Journal of the Air & Waste Management Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uawm20

Forecasts Using Neural Network versus Box-Jenkins Methodology for Ambient Air Quality Monitoring Data

Jehng-Jung Kao [a] & Shang-Shuang Huang [b]
[a] Institute of Environmental Engineering, National Chiao Tung University , Hsinchu , Taiwan
[b] Central Insurance Co., Ltd. , Taipei , Taiwan
Published online: 27 Dec 2011.

TECHNICAL PAPER

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **50**:219-226

Forecasts Using Neural Network versus Box-Jenkins Methodology for Ambient Air Quality Monitoring Data

Jehng-Jung Kao
Institute of Environmental Engineering, National Chiao Tung University, Hsinchu, Taiwan

Shang-Shuang Huang
Central Insurance Co., Ltd., Taipei, Taiwan

ABSTRACT

This study explores ambient air quality forecasts using the conventional time-series approach and a neural network. Sulfur dioxide and ozone monitoring data collected from two background stations and an industrial station are used. Various learning methods and varied numbers of hidden layer processing units of the neural network model are tested. Results obtained from the time-series and neural network models are discussed and compared on the basis of their performance for 1-step-ahead and 24-step-ahead forecasts. Although both models perform well for 1-step-ahead prediction, some neural network results reveal a slightly better forecast without manually adjusting model parameters, according to the results. For a 24-step-ahead forecast, most neural network results are as good as or superior to those of the time-series model. With the advantages of self-learning, self-adaptation, and parallel processing, the neural network approach is a promising technique for developing an automated short-term ambient air quality forecast system.

INTRODUCTION

Air pollutants exert a wide range of impacts on biological, physical, and economic systems. Their effects on human health are of particular concern. The decrease in respiratory efficiency and impaired capability to transport oxygen through the blood caused by a high concentration of air pollutants may be hazardous to those who have pre-existing respiratory and coronary artery disease.[1] Consequently, it has become a vital task to accurately keep track of the variation of ambient air pollutant levels.

> IMPLICATIONS
>
> Air pollutants adversely impact biological, physical, and human respiratory systems. Monitoring variations of ambient air quality is therefore essential. Because of the temporal, random nature of ambient air quality, a time-series approach is generally applied in order to develop a forecast model. With the merits of self-learning, self-adaptation, and parallel processing, the neural network approach adopted herein is highly promising for developing a forecast system. The work has demonstrated the compatibility of the neural network approach with the time-series approach for developing a short-term ambient air quality forecast system.

Natural phenomena are mostly a time series with some degree of randomness. Pollutants in the atmosphere may disperse or concentrate during varied time periods. Previous studies[2,3,4,5] have indicated that the data of ambient air quality are stochastic time series, thereby making it possible to make a short-term forecast on the basis of historical data. However, when applying the conventional time-series model to the ambient air pollution forecast, the pollutant level variations are generally not simple autoregressive or moving average models.[6] Analysts must employ statistical graphs of the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify an appropriate time-series model. In the model identification stage, the resulting model quality frequently relies on individual experience and knowledge of time-series statistics; in addition, different analysts might render contradictory interpretations, given the same data.

In the literature related to air pollution forecasts, the application of a time-series model is not prevalent. Kapoor and Terry[7] indicated that a time-series model requires considerable knowledge in time-series statistics. Individuals without statistical training would likely create an inappropriate model. Engineers in the field of air pollution generally lack such knowledge for time-series statistics. A program for automatically determining an appropriate model for most circumstances is unavailable, although some software packages such as SAS[8] can provide valuable assistance in identifying a time-series model. This situation creates difficulty in applying the time-series model to ambient air quality forecasting. Furthermore, a time-series model may not be applicable for varied periods of data. A model applicable in one period may require manual adjusting of its model parameters to meet the data characteristics in other time periods. These complexities make applying a time-series model to regular air quality forecast an inefficient task.

Due to its advantage of self-correction, self-learning, and parallel processing, the neural network approach is a promising alternative to substitute the conventional time-series model for developing an automatic ambient air quality forecast technique. This work explores the applicability of a neural network for air quality forecasts and compares it with the conventional time-series method.

Neural networks have found extensive applications in recent years for information processing (e.g., voice recognition, hand-written character recognition, voice synthesis, or image processing[9]). Nevertheless, applying a neural network in air pollution forecast has only received limited attention. Boznar et al.[10] established a multi-factor neural network for a 1-step-ahead forecast. In their study for Saleska, Slovenia, a regular air pollution dispersion model could not accurately forecast sulfur dioxide (SO_2) because of complex topography. In order to control the emission of SO_2 at the Sostanj thermal power plant, they developed a neural network to forecast the variation of SO_2 levels. Multiple meteorological factors such as wind velocity, wind direction, temperature, and SO_2 data collected at six monitoring stations surrounding the plant were utilized to construct the network. The meteorological factors and SO_2 data of one specific station and those of neighboring stations were normalized and input into the established neural network to forecast the SO_2 pollution level of the next period (in 30-min intervals) of the specific station. The results indicated that a neural network could keep track of the peak value of the SO_2 level. Comrie[11] compared the neural network and multiple regression models for daily ozone forecasting. Results showed that the neural network model is compatible with the multiple regression model.

Other successful examples of employing a neural network in other types of time-series forecasts have also been observed. For instance, Lapedes[12] used a backpropagation operation to forecast the variation of the Standard & Poor's 500 index, using indexes of the previous 10 weeks to forecast that of the following week. The accuracy rate could reach 61%, higher than the 53% achieved by the moving average method. Chakraborty et al.[13] studied the flour prices in Buffalo, NY; Minneapolis, MN; and Kansas City, Kansas. Prices from August 1972 to November 1980 were used as a data source to perform both a 1-step-ahead-forecast and a multi-step-ahead-forecast of the univariate and multivariate time series using a neural network and a statistical model, respectively. The neural network generally exhibited a higher accuracy than the statistical method.

The current meteorological and air pollution monitoring data acquisition and transmission system in Taiwan is not widespread. Comprehensive data are currently unavailable. A forecast system on the basis of multiple factors such as the one described by Boznar et al.[10] is not suitable for the present stage of monitoring. This work therefore attempts to develop an air quality forecast model employing a univariate time-series neural network. A univariate forecast model requires less data processing, less memory space for data storage, less staff power for data analysis, and ultimately, less cost, although it poses some limitations for complex data affected by multiple factors.

This paper is organized as follows: The monitoring data used in this study are first described, followed by the process and results of using the conventional time-series analysis in establishing a forecast model, and those of using the neural network. Finally, the performances of these two forecast models are compared for the studied data.

MONITORING DATA

Seventy-two ambient air quality-monitoring stations are operating in Taiwan and are categorized into five types: general, transportation, industrial, national park, and background. Data used in this work were collected primarily from two background and one industrial-monitoring stations. Figure 1 shows their locations. Stations A and B are background monitoring stations in Wangli and Kuanying, respectively. Station C is an industrial station in Toufen, with metal, chemical, and petrochemical factories located nearby. Data employed were those collected

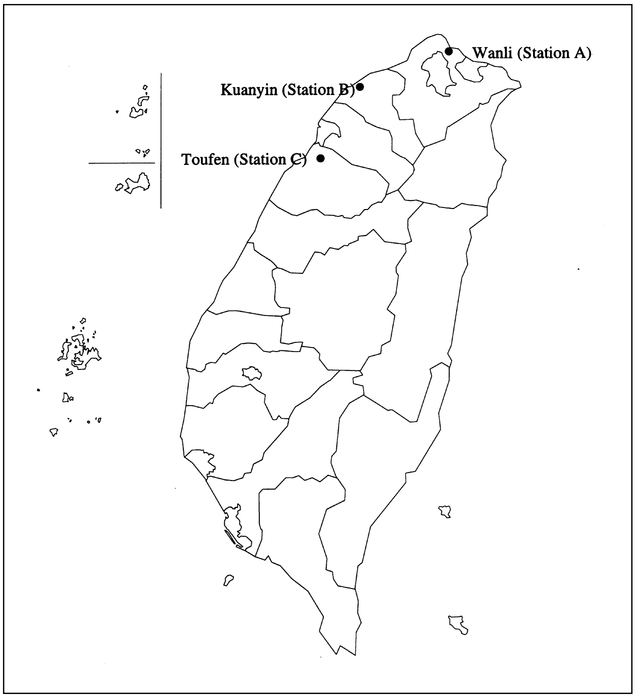

Figure 1. The location of monitoring stations in this study.

at the aforementioned three stations between September 1, 1993 and August 31, 1994. During this period, data for four weeks, as the training stage, were used to construct forecast models, and data for the week following the training stage, as the forecast stage, were used to verify and compare the forecast performances of each model. The data are hourly average values for atmospheric quality levels monitored at these stations.

FORECAST MODEL DEVELOPMENT—TIME SERIES

The auto-regressive integrated moving average model (ARIMA), since being proposed by Box and Jenkins,[13] has been applied in various fields. The model determines simulated values by a combined measure from historical data with associated temporal variation. Its general equation, as described by Bowerman and O'Connell,[15] can be expressed as

$$\phi_p(B)\phi_p(B^L)(1-B^L)^D(1-B)^d y_t = \delta + \theta_q(B)\theta_Q(B^L)a_t \quad (1)$$

where $\phi_p(B)$ is the nonseasonal auto-regressive operator with order p, $\phi_p(B^L)$ is the seasonal auto-regressive operator with order p, B is back shift operator, L is the number of seasons, D is the degree of seasonal differencing, d is the degree of nonseasonal differencing, y_t is the time-series data, transformed if necessary, δ is a constant term, $\theta_q(B)$ is the nonseasonal moving average operator with order q, $\theta_Q(B^L)$ is the seasonal moving average operator with order Q, and a_t is white noise with normal distribution $N(0,\sigma^2)$. This study employed univariate ARIMA models, and SAS/ETS[8] was used to establish the models. The process of establishing the models followed the four steps suggested by Newton:[16] model identification, parameter estimation, selection of an appropriate ARIMA model, and forecast made on the basis of the selected model.

Data stability, variation, and trend, and plots of ACF and PACF values of a studied data series were evaluated in the model identification step. If the data series is not stationary, a decision must be made whether to employ an appropriate data transformation, such as log transformation, or other difference methods. Next, plots of statistical data variations, including ACF and PACF, were used to determine an appropriate time-series model type in accordance with the ARIMA model identification standards. Q statistics, as defined by Ljung and Box[17] and reported from SAS/ETS for an auto-correlation check for white noise, were used to evaluate the suitability of a candidate ARIMA model. If the selected model was inappropriate for the data series, the entire model identification step would have to be restarted. Significant temporal variations of the studied air pollution monitoring data series were observed and made it difficult to distinctly identify an appropriate ARIMA model type from the standards, as determined from ACF and PACF and described by Box and Jenkins.[14] A trial-and-error procedure and empirical judgment were applied to determine the model type on the basis of ACF and PACF.

During the parameter estimation step, Q statistics for an auto-correlation check for residuals reported by SAS/ETS were used to determine a candidate model. Of the models determined by these two steps, more than one appropriate model is frequently available for the provided data series. Goodness-of-fit criteria of Akaie's information criterion and Schwartz's Bayesian criterion[8] were applied to determine the most appropriate model.

The final step entails making a forecast based on the selected model. The forecast methods adopted in this study were 1-step-ahead and 24-step-ahead. The former forecasts the value of the next hour on the basis of the established forecast model. When forecasting the value of the hour following the next, the actual monitoring value replaces the preceding forecast value. The latter forecasts values in the next 24 hr by repeating the 1-step-ahead forecast 24 times without replacing the actual value in each 1-step-ahead forecast.

Following the aforementioned four-step procedure, the forecast models established for data series for the three monitoring stations are listed below.

- The ARIMA model for ozone (O_3) of Station A

$$(1-B)(1-B^{24})y_t = (1-0.19583B^2 - 0.1523B^3 - 0.23088B^4 - 0.20912B^5 - 0.10748B^6)(1-0.71186B^{24})a_t \quad (2)$$

- The ARIMA model for SO_2 of Station B

$$(1-B)(1-B^{24})y_t = (1-0.23488B-0.19882B^2 - 0.21062B^3 - 0.14562B^4)(1-0.9565B^{24})a_t \quad (3)$$

- The ARIMA model for O_3 of Station B

$$(1-B^{24})y_t = (1+0.23391B - 0.158B^2 - 0.27427B^3 - 0.21670B^4 - 0.13497B^5)(1-0.92434B^{24})a_t \quad (4)$$

- The ARIMA model for SO_2 of Station C

$$(1-B^{24})y_t = (1-0.71342B - 0.3454B^2 + 0.14562B^3)(1-0.92762B^{24})a_t \quad (5)$$

- The ARIMA model for O_3 of Station C

$$(1-B)(1-B^{24})y_t = (1+0.34278B)(1-0.99975B^{24})a_t \quad (6)$$

Figure 2 illustrates typical results obtained from the above listed models in the training stage. Results from the ARIMA models closely match the observed values in

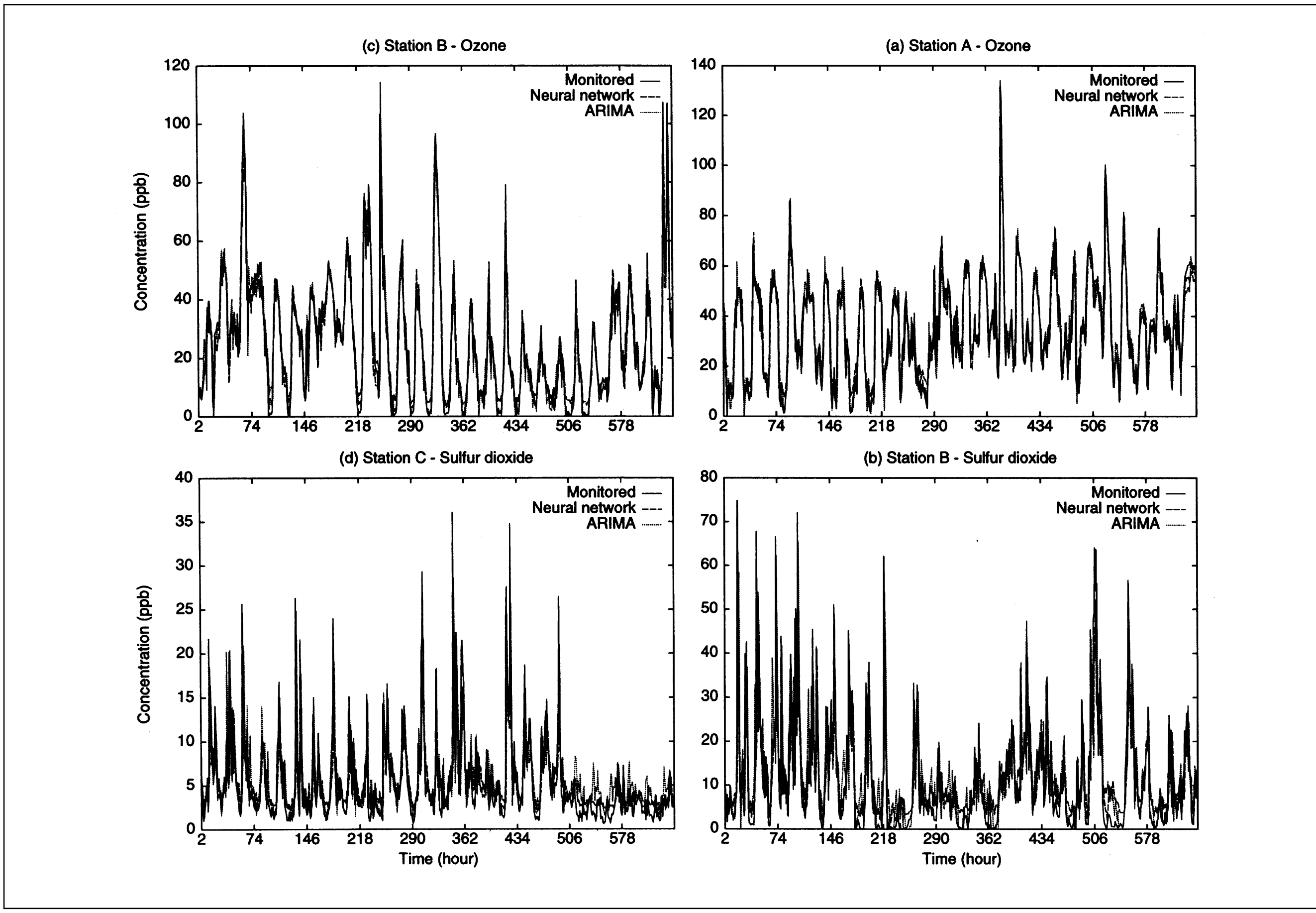

Figure 2. Simulated results of ARIMA and neural network models with observed data.

the training stage, although results for SO_2 were not as good as those for O_3. All of the ARIMA models were then applied to the data in the forecast stage for both 1-step-ahead and 24-step-ahead predictions. Figure 3 illustrates typical results of the 1-step-ahead forecast and 24-step-ahead forecast implemented with the models. The results of the 1-step-ahead forecast are obviously better than the 24-step-ahead results. The 1-step-ahead results match closely the observed values. The 24-step-ahead results, as indicated by the two samples on the left-hand side of Figure 3 for Stations A and B, reflect the general trend of the data. This is despite the fact that some unusual trends (e.g., the first two days of Station A and the unusual peak of Station B in the data series) were not well predicted. Those results are discussed later in further detail and in comparison with results obtained from the neural network models described in the next section.

FORECAST MODEL DEVELOPMENT—NEURAL NETWORKS

Network Configuration

A typical neural network is generally constructed into three layers of processing units: the input, hidden, and output layers. Processing units in each layer are connected through links, each of which is assigned a weight to depict its strength. A three-layer neural network is sufficient to define arbitrary linear decision regions under mathematical space. Cybenko[18] demonstrated that simple backpropagation neural networks, when given sufficient processing units in the hidden layer, can approximate any function. This three-layer configuration is most likely applicable for the studied data series.

Observation of the daily variations of pollutant levels provided by the monitoring data reveals that a roughly 24-hr cycle exists. In addition, among tested ARIMA models, the 24-hr model was found to be the best one. Therefore, this work used the 24-n-1 neural network configuration, as illustrated in Figure 4. The input layer used 24 input units, representing the data obtained in a consecutive 24 hr. The output layer used 1 output unit, representing the forecast value obtained 1 hr after the 24-hr interval for data used in the input layer. Regarding the number of hidden layer processing units, no appropriate theory is yet available to determine the optimal number. Following the suggestions of Weigend et al.,[19] numbers not exceeding one-tenth the size of the training patterns were used to determine the number of hidden layer processing units. In this work, 2, 5, 8, 11, 14, 17, 20, 23, 26,

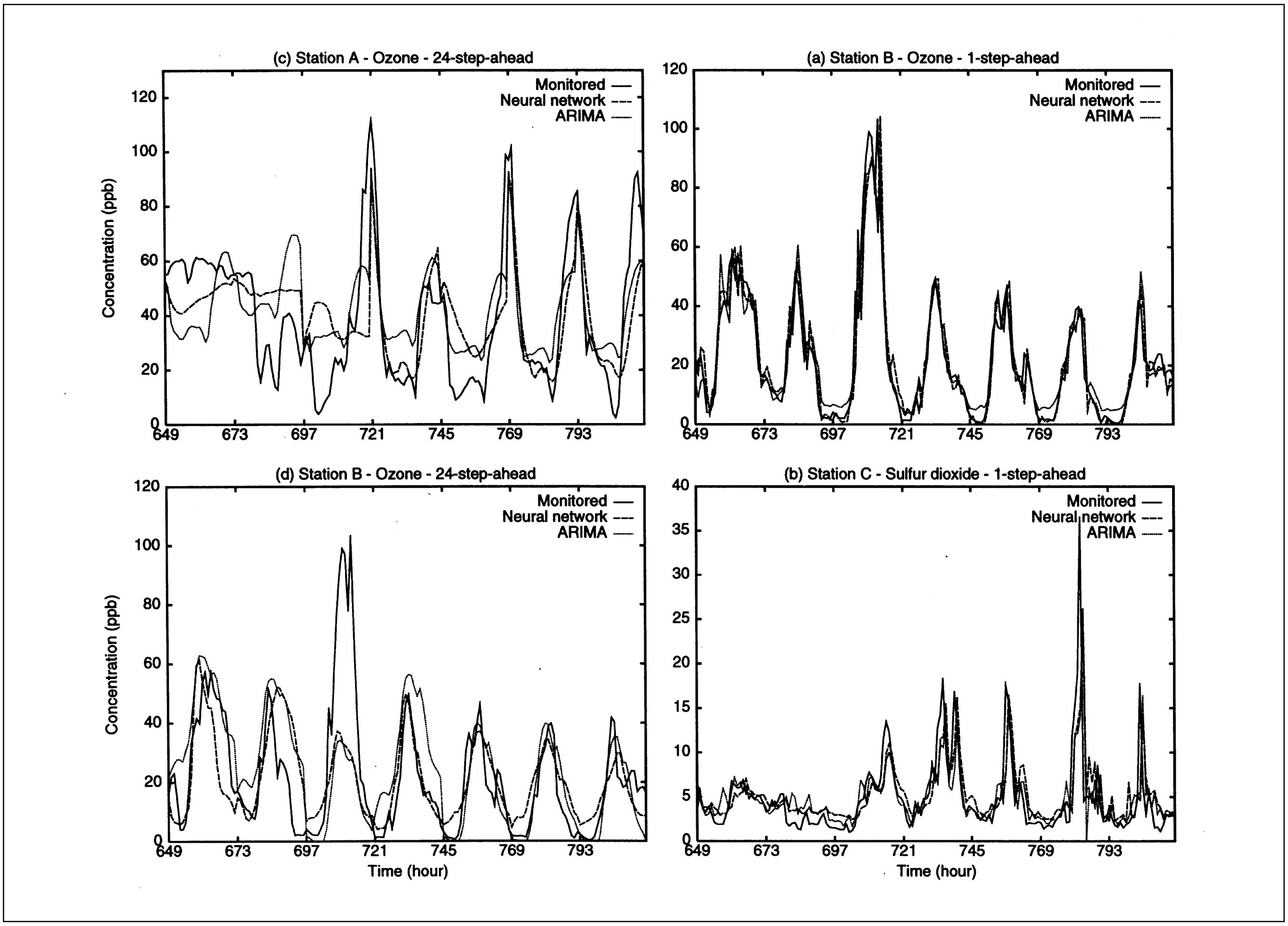

Figure 3. Typical results for 1-step-ahead and 24-step-ahead forecasts obtained from ARIMA and neural network models.

29, 32, 35, and 38 hidden layer processing units were used. In forecasting the value of x_{t+1}, the normalized data at x_{t-23}, x_{t-22}, ...x_t were taken as the input values of the processing units of the input layer, and the value at x_{t+1} was obtained from the unit of the output layer; that is, past air pollution monitoring data were input into the input layer and the value of the next period was predicted from the output layer data. Xerion,[20] a publicly accessible neural network software, was used to construct the neural network models.

Training Pattern Preparation

A neural network is established through learning. Representative known patterns or samples must be provided for training the network. An insufficient amount of patterns may render the neural network incapable of learning the properties of the system, and its forecast capability may subsequently diminish. On the other hand, too many patterns prolong its learning time. This study used normalized monitoring data obtained in four consecutive weeks as the training pattern set. The interval correlation method, probability correlation method, and the normalization method with a target range of –1 to 1 were tested in the process of normalizing the input data. Those results revealed that the air quality data for this study required only simple linear transformation by dividing the data by their maximal value to make them between 0 and 1 after normalization.

Neural Network Training

The learning algorithm adopted in this study to train a neural network was the backpropagation learning algorithm.[21] With the algorithm, the derivatives of predicted errors to network link weights were utilized to correct the weights. The algorithm attempts to find the optimal set of weights to let the network be able to produce predicted outputs that match the provided training patterns. It is basically a nonlinear programming optimization problem. The vector of all derivatives was used as a gradient vector. Each iteration determined the search direction and the moving distance in that direction, and the weight set was corrected according to the determined direction and distance. This process was repeated until all training patterns were learned. The algorithm is divided into two steps: the forward pass and the backward pass. A brief description of these two steps is provided below.

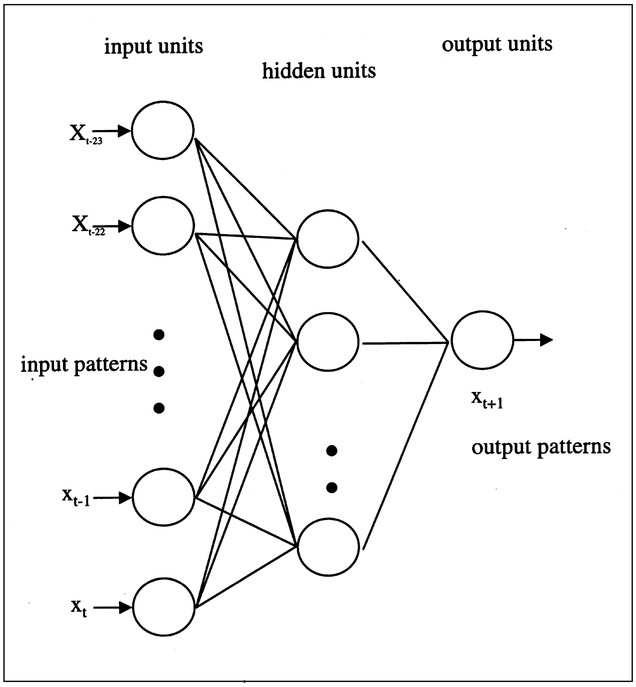

Figure 4. Structure of the univariate neural network used in this study.

Forward Pass. This step involves calculating the difference between output value and target value according to the current weight set. The net input and output values of every pattern are first computed. The output values of the preceding layer are used as the input values of the processing units linked behind them. The net input value of pattern *p* on unit *i* can be computed as

$$net_{pi} = \sum_{\substack{j \in previous \\ layer}} w_{ij} a_{pj} + b_i \tag{7}$$

where a_{pj} is the activation value of unit *j* of pattern *p*, w_{ij} is the weight of the link connecting unit *j* in the previous layer and current unit *i*, and b_i is the bias linked to unit *i*. The activation or output value of unit *i* is determined by a conversion function $f(net_{pj})$. The mathematical objective of the conversion function is to determine the output range at every point to coordinate with the data of a training pattern. Results obtained from the sigmoidal conversion function, $f(net_{pj}) = (1 + exp[-net_{pj}])^{-1}$, were reported; other functions such as identical, exponential, and negative exponential functions were also tested but not reported herein.

After the output value is determined for each input pattern, it can be compared with the real value (or target value). Assuming the training pattern set is (x_1, T_1) (x_2, T_2), ..., and (x_p, T_p), where $x_1, x_2, ..., x_p$ are input vectors, the total prediction error can then be calculated by the following equation:

$$E = \sum_p \left(T_p - O_p\right)^2 \tag{8}$$

where T_p is the target output unit value of pattern *p* and O_p is the output unit value predicted by the neural network for pattern *p*.

Backward Pass. After computing the prediction error of the network, a backward (going backward from the output layer) correction of weight values was carried out. Backward pass is a nonlinear optimization solving process with the objective of *E* being minimized. Therefore, methods used to solve a nonlinear programming optimization problem can be also applied. This study adopted steepest descent, momentum descent, and conjugate gradient, Rudi's conjugate gradient, and conjugate gradient with restart methods[20] during this step. Rudi's conjugate gradient method, developed by Rudi Mathon,[20] is a revised version of the conjugate gradient method. Of these methods, the steepest descent and momentum descent methods required a long training computational time, with results worse than those from other conjugate gradient-based methods. Therefore, results for the two methods are not discussed here.

These methods are mainly gradient-based. Each iteration of the methods determines an appropriate search direction to improve the objective function. Once the search direction is determined, the step size to move along the direction in the searching space must be determined. Three line search methods—Rudi's, Ray's, and Tap's—[20] were applied. A line search method attempts to find the minimum value of the objective function value along a given search direction. For instance, a situation is considered in which *W* is the weight set and *S* is the search direction vector. The line search methods should find a η that minimizes the error function E(W+ηS). When this pass is finished, the forward pass is initiated again to recalculate errors. Such a process is repeated until the error is reduced to a pre-specified, acceptable tolerance value.

Prediction

Table 1 lists the 16 case sets tested in this study with varied gradient and varied line search methods, excluding the cases with the steepest and momentum descent methods. With 13 different numbers of hidden layer processing units for each case set, a total of 208 cases were implemented. The seed numbers listed in the table were used for a random generator to generate an initial weight set of the neural network to be established. Figure 2 illustrates the typical results obtained by the neural networks in the training stage. In this stage, most neural networks have adequately learned the provided patterns. Following the completion of training, neural networks were utilized to perform a 1-step-ahead and 24-step-ahead forecast. Figure 3 illustrates partial results of those forecasts. The 1-step-ahead forecast gave better results than

Table 1. Tested case sets for the neural network model.

Case Set	Seed Number	Gradient Method	Line Search Method
1	197	Conjugate gradient	Rudi
2	197	Rudi conjugate gradient	Rudi
3	197	Conjugate with restarts	Rudi
4	197	Conjugate gradient	Ray
5	197	Rudi conjugate gradient	Ray
6	197	Conjugate with restarts	Ray
7	197	Conjugate gradient	Tap
8	197	Rudi conjugate gradient	Tap
9	197	Conjugate with restarts	Tap
10	1709	Conjugate gradient	Ray
11	1709	Rudi conjugate gradient	Ray
12	1709	Conjugate with restarts	Ray
13	2591	Conjugate gradient	Ray
14	2991	Rudi conjugate gradient	Ray
15	2591	Conjugate with restarts	Ray
16	197	Rudi conjugate gradient	Ray

the 24-step-ahead forecast. Figure 5 shows mean-square-error values of the 1-step-ahead forecast of the neural networks with varied numbers of units of the hidden layer. This figure reveals that the different numbers of hidden layer processing units did not significantly affect the neural network results.

COMPARISON OF THE TWO MODELS

In the training stage, the performances of neural network models were slightly better than those of time-series models, although both sets of results were generally acceptable. The forecast results indicate that the neural network performances in the forecast stage were generally as good as those of time-series models. A comparison of the forecast value with the actual value revealed the mean-square-error values of the neural network model results to be generally smaller than those of time-series models for a 1-step-ahead forecast, although the differences were insignificant. In the 24-step-ahead forecast, some neural network model results, such as the one shown in Figures 3(c) and 3(d), were superior to those

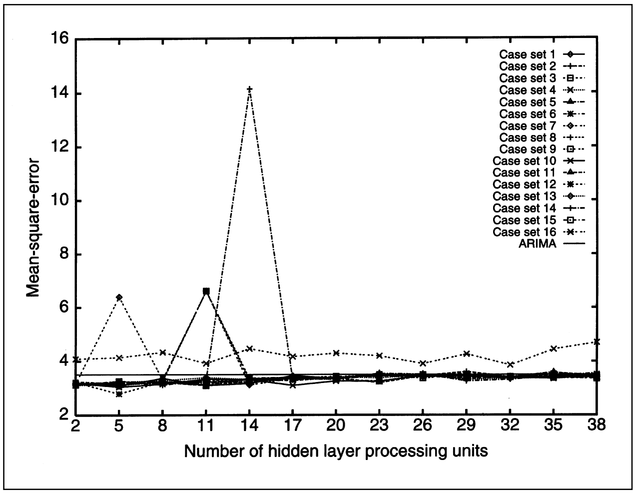

Figure 5. Mean-square-errors of ARIMA and neural network models for varied case sets with varied numbers of hidden layer processing units for monitoring station C.

for ARIMA models, although the 24-step-ahead neural network results were not as good as those for the 1-step-ahead forecast.

Table 2 summarizes the number of cases (out of 208 tested cases) in which the neural network model result with a mean-square-error (to the observed data) was smaller than that of the ARIMA model with varied training methods (cases) and varied numbers of hidden layer processing units. According to the table, the forecast performance of a neural network model can be as good as or superior to that of an ARIMA model. Some neural network results are obviously inadequate, because of the nonlinearity in the optimization searching procedure for determining the weights (or strengths) of links of the neural network, or because of network over-training.

CONCLUSION

Because of its self-learning and self-adaptation capabilities, a neural network model can automatically adjust connection weights on the basis of provided training samples. In contrast, the conventional time-series model frequently requires manual identification of model parameters by an analyst with expertise and previous experience in establishing an ARIMA model. A circumstance in which a previously manually determined model is not applicable to new data requires that a similar manual process be performed again to update the previous model. A neural network, however, can be self-adjusted without manually determining model parameters. With the neural network forecast model, human-made mistakes or erroneous judgments can be minimized and the reliability of the forecast system enhanced.

Table 2. Summary of numbers of cases[a] with neural network prediction performance better than that of ARIMA models based on mean-square-errors.

	Station A (O_3)	Station B (SO_2)	Station B (O_3)	Station C (SO_2)	Station C (O_3)
1-step-ahead	184	79	118	190	193
24-step-ahead	172	135	189	134	202

[a]The number of all tested cases is 208.

The simulation results reveal that the forecast performance of the neural network is superior to the 3conventional ARIMA in many tested cases. According to the current status of acquiring ambient air pollution monitoring data in Taiwan, a neural network requires little space for data storage without additional technical personnel for establishing the models. The neural network approach is a promising technique for developing an automated forecast system for ambient air quality.

This study applied univariate neural network models to perform the forecast. Results are acceptable for most 1-step-ahead forecasts and some 24-step-ahead forecasts. Such a univariate model requires less space to store data and less computational time to construct the model and to perform the desired forecast. However, results for some 24-step-ahead forecasts are not found to be as good as those for a 1-step-ahead forecast, especially for those data not following the general trend of the data series. Such an unusual change may be strongly influenced by external factors, which may not possess a temporal nature or are difficult to predict, thereby making it difficult for both a univariate ARIMA and neural network to provide an effective forecast.

Unfortunately, comprehensive monitoring data for multiple factors are not yet available in Taiwan. Current research is therefore focused on expanding the neural network with external multiple factors that are not continuously monitored. Unusual cases are extracted from data series and carefully inspected for possible predictable relations within these occasionally monitored data for external factors. The relations are meant to be used to improve the neural network for a 24-step-ahead forecast for unusual data variation with limited data for external factors. Any significant progress made in this endeavor will be reported in the future.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council, Taiwan R.O.C. for partial financial support of this research under Contract No. NSC-84-2211-E009-006. A number of friends on the Internet are also appreciated for their assistance in establishing the neural network models.

REFERENCES

1. California Air Resources Board. *An Evaluation of a Medical Advisory Notice to Persons with Respiratory Disease or Coronary Artery Disease*; Technical Advisory Committee Report: Sacramento, CA, 1972.
2. Merz, P.H.; Painte, L.J.; Ryason, P.R. "Aerometric data analysis—time-series analysis and forecast and an atmospheric smog," *Atmos. Environ.* **1972**, *6*, 319-342.
3. Chock, D.P.; Terrell, T.R.; Levitt, S.B. "Time-series analysis of Riverside, California air quality," *Atmos. Environ.* **1975**, *9*, 978-989.
4. Horowitz, J.; Barakat, S. "Statistical analysis of the maximum concentration of an air pollutant: effects of autocorrelation and non-stationarity," *Atmos. Environ.* **1979**, *13*, 811-818.
5. Murray, L.C.; Farber, R.H. "Time-series analysis of an historical visibility data base," *Atmos. Environ.* **1992**, *10*, 2299-2308.
6. McCollister, G.M. "Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants," *Atmos. Environ.* **1975**, *9*, 417-423.
7. Kapoor, S.G.; Terry, W.R. A Comparison of Two Automatic Systems for Building Vector Time-Series Models in Air Pollution Research. In *Time-Series Analysis: Theory and Practice*, vol. 7; Anderson, O.D., Ed.; Elsevier Science Publishers: Cincinnati, OH, 1985, 211-200.
8. SAS Institute Inc. *SAS/ETS Software: User's Guide:* Cary, NC, 1991.
9. Nelson, M.M.; Illingworth, W.T. *Neural Nets*; Addison-Wesley: Reading, MA, 1991.
10. Boznar, M.; Lesjak, M.; Mlakar, P. "A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain," *Atmos. Environ.* **1993**, *27B*, 221-230.
11. Comrie, A.C. "Comparing neural networks and regression models for ozone forecasting," *J. Air & Waste Manage. Assoc.* **1997**, *47*, 653-663.
12. Lapedes, A.; Farber, R. Nonlinear Signal Processing Using Neural Networks. Presented at *IEEE Conference on Neural Information Processing Systems—Neural and Synthetic*, Denver, CO, November 8–12, 1987.
13. Chakraborty, K.; Mehrotra, K.; Mohan, C.K.; Ranka, S. "Forecasting the behavior of multivariate time-series using neural networks," *Neural Networks* **1992**, *5*, 961-970.
14. Box, G.E.P.; Jenkins, G.M. *Time-Series Analysis Forecasting and Control*; Holden-Day: San Francisco, CA, 1977.
15. Bowerman, B.L.; O'Connell, R.T. *Forecasting and Time-series: An Applied Approach*; 3rd edition; Duxbury Press: Belmont, CA, 1993.
16. Newton, H.J. *TIMESLAB: A Time-Series Analysis Laboratory*; Wadsworth & Brooks/Cole Publishing Company: Pacific Grove, CA, 1989.
17. Ljung, G.M.; Box, G.E.P. "On a measure of lack of fit in time models," *Biometrika* **1978**, *64,* 517-522.
18. Cybenko, G. "Approximation by superposition of a sigmoidal function," *Math. Control Signal System* **1999**, *2*, 303-314.
19. Weigend, A.S.; Huberman, B.A.; Rumelhart, D.E. "Predicting the future: a connectionist approach," *Intl. J. Neural Systems* **1990**, *1*, 193-209.
20. Camp, D.; Plate, T.; Hinton, G. *The XERION Neural Network Simulator*, Version 3.1; Department of Computer Science, University of Toronto: Toronto, Canada, 1993.
21. Rumelhart, D.E.; McClelland, J.L; Parallel Distributed Processing Research Group. *Parallel Distributed Processing*; MIT Press: Cambridge, MA, 1989.

About the Authors

Dr. Jehng-Jung Kao (corresponding author) is currently the director of the Environmental Protection and Safety Center and professor in the Institute of Environmental Engineering, National Chiao Tung University, 75 Po-Ai Street, Hsinchu, Taiwan, Republic of China (or jjkao@green.ev.nctu.edu.tw). At the time this work was done, Mr. Shang-Shuang Huang was a graduate student in the Institute of Environmental Engineering, National Chiao Tung University. At present, he is working as an engineer in Central Insurance Co., Ltd., Taipei, Taiwan, R.O.C.