World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:10, 2007

Improving Air Temperature Prediction with Artificial Neural Networks

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. Previous work established that the Ward-style artificial neural network (ANN) is a suitable tool for developing such models. The current research focused on developing ANN models with reduced average prediction error by increasing the number of distinct observations used in training, adding additional input terms that describe the date of an observation, increasing the duration of prior weather data included in each observation, and reexamining the number of hidden nodes used in the network. Models were created to predict air temperature at hourly intervals from one to 12 hours ahead. Each ANN model, consisting of a network architecture and set of associated parameters, was evaluated by instantiating and training 30 networks and calculating the mean absolute error (MAE) of the resulting networks for some set of input patterns. The inclusion of seasonal input terms, up to 24 hours of prior weather information, and a larger number of processing nodes were some of the improvements that reduced average prediction error compared to previous research across all horizons. For example, the four-hour MAE of 1.40°C was 0.20°C, or 12.5%, less than the previous model. Prediction MAEs eight and 12 hours ahead improved by 0.17°C and 0.16°C, respectively, improvements of 7.4% and 5.9% over the existing model at these horizons. Networks instantiating the same model but with different initial random weights often led to different prediction errors. These results strongly suggest that ANN model developers should consider instantiating and training multiple networks with different initial weights to establish preferred model parameters.

***Keywords*—**Decision support systems, frost protection, fruit, time-series prediction, weather modeling

I. Introduction

FROST damage is a significant concern for horticultural producers in Georgia and elsewhere in the southeastern United States, especially when bud formation and flowering occur during late-winter and early-spring. For example, unseasonably cold temperatures during early 1996 and 2002 damaged floral buds and were responsible for reduced fruit harvests [1]. Growers can take steps to lessen the effects of frost by using orchard heaters or irrigation to protect their trees and bushes from the worst damage, but these methods require advance warning of freezing conditions.

The University of Georgia's Automated Environmental Monitoring Network (AEMN) was created in 1991 and currently consists of 68 automated weather stations throughout the state of Georgia. The stations cover the breadth of the state's geographic diversity, from the coastal plain in the southeast, through the Piedmont, and into the Blue Ridge Mountains in the north [2]. The solar-powered stations are primarily situated in rural areas where the National Weather Service does not provide detailed local observations. The monitoring stations collect weather data such as air temperature, relative humidity, wind speed, wind direction, solar radiation, and rainfall at one-second intervals. Since March 1996 these observations have been aggregated into 15-minute averages, totals, and extremes, depending on the nature of the variable. Previous observations were aggregated hourly.

Among the online decision support tools made available by the AEMN are short-term air temperature predictions. These hourly predictions range from one to 12 hours ahead and are available on the AEMN website, www.georgiaweather.net, during winter and early spring. The temperature predictions are generated by artificial neural network (ANN) models developed by Jain et al. [3] and Jain [4]. To predict temperature for a location, the ANNs use as inputs up to six hours of prior weather observations from the site. 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. Classification models using ANNs to predict freeze events were developed by Ramyaa [5]. These networks classify observations into one of three classes depending on whether the model predicts freezing, near-freezing, or non-freezing conditions over a 12-hour prediction period. For the 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 special frost deposition model in order to more accurately predict frost and ice on roads and bridges [6].

Manuscript received March 14, 2006. 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 GEO01654.

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

Ronald W. McClendon is a Professor with the Artificial Intelligence Center and 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).

Gerrit 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).

The previous temperature prediction and classification networks faced software constraints limiting the number of patterns used in model development to 32,000 [3], [4], [5]. These studies 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 the current research is to improve these temperature prediction models using more advanced and flexible neural network technologies. Specifically, this research explores four possible methods of improving prediction accuracy: (1) increasing the number of training patterns, (2) including input variables encoding seasonal information, (3) extending the duration of the prior data used as inputs, and (4) varying the number of nodes in the hidden layer.

II. METHODOLOGY

A. Data Sets

The previous temperature prediction work in the AEMN domain by Jain [4], trained networks using a development set drawn from sites which were selected so as to encompass a broad range of conditions. Model evaluation was performed using a data set composed of sites collectively representative of the southern and central growing regions of Georgia. The same sites and years were used herein 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 have relatively 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 from 2001 to 2005, and included patterns from Brunswick, Byron, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Homerville, Nahunta, Newton, Valdosta, and Vidalia. The previous work used the same locations for the years 2001-2003 for evaluation [4]. To allow for a direct comparison to this previous work, the evaluation data in this study was divided into two sets: the first composed of the data from 2001-2003 and the second composed of the 2004-2005 patterns. The development and evaluation sets were restricted to patterns from the first 100 days of the year, through April 9 or 10 for leap and non-leap years, respectively. This range includes winter observations and the early growing season. The data sets were restricted to "low-temperature" patterns, those 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.

Model inputs included five weather variables: temperature, relative humidity, wind speed, solar radiation, and rainfall. In addition to the current values for each observation on record, prior data, spaced at one hour intervals, were also included in each training pattern. Hourly first-difference terms for the current and prior weather variables were also included. Note that the information contained in the first-difference variables is implicit in the current and prior data, but providing this information explicitly was found to improve model performance.

Each training and evaluation pattern contained two sets of cyclic variables associated with the time and the date of the observation. Because the time of day and year are periodic variables, simply representing each with a single input fails to capture all information inherent in a measurement. To overcome this limitation, cyclic variables were constructed using fuzzy logic membership functions. For the time variable, four such triangular functions with an output range of 0 to 1 were used over the domain 0000 to 2400 hours (Fig. 1). 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.

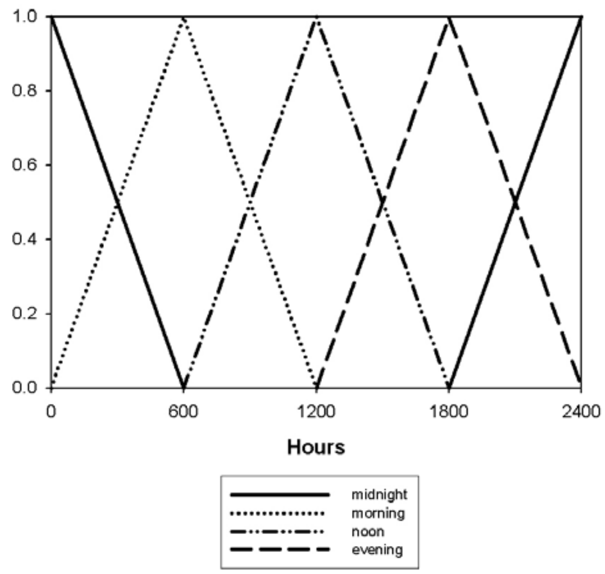

Fig. 1 Four fuzzy logic membership functions ranging over the time of day

B. Model Development

Software constraints restricted the previous AEMN temperature prediction models to 32,000 development patterns. 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. All networks were trained via the well-known error backpropagation (EBP) algorithm as described by Haykin [7]. EBP training was successfully applied in previous ANN research involving temperature prediction using AEMN data [3], [4], [5].

Throughout this paper, the term model is understood to be 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 group of potential networks differing only in the set of initial weights assigned before training and the order in which training patterns are presented. All models explored in this research were based on the Ward-style network architecture used in previous research by Jain [3], [4]. The Ward network is an ANN with multiple node types that implement multiple activation functions [8]. 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.

A linear transformation carried out by the input layer was determined for 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 three slabs using the Gaussian, Gaussian complement, and hyperbolic tangent activation functions. Fully connected, biased weight matrices connect the input layer to the hidden layer and the hidden layer to the output node.

The networks used in this research provide a mapping of a vector of I real-valued inputs, *x*, ranging over the values [0.1, 0.9] onto a real value z such that

$$z = g(\beta_0 + \sum_{j=1}^{J} \beta_j \cdot y_j)\text{, where}$$

$$g(n) = \frac{1}{1+\exp(-n)}. \tag{1}$$

y is a vector of length *J* containing the signals of the nodes in the ANN's hidden layer.

$$y_j = f_j\left(\alpha_{j0} + \sum_{i=1}^{I} \alpha_{ji} \cdot x_i\right)\text{, where}$$

$$f_j(n) = \begin{cases} \tanh(n), & \text{for } 0 < j \le j_1 \\ \exp(-n^2), & \text{for } j_1 < j \le j_2 \\ 1-\exp(-n^2), & \text{for } j_2 < j \le j_3 \end{cases} \tag{2}$$

$$\text{and } 0 < j_1 < j_2 < j_3 .$$

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 testing sets, number of hidden nodes in each slab, and the included input series. Variations in the learning rate, momentum, and the initial weight range were considered in preliminary studies, but these parameters were found to have a relatively small effect on 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.

ANN models are typically evaluated by instantiating a single network and measuring the resulting accuracy of the trained network for a set of patterns. Such an evaluation scheme assumes that the performance of a single network is an accurate measure of any network that may instantiate the model. However, due to the random nature of the initial weights and the training pattern ordering, there is no guarantee that two networks instantiating the same model will converge to the same final state from distinct starting points in the multidimensional weight space [9]. This suggests that another method of model evaluation, involving multiple networks, is warranted.

The previous temperature prediction models developed by Jain [3], [4] relied on single-network evaluation. An alternative approach was taken in this research whereby multiple ANNs were trained for identical model configurations. These networks, referred to as instantiations, differed only in the initial random weights and the order of the patterns presented. Each network was trained on a set of patterns independently constructed from all available development patterns via random selection without replacement for four million learning events prior to evaluation. Preliminary work using this approach showed that the use of a testing set to determine when to stop training was not helpful. Test set accuracies mirrored those of the training sets and it was rare for an instantiated network's accuracy to decrease. In addition, rare occurrences of increasing error for the testing set during training corresponded to the presence of increasing error for the training set as well. This phenomenon was also associated with poorly performing networks. After training, the mean absolute error (MAE) associated with each network's temperature prediction was calculated for the entire development set. Because the goal of the research was to develop a single, highly accurate ANN, the network with the minimum MAE of this group was selected as the appropriate performance measure for a model.

The error for the development set was used to decide between models so that comparisons between the final model and the previous research would not be biased in favor of the current work. A retrospective evaluation indicated that network MAEs for the development set are highly predictive of performance for the evaluation set.

Network training took place using 30 Dell Pentium IV workstations in a University of Georgia computer laboratory. Training was stopped after four million events because preliminary work suggested that epoch-by-epoch improvements were generally inconsequential by this time. Processing time was also a factor in the determination of the number of learning events. Using the threshold of four million learning events allowed the fastest of the machines used to train and evaluate two instantiated networks in a typical 12-hour run.

C. Experiments

To explore the effects of increased training set sizes on model performance, six models, differing only in the number of training patterns used, were instantiated by thirty networks each. Training set sizes of 10K, 25K, 50K, 100K, 200K, and

400K were considered. All weather variables and related first-difference series, as well as the four diurnal variables, were used as inputs. A six-hour duration of prior data was used for this experiment. 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 and parameters were the same, including the six hours of prior data. A third experiment explored the effect of variations in the number of prior hourly observations for the environmental inputs by instantiating multiple models with seasonal variables for durations of two, four, six, 12, 18, 24, 30, 36, and 48 hours to determine if increasing the duration beyond six hours improved prediction accuracy. A fourth experiment was conducted comparing the accuracy of models with seasonal inputs and hidden layer sizes of six, 15, 30, 45, 60, 75, 105, 120, 150, 180, and 225. To allow a single parameter to represent the number of nodes, the three slabs were constrained to be of equal size, so that the hidden layer sizes considered ranged between two and 75 nodes per slab.

Finally, the best-performing model was instantiated thirty times for each prediction horizon from one to 12 hours ahead. The instantiation with the lowest MAE for the development set was selected to represent the model. These final models were then run over an evaluation set consisting of all cold-weather patterns from 2001 through 2003. The relationships between target temperature, predicted temperature, and prediction error for the Ft. Valley site were evaluated for these ANNs. Additionally, the performance of the models over the damaging freeze events of late-February and early-March 2002 was examined. Each model was also evaluated over a final set that consisted of all of the low-temperature patterns from the evaluation sites during winter 2004 and 2005 (those with a temperature no greater than 20°C at the time of the prediction).

III. Results and Discussion

The results discussed here are for experiments with four-hour prediction models. Results for other horizons were qualitatively similar. Overfitting was exceedingly rare and occurred only during runs with poor prediction accuracy. An MAE for the development data was calculated for networks associated with six different models, corresponding to training set sizes that ranged from 10K and 400K unique patterns. Each model was instantiated by thirty networks (Fig. 2). The most accurate network was trained over 50K patterns and had an MAE of 1.51°C. However, the minimum MAEs associated with the most accurate instantiations of the 50K and 200K-pattern models differed by less than 0.006°C. These training set sizes were capable of yielding similar minimum MAEs over 30 network instantiations. Furthermore, there was no clear relationship between minimum network MAE and training set size for large sets. The use of single-network evaluation allows for the possibility of misleading approximations of model accuracy. In general, each model

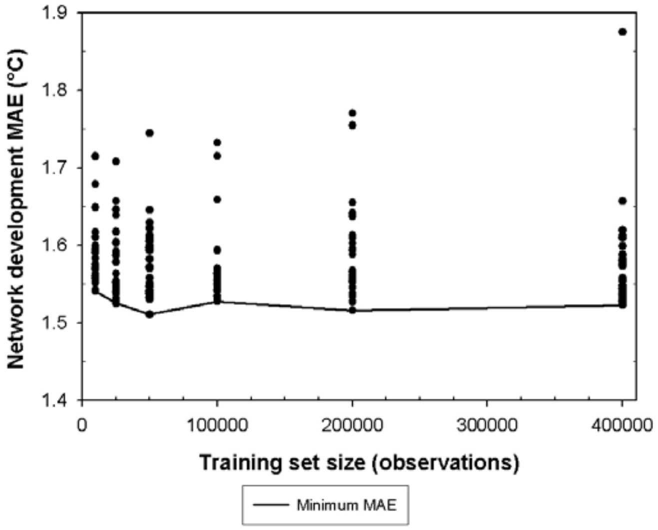

Fig. 2 Multiple-network evaluation for four-hour prediction models distinguished by training set size. Each point corresponds to the MAE, calculated for all patterns in the development set, of one of the 30 networks instantiating each model

MAE would be approximated by making a single draw from the distribution of MAEs associated with the model. However, the range of MAE values for each model is sizable. When drawing a single MAE value for each model, any combination of values is possible, many of which could suggest markedly different interpretations of the results.

The second experiment evaluated six additional models with seasonal input terms, corresponding to the six distinct

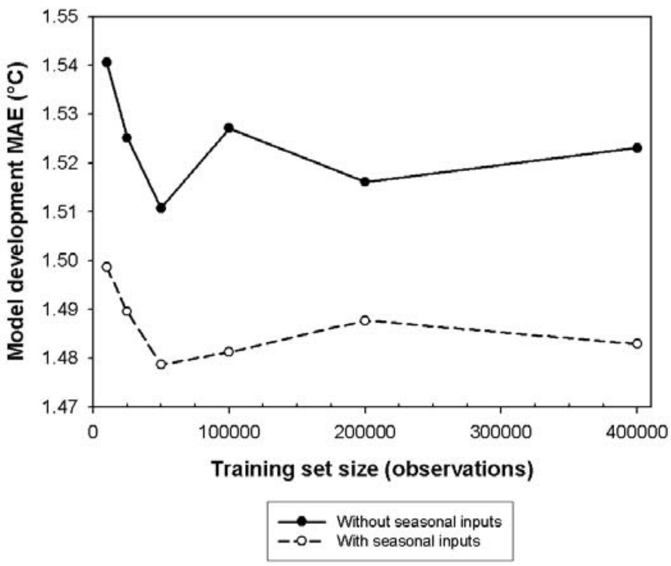

Fig. 3 A comparison of four-hour prediction models with and without seasonal input terms using minimum-error, multiple-network evaluation. Each point corresponds to the minimum MAE obtained over 30 networks instantiating each model

training set sizes. These models were more accurate than those without seasonal inputs for all the different sizes of the training set (Fig. 3). The most accurate model with seasonal inputs had an MAE of 1.48°C for the entire development set. This was an improvement of more than 0.03°C compared to the best model that did not consider seasonal inputs. Again there was no convincing evidence for a relationship between training set size and performance for sufficiently large sets. The difference between the most and least accurate model MAE was 0.02°C. Similar to the non-seasonal case, the 50K-pattern model using seasonal data exhibited higher accuracy than the other models. To explore whether this was typical, the seasonal experiment was repeated. This involved the instantiation of 30 new networks for each of the six models. In the second trial, the 50K-pattern model using seasonal data was slightly outperformed by the three seasonal models using larger training sets. With no evidence to suggest a preference for training set size, subsequent experiments made use of all available development patterns. All subsequent experiments continued to employ four million learning events during training.

Six hours was the preferred duration of prior data for prediction horizons of four hours or more in the prior temperature prediction study [4]. The current research compared various models with seasonal terms that differed only in the number of hours of prior data included as inputs. The results of the experiment indicate that a prior duration of six hours is clearly suboptimal for this horizon (Fig. 4). In fact, with an MAE of 1.48°C, the six-hour model was outperformed by all of the longer-duration models considered. The inclusion of 24 hours of prior data resulted in an MAE of 1.38°C, the lowest observed in the experiment. Models with data from more than 24 prior hourly observations were less accurate. 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. The decision in previous research to use six hours of prior data was likely due to the method of increasing the duration by short increments until evaluation errors began to increase [4]. Because that work relied on single-network evaluations and found that a network with eight hours of prior information was less accurate than a six-hour network, the experiment was stopped before exploring longer durations of prior data. The results of the research reported herein suggest that the use of multiple-network evaluation can avoid such errors.

The final experiment instantiated networks for models with seasonal inputs that differed in hidden layer size. Even if a 75-node hidden layer (25 nodes per slab) was optimal for a model with six hours of prior data, there was no guarantee that it would be best for a model with 24 hours of prior data and seasonal inputs. The results of the experiment, which evaluated each model over 30 instantiations, revealed that for models with 24 hours of prior data, a larger network with 120 hidden nodes (40 per slab) led to an instantiation with an MAE of 1.35°C, the smallest of the models considered (Fig. 5). Increasing the total number of hidden nodes beyond this

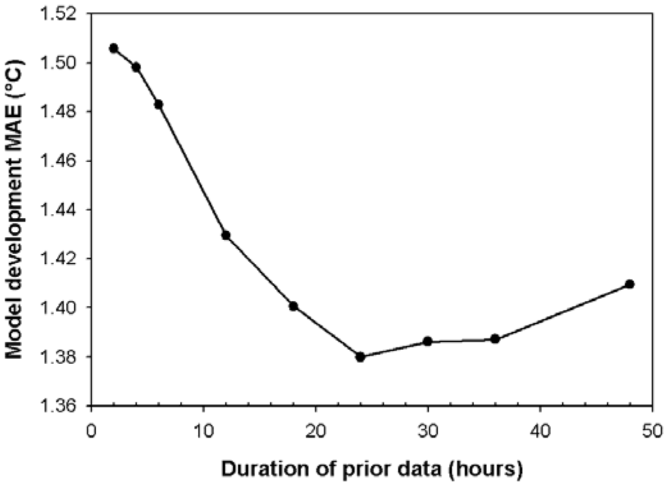

Fig. 4 A comparison of four-hour prediction models distinguished by the duration of prior data using minimum-error, multiple-network evaluation. Each point corresponds to the minimum MAE obtained over 30 networks instantiating each model

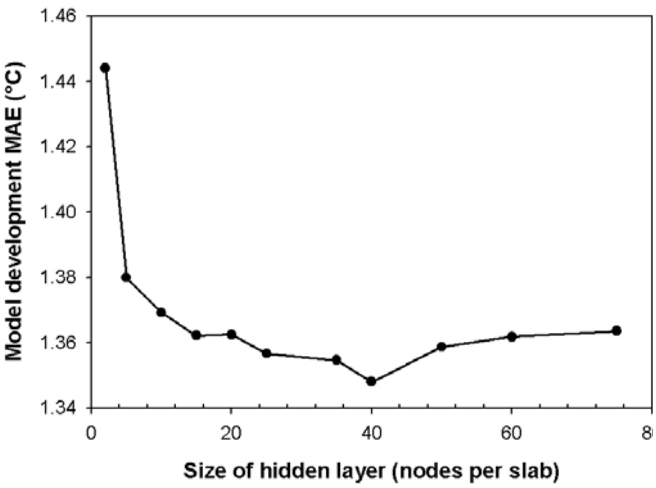

Fig. 5 A comparison of four-hour prediction models distinguished by hidden layer size using minimum-error, multiple-network evaluation. Each point corresponds to the minimum MAE obtained over 30 networks instantiating each model. All models use three equally-sized slabs per hidden layer

level did not reduce average prediction error for any of the models considered. The time-consuming nature of the training process precluded the possibility of evaluating all possible models.

To establish a direct comparison between the models developed here and those obtained by Jain [4], 30 networks were instantiated for each prediction period between one and 12 hours. For each horizon, the network having the lowest MAE for the development set was selected to represent the model. The selected networks were evaluated with the same 2001-2003 weather data for the Brunswick, Byron, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Homerville, Nahunta, Newton, Valdosta, and Vidalia sites used by Jain

[4]. These data did not include input patterns with a temperature greater than 20°C at the time of prediction.

The prediction accuracies of the best ANN models developed in this study are compared to those obtained by Jain [4] in Table I. The models developed in the current research made use of seasonal input terms, 24 hours of prior observations, and 120 hidden nodes and led to an improvement in model MAE over all horizons. For example, the four-hour prediction MAE of 1.40°C is an improvement of 0.20°C, or 12.5%, compared to the previous model. The MAE improvements at the one-, eight-, and 12-hour horizons of 0.09°C, 0.17°C, and 0.16°C respectively, do not provide a clear pattern relative to forecast horizon. However, the percent improvement in the MAE compared to the previous model decreases as the prediction period increases, from a more than 14% improvement at the one-hour horizon to less than 6% at the 12-hour horizon. The new networks were also evaluated for a data set consisting of the same sites with patterns from 2004-2005. For this set the magnitudes of the errors were, in general, slightly smaller than those associated with the 2001-2003 period.

TABLE I
COMPARISON OF MODEL PREDICTION ACCURACY OVER THE EVALUATION DATASET

Horizon length, hours	Previous model* 2001-3 MAE, °C	Current model 2001-3 MAE, °C	Improvement, °C	Improvement, %	Current model 2004-5 MAE, °C
1	0.62	0.53	0.09	14.5%	0.53
2		0.88			0.86
3		1.17			1.12
4	1.60	1.40	0.20	12.5%	1.34
5		1.62			1.55
6		1.81			1.72
7		1.99			1.87
8	2.30	2.13	0.17	7.4%	2.01
9		2.24			2.09
10		2.36			2.19
11		2.44			2.25
12	2.69	2.53	0.16	5.9%	2.33

*Jain [4]

The distribution of prediction errors across all horizons is centered near zero, while the variance of these error distributions increases relative to horizon length. The increased divergence between predicted and observed temperatures at longer horizons is apparent in the plots of Fig. 6. As prediction horizon increases, so does deviation from the line of perfect fit. The trend holds, specifically, in cases where a model fails to predict freezing temperatures. At the other extreme, the use of a logistic activation function in the output node, and the inverse of the scaling function to convert the output to a temperature, placed an upper bound on the model predictions. Because the scaling range was smaller than the output range of a logistic node, this bound was several degrees higher than the 20°C threshold used to select observations for the development and evaluation sets. As a result, models were constrained from predicting temperatures above 25°C. At temperatures near 25°C, models are more likely to under-predict. As the prediction horizon increases, the number of observed temperatures above this threshold increases. The February 25 – March 1, 2002 time frame for the Fort Valley site provides an illustration of the relative performance of the final models. This period included three freeze events during the mornings of February 27 – March 1. The first of these freezes occurred shortly after 0200 on the 27th. This freeze, however, was not predicted by the 12-hour model (Fig. 7a). Instead, the 12-hour model predicted a near freeze shortly after the observed temperature dropped below freezing. The model performed much better over the second freeze period: both the time of onset and the severity of the freeze were accurately predicted. While the predicted onset of the third freeze was off by several hours, it still indicated an approaching, sustained freeze more than six hours prior to the temperature falling below zero.

The eight-hour model predicted a brief freeze during the morning of February 27th (Fig. 7b). Though the time of onset and severity of the freeze were not perfect, the model predicted several hours of near-freezing temperatures, a noticeable improvement over the 12-hour model. The predictions for the second and third freeze events were similar to the 12-hour model. As a practical matter, the prediction of a near-freeze event by a long-horizon model would alert the user to the possibility of damaging temperatures.

The four-hour model predicted the first freeze event, though the time of onset was off by nearly three hours (Fig. 7c). Subsequently, however, the model's prediction of the severity of the first freeze event was quite close to the true minimum temperature. The four-hour model also correctly predicted the time of onset of the second freeze, which began later that evening and lasted well into the 28th. The freeze event ending March 1st was predicted with much better accuracy than either the 12- or eight-hour models managed, though time of onset was two hours late.

The most useful measure of model performance, however, comes from evaluating the sequence of 12 predictions leading to severe freeze events such as those in February and March 2002. Such a sequence is comprised of a chain of predictions generated at the same time for all 12 prediction horizons. The observed temperatures for Fort Valley during the period from 1400, February 28 to 1000, March 1, 2002 and the series of chained predictions generated at 1600 on February 28 are shown in Fig. 8. These predictions suggest a shallow freeze beginning sometime between 0300 and 0400 the following morning. In fact, overnight temperatures would dip below freezing by 2200 and bottom out below -4°C. This early, imprecise, series of predictions is subsequently refined in the presence of new data. The user, already alerted to the potential of damaging temperatures, could receive a much more accurate sequence of predictions four hours later. The

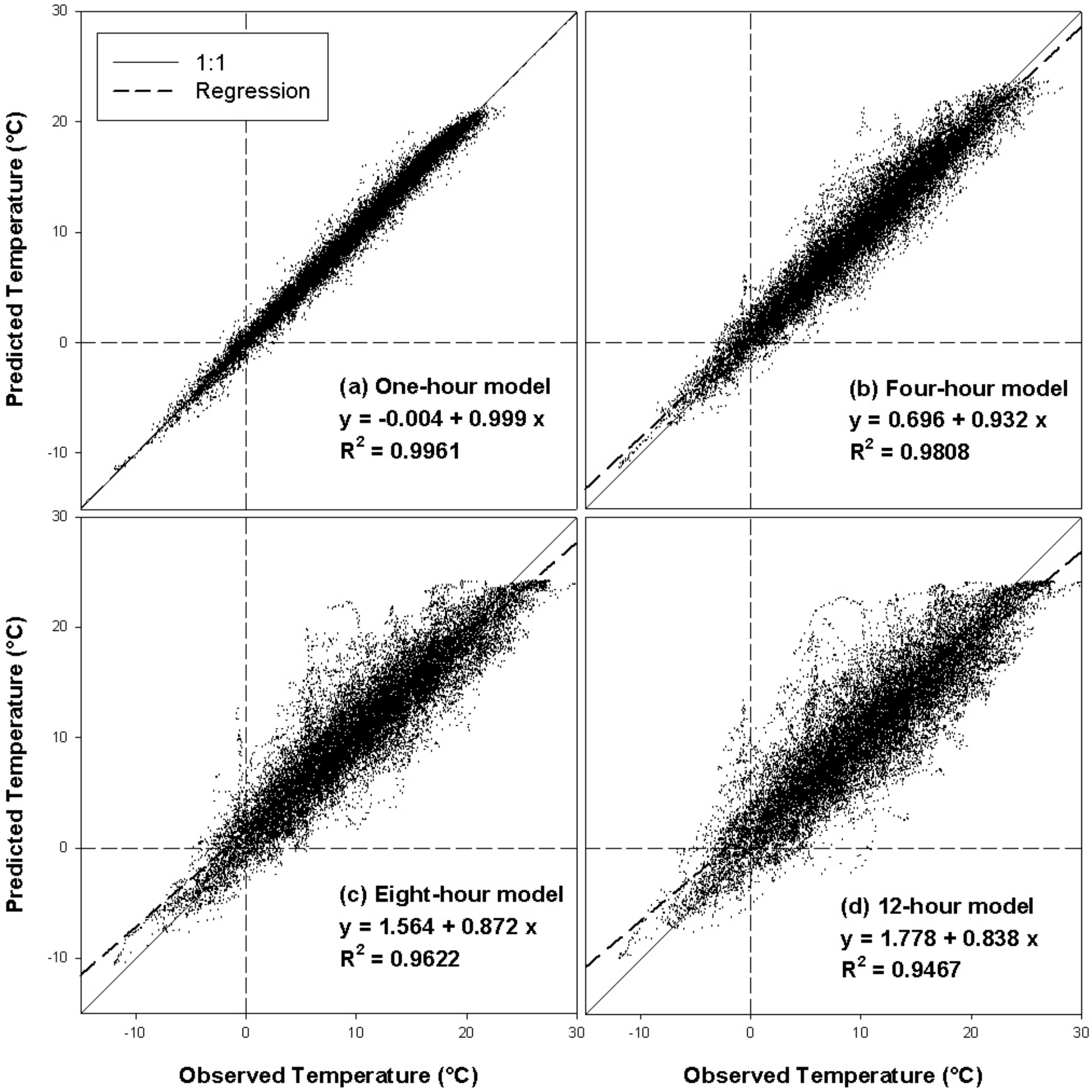

Fig. 6 A comparison of predicted and observed temperatures for the 2001-2003 evaluation set for the final (a) one-hour model, (b) four-hour model, (c) eight-hour model, and (d) 12-hour model. A solid diagonal line indicates a hypothetical perfect model

predictions made using the data available at 2000 on February 28 correctly indicate a sustained freeze lasting at least until the end of the 12-hour horizon (Fig. 8). These sequences of predictions show that the model was able to provide useful and actionable information to its users, even when early predictions were imperfect. The retrospective application of the final temperature prediction models to patterns from outside the development set suggests that users, once made aware of freezing or near-freezing temperature predictions, would be well served by checking for updated predictions throughout the day.

IV. Conclusions

The research presented in this paper explored improvements for the ANN models that are currently used to predict temperature for the Georgia AEMN data. Improvements included larger training set sizes, seasonal input terms, an increased duration of prior observations, and varying the size of the hidden layer. Increases to the size of the training set slightly reduced the prediction errors. However, the inclusion of seasonal variables corresponding to membership in the fuzzy sets winter, spring, summer, and fall did improve model accuracy, even though all observations were from the January-April period. Similar improvements resulted from extending the duration of historical data in the input vector from six to 24 hours. Models with a hidden layer with 40 nodes per slab were more accurate than other models over repeated instantiations.

The results of this work suggest avenues for further study. The introduction of seasonal terms may provide a means of implementing an accurate year-round temperature prediction model. Likewise, ensemble network approaches are worth investigating, as networks with similar MAEs over the same prediction horizon often make different predictions. Finally, when applied to data-rich environments, a clear distinction should be maintained between abstract neural network models and actual instantiations of these models. The performance of a single instantiated network is not likely to be a valid 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.

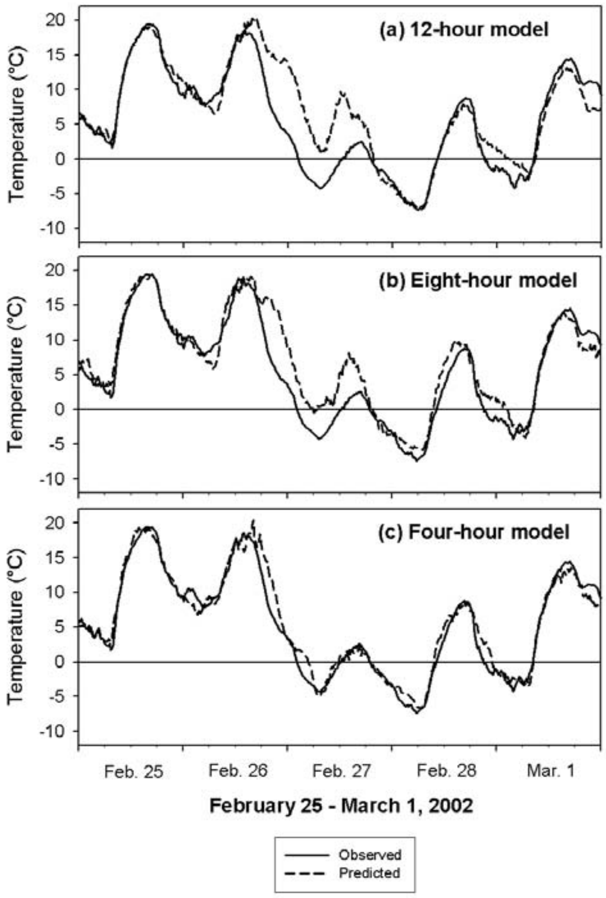

Fig. 7 Time-series plots of observed and predicted temperatures from the final (a) one-hour model, (b) four-hour model, (c) eight-hour model, and (d) 12-hour model for the period of February 25-March 1, 2002

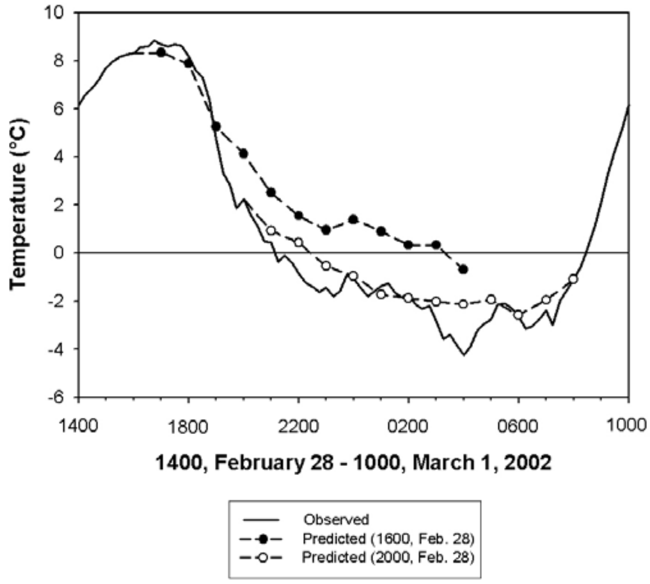

Fig. 8 A time-series plot of observed temperatures and 12-hour prediction tracks during February 28-March 1, 2002

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] 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.

[3] 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, 2003, ASAE Paper 03-3075.

[4] A. Jain, "Frost prediction using artificial neural networks: A temperature prediction approach," M.S. thesis, Artificial Intelligence Center, University of Georgia, Athens, GA, 2003.

[5] Ramyaa, "Frost prediction using artificial neural networks: A classification approach," M.S. thesis, Artificial Intelligence Center, University of Georgia, Athens, GA, 2004.

[6] 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.

[7] S. Haykin, *Neural networks: a comprehensive foundation*, 2nd edition. Upper Saddle River, NJ: Prentice Hall, 1998, pp. 161-175.

[8] *Manual of NeuroShell 2*, Ward Systems Group, Frederick, MD, 1993.

[9] 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.