

Hourly temperature forecasting using abductive networks

R.E. Abdel-Aal*

Center for Applied Physical Sciences, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Received 8 November 2003; accepted 2 April 2004

Abstract

Hourly temperature forecasts are important for electrical load forecasting and other applications in industry, agriculture, and the environment. Modern machine learning techniques including neural networks have been used for this purpose. We propose using the alternative abductive networks approach, which offers the advantages of simplified and more automated model synthesis and transparent analytical input–output models. Dedicated hourly models were developed for next-day and next-hour temperature forecasting, both with and without extreme temperature forecasts for the forecasting day, by training on hourly temperature data for 5 years and evaluation on data for the 6th year. Next-day and next-hour models using extreme temperature forecasts give an overall mean absolute error (MAE) of 1.68 °F and 1.05 °F, respectively. Next-hour models may also be used sequentially to provide next-day forecasts. Performance compares favourably with neural network models developed using the same data, and with more complex neural networks, reported in the literature, that require daily training. Performance is significantly superior to naive forecasts based on persistence and climatology.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Abductive networks; Neural networks; Neural network applications; Temperature forecasting; Forecasting; Modeling; Hourly temperatures; Artificial intelligence

1. Introduction

Accurate forecasting of hourly air temperatures has a number of important applications in industry, agriculture, and the environment. Many short-term load forecasting (STLF) schemes for power utilities require hourly temperature forecasts (Fan and McDonald, 1994; Hwang et al., 1998; Khotanzad et al., 1998; Sharif and Taylor, 2000; Xu and Chen, 1999). Such forecasts are also used for predicting the gas send-out for a gas utility (e.g., Pattern Recognition Technologies ANNGSF) and for forecasting the 1 h-ahead heat load for a district heat load network (Seppälä et al., 2000). In agriculture, hourly air temperature forecasts can be used by disease warning systems and pest management schemes to predict conditions that are favourable for disease development in crops and scheduling appropriate actions such as spraying protective fungicides (Francis, 2000; Kim et al., 2002). Road weather information systems utilize forecasted hourly air tem-

peratures for predicting road surface temperatures (Bogren and Gustavsson, 1994).

Temperature is the most important weather parameter affecting electric load generated by power utilities in many parts of the world, and therefore forecasted temperatures constitute a basic ingredient in load forecasting schemes. Forecasts for extreme (minimum and maximum) daily temperatures are provided by many weather services, but these alone are useful only for predicting the daily peak load. However, forecasting the full 24-h load curve is important for many scheduling and network analysis functions in power utilities. Since high–low temperature forecasts are usually provided without specifying the times at which they occur, this precludes their use to generate the hourly load curve through regression and interpolation (Hippert et al., 2000). Schemes for hourly temperature forecasting have been developed in the context of short-term load forecasting and in some cases form an integrated part of the load forecaster (e.g., Fan and McDonald, 1994; Khotanzad et al., 1998). In other agricultural and environmental applications, even high–low temperature forecasts that are specific to the site of

*Tel.: +966-3-860-4320; fax: +966-3-860-4281.

E-mail address: radwan@kfupm.edu.sa (R.E. Abdel-Aal).

interest may not be available, and it is often preferred that temperature forecasts rely only on parameters that are available or can be measured on-site.

Schneider et al. (1985) obtained hourly temperature forecasts by first fitting a two-harmonics Fourier model to the temperature data of the past 21 days to produce a temperature day profile. Hourly forecasts were then obtained by stretching/contracting this profile so that its minimum and maximum points coincide with those forecasted for the day by the weather service. Fan and McDonald (1994) adopt a similar approach but use a day profile that is initialized with historical average data and updated by exponential smoothing with a time constant of 28 days. The complex and nonlinear nature of temperature variations and the abundance of historical data suggest that computational intelligence data-based modeling techniques would be good candidates for solving the temperature forecasting problem. In the load forecasting arena, the use of neural networks temperature forecasting has been a natural extension to their use in load forecasting. The ANNSTLF neural network load forecasting system (Khotanzad et al., 1998) embodies a 7-day-ahead neural network hourly temperature forecaster that uses an adaptive daily update of the weights (Khotanzad et al., 1996). The day module of such forecaster uses back propagation neural networks to forecast the 24 hourly temperatures for day (d) using 28 inputs which include the minimum and maximum temperatures measured for day ($d - 2$), the 24 hourly temperatures for day ($d - 1$), and the forecasted minimum and maximum temperatures for day (d). One drawback to this approach is the large size of the neural networks involved, which implies a large number of weights to be estimated. The large input dimensionality relative to the number of training records may cause the estimation problem to be ill-posed, resulting in unstable networks for typical sizes of training sets (Hippert et al., 2000). The resulting overfitting may also degrade model generalization, thus yielding poor out-of-sample forecasts (Hippert et al., 2001). To overcome these problems, smaller networks have been proposed that tackle the simpler problem of forecasting only the next-hour temperature. Lanza and Cosme (2001) used a radial basis functions neural network to forecast the temperature on hour (h) from only three inputs including the temperature at hour ($h - 1$) and two hour indices of the sine/cosine form. With the fewer inputs, smaller databases can be used to train the network, and in this case a sliding window of 28 days was found sufficient. In many situations, however, full forecasts of the 24 next-day temperatures are required. This may be achieved through iterative use of the same next-hour forecaster, with the output forecasted for hour (h) being recycled as input for forecasting hour ($h + 1$). This approach, however, may lead to the neural network forecaster

behaving chaotically (Hippert et al., 2000). To reduce this risk, these latter authors feed the neural network instead with crude forecasts estimated using an autoregressive (AR) model. Tassadduq et al. (2002) describe a back propagation neural network that uses only the temperature at a given hour to forecast the temperature at the same hour of the following day.

In general, the neural network approach also suffers from a number of limitations, including difficulty in determining optimum network topology and training parameters (Alves da Silva et al., 2001). There are many choices to be made in determining numerous critical design parameters with little guidance available (Hippert et al., 2001), and designers often resort to trial and error approaches (Charytoniuk and Chen, 2000; Tassadduq et al., 2002) which can be tedious and time consuming. Such design parameters include the number and size of the hidden layers, the type of neuron transfer functions for the various layers, the learning rate and momentum coefficient, and training stopping criteria to avoid overfitting and ensure adequate generalization with new data. Another limitation is the black box nature of neural network models that give little insight into the modeled relationship and the relative significance of various inputs, thus providing poor explanation facilities (Matsui et al., 2001). The acceptability of, and confidence in, automated forecasting tools in operational environments appear to be related to their transparency and their ability to justify results to human experts (Lewis, III, 2001).

To overcome such limitations, we propose using abductive networks (Montgomery and Drake, 1990) as an alternative machine learning approach to hourly temperature forecasting. We have previously used this approach to model and forecast the monthly domestic energy consumption (Abdel-Aal et al., 1997) and in forecasting the minimum (Abdel-Aal and Elhadidy, 1994) and maximum (Abdel-Aal and Elhadidy, 1995) daily temperatures. The method has also been used by Fulcher and Brown (1994) in predicting temperature distributions at data-deficient sites based on detected similarities with data-rich sites. Compared to neural networks, abductive networks offer the advantages of faster model development requiring little or no user intervention, faster convergence during model synthesis without the problem of getting stuck in local minima, automatic selection of effective input variables, and automatic configuration of the model structure (Alves da Silva et al., 2001). Using the approach on a time series problem led to lower mean square errors and simpler models as compared to back propagation neural networks (Tenorio and Lee, 1989). With the model represented as a hierarchy of polynomial expressions, resulting analytical model relationships can provide insight into the modeled phenomena, highlight contributions of various inputs, and allow comparison with previously used empirical or statistical models. The

technique automatically avoids over-fitting by using a proven regularization criterion based on penalizing model complexity (Montgomery and Drake, 1990) without requiring a dedicated validation dataset during training, as is the case with many neural network paradigms.

Following a brief description of abductive network modeling in Section 2, the temperature dataset used is described in Section 3. Next-day hourly temperature forecasters, that predict the full 24-h temperature curve for a full day in one step at the end of the preceding day, are described in Section 4. Abductive network models were developed and evaluated both with and without extreme temperature forecasts for the forecasting day. Performance of representative models was compared with that of the corresponding neural network models developed using the same data. Next-hour temperature forecasters that predict the temperature hour-by-hour utilizing all data available up to the forecasting hour are presented in Section 5. Results are also given when such models are used sequentially to forecast the full next-day temperature curve.

2. Aim abductive networks

Abductory inductive mechanism (AIM) (AbTech, 1990) is a supervised inductive machine-learning tool for automatically synthesizing abductive network models from a database of inputs and outputs representing a training set of solved examples. As a group method of data handling (GMDH) algorithm (Farlow, 1984), the tool can automatically synthesize adequate models that embody the inherent structure of complex and highly nonlinear systems. The automation of model synthesis not only lessens the burden on the analyst but also safeguards the model generated from being influenced by human biases and misjudgements. The GMDH approach is a formalized paradigm for iterated (multi-phase) polynomial regression capable of producing a high-degree polynomial model in effective predictors. The process is ‘evolutionary’ in nature, using initially simple (myopic) regression relationships to derive more accurate representations in next iterations. The algorithm selects polynomial relationships and input combinations that minimize the prediction error in each phase. AIM builds networks of various types of polynomial functional elements, based on prediction performance. The network size, element types, connectivity, and coefficients for the optimum model are automatically determined using well-proven optimization criteria, thus reducing the need for user intervention compared to neural networks. This simplifies model development and considerably reduces the learning/development time and effort. The models take the form of layered feed-forward abductive networks of functional elements (nodes) (AbTech, 1990), see Fig. 1. Elements in the first layer

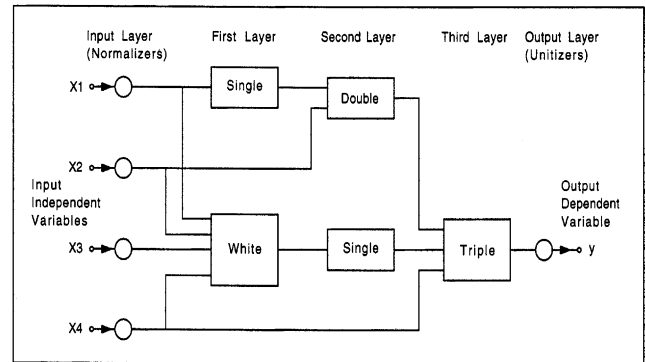


Fig. 1. A typical AIM abductive network model showing various types of functional elements.

operate on various combinations of the independent input variables (X 's) and the element in the final layer produces the predicted output for the dependent variable y . In addition to the main layers of the network, an input layer of normalizers convert the input variables into an internal representation as Z scores with zero mean and unity variance, and an output unitizer unit restores the results to the original problem space. The used version of AIM supports the following main functional elements:

(i) A white element which consists of a constant plus the linear weighted sum of all outputs of the previous layer, i.e.

$$\begin{aligned} \text{“White” Output} &= W_0 + W_1X_1 + W_2X_2 \\ &+ W_3X_3 + \dots + W_nX_n, \end{aligned} \quad (1)$$

where X_1, X_2, \dots, X_n are the inputs to the element and W_0, W_1, \dots, W_n are the element weights.

(ii) Single, double, and triple elements which implement a third-degree polynomial expression with all possible cross-terms for one, two, and three inputs, respectively, for example,

“Double” Output

$$\begin{aligned} &= W_0 + W_1X_1 + W_2X_2 + W_3X_1^2 + W_4X_2^2 \\ &+ W_5X_1X_2 + W_6X_1^3 + W_7X_2^3. \end{aligned} \quad (2)$$

The database of input–output solved examples is split into a training set and an evaluation set. AIM uses the training set to synthesize the model network layer by layer until no further improvement in performance is possible or a preset limit on the number of layers is reached. Within each layer, every element is computed and its performance scored for all combinations of allowed inputs. The best network structure, element types and coefficients, and connectivity are all determined automatically by minimizing the predicted squared error (PSE) criterion (Barron, 1984), which eliminates the problem of determining when to stop training in neural networks. This criterion selects the most accurate model that does not overfit the training data to strike a balance between the accuracy of the

model in representing the training data and its generality which allows it to fit yet unseen future data. The user may optionally control this trade-off between accuracy and generality using the complexity penalty multiplier (CPM) parameter (AbTech, 1990). Larger values than the default value of 1 lead to simpler models that are less accurate but may generalize well with previously unseen data, while lower values produce more complex networks that may overfit the training data and degrade actual prediction performance.

3. The dataset

The dataset used consists of measured hourly temperature data for the Puget power utility, Seattle, USA, as integer values over the period 1 January 1985 to 12 October 1992. The set is made available in the public domain by Professor A. M. El-Sharkawi, University of Washington, Seattle, USA (El-Sharkawi, 2002). We used the data for the first 5 years (1985–1989) for model synthesis and those of the following year (1990) for model evaluation. Data for 5 years were considered sufficient for training the required hourly temperature forecasting models, and therefore data for 1991 and the partial data available for 1992 were not utilized by the work described in this paper. A few missing values, indicated as 0's in the original dataset, were filled-in by interpolating between neighbouring values. The dataset contains no forecasted values for tomorrow's minimum and maximum daily temperatures. Table 1 lists summary statistics of the hourly temperature data for both the training and evaluation datasets, giving the overall mean and standard deviation values for each hour of the day.

4. Next-day hourly temperature forecasters

4.1. Using forecasts for next day extreme temperatures

We have developed 24 models for forecasting the hourly temperatures for the following day (d) in one step at the end of the preceding day ($d-1$). A model is dedicated for forecasting the temperature, $ET(d, h)$, for each hour of the day. Each of the 24 models was trained using 1825 data records for 5 years (1985–1989) and evaluated on 365 records for the year 1990. Unless specified otherwise, training was performed with the

Table 1

Summary statistics of the hourly temperatures data for both the training set (1825 records) and the evaluation set (365 records)

Hour of day (h)	Training dataset (1985–1989)		Evaluation dataset (1990)	
	Mean ($^{\circ}\text{F}$)	Standard deviation ($^{\circ}\text{F}$)	Mean ($^{\circ}\text{F}$)	Standard deviation ($^{\circ}\text{F}$)
1	48.97	9.74	49.26	10.50
2	48.33	9.37	48.73	10.23
3	47.73	9.13	48.14	9.87
4	47.23	8.95	47.64	9.69
5	46.80	8.76	47.25	9.48
6	46.49	8.63	47.08	9.56
7	46.97	8.95	47.52	9.78
8	47.91	9.47	48.40	10.23
9	49.46	9.90	49.79	10.82
10	51.24	10.29	51.44	11.28
11	53.07	10.70	53.15	11.73
12	54.79	11.22	54.81	12.16
13	56.25	11.81	56.05	12.62
14	57.33	12.42	57.06	13.21
15	58.00	13.02	57.63	13.76
16	58.07	13.64	57.79	14.23
17	57.57	14.13	57.20	14.67
18	56.57	14.08	56.33	14.64
19	55.31	13.56	55.16	14.11
20	53.80	12.56	53.72	13.14
21	52.18	11.34	52.36	12.02
22	51.22	10.62	51.36	11.40
23	50.43	10.33	50.61	11.11
24	49.60	10.00	49.83	10.78

default value $\text{CPM} = 1$ for the complexity penalty multiplier. All models use the same set of inputs which includes: 24 hourly temperatures on day ($d-1$) (T_1, T_2, \dots, T_{24}), the measured minimum (T_{\min}) and maximum (T_{\max}) air temperatures on day ($d-1$), the forecasted minimum (ET_{\min}) and maximum (ET_{\max}) air temperatures for day (d). T_{\min} and T_{\max} were taken as the minimum and maximum values of the 24 hourly temperatures provided for the day. In the absence of forecasted data for the minimum (ET_{\min}) and maximum (ET_{\max}) air temperatures for the forecasting day, we used actual values measured on that day instead, which would be the case with ideal forecasts for the extreme temperature. We also investigate the effect of introducing Gaussian noise depicting temperature forecasting errors that would be present in practice. A record in the training dataset for the model for hour h ($h = 1, 2, \dots, 24$) includes 28 input variables and takes the following form:

Inputs			Output
24 hourly temperatures for day ($d-1$)	Extreme Temperatures for day ($d-1$)	Forecasted Extreme Temperatures for day (d)	Temperature for hour (h) on day (d)
T_1, T_2, \dots, T_{24}	T_{\min}, T_{\max}	$\text{ET}_{\min}, \text{ET}_{\max}$	$T(d, h)$

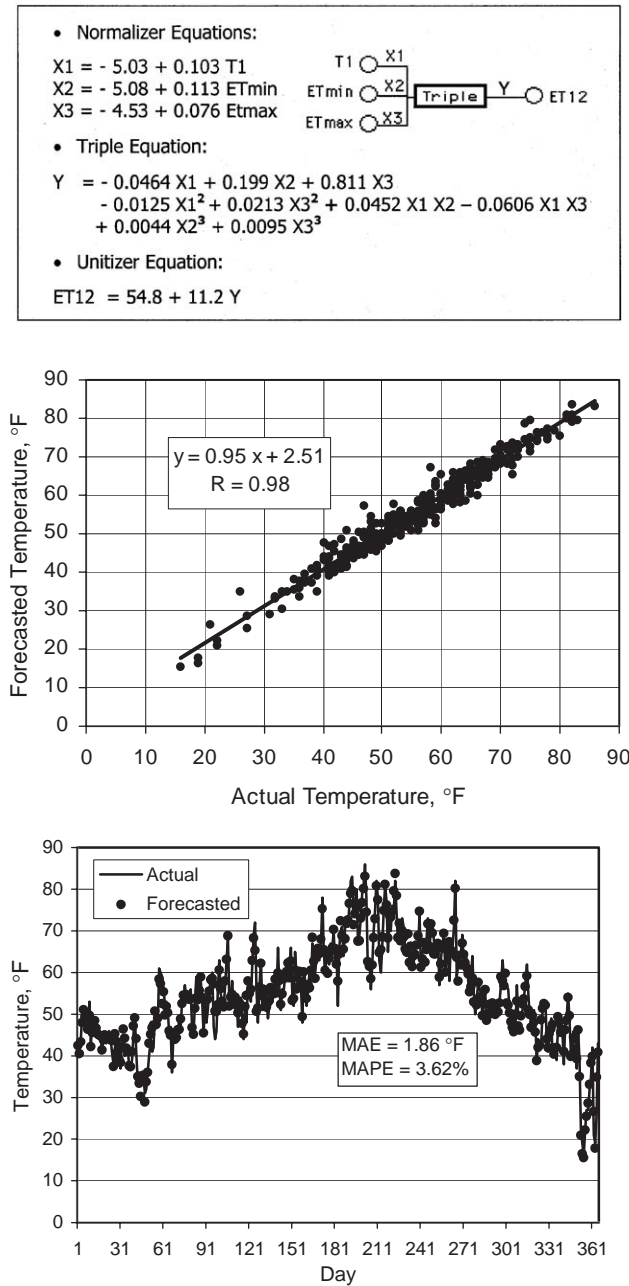


Fig. 2. Structure and performance over the evaluation year for the next-day forecaster for hour 12 with extreme temperature forecasts.

The top part of Fig. 2 shows the abductive network model synthesized for forecasting the temperature at hour 12 (midday), ET_{12} . This is a single-element nonlinear model that selects only the measured temperature at hour 1, T_1 , and the two forecasted extreme temperatures, ET_{min} and ET_{max} , as the essential predictors for ET_{12} out of the 28 inputs provided. Automatic selection of only relevant inputs simplifies the resulting models, allows them to execute faster, and prevents noise and uncertainty in the unused inputs from degrading model performance. The resulting

simpler models also provide better insight into the modeled phenomenon. A corresponding standard neural network model would require all 28 inputs, giving little indication of the relative importance of the various inputs. When required, such information must be extracted by subsequent analysis of the trained network, e.g. by summing absolute weight values (Roadknight et al., 1997). Analyzable networks were also described for this purpose (Matsui et al., 2001). To determine the importance of input parameters for speech synthesis by neural networks, Sonntag et al. (1997) adopt the manual approach of comparing the performance of individual models trained on single inputs or on groups of inputs that lack only one input.

The use of the Triple element indicates the nonlinear nature of the model. Substituting symbolically for the five equations of functional elements of the model shown at the top part of Fig. 2 gives the following analytical relationship for the forecasted temperature in terms of the three inputs:

$$ET_{12} = -12.58281 + 0.14334T_1 + 0.39523ET_{min} + 1.28334ET_{max} - 1.48526 \times 10^{-3}T_1^2 - 9.58987 \times 10^{-3}ET_{min}^2 - 6.97403 \times 10^{-3}ET_{max}^2 + 7.11060 \times 10^{-5}ET_{min}^3 + 4.67070 \times 10^{-5}ET_{max}^3 + 5.89213 \times 10^{-3}T_1ET_{min} - 5.31302 \times 10^{-3}T_1ET_{max}. \quad (3)$$

Analytical model relationships in the form of Eq. (3) help identify the model order and determine the relative contributions of various inputs and the combinatory effects between the inputs, thus enhancing model interpretation and understanding. Simpler equations can be obtained from simpler models synthesized using larger CPM values, to be presented later in this section. In the field of environmental studies, model equations were found particularly useful in generating hypothesis regarding the phenomenon under investigation, which can then be tested by performing the relevant experiments (Roadknight et al., 1997). Generating manageable analytical expressions from neural network models is a tedious process involving network pruning through iterated removal of weak weight links and testing the resulting network for adequate performance (Roadknight et al., 1997). Resulting model equations can be compared with those derived using first-principles, empirical, or statistical models. For example, simplifying Eq. (3) through omitting the nonlinear terms leaves a linear model consisting of only the first four terms. Such a simplified model can be compared with a linear regression model developed using the training data.

Fig. 2 shows also the performance of the model in the form of scatter and time series plots of the actual and forecasted temperatures for hour 12 over the evaluation year. The scatter plot shows a best line fit and the value

Table 2

Summary of the abductive network models for the 24 next-day temperature forecasters with extreme temperature forecasts

Day (<i>d</i>) forecasting hour	Model input (<i>s</i>)			Temperature time lags selected	Model structure
	Day (<i>d</i> -1) temperature at hours:	Temperature extremes			
		Day (<i>d</i> -1)	Day (<i>d</i>)		
14			ET _{max}		
15					
16					
17					
1	24		ET _{min}	1	
2				2	
3	24		ET _{min} , ET _{max}	3	
4				4	
6				6	
7				7	
8				8	
9				9	
11				33	
12				35	
13				36	
18				35	
19	35				
21	21				
22	22				
23	23				
5	24	<i>T</i> _{max}	ET _{min}	5	
10	24,1		ET _{min} , ET _{max}	10,33	
20	20,14,8			24,30,36	
24	24,15			24,33	

of Pearson's correlation coefficient as 0.98, and the time series plot shows a mean absolute error (MAE) of 1.86 °F and a mean absolute percentage error (MAPE) as 3.62%. A more elaborate assessment of the forecasting performance can be made using contingency tables and statistical criteria such as bias, the Heidke's skill score, and the Priestley skill score (Pasini et al., 2001; Abdel-Aal and Elhadidy, 1995). Performance of the abductive model for the temperature at hour 12 was compared with a back propagation neural network model developed using the PathFinder software for Windows. The model was trained and evaluated using the same data used for the abductive model, with 20% of the training data reserved for cross validation. The 28-6-1 neural model has one hidden layer containing 6 neurons with a sigmoid transfer function. MAE and MAPE values for the neural model are 2.02 °F and 3.91%, respectively, indicating inferior performance

compared to the abductive model. It should be noted that while the neural model requires all 28 inputs, the abductive model uses only three inputs as shown in Fig. 2.

Table 2 summarizes the 24 hourly models, listing the model inputs selected and the corresponding time lags in the temperature time series and showing a sketch of the model structure. Models for hours 14–17 are of the 'wire' type where the output is a linear relationship of a single input (ET_{max}). Temperature at such hours is highly correlated with *T*_{max} for the forecasting day, with values for the Pearson's correlation coefficient for the training data being around 0.99. Temperature extremes on the forecasting day are utilized by all models while those on the previous day are used only by one model. The temperature time series is utilized by 20 models, mostly using a single time lag. The last hourly temperature on the preceding day (*T*₂₄) is used by 14

Table 3
Performance of the next-day temperature forecasting models over the evaluation year

Forecasting hour (<i>h</i>)	With extreme temperature forecasts		Without extreme temperature forecasts	
	MAE (°F)	MAPE (%)	MAE (°F)	MAPE (%)
1	0.94	2.01	1.00	2.16
2	1.16	2.52	1.30	2.78
3	1.20	2.67	1.52	3.35
4	1.22	2.72	1.78	4.02
5	1.28	2.85	1.89	4.33
6	1.38	3.11	2.12	4.88
7	1.48	3.32	2.11	4.83
8	1.65	3.54	2.29	5.08
9	1.73	3.57	2.54	5.46
10	1.82	3.70	2.79	5.77
11	1.91	3.79	3.16	6.31
12	1.86	3.62	3.47	6.88
13	1.64	3.22	3.36	6.46
14	1.51	2.94	3.82	7.28
15	1.50	2.93	4.05	7.72
16	1.42	2.85	4.20	8.02
17	1.70	3.51	4.49	8.64
18	1.87	3.72	4.27	8.34
19	2.19	4.37	3.94	7.92
20	2.12	4.42	3.78	7.85
21	2.08	4.22	3.60	8.01
22	2.05	4.25	3.58	7.99
23	2.28	4.86	3.71	8.59
24	2.29	5.04	3.73	8.74
Average	1.68	3.49	3.02	6.31

models. Model structures are fairly simple, with 21 models being single element, single layer. The most complex model is a five-input, two-element, two-layer model for hour 20. The left-hand side of Table 3 lists the MAE and MAPE values for all hours, giving the overall average values for the evaluation year as 1.68 °F and 3.49%, respectively. The MAPE value compares favourably with that of 5.20% quoted for a hybrid forecasting scheme consisting of an AR model followed by a back propagation neural network, which was developed with 20 days of data and evaluated using data for the following 20 days (Hippert et al., 2000). The MAE is comparable with the value of 1.48 °F given as the average error at eight power utilities where the error values ranged from 1.04 °F to 2.03 °F using neural network hourly temperature forecasters that required daily updating of the weights (Khotanzad et al., 1996). Full-day temperature curves were forecasted using all 24 abductive network models for four days marking the beginning of the four seasons of the evaluation year, and the results are shown in Fig. 3. Fig. 4 plots the average MAE and MAPE values over each calendar month of the evaluation year. MAE values are below 2 °F for all months.

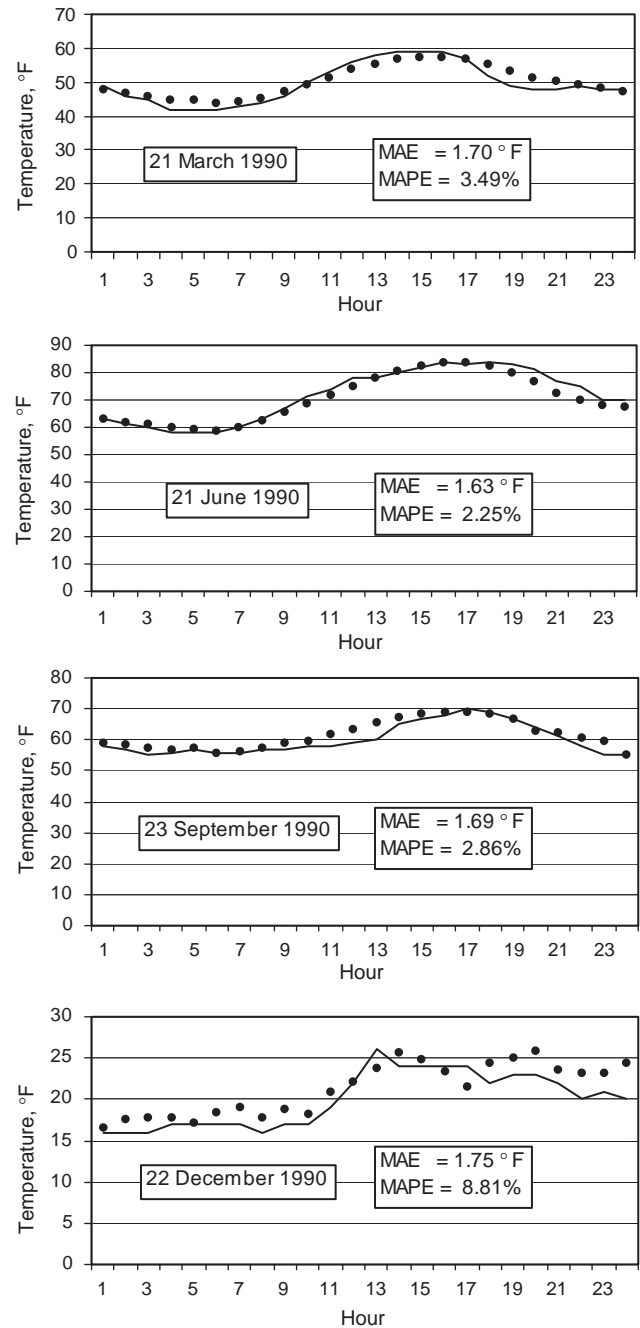


Fig. 3. Performance of the 24 next-day hourly temperature forecasters (using extreme temperature forecasts) on four days representing the beginning of the four seasons of the evaluation year. Continuous line: Actual, solid circles: Forecasted.

We have investigated the effect of simulated errors in the ideal extreme temperature forecasts ET_{min} and ET_{max} on model performance. As seen from Table 2, the model for hour 12 is an example of 17 forecasters that use both ET_{min} and ET_{max} , and would therefore be affected most by such errors. Simulated Gaussian random errors of zero mean and standard deviation σ were added to the two ideal temperature forecasts in

both the training and evaluation datasets used to develop the forecaster for hour 12. The MAE of 1.86°F for the noiseless case increased to 2.34°F for $\sigma = 2^{\circ}\text{F}$, indicating the importance of good forecasting accuracy for the extreme temperatures and that in practice errors in such forecasts are expected to degrade the performance of such models.

The effect of varying the complexity of the resulting forecasting models was investigated for the model for hour 12. Table 4 shows the structure and performance of

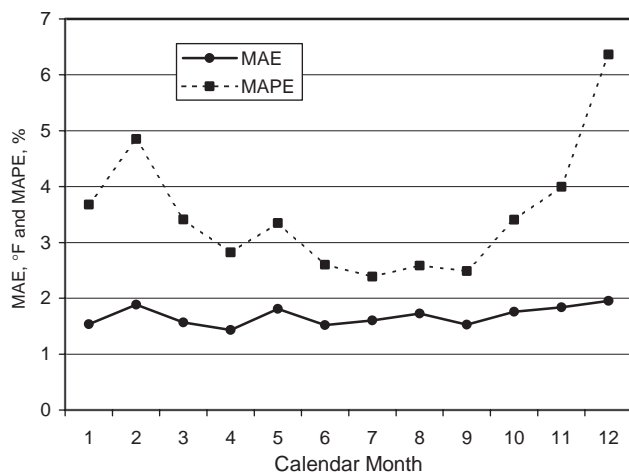


Fig. 4. Plots of average MAE and MAPE values for each calendar month of the evaluation year for next-day forecasters with extreme temperature forecasts.

the resulting more complex model with $\text{CPM} = 0.2$ and the simpler model with $\text{CPM} = 5$, in comparison with the default model synthesized with $\text{CPM} = 1$. It is noted that input ET_{\max} features in all three models which indicates its importance in explaining the modeled output. The level of model complexity varies widely from a 32-input, four-layer nonlinear model with $\text{CPM} = 0.2$ to a simple ‘wire’ model at $\text{CPM} = 5$. As expected, the default model with $\text{CPM} = 1$ gives optimum performance. The highly complex model with $\text{CPM} = 0.2$ seems to overfit the training data giving poorer generalization with the evaluation data, while that with $\text{CPM} = 5$ oversimplifies the model function. As indicated in the table, more complex models require longer train times.

Table 5 lists hourly forecasts over the evaluation year based on persistence and climatology. Persistence forecasts were obtained by assuming the temperature at a given hour of the forecasting day to be equal to the temperature measured at the preceding hour on the same day. With climatology, the forecasted value was taken as the average of the five temperatures measured at the same hour on the five calendar days similar to the forecasting day in the 5 years (1985–1989) preceding the evaluation year. These are the same years used to train the AIM models. The table indicates that abductive network models provide significantly superior forecasting accuracy compared to both persistence and climatology, with average MAE values being 1.68°F as compared to 4.17°F and 5.13°F , respectively.

Table 4

Effect of the CPM parameter on the complexity and performance of next-day temperature forecasting models for hour 12 with extreme temperature forecasts

CPM	Model structure	Relative training time	MAE, ($^{\circ}\text{F}$)	MAPE, (%)
0.2		1.52	1.89	3.64
1		1.00	1.86	3.62
5		0.81	2.13	4.30

4.2. Without using forecasts for next day extreme temperatures

In some applications, accurate extreme temperature forecasts, ET_{\min} and ET_{\max} , on the forecasting day may not be available for the region of interest. We have developed a set of 24 next-day hourly temperature forecasters that do not require these two input variables. Table 6 shows a summary of the resulting models. Compared to those in Table 2, the models are generally more complex. For example, the three-input, one-element, one-layer model described in Section 4.1 for hour 12 is now a four-input, two-element, two-layer

Table 5
Performance of persistence and climatology forecasts over the evaluation year

Forecasting hour (h)	Persistence forecast		Climatology forecast	
	MAE (°F)	MAPE (%)	MAE (°F)	MAPE (%)
1	3.76	8.69	4.75	11.62
2	3.84	8.87	4.66	11.63
3	3.83	8.99	4.52	11.50
4	3.93	9.30	4.54	11.61
5	3.95	9.40	4.50	11.54
6	4.01	9.74	4.65	12.06
7	3.83	9.18	4.64	11.95
8	3.90	9.14	4.60	11.79
9	4.00	8.99	4.79	11.82
10	4.07	8.68	4.97	11.70
11	4.32	8.73	5.22	11.51
12	4.46	8.69	5.47	11.52
13	4.44	8.43	5.60	11.26
14	4.63	8.58	5.74	11.34
15	4.76	8.77	5.94	11.68
16	4.77	8.75	5.99	11.70
17	4.94	9.20	6.24	12.38
18	4.74	9.05	5.95	12.10
19	4.36	8.75	5.61	11.82
20	4.11	8.54	5.32	11.64
21	3.83	8.29	4.98	11.39
22	3.69	8.24	4.88	11.53
23	3.90	8.86	4.77	11.57
24	3.91	8.98	4.81	11.69
Average	4.17	8.87	5.13	11.68

the temperature time series, with considerably more time lags being utilized by the models. The model for hour 20 uses six time lags. Increased model complexity degrades forecasting performance as indicated by the two columns at the right hand side of Table 3. Forecasting errors without extreme temperature forecasts are nearly double those with ideal values for such forecasts, but this margin would be narrower in practice due to errors in realistic forecasts. For hour 12, the MAE and MAPE values increase from 1.86 °F and 3.62%, respectively, for the case with extreme temperature forecasts, to 3.47 °F and 6.88%, respectively, without such forecasts. Performance of the model for hour 12 was compared with a back propagation neural network model developed using the PathFinder neural network software for Windows. The model was trained and evaluated using the same data used to develop the abductive network model, with 20% of the training data reserved for cross validation. The 26-6-1 neural model has one hidden layer containing six neurons with a sigmoid transfer function. MAE and MAPE values for the neural model are 3.38 °F and 6.59%, respectively, indicating a slightly better performance compared to the abductive model. It should be noted that while the neural model requires all 26 inputs, the abductive model for hour 12 uses only four inputs as shown in Table 6.

5. Next-hour temperature forecasters

5.1. Using forecasts for next day extreme temperatures

We have developed 24 models for forecasting the temperature at the next hour (h) during day (d) using the full hourly temperature data on day ($d - 1$) (T_1, T_2, \dots, T_{24}) together with all available hourly temperatures on day (d) up to the preceding hours hour ($h - 1$) ($NT_1, NT_2, \dots, NT(h - 1)$), in addition to the measured minimum (T_{\min}) and maximum (T_{\max}) air temperatures on day ($d - 1$) and the forecasted minimum (ET_{\min}) and maximum (ET_{\max}) air temperatures on day (d), as described in Section 4.1 above. A record in the training dataset for the model for hour h ($h = 1, 3, \dots, 24$) takes the form:

Inputs				Output
24 hourly temperatures for day ($d - 1$)	$(h - 1)$ available hourly temperatures on day (d)	Extreme Temperatures for day ($d - 1$)	Forecasted Extreme Temperatures for day (d)	Temperature for hour (h) on day (d)
T_1, T_2, \dots, T_{24}	$NT_1, NT_2, \dots, NT_{(h-1)}$	T_{\min}, T_{\max}	ET_{\min}, ET_{\max}	$T(d, h)$

model. Absence of the important ET_{\min} and ET_{\max} inputs is compensated for by increased dependence on

Contrary to the case of next-day hourly forecasters, the number of hourly temperature inputs used here is

Table 6

Summary of the abductive network models for the 24 next-day temperature forecasters without extreme temperature forecasts

Day (d) forecasting hour	Model input(s): Day ($d-1$)		Temperature time lags selected	Model structure
	Temperature at hours:	Temp. extremes		
1	24		1	
2	24,17	T_{\min}	2,9	
3	24	T_{\min}, T_{\max}	3	
4			4	
5	24,6	T_{\max}	5,23	
6			6,24	
7			7,24	
8	24,8	T_{\max}	8,24	
9	24,9,7		9,24,26	
22	24,18,14		22,28,32	
10	24,18,10,5		10,16,24,29	
11	24,18,10,6		11,17,25,29	
12	24,18,13,2		12,18,23,34	
14	24,18,16,7		14,20,22,31	
15	24,18,13,7		15,21,26,32	
16	24,18,3	T_{\max}	16,22,37	
17	24,18,5	T_{\max}	17,23,36	
18	24,23,18,17		18,19,24,25	
13	24,20,18,14,3		13,17,19,23,34	
21	24,23,19,17,15		21,22,26,28,30	
23	24,18,16,14,1		23,29,31,33,46	
24	24,18,13,8	T_{\min}	24,30,35,40	
19	24,23,19,17,3	T_{\max}	19,20,24,26,40	
20	24,23,19,16,15,3		20,21,25,28,29,41	

not fixed, but varies from 24 for hour 1 to 47 for hour 24. The maximum total number of model inputs is therefore 51 in the model for hour 24. Input T_1 was removed when training this model only in order to keep the number of model inputs within the maximum limit of 50 for the AIM version used. Table 7 summarizes the model structure for all the 24 hourly models. Compared to the corresponding models for next-day hourly temperatures (Table 2), next-hour models are simpler, reflecting the relative ease of forecasting with previous temperature data as recent as the previous hour being available. For example, the model for hour 12 is a 2-input, double element as compared to a 3-input triple element for the corresponding next-day hourly model. Dependence on previous day ($d-1$) hourly temperatures is reduced, with 15 of the 24 models totally ignoring such inputs in favour of the more recent values on the forecasting day (d). Forecasted extreme tempera-

tures on day (d) are utilized by most of the models, while no use is made at all of measured extreme temperatures on day ($d-1$). The temperature time series features strongly in all models. The time lag of 1 h is used by all but two models, indicating a strong influence for persistence. The left-hand side of Table 8 lists the MAE and MAPE values for all hours, giving the overall values for the evaluation year as 1.05°F and 2.14%, respectively, indicating the effectiveness of the technique for very short term temperature forecasting.

5.2. Without using forecasts for next day extreme temperatures

We have developed a set of 24 next-hour temperature forecasters that do not use forecasted ET_{\min} and ET_{\max} as input variables. Table 9 shows a summary of the resulting models. The models are generally of comparable

Table 7

Summary of the abductive network models for the 24 next-hour temperature forecasters with extreme temperature forecasts

Day (d) forecasting hour	Model Input (s)				Temperature time lags selected	Model structure
	Hourly temperatures		Extreme temperatures			
	Day (d-1)	Day (d)	Day (d-1)	Day (d)		
16		15			1	
23		22			1	
1	24			ET _{min}	1	
2		1		ET _{min}	1	
3		2		ET _{min}	1	
4		3		ET _{min}	1	
10		9		ET _{max}	1	
11		10		ET _{max}	1	
12		11		ET _{max}	1	
13		12		ET _{max}	1	
17	19	16			1,22	
19		18		ET _{min}	1	
21		20		ET _{min}	1	
22		21		ET _{min}	1	
5		4		ET _{min} ET _{max}	1	
6		5		ET _{min} ET _{max}	1	
7		6		ET _{min} ET _{max}	1	
8	6	7		ET _{max}	1,26	
9	19	8		ET _{max}	1,14	
14	8	12		ET _{max}	2,30	
18	7			ET _{min} ET _{max}	35	
20	11	19		ET _{min}	1,33	
24	7	23		ET _{min}	1,41	
15	19,17, 15,14	14,7,3		ET _{min} ET _{max}	1,8,12, 20,22,24,25	

complexity with those of Table 7, but with increased dependence on the temperature time series. As shown in the middle part of Table 8, forecasting performance of these models over the evaluation year (MAE = 1.06 °F and MAPE = 2.18%) is comparable with the models utilizing ET_{min} and ET_{max}. However, the former models would be more useful and more accurate in practice, as they do not require, and are not influenced by errors in, extreme temperature forecasts. Lanza and Cosme (2001) report an MAE of 0.80 °F for a next-hour radial basis functions neural network forecaster that does not use extreme temperature forecasts. The neural network required training on daily basis using a moving window of 28 days and was evaluated over a period of 90 days. Table 10 summarizes the MAPE error histograms for all forecasting hours over the evaluation year for both next-day and next-hour models, as well as persistence and

climatology forecasts. 95% of next-hour forecasts are accurate to ± 3 °F without requiring extreme temperature forecasts.

5.3. Next-day forecasts using next-hour models

The right-hand side of Table 8 lists results for next-day, day (d), forecasting obtained by sequential use of the next-hour models described in Section 5.2 above for all hours up to the forecasting hour (h). In practice, this would be performed at the end of day (d - 1), with the temperature forecasted for hour (i) being fed, along with other required inputs, to the next-hour model for hour (i + 1). As expected, performance this way is inferior to that of next-hour forecasting by the same models, due to the accumulation of forecasting errors. However, overall next-day performance of these models is comparable with that obtained by dedicated next-day forecasters

Table 8

Performance of the next-hour temperature forecasting models over the evaluation year

Forecasting hour (<i>h</i>)	With extreme temperature forecasts (Next-hour forecasts)		Without extreme temperature forecasts			
			Next-hour forecasts		Next-day forecasts (Sequential Application)	
	MAE (°F)	MAPE (%)	MAE (°F)	MAPE (%)	MAE (°F)	MAPE (%)
1	0.94	2.01	1.00	2.16	1.00	2.16
2	0.96	2.11	1.06	2.32	1.35	2.97
3	0.86	1.90	0.96	2.13	1.61	3.55
4	0.90	1.96	1.01	2.20	1.84	4.10
5	0.85	1.86	0.97	2.09	2.03	4.55
6	0.83	1.85	0.87	1.96	2.32	5.28
7	0.94	2.08	0.94	2.01	2.28	5.14
8	0.96	2.05	1.14	2.40	2.32	5.10
9	0.99	2.10	1.13	2.44	2.54	5.49
10	1.01	2.12	1.12	2.29	2.78	5.83
11	1.03	1.99	1.12	2.17	3.12	6.27
12	1.04	2.04	1.10	2.12	3.39	6.64
13	1.06	2.00	1.27	2.41	3.45	6.67
14	1.25	2.38	1.13	2.07	3.84	7.23
15	0.99	1.85	1.13	2.03	4.07	7.65
16	1.07	2.00	1.07	2.00	4.25	7.97
17	1.13	2.12	1.13	2.12	4.60	8.77
18	1.87	3.72	1.03	2.21	4.45	8.60
19	1.10	2.16	1.02	2.18	4.26	8.52
20	1.09	2.14	1.02	2.17	4.07	8.37
21	1.03	2.03	1.05	2.07	3.73	7.99
22	1.20	2.44	1.02	2.17	3.66	8.06
23	1.06	2.19	1.06	2.19	3.82	8.63
24	1.10	2.35	1.12	2.40	3.78	8.64
Average	1.05	2.14	1.06	2.18	3.11	6.42

(Table 3, right-hand side) (MAE values 3.11 °F versus 3.02 °F, respectively). The next-hour models are simpler and easier to apply and maintain in practice as compared to the corresponding next-day hourly models. Moreover, they have the advantage of forming a compact set of forecasters that serves the dual purpose of providing accurate next-hour forecasts as well as reasonably accurate next-day hourly forecasts through sequential application at the end of the day preceding the forecasting day.

6. Conclusions

Abductive network machine learning has been demonstrated as an alternative tool for next-day and next-hour hourly temperature forecasting. Models both with and without the requirement for extreme temperature forecasts have been developed. Compared to the neural networks approach, the proposed method simplifies model development, automatically selects effective inputs, gives better insight into the modeled function, and allows comparison with previously used

analytical models. Represented as a few simple equations, the models lend themselves to use in remote locations, e.g. for agricultural and environmental applications. Forecasting is performed through straight-forward substitution in such equations using simple apparatus, without the need for specialized software packages or frequent training. Performance compares favourably with neural network models developed using the same data, and with more complex neural models reported in the literature. Some of the latter neural models may require daily training or weight updates, which complicates onsite forecasting. Next-hour models do not require extreme temperature forecasts for improved accuracy, and may be used sequentially to serve the additional purpose of providing reasonably accurate next-day hourly forecasts at remote locations. In this way, such models could economically serve the dual purpose of providing both next-hour and next-day forecasts. Models described should prove useful in a variety of agricultural, environmental, and electrical load forecasting applications. Future work will attempt to further improve the forecasting accuracy by using dedicated seasonal models

Table 9

Summary of the abductive network models for the 24 next-hour temperature forecasters without extreme temperature forecasts

Day (<i>d</i>) forecasting hour	Model input (<i>s</i>)			Temperature time lags selected	Model structure
	Hourly temperatures		Extreme temperatures		
	Day (<i>d</i> -1)	Day (<i>d</i>)	Day (<i>d</i> -1):		
1	24			1	
4		3		1	
5		4		1	
6		5		1	
15		14		1	
16		15		1	
19		18		1	
23		22		1	
2		1	T_{\min}	1	
3	20	2		1,7	
8	24	7		1,8	
17	19	16		1,22	
18		17,5		1,13	
20		19,6		1,14	
21		20,5		1,16	
7	20	6,2		1,5,11	
9	13	8,6		1,3,20	
11	18	10,5		1,6,17	
12	9	11,7		1,5,27	
13	19	12,6		1,7,18	
14	17	13,6		1,8,21	
22		21,13,5		1,9,17	
24		23,15,6		1,9,18	
10	3	9,7	T_{\max}	1,3,31	

Table 10

Summary of the MAE error histograms showing percentage populations of all forecasting hours over the evaluation year in three error categories for next-day and next-hour models as well as persistence and climatology forecasts

MAE error category (°F)	Next-day models (%)		Persistence (%)	Climatology (%)	Next-hour models (%)	
	With extreme temperature forecasts	Without extreme temperature forecasts			With extreme temperature forecasts	Without extreme temperature forecasts
≤1	41	27	25	14	60	59
≤3	85	62	52	38	95	95
≥6	2.6	13	27	33	0.4	0.4

and including temperature data on a larger number of previous days.

Acknowledgements

The author wishes to acknowledge the support of the Research Institute of King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

References

- Abdel-Aal, R.E., Elhadidy, M.A., 1994. A machine-learning approach to modelling and forecasting the minimum temperature at Dhahran, Saudi Arabia. *Energy—The International Journal* 19, 739–749.
- Abdel-Aal, R.E., Elhadidy, M.A., 1995. Modeling and forecasting the maximum temperature using abductive machine learning. *Weather and Forecasting* 10, 310–325.
- Abdel-Aal, R.E., Al-Garni, A.Z., Al-Nassar, Y.N., 1997. Modelling and forecasting monthly electric energy consumption in eastern

- Saudi Arabian using abductive networks. *Energy—The International Journal* 22, 911–921.
- AbTech Corporation, Charlottesville, VA, USA, 1990. AIM User's Manual.
- Alves da Silva, A.P., Rodrigues, U.P., Rocha Reis, A.J., Moulin, L.S., 2001. NeuroDem—a neural network based short term demand forecaster. Presented at the IEEE Power Technical Conference, Portugal.
- Barron, A.R., 1984. Predicted squared error: a criterion for automatic model selection. Self-organizing. In: Farlow, S.J. (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*. Marcel-Dekker, New York, pp. 87–103.
- Bogren, J., Gustavsson, T., 1994. A combined statistical and energy balance model for prediction of road surface temperature. *Proceedings of the Seventh International Road Weather Conference*, Seefeld, Austria. (<http://www2.ceri.go.jp/sirwec2002/english/papers/bogren.pdf>).
- Charytoniuk, W., Chen, M.S., 2000. Neural network design for short-term load forecasting. *Proceedings of the International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pp. 554–561.
- El-Sharkawi, A.M., 2002. EE 559: Fundamentals of Intelligent Systems. University of Washington, Seattle, USA. (<http://www.ee.washington.edu/class/559/2002spr/>).
- Fan, J.Y., McDonald, J.D., 1994. A real-time implementation of short-term load forecasting for distribution power systems. *IEEE Transactions on Power Systems* 9, 988–994.
- Farlow, S.J., 1984. The GMDH algorithm. In: Farlow, S.J. (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*. Marcel-Dekker, New York, pp. 1–24.
- Francis, R., 2000. Early Warning System, Commercial Vegetable Notes, Vol. 2, No. 2. Cooperative Extension Service, Clemson University, South Carolina, March 2000. (<http://www.clemson.edu/coastalrec/Vegetable%20NewsMar%202000.pdf>).
- Fulcher, G.E., Brown, D.E., 1994. A polynomial network for predicting temperature distributions. *IEEE Transactions on Neural Networks* 5, 372–379.
- Hippert, H.S., Pedreira, C.E., Souza, R.C., 2000. Combining neural networks and ARIMA models for hourly temperature forecasts. *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 414–419.
- Hippert, H.S., Pedreira, C.E., Souza, R.C., 2001. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems* 16, 44–55.
- Hwang, R.-C., Huang, H.-C., Chen, Y.-J., Hsieh, J.-G., 1998. Power load forecasting by neural network with a new learning process for considering overtraining problem. *Proceedings of the International Conference on Energy Management and Power Delivery*, pp. 317–322.
- Khotanzad, A., Davis, M.H., Abaye, A., Maratukulam, D.J., 1996. An artificial neural network hourly temperature forecaster with applications in load forecasting. *IEEE Transactions on Power Systems* 11, 870–876.
- Khotanzad, A., Afkhami-Rohani, R., Maratukulam, D., 1998. ANNSTLF-Artificial Neural Network Short-Term Load Forecaster generation three. *IEEE Transactions on Power Systems* 13, 1413–1422.
- Kim, K.S., Taylor, S.E., Gleason, M.L., Koehler, K.J., 2002. Model to enhance site-specific estimation of leaf wetness duration. *Plant Disease* 86, 179–185.
- Lanza, P.N., Cosme, J.M., 2001. A short-term temperature forecaster based on a novel radial basis functions neural network. *International Journal of Neural Networks* 11, 71–77.
- Lewis, III, H.W., 2001. Intelligent hybrid load forecasting system for an electric power company. *Proceedings of the IEEE Mountain Workshop on Soft Computing in Industrial Applications*, pp. 23–27.
- Matsui, T., Iizaka, T., Fukuyama, Y., 2001. Peak load forecasting using analyzable structured neural network. *Proceedings of the IEEE Power Engineering Society Winter Meeting*, pp. 405–410.
- Montgomery, G.J., Drake, K.C., 1990. Abductive networks. *Proceedings of the SPIE Applications of Artificial Neural Networks Conference*, Orlando, FL, pp. 56–64.
- Pasini, A., Pelino, V., Potestà, S., 2001. A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *Journal of Geophysical Research* 106 (D14), 14951–14959.
- Pattern Recognition Technologies Inc., ANNGSFTM: Adaptive neural network gas sender forecaster. (<http://www.prt-inc.com/ANNGSF/ANNGSF%20Product%20Info.pdf>).
- Roadknight, C.M., Balls, G.R., Mills, G.E., Palmer-Brown, D., 1997. Modeling complex environmental data. *IEEE Transactions on Neural Networks* 8, 852–862.
- Schneider, A.M., Takenawa, T., Schiffman, D.A., 1985. 24-hour electric utility load forecasting. In: Bunn, D.W., Farmer, E.D. (Eds.), *Comparative Models for Electrical Load Forecasting*. Wiley, New York, pp. 87–108.
- Seppälä, J., Lehtoranta, O., Koivisto, H., Koivo, H., 2000. Adaptive district heat load forecasting using neural networks. Presented at the Third International Symposium on Soft Computing for Industry, Maui, Hawaii.
- Sharif, S.S., Taylor, J.H., 2000. Real-time load forecasting by artificial neural networks. *IEEE Power Engineering Society Summer Meeting*, pp. 496–501.
- Sonntag, G.P., Portele, T., Heuft, B., 1997. Prosody generation with a neural network: weighing the importance of input parameters. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 931–934.
- Tassadduq, I., Rehman, S., Bubshait, K., 2002. Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. *Renewable Energy* 25, 545–554.
- Tenorio, M.F., Lee, W.T., 1989. Self organizing neural networks for the identification problem. In: Touretzky, D.S. (Ed.), *Advances in Neural Information Processing Systems*. Morgan Kaufman, San Mateo, CA, pp. 57–64.
- Xu, L., Chen, W.J., 1999. Artificial neural network short-term electrical load forecasting techniques. *Proceedings of the IEEE Region 10 Conference 1999*, 1458–1461.