Analysis of Atmospheric Ozone Concentration Trends as Measured by Eighth Highest Values

Hsi-Hsien Yang*, Ho-Wen Chen, Tze-Wen Chi, Pao-Yu Chuang

Department of Environmental Engineering and Management, Chaoyang University of Technology, Wufong, Taichung 413, Taiwan

Abstract

The ambient air quality standard for ozone in Taiwan is 0.12 ppm (hourly average). To protect human health, this standard is not to be exceeded by the observed hourly ozone concentration. To test compliance, each day's maximum hourly ozone concentration is identified and the eighth highest value of the 365 daily hourly maxima for the entire year is calculated. To account for the uncertainty in measurement, the regulation stipulates using the three-year moving average of the eighth highest value to compare with the standard. In this study, observed ambient hourly ozone data from 1998-2002 at 9 continuous monitoring stations maintained by the government were collated and the eighth-highest concentration (MAX_8) was calculated for each site by year. For the estimate of confidence interval for MAX_8, a linear regression of ozone concentrations on their ranks was applied, as well as a quadratic logistic regression of odd ratios on ozone concentration. To estimate the confidence interval using the quadratic equation for inverse prediction, a Monte Carlo simulation was carried out in conjunction with the latter method. By taking into account the uncertainty expressed by the confidence interval, it was shown that MAX_8 did not exhibit differences statistically for all stations in the period.

Keywords: Air quality standard; Confidence interval; Monte Carlo method; Eighth highest value; Logistic regression.

INTRODUCTION

Ozone is designated as a criteria pollutant under the Air Pollution Control Act of Taiwan. High ozone concentration is a pervasive problem for the 1-h National Ambient Air Quality Standard. Tropospheric ozone is a secondary pollutant formed through the series of reactions of nitrogen oxides and active hydrocarbons under ultraviolet solar radiation (Colls, 2002). Due to its strong oxidative reactions, ozone can cause irritating symptoms on the respiratory system, such as coughing, asthma, headache, lethargy, and even lung

*Corresponding author. Tel.: 886-4-23323000 ext. 4451;
Fax: 886-4-23742365
E-*mail address*: hhyang@cyut.edu.tw

damage. Children, the elderly, patients, or persons with active outdoor activities are most vulnerable to ozone damage. Ozone can also cause damage to crops, paintings, and plastic products, such as tires (Heinsohn and Kabel, 1999). The ozone production cycle is driven by sunlight. However, other meteorological parameters, such as cloud cover, air temperature, relative humidity, atmospheric pressure and wind speeds can influence the kinetics of ozone production and distribution (Aneja *et al.*, 1994). Furthermore, this reaction is maximized during the summer time, when incoming solar radiation is greatest, together with the temperatures (Aneja *et al.*, 1999).

The eighth highest 1-h ozone concentration is used to delineate air pollution control zones and total quality control zones in Taiwan. The arithmetic average value of the eighth highest 1-h ozone concentration for 3 consecutive years is calculated, then listed in order to for each air quality monitoring station. The average first 50% of the highest values are taken, and those stations whose average values are less than 0.12 ppm shall be in compliance with air quality standards. Some uncertainties still exist despite considering the arithmetic average values of 3 consecutive years. This suggests that an ability to model the confidence intervals of MAX_8 would be desirable.

Monte Carlo simulation is a common name for a group of iterative statistical techniques. It has been used in several studies in environmental science for quantitative uncertainty analysis (Dodge, 2000; Moore and Londergan, 2001; Hanna and Davis, 2002; Zádor *et al.*, 2005). The advantages of the Monte Carlo method are that (1) it can be applied to a complete set of about 100 or more input parameters, (2) it allows useful estimates of the uncertainties in model outputs, (3) it allows use of standard nonparametric statistical tests concerning confidence intervals (Hanna *et al.*, 1998). A Monte Carlo run results in a large number of estimates, which can be displayed as a probability distribution. This highlights the fact that the final estimate is uncertain. The fundamental problem is therefore selecting probability distributions for these parameters (Rabl and Spadaro, 1999; Int Panis *et al.*, 2004).

In the literature, many researchers have investigated the ozone concentrations with different aspects. Among the papers that interest us are studies on the prediction of ozone concentration (Dodge, 2000; Koçak, 2000; Thompson *et al.*, 2001). Linear regression is the most familiar of the methodologies employed in studies. All linear regression models are open to the criticism that underlying chemical and physical processes are unlikely to be linear and additive. Bloomfield *et al.* (1996) argue that statistical linear models have difficulty capturing the complex relations between the variables and ozone. Nonlinear regression models are needed to approximate the true underlying mechanisms. Even then, if the interest is in extreme values, the regression models may be not sufficient (Smith and Huang, 1993). The inherent averaging in regression analysis often makes fitted models poor predictors of extreme values (NRC, 1991). An alternative approach, particularly useful in the context of modeling threshold exceedances, is to use extreme value theory (Thompson *et al.*, 2001). To our knowledge, no research has studied the uncertainty analysis for the prediction of ozone MAX_8. In this study, a method was established using Monte Carlo simulation to predict ozone MAX_8 and the corresponding confidence interval. The trends of expected ozone MAX_8 and the confidence intervals in central Taiwan are presented and discussed here.

METHODS

Data

An integrated air quality monitoring network was established by the Taiwan Environmental Protection Administration (TEPA) in 1993. This network of 71 stations continuously monitors the air quality in Taiwan. Ambient ozone concentration measurements provided by TEPA span the years 1998-2002. Ozone concentration is measured by using the chemiluminescence technique. In this study, nine routine air quality monitoring stations (Erlin, Dali, Jhushan, Situn, Shalu, Jhongming, Nantou, Changhua and Fongyuan) in central Taiwan were considered. Fig. 1 illustrates the locations and their geographical coverage. In addition to ozone concentration, temperature and rainfall were measured. The highest hourly ozone concentrations during each day for one year were sorted and the eighth value is the MAX_8.

Estimation of the expected MAX_8 and confidence interval by linear regression

The highest hourly ozone concentrations for each day were sorted to gain Y_x. The subscript "x" is the ranking and thus Y_1 means the maximum hourly ozone concentration of the 365 values in one year. In this study, the

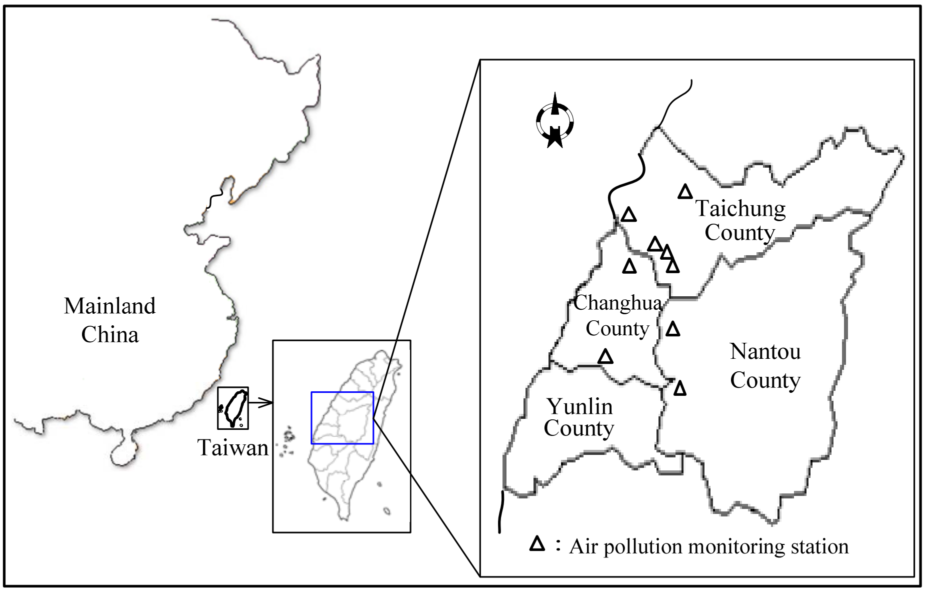

Fig. 1. Locations of the monitoring stations in central Taiwan.

ranking x is used in X-coordinate and Y_x is used in Y-coordinate. Logistic regression provides the most straightforward approach to predicting episodes of poor air quality (Dorling *et al.*, 2003). In this study, the logistic equation:

$$Y_x = \frac{a}{b + cx} \tag{1}$$

is used to simulate the distribution.
Eq. (1) is rearranged as:

$$\frac{1}{Y_x} = \frac{b}{a} + \frac{c}{a}x \tag{2}$$

where a, b and c are the parameters which are obtained by linear regression analysis. Then the simulation model of MAX_8 for the 9 air pollution monitoring stations can be gained.

The standard deviation of the expected MAX_8 can be calculated by Eq. (3).

$$Std_e = Std_R \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} \tag{3}$$

where
Std_e: standard error
Std_R: standard error of estimate
n: sample size
x: independent variable
$\bar{x}$: mean of independent variables
S_{xx}: sum of squares between samples

Then the confidence interval (CI) of the expected MAX_8 is estimated by the following formula with 95% confidence interval (Zar, 1999):

$$CI = \overline{\left(\frac{1}{Y}\right)} \pm Z_{\alpha/2} \times Std_e \tag{4}$$

where $Z_{\alpha/2}$: Z value for normal distribution of the α/2 area.

Estimation of the expected MAX_8 and confidence interval by Monte Carlo simulation

The standard deviation of the expected MAX_8 calculated by Eq. (3) was used to determine the optimum number of data. The rankings of the retained numbers were transferred to the probability P_x, where P_1 corresponds to the highest one concentration. Then a quadratic logistic regression equation was setup:

$$\ln\left(\frac{p_x}{1 - p_x}\right) = a'(Y_x)^2 + b'(Y_x) + c' \tag{5}$$

The propagation of errors through complex calculations was studied in this study. The crisp estimates of the above parameters are replaced by a probability distribution that describes the range of values that the parameters can take, as well as the probability that a certain value will actually occur. It is assumed that the three parameters reflect normal distribution. The procedure is then repeated 1000 times so that a large number of combinations of different input parameters occur. In this study, the parameters a’, b’ and c' were calculated with the commercially available SPSS software.

Let x = 8, MAX_8 can be calculated from Eq. (6).

$$Y_x = \frac{-b' \pm \sqrt{b'^2 - 4a'(c' - (\frac{p_x}{1-p_x}))}}{2a'} \quad (6)$$

RESULTS AND DISCUSSIONS

Statistical analysis of MAX_8

The MAX_8 values of ozone for the air quality monitoring stations during 1998-2002 are listed in Table 1. Except for Jhushan station, the mean MAX_8 values were all lower than the Ambient Air Quality Standard (120 ppb). Some local studies have suggested that the high ozone concentration in Jhushan was caused by the effect of terrain (TEPA, 2003). The bad airflow resulted in the accumulation of air pollutants in this area. The coefficients of variation are small (2.5%–9.2%). These show that the MAX_8 values did not change significantly for these years.

We further analyzed the occurrence time of MAX_8 for these stations. The MAX_8 appeared primarily around noon (10:00 am-3:00 pm) for all 9 air quality monitoring stations (Fig. 2). The occurrence time of MAX_8 is similar to that of the highest ozone concentration. No significant differences were found for the occurrence time of MAX_8 between these different air quality monitoring stations.

The influence of ranking number on goodness of fit

The ranking number may influence the goodness of fit of the simulation model. At Erlin station for example, the plot of ranking number vs. the inverse of MAX_8 according to Eq. (2) shows that the relation between ranking number and the inverse of MAX_8 is not linear (Fig. 3). There would be an error for the prediction of MAX_8 when linear regression analysis is performed. To improve the

Table 1. MAX_8 of ozone at 9 air quality monitoring stations in central Taiwan (ppb).

Stations	1998	1999	2000	2001	2002	Mean	Stdev	Cv (%)[1]
Erlin	93.7(362)[2]	102.8(357)	106.2(364)	100.5(357)	115.8(363)	103.8	8.1	7.8
Dali	119.2(354)	118.5(364)	113.2(365)	119.9(351)	120.5(361)	118.3	2.9	2.5
Jhushan	130.5(317)	130.1(351)	117.1(364)	124.5(355)	118.4(365)	124.1	6.3	5.1
Situn	100.2(360)	105.2(355)	100.5(361)	93.0(357)	116.7(365)	103.1	8.8	8.5
Shalu	94.4(338)	105.5(361)	93.2(358)	81.6(358)	96.4(363)	94.2	8.6	9.1
Jhongming	115.0(352)	108.9(365)	104.2(365)	106.0(358)	120.4(360)	110.9	6.7	6.1
Nantou	113.9(365)	106.3(360)	125.4(360)	117.4(355)	114.7(364)	115.5	6.9	6.0
Changhua	88.7(365)	102.5(355)	105.3(365)	108.8(357)	114.0(364)	103.9	9.5	9.2
Fongyuan	109.3(362)	118.3(359)	112.2(363)	113.1(350)	118.6(364)	114.3	4.0	3.5
Mean	107.2	110.9	108.6	107.2	115.1			
Stdev	13.9	9.4	9.5	13.7	7.4			
Cv (%)	12.9	8.5	8.8	12.8	6.4			

[1]Cv: coefficient of variation;
[2]The numbers in brackets are sample size

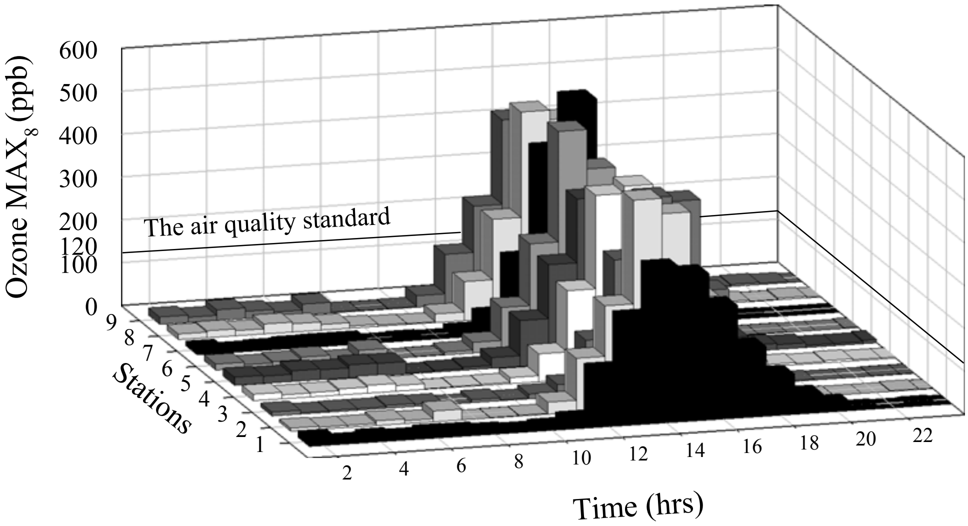

Fig. 2. MAX_8 at different occurrence time for 9 air quality monitoring stations. The station 1 is Erlin, 2 is Dali, 3 is Jhushan, 4 is Situn, 5 is Shalu, 6 is Jhongming, 7 is Nantou, 8 is Changhua, and 9 is Fongyuan.

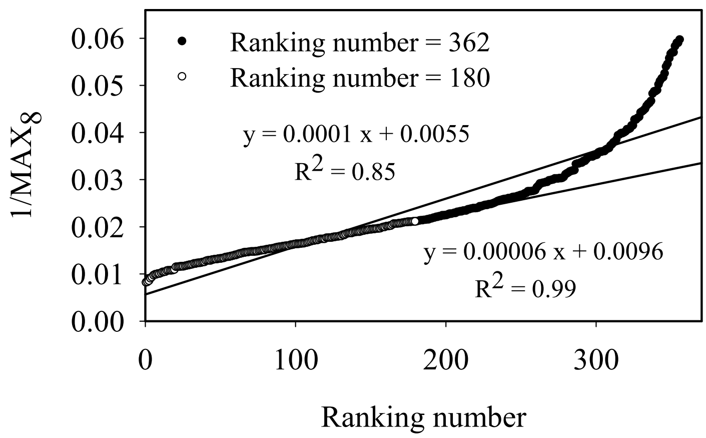

Fig. 3. The relation between ranking number and MAX_8.

goodness of fit, only some high ranking numbers were included in the simulation model.

The standard error of estimate (SEE) values of the linear regression analysis for the selected ranking numbers at Erlin station are listed in Table 2. The results indicate that there are no significant differences for SEE values when the ranking number is below 180. The results in Table 2 also show that the SEE values are small while the ranking number is below 180. Thus, the highest 180 ozone concentrations would be analyzed, as discussed in the following sections.

Prediction of MAX_8 and confidence interval by linear regression

The ranking number 180 is adopted in the

Table 2. The SEE of linear regression for different ranking numbers at Erlin station.

Ranking number	$1/MAX_8$*	1998	1999	2000	2001	2002
36	0.028	0.00032	0.00030	0.00034	0.00033	0.00035
60	0.017	0.00033	0.00031	0.00033	0.00034	0.00034
90	0.011	0.00034	0.00032	0.00036	0.00036	0.00035
120	0.008	0.00034	0.00032	0.00037	0.00037	0.00036
180	0.006	0.00035	0.00033	0.00035	0.00034	0.00037
240	0.004	0.00047	0.00049	0.00045	0.00043	0.00034
300	0.003	0.00165	0.00106	0.00111	0.00107	0.00076

*Expected value.

Table 3. SEE values by linear regression analysis.

Stations	1998	1999	2000	2001	2002	Mean
Erlin	0.000448	0.000295	0.000558	0.000341	0.000435	0.000415
Dali	0.000182	0.000297	0.000336	0.000351	0.000307	0.000295
Jhushan	0.000275	0.000283	0.000263	0.000234	0.000303	0.000271
Situn	0.000242	0.000525	0.000529	0.000461	0.000475	0.000447
Shalu	0.000708	0.000312	0.000352	0.000407	0.000367	0.000429
Jhongming	0.000274	0.000211	0.000267	0.000405	0.000415	0.000315
Nantou	0.000272	0.000163	0.000197	0.000135	0.000305	0.000214
Changhua	0.000391	0.000637	0.000417	0.000511	0.000408	0.000473
Fongyuan	0.000329	0.000255	0.000266	0.000230	0.000286	0.000273

Table 4. Expected MAX_8 values by linear regression analysis (ppb).

Stations	1998	1999	2000	2001	2002	Mean	Stdev	Cv (%)
Erlin	88.00	98.62	95.57	93.33	105.22	96.15	6.38	6.64
Dali	121.77	114.14	109.98	116.39	112.88	115.03	4.42	3.84
Jhushan	130.95	126.68	117.39	116.62	112.12	120.75	7.78	6.44
Situn	91.27	103.09	95.26	91.47	108.18	97.86	7.50	7.66
Shalu	93.06	98.71	84.62	78.41	92.06	89.37	7.92	8.86
Jhongming	108.90	106.27	99.54	102.10	115.24	106.41	6.12	5.76
Nantou	111.04	104.57	119.68	115.41	110.78	112.29	5.65	5.03
Changhua	87.47	91.82	100.77	96.83	103.48	96.07	6.51	6.78
Fongyuan	104.59	114.92	107.82	108.18	116.28	110.36	5.01	4.54

linear regression analysis. The SEE values in these analyses are quite small (Table 3). The results coincide with that proposed in the previous section. The small SEE values according to the analysis indicate the goodness of fit for the simulation model is acceptable.

The results of expected MAX_8 values by linear regression analysis are listed in Table 4. Most expected MAX_8 values are less than the actual MAX_8 values (Table 1), but the errors did not exceed 10%. This shows that the results of linear regression are similar to those of actual MAX_8 values. Therefore, the confidence intervals calculated by this model can be used for the actual MAX_8 values.

The expected MAX_8 values and their upper

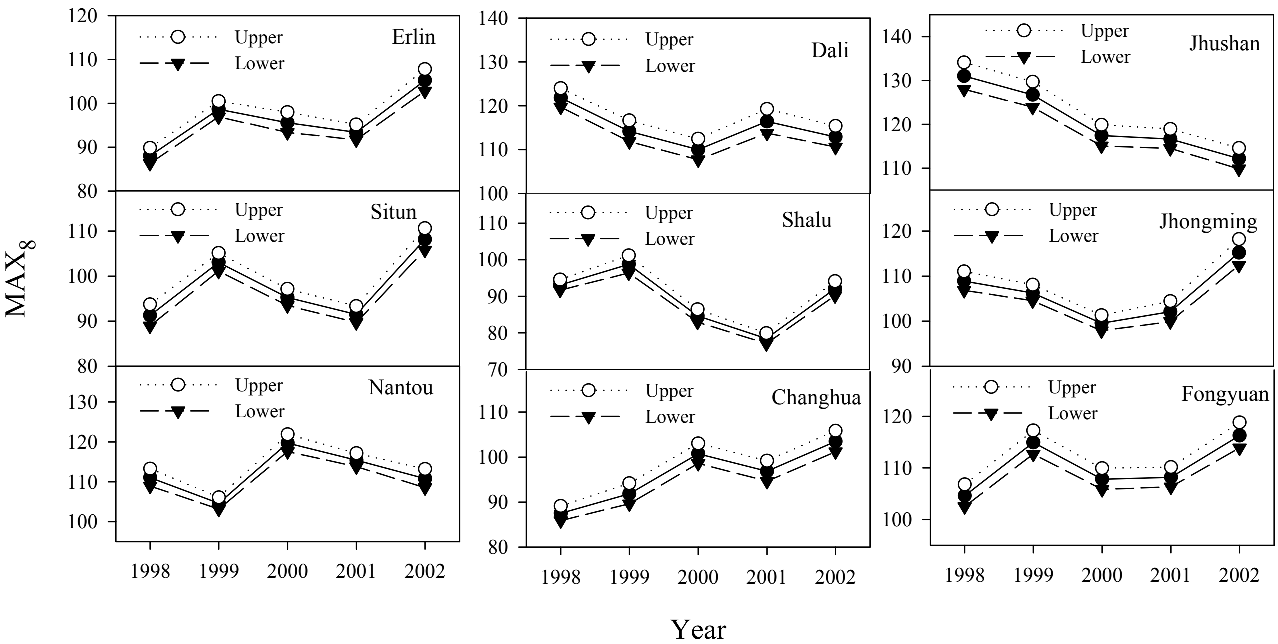

Fig. 4. The expected MAX_8 values and their upper and lower confidence intervals for the 9 air quality monitoring stations.

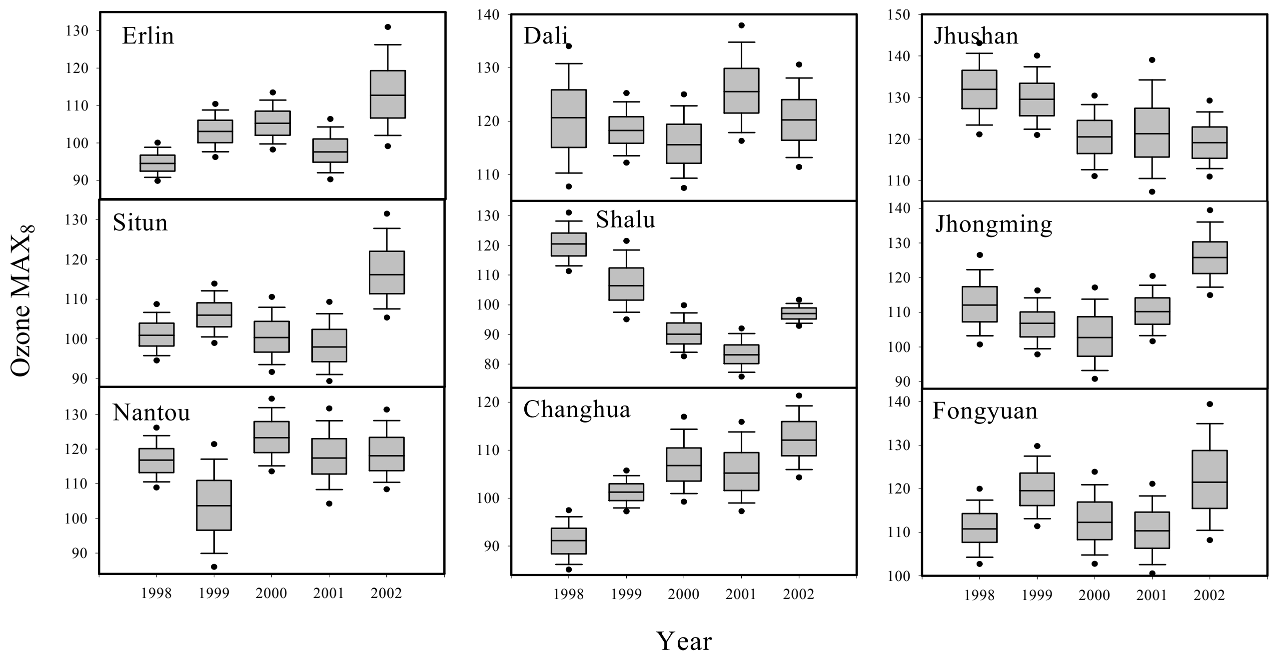

Fig. 5. Estimation of confidence intervals of MAX_8 by Monte Carlo simulation. The box-and-whiskers diagrams indicate the median, 5th, 10th, 25th, 75th, 90th and 95th percentiles.

and lower confidence intervals for the 9 air quality monitoring stations are shown in Fig. 4. Almost all the actual MAX_8 values (Table 1) are located inside the ranges of confidence intervals. The confidence interval ratios and the expected MAX_8 values are small

(1.7%–2.3%), which indicate the model can distinguish the ozone MAX_8 values well. Regarding statistics, the actual MAX_8 values located inside the confidence intervals imply the ozone MAX_8 values did not change significantly.

Prediction of MAX_8 and confidence interval by quadratic logistic regression analysis and Monte Carlo simulation

The expected MAX_8 values found by quadratic logistic regression analysis are very close to the actual values, showing that the quadratic logistic regression analysis is more suitable for the estimation of MAX_8 than linear regression analysis. The Monte Carlo simulation is performed additionally to estimate the MAX_8 values and their confidence intervals. The results are presented by a box-and-whiskers diagram (Fig. 5). The median, 5th, 10th, 25th, 75th, 90th and 95th percentiles are indicated in the diagram. The range between the 5th and 95th percentiles is recognized as the confidence interval. We found the confidence intervals are about 20% of the mean MAX_8 values for all the 9 air quality monitoring stations. The same as linear regression analysis, the high confidence intervals of Monte Carlo simulation also show that the ozone MAX_8 values did not change significantly for the analyzed years.

CONCLUSION

In this study, a method for calculating ozone MAX_8 and its confidence interval is described and demonstrated. We applied a linear regression of the ozone concentrations on their ranks, and a quadratic logistic regression of odd ratios on ozone concentration. The quadratic logistic regression was carried out in conjunction with Monte Carlo simulation to estimate the confidence interval.

The ranking number influences the goodness of fit of simulation model. The results indicate that there is no significant difference for SEE values when the ranking number is below 180. Thus, the highest 180 ozone concentrations were analyzed in this study. Most expected MAX_8 values arrived at by linear regression are less than the actual MAX_8 values, but the errors did not exceed 10%. In comparison with the results of linear regression and the quadratic logistic regression, the latter method provides better estimation of MAX_8 values. The confidence intervals calculated by Monte Carlo simulation are about 20% of the mean MAX_8 values for all the 9 air quality monitoring stations. The high confidence intervals of Monte Carlo simulation show that the ozone MAX_8 values did not change significantly for the analyzed years.

REFERENCES

Aneja, V.P., Oommen, R.G., Riordan, A.J., Arya, S.P., Wayland, R.J. and Murray, G.C. (1999). Ozone Patterns for Three Metropolitan Statistical Areas in North Carolina, USA. *Atmos. Environ.* 33: 5081-5093.

Aneja, V.P., Li, Z. and Das, M. (1994). Ozone

Case Studies at High Elevation in the Eastern United States. *Chemosphere*. 29: 1711-1733.

Bloomfield, P.J., Royle, J.A., Steinberg, L.J. and Yang, Q. (1996). Accounting for Meteorological Effects in Measuring Urban Ozone Levels and Trends. *Atmos. Environ.* 30: 3067-3077.

Colls, J. (2002). *Air Pollution.* p. 47. Spon Press. London. ISBN 0-415-25565-1.

Dodge, M.C. (2000). Chemical Oxidant Mechanisms for Air Quality Modeling: Critical Review. *Atmos. Environ.* 34: 2103-2130.

Dorling, S.R., Foxall, R.J., Mandic, D.P. and Cawley, G.C. (2003). Maximum Likelihood Cost Functions for Neural Network Models of Air Quality Data. *Atmos. Environ.* 37: 3435-3443.

Hanna, S.R., Chang, J.C. and Fernau, M.E. (1998). Monte Carlo Estimates of Uncertainties in Predictions by a Photochemical Grid Model (UAM-IV) due to Uncertainties in Input Variables. *Atmos. Environ.* 32: 3619-3628.

Hanna, S.R. and Davis, J.M. (2002). Evaluation of a Photochemical Grid Model Using Estimates of Concentration Probability Density Functions. *Atmos. Environ.* 36: 1793-1798.

Heinsohn, R.J. and Kabel, R.L. (1999). *Sources and Control of Air Pollutio*n. Prentice-Hall. New Jersey. ISBN 0-13-624834-9.

Int Panis, L., Nocker, L.D., Cornelis, E. and Torfs, R. (2004). An Uncertainty Analysis of Air Pollution Externalities from Road Transport in Belgium in 2010. *Sci. Tot. Environ.* 334-335: 287-298.

Koçak, K., Şaylan, L. and Şen, O. (2000). Nonlinear Time Series Prediction of O_3 Concentration in Istanbul. *Atmos. Environ.* 34: 1267-1271.

Moore, G.E. and Londergan, R.J. (2001). Sampled Monte Carlo Uncertainty Analysis for Photochemical Grid Models. *Atmos. Environ.* 35: 4863-4876.

NRC (National Research Council) (1991). Rethinking the Ozone Problem in Urban and Regional Air Pollution. National Academy Press, Washington.

Rabl, A. and Spadaro, J.V. (1999). Damages and Costs of Air Pollution: An Analysis of Uncertainties. *Environ. Int.* 25: 29-46.

Smith, R.L. and Huang, L. (1993). Modeling High Threshold Exceedances of Urban Ozone. National Institute for Statistical Science Technical Report #6.

TEPA, 2003. Survey of Ozone and Particulate Formation Mechanisms in Taiwan. Environmental Protection Administration, Taiwan.

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. *Atmos. Environ.* 35: 617-630.

Zádor, J., Wagner, V., Wirtz, K. and Pilling, M.J. (2005). Quantitative Assessment of Uncertainties for a Model of Tropospheric Eethane Oxidation Using the European Photoreactor (EUPHORE). *Atmos. Environ.* 39: 2805-2817.

Zar, J.H. (1999). *Biostatistical Analysis.* 4th ed. Prentice-Hall. New Jersey. ISBN 0-13-081542-X.

Received for review, July 8, 2008
Accepted, July 12, 2008