International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:7, 2007

An Enhanced Artificial Neural Network for Air Temperature Prediction

Brian A. Smith, Ronald W. McClendon, and Gerrit Hoogenboom

Abstract**—**The mitigation of crop loss due to damaging freezes requires accurate air temperature prediction models. An improved model for temperature prediction in Georgia was developed by including information on seasonality and modifying parameters of an existing artificial neural network model. Alternative models were compared by instantiating and training multiple networks for each model. The inclusion of up to 24 hours of prior weather information and inputs reflecting the day of year were among improvements that reduced average four-hour prediction error by 0.18°C compared to the prior model. Results strongly suggest model developers should instantiate and train multiple networks with different initial weights to establish appropriate model parameters.

Keywords**—**Time-series forecasting, weather modeling.

I. Introduction

FROST damage is a significant concern for fruit growers in Georgia and other southeastern states, where bud formation and flowering at the start of the growing season includes the late-winter and early-spring months. Unseasonably cold temperatures during early 1996 and 2002 killed flowers and were responsible for reduced fruit harvests [1], [2]. Growers can take steps to mitigate the effects of frost by using orchard heaters or irrigation to protect their fields from the worst damage, but these methods require local monitoring of weather conditions and advance warning of a freeze. Following a January 1997 freeze that resulted in losses of $300 million for Florida growers, the Florida Automated Weather Network was created [3].

The University of Georgia's Automated Environmental Monitoring Network (AEMN), created in 1991, currently consists of over 60 automated weather stations throughout the state of Georgia, covering the breadth of the state's geographic diversity, from the coastal plains in the southeast, through the Piedmont, and into the Blue Ridge Mountains in the north [4]. The solar-powered stations are primarily situated in rural areas where the National Weather Service does not provide detailed local observations. Every second they collect air temperature, relative humidity, wind speed, wind direction, solar radiation, rainfall, and other weather data. Since March 1996 these observations have been aggregated into 15-minute averages, totals, and extremes, depending on the nature of the series. Previous observations were aggregated hourly.

Among the online decision support tools made available by the AEMN are short-term air temperature predictions. These predictions, ranging from one to 12 hours ahead, are available on the AEMN website, www.georgiaweather.net, during the winter and early spring. Predictions available on the AEMN website are generated by artificial neural network (ANN) models developed in [5], [6]. To predict temperature for a location, these networks make use of up to six hours of lagged observations from the site as inputs. The models incorporate the time of day, as well as measurements of air temperature, humidity, wind speed, and solar radiation, and were developed for use from January through April. Additional classification models using ANNs were developed to predict freeze events. For this classification problem, the addition of recent rainfall observations as input variables was found to improve performance. ANN models have also been used to predict inputs to a separate frost deposition model in order to more accurately predict frost and ice on roads and bridges [7].

The networks created in [5], [6] faced software constraints limiting the number of observations used in model development to 32,000. The study also relied on preliminary experiments that trained and evaluated a single network to determine the effects of altering model inputs or parameters. The goal of this research was to improve these temperature prediction models using more advanced and flexible neural network technologies. Specifically, the research explored four methods of improving forecast accuracy: (1) increasing the number of training patterns, (2) including input variables encoding seasonal, or time-of-year, information, (3) extending the duration of the lagged data used as inputs, and (4) varying the number of nodes in the hidden layer.

Manuscript received July 15, 2005. This work was funded in part by a partnership between the USDA-Federal Crop Insurance Corporation through the Risk Management Agency and the University of Georgia and by state and federal funds allocated to Georgia Agricultural Experiment Stations Hatch projects GEO00877 and GEO00895.

B. A. Smith is a Graduate Research Assistant with the Artificial Intelligence Center, University of Georgia, Athens, GA 30602 USA (email: smithba@uga.edu).

R. W. McClendon is a Professor with the Department of Biological and Agricultural Engineering, Driftmier Engineering Center, University of Georgia, Athens, GA 30602 USA (corresponding author; phone: 706-542-0882; fax: 706-542-6063; e-mail: rwmc@engr.uga.edu).

G. Hoogenboom is a Professor with the Department of Biological and Agricultural Engineering, University of Georgia, Griffin, Georgia 30223 USA (e-mail: gerrit@griffin.uga.edu).

II. Methodology

A. Data Sets

Previous temperature prediction work in the AEMN domain was carried out by Jain, who trained networks using a development set in which sites were selected so as to encompass a broad range of conditions [6]. Model evaluation was undertaken with a data set composed of sites collectively representative of the southern and central growing regions of Georgia. The same sites and years were used in this work, as illustrated in Fig. 1, allowing for a comparison of these new results with the previous study. The model development sites included Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, and Savannah, which generally have long histories of weather data. For these nine stations the data up to and including the year 2000 were included in the development set. Model evaluation data were more recent, from 2001-2004, and included observations from the Brunswick, Byron, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Homerville, Nahunta, Newton, Valdosta, and Vidalia sites. Jain used the same locations for the years 2001-2003 for evaluation [6]. The development and evaluation sets were restricted to observations from the first 100 days of the year, through April 9 or 10 for leap and non-leap years respectively. This range included both a large set of winter observations and the early growing season. The data sets were restricted to "low-temperature" observations with current temperature measurements below 20°C. Temperatures above 20°C were found not to be associated with freeze events within a 12-hour prediction horizon, the longest such horizon considered in this research.

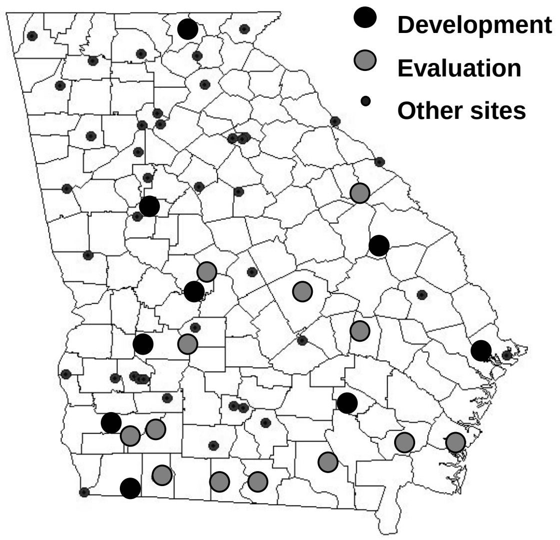

Fig. 1 Locations of AEMN weather stations

Model inputs included five weather variables: temperature, relative humidity, wind speed, solar radiation, and rainfall. While [6] determined rainfall did not improve temperature prediction accuracy, [8] found that rainfall was useful for predicting freeze events. In addition to the "current" values for each observation on record, lagged variables, spaced at one hour intervals, were also included in each observation pattern. Additionally, first-difference terms for the weather variables and their lags were derived and included. Note that the information contained in the first-difference variables is implicit in the current and lagged data. Though a model could, in principle, establish a set of weights to represent these first differences, making this information explicit has been found to improve model performance over this domain.

Each observation pattern contained two sets of cyclic variables associated with the time and the date of the observation. Because time is modular in an arithmetic sense, simply representing it with a single variable failed to capture all information inherent in a measurement. To overcome this limitation, cyclic variables were constructed using fuzzy logic membership functions. Fig. 2 shows four such functions with a range of 0 to 1 for the time variable with domain 0000 to 2400. Note that one of the variables, corresponding to the concept midnight, "wraps around" the domain's upper and lower bounds. An analogous approach was taken to convert the day-of-year for each observation to four seasonal variables.

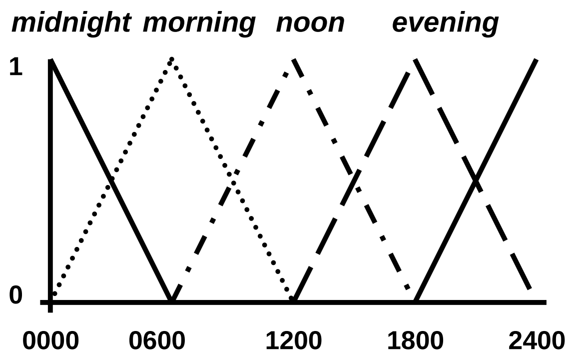

Fig. 2 Four fuzzy logic membership functions ranging over the time of day

B. Model Development

Software constraints restricted previous AEMN temperature prediction models to 32,000 development observations. To overcome this limitation, a neural network suite was written in the Java programming language. This suite placed no limits on the size of the sets used in the training or evaluation process.

Throughout this paper, the term model is to be understood as an ANN architecture and a set of associated parameters. A model is instantiated as a network by using a random seed to assign initial weight values and a training set order and subsequently training the network. That is, a model is a description of a class of potential networks. All networks were trained via the well-known error backpropagation (EBP) algorithm [9]. EBP training was successfully applied in previous ANN research in the AEMN domain.

All models explored in this research were based on the Ward-style network architecture used in [5], [6]. The Ward network is an ANN with multiple node types that implement multiple activation functions. The models used a linear input layer, three equally-sized, parallel "slabs" in the hidden layer, and a single, logistic output node, interpreted as the temperature at some prediction horizon (Fig. 3). The linear transformation carried out by the input layer was determined over the entire model development set. Each data series used as an input was transformed to the range 0.1 to 0.9. As the transformation made use of the maximum and minimum values of each series in the development set, this range may not hold when an evaluation pattern is presented. The hidden layer contained slabs implementing the Gaussian, Gaussian complement, and hyperbolic tangent activation functions [10]. Fully connected, biased weight matrixes connect the input layer to the hidden layer and the hidden layer to the output node.

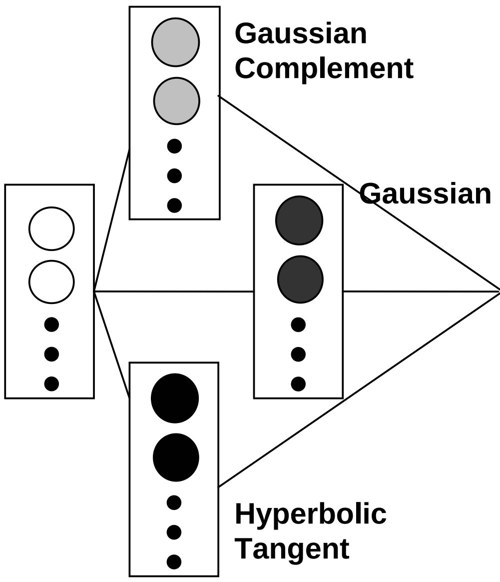

Fig. 3 A Ward-style neural network with three "slabs," each with a distinct activation function

Instantiating a Ward-style architecture requires specifying a number of network parameters including the learning rate and momentum, initial weight range, size of the training and test sets, number of hidden nodes in each slab, and the included input series. Variations to the learning rate, momentum, and the initial weight range were considered in preliminary studies, but these parameters were found to play only a small role in model accuracy. For all models considered in this research a learning rate of 0.1 and an initial weight range of -0.1 to 0.1 were used. A momentum term was not included

Models are typically evaluated by instantiating a single ANN and measuring the resulting performance of the trained network over an evaluation set. Such an evaluation scheme assumes that the performance of a single network is an accurate measure of any network that may instantiate the model. But, due to the random nature of the initial weights and training pattern ordering, there is no guarantee that two networks instantiating the same model will converge to the same final state [11]. This suggests that another method of model evaluation, involving multiple networks, is warranted.

The temperature prediction models developed in [5], [6] relied on single-network evaluation. During the course of the current research, multiple-network evaluation was applied to these aspects of the model, often with different results. A group of networks, differing only in random seeding, were instantiated for each model. Each network was trained on a training set independently constructed from all available development patterns via random selection without replacement and trained for four million learning events prior to evaluation. Preliminary work in the AEMN domain showed the use of a testing set to stop training was not helpful. Test set performances mirrored those of the training sets and it was rare for an instantiated network's performance to degrade in any reasonable amount of time. Training was stopped after four million events because preliminary work showed that epoch-by-epoch improvements were generally inconsequential by this time. After training, the mean absolute error (MAE) associated with each network's temperature prediction was calculated over the entire evaluation set. Because the goal of the research was to develop a single, highly accurate ANN, the minimum MAE of this group was selected as the appropriate performance measure for a model.

C. *Experiments*

To explore the effects of increased training set sizes on model performance, six models, differing only in the number of training observations used, were instantiated by five networks each. Training set sizes of 10K, 25K, 50K, 100K, 200K, and 400K observations were considered. All weather variables and related first-difference series, as well as the four diurnal variables, were used as inputs. A lag length of six hours was used.

Next, to determine the effect of adding time-of-year information to the input vector, these models were compared to a second group, modified to include the four seasonal variables. All other inputs were the same, including the six hours of lagged data used.

A third experiment explored the effect of variations in the lag length of the environmental inputs by instantiating multiple models with seasonal variables for lag lengths of six, eight, 10, 12, 18, 24, 36, and 48 hours to determine if increasing lag length beyond six hours improved prediction accuracy.

Finally, an experiment was conducted comparing the model accuracy of models with seasonal inputs and hidden layers of 15, 30, 75, and 225 nodes to determine a preferred hidden layer size. To allow a single parameter to represent the number of nodes, the three slabs were constrained to be of equal size, so that the four sizes considered corresponded to 5, 10, 25, and 75 nodes per slab.

III. Results

The results discussed here are for experiments with four-hour prediction models. The results for other horizons were qualitatively similar. Overfitting did not occur in any runs. Instead, the rare occurrences of increasing error over the testing set during training corresponded to periods of increasing error over the training set as well.

Fig. 4 presents the MAEs associated with thirty trained networks instantiating six different models, corresponding to training set sizes between 10K and 400K unique observations. The most accurate network was trained over all 400K observations and had an MAE of 1.55°C. But the minimum MAEs associated with the most successful models of 10K, 100K, and 400K training observations differed by less than 0.005°C. The minimum MAE curve showed no clear relationship between error and training set size. Smaller, intermediate, and larger training set sizes were capable of yielding similar performance over five network instantiations.

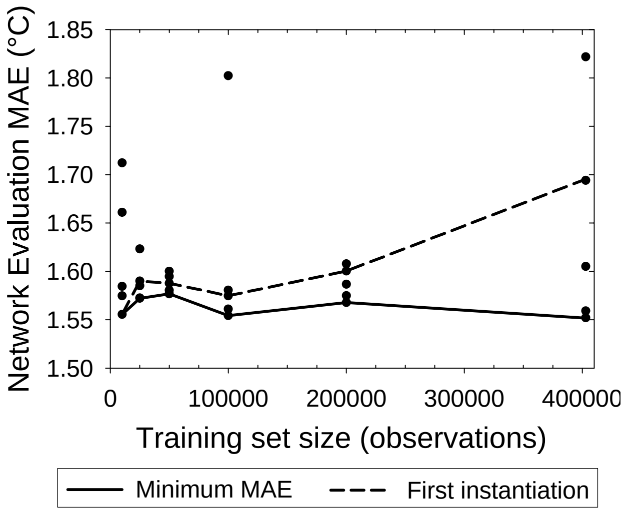

Fig. 4 A comparison of the single-network and multiple-network evaluation methods for models distinguished by training set size; each point corresponds to the evaluation MAE of an instantiated network

These results can be contrasted with single-network evaluation by considering the relative performance of the first network instantiated for each of the six models. This curve, also presented in Fig. 4, appears to show that increased training set sizes were associated with increased evaluation errors. The ANN with the smallest development set had the lowest MAE over the evaluation set (1.56°C), while the ANN trained over all available observations gave the highest (1.69°C). The use of a single-network evaluation scheme in this case would clearly lead to inaccurate conclusions regarding relative model performance. Such erroneous results will not always occur, however. In this case the first instantiations of the small-set models were relatively accurate while the first instantiations of the larger networks were relatively inaccurate, with the 400K-observation model leading to the least accurate network. The result was a deceptively suggestive curve.

The second experiment evaluated six additional models with seasonal input terms, corresponding to the six distinct training set sizes. As shown in Fig. 5, these models outperformed the models without seasonal inputs over all training set sizes. The most accurate model with seasonal inputs had an evaluation MAE of 1.50°C, an improvement of more than 0.05°C compared to best model without them. Again there was no convincing evidence for a relationship between training set size and performance. The difference between the most and least accurate model MAE was less than 0.02°C and did not trend with the number of training observations. As a result, subsequent experimental results are not presented in terms of training set size.

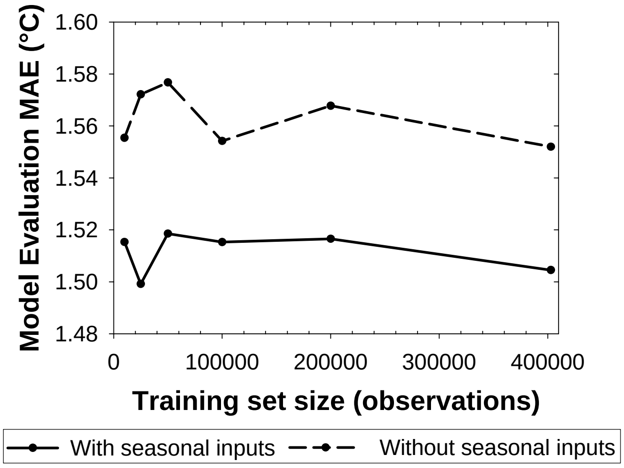

Fig. 5 A comparison of models with and without seasonal input terms using minimum-error, multiple-network evaluation; each point corresponds to the minimum MAE obtained over five networks instantiating the model

Six hours was the preferred lag length for forecast horizons of four hours or more in prior AEMN work [6]. The current study compared various models with seasonal terms that differed only in the hours of lagged data included as inputs. Fig. 6 presents the results of the experiment, which indicate that a lag length of six hours is clearly suboptimal for this forecast horizon. In fact, with an MAE of 1.50°C, the six-hour model was associated with the highest average error of any model considered here. A lag length of 24 hours resulted in an MAE of 1.41°C, the lowest observed in the experiment. The success of the 24-hour model makes intuitive sense as such a history is capable of generalizing over trends associated with the familiar daily cycle. Models with more than 24 hourly lags led to less accurate network instantiations. Presumably the information gained by the additional input terms was not sufficient to overcome the increased complexity of the search space of possible weights.

The decision in [6] to use six hours was likely due to the work's reliance on increasing the lag length by short increments until evaluation errors began to increase. Because it relied on single-network evaluation and found that a network with an eight-hour lag length was outperformed by a simpler six-hour network, the work reached an inaccurate conclusion. The results of this research suggest that the use of multiple-network evaluation can avoid such errors.

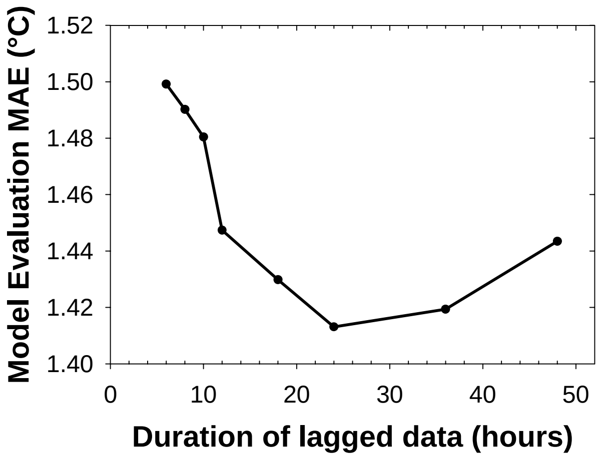

Fig. 6 A comparison of models distinguished by lag length using minimum-error, multiple-network evaluation; each point corresponds to the minimum MAE obtained over five networks instantiating the model

The final experiment instantiated networks for models with seasonal inputs which differed in hidden layer size showed that the preliminary study used in [6] to settle on a layer of 75 nodes (25 per slab) was similarly flawed. Multiple-network, minimum-error evaluation revealed that for models with six and with 24 hours of lagged data, a relatively small network with 30 hidden nodes (10 per slab) gave the smallest MAE. Improvements over the larger networks were slight. With six hours of lagged data the larger models had errors only 0.01°C higher than the 30-node model. Allowing for repeated instantiations, increasing the number of hidden nodes beyond thirty increased computational complexity but did not lead to more accurate models.

To establish a direct comparison between the models developed here and those obtained in [6], 15 new networks were instantiated for prediction horizons of one, four, eight and 12 hours. For each horizon the network with the lowest MAE over the development set was selected for comparison. The selected network was evaluated over the same sites and years used in [6] and its MAE calculated.

Table I compares the prediction accuracies of these ANN models to those obtained in [6]. This model, which made use of seasonal input terms, 24 hours of lagged observations, and 30 hidden nodes led to an improvement in model MAE over all horizons. For example, the four-hour prediction improved by 0.18°C, or 11%. As horizon length increased, the improvement as a percentage of original error decreased. The new networks were also evaluated over a data set consisting of the same sites with observations from 2004. Over this set the networks were slightly more accurate than over the 2001-2003 period.

TABLE I
COMPARISON OF MODEL PREDICTION ACCURACY OVER EVALUATION DATA

Horizon (hours)	Previous model MAE (°C), 2001-3	Current model MAE (°C), 2001-3	Improvement (°C)	Current model MAE (°C), 2004
1	0.62	0.54	0.08 (12.7%)	0.54
4	1.60	1.42	0.18 (11.1%)	1.35
8	2.30	2.13	0.17 (7.4%)	1.99
12	2.69	2.57	0.12 (4.5%)	2.33

IV. Conclusions

The research presented in this paper considered the effects of changes to the ANN models used to predict temperature over Georgia AEMN data, including larger training set sizes, seasonal input terms, increased lag lengths, and varying the size of the network. Increasing the size of the training set did not reduce forecast errors. However, the inclusion of seasonal variables corresponding to membership in the fuzzy sets winter, spring, summer, and fall did improve model performance. Similar improvements resulted from extending the duration of historical data in the input vector from six to 24 hours. Models with a hidden layer size of 30 nodes were more accurate than larger models over repeated instantiations.

The introduction of seasonal terms may provide a means of implementing an accurate year-round forecast model. Future work may compare the accuracy of such models to season-specific models such as those created in this research. The decreases in model performance associated with lag lengths greater than 24 hours and hidden layers larger than 30 nodes suggest that training temperature-predicting ANNs via error backpropagation is a process sensitive to increases in the number of weights. Feature extraction methods may be able to reduce the size of the input vector, reducing network degrees of freedom and improving performance. Additionally, different algorithms for determining weights may be less sensitive to network size.

Finally, when applied to data-rich environments, a clear distinction should be maintained between abstract neural network models and actual instances of these models. The performance of a single instantiated network is not likely to be an accurate measure of model performance. In this study, model evaluation over multiple instantiations led to better parameter selection by presenting more accurate comparisons of distinct models than those afforded by single-network evaluation. When large data sets are involved, model performance measures should make use of multiple instantiations.

References

[1] W. R. Okie, G. L. Reighard, W. C. Newall, Jr., C. J. Graham, D. J. Werner, A. Powell, G. Krewer, and T. G. Beckman, “Spring freeze damage to the 1996 peach and nectarine crop in the southeastern United States,” *HortTechnology*, vol. 8, pp. 381-386, 1998.

[2] R. W. McClendon and G. Hoogenboom, “New non-insurance risk management tools: Decision support for freeze protection using artificial neural networks,” presented at the USDA Agricultural Outlook Forum, Arlington, VA, Feb. 24, 2005.

[3] C. Woods. (1998, Jan. 18). "New Florida Automated Weather Network (FAWN) goes online today," [Press release, online]. Available: http://www.napa.ufl.edu/98news/weather.htm (Verified July 15, 2005)
[4] G. Hoogenboom, "The Georgia automated environmental monitoring network," in *Preprints of the 24th Conference On Agricultural and Forest Meteorology*, American Meteorological Society, Boston, MA, 2000, pp. 24-25.
[5] A. Jain, R. W. McClendon, G. Hoogenboom, and R. Ramyaa, "Prediction of frost for fruit protection using artificial neural networks," American Society of Agricultural Engineers, St. Joseph, MI, ASAE Paper 03-3075, 2003.
[6] A. Jain, "Frost prediction using artificial neural networks: A temperature prediction approach," M.S. thesis, Artificial Intelligence Center, University of Georgia, Athens, GA, 2003.
[7] B. R. Temeyer, W. A. Gallus, Jr., K. A. Jungbluth, D. Burkheimer, and D. McCauley, "Using an artificial neural network to predict parameters for frost deposition on Iowa bridgeways," in *Proceedings of the 2003 Mid-Continent Transportation Research Symposium*, Iowa State University, Ames, IA, 2003.
[8] Ramyaa, "Frost prediction using artificial neural networks: A classification approach," M.S. thesis, Artificial Intelligence Center, University of Georgia, Athens, GA, 2004.
[9] S. Haykin, *Neural networks: a comprehensive foundation*, 2nd edition. Upper Saddle River, NJ: Prentice Hall, 1998.
[10] *Manual of NeuroShell 2*, Ward Systems Group, Frederick, MD, 1993.
[11] F. Salehi,, R. Lacroix, and K. M. Wade, "Effects of learning parameters and data presentation on the performance of backpropagation networks for milk yield prediction," *Transactions of the ASAE*, vol. 41, pp. 253-259, 1998.