort Valley, GA.

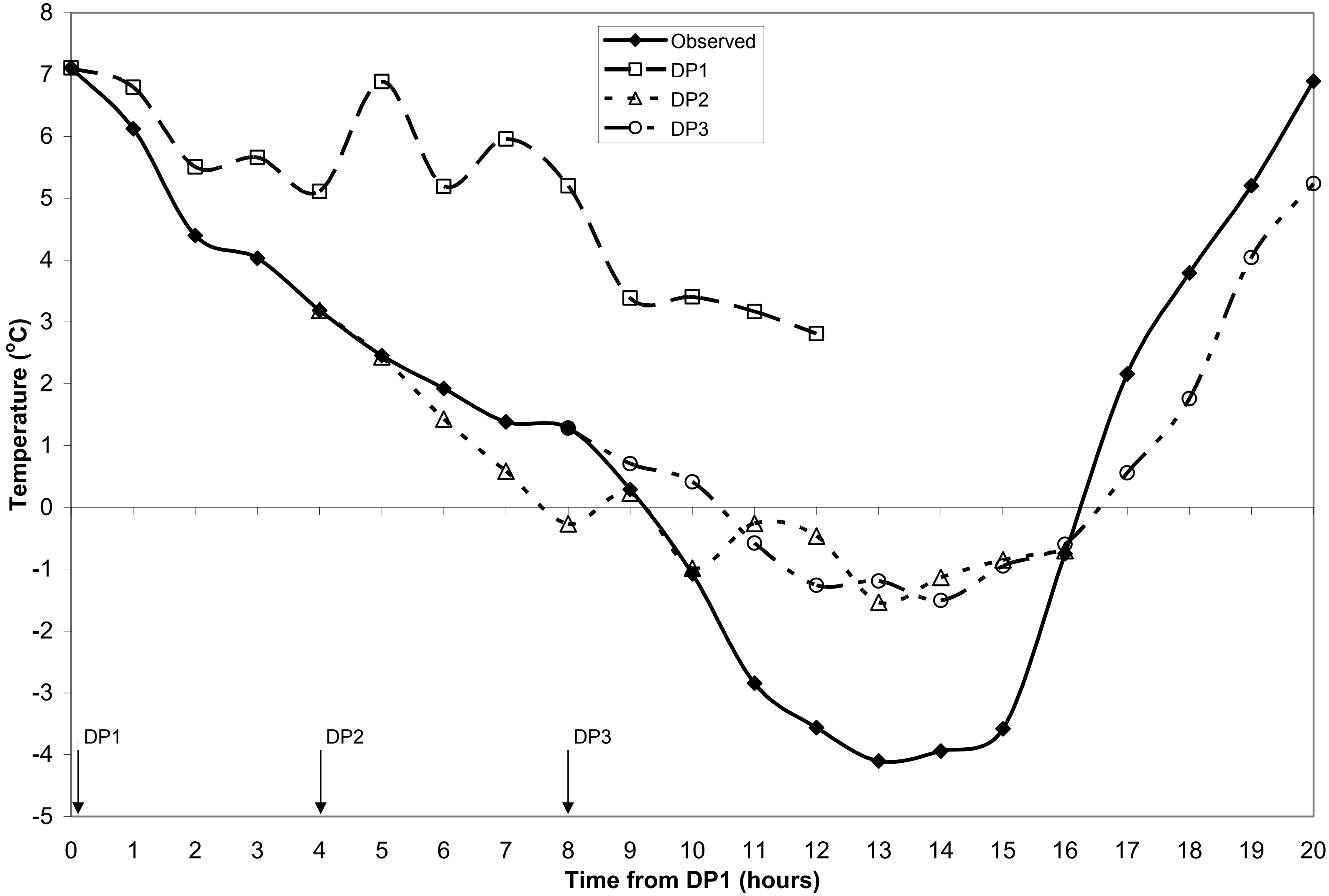

Figure 2.8: Observed and predicted temperatures, 1700 hrs on January 7, 2002 to 1300 hrs on January 8, 2002, Alma, GA.

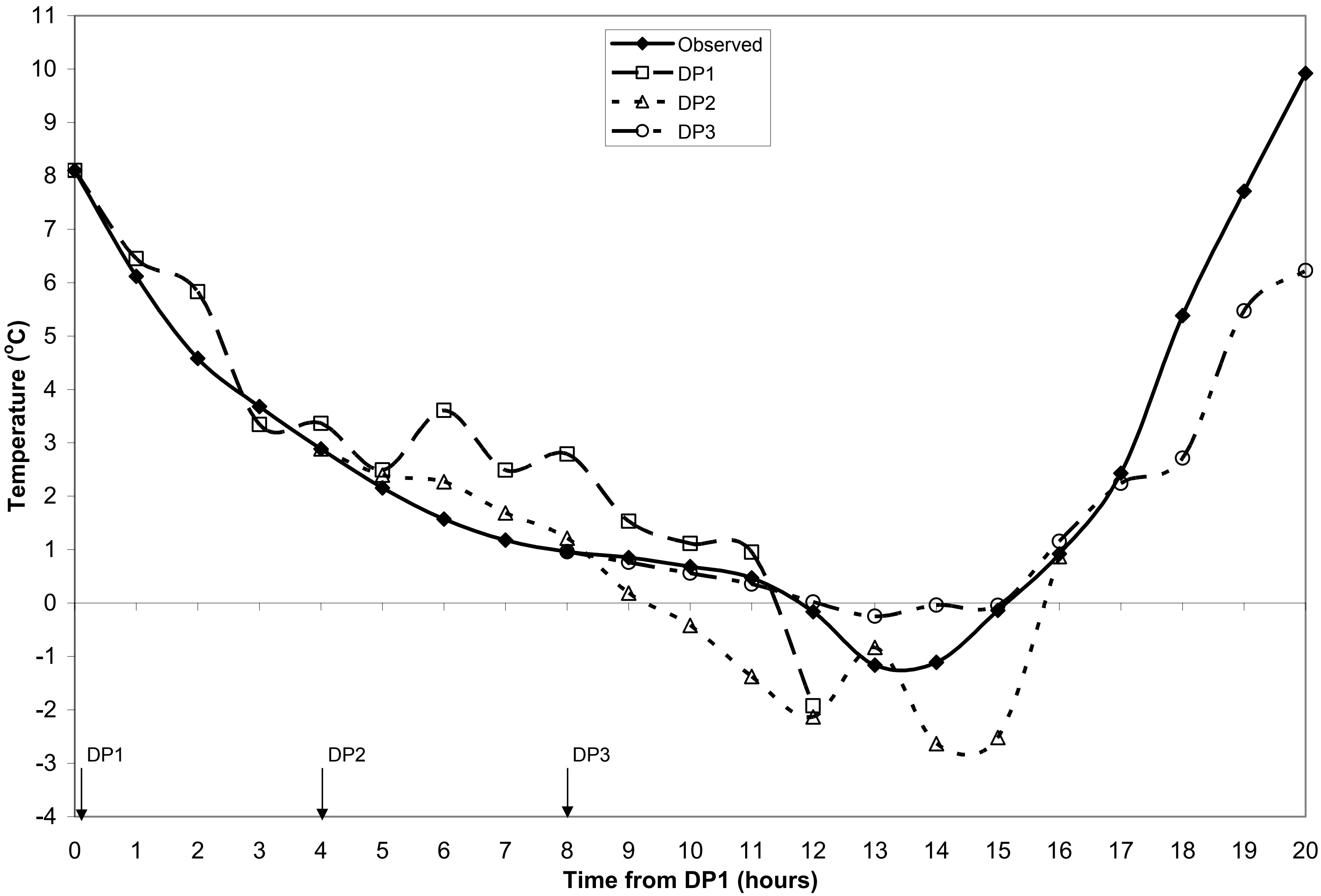

Figure 2.9: Observed and predicted temperatures, 1900 hrs on March 17, 2001 to 1500 hrs on March 18, 2001, Blairsville, GA.

CHAPTER 3

FROST PREDICTION USING ARTIFICIAL NEURAL NETWORK: A GENERAL MODEL

Jain, A., R.W. McClendon, G. Hoogenboom, to be submitted to the Agriculture and Forest Meteorology

ABSTRACT

The goal of this research was to develop Artificial Neural Networks (ANNs) to forecast temperatures in hourly increments from one to twelve hours for any location. ANNs were developed using data from multiple locations and evaluated on other locations. Thus the prime objective of this study was to determine if a general model can be developed to predict temperatures at a location without historical weather data using a model developed for multiple locations with historical data (assuming that the location in question has current weather data). It was found that the general models could predict temperatures for a given location without historical data with reasonable accuracy. For example, Byron, which lacked model development data, the Mean Absolute Error (MAE) associated with predicting temperatures one and twelve hours in the future were 0.62^{o}C and 2.51^{o}C, respectively. It was also found that the accuracy of the general model was comparable, if not better than the accuracy of the models that were developed exclusively for a given location. For example the MAE of the model developed for Dearing for predicting twelve hours in the future was 2.67^{o}C while the accuracy of the general model varied from 2.55^{o}C to 2.63^{o}C depending on the configuration of data chosen for model development. Additionally it was found that models that relied on data from locations spread across the state, rather than particular regions did slightly better.

INTRODUCTION

For agricultural production especially rain fed agricultural production systems weather is one of the most important factors that affect production. It accounts for up to 90% of the variability in the yield (Hoogenboom, 2000a). Low temperatures can harm crops such as blueberries and peaches. For these crops temperatures that are near freezing but not freezing will slow plant development. Such temperatures however do not typically cause significant damage to the plants if they are exposed to such temperatures only for a short duration. However, if the temperature is below freezing then the plants are easily damaged. The severity of damage is determined by how low the temperature was and how long were the plants exposed to such temperatures. There are other factors too that influence the severity of damage such as type of plant, variety, stage of development, amount of leaf cover and wind speed (Okie et al., 1998, Tyson et al., 2002 and Powell and Himelrick, 2003).

There are means of providing protection to the crops against frost. Some of the techniques that are used to counter frost are the use of wind machines that induce air movement or use of orchard heaters that heat the air in an orchard. However the most widely practiced technique for frost protection in southeastern US is irrigation. The water from the irrigation forms a layer of ice that keeps the temperature of the flower near freezing and prevents the temperature of the flower from dropping down further. This process however has to be initiated before the temperature drops below zero. Hence farmers need advance warning to initiate frost protection measures. Additionally, farmers need information about local wind speed, dew point or vapor pressure deficit which are also factors that determine when to initiate the freeze protection measures. Therefore, there is a need for accurate local weather information and short-term weather forecasts.

The National Weather Service (NWS) was the organization that provided weather forecasts. However, the NWS primarily collected data from urban areas, making it less useful for agricultural production decision making. Changes in the law prohibited the NWS from providing data for agriculture. The University of Georgia initiated the Georgia Automated Environmental Monitoring Network (AEMN) (Hoogenboom, 1996; 2000a; 2000b; Hoogenboom et al., 2000) as a response to this need. The AEMN is a network of over 57 automated weather stations that are located primarily in the rural areas of Georgia (www.Georgiaweather.net). Each weather station collects data every second on air temperature, relative humidity, soil temperature at 5 cm, 10 cm and 20 cm depths, wind speed, wind direction, solar radiation, vapor pressure deficit and soil moisture. Every 15 minutes the averages, or totals depending on the variable, are calculated and stored in a data logger. Daily summaries are also calculated. These two summaries are downloaded to a central computer in Griffin automatically. A website (www.Georgiaweather.net) disseminates the information collected by the AEMN program. The website also has simple calculators that can dynamically calculate degree days, chilling hours or water balance for management of irrigation (Georgiev and Hoogenboom, 1998; 1999; Hoogenboom et al., 1998). However, the AEMN system does not have a forecasting component.

An Artificial Neural Network (ANN) imitates the behavior of neurons in the brain. It is computational intelligence technique. The building blocks of an ANN are its nodes and the connections between the nodes. Nodes can be compared to neurons in a brain. A node is a computational unit that receives inputs, calculates a weighted sum and presents the sum to an activation function. The nodes are usually arranged in layers. In the Back Propagation ANN architecture the nodes of the input layer receive the inputs. They then compute and pass the results of their computations to the nodes of hidden layer by means of connections between the various nodes. The hidden layer nodes in turn sum these weighted values and use them as inputs. They then calculate activation levels and pass the results similarly to the nodes of output layer.

The usefulness of an ANN lies in learning or capturing the complex relationships between inputs and outputs. This is achieved by incrementally searching for an optimal set of the weights of the connections between the nodes. One set of inputs is presented to the ANN in the feed forward mode. The error between the ANN output and the expected output is calculated. This error is then used to adjust the weights of the connections by using the method of gradient descent. Data used in this process is called training data. Another set of data is then used as input for the ANN and the error calculated. This process first inputting training data, adjusting weights and then inputting testing data is repeated until the error on the testing data set reaches a minimum. The testing data set is used in the feed forward mode only and is not used to adjust the weights. Together the training data and testing data comprise the model development data set. Once an ANN has been developed using model development data, another set of data called model evaluation data set is used as input for the ANN. The accuracy of the ANN on model evaluation data set serves as a true measure of accuracy of the ANN. The model evaluation data is separate from the model development data.

The use of ANNs to estimate meteorological variables is widely practised. ANNs were used by Elizondo et al. (1994) for their study in which they estimated the daily solar radiation using ANNs. They used readily available daily meteorological data such as minimum and maximum air temperature, precipitation, day length and clear sky radiation, to create an ANN model to estimate the level of daily solar radiation. Felker and Han (1997) used ANNs to estimate daily soil evaporation. They used daily average relative air humidity, air temperature, wind speed and soil water content in a cactus field as inputs for a radial basis function ANN to estimate daily soil water evaporation. ANNs were also used by Bruton et al. (2000) to estimate daily pan evaporation. They used daily rainfall, temperature, relative humidity, solar radiation and wind speed as inputs. They also showed in their study that the ANNs compared favorably against other numerical methods. Mehuys et al. (1997) used ANNs to model behavior of soil temperature. In their study they used ANNs to simulate daily fluctuations of soil temperature at various depths.

ANNs have also been used to forecast air temperatures. Cook and Wolfe (1991) developed an ANN to predict long term temperatures. In their study they developed ANNs that predicted monthly average air temperature for one month, two months and three months in advance using monthly average daily maximum temperatures from the previous eleven months, the present month and appropriate months from previous two years, precipitation in the current month, relative monthly humidity at four times per day for current month and monthly daily average daily temperature as inputs. Their ANNs displayed a bias to over predict for earlier months and under predict for later months. Abdel-Aal and Elhadidy (1994) carried out a study to estimate the minimum temperature on a given day, one to three days in advance by creating a model that predicts the minimum temperature. Their inputs consisted of maximum air temperature, minimum air temperature, mean air temperature, maximum wind speed, mean wind speed, minimum wind speed, wind speed vector, wind direction, maximum barometric pressure, mean barometric pressure, minimum barometric pressure, maximum relative humidity, mean relative humidity, minimum relative humidity, prevailing wind direction, number of hours in the prevailing wind direction, direction of maximum wind, mean global solar radiation. They reported their model had 92.8% of the prediction under $3^{o}C$ MAE and had better accuracy than statistical methods.

Robinson and Mort (1996), have taken a different approach to predict frost. They chose to cast the problem as one of classification as freezing or non freezing rather than predicting the temperature. In their models the inputs consisted of previous days' minimum and maximum temperatures, cloud cover, maximum wind speed and direction, humidity, wind speed and wind direction at 1900 hrs. They chose gray code instead of continuous value variables to encode the variables and the output. The ANN developed in their study would classify the weather conditions that would lead to frost by outputting a set of binary values that were then interpreted by the same laws that were used in the training phase. A frost event was defined as any temperature below 1^{o}C. They reported that the best ANN predicted two false alarms and one failure over a course of a 50-day model evaluation set.

Li et al. (2004) conducted a study in which they developed ANNs that estimated historical daily solar radiation and minimum and maximum air temperatures. They used weather data from other neighboring locations to estimate the values of these variables. The inputs used were dependent on the variable that was being estimated. The authors reported that the results generated by the ANN were better than other statistical techniques.

Jain et al. (2003), developed site specific ANN models to estimate hourly air temperatures during the subsequent twelve hour period. The sites chosen were Blairsville located in the Georgia mountains, Fort Valley located in central Georgia and Alma located in south Georgia. The models developed for these locations were reasonably successful in predicting short term temperatures with Mean Absolute Errors (MAEs) varying between 0.56^{o}C and 2.96^{o}C, depending on the location and how far into the future the prediction was being made. However, there was a limitation that these models could not be used to predict temperatures for other areas, with the same amount of accuracy as that for the locations themselves. For example, when the ANN developed for Blairsville for predicting temperatures twelve hours in the future was used on the 2001-02 data from Fort Valley, the MAE was found to be 2.98^{o}C. The MAE for the same dataset when it was fed to the ANN developed for Fort Valley was 2.36^{o}C. Hence a need was felt to develop a general model that could predict short term temperatures for any location.

Other techniques have also been used to predict temperatures. Krasovitski et al. (1996) developed a numerical technique that used earth surface temperatures at the beginning of a night in order to predict earth temperature for the rest of the night. They then used these predictions to gauge the possibility of frost. The model was limited in intent as to only predict night time temperatures. Figuerola and Mazzeo (1997) also developed numerical models that could predict surface temperatures. Their model was again designed to predict temperatures at night and at dusk. A further limitation was that the model could be used only under clam clear sky conditions.

Another technique approach to frost forecasting is frost mapping. In this approach the objective is not to predict frost but only to find out where it could occur. Blennow and Persson (1998) using a Geographic Information System (GIS) and a stepwise linear regression model to determine where frost could occur in area measuring about 7.5km^2 covering both forests and heterogeneous region in Sweden. The model developed was able to explain between 87 and 88% of the air temperature variations in two separate validation sets.

ANNs are adept at capturing or learning relationships between outputs and inputs. In the studies that compared the accuracy of the ANNs against other methods the ANNs had better accuracy, as in the case of Abdel-Aal and El-Hadidy (1994), Li et al. (2004) and Bruton et al. (2000) when their models performed better than the statistical methods. Hence for this study ANNs were considered the best option as they have demonstrated ability to capture complex relationship between the various meteorological variables.

The overall goal of the research reported herein was to develop an ANN based Decision Support System (DSS) that would predict the occurrence of frost in a region by predicting hourly temperatures during the subsequent twelve hour period. The specific objectives of this research were to 1) determine the accuracy of a general model developed using historical data from multiple locations to predict temperatures for locations without historical data, 2) determine the number of locations needed maximize the accuracy of a general model for each period of prediction, 3) compare the accuracy of a general model developed with data from multiple locations with a model developed for a specific location.

MATERIALS AND METHODS

NeuroShell™ software (Ward System Group Inc. Frederic, MD, 1993) was used in this study to develop the ANNs. It has a user-friendly interface and provides an array of architectures to choose from. Ward ANN, as shown in Figure 3.1 was used for model development with 25 hidden nodes for each of the three hidden node slabs. It was found that minimal benefit accrued from optimizing for hidden nodes. The configuration of 25 nodes per hidden slab was chosen because it was found previously to be the most commonly occurring hidden node configuration for location specific models developed for Alma, Fort Valley and Blairsville (Jain et al., 2003) for predicting temperatures one to twelve hours in the future.

The weather data were obtained from the Georgia AEMN. The locations for consideration were selected to cover most of the state. Weather data collected prior to the year 2001 from Blairsville, Fort Valley, Alma, Attapulgus, Griffin, Arlington, Savannah, Midville and Plains were used for model development. Depending on the location, four to nine years of data were available for these locations as summarized in Table 3.1. Regardless of the location all model development data were prior to 2001. All locations had data from year 2001 onward except Homerville, which had data beginning in 2003 and Nahunta, which had data beginning on day 82 of 2002. The model development and model evaluation data were restricted to the first four months of a year, as these are the months when freezing conditions are likely to occur and cause damage to the crops. The Mean Absolute Error (MAE) was used as a measure of accuracy of the ANNs developed.The evaluation of the ANN models was based on the accuracy of the temperature predictions one, four, eight and twelve hours in the future

Our first objective was to determine the accuracy of a general model developed using historical data from multiple locations to predict temperatures for locations without historical data. Additionally these locations should be located in the fruit producing regions of south Georgia and should lack data prior to 2001. Since data prior to 2001 were used to develop ANNs, for the purpose of this study these locations are considered to lack historical data. The model development data set configurations and layouts were created as outlined in the following paragraphs.

The second objective of this study was to determine the number of locations needed maximize the accuracy of a general model for each period of prediction. Jain et al. (2003) showed that a model developed specifically for a particular location has lower accuracy when predicting temperatures at another location. For the study herein models were developed to predict the temperature for prediction periods of one, four, eight and twelve hours using data from two locations (two location data configuration). A configuration refers to how many locations were used in the model development data set. To decrease the chance that the result is biased due to unusual weather occurrences in a given year, five different data layouts were prepared for each configuration. A data layout consists of random sampling of years of data from the locations in a configuration, with total number of years restricted to nine years. Nine years of data were used in each layout because the earliest weather were from 1991, thus allowing for nine years of pre-2001 weather data. By keeping the number of years constant at nine years it was ensured that no layout or configuration was advantaged or disadvantaged by having more or less years of data available. From these five layouts (for a given configuration), five sets of ANNs are developed. A particular set of ANNs consisted of four ANNs that predict temperature one, four, eight and twelve hours in the future. All the ANNs in a set are developed using data from one particular layout.

Models were developed using three different two-location data configurations. Five sets of models were developed, using five different layouts as they were for all subsequent configurations. For the first configuration, data from Blairsville in the north and Fort Valley from the south were used. From Table 3.1, the data available for Fort Valley and Blairsville for model development was seven and eight years, respectively. But since a layout requires nine years of data for each layout, data from some years was from both locations, for example 2000 data might be from Blairsville as well as Fort Valley. The second two-location configuration models were developed using the data prior to 2001 from Alma and Blairsville. Similarly models were developed using the third two-location configuration data set. This configuration used pre-2001data from Alma and Fort Valley for model development. This was done to compare the accuracy of the models developed using data from same general region of the state.

To further investigate the second objective, ANNs that predict temperatures one, four, eight and twelve hours in the future were developed using data from four locations (four-location data configuration). Three such four-location model development data configurations were created. The first configuration consisted of model development data from Alma, Blairsville, Fort Valley and Savannah. The second configuration of model development data used data from Attapulgus, Blairsville, Fort Valley and Savannah. The third four-location configuration used data from Alma, Attapulgus, Fort Valley and Savannah. The final model development configuration consisted of data from nine locations (Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains and Savannah) which were used for developing models that predict temperature one, four, eight and twelve hours in the future.

The models developed were evaluated on the 2001 to 2003 data from Brunswick, Byron, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Homerville, Nahunta, Newton, Valdosta and Vidalia. These locations cover most of south Georgia region, the primary region of interest for the purpose of this study, and would give a fair indication of the accuracy of the models developed.

Once the ANNs (for the four prediction periods) were developed for each of the five layouts, evaluation data was presented to them in feed forward mode. Thus for each period of prediction there were now five MAEs. These were then averaged to give the MAE associated with a configuration for that period of prediction. At the end of this process the MAE associated with a particular configuration for a particular period of prediction was determined.

The MAEs obtained for the thirteen target locations from the various models developed using the seven configurations of model development data (three two-location model development data set configuration, three four-location model development data set configuration and one nine-location model development data set configuration) were averaged to determine the MAE for a given configuration for each prediction duration. If one configuration was found to have the lowest average MAE for all the locations then that configuration was chosen as the model development data configuration for that particular period of prediction.

To compare the accuracy of a general model developed with data from multiple locations with a model developed for a specific location, ANNs were developed to predict temperatures one, four, eight and twelve hours in the future, for the locations that served as the target locations i.e. Brunswick, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Newton, Valdosta and Vidalia. Models were not developed for Byron, Homerville and Nahunta as they lack data prior to 2001. The location specific ANNs were developed using the same Ward architecture of 25-25-25 hidden node configuration.

To determine the benefit of having some data from a potential target location included in the model development data, the general models developed during the previous stages were also presented with data from 2001 to 2003 for Fort Valley, Alma, Blairsville, Plains, Midville, Griffin, Arlington and Savannah. The locations listed here have been used to create the various general model development data configurations.

RESULTS AND DISCUSSION

The smallest MAE for predicting temperatures twelve hours in the future was 2.40°C for Dixie using two-location data models (Alma and Fort Valley data) and the largest MAE was 3.26°C for Nahunta using the two-location data models (Blairsville and Fort Valley) and (Table 3.2). When model development data set contained data from locations that were close to a target location, the MAE was smaller. For example, Cordele is close to Fort Valley and Plains and the smallest MAE for this location was 2.64°C with a standard deviation (SD) of 0.05 (Table 3.2). This configuration used four to five years of data from Fort Valley. Conversely the largest MAE for Cordele was 2.86°C with an SD of 0.05(Table 3.2). This configuration used data from Alma and Blairsville. Four to five years of data were from Blairsville which is located in north Georgia, a region which experiences different weather conditions than Cordele. Overall the two best configurations based on average MAE for all locations are the nine-location data models and the four-location data models (Alma, Blairsville, Fort Valley and Savannah) with MAE of 2.69^{o}C (Table 3.2). However based on the smaller SD of 0.16 associated with the four-location data models, it can be concluded that for period of prediction of twelve hours the four-location data configuration (Alma, Blairsville, Fort Valley and Savannah) is better than the other configurations. However on average for all the locations, there was no clear difference between the configurations except the two-location configuration which used data from Blairsville and Fort Valley.

For the eight hour period of prediction the smallest MAE observed for predicting temperatures eight hours in the future was 2.07^{o}C for Dixie using the nine-location data configuration and the largest MAE was 2.89^{o}C for Nahunta using the two-location data configuration (Blairsville and Fort Valley) (Table 3.3). It was observed, as in the case of Cordele discussed above, when model development data contains data from locations that are close to a target location the MAE is smaller. For example, for Camilla which, is close to Arlington, the smallest MAE is 2.20^{o}C (with SD of 0.02) associated with the nine-location data configuration which has one year of data from Arlington. Conversely the largest MAE for Camilla was 2.33^{o}C (with SD of 0.04) associated with the two-location configuration containing data from Blairsville and Fort Valley (Table 3.3). This configuration consisted of four to five years of data from Blairsville which is located in north Georgia, as well as, data from Fort Valley which is also far from Camilla. Overall, based on smallest MAE for all locations, the best configuration for the eight hour period of prediction is the nine-location data configuration with a MAE of 2.30^{o}C (Table 3.3). However the SD associated with this configuration is 0.19 and is not the smallest. The smallest SD of 0.15 is associated with the four-location data configuration containing data from Attapulgus, Blairsville, Fort Valley and Savannah, and has an MAE of 2.35^{o}C (Table 3.3).

Based on the lowest MAE and only marginally worse SD, it can be concluded that the nine-location configuration is the better configuration of the seven considered. It was also observed that on average for all the locations the MAEs were comparable except the nine-location model which had smaller MAE.

In the case of period of prediction of four hours, the smallest MAE was 1.44^{o}C for Dixie using the two-location data configuration using data from Alma and Fort Valley and the largest MAE was 1.96^{o}C for Nahunta using the two-location data configuration which had data from Blairsville and Fort Valley (Table 3.4). As was observed for the two previous periods of predictions when model development data contains data from locations that are close to a target location the MAE is smaller. In case of Newton which is closest to Arlington, the smallest MAE was 1.74^{o}C associated with the nine-location data configuration, with SD of 0.02 (Table 3.4). This configuration had one year of data from Arlington. Conversely the largest MAE for Newton was 1.81^{o}C (with SD of 0.03) associated with the two-location data configuration using data from Blairsville and Fort Valley in which both locations are located at a great distance from Newton (Table 3.4). Overall the smallest MAE averaged over all locations is associated with the nine-location data configuration. The MAE is 1.60 with an SD of 0.13 (Table 3.4). However the smallest SD is 0.11 (with MAE of 1.61^{o}C) which is associated with the four-location data configuration containing data from Attapulgus, Blairsville, Fort Valley and Savannah as shown in Table 3.4. Based on the lower MAE it can be concluded that the nine-location configuration is the better configuration of the seven considered. However, on average for all the locations there was only a slight variation in MAEs for all the configurations.

The smallest MAE observed for period of prediction of one hour was 0.54^{o}C for Vidalia using the two-location data configuration comprised of data from Alma and Fort Valley and the largest MAE was 0.72^{o}C for Nahunta using the two-location data configuration using data from Blairsville and Fort Valley as shown in Table 3.5. Again it was observed, that when model development data contains data from locations that are close to a target location the MAE is smaller. For Cordele, the smallest MAE was 0.60^{o}C (with SD of 0.01) associated with nine-location data configuration which includes data from both Fort Valley and Plains (Table 3.5). Conversely the largest MAE for Cordele was 0.62^{o}C associated with the two-location data configuration with data from Blairsville and Fort Valley (an SD of 0.01) as well as with four-location data configuration using data from Attapulgus, Blairsville, Fort Valley and Savannah (with SD of 0.02) (Table 3.5). Both these configurations gave a larger MAE despite including data from Fort Valley as they both had a large amount of data from locations that are too far from the target location (Blairsville in the case of former and Blairsville, Savannah in the case of the latter). Overall the smallest MAE averaged over all the locations was 0.62^{o}C. There are three configurations that have the same MAE, nine-location data configuration (with SD 0.03), two-location data configuration of Alma and Blairsville (with SD 0.04) and the two-location data configuration using data from Alma and Fort Valley (with SD 0.04). From this it can be concluded that based on the lowest MAE as well as the lowest SD associated with it, the nine-location data configuration is best suited for predicting temperature one hour in the future. But, on average for all the locations there was only a slightest variation in MAEs for all the configurations.

For three of the four periods of prediction the nine-location data configuration was the best configuration and in the one case where it was not, the MAE associated with it was equal to the best averaged MAE. Hence it can be concluded that the general model should be developed using the nine-location data configuration.

All the general models that were created used historical data from multiple locations to predict temperatures for locations such as Byron, Homerville and Nahunta, which did not have historical data. The MAE associated with the four periods of predictions for Byron using the nine-location data configuration was 2.49°C for predicting temperature twelve hours in the future, 2.20°C for predicting temperature eight hours in the future, 1.50°C for predicting four hours in the future and 0.62°C for predicting temperature one hour in the future. Similarly for Homerville, the MAE was 2.87°C, 2.45°C, 1.73°C and 0.63°C for predicting temperatures twelve, eight, four and one hour in the future, respectively. For Nahunta, the MAEs were 3.15°C, 2.68°C, 1.86°C and 0.67°C for predicting temperatures twelve, eight, four and one hour in the future. For locations where there was some data available for development of models specifically for that location, the MAE of the general models was comparable to the location specific models. For example, for the period of prediction of one, four, eight and twelve hours for Dearing the MAEs associated with the nine-location configuration were 0.61°C, 1.52°C, 2.17°C and 2.57°C, respectively, while the MAEs associated with the location specific models were 0.64°C, 1.54°C, 2.25°C and 2.67°C, respectively. Hence it can be concluded that a general model created by using historical data from multiple locations can be used to predict temperatures at locations that lack historical data with comparable accuracy to those models that are developed specifically for a location.

Notice that for the four periods of predictions, apart from the nine-location configuration, the best configuration based on MAE was different for each period of prediction. Hence it can be concluded that apart from the nine-location data models, the accuracy of the temperature predictions for the various periods of predictions varied with the data configuration chosen. It was also observed that if the target location was in close proximity to a location that had been used in the model development data set, then the accuracy of predictions for that target location generally improved.

It was observed that layouts that relied on data from locations from south Georgia for training did not do as well on four, eight and twelve hours prediction periods as the statewide general model that was developed from the data of nine locations spread out over the entire state of Georgia. For the period of prediction of one hour the regional models did only as good as the statewide general model. A possible explanation for this could be that by including data from the various regions of the state (Blairsville from Georgia mountains in the north and Savannah from coastal Georgia) the models were exposed to varied weather phenomena making the models robust.

The accuracy of the general models and the accuracy of the models developed specifically for individual locations were found to be comparable. However in some cases the location specific models gave higher accuracy. For example, referencing Table 3.3, the MAE of the location specific model developed for predicting temperatures eight hours in the future for Camilla was found to be 2.20^{o}C while the MAE of the best general model was found to be 2.19^{o}C. The opposite is also found to be true in some cases. For example, referencing Table 3.5, the MAE of the location specific model developed for predicting temperatures one hour in the future for Brunswick is found to have an MAE of 0.68^{o}C, while the MAE of the worst general model was found to be 0.64^{o}C. However, even in these cases the accuracies vary by small amounts.

In the nine-location data configuration models, for all the target locations which have only one year of data available for model development the MAE of the general model is smaller than that of location specific for periods of prediction. For example Brunswick, which only had one year of data available for model development, the MAE associated with the locations specific model for predicting temperature one hour in the future was 0.68^{o}C, while that associated with nine-location data models was 0.62^{o}C (Table 3.2). Similarly the location specific models had MAEs of 1.74^{o}C (Table 3.3), 2.50^{o}C (Table 3.4) and 2.98^{o}C (Table 3.5) associated with four, eight and twelve hours period of prediction, respectively. The MAEs for the same periods of predictions, for the nine-location data models, were 1.66^{o}C, 2.46^{o}C and 2.79^{o}C respectively. For those target locations that had three or more years of data available for model development, the four periods of predictions for all the locations had smaller MAEs than that of the general model.

Further experiments were conducted with data from Fort Valley wherein ANNs were created with only one, two, three, four, five, six and seven years of data to determine the improvement in accuracy with addition of years of data. It was observed that the MAE tends to become smaller as more years are added to the model development data. The reduction seems to level out at approximately four years of data. The MAEs associated with models that were developed with one or two years of data was slightly larger than the MAEs of the models that were developed with three or more years of data. For example for period of prediction of eight hours the MAE was 2.2^{o}C and 2.1^{o}C for models with one and two years of data but remained 2.0^{o}C for all other models with larger number of years of data (Table 3.6).

To determine the benefit of including some data from a potential target location into model development data, experiments were conducted as outlined in the methodology section. The smallest MAE from the general model for predicting temperatures twelve hours in the future for Alma was found to be 2.62^{o}C (Table 3.7). This model used four to five years of data from Alma, depending on the layout, as part of its model development of data. In comparison the largest MAE for the same period of prediction for Alma was 2.80^{o}C for a model that did not use any data from Alma as part of the model development data set (Table 3.7). Similarly for Blairsville the lowest MAE of 3.05^{o}C was observed for the data configuration that had four to five years of Blairsville data in the model development data set configuration, while the largest MAE of 3.28^{o}C is seen the configuration that does not use Blairsville data in the model development data set configuration (Table 3.7). For Fort Valley the same behavior is observed as well, in the configuration that does not use any data from Fort Valley the MAE was observed to be 2.48^{o}C while the configuration that has the largest amount of data from Fort Valley, the MAE was observed to be 2.38^{o}C (Table 3.7). Notice that when a configuration has four or five years of data from one particular location the accuracy of the predictions of other locations degrades. So when the layout includes data only from Alma and Fort Valley predictions for these locations increase their accuracy with MAEs of 2.62^{o}C and 2.38^{o}C respectively but the accuracy of the predictions for other locations decreases, for example the MAE for predictions for Blairsville from the same configuration is 3.28^{o}C. Hence it can be concluded that significant improvement in accuracy of predictions for a location can be achieved by including data from that location but this causes a decrease in accuracy of predictions for other locations. Effectively the model loses its generality by the inclusion of large of amount of data from any one location. This is further borne out by the fact, as listed in Table 3.8, for the four periods of predictions considered, only the nine-location data models had the smallest averaged MAEs associated with it for all the prediction durations.

Once it was determined that the nine location configuration was the best configurations, the best layout amongst the five layouts was chosen as the data set from which to train the ANNs that would predict temperatures at two, three, five, six, seven, nine, ten and eleven hours in the future. The best layout was again chosen on the basis of least MAE and SD. The models developed using the chosen layout had Ward architecture with the 25 hidden nodes per slab configuration.

The MAEs associated with all the periods of predictions for Byron when 2001 to 2003 data was used as input for the general model in the feed forward mode are listed in Table 3.9. Byron was selected as it is prototypical of a location for which the general model has been developed. It has no historical data prior to 2001 and is located in a fruit producing region of Georgia. Also listed in the table is a parameter called Mean Absolute Difference (MAD). This parameter is a measure of how much, on an average, temperature can vary in space of one to twelve hours. It was calculated by determining the absolute difference between current observed temperature and the observed temperature at a time interval 't' in the future. The absolute difference was then averaged over the three years of data starting from 2001 to 2003 to determine MAD. The ratio of MAE:MAD serves as an additional measure of accuracy of the ANN with smaller the ratio the more accurate the model. This was done to put the MAEs of the various periods of predictions in the larger perspective of over all temperature variability over fixed periods of time. As can be seen long term the predictions are much more accurate using the MAE/MAD ratio, with the lowest ratio 0.37 for seven to eleven hours periods of predictions and the highest ratio of 0.61 associated with one hour period of prediction. Figure 3.2 graphically represents the results that have been tabulated in Table 3.9. From the graph it becomes apparent that the MAE rises rapidly at short periods of predictions and then rises slowly at longer periods of predictions.

For Dixie it was again observed that short term predictions have a higher ratio than long term predictions with the lowest ratio being observed as 0.35 for eight and eleven hour periods of predictions and the highest ratio of 0.58 being observed for one hour period of prediction (Table 3.10). Figure 3.3 represents these results graphically. The same behavior as seen for Byron is observed for Dixie as well. The MAE rises quickly at smaller period of predictions and rises slowly at longer periods of predictions.

Homerville only has data available from 2003. Despite having less data Homerville was selected as it is in a fruit producing region of south Georgia and is thus prototypical of a target location for the general model. Lower ratios occur at longer periods of prediction with the lowest ratio of 0.39 occurring at ten hour period of prediction and the highest occurring at one hour period of prediction (Table 3.11). The results are graphically represented in Figure 3.4. Here too the MAE rises quickly at shorter periods of predictions and the rate of rise slows down at longer periods of prediction.

Nahunta was selected as it too is located in the fruit crop producing region of south Georgia and does not have historical data. Nahunta has data available from day 82 of 2002 and all of 2003. The ratio is again higher at shorter periods of predictions with the highest being 0.61 for one hour period of prediction and smaller at longer periods of predictions with lowest being 0.41 for ten and eleven hours period of predictions (Table 3.12). Figure 3.5 has the graphical representation of the results. Notice that Nahunta has higher MAD than the other locations, this could be due to an unstable climate, which in turn could explain the low accuracy of the general model for predicting temperatures at Nahunta.

To gauge the accuracy of the general model in a paradigm of false positives and false negatives and exercise was conducted on 2003 data from Fort Valley using the models developed for predicting temperatures one, four, eight and twelve hours in the future. It was found that there were no false positives (i.e. false alarms) and there were no false negatives (i.e. missed freeze events).

SUMMARY AND CONCLUSIONS

To determine the first objective ANN models were developed to predict temperature at a location without historical weather data using a model developed with data from locations with historical data. The general models were found to be reasonably accurate. The MAE for the twelve, eight, four and one hour period of prediction varied from 2.40°C to 3.26°C, 2.07°C to 2.89°, 1.44°C to 1.96°C and 0.54°C to 0.72°C, respectively, depending on the location and the configuration chosen.

It was determined that the nine-location data configuration was the best configuration compared to the configurations that had data from four and two locations. The nine-location configuration consisted of data from Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains and Savannah. The MAEs averaged over the thirteen location not used in model development were 2.69°C, 2.30°C, 1.60°C and 0.62°C, for the periods of prediction of twelve, eight, four and one hour respectively.

It was also observed that accuracy of the general models was comparable to the models that were developed specifically for a location. For example, for Dearing the MAEs associated with the location specific models were 0.64°C, 1.54°C, 2.25°C and 2.67°C, for the period of prediction of one, four, eight and twelve hours, respectively, while the MAEs associated with the nine-location configuration were 0.61°C, 1.52°C, 2.17°C and 2.57°C, for the corresponding periods of prediction.

Future research will focus on developing hidden node optimized ANN models that can predict temperature for all durations starting from one hour to twelve hours in the future. Once the ANNs have been developed they would be incorporated into the pre-existing web-based information dissemination programs, where they can be used as a Decision Support Tool to aid farmers in protecting their crops from frost damage.

ACKNOWLEDGEMENTS

This work was funded in part by the USDA Federal Crop Insurance Corporation Risk Management Agency and by State and Federal Funds allocated to Georgia Agricultural Experiment Stations Hatch projects GEO00895 and GEO00877.

REFERENCES

Abdelaal, R.E. and M. A. Elhadidy. 1994. A machine learning approach to modeling and forecasting the minimum temperature at Dhahran, Saudi-Arabia, *Energy 19(7): 739-749*.

Blennow, K. and P. Persson. 1998. Modeling local-scale frost variations using mobile temperature measurements with a GIS, *Agricultural and Forest Meteorology 89: 59-71*.

Bruton, J.M., G. Hoogenboom and R.W. McClendon. 2000. Estimating daily pan evaporation with artificial neural networks, *Transactions of the ASAE 43(2): 491-496*.

Cook, D.F. and M.L. Wolfe. 1991. A back propagation neural network to predict average air temperature, *AI Applications 5(1): 40-46*.

Elizondo, D., G. Hoogenboom and R.W. McClendon. 1994. Development of a neural network to predict daily solar radiation, *Agricultural and Forest Meteorology 71: 115-132*.

Felker, P. and H. Han. 1997. Estimation of daily soil water evaporation using an artificial neural network*, Journal of Arid Environments 37:251-260*.

Figuerola P.I. and N.A. Mazzeo. 1997. An analytical model for the prediction of nocturnal and dawn surface temperatures under calm, clear sky conditions, *Agriculture and Forest Meteorology, 85(3-4):229-237*

Georgiev, G. and G. Hoogenboom. 1998. Crop growth and yield estimation using current weather forecasts and weather observations, *p 69-72*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Georgiev, G. A. and G. Hoogenboom. 1999. Near real-time agricultural simulations on the web, *Simulation 73(1): 22-28*.

Hoogenboom, G. 1996. The Georgia automated environmental monitoring network, In: *Preprints of the 22nd Conf. On Agricultural and Forest Meteorology 343-346*, Boston, Mass., American Meteorological Society.

Hoogenboom, G. 2000a. Contribution of agrometeorlogy to the simulation of crop production and its applications, *Agricultural and Forest Meteorology 103(1-2): 137-157*.

Hoogenboom, G. 2000b. The Georgia automated environmental monitoring network, *p. 24-25*, In: *Preprints of the 24th Conference On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. Georgiev, D. Clarke, D.D. Gresham and G. Harbers. 1998. Internet tools for delivery of weather data and applications, *p. 251-253*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. A. Georgiev, and D.D. Gresham. 2000. Development of weather based products for agricultural and environmental applications, *p. 66-67*, In: *Preprints of the 24th Conf. On Agricultural and Forest Meteorology*, Boston, Mass., American Meteorological Society.

Jain, A., R.W. McClendon and G. Hoogenboom. 2003. Frost prediction using artificial neural networks: A temperature prediction approach, ASAE Paper no. 033075, ASAE, St. Joseph, MI.

Krasovitski, B., E. Kimmel and I. Amir. 1996. Forecasting earth surface temperature for the optimal application of frost protection methods, *Journal of Agricultural Engineering Research 63(2):93-102*

Li, B. R.W. McClendon and G. Hoogenboom. 2004. Spatial interpolation of weather data for multiple locations using artificial neural networks, *Transactions of the ASAE.* (in revision)

Mehuys, G.R., N.K. Patni, S.O. Prasher and C.C. Yang. 1997. Application of artificial neural networks for simulation of soil temperature, *Transactions of the ASAE 40(3): 649-656*.

Mort, N. and C. Robinson. 1996. A neural network system for the protection of citrus crops from frost damage, *Computer and Electronics in Agriculture 16(3): 177-187*.

Okie W.R., G.L. Reighard, W.C. Newall Jr., C.J. Graham, D.J. Werner, A.A. Powell, G. Krewer and T.G. Beckman. 1998. Spring freeze damage to the 1996 peach and nectarine crop in the southeastern United States, *HortTechanology 8(3): 381-385.*

Powell, A.A. and D.G. Himelrick. 2003. Principles of freeze protection for fruit crops, http://www.aces.edu/dept/peaches/freeze.html (Verified November 18, 2003).

Tyson, AW. , R.E. Sneed, C.M. Mainland, K.B. Perry, E.B. Polling, D.C. Sanders and R Unrath. 2002. Frost/Freeze protection by sprinkler irrigation, www.smallfruits.org/Weather/frost_freeze.htm (Verified August 13, 2002).

Ward System Group, Frederick, MD. 1993. Manual of NeuroShell 2. Frederic, MD: Ward Systems Group.

Table 3.1: Data usage for locations

Locations	Pre- 2001 Data [a]	Usage[b]
Alma	7	D
Arlington	4	D
Attapulgus	8	D
Blairsville	8	D
Brunswick	1	E
Byron	0	E
Cairo	3	E
Camilla	3	E
Cordele	3	E
Dearing	2	E
Dixie	2	E
Dublin	3	E
Fort Valley	7	D
Griffin	9	D
Homerville	0	E
Midville	9	D
Nahunta	0	E
Plains	8	D
Savannah	8	D
Valdosta	3	E
Vidalia	3	E

[a] – Number of years of data available for model development prior to the year 2001

[b] – 'D' is used to designate the data from a location was used for model development and 'E' is used to designate the data was used for model evlauation

Table 3.2: Evaluation of the twelve hour temperature predictions with independent data for 2001 to 2003 based on different model configurations.

Locations	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]		Location Spf. [h]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE (°C)	Years [i]
Brunswick	2.79	0.07	2.82	0.08	2.78	0.04	**2.77**	0.03	2.86	0.09	2.89	0.07	2.85	0.05	2.98	1
Byron	2.49	0.03	2.50	0.02	2.50	0.03	2.49	0.03	2.49	0.03	2.50	0.02	**2.47**	0.04		0
Cairo	2.54	0.05	2.57	0.04	2.53	0.04	**2.51**	0.03	2.56	0.05	2.65	0.04	2.54	0.02	2.63	3
Camilla	2.58	0.04	2.60	0.05	2.61	0.04	2.59	0.03	**2.57**	0.03	2.69	0.05	2.60	0.01	2.59	3
Cordele	2.68	0.02	2.66	0.05	2.79	0.05	2.80	0.09	**2.64**	0.05	2.77	0.05	2.86	0.05	2.64	3
Dearing	2.57	0.04	**2.55**	0.02	2.63	0.04	2.61	0.03	**2.55**	0.02	2.60	0.02	2.59	0.04	2.67	2
Dixie	2.44	0.03	2.43	0.04	2.47	0.05	2.45	0.03	**2.40**	0.03	2.52	0.06	2.45	0.02	2.44	2
Dublin	2.92	0.08	3.04	0.06	**2.85**	0.03	2.87	0.03	3.06	0.05	3.03	0.03	2.94	0.05	2.70	3
Homerville	2.87	0.05	2.92	0.06	2.83	0.05	2.79	0.05	2.90	0.03	3.02	0.07	**2.78**	0.04		0
Nahunta	3.15	0.05	3.17	0.08	3.09	0.08	**3.01**	0.07	3.15	0.05	3.26	0.08	3.03	0.06		0
Newton	2.85	0.07	2.95	0.07	**2.79**	0.03	2.81	0.03	2.95	0.06	2.95	0.03	2.85	0.04	3.02	1
Valdosta	2.56	0.03	2.50	0.05	2.60	0.05	2.62	0.07	**2.49**	0.04	2.62	0.04	2.60	0.02	2.60	3
Vidalia	2.55	0.04	**2.52**	0.05	2.60	0.07	2.64	0.05	2.54	0.04	2.61	0.03	2.65	0.02	2.53	3
AVG. **	2.69	0.21	2.71	0.24	2.70	0.18	2.69	0.16	2.70	0.24	2.78	0.23	2.71	0.18		

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah
[b] - Alma, Attapulgus, Fort Valley, Griffin
[c] - Attapulgus, Blairsville, Fort Valley, Savannah
[d] - Alma, Blairsville, Fort Valley, Savannah
[e] - Alma, Fort Valley
[f] - Blairsville, Fort Valley
[g] - Alma, Blairsville
[h] - Location Specific models used all years of data available for the location, hence no multiple layouts
[i] - No. of years of data availbale for model development for a location
* - MAE calculated by averaging all the MAEs for a location over all layouts of a given configuration
** - Average MAEs for all locations in the table for all layouts of a configuration

Table 3.3: Evaluation of the eight hour temperature predictions with independent data for 2001 to 2003 based on different model configurations.

Locations	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]		Location Spf. [h]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE (°C)	Years [i]
Brunswick	2.46	0.02	2.47	0.02	2.48	0.05	**2.45**	0.10	2.53	0.04	2.53	0.06	2.58	0.03	2.50	1
Byron	**2.08**	0.02	2.12	0.03	2.15	0.04	2.13	0.05	2.11	0.03	2.12	0.02	2.15	0.05		0
Cairo	**2.16**	0.01	2.24	0.03	2.21	0.02	2.21	0.01	2.22	0.02	2.29	0.03	2.24	0.03	2.20	3
Camilla	**2.20**	0.02	2.27	0.04	2.27	0.04	2.27	0.03	2.22	0.01	2.33	0.04	2.28	0.03	2.19	3
Cordele	2.23	0.03	2.28	0.04	2.35	0.03	2.33	0.04	**2.22**	0.02	2.38	0.05	2.44	0.05	2.20	3
Dearing	**2.17**	0.01	**2.17**	0.02	2.28	0.01	2.25	0.03	2.19	0.03	2.21	0.03	2.25	0.03	2.25	2
Dixie	**2.07**	0.03	2.12	0.04	2.16	0.04	2.14	0.03	2.08	0.01	2.18	0.04	2.15	0.04	2.06	2
Dublin	2.52	0.03	2.64	0.05	**2.45**	0.03	2.48	0.03	2.65	0.04	2.62	0.03	2.60	0.04	2.33	3
Homerville	2.45	0.03	2.56	0.07	**2.44**	0.05	2.45	0.06	2.50	0.02	2.67	0.06	2.51	0.05		0
Nahunta	2.68	0.05	2.79	0.07	**2.66**	0.08	**2.66**	0.07	2.71	0.03	2.89	0.07	2.74	0.06		0
Newton	2.49	0.02	2.60	0.06	**2.47**	0.03	2.51	0.04	2.59	0.03	2.57	0.03	2.58	0.05	2.65	1
Valdosta	2.18	0.03	2.19	0.03	2.31	0.08	2.31	0.06	**2.15**	0.03	2.29	0.03	2.29	0.04	2.23	3
Vidalia	2.17	0.05	**2.14**	0.03	2.26	0.02	2.25	0.06	2.19	0.03	2.20	0.02	2.26	0.02	2.18	3
AVG. **	2.30	0.19	2.35	0.22	2.35	0.15	2.34	0.16	2.34	0.22	2.41	0.23	2.39	0.19		

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah
[b] - Alma, Attapulgus, Fort Valley, Griffin
[c] - Attapulgus, Blairsville, Fort Valley, Savannah
[d] - Alma, Blairsville, Fort Valley, Savannah
[e] - Alma, Fort Valley
[f] - Blairsville, Fort Valley
[g] - Alma, Blairsville
[h] - Location Specific models used all years of data available for the location, hence no multiple layouts
[i] - No. of years of data availbale for model development for a location
* - MAE calculated by averaging all the MAEs for a location over all layouts of a given configuration
** - Average MAEs for all locations in the table for all layouts of a configuration

Table 3.4: Evaluation of the four hour temperature predictions with independent data for 2001 to 2003 based on different model configurations.

Locations	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]		Location Spf. [h]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE (°C)	Years [i]
Brunswick	**1.66**	0.05	1.71	0.03	1.69	0.04	1.68	0.02	1.71	0.02	1.70	0.06	1.73	0.03	1.74	1
Byron	**1.50**	0.02	1.52	0.03	1.54	0.02	1.54	0.02	**1.50**	0.01	1.53	0.03	1.54	0.02		0
Cairo	**1.53**	0.02	1.54	0.02	1.55	0.01	1.55	0.02	**1.53**	0.02	1.60	0.04	1.54	0.01	1.55	3
Camilla	1.55	0.01	1.55	0.02	1.57	0.02	1.56	0.03	**1.54**	0.02	1.61	0.03	**1.54**	0.01	1.55	3
Cordele	1.54	0.03	1.51	0.03	1.54	0.02	1.55	0.05	**1.50**	0.02	1.57	0.04	1.55	0.02	1.37	3
Dearing	1.52	0.02	**1.51**	0.02	1.56	0.02	1.57	0.03	**1.51**	0.01	1.54	0.03	1.54	0.03	1.54	2
Dixie	1.46	0.01	1.45	0.02	1.48	0.03	1.48	0.03	**1.44**	0.03	1.51	0.03	1.45	0.01	1.49	2
Dublin	1.74	0.02	1.82	0.04	**1.72**	0.02	**1.72**	0.02	1.82	0.02	1.81	0.03	1.78	0.02	1.67	3
Homerville	1.73	0.02	1.73	0.03	1.71	0.03	**1.69**	0.02	1.71	0.03	1.80	0.03	1.70	0.02		0
Nahunta	1.86	0.03	1.87	0.03	1.84	0.02	**1.80**	0.03	1.85	0.04	1.96	0.04	1.84	0.03		0
Newton	**1.74**	0.02	1.79	0.03	1.76	0.02	1.76	0.03	1.79	0.03	1.81	0.03	1.79	0.02	1.83	1
Valdosta	1.52	0.02	**1.48**	0.03	1.54	0.03	1.57	0.06	**1.48**	0.03	1.56	0.04	1.52	0.02	1.51	3
Vidalia	1.47	0.02	1.47	0.03	1.51	0.03	1.52	0.04	**1.46**	0.03	1.49	0.04	1.50	0.03	1.51	3
AVG. **	1.60	0.13	1.61	0.15	1.61	0.11	1.62	0.10	1.60	0.14	1.65	0.14	1.62	0.13		

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah
[b] - Alma, Attapulgus, Fort Valley, Griffin
[c] - Attapulgus, Blairsville, Fort Valley, Savannah
[d] - Alma, Blairsville, Fort Valley, Savannah
[e] - Alma, Fort Valley
[f] - Blairsville, Fort Valley
[g] - Alma, Blairsville
[h] - Location Specific models used all years of data available for the location, hence no multiple layouts
[i] - No. of years of data availbale for model development for a location
* - MAE calculated by averaging all the MAEs for a location over all layouts of a given configuration
** - Average MAEs for all locations in the table for all layouts of a configuration

Table 3.5: Evaluation of the one hour temperature predictions with independent data for 2001 to 2003 based on different model configurations.

Locations	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]		Location Spf. [h]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE (°C)	Years [i]
Brunswick	**0.62**	0.01	0.63	0.02	0.63	0.02	0.63	0.01	0.64	0.01	0.64	0.01	0.63	0.01	0.68	1
Byron	**0.62**	0.01	0.63	0.01	0.64	0.02	0.64	0.01	**0.62**	0.00	0.64	0.01	0.63	0.01		0
Cairo	**0.60**	0.01	0.61	0.02	0.63	0.01	0.62	0.01	**0.60**	0.01	0.63	0.01	0.61	0.02	0.64	3
Camilla	0.63	0.01	0.63	0.02	0.65	0.02	0.64	0.01	**0.62**	0.01	0.66	0.01	0.63	0.02	0.64	3
Cordele	**0.60**	0.01	0.61	0.01	0.62	0.02	0.61	0.01	0.61	0.03	0.62	0.01	0.61	0.02	0.60	3
Dearing	**0.61**	0.00	0.62	0.01	0.64	0.02	0.63	0.01	**0.61**	0.00	0.63	0.01	0.62	0.01	0.64	2
Dixie	0.59	0.01	0.59	0.02	0.62	0.02	0.60	0.01	**0.58**	0.01	0.62	0.01	0.59	0.02	0.61	2
Dublin	**0.67**	0.01	0.69	0.01	**0.67**	0.01	0.67	0.01	0.70	0.01	0.68	0.01	**0.67**	0.02	0.68	3
Homerville	**0.63**	0.01	0.65	0.01	0.65	0.01	0.64	0.01	0.64	0.01	0.66	0.01	**0.63**	0.02		0
Nahunta	**0.67**	0.01	0.68	0.02	0.71	0.02	0.68	0.01	**0.67**	0.01	0.72	0.02	**0.67**	0.02		0
Newton	**0.66**	0.01	0.68	0.02	0.68	0.02	0.67	0.01	0.67	0.01	0.70	0.04	**0.66**	0.02	0.76	1
Valdosta	0.59	0.01	0.59	0.02	0.62	0.02	0.60	0.01	**0.57**	0.01	0.63	0.01	0.59	0.02	0.60	3
Vidalia	0.55	0.01	0.55	0.02	0.57	0.02	0.56	0.01	**0.54**	0.01	0.57	0.01	0.55	0.02	0.57	3
AVG. **	0.62	0.03	0.63	0.04	0.64	0.04	0.63	0.03	0.62	0.04	0.65	0.04	0.62	0.04		

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah

[b] - Alma, Attapulgus, Fort Valley, Griffin

[c] - Attapulgus, Blairsville, Fort Valley, Savannah

[d] - Alma, Blairsville, Fort Valley, Savannah

[e] - Alma, Fort Valley

[f] - Blairsville, Fort Valley

[g] - Alma, Blairsville

[h] - Location Specific models used all years of data available for the location, hence no multiple layouts

[i] - No. of years of data availbale for model development for a location

* - MAE calculated by averaging all the MAEs for a location over all layouts of a given configuration

** - Average MAEs for all locations in the table for all layouts of a configuration

Table 3.6: Evaluation of various number of years of data in model development for Fort Valley

Number of Year of Data	Period of Prediction			
	+12 MAE (^oC)	+8 MAE (^oC)	+4 MAE (^oC)	+1 MAE (^oC)
1	2.68	2.24	1.59	0.67
2	2.52	2.10	1.44	0.63
3	2.46	2.04	1.47	0.62
4	2.40	1.97	1.43	0.60
5	2.41	1.99	1.38	0.57
6	2.43	1.99	1.37	0.58
7	2.38	1.97	1.37	0.56

Table 3.7: Evaluation of the twelve hour temperature predictions with model development data for 2001 to 2003 based on different model configurations.

Locations	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]		Location Spf. [h]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE (°C)	Years [i]
Alma	2.64	0.03	2.64	0.06	2.69	0.03	2.64	0.07	**2.62**	0.04	2.80	0.06	2.65	0.03	2.55	7
Arlington	2.58	0.02	2.60	0.02	2.60	0.02	2.59	0.02	**2.57**	0.03	2.63	0.04	2.60	0.02	2.54	4
Blairsville	3.13	0.11	3.26	0.07	3.06	0.02	**3.05**	0.03	3.28	0.07	3.08	0.03	**3.05**	0.05	2.97	8
Fort Valley	2.41	0.02	**2.38**	0.01	2.48	0.03	2.44	0.06	**2.38**	0.03	2.42	0.02	2.48	0.03	2.38	7
Griffin	2.52	0.05	**2.46**	0.01	2.62	0.07	2.63	0.06	2.48	0.02	2.51	0.03	2.66	0.05	2.40	9
Midville	2.67	0.05	**2.65**	0.04	2.72	0.06	2.74	0.03	2.68	0.03	2.74	0.02	2.75	0.03	2.66	9
Plains	2.40	0.02	2.37	0.03	2.49	0.03	2.47	0.06	**2.35**	0.03	2.41	0.03	2.49	0.04	2.36	8
Savannah	3.08	0.07	3.24	0.07	2.97	0.01	**2.96**	0.06	3.27	0.06	3.28	0.07	3.12	0.06	2.69	8
AVG. **	2.68	0.27	2.70	0.34	2.70	0.20	2.69	0.21	2.70	0.35	2.73	0.30	2.72	0.23		

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah

[b] - Alma, Attapulgus, Fort Valley, Griffin

[c] - Attapulgus, Blairsville, Fort Valley, Savannah

[d] - Alma, Blairsville, Fort Valley, Savannah

[e] - Alma, Fort Valley

[f] - Blairsville, Fort Valley

[g] - Alma, Blairsville

[h] - Location Specific models used all years of data available for the location, hence no multiple layouts

[i] - No. of years of data availbale for model development for a location

* - MAE calculated by averaging all the MAEs for a location over all layouts of a given configuration

** - Average MAEs for all locations in the table for all layouts of a configuration

Table 3.8: Evaluations of various periods of predictions with model development data for 2001 to 2003 based on different model configurations.

Period of Prediction	9 Locations [a]		4 Locations [b]		4 Locations [c]		4 Location [d]		2 Locations [e]		2 Locations [f]		2 Locations [g]	
	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev	MAE* (°C)	Std. Dev
1	0.61	0.05	0.63	0.07	0.62	0.04	0.62	0.04	0.62	0.07	0.63	0.05	0.61	0.04
4	1.57	0.18	1.60	0.23	1.59	0.14	1.59	0.14	1.59	0.24	1.61	0.19	1.59	0.17
8	2.28	0.27	2.33	0.33	2.33	0.18	2.32	0.21	2.33	0.36	2.35	0.29	2.38	0.25
12	2.68	0.27	2.70	0.34	2.70	0.20	2.69	0.21	2.70	0.35	2.73	0.30	2.72	0.23

[a]- Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, Savannah

[b] - Alma, Attapulgus, Fort Valley, Griffin

[c] - Attapulgus, Blairsville, Fort Valley, Savannah

[d] - Alma, Blairsville, Fort Valley, Savannah

[e] - Alma, Fort Valley

[f] - Blairsville, Fort Valley

[g] - Alma, Blairsville

* - MAE calculated by averaging all the MAEs for Alma, Arlington, Blairsville, Fort Valley, Griffin, Midville, Plains and Savannah over all layouts of a given configuration

Table 3.9: The results for various periods of predictions for 2001 to 2003 data from Byron, when fed to the general model

Period of Prediciton	MAE (^oC)	MAD* (^oC)	MAE/MAD
1	0.62	1.02	0.61
2	0.97	1.89	0.51
3	1.31	2.69	0.49
4	1.48	3.43	0.43
5	1.66	4.10	0.40
6	1.86	4.70	0.40
7	1.95	5.20	0.37
8	2.07	5.62	0.37
9	2.19	5.95	0.37
10	2.30	6.20	0.37
11	2.37	6.37	0.37
12	2.51	6.47	0.39

*-Mean Absolute Diffference

Table 3.10: The results for various periods of predictions for 2001 to 2003 data from Dixie, when fed to the general model

Period of Prediciton	MAE (^oC)	MAD* (^oC)	MAE/MAD
1	0.59	1.01	0.58
2	0.92	1.90	0.49
3	1.25	2.74	0.46
4	1.46	3.52	0.41
5	1.66	4.23	0.39
6	1.86	4.87	0.38
7	1.99	5.42	0.37
8	2.06	5.88	0.35
9	2.24	6.25	0.36
10	2.32	6.52	0.36
11	2.36	6.70	0.35
12	2.43	6.79	0.36

* - Mean Absolute Difference

Table 3.11: The results for various periods of predictions for 2003 data from Homerville, when fed to the general model

Period of Prediciton	MAE (°C)	MAD* (°C)	MAE/MAD
1	0.63	1.08	0.58
2	1.03	2.04	0.50
3	1.42	2.92	0.49
4	1.72	3.74	0.46
5	1.97	4.50	0.44
6	2.21	5.17	0.43
7	2.39	5.76	0.41
8	2.48	6.24	0.40
9	2.65	6.60	0.40
10	2.72	6.90	0.39
11	2.81	7.07	0.40
12	2.89	7.16	0.40

* - Mean Absolute Difference

Table 3.12: The results for various periods of predictions for 2002 to 2003 data from Nahunta, when fed to the general model

Period of Prediciton	MAE (°C)	MAD* (°C)	MAE/MAD
1	0.67	1.10	0.61
2	1.08	2.06	0.53
3	1.52	2.97	0.51
4	1.82	3.82	0.48
5	2.12	4.60	0.46
6	2.39	5.30	0.45
7	2.58	5.92	0.44
8	2.69	6.44	0.42
9	2.90	6.85	0.42
10	2.94	7.17	0.41
11	3.04	7.38	0.41
12	3.18	7.46	0.43

* - Mean Absolute Difference

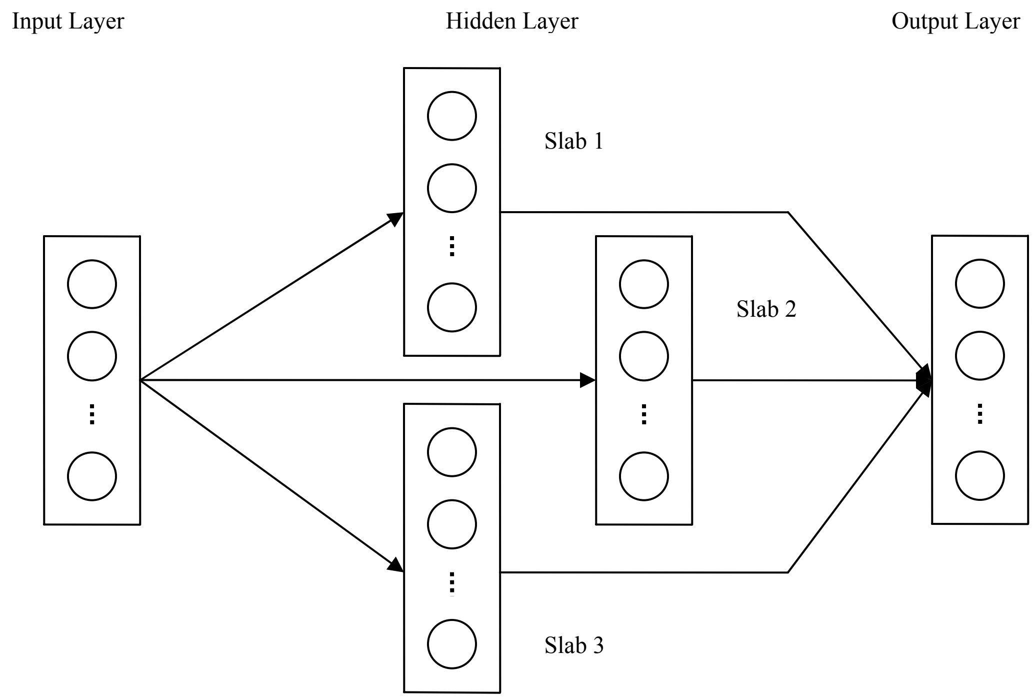

Figure 3.1: A Schematic representation of the Ward ANN.

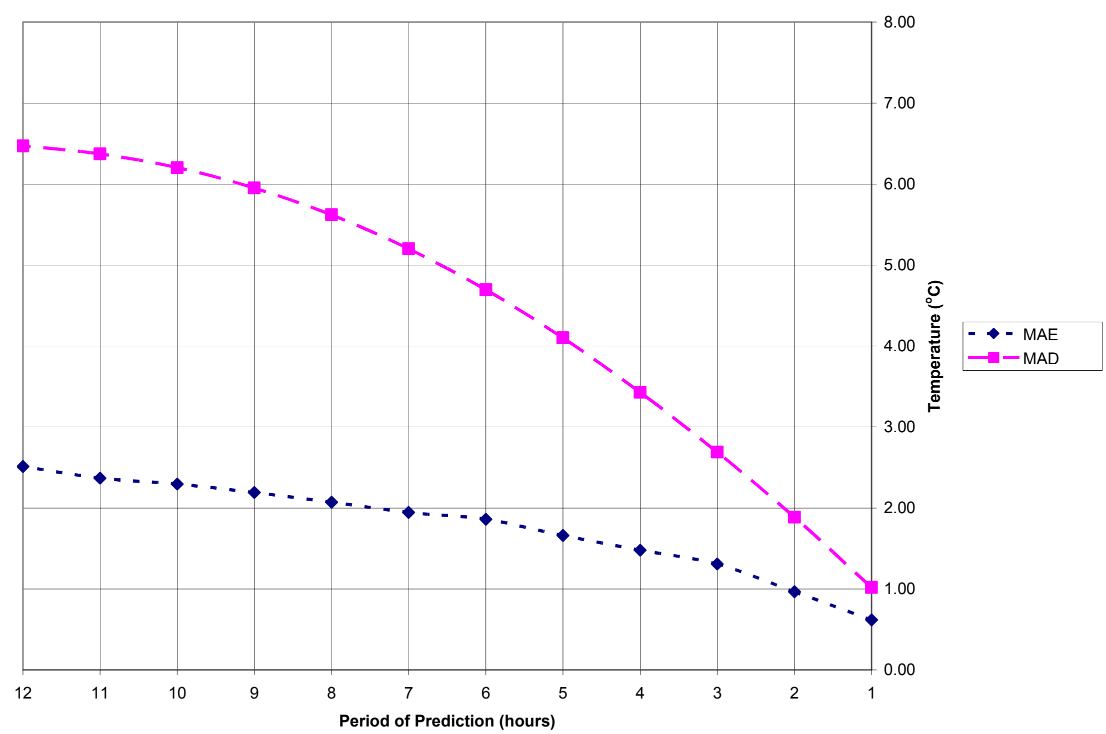

Figure 3.2: MAE/MAD v/s Period of Prediction for Byron

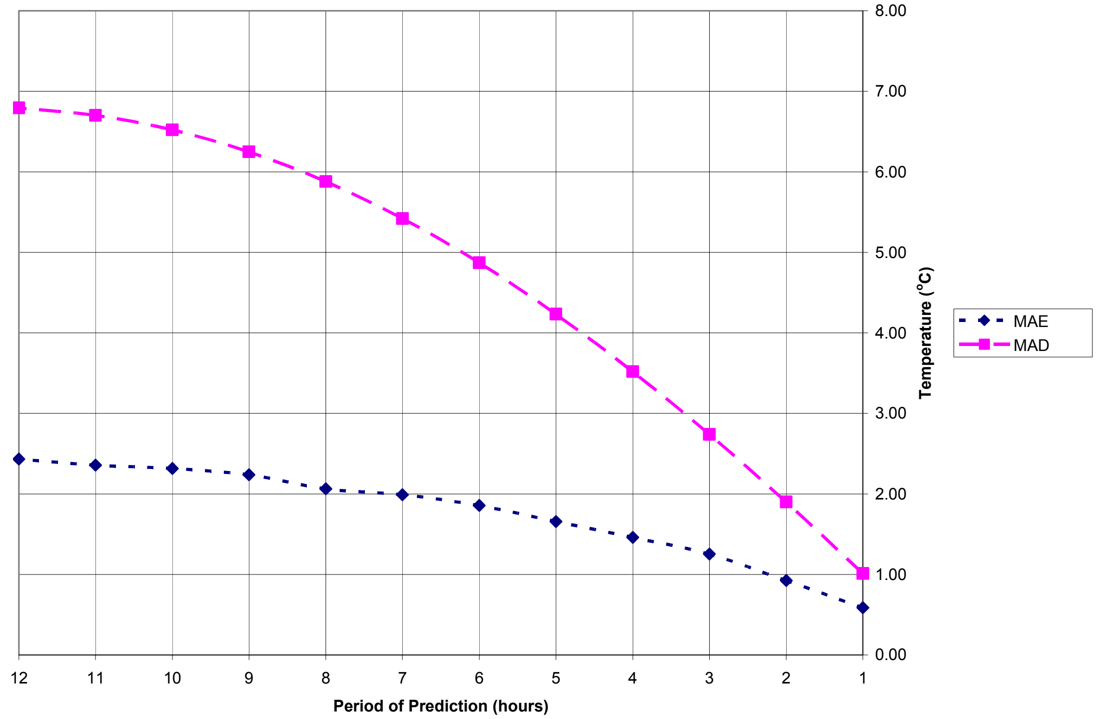

Figure 3.3: MAE/MAD v/s Period of Prediction for Dixie

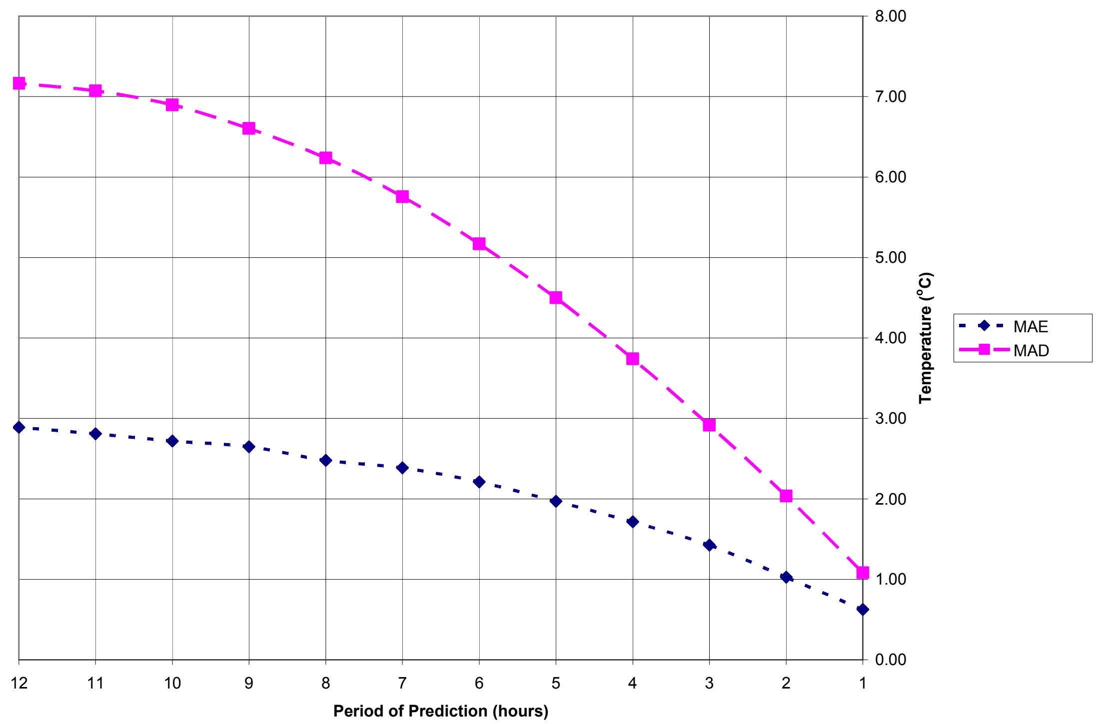

Figure 3.4: MAE/MAD v/s Period of Prediction for Homerville

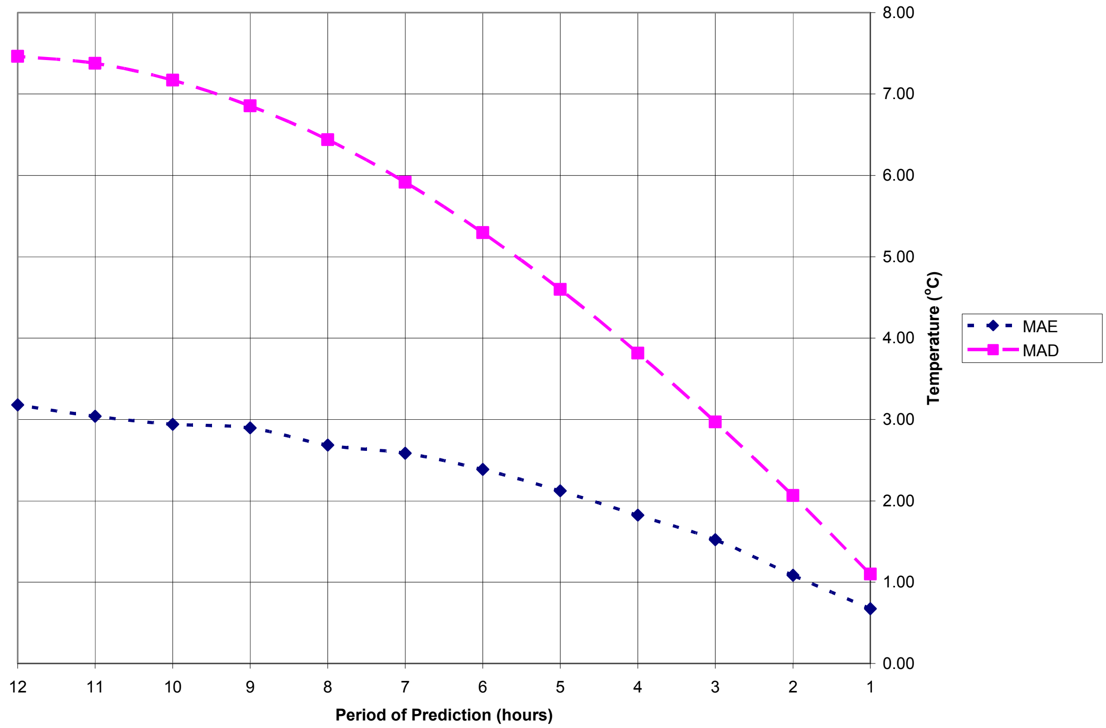

Figure 3.5: MAE/MAD v/s Period of Prediction for Nahunta

CHAPTER 4

SUMMARY AND CONCLUSION

The ultimate goal of this study was to develop an ANNs that would be incorporated in a DSS that could help farmers in protecting their crops from frost by predicting short term temperatures for a location with reasonable accuracy. Initially, ANNs were developed to predict temperatures in hourly increments starting from one hour ahead to twelve hours ahead for three locations. The locations were Fort Valley, Blairsville and Alma. Through experiments it was determined that the important weather variable inputs were, temperature, relative humidity, solar activity and wind speed. The duration of prior data needed depended on the period of prediction as well as the location for which the model was being developed. Ward ANNs were found to produce the highest accuracy and the optimal number of hidden nodes was found for each condition. It was also found that when an ANN model developed for a particular location is used to predict temperatures for a second location, the accuracy is less than when using a model developed for the second location. As a result there was a need to develop temperature prediction models that could predict temperatures independent of location.

In the second part of this study models were developed to predict temperature one, four eight and twelve hours ahead for a location without historical weather data using a model developed for locations with historical data. It was found that the models developed using data from nine-location configuration were the most consistently accurate models. The nine locations were Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains and

Savannah. The selection of these locations was dictated by the fact that these nine locations are spread all over and cover most of the state. It was observed that region specific models (south Georgia) were not as accurate as the models that relied on data from across the state for development. The accuracy of the general model was comparable to the site specific models that were developed using data only for that particular location. Following this models that predicted temperature one to twelve hours ahead were developed using the data from the nine locations discussed.

Future research will focus on developing programs that can use the ANNs developed and successfully incorporate them into the pre-existing web-based system to create the DSS that will help farmers protect their crops from frost damage by predicting temperatures.

REFERENCES

Abdelaal, R.E. and M. A. Elhadidy. 1994. A machine learning approach to modeling and forecasting the minimum temperature at Dhahran, Saudi-Arabia, *Energy 19(7): 739-749*.

Blennow, K. and P. Persson. 1998. Modeling local-scale frost variations using mobile temperature measurements with a GIS, *Agricultural and Forest Meteorology 89: 59-71*.

Bruton, J.M., G. Hoogenboom and R.W. McClendon. 2000. Estimating daily pan evaporation with artificial neural networks, *Transactions of the ASAE 43(2): 491-496*.

Cook, D.F. and M.L. Wolfe. 1991. A back propagation neural network to predict average air temperature, *AI Applications 5(1): 40-46*.

Elizondo, D., G. Hoogenboom and R.W. McClendon. 1994. Development of a neural network to predict daily solar radiation, *Agricultural and Forest Meteorology 71: 115-132*.

Felker, P. and H. Han. 1997. Estimation of daily soil water evaporation using an artificial neural network*, Journal of Arid Environments 37:251-260*.

Figuerola P.I. and N.A. Mazzeo. 1997. An analytical model for the prediction of nocturnal and dawn surface temperatures under calm, clear sky conditions, *Agriculture and Forest Meteorology, 85(3-4):229-237*

Georgiev, G. and G. Hoogenboom. 1998. Crop growth and yield estimation using current weather forecasts and weather observations, *p 69-72*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Georgiev, G. A. and G. Hoogenboom. 1999. Near real-time agricultural simulations on the web, *Simulation 73(1): 22-28*.

Hoogenboom, G. 1996. The Georgia automated environmental monitoring network, In: *Preprints of the 22nd Conf. On Agricultural and Forest Meteorology 343-346*, Boston, Mass., American Meteorological Society.

Hoogenboom, G. 2000a. Contribution of agrometeorlogy to the simulation of crop production and its applications, *Agricultural and Forest Meteorology 103(1-2): 137-157*.

Hoogenboom, G. 2000b. The Georgia automated environmental monitoring network, *p. 24-25*, In: *Preprints of the 24th Conference On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. Georgiev, D. Clarke, D.D. Gresham and G. Harbers. 1998. Internet tools for delivery of weather data and applications, *p. 251-253*, In: *Preprints of the 23rd Conf. On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, Massachusetts.

Hoogenboom, G., G. A. Georgiev, and D.D. Gresham. 2000. Development of weather based products for agricultural and environmental applications, *p. 66-67*, In: *Preprints of the 24th Conf. On Agricultural and Forest Meteorology*, Boston, Mass., American Meteorological Society.

Jain, A., R.W. McClendon and G. Hoogenboom. 2003. Frost prediction using artificial neural networks: A temperature prediction approach, ASAE Paper no. 033075, ASAE, St. Joseph, MI.

Krasovitski, B., E. Kimmel and I. Amir. 1996. Forecasting earth surface temperature for the optimal application of frost protection methods, *Journal of Agricultural Engineering Research 63(2):93-102*

Li, B. R.W. McClendon and G. Hoogenboom. 2004. Spatial interpolation of weather data for multiple locations using artificial neural networks, *Transactions of the ASAE.* (in revision)

Mehuys, G.R., N.K. Patni, S.O. Prasher and C.C. Yang. 1997. Application of artificial neural networks for simulation of soil temperature, *Transactions of the ASAE 40(3): 649-656*.

Mort, N. and C. Robinson. 1996. A neural network system for the protection of citrus crops from frost damage, *Computer and Electronics in Agriculture 16(3): 177-187*.

Okie W.R., G.L. Reighard, W.C. Newall Jr., C.J. Graham, D.J. Werner, A.A. Powell, G. Krewer and T.G. Beckman. 1998. Spring freeze damage to the 1996 peach and nectarine crop in the southeastern United States, *HortTechanology 8(3): 381-385.*

Powell, A.A. and D.G. Himelrick. 2003. Principles of freeze protection for fruit crops, http://www.aces.edu/dept/peaches/freeze.html (Verified November 18, 2003).

Tyson, AW. , R.E. Sneed, C.M. Mainland, K.B. Perry, E.B. Polling, D.C. Sanders and R Unrath. 2002. Frost/Freeze protection by sprinkler irrigation, www.smallfruits.org/Weather/frost_freeze.htm (Verified August 13, 2002).

Ward System Group, Frederick, MD. 1993. Manual of NeuroShell 2. Frederic, MD: Ward Systems Group.