Proceedings of the Third International Symposium on Air Quality Management at Urban, Regional and Global Scales. pp. 354-363, 26-30 September 2005, Istanbul – Turkey

METEOROLOGICALLY ADJUSTED GROUND LEVEL OZONE TRENDS IN SOUTHERN TAIWAN

Kuang-jung Cheng[1], Che-hui Tsai[1], Hsu-cherng Chiang[1] and Ching-wen Hsu[2]

[1] Department of Water Resources and Environmental Engineering,
Tamkang University, Tamsui, Taipei-hsien, Taiwan
[2]Department of Environmental Engineering and Science,
Tajen Institute of Technology. Yanpu, Ping-tung, Taiwan

ABSTRACT

Two methods were used to calculate the meteorologically adjusted ground level ozone trends in southern Taiwan. The first method utilized is a robust MM linear regression method. The second approach uses a multilayer perceptron (MLP) neuron network method. The observations obtained from 16 monitoring stations were analyzed and divided into six groups by hierarchical divisive clustering procedure. The daily maximum 1 h and 8 h ozone concentrations for each region are then calculated. The meteorologically adjusted trends obtained by linear regression method and MLP method are smaller than the unadjusted trends for all regions and average time. It indicts that the meteorological conditions in Taiwan tend to increase ambient ozone concentrations in recent years.

Key Words : linear regression, ANN, MLP, statistical analysis, long-term trend

1. INTRODUCTION

Photochemical air pollution in Taiwan is a serious environmental issue. In the past two decades, aggressive control strategies have been employed by the government to reduce the emissions of ozone precursor substances (NO_x and VOCs). These efforts substantially reduced the ambient concentrations of NO_x and VOCs, however, ground-level ozone concentrations still exhibit increasing trends in southern Taiwan (Chen et. al., 2004). Possible reasons of this problem include: change of emission patterns, modifications of land use, annual variations in meteorological conditions, and global warming. The objective of this study is to investigate the effects of the annual variations in meteorological conditions on long-term ozone trends in southern Taiwan.

It is well known that variations in meteorological conditions at different time scales can exert sufficiently large impacts on ozone concentrations. This makes measuring the effectiveness of a control program difficult. Great efforts have been made to separate the effects of meteorological conditions from the effects of emission reductions and other factors (Bloomfield et al., 1996; Cobourn and Lin, 2004; Cox and Chu, 1993, 1996; Fiore et al., 1998; Rao et al., 1995; Xu et al., 1996). Many statistical methods have been employed to investigate this problem, but no one method is most appropriate for all purposes and all meteorological scenarios

(Thompson et al., 2001). In this research, we used robust MM least square method and multilayer perceptron (MLP) artificial neural network (ANN) for analysis.
For linear regression model, ozone concentrations are expressed as linear functions of observed meteorological parameters and other factors. This approach is straightforward, however, it assume simple linear and additive associations between the variables are inadequate to capture interactions and nonlinearities in the ozone response. On the contrast, the ANN methods are more complex and flexible than linear statistical models. They are able to model strongly non-linear relationship between meteorological parameters and ozone concentrations. The main reason for selecting the MLP model for air quality prediction was its accuracy and reliability, compared with other available ANN model categories.

In this study, two methods were used to estimate the meteorologically-adjust ozone trends in Taiwan. In addition, the following factors were considered in this study:

(1) Because Daily maximum 1 h average concentration is the current ambient air quality standards adapted by many countries, trends associated with daily maximum 1 h concentrations are the focus of most statistical assessments. However, as noted by USEPA, daily maximum 8 h concentrations is also an important parameter from the aspect of health. Thus, two dataset (1 h and 8 h ozone concentrations) will be considered.

(2) In the past, the so called “single-site models” were used. This approach models the relationship between ozone and meteorological variables measured at the same site (Gardner et al., 2000; Thompson et al., 2001). The formation of ozone in the troposphere is a complex process, involving regional transport of ozone and its precursors. Hence, regional normalization models may be superior to single-site models. Thus, in this study we clustered of sites having similar O3 concentrations into same group by a hierarchical divisive clustering procedure. Six groups were considered.

2. METHODS

2.1 Linear regression model

A simple linear regression model was used in this study. It can be expressed as:

$$O_3 = \alpha + \beta Y + a\sin(kt) + b\cos(kt) + \sum_{i=1}^{n} c_i M_i \quad (1)$$

where α, β, *a*, *b* and c_i are coefficients to be determined by the regression procedure. The input variables include: Y a real number to represent year interval, $k = 2\pi/365$ the wave number, t the time in days starting from beginning of each year, M the meteorological parameters. Various combinations of meteorological variables were tested to determine the most appropriate form of the model. The meteorological parameters actually used in this study are shown in Table 1.

Eq. 1 consists three components: the long-term trend caused by anthropogenic emission change (βY), the seasonal term (*asin(kt)+bcos(kt)*), and meteorological effects. The coefficient β represents the slope of the long-term trend. As mentioned before, six regions were considered and two O_3 concentration targets were tested for each region. Hence, a total of 12 different ozone dataset were considered.

Table 1 input meteorological parameters

Ave. of surface wind speed at 11:00 and 14:00 LST
Ave. of u-components at 11:00 and 14:00 LST
Ave. of v-components at 11:00 and 14:00 LST
Ave. of surface temperatures at 11:00 and 14:00 LST
Ave. of surface relative humidity at 11:00 and 14:00 LST
Ave. of cloudiness at 11:00 and 14:00 LST
Change of 850mb height between 02:00 and 08:00LST
Change of temperature at 850mb between 02:00 and 08:00LST

Two regression models were run for each data set. The first model discards the meteorological parameters and obtains trends that are unadjusted. The second model considers all terms in the above equation and adjusted trend can be obtained. By comparing these two results, we can reveal the meteorological impacts on ozone trends.

The model was fitted using the robust MM regression technique in S-Plus (Insightful Co., 2001). The robust regression fit is minimally influenced by outliers in the dependent variables as well as dependent variable. This method is adopted because it has smaller RMSE than traditional linear least square method.

2.2. MLP methodology

For the MLP methodology, it is assumed that a time series of ozone $O(t)$ can be expressed as the sum of a long-term trend $T(t)$, a seasonal $S(t)$, meteorological $M(t)$ component, and error, or

$$O(t) = T(t) + S(t) + M(t) + E(t) \qquad (2)$$

The seasonal component corresponds to the annual cycle cause by solar radiation, whilst the short-term component is associated with the variations of meteorological variables. The MLP ANN was used to estimate $S(t) + M(t)$ based on daily meteorological and seasonal predictors. The residual R(t) is defined as $O(t)$ minus $S(t)+M(t)$, which is the sum of the long-trend term and random errors, or

$$R(t) = O(t) – [S(t) + M(t)] = T(t) + E(t) \qquad (3)$$

The long-term trend represents trends in ozone due to precursor emission changes and variations in the background concentrations of some related tropospheric trace gases.

We assume that:

$$R(t) = \alpha + \beta Y + E \tag{4}$$

The values of coefficients α and β were determined by a simple least square method.

Three steps are needed to determine the meteorologically-adjust ozone trend in this approach: (1) Using MLP and input meteorological data to estimate *S(t) + M(t),* (2) compute the residual *R(t)* by Eq. 3, (3) estimate adjusted-trend β by Eq. 4.

The feed-forward back-propagation MLP was used in this study; this model category will be abbreviated simply as MLP in the following. The MLP model was trained by using the trainbr algorithm in the MATLAB Neural network toolbox (Demuth and Beale, 2004). Training involves finding the set of MLP network weights, which enable the MLP model to represent the underlying patterns in the training data. As suggested by Gardner and Dorling (2001), MLP models with two hidden layers were used. There is no a standard way to decide the number of hidden neurons. A trial-and-error approach was used in this study. By changing the neurons numbers in first and second layers, the RMSE surface can be computed. Since the RMSE do not change significantly if the numbers of neurons larger than 6. Hence, six neurons in hidden layer 1 and four neurons in hidden layer 2 were adopted (see Fig. 1). The transfer function in the two hidden layer nodes was the log-sigmoid and tangent sigmoid functions, respectively, while for the output layer nodes the unbounded linear function was used.

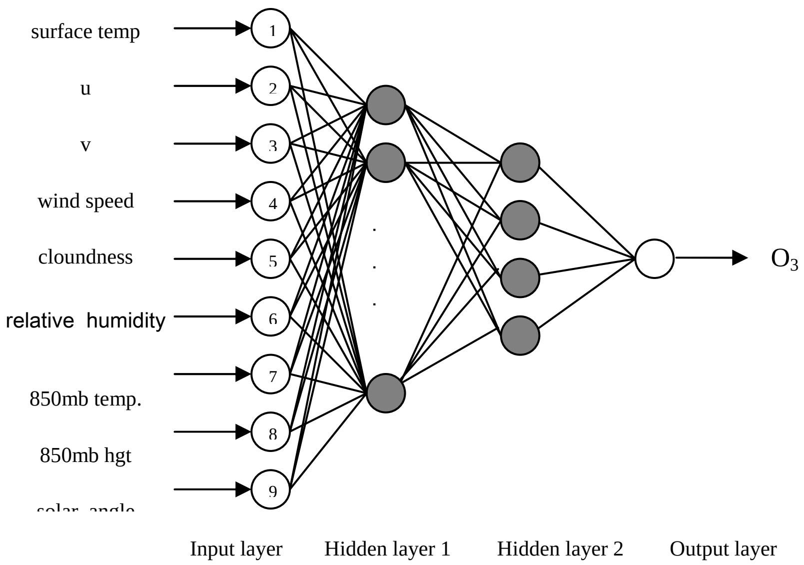

Fig 1 The architecture of a multilayer neural network

The available data were randomly divided into two subsets, training data and validation data. The training data is about half size of the dataset considered, the remaining data are used for validation. The MLP models were then trained to learn the relationship between the predictors and daily maximum ozone concentrations using training data. Validation data set is to check the performance of the MLP network to determine the epoch at which training should be stopped to avoid over-training. Typically the global minimum is not reached and a good local minimum is treated as an acceptable solution. We train MLP models 50 times and selecting the model with the best generalization performance in order to reduce the likelihood of local minima causing problems.

Following training, the model residuals were calculated and interpreted as the meteorologically adjusted long-term trends and random errors. The trends were estimated from eq. 4 by a general least square method.

3. DATA

The original air quality data utilized in this study are obtained from EPA, Taiwan. The data consist of hourly averaged concentrations of ozone and other relevant pollutants collected from 16 stations in southern Taiwan over the four years from 2000 to 2003. Meteorological data were taken from weather stations close to air quality monitoring sites. These weather stations were operated by Central Weather Bureau. The locations of air quality monitoring stations are shown in Figure 2.

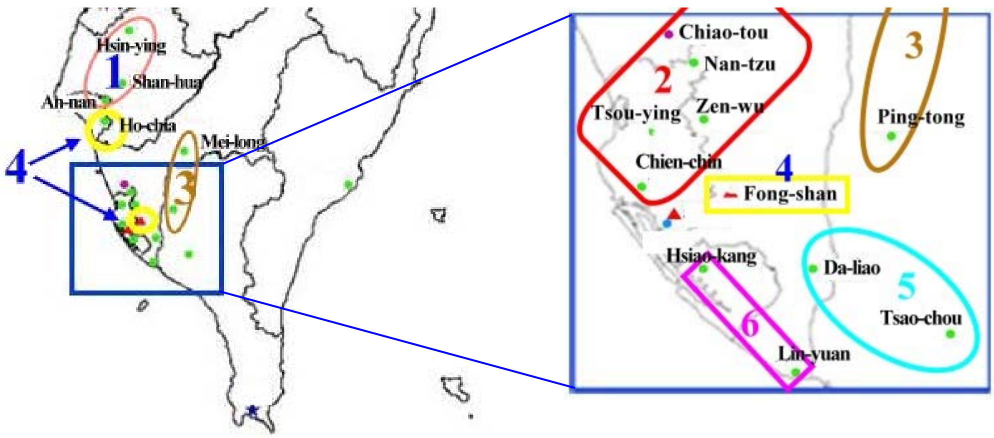

Fig 2 Locations of 16 air quality monitoring stations in southern Taiwan

The ozone response variable and meteorological predictor variables are calculated on a daily basis. If more than 6-h of data are missing on 1 day for any weather variable or for ozone, then the entire day is omitted from the analysis. The missing values typically reduce the available data by 5-10%. Most of the missing data correspond to extended down times caused by maintenance or instrument malfunction. There is no evidence that down times are related to ozone levels, so the missing values are omitted.

The hierarchical divisive clustering procedure was used to aggregate sites into several groups based on site-specific O_3 concentration data. Hierarchical divisive methods start with all observations in a single group and proceed until each observation is in a separate group. As shown in Fig. 3, air quality stations were classified into six groups by cluster analysis. Since monitoring stations in each group are similar, averaged concentrations over each region were used for analysis.

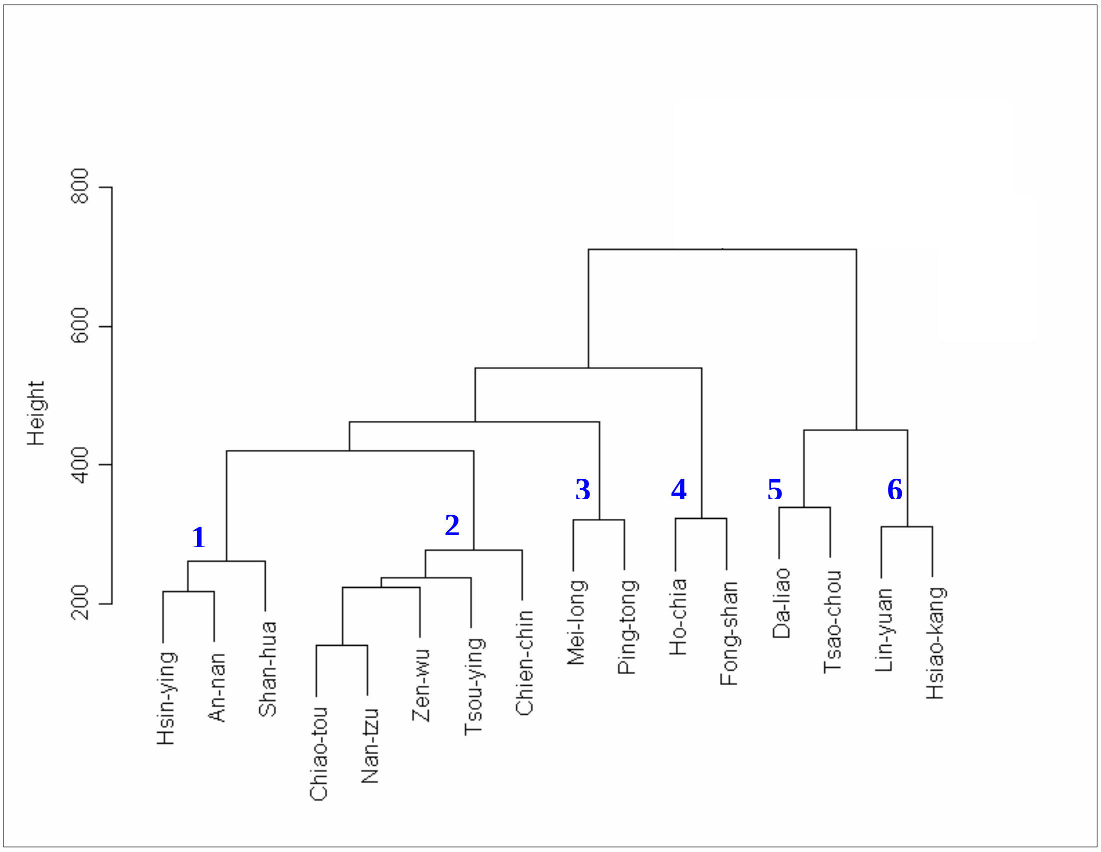

Fig 3 Air quality stations were classified into six groups by ierarchical divisive method

4. RESULTS AND DISCUSSION

(1) Trends of daily maximum 1 h ozone concentrations

Table 1 shows the meteorologically adjusted and unadjusted long-term components of the daily peak 1 h ozone time series at each region. Results including correlation coefficients, root mean square error (RMSE) and trend are listed in Table 1.

As shown in Table 1, β values for trend only analysis are all positive (except for region 2), which indicates that the unadjusted ozone trends are increasing. Having taken the meteorological effects into account, the meteorologically adjusted trends (or demeteorological trends) are most likely associated with precursor emissions. It is noted that the demeteorological trends are smaller than unadjusted trends. The implication is that if one seeks ozone trend without considering the meteorological impact, one would end up with a trend that could misrepresent the effects of the emission reduction.

Table 1 also shows that the addition of meteorological variables to a model with only seasonality and trend will reduce the prediction error (RMSE) and the standard error of the trend estimate. The values of correlation coefficients will increase if the meteorological effects are considered. The R^2 values ranged from 0.4 to 0.6. It is vary reasonable when compared with the results of other studies.

In terms of RMSE or R^2, as shown in Table 1, the linear regression model and MLP model exhibit similar levels of performance. The absolute values of the demeteorological trends obtained by MLP methods are smaller then that obtained by liner regression method.

Table 1 meteorologically adjusted and unadjusted long-term statistics of the daily peak 1 h ozone time series at each regions

Region	Trend only Linear regression			Demeteorological trend Linear regression			Demeteorological trend MLP		
	R^2	RMSE	β	R^2	RMSE	β	R^2	RMSE	β
1	0.20	22.61	1.14±1.09	0.57	16.62	0.73±0.66	0.55	16.29	0.4951
2	0.39	24.48	-0.14±0.70	0.41	20.75	-1.86±0.58	0.53	19.97	-1.4675
3	0.16	25.63	5.95±0.66	0.44	18.30	4.38±0.52	0.59	18.03	3.5523
4	0.36	24.54	4.46±1.02	0.59	19.57	5.72±0.68	0.53	18.73	2.999
5	0.31	28.76	0.08±1.09	0.57	22.81	-0.77±1.02	0.58	21.14	-0.3475
6	0.43	26.27	1.24±0.89	0.43	22.08	-0.53±0.64	0.56	20.82	0.0251

We can use two models developed in this study to ‘predict’ the daily peak ozone concentration for six regions. The results are shown in Fig. 4. This figure indicates that the differences between observed and predicted values are significant. It implies that some important factors, which have significant influence on the ozone concentration, were not well explained by the current models. These factors may include the short-term variations of emission conditions.

(2) Trends of daily maximum 8 h ozone concentrations

Since health effects research now shows that ozone affects public health over long periods of time, not just during a few 1 h peak events. A new standard based on daily maximum 8 h ozone concentration was promulgated in USA. Now, we will exam the results of daily maximum 8 h ozone concentration.

Table 2 is similar to Table1 except that daily maximum 8 h ozone concentrations were used. Since the values of maximum 8 h ozone concentrations are less than the values of daily peak 1-h ozone concentration, the values of RMSE and β appear in Table 2 are less than their corresponding values in Table 1. The correlation coefficients in Table 2 are larger. The β values would be reduced if meteorological conditions were considered.

Fig.5 shows the time series plots of the observed and predicted daily peak 8 h ozone concentrations of 2002 at six regions. The predictions of two models agree with the observations.

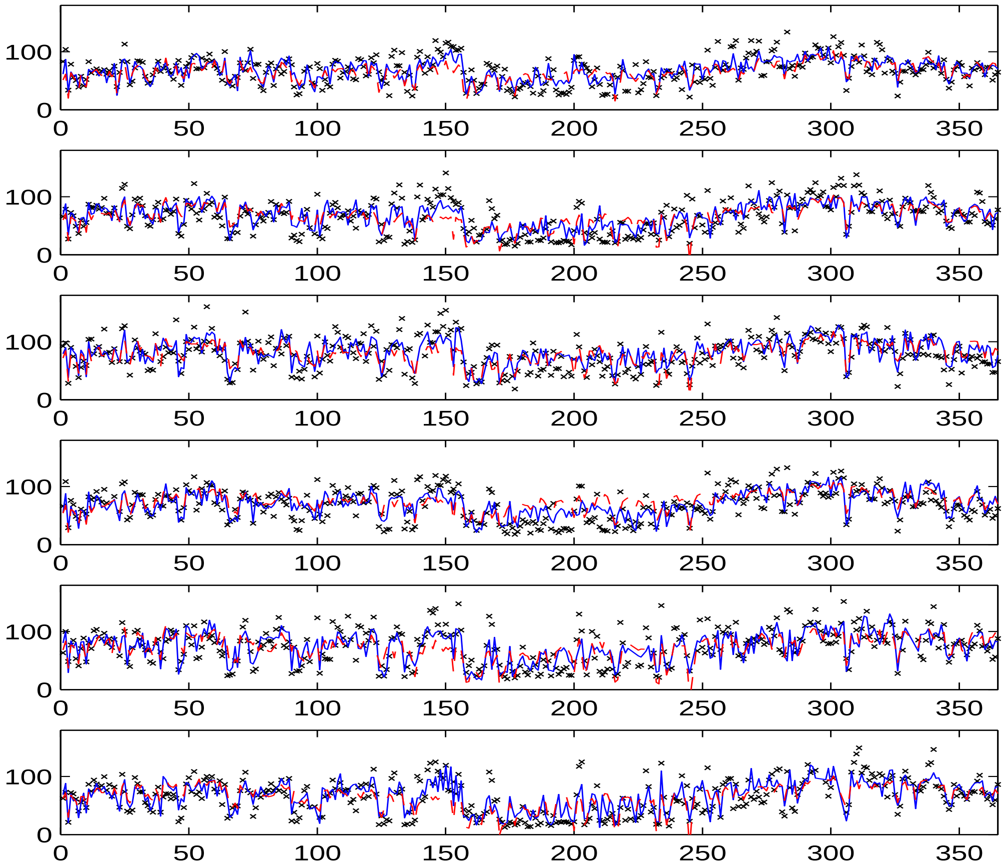

Fig. 4 The observed and predicted daily peak 1 h ozone concentrations of 2003 at six regions (blue line: least square method, red line: MLP method, x : observations)

Table2 meteorologically adjusted and unadjusted long-term statistics of the daily maximum 8 h ozone time series at each region

Region	Trend only Linear regression			Demeteorological trend Linear regression			Demeteorological trend RBFNN		
	R^2	RMSE	β	R^2	RMSE	β	R^2	RMSE	β
1	0.28	18.00	1.20±0.62	0.61	13.25	0.64±0.59	0.61	12.35	0.4477
2	0.46	18.97	0.43±0.63	0.63	15.79	-0.92±0.65	0.59	14.64	-0.5549
3	0.34	18.59	4.30±0.76	0.64	13.68	3.66±0.60	0.63	12.97	3.1609
4	0.40	19.22	3.91±0.74	0.43	15.68	3.56±0.41	0.58	14.15	2.6529
5	0.44	19.50	0.11±0.71	0.64	15.69	-1.79±0.57	0.62	14.55	-0.5677
6	0.50	19.52	1.96±0.75	0.63	16.84	0.16±0.72	0.61	14.78	0.3757

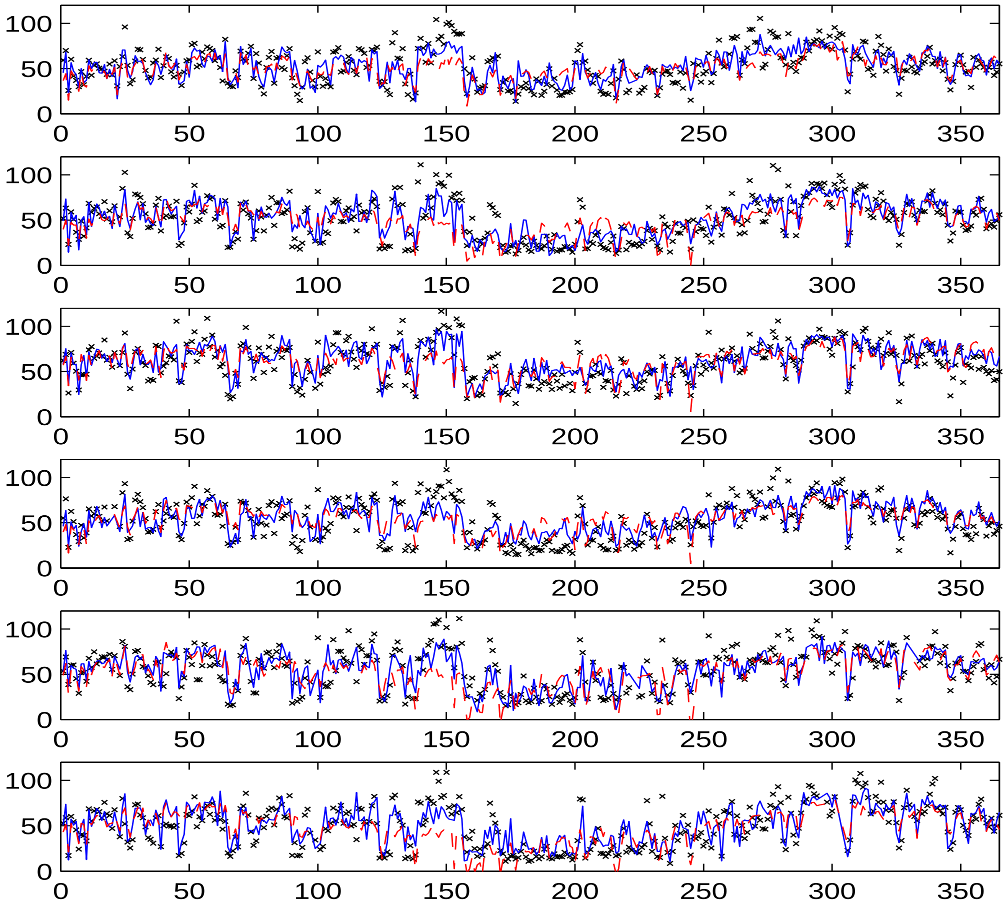

Fig. 5 The observed and predicted daily peak 8 h ozone concentrations of 2003 at six regions (blue line: least square method, red line: MLP method, + : observations)

As shown in Table 1 and 2, the meteorologically adjusted trends in region 2 and 5 are decreased. This may attribute to the effects of emission control. However, in other regions the meteorologically adjusted trends still exhibit increasing tendency. Since a new freeway was completed and operated in 2000, we believe the patterns of pollutant emission will be changed due to urban sprawl. It is well known that concentration distributions of air pollutant change while emission patterns change.

5. CONCLUSION

A robust linear regression method and a MLP neural network method were used to separate the effect of meteorological conditions on the ozone concentrations. After eliminating the meteorological factor in 2000-2003, the long-term trends (β) obtained by linear regression and MLP method are decreased. In recent years, the meteorological conditions in Taiwan tend to increase ambient ozone concentrations.

REFERENCES

Bloomfield, P.J., Royle, J.A., Steinberg, L.J., Yang, Q., 1996. Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment*, 30 (17), 3067-3077.

Chen, K-s, Ho, Y.T., Lai, C.H., Tsai, Y.A., and Chen, S-j, 2004. Trends in concentrations of ground-level ozone and meteorological conditions during high ozone episodes in the Kao-Ping Airshed, Taiwan. *J. Air & Waste Manage. Assoc.*, 54, 36-48.

Cobourn, W.G., and Lin, Y., 2004. Trends in meteorologically adjusted ozone concentrations in six Kentucky metro areas, 1998-2002. *J. Air & Waste Manage. Assoc.*, 54, 1383-1393.

Cox, W., and Chu, S., 1993. Meteorologically adjusted ozone trends in urban areas: a probabilistic approach. *Atmospheric Environment*, 27B, 425-434.

Cox, W., and Chu, S., 1996. Assessment of interannual ozone variation in urban areas from a climatological perspective. *Atmospheric Environment*, 30, 2615-2625.

Demuth, H., and Beale, M., 2005. Neural Network Toolbox User's Guide, Ver. 4., The Mathworks, Inc., Natick, MA.

Fiore, A.M., Jacob, D.J., Logan, J.A., and Yin, J.H., 1998. Long-term trends in ground level ozone over the contiguous United States, 1980-1995. *J. of Geophysical Research*, 103,1471-1480.

Gardner, M.W., and Dorling, S.R., 2000. Meteorologically adjusted trends in UK daily maximum surface ozone concentrations, *Atmospheric Environment*, 34, 171-176

Gardner, M.W., and Dorling, S.R., 2001. Artificial neural network-derived trends in daily maximum surface ozone concentrations, *J. Air & Waste Manage. Assoc.*, 51, 2001, 1202-1210

Insightful Co., 2001. S-plus 6 for Windows Guide to Statistics. Seattle, Washington.

Milanchus, M.L., Rao, T.S., Zurbenko, I.G., 1998. Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability. *J. Air & Waste Manage. Assoc.*, 48, 201-215.

Niu, X.F., 1996. Nonlinear additive models for environmental time series with applications to ground-level ozone data analysis. *Journal of the American Statistical Association*, 91, 1310-1321.

Rao, S.T., Zalewsky, E., Zurbenko, I.G., 1995. Determining temporal and spatial variations in ozone air quality. *J. Air & Waste Manage. Assoc.*, 45, 57-61.

Thompson, M.L., Reynolds, J., Cox, L.H., Guttorp, P., and Sampson, P.D., 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment*, 35, 617-630.

Xu, D., Yap, D., Taylor, P.A., 1996. Meteorologically adjusted ground level ozone trends in Ontario. *Atmospheric Environment*, 30, 1117-1124.