ELSEVIER

European Journal of Operational Research 122 (2000) 31–40

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/orms

Theory and Methodology

Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations

Victor R. Prybutok [a,*], Junsub Yi [b], David Mitchell [a]

[a] Business Computer Information Systems Department, Center for Quality and Productivity, University of North Texas, P.O. Box 13677, Denton, TX 76203-3677, USA

[b] Department of Management Information Systems, Kyungsung University, Pusan, South Korea

Received 1 November 1996; accepted 1 January 1999

Abstract

In an effort to forecast daily maximum ozone concentrations, many researchers have developed daily ozone forecasting models. However, this continuing worldwide environmental problem suggests the need for more accurate models. Development of these models is difficult because the meteorological variables and photochemical reactions involved in ozone formation are complex. In this study, a neural network model for forecasting daily maximum ozone levels is developed and compared with two conventional statistical models, regression and Box–Jenkins ARIMA. The results show that the neural network model is superior to the regression and Box–Jenkins ARIMA models we tested. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Environment; Forecasting; Neural networks; Regression

1. Introduction

Cities in all parts of our world are plagued with air pollution problems. Of these, ambient ozone is near the top of the list where danger to human health is concerned. For example, significant regional ozone problems have been noted throughout continental Europe for more than a decade (Fernandez-Bayon et al., 1993). Ozone problems have also been discovered in locations such as Brisbane, Australia (Simpson and Layton, 1983) and the Fraser Valley region of British Columbia, Canada (Robeson and Steyn, 1990). There are numerous studies supporting the adverse health impact of ambient ozone (Larsen et al., 1991; Lippmann et al., 1983; Spektor et al., 1991). In addition to its potential human health hazard, ozone adversely impacts the yields of agricultural crops and causes noticeable foliage damage. This agricultural damage is estimated to exceed several

* Corresponding author. Tel.: +1 940 565 4767.

E-mail addresses: prybutok@cobaf.unt.edu (V. R. Prybutok), junsub@star.kyungsung.ac.kr (J. Yi), mitchell@cobaf.unt.edu (D. Mitchell).

PII: S0377-2217(99)00069-7

billion dollars annually (Larsen and Heck, 1984; Enders, 1992; Fernandez-Bayon et al., 1993; Krupa et al., 1993; Slaughter et al., 1993).

To better manage ambient ozone within the United States of America, the U.S. Environmental Protection Agency (EPA) established a maximum ozone threshold of 0.120 parts per million (ppm) because this level has been deemed a threat to human health (EPA, 1992). If an area exceeds this threshold, it is considered a non-attainment area. Non-attainment areas are subject to increased costs and requirements for auto inspections, mandated commuter plans, and a loss of federal highway funds. The EPA can also establish sanctions and restrictions that hamper industrial and commercial expansion in non-attainment areas. Therefore, the development of effective prediction models of ozone concentrations in urban areas is important. Management of control and public warning strategies for ozone levels (particularly near densely populated areas) requires accurate forecasts of the concentration of ambient ozone (Robeson and Steyn, 1990). Although many ozone prediction models have been developed and some are in use, there is still a significant need for more accurate models.

The Houston area has been designated a non-attainment area by the EPA and if it fails to reduce ozone levels further sanctions are intended. In 1993, in an effort to help meet the EPA ozone level requirements, the Houston area started a campaign called 'Ozone Alert Days'. These days were intended to make the public aware of meteorological conditions that might lead to excessive ozone levels. Therefore, the development of an accurate model for forecasting daily maximum ozone levels in the Houston metropolitan area is of practical significance.

In a previous work, we developed a neural network (NN) model for forecasting daily maximum ozone levels for the Dallas–Fort Worth (DFW) metropolitan area (Yi and Prybutok, 1996). However, the DFW metropolitan area contained no days above the EPA non-attainment level (0.120). The lack of non-attainment days has the potential to influence model development and suggests that a site (Houston) with numerous days above the threshold be selected for model development. In this work, we develop and compare a NN model for forecasting maximum daily ozone levels in a non-attainment area to regression and Autoregressive Integrated Moving Average (ARIMA) models for the Houston metropolitan area.

2. Background

Ozone in urban areas varies with meteorological and vehicle emission parameters. The meteorological parameters with the highest correlation to ozone concentrations include maximum temperature, wind speed, wind direction, sky cover, humidity and mixing height (Revlett, 1978). Wolff and Lioy (1978) employed a stepwise multiple regression to develop a best fit ozone equation. Their model related daily maximum ozone concentrations to meteorological and vehicle emission parameters along a 24 hour upwind air parcel trajectory. They identified four significant variables. These variables were the upwind ozone maximums on the previous day, today's maximum temperature, the previous day's upwind maximum temperature and the mean wind speed from the surface to 1000 meters. The authors indicated that the regression model could be a potentially useful tool for predicting maximum ozone concentrations because the model was relatively simple.

Several investigators used time series techniques for forecasting daily maximum ozone levels (Chock et al., 1975; McCollister and Wilson, 1975; Robeson and Steyn, 1990). Simpson and Layton (1983) used two types of Box–Jenkins models. These models were a univariate ARIMA model and a bivariate Dynamic Regression–Noise model. The univariate model used maximum ozone data from one site to predict maximum ozone levels at another site. The bivariate model, on the other hand, used past maximum ozone levels at a site to forecast values at the same site. The prediction results for the univariate model were unsatisfactory. This was due to differences in the meteorological characteristics of the two sites. Although the bivariate model yielded statistically significant results, the authors felt that the success of the model rested more on site characteristics than on the utility of the model. The ozone in the region in which the model was applied has a well-mixed layer which leads to more consistent day-to-day ozone levels. This characteristic is not common in most high ozone urban areas and, therefore, using primarily historic maximum ozone levels is not likely to be a successful prediction strategy.

Robeson and Steyn (1990) developed three statistical forecast models of daily maximum ozone levels. The three models were a univariate deterministic/stochastic model, a univariate ARIMA model, and a bivariate temperature and persistence-based regression model. They compared the three models using the data observed in Canada. The bivariate model was superior to the two univariate models. The ARIMA model had nearly the same predictive capability as the bivariate model while the univariate deterministic/stochastic model performed worst. The study is valuable in that the authors compared the three techniques for forecasting daily maximum ozone levels. However, none of the models were sufficiently dynamic to capture the rapid variability in the ozone time series. All of the models appeared to have some difficulty forecasting high ozone values.

In spite of the existence of the models for forecasting daily maximum ozone levels, no prediction model has received general acceptance. Revlett (1978) noted that a unified model for forecasting ozone concentrations at any site would require extensive research because of the complexity of the meteorological variables and photochemical reactions involved in ozone formation.

The NN model (Fig. 1) consists of an input layer with ten input PEs, one hidden layer consisting of four PEs, and an output layer with a single output PE. The inputs to the NN include one bias value and the nine parameters used in the regression model. During training, a set of n pairs of input vectors and corresponding outputs, $((X(1), y(1)),\ (X(2), y(2)), \ldots, (X(n), y(n)))$ is presented to the network, one pair at a time (Fig. 1). A weighted sum of the inputs,

$$\mathrm{NET} = w_{i,0} + \sum (w_{i,r} x_i), \quad i = 1, \ldots, 9,$$

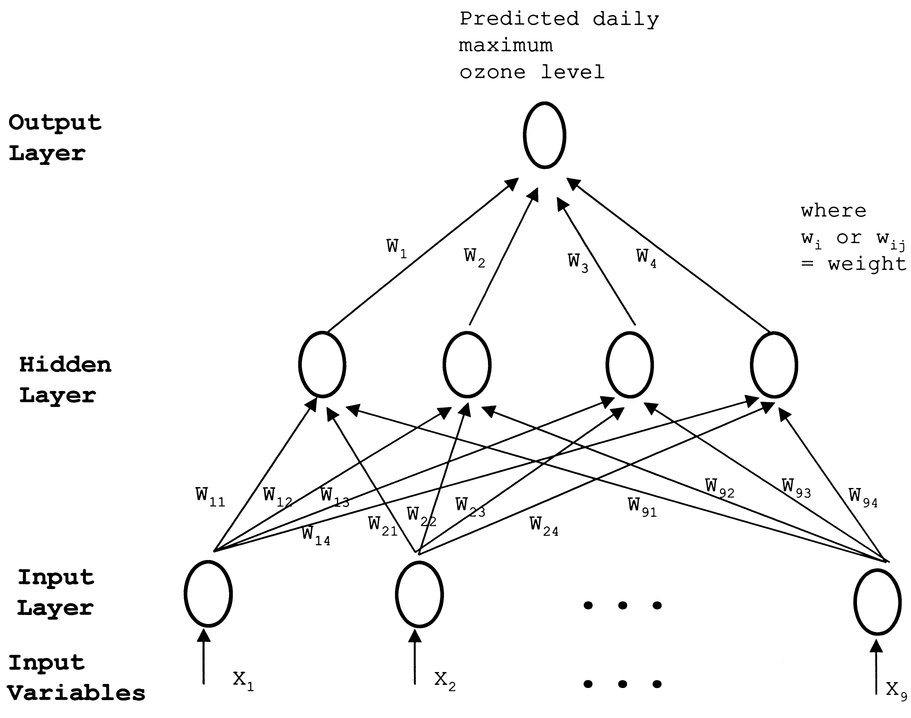

Fig. 1. Neural network model.

is calculated at each hidden layer PE. Each hidden layer PE then uses a sigmoid transfer function to generate an output,

$$z_i = [1 + \exp(-\mathrm{NET})]^{-1}, \quad i = 1, \ldots, 4,$$

between 0 and 1. The outputs from each of the four hidden-layer PEs, along with the bias input, are then sent to the output PE and again a weighted sum is calculated,

$$\mathrm{NET} = v_0 + \sum (v_i x_i), \quad i = 1, \ldots, 4.$$

The weighted sum becomes the input to the sigmoid transfer function of the output PE. The resulting output,

$$F(\mathrm{NET}) = [1 + \exp(-\mathrm{NET})]^{-1},$$

is then scaled to provide the predicted ozone level. At this point, the second phase of the BPLMS algorithm, adjustment of the connection weights, begins.

Performance is measured by looking at the degree to which the NN output matches the actual ozone level recorded for the corresponding input values.

3. Research methodology

3.1. Data collection

The data for our investigations were obtained from the Texas Natural Resource Conservation Commission. These data contain the average hourly ozone measurement and the average hourly meteorological measurements for variables such as temperature, wind speed and wind direction. These data also contain vehicle emission measurements such as nitric oxide, nitrogen dioxide, carbon dioxide and oxides of nitrogen from a number of EPA monitoring sites across Houston area.

The data for 1994 were selected from a site operating in Houston. The period we selected included only the high ozone months, 1 June–31 October, when harmful maximum ozone concentrations occur. This period was chosen to minimize the seasonal factors that affect the parameter estimates in the prediction models. The data for 1 June–30 September in 1994 from a site were used to develop the prediction models and the data October 1–10 in 1994 from the same site were used to test the forecast potential of the models. SAS was used for developing regression and ARIMA models and the NeuralWare III was used for building a NN model.

3.2. Variable specification

When selecting independent variables to forecast daily maximum ozone levels, factors that influence both photochemical production and atmospheric accumulation should be considered (Robeson and Steyn, 1990). Possibly the most important factor in predicting the daily maximum ozone levels is the expected vehicle emissions. Based on previous studies (Chock et al., 1975; Revlett, 1978; Wolff and Lioy, 1978; Prior et al., 1981; Robeson and Steyn, 1990), the input emission variables that required consideration were nitrogen dioxide, nitric oxide, carbon dioxide and oxide of nitrogen. We utilized the average concentration for these emissions between 6:00 a.m. and 9:00 a.m. in the morning of the day of interest.

It was previously found that use of the 6–9 a.m. averages of input emission and meteorological variables can be utilized to predict the maximum ozone concentrations that will occur on that day, based on analysis of measurements from Philadelphia, Washington, D.C., and Denver (Prior et al., 1981). We followed this convention in our study. As suggested by this convention, no weighting scheme was used in calculation of the 6–9 a.m. averages for the input emission variables. The data records used in this study except the dummy variable (holidays versus working days) consist of hourly averages (taken from 12 readings) of the ozone and other variables.

Previous studies showed that accurate daily maximum temperature forecasts are vital to make precise predictions of ozone levels (Revlett, 1978; Wolff and Lioy, 1978). A meteorologically stagnant condition is another vital variable for prediction of ozone levels. In this study, we incorporated average surface wind speed and wind direction between 6:00 a.m. and 9:00 a.m. in the morning of the day of interest as surrogate variables for the stagnant air conditions. The surface wind speed and its direction define the horizontal transport of the precursors of ozone, mainly vehicle emissions (Revlett, 1978; Wolff and Lioy, 1978; Prior et al., 1981). The wind speed and direction then account for the inversion pattern of the morning pollution cloud (Benarie, 1980).

3.3. Regression model building

We used a stepwise regression procedure on the data for the time period 1 June–30 September to identify significant variables for predicting the maximum ozone levels. The forecasting ability of the regression model built in this section is evaluated and compared with that of ARIMA and NN models using the holdout data (October 1–10 in 1994). Various plots showed that the data do not severely violate the major assumptions of regression analysis. The preliminary regression model for concentration of predicting daily maximum ozone levels is provided below.

$$\begin{aligned} y =& \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \\ &+ \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \epsilon, \end{aligned}$$

where y is the daily maximum ozone level (hourly average), X_1 is the dummy variable (holidays versus working days), X_2 is the ozone level at 9:00 a.m., X_3 is the actual maximum daily temperature, X_4 is the carbon dioxide, X_5 is the nitric oxide, X_6 is the nitrogen dioxide, X_7 is the oxide of nitrogen, X_8 is the surface wind speed and X_9 is the surface wind direction. All measured predictors used in this study are actual same day values.

An examination of wind speed (X_8) versus ozone (y) indicated a negative relationship for this site. Also, there appears to be slight non-linearity in the low and high ranges of wind speed. Revlett (1978) supports the use of a non-linear relationship between wind speed and ozone concentration. We transformed wind speed using logarithms to obtain a linear relationship with daily maximum ozone levels. The stepwise procedure showed that the ozone level at 9:00 a.m., maximum daily temperature, nitrogen dioxide and surface wind speed were important to predict daily maximum ozone levels ($R^2 = 0.4713$, $\text{MSE} = 0.0005$). However, the regression analysis indicated that the dummy variable (holidays versus working days), carbon dioxide, nitric oxide and oxide of nitrogen were not significant for this site in the presence of the other variables. The best single variable among the nine independent variables was the logarithm of surface wind speed ($R^2 = 0.2893$, $\text{MSE} = 0.00065612$). The second-best single variable was the maximum daily temperature as shown in Table 1. There are two factors that attribute the strength of the correlation between temperature and ozone concentration. High air temperature is an excellent indication of environmental conditions conducive to ozone formation and accumulation. In addition, the photochemical reaction rates are highly temperature dependent (Robeson and Steyn, 1990).

Each step of our forward stepwise regression procedure is shown in the Table 1. Multicollinearity is not considered to be a problem in this study, since we used the regression model just for forecasting purposes.

The final regression model using the criteria of both accuracy and parsimony is

$$y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_6 + \beta_8 \log(X_8) + \epsilon.$$

3.4. ARIMA model building

An autoregressive integrated moving average (ARIMA) model was used to predict daily maximum ozone levels. After some initial analyses, we have found that an ARIMA model with no differencing worked better than that with differencing. With no differencing, we had an autocor-

Table 1
Stepwise regression results

Steps	Set of variables	R^2	MSE
1	$\log(X_8)$	0.2893	0.00065612
2	$\log(X_8)$, X_3	0.3650	0.00059125
3	$\log(X_8)$, X_3, X_2	0.4129	0.00055135
4	$\log(X_8)$, X_3, X_2, X_6	0.4713	0.00050000

relation plot with a spike at lag 1 and damped sine waves after the lag 1. We had also a partial autocorrelation plot with a spike at lag 1 and cutting off after the lag 1. It seemed that there was no seasonality. Therefore, we used the non-seasonal autoregressive operator of order 1. We included a constant term in this model as shown in the following tentative model (Bowerman and O'Connell, 1987):

$$(1 - \phi_1 B)y_t = \delta + a_t.$$

The Ljung–Box statistics were used to check the adequacy of the tentatively identified model. The p-value for the Ljung–Box statistics was 0.489, 0.612, 0.583 and 0.605 for values of lags equal to 6, 12, 18 and 24, respectively. This fact indicated that the model is adequate. We also examined the individual sample autocorrelations and individual sample partial autocorrelations of the residuals to further investigate the adequacy of the model. The results supported the adequacy of the model.

The resulting ARIMA model without seasonality can be supported by some previous ozone forecasting research. For example, Simpson and Layton (1983) used to predict maximum ozone levels for the data sets from three monitoring sites in Brisbane, Queensland and Australia. They identified the following models as suitable:

$$(1 - \phi_1 B^1)y_{1t} = \delta_1 + a_{1t} \quad \text{for Brisbane,}$$

$$(1 - \phi_1 B^1 - \phi_7 B^7)y_{2t} = \delta_2 + a_{2t} \quad \text{for Queensland,}$$

$$(1 - \phi_1 B^1)y_{3t} = \delta_3 + a_{3t} \quad \text{for Australia.}$$

They concluded in their paper published on *Atmospheric Environment* that using historic maximum ozone levels is not likely to be a successful prediction strategy.

The special characteristics of ozone formation and variability explain why using historic maximum ozone levels is not likely to be successful and why there is no seasonality in the model. Ozone formation and variability result from photochemical reactions involving the interplay of a number of factors such as meteorological, topographical and vehicle emission variables. Ozone formation and variability are directly related to the meteorological variables and chemical reactions in the morning of the day of interest (less related to the ozone levels of previous days). Ozone level goes down to almost zero level at night when it is cooler, without sunlight and with less car emission. It starts going up as it gets warmer and gets a peak level in the afternoon usually. A high level of sunlight and high temperatures are prime contributors to the formation of ozone. When it rains, the ozone level will not go up.

The forecasting ability of the ARIMA model built in this section is evaluated and compared with those of regression and NN models using the holdout data (October 1–10 in 1994). The comparison is shown in the comparative analysis section.

3.5. Neural network model building

As early as 1975, McCollister and Wilson noted that possible techniques for forecasting maximum ozone levels included time series analysis and pattern recognition techniques. A number of studies have used time series techniques for forecasting maximum ozone levels. However, the literature does not contain studies that utilize a pattern recognition technique. It is well known that a NN shows excellent performance on pattern recognition tasks. The literature (Ball and Jur, 1993; Boznar et al., 1993; Hill et al., 1994) suggests potential advantages of NNs over statistical methods. One such advantage is the better performance of the NN when many extreme values exist like in the ozone data in this study. Another advantage of NNs is that the estimation of a NN can be automated, while the regression and ARIMA models must be reestimated periodically whenever new data arrives.

The inputs in the NN experiments consisted of the nine independent variables used in the regression analysis. The same independent variables were used so that a direct comparison could be made with the results obtained from the regression model. The network was trained with the data from one site in Houston for 1 June–30 September in 1994. In building a NN model, the selection of the training set is important because it is from this set that the NN will learn to generalize. The training set should contain as many significant meteorological situations as possible. When a representative training set of patterns is not available, accurate prediction cannot be obtained from the NN model (Boznar et al., 1993). We believe that our Houston training set was representative of the various patterns of daily maximum ozone levels found in Houston. The back-propagation learning algorithm was used in the training process. More than 50 experiments were performed to determine the best combination of the learning rates, momentum, number of hidden layers, number of hidden layer PEs, learning rule and transfer function to utilize. The two most popular learning rules, generalized and cumulative delta rules, were tested.

As discussed above, we used a fully connected feed-forward network with nine input PEs, 4 PEs in a single hidden layer and 1 output PE. Each of the nine input PEs represents one of the nine independent variables used in the analysis. The one output PE produced the prediction estimates for the daily maximum ozone levels. For the fully-connected, feed-forward network, there is a general guideline for deciding how many PEs should be placed in the hidden layer (Neural Ware, 1993). One such guideline is that h = number of training cases$/5 \times (I + o)$, where h, I and o are numbers of PEs in hidden, input and output layers, respectively. For this study, $h = 120/5 \times (9 + 1) = 3$ hidden PEs. We used this resulting number of hidden PEs as our starting point, and trained the data with 2, 3, 4, 5 and 6 hidden PEs.

Throughout the training, the NeuralWare utility, 'SAVEBEST' was used to monitor and save the lowest root mean square (RMS) error value from the cross-validation set. The best RMS error result was obtained using a learning rate of 0.1, a momentum of 0.6 and 4 PEs in a single hidden layer that used the generalized delta learning rule and a sigmoid transfer function. The architecture of the best network contains 9 input layer PEs, 4 hidden PEs and 1 output layer PE (9:4:1 architecture).

4. Comparative analysis of the models

Once the three models to predict daily maximum ozone levels were developed, we empirically examined the relative effectiveness of the models in predicting ozone levels using the data for October 1–10 in 1994. The forecast performance of each model was evaluated using graphical and statistical comparisons as shown in Fig. 2.

Fig. 2 shows the forecasting ability of the regression, ARIMA, and NN models for the testing data. The mean absolute deviations (MAD) of the three models calculated from the estimates are 0.025741, 0.02879 and 0.012945, respectively. The RMS errors are 0.031239, 0.033023 and 0.016418, respectively. In order to compare the relative effectiveness of the three empirical models, we performed the Friedman test, a non-parametric technique for comparing k related samples (Siegel and Castellan, 1988).

Absolute differences of the actual maximum daily ozone levels minus those predicted by each of the three methods were calculated. The Friedman test (as detailed in Conover, 1980) was then used. The result indicated that at least one significant difference did exist. Multiple comparisons, using total ranks (Conover, 1980), indicated that the NN model was better than both the regression model and the ARIMA model. Fig. 2 graphically supports these results.

5. Discussion

The methodology developed in this study for the Houston site is somewhat generalizable. This generalizability is extendable to international as well as domestic sites. In addition, this model building approach may apply to rural areas in spite of differences in topographical features. Furthermore, the seasonal variability of ozone formation is not a likely consideration in developing a model for prediction of potential health risk days because like the Houston site, it is possible that high ozone days occur only in a single season (summer). Alternatively, if necessary, a seasonal component may be added.

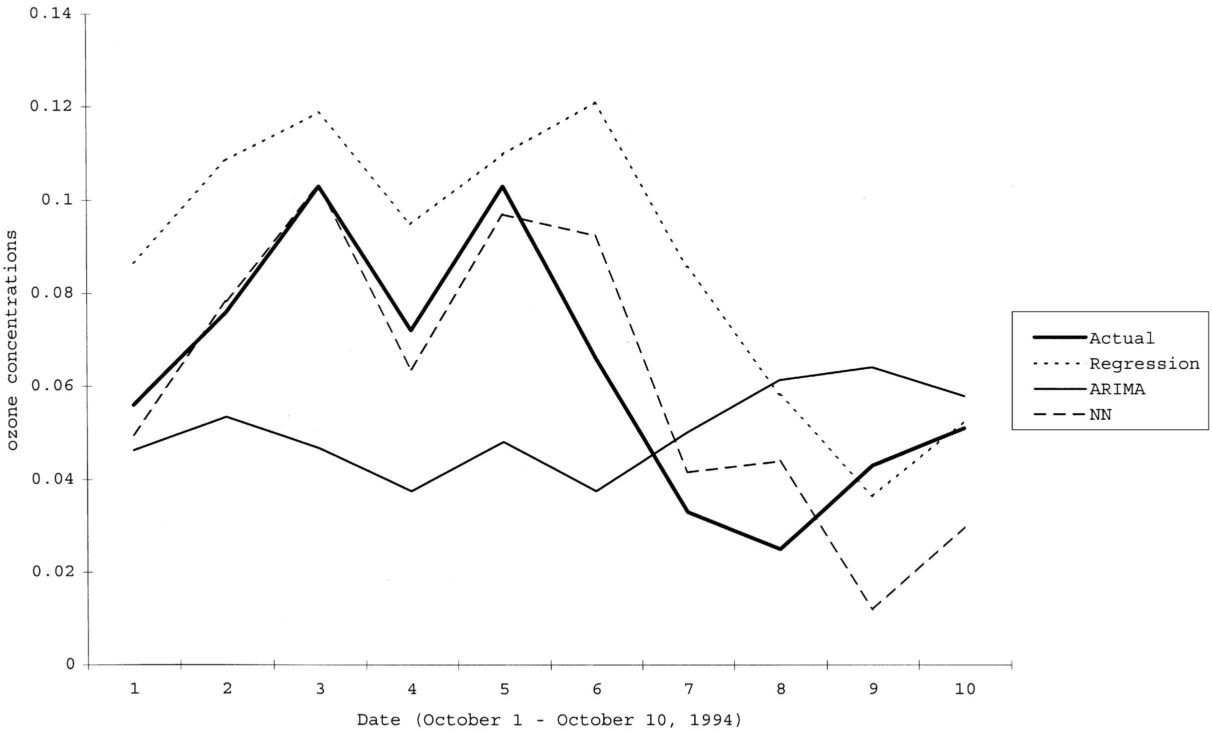

Fig. 2. Comparison of Houston ozone prediction models (1–10 October, 1994).

Ozone concentrations are highest in summer in most places. The trend is usually a steady increase in ozone level from the start of the year to the summer months, followed by a tapering off through fall and winter. The seasonal peaks in ozone concentration occur in summer, without significant differences between urban and rural sites. However, not all sites or all years fit into the same trend. Seasonal distributions at a few sites are more individual. For example, New Orleans, LA, has an early spring peak followed by lower concentrations in the summer with another peak again in the fall. The city is located near the coast of the Gulf of Mexico on the Mississippi River delta, where local meteorology, such as summertime afternoon thunderstorms, may serve to scavenge the ozone out of the area (Ajena et al., 1992). Despite the fact that the model development in this study is generalizable, the developed model is site specific to Houston. Moreover, there is no guaranty as to the reliability of the model once it is extrapolated beyond the range of the input data used to construct it. That is, the model is valid only in the period and may not be appropriate if the meteorological conditions are unusual in this period. However, the period chosen in this study is of interest because it is when harmful maximum ozone concentrations occur in the Houston area.

Another consideration in applying our model is the representativeness of the variable constructs (Kerlinger, 1986; Buckley et al., 1986). When selecting independent variables for forecasting daily maximum ozone levels, factors that influence both photochemical production and atmospheric accumulation of ozone should be considered. Ozone formation and variability result from photochemical reactions involving the interplay of a number of factors such as meteorological, topographical, and vehicle emission variables. This study included only four vehicle emission variables and three meteorological variables. While these variables represent an excellent starting point, models that consider other variable constructs, topographical features that can affect the movement of air, and air inversion, are appropriate areas for future works. However, this work demonstrates that NN models are worthy of further development with more complex variable constructs.

This work also demonstrates that NN models are viable in comparison to traditional statistical methodologies. However, a general limitation of NN models is their potential convergence to a local minimum rather than to a global minimum (Archer and Wang, 1993). The avoidance of such convergence is dependent on the model building of the developer. Therefore, like all quantitative methodologies, NNs require an experienced or informed practitioner (See for example, Boznar et al., 1993).

The prediction models developed in this study employed the independent variables measured in the morning on the forecast day. The use of the forecast day's values does not seriously limit the usefulness of the models for predicting 'Ozone Alert Days' in advance because the models can be modified by using surrogate variables. In general, it is possible to predict the relevant meteorological conditions and vehicle emission measurements based on past data and to use these data in the ozone prediction models. Areas for our future research include the modification of these prediction models using such surrogate variables.

We also intend to investigate the use of NN, multiple discriminant analysis and logistic regression as means for separating high from low maximum ozone days. If this can be done reliably, then authorities will have the information needed to warn the public and thereby help reduce maximum ozone levels. In addition, further comparison between site specific models is an important area for further work because such a comparison may aid in developing site characterizations. These characterizations may prove important for extensions of the developed models to new sites around the world and may aid in creating more guidelines for site specific model development.

Acknowledgements

We are pleased to acknowledge the gracious cooperation of the personnel at the Texas Natural Resource Conservation Commission. We are especially grateful to Larry Butts and Tom Larkins for their assistance in obtaining the data. We also thank the University of North Texas for the faculty research grant that provided the Neural Network software utilized in our investigations.

References

Ajena, V.P., Yoder, G.T., Arya, S.P., 1992. Ozone in the urban southeastern United States. Environmental Pollution 75, 39–44.

Archer, N.P., Wang, S., 1993. Application of the back propagation neural network algorithm with monoticity constraints for two-group classification problems. Decision Sciences 24, 60–75.

Ball, J.W., Jur, P.C., 1993. Simulation of polysaccharide C nuclear magnetic resonance spectra using regression analysis and neural networks. Analytical Chemistry 65 (24), 3615–3621.

Benarie, M.M., 1980. Urban Air Pollution Modelling, MIT Press, Cambridge, MA.

Bowerman, B.L., O'Connell, R.T., 1987. Time Series Forecasting, 2nd ed., Duxbury, Pacific Grove, CA.

Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment 27 (2), 221–230.

Buckley, J.W., Buckley, M.H., Chiang, H.-F., 1986. Research Methodology and Business Decisions, National Association of Accountants and The Society of Industrial Accountants of Canada.

Chock, D.P., Terrell, T.R., Levitt, S.B., 1975. Time-series analysis of Riverside, California air quality data. Atmospheric Environment 9, 978–989.

Conover, W.J., 1980. Practical Nonparametric Statistics, 2nd edn. Wiley, New York.

Enders, G., 1992. Deposition of ozone to a mature spruce forest: Measurements and comparison to models. Environmental Pollution 75, 61–67.

Environmental Protection Agency, 1992. National Air Quality and Emission Trends Report, October, pp. 1–11.

Fernandez-Bayon, J.M., Barnes, J.D., Ollerenshaw, J.H., Davison, A.W., 1993. Physiological effects of ozone on cultivars of watermelon and muskmelon widely grown in Spain. Environmental Pollution 81, 199–206.

Hill, T., Marquez, L., O'Connor, M., Remus, W., 1994. Artificial neural network models for forecasting and decision making, International Journal of Forecasting (10) 5–15.

Kerlinger, F.N., 1986. Foundations of Behavioral Research, 3rd ed., Harcourt, Brace, Jovanovich, New York.

Krupa, S.V., Manning, W.J., Nosal, M., 1993. Use of tobacco cultivars as biological indicators of ambient ozone pollution: An analysis of exposure–response relationships. Environmental Pollution 81, 137–146.

Larsen, R.I., Heck, W.W., 1984. An air quality analysis system for interrelating effects, standards, and needed source reductions: Part 8. An effect mean ozone crop reduction mathematical model. Journal of Air Pollution Control Association 34, 1023–1034.

Larsen, R.I., McDonnell, W.F., Horstman, D.H., 1991. An air quality data analysis system for interrelating effects, standards, and needed source reductions: Part 11. A lognormal model relating human lung function decrease to ozone exposure. Journal of Waste Management Association 41 (4), 55–459.

Lippmann, M., Lioy, P.J., Leikuaf, G., Green, K.B., Baxter, D., Morandi, M., Pasternack, B., Fife, D., Speizer, F.E., 1983. Effects of ozone on the pulmonary function of children. Advances Modern Environment Toxicology 5, 423–446.

McCollister, G.M., Wilson, K.R., 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. Atmospheric Environment 9, 417–423.

Neural Ware, 1993. Neural Computing: Neural works Professional II/ PLUS and Neuralworks Explorer, NeuralWare.

Prior, E.J., Schiess, J.R., McDougal, D.S., 1981. Approach to forecasting daily maximum ozone levels in St. Louis. Environmental Science & Technology 15 (4), 430–436.

Revlett, G.H., 1978. Ozone forecasting using empirical modeling. Journal of Air Pollution Control Association 28, 338–343.

Robeson, S.M., Steyn, D.G., 1990. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. Atmospheric Environment 24B, 303–312.

Siegel, S., Castellan, N., John Jr. 1988. Nonparametric Statistics, 2nd ed. McGraw-Hill, New York.

Simpson, R.W., Layton, A.P., 1983. Forecasting peak ozone levels. Atmospheric Environment 17, 1649–1654.

Slaughter, L.H., Mulchi, C.L., Lee, E.H., 1993. Wheat-kernel growth characteristics during exposure to chronic ozone pollution. Environmental Pollution 80, 73–79.

Spektor, D.M., Thurston, G.D., Mao, J., He, D., Hayes, C., Lippman, M., 1991. Effects of single and multiday ozone exposures on respiratory function in active normal children. Environmental Research 55, 107–122.

Wolff, G.T., Lioy, P.J., 1978. An empirical model for forecasting maximum daily ozone levels in the northeastern U.S.. Journal of Air Pollution Control Association 28, 1034–1038.

Yi, J., Prybutok, V.R., A neural network forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, Environmental Pollution, forthcoming.